TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Computation,
Information and Technology

Evolution Equations in Fourier Phase Retrieval

Arseniy Tsipenyuk

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Daniel Matthes
Prüfer der Dissertation:	1.	Prof. Dr. Gero Friesecke
	2.	Prof. Dr. D. Russel Luke
	3.	Prof. Dr. Felix Krahmer

Die Dissertation wurde am 17.01.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 16.11.2022 angenommen.

Dedicated to my mother

ABSTRACT

Projection algorithms are successfully used for phase retrieval in the high-dimensional setting of X-ray crystallography, but the reason of this success is not well-understood. This complicates systematic development of better reconstruction algorithms. This dissertation studies variational structure of two prominent algorithms: it shows that the Error-Reduction algorithm is a discretized gradient (or, more generally, subdifferential) flow, and that the Douglas-Rachford algorithm is related to same flow through an appropriate selection of resolvents. Analysis of this gradient flow and of the corresponding energy functional yields new insights on said algorithms and can serve as a framework for variational analysis of infinite-dimensional non-convex feasibility problems.

ZUSAMMENFASSUNG

Die hochdimensionalen Phasenrobleme in der Röntgenstrahlenkristallographie werden in der Regel mit Projektionsalgorithmen gelöst. Viele Eigenschaften dieser Verfahren sind bisher nur heuristisch verstanden. Dies erschwert eine systematische Entwicklung besserer Rekonstruktionsalgorithmen. Die vorliegende Arbeit untersucht variationelle Eigenschaften von zwei prominenten Algorithmen. Es wird gezeigt, dass der Error-Reduction Algorithmus die Diskretisierung eines Gradientenflusses (oder, im allgemeinen Fall, eines Subdiffertialflusses) ist. Der Douglas-Rachford Algorithmus kann von demselben Gradientenfluss durch eine geschickte Wahl der Resolventen hergeleitet werden. Die variationelle Analyse des zu diesem Gradientenfluss zugehörigen Energiefunktionals liefert neue Einsichten zu den genannten Algorithmen und kann auf weitere unendlichdimensionale nicht-konvexe Schnittmengenprobleme verallgemeinert werden.

ACKNOWLEDGMENTS

First and foremost, I thank my supervisor Prof. Gero Friesecke for the possibility to work on this project. I thank him for his helpful feedback and patience, for opportunities and creative freedoms he granted me, and for the mathematical and non-mathematical insights he shared with me during my studies.

I especially thank my mentor Dominik Jüstel for his inspiring and joyful approach to mathematical problems, for many mathematical discussions and for his guidance in soft skill and organizational matters.

I would like to thank Prof. D. Russel Luke and Prof. Felix Krahmer for refereeing my thesis, and Prof. Daniel Matthes for chairing my defense committee.

I gratefully acknowledge the funding provided by the German Research Foundation (DFG) through the SFB/TRR 109 "Discretization in Geometry and Dynamics".

Thanks to my colleagues and friends Benedikt Grasswald, Sören Behr, Rufat Badal, Mi-Song Dupuy, Daniela Vögler and all other members of the analysis chair for helpful discussions on many different topics, and for making my experience at TUM so enjoyable. Thanks to Frank Filbir, Oleh Melnyk and other members of the Phase Retrieval Journal Club for sharing their insights and enticing ideas on phase retrieval.

Thanks to Frauke Bäcker and Diane Clayton-Winter for their help in administrative matters, and for many pleasant chats.

Finally, I thank family — Anna, Irina, Liya and Ilya, — for their love, patience and support.

CONTENTS

1 INTRODUCTION

1.1 DESCRIPTION OF THE STUDIED PROBLEM

Phase retrieval is a generic term used to describe a wide class of problems. These problems — often connected to imaging — are unified by the following setting: an object g must be reconstructed from a given measurement $|F[g]|$ using additional information $\mathcal{A}$. The transformation F is usually linear and bijective; thus, to recover g, it is sufficient to recover the phase of $F[g]$ using $\mathcal{A}$, giving phase retrieval its name.

The application that motivates the setting of this thesis is X-ray crystallography. It is the most widely used method to determine molecular structures to date: as of July 2020, ca. 89% of all structures found using query "protein" in the Protein Data Bank [Ber+00] were solved using X-ray crystallography.

In X-ray crystallography, one endeavors to reconstruct the electron density g of a complex molecule (like a protein or a virus) — meaning, one wishes to determine the positions of atoms relative to each other — from the absolute value of its Fourier transform $|\hat{g}|$. The measurement $|\hat{g}| \approx \sqrt{I}$ is related to the square-root of the diffraction intensity I. This diffraction pattern emerges from X-rays scattered by the crystallized sample of the molecule in question, see Figure 1.1. Additional information $\mathcal{A}$ comes from the fact that the electron density g is non-negative, or from other constraints on the support or sparsity of g.

This problem is an example of *Fourier phase retrieval*, since the transformation F is the Fourier transform of the object.

A solution of phase retrieval is an object g that — provided transformation F, measurement $\sqrt{I}$ and additional information $\mathcal{A}$ — satisfies $|F[g]| = \sqrt{I}$ and complies with $\mathcal{A}$ within the desired degree of accuracy. In this thesis, we distinguish between *phase problem* — the task of determining whether a solution exists, is unique, is stable under perturbations of the provided data — and *phase retrieval* — the task of finding any solution, assuming that at least one exists.

This thesis primarily discusses phase retrieval in the latter sense. Specifically, it discusses Fourier phase retrieval on (possibly infinite-dimensional) Hilbert spaces, assuming non-negativity of the object as the main additional information.

More information on phase retrieval settings, applications and results can be found, for example, in surveys [JEH15; She+15; GKR20], or in a less formal essay [Luk17].

The main goal of this thesis is to study Fourier phase retrieval by examining connections between:

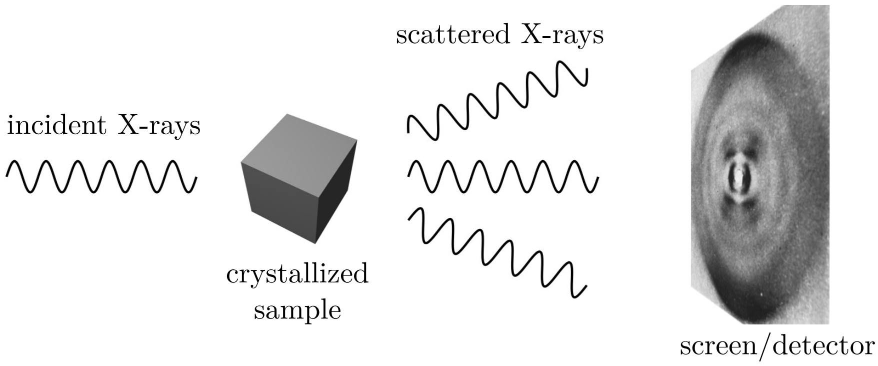

Figure 1.1: Sketch of a crystallographic measurement.
In X-ray crystallography, the unit cell electron density of a crystal must be reconstructed from the corresponding diffraction pattern. On the right: (linearly transformed) diffraction pattern of a crystallized DNA molecule from the celebrated paper [WC53]. The authors J.D. Watson and F.H.C. Crick received the Nobel Prize in 1962 for the discovery of the double helix structure of the DNA. The corresponding phase problem was solved by an explicit calculation of X-ray diffraction patterns corresponding to helical structures in [CCV52].

— certain formulations of phase retrieval (set intersection formulation, energy minimization formulation);

— certain evolution equations derived using variation of functionals (Error-Reduction Flow, Douglas-Rachford Flow);

— certain algorithms used for phase retrieval (Error-Reduction / Alternating Projections algorithm, Hybrid Input-Output / Douglas-Rachford algorithm),

see Figure 1.2. The precise nature of the studied connections is outlied in the following sections.

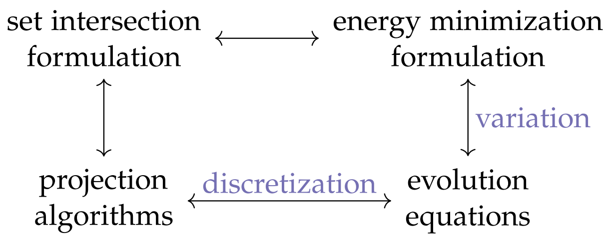

Figure 1.2: Various formulations of phase retrieval.

1.2 FORMULATIONS OF PHASE RETRIEVAL

1.2.1 *Set intersection formulation and algorithms*

A common way to formulate phase retrieval is to cast it as a set intersection problem (feasibility problem). To do so, let $\mathcal{M}$ denote the set of all objects that comply with the measurement, and let $\mathcal{A}$ denote the set of all objects that comply with the additional information. Then, phase retrieval states:

$$\text{assuming } \mathcal{M} \cap \mathcal{A} \neq \varnothing, \text{ find any } f \in \mathcal{M} \cap \mathcal{A}.$$

This formulation is used for phase retrieval, because for many additional constraints $\mathcal{A}$ (such non-negativity, support, sparsity) one can efficiently — in $O(N \log N)$ steps for images with N pixels — calculate single-valued projecton selections (selections of set-valued distance-minimizing projection operators) $P_{\mathcal{M}}$ and $P_{\mathcal{A}}$ onto the sets $\mathcal{M}$ and $\mathcal{A}$, respectively. These projecton selections are the basic constituents of many phase retrieval algorithms. For example, the Error-Reduction algorithm

$$g_{n+1} = P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n],$$

also known as the Gerchberg-Saxton algorithm [GS72], is the basic algorithm used for phase retrieval, and the following variant of the Hybrid-Input-Output algorithm

$$g_{n+1} = g_n - P_{\mathcal{M}} \circ \left[2P_{\mathcal{A}}[g_n] - g_n\right] - P_{\mathcal{A}}[g_n], \tag{1.1}$$

is a baseline state-of-the-art algorithm used for phase retrieval [ELB18]. These algorithms received wide attention after their systematic discussion in the seminal paper by Fienup [Fie82].

Set intersection formulation is well-understood in convex optimization, where it corresponds to the task of determining

$$\mathcal{X} \cap \mathcal{Y} \text{ for two convex sets } \mathcal{X} \text{ and } \mathcal{Y}.$$

In convex optimization, single-valued projecton selections $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ onto the sets $\mathcal{X}$ and $\mathcal{Y}$ are unique and can be used to find points in $\mathcal{X} \cap \mathcal{Y}$. Some of the algorithms used for this task correspond to algorithms independently developed for phase retrieval; notably, the Alternating Projections algorithm

$$g_{n+1} = P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g_n] \tag{1.2}$$

corresponds to the Error-Reduction algorithm, and the Douglas-Rachford algorithm

$$g_{n+1} = g_n - P_{\mathcal{Y}} \circ \left[2P_{\mathcal{X}}[g_n] - g_n\right] - P_{\mathcal{X}}[g_n], \tag{1.3}$$

corresponds to the Hybrid-Input-Output algorithm under certain assumptions. These connections were established in [BCL02].

Unfortunately, general results from the convex setting do not carry over to phase retrieval, as the set $\mathcal{M}$ is non-convex. Thus, in many aspects, behavior of the Error-Reduction and Hybrid-Input-Output algorithms remains open and highly relevant for applications.

This work focuses on these two particular algoritms; particularly, on the Error-Reduction algorithm. It does so for the following reasons: Error-Reduction is arguably the simplest algorithm for phase retrieval, and it forms the basis for other projection algorithms; Hybrid-Input-Output (or its Douglas-Rachford variant) is a state-of-the-art algorithm used for benchmarking in crystallographic phase retrieval.

1.2.2 *Energy minimization formulation*

In general, it is very difficult to quantify whether an approximation g is close to the intersection $\mathcal{M} \cap \mathcal{A}$. This difficulty is exacerbated by the fact that $\mathcal{M} \cap \mathcal{A}$ can contain multiple non-trivially distinct elements.

In practice, one must resort to estimating square distances

$$E_{\mathcal{M}}[g] := \frac{1}{2}\|g - P_{\mathcal{M}}[g]\|_2^2 \text{ and } E_{\mathcal{A}}[g] := \frac{1}{2}\|g - P_{\mathcal{A}}[g]\|_2,$$

assuming phase retrieval on some Hilbert space $\mathcal{H}$ with norm $\|\cdot\|_2$. If smallness of $E_{\mathcal{M}}[g]$ and $E_{\mathcal{A}}[g]$ implies that $\|g - f\|_2$ is small for at least one element $f \in \mathcal{M} \cap \mathcal{A}$, the intersection of $\mathcal{M}$ and $\mathcal{A}$ is called regular. In general, one can not expect the intersection of $\mathcal{M}$ and $\mathcal{A}$ to be regular, but this regularity assumption is often required to treat phase retrieval in its set intersection formulation.

An alternative approach is to formulate phase retrieval as an energy minimization problem: for given data $\sqrt{I}$ and constraint $\mathcal{A}$, find any

$$g \in \arg\min E_{\mathcal{M}}[g] + E_{\mathcal{A}}[g]. \tag{1.4}$$

This formulation can be more practical than the set intersection formulation, as, generally, existence of a solution can be guaranteed by the direct method in the calculus of variations.

The choice of functionals in Equation (1.4) is only one of many possibile variants. However, square distance functionals $E_{\mathcal{M}}$ and $E_{\mathcal{A}}$ posess certain specific properties. Let us outline two such properties and illustrate how they were used in [Fie82].

First, for any proximal set $\mathcal{X} \subset \mathcal{H}$ there exists — by definition of proximality — a well-defined single-valued projecton selection $P_{\mathcal{X}}$ such that

$$\|g - P_{\mathcal{X}}[g]\|_2 \leqslant \|g - P_{\mathcal{X}}[f]\|_2$$

for any $g, f \in \mathcal{H}$. This property is highly useful for calculations with functionals containing terms like $\|g - P_{\mathcal{X}}[g]\|_2$.

In [Fie82], this property was used to show that that for all iterates g_n generated by the Error-Reduction algorithm holds

$$E_{\mathcal{M}}[g_{n+1}] \leqslant E_{\mathcal{M}}[g_n], \tag{1.5}$$

giving the algorithm its name, as it does not increase the error $E_{\mathcal{M}}$. (And $E_{\mathcal{A}}[g_n] = 0$ for all n by definition of Error-Reduction.)

Second, for $P_{\mathcal{X}}$ as described above and for $E_{\mathcal{X}}[g] := \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2$, the formal derivative of $E_{\mathcal{X}}$ at g is given by $g - P_{\mathcal{X}}[g]$.

In [Fie82], this property was used to interpret Error-Reduction as a projected gradient descent with respect to the energy $E_{\mathcal{M}}$ in the following sense. If for an iterate g_n one takes a gradient descent step

$$g_{n+1/2} := g_n - \nabla E_{\mathcal{M}}[g_n] = P_{\mathcal{M}}[g_n]$$

and then explicitely takes the projection $P_{\mathcal{A}}$ at $g_{n+1/2}$, one recovers the Error-Reduction algorithm.

In a more recent example, a similar property was used in [ELB18] to interpret variants of Error-Reduction as minimization-majorization algorithms for functionals of the form $E_{\mathcal{M}}[g] + F[g]$, for certain convex functionals F.

To our knowledge, the specific form $E_{\mathcal{M}}[g] + E_{\mathcal{A}}[g]$ studied in this thesis is not prevalent in phase retrieval literature, yet it does exhibit certain remarkable properties described further below.

1.2.3 *Evolution equations*

In convex optimization, minimization problems of the form

$$f \in \arg\min F_a[f] + F_b[f], \text{ where } F_a \text{ and } F_b \text{ are convex} \tag{1.6}$$

are often reduced to the multivalued equation

$$0 \in A[f] + B[f], \tag{1.7}$$

where A is the subdifferential of F_a, and B is the subdifferential of F_b. Then, solutions of Equation (1.7) are minimizers of Equation (1.6).

Equation (1.7) can be analyzed independently of the minimization problem (1.6), and for a larger class of operators A and B.

For example, the seminal paper [LM79] investigated Equation (1.7) for the case when operators A and B are maximal monotone. They reformulated the Douglas-Rachford algorithm from [DR56] in terms of resolvents:

$$g_{n+1} = g_n + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g_n] - J_{\lambda B}[g_n], \tag{1.8}$$

where $\lambda > 0$, $J_{\lambda A} = (I + \lambda A)^{-1}$ is the resolvent of λA, and $J_{\lambda B}$ is the resolvent of λB. They have shown that Douglas-Rachford converges weakly to some fixed point g_*, and that $f := J_{\lambda B}[g_*]$ is a solution of Equation (1.7).

For convex feasibility problems one can show the equivalence of Douglas-Rachford forms Equation (1.3) and Equation (1.8). Indeed, let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be weakly closed and convex. Then, projections $P_{\mathcal{X}}, P_{\mathcal{Y}}$ are well-defined, and the minimization problem

$$\arg\min \underbrace{\frac{1}{2}\|f - P_{\mathcal{X}}[f]\|_2^2}_{=:E_{\mathcal{X}}[g]} + \underbrace{\frac{1}{2}\|f - P_{\mathcal{X}}[f]\|_2^2}_{=:E_{\mathcal{Y}}[g]} \tag{1.9}$$

leads to the equation

$$0 \in -(f - P_{\mathcal{X}}[f]) - (f - P_{\mathcal{Y}}[f]), \tag{1.10}$$

since $f - P_{\mathcal{X}}[f]$ belongs to the convex subdifferential $\partial_{\mathrm{conv}} E_{\mathcal{X}}[f]$, and likewise for $\mathcal{Y}$. One can show that for $A[f] := f - P_{\mathcal{X}}[f]$ holds $J_{\lambda A}[f] = P_{\mathcal{X}}[f]$ for all $\lambda > 0$; likewise, for $B[f] := f - P_{\mathcal{Y}}[f]$ holds $J_{\lambda B}[f] = P_{\mathcal{Y}}[f]$.

This argument relates the minimization problem (1.9) to the evolution equation (1.10) — which, in turn, is related to a specific variant of the Douglas-Rachford algorithm for convex feasibility problems.

Connections like this one can be formulated for the Alternating Projections and Douglas-Rachford algorithms in the non-convex setting of phase retrieval. These connections, analysis of resulting evolution equations, algorithms, and insights for phase retrieval applications shape main ideas developed in this thesis.

1.3 RESULTS OF THE THESIS

We would like to highlight the following results of this thesis.

ER is a discretization of a formal gradient flow equation we call ERF. We establish that the Error-Reduction algorithm corresponds to two consecutive steps of gradient descent with energy $E_{\mathcal{M}} + E_{\mathcal{A}}$. What distinguishes this result from related known results is the fact that no explicit imposition of the constraint $\mathcal{A}$ (as in Fienup's projected gradient descent), or no splitting techinque is required. The correspondence is based on the following observation: for any $g \in \mathcal{H}$

where $E_{\mathcal{M}}[g] + E_{\mathcal{A}}[g]$ is Fréchet-differentiable, the discrete Euler gradient descent update with step size $\varepsilon = 1$ is

$$\begin{aligned} g_{n+1} &= g_n - \varepsilon\nabla(E_{\mathcal{M}}[g_n] + E_{\mathcal{A}}[g_n]) \\ &= g_n - (g_n - P_{\mathcal{M}}[g_n] + g_n - P_{\mathcal{A}}[g_n]) = P_{\mathcal{M}}[g_n] + P_{\mathcal{A}}[g_n] - g_n. \end{aligned} \tag{1.11}$$

In particular, if $g_n \in \mathcal{A}$, then $g_{n+1} = P_{\mathcal{M}}[g_n] \in \mathcal{M}$, and

$$g_{n+2} = P_{\mathcal{A}}[g_{n+1}] = P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n].$$

This observation connects Error-Reduction to the study of the equation we call Error-Reduction Flow:

$$\partial_t g(t) = -2g(t) + P_{\mathcal{M}}[g(t)] + P_{\mathcal{A}}[g(t)]. \tag{1.12}$$

To make this equation well-defined, one must pick a single-valued projecton selection $P_{\mathcal{M}}$. In general, such a choice is not unique.

On bounded domains, the modulus set is weakly closed, and ERF is a rigorous subdifferential flow. In general, the functional $g \mapsto (E_{\mathcal{M}} + E_{\mathcal{A}})[g]$ is not Fréchet-differentiable, and Equation (1.12) is only formally a gradient flow.

To establish the corresponding rigorous result, we use non-convex subdifferential analysis ideas based on and inspired by the paper [BL03], which calculates the Kruger-Mordukhovich [KM80; MS96] subdifferential of $E_{\mathcal{M}}$ on unbounded domains, where the set $\mathcal{M}$ is not weakly closed.

Specifically, we use compactness results of [Peg85] to establish that the set $\mathcal{M}$ is weakly closed on bounded domains (joint work with Gero Friesecke). Further, we show that for weakly closed sets $\mathcal{X}$, considered on separable Hilbert spaces, the generalized subdiffential of the functional $g \mapsto E_{\mathcal{X}}[g]$ equals $g - \Pi_{\mathcal{X}}[g]$, where $\Pi_{\mathcal{X}}$ is the multivalued projection operator. This implies that on bounded domains and assuming that $E_{\mathcal{A}}$ is weakly closed, the multivalued analogon of Equation (1.12)

$$\partial_t g(t) \in -2g(t) + \Pi_{\mathcal{M}}[g(t)] + \Pi_{\mathcal{A}}[g(t)] \subseteq \partial_{\mathrm{KM}}(E_{\mathcal{M}} + E_{\mathcal{A}})[g]$$

is a selection of the subdifferential flow of the energy $E_{\mathcal{M}} + E_{\mathcal{A}}$, where ∂_{KM} on the right-hand side denotes the aforementioned non-convex Kruger-Mordukhovich subdifferential.

DR can be derived from ERF using resolvent selection. It is established that for operators $A := g - P_{\mathcal{M}}[g(t)]$, $B := g - P_{\mathcal{A}}[g(t)]$ — provided minor additional assumptions — holds

$$\begin{aligned} &g + P_{\mathcal{M}}[2P_{\mathcal{A}}[g] - g] - P_{\mathcal{A}}[g] \\ &\qquad\qquad \in \operatorname*{Li}_{\lambda\to\infty}\Big(g + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g] - J_{\lambda B}[g]\Big), \end{aligned}$$

where Li is the Kuratowski limit inferior. This result shows that one can make (single-valued) selections of (multi-valued) resolvents $J_{\lambda A}$, $J_{\lambda B}$, such that — after an appropriate limiting procedure — Douglas-Rachford in its resolvent form (1.8) is reduced to the HIO/DR form (1.1). This shows how the HIO/DR form (1.1) can be derived directly from ERF (1.12).

Discretized version of ERF dissipates energy and has subsequences strongly convergent to fixed points. It is shown that for the discretized Error-Reduction Flow (1.11) with $\varepsilon \in (0,1]$ a generalized version of Fienup's Error-Reduction property (1.5) is true, namely that

$$\frac{E[g_{n+1}] - E[g_n]}{\varepsilon} = -(1-\varepsilon)\left\|-2g_n + P_{\mathcal{M}}[g_n] + P_{\mathcal{A}}[g_n]\right\|_2^2, \qquad (1.13)$$

where $E = E_{\mathcal{M}} + E_{\mathcal{A}}$. In fact, this result is true not only for $\mathcal{M}$ and $\mathcal{A}$, but for any weakly closed sets $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$, and does not require differentiability of $E_{\mathcal{X}}$ or $E_{\mathcal{Y}}$ at g_n.

It is shown that on bounded domains, for a sequence $(g_n)_n$ generated by the discretized Error-Reduction Flow (1.11), fixed points exist, and there exist subsequences that converge to these fixed points.

ERF has global weak solutions on bounded domains. One of the central contributions of this dissertation, derived in collaboration with Gero Friesecke, is the proof that ERF (1.12) has global weak solutions. The result holds on bounded domains, under the assumption that $\mathcal{A}$ is the set of non-negative functions. The main idea of the proof, inspired by the approach described in [FD97], is as follows.

First, a solution candidate is constructed by means of the Aubin-Lions lemma. Namely, approximate solutions $g^{(\varepsilon)}$ are constructed by taking updates (1.11) and linearly interpolating between them in time. A solution candidate is obtained from $(g^{\varepsilon})_{\varepsilon}$ by extracting an appropriate subsequence and passing to the limit $\varepsilon \to 0$.

Second, it is shown that the solution candidate formally solves equation (1.12); the key to this is the fact that $g^{(\varepsilon)}$ is constructed using explicit discretization update (1.11).

Third, it is shown that the solution candidate rigorously solves equation (1.12). This last step relies on a generalized version of Rademacher's theorem to show that solution candidate is a. e. differentiable in time, and prescribes a specific selection of $P_{\mathcal{M}}$ that needs to be chosen in (1.12).

There exists a correspondence between fixed points of ER and ERF. This correspondence is established in the thesis and motivates a closer study of ERF dynamics. Certain conditions for ERF fixed point instability are developed. Obtained insights are illustrated using numerical simulations (see below). Numerical simulations illustrate that ERF exhibits certain patterns of chaotic behavior (change of direction).

This provides a heuristic argument that may explain why acceleration of ERF is not fruitful for certain cases.

Numerical simulations illustrate that ER dynamics can be dominated by abundance of unstable fixed points. This provides a heuristic argument that indicates the following.

i) Theoretical ER convergence radius — for known ER local convergence results — shrinks extremely rapidly with finer object discretization.

ii) Numerical ER convergence radius — for certain problems — can be much larger than the theoretical one, due to prevalence of unstable fixed points in the solution's vicitity.

An important numerical technique that highlights these points is the restriction of phase retrieval to even functions. Following reasons grant significance to this case.

i) ER, DR, and ERF all preserve evenness of the argument g.

ii) Due to the fact that Fourier transforms of even real-valued functions are real-valued, projection $P_{\mathcal{M}}$ operates in a special regime.

iii) In general, solutions of phase retrieval are trivially not locally unique: for any solution g, its translation is also a solution. However, translation of a generic even function is no longer even. Thus, by considering phase retrieval for even functions, one can ignore the aforementioned translation ambiguity.

There exists a variational connection between the Alternating Projections, Dykstra and DR algorithms. It is shown that — just as ERF is an equation that can be used to analyze the dynamics of ER — there exists a system of equations that potentially may be used to analyze the dynamics of DR.

Namely, it is shown that for the functional

$$F[s,d] = E_{\mathcal{M}}[s+d] + E_{\mathcal{A}}[s-d] - \frac{1}{2}\|d\|_2^2,$$

(which is closely related to the functional used in [LP16] to prove local convergence of DR), one can use the system of equations

$$\partial_t \begin{pmatrix} s \\ d \end{pmatrix} = M \cdot \begin{pmatrix} \frac{\delta}{\delta s} \\ \frac{\delta}{\delta d} \end{pmatrix} F[s,d]$$

to recover — using explicit Euler discretization —

$$\text{Error-Reduction, if} \qquad M = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix};$$

$$\text{Dykstra's algorithm, if} \qquad M = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix};$$

$$\text{a variant of DR, if} \qquad M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

These formulations provide new heuristic interpretations of said algorithms. Specifically for DR, the resulting equations have the form

$$\begin{aligned} \partial_t s &= -s + \frac{P_{\mathcal{A}}[s+d] + P_{\mathcal{M}}[s-d]}{2}; \\ \partial_t d &= \quad - \frac{P_{\mathcal{A}}[s+d] - P_{\mathcal{M}}[s-d]}{2}; \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} \partial_t p &= -\frac{p+q}{2} + P_{\mathcal{A}}[p]; \\ \partial_t q &= -\frac{p+q}{2} + P_{\mathcal{M}}[q], \end{aligned} \tag{1.14}$$

where $s = \frac{p+q}{2}$, $d = \frac{p-q}{2}$. If (s, d) is a fixed point, one can show that $s \in \mathcal{A} \cap \mathcal{M}$. The energy $E_{\mathcal{M}}[s] + E_{\mathcal{A}}[s]$ — the same as appears in ERF — can be used as termination criterion for the resulting discretized DR variant. It is shown that — by the same argument as for ERF — **the system of equations** (1.14) **admits global weak solutions**.

1.4 BIBLIOGRAPHIC CONTEXT

Some of the commonly listed phase retrieval applications are astronomy [DF87; Luk17], radar ambiguity [Jam99], speech recognition [RJ93; MSS14], quantum mechanics [GKR20; Pau47], and diffraction imaging [She+15; GKR20] — the latter being the most prominent example.

In diffraction imaging (optical imaging), an object is placed in front of a source of electromagnetic radiation. The incident electromagnetic wave is diffracted by the object. At a sufficiently large distance, the outgoing electromagnetic field can be modeled by a Fourier transform of some kind (e.g. by the Fresnel transform, or some modification of the Fraunhofer transform); see [Goo04] for an introduction to diffraction theory.

Depending on the application, phase retrieval allows a degree of flexibility in the way measurements are acquired. Some examples (in no particular order) include (from [Luk17]):

— ptychography, where multiple diffraction images are taken by illuminating small overlapping regions of the object;

— single-shot X-ray imaging, where the object is illuminated by a short-time electromagnetic pulse;

— X-ray crystallography, where the object must be crystallized before measurement to produce a clear diffraction pattern.

Some other setups investigated in literature include measurements where random masks are placed in front of the object [CSV11], or measurements that probe non-crystalline symmetries of the object using specially designed electromagnetic waves [FJJ16].

A variety of methods has been employed to adress these different settings; see [GKR20; Luk17; She+15; JEH15] for some recent overviews on phase retrieval.

Two phase retrieval methods mentioned above — Error-Reduction and Douglas-Rachford/Hybrid-Input-Output — stand out among many other methods, having been around for a long time and remaining highly relevant to date. See, e. g., [Fie13] for a brief overview of various applications where these algorithms are used.

Recent results have expanded understanding of these algorithms (cf. [Luk17]). For example, [HL13; NR16; Pau+18] demonstrated local convergence of Error-Reduction, and [Pha15; LP16] demonstrated local convergence of Douglas-Rachford variants in finite-dimensional spaces. Further references that discuss convergence of Hybrid-Input-Output can be found in [Fie13]. For references on Douglas-Rachford in a more generic setting, the interested reader is directed to the recent survey [LS19].

This thesis attempts to contribute to the analysis of Error-Reduction and — to lesser degree — Douglas-Rachford algorithms in an infinite-dimensional setting by linking them to appropriate evolution equations. The aspiration is that — by disconnecting the algorithms from discretization in time and space — one can reveal new features about the dynamics of said algorithms.

1.5 READER'S GUIDE

Chapter 2 formalizes the setting of phase retrieval used throughout the work; it contains basic definitions of modulus and non-negativity sets as well as some other constraints like support and sparsity.

Chapter 3 discusses projection operators and some of their relevant properties, like the continuity properties Propositions 3.11 and 3.12 (known in literature, but adapted to the purposes of the thesis). Further, Section 3.3 shows that on a bounded domain, the modulus set is weakly closed. This result is of particular importance, as it is necessary to rigorously derive the evolution equations analyzed in Chapters 7 and 9. Finally, Section 3.5 formalizes the notion of a local projection that is heuristically well-known in literature. Local projections are a useful tool to rigorously discuss various formulations of projection algorithms in phase retrieval (cf. local vs. global algorithm formulations in Chapter 5).

Chapter 4 studies square distance functionals (energy functionals) of the form $g \mapsto E_{\mathcal{X}}[g] := \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2$, which are well-defined for any proximal set $\mathcal{X}$. Using the direct method in the calculus of variations, phase retrieval is reformulated as an energy-minimization algorithm. A sufficient condition for the Fréchet-differentiability of the modulus energy $E_{\mathcal{M}}$ is established. It is shown that for weakly closed sets $\mathcal{X}$, the Mordukhovich-Kruger subdifferential of $E_{\mathcal{X}}$ is given by $g - \Pi_{\mathcal{X}}$, where $\Pi_{\mathcal{X}}$ is the (multi-valued) projection operator. It is also shown that the Clarke subdifferential of $E_{\mathcal{X}}$ is given by the convex closure of $g - \Pi_{\mathcal{X}}$.

Chapter 5 describes some common projection-based algorithms used for phase retrieval, and common reformulations of said algorithms. Important novel contributions in this chapter are Remark 5.17 that establishes gradient descent of Error-Reduction and motivates Error-Reduction Flow (ERF), and Section 5.3.5 that connects Error-Reduction Flow to the resolvent form of the Douglas-Rachford algorithm.

Chapter 6 describes energy dissipation properties and existence of fixed points for the explicit discretization of ERF. Whenever possible, results are stated in a more general setting (i. e. for generic weakly closed sets).

Chapter 7 demonstrates existence of global weak solutions of ERF. This part of the dissertation is joint work with Gero Friesecke.

Chapter 8 establishes correspondence between fixed points of ER and ERF. Further, the chapter investigates certain stability conditions for these fixed points.

Chapter 9 presents how ER, Dykstra and DR can be derived from $F[s,d]$ described above. The resulting interpretation of dynamics is illustrated using a simple example (searching for intersection of balls in a 2D plane). It is shown that the system of equations arising from the variation of $F[s,d]$ admits global weak solutions for the DR phase retrieval case.

Chapter 10 illustrates the content of Chapters 8 and 9 by providing some numerical examples using a toy model.

To improve the readability of the text, an index of the used notations is provided at the very end. A brief overview of the key algorithms referenced throughout the thesis can be found in Remarks 5.1 and 5.2.

Part I

BACKGROUND AND PROJECTION ALGORITHMS

This part establishes the necessary background and demonstrates how the Error-Reduction Flow — an integro-differential equation that is formally a gradient flow, and rigorously a generalized (Kruger-Mordukhovich) subdifferential flow — is connected to the Error-Reduction and Douglas-Rachford algorithms.

Chapters 2 and 3 describe standard notions and some well-known results used in phase retrieval and best approximation theory. They contain very minor original contibutions of didactical and technical nature, with the following exception: Section 3.3 demonstrates thot the modulus set is weakly closed on a bounded domain, which appears to be a novel result.

Chapter 4 studies differentiability conditions of energy functionals. Of note are the results on generalized differentiation, established in Section 4.3.

Chapter 5 is an attempt to systematically describe some projection algorithms used in Fourier phase retrieval, and to highlight connections between these algorithms. It contains two noteworthy original contributions. Remark 5.17 shows that the Error-Reduction algorithm is a discretization of a gradient flow. Section 5.3.5 shows how the Douglas-Rachford algorithm is connected to this flow through an appropriate selection of multivalued resolvents.

2 SET INTERSECTION FORMULATION

First section of this chapter formalizes mathematical setting and some basic notions of phase retrieval used in this work. Second section defines phase retrieval as a set intersection problem, and discusses relevant examples of constraint sets.

The described setting is essentially the same as the setting in [BCL02] (phase retrieval on generic Hilbert spaces), but with a more explicit specification of the physical space. This restriction facilitates the discussion of smoothness or regularity of the objects in later chapters. The setting is

— more rigorous than the finite-dimensional setting common for applied phase retrieval (e. g. [She+15]);

— less rigorous than the frame theory setting common for generalized phase retrieval (e. g. [GKR20]).

As such, most of what is described in this chapter is known to the readers familiar with phase retrieval. The main points of this chapter are the definition of constraint difficulties in Definition 2.1 and their discussion in Remark 2.5.

2.1 MATHEMATICAL FRAMEWORK

2.1.1 *Definition of the framework*

Throughout this work and unless explicitely stated otherwise, the object to be reconstructed is modeled as an element of the real-valued Hilbert space $\mathcal{H}(\Omega) = L^2(\Omega;\mathbb{R})$, where $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$, or where Ω is a bounded Lebesgue-measurable subset of $\mathbb{R}^d$, for $d \in \mathbb{N}$. Here, $\mathbb{T}^d$ denotes the d-dimensional torus, and $\mathbb{T}^d_N$ denotes the discretized d-dimensional torus with $N = (N_1, \ldots, N_d) \in \mathbb{N}^d$ discretization points along each dimension. In particular, $\mathcal{H}(\mathbb{T}^d_N)$ is a finite space with Euclidean metric, and it is isomorph to $\mathbb{R}^{N_1 \cdot \ldots \cdot N_d}$. Coordinates on $\mathbb{T}^d$ and $\mathbb{T}^d_N$ are added and subtracted using the usual modular arithmetic. For example, for $f \in \mathcal{H}(\mathbb{T}^d_N)$ one has $f(-k) = f(q)$, where $q \in \mathbb{T}^d_N$ with $q_i = -k_i \mod N_i$ for all $i \in \{1, \ldots, d\}$. Integration over $\mathbb{R}^d$ and $\mathbb{T}^d$ is performed using the Lebesgue measure; integration over $\mathbb{T}^d_N$ is performed using the counting measure. The set Ω is called *physical space*; the space $\mathcal{H}(\Omega)$ is called *object space*.

We use the abbreviation $\mathcal{H} = \mathcal{H}(\Omega)$ for statements that are essentially the same for all mentioned physical spaces, or where Ω is clear from the context. The inner product on $\mathcal{H}$ is denoted by $\langle \cdot, \cdot \rangle$; the corresponding induced metric is denoted by $\|\cdot\|_2$. The uniform norm on $\mathcal{H}$ is denoted by $\|\cdot\|_\infty$. The derivatives of operators and functionals on $\mathcal{H}$ are always taken in Fréchet sense unless explicitely stated otherwise.

The Fourier transformation maps $\mathcal{H}(\Omega)$ to its Fourier dual $\widehat{\mathcal{H}}(\Omega_F)$, which is called *Fourier space*. For $\mathcal{H}(\mathbb{R}^d)$, the Fourier space is given by

$$\left\{f \in L^2(\Omega_F;\mathbb{C}) \mid \Omega_F = \mathbb{R}^d,\ f(-k) = f^*(k) \text{ for a. a. } k \in \mathbb{R}^d\right\}, \tag{2.1}$$

where z^* denotes the complex conjugate of $z \in \mathbb{C}$. For $\Omega = \mathbb{T}^d$ the definition remains the same, except that $\Omega_F = \mathbb{Z}^d$. Likewise, for $\Omega = \mathbb{T}_N^d$ one must take $\Omega_F = \mathbb{T}_N^d$. If Ω is a measurable bounded subset of $\mathbb{R}^d$, then $\Omega_F = \mathbb{R}^d$, but the Fourier dual of $\mathcal{H}(\Omega)$ is a subset of $L^2(\mathbb{R}^d;\mathbb{C})$ that we can not write down explicitely.

In a minor abuse of notation, inner products and norms on Fourier spaces are denoted exactly as inner products and norms on object spaces. The Fourier transform is such that Plancherel's theorem states $\|f\|_2^2 = C_{\mathcal{F}}\|\hat{f}\|_2^2$, where the constant $C_{\mathcal{F}} = 1/(2\pi)^d$ if $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d\}$ or a measurable bounded subset of $\mathbb{R}^d$, and $C_{\mathcal{F}} = 1/|N|$ with $|N| = N_1 \cdot \ldots \cdot N_d$ if $\Omega = \mathbb{T}_N^d$. See Appendix A for more details on Fourier transform.

2.1.2 *Discussion of the framework*

As phase retrieval is an actively developed problem with many open issues, there does not exist a canonical framework that connects basic notions of phase retrieval to mathematical objects. The setting we use is very close to the one employed in [BCL02], which considered phase retrieval algorithms on generic Hilbert spaces.

The choice of our setting was influenced by following goals:

i) to appropriately model the studied phenomenon;

ii) to allow the use of specific mathematical tools;

iii) to be transferrable to applications;

iv) to remain as simple as possible for didactical purposes while keeping key features of the problem intact.

The influence of these goals is briefly discussed below.

On the choice of object and Fourier space

In X-ray crystallography, the object of interest is an electron density of a crystallized molecule. Common candidate spaces describing molecular structures would include L^p-spaces, spaces of (possibly signed) measures, space of tempered distributions. Of these spaces, L^1 is the most natural candidate to model electron density, as one would expect integrability of a density function. Additionally, for non-negative integrable functions f one has $\|f\|_{L^1} = \int f = |\hat{f}(0)|$. This means that integrability can showcase important properties of phase problem in object and Fourier space. This latter point is accentuated by the fact that $|\hat{f}(0)|$ is not measured in X-ray crystallography, but its correct determination is highly relevant for successful reconstruction.

Another example of a space used to model electron densities is the space containing functions $\rho \in L^1(\mathbb{R}^3)$ that satisfy $\sqrt{\rho} \in H^1(\mathbb{R}^3)$; this space can be embedded into $L^1(\mathbb{R}^3) \cap L^3(\mathbb{R}^3)$ by the Sobolev's embedding theorem and guarantees that the kinetic energy of an electron density is finite.

From the mathematical point of view, it is much more convenient to consider phase problem on L^2: it simplifies differentiation of energy functionals and transformations between object and Fourier spaces (admitting, for example, Plancherel's theorem). Therefore, the space L^2 is the space of choice as it is a Hilbert space and admits Plancherel's theorem. Some exceptions to this choice appear: i) in Section 6.3.3, where we use L^p spaces to demonstrate existence of fixed points of the Error-Reduction algorithm; ii) in Chapter 7, where we use the Sobolev space $H^1(\Omega)$ with $\Omega = \mathbb{T}^d$ or Ω a bounded measurable subset of $\mathbb{R}^d$ (as it embeds compactly into $L^2(\Omega)$) to demonstrate global weak solutions of a PDE corresponding to the Error-Reduction algorithm.

On the choice of physical space and object codomain

The physical space dimension d belongs to $\{1, 2, 3\}$ for most applications. There is a notable phenomenological difference between one-dimensional and higher-dimensional physical space settings. This difference is well-understood for certain finite-dimensional phase problem variants and is connected to uniqueness of phase problem solutions (e.g., see review [GKR20]). The analysis presented in this thesis is valid for any $d \in \mathbb{N}$ (unless explicitely specified otherwise), and obtained theoretical results do not suggest any difference between $d = 1$ and $d > 1$. Nevertheless, the focus of computational examples is placed on physical space dimension $d = 2$, since $d = 3$ is the case for X-ray crystallography, and for $d = 2$ simulations seem to remain phenomenologically close to $d = 3$ and are easier to visualize.

As for the choice of physical space itself, the taken approach remains flexible: $\Omega \subseteq \mathbb{R}^d$ is appropriate to model single molecules; $\Omega = \mathbb{T}^d$ is appropriate to model crystalline structures; $\Omega = \mathbb{T}^d_N$ is appropriate to investigate numerical properties. From the mathematical standpoint, the domain Ω must be bounded for results of Chapter 7. The choice $\Omega = \mathbb{T}^d_N$ is essential in Section 8.2 and Section 8.3 to investigate stability properties of phase problem, as second derivatives of essential functionals do not exist otherwise.

The codomain of object space functions is commonly set to be real. This choice is consistent with electron density modeling and is used in applications, where measured data is often symmetrized so that it belongs to Fourier space (e. g., see CIF data files in [Ber+00]). An argument can be made that complex codomain of the object space functions is necessary for a more detailed analysis of noisy phase retrieval. We conjecture that a majority of this thesis' results can be extended to this case; however, such generalization lies beyond the scope of this work.

2.2 PHASE RETRIEVAL AS A SET INTERSECTION PROBLEM

At its most general, phase problem is a task of reconstructing an object that satisfies two conditions.

- The object must be compatible with a measurement. The measurement is given by the absolute value of a transformation of the object; the transformation is usually usually complex-valued, linear and bijective. This condition is called the *modulus constraint*; the set of all objects satisfying modulus constraint is called the *modulus constraint set*. Thus, the absolute value of this bijective transformation is known, and the phase is not known, leading to the name "phase retrieval": if the phase is recovered, one can uniquely reconstruct the object.
- The object must satisfy certain modeling requirements. For example, it may be non-negative, sparse, or have a prescribed support. These conditions are called the *additional constraints*.

There are two types of common modulus constraints: 1) if the object must be reconstructed from the absolute value of its Fourier transform, one commonly speaks of *Fourier phase problem*; 2) if the object must be reconstructed from the absolute value of some other measurement, one commonly speaks of *generalized phase problem*.

There are many ways to provide additional constraints. Roughly, additional constraints may be divided into three categories: 1) requirements on the framework of the problem; 2) requirements on the measurement; 3) requirements on the object itself. These requirements may be described as follows.

1) Requirements on the framework address questions such as: is the object real-valued, or is the object finite-dimensional.
2) Requirements on the measurement address questions such as: what transformation describes the measurement (see, e. g., the approach in [CSV11], where randomness in the measurement is utilized to analyze generalized phase problem), or what is the resolution of the measurement (e. g., it is long known that with sufficient oversampling, phase problem can admit unique solutions [Hay82]).

3) Requirements on the object itself are properties such as non-negativity, sparsity, or support. They are usually prescribed by the specific application.

This last class of additional constraints is of primary importance for X-ray crystallography and can be conveniently represented by sets (of all objects satisfying these constraints). Phase problem is then formulated as a set intersection problem (feasibility problem): find the intersection of the modulus constraint set with additional constraint sets. In generic mathematical contexts, a set intersection problem is usually called feasible, if there exists a unique solution, i. e. an element that belongs to the set intersection and is unique, possibly up to some trivial ambiguities. We use the phrase "feasible phase problem" without any precise mathematical meaning, but as applied description of phase problem for which there exists at least one, not necessarily unique, element that satisfies given constraints to a certain meaningful degree.

Set intersection formulation is most useful if one can efficiently calculate projections onto the constraint sets. This is the case for phase retrieval (cf. Remark 3.28). The definitions below present some common crystallographic constraints and formalize Fourier phase retrieval as a feasibility problem.

Definition 2.1 (constraint sets). *Let $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$, or let $\Omega \subset \mathbb{R}^d$ bounded and Lebesgue-measurable. Let the measured intensity $I \in L^1(\Omega_F)$ with $I \geqslant 0$, (so that $\sqrt{I}$ belongs to $\L^2(\Omega_f)$). The following sets will be of primary importance for this work:*

$$\begin{aligned} \mathcal{M}(\sqrt{I}) &= \{f \in \mathcal{H} \mid |\hat{f}| = \sqrt{I}\}; && \textit{(modulus)} \\ \mathcal{P} &= \{f \in \mathcal{H} \mid f \geqslant 0\}. && \textit{(positivity)} \end{aligned}$$

Further, let S be a measurable subset of Ω; let S_F be a measurable subset of Ω_F. The following sets are example sets that can be used for crystallographic phase retrieval:

$$\begin{aligned} \mathcal{M}^{(i)}(\sqrt{I}; S_F) &= \{f \in \mathcal{H} \mid \mathbb{1}_{S_F}|\hat{f}| = \mathbb{1}_{S_F}\sqrt{I}\}; && \textit{(incomplete modulus)} \\ \mathcal{S}(S) &= \{f \in \mathcal{H} \mid \operatorname{supp} f = S\}; && \textit{(support)} \\ \mathcal{T}_a(\alpha) &= \{f \in \mathcal{P} \mid f(x) \geqslant \alpha \text{ for a. a. } x \in \operatorname{supp} f\}; && \textit{(amplitude thresholding)} \\ \mathcal{T}_s(\nu) &= \{f \in \mathcal{P} \mid \lambda(\operatorname{supp} f) \leqslant \nu\}, && \textit{(support size)} \end{aligned}$$

where $\alpha \geqslant 0$ is the thresholding level of an object, $\nu > 0$ is the support size of an object, and λ is the Lebesgue measure for $\Omega \in \mathbb{R}^d, \mathbb{T}^d$, and the counting measure for $\Omega = \mathbb{T}^d_N$.

We abbreviate $\mathcal{M} = \mathcal{M}(\sqrt{I})$, $\mathcal{M}^{(i)} = \mathcal{M}^{(i)}(\sqrt{I}; S_F)$ and $\mathcal{S} = \mathcal{S}(S)$ where it can not cause confusion. Whenever the sets $\mathcal{M}$, $\mathcal{M}^{(i)}$ and $\mathcal{S}$ are used without arguments, we imply that there exist corresponding $\sqrt{I}$, S_F and S such as in this definition. We call $\mathcal{T}_a(\alpha)$ the amplitude thresholding constraint and

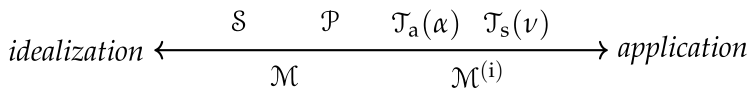

Figure 2.1: Sketch of relative constraint difficulty discussed in Remark 2.5.

$\mathcal{T}_s(\nu)$ *the support size constraint; note that in our notation these constraints automatically imply non-negativity.*

Remark 2.2 ($\mathcal{M}$ on a bounded domain). Whenever Ω is a measurable bounded subset of $\mathbb{R}^d$, the modulus set is

$$\mathcal{M}(\sqrt{I}) = \{f \in L^2(\Omega) \,|\, |\hat{f}| = \sqrt{I}\},$$

where $\hat{f}$ is the Fourier transform of the extension of f by zero to all of $\mathbb{R}^d$. In particular, $\mathcal{M}(\sqrt{I})$ may be empty.

Definition 2.3 (Phase problem and phase retrieval). *Let* $\sqrt{I} \in \widehat{\mathcal{H}}$ *be non-negative. Define the following set intersection problems.*

Phase problem:	*determine* $\mathcal{M} \cap \mathcal{P}$.
Phase retrieval:	*assuming* $\mathcal{M} \cap \mathcal{P} \neq \varnothing$, *find any element of* $\mathcal{M} \cap \mathcal{P}$.

Remark 2.4 (discussing phase retrieval definition). In applications, it is common to combine various additional constraints. The positivity set is used throughout this thesis as a generic representative of an additional constraint. This choice is discussed below in Remark 2.5.

Note the difference between *phase problem* and *phase retrieval*. The former is more general, implies questions of existence and uniqueness, and lies beyond the scope of this thesis. The latter is more applied, dispenses with questions of existence and uniqueness, and is of primary interest for this thesis.

The set intersection formulation is convenient but has the drawback of being sensitive to noise: it may happen that $\mathcal{M}(\sqrt{I}) \cap \mathcal{P} \neq \varnothing$ for some measurement $\sqrt{I}$, but $\mathcal{M}(\sqrt{I} + \delta\sqrt{I}) \cap \mathcal{P} = \varnothing$ for some noisy measurement $\sqrt{I} + \delta\sqrt{I}$, no matter how small $\delta\sqrt{I}$. In practice, set intersection formulation must be appropriately relaxed to recover phase from noisy measurements, or to analyze noise stability.

Remark 2.5 (discussing constraint sets). This remark discusses constraint sets, cf. Figure 2.1.

In object space, support constraint set $\mathcal{S}$ is arguably the easiest to analyze, as projection onto $\mathcal{S}$ is linear (see Example 3.14). However, it is difficult to get an estimate on $\mathcal{S}$ that is sufficiently strict for feasible phase retrieval in a crystallographic setting. (I. e., if the support is not estimated tight enough, considering $\mathcal{S}$ as a sole additional constraint can yield non-physical solutions. The problem of estimating the object support from the Fourier modulus data has been studied, for example, in [CFT90].)

Positivity set is integral to crystallographic modeling and is widely used, either alone or in conjunction with other constraints. It would be more precise to call the set $\mathcal{P}$ "non-negativity set", but we use the name "positivity set" as more didactically intuitive. From mathematical point of view, the set $\mathcal{P}$ is a favorable additional constraint, since it is convex and the corresponding error functional is differentiable (see Example 4.6, Lemma 4.14). This set also has a certain aesthetical appeal since positivity is nothing other than constant zero phase in object space. While simulations indicate that $\mathcal{P}$ can be sufficiently strict for feasible phase retrieval for certain objects (e. g., those that decay sufficiently fast in object space), it is, in general, not sufficiently strict for feasible phase retrieval in a crystallographic setting.

The amplitude thresholding set $\mathcal{T}_a(\alpha)$ is a subset of $\mathcal{P}$ that selects positive objects with moduli above some positive threshold α. This set is not convex and the corresponding error functional is not Fréchet-differentiable at many points of interest, see Remark 4.15. It also can violate modeling, since molecular electron densities a priori have values in the interval $(0, \alpha)$. This set has the locality property (cf. Section 3.5, Definition 3.29), which is convenient for reformulation of certain reconstruction algorithms (see Section 5.1). To our knowledge, this form of the sparsity constraint has not been investigated in literature. This constraint is (arguably) a more theoretically accessible version of the support size constraint $\mathcal{T}_s(\nu)$.

The support size constraint $\mathcal{T}_s(\nu)$ is even more strict than $\mathcal{T}_a(\alpha)$ in the sense that it does not have the locality property. It does not require the thresholding parameter and uses instead a support size parameter ν, which is easier to estimate in an experimental setup. This constraint is sometimes called the histogram constraint [Els03]. It is used in [ELB18] as an additional constraint to analyze a set of benchmark crystallographic phase problems.

Definition 2.1 shows only few of many possible additional constraints. Notably, the *atomicity* constraint is sometimes used in crystallographic phase retrieval [Els03]. This constraint describes sparse non-overlapping atoms with prescribed supports. Additionally, it can fix minimal allowed distance between two atoms (such restriction is used in [ELB18] to generate a set of benchmark problems). The atomicity constraint requires involved phenomenological assumptions and lies beyond the scope of this work.

Another common way to model sparsity is by minimizing the L^1-norm of an approximation; see [Pau+18].

In Fourier space, the modulus set $\mathcal{M}$ is the key set for Fourier phase retrieval. While additional constraints are shared by objects of interest, the set $\mathcal{M}$ makes solutions distinct. It is natural to conjecture that properties of this set have decisive impact on stability, existense, and uniqueness of phase problem. The fact that $\mathcal{M}$ is non-convex consitutes one of the main difficulties characterising phase problem. Striking experimental success of heuristic algorithms such as HIO (see Section 5.1) indicates that structure of $\mathcal{M}$ — which is isomorph to a torus in Fourier space — has remarkable intrinsic properties.

The more realistic variant of the modulus set $\mathcal{M}^{(i)}$ takes into account the fact that in an experimental setting one can not measure intensity values at certain coordinates. For example, the value $\sqrt{I}(0)$ is not measured, since the corresponding detector location is shielded by a backstop to protect the detector from incident unscattered X-ray beam. This constraint is used in [ELB18] to analyze a set of benchmark crystallographic phase problems.

3 PROJECTION OPERATORS

This chapter covers definitions of projections (set-valued distance-minimizing operators, also known as projectors) and their single-valued selections, as well as some standard results from best approximation theory, and shows that the modulus set is weakly closed on bounded domains.

Section 3.1 formally introduces projections (also known as best approximation operators, or projectiors), the corresponding single-valued selections, and recalls some of their known properties like the continuity results Propositions 3.11 and 3.12.

Section 3.2 recalls the well-known explicit forms of projectors onto additional constraint sets from the previous chapter.

Section 3.3 is the main novel contribution of this chapter: it demonstrates that the modulus set is weakly closed on bounded domains (joint work with Gero Friesecke). Inspired by variational analysis of [BL03], the result is based on compactness theorems of [Peg85] and is crucial for later results (such as subdifferential calculus in Section 4.3 or fixed point results in Chapter 6, which use continuity properties from Section 3.1).

Section 3.4 presents the well-known explicit form of the modulus projection operator.

Section 3.5 notes that under a certain locality condition that is met for many phase retrieval sets, very similar arguments are used to prove that an operator is a projection. The section describes a formalism that exploits this similarity. The main result of the section is Proposition 3.33: it shows how the form of a projection on $\mathcal{X} \subset L^2(\Omega)$ can be deduced from the form of an appropriate projection on the much smaller set $\mathcal{X}_{\text{loc}} \subset \Omega \times \mathbb{R}$, provided certain assumptions.

3.1 DEFINITION AND RELEVANT PROPERTIES

This section starts with the standard material from best approximation theory (see, e. g., [BC17], [Deu01]); also discussed in [BCL02] in context of phase retrieval).

The continuity properties Proposition 3.11 and Proposition 3.12 are known in literature (e.g. [RW09, Example 1.20]), but are adapted to purposes of this thesis and will later be used in Section 4.3 and Proposition 6.17, respectively.

Remark 3.1 (Multivalued projections). A single-valued selection of projecton onto $\mathcal{X}$ will be defined by the following distance-minimizing property:

$$P_{\mathcal{X}}[g] \in \Pi_{\mathcal{X}}[g] := \arg\min_{f \in \mathcal{X}} \|g - f\|_2.$$

The set $\Pi_{\mathcal{X}}[g]$, also known as the (set-valued) projection or projector, consists of points $f \in \mathcal{X}$ that are closest to g. For any g, the set $\Pi_{\mathcal{X}}[g]$ can be empty, or it can contain one or more elements. The set $\mathcal{X}$ is called proximal, if $\Pi_{\mathcal{X}}[g]$ is not empty for all g in $\mathcal{H}$; $\mathcal{X}$ is called Chebyshev, if $\Pi_{\mathcal{X}}[g]$ contains exactly one element for all g in $\mathcal{H}$. All nonempty weakly closed subsets of $\mathcal{H}$ are proximal (see Proposition 3.4). In particular, all sets introduced in Definition 2.1 are proximal (see Examples 3.14, 3.15 and 3.24). The multivalued operator $\Pi_{\mathcal{X}} \colon \mathcal{H} \rightrightarrows \mathcal{X}$ is called projection operator, projector, metric projection, nearest point mapping, or best approximation operator. For purposes of this thesis, it will be more convenient to work with single-valued selections of $\Pi_{\mathcal{X}}$. It is important to remember that, in general, single-valued selections are not uniquely defined unless $\mathcal{X}$ is Chebyshev. Following [BCL03], if $\mathcal{X}$ is Chebyshev, we call the unique single-valued projection selection $P_{\mathcal{X}}$ itself a projector.

Definition 3.2 (Projection selections and reflectors).
i) Let $\mathcal{X} \subset \mathcal{H}$, let $\mathcal{D} \subset \mathcal{H}$ be the set of points $g \in \mathcal{H}$ where $\arg\min_{f \in \mathcal{X}} \|g - f\|_2$ is not empty. An operator $P_{\mathcal{X}} \colon \mathcal{D} \to \mathcal{H}$ is called a single-valued selection of a projection onto $\mathcal{X}$, if

$$P_{\mathcal{X}}[g] \in \arg\min_{f \in \mathcal{X}} \|g - f\|_2$$

for all $g \in \mathcal{D}$.
ii) Given a single-valued projection selection $P_{\mathcal{X}}$, the corresponding reflector is defined as $R_{\mathcal{X}} \colon \mathcal{D} \to \mathcal{H}$ with $R_{\mathcal{X}}[g] = 2P_{\mathcal{X}}[g] - g$.

Remark 3.3 (Name of the reflector). Observe that

$$\begin{aligned} P_{\mathcal{X}}[g] &= g + \beta(P_{\mathcal{X}}[g] - g) && \text{for } \beta = 1, \text{ while} \\ R_{\mathcal{X}}[g] &= g + \beta(P_{\mathcal{X}}[g] - g) && \text{for } \beta = 2. \end{aligned}$$

One can easily check that $R_{\mathcal{X}}[g]$ is the reflection of g with respect to the point $P_{\mathcal{X}}[g]$, meaning that

$$\begin{aligned} g &= P_{\mathcal{X}}[g] - (P_{\mathcal{X}}[g] - g) \quad \text{and} \\ R_{\mathcal{X}}[g] &= P_{\mathcal{X}}[g] + (P_{\mathcal{X}}[g] - g). \end{aligned}$$

The following proposition shows that weakly closed sets are proximal in Hilbert spaces and can be found, for examlpe, in [BC17, Thm. 3.14].

PROPOSITION 3.4. *Let $\mathcal{X}$ be a nonempty weakly closed subset of $\mathcal{H}$. Then, $\mathcal{X}$ is proximal, i. e.* $\arg\min_{f\in\mathcal{X}} \|g-f\|_2$ *is not empty for all* $g \in \mathcal{H}$.

Proof. For all $g \in \mathcal{X}$, the set $\arg\min_{f\in\mathcal{X}} \|g-f\|_2 = \{g\}$ is not empty. Let $g \in \mathcal{H}\backslash\mathcal{X}$, let $(f_n)_{n\in\mathbb{N}}$ be a sequence such that

$$\lim_{n\to\infty} \|g-f_n\|_2 = \inf_{f\in\mathcal{X}} \|g-f\|_2. \tag{3.1}$$

Pick an arbitrary element $a \in \mathcal{X}$; then, $\inf_{f\in\mathcal{X}} \|g-f\|_2 \leqslant \|g-a\|_2$. If this inequality is an equality, then $a \in \arg\min_{f\in\mathcal{X}} \|g-f\|_2$ and the proof is complete.

If $\inf_{f\in\mathcal{X}} \|g-f\|_2 < \|g-a\|_2$, then $\|g-f_n\|_2 < \|g-a\|_2$ for infinitely many $n \in \mathbb{N}$. In other words, there exists a subsequence $(f_n)_{n\in M\subset\mathbb{N}}$ such that $\lim_{\substack{n\in M\\ n\to\infty}} \|g-f_n\|_2 = \inf_{f\in\mathcal{X}} \|g-f\|_2$ and $f_n \in B_{\|g-a\|_2}(g)$ for all $n \in M$.

By the Banach-Alaoglu theorem, $B_{\|g-a\|_2}(g)$ is weakly compact, which means that there exists a subsubsequence $(f_n)_{n\in\tilde{M}\subset M}$ that converges weakly to some $f_* \in \mathcal{H}$.

Then, $f_* \in \arg\min_{f\in\mathcal{X}} \|g-f\|_2$. Indeed, on one hand,

$$\inf_{f\in\mathcal{X}} \|g-f\|_2 \leqslant \|g-f_*\|_2;$$

on the other hand,

$$\|g-f_*\|_2 \overset{(*)}{\leqslant} \liminf_{\substack{n\in\tilde{M}\\ n\to\infty}} \|g-f_n\|_2 = \lim_{\substack{n\in\tilde{M}\\ n\to\infty}} \|g-f_n\|_2 = \inf_{f\in\mathcal{X}} \|g-f\|_2,$$

where in $(*)$ we used the fact that $f \mapsto \|g-f\|_2$ is weakly lower semicontinuous (see [BC17, Lemma 2.42]). □

Remark 3.5 (Proximal and Chebyshev sets). Proximal and Chebyshev sets are studied in best approximation theory (see, e. g., [Deu01].). An example of a famous open problem in best approximation theory is the Chebyshev set problem. It asks: is every Chebyshev set in a Hilbert space convex? Some results on this problem, as well as other results on proximality can be found, e. g., in [Bor07] or [FM14].

The following property holds trivially by definition of a projection. Nontheless, the property is extremely useful and is written out for ease of reference.

COROLLARY 3.6 (DISTANCE MINIMIZING PROPERTY). *Let $\mathcal{D} \subset \mathcal{H}$, let $P_{\mathcal{X}}\colon \mathcal{D} \to \mathcal{H}$ be single-valued selection of a projecton onto $\mathcal{X} \subset \mathcal{H}$. Then, $\|g-P_{\mathcal{X}}[g]\|_2 \leqslant \|g-f\|_2$ for any $f \in \mathcal{X}$.*

Remark 3.7. Definition 3.2 and Corollary 3.6 can be formulated on a generic metric space. One merely has to replace the norm $\|g-f\|_2$ by the appropriate metric $d(g, f)$.

The following is another well-known property of projections, see, e.g., [Deu01].

LEMMA 3.8 (INTERPOLATION PROJECTON PROPERTY). *Let $P_{\mathcal{X}}\colon \mathcal{H} \to \mathcal{H}$ be a single-valued projection selection onto a proximal set $\mathcal{X}$. Then, for any $g \in \mathcal{H}$, $\varepsilon \in [0,1]$,*

$$P_{\mathcal{X}}\big[(1-\varepsilon)g + \varepsilon P_{\mathcal{X}}[g]\big] = P_{\mathcal{X}}[g]. \tag{3.2}$$

Proof. We use the definition of a projecton and the fact that $\mathcal{H}$ is a Hilbert space to show that the left-hand side in Equation (3.2) can not be equal to any other point but $P_{\mathcal{X}}[g]$ (see Figure 3.1).

Let $g_\varepsilon = (1-\varepsilon)g + \varepsilon P_{\mathcal{X}}[g]$, let $R = \operatorname{dist}(g,\mathcal{X}) = \inf_{f\in\mathcal{X}} \|g-f\|_2$. We calculate $P_{\mathcal{X}}[g_\varepsilon]$ in two steps: first, we show that $P_{\mathcal{X}}[g_\varepsilon]$ is not inside of the open ball $\mathring{B}_{(1-\varepsilon)R}(g_\varepsilon)$; then, we show that the only candidate for $P_{\mathcal{X}}[g_\varepsilon]$ is $P_{\mathcal{X}}[g]$, since $\partial\mathring{B}_R(g) \cap \partial\mathring{B}_{(1-\varepsilon)R}(g_\varepsilon) = \{P_{\mathcal{X}}[g]\}$. (Throughout this proof, ∂ denotes the boundary of a set.) The claim then follows from the definition of a projection.

First, let us show that $P_{\mathcal{X}}[g_\varepsilon] \notin \mathring{B}_{(1-\varepsilon)R}(g_\varepsilon)$. By definition of R, $\mathring{B}_R(g) \cap \mathcal{X} = \varnothing$. Further, by triangle inequality, $\mathring{B}_{(1-\varepsilon)R}(g_\varepsilon) \subset \mathring{B}_R(g)$, since for any $f \in \mathring{B}_{(1-\varepsilon)R}(g_\varepsilon)$ holds

$$\|g-f\|_2 = \|g-g_\varepsilon+g_\varepsilon-f\|_2 \leqslant \underbrace{\|g-g_\varepsilon\|_2}_{=\varepsilon R} + \underbrace{\|g_\varepsilon-f\|_2}_{<(1-\varepsilon)R} < R.$$

Hence, $P_{\mathcal{X}}[g_\varepsilon] \notin \mathring{B}_{(1-\varepsilon)R}(g_\varepsilon)$.

Second, let us show that $\partial\mathring{B}_R(g) \cap \partial\mathring{B}_{(1-\varepsilon)R}(g_\varepsilon) = \{P_{\mathcal{X}}[g]\}$. Assume that $f \in \partial\mathring{B}_R(g) \cap \partial\mathring{B}_{(1-\varepsilon)R}(g_\varepsilon)$. The points g, g_ε and f satisfy triangle equality

$$\|g-f\|_2 = \|g-g_\varepsilon\|_2 + \|g_\varepsilon - f\|,$$

since

$$\|g-g_\varepsilon\|_2 = \varepsilon R; \quad \|g_\varepsilon - f\|_2 = (1-\varepsilon R); \quad \|g-f\|_2 = R.$$

The space $\mathcal{H}$ is a Hilbert space and therefore it is strictly convex, hence all three points lie on the same line, and

$$f = g + \frac{\|g-f\|}{\|g-g_\varepsilon\|}(g-g_\varepsilon) = g + P_{\mathcal{X}}[g] - g = P_{\mathcal{X}}[g].$$

□

The following is a standard result from best approximation theory (see, e.g., [BC17, Ch. 4, Ch. 20] or [Deu01]).

PROPOSITION 3.9 (PROJECTON SELECTIONS ONTO CONVEX SETS ARE UNIQUE). *Let $\mathcal{X} \subset \mathcal{H}$ be a nonempty weakly closed convex set. Then, the single-valued projection selection $P_{\mathcal{X}}\colon \mathcal{H} \to \mathcal{H}$ exists and is unique.*

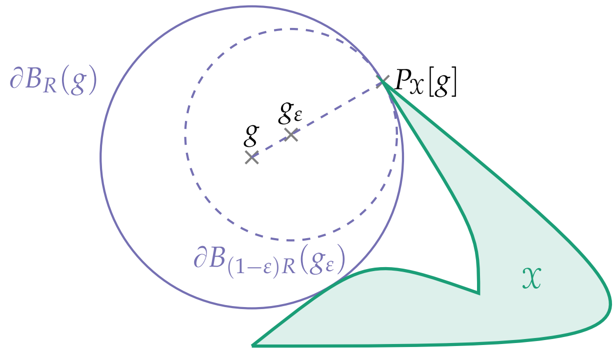

Figure 3.1: Illustration to Lemma 3.8 (interpolation projecton property). One has $P_{\mathcal{X}}[g_\varepsilon] = P_{\mathcal{X}}[g]$, even if the set $\arg\min_{f\in\mathcal{X}} \|g - f\|_2$ contains more than one element.

Proof. By Proposition 3.4, $\mathcal{X}$ is proximal; therefore, at least one selection P_1 onto $\mathcal{X}$ is well-defined.

Further, were P_1 not unique, there would exist two distinct projection selections with values $P_1[g], P_2[g] \in \mathcal{X}$ equidistant from some $g \in \mathcal{H}$, meaning $P_1[g]$ and $P_2[g]$ would lie on the sphere $\partial B_{\|g-P_1[g]\|_2}(g)$. Then, any point on the line segment between $P_1[g]$ and $P_2[g]$ would belong to $\mathcal{X}$ by convexity and be closer to g than $P_1[g]$. This would contradict to P_1 being a projection selection.

Formally, this argument can be written down as follows. Assume that P_1 is not unique, i. e. assume there exist another projection selection P_2 and $g \in \mathcal{H}$ such that $p_1 := P_1[g] \neq P_2[g] =: p_2$. By the parallelogram law,

$$2\|g - p_1\|_2^2 + 2\|g - p_2\|_2^2 = \|2g - p_1 - p_2\|_2^2 + \|p_1 - p_2\|_2^2.$$

Since $\|g - p_1\|_2 = \|g - p_2\|^2$,

$$4\left\|g - \frac{p_1 + p_2}{2}\right\|_2^2 = 4\|g - p_1\|_2^2 - \|p_1 - p_2\|_2^2.$$

Therefore,

$$\left\|g - \frac{p_1 + p_2}{2}\right\|_2 < \|g - p_1\|_2,$$

which contradicts Corollary 3.6, since $\frac{p_1+p_2}{2} \in \mathcal{X}$ by convexity of $\mathcal{X}$. □

The following lemma is also a standard result (see, e.g., [BC17, Ch. 4, Ch. 20] or [Deu01]).

Lemma 3.10 (Angle property for convex projectors). *Let $\mathcal{C} \subset \mathcal{H}$ be convex, let $P_{\mathcal{C}}$ be the corresponding projector. Then,*

$$\langle f - P_{\mathcal{C}}[f], P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]\rangle \leqslant 0$$

for all $f, g \in \mathcal{H}$.

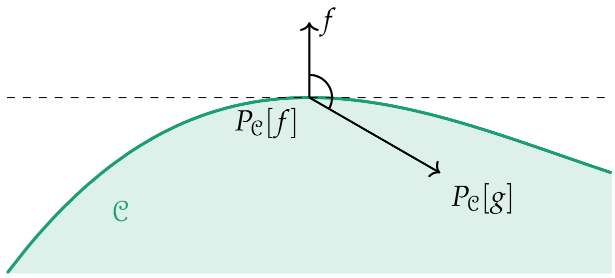

Figure 3.2: Illustration to Lemma 3.10 (angle property for convex projectors). For a convex set $\mathcal{C}$, the angle between $f - P_{\mathcal{C}}[f]$ and $P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]$ is larger that $\pi/2$ for any $g \in \mathcal{H}$, i. e. $\langle f - P_{\mathcal{C}}[f], P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]\rangle \leqslant 0$.

Proof. Idea of the proof: assume the contrary, show that $P_{\mathcal{C}}$ is not unique in contradiction to Corollary 3.6 (cf. Figure 3.2). The argument can be formalized as follows.

If $f \in \mathcal{C}$, the claim is true since $f - P_{\mathcal{C}}[f] = 0$. If $P_{\mathcal{C}}[f] = P_{\mathcal{C}}[g]$, the claim is true since $P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f] = 0$.

Let $f \notin \mathcal{C}$, let $P_{\mathcal{C}}[g] \neq P_{\mathcal{C}}[f]$. Assume that $\langle f - P_{\mathcal{C}}[f], P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]\rangle > 0$. Define $f_\varepsilon = (1-\varepsilon)P_{\mathcal{C}}[f] + \varepsilon P_{\mathcal{C}}[g]$ for $\varepsilon \in (0,1)$. Note that $f_\varepsilon \in \mathcal{C}$ by convexity of $\mathcal{C}$. Then,

$$\begin{aligned}
\|f - f_\varepsilon\|_2^2 &= \|f - P_{\mathcal{C}}[f] - \varepsilon(P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f])\|_2^2 \\
&= \|f - P_{\mathcal{C}}[f]\|_2^2 \underbrace{-2\varepsilon\langle f - P_{\mathcal{C}}[f], P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]\rangle + \varepsilon^2\|P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]\|_2^2}_{<0} \\
&< \|f - P_{\mathcal{C}}[f]\|_2^2
\end{aligned}$$

as long as one picks

$$\varepsilon \in \left(0, 2\frac{\langle f - P_{\mathcal{C}}[f], P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]\rangle}{\|P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]\|_2^2}\right),$$

where the upper bound is strictly larger than zero by assumption. Therefore, there exist $\varepsilon > 0$ such that $\|f - f_\varepsilon\|_2 < \|f - P_{\mathcal{C}}[f]\|_2$, which is in contradiction to Corollary 3.6 since $f_\varepsilon \in \mathcal{C}$. □

The following proposition is in essense the same as [RW09, Example 1.20]. For later use, we formulate it in a slightly different manner, and prove it (since we use it in an infinite-dimensional space, in contrast to the setting of [RW09]).

PROPOSITION 3.11 (CONTINUITY OF PROJECTON SELECTIONS). *Let $\mathcal{X} \subset \mathcal{H}$ be non-empty and weakly closed. Then, any projecton selection $P_{\mathcal{X}}$ is continuous at $g \in \mathcal{H}$ if and only if $\Pi_{\mathcal{X}}$ is single-valued at g.*

Proof. First, we show by contradiction that if $\Pi_{\mathcal{X}}$ is single-valued at some $g \in \mathcal{H}$, then any $P_{\mathcal{X}}$ is continuous at g. Assume that some selection $P_{\mathcal{X}}$ is not continuous at g. Let us show that there exists $q \in \Pi_{\mathcal{X}}[g]$ with $q \neq \Pi_{\mathcal{X}}[g]$; this would be in contradiction to $\Pi_{\mathcal{X}}$ being single-valued at g.

Constructing q. Since $P_{\mathcal{X}}$ is not continuous at g, there exists a sequence $(g_n)_{n_i n \mathbb{N}}$ in $\mathcal{H}$ that converges to some $g \in \mathcal{H}$ as $n \to \infty$, but the corresponding sequence $(p_n)_{n \in \mathbb{N}}$, defined by $p_n := P_{\mathcal{X}}[g_n]$, does not converge to $p := P_{\mathcal{X}}[g]$.

Then, there exists an $\varepsilon \geqslant 0$ and a subsequence of $(g_n)_{n \in \mathbb{N}}$, again denoted by $(g_n)_{n \in \mathbb{N}}$, such that

$$\|p_n - p\|_2 > \varepsilon \tag{3.3}$$

for all $p_n := P_{\mathcal{X}}[g_n]$, $n \in \mathbb{N}$.

It is straghtforward to show that the sequence $(p_n)_{n \in \mathbb{N}}$ is bounded: for any $n \in \mathbb{N}$,

$$\|p_n\|_2 \leqslant \|p_n - g_n\|_2 + \|g_n\|_2 \leqslant \|g_n - p\|_2 + \|g_n\|_2 \leqslant 2\|g_n\|_2 + \|p\|_2,$$

which is bounded since $\mathcal{X}$ is non-empty and since $g_n \to g$ as $n \to \infty$. Therefore, there exists a subsequence of $(g_n)_{n \in \mathbb{N}}$, again denoted by $(g_n)_{n \in \mathbb{N}}$, such that for $p_n := P_{\mathcal{X}}[g_n]$ have $p_n \rightharpoonup q$ as $n \to \infty$ for some $q \in \mathcal{X}$.

Our goal is to show that

$$q \neq p \quad \text{and that} \tag{3.4}$$

$$q \in \Pi_{\mathcal{X}}[g]. \tag{3.5}$$

To show that, we may assume w.l.o.g. that $\|g - p\|_2 > 0$. Indeed, if p were equal to g, then

$$\begin{aligned} \varepsilon &\leqslant \liminf_{n\to\infty} \|g - p_n\|_2 \leqslant \liminf_{n\to\infty} \big(\|g - g_n\|_2 + \underbrace{\|g_n - p_n\|_2}_{\to 0 as n \to \infty}\big) \\ &\leqslant \liminf_{n\to\infty} \|g - p_n\|_2 \overset{(*)}{\leqslant} \liminf_{n\to\infty} \|g - p\|_2 = \liminf_{n\to\infty} \|g - g\|_2 = 0, \end{aligned}$$

which would be a contradiction (we used the distance-minimizing projection property Corollary 3.6 in $(*)$).

Thus, we can assume that $\|g - p\|_2 > 0$ and we want to show Equation (3.4) and Equation (3.5).

Establishing $q \neq p$. By Mazur's lemma, we know that

$$q \in \overline{\operatorname{conv}}_{n \geqslant N}\{p_n\},$$

where we can pick any $N \in \mathbb{N}$. For reasons that will be apparent later, pick $N \in \mathbb{N}$ such that

$$\|g - p_n\| \leqslant (1 + \alpha)s \qquad \text{for all } n \geqslant N,$$

where $s := \|g - p\|_2 > 0$, and where $\alpha > 0$ is picked such that $2\alpha + \alpha^2 = \frac{\varepsilon}{2s^2}$ (i.e. $\alpha := 1 - \sqrt{1 - \frac{\varepsilon}{2s^2}}$, where we can w.l.o.g. assume that $\frac{\varepsilon}{2s^2} < 1$; otherwise, we can redefine ε to be smaller in the beginning of the proof). The choice of α and N will be justified in Equation (3.7) below; the idea is to show that for such large N, $\overline{\operatorname{conv}}_{n \geqslant N}\{p_n\}$ will belong to a closed affine half-space of $\mathcal{H}$ that does not contain p.

By definition of α and N, p_n belongs to the ball $B_{(1+\alpha)s}[g]$ for all $n \geqslant N$; further, by construction of p_n, p_n is outside of $B_\varepsilon[p]$. Overall, for $n \geqslant N$,

$$\text{all } p_n \text{ are contained in the ball difference } B_{(1+\alpha)s}[g] \backslash B_\varepsilon[p]. \tag{3.6}$$

Let us show the geometric fact that this ball difference can be separated by a hyperplane from p.

Specifically, let us show that for an appropriately chosen $c < 1$,

$$\begin{cases} \langle y-g, p-g \rangle \geqslant c\|p-g\|_2^2 \\ y \in B_{(1+\alpha)s}[g] \end{cases} \qquad \rightarrow \qquad y \in B_\varepsilon[p]; \tag{3.7}$$

from this, it will be straigtforward to show that $q \neq p$. To establish (3.7), choose $c := \frac{1}{2}\big((1+\alpha)^2 + 1 - \frac{\varepsilon}{s^2}\big)$. Then, by definition of α,

$$c = \frac{1}{2}(2 + 2\alpha + \alpha^2 - \frac{\varepsilon}{s^2}) \tag{3.8}$$
$$= \frac{1}{2}(2 + \frac{\varepsilon}{2s^2} - \frac{\varepsilon}{s^2}) \tag{3.9}$$
$$= 1 - \frac{\varepsilon}{4s^2} < 1. \tag{3.10}$$

Further, w.l.o.g $c > 0$, otherwise one can pick g a smaller ε in the beginning of the proof. Then, for any $y \in B_{(1+\alpha)s}[g]$ with the hyperplane condition $\langle y-g, p-g \rangle \geqslant c\|p-g\|_2^2$ have

$$\begin{aligned} \|y-p\|_2^2 &= \|y-g\|_2^2 + \|g-p\|_2^2 + 2\langle y-g, g-p \rangle \\ &\leqslant (1+\alpha)^2 s^2 + s^2 - 2\langle y-g, p-g \rangle \\ &\leqslant (1+\alpha)^2 s^2 + s^2 - 2c\|p-g\|_2^2 \\ &= (1+\alpha)^2 s^2 + s^2 - 2cs^2 \\ &= \left((1+\alpha)^2 + 1 - 2 + \frac{\varepsilon}{2s^2}\right) s^2 \\ &= (\frac{\varepsilon}{2s^2} + \frac{\varepsilon}{2s^2})s^2 = \varepsilon; \end{aligned}$$

thus, $y \in B_\varepsilon[p]$ and (3.7) is proven.

From (3.7) follow: for all $y \in B_{(1+\alpha)s}[g] \backslash B_\varepsilon[p]$ holds

$$\langle y-g, p-g \rangle \leqslant c\|p-g\|_2^2.$$

Together with Equation (3.6) this implies

$$\langle p_n - g, p-g \rangle \leqslant c\|p-g\|_2^2.$$

Since the affine hyperplane

$$\{y \in \mathcal{H} \mid \langle y - g, p - g\rangle \leqslant c\|p - g\|_2^2\}$$

is closed and convex, have

$$\overline{\operatorname{conv}}_{n \geqslant \mathbb{N}} \langle p_n - g, p - g\rangle \leqslant c\|p - g\|_2^2,$$

which — since $p_n \rightharpoonup q$ as $n \to \infty$ — implies

$$\langle q - g, p - g\rangle \leqslant c\|p - g\|_2^2,$$

If q were equal to p,

$$\|p - q\|_2^2 \leqslant c\|p - g\|_2^2$$

would be a contradiction since $c < 1$; thus, $q \neq p$.

Demonstrating that $q \in \Pi_{\mathcal{X}}[g]$. Having constructed q as a weak limit of $p_n = P_{\mathcal{X}}[g_n]$, where $g_n \to g$, and having shown that $q \neq p = P_{\mathcal{X}}[g]$, we can finally estimate

$$\begin{aligned}
\|g - q\|_2^2 &= \|g\|_2^2 - \underbrace{\langle g, q\rangle}_{=\liminf_{n\to\infty}\langle g, p_n\rangle} + \underbrace{\|q\|_2^2}_{\leqslant \|p_n\|_2^2} \\
&\leqslant \liminf_{n\to\infty} \|g\|_2^2 - 2\langle g, p_n\rangle + \|p_n\|_2^2 \\
&= \liminf_{n\to\infty} \|g\|_2^2 - \|g_n\|_2^2 + \|g_n\|_2^2 \\
&\qquad - 2\langle g - g_n + g_n, p_n\rangle + \|p_n\|_2^2 \\
&= \underbrace{\liminf_{n\to\infty} \|g\|_2^2 - \|g_n\|_2^2 - 2\langle g - g_n, p_n\rangle + \|p_n\|_2^2}_{=0 \text{ since } g_n \to g \text{ and since } (p_n) \text{ is bounded}} \\
&= \liminf_{n\to\infty} \|g_n - p_n\|_2^2 \\
&= \liminf_{n\to\infty} 2E_{\mathcal{X}}[g_n] = 2E_{\mathcal{X}}[g],
\end{aligned}$$

where we have used the continuity of $E_{\mathcal{X}}$ (Lemma 4.5) in the last step. Thus, $q \in \Pi_{\mathcal{X}}[g]$ and $q \neq p$, meaning that $\Pi_{\mathcal{X}}$ is not single-valued at g, concluding one direction of the proposition claim.

Other direction — if $\Pi_{\mathcal{X}}$ is not single-valued at g, then $P_{\mathcal{X}}$ is not continuous at g — is straightforward: if $p, q \in \Pi_{\mathcal{X}}[g]$ with $p \neq q$, let

$$g_n := \begin{cases} (1 - \frac{1}{n})g + \frac{1}{n}p \text{ for odd } n \in \mathbb{N}; \\ (1 - \frac{1}{n})g + \frac{1}{n}q \text{ for even } n \in \mathbb{N}, \end{cases}$$

let $P_{\mathcal{X}}$ be a projecton selection onto $\mathcal{X}$. By the interpolation property Lemma 3.8, $P_{\mathcal{X}}[g_n] = p$ for odd n, and $P_{\mathcal{X}}[g_n] = q$ for even n, meaning that $P_{\mathcal{X}}$ can not be continuous at $g = \lim_{n\to\infty} g_n$. □

The following property of weakly closed projections will be used in Proposition 6.17 to establish sufficient conditions for accumulation points of a generalized alternating projections algorithm to be fixed points.

Proposition 3.12 ("$g_n \to g \Rightarrow P_{\mathcal{X}}[g_n] \rightharpoonup P_{\mathcal{X}}[g]$"). *Let $\mathcal{X} \subset \mathcal{H}$ be weakly closed, let $(g_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ that converges to $g \in \mathcal{H}$.*

Then, there exists a selection $P_{\mathcal{X}}$ and a subsequence of $(g_n)_n$, again denoted by $(g_n)_n$, such that

$$P_{\mathcal{X}}[g_n] \rightharpoonup P_{\mathcal{X}}[g] \quad \text{as } n \to \infty.$$

Proof. Pick any projecton selection $P_{\mathcal{X}}$ onto $\mathcal{X}$.

The sequence $(P_{\mathcal{X}}[g_n])_n$ is bounded. Indeed, $(g_n)_{n\in\mathbb{N}}$ is bounded (since it is convergent to g). Therefore, there exists $C_1 < \infty$ such that $\|g_n\|_2 \leqslant C_1$ for all $n \in \mathbb{N}$. Further, since $\mathcal{X}$ is not empty, there exists $C_2 < \infty$ such that $B_{C_2}(0) \cap \mathcal{X}$ is not empty and contains some element $\tilde{g} \in \mathcal{H}$. Therefore, for all $n \in \mathbb{N}$,

$$\|P_{\mathcal{X}}[g_n]\|_2 \leqslant \|P_{\mathcal{X}}[g_n] - g_n\|_2 + \|g_n\|_2 \leqslant \|\tilde{g} - g_n\|_2 + \|g_n\|_2 \leqslant C_2 + 2C_1. \tag{3.11}$$

Since $(P_{\mathcal{X}}[g_n])_n$ is bounded and contained in the weakly closed set $\mathcal{X}$, there exists a weakly convergent subsequence, again denoted by $(P_{\mathcal{X}}[g_n])_n$, with

$$P_{\mathcal{X}}[g_n] \rightharpoonup p \in \mathcal{X}.$$

Let us show that $p \in \Pi_{\mathcal{X}}[g]$. Indeed, since $\|\cdot\|_2^2$ is w.l.s.c.,

$$\begin{aligned}
\|g - p\|_2^2 &\leqslant \|g\|_2^2 - \lim_{n\to\infty} 2\langle g, P_{\mathcal{X}}[g_n]\rangle + \liminf_{n\to\infty} \|P_{\mathcal{X}}[g_n]\|_2^2 \\
&= \liminf_{n\to\infty}(\|g_n\|_2^2 - 2\langle g_n, P_{\mathcal{X}}[g_n]\rangle + \|P_{\mathcal{X}}[g_n]\|_2^2) \\
&= \liminf_{n\to\infty} \|g_n - P_{\mathcal{X}}[g_n]\|_2^2
\end{aligned}$$

— and by the distance-minimizing property of projections —

$$\leqslant \liminf_{n\to\infty} \|g_n - P_{\mathcal{X}}[g]\|_2^2 = \|g - P_{\mathcal{X}}[g]\|_2^2,$$

since g_n converges to g.

Thus, $p \in \Pi_{\mathcal{X}}[g]$. Pick a projecton selection that equals to p at g, and to $P_{\mathcal{X}}[f]$ at any $f \neq g$, to follow the claim. □

3.2 PROJECTONS ONTO ADDITIONAL CONSTRAINT SETS

The section continues with a description of projection operators for phase retrieval sets in Examples 3.14, 3.15 and 3.24; it shows the proofs of explicit forms of projection operators. These examples are widely known in phase retrieval literature. For purposes of mathematical rigor, Examples 3.14, 3.15 and 3.24 present proofs that defined operators are, indeed, projections.

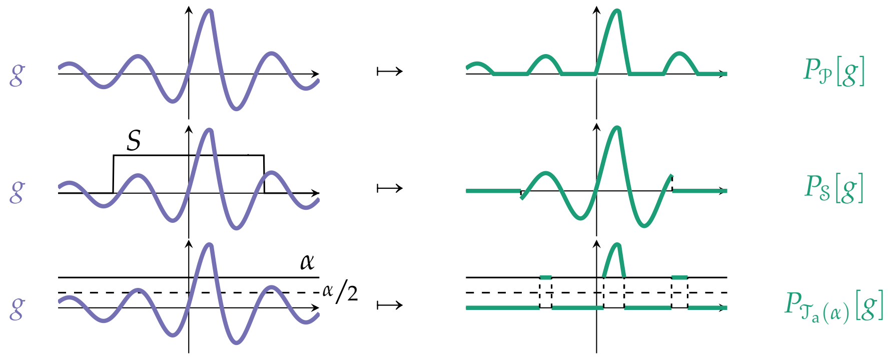

Figure 3.3: Illustration to Example 3.14 (positivity, support, amplitude thresholding projectons).
The sketch illustrates positivity, support, and amplitude thresholding projectons for a function $g \in \mathcal{H}(\mathbb{R})$. In the sketch, the values of $P_{\mathcal{S}}[g]$ and $P_{\mathcal{T}_a(\alpha)}[g]$ at discontinuities can be chosen freely, as such points constitute a Lebesgue null set.

Notation 3.13. We use the following shorthand notation for indicator functions on Ω:

$$\mathbb{1}_{\{\text{condition}\}}(x) = \mathbb{1}_{\{x \in \Omega | \text{condition is valid at } x\}}(x)$$

for a. a. $x \in \Omega$. For example, $\mathbb{1}_{\{g \geqslant 0\}}(x) = \mathbb{1}_{\{x \in \Omega | g(x) \geqslant 0\}}(x)$.

Example 3.14 (Positivity, support, amplitude thresholding projectons). For sets from Definition 2.1, define the following operators with domain $\mathcal{D} = \mathcal{H}$:

$$\begin{aligned} A_{\mathcal{P}}[g] &= \mathbb{1}_{\{g \geqslant 0\}} g; && \text{(positivity)} \\ A_{\mathcal{S}}[g] &= \mathbb{1}_{S} g; && \text{(support)} \\ A_{\mathcal{T}_a(\alpha)}[g] &= \mathbb{1}_{\{g \geqslant \alpha\}} g + \mathbb{1}_{\{\alpha > g > \frac{\alpha}{2}\} \cup \tilde{S}}\, \alpha, && \text{(amplitude thresholding)} \end{aligned}$$

where $\tilde{S}$ is any measurable subset of $\{g = \frac{\alpha}{2}\}$. For simplicity, we use the choice $\tilde{S} = \{g = \frac{\alpha}{2}\}$ unless mentioned otherwise.

These operators are projecton selections onto corresponding sets: $P_{\mathcal{P}} = A_{\mathcal{P}}$, $P_{\mathcal{S}} = A_{\mathcal{S}}$, $P_{\mathcal{T}_a(\alpha)} = A_{\mathcal{T}_a(\alpha)}$. Further, the positivity and support selections are unique, and the amplitude thersholding ambiguous only in the choice of $\tilde{S} \subset \{g = \frac{\alpha}{2}\}$.

Proof (that $P_{\mathcal{P}} = A_{\mathcal{P}}$, $P_{\mathcal{T}_a(\alpha)} = A_{\mathcal{T}_a(\alpha)}$ and $P_{\mathcal{S}} = A_{\mathcal{S}}$). Let us prove by contradiction that $P_{\mathcal{T}_a(\alpha)} = A_{\mathcal{T}_a(\alpha)}$. Note that this case implies that $P_{\mathcal{P}} = A_{\mathcal{P}}$, since $P_{\mathcal{P}} = P_{\mathcal{T}_a(\alpha)}$ and $A_{\mathcal{P}} = A_{\mathcal{T}_a(\alpha)}$ for $\alpha = 0$.

Assume that $A_{\mathcal{T}_a(\alpha)}$ is not a projection selection, meaning there exist $g \in \mathcal{H}$ and $f \in \mathcal{T}_a(\alpha)$ such that

$$\|g - f\|_2 < \|g - A_{\mathcal{T}_a(\alpha)}[g]\|_2. \tag{3.12}$$

This leads to a contradiction, since one can observe reverse inequalities pointwise a. e. in Ω as follows.

First, for almost all $x \in \Omega$ with $g(x) \geqslant \alpha$ one can assume that $f(x) = A_{\mathcal{T}_a(\alpha)}[g](x) = g(x)$. Otherwise, one can redefine $f(x)$ to equal $g(x)$ at such points, as this will only decrease the value of the left-hand side of Equation (3.12).

Second, for almost all $x \in \Omega$ with $g(x) \in [\frac{\alpha}{2}, \alpha)$ holds

$$|g(x) - \alpha| \leqslant \frac{\alpha}{2} \leqslant |g(x)| = |g(x) - f(x)|$$

if $f(x) = 0$, and

$$|g(x) - \alpha| = \alpha - g(x) \leqslant f(x) - g(x) = |g(x) - f(x)|$$

if $f(x) \geqslant \alpha$.

Third, for all $x \in \Omega$ with $g(x) \in (0, \frac{\alpha}{2}]$ holds

$$|g(x)| = |g(x) - 0| \leqslant |g(x) - f(x)|$$

if $f(x) = 0$, and

$$|g(x)| \leqslant \frac{\alpha}{2} \leqslant f(x) - g(x) \leqslant |g(x) - f(x)|$$

if $f(x) \geqslant \alpha$.

Finally, for all $x \in \Omega$ with $g(x) \in (-\infty, 0]$ holds

$$|g(x)| = |g(x) - 0| \leqslant |g(x) - f(x)|$$

if $f(x) = 0$, and

$$|g(x)| = -g(x) \leqslant f(x) - g(x) \leqslant |g(x) - f(x)|$$

if $f(x) \geqslant \alpha$.

Overall, obtain

$$\|g - A_{\mathcal{T}_a(\alpha)}[g]\|_2 = \|\mathbb{1}_{\{\frac{\alpha}{2} < g < \alpha\} \cup \tilde{S}} (g - \alpha)\|_2 + \|\mathbb{1}_{\{g \leqslant \frac{\alpha}{2}\} \setminus \tilde{S}}\, g\|_2 \leqslant \|g - f\|_2$$

in contradiction to Equation (3.12). Therefore, $P_{\mathcal{T}_a(\alpha)} = A_{\mathcal{T}_a(\alpha)}$; setting $\alpha = 0$ implies that $P_{\mathcal{P}} = A_{\mathcal{P}}$. In particular, this argument shows that the operator $P_{\mathcal{T}_a(\alpha)}$ is ambiguous only in the choice of $\tilde{S} \subset \{g = \frac{\alpha}{2}\}$.

Let us now prove that $P_{\mathcal{S}} = A_{\mathcal{S}}$. Assume that there exists $f \in \mathcal{S}$ such that

$$\|g - f\|_2 \leqslant \|g - A_{\mathcal{S}}[g]\|_2. \tag{3.13}$$

But

$$\|g - A_{\mathcal{S}}[g]\|_2 = \|\mathbb{1}_{\Omega \setminus S}\, g\|_2 = \|\mathbb{1}_{\Omega \setminus S}(g - f)\|_2,$$

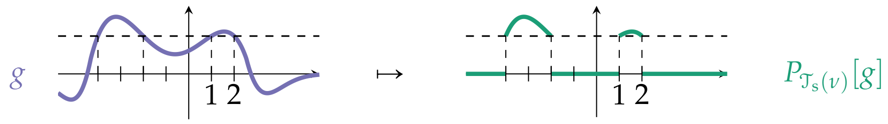

Figure 3.4: Illustration to Example 3.15 (support size projecton selection). The sketch illustrates $P_{\mathcal{T}_s(\nu)}$ for $\nu = 3$; i. e. the Lebesgue measure of $\operatorname{supp} P_{\mathcal{T}_s(\nu)}[g]$ equals 3. In the sketch, the value of $P_{\mathcal{T}_s(\nu)}[g]$ at discontinuities can be chosen freely, as such points constitute a Lebesgue null set.

since $\operatorname{supp} f \in S$; therefore,

$$\|g - A_{\mathcal{S}}[g]\|_2 = \|\mathbb{1}_{\Omega \setminus S}(g - f)\|_2 \leqslant \|g - f\|_2$$

in contradiction to Equation (3.13). Therefore, $P_{\mathcal{S}} = A_{\mathcal{S}}$.

Projection selections $P_{\mathcal{P}}$ and $P_{\mathcal{S}}$ are unique by Proposition 3.9, since the sets $\mathcal{P}$ and $\mathcal{S}$ are convex. □

Example 3.15 (Support size projecton). A projecton onto the support size constraint $\mathcal{T}_s(\nu)$ can be elegantly described for finite-dimensional physical space $\Omega = \mathbb{T}_N^d$: to calculate $P_{\mathcal{T}_s(\nu)}$ at $g \in \mathcal{H}(\Omega)$, set g to zero at all but ν largest non-negative values. Note that the choice of ν largest non-negative values may not be unique; in that case, it suffices to take any combination of them. For example, if $g \equiv 1$, any element is a largest element of g, and one has to resolve this freedom of choice by picking a set of coordinates $\tilde{S}$ of size ν and setting g to zero at all $x \notin \tilde{S}$.

To describe the case of the more general domains $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d\}$, one may generalize as follows: estimate a thresholding parameter $\tilde{\alpha} \geqslant 0$ for which the support of $\{g \geqslant \tilde{\alpha}\}$ has an appropriate measure; then, resolve any remaining freedom of choice by picking an appropriate $\tilde{S}$. Formally, for $\nu > 0$, $g \in \mathcal{H}(\Omega)$ define

$$A_{\mathcal{T}_s(\nu)}[g] := \mathbb{1}_{S \cup \tilde{S}}\, g.$$

Here,

$$S := \bigcup_{\alpha > \tilde{\alpha}} \{g \geqslant \alpha\},$$

and $\tilde{\alpha}$ is the smallest non-negative number for which $\lambda(\{g \geqslant \alpha\})$ is not larger than ν, i. e.

$$\tilde{\alpha} := \inf\{\alpha \geqslant 0 \mid \lambda(\{g \geqslant \alpha\}) > \nu\}$$

for the Lebesgue measure λ.

As for the set $\tilde{S}$, it denotes any measurable subset of $\{g = \tilde{\alpha}\}$ of measure $\nu - \lambda(S)$. The set $\tilde{S}$ is empty, if $\lambda(S) = \nu$, i. e. if the choice of largest elements is unique. Such case is depicted in Figure 3.4.

Then, $P_{\mathcal{T}_s(\nu)} = A_{\mathcal{T}_s(\nu)}$.

Proof (that $P_{\mathcal{T}_s(\nu)} = A_{\mathcal{T}_s(\nu)}$). Observe that by construction of $A_{\mathcal{T}_s(\nu)}$ holds

$$A_{\mathcal{T}_s(\nu)}[g] \in \arg\max\Big\{ \big\| \mathbb{1}_Q \mathbb{1}_{\{g \geqslant 0\}} g \big\|_2 \Big| Q \subset \Omega \text{ measurable}, \lambda(Q) \leqslant \nu \Big\}. \tag{3.14}$$

Let us show the claim by contradiction: assume that there exists $\tilde{f} \in \mathcal{T}_s(\nu)$ such that

$$\|g - \tilde{f}\|_2 < \|g - A_{\mathcal{T}_s(\nu)}[g]\|_2. \tag{3.15}$$

Let $Q = \operatorname{supp} \tilde{f}$, let $f = \mathbb{1}_Q g$; then, trivially,

$$\|g - f\|_2 \leqslant \|g - \tilde{f}\|_2 < \|g - A_{\mathcal{T}_s(\nu)}[g]\|_2. \tag{3.16}$$

Further, from the assumption $\tilde{f} \in \mathcal{T}_s(\nu)$ follows $\lambda(Q) \leqslant \nu$. From Equation (3.14) we know that

$$\|A_{\mathcal{T}_s(\nu)}[g]\|_2 = \|\mathbb{1}_{S \cup \tilde{S}} g\|_2 \geqslant \|\mathbb{1}_Q g\|_2 = \|f\|_2.$$

But then

$$\begin{aligned} \|\mathbb{1}_{\Omega \setminus (S \cup \tilde{S})} g\|_2 &\leqslant \|\mathbb{1}_{\Omega \setminus Q} g\|_2 && \Rightarrow \\ \|g - \mathbb{1}_{S \cup \tilde{S}} g\|_2 &\leqslant \|g - \mathbb{1}_Q g\|_2 && \Rightarrow \\ \|g - A_{\mathcal{T}_s(\nu)}[g]\|_2 &\leqslant \|g - f\|_2 \end{aligned}$$

in contradiction to Equation (3.16). Therefore, $A_{\mathcal{T}_s(\nu)}$ is a projecton selection onto $\mathcal{T}_s(\nu)$, i. e. $P_{\mathcal{T}_s(\nu)} = A_{\mathcal{T}_s(\nu)}$.

The definition of $P_{\mathcal{T}_s(\nu)}$ implies that a hard sparsity projector selection is not unique. However, it nessesarily has the form $P_{\mathcal{T}_s(\nu)}[g] = \mathbb{1}_{S \cup \tilde{S}}$. Indeed, if f is a projecton selection of g onto $\mathcal{T}_s(\nu)$, it must have the form $f = \mathbb{1}_Q g$ for $Q = \operatorname{supp} f$ by the same argument as above; and by construction of $P_{\mathcal{T}_s(\nu)}[g]$, cf. Equation (3.14), f can differ from $P_{\mathcal{T}_s(\nu)}[g]$ only in the choice of $\tilde{S}$. □

3.3 CONDITIONS FOR THE MODULUS CONSTRAINT SET TO BE WEAKLY CLOSED

This section demonstrates that the modulus set is weakly closed on bounded domains (joint work with Gero Friesecke). The weak closedness of a set is a powerful property in best approximation theory. This property underpins several important results of the thesis: it is used in Section 4.3 for subdifferential calculus, and in Chapter 6 for fixed point results. These, in turn, enable rigorous analysis of evolution equations done in Chapters 7 and 9. The section proceeds as follows.

Definition 3.16 and Lemma 3.18 recall relevant topological definitions and properties. Remark 3.19 recalls the observation from [LBL02, Property 4.1] that the modulus set is not weakly closed on $L^2(\mathbb{R}^d)$.

The key component in the result of the section is the compactness theorem by Pego [Peg85, Theorem 3], recalled in Lemma 3.20.

The main result of the section is Theorem 3.21, demonstrating weak closedness of $\mathcal{M} \subset L^2(\Omega)$ on bounded domains Ω. The particular case of $\mathcal{M}$ on a torus domain is very similar, it is demonstrated in Theorem 3.22.

DEFINITION 3.16 (WEAK CLOSEDNESS AND WEAK COMPACTNESS). *Let $\mathcal{H}$ be a Hilbert space. A subset $\mathcal{X} \subseteq \mathcal{H}$ is called*

1. weakly closed *if*
   it is closed in the weak topology;

2. weakly sequentially closed *if*
   for every weakly convergent subsequence in $\mathcal{X}$, its weak limit is also in $\mathcal{X}$;

3. weakly compact *if*
   it is compact in the weak topology (every weakly open cover of $\mathcal{X}$ has a finite subcover);

4. weakly sequentially compact *if*
   every sequence in $\mathcal{X}$ has a subsequence that is weakly converging to some element in $\mathcal{X}$;

5. weakly relatively sequentially compact *if*
   every sequence in $\mathcal{X}$ has a subsequence that is weakly converging to some element in $\mathcal{H}$;

6. relatively sequentially compact *if*
   every sequence in $\mathcal{X}$ has a subsequence that is (strongly) converging to some element in $\mathcal{H}$.

In this section we demonstrate that the modulus set is weakly sequentially compact and thus weakly sequentially closed. As a sidenote, recall that weak sequential closedness by itself does not imply weak closedness:

Remark 3.17. Let $\mathcal{H}$ be a Hilbert space. Recall that

$$\mathcal{X} \subset \mathcal{H} \text{ weakly sequentially closed} \quad \not\Rightarrow \quad \mathcal{X} \subset \mathcal{H} \text{ weakly closed.}$$

For proof, see, e.g., [BC17, Example 3.33].

However, in Hilbert spaces weak sequential compactness implies weak closedness, which will be sufficient for our purposes:

LEMMA 3.18 (SEE E.G. [BC17, COROLLARY 2.38]). *Let $\mathcal{X}$ be a subset of a Hilbert space $\mathcal{H}$. Then the following are equivalent:*

1) $\mathcal{X}$ is weakly compact.

2) $\mathcal{X}$ is weakly sequentially compact.

3) $\mathcal{X}$ is weakly closed and bounded.

It is known that the set $\mathcal{M}$ is not weakly closed on $\mathbb{R}^d$:

Remark 3.19 ($\mathcal{M}$ is not weakly closed on $\mathbb{R}^d$). As demonstrated in [LBL02, Property 4.1], the modulus set $\mathcal{M}(\sqrt{I}) \subset L^2(\mathbb{R}^d)$ is neither convex nor weakly closed if $\sqrt{I}$ is not identically equal to zero.

$\mathcal{M}$ is not convex. Indeed, if $\sqrt{I} \not\equiv 0$, then for any $f \in \mathcal{M}(\sqrt{I})$ one has $-f \in \mathcal{M}(\sqrt{I})$, but $0.5f + 0.5(-f) = 0 \notin \mathcal{M}(\sqrt{I})$; thus, $\mathcal{M}(\sqrt{I})$ is not convex. In fact, this is true for any domain $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$ as long as $\sqrt{I} \not\equiv 0$.

$\mathcal{M}$ is not weakly closed. Further, if $\sqrt{I} \not\equiv 0$, then define the sequence $(f_n)_{n\in\mathbb{N}}$, $f_n \in \mathcal{M}(\sqrt{I})$, using $\hat{f}_n(k) := \sqrt{I}(k)e^{ikn}$ for all $n \in \mathbb{N}$ and almost all $k \in \mathbb{R}^d$. Then, for any $g \in L^2(\mathbb{R}^d)$ have

$$\langle \hat{f}_n, \hat{g} \rangle = \int_{\mathbb{R}^d} \sqrt{I}(k)\hat{g}(k)e^{ikn}dk \to 0 \text{ as } n \to \infty$$

by the Riemann-Lebesgue lemma, since $\sqrt{I}\hat{g} \in L^1(\mathbb{R}^d)$ (by Hölder's inequality and because $\sqrt{I}$ and $\hat{g}$ belong to $L^2(\mathbb{R}^d)$). Thus, f_n is weakly convergent to the zero function, which is not in $\mathcal{M}$. Therefore, $\mathcal{M}$ is not weakly sequentially compact, and by Lemma 3.18 $\mathcal{M}$ is not weakly closed.

However — as will be shown below — the set $\mathcal{M}$ is weakly closed if all its elements are defined on a bounded domain.

Indeed, in certain cases $\mathcal{M}$ is weakly sequentially compact and thus weakly closed by Lemma 3.18. To show weak sequential compactness of $\mathcal{M}$, we use the following lemma by Pego which is not as well known as it deserves to be.

LEMMA 3.20 ([Peg85, THEOREM 3]). *A bounded subset $\mathcal{X}$ of $L^2(\mathbb{R}^d)$ is relatively sequentially compact if and only if*

$$\sup_{f\in\mathcal{X}} \int_{|x|>R} |f(x)|^2 dx \to 0$$

and

$$\sup_{f\in\mathcal{X}} \int_{|k|>R} |\hat{f}(k)|^2 dk \to 0$$

as $R \to \infty$.

This lemma can be used to show weak closedness of $\mathcal{M}$ on bounded domains in the following manner.

THEOREM 3.21 ($\mathcal{M}(\sqrt{I})$ IS WEAKLY CLOSED ON A BOUNDED DOMAIN). *Let $\Omega \subset \mathbb{R}^d$ bounded and Lebesgue-measurable, and let $I \in L^1(\mathbb{R}^d)$ with $I \geqslant 0$*

(so that $\sqrt{I}$ belongs to $L^2(\mathbb{R}^d)$). Then the modulus set $\mathcal{M}(\sqrt{I}) = \{f \in L^2(\Omega) \mid |\hat{f}| = \sqrt{I}\}$ is weakly closed (where $\hat{f}$ is the Fourier transform of the extension of f by zero to all of $\mathbb{R}^d$).

Proof. We may assume $\mathcal{M}(\sqrt{I})$ is nonempty (otherwise there is nothing to show). By assumption on Ω, there exists an $R > 0$ such that $\Omega \subset \{|x| \leqslant R\}$. Therefore,

$$\sup_{f \in \mathcal{M}(\sqrt{I})} \int_{|x|>\tilde{R}} |f(x)|^2 dx = 0$$

for all $\tilde{R} > R$. Moreover, by definition of $\mathcal{M}(\sqrt{I})$,

$$\sup_{f \in \mathcal{M}(\sqrt{I})} \int_{|k|>R} |\hat{f}(k)|^2 dk = \int_{|k|>R} |\sqrt{I}(k)|^2 dk \to 0 \text{ as } R \to \infty. \tag{3.17}$$

Thus, $\mathcal{M}(\sqrt{I})$ is relatively sequentially compact, by Pego's lemma (Lemma 3.20).

Let us deduce that $\mathcal{M}(\sqrt{I})$ is sequentially compact. Let $(f_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{M}(\sqrt{I})$; by definition of relative sequential compactness, there exists a subsequence $(f_{n_\ell})_{\ell \in \mathbb{N}}$ that converges strongly to some $f \in L^2(\mathbb{R}^d)$. By L^2-continuity of the Fourier transform, $(\hat{f}_{n_\ell})_{\ell \in \mathbb{N}}$ converges strongly to $\hat{f} \in L^2(\mathbb{R}^d)$. Therefore, there exists a further subsequence, again denoted $(\hat{f}_{n_\ell})_{\ell \in \mathbb{N}}$, that converges almost everywhere to $\hat{f}$. Since the set $\{z \in \mathbb{C} \mid |z| = \sqrt{I}(k)\}$ is closed, the limit of $(\hat{f}_{n_\ell})_{\ell \in \mathbb{N}}$ has absolute value $\sqrt{I}$ almost everywhere. Thus $\hat{f} \in \hat{\mathcal{M}}$, establishing the asserted sequential compactness.

Since $f_{n_\ell} \to f$, one also has $f_{n_\ell} \rightharpoonup f$; thus, $\mathcal{M}$ is also weakly sequentially compact, and therefore weakly closed by Lemma 3.18. □

The above case is applicable to non-crystallographic applications. For crystallographic applications, it is necessary to consider the case of the torus $\mathbb{T}^d = \mathbb{R}^d / 2\pi\mathbb{Z}^d$, which is easier. In this case, the Fourier transform $\hat{f}$ of $f \in L^2(\mathbb{T}^d)$ (in the sense of abelian groups) is a function on the dual group $\mathbb{Z}^d$, coinciding with the vector $(\hat{f}(k))_{k \in \mathbb{Z}^d}$ of Fourier coefficients of f.

Theorem 3.22 ($\mathcal{M}(\sqrt{I})$ is weakly closed on a torus). *Let $I \in \ell^1(\mathbb{Z}^d)$ with $I \geqslant 0$ (so that $\sqrt{I}$ belongs to $\ell^2(\mathbb{Z}^d)$). Then the modulus set*

$$\mathcal{M}(\sqrt{I}) = \{f \in L^2(\mathbb{T}^d) \mid |\hat{f}| = \sqrt{I}\}$$

is weakly closed. (Here, as explained above, $(\hat{f}(k))_{k \in \mathbb{Z}^d}$ is the vector of Fourier coefficients of f.)

Proof. By Lemma 3.23 (see below), the modulus set $\mathcal{M}(\sqrt{I})$ is weakly sequentially compact. It is thus weakly closed, by Lemma 3.18. □

Lemma 3.23. *Let $\sqrt{I}$ be as in Theorem 3.22. Then the modulus set $\mathcal{M}(\sqrt{I})$ is weakly sequentially compact.*

Proof. We first claim that $\mathcal{M}(\sqrt{I})$ is weakly relatively sequentially compact. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\sqrt{I})$. By Parseval's equation,

$$\|f_n\|^2_{L^2(\mathbb{T}^d)} = (2\pi)^d\|\hat{f}\|^2_{\ell^2(\mathbb{Z}^d)} = (2\pi)^d\|I\|_{\ell^1(\mathbb{Z}^d)}.$$

Since the right hand side is finite by assumption, the sequence is bounded in L^2, and hence possesses a subsequence $(f_{n_\ell})_{\ell\in\mathbb{N}}$ converging weakly in L^2 to some $f \in L^2(\mathbb{T}^d)$, establishing our claim.

We now show that $\mathcal{M}(\sqrt{I})$ is in fact weakly sequentially compact. Indeed, the above subsequence satisfies

$$\int_{\mathbb{T}^d} f_{n_\ell}(x)e^{-ik\cdot x}dx = \hat{f}_{n_\ell}(k) \to \hat{f}(k) \ \ \forall k \in \mathbb{Z}^d.$$

Consequently

$$\sqrt{I(k)} = |\hat{f}_{n_\ell}(k)| \to |\hat{f}(k)| \ \ \forall k \in \mathbb{Z}^d,$$

and so the limit f belongs to $\mathcal{M}(\sqrt{I})$, establishing weak sequential compactness. □

On a side note, $\mathcal{M}$ is in fact strongly compact on a torus (see Proposition D.2).

3.4 PROJECTONS ONTO THE MODULUS CONSTRAINT SET

The form of a modulus projecton selection $P_\mathcal{M}$ — described in the example below — is well-known in literature. It is difficult to locate the first rigorous proof demonstrating the form of $P_\mathcal{M}$. Such a proof can be found, for example, in [LBL02, Thm.4.2]. For reader's convenience, a proof is presented below. It is only slightly different from the proof of [LBL02, Thm.4.2], as our notation is tailored to single-valued projection selections (instead of projections); also, eq. (4.8) of [LBL02, Thm.4.2] that states

$$\|g - P_{\mathcal{M};\varphi}[g]\|_2 = \sqrt{C_\mathcal{F}}\left\||\hat{g}(k)| - \sqrt{I}(k)\right\|_2$$

is postponed to Example 4.6.

Example 3.24 (Modulus and incomplete modulus projectons). Let $\sqrt{I}, S_F$ be as in Definition 2.1. Let $\varphi\colon \Omega_F \to [0;2\pi)$ be such that $\sin\varphi$ is odd. Define the following operators with domain $\mathcal{D} = \mathcal{H}$:

$$A_{\mathcal{M};\varphi}[g] = \mathcal{F}^{-1}\Big(\sqrt{I}\frac{\hat{g}}{|\hat{g}|}\mathbb{1}_{\{\hat{g}\neq 0\}} + \sqrt{I}e^{i\varphi}\mathbb{1}_{\{\hat{g}=0\}}\Big); \quad \text{(modulus)}$$

$$A_{\mathcal{M}^{(i)};\varphi}[g] = \mathcal{F}^{-1}\Big(\sqrt{I}\frac{\hat{g}}{|\hat{g}|}\mathbb{1}_{\{\hat{g}\neq 0\}\cap S_F} + \sqrt{I}e^{i\varphi}\mathbb{1}_{\{\hat{g}=0\}\cap S_F} + \hat{g}\mathbb{1}_{\Omega_F\setminus S_F}\Big). \quad \text{(incomplete modulus)}$$

Then, $P_{\mathcal{M};\varphi}[g] = A_{\mathcal{M};\varphi}[g]$ is a projecton selection onto $\mathcal{M}(\sqrt{I})$, and $P_{\mathcal{M}^{(i)};\varphi}[g] = A_{\mathcal{M}^{(i)};\varphi}[g]$ is a projecton selection onto $\mathcal{M}^{(i)}(\sqrt{I}; S_f)$. We abbreviate $P_{\mathcal{M}} = P_{\mathcal{M};\varphi}$ and $P_{\mathcal{M}^{(i)}} = P_{\mathcal{M}^{(i)};\varphi}$ whenever it can not cause confusion.

It is convenient to introduce versions of these operators that act on the Fourier space: let $A^F_{\mathcal{M};\varphi}$, $A^F_{\mathcal{M}^{(i)};\varphi}$ map $\widehat{\mathcal{H}}(\Omega_F)$ to itself with

$$P^F_{\mathcal{M};\varphi}[\hat{g}] = \sqrt{I}\frac{\hat{g}}{|\hat{g}|}\mathbb{1}_{\{\hat{g}\neq 0\}} + \sqrt{I}e^{i\varphi}\mathbb{1}_{\{\hat{g}=0\}};$$
$$P^F_{\mathcal{M}^{(i)};\varphi}[\hat{g}] = \sqrt{I}\frac{\hat{g}}{|\hat{g}|}\mathbb{1}_{\{\hat{g}\neq 0\}\cap S_F} + \sqrt{I}e^{i\varphi}\mathbb{1}_{\{\hat{g}=0\}\cap S_F} + \hat{g}\mathbb{1}_{\Omega_F\setminus S_F}.$$

Then, $P^F_{\mathcal{M};\varphi}[g] := A_{\mathcal{M};\varphi}[g]$ is a projecton selection onto $\hat{\mathcal{M}} := \{h \in \mathcal{H} \mid \check{h} \in \mathcal{M}(\sqrt{I})\}$, and mutatis mutandis for $P^F_{\mathcal{M}^{(i)};\varphi} = A^F_{\mathcal{M}^{(i)};\varphi}$.

For these operators we also drop the subscript φ where it can not cause confusion. Whenever the operators $P_{\mathcal{M}}$, $P_{\mathcal{M}^{(i)}}$ and $P^F_{\mathcal{M}}$ are used without the subscript φ, we imply that there exists a corresponding φ such as in this definition.

Proof (that $P_{\mathcal{M}} = A_{\mathcal{M}}$ and $P_{\mathcal{M}^{(i)}} = A_{\mathcal{M}^{(i)}}$). Choose φ_g such that $\widehat{A_{\mathcal{M}}[g]} = \sqrt{I}e^{i\varphi_g}$. Assume there exists $f \in \mathcal{M}$ satisfying

$$\|g - f\|_2 < \|g - A_{\mathcal{M}}[g]\|_2. \tag{3.18}$$

Choose φ_f such that $\hat{f} = \sqrt{I}e^{i\varphi_f}$. By Plancherel's theorem,

$$\begin{aligned}
\|g - A_{\mathcal{M}}[g]\|_2 &= \sqrt{C_{\mathcal{F}}}\left\||\hat{g}|e^{i\varphi_g} - \sqrt{I}e^{i\varphi_g}\right\|_2 = \\
&= \sqrt{C_{\mathcal{F}}}\left\||\hat{g}| - \sqrt{I}\right\|_2 < \\
&\overset{(*)}{<} \sqrt{C_{\mathcal{F}}}\left\||\hat{g}| - \sqrt{I}e^{i(\varphi_g-\varphi_f)}\right\|_2 = \\
&= \sqrt{C_{\mathcal{F}}}\left\||\hat{g}|e^{i\varphi_g} - \sqrt{I}e^{i\varphi_f}\right\|_2 = \|g - f\|_2.
\end{aligned} \tag{3.19}$$

Strict inequality $(*)$ holds, since $f \neq A_{\mathcal{M}}[g]$, and since for any positive numbers a, b and any phase $\varphi_b \in (0, 2\pi)$ one has

$$\begin{aligned}
|a - be^{i\varphi_b}| &= \sqrt{(a - b\cos\varphi_b)^2 + (b\sin\varphi_b)^2} = \\
&= \sqrt{a^2 - 2ab\cos\varphi_b + b^2} > \\
&> \sqrt{a^2 - 2ab + b^2} = |a - b|.
\end{aligned}$$

Equation (3.19) contradicts Equation (3.18); therefore, $A_{\mathcal{M}}$ is a projection selection, i. e. $P_{\mathcal{M}} = A_{\mathcal{M}}$.

The definition of $A_{\mathcal{M}}$ implies that a modulus projection selection is, in general, not unique. However, it nessesarily has the form $P_{\mathcal{M};\varphi}[g]$ for an appropriate φ. Indeed, if one assumes "$\leqslant$" in Equation (3.18), from Equation (3.19) follows that phases of $\hat{f}$ and $\widehat{P_{\mathcal{M};\varphi}[g]}$ can differ

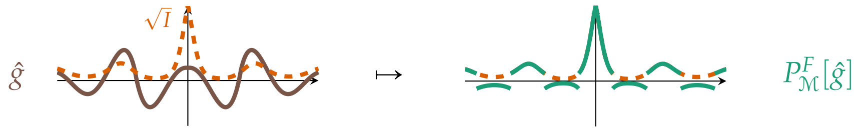

Figure 3.5: Illustration to Example 3.24 (modulus projecton). For illustrative purposes, choose $g \in \mathcal{H}(\mathbb{R})$ even and real-valued, so that $\hat{g}$ (colored brown) is also even and real-valued. The map $P^F_{\mathcal{M};\varphi}$ preserves the phase of $\hat{g}$, but sets the modulus of $\hat{g}$ to $\sqrt{I}$ (dashed orange). At discontinuities — i. e. at the points k where $\sqrt{I}(k) \neq 0$, but $\hat{g}(k) = 0$ — the phase of $P^F_{\mathcal{M};\varphi}[\hat{g}](k)$ is prescribed by $\varphi(k)$.

only at $k \in \Omega_F$ where $\sqrt{I}(k) = 0$. In particular, if $\sqrt{I}$ is supported on the whole of Ω_F, then $P_{\mathcal{M};\varphi}[g]$ is unique and independent of φ.

The proofs for $P_{\mathcal{M}^{(\mathrm{i})};\varphi}$ is the same mutatis mutandis. □

Remark 3.25 (Sine of Fourier phase is odd). The assumption

"let $\varphi\colon \Omega_F \to [0; 2\pi)$ be such that $\sin\varphi$ is odd"

is a convenient way to ensure that functions in range of $P_{\mathcal{M}}$ and $P_{\mathcal{M}^{(\mathrm{i})}}$ remain real-valued, cf. Equation (2.1) or Appendix A.

Remark 3.26 (Multi-valuedness of the modulus projection). The fact that

a modulus projecton selection is not unique
and is specified by a certain phase φ

is a crucial property connected to non-convexity of phase-retrieval. (Indeed, recall that if a projection selection $P_{\mathcal{X}}\colon \mathcal{H} \to \mathcal{H}$ is unique, then the set $\mathcal{X}$ is Chebyshev. In certain cases — for example, if $\mathcal{H}$ is finite-dimensional, — this is equivalent to $\mathcal{X}$ being convex; see references in Remark 3.5).

In computational phase retrieval, explicit calls to φ are extremely rare, since $\hat{f}(k) = 0$ almost never occurs within computer precision (for any approximation f and coordinate k). It is therefore common to select

$$\varphi \equiv 0 \tag{3.20}$$

which is arguably the simplest choice.

A rigorous way to treat $P_{\mathcal{M}}$ is to consider it as a multi-valued operator (see, for example, [LBL02] or [BL03]). However, in this work, we use single-valued selections: this allows us to formulate many results in a more concise and readable form. For example, Chapter 7 demonstrates existence of solutions for an integro-differential equation that contains $P_{\mathcal{M};\varphi_t}$, where the selection φ_t changes with time t. While we conjecture that the corresponding results can be extended to a more general form (using multi-valued operators), such a generalization lies beyond the scope of this work.

Such a generalization would require to consider the set

$$\left\{P_{\mathcal{M};\varphi} \text{ for all measurable } \varphi\colon \Omega_F \to [0;2\pi) \text{ such that } \sin\varphi \text{ is odd}\right\}.$$

in places where we consider the single-valued operator

$$P_{\mathcal{M};\varphi} \text{ for any measurable } \varphi\colon \Omega_F \to [0;2\pi) \text{ such that } \sin\varphi \text{ is odd},$$

The arising differences are of minor importance to this work, since both single-valued and multi-valued formulations can be used to analyze the non-convexity of phase retrieval.

Remark 3.27 (Importance of multiplicity in applications). In this remark, we call f a multiplicity point if $f \in \mathcal{H}$ is such that $\hat{f}(k) = 0$ for at least one (Lebesgue point) $k \in \Omega_F$. At the multiplicity points, the modulus projecton is not uniquely defined and must be selected to equal some $P_{\mathcal{M};\varphi}$.

Note that $P_{\mathcal{M}}$ is not continuous at these multiplicity points. Therefore, behavior of any phase retrieval algorithm becomes difficult to analyze in the neighborhood of multiplicity points.

While orbits of many algorithms (e.g. any common variant of ER and DR) rarely contain multiplicity points, they commonly pass through the neighborhood of multiplicity points. (I.e., for a generated sequence of approximations $(g_n)_n$, it is common to observe $\hat{g}_n(k) \ll \sqrt{I}(k)$ for at least one $n \in \mathbb{N}$, for at least one $k \in \Omega_f$.)

It is therefore important to analyze and improve the behaviour of algorithms near the multiplicity points. Informally speaking, a "fortunate guess" of the phase φ (possibly dependent on f) at a multiplicity f could render phase retrieval trivial, cf. Remark 5.20 and Example 7.23.

Remark 3.28 (Computational efficiency of projectons). It is easy to check that for $\Omega = \mathbb{T}_N^d$, evaluation of positivity, support, and amplitude thresholding projectons requires $O(|N|)$ calculations. The support size projecton requires $O(|N|\nu)$ calculations to get ν largest elements of the argument, or $O\big(|N|\log|N|\big)$ calculations if one wishes to sort all values of the argument. Finally, the modulus and incomplete modulus projectons require $O\big(|N|\log|N|\big)$ for fast Fourier transform and $O(|N|)$ steps in Fourier space.

Thus, all defined projectons require at most $O\big(|N|\log|N|\big)$ calculations. With desired resolutions of order $10^4 - 10^6$ pixels (voxels), projectons – or, in general, reconstruction algorithms – that require $O(|N|^2)$ calculations per update step quickly become unfeasible for crystallographic phase retrieval.

The $O\big(|N|\log|N|\big)$ complexity of the fast Fourier transform is crucial for phase retrieval algorithms. Before the discovery of the fast

Fourier transform one had to devise phase retrieval algorithms operating solely in Fourier space. For such algorithms, object space constraints were reformulated analytically to be applied directly in the Fourier space. An example is the Sayre's equation, developed in [Say52] and more recently discussed in [Els03]. In general, it is noteworthy that X-ray crystallography was one of the applications driving the developement of Fourier transformation algorithms that eventually led to the discovery of the fast Fourier transform [CLW67].

3.5 LOCAL PROJECTION OPERATORS

This section formalizes certain ideas that are implictly well-known in phase retrieval literature, but — to the best of our knowledge — have not been formulated rigorously. Namely, it introduces and establishes the properties — notably, Proposition 3.33 and Corollary 3.35, — underlying the transformation between local and global versions of phase retrieval algorithms in Section 5.1.

The examples in this chapter proving that projections have a certain form are very similar. The purpose of this section is to demonstrate how to prove the form of projection operators for $\{P_{\mathcal{P}}, P_{\mathcal{S}}, P_{\mathcal{T}_a(\alpha)}\}$ and $\mathcal{M}$ with a common proof, that relies on the following locality property.

An operator $P_{\mathcal{X}} \in \{P_{\mathcal{P}}, P_{\mathcal{S}}, P_{\mathcal{T}_a(\alpha)}\}$ is local in the sense that the value $P_{\mathcal{X}}[g](x)$ is determined solely by x and $g(x)$ and does not depend on values of g at other coordinates (pointwise evaluation $g(x)$ is meant for any representant g at almost all $x \in \Omega$). The same is true for operators $P^F_{\mathcal{M}}, P^F_{\mathcal{M}^{(i)}}$ acting in Fourier space. For example, $P^F_{\mathcal{M}}[\hat{g}](k)$ is determined solely by k and $\hat{g}(k)$ and does not depend on values of $\hat{g}$ at other Fourier coordinates.

Thus, instead of considering an operator $P_{\mathcal{X}}$ acting on $\mathcal{H}(\Omega)$, one can consider an appropriate local version $P^{(\mathrm{loc})}_{\mathcal{X}}$ that takes pairs $(x, g(x))$ as arguments and acts in a way such that $P_{\mathcal{X}}[g](x) = P^{(\mathrm{loc})}_{\mathcal{X}}(x, g(x))$ for almost all $x \in \Omega$; this is done in Definition 3.29. Proposition 3.33 demonstrates that if $P^{(\mathrm{loc})}_{\mathcal{X}}$ is a projection selection, then $P_{\mathcal{X}}$ is a projection selection. This argument is used in Corollary 3.35 to show that $P_{\mathcal{P}}$, $P_{\mathcal{S}}$, $P_{\mathcal{T}_a(\alpha)}$, $P_{\mathcal{M}}$, and $P_{\mathcal{M}^{(i)}}$ are projection selections.

For readability, the following definition is split in two very similar cases: locality in object space and locality in Fourier space.

DEFINITION 3.29 (LOCAL OPERATOR AND ITS LOCAL VERSION).

1) *Let $\mathcal{D} \subset \mathcal{H}(\Omega)$. An operator $A\colon \mathcal{D} \to \mathcal{H}(\Omega)$ is called local in object space if there exists a corresponding operator $A^{(\mathrm{loc})}$ from $\Omega \times \mathbb{R}$ to itself such that*

$$A^{(\mathrm{loc})}(x, g(x)) = \big(x, A[g](x)\big)$$

for almost all $x \in \Omega$.

Local operators $A^{(\mathrm{loc})}$ are identified with their equivalence classes (analogous to identification of functions in L^p-spaces). Two operators $A_1^{(\mathrm{loc})}$ and $A_2^{(\mathrm{loc})}$ are called equivalent, if for all $g \in \mathcal{H}$, the equality

$$A_1^{(\mathrm{loc})}[g](x) = A_2^{(\mathrm{loc})}[g](x)$$

holds for almost all $x \in \Omega$. For a local operator A, the corresponding operator $A^{(\mathrm{loc})}$ is unique up to the aforementioned equivalence relationship.

2) *Let $\widehat{\mathcal{D}} \subset \widehat{\mathcal{H}}(\Omega_F)$. An operator $A\colon \widehat{\mathcal{D}} \to \mathcal{H}(\Omega_F)$ is called local in Fourier space if there exists a corresponding operator $A^{(\mathrm{loc})}$ from $\Omega_F \times \mathbb{C}$ to itself such that*

$$A^{(\mathrm{loc})}(k, \hat{g}(k)) = \big(k, A[\hat{g}](k)\big)$$

for almost all $k \in \Omega_F$. Similarly to case 1) above, $A^{(\mathrm{loc})}$ is identified with the corresponding equivalence class.

The operator $A^{(\mathrm{loc})}$ is called the local version of A.

Example 3.30 (Local versions of $P_{\mathcal{P}}$, $P_{\mathcal{S}}$, $P_{\mathcal{T}_a(\alpha)}$, and $P_{\mathcal{M}}$). By their respective definitions, it is clear that for any $(x, a) \in \Omega \times \mathbb{R}$

$$P_{\mathcal{P}}^{(\mathrm{loc})}(x,a) = \begin{cases} (x,a) & \text{if } a \geqslant 0, \\ (x,0) & \text{else;} \end{cases}$$

$$P_{\mathcal{S}(S)}^{(\mathrm{loc})}(x,a) = \begin{cases} (x,a) & \text{if } x \in S, \\ (x,0) & \text{else;} \end{cases}$$

$$P_{\mathcal{T}_a(\alpha)}^{(\mathrm{loc})}(x,a) = \begin{cases} (x,a) & \text{if } a \geqslant \alpha, \\ (x,\alpha) & \text{if } a \in [\frac{\alpha}{2}, \alpha), \\ (x,0) & \text{else} \end{cases}$$

define local versions of $P_{\mathcal{P}}$, $P_{\mathcal{S}}$, and $P_{\mathcal{T}_a(\alpha)}$. Further, for any $(k, a) \in \Omega_F \times \mathbb{C}$

$$P_{\mathcal{M};\varphi}^{F\,(\mathrm{loc})}(k,a) = \begin{cases} \left(k, \sqrt{I}(k)\frac{a}{|a|}\right) & \text{if } a \neq 0, \\ \left(k, \sqrt{I}(k)e^{i\varphi(k)}\right) & \text{else;} \end{cases}$$

$$P_{\mathcal{M}^{(i)};\varphi}^{F\,(\mathrm{loc})}(k,a) = \begin{cases} \left(k, \sqrt{I}(k)\frac{a}{|a|}\right) & \text{if } k \in S_F \text{ and } a \neq 0, \\ \left(k, \sqrt{I}(k)e^{i\varphi(k)}\right) & \text{if } k \in S_F \text{ and } a = 0, \\ (k,a) & \text{if } k \notin S_F. \end{cases}$$

define local versions of $P_{\mathcal{M};\varphi}^{F}$ and $P_{\mathcal{M}^{(i)};\varphi}^{F}$.

Let us show that $P^{(\mathrm{loc})}_{\mathcal{P}}$, $P^{(\mathrm{loc})}_{\mathcal{S}}$, $P^{(\mathrm{loc})}_{\mathcal{T}_a(\alpha)}$ are themselves projection selections onto appropriate subsets of $\Omega \times \mathbb{R}$, and that $P^{F\,(\mathrm{loc})}_{\mathcal{M};\varphi}$, $P^{F\,(\mathrm{loc})}_{\mathcal{M}^{(\mathrm{i})};\varphi}$ are projection selections onto appropriate subsets of $\Omega_F \times \mathbb{C}$.

LEMMA 3.31 (PROJECTIONS: BOX; SINGLETON; TORUS).

1) *Endow $\Omega \times \mathbb{R}$ with the metric*

$$d_{\Omega\times\mathbb{R}}\big((x,a),(\tilde{x},\tilde{a})\big) = \|x-\tilde{x}\|_{dsc.} + |a-\tilde{a}|.$$

Here, $\|x-\tilde{x}\|_{dsc.}$ denotes the disconnected norm

$$\|x-\tilde{x}\|_{dsc.} = \begin{cases} 0 & \textit{if } x=\tilde{x}; \\ \infty & \textit{else.} \end{cases}$$

Let $\Theta = \bigcup_{x\in\Omega}\{(x,L(x))\}$, where $L(x)$ is a closed interval in $\mathbb{R}$. In particular, $L(x)$ can be unbounded or a singleton, but not empty. Then, the projection onto Θ is given by

$$P_\Theta\big((x,a)\big) = \begin{cases} (x,\inf L(x)) & \textit{if } a<\inf L(x); \\ (x,\sup L(x)) & \textit{if } a>\sup L(x); \\ (x,a) & \textit{else.} \end{cases}$$

(We write $\inf L(x)$ instead of $\min L(x)$ to allow the case when $L(x)$ is unbounded from below; analogously for $\sup$.)

2) *Endow $\Omega \times \mathbb{R}$ with the metric from 1). Let $\alpha > 0$, let $\Theta = \Omega \times \big(\{0\} \cup [\alpha,\infty)\big)$. Then, a projection onto Θ is given by*

$$P_\Theta\big((x,b)\big) = \begin{cases} (x,0) & \textit{if } b<\frac{\alpha}{2}; \\ (x,\alpha_*) & \textit{if } b=\frac{\alpha}{2}; \\ (x,\alpha) & \textit{if } b\in(\frac{\alpha}{2},\alpha); \\ (x,b) & \textit{else} \end{cases}$$

with ambiguity $\alpha_* \in \{0,\alpha\}$. For convenience of later usage, choose $\alpha_* = \alpha$ (to remain consistent with the default choice of amplitude thresholding projection in Example 3.14).

3) *Endow $\Omega_F \times \mathbb{C}$ with the metric*

$$d_{\Omega\times\mathbb{R}}\big((k,a),(\tilde{k},\tilde{a})\big) = \|k-\tilde{k}\|_{dsc.} + |a-\tilde{a}|.$$

Let $\sqrt{I} \in \widehat{\mathcal{H}}$ be real-valued and non-negative, let S_F be a measurable subset of Ω_F. Let

$$\Theta = \bigcup_{k\in S_F}\{\tilde{\varphi}\in[0,2\pi) \mid (k,\sqrt{I}(k)e^{i\tilde{\varphi}})\} \cup \bigcup_{k\in\Omega_F\setminus S_F}\{(k,\mathbb{C})\}$$

— at each point k, this set is constructed from circumferences of radius $\sqrt{I}(k)$ or whole complex planes. Then,

$$P_\Theta\big((k,a)\big) = \begin{cases} (k, \sqrt{I}\frac{a}{|a|}) & \text{if } a \neq 0 \text{ and } k \in S_F; \\ (k, \sqrt{I}e^{i\varphi(k)}) & \text{if } a \neq 0 \text{ and } k \notin S_F; \\ (k,a) & \text{else,} \end{cases}$$

for any measurable $\varphi_*\colon \Omega_F \to [0, 2\pi)$. Here, φ_* is the ambiguity in the choice of projection selection.

Proof. 1) Note that $P_\Theta\big((x,a)\big) \in \Theta$, since $L(x)$ is closed for all $x \in \Omega$. Assume there exists $(y,b) \in \Theta$ satisfying

$$d_{\Omega\times\mathbb{R}}\Big((x,a),(y,b)\Big) \leqslant d_{\Omega\times\mathbb{R}}\Big((x,a), P_\theta\big((x,a)\big)\Big).$$

Let us show that from this follows $(y,b) = (x,a)$. Indeed, if $y \neq x$, then $d_{\Omega\times\mathbb{R}}\Big((x,a),(y,b)\Big)$ is infinite, but $d_{\Omega\times\mathbb{R}}\Big((x,a), P_\theta\big((x,a)\big)\Big)$ is finite for any $a \in \mathbb{R}$; hence, $y = x$. Further, $b = a$ is staightforward since $b \in L(x)$:

a) if $a < \inf L(x)$, then b can not be closer to a than $\inf L(x)$;
b) if $a > \sup L(x)$, then b can not be closer to a than $\sup L(x)$;
c) if $a \in L(x)$, then b can not be closer to a than a.

2) The proof is similar to case 1) with a minor difference that if $b = \frac{\alpha}{2}$ for some (x,b), then $(x,0)$ and (x,α) both belong to Θ and are equidistant from (x,b), resulting in the multivaluedness of P_Θ; this multiplicity can be resolved by the choice of any candidate. In our case, the choice (x,α) is more convenient for the following Corollary 3.32.
3) Assume there exists $(q,b) \in \Theta$ satisfying

$$d_{\Omega_F\times\mathbb{C}}\Big((k,a),(q,b)\Big) \leqslant d_{\Omega_F\times\mathbb{C}}\Big((k,a), P_\theta\big((k,a)\big)\Big). \tag{3.21}$$

One has $q = k$ by the same argument as above. If $k \in \Omega\backslash S_F$, then $P_\Theta\big((k,a)\big) = (k,a)$ is obviosly the closest element in Θ to (k,a).

Consider the case $k \in S_F$. Since $(q,b) \in \Theta$, there exists φ_b such that $b = \sqrt{I}(k)e^{i\varphi_b}$. For $a \neq 0$, choose φ_a such that $e^{i\varphi_a} = \frac{a}{|a|}$. Then,

$$\begin{aligned} |a-b| &= \left|a - \sqrt{I}e^{i\varphi_b}\right| = \left||a| - \sqrt{I}e^{i(\varphi_b-\varphi_a)}\right| = \\ &= \sqrt{|a|^2 - 2|a|\sqrt{I}\cos(\varphi_b-\varphi_a) + \sqrt{I}^2} \leqslant \left||a| - \sqrt{I}\right| = \left|a - \sqrt{I}\frac{a}{|a|}\right|. \end{aligned}$$

In conjunction with Equation (3.21) — use definition of $d_{\Omega_F\times\mathbb{C}}$ and $k = q$ in the latter — follows that this inequality must be an equality. The equality holds only if $\cos(\varphi_b - \varphi_a) = 1$, meaning $e^{i\varphi_b} = e^{i\varphi_a}$. Therefore, $(q,b) = P_\Theta\big((k,a)\big)$.

For $a = 0$, any $b = \sqrt{I}e^{i\varphi_b}$ has the same distance to a, and $(q, b) = P_\Theta\big((k, a)\big)$ for an appropriate choice of φ, thus concluding the proof. □

Corollary 3.32. *From previous lemma immediately follows that $P^{(\text{loc})}_{\mathcal{P}}$, $P^{(\text{loc})}_{\mathcal{S}}$, $P^{(\text{loc})}_{\mathcal{T}_a(\alpha)}$, $P^{F\,(\text{loc})}_{\mathcal{M};\varphi}$, $P^{F\,(\text{loc})}_{\mathcal{M}^{(i)};\varphi}$ are projection selections. Namely:*

1) $P^{(\text{loc})}_{\mathcal{P}} = P_\Theta$*, where* $\Theta = \bigcup_{x\in\Omega}\{(x, \mathbb{R}_{\geqslant 0})\} = \Omega \times \mathbb{R}_{\geqslant 0}$.

2) $P^{(\text{loc})}_{\mathcal{S}} = P_\Theta$*, where*

$$\Theta = \bigcup_{x\in S}(x, \mathbb{R}) \cup \bigcup_{x\in\Omega\setminus S}\{(x, 0)\}.$$

3) $P^{(\text{loc})}_{\mathcal{T}_a(\alpha)} = P_\Theta$*, where* $\Theta = \bigcup_{x\in\Omega}\{(x, \{0\} \cup \mathbb{R}_{\geqslant\alpha})\} = \Omega \times \big(\{0\} \cup \mathbb{R}_{\geqslant\alpha}\big)$, *with ambiguity resolution* $\alpha_* := \alpha$.

4) $P^{F\,(\text{loc})}_{\mathcal{M};\varphi} = P_\Theta$*, where* $\Theta = \bigcup_{k\in\Omega_F}\{(k, \sqrt{I}(k))\}$, *with ambiguity resolution* $\varphi_* := \varphi$.

5) $P^{F\,(\text{loc})}_{\mathcal{M}^{(i)};\varphi} = P_\Theta$*, where*

$$\Theta = \bigcup_{k\in S_F}\big\{\big(k, \sqrt{I}(k)\big)\big\} \cup \bigcup_{k\in\Omega_F\setminus S_F}\{(k, \mathbb{C})\},$$

with ambiguity resolution $\varphi_* := \varphi$.

Equality $A^{(\text{loc})} = P_\Theta$ is understood in the following sense: in the equivalence class of $A^{(\text{loc})}$ there exists an operator $A^{(\text{loc})}_*$ such that $A^{(\text{loc})}_*(x, a) = P_\Theta(x, a)$ for all $(x, a) \in \Theta$.

Proposition 3.33 (Criterion for local projection selections). *Let $\mathcal{D} \subset \mathcal{H}(\Omega)$, let $A\colon \mathcal{D} \to \mathcal{H}(\Omega)$ be a local operator with the local version $A^{(\text{loc})}\colon \Omega \times \mathbb{R} \to \Omega \times \mathbb{R}$. Endow $\Omega \times \mathbb{R}$ with the metric*

$$d_{\Omega\times\mathbb{R}}\big((x, a), (y, b)\big) = d_\Omega(x, y) + |a - b|,$$

where d_Ω is any metric on Ω. If $A^{(\text{loc})} = P_\Theta$ is a projection selection onto some set $\Theta \subset \Omega \times \mathbb{R}$, then $A = P_{\mathcal{X}}$ is a projection selection onto

$$\mathcal{X} = \big\{g \in \mathcal{H}(\Omega) \mid \big(x, g(x)\big) \in \Theta \text{ for a. a. } x \in \Omega\big\}.$$

Proof. The proof is straightforward: first, we show that $A[g] \in \mathcal{X}$ for all $g \in \mathcal{D}$; second, we show that $A[g]$ is the closest point in $\mathcal{X}$ to g.

First: by definition of the local operator holds

$$\big(x, A[g](x)\big) = P_\Theta\big(x, g(x)\big)$$

almost everywhere in Ω; therefore, $\big(x, A[g](x)\big) \in \Theta$ for a. a. $x \in \Omega$; therefore, $A[g] \in \mathcal{X}$ by definition of $\mathcal{X}$.

Second: assume that $A[g]$ is not the closest point in $\mathcal{X}$ to g, i.e. assume that there exists $f \in \mathcal{X}$ such that $\|g - f\|_2 < \|g - A[g]\|_2$.

But this would contradict the inequality

$$\begin{aligned}\|g - A[g]\|_2^2 = \int (d_\Omega(x, x) + |g(x) - A[g](x)|)^2 dx = \\ = \int d_{\Omega\times\mathbb{R}}\Big((x, g(x)), (x, A[g](x))\Big)^2 dx = \\ = \int d_{\Omega\times\mathbb{R}}\Big((x, g(x)), P_\Theta(x, g(x))\Big)^2 dx = \\ \overset{(*)}{\leqslant} \int d_{\Omega\times\mathbb{R}}\Big((x, g(x)), (x, f(x))\Big)^2 dx = \\ = \int |g(x) - f(x)|^2 dx = \|g - f\|_2^2.\end{aligned}$$

In step $(*)$ we used a variant of Corollary 3.6 applied on the metric space $(\Omega \times \mathbb{R}, d_{\Omega\times\mathbb{R}})$, cf. Remark 3.7.

This contradiction completes the proof. □

Remark 3.34. Proposition 3.33 holds mutatis mutandis in Fourier space. In that case, $\Omega_F \times \mathbb{C}$ is endowed with the metric

$$d_{\Omega\times\mathbb{R}}\big((k, a), (q, b)\big) = d_{\Omega_F}(k, q) + |a - b|,$$

where d_{Ω_F} is any metric on Ω_F. The proof remains essentially the same.

Corollary 3.35 (Alternative proof for projection selections). *The operators $P_\mathcal{P}$, $P_\mathcal{S}$, $P_{\mathcal{T}_a(\alpha)}$, $P_\mathcal{M}$, and $P_{\mathcal{M}^{(i)}}$ — as explicitly defined in Example 3.14 and Example 3.24 — are projection selections.*

Proof. The operators $P_\mathcal{P}$, $P_\mathcal{S}$, $P_{\mathcal{T}_a(\alpha)}$, $P^F_\mathcal{M}$, and $P^F_{\mathcal{M}^{(i)}}$ are projection selections by Proposition 3.33, since their local versions from Example 3.30 are themselves projection selections by Corollary 3.32. The operators $P_\mathcal{M}$ and $P_{\mathcal{M}^{(i)}}$ are projection selections by Plancherel's theorem, since

$$\|g - P_\mathcal{M}[g]\|_2 = \sqrt{C_\mathcal{F}}\|\hat{g} - P^F_\mathcal{M}[\hat{g}]\|_2 \leqslant \sqrt{C_\mathcal{F}}\|\hat{g} - \hat{f}\|_2 = \|g - f\|_2$$

for any $f, g \in \mathcal{H}$; likewise for $P_{\mathcal{M}^{(i)}}$. □

4 SQUARE DISTANCE ENERGY FUNCTIONALS

This chapter is devoted to the study of the square distance energy functionals, i.e. the functionals of the form

$$E_{\mathcal{X}}[g] = \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2,$$

where $P_{\mathcal{X}}$ is a single-valued projecton selection onto some proximal set $\mathcal{X} \subset \mathcal{H}$.

Section 4.1 reformulates phase retrieval as an energy minimization problem. While similar reformulations are common in phase retrieval literature, we use a very specific choice of the corresponding functional, namely

$$g \mapsto E_{\mathcal{A}}[g] + E_{\mathcal{M}}[g].$$

This choice is symmetric in the modulus and additional constraints, admits a specific choice of the generalized Kruger-Mordukhovich subdifferential ($g - \Pi_{\mathcal{A}}[g] + g - \Pi_{\mathcal{M}}[g]$), and will play a crucial role in Chapter 5, where we explore how this exact functional is connected to the Error-Reduction and Douglas-Rachford phase retrieval algorithms.

Section 4.1 discusses some useful basic properties of energy functionals like lower weaker semicontinuity for weakly closed $\mathcal{X}$ or Lipshitz-continuity for proximal $\mathcal{X}$. Using the direct method in calculus of variations, we show that the energy minimization formulation of phase retrieval always admits solutions provided that $\mathcal{A}$ and $\mathcal{M}$ are weakly closed.

Section 4.2 studies conditions under which $\mathcal{A}$ and $\mathcal{M}$ are Fréchet-differentiable. While colloquially known in phase retrieval literature, the corresponding rigorous proofs are more difficult to encounter. Notably, we show that $E_{\mathcal{M}}$ is Fréchet-differentiable at g, if $\|\sqrt{I}(k)/|\hat{g}|\| < \infty$.

This condition is rather restrictive; one can not expect it to hold in the applied setting.

To address this issue, Section 4.3 is devoted do the rigorous variational analysis of $E_{\mathcal{X}}$ for weakly closed sets $\mathcal{X}$. It demonstrates Clarke, Dini, and generalized (in the sense of Kruger-Mordukhovich [KM80]) subdifferentials of $E_{\mathcal{X}}$ for weakly closed sets $\mathcal{X}$. This analysis is applicable to $E_{\mathcal{M}}$, when $\mathcal{M}$ is considered on bounded domains (and thus weakly closed as is demonstrated in Section 3.3). The analysis is inspired by the results of Burke and Luke [BL03], where the generalized subdifferential of $E_{\mathcal{M}}$ was established on unbounded domains (using the Aumann theorem on integration of set-valued mappings [Aum65, Thms.3, 4]), and it was established to coincide with the (convex) Clarke subdifferential:

$$\overline{\partial} E_{\mathcal{M}}[g] = \partial_{\text{KM}} E_{\mathcal{M}}[g] = \text{weak closure of convex hull of } (g - \Pi_{\mathcal{M}}[g]),$$

where $\overline{\partial}$ denotes the Clarke and ∂_{KM} denotes the generalized subdifferentials.

In contrary, Section 4.3 requires a stronger assumption — weak closedness of $\mathcal{X}$, which holds for $\mathcal{M}$ on bounded domains — to show that

$$\begin{aligned} \overline{\partial} E_{\mathcal{X}}[g] &= \text{weak closure of convex hull of } (g - \Pi_{\mathcal{X}}[g]), \\ \partial_{\text{KM}} E_{\mathcal{X}}[g] &= g - \Pi_{\mathcal{X}}[g]. \end{aligned}$$

The difference in results can be connected to the following difference in the setting. The modulus functional is weakly closed on $\mathcal{M} \subset L^2(\Omega)$ for bounded Ω. For such Ω, $L^2(\Omega)$ is homeomorphic to $\ell^2(\mathbb{Z}^d)$ (through coordinate scaling and Fourier transform, cf. Corollary D.3), and $\mathbb{Z}^d$ (with the counting measure) is an atomic measure space. This diverges from the setting of [BL03] which considers only non-atomic domains.

For purposes of this thesis, the non-convex form $g - \Pi_{\mathcal{X}}[g]$ (that coincides with the formal Fréchet-derivative) — rather than its convexification — is better suited to the purposes of Chapter 5, which establishes the connection between the subgradient flow of $E_{\mathcal{M}} + E_{\mathcal{A}}$ and Error-Reduction and Douglas-Rachford algorithms.

In the very end of this chapter, we also briefly discuss the slope of $E_{\mathcal{M}}$ as yet another way to go beyond Fréchet-differentiation.

4.1 PHASE RETRIEVAL VIA ENERGY MINIMIZATION

This section formalizes phase retrieval as an energy minimization problem, which is a common formulation of phase retrieval. Less common is the particular choice of energy functionals: it is symmetric with respect to the exchange of modulus and additional constraints. Of note are Proposition 4.4 (weak l.s.c. of $E_{\mathcal{X}}$ for weakly closed $\mathcal{X}$) and Remark 4.10 (direct method in calculus of variation is applicable) that justify the energy minimization formulation of phase retrieval using the direct method in calculus of variations.

By the very definition of a projection onto set $\mathcal{X}$ at $g \in \mathcal{H}$,

$$P_{\mathcal{X}}[g] \in \operatorname*{arg\,min}_{f \in \mathcal{X}} \|g - f\|_2^p$$

for any $p \in (0, \infty)$ and any single-valued projection selection $P_{\mathcal{X}}$. Therefore, properties of selections like $P_{\mathcal{X}}$ are naturally connected to the functional

$$g \mapsto \min_{f \in \mathcal{X}} \|g - P_{\mathcal{X}}[g]\|_2^p. \tag{4.1}$$

The choice $p = 1$ is the easiest for geometric interpretation: the resulting functional measures the distance between point g and set $\mathcal{X}$. The choice $p = 2$ yields a more regular functional with better differentiability properties. Such functionals — and their combinations — will be of primary interest for this work.

DEFINITION 4.1 (PROJECTION ENERGY FUNCTIONAL). *Let $\mathcal{X} \subset \mathcal{H}$ be weakly closed, let $P_{\mathcal{X}} \colon \mathcal{H} \to \mathcal{H}$ be a projection. The functional*

$$E_{\mathcal{X}}^{(p)} \colon \mathcal{H} \to \mathbb{R}, \quad g \mapsto \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^p$$

is called the corresponding energy functional with exponent p. When $p = 2$, we use the notation $E_{\mathcal{X}} := E_{\mathcal{X}}^{(2)}$. The functional $E_{\mathcal{X}}$ is also known as the squared distance functional.

Remark 4.2 (Moreau envelopes). In best approximation theory, the above functionals are also known as a particular case of Moreau envelopes, see, e.g., [RW09, Chapter 1.G]. The study of proximal mappings (that generalize projections) and Moreau envelopes (that generalize square distance functionals) lies beyond the scope of this work.

Remark 4.3 (Notation via projections). It is common to use the notation like $\frac{1}{2}d_{\mathcal{X}}^2$ for square distance functionals (see, e.g., [RW09, Chapter 1.G]).

The notation we use is less general (it is specifically tailored to Hilbert spaces), but has the benefit of being explicitely defined using projections. This form highlights the similarity between the square distance functional

$$g \mapsto \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2$$

and its (formal) Fréchet-derivative (or rigorous Mordukhovich-Kruger subdifferential selection)

$$g \mapsto g - P_{\mathcal{X}}[g].$$

This connection is important for the main idea of this thesis, which is to study projection algorithms using evolution equations derived from square distance functionals.

The following proposition is necessary to reformulate phase retrieval as energy minimization problem in Definition 4.9 and Remark 4.10.

Proposition 4.4 (Energy functionals are weakly sequentially lsc). *Let $\mathcal{X} \subset \mathcal{H}$ be nonempty and weakly closed, let $P_{\mathcal{X}} \colon \mathcal{H} \to \mathcal{H}$ be a single-valued projecton selection. The corresponding energy functional is boundedly weakly sequentially lower semicontinuous, i.e. for every bounded sequence $(g_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ that converges weakly to some $g \in \mathcal{H}$ holds*

$$E_{\mathcal{X}}[g] \leqslant \liminf_{n\to\infty} E_{\mathcal{X}}[g_n].$$

Proof (by contradiction). Let $(g_n)_n$ be a bounded sequence in $\mathcal{H}$ that converges weakly to g. Assume that $E_{\mathcal{X}}[g] > \liminf_{n\to\infty} E_{\mathcal{X}}[g_n]$.

The sequence $P_{\mathcal{X}}[g_n]$ is bounded (see the argument immediately before Equation (3.11)).

Since the sequence $(P_{\mathcal{X}}[g_n])_{n\in\mathbb{N}}$ is bounded, by the Banach-Alaogly theorem there exists a weakly convergent subsequence $(P_{\mathcal{X}}[g_{n_m}])_{m\in\mathbb{N}}$ such that $g_{n_m} \rightharpoonup g$, such that $E_{\mathcal{X}}[g] > \liminf_{m\to\infty} E_{\mathcal{X}}[g_{n_m}]$, and $P_{\mathcal{X}}[g_{n_m}] \rightharpoonup f$. Also, $f \in \mathcal{X}$ since $\mathcal{X}$ is weakly closed.

Finally, since the Hilbert space norm is sequentially weakly lsc (e.g. [BC17, Lemma 2.42]), we have

$$\begin{aligned}\frac{1}{2}\|g-f\|_2^2 &\leqslant \liminf_{m\to\infty} \frac{1}{2}\|g_{n_m} - P_{\mathcal{X}}[g_{n_m}]\|_2^2 \\ &= \liminf_{m\to\infty} E_{\mathcal{X}}[g_{n_m}] < E_{\mathcal{X}}[g] = \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2\end{aligned}$$

by assumption. This inequality contradicts Corollary 3.6, since $f \in \mathcal{X}$. □

Lemma 4.5. *Let $\mathcal{X} \subset \mathcal{H}$ be not empty and proximal. Then, $E_{\mathcal{X}}$ is locally Lipshitz-continuous with*

$$|E_{\mathcal{X}}[f] - E_{\mathcal{X}}[g]| \leqslant \left(\frac{1}{2}\|f-g\|_2 + \sqrt{2E_{\mathcal{X}}[g]}\right)\|f-g\|_2$$

for all $f, g \in \mathcal{H}$.

Proof. Let $f, g \in \mathcal{H}$. Without loss of generality, assume that $E_{\mathcal{X}}[f] \geqslant E_{\mathcal{X}}[g]$. By a straightforward calculation, for any $P_{\mathcal{X}}$ have

$$\begin{aligned}2(E_{\mathcal{X}}[f] - E_{\mathcal{X}}[g]) &= \|f - P_{\mathcal{X}}[f]\|_2^2 - \|g - P_{\mathcal{X}}[g]\|_2^2 \overset{(*1)}{\leqslant} \|f - P_{\mathcal{X}}[g]\|_2^2 - \|g - P_{\mathcal{X}}[g]\|_2^2 \\ &= \int (f + g - 2P_{\mathcal{X}}[g])(f-g) \overset{(*2)}{\leqslant} \|f + g - 2P_{\mathcal{X}}[g]\|_2 \|f-g\|_2 \\ &\leqslant (\|f-g\|_2 + 2\|g - P_{\mathcal{X}}[g]\|_2)\|f-g\|_2 = (\|f-g\|_2 + 2\sqrt{2E_{\mathcal{X}}[g]})\|f-g\|_2,\end{aligned}$$

where we use Corollary 3.6 in $(*1)$ and Hölder's inequality in $(*2)$. □

Example 4.6 (Positivity and modulus energy functionals). Let $g \in \mathcal{H}$. By definitions of positivity and modulus projections,

$$E_{\mathcal{P}}[g] = \frac{1}{2}\int_{\{g\geqslant 0\}} g(x)^2 dx,$$

and

$$E_{\mathcal{M}}[g] = \frac{1}{2}\|g - P_{\mathcal{M};\varphi}[g]\|_2^2 = \frac{C_{\mathcal{F}}}{2}\int \left(|\hat{g}(k)| - \sqrt{I}(k)\right)^2 dk. \tag{4.2}$$

This reformulation is well-known and established, e.g., in [LBL02, Cor. 4.3]. The reformulation follows from the Plancherel theorem and explicitely shows that $E_{\mathcal{M}}[g]$ does not depend on φ. In general, by the definition of the functional $E_{\mathcal{X}}$ it is clear that it does not depend on the possible multivaluedness of the projection $\Pi_{\mathcal{X}}$.

The following example is not necessary for the main discussion of this section. Rather, it demonstrates a calculation common for energy functionals; it is very similar to the one used later in Proposition 6.4.

Example 4.7 (Convex energy functionals). Let $\mathcal{C} \subset \mathcal{H}$ be weakly closed and convex. Then, for any $f, g \in \mathcal{H}$ and $t \in [0,1]$ holds

$$\begin{aligned} E_{\mathcal{C}}[tf + (1-t)g] \leqslant{}& tE_{\mathcal{C}}[f] + (1-t)E_{\mathcal{C}}[g] \\ & - (1-t)t\|(f - P_{\mathcal{C}}[f]) - (g - P_{\mathcal{C}}[g])\|_2^2. \end{aligned}$$

In particular, $E_{\mathcal{C}}$ is convex.

Proof. Let $f, g \in \mathcal{H}$, let $t \in [0,1]$. The key component of the proof is the inequality

$$\begin{aligned} E_{\mathcal{C}}[tf + (1-t)g] &= \frac{1}{2}\|tf + (1-t)g - P_{\mathcal{C}}[tf + (1-t)g]\|_2^2 \\ &\leqslant \frac{1}{2}\left\|tf + (1-t)g - \left(tP_{\mathcal{C}}[f] + (1-t)P_{\mathcal{C}}[g]\right)\right\|_2^2, \end{aligned} \tag{4.3}$$

which follows from Corollary 3.6 and the fact that $tP_{\mathcal{C}}[f] + (1-t)P_{\mathcal{C}}[g]$ belongs to $\mathcal{C}$ by convexity of $\mathcal{C}$. After that, the proof follows by a typical calculation (e. g., such calculation can be used to show that the functional $f \mapsto \|f\|_2^2$ is λ-convex). Use the binomial formula on the right-hand side of Equation (4.3) to split the expression into terms with $(g - P_{\mathcal{C}}[g])$ and $(f - P_{\mathcal{C}}[f])$ to obtain

$$t^2\frac{1}{2}\|f - P_{\mathcal{C}}[f]\|_2^2 - t(1-t)\langle f - P_{\mathcal{C}}[f], g - P_{\mathcal{C}}[g]\rangle + (1-t)^2\|g - P_{\mathcal{C}}[g]\|_2^2.$$

Use $t^2 = t - t(1-t)$ in the first term and $(1-t)^2 = (1-t) - t(1-t)$ in the third term to get

$$\begin{aligned} t\frac{1}{2}\|f - P_{\mathcal{C}}[f]\|_2^2 - t(1-t)\frac{1}{2}\|f - P_{\mathcal{C}}[f]\|_2^2 \\ + t(1-t)\langle f - P_{\mathcal{C}}[f], g - P_{\mathcal{C}}[g]\rangle \\ + (1-t)\|g - P_{\mathcal{C}}[g]\|_2^2 - t(1-t)\|g - P_{\mathcal{C}}[g]\|_2^2. \end{aligned}$$

Apply the binomial formula to all terms containing the factor $t(1-t)$ and obtain the desired inequality. □

Remark 4.8 (Variants of energy functionals). One of the main roles of energy functionals is to determine whether an algorithm is near a solution. (This implicitely assumes that the underlying feasibility problem is regular (see Definition 9.6), which is not necessarily the case.) Depending on application and algorithm, one may use other functionals to this end.

For example, [CLS15] uses the modulus functional

$$g \mapsto \int \left(|\hat{g}|^2 - \sqrt{I}^2\right)^2.$$

This functional is more regular than $E_{\mathcal{M}}[g]$, but lacks a direct connection to projection operators and resulting properties (like the energy dissipation shown in Proposition 6.4).

Energy functionals can be used to reformulate phase retrieval as an unconstrained minimization problem.

Definition 4.9 (Energy minimization phase retrieval). *Let $\sqrt{I} \in \widehat{\mathcal{H}}$ be as in Definition 2.1, let $\mathcal{A} \subset \mathcal{H}$ be a proximal nonempty set representing the additional constraint. Define* energy minimization phase retrieval *as the task of finding any*

$$g \in \arg\min_{f \in \mathcal{H}} E_{\mathcal{M}}[f] + E_{\mathcal{A}}[f]. \tag{4.4}$$

Remark 4.10 (Existence of solutions on bounded domains). It is straightforward to check that g is a solution of phase problem in the set intersection sense (Definition 2.3) if and only if $E_{\mathcal{M}}[g] + E_{\mathcal{A}}[g] = 0$.

Thus, Definition 4.9 can be used as a generalization of Definition 2.3. With the feasibility definition 2.3, phase retrieval is not well-defined if $\mathcal{M} \cap \mathcal{X}$ is empty, which makes it very susceptible to measurement errors. Meanwhile, with the energy minimization definition, phase retrieval always admits a solution by the direct method in calculus of variations, if it considered on bounded domains ($\mathcal{H} = L^2(\Omega)$ where Ω is bounded) on which $\mathcal{M}$ is weakly closed (Theorem 3.21).

Indeed, by definition $E_{\mathcal{M}} + E_{\mathcal{A}}$ is bounded from below by zero. Further, use Equation (4.2) to observe that

$$\begin{aligned} E_{\mathcal{M}}[g] = \frac{C_{\mathcal{F}}}{2}\|\hat{g}\|_2^2 - C_{\mathcal{F}} \int |\hat{g}|\sqrt{I} + \frac{C_{\mathcal{F}}}{2}\|\sqrt{I}\|_2^2 \\ \leqslant \frac{\|g\|_2^2}{2} - \|g\|_2 \,\|\check{\sqrt{I}}\|_2 + \frac{\|\check{\sqrt{I}}\|_2^2}{2} \end{aligned}$$

is coercive. Since $\mathcal{A}$ is assumed to be not empty, $E_{\mathcal{A}}[g] < \infty$ for all $g \in \mathcal{H}$. Overall, $E_{\mathcal{M}} + E_{\mathcal{A}}$ is coercive and less than infinity at any $g \in \mathcal{H}$. Therefore, minimizing sequences exist, and any minimizing sequence $(g_n)_{n \in \mathbb{N}}$ is bounded. By the Banach-Alaoglu theorem, $(g_n)_{n \in \mathbb{N}}$ has a weakly convergent subsequence. By the lower semicontinuity of energy functionals (Proposition 4.4 applies since $\mathcal{A}$ and $\mathcal{M}$ are weakly closed), the infimum is attained at any accumulation point of $(g_n)_{n\in\mathbb{N}}$.

In practice, one typically uses $E_{\mathcal{M}}$ to track algorithms if approximation belongs to $\mathcal{A}$, and vice versa (cf. Proposition 5.14 or Remark 5.54). The functional $E_{\mathcal{M}} + E_{\mathcal{A}}$ provides a unified error estimate for such cases.

4.2 MINIMIZERS AND FRÉCHET DERIVATIVES

This section recalls energy minimization properties Lemma 4.11, Remark 4.12 that are well-known in phase retrieval and variational analysis. These properties can be demonstrated in more generic settings (e. g., see [Mor18, Ch. 1.3.6]).

The section is concluded with Lemma 4.14 which demonstrates rigorous Fréchet-differentiability conditions for $E_{\mathcal{M}}$ and $E_{\mathcal{P}}$. While the formal calculation of these derivatives is common in phase retrieval, the rigorous derivation of a sufficient condition under which $E_{\mathcal{M}}$ is Fréchet-differentiable ($\|\sqrt{I}/|\hat{g}|\|_{L^\infty} < \infty$) can be seen as a minor novel contribution of this thesis. A sufficient and necessary condition for $E_{\mathcal{M}}$ to be Fréchet-differentiable (on a Hilbert space with a bounded domain) is developed later in Lemma 4.19.

Projection energy functionals of the form (4.1) share the following important property: for any point $g \notin \mathcal{X}$ there exists a neighborhood such that one can explicitely calculate the minimizer of $E_{\mathcal{X}}$ in this neighborhood.

LEMMA 4.11 (EXPLICIT LOCAL MINIMIZER). *Let $\mathcal{D} \subset \mathcal{H}$, let $P_{\mathcal{X}}\colon \mathcal{D} \to \mathcal{H}$ be a single-valued projection selection onto $\mathcal{X} \subset \mathcal{H}$, let $E_{\mathcal{X}}$ be the corresponding energy functional. Let $g \in \mathcal{D}\backslash\mathcal{X}$; this implies $\|g - P_{\mathcal{X}}[g]\|_2 > 0$. Let $\varepsilon \in (0, 1]$, let $p \in [1, \infty)$. Then,*

$$-\varepsilon(g - P_{\mathcal{X}}[g]) \in \underset{\|h\|_2 \leqslant \varepsilon\|g-P_{\mathcal{X}}[g]\|_2}{\arg\min} E_{\mathcal{X}}^{(p)}[g + h]. \tag{4.5}$$

Proof (by contradiciton). Let $(h_n)_{n\in\mathbb{N}}$ be a minimizing sequence, i. e. let $\|h_n\|_2 \leqslant \varepsilon\|g - P_{\mathcal{X}}[g]\|_2$ be such that

$$E_{\mathcal{X}}^{(p)}[g + h_n] \to \inf_{\|h\|_2 \leqslant \varepsilon\|g-P_{\mathcal{X}}[g]\|_2} E_{\mathcal{X}}^{(p)}[g + h] \quad \text{as } n \to \infty.$$

Assume Equation (4.5) is not true. Then, there exists $n \in \mathbb{N}$ such that

$$E_{\mathcal{X}}^{(p)}[g + h_n] < E_{\mathcal{X}}^{(p)}[g - \varepsilon(g - P_{\mathcal{X}}[g])];$$

this is equivalent to

$$\|g + h_n - P_{\mathcal{X}}[g + h_n]\|_2 < \left\|g - \varepsilon(g - P_{\mathcal{X}}[g]) - P_{\mathcal{X}}\left[g - \varepsilon(g - P_{\mathcal{X}}[g])\right]\right\|_2.$$

Since by Corollary 3.6

$$\begin{aligned}&\left\|g-\varepsilon(g-P_{\mathcal{X}}[g])-P_{\mathcal{X}}\left[g-\varepsilon(g-P_{\mathcal{X}}[g])\right]\right\|_2 \leqslant\\ \leqslant &\left\|g-\varepsilon(g-P_{\mathcal{X}}[g])-\qquad P_{\mathcal{X}}[g]\qquad\right\|_2 = (1-\varepsilon)\left\|g-P_{\mathcal{X}}[g]\right\|_2,\end{aligned}$$

we get

$$\|g+h_n-P_{\mathcal{X}}[g+h_n]\|_2 < (1-\varepsilon)\|g-P_{\mathcal{X}}[g]\|_2.$$

However, this leads to the following contradiction:

$$\begin{aligned}\|g-P_{\mathcal{X}}[g]\|_2 &\overset{(*)}{\leqslant} \|g-P_{\mathcal{X}}[g+h_n]\|_2\\ &\leqslant \|g+h_n-P_{\mathcal{X}}[g+h_n]\|_2+\|h_n\|_2\\ &< (1-\varepsilon)\|g-P_{\mathcal{X}}[g]\|_2+\varepsilon\|g-P_{\mathcal{X}}[g]\|_2 = \|g-P_{\mathcal{X}}[g]\|_2,\end{aligned}$$

where we used Corollary 3.6 in $(*)$. □

Lemma 4.11 indicates that the derivative of $E^{(p)}_{\mathcal{X}}$, if it exists, is parallel to $g-P_{\mathcal{X}}[g]$. This statement is well-known, but its precise formulation varies depending on the mathematical context. In general, $E_{\mathcal{X}}$ is differentiable only in a suitably weak sense (see Remark 4.13 below).

The following remark demonstrates that if for proximal $\mathcal{X}$ the Fréchet derivative of $E_{\mathcal{X}}$ exists at a point g, then it must be equal to $g-P_{\mathcal{X}}[g]$. If $\mathcal{X}$ is known to be not only proximal but also weakly closed, than a stronger statement (Lemma 4.19) is possible.

Remark 4.12 (Fréchet derivatives of energy functionals).
Let $\mathcal{X}\subset\mathcal{H}$ be proximal. Assume that $E_{\mathcal{X}}$ is Fréchet-differentiable at $g\in\mathcal{X}$. Then,

$$\nabla E_{\mathcal{X}}[g] = g-P_{\mathcal{X}}[g].$$

Proof. Treat the cases $g\notin\mathcal{X}$ and $g\in\mathcal{X}$ separately.

(i) Assume that $g\notin\mathcal{X}$, i.e. $\|g-P_{\mathcal{X}}[g]\|_2>0$. Let $f=g-P_{\mathcal{X}}[g]$. Since $E_{\mathcal{X}}$ is assumed to be differentiable at g, there exists an $\varepsilon>0$ such that by Lemma 4.11 holds:

$$E_{\mathcal{X}}\left[g-\varepsilon f\frac{\|\nabla E_{\mathcal{X}}[g]\|_2}{\|f\|_2}\right] \leqslant E_{\mathcal{X}}\left[g-\varepsilon\nabla E_{\mathcal{X}}[g]\right]$$

By Taylor's theorem,

$$E_{\mathcal{X}}[g]-\varepsilon\frac{\|\nabla E_{\mathcal{X}}[g]\|_2}{\|f\|_2}\langle\nabla E_{\mathcal{X}}[g],f\rangle+o(\varepsilon) \leqslant E_{\mathcal{X}}[g]-\varepsilon\|\nabla E_{\mathcal{X}}[g]\|_2^2+o(\varepsilon).$$

Taking limit $\varepsilon\to 0$, obtain

$$\langle\nabla E_{\mathcal{X}}[g],f\rangle \geqslant \|f\|_2\|\nabla E_{\mathcal{X}}[g]\|_2.$$

Since the converse inequality is also true (Cauchy's inequality), one has $\langle \nabla E_{\mathcal{X}}[g], f\rangle = \|f\|_2 \|\nabla E_{\mathcal{X}}[g]\|_2$. Therefore,

$$\begin{aligned}\left\| \frac{\|\nabla E_{\mathcal{X}}[g]\|_2}{\|f\|_2} f - \nabla E_{\mathcal{X}} \right\|_2^2 \\ &= \|\nabla E_{\mathcal{X}}\|_2^2 - 2\frac{\|\nabla E_{\mathcal{X}}\|_2}{\|f\|_2}\langle E_{\mathcal{X}}[g], f\rangle + \|\nabla E_{\mathcal{X}}\|_2^2 \\ &= 2\|\nabla E_{\mathcal{X}}\|_2^2 - 2\|\nabla E_{\mathcal{X}}\|_2^2 = 0,\end{aligned}$$

meaning that

$$\nabla E_{\mathcal{X}} = \frac{\|\nabla E_{\mathcal{X}}[g]\|_2}{\|f\|_2} f. \tag{4.6}$$

Let us now show that $\|\nabla E_{\mathcal{X}}\|_2 = \|f\|_2$. By definition of $E_{\mathcal{X}}$ and f,

$$\begin{aligned} E_{\mathcal{X}}[g - \varepsilon f] &= \frac{1}{2}(1-\varepsilon)^2 \|g - P_{\mathcal{X}}[g]\|_2^2 && \Leftrightarrow \\ E_{\mathcal{X}}[g] - \varepsilon\langle \nabla E_{\mathcal{X}}[g], f\rangle + o(\varepsilon) &= E_{\mathcal{X}}[g] - \|g - P_{\mathcal{X}}[g]\|_2^2 + o(\varepsilon) && \overset{\varepsilon\to 0}{\Longrightarrow} \\ \langle \nabla E_{\mathcal{X}}[g], f\rangle &= \|g - P_{\mathcal{X}}[g]\|_2^2 && \Rightarrow \\ \|\nabla E_{\mathcal{X}}[g]\|_2 &= \|g - P_{\mathcal{X}}[g]\|_2. \end{aligned}$$

In the last equivalence we have used Equation (4.6). Overall, one has $\nabla E_{\mathcal{X}}[g] = g - P_{\mathcal{X}}[g]$.

(ii) Assume that $g \in \mathcal{X}$; hence, $E_{\mathcal{X}}[g] = 0$. Assume that $\nabla E_{\mathcal{X}}[g] \neq 0$; then, there exists an $\varepsilon > 0$ such that

$$E_{\mathcal{X}}[g - \varepsilon \nabla E_{\mathcal{X}}[g]] = \underbrace{E_{\mathcal{X}}[g]}_{=0} - \varepsilon \|\nabla E_{\mathcal{X}}[g]\|_2^2 + o(\varepsilon) < 0,$$

in contradiction to $E_{\mathcal{X}}[f] \geqslant 0$ for all $f \in \mathcal{H}$. Therefore, $\nabla E_{\mathcal{X}}[g] = 0 = g - P_{\mathcal{X}}[g]$. □

Remark 4.13 (Existence of derivatives for phase retrieval). The following Lemma 4.14 specifies under which conditions derivatives of $E_{\mathcal{P}}$ and $E_{\mathcal{M}}$ exist, and gives explicit bounds on the error terms of functional expansions.

If the Fréchet derivative does not exist, one may consider weaker notions of differentiability. In Section 4.3, we calculate the Clarke subdifferential ($\overline{\text{conv}}^*\{g - \Pi_{\mathcal{X}}[g]\}$), the generalized subdifferential ($g - \Pi_{\mathcal{X}}[g]$), and the slope ($\|g - P_{\mathcal{X}}[g]\|_2$) of $E_{\mathcal{X}}$ at any $g \in \mathcal{H}$ for weakly closed g. These notions can be used for more rigorous variational analysis of phase retrieval.

LEMMA 4.14 (FRÉCHET DERIVATIVES OF $E_{\mathcal{M}}$ AND $E_{\mathcal{P}}$).

(i) *(Modulus) Let* $\sqrt{I} \in \widehat{\mathcal{H}}(\Omega)$ *be non-negative. If* $g \in \mathcal{H}(\Omega)$ *is such that* $C_{\sqrt{I}} := \left\|\sqrt{I}/|\hat{g}|\right\|_\infty < \infty$*, then* $E_\mathcal{M}$ *is Fréchet-differentiable at* g*, and for all* $\varepsilon > 0, h \in \mathcal{H}$ *holds*

$$\left|E_\mathcal{M}[g+\varepsilon h] - E_\mathcal{M}[g] - \varepsilon \int \nabla E_\mathcal{M}[g]h\right| \leqslant \frac{\varepsilon^2 \|h\|_2^2}{2}\left(22 + 28C_{\sqrt{I}} + 8C_{\sqrt{I}}^{\ 2}\right).$$

In particular, if $\Omega = \mathbb{T}_N^d$ *and* $|\hat{g}(k)| \neq 0$ *for all* $k \in \operatorname{supp}\sqrt{I}$*, then* $g \mapsto E_\mathcal{M}[g]$ *is differentiable at* g*.*

(ii) *(Positivity) The functional* $g \mapsto E_\mathcal{P}[g]$ *is differentiable for all* $g \in \mathcal{H}$*, and for all* $\varepsilon > 0, h \in \mathcal{H}$ *holds*

$$E_\mathcal{P}[g+\varepsilon h] - E_\mathcal{P}[g] - \varepsilon \int \nabla E_\mathcal{P}[g]h = C[h],$$

where $C\colon \mathcal{H} \to \mathcal{H}$ *satisfies* $0 \leqslant C[h] \leqslant \frac{\varepsilon^2}{2}\|h\|_2^2$*.*

Proof. (i) The proof is straightforward but somewhat lengthy. The Taylor-expansion of $E_\mathcal{M}[g+\varepsilon h]$ in ε is performed in Fourier space (cf. Equation (4.2)) using the following steps.

Step 1) Assume $\sqrt{I}(k) \neq 0$ for a. a. $k \in \Omega_F$. Obtain pointwise second-order Taylor-expansion of the integrand $(|\hat{g} + \varepsilon\hat{h}| + \sqrt{I})^2$.

- 1a) Calculate $\frac{d}{d\varepsilon}$(integrand).
- 1b) Calculate $\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0}$(integrand).
- 1c) Observe that integral over terms from 1b) can be simplified; this is used later in Step 2).
- 1d) Calculate $\frac{d^2}{d\varepsilon^2}$(integrand).
- 1e) State pointwise first order Taylor-expansion with Lagrange remainder; show that it remains valid at points where $\sqrt{I}(k) = 0$.

Step 2) Estimate integrals over individual Taylor terms.

- 2a) Split the integration domain into two sets: the "good" set S_G where fractions appearing in integrals are well-behaved, and the "bad" set S_B where fractions appearing in integrals blow up.
- 2b) Estimate integral of the Taylor exansion from 1e) over the "good set".
- 2c) Estimate integral over the "bad set" directly, without using the Taylor expansion.
- 2d) Combine all the estimates.

Step 3) Generalize to the case where $\sqrt{I}(k) = 0$ on a set of positive measure.

For readability, we omit the argument k in all appearing functions if k remains unchanged throughout the calculation.

Step 1). Assume $\sqrt{I}(k) \neq 0$ for a. a.$k \in \Omega_F$. Then, the pointwise expansion of the integrand is well-defined at almost all $k \in \Omega_F$. Indeed, $\hat{g}(k) \neq 0$ for a. a. $k \in \Omega_F$, since $C_{\sqrt{I}} < \infty$. Further, since $|\hat{h}(k)| < \infty$ for a. a. $k \in \Omega_F$, the following pointwise identities remain valid as long as ε is small enough.

Step 1a): First order pointwise derivative.

$$\frac{d}{d\varepsilon}\frac{1}{2}\left(|\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right)^2 = \left(1-\frac{\sqrt{I}}{|\hat{g}+\varepsilon\hat{h}|}\right)\left(\mathrm{Re}(\hat{g}^*\hat{h})+\varepsilon|\hat{h}|^2\right). \tag{4.7}$$

Step 1b): First order pointwise derivative at $\varepsilon = 0$.

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0}\frac{1}{2}\left(|\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right)^2 = \mathrm{Re}(\hat{g}^*\hat{h}) - \sqrt{I}\,\mathrm{Re}\left(\frac{\hat{g}^*}{|\hat{g}|}\hat{h}\right).$$

Step 1c): Integral simplification for later use. Since g is real-valued, $\hat{g}(-k) = \hat{g}^*(k)$ for a. a. $k \in \Omega_F$, and the same is true for h. Therefore, for any symmetric measurable set $S \subset \Omega_F$ holds

$$\begin{aligned}\int_S \mathrm{Re}(\hat{g}^*\hat{h}) &= \int_S \frac{1}{2}\left(\hat{g}^*(k)\hat{h}(k)+\hat{g}(k)\hat{h}^*(k)\right)dk = \\ &= \int_S \frac{1}{2}\left(\hat{g}^*(k)\hat{h}(k)+\hat{g}(-k)\hat{h}^*(-k)\right)dk = \\ &= \int_S \hat{g}^*(k)\hat{h}(k)\,dk = \int_S \hat{g}^*\hat{h}.\end{aligned} \tag{4.8}$$

Analogously, since $\sqrt{I}(k) = \sqrt{I}(-k)$ and $|\hat{g}(k)| = |\hat{g}(-k)|$,

$$\int_S \sqrt{I}\,\mathrm{Re}\left(\frac{\hat{g}^*}{|\hat{g}|}\hat{h}\right) = \int_S \sqrt{I}\frac{\hat{g}^*}{|\hat{g}|}\hat{h}. \tag{4.9}$$

Step 1e): Second order pointwise derivative. From *Step 1a)* follows

$$\begin{aligned}\frac{d^2}{d\varepsilon^2}\frac{1}{2}&\left(|\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right)^2 = \\ &= \frac{1}{2}\frac{\sqrt{I}}{|\hat{g}+\varepsilon\hat{h}|^3}\big(\underbrace{\mathrm{Re}(\hat{g}^*\hat{h})+\varepsilon|\hat{h}|^2}_{=\mathrm{Re}\left((\hat{g}+\varepsilon\hat{h})^*\hat{h}\right)}\big)^2 + \left(1-\frac{\sqrt{I}}{|\hat{g}+\varepsilon\hat{h}|}\right)|\hat{h}|^2 \\ &= |\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{g}+\varepsilon\hat{h}|}|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{g}+\varepsilon\hat{h}|}\,\mathrm{Im}\left(\frac{(\hat{g}+\varepsilon\hat{h})^*}{|\hat{g}+\varepsilon\hat{h}|}\hat{h}\right)^2.\end{aligned}$$

Step 1d): First order Taylor-expansion with Lagrange remainder. By Taylor's theorem, for a. a. $k \in \Omega_F$ there exists $\varepsilon(k)$ such that

$$\frac{1}{2}\left(|\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right)^2 = \frac{1}{2}(|\hat{g}|-\sqrt{I})^2 + \varepsilon\left(\operatorname{Re}(\hat{g}^*\hat{h}) - \sqrt{I}\operatorname{Re}\left(\frac{\hat{g}^*}{|\hat{g}|}\hat{h}\right)\right)$$
$$+\frac{\varepsilon^2}{2}\underbrace{\left(|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{g}+\tilde{\varepsilon}\hat{h}|}|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{g}+\tilde{\varepsilon}\hat{h}|}\operatorname{Im}\left(\frac{(\hat{g}+\tilde{\varepsilon}\hat{h})^*}{|\hat{g}+\tilde{\varepsilon}\hat{h}|}\hat{h}\right)^2\right)}_{:=T_{\varepsilon,k}},$$

where $\tilde{\varepsilon}(k) \in [0, \varepsilon(k)]$ for a. a. $k \in \Omega_F$.

This expansion is well-defined at k as long as

$$|\hat{g}(k) + \tilde{\varepsilon}(k)\hat{h}(k)| \geqslant a(k) > 0$$

for all $\tilde{\varepsilon}(k) \in [0, \varepsilon(k)]$ for some lower bound $a(k)$. Keeping in mind *Step 3)* that comes below, it is appropriate to notice here that if $\sqrt{I}(k)$ is zero, this Taylor-expansion remains (trivially) valid for any possible values of $\varepsilon, \hat{g}$ and $\hat{h}$ — one merely has to drop all terms where $\sqrt{I}$ appears.

Step 2). Keep using the assumption $\sqrt{I}(k) \neq 0$ for a. a.$k \in \Omega_F$.
Step 2a). In *Step 1)*, $\varepsilon(k)$ depended on k and was assumed to be small enough for expansion to hold. Frow now on, we pick a small positive number $\varepsilon \in \mathbb{R}_{>0}$ — that does not depend on k — and use the expansion from Step 1d) only at those k where it is well-defined. Split the integration domain:

$$\Omega_F = \underbrace{\left\{\varepsilon|\hat{h}| < \frac{|\hat{g}|}{2}\right\}}_{=:S_G} \cup \underbrace{\left\{\varepsilon|\hat{h}| \geqslant \frac{|\hat{g}|}{2}\right\}}_{=:S_B}.$$

The "good" set S_G is the set where

$$|\hat{g}+\tilde{\varepsilon}\hat{h}| \geqslant |\hat{g}| - \tilde{\varepsilon}|\hat{h}| \geqslant |\hat{g}| - \varepsilon|\hat{h}| \geqslant \frac{|\hat{g}|}{2},$$

thus, the Taylor-expansion from *Step 1d)* is well-defined for a. a. $k \in S_G$, and one can estimate the integral over $T_{\varepsilon,k}$ directly.

The "bad" set S_B is the set where $|\hat{g}+\tilde{\varepsilon}\hat{h}|$ can become small, thus the denominator in $T_{\varepsilon,k}$ can grow large. However, in this case one can estimate integrals over all other terms in the Taylor-expansion:

$$\int_{S_B} \frac{1}{2}\left(|\hat{g}+\varepsilon\hat{h}| - \sqrt{I}\right)^2 = O(\varepsilon^2);$$
$$\int_{S_B} \frac{1}{2}\left(|\hat{g}| - \sqrt{I}\right)^2 = O(\varepsilon^2);$$
$$\varepsilon\int_{S_B} \left|\hat{g}^* - \sqrt{I}\frac{\hat{g}^*}{|\hat{g}|}\right| |\hat{h}| = O(\varepsilon^2);$$

these estimates will be established in *Step 2c)* below. Note that the Taylor-expansion itself is not necessarily valid at $k \in S_B$, but — see *Step 2d)* below — we do not use it at such points k.

Step 2b). Exploit the definition of the "good" set to estimate the last term in the Taylor expansion:

$$\begin{aligned}\int_{S_G} &\left(|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{g}+\tilde{\varepsilon}\hat{h}|}|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{g}+\tilde{\varepsilon}\hat{h}|}\operatorname{Im}\left(\frac{(\hat{g}+\tilde{\varepsilon}\hat{h})^*}{|\hat{g}+\tilde{\varepsilon}\hat{h}|}\hat{h}\right)^2\right) \\ &\leqslant \|\hat{h}\|_2^2 + \int_{S_G} \frac{\sqrt{I}}{|\hat{g}+\tilde{\varepsilon}\hat{h}|}|\hat{h}|^2 \\ &\leqslant \|\hat{h}\|_2^2 + \int_{S_G} \frac{2\sqrt{I}}{|\hat{g}|}|\hat{h}|^2 \leqslant \|\hat{h}\|_2^2(1+2C_{\sqrt{I}}). \qquad (4.10)\end{aligned}$$

Step 2c). To estimate the integrals over the "bad set" S_B, observe that at any $k \in S_B$ holds

$$|\hat{g}(k)| < 2\varepsilon|\hat{h}(k)| \quad \text{and} \quad \sqrt{I}(k) < C_{\sqrt{I}}|\hat{g}(k)| < 2\varepsilon C_{\sqrt{I}}|\hat{h}(k)|.$$

Therefore,

$$\begin{aligned}\frac{1}{2}\int_{S_B} \left(|\hat{g}+\varepsilon\hat{h}| - \sqrt{I}\right)^2 &\leqslant \frac{1}{2}\left\|\mathbb{1}_{S_B}|\hat{g}+\varepsilon\hat{h}| - \mathbb{1}_{S_B}\sqrt{I}\right\|_2^2 \\ &\leqslant \frac{1}{2}\left(\left\|\mathbb{1}_{S_B}|\hat{g}|\right\|_2 + \varepsilon\left\|\mathbb{1}_{S_B}\hat{h}|\right\|_2 + \left\|\mathbb{1}_{S_B}\sqrt{I}\right\|_2\right)^2 \\ &\leqslant \frac{1}{2}\left(2\varepsilon\left\|\mathbb{1}_{S_B}|\hat{h}|\right\|_2 + \varepsilon\left\|\mathbb{1}_{S_B}\hat{h}|\right\|_2 + 2\varepsilon C_{\sqrt{I}}\left\|\mathbb{1}_{S_B}\hat{h}\right\|_2\right)^2 \\ &\leqslant \frac{\varepsilon^2}{2}\|\mathbb{1}_{S_B}\hat{h}\|_2^2(3+2C_{\sqrt{I}})^2. \qquad (4.11)\end{aligned}$$

Analogously,

$$\frac{1}{2}\int_{S_B} (|\hat{g}| - \sqrt{I})^2 \leqslant \frac{\varepsilon^2}{2}\|\mathbb{1}_{S_B}\hat{h}\|_2^2(2+2C_{\sqrt{I}})^2. \qquad (4.12)$$

$$\varepsilon\int_{S_B}\left|\hat{g}^* - \sqrt{I}\frac{\hat{g}^*}{|\hat{g}|}\right||\hat{h}| \leqslant \frac{\varepsilon^2}{2}\|\mathbb{1}_{S_B}\hat{h}\|_2^2(4+4C_{\sqrt{I}}). \qquad (4.13)$$

Step 2d). Gathering all these estimates together, obtain

$$\begin{aligned}
&\frac{1}{2}\left\||\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right\|_2^2 = \frac{1}{2}\int_{S_G}\left(|\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right)^2 + \frac{1}{2}\int_{S_B}\left(|\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right)^2 = \\
&\quad\leqslant \frac{1}{2}\int_{S_G}(|\hat{g}|-\sqrt{I})^2 \\
&\quad + \varepsilon\int_{S_G}\left(\operatorname{Re}(\hat{g}^*\hat{h}) - \sqrt{I}\operatorname{Re}\left(\frac{\hat{g}^*}{|\hat{g}|}\hat{h}\right)\right) \\
&\quad + \frac{\varepsilon^2}{2}\int_{S_G}\left(|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{g}+\tilde{\varepsilon}\hat{h}|}|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{g}+\tilde{\varepsilon}\hat{h}|}\operatorname{Im}\left(\frac{(\hat{g}+\tilde{\varepsilon}\hat{h})^*}{|\hat{g}+\tilde{\varepsilon}\hat{h}|}\hat{h}\right)^2\right) \\
&\qquad\qquad + \frac{1}{2}\int_{S_B}\left(|\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right)^2 = \\
&\quad = \underbrace{\frac{1}{2}\int_{\Omega_F}(|\hat{g}|-\sqrt{I})^2}_{=(2\pi)^d E_{\mathcal{M}}[g]} - \underbrace{\frac{1}{2}\int_{S_B}(|\hat{g}|-\sqrt{I})^2}_{\text{Use (4.12)}} \\
&\quad + \underbrace{\varepsilon\int_{\Omega_F}\left(\operatorname{Re}(\hat{g}^*\hat{h}) - \sqrt{I}\operatorname{Re}\left(\frac{\hat{g}^*}{|\hat{g}|}\hat{h}\right)\right)}_{=(2\pi)^d\int\nabla E_{\mathcal{M}}[g]h;\ \text{cf. (4.8), (4.9)}} - \underbrace{\varepsilon\int_{S_B}\left(\operatorname{Re}(\hat{g}^*\hat{h}) - \sqrt{I}\operatorname{Re}\left(\frac{\hat{g}^*}{|\hat{g}|}\hat{h}\right)\right)}_{\text{Use (4.13)}} \\
&\quad + \underbrace{\frac{\varepsilon^2}{2}\int_{S_G}\left(\frac{\|\hat{h}\|^2}{2} - \frac{\sqrt{I}}{2|\hat{g}+\tilde{\varepsilon}\hat{h}|}\operatorname{Im}\left(\frac{\hat{g}^*}{|\hat{g}+\tilde{\varepsilon}\hat{h}|}\hat{h}\right)^2\right)}_{\text{Use (4.10)}} \\
&\qquad\qquad + \underbrace{\frac{1}{2}\int_{S_B}\left(|\hat{g}+\varepsilon\hat{h}|-\sqrt{I}\right)^2}_{\text{Use (4.11)}};
\end{aligned}$$

from which follows

$$\left|E_{\mathcal{M}}[g+\varepsilon h] - E_{\mathcal{M}}[g] - \varepsilon\int\nabla E_{\mathcal{M}}[g]h\right| \leqslant \frac{\varepsilon^2\|h\|_2^2}{2}\left(21 + 28C_{\sqrt{I}} + 8{C_{\sqrt{I}}}^2\right). \quad (4.14)$$

Step 3). To lift the assumption $\operatorname{supp}\sqrt{I} = \Omega_F$, split Ω_F into $S_1 = \operatorname{supp}\sqrt{I}$ and $S_2 = \Omega_F \backslash S_1$. Steps 1),2) apply to integration over S_1 with the same estimates. For $k \in S_2$, the second term in the Taylor-expansion is reduced to $\frac{\varepsilon^2\|h\|_2^2}{2}$; adding this term to the right-hand side of the estimation yields the desired result.

(ii) Positivity estimate holds, since at the points where the operator $g + \varepsilon h - P_{\mathcal{P}}[g+\varepsilon h]$ is ill-behaved, i. e. at points where the sign of

$g + \varepsilon h$ differs from the sign of g, $|g + \varepsilon h|$ can be bounded by εh. Indeed, by a straightforward calculation,

$$\begin{aligned} E_{\mathcal{P}}[g + \varepsilon h] &= \frac{1}{2} \int \mathbb{1}_{\{g+\varepsilon h<0\}} (g + \varepsilon h)^2 \\ &= \frac{1}{2} \int \mathbb{1}_{\{g<0\}} (g + \varepsilon h)^2 + \mathbb{1}_{\{g\geqslant 0\}\cap\{g+\varepsilon h<0\}} (g + \varepsilon h)^2 \\ &\qquad\qquad - \mathbb{1}_{\{g<0\}\cap\{g+\varepsilon h\geqslant 0\}} (g + \varepsilon h)^2 \\ &= \frac{1}{2} \int \mathbb{1}_{\{g<0\}} g^2 + 2\varepsilon \mathbb{1}_{\{g<0\}} gh + \mathbb{1}_{\{g<0\}} \varepsilon^2 h^2 \\ &\qquad\qquad + \mathbb{1}_{\{g\geqslant 0\}\cap\{g+\varepsilon h<0\}} (g + \varepsilon h)^2 \\ &\qquad\qquad - \mathbb{1}_{\{g<0\}\cap\{g+\varepsilon h\geqslant 0\}} (g + \varepsilon h)^2. \end{aligned}$$

The claim then follows with

$$\begin{aligned} C[h] = \frac{1}{2} \int \mathbb{1}_{\{g<0\}} \varepsilon^2 h^2 + \mathbb{1}_{\{g\geqslant 0\}\cap\{g+\varepsilon h<0\}} \underbrace{(g + \varepsilon h)^2}_{<\varepsilon^2 h^2} \\ - \mathbb{1}_{\{g<0\}\cap\{g+\varepsilon h\geqslant 0\}} \underbrace{(g + \varepsilon h)^2}_{\leqslant \varepsilon^2 h^2}. \end{aligned}$$

Combine the first and the third summands to estimate $C[h]$ from below:

$$\begin{aligned} C[h] \geqslant \frac{1}{2} \int \mathbb{1}_{\{g<0\}\cap\{g+\varepsilon h<0\}} \varepsilon^2 h^2 + \mathbb{1}_{\{g\geqslant 0\}\cap\{g+\varepsilon h<0\}} \varepsilon^2 h^2 \\ = \frac{1}{2} \int \mathbb{1}_{\{g+\varepsilon h<0\}} \varepsilon^2 h^2. \qquad \square \end{aligned}$$

Remark 4.15. While the differentiability condition on $E_{\mathcal{M}}$ is somewhat restrictive, it holds for certain points of interest (notably, certain points corresponding to fixed points of Error-Reduction algorithm, see Corollary 8.2.) On the fundamental level, the points at which $E_{\mathcal{M}}$ is non-differentiable play a crucial role for phase retrieval and can not be dismissed, cf. Remark 5.20.

Since support projector $P_{\mathcal{S}}$ is linear, one can easily show that $E_{\mathcal{S}}$ is differentiable on the whole $\mathcal{H}$.

One can show that the amplitude thresholding energy $E_{\mathcal{T}_a(\alpha)}$ will be differentiable if $\inf_{x\in\Omega} |g(x) - \frac{\alpha}{2}| > 0$. This condition is rather restrictive, especially if one is working on the space $\Omega = \mathbb{R}^d$ and assumes reasonable decay of $\sqrt{I}$. Indeed, decay of $\sqrt{I}$ implies smoothness of $P_{\mathcal{M}}[g]$ for any $g \in \mathcal{H}$, and this in turn implies that $\inf_{x\in\Omega} |P_{\mathcal{M}}[g](x) - \frac{\alpha}{2}| = 0$ as long as $P_{\mathcal{M}}[g](x) \geqslant \frac{\alpha}{2}$ at at least one Lebesgue point $x \in \mathbb{R}^d$. (And $P_{\mathcal{M}}[g](x) \geqslant \alpha$ will be true for at least one such x in a vicinity of a solution that is not identically equal to 0.)

The functional $E_{\mathcal{T}_s(\nu)}$ exhibits similar differentiability issues.

To work with such functionals, one must regularize them, or explore tools beyond Fréchet-differentiability.

4.3 GENERALIZED DIFFERENTIATION FOR WEAKLY CLOSED SETS

The goal of this section is to establish certain subdifferentials and subgradients of square distance functionals. Main result is that for a weakly closed set $\mathcal{X}$, the generalized subgradient $\partial_{\mathrm{KM}} E_{\mathcal{X}}[g] = g - \Pi_{\mathcal{X}}[g]$ for all $g \in \mathcal{H}$. Thus, for bounded $\Omega \subset \mathbb{R}^d$ with $\mathcal{M} \subset L^2(\Omega)$ holds $\partial_{\mathrm{KM}} E_{\mathcal{M}}[g] \ni g - P_{\mathcal{M};\varphi}[g]$ for any measurable phase $\varphi\colon \mathbb{R}^d \to [0; 2\pi)$ where $\sin\varphi$ is odd. This holds even if $E_{\mathcal{M}}$ is not Fréchet-differentiable at g.

Typically, results of generalized differentiation involve weak-star-closure (e.g. of Fréchet normals), see [MS96]. On separable Hilbert spaces, the weak closure $\overline{\cdot}^*$ is equivalent to the weak-star-closure ([BL03, Sec. 3]). Thus, it is sufficient to work with the weak closure for our purposes.

Another benefit of the separable Hilbert space phase retrieval is that many of the possible definitions for the subdifferential coincide [MS96, Thm.9.2].

The results of this section are inspired by [BL03]. Specifically, [BL03, Prop.3.19] establishes subdifferential regularity of the modulus projection on an unbounded domain (demonstrating that the Clarke and generalized subdifferentials of $E_{\mathcal{M}}$ coincide in that case). Further, [BL03, Thm.3.1] demonstrates that

$$\partial_{\mathrm{KM}} E_{\mathcal{M}}[g] = \overline{\mathrm{conv}}^* \left\{ g - \Pi_{\mathcal{M}}[g] \right\},$$

where $\mathcal{M} \subset L^2(\mathbb{R}^2; \mathbb{R}^2)$ (which is not weakly closed), and convexification comes from [BL03, Prop.3.19].

Our results are different from [BL03] in the following aspects:

- Our results apply only to weakly closed sets $\mathcal{X}$ such as $\mathcal{M} \subset L^2(\Omega)$ but not $\mathcal{M} \subset L^2(\mathbb{R}^d)$. They do not rely on the specific form of the modulus projection but apply to any weakly closed sets $\mathcal{X}$.

- For bounded Ω, $L^2(\Omega)$ is homeomorphic to $\ell^2(\mathbb{Z}^d)$ (through coordinate scaling and Fourier transform, cf. Corollary D.3). Since $\mathbb{Z}^d$ (with the counting measure) is an atomic measure space, which makes our setting different from the setting of [BL03] which considers only non-atomic domains. This is one of the underlying reasons in the difference of the obtained results (the subdifferential of $E_{\mathcal{M}}$ is not convex in our case).

- Our results are developed using a different approach. Namely, [BL03] uses theory of integrals of multi-valued functions to establish conditions that allow the interchange of subdifferentiation and integration [BL03, Lem.3.18, Lem.4.1]. We calculate the subdifferential of $E_{\mathcal{X}}$ from elementary definitions using properties of the Hilbert norm $\|\cdot\|_2$. This approach does not require any use of integrals of multi-functions.

A discussion of distance functional derivatives can also be found in [Mor18, Ch. 1.3.6], where the derivative of $g \mapsto \|g - P_{\mathcal{X}}[g]\|_2$ is calculated on the basis of sequential limits of Fréchet-ε-normals and subdifferentials.

For our purposes, it is possible to describe relevant aspects of generalized differentiation in a relatively brief self-contained manner, without delving deep into the vast subject of generalized differentiation. Specifically, it is possible to establish relevant subdifferentials from elementary definitions, using the mild setting of separable Hilbert spaces as well as properties of projection operators.

4.3.1 *Definitions and main result*

Recall the standard notions and results on the Clarke subdifferential and generalized subdifferential for locally Lipschitz functions ([RW09; MS96; BL03]).

Definition 4.16 (Regular subderivative and Clarke subdifferential). *Let $\mathcal{H}$ be a separable Hilbert space, let $F\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be locally Lipschitz continuous at some $g \in \mathcal{H}$ such that $F[g]$ is finite.*

i) The regular subderivative function $\widehat{d}F[g]\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\widehat{d}F[g][w] := \limsup_{f \to g, \varepsilon \searrow 0} \frac{F[f + \varepsilon w] - F[f]}{\varepsilon}.$$

ii) $v \in \mathcal{H}$ is called a Clarke subgradient of F at g, if F is l.s.c. on a neighborhood of g and v satisfies

$$\langle v, w \rangle \leqslant \widehat{d}F[g][w] \text{ for all } w \in \mathcal{H}.$$

iii) The set of Clarke subgradients of F at g is called the Clarke subdifferential and is denoted by $\overline{\partial} F[g]$.

Definition 4.17 (Subdifferential). *Let $\mathcal{H}$ be a separable Hilbert space, let $F\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be locally Lipschitz continuous at some $g \in \mathcal{H}$ such that $F[g]$ is finite. Let $v \in \mathcal{H}$.*

i) v is a Dini ε-subgradient of F at g, if

$$\liminf_{t \searrow 0} \frac{F[g + tw] - F[g]}{t} - \langle v, w \rangle \geqslant -\varepsilon \|w\|_2 \text{ for all } w \in \mathcal{H}.$$

The set of Dini ε-subgradients v is called the Dini-ε-subdifferential of F at g and is denoted by $\partial_{\varepsilon}^{-} F[g]$.

ii) v is a subgradient of F at g if there are sequences $\varepsilon_n \searrow 0$, $g_n \to g$ and $v_n \in \partial_{\varepsilon_n}^{-} F[g]$ with $F[g_n] \to F[g]$ and $v_n \rightharpoonup v$. We call the set of subgradients v the subdifferential of F at g and denote this set by $\partial_{KM} F[g]$.

THEOREM 4.18 (RESULTS ON GENERALIZED DIFFERENTIATION). *Let $\mathcal{X} \subset \mathcal{H}$ be weakly closed. Then:*

$$\overline{\partial} E_{\mathcal{X}}[g] = \overline{\operatorname{conv} g - \Pi_{\mathcal{X}}[g]}^{*} \quad \text{(Clarke subdifferential), and} \tag{4.15}$$

$$\partial_{KM} E_{\mathcal{X}}[g] = g - \Pi_{\mathcal{X}}[g] \quad \text{(generalized subdifferential).} \tag{4.16}$$

These claims are proven in Theorem 4.22 and Theorem 4.24, respectively.

The following lemma is particularly useful in conjunction with Proposition 3.11 that states necessary and sufficient conditions for $P_{\mathcal{X}}$ to be locally continuous.

LEMMA 4.19. *Let $\mathcal{X} \subset \mathcal{H}$ be non-empty, proximal (but not necessarily weakly closed). Let $P_{\mathcal{X}}$ be any single-valued projection onto $\mathcal{X}$. Then, for any $g, h \in \mathcal{H}$,*

$$E_{\mathcal{X}}[g+h] - E_{\mathcal{X}}[g] \geqslant \langle g - P_{\mathcal{X}}[g+h], h \rangle + \frac{1}{2}\|h\|_2^2; \tag{4.17}$$

$$E_{\mathcal{X}}[g+h] - E_{\mathcal{X}}[g] \leqslant \langle g - \quad P_{\mathcal{X}}[g] \quad , h \rangle + \frac{1}{2}\|h\|_2^2. \tag{4.18}$$

In particular, if $P_{\mathcal{X}}$ is continuous at g, then $E_{\mathcal{X}}$ is Fréchet-differentiable at g with

$$\nabla E_{\mathcal{X}}[g] = g - P_{\mathcal{X}}[g].$$

Proof. "$\geqslant$". To establish (4.19), use the distance-minimizing projection property (3.6) to estimate $-E_{\mathcal{X}}[g]$ from below:

$$\begin{aligned} &\frac{1}{2}\|g+h-P_{\mathcal{X}}[g+h]\|_2^2 - \frac{1}{2}\|g - P_{\mathcal{X}}[\quad g \quad]\|_2^2 \\ \geqslant &\frac{1}{2}\|g+h-P_{\mathcal{X}}[g+h]\|_2^2 - \frac{1}{2}\|g - P_{\mathcal{X}}[g+h]\|_2^2. \end{aligned}$$

Open the squares and simplify the right-hand side:

$$\begin{aligned} \frac{1}{2}\|g+h\|_2^2 - \langle g+h, P_{\mathcal{X}}[g+h]\rangle + \frac{1}{2}\|P_{\mathcal{X}}[g+h]\|_2^2 \\ -\frac{1}{2}\|g\|_2^2 + \langle g, P_{\mathcal{X}}[g+h]\rangle - \frac{1}{2}\|P_{\mathcal{X}}[g+h]\|_2^2 \\ = \frac{1}{2}\|g+h\|_2^2 - \frac{1}{2}\|g\|_2^2 + \langle -P_{\mathcal{X}}[g+h], h\rangle \\ = \langle g - P_{\mathcal{X}}[g+h], h\rangle + \frac{1}{2}\|h\|_2^2. \end{aligned} \tag{4.19}$$

"$\leqslant$" Again, use the distance-minimizing projection property Corollary 3.6, now to estimate $E_{\mathcal{X}}[g+h]$ from above:

$$\begin{aligned} &\frac{1}{2}\|g+h-P_{\mathcal{X}}[g+h]\|_2^2 - \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2 \\ \leqslant &\frac{1}{2}\|g+h-P_{\mathcal{X}}[\quad g \quad]\|_2^2 - \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2. \end{aligned}$$

Similarly to (4.19), open the squares and simplify the right-hand side to the form

$$\langle g - P_{\mathcal{X}}[g], h\rangle + \frac{1}{2}\|h\|_2^2,$$

Finally, if $P_{\mathcal{X}}$ is continuous at g, then the Fréchet-differentiability of $E_{\mathcal{X}}$ at g follows by definition from (4.17) and (4.18). □

4.3.2 *Clarke subdifferential of energy functionals*

This section gives an explicit formula for the Clarke subdifferential of $E_{\mathcal{X}}$ for weakly closed $\mathcal{X}$.

PROPOSITION 4.20. *Let $\mathcal{X} \subset \mathcal{H}$ be non-empty and weakly closed. Then, for all $g_*, w \in \mathcal{H}$,*

$$\widehat{d}E_{\mathcal{X}}[g_*][w] = \sup_{p \in \Pi_{\mathcal{X}}[g_*]} \langle g_* - p, w\rangle.$$

Proof. By definition, for any $w \in \mathcal{H}$ have

$$\widehat{d}E_{\mathcal{X}}[g_*][w] = \limsup_{g \to g_*, \varepsilon \searrow 0} \frac{E_{\mathcal{X}}[g + \varepsilon w] - E_{\mathcal{X}}[g]}{\varepsilon}. \tag{4.20}$$

Let $P_{\mathcal{X}}$ be a single-valued selection of the (in general, multi-valued) operator $\Pi_{\mathcal{X}}$.

"$\geqslant$". By Lemma 4.19,

$$\frac{1}{2}\|g + \varepsilon w - P_{\mathcal{X}}[g + \varepsilon w]\|_2^2 - \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2 \geqslant \varepsilon\langle g - P_{\mathcal{X}}[g + \varepsilon w], w\rangle + \frac{\varepsilon^2}{2}\|w\|_2^2. \tag{4.21}$$

Inserting this into the definition of the subdifferential, have

$$\widehat{d}E_{\mathcal{X}}[g_*][w] \geqslant \limsup_{g \to g_*, \varepsilon \searrow 0} \left(\langle g - P_{\mathcal{X}}[g + \varepsilon w], w\rangle + \frac{\varepsilon}{2}\|w\|_2^2\right). \tag{4.22}$$

To estimate the limes superior in (4.22) from below, consider the particular sequence $(g_n)_n$ defined by $g_n = g_* - \varepsilon_n w$ with $\varepsilon_n = \frac{1}{n}$ for all $n \in \mathbb{N}$:

$$\widehat{d}E_{\mathcal{X}}[g_*][w] \geqslant \lim_{n\to\infty}\Big(\langle \underbrace{g_n}_{\to g_* \text{ as } n\to\infty} - P_{\mathcal{X}}[\underbrace{g_n + \varepsilon_n w}_{=g_*}], w\rangle + \underbrace{\frac{\varepsilon_n}{2}\|w\|_2^2}_{\to 0 \text{ as } n\to\infty}\Big) = \langle g_* - P_{\mathcal{X}}[g_*], w\rangle.$$

This inequality holds for any and all selections $P_{\mathcal{X}}$ of $\Pi_{\mathcal{X}}$, i.e.

$$\widehat{d}E_{\mathcal{X}}[g_*][w] \geqslant \langle g_* - p, w\rangle$$

for all $p \in \Pi_{\mathcal{X}}[g]$, from which follows the desired

$$\widehat{d}E_{\mathcal{X}}[g_*][w] \geqslant \sup_{p \in \Pi_{\mathcal{X}}[g]} \langle g_* - p, w \rangle.$$

"$\leqslant$" Again, by Lemma 4.19,

$$\frac{1}{2}\|g + \varepsilon w - P_{\mathcal{X}}[g + \varepsilon w]\|_2^2 - \frac{1}{2}\|g - P_{\mathcal{X}}[g]\|_2^2 \leqslant \varepsilon \langle g - P_{\mathcal{X}}[g], w \rangle + \frac{\varepsilon^2}{2}\|w\|_2^2,$$

which implies

$$\widehat{d}E_{\mathcal{X}}[g_*][w] \leqslant \limsup_{g \to g_*} \langle g - P_{\mathcal{X}}[g], w \rangle \text{ for all } w \in \mathcal{H} \tag{4.23}$$

(as the term $\frac{\varepsilon^2}{2}\|w\|_2^2$ can be separated by subadditivity of lim sup) and vanishes.

Thus, it remains to show that

$$\limsup_{g \to g_*} \langle g - P_{\mathcal{X}}[g], w \rangle \leqslant \sup_{p \in \Pi_{\mathcal{X}}[g]} \langle g - p, w \rangle \text{ for all } w \in \mathcal{H}.$$

Let $(g_n)_{n \in \mathbb{N}}$ be the sequence that attains the supremum on the left-hand side. Since $\mathcal{X}$ is weakly closed and since $(P_{\mathcal{X}}[g_n])_{n \in \mathbb{N}}$ is bounded (similarly to the argument before Equation (3.11)), the sequence $(P_{\mathcal{X}}[g_n])_{n \in \mathbb{N}}$ has a weakly convergent subsequence, again denoted by $(P_{\mathcal{X}}[g_n])_{n \in \mathbb{N}}$, such that

$$P_{\mathcal{X}}[g_n] \rightharpoonup q \in \mathcal{X} \text{ as } n \to \infty.$$

For this subsequence,

$$\limsup_{g \to g_*} \langle g - P_{\mathcal{X}}[g], w \rangle = \lim_{n \to \infty} \langle g_n - P_{\mathcal{X}}[g_n], w \rangle = \langle g_* - q, w \rangle \text{ for all } w \in \mathcal{H}.$$

Further, using the same subsequence,

$$\begin{aligned} \|g_* - P_{\mathcal{X}}[g_*]\|_2 &\overset{(*1)}{\leqslant} \|g_* - q\|_2 \overset{(*2)}{\leqslant} \liminf_{n \to \infty} \|g_n - P_{\mathcal{X}}[g_n]\|_2 \\ &\overset{(*3)}{\leqslant} \liminf_{n \to \infty} \|g_n - P_{\mathcal{X}}[g_*]\|_2 = \|g_* - P_{\mathcal{X}}[g_*]\|_2, \end{aligned}$$

where $(*1)$ and $(*3)$ hold by projection property (3.6), and $(*2)$ holds since $E_{\mathcal{X}}$ is weakly sequentially l.s.c. (Proposition 4.4). Thus, all inequalities are, in fact, equalities, implying that $q \in \Pi_{\mathcal{X}}[g_*]$, meaning that

$$\limsup_{g \to g_*} \langle g - P_{\mathcal{X}}[g], w \rangle = \langle g_* - q, w \rangle \leqslant \sup_{p \in P_{\mathcal{X}}[g_*]} \langle g_* - p, w \rangle \text{ for all } w \in \mathcal{H}.$$

and for any selection $P_{\mathcal{X}} \in \Pi_{\mathcal{X}}$. Combining with (4.23), have the desired

$$\widehat{d}E_{\mathcal{X}}[g_*][w] \leqslant \sup_{p \in P_{\mathcal{X}}[g_*]} \langle g_* - p, w \rangle \text{ for all } w \in \mathcal{H}.$$

□

To calculate the Clarke subdifferential of $E_{\mathcal{X}}$, we use the following technical

LEMMA 4.21. *Let $\mathcal{Y} \subset \mathcal{H}$. Then,*

$$\sup_{u \in \overline{\operatorname{conv} \mathcal{Y}}^*} \langle u, w \rangle = \sup_{v \in \mathcal{Y}} \langle v, w \rangle \textit{ for all } w \in \mathcal{H}.$$

Proof. The direction $\geqslant$ is trivially true; let us show $\leqslant$.

1) First, show that

$$\sup_{u \in \operatorname{conv} \mathcal{Y}} \langle u, w \rangle \leqslant \sup_{v \in \mathcal{Y}} \langle v, w \rangle \text{ for all } w \in \mathcal{H}. \tag{4.24}$$

Let $(u_n)_{n \in \mathbb{N}}$ be the sequence in $\operatorname{conv} \mathcal{Y}$ that attains the supremum on the left-hand side of (4.24). By definition of the convex hull, for all $n \in \mathbb{N}$ there exist $(u_{n,m})_{m \in \{1,\ldots,M\}}$ with $u_{n,m} \in \mathcal{Y}[g]$, $\alpha_{n,m} \in [0,1]$ for all $m \in \{1, \ldots, M\}$, and with $\sum_{m=1}^{M} \alpha_{n,m} = 1$, such that $u_n = \sum_{m=1}^{M} \alpha_{n,m} u_{n,m}$.

For all $n \in \mathbb{N}$, let $\tilde{u}_n \in \arg\max\big\{\langle u_{n,m}, w \rangle \mid m \in \{1, \ldots, M\}\big\}$. Then,

$$\begin{aligned} \sup_{u \in \operatorname{conv} \mathcal{Y}} \langle u, w \rangle &= \lim_{n \to \infty} \langle u_n, w \rangle = \lim_{n \to \infty} \Big\langle \sum_{m=1}^{M} \alpha_{n,m} u_{n,m}, w \Big\rangle \\ &\leqslant \limsup_{n \to \infty} \sum_{m=1}^{M} \alpha_{n,m} \langle \tilde{u}_n, w \rangle \leqslant \sup_{v \in \mathcal{Y}} \langle v, w \rangle, \end{aligned}$$

demonstrating (4.24).

2) Second, show that

$$\sup_{u \in \overline{\operatorname{conv} \mathcal{Y}}^*} \langle u, w \rangle \leqslant \sup_{v \in \operatorname{conv} \mathcal{Y}} \langle v, w \rangle \text{ for all } w \in \mathcal{H}. \tag{4.25}$$

Let $w \in \mathcal{H}$. Let $(u_n)_{n \in \mathbb{N}}$ with $u_n \in \overline{\operatorname{conv} \mathcal{Y}}^*$ for all $n \in \mathbb{N}$ be a sequence that attains the supremum on the left-hand side of (4.25). Further, since $u_n \in \overline{\operatorname{conv} \mathcal{Y}}^*$ for all $n \in \mathbb{N}$, for any $n \in \mathbb{N}$ there exists a sequence $(u_{n,m})_{m \in \mathbb{N}}$ such that $u_{n,m} \in \operatorname{conv} \mathcal{Y}$ with $u_{n,m} \rightharpoonup u_n$ as $m \to \infty$.

Construct the sequence $(z_n)_{n\in\mathbb{N}}$ by picking $z_n := u_{n,m}$, where m is the smallest integer such that $\langle u_{n,\tilde{m}} - u_n, w\rangle < \frac{1}{n}$ for all integers $\tilde{m} \geqslant m$. Then, $z_n \in \operatorname{conv} \mathcal{Y}$ for all $n \in \mathbb{N}$, and

$$\begin{aligned}
\sup_{u\in\overline{\operatorname{conv}\mathcal{Y}}^*} \langle u, w\rangle &= \limsup_{n\to\infty} \langle u_n, w\rangle \\
&\leqslant \limsup_{n\to\infty} \langle z_n, w\rangle + \limsup_{n\to\infty} \underbrace{\langle u_n - z_n, w\rangle}_{<\frac{1}{n}} \\
&= \limsup_{n\to\infty} \langle z_n, w\rangle \leqslant \sup_{v\in\operatorname{conv}\mathcal{Y}} \langle v, w\rangle
\end{aligned}$$

for all $w \in \mathcal{H}$, showing (4.25) and concluding the proof. □

Theorem 4.22. *Let $\mathcal{X} \subset \mathcal{H}$ be weakly closed. Then,*

$$\overline{\partial} E_{\mathcal{X}}[g] = \overline{\partial}\left(\frac{1}{2}\|g - \Pi_{\mathcal{X}}[g]\|_2^2\right) = \overline{\operatorname{conv} g - \Pi_{\mathcal{X}}[g]}^*.$$

Proof. By definition, $v \in \overline{\partial} E_{\mathcal{X}}[g]$ means

$$\langle v, w\rangle \leqslant \hat{d} E_{\mathcal{X}}[g][w] \text{ for all } w \in \mathcal{H}.$$

By Proposition 4.20 and Lemma 4.21,

$$\widehat{d} E_{\mathcal{X}}[g][w] = \sup_{p\in\Pi_{\mathcal{X}}[g]} \langle g - p, w\rangle = \sup_{v\in\overline{\operatorname{conv} g - \Pi_{\mathcal{X}}[g]}^*} \langle v, w\rangle,$$

meaning that $v \in \overline{\operatorname{conv} g - \Pi_{\mathcal{X}}[g]}^*$ implies $v \in \overline{\partial} E_{\mathcal{X}}[g]$. To conclude the proof, let us show that $v \notin \overline{\operatorname{conv} g - \Pi_{\mathcal{X}}[g]}^*$ implies $v \notin \overline{\partial} E_{\mathcal{X}}[g]$.

Let $\mathcal{Y} := \overline{\operatorname{conv} g - \Pi_{\mathcal{X}}[g]}^*$, assume that $v \notin \mathcal{Y}$. Since $\mathcal{Y}$ is a weakly closed convex set, the projection $\Pi_{\mathcal{Y}}$ is well-defined and has the unique projecton selection $P_{\mathcal{Y}} \colon \mathcal{H} \to \mathcal{H}$ (e.g. Proposition 3.9). Further, since $\mathcal{Y}$ is convex, by Lemma 3.10

$$\langle v - P_{\mathcal{Y}}[v], P_{\mathcal{Y}}[u] - P_{\mathcal{Y}}[v]\rangle \leqslant 0 \text{ for all } u \in \mathcal{H}. \tag{4.26}$$

We want to show that $\langle v, w\rangle > \sup_{u\in\overline{\operatorname{conv}}^*Y} \langle u, w\rangle$ for an appropriately chosen $w \in \mathcal{H}$. Pick $w = v - P_{\mathcal{Y}}[v]$; then, $\|w\|_2 > 0$ since $v \notin \mathcal{Y}$. Further, for all $u \in \mathcal{Y}$ have

$$\begin{aligned}
\langle v - u, w\rangle &= \langle v - P_{\mathcal{Y}}[v], w\rangle + \langle P_{\mathcal{Y}}[v] - u, w\rangle \\
&= \|w\|_2^2 + \underbrace{\langle P_{\mathcal{Y}}[v] - u, v - P_{\mathcal{Y}}[v]\rangle}_{\overset{(*)}{\geqslant} 0} \geqslant \|w\|_2^2,
\end{aligned}$$

where $(*)$ holds by (4.26) since $u = P_{\mathcal{Y}}[u]$. Therefore,

$$\begin{gathered}
\langle v, w\rangle \geqslant \langle u, w\rangle + \|w\|_2^2 \text{ for all } u \in \mathcal{Y}, \text{ thus} \\
\langle v, w\rangle \geqslant \sup_{u\in\mathcal{Y}} \langle u, w\rangle + \|w\|_2^2 > \sup_{u\in\mathcal{Y}} \langle u, w\rangle.
\end{gathered}$$

□

4.3.3 *Generalized subdifferential of energy functionals*

PROPOSITION 4.23 (DINI-ε-SUBDIFFERENTIAL). *Let $\mathcal{X} \subset \mathcal{H}$ be weakly closed. Then, for any $g \in \mathcal{H}$ there exists an $\varepsilon_* > 0$ such that*

$$\partial_\varepsilon^- E_{\mathcal{X}}[g] = \begin{cases} \{v \in \mathcal{H} \mid \|g - P_{\mathcal{X}}[g] - v\|_2 \leqslant \varepsilon\} \text{ if } \Pi_{\mathcal{X}} \text{ is single-valued at } g \text{, and} \\ \varnothing \text{ else} \end{cases}$$

for all $\varepsilon \in [0, \varepsilon_*)$. Specifically, the single-valued case holds for any $\varepsilon_* \in \mathbb{R}_{>0}$, and the multi-valued case holds for any

$$\varepsilon_* < \sup_{p,q \in \Pi_{\mathcal{X}}[g]} \|p - q\|_2 / 2. \tag{4.27}$$

Proof. Let $g \in \mathcal{H}$. First, consider the case when $\Pi_{\mathcal{X}}$ is single-valued at g, i.e. $\Pi_{\mathcal{X}}[g] = \{P_{\mathcal{X}}[g]\}$ for any single-valued selection $P_{\mathcal{X}}$.

"$\supseteq$". Assume that

$$v \in \{\tilde{v} \in \mathcal{H} \mid \|g - P_{\mathcal{X}}[g] - \tilde{v}\|_2 \leqslant \varepsilon\}.$$

Then,

$$\begin{aligned} &\liminf_{t \searrow 0} \frac{E_{\mathcal{X}}[g + tw] - E_{\mathcal{X}}[g]}{t} - \langle v, w \rangle \\ &\quad \overset{(*1)}{\geqslant} \liminf_{t \searrow 0} \langle g - P_{\mathcal{X}}[g + tw], w \rangle - \langle v, w \rangle \\ &\qquad \overset{(*2)}{=} \langle g - P_{\mathcal{X}}[g] - v, w \rangle \overset{(*3)}{\geqslant} -\varepsilon \|w\|_2, \end{aligned}$$

where we used Lemma 4.19 in $(*1)$, continuity of $P_{\mathcal{X}}$ due to single-valuedness of $\Pi_{\mathcal{X}}$ at g (Proposition 3.11) in $(*2)$, and definition of v in $(*3)$. Thus, $v \in \partial_\varepsilon^- E_{\mathcal{X}}[g]$.

"$\subseteq$". Let $v \in \partial_\varepsilon^- E_{\mathcal{X}}[g]$. Assume that for $y := g - P_{\mathcal{X}}[g] - v$ holds $\|y\|_2 = \alpha > \varepsilon$; let us show that this leads to a contradiction. Indeed, by Lemma 4.19 have

$$\begin{aligned} &\liminf_{t \searrow 0} \frac{E_{\mathcal{X}}[g + tw] - E_{\mathcal{X}}[g]}{\varepsilon} - \langle v, w \rangle \\ &\qquad \leqslant \langle g - P_{\mathcal{X}}[g], w \rangle - \langle v, w \rangle = \langle y, w \rangle \end{aligned} \tag{4.28}$$

for all $w \in \mathcal{H}$. For $w = -y$ have

$$\begin{aligned} &\liminf_{\varepsilon \searrow 0} \frac{E_{\mathcal{X}}[g + \varepsilon w] - E_{\mathcal{X}}[g]}{\varepsilon} - \langle v, w \rangle \\ &\leqslant -\|y\|_2^2 = -\alpha \|w\|_2 < -\varepsilon \|w\|_2 \end{aligned} \tag{4.29}$$

in contradiction to $v \in \partial_\varepsilon^- E_{\mathcal{X}}[g]$, concluding the proof for the single-valued case.

Second, consider the case when $\Pi_{\mathcal{X}}$ is multi-valued at g, i.e. there exist $p, q \in \Pi_{\mathcal{X}}[g]$ such that $p \neq q$. Let $\varepsilon < \|p-q\|/2$. Assume that $v \in \partial_{\varepsilon}^{-} E_{\mathcal{X}}[g]$. Proceed analogously to case "$\subseteq$" above to show that $\|g-p-v\|_2 \leqslant \varepsilon$ and $\|g-q-v\|_2 \leqslant \varepsilon$. For example, if one assumes that $\|g-p-v\|_2 > \varepsilon$, then for $w_p = -(g-p-v)$ get — analogously to Equation (4.28) and Equation (4.29) —

$$\liminf_{t \searrow 0} \frac{E_{\mathcal{X}}[g+tw_p] - E_{\mathcal{X}}[g]}{\varepsilon} - \langle v, w_p \rangle < -\varepsilon \|w_p\|_2$$

in contradiction to $v \in \partial_{\varepsilon}^{-} E_{\mathcal{X}}[g]$, showing $\|g-p-v\|_2 \leqslant \varepsilon$, and similarly for q.

Thus,

$$\begin{aligned} \|p-q\|_2 &= \|g-q-v-(g-p-v)\|_2 \\ &\leqslant \|g-q-v\|_2 + \|g-p-v\|_2 \leqslant 2\varepsilon < \|p-q\|_2 \end{aligned}$$

by choice of ε, which is a contradiciton. Thus, there do not exist $v \in \partial_{\varepsilon}^{-} E_{\mathcal{X}}[g]$, from which it is straight-forward to follow (4.27). ◻

Theorem 4.24 (Generalized subdifferential). *Let $\mathcal{X} \subset \mathcal{H}$ be weakly closed. Then, $\partial_{KM} E_{\mathcal{X}}[g] = g - \Pi_{\mathcal{X}}[g]$.*

Proof. "$\supseteq$". Let $p \in \Pi_{\mathcal{X}}[g]$, let $g_n = (1 - \frac{1}{n} g) + \frac{1}{n} p$ for $n \in \mathbb{N}$. By Lemma 3.8 (interpolation projection property),$P_{\mathcal{X}}[g_n] = \{p\}$, and $g_n - p \in \partial_{1/n}^{-} E_{\mathcal{X}}[g_n]$ by Proposition 4.23. Thus, $g_n \to g$ as $n \to \infty$, and $(g_n - p)$ is a sequence in $\partial_{1/n}^{-} E_{\mathcal{X}}[g_n]$ that converges to $g - p$ as $n \to \infty$. Thus, $g - p \in \partial_{KM} E_{\mathcal{X}}[g]$.

"$\subseteq$". Let (g_n) be a sequence converging to $g \in \mathcal{H}$, $(v_n) \in \partial_{1/n}^{-} E_{\mathcal{X}}[g_n]$ with $v_n \rightharpoonup v \in \mathcal{H}$ and $\varepsilon_n \to 0$ as $n \to \infty$. (W.l.o.g. we can assume that the sequence (ε_n) is non-increasing with $\varepsilon_0 \leqslant \varepsilon_*$, where ε_* is determined in By Proposition 4.23 and depends on g.) By Proposition 4.23, $v_n \in \partial_{1/n}^{-} E_{\mathcal{X}}[g_n]$ implies that $\Pi_{\mathcal{X}}$ is single-valued at g_n, and that $\|v_n - (g_n - P_{\mathcal{X}}[g_n])\|_2 \leqslant \varepsilon_n$ for any single-valued selection $P_{\mathcal{X}}$. Define $q := g - v$, trivially meaning that $v = g - q$. Then, using w.l.s.c. of the norm, have

$$\begin{aligned} \|g-q\|_2 = \|v\|_2 \leqslant \lim_{n\to\infty} \|v_n\|_2 &\leqslant \lim_{n\to\infty} (\|g_n - P_{\mathcal{X}}[g_n]\|_2 + \varepsilon_n) \\ &\leqslant \lim_{n\to\infty} \|g_n - P_{\mathcal{X}}[g]\|_2 = \|g - P_{\mathcal{X}}[g]\|_2 \end{aligned}$$

for any single-valued selection $P_{\mathcal{X}}$, meaning that $q \in \Pi_{\mathcal{X}}[g]$, and that $v = g - q \in g - \Pi_{\mathcal{X}}[g]$. ◻

4.3.4 *Slopes*

Another possible way to generalize Fréchet-differentiability of $E_{\mathcal{M}}$ is to do with the notion of slopes. Slopes are commonly used in the theory of minimizing movements to describe generalized gradient flows [AG13; AGS01].

Definition 4.25 (Slope). *Let* $E\colon \mathcal{H} \to \mathbb{R} \cup \{\infty\}$, *let* $g \in \mathcal{H}$ *be such that* $E[g] < \infty$.

The slope $|\nabla E|\colon \mathcal{H} \to \mathbb{R}$ *is defined by*

$$|\nabla E|[g] := \max \left\{ \limsup_{h \to 0} -\frac{E[g+h] - E[g]}{\|h\|_2}, 0 \right\}.$$

Slopes are an important tool that can be used to generalize gradient flows. Informally speaking, in the classical setting the flow $g\colon [0,T] \to \mathcal{H}$ is a gradient flow, if it satisfies $\partial_t g(t) = -\nabla E[g(t)]$ for some energy E. In the generalized setting, the flow $g\colon [0,T] \to \mathcal{H}$ is considered a gradient flow if $\|\partial_t g(t)\|_2$ is suitably related to $-|\nabla E|[g(t)]$; see [AG13] for details. For example, if the Fréchet-derivative $\nabla E_{\mathcal{X}}$ is ill-defined, and if $E_{\mathcal{X}}$ is non-convex (such that subgradients are ill-defined as well), the slope may still exist. This is illustrated by the following example.

Example 4.26 (Slope of the modulus energy at a single point). Let $\sqrt{I_k} > 0$. Consider the function $E[x] = \frac{1}{2}(|x| - \sqrt{I_k})^2$. This is a pointwise analogue of the modulus constraint in the Fourier space. The function $E[x]$ is non-differentiable at the point $x = 0$, see Figure 4.1. The slope of E can be calculated as follows.

Observe that for $x \leqslant 0$ the energy E coincides with the differentiable function $E_{\leqslant}[x] = \frac{1}{2}(x + \sqrt{I_k})^2$, and for $x \geqslant 0$ the energy E coincides with the differentiable function $E_{\geqslant}[x] = \frac{1}{2}(x - \sqrt{I_k})^2$.

Thus,

$$\limsup_{x \to 0} -\frac{E[x] - E[0]}{\|x\|_2}$$

is equal to largest of

$$\limsup_{x \nearrow 0} -\frac{E[x] - E[0]}{\|x\|_2} \qquad \text{and} \qquad \limsup_{x \searrow 0} -\frac{E[x] - E[0]}{\|x\|_2};$$

for these subdomains, the energy E coincides, respectively, with differentiable functions

$$\limsup_{x \nearrow 0} -\frac{E_{\leqslant}[x] - E_{\leqslant}[0]}{\|x\|_2} \qquad \text{and} \qquad \limsup_{x \searrow 0} -\frac{E_{\geqslant}[x] - E_{\geqslant}[0]}{\|x\|_2}$$

which, by differential calculus, are both equal to $\sqrt{I_k}$.

Therefore, $|\nabla E|[0] = \sqrt{I_k}$; cf. the affine function $A(x) = -\sqrt{I_k}x + \frac{1}{2}\sqrt{I_k}^2$ on Figure 4.1.

The energy E is locally concave at 0; therefore, one could also consider the local supgradient of E at 0. However, the supgradient approach would not work for the infinitely-dimensional modulus en-

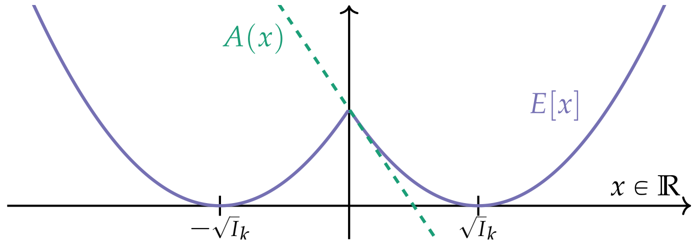

Figure 4.1: Illustration to Example 4.26 (Slope of the modulus energy). The function $E[x] = \frac{1}{2}(|x| - \sqrt{I_k})^2$ is non-differentiable at the point $x = 0$. Its slope at 0 equals $\sqrt{I_k}$ (cf. the tangent $A(x) = -\sqrt{I_k}x + \frac{1}{2}\sqrt{I_k^2}$).

ergy $E_{\mathcal{M}}$ in Fourier space: the energy $E_{\mathcal{M}}$ is not locally concave, as the values $\sqrt{I}(k)$ can get arbitrarily small.

Proposition 4.27 (Slopes of energy functionals). *Let $\mathcal{D} \subset \mathcal{H}$, let $P_{\mathcal{X}} \colon \mathcal{D} \to \mathcal{H}$ be a single-valued projection selection onto $\mathcal{X} \subset \mathcal{H}$, let $E_{\mathcal{X}}$ be the corresponding energy functional. Then, $|\nabla E_{\mathcal{X}}|[g] = \|g - P_{\mathcal{X}}[g]\|_2$ for any $g \in \mathcal{D}$.*

Proof. Distinguish between the cases $g \in \mathcal{X}$ and $g \notin \mathcal{X}$.

If $g \in \mathcal{X}$, then $E_{\mathcal{X}}[g] = 0$, and

$$-\frac{E_{\mathcal{X}}[g+h] - E_{\mathcal{X}}[g]}{\|h\|_2} = -\frac{E_{\mathcal{X}}[g+h]}{\|h\|_2} < 0$$

for any h with $g + h \in \mathcal{D}$, hence $|\nabla E_{\mathcal{X}}|[g] = 0 = \|g - P_{\mathcal{P}}[g]\|_2$.

If $g \in \mathcal{D} \backslash \mathcal{X}$, then $\|g - P_{\mathcal{P}}[g]\|_2 > 0$, and

$$\begin{aligned}
&\limsup_{h \to 0} -\frac{1}{\|h\|_2}(E_{\mathcal{X}}[g+h] - E_{\mathcal{X}}[g]) \\
&= \lim_{\varepsilon \to 0} \sup_{\substack{g+h \in \mathcal{D} \\ \|h\|_2 = \varepsilon \|g - P_{\mathcal{X}}[g]\|_2}} \frac{1}{\|h\|_2}(E_{\mathcal{X}}[g] - E_{\mathcal{X}}[g+h]) \\
&\overset{(a)}{=} \lim_{\varepsilon \to 0} \frac{1}{\varepsilon \|g - P_{\mathcal{X}}[g]\|_2}\Big(E_{\mathcal{X}}[g] - E_{\mathcal{X}}\left[g - \varepsilon(g - P_{\mathcal{X}}[g])\right]\Big) \\
&= \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon \|g - P_{\mathcal{X}}[g]\|_2}\Big(\|g - P_{\mathcal{X}}[g]\|_2^2 \\
&\qquad - \Big\|(1-\varepsilon)g + \varepsilon P_{\mathcal{X}}[g] - P_{\mathcal{X}}\Big[(1 - \varepsilon g) + \varepsilon P_{\mathcal{X}}[g]\Big]\Big\|\Big) \\
&\overset{(b)}{=} \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon \|g - P_{\mathcal{X}}[g]\|_2}(\|g - P_{\mathcal{X}}[g]\|_2^2 - (1-\varepsilon)^2\|g - P_{\mathcal{X}}[g]\|_2^2) \\
&= \|g - P_{\mathcal{X}}[g]\|_2.
\end{aligned}$$

Lemma 4.11 (explicit local minimizer) was used in (a) and Lemma 3.8 (interpolation projection property) was used in (b). □

5

CONNECTIONS BETWEEN PROJECTION ALGORITHMS

This chapter aims to systematically present some projection-based algorithms used for crystallographic phase retrieval, and to connect the Error-Reduction (ER) and Douglas-Rachford (DR) algorithms to the equation we call Error-Reduction Flow. To this end, we proceed as follows. We start with a brief discussion on local and global formulations between algorithms, and with naming conventions used throughout the thesis. Namely,

i) Section 5.1 states some of Fienup variants and highlights some connections between different formulations;

ii) Section 5.2 formalizes the setting of the AP (alternating projections) algorithm and its specific phase retrieval instance ER, and shows that ER is a discretization of a formal gradient flow with energy $E_{\mathcal{M}} + E_{\mathcal{A}}$, or a rigorous selection of the generalized subdifferential flow with the same energy;

iii) Section 5.3 formalizes the setting of the DR-LM algorithm, presents the well-known connection between DR-LM and DR-cf in convex optimization, and demonstrates that this connection persists between DR-LM and DR-HIO.

Two main original contributions of this chapter are i) Remark 5.17 that motivates introduction of equation ERF by establishing its connection to ER through time discretization, and ii) the argument outlined in Section 5.3.5 and formalized in Proposition 5.52 that establishes connection between DR-LM and DR-HIO.

This connection is conceptually different from the connection between DR and HIO that was established in [BCL02]. The argument from [BCL02] is sketched on p. 85 in the transformations between HIO, HPR, and DR.

Local and global forms of Fienup variants; naming conventions

This chapter describes certain projection-based algorithms used for crystallographic phase retrieval. These algorithms are sometimes known as Fienup variants, since many of them are variants of algorithms that were systematically studied in Fienup's celebrated paper [Fie82].

Fienup variants generate a sequence of approximations $(g_n)_{n\in\mathbb{N}}$; the iterate g_{n+1} is explicitely dependent only on the previous approximation g_n.

There are two common ways of writing down Fienup variants. The first, local way typically has the form

$$g_{n+1}(x) = \begin{cases} \text{expression that depends on } x \text{ and } g_n(x) \text{ for certain } x; \\ \text{expression that depends on } x \text{ and } g_n(x) \text{ for other } x. \end{cases}$$

This form is favoured in the optics community and is considered more physically intuitive [Luk05, pp. 40-41]. The second, global way to formulate Fienup variants typically has the form

$$g_{n+1} = T[g_n],$$

where the update operator T combines various projection operators in a manner that capitalizes on the geometric structure of the underlying sets.

For the most part, update operators are heuristically developed. For example, they can be inspired by local analysis of the linearized problem near the solution [Els03] or have connections to other feasibility problems — for example, to convex feasibility problems [BCL02]. The behavior of projection algorithms in convex setting is better understood (e. g., see monography [BC17]), but convex results generally do not carry over to phase retrieval.

One of the main ideas of this work is to provide new insights to Fienup variants by studying the corresponding evolution equations. The study is focused on ER — the most basic algorithm used in phase retrieval — and DR-HIO version of the HIO algorithm — one of the state-of-the art algorithms for crystallographic phase retrieval (cf. [ELB18]).

There are many names for Error-Reduction and Douglas-Rachford variants that depend on the exact setting and formulation of these algorithms.

For reader's convenience, the following remarks summarize the most important names of algorithms that we use throughout this thesis.

Remark 5.1 (Naming conventions). We use the name AP (Alternating Projections), see Definition 5.11, to describe the generic feasibility problem algorithm with the update $g_{n+1} = P_{\mathcal{Y}} \circ P_{\mathcal{X}}[g_n]$ for two proximal sets $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$.

We use the name ER (Error-Reduction), see Definition 5.12, to describe AP for the phase retrieval case with $\mathcal{Y} = \mathcal{A}$ and $\mathcal{X} = \mathcal{M}$.

We use the name DR-LM (Douglas-Rachford in Lions-Mercier formulation), see Definition 5.41, to describe the algorithm for finding zeros of maximal monotone operators A and B with the update

$$g_{n+1} = g_n + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g_n] - J_{\lambda B}[g_n],$$

where $\lambda > 0$, $J_{\lambda A} = (I + \lambda A)^{-1}$ is the resolvent of λA, and $J_{\lambda B}$ is a the resolvent of λB.

We use the name DR (Douglas-Rachford), see Definition 5.53, to to describe the generic feasibility problem algorithm with the update

$$g_{n+1} = g_n + P_{\mathcal{Y}} \circ (2P_{\mathcal{X}}[g_n] - g_n) - P_{\mathcal{X}}[g_n].$$

for two proximal sets $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$. We use the name DR-cf (Douglas-Rachford in convex formulation), see Definition 5.47, for the particular case of DR when $\mathcal{X}$ and $\mathcal{Y}$ are convex. We use the name DR-HIO (Douglas-Rachford variant of Hybrid-Input-Output), see Definition 5.49, for the particular case of DR for phase retrieval, when $\mathcal{X}$ and $\mathcal{Y}$ belong to $\{\mathcal{A}, \mathcal{M}\}$.

We use the name HIO (Hybrid-Input-Output), described on p. 84, for the phase retrieval algorithm that was introduced in [Fie82] by the same name. HIO is related to DR through the HPR (Hybrid-Projection-Reflection) algorithm as described on p. 85; this connection was first established in [BCL02].

Remark 5.2 (Naming conventions for variants introduced in this thesis). We use the name APF (AP Flow), introduced in Section 6.1, to describe the formal evolution equation

$$\partial_t g = -(g - P_{\mathcal{X}}[g]) - (g - P_{\mathcal{Y}}[g]),$$

and the name dAPF (discretized APF), see Definition 6.1, to describe the connected algorithm with the update

$$g_{n+1}^{(\varepsilon)} = g_n^{(\varepsilon)} + \varepsilon\Big(-(g_n^{(\varepsilon)} - P_{\mathcal{X}}[g_n^{(\varepsilon)}]) - (g_n^{(\varepsilon)} - P_{\mathcal{Y}}[g_n^{(\varepsilon)}])\Big),$$

for two weakly closed sets $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$.

We use the name ERF (ER Flow), introduced in Remark 5.17, to describe APF for the phase retrieval case $\mathcal{Y} = \mathcal{A}$ and $\mathcal{X} = \mathcal{M}$. We use the name dERF (discretized ERF), see Definition 6.19, to describe dAPF for the phase retrieval case $\mathcal{Y} = \mathcal{A}$, $\mathcal{X} = \mathcal{M}$.

We use the name 2v-FPF (two-variable Feasibility Problem Flow), see Chapter 9, to describe a system of equations

$$\partial_t \begin{pmatrix} s \\ d \end{pmatrix} = M \cdot \begin{pmatrix} \frac{\delta}{\delta s} \\ \frac{\delta}{\delta d} \end{pmatrix} \left(\frac{1}{2} E_{\mathcal{X}}[s+d] + \frac{1}{2} E_{\mathcal{Y}}[s-d] - \frac{1}{2}\|d\|_2^2 \right),$$

where $M \in \mathbb{R}^{2\times 2}$, and the name d2v-FPF (discretized 2v-FPF), see Definition 9.8, to describe the corresponding explicitely discretized algorithms.

For the particular case of 2v-FPF and d2v-FPF with $M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, we use the names DRF (DR Flow) and dDRF (discretized DRF) to highlight their connection to DR. We use the name DR/HIO-F for the particular phase retrieval case of DRF when $\mathcal{X}$ and $\mathcal{Y}$ belong to $\{\mathcal{A}, \mathcal{M}\}$.

All the algorithms named above will be rigorously introduced in due course of the thesis.

5.1 FIENUP VARIANTS

This section presents some Fienup variants used in literature. First, we write down the definition and establish some properties of a local indicator projecton; these are used in transformations between local and global forms of Fienup variants. Then, algorithms ER, BIO and HIO are presented in their local forms (as in [Fie82]) and subsequently reformulated to their global form (as in [BCL02]). Finally, for reader's convenience we write down Hybrid-Projection-Reflection, Douglas-Rachford, Relaxed Averaged Alternating Reflections, and Alternating Direction Method of Multipliers algoritms.

5.1.1 *Properties of local projections*

This subsection presents properties of local and indicator projections (Corollary 5.5, Lemma 5.7) that are used later in the work.

In particular, the following definition and corollary are used to establish equivalence of local and global algorithm formulations.

DEFINITION 5.3 (INDICATOR PROJECTION). *We call a projection selection $P_\mathcal{X}\colon \mathcal{H} \to \mathcal{H}$ an indicator projection selection, if for every $g \in \mathcal{H}$ there exists a measurable indicator set $S[g] \subset \Omega$ such that*

$$P_\mathcal{X}[g] = \mathbb{1}_{S[g]} g.$$

Example 5.4 (Indicator projection). The operators $P_\mathcal{P}, P_\mathcal{S}$ and $P_{\mathcal{T}_s(\nu)}$ are indicator projection selections, while $P_{\mathcal{T}_a(\alpha)}$ and $P_\mathcal{M}$ are not.

If $P_\mathcal{X}$ is an indicator projection selection, the operator $g \mapsto g - P_\mathcal{X}[g]$ is also an indicator projection selection:

$$g - P_\mathcal{X}[g] = g - \mathbb{1}_{S[g]} g = \mathbb{1}_{\Omega \setminus S[g]} g.$$

COROLLARY 5.5. *Let projection $P_\mathcal{X}$ be an indicator projection with the indicator set $S[g]$ at $g \in \mathcal{H}$. Further, let $P_\mathcal{X}$ be local with a local version P_Θ for an appropriate set Θ. Then,*

$$\underbrace{\{x \in \Omega \mid (x, g(x)) \in \Theta\}}_{=:S_1} = \underbrace{\{x \in \Omega \mid x \in S[g]\}}_{=:S_2},$$

where the equality holds for all $g \in \mathcal{H}$ and up to a Lebesgue null-set, i.e. $\lambda(S_1 \setminus S_2) + \lambda(S_2 \setminus S_1) = 0$.

Proof. For almost all $x \in \Omega$ and for all $g \in \mathcal{H}$ the equality

$$P_\Theta\Big((x, g(x))\Big) \overset{(*)}{=} (x, P_\mathcal{X}[g](x)) \overset{(**)}{=} (x, \mathbb{1}_{S[g]}(x)g(x)) \tag{5.1}$$

follows direct from definitions of local $(*)$ and indicator $(**)$ projections.

"$\subseteq$". Let $x \in \Omega$ be such that $(x, g(x)) \in \Theta$. Then, by definition of projection, $P_\Theta(x, g(x)) = (x, g(x))$. From Equation (5.1) then follows that $x \in S[g]$ for a. a. $x \in \Omega$ that satisfy $(x, g(x)) \in \Theta$.

"$\supseteq$". Let $x \in \Omega$ be such that $(x, g(x)) \notin \Theta$. Then, by definition of projection, $P_\Theta(x, g(x)) \neq (x, g(x))$. From Equation (5.1) then follows that $g(x) = 0$ and that $x \notin S[g]$ for a. a. $x \in \Omega$ that satisfy $(x, g(x)) \notin \Theta$. □

Example 5.6 (Positivity). The positivity projector $P_\mathcal{P}[g] = \mathbb{1}_{g \geqslant 0} g$ is an indicator projector with an indicator set $S[g] = \{g \geqslant 0\} = \{x \in \Omega \mid g(x) \geqslant 0\}$ at g. Further, it is a local projector with local version $P_\mathcal{P}^{(\mathrm{loc})} = P_\Theta$ with $\Theta = \Omega \times \mathbb{R}_{\geqslant 0}$ by Example 3.30. Corollary 5.5 states that

$$(x, g(x)) \in \Omega \times \mathbb{R}_{\geqslant 0} \Leftrightarrow x \in \{g \geqslant 0\}$$

for a. a. $x \in \Omega$ and for all $g \in \mathcal{H}$.

The following property of local indicator projections is relevant for characterization of ER fixed points in Section 8.1.

LEMMA 5.7. *Let projecton selection $P_\mathcal{X}$ be an indicator projection selection with the indicator set $S[g]$ at $g \in \mathcal{H}(\Omega)$. Further, let $P_\mathcal{X}$ be local with a local version P_Θ for an appropriate set Θ. Assume that $P_\mathcal{X}[2g] = 2P_\mathcal{X}[g]$ for all $g \in \mathcal{H}$.*

Then,

$$P_\mathcal{X}\big[2g - P_\mathcal{X}[g]\big] = P_\mathcal{X}[g]$$

for all $g \in \mathcal{H}$.

Proof. Let $g \in \mathcal{H}$. For almost all $x \in S[g]$ have

$$2g(x) - P_\mathcal{X}[g](x) = g(x) + \mathbb{1}_{\Omega \setminus S[g]}(x)g(x) = g(x);$$

therefore,

$$\begin{aligned}\Big(x, P_\mathcal{X}\big[2g - P_\mathcal{X}[g]\big](x)\Big) &= P_\Theta\Big((x, 2g(x) - P_\mathcal{X}[g](x))\Big)\\ &= P_\Theta\Big((x, g(x))\Big) = (x, P_\mathcal{X}[g](x))\end{aligned}$$

for a. a. $x \in S[g]$.

Further, for almost all $x \in \Omega \setminus S[g]$ have

$$2g(x) - P_\mathcal{X}[g](x) = 2g(x) - \mathbb{1}_{S[g]}(x)g(x) = 2g(x);$$

therefore,

$$\begin{aligned}
&\Big(x, P_{\mathcal{X}}\big[2g - P_{\mathcal{X}}[g]\big](x)\Big) \\
&\qquad = P_{\Theta}\Big(\big(x, 2g(x) - P_{\mathcal{X}}[g](x)\big)\Big) = P_{\Theta}\Big(\big(x, 2g(x)\big)\Big) \\
&\qquad = \big(x, P_{\mathcal{X}}[2g](x)\big) = \big(x, 2P_{\mathcal{X}}[g](x)\big) \\
&\qquad = \Big(x, 2\mathbb{1}_{x \in S[g]} g(x)\Big) = (x, 0) \\
&\qquad\qquad\qquad\qquad = \big(x, \mathbb{1}_{x \in S[g]} g(x)\big) = \big(x, P_{\mathcal{X}}[g](x)\big)
\end{aligned}$$

for a. a. $x \in \Omega \backslash S[g]$. □

Example 5.8. It is straightforward to check that for $P_{\mathcal{X}} \in \{P_{\mathcal{P}}, P_{\mathcal{S}}, P_{\mathcal{S} \cap \mathcal{P}} = P_{\mathcal{S}} \circ P_{\mathcal{P}}\}$ the requirements of Lemma 5.7 are satisfied. Therefore, $P_{\mathcal{X}}\big[2g - P_{\mathcal{X}}[g]\big] = P_{\mathcal{X}}[g]$ for these $\mathcal{X}$.

The projecton selection $P_{\mathcal{T}_s(\nu)}$ is not local and therefore does not satisfy the requirements of Lemma 5.7. It is easy to construct an example for which $P_{\mathcal{T}_s(\nu)}\big[2g - P_{\mathcal{T}_s(\nu)}[g]\big] = P_{\mathcal{T}_s(\nu)}[g]$. Indeed, let $\Omega = \mathbb{T} = [0, 2\pi)$, let $g(x) = 1$ if $x \in [0, \pi)$, let $g(x) = 2/3$ if $x \in [\pi, 2\pi)$. Then, for $\nu = \pi$ have $P_{\mathcal{T}_s(\nu)}[g](x) = \mathbb{1}_{[0,\pi)}(x) g(x)$, and $P_{\mathcal{T}_s(\nu)}\big[2g - P_{\mathcal{T}_s(\nu)}[g]\big](x) = 4/3\mathbb{1}_{[\pi,2\pi)}$.

5.1.2 *ER, BIO, HIO: local and global*

This subsection contains standard algorithm reformulations that are well-known in phase retrieval. Here, these reformulations are written down using the properties of local projections. The presented transformations resemble very closely the standard ones known in literature (see [BCL02]), but are presented in a more general form using the properies from above.

Let us formally fix the notion of an approximation sequence generated by an algorithm.

DEFINITION 5.9 (APPROXIMATION SEQUENCE). *Let $g_0 \in \mathcal{H}$, $T \colon \mathcal{H} \to \mathcal{H}$. The sequence $(g_n)_{n \in \mathbb{N}_0}$ is generated by T with initial value g_0, if the update*

$$g_{n+1} = T[g_n] \tag{5.2}$$

holds for all $n \in \mathbb{N}_0$. The sequence $(g_n)_{n \in \mathbb{N}_0}$ is called the approximation sequence; the operator T is called the update operator. As is common in literature, we use the update equation (5.2) to define the operator T implying the setting of this definition.

Error-Reduction algorithm

Local formulation. Let $P_{\mathcal{A}}$ be local with the local version P_Θ for an appropriate $\Theta \subset \Omega \times \mathbb{R}$. Then, the local ER update operator is defined by the update

$$g_{n+1}(x) = \begin{cases} P_{\mathcal{M}}[g_n](x) & \text{if } \big(x, P_{\mathcal{M}}[g_n](x)\big) \in \Theta; \\ 0 & \text{else} \end{cases} \tag{ER-local}$$

for almost all $x \in \Omega$.
Transformation to global formulation. Further, assume that $P_{\mathcal{A}}$ is an indicator projecton selection with the indicator set $S\big[P_{\mathcal{M}}[g]\big]$ at $P_{\mathcal{M}}[g]$. Then,

$$\begin{aligned} g_{n+1}(x) &= \mathbb{1}_{\{x \in \Omega | \left(x, P_{\mathcal{M}}[g_n](x)\right) \in \Theta\}}(x) P_{\mathcal{M}}[g_n](x) \\ &\overset{(*)}{=} \mathbb{1}_{S\left[P_{\mathcal{M}}[g]\right]} P_{\mathcal{M}}[g_n](x) = P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n](x), \end{aligned}$$

with Corollary 5.5 used in $(*)$.

For example, this transformation is valid for positivity and support projections, but it is not valid for amplitude thresholding since any projection selection $P_{\mathcal{T}_a(\alpha)}$ is not indicator for $\alpha > 0$, and it is not valid for support size projecton since any projection selection $P_{\mathcal{T}_s(\nu)}$ is not local and can not be written down in the form ER-local.

The resulting form of ER,

$$g_{n+1} = P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n], \tag{ER}$$

is discussed in Section 5.2 in more detail.

Basic-Input-Output algorithm

Local formulation. Let $P_{\mathcal{A}}$ be local with the local version P_Θ for an appropriate $\Theta \subset \Omega \times \mathbb{R}$; let $\beta > 0$. The Basic-Input-Output (BIO) update operator is defined by the update

$$g_{n+1}(x) = \begin{cases} g_n(x) & \text{if } \big(x, P_{\mathcal{M}}[g_n](x)\big) \in \Theta; \\ g_n(x) - \beta P_{\mathcal{M}}[g_n](x) & \text{else.} \end{cases} \tag{BIO-local}$$

Transformation to global formulation. Further, assume that $P_{\mathcal{A}}$ is the support projector: $P_{\mathcal{A}} = P_{\mathcal{S}}(S)$ for measurable subset $S \subset \Omega$, with Θ as in

Corollary 3.32. (In other words, $P_{\mathcal{A}}$ is an indicator projecton selection with the constant indicator set $S[g] = S$.) Then,

$$\begin{aligned} g_{n+1}(x) &= \mathbb{1}_{\{(x,P_{\mathcal{M}}[g_n](x))\in\Theta\}}(x)g_n(x) \\ &\quad + (1-\mathbb{1}_{\{(x,P_{\mathcal{M}}[g_n](x))\in\Theta\}}(x))(g_n(x)-\beta P_{\mathcal{M}}[g_n](x)) \\ &\overset{(*)}{=} \mathbb{1}_S(x)g_n(x) \\ &\quad + (1-\mathbb{1}_S(x))(g_n(x)-\beta P_{\mathcal{M}}[g_n](x)) \\ &= g_n - \beta P_{\mathcal{M}}[g_n] + \beta P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n]. \end{aligned}$$

with Corollary 5.5 used in $(*)$.

The resulting form of BIO,

$$g_{n+1} = g_n - \beta P_{\mathcal{M}}[g_n] + \beta P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n], \tag{BIO}$$

corresponds to a nonconvex version of the Dykstra algorithm, see Definition 6.35. This correspondence was first established in [BCL02]. In convex analysis, Dykstra's algorithm can be used to calculate projectors onto nonempty intersections of closed convex sets, see Theorem 6.36. In phase retrieval setting, it shows similar stagnation issues as ER [Fie82], thus it is, to our knowledge, not frequently studied in that context.

Hybrid-Input-Output algorithm

Local formulation. Let $P_{\mathcal{A}}$ be local with the local version P_Θ for an appropriate $\Theta \subset \Omega \times \mathbb{R}$; let $\beta \in [-1, 1]$. The Hybrid-Input-Output (HIO) update operator is defined by the update

$$g_{n+1}(x) = \begin{cases} P_{\mathcal{M}}[g_n](x) & \text{if } (x, P_{\mathcal{M}}[g_n](x)) \in \Theta; \\ g_n(x) - \beta P_{\mathcal{M}}[g_n](x) & \text{else.} \end{cases} \tag{HIO-local}$$

There is a minor variation of HIO-local that treats the case $(x, 0)$ differently, cf. [BCL02, Remark 4.1].

Transformation to global formulation. Further, assume that $P_{\mathcal{A}}$ is the support projector: $P_{\mathcal{A}} = P_{\mathcal{S}}(S)$ for measurable subset $S \subset \Omega$, with Θ as in Corollary 3.32. Similarly to the BIO-local case,

$$\begin{aligned} g_{n+1}(x) &= \mathbb{1}_{\{(x,P_{\mathcal{M}}[g_n](x))\in\Theta\}}(x)P_{\mathcal{M}}[g_n](x) \\ &\quad + (1-\mathbb{1}_{\{(x,P_{\mathcal{M}}[g_n](x))\in\Theta\}}(x))(g_n(x)-\beta P_{\mathcal{M}}[g_n](x)) \\ &= \mathbb{1}_S(x)P_{\mathcal{M}}[g_n](x) \\ &\quad + (1-\mathbb{1}_S(x))(g_n(x)-\beta P_{\mathcal{M}}[g_n](x)) \\ &= g_n - P_{\mathcal{A}}[g_n] - \beta P_{\mathcal{M}}[g_n] + (1+\beta)P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n]. \end{aligned} \tag{HIO}$$

5.1.3 *Other Fienup variants (HPR, RAAR, DM, ADMM)*

This subsection presents some further Fienup variants — only in their global form — that are related to HIO; it does not contain any original results.

The relationships presented below are well-known in phase retrieval (see each particular algorithm below for bibliographic references).

Transformation to Hybrid-Projection-Reflection form. Using the linearity of $P_{\mathcal{A}} = P_{\mathcal{S}}$, HIO can be rewritten in the following manner:

$$\begin{aligned}
&P_{\mathcal{A}}\big((1+\beta)P_{\mathcal{M}}[g_n]-g_n\big)+g_n-\beta P_{\mathcal{M}}[g_n]\\
&\quad=\frac{1}{2}\Big(2P_{\mathcal{A}}\big((1+\beta)P_{\mathcal{M}}[g_n]-g_n\big)+2g_n-2\beta P_{\mathcal{M}}[g_n]\Big)\\
&\quad=\frac{1}{2}\Big(2P_{\mathcal{A}}\big((1+\beta)P_{\mathcal{M}}[g_n]-g_n\big)-(1+\beta)P_{\mathcal{M}}[g_n]+g_n\\
&\qquad\qquad+(1+\beta)P_{\mathcal{M}}[g_n]-g_n+2g_n-2\beta P_{\mathcal{M}}[g_n]\Big)\\
&\quad=\frac{1}{2}\Big(R_{\mathcal{A}}\big((1+\beta)P_{\mathcal{M}}[g_n]-g_n\big)+(1+\beta)P_{\mathcal{M}}[g_n]-g_n+2g_n-2\beta P_{\mathcal{M}}[g_n]\Big)\\
&\quad=\frac{1}{2}\Big(R_{\mathcal{A}}\big((1+\beta)P_{\mathcal{M}}[g_n]-g_n\big)+g_n+(1-\beta)P_{\mathcal{M}}[g_n]\Big)\\
&\quad=\frac{1}{2}\Big(R_{\mathcal{A}}\big(2P_{\mathcal{M}}[g_n]-g_n-(1-\beta)P_{\mathcal{M}}[g_n]\big)+g_n+(1-\beta)P_{\mathcal{M}}[g_n]\Big)\\
&\quad=\frac{1}{2}\Big(R_{\mathcal{A}}\big(R_{\mathcal{M}}[g_n]-(1-\beta)P_{\mathcal{M}}[g_n]\big)+g_n+(1-\beta)P_{\mathcal{M}}[g_n]\Big).
\end{aligned} \tag{HPR}$$

The transformation is valid only if the operator $P_{\mathcal{A}}$ is linear. For non-linear operators $P_{\mathcal{A}}$ this last form is known as HPR (Hybrid-Projection-Reflection) algorithm, first discussed — along with the above transformation — in [BCL03] (see therein for the corresponding local formulation of HPR).

Transformation to Douglas-Rachford form. Setting $\beta = 1$ in HPR, obtain the Relaxed-Reflect-Reflect update

$$g_{n+1}(x) = \frac{1}{2}(g_n + R_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n]). \tag{DR}$$

This update is known the Douglas-Rachford algorithm in convex optimization. It is also known as Averaged-Alternating-Reflection in [Luk05].

The word "relaxed" in the name "Relaxed-Reflect-Reflect" indicates the averaging between g_n and $R_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n]$; the update $g_{n+1} = R_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n]$ could be called "non-relaxed", and the update

$$g_{n+1} = \gamma g_n + (1-\gamma) R_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n]$$

could be called "relaxed with parameter $\gamma \in (0,1)$".

This relaxation can be rewritten as follows:

$$\begin{aligned} g_{n+1} &= \gamma g_n + (1-\gamma) R_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n] \\ &= \gamma g_n + 2(1-\gamma) P_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n] - (1-\gamma) R_{\mathcal{M}}[g_n]) \\ &= \gamma g_n + 2(1-\gamma) P_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n] - 2(1-\gamma) P_{\mathcal{M}}[g_n] + (1-\gamma) g_n \\ &= g_n + \beta (P_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n] - P_{\mathcal{M}}[g_n]), \end{aligned} \tag{β-DR}$$

where $\beta = 2(1-\gamma)$. In particular, [ELB18] uses this variant with $\beta = 0.5$ — albeit with reversed roles of $R_{\mathcal{M}}$ and $R_{\mathcal{A}}$ — as a baseline benchmark algorithm for phase retrieval. The parameter β looks like a time step, but is usually not interpreted as such; see [ELB18, pp. 2440-2441] for a more detailed discussion on β.

Relaxed Averaged Alternating Reflections (RAAR)

The following modification of DR is proposed and discussed in [Luk05]:

$$g_{n+1}(x) = \frac{\beta}{2}(g_n + R_{\mathcal{A}} \circ R_{\mathcal{M}}[g_n]) + (1-\beta) P_{\mathcal{M}}[g_n]. \tag{RAAR}$$

In this algorithm, the relaxation parameter $\beta \in \mathbb{R}$ interpolates between the DR update and $P_{\mathcal{M}}[g_n]$; for $\beta = 1$ this algorithm coincides with DR. Convergence of this algorithm, also known as DRλ, has been recently established for different settings in [LP16] and [LM20].

Difference Map algorithm

A broad class of relaxation strategies can be unified by the DM algorithm, introduced and analyzed in [Els03].

Let $\beta_{\text{DM}} \in \mathbb{R}$ be distinct from zero, let $\gamma_{\mathcal{A}}, \gamma_{\mathcal{M}} \in \mathbb{R}$. Define

$$\begin{aligned} T[g_n] &= g_n + \beta_{\text{DM}} (P_{\mathcal{A}} \circ F_{\mathcal{M}}[g_n] - P_{\mathcal{M}} \circ F_{\mathcal{A}}[g_n]), \quad \text{where} \\ F_{\mathcal{M}}[g_n] &= (1+\gamma_{\mathcal{M}}) P_{\mathcal{M}}[g_n] - \gamma_{\mathcal{M}} g_n; \\ F_{\mathcal{A}}[g_n] &= (1+\gamma_{\mathcal{A}}) P_{\mathcal{A}}[g_n] - \gamma_{\mathcal{A}} g_n. \end{aligned} \tag{DM}$$

If $P_{\mathcal{A}}$ and $P_{\mathcal{M}}$ were to intersect transversally, the choice $\gamma_{\mathcal{M}} = -1/\beta_{\text{DM}}$, $\gamma_{\mathcal{A}} = 1/\beta_{\text{DM}}$ would be locally optimal (contracting to the solution) in an appropriate sense (see [Els03]).

If $P_{\mathcal{A}}$ is linear, then, for $\beta_{\text{DM}} = \beta, \gamma_{\mathcal{A}} = -1, \gamma_{\mathcal{M}} = 1/\beta$ the DM update is reduced to the global form of (HIO):

$$\begin{aligned} T[g_n] &= g_n + \beta \left(P_{\mathcal{A}} \left[\left(1 + \frac{1}{\beta} \right) P_{\mathcal{M}}[g_n] - \frac{1}{\beta} g_n \right] - P_{\mathcal{M}}[g_n] \right) \\ &= g_n - P_{\mathcal{A}}[g_n] + (\beta+1) P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n] - P_{\mathcal{M}}[g_n]. \end{aligned}$$

Alternating Direction Method of Multipliers

Yet another method used in phase retrieval is the Alternating Direction Method of Multipliers (ADMM) [LS19; Wen+12; Els17].

It considers phase problem as the constrained minimization problem:

$$\text{find } p \text{ and } q \text{ in } \mathcal{H} \text{ such that } p = q, \text{ with } p \in \mathcal{A} \text{ and } q \in \mathcal{M},$$

and uses the method of Lagrangian multipliers. One way to define the corresponding augmented Lagrangian is

$$\mathcal{L}(p, q, \lambda) = \langle \lambda, p - q \rangle + \frac{1}{2}\|p - q\|_2^2,$$

where $\lambda \in \mathcal{H}$ is the Lagrangian multiplier associated with the constraint $\|p - q\|_2^2$. The constraints $p \in \mathcal{A}$ and $q \in \mathcal{M}$ are not present in the Lagrangian as they are enforced directly by taking projections (somewhat similar to projected gradient descent methods, cf. Remark 5.16 below). The algorithm is formulated as follows.

Pick $\lambda_0 = 0$ and $q_0 \in \mathcal{M}$, $\beta > 0$. Generate the sequences $(p)_{n \in \mathbb{N}_{>0}}$, $(q)_{n \in \mathbb{N}_{\geqslant 0}}$, $(\lambda)_{n \in \mathbb{N}_{\geqslant 0}}$ by splitting the minimization problem into three consecutive steps:

$$\begin{aligned} p_{n+1} &= \arg\min_{p \in \mathcal{A}} \mathcal{L}(p, q_n, \lambda_n); \\ q_{n+1} &= \arg\min_{q \in \mathcal{M}} \mathcal{L}(p_{n+1}, q, \lambda_n); \\ \lambda_{n+1} &= \lambda_n + \beta(p_{n+1} - q_{n+1}). \end{aligned}$$

The minimizers of the first two steps can be calculated explicitely. For the first step the expression

$$\arg\min_{p \in \mathcal{A}} \mathcal{L}(p, q_n, \lambda_n) = \arg\min_{p \in \mathcal{A}} \|p - (q_n - \lambda_n)\|_2^2 - \frac{1}{2}\|\lambda\|_2^2$$

is minimized for $p = P_{\mathcal{A}}[q_n - \lambda_n]$ by definition of a projector. With a similar equation for the second step, the overall algorithm states

$$\begin{aligned} p_{n+1} &= P_{\mathcal{A}}[q_n - \lambda_n]; \\ q_{n+1} &= P_{\mathcal{M}}[p_{n+1} + \lambda_n]; \\ \lambda_{n+1} &= \lambda_n + \beta(p_{n+1} - q_{n+1}). \end{aligned}$$

Choose $\beta = 1$ and $g_{n+1} = p_{n+2} + \lambda_{n+1}$ to recover the (DR) update:

$$\begin{aligned} g_{n+1} &= p_{n+2} + \lambda_{n+1} \\ &= P_{\mathcal{A}}[q_{n+1} - \lambda_{n+1}] + \lambda_{n+1} \\ &= P_{\mathcal{A}}[\underbrace{q_{n+1}}_{=P_{\mathcal{M}}[g_n]} - \underbrace{(\lambda_n + p_{n+1})}_{=g_n} + q_{n+1}] + \underbrace{\lambda_n + p_{n+1}}_{=g_n} - q_{n+1} \\ &= P_{\mathcal{A}}[2P_{\mathcal{M}}[g_n] - g_n] + g_n - P_{\mathcal{M}}[g_n]. \end{aligned}$$

Note that linearity of $P_{\mathcal{A}}$ is not needed for this transformation, and the relationship persists if the roles of $P_{\mathcal{A}}$ and $P_{\mathcal{M}}$ are exchanged.

We refer the reader to the recent survey [LS19] for more information on the connection between ADMM and DR. The setup of ADMM exhibits certain similarities to the DR description in Chapter 9, and to the DR variant used in [LP16] to analyze local convergence of DR, see Remark 9.23.

Remark 5.10. Given the non-convex nature of phase retrieval, the overall numerical success of certain Fienup variants is remarkable. The variety of proposed heuristic variants with no clear favorite indicates that there may be an underlying reason for this success. Since there is no obvious way to pinpoint the exact reason, this work focuses on the two arguably simplest algorithms used in phase retrieval: on the ER algorithm and on DR formulation of HIO.

5.2 ERROR-REDUCTION AS A SUBDIFFERENTIAL FLOW

This section discusses the alternating projections (AP) algorithm, its phase retrieval variant (ER), and some of their properties from [Fie82]. The section continues with Remark 5.17, showing that ER corresponds to a selection of the (Mordukhovich-Kruger) generalized subdifferential flow (which is also a formal gradient flow) we call ERF. While closely related to results developed in [BL03], this appears to be a novel observation, and it is essential to the discussion in Chapter 7.

The following points are essential to the explored connection between AP and APF:

- For weakly closed $\mathcal{X}, \mathcal{Y}$, APF is a selection of the generalized subdifferential flow of $E_{\mathcal{X}} + E_{\mathcal{Y}}$.
- For proximal $\mathcal{X}, \mathcal{Y}$, APF is a formal gradient flow (Fréchet-derivative is not necessarily well-defined).
- As will be established later, APF dissipates energy very similarly to any rigorously defined gradient flow (Proposition 6.4, Corollary 7.21).

DEFINITION 5.11 (ALTERNATING PROJECTIONS). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal with projecton selections $P_{\mathcal{X}}, P_{\mathcal{Y}}$, let let $g_0 \in \mathcal{X}$. The sequence $(g_n)_{n \in \mathbb{N}_0}$ is generated by the AP algorithm with initial value g_0, if*

$$g_{n+1} = P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g_n] \tag{AP}$$

for all $n \in \mathbb{N}_0$. In this case, we call $(g_n)_{n \in \mathbb{N}_0}$ an AP sequence with initial value g_0.

DEFINITION 5.12 (ERROR-REDUCTION ALGORITHM). *Let $g_0 \in \mathcal{A}$ for the proximal additional constraint $\mathcal{A} \subset \mathcal{H}$. The sequence $(g_n)_{n \in \mathbb{N}_0}$ is generated by the ER algorithm, if*

$$g_{n+1} = P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n] \tag{ER}$$

for all $n \in \mathbb{N}_0$. In this case, we call $(g_n)_{n \in \mathbb{N}_0}$ an ER sequence with initial value g_0.

Alternating projections algorithm dates back to 1870's: in [Sch70], it was used to solve the Dirichlet problem for the Laplace equation on a composite domain by alternating between solutions on simpler subdomains. AP is used in a variety of applications, such as computer tomography, Navier-Stokes equations, pattern reckognition, image restoration, and others [ER11]. A generalized variant of AP — where projections are replaced by proximal mappings — is called PPA (proximal point algorithm). In diffraction imaging, alternating projections is also known as Gerchberg-Saxton algorithm [GS72] and as Error-Reduction algorithm [Fie82].

When both sets $\mathcal{X}, \mathcal{Y}$ are convex, AP is well-undrestood; this case was essentially settled in a paper by Bauschke and Borwein [BB93] (see, e. g., [BB96] for a survey on the matter).

The general non-convex case still poses open questions. A local convergence result was proven by Combettes and Trussel in [CT90]: it demonstrates that if sequences $(P_{\mathcal{X}}[g_n])_n$ and $(P_{\mathcal{Y}}[g_n])_n$ are bounded and if $P_{\mathcal{X}}[g_n] - P_{\mathcal{Y}}[g_n] \to 0$, the set of accumulation points of (g_n) converges to a singleton or a nontrivial compact continuum. An example from [BN13] demonstrated that the case of a nontrivial compact continuum can occur. Further papers — such as [LLM09] and [Bau+13] among others — investigated, under which additional assumptions on the sets $\mathcal{X}$ and $\mathcal{Y}$ one can establish local convergence.

Error-Reduction algorithm was shown to converge locally to a solution in finite-dimensional case in [NR16] and [Pau+18]; the proofs exploited the Łojasiewicz inequality.

The success of alternating projection in applied diffraction imaging indicates that the convergence radius of phase retrieval, is, in practice, much better than one could expect for a such high-dimensional problem. In X-ray applications, the difficulty of phase retrieval seems to be more tightly linked to properties of the solution rather than the discretization dimension of the problem [ELB18].

It is therefore natural to ask, to which extent inherent properties of the set $\mathcal{M}$ and object space information sets — such as $\mathcal{P}, \mathcal{S}, \mathcal{T}_s(\nu)$ — affect the performance of Error-Reduction in high-dimensional spaces.

While there exist remarkable results on stability of generalized phase retrieval, primary in context of frame theory — see, e. g., [GKR20] for a recent survey — to the extent of our knowledge, there are no results concerning local convergence of Error-Reduction in an an infinite-dimensional space; nor an explicit example, akin to the one in [BN13], demonstrating the compact continuum case.

One of the goals of this work is to provide a new perspective for analysis of Error-Reduction algorithm, see Chapter 7. The following discussion of ER will be relevant in that regard.

Remark 5.13. By definition, if the sequence $(g_n)_{n\in\mathbb{N}}$ is generated by ER, then $g_n \in \mathcal{A}$ for all $n \in \mathbb{N}$. In particular, one can use the functional $g_n \mapsto E_{\mathcal{M}}[g_n]$ to track the progress of the algorithm, since $E_{\mathcal{M}}[g_n] = 0$ implies $g_n \in \mathcal{M}$ and thus belongs to $\mathcal{A} \cap \mathcal{M}$.

Note that for any $\varepsilon > 0$ the condition $E_{\mathcal{M}}[g_n] < \varepsilon$ does not necessarily imply that the distance from g to $\mathcal{A} \cap \mathcal{M}$ is small. (For this to be true, one must assume that the intersection $\mathcal{A} \cap \mathcal{M}$ is regular; see, for examlpe, how regularity is used in Definition 9.6 and Proposition 9.7.) Nonetheless, in applications $E_{\mathcal{M}}$ is used as a termination criterion for ER, and the quality of the obtained solution could be additionaly judged by other means.

The following result of [Fie82] gives name to the Error-Reduction algorithm.

Proposition 5.14 (Fienup [Fie82]: ER does not increase error). *Let $(g_n)_{n\in\mathbb{N}_0}$ be an ER sequence with initial value g_0. Then, for all $n \in \mathbb{N}_0$,*

$$E_{\mathcal{M}}[g_{n+1}] \leqslant E_{\mathcal{M}}[g_n].$$

Proof.

$$\begin{aligned} E_{\mathcal{M}}[g_{n+1}] &= \frac{1}{2}\|g_{n+1} - P_{\mathcal{M}}[g_{n+1}]\|_2^2 \overset{(*)}{\leqslant} \frac{1}{2}\|g_{n+1} - P_{\mathcal{M}}[g_n]\|_2^2 \\ &= \frac{1}{2}\|P_{\mathcal{A}}[P_{\mathcal{M}}[g_n]] - P_{\mathcal{M}}[g_n]\|_2^2 \overset{(**)}{\leqslant} \frac{1}{2}\|P_{\mathcal{A}}[g_n] - P_{\mathcal{M}}[g_n]\|_2^2 \\ &= \frac{1}{2}\|g_n - P_{\mathcal{M}}[g_n]\|_2^2 = E_{\mathcal{M}}[g_n]. \end{aligned}$$

In $(*)$, use Corollary 3.6 with $g = g_{n+1}$, $f = g_n$, $P_{\mathcal{X}} = P_{\mathcal{M}}$. In $(**)$, use Corollary 3.6 with $g = P_{\mathcal{M}}[g_n]$, $f = g_n$, $P_{\mathcal{X}} = P_{\mathcal{A}}$. In the second to last equality, use $g_n = P_{\mathcal{A}}[g_n]$, which holds by definition of g_n. □

Remark 5.15. This proposition holds for a generic AP sequence $(g_n)_{n\in\mathbb{N}_0}$ and energy functional $E_{\mathcal{Y}}$ by the same argument.

Remark 5.16 (ER is a projected gradient descent [Fie82]). Let $(g_n)_{n\in\mathbb{N}_0}$ be an ER sequence with initial value g_0. Let $f_0 = g_0$, let $(f_m)_{m\in\mathbb{N}_0/2}$ be a sequence generated by the following rule:

$$\begin{aligned} f_{n+1/2} &= f_n - \varepsilon\nabla E_{\mathcal{M}}[f_n] = f_n - \varepsilon(f_n - P_{\mathcal{M}}[f_n]) = P_{\mathcal{M}}[f_n]; \\ f_{n+1} &= P_{\mathcal{A}}[f_{n+1/2}] \end{aligned}$$

for step size $\varepsilon = 1$ for all $n \in \mathbb{N}_0$. (The gradient $\nabla E_{\mathcal{M}}$ is taken in the formal sense.) Then, $f_n = g_n$ for all $n \in \mathbb{N}_0$.

This means that ER is a projected gradient descent algorithm, where explicit gradient descent of the energy $E_{\mathcal{M}}$ with step size $\varepsilon = 1$ is followed by explicit projection onto the constraint set $\mathcal{A}$. Note that the fact that $g_0 \in \mathcal{A}$ is not essential for the equivalence of ER and this projected gradient scheme.

The following argument appears to be new; it establishes that ER can be viewed as a gradient flow algorithm with respect to the combined energy $E_{\mathcal{M}}[g] + E_{\mathcal{A}}[g]$.

Remark 5.17 (ER is a formal gradient descent). Recall that formal derivatives of $E_{\mathcal{M}}$ and $E_{\mathcal{A}}$ are given by $g - P_{\mathcal{M}}[g]$ and $g - P_{\mathcal{A}}[g]$ respectively. Let $(g_n)_{n\in\mathbb{N}_0}$ be an ER sequence with initial value $g_0 \in \mathcal{A}$. Let $f_0 = g_0$, let $(f_m)_{m\in\mathbb{N}_0/2}$ be a sequence generated by the following rule:

$$f_{m+1/2} = f_m - \varepsilon\nabla(E_{\mathcal{M}}[f_m] + E_{\mathcal{A}}[f_m])$$

for step size $\varepsilon = 1$ and for all $m \in \mathbb{N}_0/2$. Then, $f_n = g_n$ for all $n \in \mathbb{N}_0$.

Indeed, since $f_0 \in \mathcal{A}$,

$$f_{1/2} = f_0 - \nabla E[f_0] = f_0 - (f_0 - P_{\mathcal{M}}[f_0]) - \underbrace{(f_0 - P_{\mathcal{A}}[f_0])}_{=0} = P_{\mathcal{M}}[f_0],$$

and since $f_{1/2} \in \mathcal{M}$,

$$f_1 = f_{1/2} - \nabla E[f_{1/2}] = f_{1/2} - \underbrace{(f_{1/2} - P_{\mathcal{M}}[f_{1/2}])}_{=0} - (f_{1/2} - P_{\mathcal{A}}[f_{1/2}]) = P_{\mathcal{A}}[f_{1/2}].$$

Overall, $f_1 = P_{\mathcal{A}}[P_{\mathcal{M}}[f_0]] = g_1$, and the argument can be repeated inductively to show $f_n = g_n$ for $n \in \mathbb{N}_0$.

This means that ER can be viewed as two consecutive updates of a (formal) gradient flow. This observation motivates a closer study of the equation

$$\begin{aligned} \partial_t g_t &= -\nabla(E_{\mathcal{M}}[g_t] + E_{\mathcal{A}}[g_t]) \\ &= -(g_t - P_{\mathcal{M}}[g_t]) - (g_t - P_{\mathcal{A}}[g_t]), \end{aligned} \tag{ERF}$$

where the first line is formal since $\nabla(E_{\mathcal{M}} + E_{\mathcal{A}})[g]$ is not necessarily well-defined for all $g \in \mathcal{H}$, but the second line presents a rigorously well-defined integro-differential equation we call ER Flow, investigated in more detail in Chapter 7.

Remark 5.18 (ER is a rigorous subdifferential descent selection). Let $\mathcal{H} = L^2(\Omega)$ for some bounded measurable $\Omega \subset \mathbb{R}^d$, so that $\mathcal{M}$ is weakly closed (Section 3.3). Assume that the additional constraint $\mathcal{A}$ is weakly closed. Then, by Theorem 4.18 holds $\partial_{\text{KM}} E_{\mathcal{A}}[g] = g - \Pi_{\mathcal{A}}[g]$ and $\partial_{\text{KM}} E_{\mathcal{M}}[g] = g - \Pi_{\mathcal{M}}[g]$. Further, by [MS96, Thm. 4.1],

$$\partial_{\text{KM}}(E_{\mathcal{M}} + E_{\mathcal{A}})[g] \subseteq \partial_{\text{KM}} E_{\mathcal{M}}[g] + \partial_{\text{KM}} E_{\mathcal{A}}[g]$$

(where $E_{\mathcal{M}}, E_{\mathcal{A}}$ are normally compact since they are Lipschitz, see Lemma 4.5 and the introductory text in Section 4 of [MS96]).

Thus, analogously to Remark 5.17, but rigorously,

$$-(g_t - \Pi_{\mathcal{M}}[g_t]) - (g_t - P_{\mathcal{A}}[g_t]) \subseteq -\partial_{\text{KM}}(E_{\mathcal{M}} + E_{\mathcal{A}})[g_t],$$

meaning that ERF is a selection of the subdifferential of $E_{\mathcal{M}} + E_{\mathcal{A}}$.

Similarly, the same argument can be applied to any weakly closed sets $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ on any separable Hilbert space $\mathcal{H}$.

Remark 5.19 (Composition of projecton selections is not a derivative). The connection between ER and a gradient of a functional is not obvious from the update $g_{n+1} = P_{\mathcal{A}} \circ P_{\mathcal{M}}[g_n]$. To our knowledge, there does not exist a functional $F[g]$ with the derivative of the form $g - P_{\mathcal{A}} \circ P_{\mathcal{M}}[g]$. For example, formal derivative of the functional $F[g] := \frac{1}{2}\|g - P_{\mathcal{A}} \circ P_{\mathcal{M}}[g]\|_2^2$ contains artifacts arising from the Fourier transform that is present in $P_{\mathcal{M}}$.

Remark 5.20 (Non-differentiability of $E_{\mathcal{M}}$). The functional $E_{\mathcal{M}}$ is non-differentiable at points f where $\hat{f}(k) = 0$ for at least one (Lebesgue-point) $k \in \operatorname{supp} \sqrt{I}$ (Remark 4.15). These are precisely the points where the multiplicity of the operator $P_{\mathcal{M}}$ needs to be resolved explicitely (Remark 3.26) by choice of the multiplicity resolution phase φ.

While it is common to pick $\varphi \equiv 0$ for most applications, it is not clear whether a more efficient choice is possible. In an idealized scenario, such choice could have a dramatic effect on the reduction of phase retrieval complexity.

Indeed, let g be a non-negative solution to a phase problem with $|\hat{g}| = \sqrt{I}$, with phase of $\hat{g}$ being equal to φ_g. If one were to set the multiplicity resolution phase in $P_{\mathcal{M};\varphi}$ to $\varphi = \varphi_g$ and choose the starting value 0, then $P_{\mathcal{M};\varphi_g}[0] = g$ would be a solution of the phase problem. In the energy minimization framework, the gradient flow starting at 0 with multiplicity resolution phase φ_g would (trivially) converge to g, cf. Example 7.23.

Thus, in an extreme case, choosing the "correct" multiplicity resolution phase φ is exactly the same as solving phase retrieval.

Resolving phase at non-differentiable points is highly relevant for applications. This is not obvious: at a first glance, it is very unlikely that $\hat{f}_n(k)$ equals exactly 0 at any pixel k due to computational artifacts. However, while $\hat{f}_n(k) = 0$ is very unlikely, the case when $\hat{f}_n(k)$ is small (meaning that

$$C_{\mathcal{F}}|\hat{f}_n(k)|^2 \ll \|f_n - f_{n-1}\|_2^2$$

for two consecutive iterates, keeping in mind that the correlation between phases $\hat{f}_{n+1}(k)$ and $\hat{f}_n(k)$ is non-local) is abundantly present in applications.

This case — stemming from discontinuity of $P_{\mathcal{M}}$ — is a phase retrieval issue that has not received much attention in literature.

However, if a subtle choice of φ is found, it can yield notable benefits for applications. Such choice can be used in applications through an appropriate regularization of $E_{\mathcal{M}}$ at non-differentiable points.

5.3 CONNECTION BETWEEN DR ALGORITHM AND ER FLOW

This section discusses the relationship between the Douglas-Rachford algorithm in the setting of maximal monotone operators (DR-LM) and its phase retrieval variant (DR-HIO). Notably, we show that the relationship

$$\text{evolution equation} \quad \leftrightsquigarrow \quad \text{Douglas-Rachford algorithm}$$

that is well-known for the convex case, persists for phase retrieval, albeit with some modifications, cf. Figure 5.1.

The Douglas-Rachford algorithm was originally introduced in [DR56] in connection with non-linear heat flow problems. A modern form of the algorithm, established in [LM79] in the setting of maximal monotone operators, looks quite different from the original. The connection between these two forms lies beyond the scope of this work; the reader is refered to [Mou16, Chapter 5.6] for a detailed discussion on the matter.

The research of DR-LM for the non-convex case has attracted considerable interest in the recent years, see [LS19, Sec. 3] for a survey of applications and results. Nevertheless, the questions connected to remarkable success of DR and its variants for the phase retrieval case remain, to the extent of our knowledge, open. Notably, such questions include: i) computationally observed global convergence (for "generic" cases); and ii) connection between the difficulty of a specific phase retrieval instance and expected number of iterations until an approximate solution is found, cf. [ELB18].

The section starts with a more detailed outline of various DR-LM aspects. It proceeds by recalling some standard results from convex analysis. Thereafter, DR-LM is discussed in three following contexts.

i) The formulation DR-LM in the setting of maximal monotone operators from [LM79] is presented in Section 5.3.3.

ii) In the setting of convex feasibility problems, DR-LM can be simplified to the form DR-cf, presented in Section 5.3.4.

iii) In the phase retrieval setting, DR-LM can be simplified to the form DR-HIO, presented in Section 5.3.5.

Figure 5.1 compares the setting of different Douglas-Rachford variants. For more information on DR-LM and DR-cf, see [BC17], [LS19]. The established connection between DR-LM and DR-HIO appears to be new. Along with the previously established Remark 5.17, it motivates a closer study of ERF.

5.3.1 *From functional minimization to DR*

Before formally writing down the mathematical arguments, let us outline: i), ii), iii) which problems can be addressed by DR-LM; iii) a sufficient condition for the convergence of DR-LM; iv), v) how one can simplify DR-LM for feasibility problems.

i) Consider the task of minimizing the sum $F_A + F_B$ of two Fréchet-differentiable functionals F_A, F_B. It is connected to the study of the corresponding gradient flow $\partial g_t = -\nabla(F_A[g] + F_B[g])$. DR-LM can be used to search for the extremal points g at which $0 = \nabla(F_A[g] + F_B[g])$.

ii) Consider the task of minimizing the sum $F_A + F_B$ of two convex (but not necessarily differentiable) functionals F_A, F_B. It is connected to the study of the multivalued evolution equation $\partial_t g \in -\partial_c(F_A + F_B)$, where $\partial_c(F_A + F_B)$ is the convex subdifferential of $F_A + F_B$. DR-LM can be used to search for the extremal points g at which $0 \in -\partial_c(F_A[g] + F_B[g])$.

iii) Generally, DR-LM is not necessarily connected to a minimization problem, but can be used to find extremal points of any evolution equation $\partial_t g \in -(A[g] + B[g])$, where A, B are maximal monotone operators.

This equation does not imply case i) — gradients of Fréchet-differentiable functionals are not necessarily monotone — but it implies case ii): if two convex functionals F_A, F_B are proper lower semicontinuous and map $\mathcal{H}$ to $(-\infty, \infty]$, then their subdifferentials $A = \partial_c F_A$, $B = \partial_c F_B$ are maximal monotone.

In this maximal monotone case, DR-LM is formulated in terms of resolvents of A, B. Note that an operator being maximal monotone is equivalent to its resolvent being firmly nonexpansive. This is a crucial assumption for the weak convergence of DR-LM, the proof of which relies on a fixed point argument ([LM79]).

iv) If the problem in question is a convex feasibility problem — i. e. if one searches for the intersection of two weakly closed convex sets $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ — the corresponding energy functionals $F_A = E_{\mathcal{X}}, F_B = E_{\mathcal{Y}}$ are convex, and the subgradient evolution equation has the form

$$\partial_t g \in -N_{\mathcal{X}}[g] - N_{\mathcal{Y}}[g],$$

where $N_{\mathcal{X}}$ is the normal cone to $\mathcal{X}$ at g, and likewise for $\mathcal{Y}$. Further, for convex $\mathcal{X}$, the resolvent of the normal cone $N_{\mathcal{X}}$ is given by $P_{\mathcal{X}}$, and likewise for $\mathcal{Y}$. Therefore, one can explicitely calculate the resolvents and simplify generic form DR-LM to obtain the

	Energy minimization	(Multivalued) Equation	DR update
1)	—	$0 \in -A[g] \quad -B[g]$	$g + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g] - J_{\lambda B}[g]$
2a)	$E_{\mathcal{X}}[g] + E_{\mathcal{Y}}[g]$	$\partial_t g = -(g - P_{\mathcal{X}}[g]) - (g - P_{\mathcal{Y}}[g])$	$g + P_{\mathcal{X}} \circ (2P_{\mathcal{Y}} - \mathrm{Id})[g] - P_{\mathcal{X}}[g]$
2b)	$e^{\infty}_{\mathcal{X}}[g] + e^{\infty}_{\mathcal{Y}}[g]$	$0 \in -N_{\mathcal{X}}[g] \quad -N_{\mathcal{Y}}[g]$	
3)	$E_{\mathcal{M}}[g] + E_{\mathcal{P}}[g]$	$\partial_t g = -(g - P_{\mathcal{M}}[g]) - (g - P_{\mathcal{P}}[g])$	$g + P_{\mathcal{M}} \circ (2P_{\mathcal{P}} - \mathrm{Id})[g] - P_{\mathcal{P}}[g]$

Figure 5.1: From minimization problems to specific variants of DR
1) Generally, (DR-LM) does not necessarily correspond to a minimization problem. It searches for zeros of sums of two maximal monotone operators A, B, and converges under certain assumptions. See Section 5.3.3.
2) For convex feasibility problems, one can consider quadratic energy functional minimization — as in 2a) — or infinite well functional minimization — as in 2b). Take subdifferential to derive equations such the ones presented in the table. The corresponding resolvents are single-valued and lead to the presented update operator. Convergence results for (DR-LM) apply. See Section 5.3.4.
3) Phase problem can be formulated as energy minimization problem. The energy is not convex; instead of taking the subdifferential, one can take the (Kruger-Mordukhovich) generalized subdifferential to motivate the corresponding evolution equation. The subdifferential exists if both sets are weakly closed, cf. Theorem 4.18, Remark 5.18. The resulting equation admits global weak solutions, as shown in Chapter 7. The corresponding resolvents are not necesserily single-valued and lead, among others, to the presented update operator. See Section 5.3.5.

convex form DR-cf; the update then essentially coincides with the form DR.

Note that one has $g - P_{\mathcal{X}}[g] \in N_{\mathcal{X}}[g]$, and likewise for $\mathcal{Y}$. This means that one particular selection of the evolution equation has the form we call APF (AP Flow)

$$\partial_t g = -(g - P_{\mathcal{X}}[g]) - (g - P_{\mathcal{Y}}[g]).$$

(See [BC17], [LS19].)

v) The main point of this section is the following: if one considers the evolution equation

$$\partial_t g = -(g - P_{\mathcal{P}}[g]) - (g - P_{\mathcal{M}}[g]),$$

— note that $g - P_{\mathcal{M}}[g]$ is not maximal monotone, so the setting of iii) does not apply; and $\mathcal{M}$ is not convex, so the setting of iv) does not apply; — it is still possible to explicitely calculate selections of resolvents and simplify DR-LM to the form DR-HIO.

5.3.2 *Background from convex analysis*

This subsection briefly recalls some well-known relevant notions and results from convex analysis — notably, such notions as normal cones, (firmly) nonexpansive and (maximally) monotone multivalued operators and their resolvents.

A much more thorough systematic discussion on these topics may be found, e. g., in [BC17]; see, in particular, Chapters 4, 16, 20, 23, and 26.3 therein.

To the reader familiar with notions of convex analysis, of interest could be Case 3) of Example 5.26 that highlights some monotone operators that appear in phase retrieval.

Let $\mathcal{D} \subset \mathcal{H}$; a map $A\colon \mathcal{D} \rightrightarrows \mathcal{H}$, is called multivalued (set-valued), if it maps an element of $\mathcal{D}$ to a subset of $\mathcal{H}$. For $f, g \in \mathcal{D}$, define

$$A[f] + g := \{a_f + g \mid \text{ for all } a_f \in A[f]\}.$$

Additionally, for $B\colon \mathcal{D} \rightrightarrows \mathcal{H}$ define

$$A[f] + B[g] := \{a_f + b_g \mid \text{for all } a_f \in A[f], b_g \in B[g]\}.$$

For a functional $F\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, its domain is defined as

$$\operatorname{domain} F = \{g \in \mathcal{H} \mid F[g] < \infty\}.$$

The functional F is called proper, if domain $F \neq \varnothing$. The subdifferential of a proper functional F is a multivalued mapping

$$\begin{aligned} &\partial_c F\colon \mathcal{H} \rightrightarrows \mathcal{H} \\ &g \mapsto \{v \in \mathcal{H} \mid F[g] + \langle v, f - g\rangle \leqslant F[f] \text{ for all } f \in \mathcal{H}\}. \end{aligned} \tag{5.3}$$

It is common to consider subdifferentials of convex functionals, as for convex functionals there exist results that ensure that $\partial_c F[g]$ is not empty (see, e. g. [BC17, Prop. 16.17, Prop. 16.20]).

Proposition 5.21 (Subdifferentials of convex energy functionals). *Let $\mathcal{C} \subset \mathcal{H}$ be nonempty, closed, convex, let $E_\mathcal{C}$ be the corresponding energy functional. Then, $g - P_\mathcal{C}[g] \subset \partial_c E_\mathcal{C}[g]$ for all $g \in \mathcal{H}$.*

Proof. It is necessary to show that for all $f \in \mathcal{H}$

$$E_\mathcal{C}[g] + \langle g - P_\mathcal{C}[g], f - g\rangle \leqslant E_\mathcal{C}[f].$$

Multiply both sides by 2 and add $\|f - g\|_2^2$ to obtain the equivalent expression

$$\underbrace{\|g - P_\mathcal{C}[g]\|_2^2 + 2\langle g - P_\mathcal{C}[g], f - g\rangle + \|f - g\|_2^2}_{= \|f - P_\mathcal{C}[g]\|_2^2} \leqslant \|f - P_\mathcal{C}[f]\|_2^2 + \|f - g\|_2^2 \tag{5.4}$$

illustrated in Figure 5.2.

If $g \in \mathcal{C}$, Equation (5.4) is reduced to $\|f-g\|_2^2 \leqslant \|f-P_{\mathcal{C}}[f]\|_2^2 + \|f-g\|_2^2$, which is always satisfied.

If $g \notin \mathcal{C}$ but $f \in \mathcal{C}$, then by angle property (Lemma 3.10) holds

$$\langle g - P_{\mathcal{C}}[g], P_{\mathcal{C}}[f] - P_{\mathcal{C}}[g]\rangle = \langle g - P_{\mathcal{C}}[g], f - P_{\mathcal{C}}[g]\rangle \leqslant 0$$

which — adding $-\|g-P_{\mathcal{C}}[g]\|_2^2 - \|f-P_{\mathcal{C}}[g]\|_2^2$ on both sides — is equivalent to

$$\begin{aligned} \|g-f\|_2^2 &\geqslant \|g-P_{\mathcal{C}}[g]\|_2^2 + \|f-P_{\mathcal{C}}[g]\|_2^2 \quad \Rightarrow \\ \|f-g\|_2^2 + \underbrace{\|f-P_{\mathcal{C}}[f]\|_2^2}_{=0} &\geqslant \|g-P_{\mathcal{C}}[g]\|_2^2 + \|f-P_{\mathcal{C}}[g]\|_2^2 \geqslant \|f-P_{\mathcal{C}}[g]\|_2^2, \end{aligned}$$

so that Equation (5.4) is demonstrated.

If $f = g$, Equation (5.4) is reduced to $\|g-P_{\mathcal{C}}[g]\|_2^2 \leqslant \|g-P_{\mathcal{C}}[g]\|_2^2 + 0$, which is always satisfied.

Finally, let $f, g \notin \mathcal{C}$ with $f \neq g$. The following proof is easy to summarize (see Figure 5.2), but somewhat technical when written down. Let $p_f := P_{\mathcal{C}}[f], p_g := P_{\mathcal{C}}[g]$, let l denote the line containing the segment between p_f and p_g. Define the altitude from point f onto line l as

$$f^{\perp} := p_g + t^{\perp}(p_g - p_f), \quad t^{\perp} := \frac{\langle f - p_g, p_g - p_f\rangle}{\|p_g - p_f\|_2^2}.$$

It is easy to verify that $f - f^{\perp}$ is orthogonal to $p_g - p_f$:

$$\begin{aligned} \langle f - f^{\perp}, p_g - p_f\rangle &= \langle f - p_g - t^{\perp}(p_g - p_f), p_g - p_f\rangle \\ &= \langle f - p_g, p_g - p_f\rangle - \langle f - p_g, p_g - p_f\rangle = 0, \end{aligned}$$

and that $f^{\perp}$ belongs to line l. Let us show that

$$\|f - p_g\|_2^2 = \underbrace{\|f - f^{\perp}\|_2^2}_{\overset{(*)}{\leqslant} \|f-p_g\|_2^2} + \underbrace{\|p_g - f^{\perp}\|_2^2}_{\overset{(**)}{\leqslant} \|g-f\|_2^2}.$$

The equality is true by Pythagoras' theorem. Also by Pythagoras, the point $f^{\perp}$ is closer to f than any other point on l (since $f - f^{\perp}$ is an altitude to l); hence, the inequality $(*)$ is true. Finally, the inequality $(**)$ follows from the Lemma 5.22 below that states that for two diverging rays — here, the ray $P_{\mathcal{C}}[f] + t(f - P_{\mathcal{C}}[f]), t \in [0, \infty)$ and the ray $P_{\mathcal{C}}[g] + t(g - P_{\mathcal{C}}[g]), t \in [0, \infty)$; these rays are diverging by the angle property for convex sets (Lemma 3.10) — two closest points on the rays are the origins of the rays. □

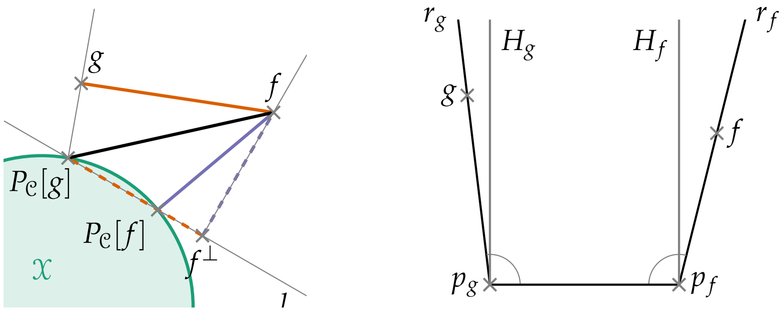

Figure 5.2: Illustration to Proposition 5.21 and Lemma 5.22 (subdifferentials of convex energy functionals).

On the left: to show that

$$\|f - P_{\mathcal{C}}[g]\|_2^2 \leqslant \|f - P_{\mathcal{C}}[f]\|_2^2 + \|f - g\|_2^2 ,$$

construct an altitude $f^{\perp}$ from the point f to the line l that goes through $P_{\mathcal{C}}[g]$ and $P_{\mathcal{C}}[f]$, and use

$$\|f - f^{\perp}\|_2 \leqslant \|f - P_{\mathcal{C}}[f]\|_2 \quad \text{and}$$
$$\|P_{\mathcal{C}}[g] - f^{\perp}\|_2 \leqslant \|f - g\|_2 . \tag{5.5}$$

Then, by Pythagoras' theorem,

$$\|f - P_{\mathcal{C}}[g]\|_2^2 = \|f - f^{\perp}\|_2 + \|P_{\mathcal{C}}[g] - f^{\perp}\|_2 \leqslant \|f - P_{\mathcal{C}}[f]\|_2^2 + \|f - g\|_2^2 .$$

On the right: Equation (5.5) is shown in Lemma 5.22. If two rays r_f and r_g are diverging — meaning that angles $\angle f p_f p_g$ and $\angle g p_g p_f$ are obtuse — the distance between any two points on the rays is not smaller than the distance between half-planes H_f and H_g, which equals $\|p_f - p_g\|_2$.

LEMMA 5.22. *Distance between two diverging rays in a Hilbert space is smallest at the origins of the rays. I. e., let points $f, g, p_f, p_g \in \mathcal{H}$; let*

$$r_f = \{p_f + t(f - p_f) \mid t \geqslant 0\};$$
$$r_g = \{p_g + t(g - p_g) \mid t \geqslant 0\}$$

denote two rays with origins in p_f and p_g. Assume that the rays are diverging, meaning that angles $\angle f p_f p_g$ and $\angle g p_g p_f$ are obtuse:

$$\langle f - p_f, p_g - p_f \rangle \leqslant 0; \qquad \langle g - p_g, p_f - p_g \rangle \leqslant 0.$$

Then,

$$\inf_{\substack{\tilde{f} \in r_f \\ \tilde{g} \in r_g}} \|\tilde{f} - \tilde{g}\|_2^2 = \|p_f - p_g\|_2^2.$$

Proof. Define the affine half-planes

$$H_f = p_f + \{v \mid v \in \mathcal{H}, \langle p_f - p_g, v \rangle \geqslant 0\},$$
$$H_g = p_g + \{v \mid v \in \mathcal{H}, \langle p_g - p_f, v \rangle \geqslant 0\},$$

cf. Figure 5.2. It is easy to verify that the distance between the half-planes equals $\|p_f - p_g\|_2^2$:

$$\begin{aligned}\inf_{\substack{\tilde{f} \in H_f \\ \tilde{g} \in H_g}} \|\tilde{f} - \tilde{g}\|_2^2 &= \inf_{S_{HH}} \|p_f + v_f - p_g - v_g\|_2^2 \\ &= \inf_{S_{HH}} \|p_f - p_g\|_2^2 + 2\langle p_f - p_g, v_f - v_g \rangle + \|v_f - v_g\|_2^2 \\ &= \inf_{S_{HH}} \|p_f - p_g\|_2^2 + 2\underbrace{\langle p_f - p_g, v_f \rangle}_{\geqslant 0} + 2\underbrace{\langle p_g - p_f, v_f \rangle}_{\geqslant 0} + \|v_f - v_g\|_2^2 \\ &\geqslant \|p_f - p_g\|_2^2,\end{aligned}$$

where S_{HH} is the set of all pairs $(v_f, v_g) \in \mathcal{H} \times \mathcal{H}$ such that such that $\langle p_f - p_g, v_f \rangle \geqslant 0$, and $\langle p_g - p_f, v_g \rangle \geqslant 0$.

The ray r_f is contained in H_f, since for any point on the ray $p_f + t(f - p_f)$ holds

$$\langle p_f - p_g, t(f - p_f) \rangle = -t \langle f - p_f, p_g - p_f \rangle \geqslant 0.$$

Similarly, r_g is contained in H_g.

Therefore, the distance between the rays is not less than $\|p_f - p_g\|_2$, and the claim follows since $p_f \in r_f$ and $p_g \in r_g$. □

The following definition denotes a multivalued operator with the letter T — instead of A, as before — for the following didactical reason. In this section, we discuss multivalued operators that can be split into two groups: operators that appear in evolution equations (A, B, $I - P_{\mathcal{C}}$, N for normal cones, typically monotone), and operators that appear in the Douglas-Rachford algorithm (T, $P_{\mathcal{C}}$, J for resolvents, typically non-expansive). The choice of different letters aims to ease the distinction between the groups.

Definition 5.23 (Firmly nonexpansive and nonexpansive operators). *Let $\mathcal{D} \subseteq \mathcal{H}$. 1) A (multivalued) operator $T\colon \mathcal{D} \rightrightarrows \mathcal{H}$ is called firmly nonexpansive, if for all $f, g \in \mathcal{D}$, for all $t_f \in T[f]$, $t_g \in T[g]$*

$$\|(f - t_f) - (g - t_g)\|_2^2 + \|t_f - t_g\|_2^2 \leqslant \|f - g\|_2^2, \tag{5.6}$$

which is equivalent to

$$\|t_f - t_g\|_2^2 \leqslant \langle t_f - t_g, f - g\rangle. \tag{5.7}$$

2) A (multivalued) operator $T\colon \mathcal{D} \rightrightarrows \mathcal{H}$ is called nonexpansive, if for all $f, g \in \mathcal{D}$, for all $t_f \in T[f]$, $t_g \in T[g]$

$$\|t_f - t_g\|_2 \leqslant \|f - g\|_2.$$

From the definition immediately follows that every firmly nonexpansive operator is nonexpansive.

Proof (equivalence of (5.6) and (5.7)). Let $f, g \in \mathcal{D}$, let $t_f \in T[f]$, $t_g \in T[g]$.

$$\begin{aligned}
\|(f - t_f) - (g - t_g)\|_2^2 + \|t_f - t_g\|_2^2 \leqslant \|f - g\|_2^2 \quad &\Leftrightarrow \\
\|f - g\|_2^2 + 2\langle f - g, t_g - t_f\rangle + 2\|t_f - t_g\|_2^2 \leqslant \|f - g\|_2^2 \quad &\Leftrightarrow \\
\|t_f - t_g\|_2^2 \leqslant \langle f - g, t_f - t_g\rangle. \quad & \qquad \square
\end{aligned}$$

Example 5.24 (Convex projections are firmly nonexpansive). Let $\mathcal{X} \subset \mathcal{H}$ be convex, let $P_{\mathcal{X}}$ be the corresponding projection. Then, $P_{\mathcal{X}}$ is firmly nonexpansive.

Proof. By Lemma 3.10,

$$\begin{aligned}
\langle g - P_{\mathcal{X}}[g], P_{\mathcal{X}}[f] - P_{\mathcal{X}}[g]\rangle \leqslant 0, \\
\langle f - P_{\mathcal{X}}[f], P_{\mathcal{X}}[g] - P_{\mathcal{X}}[f]\rangle \leqslant 0.
\end{aligned}$$

Adding the inequalities together yields

$$\langle g - P_{\mathcal{X}}[g] - f + P_{\mathcal{X}}[f], P_{\mathcal{X}}[f] - P_{\mathcal{X}}[g]\rangle \leqslant 0,$$

which is equivalent to

$$\|P_{\mathcal{X}}[f] - P_{\mathcal{X}}[g]\|_2^2 \leqslant \langle g - f, P_{\mathcal{X}}[g] - P_{\mathcal{X}}[f]\rangle. \qquad \square$$

Definition 5.25 (Monotone operators). *Let $\mathcal{D} \subset \mathcal{H}$. A multivalued map $A\colon \mathcal{D} \rightrightarrows \mathcal{H}$ is called monotone, if for all $f, g \in \mathcal{D}$ and for all $a_f \in A[f]$ and $a_g \in A[g]$ one has*

$$\langle a_f - a_g, f - g\rangle \geqslant 0.$$

If $A\colon \mathcal{H} \to \mathcal{H}$ is single-valued and $f = g + h$ for some non-zero $h \in \mathcal{H}$, this equation can be rewritten to a (perhaps more intuitive) condition

$$\left\langle \frac{A[g+h] - A[g]}{\|h\|_2}, h \right\rangle \geqslant 0$$

for all $f, g \in \mathcal{H}$. For example, for $c \in \mathbb{R}$, the linear function $A[g] := cg$ is monotone if and only if $c \geqslant 0$.

Example 5.26 (Monotone operators).

1) Let $E\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be convex; then, its subdifferential is a monotone operator, since for all $f, g \in \operatorname{domain} \partial_c E$

$$\langle \partial_c E[g], f - g \rangle \leqslant E[f] - E[g] \quad \text{and} \quad \langle \partial_c E[f], g - f \rangle \leqslant E[g] - E[f];$$

adding both inequalities yields $\langle \partial_c E[g] - \partial_c E[f], f - g \rangle \leqslant 0$, equivalent to monotonicity of $\partial_c E$.

2) Monotone operators also arise in non-convex contexts. Let $\mathcal{X}$ be a weakly closed, not necessarily convex, subset of $\mathcal{H}$. A projection operator $P_{\mathcal{X}}$ is monotone, since by Corollary 3.6

$$\|g - P_{\mathcal{X}}[g]\|_2^2 \leqslant \|g - P_{\mathcal{X}}[f]\|_2^2 \quad \text{and} \quad \|f - P_{\mathcal{X}}[f]\|_2^2 \leqslant \|f - P_{\mathcal{X}}[g]\|_2^2.$$

Expand the squares and add both inequalities to obtain

$$\begin{array}{rcll} -2\langle g, P_{\mathcal{X}}[g]\rangle - 2\langle f, P_{\mathcal{X}}[f]\rangle & \leqslant & -2\langle g, P_{\mathcal{X}}[f]\rangle - 2\langle f, P_{\mathcal{X}}[g]\rangle & \Leftrightarrow \\ 0 & \leqslant & 2\langle f - g, P_{\mathcal{X}}[f]\rangle + 2\langle g - f, P_{\mathcal{X}}[g]\rangle & \Leftrightarrow \\ 0 & \leqslant & 2\langle f - g, P_{\mathcal{X}}[f] - P_{\mathcal{X}}[g]\rangle. & \end{array}$$

3a) Let $\mathcal{X}$ be a convex subset of $\mathcal{H}$. Then, the operator $g \mapsto g - P_{\mathcal{X}}[g]$ is monotone, since for all $f, g \in \mathcal{H}$

$$\begin{aligned} &\langle g - P_{\mathcal{X}}[g] - (f - P_{\mathcal{X}}[f]), g - f \rangle \\ &\qquad = \|g - f\|_2^2 - \langle P_{\mathcal{X}}[g] - P_{\mathcal{X}}[f], g - f \rangle \\ &\qquad \geqslant \|g - f\|_2^2 - \|P_{\mathcal{X}}[g] - P_{\mathcal{X}}[f]\|_2 \|g - f\|_2 \geqslant 0, \end{aligned}$$

since $\|P_{\mathcal{X}}[g] - P_{\mathcal{X}}[f]\|_2 \leqslant \|g - f\|_2$ for convex $\mathcal{X}$ by Example 5.24.

3b) Let $\mathcal{X} \subset \mathcal{H}$ be such that a corresponding projection $P_{\mathcal{X}}$ is an indicator projection. Then, the operator $g \mapsto g - P_{\mathcal{X}}[g]$ is also an indicator projection by Example 5.4, and thus monotone by Example 5.26 2).

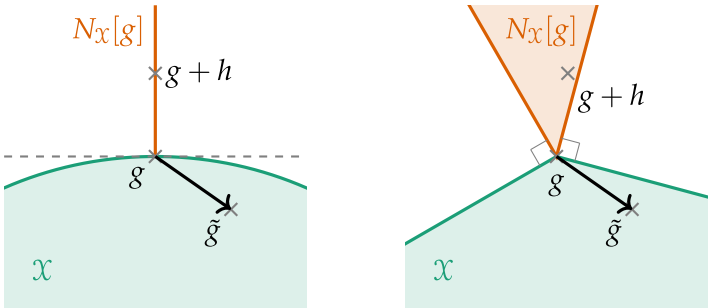

Figure 5.3: Illustration to Definition 5.30: normal cones to convex sets. Two examples of convex cones $N_{\mathcal{X}}$ to convex set $\mathcal{X}$ at $g \in \mathcal{X}$. For any $h \in N_{\mathcal{X}}$, the angle between $\tilde{g} - g$ and $(g+h) - g$ is greater or equal than $\pi/2$ for all $\tilde{g} \in \mathcal{X}$, i. e. $\langle \tilde{g} - g, h \rangle \leqslant 0$.
On the left, $N_{\mathcal{X}}[g]$ given by a ray; on the right, $N_{\mathcal{X}}[g]$ has non-empty interior.

3c) The operator $g \mapsto g - P_{\mathcal{M}}[g]$ is not monotone: pick $g \in \mathcal{H}$ with $|\hat{g}| = \sqrt{I}/2$, let $f = -g$. Then, $P_{\mathcal{M}}[f] = -P_{\mathcal{M}}[g]$, and

$$\begin{aligned}\langle g - P_{\mathcal{M}}[g] - (f - P_{\mathcal{M}}[f]), g - f \rangle \\ = \langle g - P_{\mathcal{M}}[g] + (g - P_{\mathcal{M}}[g]), g + g \rangle = 4\langle g - P_{\mathcal{M}}[g], g \rangle \\ = 4C_{\mathcal{F}} \int (|\hat{g}| - \sqrt{I})|\hat{g}| = -C_{\mathcal{F}} \|\sqrt{I}\|_2^2 < 0.\end{aligned}$$

The focus on operators of the form $g \mapsto g - P_{\mathcal{X}}[g]$ in these examples will become apparent in light of Equation (5.21) in Section 5.3.5.

Definition 5.27 (Maximal monotone operators). *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone. The operator A is called maximal monotone, if it can not be extended to a different monotone operator, i. e., if for any $g, a_* \in \mathcal{H}$ the monotonicity condition*

$$(\forall f \in \mathcal{H}, a_f \in A[f]) \quad \langle a_f - a_*, f - g \rangle \geqslant 0$$

implies $a_* \in A[g]$.

Recall the following well-known results on maximal monotone operators (presented here without proofs).

Theorem 5.28 (Monotone extension [BC17, Thm. 20.21]). *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone. Then there exists a maximally monotone extension of A.*

Theorem 5.29 (Moreau [BC17, Thm. 20.25]). *Let F be a proper lower semicontinuous functional from $\mathcal{H}$ to $(-\infty, \infty]$. Then, its subdifferential $\partial_c F$ is a maximal monotone operator.*

An important example of maximal monotone operators are normal cones.

DEFINITION 5.30 (NORMAL CONES). *Let $\mathcal{X} \in \mathcal{H}$ be nonempty and convex, let $g \in \mathcal{H}$. The normal cone to $\mathcal{X}$ at g is defined as*

$$N_{\mathcal{X}}[g] = \begin{cases} \{h \in \mathcal{H} \mid \sup_{\tilde{g}\in\mathcal{X}}\langle h, \tilde{g} - g\rangle \leqslant 0\} & \text{if } g \in \mathcal{X}; \\ \varnothing & \text{otherwise,} \end{cases}$$

see Figure 5.3.

Example 5.31 (Normal cones and convex set wells). Let $\mathcal{C} \subset \mathcal{H}$ be closed, convex, non-epmty. Define the infinite well functional

$$e^{\infty}_{\mathcal{C}} : \mathcal{H} \mapsto (-\infty, \infty], \quad g \mapsto \begin{cases} \infty & \text{if } g \notin \mathcal{C} \\ 0 & \text{else.} \end{cases}$$

Then, $\partial_c e^{\infty}_{\mathcal{C}}[g] = N_C[g]$ for all $g \in \mathcal{H}$.

Proof. Follows directly from the definition of the normal cone (Definition 5.30) and the definition of the subdifferential (Equation (5.3)).
"$\subseteq$" Let $g \in \mathcal{C}$. If $v \in \partial_c e^{\infty}_{\mathcal{C}}[g]$, then

$$\underbrace{e^{\infty}_{\mathcal{C}}[g]}_{=0} + \langle v, f - g\rangle \leqslant e^{\infty}_{\mathcal{C}}[f]$$

for all $f \in \mathcal{H}$. In particular, for all $f \in \mathcal{C}$, right-hand side equals zero and $\langle v, f - g\rangle \leqslant 0$; thus, $v \in N_{\mathcal{C}}[g]$. Let $g \notin \mathcal{C}$; pick $f \in \mathcal{C}$ (which is non-empty by assumption). Then, for any $v \in \mathcal{H}$

$$\underbrace{e^{\infty}_{\mathcal{C}}[g]}_{=\infty} + \langle v, f - g\rangle > \underbrace{e^{\infty}_{\mathcal{C}}[f]}_{=0};$$

thus, $\partial_c e^{\infty}_{\mathcal{C}}[g] = \varnothing$. "$\supseteq$" follows in a similar straightforward fashion. □

LEMMA 5.32 (NORMAL CONES IN TERMS OF PROJECTIONS). *Let $\mathcal{C} \in \mathcal{H}$ be nonempty and convex, let $g \in \mathcal{H}$. The normal cone can be expressed in terms of the projection onto $\mathcal{C}$ as follows:*

$$N_{\mathcal{C}}[g] = \bigcup \{\alpha(f - P_{\mathcal{C}}[f]) \mid \alpha \geqslant 0\}, \tag{5.8}$$

where the union is taken over all $f \in \mathcal{H}$ such that $P_{\mathcal{C}}[f] = g$. A following variation on this equivalence is also true:

$$N_{\mathcal{C}}[g] = \{f - P_{\mathcal{C}}[f] \mid f \in \mathcal{H}, P_{\mathcal{C}}[f] = g\}. \tag{5.9}$$

Proof. If $g \notin \mathcal{C}$, both $N_{\mathcal{C}}[g]$ and the union on the right-hand side of Equation (5.8) are empty. Therefore, let $g \in \mathcal{C}$ for the rest of the proof.

$\subseteq$. Let $h \in N_{\mathcal{C}}[g]$. Choose $\alpha = 1$; let us show that for $f := g + h$ holds $P_{\mathcal{C}}[f] = g$. Indeed, by Lemma 3.10

$$\langle f - P_{\mathcal{C}}[f], P_{\mathcal{C}}[g] - P_{\mathcal{C}}[f]\rangle \leqslant 0,$$

and by definition of $N_{\mathcal{C}}$

$$\langle h, P_{\mathcal{C}}[f] - g \rangle \leqslant 0.$$

Since $P_{\mathcal{C}}[g] = g$ and $h = f - g$, adding both inequalities yields

$$\begin{aligned} \langle f - P_{\mathcal{C}}[f], g - P_{\mathcal{C}}[f] \rangle + \langle f - g, P_{\mathcal{C}}[f] - g \rangle &\leqslant 0 \quad \Rightarrow \\ \langle f - P_{\mathcal{C}}[f] - f + g, g - P_{\mathcal{C}}[f] \rangle &\leqslant 0 \quad \Rightarrow \\ \|g - P_{\mathcal{C}}[f]\|_2^2 &\leqslant 0, \end{aligned}$$

so $P_{\mathcal{C}}[f] = g$. From this follows $h = f - g = f - P_{\mathcal{C}}[f]$, meaning that h belongs to the right-hand side of Equation (5.8).

$\supseteq$. Let $f \in \mathcal{H}$ be such that $P_{\mathcal{C}}[f] = g$, let $h := \alpha(f - P_{\mathcal{C}}[f])$ for $\alpha \geqslant 0$. Then,

$$\begin{aligned} \sup_{\tilde{g} \in \mathcal{C}} \langle h, \tilde{g} - g \rangle &= \sup_{\tilde{g} \in \mathcal{C}} \alpha \langle f - P_{\mathcal{C}}[f], \tilde{g} - g \rangle \\ &= \sup_{\tilde{g} \in \mathcal{C}} \alpha \langle f - P_{\mathcal{C}}[f], P_{\mathcal{C}}[\tilde{g}] - P_{\mathcal{C}}[f] \rangle \leqslant 0 \end{aligned}$$

by Lemma 3.10, so $h \in N_{\mathcal{C}}[g]$.

Expression Equation (5.9) is shown using exactly the same argumentation. □

Proposition 5.33 (Normal cones are maximal monotone). *Let $\mathcal{C} \subset \mathcal{H}$ be nonempty, closed, convex. Then, $N_{\mathcal{C}}$ is maximal monotone.*

Proof. Let $g, a \in \mathcal{H}$, assume that $\forall f \in \mathcal{H}$, $a_f \in N_{\mathcal{C}}[f]$ holds the monotonicity condition $\langle a_f - a_*, f - g \rangle \geqslant 0$. We need to show that $a_* \in N_{\mathcal{C}}[g]$.

From the monotonicity condition follows

$$\sup_{f \in \mathcal{C}} \langle a_*, f - g \rangle \leqslant \sup_{\substack{f \in \mathcal{C} \\ a_f \in N_{\mathcal{C}}[f]}} \langle a_f, f - g \rangle \overset{(*1)}{\leqslant} 0, \qquad (*2)$$

where $(*1)$ is true since $a_f \in N_{\mathcal{C}}[f]$. From $(*2)$ follows $a_* \in N_{\mathcal{C}}[g]$. □

Definition 5.34 (Inverse of a multivalued map). *Let $\mathcal{D} \subset \mathcal{H}$. The inverse of a multivalued map $B\colon \mathcal{D} \rightrightarrows \mathcal{H}$ at $f \in \mathcal{D}$ is defined as*

$$\begin{aligned} B^{-1}\colon \operatorname{range}(B) &\rightrightarrows \mathcal{D} \\ f &\mapsto B^{-1}[f] = \{ g \in \mathcal{H} \mid B[g] \ni f \}, \end{aligned}$$

i.e. $g \in B^{-1}[f] \Leftrightarrow f \in B[g]$.

Definition 5.35 (Resolvent). *Let $\mathcal{D} \subset \mathcal{H}$, let $A\colon \mathcal{D} \rightrightarrows \mathcal{H}$. The resolvent of A is defined as*

$$\begin{aligned} J_A\colon \operatorname{range}(I + A) &\rightrightarrows \mathcal{D} \\ g &\mapsto J_A[g] = (I + A)^{-1}[g], \end{aligned}$$

where $(I + A)[g] = g + A[g] = \{g + a_g \mid a_g \in A[g]\}$.

Corollary 5.36 (Resolvents of monotone operators are single-valued). *Let* $\mathcal{D} \subset \mathcal{H}$*, let* $A\colon \mathcal{D} \rightrightarrows \mathcal{H}$ *be monontone. Then,* $J_A\colon \operatorname{range}(I + A) \to \mathcal{D}$ *is single-valued.*

Proof. Let $q \in \operatorname{range}(I + A)$. Let us show that if $f, g \in J_A[q]$, then $f = g$. Indeed, since $f \in J_A[q] = (I + A)^{-1}(q)$, there exists $a_f \in A[f]$ such that $f + a_f = q$. Analogously, there exists $a_g \in A[g]$ such that $g + a_g = q$. Therefore,

$$\begin{aligned} f + a_f &= g + a_g && && \Rightarrow \\ a_f - a_g &= -(f - g) && \mid \langle \cdot, f - g \rangle && \Rightarrow \\ 0 \overset{(*)}{\leqslant} \langle a_f - a_g, f - g \rangle &= -\|f - g\|_2^2 \leqslant 0, \end{aligned}$$

where $(*)$ follows from monotonicity of A. Therefore, $f = g$ and J_A is single-valued. □

Example 5.37 (Resolvents of normal cones). Let $\mathcal{C} \subset \mathcal{H}$ be nonempty, closed, convex. Then, $J_{N_\mathcal{C}}[g] = P_\mathcal{C}[g]$ for all $g \in \mathcal{H}$.

Proof. By Lemma 5.32, for $g \in \mathcal{H}$

$$\begin{aligned} N_\mathcal{C}[g] &= \{f - P_\mathcal{C}[f] \mid f \in \mathcal{H}, P_\mathcal{C}[f] = g\} && \Rightarrow \\ g + N_\mathcal{C}[g] &= \{ \quad f \quad \mid f \in \mathcal{H}, P_\mathcal{C}[f] = g\} = P_\mathcal{C}^{-1}[g]. \end{aligned}$$

Therefore, for $f \in \mathcal{H}$

$$g \in (I + N_\mathcal{C})^{-1}[f] \Leftrightarrow f \in g + N_\mathcal{C}[g] \Leftrightarrow f \in P_\mathcal{C}^{-1}[g] \Leftrightarrow g \in P_\mathcal{C}[f],$$

where the last expression means $g = P_\mathcal{C}[f]$ (in a minor abuse of notation, single-valued sets in this proof are identified with their elements). □

Theorem 5.38 (Monotonicity and firm nonexpansiveness). *Let* $\mathcal{D}$ *be a nonempty subset of* $\mathcal{H}$*, let* $A\colon \mathcal{D} \rightrightarrows \mathcal{H}$*. Then,* A *is monotone if and only if* J_A *is firmly nonexpansive.*

Proof. The essence of the proof is calculation (5.10), but for ease of domain and range bookkeeping the proof is split in two parts.

"$\Leftarrow$". Let J_A be firmly nonexpansive. We need to show that

$$0 \leqslant \langle j_f - j_g, a_{j_f} - a_{j_g} \rangle.$$

for all $j_f, j_g \in \mathcal{D}$, for all $a_{j_f} \in A[j_f]$, $a_{j_g} \in A[j_g]$.

Let $f := j_f + a_{j_f}$, let $g := j_g + a_{j_g}$. Then, $f, g \in \text{range}(I + A) = \text{domain}\, J_A$, and by Equation (5.7)

$$\begin{aligned}
\|j_f - j_g\|_2^2 &\leqslant \langle j_f - j_g, f - g\rangle && \Leftrightarrow \\
\|j_f - j_g\|_2^2 &\leqslant \langle j_f - j_g, j_f + a_{j_f} - j_g - a_{j_g}\rangle && \Leftrightarrow \\
\|j_f - j_g\|_2^2 &\leqslant \|j_f - j_g\|_2^2 + \langle j_f - j_g, a_{j_f} - a_{j_g}\rangle && \Leftrightarrow \\
0 &\leqslant \langle j_f - j_g, a_{j_f} - a_{j_g}\rangle && \qquad (5.10)
\end{aligned}$$

for all $j_f, j_g \in \mathcal{D}$, for all $a_{j_f} \in A[j_f]$, $a_{j_g} \in A[j_g]$.

"$\Rightarrow$". Let A be monotone. We need to show that

$$\|j_f - j_g\|_2^2 \leqslant \langle j_f - j_g, f - g\rangle$$

for all $f, g \in \text{range}(I + A)$, for all $j_f \in J_A[f]$, $j_g \in J_A[g]$. By definition of the resolvent, $j_f \in J_A[f]$ imlies $f \in j_f + A[j_f]$, which means that there exists $a_{j_f} \in A[j_f]$ such that $f = j_f + a_{j_f}$; likewise, there exists $a_{j_g} \in A[j_g]$ such that $g = j_g + a_{j_g}$. Then, calculation (5.10) applies for all $f, g \in \text{range}(I + A)$, for all $j_f \in J_A[f]$, $j_g \in J_A[g]$, meaning that J_A is nonexpansive. □

In applications where one can use maximal monotone operators instead of monotone operators, the choices of domains and ranges can be sidestepped. This is shown by the following theorem, stated here without proof.

THEOREM 5.39 (MINTY [BC17, THM. 21.1]). *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone. Then, A is maximally monotone if and only if* $\text{range}(I + A) = \mathcal{H}$.

COROLLARY 5.40 (MAXIMAL MONOTONICITY AND FIRM NONEXPANSIVENESS). *An operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone if and only if J_A is firmly nonexpansive, and* $\text{domain}\, J_A = \mathcal{H}$.

Proof. By Theorem 5.38 and Theorem 5.39, A is maximal monotone if and only if $\mathcal{H} = \text{range}(I + A) = \text{domain}\, J_A$. □

5.3.3 *DR for maximal monotone operators*

This subsection recalls well-known results on the Lions-Mercier formulation of the Douglas-Rachford from [LM79]. For a survey of more recent results on convergence of DR-LM, the reader is referred to [LS19].

Consider the nonlinear multivalued evolution equation

$$\partial_t f \in -A[f] - B[f], \qquad (5.11)$$

where A, B are maximal monotone (possibly multivalued) operators on $\mathcal{H}$.

Definition 5.41 (DR for maximal monotone operators [LM79]). *Let $g_0 \in \mathcal{H}$, let A, B be maximal monotone (possibly multivalued) operators on $\mathcal{H}$, let $\lambda > 0$. The sequence $(g_n)_{n\in\mathbb{N}}$ is generated by the DR-LM algorithm (Douglas-Rachford in Lions-Mercier formulation), if*

$$\begin{aligned} g_{n+1} &= \frac{1}{2}\big(\mathrm{Id} + (2J_{\lambda A} - \mathrm{Id}) \circ (2J_{\lambda B} - \mathrm{Id})\big)[g_n] \\ &= g_n + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g_n] - J_{\lambda B}[g_n]. \end{aligned} \tag{DR-LM}$$

Remark 5.42. In previous definition, $J_{\lambda A}, J_{\lambda B}$ are single-valued by Corollary 5.36, and therefore update DR-LM is uniquely defined, even though the evolution equation (5.11) is multivalued.

In applications, the algorithm may prove itself successful even if A or B are not monotone. In this case, the iterate g_{n+1} is chosen from any value in the set on the right-hand side of DR-LM.

The following proposition shows the one-to one correspondence between fixed points of the evolution equation (5.11) and fixed points of DR-LM.

Proposition 5.43 (Fixed points correspondence [LM79]). *Let $g_0 \in \mathcal{H}$, let A, B be maximal monotone (possibly multivalued) operators on $\mathcal{H}$, let $\lambda > 0$.*

i) If $g \in \mathcal{H}$ is such that

$$g = g + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g] - J_{\lambda B}[g],$$

then for $f = J_{\lambda B}[g]$ holds $0 \in -A[f] - B[f]$.

ii) If $f \in \mathcal{H}$ is such that $0 \in -A[f] - B[f]$, then there exists $g \in (I + B)[f]$ such that

$$g \in g + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g] - J_{\lambda B}[g].$$

Proof. i) Since g is a fixed point, one has

$$\begin{aligned} 0 &= J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g] - \underbrace{J_{\lambda B}[g]}_{=f} && \Rightarrow \\ f &= J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g] && \Rightarrow \\ (I + \lambda A)[f] &\ni (2J_{\lambda B} - \mathrm{Id})[g] && \Rightarrow \\ f + \lambda A[f] &\ni 2f - g && \Rightarrow \\ f - g &\in \lambda A[f]. \end{aligned}$$

Further, by definition of f one has

$$f = J_{\lambda B}[g] \Rightarrow g \in (I + \lambda B)[f] \Rightarrow g - f \in \lambda B[f].$$

Combine both inclusions to obtain $0 \in A[f] + B[f]$, since $\lambda \neq 0$.

ii) Since f is a fixed point, there exist $a_f \in A[f]$ and $b_f \in B[f]$ such that $0 = a_f + b_f$. Choose $g := f + b_f$; then,

$$0 = -f + a_f + f + b_f = -f + a_f + g \Rightarrow f - g = a_f. \qquad (5.12)$$

To prove that g is a fixed point of DR-LM, it is sufficient to show that

$$J_A \circ (2J_B - I)[g] - J_B[g] = 0 \Leftrightarrow J_A[2f - g] - f = 0, \qquad (5.13)$$

where the equivalence holds since by definition of g have $J_B[g] = f$. Adding f to both sides of Equation (5.12) and applying J_a,

$$J_A[2f - g] = J_A[f + a_f] \overset{(*)}{=} f,$$

where $(*)$ is true by definition of J_A. Therefore, Equation (5.13) holds and g is a fixed point of DR-LM. □

While we conjecture that similar correspondence persists for multivalued resolvents, the resulting rigor lies beyond the applications of this work. In practice, it is easier to verify such correspondence explicitely for the feasibility problem formulation Definition 5.53, see Remark 5.54.

The following result from [BP67] (stated here without proof) is the key component for the proof of DR-LM convergence.

Theorem 5.44 (Browder [BP67]). *Let $\mathcal{D}$ be a closed convex subset of $\mathcal{H}$. If $T\colon \mathcal{H} \to \mathcal{H}$ is nonexpansive, and if T has at least one fixed point $f \in \mathcal{D}$, then for all $g \in \mathcal{D}$ holds:*

i) $T^{n+1}[g] - T^n[g] \to 0$ strongly as $n \to \infty$;

ii) $T^n[g] \rightharpoonup g_* \in \mathcal{D}$ weakly,

where $g_* = T[g_*]$ is a fixed point of T.

Lemma 5.45 (DR-LM update is firmly nonexpansive [LM79]). *Let $\mathcal{D}$ be a closed convex subset of $\mathcal{H}$, let $T_1, T_2\colon \mathcal{D} \to \mathcal{D}$ be firmly nonexpansive; then, $T = T_1[2T_2 - I] + I - T_2$ is firmly nonexpansive.*

Theorem 5.46 (DR-LM convergence [LM79, Prop. 2]). *Let $g_0 \in \mathcal{H}$, let A, B be maximal monotone (possibly multivalued) operators on $\mathcal{H}$, let $\lambda > 0$. Assume that there exists a fixed point $g \in \mathcal{H}$ of the DR-LM update. Let $(g_n)_{n\in\mathbb{N}}$ be generated by the DR-LM algorithm. Then, g_n weakly converges to a fixed point g_* of the DR-LM update, and for $f_* = J_{\lambda B}[g_*]$ holds $0 \in -A[f_*] - B[f_*]$.*

Proof. Use Lemma 5.45 to show that DR-LM update is nonexpansive; use Theorem 5.44 to follow the weak convergence; use Proposition 5.43 to follow that f_* is a fixed point of the evolution equation. □

5.3.4 *DR for convex feasibility problems*

This presents a connection between energy minimization, and DR-LM reformulation for convex feasibility problems. This connection is well known in convex analysis; however, in the convex setting it is more common to work directly with the simplified form DR-cf, sidestepping the energy minimization discussion. The added value of the discussion presented here lies in the comparisson between the convex case and the phase retrieval case discussed in Section 5.3.5.

An important particular case of DR-LM is the case of convex feasibility problems; see [LS19, Sec. 2.2] for numerous applications of the convex case.

Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be nonempty closed convex sets. Consider the minimization problem

$$\arg\min_{f \in \mathcal{H}} E_{\mathcal{X}}[f] + E_{\mathcal{Y}}[f]. \tag{5.14}$$

For $E_{\mathcal{X}} + E_{\mathcal{Y}}$ to have a minimum at f, it is necessary — this is known as Fermat's rule, cf. [BC17, Thm. 16.3] — that

$$0 \in \partial_c (E_{\mathcal{X}} + E_{\mathcal{Y}})[f] = \partial_c E_{\mathcal{X}}[f] + \partial_c E_{\mathcal{X}}[g], \tag{5.15}$$

where we can use the additivity of the subdifferential (e. g. [BC17, Rem. 16.46]) since $E_{\mathcal{X}}, E_{\mathcal{Y}}$ are proper lower semi-continuous convex functions with domain $E_{\mathcal{X}} =$ domain $E_{\mathcal{Y}} = \mathcal{H}$. From Proposition 5.21 we know that $f - P_{\mathcal{X}}[f] \in \partial_c E_{\mathcal{X}}[f]$, and likewise for $E_{\mathcal{Y}}$. Therefore, the problem is equivalent to searching fixed points of the evolution equation

$$\partial_t f = -\underbrace{(f - P_{\mathcal{X}}[f])}_{=:A[f]} - \underbrace{(f - P_{\mathcal{Y}}[f])}_{=:B[f]}. \tag{5.16}$$

For any $\lambda > 0$, the map $f \mapsto \lambda(f - P_{\mathcal{X}}[f])$ is contained in $N_{\mathcal{X}}[f]$, and its resolvent can be calculated using the same argumentation as in Example 5.37. Namely, $J_{\lambda A}[f] = P_{\mathcal{X}}[f]$ for any $\lambda > 0$. Likewise, $J_{\lambda B}[f] = P_{\mathcal{Y}}[f]$ for any $\lambda > 0$. Therefore, Definition 5.41 reads as follows.

Definition 5.47 (DR for convex feasibility). *Let $g_0 \in \mathcal{H}$, let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be nonempty closed convex sets, let $\lambda > 0$. The sequence $(g_n)_{n \in \mathbb{N}}$ is generated by the DR-cf variant (Douglas-Rachford for convex feasibility), if*

$$\begin{aligned} g_{n+1} &= \frac{1}{2}\big(\mathrm{Id} + (2P_{\mathcal{X}} - \mathrm{Id}) \circ (2P_{\mathcal{Y}} - \mathrm{Id})\big)[g_n] \\ &= \frac{1}{2}\big(g_n + R_{\mathcal{X}} \circ R_{\mathcal{Y}}[g_n]\big) \\ &= g_n + P_{\mathcal{X}}[R_{\mathcal{Y}}[g_n]] - P_{\mathcal{Y}}[g_n]. \end{aligned} \tag{DR-cf}$$

The resulting algorithm is invariant with respect to rescaling of the time variable in Equation (5.16). Indeed, such rescaling would lead to the different choice of the constant λ in DR-LM that does not alter the resulting form DR-cf.

This invariance can be highlighted by the following minimization problem that also leads to Equation (DR-cf): as before, let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be nonempty closed convex sets; consider the infinite well minimization problem

$$\underset{f \in \mathcal{H}}{\arg\min}\, e^{\infty}_{\mathcal{X}}[f] + e^{\infty}_{\mathcal{Y}}[f] \tag{5.17}$$

with well functionals from Example 5.31. Again, by Fermat's rule, f is a minimum, only if

$$0 \in \partial_{\mathrm{c}}(e^{\infty}_{\mathcal{X}} + e^{\infty}_{\mathcal{Y}})[f] = \partial_{\mathrm{c}} e^{\infty}_{\mathcal{X}}[f] + \partial_{\mathrm{c}} e^{\infty}_{\mathcal{Y}}[f]. \tag{5.18}$$

Again, we can use the additivity of the subdifferential (e. g. [BC17, Rem. 16.46]) since $E_{\mathcal{X}}, E_{\mathcal{Y}}$ are proper lower semi-continuous convex functions with domain $E_{\mathcal{X}} \cap$ domain $E_{\mathcal{Y}} \neq \varnothing$. Consider the restricted equation

$$0 \in \partial_{\mathrm{c}} e^{\infty}_{\mathcal{X}}[f] + \partial_{\mathrm{c}} e^{\infty}_{\mathcal{Y}}[f] = N_{\mathcal{X}}[f] + N_{\mathcal{Y}}[f], \tag{5.19}$$

where we used Example 5.31. Note that the right-hand side is not empty if and only if $f \in \mathcal{X} \cap \mathcal{Y}$, yet it leads to the same form DR-cf, since $J_{N_{\mathcal{X}}}[g] = P_{\mathcal{X}}[g]$, and likewise for $\mathcal{Y}$, due to Example 5.37.

Thus, in convex case the algoritm DR-LM has the following notable features:

i) Resulting form Equation (DR-cf) is invariant under time rescaling of the evolution equation (5.15), i. e. does not depend on parameter λ that appears in DR-LM;

ii) Energy functionals in (5.14) and much less regular infinite wells in (5.17) lead to exactly the same formulation (DR-cf).

5.3.5 *DR for phase retrieval*

Similarly to the convex case, Douglas-Rachford for phase retrieval can be connected to a minimization problem, but the precise nature of the connection is different due to the fact that $E_{\mathcal{M}}$ is non-convex. This connection is explored in this subsection. To the best of our knowledge, this connection previously has not observed for phase retrieval.

Consider the energy minimization phase retrieval with positivity

$$\underset{f \in \mathcal{H}}{\arg\min}\, E_{\mathcal{M}}[f] + E_{\mathcal{A}}[f] \tag{5.20}$$

on a bounded domain, so that $\mathcal{M}$ is weakly closed, and assuming that $\mathcal{A}$ is closed and convex. The reason for this strong assumption on $\mathcal{A}$ will become apparent in Proposition 5.52. As discussed in Remark 5.18, one possible selection of the corresponding subdifferential flow of $E_{\mathcal{M}} + E_{\mathcal{A}}$ results in the evolution equation

$$\partial_t f = -\gamma(g - P_{\mathcal{M}}[f]) - \gamma(g - P_{\mathcal{A}}[f]) \in -\gamma\partial_{\mathrm{KM}}(E_{\mathcal{M}} + E_{\mathcal{A}})[f], \quad (5.21)$$

where we have introduced a time scaling parameter γ; its role will become apparent in the end of the section. For now, consider the case $\gamma = 1$.

As demonstrated in Example 5.26 3c), the operator $A[f] := f - P_{\mathcal{M}}[f]$ is not monotone, and therefore its resolvent is not necessarily single-valued. This can be demonstrated explicitely by showing that the operator $I + A$ is not injective. For any data $\sqrt{I}$, let f be such that

$$\begin{cases} |\hat{f}(k)| < \sqrt{I}(k) \text{ and } \hat{f}(k) \neq 0 & \text{if } \sqrt{I}(k) \neq 0, \\ |\hat{f}(k)| = 0 & \text{otherwise.} \end{cases}$$

Let

$$\hat{g}_1(k) := +\frac{\sqrt{I}(k) + |\hat{f}(k)|}{2}\frac{\hat{f}(k)}{|\hat{f}(k)|} \qquad \Rightarrow \qquad g_1 = +\frac{1}{2}(P_{\mathcal{M}}[f] + f);$$
$$\hat{g}_2(k) := -\frac{\sqrt{I}(k) - |\hat{f}(k)|}{2}\frac{\hat{f}(k)}{|\hat{f}(k)|} \qquad \Rightarrow \qquad g_2 = -\frac{1}{2}(P_{\mathcal{M}}[f] - f),$$

with the convention that $\hat{g}_i(k) = 0$ at k where $\hat{f}(k) = \sqrt{I}(k) = 0$ for $i \in \{1, 2\}$. Since $|\hat{f}| \leqslant \sqrt{I}$, we have $P_{\mathcal{M}}[g_1] = P_{\mathcal{M}}[f]$, $P_{\mathcal{M}}[g_2] = -P_{\mathcal{M}}[f]$, and

$$2g_1 - P_{\mathcal{M}}[g_1] = f = 2g_2 - P_{\mathcal{M}}[g_2],$$

meaning that the resolvent of $A[f]$ is not single-valued.

Therefore, one is free to choose any element of the resolvent J_A in DR-LM. One of the possible choices is demonstrated in the following lemma.

Lemma 5.48. *Let $\mathcal{X}$ be a nonempty proximal subset of $\mathcal{H}$, let $\lambda > 0$, let $A = \mathrm{Id} - P_{\mathcal{X}}$. Then, for any $f \in \mathcal{H}$,*

$$\frac{1}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{X}})[f] \in J_{\lambda A}[f]. \quad (5.22)$$

Proof. Let $f \in \mathcal{H}$, let $g = \frac{1}{1+\lambda}(f + \lambda P_{\mathcal{X}}[f])$. By the interpolation property (Lemma 3.8), $P_{\mathcal{X}}[g] = P_{\mathcal{X}}[f]$. Therefore,

$$\begin{aligned}(\mathrm{Id} + \lambda A)[g] &= g + \lambda A[g] = g + \lambda g - \lambda P_{\mathcal{X}}[g] \\ &= f + \lambda P_{\mathcal{X}}[f] - \lambda P_{\mathcal{X}}[f] = f,\end{aligned}$$

from which follows $g \in J_{\lambda A}[f]$. □

This motivates using $\frac{1}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{M}}) \in J_{\lambda A}$ and $P_{\mathcal{A}} = J_{\lambda B}$, where $A[f] = f - P_{\mathcal{M}}[f]$, $B[f] = f - P_{\mathcal{A}}[f]$. The pointwise limit of the operator $\frac{1}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{X}})$ as $\lambda \to \infty$ is well-defined: for any $f \in \mathcal{H}$,

$$\lim_{\lambda \to \infty} \frac{1}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{X}})[g] = P_{\mathcal{X}}[g].$$

Therefore, one can take the limit $\lambda \to \infty$ in Douglas-Rachford and simplify it to the following form.

DEFINITION 5.49 (DOUGLAS-RACHFORD FOR PHASE RETRIEVAL). *Let $g_0 \in \mathcal{H}$, let $\sqrt{I}$ be as in Definition 2.1, let the additional constraint $\mathcal{A}$ be proximal. The sequence $(g_n)_{n \in \mathbb{N}}$ is generated by the DR-HIO algorithm in the $P_{\mathcal{M}} \circ R_{\mathcal{A}}$ order, if*

$$g_{n+1} = \frac{1}{2}(g_n + R_{\mathcal{M}} \circ R_{\mathcal{A}}[g_n]) = g_n + P_{\mathcal{M}} \circ R_{\mathcal{A}}[g_n] - P_{\mathcal{A}}[g_n]. \tag{DR-HIO}$$

Remark 5.50. If the roles of $\mathcal{M}$ and $\mathcal{A}$ are reversed in DR-HIO, we call the resulting algorithm DR-HIO in the $P_{\mathcal{A}} \circ R_{\mathcal{M}}$ order. As demonstrated in Section 5.1, the DR form of HIO is usually derived in this order — the linearity of the object space operator is necessary for the derivation from the local form. The order $P_{\mathcal{A}} \circ R_{\mathcal{M}}$ is used, for example, in [BCL02].

Following [ELB18], we use the formulation DR-HIO with $P_{\mathcal{M}} \circ R_{\mathcal{A}}$ order unless mentioned otherwise.

See [Mou16] for a more detailed discussion on the order of operators in DR.

Remark 5.51. It is important to point out that $P_{\mathcal{M}} \notin J_{\lambda A}$ for $A[g] := g - P_{\mathcal{M}}[g]$, since

$$(I + \lambda A)[P_{\mathcal{M}}[g]] = ((1+\lambda)I - P_{\mathcal{M}})[P_{\mathcal{M}}[g]] = (1+\lambda)P_{\mathcal{M}}[g] - P_{\mathcal{M}}[g] = \lambda P_{\mathcal{M}}[g]$$

is, in general, not equal to g. Therefore, DR-HIO is not directly a special case of DR-LM. Rather, DR-HIO coincides with the limit of DR-LM as $\lambda \to \infty$ — provided the appropriate (as in Lemma 5.48) selection of the resolvent.

The DR limit $\lambda \to \infty$ may be interpreted alternatively as the time scale limit $\gamma \to \infty$ in Equation (5.20).

Indeed, if one keeps $\lambda = 1$ in the definition of DR-LM and uses the operator $\tilde{A}[g] = \gamma(g - P_{\mathcal{M}}[g])$ instead of $A = g - P_{\mathcal{M}}[g]$, the DR scale constant λ is effectively replaced by γ throughout all calculations.:

$$\lambda \tilde{A}[g] = 1\tilde{A}[g] = \gamma A[g].$$

Therefore, DR-HIO may be interpreted as algorithm corresponding to the flow (5.20) "infinitely accelerated in time".

While such interpretation is nonsensical from the dynamical point of view, it can be considered in a geometrical context akin to the normal cone equation (5.19) or the equation (5.16) in the convex case. Specifically, — cf. explanation immediately after Equation (5.16) — in the convex case, all values $\lambda > 0$ for (5.19) lead to exactly the same form DR-cf, and taking the limit $\lambda \to \infty$ does not alter the resulting DR algorithm.

Thus, if one would define DR-LM not for $\lambda > 0$, but by an update operator defined pointwise through the limit $\lambda \to \infty$, then both DR-cf and DR-HIO would be special cases of DR-LM.

The argument preceeding Definition 5.49 can be generalized to other non-convex problems in the following way.

PROPOSITION 5.52 (DR FOR NON-CONVEX CASE). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal, define $A[g] := g - P_{\mathcal{X}}[g], B[g] := g - P_{\mathcal{Y}}[g]$ for all $g \in \mathcal{H}$. Assume that at least one of the following holds:*

i) $P_{\mathcal{X}}$ is continuous (or $\mathcal{X}$ is Chebyshev, see Proposition 3.11);

ii) $\mathcal{Y}$ is non-empty, closed and convex.

Then,

$$g + P_{\mathcal{X}}[2P_{\mathcal{Y}}[g] - g] - P_{\mathcal{Y}}[g] \in \operatorname*{Li}_{\lambda\to\infty}\Big(g + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g] - J_{\lambda B}[g]\Big)$$

where Li *is the Kuratowski limit inferior.*

Proof. i) Assume that $P_{\mathcal{X}}$ is continuous. By Lemma 5.48, for any $g \in \mathcal{H}$ and any $\lambda > 0$ have

$$\frac{1}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{X}})[g] \in J_{\lambda A}[g], \quad \text{and}$$
$$\frac{1}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{Y}})[g] \in J_{\lambda B}[g].$$

Therefore, for any $g \in \mathcal{H}$ and any $\lambda > 0$

$$g + \frac{1}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{X}})\left[\frac{2}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{Y}})[g] - g\right] - \frac{1}{1+\lambda}(\mathrm{Id} + \lambda P_{\mathcal{Y}})[g] \in g + J_{\lambda A} \circ (2J_{\lambda B} - \mathrm{Id})[g] - J_{\lambda B}[g].$$

Using continuity of $P_{\mathcal{X}}$ on the left-hand side, one can take the limit $\lambda \to \infty$ to obtain the desired statement.

ii) If $\mathcal{Y}$ is non-empty, closed and convex, then $J_{\lambda B} = P_{\mathcal{Y}}$ for all $\lambda > 0$ (see Example 5.37 and observe that $B[g] \in N_{\mathcal{Y}}[g]$ for all $g \in \mathcal{H}$). Therefore, the argument from part i) of the proof holds, with the minor difference that continuity of $P_{\mathcal{X}}$ is no longer required, as λ is no longer present in the argument of $J_{\lambda A}$. □

For ease of reference, we restate Douglas-Rachford for generic feasibility problems.

DEFINITION 5.53 (DOUGLAS-RACHFORD FOR FEASIBILITY PROBLEMS). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal. The sequence $(g_n)_{n\in\mathbb{N}}$ is generated by the DR algorithm, if*

$$g_{n+1} = \frac{1}{2}\left(g_n + R_{\mathcal{Y}} \circ R_{\mathcal{X}}[g_n]\right) = g_n + P_{\mathcal{Y}} \circ R_{\mathcal{X}}[g_n] - P_{\mathcal{X}}[g_n]. \qquad \text{(DR)}$$

Proposition 5.52 establishes sufficient conditions for DR to be a subcase of DR-LM.

Remark 5.54 (Error estimation for Douglas-Rachford). Let $g \in \mathcal{H}$ be a fixed point of DR for a generic feasibility problem. At this fixed point holds

$$P_{\mathcal{X}} \circ R_{\mathcal{Y}}[g] = P_{\mathcal{Y}}[g],$$

so that $P_{\mathcal{Y}}[g] \in \mathcal{X} \cap \mathcal{Y}$ is a solution. In particular, one can use the quantities $E_{\mathcal{Y}}\left[P_{\mathcal{X}} \circ R_{\mathcal{Y}}[g_n]\right]$ or $E_{\mathcal{X}}\left[P_{\mathcal{Y}}[g_n]\right]$ to track convergence of the Douglas-Rachford algorithm.

In [ELB18], $E_{\mathcal{Y}}\left[P_{\mathcal{X}} \circ R_{\mathcal{Y}}[g_n]\right]$ with $\mathcal{X} = \mathcal{M}^{(i)}$ (incomplete modulus) and $\mathcal{Y} = \mathcal{T}_s(\nu)$ (non-negativity + support size) is used to track progress of DR for phase retrieval: the algorithm is terminated, when the energy drops below $0.05\|g_n\|_2^2$.

Remark 5.55 (Alternating Reflection Flow). ER is an explicit discretization of ERF, while the connection between DR and ERF is more intricate.

Inspecting similarities ER between and DR updates — the first contains alternating projections, the second contains relaxed alternating reflections — one may ask whether it is possible to find an equation that yields DR directly as an explicit discretization.

For example, it is natural to consider the flow

$$\partial_t g = -2g + R_{\mathcal{X}}[g] + R_{\mathcal{Y}}[g];$$

however, on further inspection one can see that it equals

$$\begin{aligned} & -2g + 2P_{\mathcal{X}}[g] - g + 2P_{\mathcal{Y}}[g] - g \\ =& 2(-2g + P_{\mathcal{X}}[g] - P_{\mathcal{Y}}[g]); \end{aligned}$$

in other words, such flow is merely an accelerated version of APF.

Chapter 9 shows that DR can be written as the explicit discretization of a system of equations, albeit in two variables.

Part II

ANALYZING THE ERROR-REDUCTION FLOW

This part studies the Error-Reduction Flow and its explicit discretization dERF (which can be seen as a generalization of the Error-Reduction algorithm). It contains the following main original contributions.

Chapter 6 establishes energy dissipation properties of dERF, and demonstrates sufficient conditions under which dERF has fixed points.

Chapter 7 demonstrates that the Error-Reduction Flow admits global weak solutions.

Chapter 8 establishes a correspondence between fixed points of the Error-Reduction algorithm and the Error-Reduction Flow. It analyses stability of these fixed points.

6 DISCRETIZATIONS OF AP AND ER FLOWS

This chapter discusses properties of explicit Euler discretizations of ERF and its generic version APF. Section 6.1 demonstrates that discretized APF (dAPF) dissipates energy, which is a generalization of Proposition 5.14 (Error-Reduction energy does not increase). Section 6.2 shows sufficient conditions under which accumulation points of dAPF exist. Section 6.3 uses specific properties of projections $\mathcal{P}$ and $\mathcal{M}$ to show that accumulation points of dERF exist and are fixed points. It also discusses how fixed point results can be generalized to other object space constraints $\mathcal{A}$, namely, by by intersecting $\mathcal{A}$ with other sets.

6.1 ENERGY DISSIPATION OF THE APF

The highlights of this section are Corollary 6.5 — demonstrating that dAPF dissipates energy — and Corollary 6.9 — demonstrating that the distance functional $\|P_{\mathcal{X}}[g_n] - P_{\mathcal{Y}}[g_n]\|_2$ does not increase for dAPF iterates g_n.

The first of these results (Corollary 6.5) is a generalization of the well-known Fienup's Error-Reduction property (Proposition 5.14).

Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal nonempty sets. It is easy to verify that the argument presented in Remark 5.17 remains valid in the following sense: minimization problem

$$\arg\min_{g \in \mathcal{H}} E_{\mathcal{X}}[g] + E_{\mathcal{Y}}[g]$$

leads to the formal gradient descent flow

$$\partial_t g = -(g - P_{\mathcal{X}}[g]) - (g - P_{\mathcal{Y}}[g]), \qquad \text{(APF)}$$

which — discretized (using explicit Euler) with step size $\varepsilon = 1$ and initialized at some $g_0 \in \mathcal{X}$ — leads to AP. If, additionally, $\mathcal{X}$ and $\mathcal{Y}$ are weakly closed, then APF is a selection of the (Mordukhovich-Kruger) subgradient of $E_{\mathcal{X}} + E_{\mathcal{Y}}$ (Remark 5.18).

The goal of this section is to show that the subdifferential selection (APF) — discretized (using explicit Euler) with step size $\varepsilon > 0$ and initialized at any $g_0 \in \mathcal{H}$ — leads to a generalization of AP we call dAPF.

This generalized algorithm exhibits energy dissipation properties similar to a gradient flow. Namely, since APF is a formal gradient descent — and assuming that $E := E_{\mathcal{X}} + E_{\mathcal{Y}}$ is Fréchet-differentiable at g that solves APF — one has

$$\frac{d}{dt}E[g] = -\|\nabla E[g]\|_2^2 = \|-2g + P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g]\|_2^2. \tag{6.1}$$

For a sequence $(g_n^{(\varepsilon)})_{n\in\mathbb{N}_0}$ generated by dAPF, we show that — even if E is not differentiable at g_n — one has

$$\frac{E[g_{n+1}^{(\varepsilon)}] - E[g_n^{(\varepsilon)}]}{\varepsilon} = -(1-\varepsilon)\left\|-2g_n^{(\varepsilon)} + P_{\mathcal{X}}[g_n^{(\varepsilon)}] + P_{\mathcal{Y}}[g_n^{(\varepsilon)}]\right\|_2^2. \tag{6.2}$$

for $\varepsilon \in (0, 1]$. This energy dissipation is a generalization of Fienup's error-reduction property (Proposition 5.14). Remarkably, it does not require weak closedness of $\mathcal{X}$ or $\mathcal{Y}$.

Moreover, as will de shown further below, for $\varepsilon \in (0, \frac{1}{2}]$, the functional $g \mapsto \|P_{\mathcal{X}}[g_n^{(\varepsilon)}] - P_{\mathcal{Y}}[g_n^{(\varepsilon)}]\|_2$ is Lyapunov (i. e. $\|P_{\mathcal{X}}[g_n^{(\varepsilon)}] - P_{\mathcal{Y}}[g_n^{(\varepsilon)}]\|_2$ is non-increasing with n).

DEFINITION 6.1 (DISCRETIZED AP FLOW). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal with projecton selections $P_{\mathcal{X}}, P_{\mathcal{Y}}$, let $g_0 \in \mathcal{H}$, let $\varepsilon > 0$. The sequence $(g_n^{(\varepsilon)})_{n\in\mathbb{N}}$ is generated by the dAPF algorithm, if $g_0^{(\varepsilon)} = g_0$ and*

$$g_{n+1}^{(\varepsilon)} = g_n^{(\varepsilon)} + \varepsilon(-2g_n^{(\varepsilon)} + P_{\mathcal{X}}[g_n^{(\varepsilon)}] + P_{\mathcal{Y}}[g_n^{(\varepsilon)}]) \tag{dAPF}$$

for all $n \in \mathbb{N}$. In this case, we call $(g_n^{(\varepsilon)})_{n\in\mathbb{N}_0}$ a dAPF sequence with initial value $g_0 \in \mathcal{H}$ and step size ε.

In certain cases it is more convenient to consider the multi-valued variant of the dAPF, defined by

$$g_{n+1}^{(\varepsilon)} \in g_n^{(\varepsilon)} + \varepsilon(-2g_n^{(\varepsilon)} + \Pi_{\mathcal{X}}[g_n^{(\varepsilon)}] + \Pi_{\mathcal{Y}}[g_n^{(\varepsilon)}])$$

for multi-valued projections $\Pi_{\mathcal{X}}, \Pi_{\mathcal{Y}}$.

Remark 6.2 (Relationship to Halpern's algorithm). In case when $\mathcal{X}$ and $\mathcal{Y}$ are convex, dAPF is a particular instance of the Halpern algorithm, see [BC17, Ch. 30]. Another particular instance of the Halpern algorithm is used below to establish continuity of projections for regularized additional constraints, see Definition 6.33, Theorem 6.34, and Lemma 6.37.

Remark 6.3. Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal, let $(f_n^{(\varepsilon)})_{n\in\mathbb{N}_0/2}$ be a dAPF sequence with initial value $f_0 \in \mathcal{X}$ and step size $\varepsilon = 1$. (The indexing of the sequence is chosen to be consistent with Remarks 5.16 and 5.17.) Let $(g_n)_{n\in\mathbb{N}_0}$ be an AP sequence with initial value f_0. By the same argument as in Remark 5.17, $g_n = f_n$ for all $n \in \mathbb{N}$.

PROPOSITION 6.4 (ENERGY DISSIPATION IN DAPF UPDATE). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal, let $g \in \mathcal{H}$, let $\varepsilon \in (0, 1]$. Use the dAPF update to define*

$$g^{(\varepsilon)} := g + \varepsilon(-2g + P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g]).$$

Use the notation $E := E_{\mathcal{X}} + E_{\mathcal{Y}}$. Then,

$$E[g^{(\varepsilon)}] - E[g] \leqslant -\varepsilon(1-\varepsilon)\,\|2g - P_{\mathcal{X}}[g] - P_{\mathcal{Y}}[g]\|_2^2. \tag{6.3}$$

Proof. The proof is a straightforward calculation. Start with the definition of $E[g^{(\varepsilon)}]$, use distance minimizing property (Corollary 3.6) to estimate

$$\begin{aligned}\|g^{(\varepsilon)} - P_{\mathcal{X}}[g^{(\varepsilon)}]\|_2^2 &\leqslant \|g^{(\varepsilon)} - P_{\mathcal{X}}[g]\|_2^2\\ \|g^{(\varepsilon)} - P_{\mathcal{Y}}[g^{(\varepsilon)}]\|_2^2 &\leqslant \|g^{(\varepsilon)} - P_{\mathcal{Y}}[g]\|_2^2.\end{aligned}$$

Insert the definition of $g^{(\varepsilon)}$ into E, get

$$\begin{aligned}E[g^{(\varepsilon)}] \leqslant\ & \frac{1}{2}\|\underbrace{(1-2\varepsilon)g + \varepsilon P_{\mathcal{X}}[g] + \varepsilon P_{\mathcal{Y}}[g] - P_{\mathcal{X}}[g]}_{=:\,T_1}\|_2^2\\ &+ \frac{1}{2}\|\underbrace{(1-2\varepsilon)g + \varepsilon P_{\mathcal{X}}[g] + \varepsilon P_{\mathcal{Y}}[g] - P_{\mathcal{Y}}[g]}_{=:\,T_2}\|_2^2.\end{aligned}$$

Use the rearrangements

$$\begin{aligned}T_1 &= (1-\varepsilon)\big(g - P_{\mathcal{X}}[g]\big) - \varepsilon\ \big(g - P_{\mathcal{Y}}[g]\big),\\ T_2 &= \ -\varepsilon\ \big(\underbrace{g - P_{\mathcal{X}}[g]}_{=:\,d_{\mathcal{X}}}\big) + (1-\varepsilon)\big(\underbrace{g - P_{\mathcal{Y}}[g]}_{=:\,d_{\mathcal{Y}}}\big)\end{aligned}$$

to expand the squares. The terms containing the factor $\frac{1}{2}\|d_{\mathcal{X}}\|_2^2$ will be

$$\big((1-\varepsilon)^2 + \varepsilon^2\big)\,\frac{1}{2}\|d_{\mathcal{X}}\|_2^2 = \frac{1}{2}\|d_{\mathcal{X}}\|_2^2 - \varepsilon(1-\varepsilon)\|d_{\mathcal{X}}\|_2^2;$$

the terms containing the factor $\frac{1}{2}\|d_{\mathcal{Y}}\|_2^2$ will be

$$\big((1-\varepsilon)^2 + \varepsilon^2\big)\,\frac{1}{2}\|d_{\mathcal{Y}}\|_2^2 = \frac{1}{2}\|d_{\mathcal{Y}}\|_2^2 - \varepsilon(1-\varepsilon)\|d_{\mathcal{Y}}\|_2^2;$$

the terms containing the factor $\langle d_{\mathcal{X}}, d_{\mathcal{Y}}\rangle$ will be

$$-\frac{1}{2}2\varepsilon(1-\varepsilon)\langle d_{\mathcal{X}}, d_{\mathcal{Y}}\rangle - \frac{1}{2}2\varepsilon(1-\varepsilon)\langle d_{\mathcal{X}}, d_{\mathcal{Y}}\rangle = -2\varepsilon(1-\varepsilon)\langle d_{\mathcal{X}}, d_{\mathcal{Y}}\rangle.$$

Combine together all terms with the factor $\varepsilon(1-\varepsilon)$ to obtain

$$\begin{aligned}E[g^{(\varepsilon)}] &\leqslant \frac{1}{2}\|d_{\mathcal{X}}\|_2^2 + \frac{1}{2}\|d_{\mathcal{Y}}\|_2^2 - \varepsilon(1-\varepsilon)\left(\|d_{\mathcal{X}}\|_2^2 + \|d_{\mathcal{Y}}\|_2^2 + 2\langle d_{\mathcal{X}}, d_{\mathcal{Y}}\rangle\right)\\ &= \qquad E[g] \qquad - \varepsilon(1-\varepsilon)\|2g - P_{\mathcal{X}}[g] - P_{\mathcal{Y}}[g]\|_2^2.\end{aligned}$$

□

From the previous proposition immediately follows

COROLLARY 6.5 (ENERGY DISSIPATION IN dAPF). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal, let $(g_n^{(\varepsilon)})_{n\in\mathbb{N}_0}$ be a dAPF sequence with initial value $g_0 \in \mathcal{H}$ and step size $\varepsilon \in (0,1]$. Use the notation $E := E_{\mathcal{X}} + E_{\mathcal{Y}}$. Then,*

$$E[g_{n+1}^{(\varepsilon)}] - E[g_n^{(\varepsilon)}] \leqslant -\varepsilon(1-\varepsilon)\left\|2g_n^{(\varepsilon)} - P_{\mathcal{X}}[g_n^{(\varepsilon)}] - P_{\mathcal{Y}}[g_n^{(\varepsilon)}]\right\|_2^2 \tag{6.4}$$

for all $n \in \mathbb{N}_0$. In particular, if E is differentiable at $g_n^{(\varepsilon)}$, this is equivalent to

$$\frac{E[g_{n+1}^{(\varepsilon)}] - E[g_n^{(\varepsilon)}]}{\varepsilon} \leqslant -(1-\varepsilon)\|\nabla E[g_n^{(\varepsilon)}]\|_2^2. \tag{6.5}$$

Remark 6.6 (Recovering Fienup's dissipation theorem). The previous corollary is a generalization of Fienup's Error-Reduction property (Proposition 5.14). Indeed, let $(g_n)_{n\in\mathbb{N}_0}$ be an ER sequence with initial value $g_0 \in \mathcal{P}$; let $(f_n^{(\varepsilon)})_{n\in\mathbb{N}_0/2}$ be a dAPF sequence with initial value $f_0 \in \mathcal{X}$, with step size $\varepsilon = 1$, and with $\mathcal{X} = \mathcal{M}$ and $\mathcal{Y} = \mathcal{P}$. By Remark 5.17, $g_n = f_n$, $f_n \in \mathcal{P}$ and $f_{n+1/2} \in \mathcal{M}$ for all $n \in \mathbb{N}_0$. Use the notation $E := E_{\mathcal{M}} + E_{\mathcal{P}}$. Then, by Corollary 6.5,

$$\begin{aligned} E_{\mathcal{M}}[g_{n+1}] - E_{\mathcal{M}}[g_n] &= E_{\mathcal{M}}[f_{n+1}] - E_{\mathcal{M}}[f_n] \\ &= E_{\mathcal{M}}[f_{n+1}] + \underbrace{E_{\mathcal{P}}[f_{n+1}]}_{=0} - E_{\mathcal{M}}[f_n] - \underbrace{E_{\mathcal{P}}[f_n]}_{=0} \\ &= E[f_{n+1}] - E[f_{n+1/2}] + E[f_{n+1/2}] - E[f_n] \\ &\leqslant 0, \end{aligned}$$

recovering Proposition 5.14.

COROLLARY 6.7 (QUADRATIC SUMMABILITY IN dAPF). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal, let $(g_n^{(\varepsilon)})_{n\in\mathbb{N}_0}$ be a dAPF sequence with initial value $g_0 \in \mathcal{H}$ and step size $\varepsilon \in (0,1)$. Use the notation $E := E_{\mathcal{X}} + E_{\mathcal{Y}}$. Then,*

$$\frac{1}{\varepsilon}\sum_{n=0}^{\infty}\|g_{n+1}^{(\varepsilon)} - g_n^{(\varepsilon)}\|_2^2 \leqslant \frac{1}{1-\varepsilon}E[g_0].$$

Proof. By a straightforward calculation (using Proposition 6.4 in $(*)$),

$$\begin{aligned} \sum_{n=0}^{\infty}\|g_{n+1}^{(\varepsilon)} - g_n^{(\varepsilon)}\|_2^2 &= \sum_{n=0}^{\infty}\varepsilon^2\|2g_n^{(\varepsilon)} - P_{\mathcal{X}}[g_n^{(\varepsilon)}] - P_{\mathcal{Y}}[g_n^{(\varepsilon)}]\|_2^2 \\ &= \sum_{n=0}^{\infty}\frac{\varepsilon}{1-\varepsilon}\varepsilon(1-\varepsilon)\|2g_n^{(\varepsilon)} - P_{\mathcal{X}}[g_n^{(\varepsilon)}] - P_{\mathcal{Y}}[g_n^{(\varepsilon)}]\|_2^2 \\ &\overset{(*)}{\leqslant} \frac{\varepsilon}{1-\varepsilon}\sum_{n=0}^{\infty}(E[g_n] - E[g_{n+1}]) \\ &\leqslant \frac{\varepsilon}{1-\varepsilon}(E[g_0] - \limsup_{n\to\infty} E[g_n]) \leqslant \frac{\varepsilon}{1-\varepsilon}E[g_0], \end{aligned}$$

since $E[g_n] \geqslant 0$ for all $n \in \mathbb{N}_0$. □

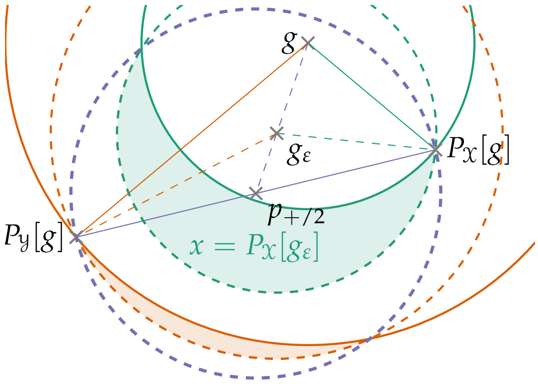

Figure 6.1: Illustration to Proposition 6.8 (dAPF does not increase projection difference).
Let the point g have projections $P_{\mathcal{X}}[g]$ and $P_{\mathcal{Y}}[g]$. Equations (6.7) and (6.8) state that projection $x = P_{\mathcal{X}}[g^{(\varepsilon)}]$ is not closer to g than $\|g - P_{\mathcal{X}}[g]\|_2$, – meaning that x is outside the cirle bound by solid green circumference, – and not farther from $g^{(\varepsilon)}$ than $\|g^{(\varepsilon)} - P_{\mathcal{X}}[g]\|_2$, – meaning x is inside the circle bound by dashed green circumference. Overall, this means that x is inside the green filled area. Analogous relations for $P_{\mathcal{Y}}[g^{(\varepsilon)}]$ are marked orange. The proposition shows that $P_{\mathcal{X}}[g^{(\varepsilon)}]$ and $P_{\mathcal{Y}}[g^{(\varepsilon)}]$ are within $\|P_{\mathcal{X}}[g] - P_{\mathcal{Y}}[g]\|_2/2$ distance to the midpoint $p_{+/2} = (P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g])/2$, – meaning that areas filled green and orange are inside the violet dashed circle.

Corollary 6.5 seems remarkable since it does not require any differentiability from $E_{\mathcal{X}} + E_{\mathcal{Y}}$, but instead uses distance-minimizing properties of a projection. A similar argument is exploited in the following proposition. It shows that projection difference $\|P_{\mathcal{X}}[g_n^{(\varepsilon)}] - P_{\mathcal{Y}}[g_n^{(\varepsilon)}]\|_2$ — which, in general, exhibits very poor Fréchet-differentiability — is a non-increasing quantity, albeit only for step sizes $\varepsilon \in (0, \frac{1}{2}]$.

Proposition 6.8 (dAPF update does not increase projection difference). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal, let $g \in \mathcal{H}$, let $\varepsilon \in (0, \frac{1}{2}]$. Use the dAPF update to define*

$$g^{(\varepsilon)} := g + \varepsilon(-2g + P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g]).$$

Use the notation $E := E_{\mathcal{X}} + E_{\mathcal{Y}}$. Then,

$$\left\|P_{\mathcal{X}}[g^{(\varepsilon)}] - P_{\mathcal{Y}}[g^{(\varepsilon)}]\right\|_2 \leqslant \left\|P_{\mathcal{X}}[g] - P_{\mathcal{Y}}[g]\right\|_2$$

for all $n \in \mathbb{N}$.

Proof. Let us show that

$$P_{\mathcal{X}}[g^{(\varepsilon)}] \in \overline{B_{\left\|\frac{P_{\mathcal{X}}[g] - P_{\mathcal{Y}}[g]}{2}\right\|_2}\left(\frac{P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g]}{2}\right)}, \tag{6.6}$$

and so does $P_{\mathcal{Y}}[g^{(\varepsilon)}]$. From this, the desired result will immediatly follow by triangle inequality.

Observe that by distance minimizing property (Corollary 3.6) holds:

$$\|g - P_{\mathcal{X}}[g]\|_2 \leqslant \|g - P_{\mathcal{X}}[g^{(\varepsilon)}]\|_2. \tag{6.7}$$

$$\|g^{(\varepsilon)} - P_{\mathcal{X}}[g^{(\varepsilon)}]\|_2 \leqslant \|g^{(\varepsilon)} - P_{\mathcal{X}}[g]\|, \tag{6.8}$$

see Figure 6.1. Our goal is to use these inequalities to show that

$$\left\| P_{\mathcal{X}}[g^{(\varepsilon)}] - \frac{P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g]}{2} \right\|_2 \leqslant \left\| \frac{P_{\mathcal{X}}[g] - P_{\mathcal{Y}}[g]}{2} \right\|, \quad \text{or}$$

$$\|x - p_{+/2}\|_2 \leqslant \|p_{\mathcal{X}} - p_{+/2}\|, \tag{6.9}$$

where we use the notation $p_{\mathcal{X}} = P_{\mathcal{X}}[g], p_{\mathcal{Y}} = P_{\mathcal{Y}}[g], p_{+/2} = (p_{\mathcal{X}} + p_{\mathcal{Y}})/2$ and $x = P_{\mathcal{X}}[g^{(\varepsilon)}]$ for brevity. Start with Equation (6.8) in the form

$$0 \leqslant \|g^{(\varepsilon)} - p_{\mathcal{X}}\|_2^2 - \|g^{(\varepsilon)} - x\|_2^2.$$

Insert definition of $g^{(\varepsilon)}$, rearrange the terms by 2ε and $1 - 2\varepsilon$ to get

$$\begin{aligned} 0 \leqslant \quad & \|(1-2\varepsilon)g + \varepsilon p_{\mathcal{X}} + \varepsilon p_{\mathcal{Y}} - p_{\mathcal{X}}\|_2^2 \quad - \quad \|(1-2\varepsilon)g + \varepsilon p_{\mathcal{X}} + \varepsilon p_{\mathcal{Y}} - x\|_2^2 \\ \leqslant \quad & \|(1-2\varepsilon)(g - p_{\mathcal{X}}) + \varepsilon(p_{\mathcal{Y}} - p_{\mathcal{X}})\|_2^2 \quad - \|(1-2\varepsilon)(g-x) + \varepsilon(p_{\mathcal{X}} + p_{\mathcal{Y}} - 2x)\|_2^2 \\ = & \|(1-2\varepsilon)(g - p_{\mathcal{X}}) + 2\varepsilon(p_{+/2} - p_{\mathcal{X}})\|_2^2 - \|(1-2\varepsilon)(g-x) + 2\varepsilon(p_{+/2} - x)\|_2^2. \end{aligned}$$

Apply cosine triangle theorem to open the brackets, rearrange the terms:

$$\begin{aligned} 0 \leqslant\ & (1-2\varepsilon)^2\left(\|g - p_{\mathcal{X}}\|^2 - \|g - x\|^2\right) + (2\varepsilon)^2\left(\|p_{+/2} - p_{\mathcal{X}}\|_2^2 - \|p_{+/2} - x\|_2^2\right) \\ & + 2(1-2\varepsilon)2\varepsilon\langle g - p_{\mathcal{X}}, p_{+/2} - p_{\mathcal{X}}\rangle - 2(1-2\varepsilon)2\varepsilon\langle g - x, p_{+/2} - x\rangle. \end{aligned}$$

By cosine theorem, for $(1-2\varepsilon)2\varepsilon$-terms holds

$$\begin{aligned} 2\langle g - p_{\mathcal{X}}, p_{+/2} - p_{\mathcal{X}}\rangle &= \|g - p_{\mathcal{X}}\|_2^2 + \|p_{+/2} - p_{\mathcal{X}}\|_2^2 - \|g - p_{+/2}\|_2^2, \\ -2\langle g - x, p_{+/2} - x\rangle &= -\|g - x\|_2^2 - \|p_{+/2} - x\|_2^2 + \|g - p_{+/2}\|_2^2. \end{aligned}$$

When inserted into estimation, the summands $\|g - p_{+/2}\|_2^2$ cancel out, and the remaining terms combine to

$$\begin{aligned} 0 \leqslant\ & (1-2\varepsilon)^2\left(\|g - p_{\mathcal{X}}\|^2 - \|g - x\|^2\right) + (2\varepsilon)^2\left(\|p_{+/2} - p_{\mathcal{X}}\|_2^2 - \|p_{+/2} - x\|_2^2\right) \\ & + (1-2\varepsilon)2\varepsilon\left(\|g - p_{\mathcal{X}}\|_2^2 + \|p_{+/2} - p_{\mathcal{X}}\|_2^2\right) - (1-2\varepsilon)2\varepsilon\left(\|g - x\|_2^2 + \|p_{+/2} - x\|_2^2\right) \\ & = (1-2\varepsilon)\left(\|g - p_{\mathcal{X}}\|^2 - \|g - x\|^2\right) + 2\varepsilon\left(\|p_{+/2} - p_{\mathcal{X}}\|_2 - \|p_{+/2} - x\|_2^2\right), \end{aligned}$$

where we use $(1-2\varepsilon)^2 + (1-2\varepsilon)\varepsilon = (1-2\varepsilon)$ and $(2\varepsilon)^2 + (1-2\varepsilon)\varepsilon = 2\varepsilon$ in the last equality. Therefore,

$$\|p_{+/2} - p_{\mathcal{X}}\|_2^2 - \|p_{+/2} - x\|_2^2 \geqslant \frac{1-2\varepsilon}{2\varepsilon}\left(-\|g - p_{\mathcal{X}}\|_2^2 + \|g - x\|_2^2\right) \geqslant 0,$$

where the last inequality holds by Equation (6.7) for all $2\varepsilon \in (0,1]$. Therefore, Equation (6.9) is true; therefore, inclusion (6.6) is true. Analogously, $P_{\mathcal{Y}}[g^{(\varepsilon)}] \in \overline{B_{\|\frac{p_{\mathcal{X}}-p_{\mathcal{Y}}}{2}\|_2}\left(\frac{p_{\mathcal{X}}+p_{\mathcal{Y}}}{2}\right)}$. □

From the previous proposition immediately follows

Corollary 6.9 (dAPF does not increase projection difference). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be weakly closed, let $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ be a dAPF sequence with initial value $g_0 \in \mathcal{H}$ and step size $\varepsilon \in (0, \frac{1}{2}]$. Then,*

$$\left\|P_{\mathcal{X}}[g_{n+1}^{(\varepsilon)}] - P_{\mathcal{Y}}[g_{n+1}^{(\varepsilon)}]\right\|_2 \leqslant \left\|P_{\mathcal{X}}[g_n^{(\varepsilon)}] - P_{\mathcal{Y}}[g_n^{(\varepsilon)}]\right\|_2$$

for all $n \in \mathbb{N}_0$.

6.2 accumulation points of dapf

This section contains minor technical results sufficient for boundedness of dAPF. Namely, it introduces Definition 6.10 which is a sufficient assumption for dAPF to be bounded (Proposition 6.13).

Definition 6.10 (AP-bounded operators). *Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a Banach space. We call the operator $P\colon \mathcal{B} \to \mathcal{B}$ AP-bounded on $\mathcal{B}$, if at least one of the following conditions holds:*

i) there exists $C \in [0;\infty)$ such that $\|P[f]\|_{\mathcal{B}} \leqslant C$ for all $f \in \mathcal{B}$;

ii) $\|P[f]\|_{\mathcal{B}} \leqslant \|f\|_{\mathcal{B}}$ for all $f \in \mathcal{B}$.

Remark 6.11 (Why on a Banach space). For the purposes of this section, the choice $(\mathcal{B}, \|\cdot\|_{\mathcal{B}}) = (\mathcal{H}, \|\cdot\|_2)$ is sufficient. In Section 6.3, AP-boundedness is used to establish properties of ER on more general spaces — for example, on Sobolev spaces: $(\mathcal{B}, \|\cdot\|_{\mathcal{B}}) = (W^{1,p}(\mathbb{T}^d), \|\cdot\|_{W^{1,p}})$ for $p \in [1, \infty)$.

Example 6.12.

i) Let $\sqrt{I} \in \widehat{\mathcal{H}}$ be non-negative. The projection $P_{\mathcal{M}(\sqrt{I})}$ is AP-bounded on $\mathcal{H}$, since for all $f \in \mathcal{H}$ holds

$$\|P_{\mathcal{M}}[f]\|_2^2 = C_{\mathcal{F}}\|\sqrt{I}\|_2^2 < \infty$$

by Plancherel's theorem.

ii) Let $\mathcal{X}$ be such that $P_{\mathcal{X}}$ is an indicator projection (Definition 5.3). Then, $P_{\mathcal{X}}$ is AP-bounded on $\mathcal{H}$, since for any $f \in \mathcal{H}$ holds

$$\|P_{\mathcal{X}}[f]\|_2 = \|\mathbb{1}_{S[f]} f\|_2 \leqslant \|f\|_2.$$

In particular $P_{\mathcal{P}}, P_{\mathcal{S}}$ and $P_{\mathcal{T}_s(\nu)}$ are AP-bounded on $\mathcal{H}$.

iii) Let $\alpha > 0$, let $\Omega = \mathbb{R}$. The projection $P_{\mathcal{T}_a(\alpha)}$ is not AP-bounded on $\mathcal{H}$, since for any $R > 0$,

$$\left\| P_{\mathcal{T}_a(\alpha)} \left[\frac{3\alpha}{4} \mathbb{1}_{[-R/2,R/2]} \right] \right\|_2 = \|\alpha \mathbb{1}_{[-R/2,R/2]}\|_2 = \alpha R,$$

which is i) larger than any $C < \infty$ for large enough R; and ii) larger than

$$\frac{3}{4}\alpha R = \left\| \frac{3\alpha}{4} \mathbb{1}_{[-R/2,R/2]} \right\|_2 .$$

PROPOSITION 6.13 (DAPF IS BOUNDED). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal. Let $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ be a dAPF sequence with initial value $g_0 \in \mathcal{H}$ and step size $\varepsilon \in (0, 1/2]$.*

i) If $P_\mathcal{X}$ and $P_\mathcal{Y}$ are AP-bounded on $\mathcal{H}$, then $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ is uniformly bounded in n, i.e. there exists $C < \infty$ such that $\|g_n^{(\varepsilon)}\|_2 \leqslant C$ for all $n \in \mathbb{N}_0$.

ii) Let $(\mathcal{B}, \|\cdot\|_\mathcal{B})$ is a Banach space with $\mathcal{B} \subset \mathcal{H}$. If $P_\mathcal{X}$ and $P_\mathcal{Y}$ map $\mathcal{B}$ to $\mathcal{B}$ and are AP-bounded on $\mathcal{B}$, and if $g_0 \in \mathcal{B}$, then $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ is uniformly bounded in n on $\mathcal{B}$, i.e. there exists $C < \infty$ such that $\|g_n^{(\varepsilon)}\|_\mathcal{B} \leqslant C$ for all $n \in \mathbb{N}_0$.

Proof. i) Assume that $\|P_\mathcal{X}[f]\|_2 \leqslant C_\mathcal{X}$ for all $f \in \mathcal{H}$, assume that $\|P_\mathcal{Y}[f]\|_2 \leqslant \|f\|_2$ for all $f \in \mathcal{H}$. (The proof is similar for any other combination of AP-boundedness conditions.)

We claim that for all $n \in \mathbb{N}_0$

$$\|g_n^{(\varepsilon)}\|_2 \leqslant C := \max\{\|g_0\|_2, C_\mathcal{X}\}.$$

The claim is proven by induction. It is obviously true for $n = 0$. Assuming the claim is true for some $n \in \mathbb{N}$, have

$$\begin{aligned} \|g_{n+1}^{(\varepsilon)}\|_2 &\leqslant (1-2\varepsilon)\|g_n^{(\varepsilon)}\|_2 + \varepsilon\|P_\mathcal{X}[g_n^{(\varepsilon)}]\|_2 + \varepsilon\|P_\mathcal{Y}[g_n^{(\varepsilon)}]\|_2 \\ &\leqslant (1-2\varepsilon)\|g_n^{(\varepsilon)}\|_2 + \varepsilon C_\mathcal{X} + \varepsilon\|P_\mathcal{Y}[g_n^{(\varepsilon)}]\|_2 \\ &\leqslant \max\{\|g_n\|_2, C_\mathcal{X}\}. \end{aligned} \tag{6.10}$$

Since the induction step Equation (6.10) uses only triangle inequality, the proof works analogously for case ii). □

COROLLARY 6.14 (WEAK ACCUMULATION POINTS OF DAPF EXIST). *Let $\mathcal{X}$, $\mathcal{Y} \subset \mathcal{H}$ be proximal; assume that $P_\mathcal{X}$, $P_\mathcal{Y}$ are AP-bounded. Let $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ be a dAPF sequence with initial value $g_0 \in \mathcal{H}$ and step size $\varepsilon \in (0, 1/2]$.*

Then, there exists a weak accumulation point $g \in \mathcal{H}$, such that $g_n^{(\varepsilon)} \rightharpoonup g$ (convergences weakly to g) as $n \in M$ for an appropriate infinite $M \subset \mathbb{N}$.

Proof. The claim follows from Proposition 6.13 and the Banach-Alaoglu theorem (closed balls are weakly compact in Hilbert spaces). □

Example 6.15. Let $\sqrt{I} \in \widehat{\mathcal{H}}$ be non-negative, let $g_0 \in \mathcal{P}$, let $\varepsilon \in (0, \frac{1}{2}] \cup \{1\}$. Let $\mathcal{X} = \mathcal{M}, \mathcal{Y} = \mathcal{P}$. The resulting dAPF sequence $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ is bounded by $\max\{\|g_0\|_2, \|\check{\sqrt{I}}\|_2\}$ and has weak accumulation points.

Remark 6.16 (Fixed points of dAPF). If one assumes that $P_\mathcal{X}, P_\mathcal{Y}$ are weakly sequentially continuous, then one can show that weak accumulation points of dAPF are fixed points. The proof is essentially the same as the one demonstrated for the strong topology in Proposition 6.32.

However, the assumption for $P_\mathcal{X}, P_\mathcal{Y}$ to be weakly sequentially continuous is unrealistic in praxis. For example, let $\Omega = \mathbb{T}^1$, consider the sequence $(g_n)_{n \in \mathbb{N}}$ with $g_n(x) = \sin(nx)$ for all $n \in \mathbb{N}$. Then, $g_n \rightharpoonup 0$, but

$$\begin{aligned}\langle P_\mathcal{P}[g_n], \mathbb{1}_{\mathbb{T}^1} \rangle &= \int_0^{2\pi} P_\mathcal{P}[\sin(nx)]\, dx = n \int_0^{\pi/n} \sin(nx)\, dx = \\ &= n \cdot \left(-\frac{\cos(nx)}{n} \right)\Big|_0^{\pi/n} = 2 \neq \langle 0, \mathbb{1}_{\mathbb{T}^1} \rangle,\end{aligned}$$

meaning that the positivity functional is not weakly sequentially continuous.

The following proposition shows that if one has strong convergence to an accumulation point, then it is a fixed point. This result will be applicable for phase retrieval on bounded domains (demonstrated in the next section) due to strong properties of $P_\mathcal{P}$ and $P_\mathcal{M}$ (in this setting of a bounded domain).

PROPOSITION 6.17 (FIXED POINTS OF DAPF). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be proximal, let $\varepsilon \in (0, 1]$, let the sequence $(g_n)_{n \in \mathbb{N}}$ be generated by the multi-valued dAPF algorithm (6.1).*

If some $g \in \mathcal{H}$ is a strong accumulation point of (g_n), then it is a fixed point, meaning that

$$0 \in 2g - \Pi_\mathcal{X}[g] - \Pi_\mathcal{Y}[g].$$

Proof. Consider a subsequence of $(g_n)_n$, again denoted by $(g_n)_n$, such that $g_n \to g$ as $n \to \infty$. Then, by Proposition 3.12, there exist projecton selections $P_\mathcal{X}$, $P_\mathcal{Y}$, and a further subsequence, again denoted by $(g_n)_n$, such that

$$\begin{aligned} P_\mathcal{X}[g_n] &\rightharpoonup P_\mathcal{X}[g], &&\text{and} \\ P_\mathcal{Y}[g_n] &\rightharpoonup P_\mathcal{Y}[g] &&\text{as } n \to \infty. \end{aligned}$$

Assume that $2g \neq P_\mathcal{X}[g] + P_\mathcal{Y}[g]$, meaning that

$$\|2g - P_\mathcal{X}[g] - P_\mathcal{Y}[g]\|_2^2 = \alpha > 0$$

for some α. Since $2g_n - P_\mathcal{X}[g_n] - P_\mathcal{Y}[g_n] \rightharpoonup 2g - P_\mathcal{X}[g] - P_\mathcal{Y}[g]$, by weak l.s.c. of the squared norm $\|\cdot\|^2$ have

$$\alpha = \|2g - P_\mathcal{X}[g] - P_\mathcal{Y}[g]\|_2^2 = \liminf_{n \to \infty} \|2g_n - P_\mathcal{X}[g_n] - P_\mathcal{Y}[g_n]\|_2^2,$$

meaning that the sum

$$\sum_{n=0}^{\infty} \left\| 2g_n^{(\varepsilon)} - P_{\mathcal{X}}[g_n^{(\varepsilon)}] - P_{\mathcal{Y}}[g_n^{(\varepsilon)}] \right\|_2^2$$

is unbounded, in contradiction to Corollary 6.5. Thus, $\|2g - P_{\mathcal{X}}[g] - P_{\mathcal{Y}}[g]\|_2^2 = 0$, and the claim follows. □

6.3 FIXED POINTS OF DERF

This section shows how obtained results for dAPF can be strengthened for the particular case of phase retrieval.

Namely, Section 6.3.1 formalizes the setting required to analyse dERF on $W^{1,p}(\Omega)$ where $\Omega = \mathbb{T}^d$ or $\Omega \subset \mathbb{R}^d$ is measurable and bounded. Section 6.3.2 establishes technical properties of $P_{\mathcal{P}}$ and $P_{\mathcal{M}}$ that ensure boundedness of dERF. Section 6.3.3 shows that accumulation points of dERF are fixed points. In conclusion, Section 6.3.4 demonstrates an examlple on how the above results can be generalized to additional constraints other than $P_{\mathcal{P}}$.

6.3.1 *Discretized ERF*

This subsection formalizes the precise setting of dERF.

DEFINITION 6.18 (POSITIVITY AND MODULUS FUNCTIONALS ON L^p).

i) Let $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d\}$, let $p \in [1, \infty]$. Then, $P_{\mathcal{P}}$, as defined in Example 3.14, maps $L^p(\Omega)$ to $L^p(\Omega)$.

ii) Let $\Omega = \mathbb{T}^d$ or let $\Omega \subset \mathbb{R}^d$ be measurable and bounded, let $p \in [1, \infty]$, let $\sqrt{I} \in L^1(\Omega_F)$, let $\varphi\colon \Omega_F \to [0; 2\pi)$ be measurable and such that $\sin\varphi$ is odd.

Then, $P_{\mathcal{M};\varphi}$, as defined in Example 3.24, maps $L^p(\Omega)$ to $L^p(\Omega)$.

Proof. The statement i) follows from $P_{\mathcal{P}}$ being an indicator projection. Indeed, for any indicator projection $P_{\mathcal{X}}[g] = \mathbb{1}_{S[g]}g$ have $\|P_{\mathcal{X}}[g]\|_{L^p} \leqslant \|g\|_{L^p}$.

For $P_{\mathcal{M};\varphi}$, recall that $L^p(\Omega) \subseteq L^1(\Omega)$ for bounded Ω; thus, Fourier transform is well-defined. Further, for any $g \in L^p(\mathbb{T}^d)$

$$\|P_{\mathcal{M};\varphi}[g]\|_p^p \leqslant (2\pi)^d \|P_{\mathcal{M};\varphi}[g]\|_\infty^p \leqslant (2\pi)^d \left\| \sum_{k\in\mathbb{Z}^d} |\sqrt{I}(k)| \right\|_\infty^p < \infty. \quad \square$$

DEFINITION 6.19 (DISCRETIZED ER FLOW ON A BOUNDED DOMAIN). *Let $\Omega = \mathbb{T}^d$ or let Ω be a bounded measurable subset of $\mathbb{R}^d$, let $\sqrt{I} \in L^1(\Omega)$, let $g_0 \in L^p(\Omega)$ for $p \in [1, \infty]$. Let $\varepsilon > 0$; let $\mathcal{A}$ represent an additional con-*

straint such that $P_{\mathcal{A}}$ is well-defined as an operator on $L^p(\Omega)$. The sequence $(g_n^{(\varepsilon)})_{n\in\mathbb{N}}$ is generated by the dERF algorithm, if $g_0^{(\varepsilon)} = g_0$ and

$$g_{n+1}^{(\varepsilon)} = g_n^{(\varepsilon)} + \varepsilon(-2g_n^{(\varepsilon)} + P_{\mathcal{M}}[g_n^{(\varepsilon)}] + P_{\mathcal{A}}[g_n^{(\varepsilon)}]) \tag{dERF}$$

for all $n \in \mathbb{N}$. In this case, we call $(g_n^{(\varepsilon)})_{n\in\mathbb{N}_0}$ a dERF sequence with initial value $g_0 \in \mathcal{H}$ and step size ε.

Unless explicitly mentioned otherwise, we use positivity as the additional constraint: $\mathcal{A} = \mathcal{P}$.

For $p = 2$, dERF is a particular case of dAPF (for $\Omega = \mathbb{T}^d$, $\mathcal{X} = \mathcal{M}$, and $\mathcal{Y} = \mathcal{P}$).

Remark 6.20 (Why bounded domain). When ERF is considered on an unbounded domain such as $\Omega = \mathbb{R}^d$, lacking regularity of $P_{\mathcal{M}}$ in Fourier space (lacking decay of $P_{\mathcal{M}}$ in object space) causes difficulties.

Indeed, if at some region in Fourier space $|\hat{f}| \ll \sqrt{I}$ — meaning that $|\hat{f}(k)| \ll \sqrt{I}(k)$ for a. a. $k \in U$, where U is a measurable subset of Ω_F with $\lambda(U) > 0$ — then $\sqrt{I}\frac{\hat{f}}{|\hat{f}|}$ can be much less regular than $\hat{f}$.

This difficulty can be tackled in several ways.

One can attempt to regularize $P_{\mathcal{M}}$ in the appropriate region. The challenge in this case is to keep the relevant geometric properties (projection properties) intact.

Alternatively, one can modify the object space constraint in a way that leads dERF to avoid regions where $|\hat{f}| \ll \sqrt{I}$. For example, if the positivity projection is restricted to an operator $P_{\tilde{\mathcal{P}}}$ which is pointwise parallel to $\hat{f}$ in Fourier space — i. e.,

$$\mathrm{Re}\big(\widehat{P_{\tilde{\mathcal{P}}}}[f](k)^* \hat{f}(k)\big) \geqslant 0, \qquad \text{for all } k \in \mathbb{Z}^d$$

— then

$$|\widehat{P_{\tilde{\mathcal{P}}}}[f](k) + \widehat{P_{\mathcal{M}}}[f](k)|^2 \geqslant |\widehat{P_{\tilde{\mathcal{P}}}}[f](k)|^2 + |\sqrt{I}(k)|^2,$$

avoiding the region where $|\hat{f}| \ll \sqrt{I}$ altogether. However, in this case $\tilde{\mathcal{P}}$ depends on f and changes with each iteration, but this does not necessarily breaks properties such as energy dissipation.

If an appropriate regularization is established, we conjecture that tools such as Pego's compactness criteria [Peg85] can extend results of this section and Chapter 7 to unbounded domains.

6.3.2 *Positivity and modulus projection properties*

This subsection contains technical results regarding properties of $P_{\mathcal{P}}$ and $P_{\mathcal{M}}$, and demonstrates the boundedness of dERF in the Sobolev space $W^{1,p}(\Omega)$ (Proposition 6.27).

The technical results on $P_{\mathcal{P}}$ and $P_{\mathcal{M}}$ are simple in nature and folklorically known. For example, the statement of Lemma 6.24 is taken from [Eva10].

Lemma 6.21 ($P_{\mathcal{P}}$ is sequentially continuous). *Let $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$, or let $\Omega \subset \mathbb{R}^d$ be measurable and bounded. Let $p \in [1, \infty]$, let $(g_n)_n$ be a sequence in $L^p(\Omega)$ that converges to some $g \in L^p(\Omega)$. Then, $\lim_{n\to\infty} P_{\mathcal{P}}[g_n] = P_{\mathcal{P}}[g]$.*

Proof. Proof by a direct calculation: observe that $2P_{\mathcal{P}}[g](x) = g(x) + |g|(x)$ for a. a. $x \in \Omega$. Further, observe that for any $a, b \in \mathbb{R}$ holds

$$|a| - |b| \leqslant |a - b| \quad \text{and} \quad |b| - |a| \leqslant |a - b| \quad \Rightarrow \quad \big||a| - |b|\big| \leqslant |a - b|$$

by triangle inequality. Therefore,

$$2\|P_{\mathcal{P}}[g_n] - P_{\mathcal{P}}[g]\|_{L^p} \leqslant \|g_n - g\|_{L^p} + \big\||g_n| - |g|\big\|_{L^p} \leqslant 2\|g_n - g\|_{L^p} \to 0$$

as $n \to \infty$.

Alternatively, for $p = 2$ the claim follows from $P_{\mathcal{P}}$ being firmly nonexpansive as a projection onto a convex set (Example 5.24). □

For later use, observe the following property of the modulus functional.

Lemma 6.22 ($P_{\mathcal{M}}$ is Lipshitz continuous). *Let $\Omega, \sqrt{I}$ be as in Definition 2.1; assume that $C_{\sqrt{I}/|\hat{g}|} := \left\|\frac{\sqrt{I}}{\hat{g}}\right\|_\infty < \infty$. Let $\varphi\colon \Omega_F \to [0; 2\pi)$ be such that $\sin\varphi$ is odd. Then, $\|P_{\mathcal{M};\varphi}[g] - P_{\mathcal{M};\varphi}[f]\|_2 \leqslant 2C_{\sqrt{I}/|\hat{g}|}\|g - f\|_2$ for all $f \in L^2(\Omega)$.*

Proof. Note that the statement, considered pointwise in Fourier space, is trivial for k where $\sqrt{I}(k) = 0$. Further note that for k where $\sqrt{I}(k) \neq 0$ one has $\hat{g}(k) \neq 0$ due to the assumption $C_{\sqrt{I}/|\hat{g}|} < \infty$.

For a. a. $k \in \Omega_F$ where $\sqrt{I}(k) \neq 0$ and $\hat{f}(k) \neq 0$ have

$$\begin{aligned}\left|\sqrt{I}(k)\frac{\hat{g}(k)}{|\hat{g}(k)|} - \sqrt{I}(k)\frac{\hat{f}(k)}{|\hat{f}(k)|}\right| \\ \leqslant \left\|\frac{\sqrt{I}}{|\hat{g}|}\right\|_\infty \left|\hat{g}(k) - |\hat{g}(k)|\frac{\hat{f}(k)}{|\hat{f}(k)|}\right| \leqslant 2\left\|\frac{\sqrt{I}}{|\hat{g}|}\right\|_\infty \left|\hat{g}(k) - \hat{f}(k)\right|,\end{aligned}$$

where in the last inequality we used

$$\left|a - \frac{b}{|b|}|a|\right| \leqslant 2|a - b| \qquad \text{for } a, b \in \mathbb{C}, b \neq 0,$$

which is proved in Lemma D.1.

For a. a. $k \in \Omega_F$ where $\sqrt{I}(k) \neq 0$ and $\hat{f}(k) = 0$ have

$$\begin{aligned}\left|\sqrt{I}(k)\frac{\hat{g}(k)}{|\hat{g}(k)|} - \sqrt{I}(k)e^{i\varphi(k)}\right| \\ \leqslant \left\|\frac{\sqrt{I}}{|\hat{g}|}\right\|_\infty \underbrace{\left|\hat{g}(k) - |\hat{g}(k)|e^{i\varphi(k)}\right|}_{\leqslant 2|\hat{g}(k)| = 2|\hat{g}(k) - \hat{f}(k)|} \leqslant 2\left\|\frac{\sqrt{I}}{|\hat{g}|}\right\|_\infty \left|\hat{g}(k) - \hat{f}(k)\right|.\end{aligned}$$

Therefore, and by Plancherel's theorem, have

$$\begin{aligned}
&\|P_{\mathcal{M}}[g] - P_{\mathcal{M}}[f]\|_2 \\
&= C_{\mathcal{F}} \left\| \sqrt{I}\frac{\hat{g}}{|\hat{g}|}\mathbb{1}_{\operatorname{supp}\sqrt{I}} - \sqrt{I}\frac{\hat{f}}{|\hat{f}|}\mathbb{1}_{\operatorname{supp}\sqrt{I}\cap\{\hat{f}\neq 0\}} - \sqrt{I}e^{i\varphi}\mathbb{1}_{\operatorname{supp}\sqrt{I}\cap\{\hat{f}=0\}} \right\|_2 \\
&\leqslant 2C_{\mathcal{F}} \left\| \frac{\sqrt{I}}{|\hat{g}|} \right\|_\infty \left\| \hat{g} - \hat{f} \right\|_2 \leqslant 2 \left\| \frac{\sqrt{I}}{|\hat{g}|} \right\|_\infty \|g - f\|_2 \qquad \square
\end{aligned}$$

Lemma 6.23 ($P_{\mathcal{M}}$ is subsequentially continuous: Hilbert space case). *Let one of the following hold:*

- *Let $\Omega = \mathbb{T}^d$, $\Omega_F = \mathbb{Z}^d$, let $\sqrt{I} \in \ell^2(\mathbb{Z}^d)$.*
- *Let Ω be a bounded measurable subset of $\mathbb{R}^d$, let $\Omega_F = \mathbb{R}^d$, let $\sqrt{I} \in L^2(\mathbb{R}^d)$.*

Further, let $\varphi\colon \Omega \to [0; 2\pi)$ be measurable and such that $\sin\varphi$ is odd.

Let $(g_n)_{n\in\mathbb{N}}$ converge to g in $\mathcal{H} = L^2(\Omega)$. Then, there exist:

i) a measurable $\psi\colon \Omega \to [0; 2\pi)$ such that $\sin\psi$ is odd, and

ii) an subsequence of $(P_{\mathcal{M};\varphi}[g_n])_{n\in\mathbb{N}}$ that converges to $P_{\mathcal{M};\psi}[g]$ in $\mathcal{H}$.

Proof. For bounded Ω, the set $\mathcal{M}(\sqrt{I})$ is (strongly) compact (Proposition D.2, Corollary D.3). Thus, there exists a subsequence of $(g_n)_n$, again denoted by $(g_n)_n$, such that $(P_{\mathcal{M};\varphi}[g_n])_{n\in\mathbb{N}}$ that converges to some $f \in \mathcal{M}(\sqrt{I})$.

Further, by Proposition 3.12, there exists a subsubsequence, again denoted by $(g_n)_n$, that converges to some $\tilde{f} \in \Pi_{\mathcal{M}}[g]$, meaning that $\tilde{f} = P_{\mathcal{M};\psi}[g]$ for some measurable $\psi\colon \Omega \to [0; 2\pi)$ such that $\sin\psi$ is odd.

Thus, (g_n) converges strongly to f and weakly to $\tilde{f}$, meaning that $f = \tilde{f}$. $\square$

The following lemma is from [Eva10, Section 5.10, Problem 17].

Lemma 6.24 (Regularity of $P_{\mathcal{P}}$). *Let $p \in [1, \infty)$, let $g \in W^{1,p}(\Omega)$, where $\Omega = \mathbb{T}^d$, or Ω is a bounded subset of $\mathbb{R}^d$. Then, $P_{\mathcal{P}}[g] \in W^{1,p}(\Omega)$, and*

$$\nabla P_{\mathcal{P}}[g] = \mathbb{1}_{g>0}\nabla g.$$

In particular, $\|P_{\mathcal{P}}[g]\|_{W^{1,p}} \leqslant \|g\|_{W^{1,p}}$ for all $g \in W^{1,p}$, meaning that $P_{\mathcal{P}}$ is AP-bounded on $W^{1,p}$.

Proof. For $\varepsilon > 0$, define

$$F^\varepsilon[g] := \begin{cases} \sqrt{g^2+\varepsilon^2} - \varepsilon & \text{if } g \geqslant 0; \\ 0 & \text{else.} \end{cases}$$

Then, $F^\varepsilon[g] \in W^{1,p}(\Omega)$. Indeed,

$$\int_\Omega \left(\sqrt{g^2(x)+\varepsilon^2}-\varepsilon\right)^p dx = \int_\Omega |g|^p \leqslant \|g\|^p_{L^p(\Omega)},$$

since $\sqrt{g^2(x)+\varepsilon^2}-\varepsilon > 0$ and since $\sqrt{g^2(x)+\varepsilon^2} \leqslant g(x)+\varepsilon$ pointwise a. e. by concavity of $\sqrt{\ }$. Further, at any $x \in \Omega$ with $g(x) > 0$ holds

$$\partial_{x_i} F^\varepsilon[g](x) = \frac{1}{2\sqrt{g^2(x)+\varepsilon^2}} 2g(x)\partial_{x_i} g(x),$$

where $\partial_{x_i} g$ is the weak derivative of g for $i \in \{1,\dots,d\}$. Therefore,

$$\int_\Omega (\partial_{x_i} F^\varepsilon[g](x))^p = \int_\Omega \Big(\underbrace{\frac{g(x)}{\sqrt{g^2(x)+\varepsilon^2}}}_{\leqslant 1} \partial_{x_i} g(x)\Big)^p dx \leqslant \|\partial_{x_i} g\|^p_{L^p},$$

so that $g \in W^{1,p}(\Omega)$.

It is obvious that $F^\varepsilon[g] \to P_{\mathcal{P}}[g]$ pointwise a. e. as $\varepsilon \to 0$. Let us show that $F^\varepsilon[g] \to P_{\mathcal{P}}[g]$ in L^p, and that $\partial_{x_i} F^\varepsilon[g] \to \mathbb{1}_{g>0}\partial_{x_i} g$ in L^p for $i \in \{1,\dots,d\}$ as $\varepsilon \to 0$.

Indeed, using $\sqrt{g^2(x)+\varepsilon^2} - g(x) \leqslant \varepsilon$ and boundedness of Ω, obtain

$$\begin{aligned}
&\int_\Omega \left(\sqrt{g^2(x)+\varepsilon^2}-\varepsilon-P_{\mathcal{P}}[g](x)\right)^p dx \\
&\qquad = \int_{\{g>0\}} \left(\sqrt{g^2(x)+\varepsilon^2}-\varepsilon-g(x)\right)^p dx \\
&\qquad\qquad = \int_{\{g>0\}} (2\varepsilon)^p\, dx \leqslant (2\pi)^d (2\varepsilon)^p \to 0
\end{aligned}$$

as $\varepsilon \to 0$. Further,

$$\begin{aligned}
&\int_\Omega \left(\partial_{x_i} F^\varepsilon[g](x) - \mathbb{1}_{g>0}\partial_{x_i} g(x)\right)^p dx \\
&\qquad = \int_{\{g>0\}} \left(\frac{g(x)}{\sqrt{g^2(x)+\varepsilon^2}}\partial_{x_i} g(x) - \partial_{x_i} g\right)^p dx \to 0
\end{aligned}$$

as $\varepsilon \to 0$ by the dominated convergence theorem with the majorant $2\partial_{x_i} g(x)$, since $\frac{g}{\sqrt{g^2(x)+\varepsilon^2}} \leqslant 1$ for a. a. $x \in \Omega$. □

Remark 6.25 (Positivity functional does not belong to $W^{1,\infty}$). Note that the last step of the proof is not valid for $p = \infty$, as the dominated convergence theorem is not applicable, and there exist $g \in W^{1,p}(\Omega)$ such that

$$\limsup_{\varepsilon\to 0} \left\| \frac{g(x)}{\sqrt{g^2(x)+\varepsilon^2}} - 1 \right\|_\infty \geqslant 1. \tag{6.11}$$

Indeed, observe that

$$\lim_{\varepsilon\to 0} \frac{\varepsilon^2}{\sqrt{\varepsilon^4+\varepsilon^2}} - 1 = \lim_{\varepsilon\to 0} \frac{\varepsilon}{\sqrt{\varepsilon^2+1}} - 1 = -1,$$

which means that if ε^2 is in range of g — which is possible for g which range includes a neighborhood of zero — inequality (6.11) holds, and $\|\nabla F^\varepsilon[g] - \mathbb{1}_{g>0}\nabla g\|_\infty \nrightarrow 0$ as $\varepsilon \to 0$.

LEMMA 6.26 (REGULARITY OF $P_{\mathcal{M}}$). *Let $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$, or let $\Omega \subset \mathbb{R}^d$ be measurable and bounded. Use $\widehat{\mathcal{H}}(\Omega_F)$ to denote the Fourier dual of $\mathcal{H}(\Omega)$. Let $\sqrt{I} \in \widehat{\mathcal{H}}(\Omega_F)$.*

i) If $\sqrt{I}$ satisfies

$$\|\check{\sqrt{I}}\|^2_{H^1} = C_{\mathcal{F}} \int_{\Omega_F} (1+|k|^2)\sqrt{I}^2(k)\,dk < \infty,$$

then $P_{\mathcal{M}}[f] \in H^1(\Omega)$ with $\|P_{\mathcal{M}}[f]\|_{H^1} = \|\check{\sqrt{I}}\|_{H^1}$ for any $f \in \mathcal{H}(\Omega)$.

ii) If $\sqrt{I}$ satisfies

$$\int_{\Omega_F} (1+|k|)\sqrt{I}(k)\,dk < \infty,$$

then $P_{\mathcal{M}}[f] \in W^{1,\infty}(\Omega)$ for any $f \in \mathcal{H}(\Omega)$.

iii) If Ω is a bounded subset of $\mathbb{R}^d$ or if $\Omega = \mathbb{T}^d$, and if $\sqrt{I}$ satisfies $\int(1+|k|)\sqrt{I}(k)\,dk < \infty$, then $P_{\mathcal{M}}[f] \in W^{1,p}(\Omega)$ for any $f \in \mathcal{H}(\Omega)$ and $p \in [1,\infty)$.

In particular, $P_{\mathcal{M}}$ is AP-bounded on $W^{1,p}(\Omega)$ for bounded Ω, for $p \in [1,\infty]$ with the constant $\lambda(\Omega)^{1/p}\int(1+|k|)\sqrt{I}(k)\,dk$.

Proof. All statements follow directly by Fourier calculus.

i) By Plancherel's theorem have

$$\int |P_{\mathcal{M}}[f](x)|^2 dx = C_{\mathcal{F}} \int |\sqrt{I}(k)|^2 dk < \infty, \quad \text{and}$$
$$\int |\nabla P_{\mathcal{M}}[f](x)|^2 dx = C_{\mathcal{F}} \int |k|^2 \sqrt{I}^2(k)\,dk < \infty,$$

hence $P_{\mathcal{M}}[f] \in H^1(\Omega)$ with desired norm.

ii) By definition of Fourier transform,

$$\|P_{\mathcal{M}}[f](x)\|_\infty \leqslant \sup_{x\in\Omega} \int_{\Omega_F} |\sqrt{I}(k)|\,dk < \infty, \quad \text{and}$$
$$\|\nabla P_{\mathcal{M}}[f](x)\|_\infty \leqslant \sup_{x\in\Omega} \int_{\Omega_F} |k\sqrt{I}(k)|\,dk < \infty,$$

hence $P_{\mathcal{M}}[f] \in W^{1,\infty}(\Omega)$.

iii) Since Ω is bounded, its Lebesgue measure $\lambda(\Omega)$ is less than infinity, and

$$\int_\Omega |P_{\mathcal{M}}[f]|^p \leqslant \lambda(\Omega)\|P_{\mathcal{M}}[f]\|_\infty^p < \infty, \quad \text{and}$$
$$\int_\Omega |\nabla P_{\mathcal{M}}[f]|^p \leqslant \lambda(\Omega)\|\nabla P_{\mathcal{M}}[f]\|_\infty^p < \infty$$

by part ii).

By ii) and iii), $P_{\mathcal{M}}$ is AP-bounded on $W^{1,p}(\Omega)$ for bounded Ω and any $p \in [1, \infty]$. □

Proposition 6.27 (dERF/ER is $W^{1,p}$-bounded on bounded domains). *Let $p \in [1, \infty)$, let $\mathcal{H} = L^2(\Omega)$ where $\Omega = \mathbb{T}^d$ or $\Omega \subset \mathbb{R}^d$ is measurable and bounded. Let $\sqrt{I} \in \widehat{\mathcal{H}}$ be non-negative with $\int \sqrt{I}(k)\, dk + \int |k|\sqrt{I}(k)\, dk < \infty$.*

Let $(g_n^{(\varepsilon)})_{n\in\mathbb{N}_0}$ be a dERF sequence with initial value $g_0 \in W^{1,p}(\Omega)$ and step size $\varepsilon > 0$.

i) If $\varepsilon \in (0, \frac{1}{2}]$, then $g_n^{(\varepsilon)} \in W^{1,p}(\Omega)$ for all $n \in \mathbb{N}_0$, and

$$\|g_n^{(\varepsilon)}\|_{W^{1,p}} \leqslant \max\left\{\|g_0\|_{W^{1,p}},\ \lambda(\Omega)^{1/d}\int(1+|k|)\sqrt{I}(k)\, dk\right\}.$$

ii) If $\varepsilon = 1$ and $g_0 \in \mathcal{P} \cap W^{1,p}(\Omega)$ — in other words, if $(g_n^{(\varepsilon)})$ is an ER sequence — then, $g_n \in W^{1,p}(\Omega)$ for all $n \in \mathbb{N}_0$, and

$$\|g_n^{(\varepsilon)}\|_{W^{1,p}} \leqslant \max\left\{\|g_0\|_{W^{1,p}},\ \lambda(\Omega)^{1/d}\int(1+|k|)\sqrt{I}(k)\, dk\right\},$$

where λ is the Lebesgue measure on $\mathbb{T}^d$ (if $\Omega = \mathbb{T}^d$) or on $\mathbb{R}^d$ (if $\Omega \subseteq \mathbb{T}^d$).

Proof. i) The claim follows from Proposition 6.13 with $\mathcal{B} = W^{1,p}(\Omega)$, applicable by Lemmas 6.24 and 6.26. ii) The claim follows by induction from Lemmas 6.24 and 6.26. □

Remark 6.28 (Two cases of Proposition 6.27). In the presented form, the case i) of Proposition 6.27 can not be extended to $\varepsilon > \frac{1}{2}$. For example, for generic $g \in W^{1,p}(\mathbb{T}^d)$ and $\varepsilon = 1$ have

$$\|g + (-2g + P_{\mathcal{P}}[g] + P_{\mathcal{M}}[g])\|_{W^{1,p}} = \|P_{\mathcal{M}}[g] + P_{\mathcal{P}}[g] - g\|_{W^{1,p}},$$

which can, in general, exceed $\|g\|_{W^{1,p}}$ and $\|P_{\mathcal{M}}[g]\|_{W^{1,p}}$. However, one still has boundedness of ER in $W^{1,p}$ due to ER's specific form.

6.3.3 *Fixed points of dERF*

The main contribution of this subsection is Proposition 6.32; it demonstrates that since dERF dissipates energy, its accumulation points are fixed points.

Recall the following variant of the Rellich-Kondrachov compactness theorem, see [Leo09, Theorem 11.21]. (Note that in [Leo09], the theorem is stated for a bounded domain $U \subseteq \mathbb{R}^d$ with continuous boundary ∂U. The results of the theorem are still valid on the torus $\mathbb{T}^d$ which is an open and bounded set with an empty boundary.)

Theorem 6.29 (Rellich-Kondrachov variant). *Let $p \in [1, \infty)$, let $\Omega = \mathbb{T}^d$ or let $\Omega \subset \mathbb{R}^d$ be measurable and bounded with continuous boundary $\partial\Omega$. The space $W^{1,p}(\Omega)$ is compactly embedded in $L^p(\Omega)$.*

Corollary 6.30 (Accumulation points exist). *Let $\Omega = \mathbb{T}^d$ or let $\Omega \subset \mathbb{R}^d$ be measurable and bounded with continuous boundary $\partial\Omega$. Let $p \in [1, \infty)$, let $\sqrt{I}\colon \Omega_F \to \mathbb{C}$ be non-negative with $\int \sqrt{I}(k)\, dk + \int |k| \sqrt{I}(k)\, dk < \infty$. Let $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ be a dERF sequence with initial value $g_0 \in W^{1,p}(\Omega)$ and step size $\varepsilon > 0$.*

If either i) $\varepsilon \in (0; \frac{1}{2}]$, or ii) $\varepsilon = 1$ and $g_0 \in W^{1,p}(\Omega) \cap \mathcal{P}$, then, there exists $g \in L^p(\Omega)$ such that a subsequence of $(g_n^{(\varepsilon)})_{n \in \mathbb{N}}$ converges to g in $L^p(\Omega)$.

Proof. By Proposition 6.27, $(g_n^{(\varepsilon)})_n$ is bounded in $W^{1,p}(\Omega)$ and thus compact in $L^p(\Omega)$ by Theorem 6.29. □

Example 6.31. In particular, for any $d \in \mathbb{N}$ and $g_0 \in H^1(\Omega)$ with a bounded $\Omega \subset \mathbb{R}^d$, there exists a fixed point $g \in L^2(\Omega)$ such that $g_n^{(\varepsilon)} \to g$ in L^2 as $n \to \infty$, where $n \in M$ for an appropriate unbounded $M \subset \mathbb{N}$.

The following proposition follows from Proposition 6.17, the core step of which was the energy dissipation inequality from Corollary 6.5.

Proposition 6.32 (Accumulation points are fixed points). *Let $\Omega = \mathbb{T}^d$ or let $\Omega \subset \mathbb{R}^d$ be measurable and bounded with continuous boundary $\partial\Omega$. Let $\mathcal{A}$ represent an additional constraint such that $P_{\mathcal{A}}\colon L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ is continuous. Let $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ be a dERF sequence with initial value $g_0 \in L^2(\mathbb{T}^d)$, with additional constraint $\mathcal{A}$, and with step size $\varepsilon > 0$. Let $\sqrt{I}\colon \Omega_F \to \mathbb{C}$ be non-negative with $\int \sqrt{I}(k)\, dk + \int |k| \sqrt{I}(k)\, dk < \infty$.*

Then, if g is an accumulation point of $(g_n^{(\varepsilon)})_{n \in \mathbb{N}}$, there exists $\psi\colon \mathbb{Z}^d \to [0, 2\pi)$ with $\sin\psi$ odd such that g a fixed point of dERF, i. e.

$$2g - P_{\mathcal{M};\psi}[g] - P_{\mathcal{A}}[g] = 0$$

Proof. By Proposition 6.27, $(g_n^{(\varepsilon)})_n$ is bounded in $H^1(\Omega)$ and thus compact in $L^2(\Omega)$ by Theorem 6.29. Thus, there exists a strong accumulation point g of $(g_n^{(\varepsilon)})_n$. By Proposition 6.17, there exists an appropriate ψ such that g is the desired fixed point. □

6.3.4 *Regularization of additional constraints*

This subsection demonstrates one of the possible ways to regularize additional constraints, such that an extension of Corollary 6.30 becomes possible. The highlight of this section is Proposition 6.43; it presents an example on how (convex) additional constraints can be regularized so that existence results that will be derived in Chapter 7 remain valid.

Results of the previous subsection rely on $\|P_{\mathcal{P}}[g]\|_{W^{1,p}} \leqslant \|g\|_{W^{1,p}}$ for all $g \in W^{1,p}$, and on $W^{1,p}$ being compact in L^p. In general, it can not be expected for other additional constraints. For example, when positivity is combined with support constraint, the resulting projection $P_{\mathcal{P}\cap\mathcal{S}}$ does not map $W^{1,p}(\mathbb{T}^d)$ to $W^{1,p}(\mathbb{T}^d)$.

6.3.4.1 *Regularization via intersection of closed convex sets*

In practice, it is most convenient to regularize convex additional constraints $\mathcal{A}$ using some convex regularization set $\mathcal{X}$, since one can use methods from convex optimization to calculate $P_{\mathcal{A}\cap\mathcal{X}}$. Recall the Halpern and the Dykstra algorithms, cited here from [BC17, Ch. 30] and [BCL02]; these algorithms may be used to calculate $P_{\mathcal{X}_1\cap\ldots\cap\mathcal{X}_m}$ for closed convex sets $\mathcal{X}_1, \ldots, \mathcal{X}_m$.

DEFINITION 6.33 (A SPECIFIC CASE OF THE HALPERN ALGORITHM). *Let $M \in \mathbb{N}$, let $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M \subset \mathcal{H}$ be closed and convex, let P_i denote the projection onto $\mathcal{X}_i$ for $i \in \{1, \ldots, M\}$. Let $g_0 \in \mathcal{H}$. Generate the sequence $(g_n)_{n\in\mathbb{N}}$ using the following update:*

$$g_{n+1} = \frac{1}{n+2} g_0 + \frac{n+1}{n+2} \left(\frac{1}{M} \sum_{i=1}^{M} P_i[g_n] \right). \tag{Halpern}$$

In this case, call $(g_n)_{n\in\mathbb{N}_0}$ the Halpern sequence with initial value g_0.

THEOREM 6.34 (HALPERN'S CONVERGENCE [BC17, Ex. 30.4]). *Let $M \in \mathbb{N}$, let $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M \subset \mathcal{H}$ be closed and convex with $C := \bigcap_{i=1}^{M} \mathcal{X}_i \neq \varnothing$. let P_i denote the projection onto $\mathcal{X}_i$ for $i \in \{1, \ldots, M\}$. Let $g_0 \in \mathcal{H}$, let $(g_n)_{n\in\mathbb{N}}$ be the corresponding Halpern sequence. Then, $g_n \to P_{\mathcal{X}\cap\mathcal{Y}}[g_0]$.*

The following variant of the Dykstra algorithm for two sets is taken from [BCL02]. A formulation for intersection of $M \in \mathbb{N}$ sets can be found in [BC17, Ch. 30].

DEFINITION 6.35 (DYKSTRA'S ALGORITHM). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be closed and convex. Let $g_0 \in \mathcal{H}$, let $h_{-1} = \tilde{h}_0 = 0$. Generate the sequences $(g_n)_{n\in\mathbb{N}}$, $(\tilde{g}_n)_{n\in\mathbb{N}}$, $(h_n)_{n\in\mathbb{N}}$, and $(\tilde{h}_n)_{n\in\mathbb{N}}$ using the following update:*

$$\begin{aligned} \tilde{g}_n &= P_{\mathcal{Y}}[g_n + h_{n-1}] \\ h_n &= g_n + h_{n-1} - \tilde{g}_n \\ g_{n+1} &= P_{\mathcal{X}}[\tilde{g}_n + \tilde{h}_n] \\ \tilde{h}_{n+1} &= \tilde{g}_n + \tilde{h}_n - g_{n+1} \end{aligned} \tag{Dykstra}$$

In this case, call $(g_n)_{n\in\mathbb{N}_0}$ the Dykstra sequence with initial value g_0.

THEOREM 6.36 (DYKSTRA'S CONVERGENCE [BC17, THM. 30.7]). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be closed and convex with non-empty $\mathcal{X} \cap \mathcal{Y} \neq \varnothing$. Let $g_0 \in \mathcal{H}$, let $(g_n)_{n\in\mathbb{N}}$ be the corresponding Dykstra sequence. Then, $g_n \to P_{\mathcal{X}\cap\mathcal{Y}}[g_0]$.*

The following lemma shows sufficient conditions for continuity of regularized projections.

LEMMA 6.37 (CONTINUITY OF REGULARIZED PROJECTIONS). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be closed and convex with non-empty intersection. If $P_\mathcal{X}$ and $P_\mathcal{Y}$ are continuous, then so is $P_{\mathcal{X}\cap\mathcal{Y}}$.*

Proof. Let $g^{(m)} \to g$ an $g \to \infty$. Let $p_0 \in \mathcal{H}$, let $(p_n^{(m)})_{n\in\mathbb{N}}$ be the Halpern sequence with initial value p_0 that converges to $P_{\mathcal{X}\cap\mathcal{Y}}[g^{(m)}]$ as $n \to \infty$; let $(p_n)_{n\in\mathbb{N}}$ be the Halpern sequence with initial value p_0 that converges to $P_{\mathcal{X}\cap\mathcal{Y}}[g]$ as $n \to \infty$.

Using continuity of $P_\mathcal{X}$ and $P_\mathcal{Y}$, have

$$\begin{aligned}
\|P_{\mathcal{X}\cap\mathcal{Y}}[g^{(m)}] - P_{\mathcal{X}\cap\mathcal{Y}}[g]\|_2 &= \lim_{n\to\infty} \|p_{n+1}^{(m)} - p_{n+1}\|_2 \\
&= \lim_{n\to\infty} \left\| \frac{1}{n+2} p_0 + \frac{n+1}{n+2} \left(P_\mathcal{X}[p_n^{(m)}] + P_\mathcal{Y}[p_n^{(m)}] \right) \right. \\
&\qquad \left. - \frac{1}{n+2} p_0 - \frac{n+1}{n+2} \left(P_\mathcal{X}[p_n] + P_\mathcal{Y}[p_n] \right) \right\|_2 \\
&= \left\| P_\mathcal{X}[g^{(m)}] + P_\mathcal{Y}[g_n^{(m)}] - P_\mathcal{X}[g] - P_\mathcal{Y}[g] \right\|_2 \to 0
\end{aligned}$$

as $m \to \infty$. □

6.3.4.2 *Hilbert cube regularization*

One way to regularize a function in object space is to enforce its decay in Fourier space. For example, one may bound all values of $|\hat{g}|$ by some upper bound $\rho \in L^1(\mathbb{Z}^d)$ as follows.

DEFINITION 6.38 (HILBERT CUBE). *i) For $\rho \in l^1(\mathbb{Z}^d; \mathbb{R}_{\geqslant 0})$, define the corresponding Hilbert cube (in Fourier space)*

$$\mathcal{Q}_\rho^\mathrm{F} = \left\{ f \in L^1(\mathbb{T}^d) \mid |\hat{f}(k)| \leqslant \rho(k) \quad \textit{for all } k \in \mathbb{Z}^d \right\}.$$

ii) For $\rho \in L^2(\mathbb{R}^d; \mathbb{R}_{\geqslant 0})$, define the corresponding Hilbert cube (in Fourier space)

$$\mathcal{Q}_\rho^\mathrm{F} = \left\{ f \in L^2(\mathbb{R}^d) \mid |\hat{f}(k)| \leqslant \rho(k) \quad \textit{for a.a. } k \in \mathbb{R}^d \right\}.$$

Given an additional constraint $\mathcal{A}$, we call $\mathcal{A} \cap \mathcal{Q}_\rho^\mathrm{F}$ the regularized version of $\mathcal{A}$.

Example 6.39 (Some Hilbert cubes). Let $C > 0$, let $p \in [1, \infty]$. Let $\|\cdot\|_{2;\mathbb{R}^d}$ denote the Euclidean norm of a d-dimensional vector. The following choices of $\rho \in l^1(\mathbb{Z}^d; \mathbb{R}_{\geqslant 0})$ will be used in this section.

i) Let $r \in (d; \infty]$. The choice $\rho(k) = \frac{C}{1+\|k\|^r_{2;\mathbb{R}^d}}$ ensures that $\mathcal{Q}^{\mathrm{F}}_\rho$ is a compact subset of $L^p(\mathbb{T}^d)$ (see Corollary D.7). We use the shorthand notation $\mathcal{Q}^{\mathrm{F}}_{C;r} := \mathcal{Q}^{\mathrm{F}}_{\frac{C}{1+\|\cdot\|^r_{2;\mathbb{R}^d}}}$.

ii) Let $\sqrt{I} \in l^1(\mathbb{Z}^d; \mathbb{R}_{\geqslant 0})$ represent the phase retrieval measurement. The choice $\rho = C\sqrt{I}$ is appropriate to use when regularity of the additional constraint needs to be tied to the regularity of the modulus constraint.

Remark 6.40. Let ρ be in $l^1(\mathbb{Z}^d; \mathbb{R}_{\geqslant 0})$ or in $L^2(\mathbb{R}^d; \mathbb{R}_{\geqslant 0})$. From its definition immediately follows that $\mathcal{Q}^{\mathrm{F}}_\rho$ is convex and closed, and therefore weakly closed.

PROPOSITION 6.41 (PROJECTION ONTO $\mathcal{Q}^{\mathrm{F}}_\rho$). *Let ρ be in $l^1(\mathbb{Z}^d; \mathbb{R}_{\geqslant 0})$ or in $L^2(\mathbb{R}^d; \mathbb{R}_{\geqslant 0})$. Since $\mathcal{Q}^{\mathrm{F}}_\rho$ is weakly closed, a projection onto $\mathcal{Q}^{\mathrm{F}}_\rho$ is well-defined by Proposition 3.4. The projection onto $\mathcal{Q}^{\mathrm{F}}_\rho$ — with respect to metric of $L^2(\mathbb{T}^d)$ if $\rho \in l^1$, and with respect to metric of $L^2(\mathbb{R}^d)$ if $\rho \in L^2$ — is given by*

$$P_{\mathcal{Q}^{\mathrm{F}}_\rho}[g] = \mathcal{F}^{-1}\left(\rho \frac{\hat{g}}{|\hat{g}|} \mathbb{1}_{|\hat{g}|>\rho} + \hat{g}\mathbb{1}_{|\hat{g}|\leqslant\rho}\right).$$

Proof. The projection of $\mathcal{Q}^{\mathrm{F}}_\rho$ is unique by Proposition 3.9. One can check that $P_{\mathcal{Q}^{\mathrm{F}}_\rho}$ has the presented form, either by a direct calculation (similarly to Example 3.24), or using an appropriate local version and Proposition 3.33. □

Remark 6.42 (Calculating $P_\mathcal{A} \cap P_{\mathcal{Q}^{\mathrm{F}}_\rho}$). For any convex, closed additional constraint $\mathcal{A}$ with non-empty $\mathcal{A} \cap \mathcal{Q}^{\mathrm{F}}_\rho$, regularized projection $P_{\mathcal{A}\cap\mathcal{Q}^{\mathrm{F}}_\rho}$ can be calculated using Dykstra's algorithm or Halpern's algorithm.

An example can be found in Figure 6.2, where Dykstra's algorithm demonstrates $P_\mathcal{A}$ for $\mathcal{A} = (\mathcal{P} \cap \mathcal{S}) \cap \mathcal{Q}^{\mathrm{F}}_{C;r}$, for $C \in \{0.5, 1, 2\}$ and $r = 2$.

PROPOSITION 6.43 (REGULARIZED ACCUMULATION POINTS EXIST). *Let $\Omega = \mathbb{T}^d$, let $p \in [1, \infty]$, let $C > 0$, let $r > d$.*

Further, let $\sqrt{I} \in \hat{\mathcal{H}}(\mathbb{Z}^d)$ be non-negative with $\check{\sqrt{I}} \in \mathcal{Q}^{\mathrm{F}}_{C;r}$, let $\mathcal{A} \subset \mathcal{Q}^{\mathrm{F}}_{C;r}$ be closed. Let $(g^{(\varepsilon)}_n)_{n\in\mathbb{N}_0}$ be a dERF sequence with additional constraint $\mathcal{A}$, with initial value $g_0 \in \mathcal{Q}^{\mathrm{F}}_{C;r}$, with step size $\varepsilon \in (0, \frac{1}{2}]$.

Then, $g^{(\varepsilon)}_n \in \mathcal{Q}^{\mathrm{F}}_{C;r}$ for all $n \in \mathbb{N}_0$.

In particular, for any $p \in [1, \infty]$ there exists $g \in L^p(\mathbb{T}^d)$ such that a subsequence of $(g^{(\varepsilon)}_n)_{n\in\mathbb{N}}$ converges to g in $L^p(\mathbb{T}^d)$.

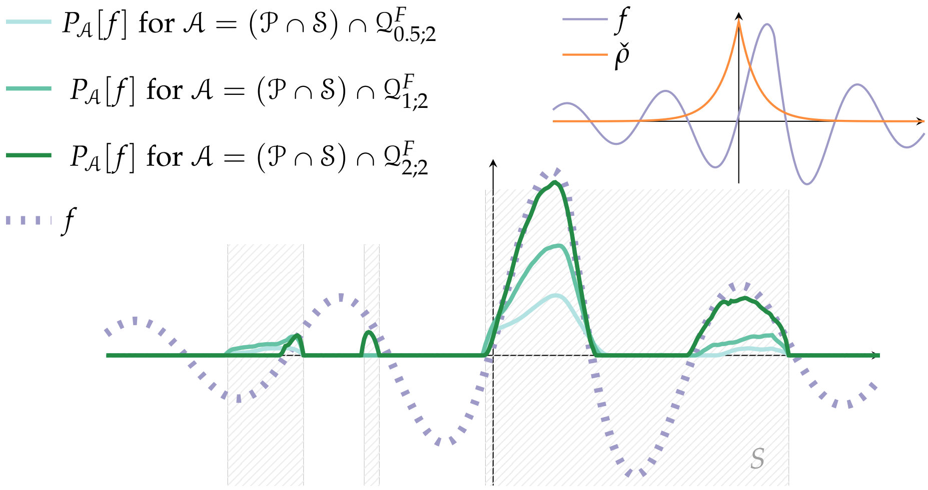

Figure 6.2: Illustration to Remark 6.42 (regularized $\mathcal{P} \cap \mathcal{S}$). Regularized "support+positivity" projection $P_{\mathcal{A}}[f]$ of a function f for $\mathcal{A} = (\mathcal{P} \cap \mathcal{S}) \cap \mathcal{Q}^{\mathrm{F}}_{C;r}$, for $r = 2$ and $C \in \{0.5, 1, 2\}$. The support set S is indicated by the grey hatched region. The regularization bound ρ is of the form $1/(|k|^2 + 1)$, rescaled so that $\check{\rho}$ — being of the form $\exp(-|x|)$ — is comparable to f, see the small plot in the upper right corner.
The set $\mathcal{Q}^{\mathrm{F}}_{C;2}$ is a compact subset of $L^p(\mathbb{T}^1)$, and Corollary 6.30 applies.
The "support+positivity" projection can be calculated explicitely: $P_{\mathcal{P}\cap\mathcal{S}(S)}[g] = \mathbb{1}_{S\cap\{g\geqslant 0\}}g$.
For $C = 4$ (not plotted), differences between the regularized projection and $P_{\mathcal{P}\cap\mathcal{S}}$ are unvisible to the naked eye.
As described in Remark 6.42, $P_{\mathcal{A}}$ can be calculated using Dykstra's algorithm. Differences between iterates become invisible to the naked eye after the approximation $p_f \approx P_{\mathcal{A}}[f]$ satisfies $E_{\mathcal{P}\cap\mathcal{S}}[p_f] + E_{\mathcal{Q}^{\mathrm{F}}_{C;r}}[p_f] \leqslant 10^{-3} \min\{\|\check{\rho}\|_2, \|f\|_2\}$. For $C = 2$ (mild regularization), this precision was achieved after 17 iterations. For $C = 0.5$ (very restrictive regularization), this presicion was achieved after 41 iterations. (The space was discretized using 511 grid points.)

Proof. Since $\check{\sqrt{I}} \in \mathcal{Q}^{\mathrm{F}}_{C;r}$, one has $P_{\mathcal{M}}[f] \in \mathcal{Q}^{\mathrm{F}}_{C;r}$ for any $f \in L^p(\mathbb{T}^d)$. By assumption, $P_{\mathcal{A}}[f] \in \mathcal{Q}^{\mathrm{F}}_{C;r}$ for any $f \in L^p(\mathbb{T}^d)$. If $g_n^{(\varepsilon)} \in \mathcal{Q}^{\mathrm{F}}_{C;r}$, then at any $k \in \mathbb{Z}^d$

$$|\hat{g}_{n+1}^{(\varepsilon)}(k)| = |(1-2\varepsilon) \underbrace{\hat{g}_n^{(\varepsilon)}(k)}_{\leqslant \frac{C}{\|k\|^r_{2;\mathbb{R}^d}}} + \varepsilon \underbrace{\widehat{P_{\mathcal{M}}[f]}(k)}_{\leqslant \frac{C}{\|k\|^r_{2;\mathbb{R}^d}}} + \varepsilon \underbrace{\widehat{P_{\mathcal{P}\cap\mathcal{Q}^{\mathrm{F}}_{C;r}}[f]}}_{\leqslant \frac{C}{\|k\|^r_{2;\mathbb{R}^d}}}| \leqslant \frac{C}{\|k\|^r_{2;\mathbb{R}^d}}.$$

By induction, $g_n^{(\varepsilon)} \in \mathcal{Q}^{\mathrm{F}}_{C;r}$ for all $n \in \mathbb{N}$, and since $\mathcal{Q}^{\mathrm{F}}_{C;r}$ is compact in $L^\infty(\mathbb{T}^d)$ (see Corollary D.7), there exists an unbounded $M \subset \mathbb{N}$ and $g \in L^\infty(\mathbb{T}^d)$ such that

$$\lim_{\substack{n\to\infty\\ n\in M}} \|g_n^{(\varepsilon)} - g\|_{L^p(\mathbb{T}^d)} \leqslant \lim_{\substack{n\to\infty\\ n\in M}} (2\pi)^{d/p} \|g_n^{(\varepsilon)} - g\|_\infty = 0,$$

and $g \in L^p(\mathbb{T}^d)$. □

Example 6.44. For any $d \in \mathbb{N}$, if $r > d$ and if assumptions of Proposition 6.43 are satisfied, accumulation points exist. In particular, for $\mathcal{A} = (\mathcal{P} \cap \mathcal{S}) \cap \mathcal{Q}^{\mathrm{F}}_{C;r}$, $g_0 \in L^2(\mathbb{T}^d) \cap \mathcal{Q}^{\mathrm{F}}_{C;r}$ there exists a fixed point $g \in L^2(\mathbb{T}^d) \cap \mathcal{Q}^{\mathrm{F}}_{C;r}$ such that $g_n^{(\varepsilon)} \to g$ in $\mathcal{H}$ as $n \to \infty$, where $n \in M$ for an appropriate unbounded $M \subset \mathbb{N}$.

For future references — for example, for application of the Aubin-Lions lemma — it is convenient not to use $\mathcal{Q}^{\mathrm{F}}_{C;r}$ directly, but to introduce a corresponding Banach space, a subspace of $L^p(\mathbb{T}^d)$, where $\mathcal{Q}^{\mathrm{F}}_{C;r}$ corresponds to a closed ball of radius C.

LEMMA 6.45 (BANACH SPACE WHERE CLOSED BALLS ARE HILBERT CUBES). *Let $\rho \in l^1(\mathbb{Z}^d; \mathbb{R}_{\geqslant 0})$, let $p \in [1, \infty]$. Define*

$$\begin{aligned} \|\cdot\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_\rho]} \colon L^p(\mathbb{T}^d) &\to [0,\infty], \\ f &\mapsto \inf\{C > 0 \mid |\hat{f}(k)| \leqslant C\rho(k) \text{ for all } k \in \mathbb{Z}^d\}, \end{aligned}$$

with convention $\inf \varnothing = \infty$. *Let* $\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_\rho] = \{f \in L^p(\mathbb{T}^d) \mid \|f\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_\rho]} < \infty\}$.

Then, $(\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_\rho], \|\cdot\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_\rho]})$ *is a Banach space. It is compactly embedded in* $L^p(\mathbb{T}^d)$.

Proof. It is straightforward to verify that $\|\cdot\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_\rho]}$ is a norm, since for all $f, g \in L^p(\mathbb{T}^d)$ have

i) $\|\cdot\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_\rho]}$ is non-negative and $\|f\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_\rho]} = 0$ implies $\hat{f}(k) = 0$ for all $k \in \mathbb{Z}^d$;

ii) if $|\hat{f}(k)| \leqslant C\rho(k)$ for all $k \in \mathbb{Z}^d$, then $|a\hat{f}(k)| \leqslant aC\rho(k)$ for all $a \geqslant 0$;

iii) if $|\hat{f}(k)| \leqslant C_f\rho(k)$ and $|\hat{g}(k)| \leqslant C_g\rho(k)$, then $|\hat{f}(k) + \hat{g}(k)| \leqslant (C_f + C_g)\rho(k)$.

The set $\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]$ is a vector space with addition and scalar multiplication from $L^p(\mathbb{T}^d)$.

It is straightforward to show that $\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]$ is complete. Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]$; then, $(f_n)_{n\in\mathbb{N}}$ is bounded by some $C \in \mathbb{R}$. Therefore, $f_n \in \mathcal{Q}^{\mathrm{F}}_{C\rho}$ for all $n \in \mathbb{N}$. Thes set $\mathcal{Q}^{\mathrm{F}}_{C\rho}$ is compact in $L^p(\mathbb{Z}^d)$ by Proposition D.6, so $(f_n)_{n\in\mathbb{N}}$ has a convergent subsequence, i. e. $f_m \to f \in \mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]$ as $m \to \infty$, $m \in M$ for some unbounded $M \subset \mathbb{N}$.

For any $\varepsilon > 0$, there exists $N_1 \in \mathbb{N}$ such that $\|f_l - f_n\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]} < \varepsilon/2$ for all $l, n \geqslant N_1$. Further, there exists $N_2 \in \mathbb{N}$ such that $\|f_m - f\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]} < \varepsilon/2$ for all $m \in M$, $m \geqslant N_2$. Therefore, for all $n \in \mathbb{N}$, $n \geqslant \max\{N_1, N_2\}$ holds

$$\|f_n - f\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]} \leqslant \|f_m - f\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]} + \|f_n - f_m\|_{\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]} < \varepsilon,$$

where m is any element of M that is larger than n (which exists since M is unbounded). Therefore, $f_n \to f$ in $\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]$.

To show that $\mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]$ is compact, consider any closed bounded set $\mathcal{X} \subset \mathcal{B}[\mathcal{Q}^{\mathrm{F}}_{\rho}]$. Let $\mathcal{X}$ be contained inside the ball with radius $C < \infty$. Then, $\mathcal{X} \subset \mathcal{Q}^{\mathrm{F}}_{C\rho}$, which is is compact in $L^p(\mathbb{Z}^d)$ by Proposition D.6, and therefore $\mathcal{X}$ is compact as a closed subset of $\mathcal{Q}^{\mathrm{F}}_{C\rho}$. □

7 GLOBAL WEAK SOLUTIONS OF ERF

Remark 5.17 and Section 5.3.5 show that the flow

$$\partial_t g = -(g - P_{\mathcal{M}}[g]) - (g - P_{\mathcal{P}}[g])$$

can be derived from the minimization of the energy $E_{\mathcal{M}} + E_{\mathcal{P}}$ and is connected to ER and DR-HIO algorithms. The main result of this chapter is to show existence of global weak solutions of this equation; these results are obtained in joint work with Gero Friesecke. The general approach is inspired by the existence proof demonstrated in [FD97].

Sections 7.1 and 7.2 establish condtitions that are sufficient to guarantee the existence of a global weak solution, and Section 7.3 establishes certain properties of the solution.

Throughout the chapter we assume the following setting.

Definition 7.1 (Setting). *Assume one of the following:*

1) *Let $\Omega = \mathbb{T}^d$, let $\Omega_F = \mathbb{Z}^d$, let $I \colon \mathbb{Z}^d \to \mathbb{R}$ with $I \geqslant 0$ such that $\sqrt{I}$ belongs to $\ell^1(\mathbb{Z}^d)$.*

2) *Let Ω be a bounded measurable subset of $\mathbb{R}^d$ with continuous boundary $\partial\Omega$, let $\Omega_F = \mathbb{R}^d$, let $I \colon \mathbb{R}^d \to \mathbb{R}$ be a measurable function with $I \geqslant 0$ such that $\sqrt{I}$ belongs to $L^1(\mathbb{R}^d)$.*

Further, assume that $\int_{\Omega_F}(1 + |k|)\sqrt{I}(k) < \infty$ (which can be considered as a smoothness condition for $\check{\sqrt{I}}$). Unless explicitly stated otherwise, we use positivity $\mathcal{P}$ as the additional constraint $\mathcal{A}$ of the phase problem.

Definition 7.2 (Solutions of ERF). *Let $\Omega, \Omega_F, \sqrt{I}$ be as in Definition 7.1. Let $p \in [1, \infty]$, let $\tilde{g}_0 \in L^2(\Omega)$, let $T \in (0, \infty)$. Let $g \in L^p([0,T], L^2(\mathbb{T}^d))$ with $g_t := g(t)$, let*

$$\begin{aligned} \varphi \colon \mathbb{R}_{\geqslant 0} &\to \{\tilde{\varphi} \colon \Omega_F \to [0, 2\pi) \mid \tilde{\varphi} \text{ measurable, } \sin\tilde{\varphi} \text{ is odd}\} \\ t &\mapsto \varphi_t := \varphi(t) \end{aligned}$$

be measurable in t.

In that case, g is a weak solution of the Error-Reduction Flow with phase φ and initial value $\tilde{g}_0$, if

$$\partial_t g_t = -2g_t + P_{\mathcal{M};\varphi_t}[g_t] + P_{\mathcal{P}}[g_t] \tag{ERF}$$

holds in the weak sense, and $\lim_{t\to 0} g_t = \tilde{g}_0$.

The goal of Sections 7.1 and 7.2 is to show that for any starting value $\tilde{g}_0$ there exists a phase φ required by Definition 7.2, such that ERF has a global weak solution g, and that

$$g \in L^p\big((0,T); L^2(\Omega)\big)$$

for any $T > 0$, for any $p \in [1, \infty)$, as well as

$$g \in L^\infty\big((0,\infty); L^2(\Omega)\big) \cap C\big([0,\infty); L^2(\Omega)\big).$$

Remark 7.3 (Measurable selection instead of multi-valuedness; regularity in time). The presented equation (ERF) is only a particular measurable selection of a generic subdifferential flow

$$\partial_t g_t \in -2g_t + \Pi_{\mathcal{M}}[g_t] + P_{\mathcal{P}}[g_t]. \tag{7.1}$$

that was derived in Remark 5.18.

Working with a measurable selection rather than with a multi-valued flow simplifies our analysis. While being simpler, it retains noteworthy features of the underlying phase retrieval problem.

For example, existence of a solution is by definition connected to regularity in time. This regularity in time is closely tied to the time-dependent multiplicity resolution $P_{\mathcal{M};\varphi_t}$.

The selection of the multiplicity resolution $P_{\mathcal{M};\varphi_t}$, as introduced in Definition 7.2, is not prescribed beforehand. In our approach, it emerges from the construction of the solution g_t through an explicit limiting procedure.

This procedure will use the selection $P_{\mathcal{M};0}$ to construct the sequence of approximate solutions (and then form an appropriate limit). The selection $P_{\mathcal{M};0}$ is chosen as arguably the simplest selection (and the one most commonly used in literature). The presented construction applies to any other selection $P_{\mathcal{M};\psi}$ as long as ψ is not varied during the construction of approximate solutions.

When passing from approximate solutions to the resulting limit, the latter will satisfy (ERF) with selection $P_{\mathcal{M};\varphi_t}$ — and the selection $P_{\mathcal{M};\varphi_t}$ will be varied in time.

This resulting limit will satisfy the multivalued subdifferential equation Equation (7.1) for almost all $t > 0$, and it will be continuous in time.

The general approach is similar to the one used in [FD97] and reads as follows:

(i) Construct a sequence of approximating solutions using dERF with ever smaller step sizes (Definition 7.5).

(ii) In Corollary 7.6, show that the sequence of approximating solutions has necessary compactness properties for a known Aubin-type theorem (recalled on Page 147); the boundedness of Ω is required here.

(iii) From this follows existence of a solution candidate (Corollary 7.10).

(iv) A short technical result (Lemma 7.13) allows us to show that the solution candidate formally solves ERF (Remark 7.14).

(v) This formal calculation is made general in Theorem 7.18, where we exploit the fact that the solution candidate is sufficiently smooth (Corollary 7.16), which, in turn, follows from a known version of Rademacher's theorem (recalled on Page 151).

7.1 construct solution candidate

Notation 7.4 (Time discretization). Let $\varepsilon > 0$. For $t \in \mathbb{R}$, use

$$\lfloor t \rfloor_\varepsilon := \sup_{n \in \mathbb{N}_0} \{ n\varepsilon \mid n\varepsilon < t \}$$

to denote the latest discretization point in the time interval $[0, t)$ and $N_\varepsilon(t) := \lfloor t \rfloor_\varepsilon / \varepsilon$ to denote the number of discrete steps that fits in $[0, t)$.

Definition 7.5 (Approximation sequences). *Let $\Omega, \Omega_F, \sqrt{I}$ be as in Definition 7.1, let $g_0 \in L^2(\Omega)$. Let $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ be the dERF sequence with initial value g_0, additional constraint $\mathcal{A} = \mathcal{P}$, and step size $\varepsilon > 0$ (see Figure 7.1).*

i) Define the piecewise constant interpolation

$$\begin{aligned} &\bar{g}^{(\varepsilon)} \colon [0, +\infty) \mapsto L^2(\Omega) \\ &\bar{g}^{(\varepsilon)}(t) = g_n^{(\varepsilon)} \quad \text{for } t \in \big[n\varepsilon, (n+1)\varepsilon \big), n \in \mathbb{N}_0. \end{aligned}$$

ii) Define the piecewise linear interpolation

$$\begin{aligned} &\tilde{g}^{(\varepsilon)} \colon [0, +\infty) \mapsto L^2(\Omega) \\ &\tilde{g}^{(\varepsilon)}(t) = \left(1 - \frac{\tau}{\varepsilon}\right) g_n^{(\varepsilon)} + \frac{\tau}{\varepsilon} g_{n+1}^{(\varepsilon)} \quad \text{for } t \in \big[n\varepsilon, (n+1)\varepsilon \big), n \in \mathbb{N}_0, \end{aligned}$$

where $\tau = t - \lfloor t \rfloor_\varepsilon$.

Corollary 7.6 (Interpolations are well-defined and bounded). *Let $\Omega, \Omega_F, \sqrt{I}$ be as in Definition 7.1. Let $g_0 \in H^1(\Omega)$, let $\varepsilon \in (0; \frac{1}{2}]$, let $p \in [1, \infty]$. Let $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$, $\tilde{g}^{(\varepsilon)}$ and $\bar{g}^{(\varepsilon)}$ be as in Definition 7.5.*

(i) Let $\mathcal{H} \in \{H^1(\Omega), L^2(\Omega)\}$. For the discrete sequence $(g_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ holds:

$$\| g_n^{(\varepsilon)} \|_{\mathcal{H}} \leqslant \max\{ \| g_0 \|_{\mathcal{H}}, \| \check{\sqrt{I}} \|_{\mathcal{H}} \}.$$

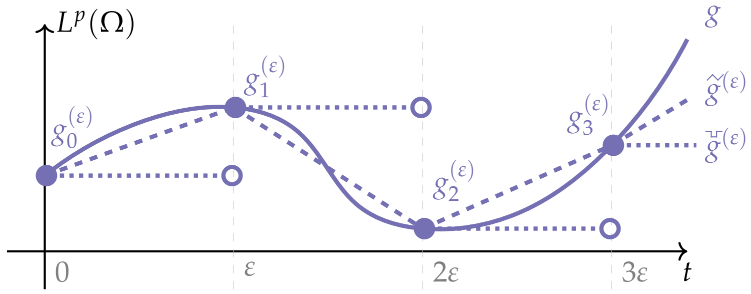

Figure 7.1: Illustration to Chapter 7 (Global solution of ERF). To show existence of the global solution g (Theorem 7.18), construct a sequence of linear approximations $(\widetilde{g}^{(\varepsilon)})_{\varepsilon>0}$ (Definition 7.5). For $\varepsilon \to 0$, this sequence has a subsequence that converges to some solution candidate g (Corollary 7.10).
Use $\bar{g}^{(\varepsilon)} \to g$ (Lemma 7.13) and $\partial_t \widetilde{g}^{(\varepsilon)} = \bar{g}^{(\varepsilon)}$ to show that $\partial_t g$, if exists, must have a form close to the desired solution (Remark 7.14).
Use Rademacher's theorem to establish that $\partial_t g$ exists (Corollary 7.16), and follow that there exists a phase φ required in Definition 7.5 to solve ERF (Lemma 7.17).

(ii) *The interpolations* $\widetilde{g}^{(\varepsilon)}$ *and* $\bar{g}^{(\varepsilon)}$ *belong to* $L^p(0,T;H^1(\Omega))$ *with the following estimates:*

$$\sup_{t\in(0,T)} \|\widetilde{g}^{(\varepsilon)}(t)\|_{H^1} \leqslant \max\{\|g_0\|_{H^1}, \|\widecheck{\sqrt{I}}\|_{H^1}\}, \tag{7.2}$$

$$\sup_{t\in(0,T)} \|\bar{g}^{(\varepsilon)}(t)\|_{H^1} \leqslant \max\{\|g_0\|_{H^1}, \|\widecheck{\sqrt{I}}\|_{H^1}\}, \tag{7.3}$$

$$\sup_{t\in(0,T)} \left\|\frac{d}{dt}\widetilde{g}^{(\varepsilon)}(t)\right\|_{L^2} \leqslant 4\max\{\|g_0\|_{L^2}, \|\widecheck{\sqrt{I}}\|_{L^2}\}; \tag{7.4}$$

$$\int_0^T \|\widetilde{g}^{(\varepsilon)}(t)\|_{H^1}^p dt \leqslant T\max\{\|g_0\|_{H^1}^p, \|\widecheck{\sqrt{I}}\|_{H^1}^p\}, \tag{7.5}$$

$$\int_0^T \|\bar{g}^{(\varepsilon)}(t)\|_{H^1}^p dt \leqslant T\max\{\|g_0\|_{H^1}^p, \|\widecheck{\sqrt{I}}\|_{H^1}^p\}, \tag{7.6}$$

$$\int_0^T \left\|\frac{d}{dt}\widetilde{g}^{(\varepsilon)}(t)\right\|_{L^2}^p dt \leqslant T(4\max\{\|g_0\|_{L^2}, \|\widecheck{\sqrt{I}}\|_{L^2}\})^p. \tag{7.7}$$

Proof. (i) We show the statement for $\mathcal{H} = H^1(\Omega)$; the other case follows analogously. The proof is extremely similar to Proposition 6.27, with improved estimates due to the Plancherel theorem in L^2. Namely,

$$\begin{aligned}\|g_{n+1}^{(\varepsilon)}\|_{H^1} &= \|(1-2\varepsilon)g_n^{(\varepsilon)} + \varepsilon P_{\mathcal{P}}[g_n^{(\varepsilon)}] + \varepsilon P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\|_{H^1} \\ &\leqslant (1-2\varepsilon)\|g_n^{(\varepsilon)}\|_{H^1} + \varepsilon\|P_{\mathcal{P}}[g_n^{(\varepsilon)}]\|_{H^1} + \varepsilon\|P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\|_{H^1}\end{aligned}$$

by triangle inequality. By Lemma 6.26, $\|P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\|_{H^1} = \|\widecheck{\sqrt{I}}\|_{H^1}$, and by Lemma 6.24 have $\|P_{\mathcal{P}}[g]\|_{H^1} \leqslant \|g\|_{H^1}$. The claim follows by induction in n.

(ii) Equations (7.2) and (7.3) follow immediately from the definition of $\widetilde{g}^{(\varepsilon)}, \overline{g}^{(\varepsilon)}$ and (i). Equation (7.4) also follows from the definition and (i) with $\mathcal{H} = L^2$:

$$\left\| \frac{d}{dt}\widetilde{g}^{(\varepsilon)}(t) \right\|_2 = \left\| -\frac{1}{\varepsilon} g_n^{(\varepsilon)} + \frac{1}{\varepsilon} g_{n+1}^{(\varepsilon)} \right\|_{L^2} =$$
$$= \left\| -2g_n^{(\varepsilon)} + P_{\mathcal{P}}[g_n^{(\varepsilon)}] + P_{\mathcal{M}}[g_n^{(\varepsilon)}] \right\|_{L^2} \leqslant 4 \max\{\|g_n^{(\varepsilon)}\|_{L^2}, \|\widecheck{\sqrt{I}}\|_{L^2}\}.$$

Equations (7.5) to (7.7) follow from Equations (7.2) to (7.4), respectively. □

Remark 7.7 (Other additional constaints). With the same argument, previous corollary can be extended to any other additional constraint $P_{\mathcal{A}}$ — provided that $P_{\mathcal{A}}$ is AP-bounded on $H^1(\Omega)$ (see Definition 6.10). For example, positivity with restricted support $\mathcal{A} = \mathcal{P} \cap \mathcal{S}$ is not necessarily bounded on $H^1(\Omega)$ (since $P_{\mathcal{S}}$ lacks regularity), but its regularized variant $\mathcal{A} = \mathcal{P} \cap \mathcal{S} \cap \mathcal{Q}_\rho^{\mathrm{F}}$ is AP-bounded by $\|\check{\rho}\|_{H^1}$. Then, Corollary 7.6 holds with the estimate $\max\{\|g_n^{(\varepsilon)}\|_{\mathcal{H}}, \|\widecheck{\sqrt{I}}\|_{\mathcal{H}}, \|\check{\rho}\|_{\mathcal{H}}\}$ instead of $\max\{\|g_n^{(\varepsilon)}\|_{\mathcal{H}}, \|\widecheck{\sqrt{I}}\|_{\mathcal{H}}\}$ for $\mathcal{H} \in \{L^2, H^1\}$.

Recall the well-known Aubin-type result which can be found, for example, in [BF13, Thm. II.5.16].

Theorem 7.8 (Aubin-Lions-Simon Lemma). *Let X_s, X and X_w be Banach spaces such that the following inclusions hold:*

$$X_s \hookrightarrow (compact) X \hookrightarrow (continuous) X_w.$$

Let p, r be such that $1 \leqslant p, r \leqslant \infty$. For $T > 0$, define

$$W_{p,r} = \left\{ v \in L^p\big((0,T), X_s\big), \frac{dv}{dt} \in L^r\big((0,T), X_w\big) \right\}$$

(i) If $p < \infty$, the embedding of $W_{p,r}$ in $L^p\big((0,T), X\big)$ is compact.

(ii) If $p = \infty$ and $r > 1$, the embedding of $W_{p,r}$ in $C\big([0,T], X\big)$ is compact.

Notation 7.9. For $p \in [1, \infty]$, $T \in (0, \infty]$, use shorthand

$$L^p_{t,L^2} \cap C_{t,L^2} = L^p\big((0;T); L^2(\Omega)\big) \cap C\big([0,T], L^2(\Omega)\big).$$

A sequence $(g_n)_{n\in\mathbb{N}}$ is said to converge strongly to g in $L^p_{t,L^2} \cap C_{t,L^2}$, if $g \in L^p_{t,L^2} \cap C_{t,L^2}$, and if

$$\int \|g_n(t) - g(t)\|^p_{L^2(\Omega)} dt \to 0 \quad \text{and}$$
$$\sup_{t\in(0,T)} \|g_n(t) - g(t)\|_{L^2(\Omega)} \to 0 \quad \text{as } n \to \infty. \tag{7.8}$$

The setting of the following corollary will be used throughout the remaining chapter.

Corollary 7.10 (Strongly convergent subsequence exists). *Let the domains Ω, Ω_F and the measurement $\sqrt{I}$ be as in Definition 7.1. Let the initial value $\tilde{g}_0$ be in $H^1(\Omega)$, let the time step $\varepsilon \in (0; \frac{1}{2}]$, let $p \in [1, \infty]$, let the time $T \in (0, \infty)$. Let the approximate solutions $g_n^{(\varepsilon)}$, $\overline{g}^{(\varepsilon)}$ and $\widetilde{g}^{(\varepsilon)}$ be as in Definition 7.5 with initial value $\tilde{g}_0$. Let $J_1 \subset (0, \frac{1}{2}]$ be the set of discretization step sizes with $0 \in \overline{J_1}$.*

Then, there exists a set $J_2 \subset J_1$ with $0 \in \overline{J_2}$ such that $\widetilde{g}^{(\varepsilon)}$ converges strongly to some $g \in L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.

Proof. Follows from Corollary 7.6 and Aubin-Lions-Simon lemma with $X_s = H^1(\Omega)$, $X = L^2(\Omega)$, $X_w = L^2(\Omega)$. □

Remark 7.11 (Difficulties with unbounded domains). The proposition above relies on $H^1(\Omega)$ being compactly embedded into $L^2(\Omega)$. In order to generalize this result to unbounded $\tilde{\Omega} \subset \mathbb{R}^d$, the set $\{g_n^{(\varepsilon)} \mid n \in \mathbb{N}\}$ must be relatively compact in $L^2(\tilde{\Omega})$. As discussed in Remark 6.20, the modulus operator as-is lacks necessary decay in object space: convergence of $\int_{|x|>R} |P_{\mathcal{M}}[g_n^{(\varepsilon)}](x)|^2 \, dx \to 0$ as $R \to \infty$ is not necessarily uniform in n.

While from Corollary 7.10 follows that $g(t)$ is continuous on $[0, T]$, it must be shown that $g(0) = \tilde{g}_0$.

Lemma 7.12 (Initial value of the limit). *Under assumptions of Corollary 7.10 holds $\lim_{t \to 0} g(t) = \tilde{g}_0$, the limit being taken in $\|\cdot\|_{L^2}$.*

Proof. Throughout this proof, use $\|\cdot\|_2 = \|\cdot\|_{L^2}$. Observe that for any $\varepsilon \in J_2$ and almost all $t > 0$

$$\|g(t) - \tilde{g}_0\|_2 \leqslant \|g(t) - \widetilde{g}^{(\varepsilon)}(t)\|_2 + \|\widetilde{g}^{(\varepsilon)}(t) - \widetilde{g}^{(\varepsilon)}(0)\|_2,$$

hence

$$\|g(t) - \tilde{g}_0\|_2 \leqslant \underbrace{\inf_{\varepsilon \in J_2} \|g(t) - \widetilde{g}^{(\varepsilon)}(t)\|_2}_{=0 \text{ due to } (7.8)} + \inf_{\varepsilon \in J_2} \|\widetilde{g}^{(\varepsilon)}(t) - \widetilde{g}^{(\varepsilon)}(0)\|_2.$$

Since

$$\begin{aligned}
\|\widetilde{g}^{(\varepsilon)}(t) - \widetilde{g}^{(\varepsilon)}(0)\|_2 &\leqslant \sum_{n=0}^{N_\varepsilon(t)} \|\widetilde{g}^{(\varepsilon)}\big(\varepsilon(n+1)\big) - \widetilde{g}^{(\varepsilon)}(\varepsilon n)\|_2 \\
&\leqslant \sum_{n=0}^{N_\varepsilon(t)} \varepsilon \|2\widetilde{g}^{(\varepsilon)}(\varepsilon n) - P_{\mathcal{P}}[\widetilde{g}^{(\varepsilon)}(\varepsilon n)] - P_{\mathcal{M};0}[\widetilde{g}^{(\varepsilon)}(\varepsilon n)]\|_2 \\
&\leqslant \sum_{n=0}^{N_\varepsilon(t)} \varepsilon (3\|\widetilde{g}^{(\varepsilon)}(\varepsilon n)\|_2 + \|\widecheck{\sqrt{I}}\|_2) \leqslant \sum_{n=0}^{N_\varepsilon(t)} 4\varepsilon \max\{\|\widetilde{g}^{(\varepsilon)}(\varepsilon n)\|_2, \|\widecheck{\sqrt{I}}\|_2\} \\
&\leqslant (N_\varepsilon(t) + 1) 4\varepsilon \max\{\|\tilde{g}_0\|_2, \|\widecheck{\sqrt{I}}\|_2\} \leqslant 4(t + \varepsilon) \max\{\|\tilde{g}_0\|_2, \|\widecheck{\sqrt{I}}\|_2\},
\end{aligned}$$

one has

$$\inf_{\varepsilon\in J_2} \|\tilde{\check{g}}^{(\varepsilon)}(t) - \tilde{\check{g}}^{(\varepsilon)}(0)\| \leqslant \inf_{\varepsilon\in J_2}(t+\varepsilon)4\max\{\|\tilde{g}_0\|_2, \|\check{\sqrt{I}}\|_2\}$$
$$= 4t\max\{\|\tilde{g}_0\|_2, \|\check{\sqrt{I}}\|_2\}.$$

Therefore,

$$\lim_{t\to 0} \|\tilde{\check{g}}^{(\varepsilon)}(t) - \tilde{\check{g}}^{(\varepsilon)}(0)\|_2 \leqslant \lim_{t\to 0}\inf_{\varepsilon\in J_2} \|\tilde{\check{g}}^{(\varepsilon)}(t) - \tilde{\check{g}}^{(\varepsilon)}(0)\|$$
$$\leqslant \lim_{t\to 0} 4t\max\{\|\tilde{g}_0\|_2, \|\check{\sqrt{I}}\|_2\} = 0. \quad \square$$

7.2 SOLUTION CANDIDATE SOLVES ERF

Let us show that the solution candidate g provided by Corollary 7.10 solves ERF.

First, observe the following technical result.

Lemma 7.13 (Convergence of $\tilde{\check{g}}^{(\varepsilon)}$ implies convergence of $\bar{\breve{g}}^{(\varepsilon)}$). *Consider the setting of Corollary 7.10, meaning that $\tilde{\check{g}}^{(\varepsilon)} \to g$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.*

Then, $\bar{\breve{g}}^{(\varepsilon)} \to g$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.

Proof. First, check the convergence in L^p_{t,L^2} for $p \in [1,\infty)$. To do so, split the time integration into intervals of size ε: on these intervals, the distance between $\tilde{\check{g}}^{(\varepsilon)}$ and $\bar{\breve{g}}^{(\varepsilon)}$ can be calculated explicitly. To this end, recall that $\lfloor T\rfloor_\varepsilon \to T$ as $\varepsilon \to 0$ (Notation 7.4), whence by dominated convergence theorem — existence of the necessary upper bound follows from Corollary 7.6 — have

$$\int_0^T \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \|\tilde{\check{g}}^{(\varepsilon)}(t) - \bar{\breve{g}}^{(\varepsilon)}(t)\|^p_{L^2} dt = \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \int_0^{\lfloor T\rfloor_\varepsilon} \|\tilde{\check{g}}^{(\varepsilon)}(t) - \bar{\breve{g}}^{(\varepsilon)}(t)\|^p_{L^2} dt$$
$$= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon-1} \int_{n\varepsilon}^{(n+1)\varepsilon} \|\tilde{\check{g}}^{(\varepsilon)}(t) - \bar{\breve{g}}^{(\varepsilon)}(t)\|^p_{L^2} dt,$$

which, inserting the definition of $\tilde{\check{g}}^{(\varepsilon)}$ and $\bar{\breve{g}}^{(\varepsilon)}$, equals

$$\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon-1} \int_0^\varepsilon \Bigg\| \left(1 - \frac{\tau}{\varepsilon}\right) g_n^{(\varepsilon)} + \frac{\tau}{\varepsilon}\Big(\underbrace{(1-2\varepsilon)g_n^{(\varepsilon)} + \varepsilon P_{\mathcal{P}}[g_n^{(\varepsilon)}] + \varepsilon P_{\mathcal{M};0}[g_n^{(\varepsilon)}]}_{\text{do not cancel out}}\Big) - g_n^{(\varepsilon)}\Bigg\|^p_{L^2} d\tau$$
$$= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon-1} \int_0^\varepsilon \tau^{p/2} \left\| -2g_n^{(\varepsilon)} + P_{\mathcal{P}}[g_n^{(\varepsilon)}] + P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\right\|^p_{L^2} =$$

$$= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon - 1} \frac{\varepsilon^{p/2+1}}{\frac{p}{2}+1} \left\| -2g_n^{(\varepsilon)} + P_{\mathcal{P}}[g_n^{(\varepsilon)}] + P_{\mathcal{M};0}[g_n^{(\varepsilon)}] \right\|_{L^2}^p =$$

$$= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \underbrace{\varepsilon^{p/2}}_{\to 0} \underbrace{N_\varepsilon(T)\varepsilon}_{\to T} \frac{1}{\frac{p}{2}+1} \left(3\max\{\|g_0\|_{L^2}, \|\widecheck{\sqrt{I}}\|_{L^2}\} + \|\widecheck{\sqrt{I}}\|_{L^2} \right)^p = 0.$$

Second, check the convergence in L^∞_{t,L^2} and C_{t,L^2}; the calculation is analogous to the previous case.

$$\sup_{t\in(0,T)} \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \|\tilde{g}^{(\varepsilon)}(t) - \check{\tilde{g}}^{(\varepsilon)}(t)\|_{L^2}^p dt =$$

$$\leqslant \sup_{\substack{\tau\in(0,\varepsilon]\\ n\in\{1,\dots,N_\varepsilon(T)\}}} \left\| \left(1-\frac{\tau}{\varepsilon}\right) g_n^{(\varepsilon)} + \frac{\tau}{\varepsilon}\left((1-2\varepsilon) g_n^{(\varepsilon)} + \varepsilon P_{\mathcal{P}}[g_n^{(\varepsilon)}] + \varepsilon P_{\mathcal{M};0}[g_n^{(\varepsilon)}] \right) - g_n^{(\varepsilon)} \right\|_{L^2}$$

$$\leqslant \sup_{\substack{\tau\in(0,\varepsilon]\\ n\in\{1,\dots,N_\varepsilon(T)\}}} \|-2\tau g_n^{(\varepsilon)} + \tau P_{\mathcal{P}}[g_n^{(\varepsilon)}] + \tau P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\|_{L^2}$$

$$\leqslant \varepsilon(3\max\{\|g_0\|_{L^2}, \|\widecheck{\sqrt{I}}\|_{L^2}\} + \|\widecheck{\sqrt{I}}\|_{L^2}) \to 0$$

as $\varepsilon \to 0$.

Both cases imply the desired result, since for any norm $\|\cdot\|$ holds $\|\check{\tilde{g}}^{(\varepsilon)} - g\| \leqslant \|\check{\tilde{g}}^{(\varepsilon)} - \tilde{g}^{(\varepsilon)}\| + \|\tilde{g}^{(\varepsilon)} - g\| \to 0$ as $\varepsilon \to 0$. □

The following remark demonstrates that, formally, the weak derivative of the candidate g solves ERF.

Remark 7.14 (Formal derivative of the solution candidate). Consider the setting of Corollary 7.10, meaning that $\tilde{g}^{(\varepsilon)} \to g$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.

Then, for a test function $\zeta \in C_0^\infty([0,T], L^2(\Omega))$ have

$$\int_0^T \langle \partial_t g(t), \zeta(t)\rangle dt = -\int_0^T \langle g(t), \partial_t\zeta(t)\rangle dt$$

$$= -\int_0^T \left\langle \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \tilde{g}^{(\varepsilon)}(t), \partial_t\zeta(t) \right\rangle dt \overset{(*)}{=} -\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \int_0^{[T]_\varepsilon} \left\langle \tilde{g}^{(\varepsilon)}(t), \partial_t\zeta(t) \right\rangle dt,$$

using the dominated convergence theorem in $(*)$; the necessary upper bound exists due to Corollary 7.6 Equation (7.2). Inserting the definition of $\tilde{g}^{(\varepsilon)}$ and using Notation 7.4, last expression equals

$$-\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon(T)-1} \int_0^\varepsilon \left\langle \frac{\tau}{\varepsilon}(g_{n+1}^{(\varepsilon)} - g_n^{(\varepsilon)}), \partial_t\zeta(n\varepsilon+\tau) \right\rangle dt$$

$$= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon(T)-1} \int_{n\varepsilon}^{(n+1)\varepsilon} \left\langle \frac{1}{\varepsilon}(g_{n+1}^{(\varepsilon)} - g_n^{(\varepsilon)}), \zeta(t) \right\rangle dt$$

by partial integration (the boundary summand is a telescopic sum and cancels itself out exept at the endpoints, where it vanishes due to ζ being continuous and equal to zero at $t = 0$ and $t = T$). Inserting definition of $g^{(\varepsilon)}_{n+1}$, continue with

$$\begin{aligned}
&\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon(T)-1} \int_{n\varepsilon}^{(n+1)\varepsilon} \left\langle \frac{1}{\varepsilon}(-2\varepsilon g_n^{(\varepsilon)} + \varepsilon P_{\mathcal{P}}[g_n^{(\varepsilon)}] + \varepsilon P_{\mathcal{M};0}[g_n^{(\varepsilon)}]), \zeta(t) \right\rangle dt \\
={}& \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \int_0^{\lfloor T\rfloor_\varepsilon} \left\langle -2\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t) + P_{\mathcal{P}}[\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t)] + P_{\mathcal{M};0}[\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t)], \zeta(t) \right\rangle dt
\end{aligned}$$

where we used definition of $\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}$ to combine all integration domains of size ε. Finally, by the dominated convegence theorem, applicable due to Corollary 7.6 Equation (7.4), the limit can be brought back into the integral

$$= \int_0^T \left\langle -2 \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t) + \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{P}}[\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t)] + \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t)], \zeta(t) \right\rangle dt. \tag{7.9}$$

From Lemma 7.13 and continuity of $P_{\mathcal{P}}$ we know that

$$\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t) = g(t) \qquad \text{and} \qquad \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{P}}[\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t)] = P_{\mathcal{P}}[g(t)],$$

The presented calculation is formal in the following sense: at this point we do not know whether $\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t)]$ exists; thus, we do not know whether $\partial_t g$ is well-defined. Rather than establishing existence of $\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t)]$ directly, we shall show that $\partial_t g$ exists and belongs to $L^2(\Omega)$ for Lebesgue–almost all $t \in (0, T)$. From this will follow existence of $\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\overset{\sqcap\!\sqcup}{g}^{(\varepsilon)}(t)]$; and Equation (7.9) will become rigorous.

To establish existence of $\partial_t g$, we show that $t \mapsto g(t)$ is Lipshitz and therefore a. e. Fréchet-differentiable by the general case of the Rademacher theorem from [AK00] (therein, see Thm. 3.5 and Fréchet differentiability remark on p. 528).

Theorem 7.15 (Generalized Rademacher theorem [AK00]). *Let $\mathcal{H}$ be a Hilbert space, let $f : \mathbb{R} \to \mathcal{H}$ be Lipschitz. Then, f is Fréchet differentiable at Lebesgue–almost all $t \in \mathbb{R}$.*

Corollary 7.16. *Consider the setting of Corollary 7.10, meaning that $\widetilde{g}^{(\varepsilon)} \to g$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.*

Then, there exists a function $\partial_t g : (0, T) \to L^2(\Omega)$ such that

$$\lim_{\varepsilon\to 0} \frac{\|g(t+\varepsilon) - g(t) - \varepsilon\partial_t g(t)\|_{L^2}}{\varepsilon} = 0$$

for Lebesgue–almost all $t \in (0, T)$.

Proof. To show that g is Lipshitz, let $t, s \in (0, T)$ with $t < s$. Then,

$$\begin{aligned}
\|g(s) - g(t)\|_2 &= \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} \|\bar{g}^{(\varepsilon)}(s) - \bar{g}^{(\varepsilon)}(t)\|_2 \\
&\leqslant \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} \sum_{n=N_\varepsilon(t)}^{N_\varepsilon(s)+1} \varepsilon \| - 2g_n^{(\varepsilon)} + P_{\mathcal{P}}[g_n^{(\varepsilon)}] + P_{\mathcal{M};0}[g_n^{(\varepsilon)}] \| \\
&\leqslant \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} \underbrace{(N_\varepsilon(s) + 1 - N_\varepsilon(t))\varepsilon}_{= \lfloor s \rfloor_\varepsilon - \lfloor t \rfloor_\varepsilon + \varepsilon} (3 \max\{\|g_0\|_2, \|\check{\sqrt{I}}\|_2\} + \|\check{\sqrt{I}}\|_2) \\
&\leqslant \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} (s - t + 3\varepsilon) 4 \max\{\|g_0\|_2, \|\check{\sqrt{I}}\|_2\} \\
&\leqslant (s - t) 4 \max\{\|g_0\|_2, \|\check{\sqrt{I}}\|_2\}.
\end{aligned}$$

The claim follows by the aforementioned Rademacher theorem. □

We are now ready to show existence of the limit $\lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} P_{\mathcal{M};0}[\bar{g}^{(\varepsilon)}(t)]$. This limit is required by Remark 7.14 to turn derivation of $\partial_t g$ from formal to rigorous.

Lemma 7.17 (Limit of $P_{\mathcal{M}}[\bar{g}^{(\varepsilon)}]$ exists). *Consider the setting of Corollary 7.10, meaning that* $\tilde{g}^{(\varepsilon)} \to g$ *in* $L^p_{t,L^2} \cap C_{t,L^2}$ *as* $\varepsilon \to 0$ *for* $\varepsilon \in J_2$.

Then, there exists a (measurable) function

$$\varphi \colon (0, T) \to \{\tilde{\varphi} \colon \Omega_F \to [0, 2\pi) \mid \sin \tilde{\varphi}(t) \text{ odd for a. a. } t \in (0, T)\}$$

such that

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} P_{\mathcal{M};0}[\bar{g}^{(\varepsilon)}(t)] = P_{\mathcal{M};\varphi(t)}[g(t)]$$

in $L^2(\Omega)$ *for Lebesgue–almost all* $t \in (0, T)$.

Proof. By Corollary 7.16, $\partial_t g$ exists and belongs to $L^2(\Omega)$ for almost all $t \in (0, T)$. Define

$$m_t := \partial_t g(t) + 2g(t) - P_{\mathcal{P}}[g(t)].$$

By Equation (7.9) from Remark 7.14, for any test function $\zeta \in C_0^\infty([0, T], L^2(\Omega))$ have

$$\int_0^T \langle m_t, \zeta(t) \rangle \, dt = \int_0^T \left\langle \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} P_{\mathcal{M};0}[\bar{g}^{(\varepsilon)}(t)], \zeta(t) \right\rangle dt;$$

therefore, $\lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} P_{\mathcal{M};0}[\bar{g}^{(\varepsilon)}(t)]$ exists and equals m_t in $L^2(\Omega)$ for Lebesgue–almost all $t \in (0, T)$.

Let us show that there exists $\varphi(t)$ such that $m_t = P_{\mathcal{M};\varphi}[g(t)]$ for a. a. $t \in (0, T)$ in the following three steps:

1) show that $|\hat{m}_t| = \sqrt{I}$;

2) show that $\hat{m}_t(k) = \widehat{g(t)}(k)$ at those k where $\widehat{g(t)}(k)$ is not equal zero;

3) set $\varphi(t,k) = \arg \hat{m}_t(k)$ at all other k.

1) One has $|\hat{m}_t| = \sqrt{I}$, since

$$\big\| |\sqrt{I}| - |\hat{m}_t| \big\|_{L^2} = \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \big\| |\sqrt{I}| - |\hat{m}_t| \big\|_{L^2}$$

$$= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \Big\| \Big| \mathcal{F}\left[P_{\mathcal{M};0}[\breve{g}^{(\varepsilon)}(t)] \right] \Big| - |\hat{m}_t| \Big\|_{L^2} \leqslant \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \| P_{\mathcal{M};0}[\breve{g}^{(\varepsilon)}(t)] - m_t \|_{L^2} = 0,$$

because for any complex numbers a, b holds

$$\big| |a| - |b| \big|^2 = |a|^2 + |b|^2 - 2|ab| \leqslant |a|^2 + |b|^2 - 2\langle a, b\rangle_{\mathbb{C}} \leqslant |a-b|^2.$$

2) Let $k \in \Omega_F$ be such that $\hat{g}(t,k) \neq 0$. Observe that for any $\varphi_1, \varphi_2 \colon \Omega_F \to [0, 2\pi)$ holds

$$\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \mathcal{F}\left[P_{\mathcal{M};\varphi_1}[\breve{g}^{(\varepsilon)}(t)] \right](k) - \mathcal{F}\left[P_{\mathcal{M};\varphi_2}[g(t)] \right](k) = 0. \tag{7.10}$$

Indeed, since $\hat{g}(t,k) \neq 0$ and $\mathcal{F}[\breve{g}^{(\varepsilon)}(t)](k) \to \hat{g}(t,k)$, there exists $\tilde{\varepsilon} > 0$ such that for all $\varepsilon < \tilde{\varepsilon}, \varepsilon \in J_2$, holds $\mathcal{F}[\breve{g}^{(\varepsilon)}(t)](k) \neq 0$. Therefore, one can estimate the expression in the left-hand side of Equation (7.10) by

$$\left| \sqrt{I}(k) \frac{\hat{g}(t,k)}{|\hat{g}(t,k)|} - \sqrt{I}(k) \frac{\mathcal{F}[\breve{g}^{(\varepsilon)}(t)](k)}{|\mathcal{F}[\breve{g}^{(\varepsilon)}(t)](k)|} \right|$$

$$\leqslant \frac{\sqrt{I}(k)}{|\hat{g}(t,k)|} \left| \hat{g}(t,k) - |\hat{g}(t,k)| \frac{\mathcal{F}[\breve{g}^{(\varepsilon)}(t)](k)}{|\mathcal{F}[\breve{g}^{(\varepsilon)}(t)](k)|} \right|$$

$$\leqslant 2 \frac{\sqrt{I}(k)}{|\hat{g}(t,k)|} \left| \hat{g}(t,k) - \mathcal{F}[\breve{g}^{(\varepsilon)}(t)](k) \right| \to 0 \quad \text{as } \varepsilon \to 0, \varepsilon \in J_2.$$

In this calculation we used Lemma D.1, according to which for two complex numbers a, b with $b \neq 0$ holds $\left| a - \frac{b}{|b|}|a| \right| \leqslant 2|a-b|$.

3) Let $\varphi(t,k) = \arg \hat{m}_t(k)$ if $\widehat{g(t)}(k) = 0$ and $\varphi(t,k) = 0$ else. Then, by construction, $m_t = P_{\mathcal{M};\varphi(t)}[g(t)]$.

From this construction also follows that $\sin \varphi(t)$ is odd for a. a. $t \in (0, T)$.

Overall,

$$\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\breve{g}^{(\varepsilon)}(t)] = P_{\mathcal{M};\varphi(t)}[g(t)]$$

in $L^2(\Omega)$ for Lebesgue–almost all $t \in (0, T)$. □

Theorem 7.18 (Global solutions exist). *Consider the setting of Corollary 7.10, meaning that: the domain Ω is bounded and $\sqrt{I}$ is sufficiently* *smooth (as specified in Definition 7.1); the initial value $\tilde{g}_0 \in L^2(\Omega)$; the time exponent $p \in (1, \infty)$.*

Then, there exist φ (specified in Definition 7.2) and

$$g \in L^\infty\big((0,\infty); L^2(\Omega)\big) \cap C\big([0,\infty); L^2(\Omega)\big)$$

such that g belongs to $L^p\big((0,T); L^2(\Omega)\big)$ and is a weak solution of the equation

$$\partial_t g_t = -2g_t + P_{\mathcal{M};\varphi_t}[g_t] + P_{\mathcal{P}}[g_t]$$

for any $T \in (0,\infty)$.

Proof. Let $T > 0$. First, construct g on $[0,T]$; then, inductively extend the domain of g to $[0,\infty)$.

Let $J_1 \subset (0, \frac{1}{2}]$ be the set of discretization step sizes with $0 \in \overline{J_1}$. Let $\widetilde{g}^{(\varepsilon)}$ be as in Definition 7.5, with initial value $\tilde{g}_0$. By Corollary 7.10, there exists a set $J_2 \subset J_1$ with $0 \in \overline{J_2}$ such that $\widetilde{g}^{(\varepsilon)}$ converges strongly to some $g \in L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.

By Corollary 7.16, $\partial_t g$ exists. By Lemma 7.17, there exists a function

$$\varphi\colon (0,T) \to \big\{\tilde{\varphi}\colon \Omega_F \to [0,2\pi) \mid \sin\tilde{\varphi}(t) \text{ odd for a. a. } t \in (0,T)\big\}$$

such that

$$\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\overline{\widetilde{g}}^{(\varepsilon)}(t)] = P_{\mathcal{M};\varphi(t)}[g(t)]$$

in $L^2(\Omega)$ for Lebesgue–almost all $t \in (0,T)$. Therefore, Equation (7.9) from Remark 7.14 holds rigorously, and g solves ERF on $(0,T)$.

By Lemma 7.12, $g(0) = \tilde{g}_0$.

It remains to show that g can be extended to

$$g \in L^\infty\big((0,\infty); L^2(\Omega)\big) \cap C\big([0,\infty); L^2(\Omega)\big).$$

Apply the above argumentation using starting value $g(T)$ to extend g to $(0,2T)$. By Corollary 7.6,

$$\|g(2T)\|_{L^2} \leqslant \max\{\|g(T)\|_{L^2}, \sqrt{I}\} \leqslant \max\{\|g(0)\|_{L^2}, \sqrt{I}\}.$$

Proceed inductively to show the desired claim. □

Remark 7.19 (Other convex constraints). One can generalize Theorem 7.18 to other additional constraints $\mathcal{A}$, provided $P_{\mathcal{A}}$ is AP-bounded as discussed in Remark 7.7, and continuous, which is necessary for Remark 7.14.

For example, Theorem 7.18 holds for $\mathcal{A} = (\mathcal{P} \cap \mathcal{S}) \cap \mathcal{Q}^{\mathrm{F}}_{C;r}$ (regularized "support + positivity") presented in Section 6.3.4.2 and shown in Figure 6.2. Indeed, such $\mathcal{A}$ will be AP-bounded on $H^1(\Omega)$ by definition of $\mathcal{Q}^{\mathrm{F}}_{C;r}$, and $P_{\mathcal{A}}$ will be continuous by Lemma 6.37 applied to $(\mathcal{P} \cap \mathcal{S})$ and $\mathcal{Q}^{\mathrm{F}}_{C;r}$.

A higher regularity of solutions in space may be achieved, if one uses Aubin-Lions-Simon with more regular Banach spaces than H^1; for example, Banach spaces presented in Lemma 6.45.

Remark 7.20 (Other non-convex constraints). The question whether one can generalize Theorem 7.18 to additional constraints with non-continuous projections remains open.

For example, for the regularized constraint $\mathcal{A} = \mathcal{T}_s(\nu) \cap \mathcal{Q}^F_{C;r}$ (regularized non-negativity + support size), one must first establish that projection onto $\mathcal{A}$ can be efficiently calculated — for example, using the Dykstra algorithm (to remain relevant for applications).

The regularized $\mathcal{A}$ will be AP-bounded on $H^1(\Omega)$ by definition of $\mathcal{Q}^F_{C;r}$.

However, the resulting projection operator will be multivalued, and any corresponding projecton selection will not be continuous. To prove Theorem 7.18 as presented here, one will have to establish that

$$\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \left(P_{\mathcal{A}}[\tilde{g}^{(\varepsilon)}(t)] + P_{\mathcal{M};0}[\tilde{g}^{(\varepsilon)}(t)]\right) = P_{\mathcal{A}}[\tilde{g}^{(\varepsilon)}(t)] + P_{\mathcal{M};\varphi}[\tilde{g}^{(\varepsilon)}(t)]. \quad (7.11)$$

While existence of the limit on the left-hand side follows from Rademacher's theorem as in Lemma 7.17, it is less obvious how to establish existence of $\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{A}}[\tilde{g}^{(\varepsilon)}(t)]$ or existence of $\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};\varphi}[\tilde{g}^{(\varepsilon)}(t)]$ separately.

7.3 SOLUTION PROPERTIES

This section establishes some properties of an ERF solution. Notably, Proposition 7.22 shows that any solution exponentially decreases $|E_{\mathcal{M}}[g(t)] - E_{\mathcal{P}}[g(t)]|$ in time. Remark 7.25 discusses difficulties that need to be overcome to prove convergence of $g(t)$ to a fixed point.

The following corollary shows that the solution g of ERF dissipates energy. It is noteworthy that the corollary does not require Fréchet-differentiability of $E_{\mathcal{M}}$ at any $g(t)$.

COROLLARY 7.21 (ENERGY DISSIPATION). *Let $\sqrt{I}, \tilde{g}_0, g$ and φ be as in Theorem 7.18. Then,*

$$E[g(T)] - E[g(0)] \leqslant \int_0^T \|2g(t) - P_{\mathcal{P}}[g(t)] - P_{\mathcal{M};\varphi}[g(t)]\|_2^2 dt,$$

where $E[g] = E_{\mathcal{P}}[g] + E_{\mathcal{M}}[g]$. (Recall that $E_{\mathcal{M}}$ does not depend on φ, see Example 4.6.)

Proof. The idea of the proof is to show that one can rigorously take the limit $\varepsilon \to 0$ in Proposition 6.4.

Let J_2 and $\breve{g}^{(\varepsilon)}$ be as in proof of Theorem 7.18. By Proposition 4.4 and Proposition 6.4,

$$\begin{aligned}E[g(T)] - E[g(0)] &\leqslant \liminf_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} E[\breve{g}^{(\varepsilon)}(T)] - E[\breve{g}^{(\varepsilon)}(0)]\\ &\leqslant -\limsup_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \Bigg(\sum_{n=0}^{N_\varepsilon(T)-1} \varepsilon(1-\varepsilon)\|2g_n^{(\varepsilon)} - P_{\mathcal{P}}[g_n^{(\varepsilon)}] - P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\|_2^2\\ &+ (T-\lfloor T\rfloor_\varepsilon)(1-T+\lfloor T\rfloor_\varepsilon)\|2g_{N_\varepsilon(T)}^{(\varepsilon)} - P_{\mathcal{P}}[g_{N_\varepsilon(T)}^{(\varepsilon)}] - P_{\mathcal{M};0}[g_{N_\varepsilon(T)}^{(\varepsilon)}]\|_2^2\Bigg).\end{aligned}$$

Split the first term in the limit by splitting $\varepsilon(1-\varepsilon)$ into ε and $-\varepsilon^2$, and the second term in the limit by splitting $(T-\lfloor T\rfloor_\varepsilon)(1-T+\lfloor T\rfloor_\varepsilon)$ into $(T-\lfloor T\rfloor_\varepsilon)$ and $-(T-\lfloor T\rfloor_\varepsilon)^2$. Observe that by definition of $\breve{g}^{(\varepsilon)}$, for terms with ε and $(T-\lfloor T\rfloor_\varepsilon)$ holds

$$\begin{aligned}&\sum_{n=0}^{N_\varepsilon(T)} \varepsilon\|2g_n^{(\varepsilon)} - P_{\mathcal{P}}[g_n^{(\varepsilon)}] - P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\|_2^2\\ &\quad + (T-\lfloor T\rfloor_\varepsilon)\|2g_{N_\varepsilon(T)}^{(\varepsilon)} - P_{\mathcal{P}}[g_{N_\varepsilon(T)}^{(\varepsilon)}] - P_{\mathcal{M};0}[g_{N_\varepsilon(T)}^{(\varepsilon)}]\|_2^2\\ &\quad = \int_0^T \|2\breve{g}^{(\varepsilon)}(t) - P_{\mathcal{P}}[\breve{g}^{(\varepsilon)}(t)] - P_{\mathcal{M};0}[\breve{g}^{(\varepsilon)}(t)]\|_2^2 dt\\ &\quad\quad \to \int_0^T \|2g(t) - P_{\mathcal{P}}[g(t)] - P_{\mathcal{M};\varphi(t)}[g(t)]\|_2^2 dt\end{aligned}$$

as $\varepsilon\to 0$ by Remark 7.14 and Lemma 7.17. As for the terms with $-\varepsilon^2$ and $-(T-\lfloor T\rfloor_\varepsilon)^2$,

$$\begin{aligned}&\Bigg|\sum_{n=0}^{N_\varepsilon(T)} -\varepsilon^2 \left\|2g_n^{(\varepsilon)} - P_{\mathcal{P}}[g_n^{(\varepsilon)}] - P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\right\|_2^2\\ &\quad - (T-\lfloor T\rfloor_\varepsilon)^2 \left\|2g_{N_\varepsilon(T)}^{(\varepsilon)} - P_{\mathcal{P}}[g_{N_\varepsilon(T)}^{(\varepsilon)}] - P_{\mathcal{M};0}[g_{N_\varepsilon(T)}^{(\varepsilon)}]\right\|_2^2\Bigg|\\ &\quad \leqslant N_\varepsilon(T)\varepsilon^2 \sup_{n\in\{0,\dots,N_\varepsilon(T)\}} \left\|2g_n^{(\varepsilon)} - P_{\mathcal{P}}[g_n^{(\varepsilon)}] - P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\right\|_2^2\\ &\quad\quad + \varepsilon^2 \left\|2g_{N_\varepsilon(T)+1}^{(\varepsilon)} - P_{\mathcal{P}}[g_{N_\varepsilon(T)+1}^{(\varepsilon)}] - P_{\mathcal{M};0}[g_{N_\varepsilon(T)+1}^{(\varepsilon)}]\right\|_2^2 \to 0\end{aligned}$$

as $\varepsilon\to 0$, since $N_\varepsilon(T)\varepsilon\leqslant T$ by definition, and since $\|2g_n^{(\varepsilon)} - P_{\mathcal{P}}[g_n^{(\varepsilon)}] - P_{\mathcal{M};0}[g_n^{(\varepsilon)}]\|_2^2$ is uniformly bounded from above for all $n\in\mathbb{N}_0$ by Corollary 7.6. □

The following proposition demonstrates that a solution tends to an energy equilibrium manifold where $E_{\mathcal{P}}[g(t)] = E_{\mathcal{M}}[g(t)]$.

Proposition 7.22 (Energy balance). *Let $\sqrt{I}, \tilde{g}_0, g$ and φ be as in Theorem 7.18. Let $T>0$.*

Define $D(t) := E_{\mathcal{P}}[g(t)] - E_{\mathcal{M}}[g(t)]$. *Then,*

$$D(t) = D(0)\exp(-2t) \quad \text{for all } t \in (0, T).$$

In particular, if $D(0) = 0$*, then* $D(t) = 0$ *for almost all* $t \in (0, T)$*, and if* $D(0) > 0$*, then* $D(t) > 0$ *for almost all* $t \in (0, T)$*.*

Proof. By the chain rule,

$$\begin{aligned}\frac{d}{dt}D(t) &= \langle g - P_{\mathcal{P}}[g], -\partial_t g\rangle - \langle g - P_{\mathcal{M}}[g], -\partial_t g\rangle \\ &= \langle g - P_{\mathcal{M}}[g] - (g - P_{\mathcal{P}}[g]), 2g - P_{\mathcal{M}}[g] - P_{\mathcal{P}}[g]\rangle \\ &= \|g - P_{\mathcal{M}}[g]\|_2^2 - \|g - P_{\mathcal{P}}[g]\|_2^2 = -2D(t),\end{aligned}$$

where we used the shorthand notations $g = g(t)$ and $P_{\mathcal{M}} = P_{\mathcal{M};\varphi}$ for readability.

The claim follows from the differential equation for $D(t)$. □

The following example shows that in a trivial case when the phase φ of a solution f is known, one can construct an explicit solution for the starting value 0.

Example 7.23 (An explicit solution). Let $f \in \mathcal{P}$, let $\sqrt{I} = |\hat{f}|$, let φ be the phase of $\hat{f}$.

Let $g(t) = (1 - e^{-t})f$ for $t \in [0, \infty)$. Then, g solves ERF with $g(0) \equiv 0$, i. e.

$$\partial_t g(t) = -2g(t) + P_{\mathcal{P}}[g(t)] + P_{\mathcal{M};\varphi}[g(t)]$$

for all $t > 0$.

Indeed, since $f \in \mathcal{P}$, we have $g(t) \in \mathcal{P}$, meaning that $P_{\mathcal{P}}[g(t)] = g(t)$. Further, since the phase of $\widehat{g(t)}$ is equal to φ, and since $P_{\mathcal{M};\varphi}[0] = f$ by the choice of φ, have $P_{\mathcal{M};g(t)} = f$ for all $t \geqslant 0$. Overall,

$$\partial_t g(t) = e^{-t}f = -f + e^{-t}f + f = -g + P_{\mathcal{M};\varphi}[g] = -2g + P_{\mathcal{P}}[g] + P_{\mathcal{M};\varphi}[g].$$

Lemma 7.24 (Evenness invariance). *Let* $\sqrt{I}, \tilde{g}_0, g$ *and* φ *be as in Theorem 7.18.*

Assume that Ω *is symmetric (meaning that* $\Omega = -\Omega$*), and that* $\tilde{g}_0$ *is even. Then,* $g(t)$ *is even for all* $T \in (0, \infty)$*.*

Proof. Follows by construction of $g(t)$. If $\tilde{g}_0 \in L^2(\Omega)$ is even, then all $g_n^{(\varepsilon)}$ are even, and so are $\tilde{g}^{(\varepsilon)}(t)$ and $g(t)$ for all $t \in (0, \infty)$. □

Remark 7.25 (Difficulties with convergence of an ERF solution). Let $\sqrt{I}, \tilde{g}_0, g$ and φ be as in Theorem 7.18.

Just as from energy dissipation Corollary 6.5 follows existence of fixed points of dERF (Proposition 6.32), from energy dissipation Corollary 7.21 follows that any

$$f \in \operatorname*{Ls}_{t\to\infty}\{g(t)\}$$

is a fixed point of ERF, i. e. $2f = P_{\mathcal{P}}[f] + P_{\mathcal{M}}[f]$. (Here, Ls is the Kuratowski limit superior.)

The question whether $g(t)$ converges to such f remains open.

Using $2f = P_{\mathcal{P}}[f] + P_{\mathcal{M}}[f]$, one has

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\|g(t)-f\|_2^2 &= -\langle g(t)-f, 2g(t) - P_{\mathcal{M}}[g(t)] - P_{\mathcal{P}}[g(t)]\rangle \\ &= -\langle g(t)-f, 2g(t) - 2f + 2f - P_{\mathcal{M}}[g(t)] - P_{\mathcal{P}}[g(t)]\rangle \\ &= -2\|g(t)-f\|_2^2 + 2\langle g(t)-f, (P_{\mathcal{M}}+P_{\mathcal{P}})[g(t)] - (P_{\mathcal{M}}+P_{\mathcal{P}})[f]\rangle.\end{aligned}$$

Therefore, if

$$\langle g(t)-f, (P_{\mathcal{M}}+P_{\mathcal{P}})[g(t)] - (P_{\mathcal{M}}+P_{\mathcal{P}})[f]\rangle \leqslant \|g(t)-f\|_2^2 \qquad (7.12)$$

for all sufficiently large t, then $g(t)$ converges to f.

Since $P_{\mathcal{P}}$ is nonexpansive, $\|P_{\mathcal{P}}[g(t)] - P_{\mathcal{P}}[f]\|_2 \leqslant \|g-f\|$. In simulations we observed

$$\|(P_{\mathcal{M}}+P_{\mathcal{P}})[g(t)] - (P_{\mathcal{M}}+P_{\mathcal{P}})[f]\|_2 \leqslant \|g(t)-f\|_2$$

for large enough t, which implies Equation (7.12).

However, it is not obvious how to show this inequality, since for $P_{\mathcal{M}}$, for any $f \in \mathcal{H}$ one can find $g_* \in \mathcal{H}$ with

$$\|P_{\mathcal{M}}[g_*] - P_{\mathcal{M}}[f]\|_2 \nleqslant \|g_* - f\|_2.$$

In particular, in simulations we often observed

$$\|P_{\mathcal{M}}[g(t)] - P_{\mathcal{M}}[f]\|_2 > \|g(t)-f\|_2$$

when g was converging to f.

A simpler subcase of this problem is a question of convergence for even functions. Indeed, for any even $f, g \in \mathcal{H}(\Omega)$ with symmetric Ω, their Fourier transforms are real-valued. Therefore, for $k \in \Omega_F$, one has

$$|\widehat{P_{\mathcal{M}}[g]}(k) - \widehat{P_{\mathcal{M}}[f]}(k)| = \begin{cases} 0 & \text{if } \hat{g}(k) \text{ and } \hat{f}(k) \text{ have the same sign,} \\ 2\sqrt{I}(k) & \text{else.}\end{cases}$$

To show convergence, one has to show that the contraction effect dominates over the expansion effect. In even-restricted simulations, we observed

$$\|P_{\mathcal{M}}[g(t)] - P_{\mathcal{M}}[f]\|_2 \leqslant \|g(t)-f\|_2,$$

which correlates with rapid convergence to fixed points that one can observe for the even-restricted dERF (cf. the bottom row of Figure 10.7).

8 FIXED POINTS

The goal of this chapter is to analyze fixed points of ERF.

To this end, Section 8.1 establishes that there exists a correspondence between fixed points of AP and APF. In case of ERF, this correspondence essentially is conditional on $\|\sqrt{I}/\hat{f}\|_\infty$:

$$g = P_{\mathcal{P}}[f] \text{ is a fixed point of ER} \quad \Leftrightarrow \quad f = 0.5(P_{\mathcal{P}}[g] + P_{\mathcal{M}}[g]) \text{ is a fixed point of ERF and } \|\sqrt{I}/\hat{f}\|_\infty \leqslant 2,$$

where the exact formulation of latter condition is slightly more complicated due to the phase multiplicity in $P_{\mathcal{M};\varphi}$.

Section 8.2 presents criteria that are sufficient to establish stability or instability of ERF fixed points for certain cases. Roughly speaking, a fixed point f of ERF is unstable, if $2 < \|\sqrt{I}/\hat{f}\|_\infty < \infty$, see Proposition 8.6. Further, a fixed point f is stable along all even directions, if it is even and if $\|\sqrt{I}/\hat{f}\|_\infty < \infty$, see Corollary 8.10. This result allows one to numerically generate unstable fixed points.

At the end of the chapter, the formal Hessian at fixed points is presented; in Chapter 10, it is used to discuss numerical stability of fixed points.

8.1 FIXED POINT CORRESPONDENCE OF AP AND APF

This section contains two results: Proposition 8.1 demonstrates correspondence between fixed points of AP and APF, and Corollary 8.2 reformulates this correspondence for the particular case of phase retrieval. These results justify that to understand fixed points of ER, one can study fixed points of ERF.

Proposition 8.1 (Fixed points of AP and APF). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be weakly closed. Consider the following equations.*

$$g = P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g]; \tag{8.1}$$

$$0 = -2f + P_{\mathcal{X}}[f] + P_{\mathcal{Y}}[f]; \tag{8.2}$$

$$\begin{cases} P_{\mathcal{X}}\left[2f - P_{\mathcal{X}}[f]\right] = P_{\mathcal{X}}[f]; \\ P_{\mathcal{Y}}\left[2f - P_{\mathcal{Y}}[f]\right] = P_{\mathcal{Y}}[f]. \end{cases} \tag{8.3}$$

Then:

i) If g satisfies Equation (8.1), then $f := \frac{P_{\mathcal{X}}[g]+P_{\mathcal{Y}}[g]}{2}$ satisfies Equations (8.2) and (8.3).

ii) If f satisfies Equations (8.2) and (8.3), then $g := P_{\mathcal{X}}[f]$ satisfies Equation (8.1), *see Figure 8.1.*

Proof. i) If $g \in \mathcal{H}$ satisfies (8.1), then

$$P_{\mathcal{X}}[g] = P_{\mathcal{X}}\left[P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g]\right] = P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g] = g.$$

Therefore,

$$\begin{aligned} P_{\mathcal{X}}[f] &= P_{\mathcal{X}}\left[\frac{P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g]}{2}\right] = P_{\mathcal{X}}\left[\frac{g + P_{\mathcal{Y}}[g]}{2}\right] \\ &= P_{\mathcal{X}}\left[\frac{P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g] + P_{\mathcal{Y}}[g]}{2}\right] = P_{\mathcal{X}}\left[P_{\mathcal{Y}}[g]\right] = g = P_{\mathcal{X}}[g], \end{aligned}$$

where we have used Lemma 3.8 with $\varepsilon = 1/2$ in the third to last equality. Furthermore,

$$P_{\mathcal{Y}}[f] = P_{\mathcal{Y}}\left[\frac{P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g]}{2}\right] = P_{\mathcal{Y}}\left[\frac{g + P_{\mathcal{Y}}[g]}{2}\right] = P_{\mathcal{Y}}[g],$$

where we have used Lemma 3.8 with $\varepsilon = 1/2$ in the last equality. Overall,

$$P_{\mathcal{X}}[f] + P_{\mathcal{Y}}[f] = P_{\mathcal{X}}[g] + P_{\mathcal{Y}}[g] = 2f,$$

meaning that (8.2) is satisfied. In particular,

$$P_{\mathcal{X}}\left[2f - P_{\mathcal{X}}[f]\right] = P_{\mathcal{X}}\left[P_{\mathcal{Y}}[f]\right] = P_{\mathcal{X}}\left[P_{\mathcal{Y}}[g]\right] = g = P_{\mathcal{X}}[f],$$

and

$$P_{\mathcal{Y}}\left[2f - P_{\mathcal{Y}}[f]\right] = P_{\mathcal{Y}}\left[P_{\mathcal{X}}[f]\right] = P_{\mathcal{Y}}[g] = P_{\mathcal{Y}}[f].$$

ii) If $f \in \mathcal{H}$ satisfies (8.2) and (8.3), then

$$\begin{aligned} P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g] &= P_{\mathcal{X}} \circ P_{\mathcal{Y}}\left[P_{\mathcal{X}}[f]\right] = P_{\mathcal{X}} \circ P_{\mathcal{Y}}\left[2f - P_{\mathcal{Y}}[f]\right] \\ &= P_{\mathcal{X}} \circ P_{\mathcal{Y}}[f] = P_{\mathcal{X}}\left[2f - P_{\mathcal{X}}[f]\right] = P_{\mathcal{X}}[f] = g. \end{aligned}$$ □

Corollary 8.2 (Fixed points of ER). *Let $\mathcal{A} \subset \mathcal{H}(\Omega)$ be weakly closed, with $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$. Let $P_{\mathcal{A}}$ be such that $P_{\mathcal{A}}\left[2g - P_{\mathcal{A}}[g]\right] = P_{\mathcal{A}}[g]$ (for example, $P_{\mathcal{A}} \in \{P_{\mathcal{P}}, P_{\mathcal{S}} \cap P_{\mathcal{P}}\}$, see Lemma 5.7 and Example 5.8). Let $\sqrt{I} \in \widehat{\mathcal{H}}(\Omega_F)$ be non-negative, let $\varphi\colon \Omega_F \to [0, 2\pi)$ be measurable and such that $\sin\varphi$ is odd. Then,*

$$P_{\mathcal{A}} \circ P_{\mathcal{M};\varphi}[g] = g, \tag{8.4}$$

if and only if

$$\begin{gathered} 2f = P_{\mathcal{A}}[f] + P_{\mathcal{M}}[f]; \left\| \tfrac{\sqrt{I}}{|\hat{f}|} \mathbb{1}_{\operatorname{supp}\sqrt{I}} \right\|_{\infty} \leqslant 2, \\ \text{and } \sqrt{I}(k) = 2|\hat{f}(k)| \text{ for } k \in \operatorname{supp}\sqrt{I} \text{ implies } e^{i\varphi(k)} = \tfrac{\hat{f}(k)}{|\hat{f}(k)|}, \end{gathered} \tag{8.5}$$

where

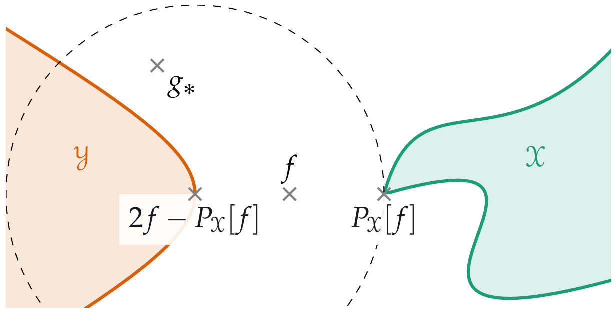

Figure 8.1: Illustration to Proposition 8.1 (Fixed points of AP and APF). The point $g = P_\mathcal{X}[f]$ is a fixed point of AP if and only if: i) $f = \frac{P_\mathcal{X}[g]+P_\mathcal{Y}[g]}{2}$ is a fixed point of APF; ii) $P_\mathcal{X}[2f - P_\mathcal{X}[f]] = P_\mathcal{X}[f]$. The condition ii) is can not be omitted. For example, if g_* were to belong to $\mathcal{X}$, then one would have $P_\mathcal{X}[2f - P_\mathcal{X}[f]] = g_*$. Then, f would be a fixed point of APF, but $P_\mathcal{X}[f]$ would not be a fixed point of AP.

(i) $f = \frac{1}{2}\left(P_\mathcal{A}[g] + P_{\mathcal{M};\varphi}[g]\right)$*, if* (8.4) *is known;*

(ii) $g = P_\mathcal{A}[g]$*, if* (8.5) *is known.*

Proof. Check the additional conditions (8.3) for the case $P_\mathcal{X} = P_\mathcal{A}, P_\mathcal{Y} = P_{\mathcal{M};\varphi}$. Let $f \in \mathcal{H}$. For $P_\mathcal{A}$, the necessary condition holds by assumption. For $P_{\mathcal{M};\varphi}$, equation

$$P_{\mathcal{M};\varphi}\left[2f - P_{\mathcal{M};\varphi}[f]\right] = P_{\mathcal{M};\varphi}[f] \tag{8.6}$$

holds if and only if

$$2\hat{f} - \sqrt{I}\frac{\hat{f}}{|\hat{f}|}\mathbb{1}_{\{\hat{f}\neq 0\}} - \sqrt{I}e^{i\varphi}\mathbb{1}_{\{\hat{f}=0\}} \quad \text{and} \quad \sqrt{I}\frac{\hat{f}}{|\hat{f}|}\mathbb{1}_{\{\hat{f}\neq 0\}} + \sqrt{I}e^{i\varphi}\mathbb{1}_{\{\hat{f}=0\}}$$

satisfy one of the following conditions at every point $k \in \operatorname{supp}\sqrt{I}$:

(i) if both sides are different from 0 at k, they have the same phase;

(ii) if only one of the sides is different from 0 at k, its phase equals $\varphi(k)$;

(iii) both sides are equal to 0 at k,

cf. Figure 8.2. Note that if $k \notin \operatorname{supp}\sqrt{I}$, Fourier transform of Equation (8.6) at k is trivially true since both sides are equal to 0. Let us check conditions (i)-(iii) for $k \in \operatorname{supp}\sqrt{I}$. Condition (i) means that

$$2|\hat{f}| - \sqrt{I}\mathbb{1}_{\{\hat{f}\neq 0\}} \quad \text{and} \quad \sqrt{I}\mathbb{1}_{\{\hat{f}\neq 0\}} \tag{8.7}$$

have the same phase at k as long as $\sqrt{I}(k) \neq 2|\hat{f}(k)|$. Condition (ii) means that

$$-\sqrt{I}\mathbb{1}_{\{\hat{f}=0\}}e^{i\varphi} = \sqrt{I}\mathbb{1}_{\{\hat{f}=0\}}e^{i\varphi} \tag{8.8}$$

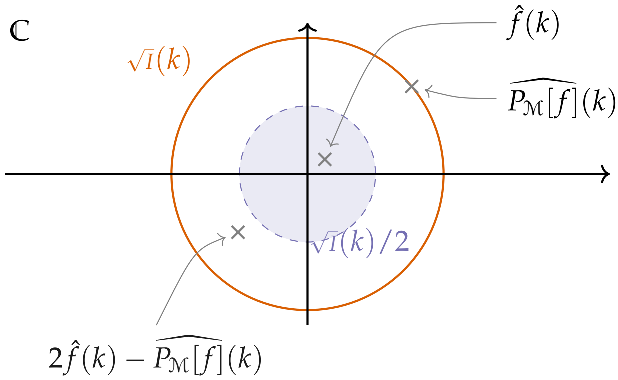

Figure 8.2: Illustration to Corollary 8.2 (ER fixed point correspondence). Let f be a fixed point of ERF. If even for a single pixel $k \in \mathbb{Z}^d$ holds $\hat{f}(k) < \sqrt{I}(k)/2$ — meaning that $\hat{f}(k)$ is inside the violet filled ball — then $\mathcal{F}\left(P_{\mathcal{M}}\left[2f - P_{\mathcal{M}}[f]\right]\right)(k) \neq \widehat{P_{\mathcal{M}}[f]}(k)$, and $P_{\mathcal{A}}[f]$ is not a fixed point of ER.
For $P_{\mathcal{A}}[f]$ to be a fixed point of ER, $\hat{f}(k)$ must be outside the violet filled ball for all k, or belong to the boundary of the ball provided that the corresponding condition from Corollary 8.2 is satisifed.

at k as long as $\hat{f}(k) = 0$, and that

$$e^{i\varphi(k)} = \frac{\hat{f}(k)}{|\hat{f}(k)|}$$

as long as $\sqrt{I}(k) = 2|\hat{f}(k)|$.

Expressions in Equation (8.7) have the same phase if and only if

$$2|\hat{f}| < \sqrt{I}\mathbb{1}_{\{\hat{f}\neq 0\}};$$

expressions in Equation (8.8) are equal if and only if $\hat{f} = 0$ implies $\sqrt{I} = 0$. In other words, (8.3) holds for $P_{\mathcal{M}}$ if and only if

$$\left\| \frac{\sqrt{I}}{|\hat{f}|} \mathbb{1}_{\operatorname{supp}\sqrt{I}} \right\|_{\infty} \leqslant 2,$$

and $\sqrt{I}(k) = 2|\hat{f}(k)|$ for $k \in \operatorname{supp}\sqrt{I}$ implies $e^{i\varphi(k)} = \frac{\hat{f}(k)}{|\hat{f}(k)|}$.

With this result, the corollary follows from Proposition 8.1. □

8.2 FIXED POINT STABILITY OF ERF

This section presents certain results on stability of fixed points. Notably, Proposition 8.6 shows a criterion for instability of fixed points, and Corollary 8.10 shows a criterion for stability along even directions. These results are illustrated in Figure 8.3. Particularly Corollary 8.10 can be of practical significance, as it states that fixed points of ER on the subspace of even functions are likely to be stable. The study of phase retrieval in the artificially restrictive setting of problems on the even subspace of $\mathcal{H}(\Omega)$ may yield new insights on the dynamic of phase retrieval algorithms.

For any translation-invariant additional constraint $\mathcal{A}$ (such as $\mathcal{P}$ or $\mathcal{T}_s(\nu)$) and any fixed ERF point f, the translation $f(x_0 + \cdot)$ is also a fixed point of ERF with exactly the same energy. It is therefore not possible — unless $\mathcal{A}$ is strengthened, or the metric distance of L^2 is modified — for a fixed point f to be attractive. Rather, it makes sense to study fixed points that are unstable or not unstable. To enhance readability, we call "not unstable" fixed points "stable"; with this convention, *stable fixed points are not attractive.*

DEFINITION 8.3 (STABLE AND UNSTABLE FIXED POINTS). *Let $\mathcal{A} \subset \mathcal{H}$ be weakly closed.*

i) Let $g \in \mathcal{H}$ be a fixed point of ER, i. e. let g satisfy $g = P_{\mathcal{A}} \circ P_{\mathcal{M}}[g]$. The point g is called unstable, if for all $\varepsilon_* > 0$ there exists $\varepsilon \in (0, \varepsilon_*]$ and direction $h \in \mathcal{H}$ with $\|h\|_2 = 1$, such that $g + \varepsilon h \in \mathcal{A}$, and $E_{\mathcal{M}}[g + \varepsilon h] < E_{\mathcal{M}}[g]$.

ii) Let $f \in \mathcal{H}$ be a fixed point of ERF, i. e. let f satisfy $2f = P_{\mathcal{A}}[f] + P_{\mathcal{M}}[f]$. The point f is called unstable, if for all $\varepsilon_* > 0$ there exists $\varepsilon \in (0, \varepsilon_*]$ and direction $h \in \mathcal{H}$ with $\|h\|_2 = 1$, such that

$$E_{\mathcal{M}}[f + \varepsilon h] + E_{\mathcal{A}}[f + \varepsilon h] < E_{\mathcal{M}}[f] + E_{\mathcal{A}}[f].$$

A fixed point is called stable, if it is not unstable.

One can now investigate stability properties of ER-GF fixed points using the following property of the modulus projection.

LEMMA 8.4 (MODULUS ENERGY EXPANSION). *Let $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$, let $\mathcal{A} = \mathcal{P} \in \mathcal{H}(\Omega)$, let $\sqrt{I} \in \widehat{\mathcal{H}}(\Omega_F)$ be non-negative. Further, let $f, h \in \mathcal{H}$ such that $\left\|\frac{\sqrt{I}}{|\hat{f}|}\right\|_\infty =: C_{\sqrt{I}/|\hat{f}|} < \infty$ and $\left\|\frac{|\hat{h}|}{\sqrt{I}}\right\|_\infty =: C_{|\hat{h}|/\sqrt{I}} < \infty$. Then, for all $\varepsilon \in (0, C_{\sqrt{I}/|\hat{f}|}/C_{|\hat{h}|/\sqrt{I}})$ there exists a function $\tilde{\varepsilon}\colon \mathcal{H} \to [0; \varepsilon]$ such that*

$$\begin{aligned} &E_{\mathcal{M}}[f + \varepsilon h] - E_{\mathcal{M}}[f] - \int (f - P_{\mathcal{M}}[f])h = \\ &\quad = \frac{\varepsilon^2}{2(2\pi)^d} \int \left(|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{f} + \tilde{\varepsilon}\hat{h}|} |\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{f} + \tilde{\varepsilon}\hat{h}|} \operatorname{Im}\left(\frac{(\hat{f} + \tilde{\varepsilon}\hat{h})^*}{|\hat{f} + \tilde{\varepsilon}\hat{h}|} \hat{h} \right)^2 \right), \end{aligned} \tag{8.9}$$

and the integral on the right-hand side exists and is finite. (The arguments of $\hat{f}, \hat{h}, \tilde{\varepsilon}, \sqrt{I}$ in the integral are omitted for readability.)

Proof. For a. a. $k \in \Omega_F$, have

$$|\hat{f}(k) + \tilde{\varepsilon}(k)\hat{h}(k)| \geqslant |\hat{f}(k)| - \varepsilon|\hat{h}(k)| \geqslant C_{\sqrt{I}/|\hat{f}|}\sqrt{I}(k) - \varepsilon C_{|\hat{h}|/\sqrt{I}}\sqrt{I}(k) \geqslant 0.$$

Therefore, the Taylor expansion from the proof of Lemma 4.14, Step 1d) on p. 62, holds for almost all $k \in \Omega_F$. The result of Lemma 4.14 justifies that integral over the Taylor expansion is well-defined, and — using equations (4.8), (4.9) from the proof of Lemma 4.14 — integral over the Taylor expansion can be transformed to the desired form. □

Remark 8.5. Condition $\left\|\frac{|\hat{h}|}{\sqrt{I}}\right\|_\infty$ is very restrictive when considered on the domain $\Omega_F \in \{\mathbb{R}^d, \mathbb{Z}^d\}$. It is less restrictive on the finite domain $\Omega_F = \mathbb{T}_N^d$: if $\operatorname{supp}\sqrt{I} = \mathbb{T}_N^d$ and $\hat{g}(k) \neq 0$ for all $k \in \mathbb{T}_N^d$, then assumptions of Lemma 8.4 are satisfied for all h, but the resulting upper bound on ε can become extremely small.

Therefore, Lemma 8.4 is much more useful to establish unstability of a fixed point rather than its stability: fore some fixed points, the lemma allows to find a direction h that satisfies $\left\|\frac{|\hat{h}|}{\sqrt{I}}\right\|_\infty < \infty$ and diminishes energy.

The next proposition shows that — under relatively mild assumptions — fixed points of (ERF) that do not correspond to fixed points of (ER) are unstable.

PROPOSITION 8.6 (A CONDITION FOR UNSTABLE FIXED POINTS). *Let $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}_N^d\}$, let $\mathcal{A} = \mathcal{P} \in \mathcal{H}(\Omega)$, let $\sqrt{I} \in \widehat{\mathcal{H}}(\Omega_F)$ be non-negative. Assume that f is a fixed point of ERF and that $2 < \left\|\frac{\sqrt{I}}{|\hat{f}|}\mathbb{1}_{\operatorname{supp}\sqrt{I}}\right\|_\infty < \infty$. Further, assume that $\left\|\frac{|\hat{f}|}{\sqrt{I}}\mathbb{1}_{\operatorname{supp}\sqrt{I}}\right\|_\infty < \infty$.*

Then, f is unstable.

Proof. Let us show that for all $\varepsilon > 0$ one can construct $h \in \mathcal{H}$ such that $E[f] - E[f + \varepsilon h] < 0$.

The goal is to choose a function h such that

$$\operatorname{Im}\big((\hat{f} + \varepsilon\hat{h})^*\hat{h}\big) = \operatorname{Im}\big(\hat{f}^*\hat{h}\big) = |\hat{f}|\,|\hat{h}|, \tag{8.10}$$

(the first equality being trivially true for all $h \in \mathcal{H}$), such that

$$\mathbb{1}_{\operatorname{supp}\hat{h}}\frac{\sqrt{I}}{|\hat{f} + \varepsilon\hat{h}|} > 2,$$

and such that $\left\|\frac{|\hat{h}|}{\sqrt{I}}\mathbb{1}_{\operatorname{supp}\sqrt{I}}\right\|_\infty < \infty$.

Then, one can use Equation (8.9) to follow

$$E[f+\varepsilon h]-E[f] \leqslant \underbrace{\frac{\varepsilon^2}{2(2\pi)^d}\int\left(1-\frac{\sqrt{I}}{|\hat{f}+\tilde{\varepsilon}\hat{h}|}\right)|\hat{h}|^2}_{\substack{\text{Exact equality from the}\\ \text{expansion of } E_{\mathcal{M}}}} + \underbrace{\frac{\varepsilon^2}{2(2\pi)^d}\|\hat{h}\|_2^2}_{\substack{\text{Worst-case}\\ \text{estimation from the}\\ \text{expansion of } E_{\mathcal{A}}}}$$

$$\overset{(*)}{=} \frac{\varepsilon^2}{2(2\pi)^d}\int\left(1-\frac{\sqrt{I}}{\sqrt{|\hat{f}|^2+\tilde{\varepsilon}^2|\hat{h}|^2}}\right)|\hat{h}|^2+\frac{\varepsilon^2}{2(2\pi)^d}\|\hat{h}\|_2^2$$

$$\leqslant \frac{\varepsilon^2}{2(2\pi)^d}\int\left(1-\frac{\sqrt{I}}{\sqrt{|\hat{f}|^2+\varepsilon^2|\hat{h}|^2}}\right)|\hat{h}|^2+\frac{\varepsilon^2}{2(2\pi)^d}\|\hat{h}\|_2^2$$

$$= \frac{\varepsilon^2}{2(2\pi)^d}\int\left(2-\frac{\sqrt{I}}{|\hat{f}+\varepsilon\hat{h}|}\right)|\hat{h}|^2 < 0,$$

and in $(*)$ we have used the construction requirement (8.10).

The desired h may be chosen as follows. Since $\left\|\frac{\sqrt{I}}{|\hat{f}|}\mathbb{1}_{\operatorname{supp}\sqrt{I}}\right\|_\infty > 2$, there exists an $\alpha > 0$ such that the set

$$S_> = \left\{k \in \Omega_F \text{ for which } \frac{\sqrt{I}(k)}{|\hat{f}(k)|}\mathbb{1}_{\operatorname{supp}\sqrt{I}} > 2+\alpha\right\}$$

has Lebesgue measure larger than 0.

Note that $0 \notin S_>$ since $\hat{f}(0) \geqslant \sqrt{I}(0)$ for fixed point f: indeed,

$$2f = P_{\mathcal{P}}[f]+P_{\mathcal{M}}[f] \quad \Rightarrow \quad 2\hat{f}(0) = \widehat{P_{\mathcal{P}}[f]}(0)+\widehat{P_{\mathcal{M}}[f]}(0) \quad \Rightarrow f(0) \geqslant \widehat{P_{\mathcal{M}}[f]}(0),$$

since $\widehat{P_{\mathcal{P}}[f]}(0) \geqslant \hat{f}(0)$ due to the definition of the positivity operator. Let $\xi\colon \Omega_F \to \{0, \pm i\}$ be an odd measurable function that equals 0 only at the point $0 \in \Omega_F$.

Define

$$h\colon \mathcal{H}(\Omega) \to \mathbb{R};$$
$$h = \mathcal{F}^{-1}\left(\frac{\sqrt{\alpha}}{\sqrt{2}\varepsilon}\hat{f}\xi\mathbb{1}_{S_>}\right).$$

Note that h is indeed real-valued as the Fourier inverse of a function with even modulus and odd phase. Further,

$$\left\|\frac{|\hat{h}|}{\sqrt{I}}\mathbb{1}_{\operatorname{supp}\sqrt{I}}\right\|_\infty \leqslant \left\|\frac{\sqrt{\alpha}}{\sqrt{2}\varepsilon}\frac{|\hat{f}|}{\sqrt{I}}\mathbb{1}_{\operatorname{supp}\sqrt{I}}\right\|_\infty \leqslant \frac{\sqrt{\alpha}}{\sqrt{2}\varepsilon} < \infty.$$

Also, by construction, $S_>$ is not empty and $|\mathbb{1}_{S_>}\xi| = |\mathbb{1}_{S_>}|$; therefore, h is not identically equal zero. Obviously, $\hat{f}$ and $\hat{h}$ are pointwise perpendicular due to the factor ξ, so that (8.10) is true. Further, for all $k \in \operatorname{supp} \sqrt{I}$

$$\mathbb{1}_{\operatorname{supp}\hat{h}}(k)\frac{\sqrt{I}(k)}{|\hat{f}(k)+\varepsilon\hat{h}(k)|} = \mathbb{1}_{\operatorname{supp}\hat{h}}(k)\frac{\sqrt{I}(k)}{\sqrt{|\hat{f}(k)|^2+\varepsilon^2|\hat{h}(k)|^2}}$$
$$= \mathbb{1}_{\operatorname{supp}\hat{h}}(k)\frac{\sqrt{I}(k)}{\sqrt{(1+\varepsilon^2\frac{\alpha}{2\varepsilon^2})|\hat{f}(k)|^2}} = \mathbb{1}_{S_>}(k)\frac{\sqrt{I}(k)}{(1+\frac{\alpha}{2})|\hat{f}(k)|} > \frac{2+\alpha}{1+\frac{\alpha}{2}} = 2,$$

justifying the claim. □

Remark 8.7 (Generalization to other $\mathcal{A}$). It is straightforward to verify that Proposition 8.6 holds for all weakly closed additional constraints $\mathcal{A}$, as long as

$$\left| E_{\mathcal{A}}[g+\varepsilon h] - E_{\mathcal{A}}[g] - \varepsilon \int \nabla E_{\mathcal{A}}[g]h \right| \leqslant \frac{\varepsilon^2}{2}\|h\|_2^2.$$

In particular, support constraint $\mathcal{S}(S)$ satisifies this condition, since $P_{\mathcal{S}}$ is linear and

$$E_{\mathcal{S}}[g+\varepsilon h] = \frac{1}{2}\|g+\varepsilon h - P_{\mathcal{S}}[g] - \varepsilon P_{\mathcal{S}}[h]\|_2^2 = E_{\mathcal{S}}[g] - \varepsilon \int (g - P_{\mathcal{S}}[g])h + \frac{\varepsilon^2}{2}\|P_{\mathcal{S}}[h]\|_2^2,$$

and $\|P_{\mathcal{S}}[h]\|_2 \leqslant \|h\|_2$ for the indicator projection $P_{\mathcal{S}}$. Analogously, this condition can be shown for $\mathcal{P} \cap \mathcal{S}$.

As is easy to check, this condition does not hold, for example, for sparsity constraints $\mathcal{T}_a(\alpha)$ and $\mathcal{T}_s(\nu)$.

Remark 8.8 (Informal interpretation of $P_{\mathcal{M}}$ expansion). Lemma 8.4 shows the conditions under which Taylor-expansion of $P_{\mathcal{M}}$ is applicable. Informally, behavior of $\mathcal{F}(P_{\mathcal{M}}[f+\varepsilon h])(k)$ can be qualitatively categorized as follows, see Figure 8.3.

Case I) Let $\varepsilon\hat{h}(k)$ is parallel or antiparallel to $\hat{f}(k)$, i.e.
$\operatorname{Re}(\varepsilon\hat{h}^*(k)\hat{f}(k)) \approx \varepsilon\hat{h}^*(k)\hat{f}(k)$.
Case Ia). If, additionally, $\hat{f}(k)+\varepsilon\hat{h}(k)$ is parallel to $\hat{f}(k)$, the phase is not changed, meaning that

$$\mathcal{F}(P_{\mathcal{M}}[f+\varepsilon h])(k) \approx \mathcal{F}(P_{\mathcal{M}}[f])(k).$$

For such arguments f, h and points $k \in \Omega_F$, the operator $P_{\mathcal{M}}$ is correctly described by Taylor-expansion and well-behaved (contractive to the point $P_{\mathcal{M}}[g]$ undependent on h).
Case Ib). If $\hat{f}(k)+\varepsilon\hat{h}(k)$ is antiparallel to $\hat{f}(k)$, the phase is flipped, meaning that

$$\mathcal{F}(P_{\mathcal{M}}[f+\varepsilon h])(k) \approx -\mathcal{F}(P_{\mathcal{M}}[f])(k).$$

For such arguments f, h and points $k \in \Omega_F$, the operator $P_{\mathcal{M}}$ is poorly approximated by the Taylor expansion.

Case II). If $\varepsilon\hat{h}(k)$ is perpendicular to $\hat{f}(k)$, i. e. $\mathrm{Im}(\varepsilon\hat{h}^*(k)\hat{f}(k)) \approx \varepsilon\hat{h}^*(k)\hat{f}(k)$, the behavior of $P_{\mathcal{M}}$ — the change in phase — can be efficiently captured by the Taylor expansion and relatively well-behaved (linearization in h approximates $P_{\mathcal{M}}[g + \varepsilon h]$ fairly well).
Case III) If $\hat{f}(k) + \varepsilon\hat{h}(k) \approx 0$, $\mathcal{F}\left(P_{\mathcal{M}}[f + \varepsilon h]\right)(k)$ is highly susceptible to noise, and the Taylor approximation is not well-defined.

Proposition 8.6 assumes that points of Case Ib) are not present; exploits condition (8.10) to avoid points of Case II); and shows that for remaining points of Case Ia), operator $P_{\mathcal{M}}$ is sufficiently well-behaved to achieve the desired estimate.

However, in general, points of Case Ib) also consitute an integral part of phase retrieval problems. The points of Case II) are not present for thef even-restricted ERF, which simplifies the analysis of $P_{\mathcal{M}}$ to a certain extent.

To the best of our knowledge, the task to find a description of $P_{\mathcal{M}}$ that allows a more detailed analysis of algorithm behavior on points of Case Ib) remains an open challenge — a crucial one for description of phase retrieval algorithms.

The following corollary applies Proposition 8.6 to the phase retrieval setting. It shows that the fixed points of ERF that do not correspond to fixed points of ER are unstable.

COROLLARY 8.9. *Let $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$, let $\mathcal{A} = \mathcal{P} \in \mathcal{H}(\Omega)$, let $\sqrt{I} \in \widehat{\mathcal{H}}(\Omega_F)$ be non-negative. Assume that f is a fixed point of ERF (i. e. $2f = P_{\mathcal{P}}[f] + P_{\mathcal{M}}[f]$) with $\left\|\frac{|\hat{h}|}{\sqrt{I}}\mathbb{1}_{\operatorname{supp}\sqrt{I}}\right\|_\infty < \infty$. Further, assume that for any measurable phase $\varphi\colon \Omega_F \to [0, 2\pi)$, $P_{\mathcal{P}} \circ P_{\mathcal{M};\varphi} \circ P_{\mathcal{P}}[f] \neq P_{\mathcal{P}}[f]$ (i. e. $P_{\mathcal{P}}[f]$ is not a fixed point of ER).*

Then, f is unstable.

Proof. Use Corollary 8.2 in the following manner. Since for any measurable phase $\varphi\colon \Omega_F \to [0, 2\pi)$, $P_{\mathcal{P}} \circ P_{\mathcal{M};\varphi} \circ P_{\mathcal{P}}[f] \neq P_{\mathcal{P}}[f]$, condition

$$\sqrt{I}(k) = 2|\hat{f}(k)| \text{ for } k \in \operatorname{supp}\sqrt{I} \text{ implies } e^{i\varphi(k)} = \frac{\hat{f}(k)}{|\hat{f}(k)|}$$

from Corollary 8.2 can not be satisfied, and $\left\|\frac{|\hat{h}|}{\sqrt{I}}\mathbb{1}_{\operatorname{supp}\sqrt{I}}\right\|_\infty > 2$.

Then, f is unstable by Proposition 8.6. □

Informally speaking, Lemma 8.4 indicates that an ERF fixed point f is more likely to be stable in the direction h, if $\mathrm{Im}(\hat{f} + \varepsilon\hat{h})^*\hat{h} = 0$. This is true in the particular case when f and h are even, since for such f, h have $\mathrm{Im}(\hat{f}) = \mathrm{Im}(\hat{f}) = 0$.

Note that the modulus projection and many additional projections (such as positivity and sparsity) preserve evenness. Thus,

the study ERF for even functions

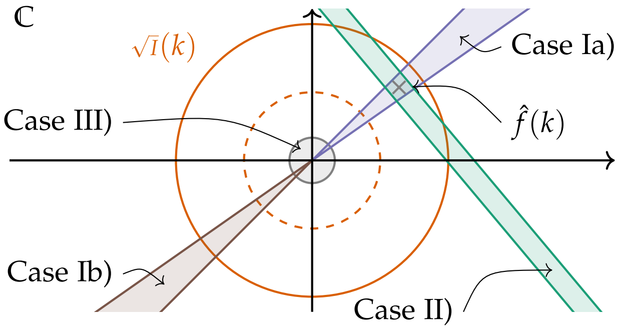

Figure 8.3: Illustration to Remark 8.8 (Interpretation of $\widehat{P_{\mathcal{M}}}$ at k). Dashed orange circle is of radius $\sqrt{I}(k)/2$.

Case Ia) If $\hat{f}(k) + \varepsilon\hat{h}(k)$ is approximately parallel to $\hat{f}(k)$, then $\widehat{P_{\mathcal{M}}}[f + \varepsilon\hat{h}](k) \approx \widehat{P_{\mathcal{M}}}[\hat{f}](k)$; "phase is not changed".

Case Ib) If $\hat{f}(k) + \varepsilon\hat{h}(k)$ is approximately antiparallel to $\hat{f}(k)$, then $\widehat{P_{\mathcal{M}}}[f + \varepsilon\hat{h}](k) \approx -\widehat{P_{\mathcal{M}}}[\hat{f}](k)$; "phase is flipped".

Case II) If $\hat{f}(k) + \varepsilon\hat{h}(k)$ is approximately perpendicular to $\hat{f}(k)$, "phase is continuously rotated".

Case III) if $\hat{f}(k) + \varepsilon\hat{h}(k) \approx 0$, "phase is discontinuous".

Proposition 8.6 assumes that regions of Case Ib) and Case III) are not present for suitable directions h and small enough ε. It uses condition (8.10) to construct h that avoids Case II). It shows that if there exist regions of Case Ia) with $|\hat{f}(k)| < \sqrt{I}(k)/2$, second order term in the expansion of $P_{\mathcal{M}}$ makes the fixed point f unstable along an appropriate direction h.
The regions of Case II) are not present for the even-restricted ERF.
Corollary 8.10 assumes that the regions of Case Ib) and Case III) are not present for suitable directions h and small enough ε. It excludes the regions of Case II) by restricting ERF to the even case. It shows that if $|\hat{f}(k)| \geqslant \sqrt{I}(k)/2$ for almost all $k \in \Omega_F$, then f is a stable fixed point.

relates to

the study of ERF behavior in absence of Case II) region from Remark 8.8.

The following proposition states that even fixed points are stable along all even directions if phase flips — Case Ib) from Remark 8.8 — are not allowed.

COROLLARY 8.10 (FIXED POINT STABILITY OF EVEN-RESTRICTED ERF AND ER). *Let* $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}^d_N\}$, *let* $\mathcal{A} = \mathcal{P} \in \mathcal{H}(\Omega)$, *let* $\sqrt{I} \in \widehat{\mathcal{H}}(\Omega_F)$ *be even.*

(i) If f is an even fixed point of ERF with $C_{\sqrt{I}/|\hat{f}|} := \left\|\frac{\sqrt{I}}{\hat{f}}\right\|_\infty < 2$, *then it is stable along all even directions $h \in \mathcal{H}$ if they satisfy* $C_{|\hat{h}|/|\hat{f}|} := \left\|\frac{\hat{h}}{\hat{f}}\right\|_\infty < \infty$. *More precisely, for such h holds*

$$E[f + \varepsilon h] - E[f] \geqslant 0$$

for all $\varepsilon \in \left(0, \frac{1}{2C_{|\hat{h}|/|\hat{f}|}(2 - C_{\sqrt{I}/|\hat{f}|})}\right) \cap (0, C_{\sqrt{I}/|\hat{f}|}/C_{|\hat{h}|/\sqrt{I}})$.

ii) Assume g is an even fixed point of ER with $C_{\sqrt{I}/|\hat{g}|} := \left\|\frac{\sqrt{I}}{\hat{g}}\right\|_\infty < \infty$. *Then, there exists an $\varepsilon > 0$ such that*

$$E_{\mathcal{M}}[g + h] - E_{\mathcal{M}}[g] \geqslant 0$$

for all directions $h \in \mathcal{H}$ that satisfy the following assumptions:

a) h is even and $\|h\| < \varepsilon$;

b) $C_{|\hat{h}|/\sqrt{I}} := \left\|\frac{\hat{h}}{\sqrt{I}}\right\|_\infty < \infty$ *or* $C_{|\hat{h}|/|\hat{g}|} := \left\|\frac{\hat{h}}{\hat{g}}\right\|_\infty < \infty$ *is true;*

c) $g + h \in \mathcal{P}$.

Proof. i) By Lemma 4.14 ii),

$$E_{\mathcal{P}}[f] - E_{\mathcal{P}}[h] - \int (f - P_{\mathcal{P}}[f])h \geqslant 0 \tag{8.11}$$

for all h in $\mathcal{H}$. Further, by Lemma 8.4, — and since f and h are even and real-valued, and so are $\hat{f}$ and $\hat{h}$ — have

$$E_{\mathcal{M}}[f] - E_{\mathcal{M}}[h] - \int (f - P_{\mathcal{M}}[f])h = \frac{\varepsilon^2}{2(2\pi)^d} \int \left(|\hat{h}|^2 - \frac{\sqrt{I}}{2|\hat{f} + \tilde{\varepsilon}\hat{h}|} |\hat{h}|^2 \right) \tag{8.12}$$

for all $\varepsilon \in (0, C_{\sqrt{I}/|\hat{f}|}/C_{|\hat{h}|/\sqrt{I}})$. By the definition of $C_{\sqrt{I}/|\hat{f}|}$ and $C_{|\hat{h}|/|\hat{f}|}$, for a. a. $k \in \Omega_F$ have

$$\sqrt{I}(k) \leqslant C_{\sqrt{I}/|\hat{f}|} |\hat{f}(k)|,$$
$$|\hat{h}(k)| \leqslant C_{|\hat{h}|/|\hat{f}|} |\hat{f}(k)|.$$

Therefore, for a. a. $k \in \Omega_F$, for $\varepsilon \in \left(0, \frac{2-C_{\sqrt{I},\hat{f}}}{2C_{\hat{h},\hat{f}}}\right)$ holds

$$\frac{\sqrt{I}}{|\hat{f}+\varepsilon\hat{h}|} \leqslant \frac{\sqrt{I}}{|\hat{f}|-\varepsilon|\hat{h}|} \overset{(*)}{\leqslant} \frac{\sqrt{I}}{|\hat{f}|-\varepsilon C_{\hat{h},\hat{f}}|\hat{f}|}$$
$$\leqslant \frac{\sqrt{I}}{|\hat{f}|(1-\varepsilon C_{\hat{h},\hat{f}})|} \leqslant \frac{C_{\sqrt{I},\hat{f}}}{1-\varepsilon C_{\hat{h},\hat{f}}} \leqslant 2, \tag{8.13}$$

where k was omitted for readability, and inequality $(*)$ holds since the assumption $\varepsilon < \frac{2-C_{\sqrt{I},\hat{f}}}{2C_{\hat{h},\hat{f}}}$ implies $1-\varepsilon C_{\hat{h},\hat{f}} > 0$. The result follows by inserting inequality (8.13) into (8.12) and combining it with (8.11).

ii) To show the statement, assume the contrary and construct a contradiction to Case i). That is, assume there exists an even fixed point of ER $g \in \mathcal{H}$ with $C_{\sqrt{I}/|\hat{g}|} < \infty$, such that for any arbitrarily small $\varepsilon > 0$ there exists an even direction $h \in \mathcal{H}$ with $\|h\|_2 < \varepsilon$, with $C_{|\hat{h}|/|\hat{g}|} < \infty$, with $g+h \in \mathcal{P}$, and with

$$E[g+h] - E[g] < 0.$$

Let $\tilde{g} := g+h$, let $f := \frac{1}{2}(g + P_{\mathcal{M}}[g])$. We shall show that for all $\varepsilon_f > 0$ there exists an even $\tilde{f} := \frac{1}{2}(\tilde{g} + P_{\mathcal{M}}[\tilde{g}])$ such that $\|\tilde{f} - f\|_2 < \varepsilon_f$,

$$E[\tilde{f}] - E[f] < 0, \tag{8.14}$$

which will be a contradiction to i). Let $h_f := \tilde{f} - f$.

Let us verify that all assumptions for i) are satisfied for f defined above. By Corollary 8.2, f is a fixed point of (ERF). It satisfies

$$\left\|\frac{\sqrt{I}}{\hat{f}}\right\|_\infty = \left\|\frac{2\sqrt{I}}{|\hat{g}|+\sqrt{I}}\right\|_\infty < 2,$$

since

$$\sup_{k\in\Omega_F} \frac{\sqrt{I}(k)}{|\hat{g}(k)|} < \infty \Rightarrow \inf_{k\in\Omega_F} \frac{|\hat{g}(k)|}{\sqrt{I}(k)} > 0.$$

As for the final assumption of Case i), is true since

$$\left\|\frac{\hat{h}_f}{\hat{f}}\right\|_\infty = \left\|\frac{\hat{g}+\hat{h}+\mathcal{F}(P_{\mathcal{M}}[g+h]) - \hat{g} - \widehat{P_{\mathcal{M}}[g]}}{|\hat{g}|+\sqrt{I}}\right\|_\infty$$
$$\leqslant \left\|\frac{\hat{h}}{|\hat{g}|+\sqrt{I}}\right\|_\infty + \left\|\frac{2\sqrt{I}}{|\hat{g}|+\sqrt{I}}\right\|_\infty \leqslant \min\{C_{|\hat{h}|/|\hat{g}|}, C_{|\hat{h}|/\sqrt{I}}\} + 2 < \infty.$$

All assumptions of i) are satisfied, and we need to show Equation (8.14) to establish a contradiction. By the definition of f and $\tilde{f}$ and by Lemma 3.8,

$$P_{\mathcal{P}}[f] = P_{\mathcal{P}}[g], \qquad P_{\mathcal{M}}[f] = P_{\mathcal{M}}[g], \qquad P_{\mathcal{M}}[\tilde{f}] = P_{\mathcal{M}}[\tilde{g}].$$

Note that $P_{\mathcal{P}}[\tilde{f}] = P_{\mathcal{P}}[\tilde{g}]$ is not necessarily true, since $\tilde{g}$ is not necessarily a fixed point of ER, and $\tilde{f}$ is not necessarily a fixed point of ERF. By a straightforward calculation, $E[f] = \frac{1}{2}E[g]$. Further,

$$\begin{aligned} E[\tilde{f}] &= \frac{1}{2}\|\tilde{f} - P_{\mathcal{P}}[\tilde{f}]\|_2^2 + \frac{1}{2}\|\tilde{f} - P_{\mathcal{M}}[\tilde{f}]\|_2^2 \\ &\overset{(*)}{\leqslant} \frac{1}{2}\|\tilde{f} - \tilde{g}\|_2^2 + \frac{1}{2}\|\tilde{f} - P_{\mathcal{M}}[g]\|_2^2 \\ &= \frac{1}{8}\|\tilde{g} - P_{\mathcal{M}}[\tilde{g}]\|_2^2 + \frac{1}{8}\|\tilde{g} - P_{\mathcal{M}}[\tilde{g}]\|_2^2 \\ &= \frac{1}{2}E_{\mathcal{M}}[\tilde{g}] < \frac{1}{2}E_{\mathcal{M}}[g] = E[f], \end{aligned}$$

where we used $\tilde{g} \in \mathcal{P}$ and Corollary 3.6 in $(*)$. Therefore,

$$E[\tilde{f}] - E[f] < 0.$$

Lastly, we need to show that chosen $\tilde{f}$ satisfies $\|\tilde{f} - f\| < \varepsilon_f$. Since $P_{\mathcal{M}}$ is Lipshitz at g with constant $2C_{\sqrt{1}/|\hat{g}|}$ by Lemma 6.22, one has

$$\|f - \tilde{f}\|_2 \leqslant \frac{1}{2}\|g - \tilde{g}\|_2 + C_{\sqrt{1},\hat{g}}\|g - \tilde{g}\|_2 \leqslant \left(\frac{1}{2} + C_{\sqrt{1},\hat{g}}\right)\varepsilon < \varepsilon_f,$$

as long as $\varepsilon < \frac{\varepsilon_f}{\frac{1}{2}+C_{\sqrt{1},\hat{g}}}$. This choice is possible, since ε can be chosen arbitrarily ($\tilde{g}$ can be chosen arbitrarily close to g). Thus, we have a contradiction to Case i), concluding the proof. □

Remark 8.11. On a finite space (i. e. for $\Omega = \mathbb{T}_N^d$), Corollary 8.10 holds for all even directions h at the point g provided that $\hat{g}(k) \neq 0$ at all coordinates k. Thus, the restriction to even case provides an experimental tool for studying fixed points of E. In simulations we observed that even-restricted ER typically takes about order of 50 steps to converge to a numerically stable fixed point (stable up to machine precision errors).

8.3 FORMAL HESSIAN OF ERF ENERGY

This section writes down the formal Hessian of the energy $E = E_{\mathcal{M}} + E_{\mathcal{P}}$; it is used in Chapter 10 to investigate stability of fixed points of ERF.

This section does not contain original results, as these types of Hessians are often studied in phase retrieval literature (for example, prominent paper [CLS15] studies the Hessian of the similar functional $g \mapsto \int \left(|\hat{g}|^2 - \sqrt{I}^2\right)^2$, which is more regular but similar to $E_{\mathcal{M}}$).

As shown in Remark D.8, the Hessian of E at g is formally equal to

$$H(x,y) = \mathbb{1}_{g<0}(x)\delta(x-y) + \\ 1 - \mathcal{F}^{-1}\left(\frac{\sqrt{I}}{2|\hat{g}|}\right)(x-y) + \mathcal{F}^{-1}\left(\frac{\sqrt{I}}{2|\hat{g}|}\frac{(\hat{g}^*)^2}{|\hat{g}|^2}\right)(x+y)$$

in object space, or to

$$H_F(k,q) = \widehat{\mathbb{1}}_{g<0}(k-q) \\ + \frac{1}{(2\pi)^d}\left(1 - \frac{\sqrt{I}(k)}{2|\hat{g}(k)|}\right)\delta(k-q) + \frac{1}{(2\pi)^d}\frac{\sqrt{I}(k)}{2|\hat{g}(k)|}\frac{(\hat{g}(k)^*)^2}{|\hat{g}(k)|^2}\delta(k+q)$$

in Fourier space. It is not mathematically rigorous unless considered on a finite space at points g satisfying $g(x) \neq 0$ for all $x \in \mathbb{T}_N^d$, and $|\hat{g}(k)| \neq 0$ for all $k \in \operatorname{supp} \sqrt{I}$. Fortunately, this condition seems to be almost always satisfied for numerical simulations, justifying a more detailed look on $H(x,y)$.

Of interest are the smallest eigenvalues of H at a fixed point g: if any of them are negative, the fixed point g is unstable.

One can reliably generate even fixed points of ER and ERF by restricting them to even subspaces; this is motivated by Corollary 8.10. In practice, such fixed points are usually unstable and can be found both far from and close to solutions. Theoretical convergence radius can not exceed the distance between solutions and closest fixed points that are distinct from solutions. This is why existence of fixed points near solutions — cf. bottom row of Figure 10.7 — indicates how small the convergence radius can be for phase retrieval. For the unstable fixed points generated using the evenness restriction, the algorithm typically leaves the fixed point by accumulating odd disturbances after the evenness restriction is lifted.

As for the non-even fixed points of ER, it is unclear how to generate them in general. For certain problems, if one runs ER for long enough, the algorithm arrives at a stable fixed point, see Figure 10.10. Existence of such fixed points is a key for understanding convergence properties of phase retrieval.

Part III

BEYOND THE ERROR-REDUCTION FLOW

This part introduces a generalization of the Error-Reduction Flow, and provides some numerical examples.

Namely, Chapter 9 introduces a system of equations that generalizes the Error-Reduction Flow and can be used to analyze the Douglas-Rachford algorithm. This system of equations can be derived as a variation of a functional. The functional reveals a connection between the Error-Reduction, Dykstra, and Douglas-Rachford algorithms. The connection appears to be new, and is valid for a general setting of two-set feasibility problems. In phase retrieval setting, this system of equations is shown to posess global weak solutions.

Chapter 10 provides numerical examples that illustrate behaviour of classical ER and DR algorithms, and of the corresponding discretized flows introduced in this work. The examples demonstrate that classical algorithms and corresponding disretized flows have similar dynamics. Further, the examples demonstrate that for certain problems, large local convergence radius of ER may be heuristically explained by the fact that many of its fixed points are unstable.

The thesis is concluded in Chapter 11, which contains an outlook describing some open questions connected to this work.

9 AP, DYKSTRA AND DR VIA TWO-VARIABLES FLOW

This chapter introduces a system of equations that generalizes the Error-Reduction Flow and is better suited to describe the Douglas-Rachford algorithm. The main results of this chapter are presented in Table 9.1 and illustrated in Figures 9.2 to 9.5.

Namely, as demonstrated in Chapter 5, the AP and DR algorithms can be interpreted to be discretizations of the same APF

$$\partial_t g = -(g - P_{\mathcal{X}}[g]) - (g - P_{\mathcal{Y}}[g]).$$

We used explicit discretization to obtain AP from APF, and DR-LM to obtain DR from APF.

AP and DR can be connected to other equations through other discretization procedures. Such connections can yield new insights, if the corresponding equations posess interesting properties.

For example, a remarkable property of APF is that it is a (formal) gradient flow, and the corresponding energy dissipation transfers to AP (Chapter 6).

The goal of this chapter is to present a system of equations that is connected to the AP, DR and Dykstra algorithms. (The latter is mentioned in Section 5.1.2 in connection to BIO and is formally introduced in Section 6.3.4.1.) Namely, we demonstrate that AP, Dykstra and DR correspond to the following system of equations:

$$\partial_t \begin{pmatrix} s \\ d \end{pmatrix} = M \cdot \begin{pmatrix} \frac{\delta}{\delta s} \\ \frac{\delta}{\delta d} \end{pmatrix} F[s, d], \qquad \text{(2v-FPF)}$$

where $s, d \in L^2(\mathbb{R}^d)$ are two variables that describe the approximation (their interpretation is discussed below in more detail). The functional

$$F[s, d] := \frac{1}{2} E_{\mathcal{X}}[s + d] + \frac{1}{2} E_{\mathcal{Y}}[s - d] - \frac{1}{2} \|d\|_2^2$$

is the same for AP, Dykstra and DR. The matrix $M \in \mathbb{R}^{2x2}$ must be chosen dependent on the algorithm:

$$M = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \text{ for AP}, \qquad M = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ for Dykstra},$$

$$\text{and } M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for DR}.$$

We call the equation (2v-FPF) the two-variables Feasibility Problem Flow, or the two-variables flow (2v-FPF). It is noteworthy that 2v-FPF has a gradient structure in the sense of [LM13].

The differentiation on the right-hand side of Equation (2v-FPF) is a formal Fréchet-differentiation in variables s, d, or a selection of the rigorous generalized (Mordukhovich-Kruger) subdifferential in s, d, provided $\mathcal{X}$ and $\mathcal{Y}$ are weakly closed (similarly to Section 5.2). For ease of readability, we use formal Fréchet-differentiation of functionals throughout the chapter, but reformulation to selections of subdifferentials is possible for all presented results, provided $\mathcal{X}$ and $\mathcal{Y}$ are weakly closed.

This chapter proceeds as follows. First, we provide motivation for 2v-FPF and interpret its terms. Second, we establish its connection to AP, Dykstra, and DR algorithms (summarized in Table 9.1). Third, we describe the dynamics of 2v-FPF using examples where $\mathcal{X}$ and $\mathcal{Y}$ are particularly simple (unions of two-dimensional balls). We conclude the chapter by discussing certain open problems connected to the 2v-FPF DR algorithm.

9.1 MOTIVATION OF 2V-FPF

This section argues why of many possible continuous formulations of the Douglas-Rachford algorithm we favor 2v-FPF.

For two proximal sets $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$, the feasibility problem — the task of finding an element $g \in \mathcal{X} \cap \mathcal{Y}$ — can be adressed by various algorithms, such as AP, Dykstra, and DR. Our search of equations — that describe behaviours of these algorithms — was influenced by two following principles: i) avoid composition of operators; ii) preserve symmetry.

First, the reason to avoid composition of operators — say, $P_\mathcal{X} \circ P_\mathcal{Y}$ — is the following. Terms like $P_\mathcal{X} \circ P_\mathcal{Y}$ rarely arise as derivatives of functionals (cf. Remark 5.19). Therefore, it is unlikely that equations containing such terms provide viable alternative descriptions of the algorithms one strives to understand.

Second, the discussed feasibility problem is — at least, nominally — symmetric with respect to exchange of $\mathcal{X}$ and $\mathcal{Y}$. This symmetry is broken by the algorithms: neither the AP update $g_{n+1} = P_\mathcal{X} \circ P_\mathcal{Y}[g_n]$ nor the DR update $g_{n+1} = 0.5(g_n + R_\mathcal{X} \circ R_\mathcal{Y}[g_n])$ are symmetric under the exchange of $\mathcal{X}$ and $\mathcal{Y}$. This symmetry breaking can be analyzed and exploited, see [Mou16], but it is not an intrinsic property of the feasibility problem.

APF maintains this symmetry: $\partial_t g = -2g + P_\mathcal{X}[g] + P_\mathcal{Y}[g]$ is symmetric in $\mathcal{X}$ and $\mathcal{Y}$. Note how g appears as an argument of both $P_\mathcal{X}$ and $P_\mathcal{Y}$. While APF is connected to AP and DR (Section 5.3), its dynamics — at least in phase retrieval — is closer to AP (stagnates, dissipates energy) and differs significantly from DR (globally convergent in simulations, escapes local minima).

Our goal was to find equations whose behavior is comparable to the behavior of DR, while avoiding operator compositions and maintaining $\mathcal{X}$–$\mathcal{Y}$-symmetry. The connection to AP and Dykstra was obtained as a by-product of this search.

An obvious way to generalize APF is to consider two independent variables $p, q \in \mathcal{H}$ instead of g; each variable corresponding to $\mathcal{X}$ or to $\mathcal{Y}$.

As shown below, AP, Dykstra and DR can all be written in a similar fashion, using exclusively sums and differences of the following terms: p, q, $P_{\mathcal{X}}[p]$, and $P_{\mathcal{Y}}[q]$. Note how p appears as an argument only of $P_{\mathcal{X}}$, and q appears as an argument only of $P_{\mathcal{Y}}$. The resulting equations, shown below, are symmetric, if $\mathcal{X}$ is exchanged with $\mathcal{Y}$ and p is exchanged with q. The variables $s = \frac{p+q}{2}$ and $d = \frac{p-q}{2}$ — used in the beginning of this chapter — correspond to center-of-mass coordinates for variables p and q.

9.2 CONNECTION BETWEEN 2V-FPF AND AP, DYKSTRA AND DR

This section describes how AP, Dykstra and DR can be recovered from 2v-FPF using explicit discretization.

Canonical formulations of AP, Dykstra, and DR can be found in the first row of Table 9.1. Their reformulations in terms of variables p, q can be found in the second row of Table 9.1.

PROPOSITION 9.1. *Reformulations of AP, Dykstra, and DR (second row of Table 9.1) are equivalent to canonical formulations of AP, Dykstra, and DR, in the sense specified in the second row of Table 9.1.*

Proof. All cases (AP, Dykstra, and DR) are shown by induction.

AP. Let $g_0 \in \mathcal{X}$, let the sequences (g_n), (p_n), and (q_n) be generated as described in Table 9.1. We need to show that

$$q_n = g_n = P_{\mathcal{X}}[p_n] \text{ and } p_n = P_{\mathcal{Y}}[g_{n-1}] = P_{\mathcal{Y}}[q_{n-1}]$$

for all $n \in \mathbb{N}$. By construction, $q_1 = g_1 = P_{\mathcal{Y}}[p_1]$, $p_1 = P_{\mathcal{Y}}[g_0] = P_{\mathcal{Y}}[q_0]$. Induction step: if the assumption is true for an $n \in \mathbb{N}$,

$$\begin{aligned} p_{n+1} = p_n + 2\left(-\frac{p_n \quad + q_n}{2} + \frac{P_{\mathcal{X}}[p_n \quad] + P_{\mathcal{Y}}[q_n]}{2}\right) \\ = -q_n + P_{\mathcal{X}}[p_n] + P_{\mathcal{Y}}[q_n] = P_{\mathcal{Y}}[q_n] = P_{\mathcal{Y}}[g_n], \end{aligned}$$

since $P_{\mathcal{X}}[p_n] = P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g_{n-1}] = g_n = q_n$. Further,

$$\begin{aligned} q_{n+1} = q_n + 2\left(-\frac{p_{n+1} + q_n}{2} + \frac{P_{\mathcal{X}}[p_{n+1}] + P_{\mathcal{Y}}[q_n]}{2}\right) \\ = -p_{n+1} + P_{\mathcal{X}}[p_{n+1}] + P_{\mathcal{Y}}[q_n] = P_{\mathcal{X}}[p_{n+1}] = P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g_n] = g_{n+1}, \end{aligned}$$

since $P_{\mathcal{Y}}[q_n] = p_{n+1}$, completing the proof.

Dykstra. Let $g_0 \in \mathcal{X}$, let the sequences (g_n), (h_n), $(\tilde{g}_n)$, $(\tilde{h}_n)$, (p_n), and (q_n) be generated as described in Table 9.1. We need to show that $p_n = \tilde{g}_n + \tilde{h}_n$, $q_n = g_n + h_{n-1}$, and $P_{\mathcal{X}}[p_n] = g_{n+1}$ for all $n \in \mathbb{N}_0$. Base case: by construction,

$$\begin{aligned}
&\tilde{g}_0 + \tilde{h}_0 = P_{\mathcal{Y}}[g_0 + h_{-1}] + 0 = P_{\mathcal{Y}}[g_0] = p_0;\\
&g_0 + h_{-1} = g_0 = q_0;\\
&g_1 = P_{\mathcal{X}}[\tilde{g}_0 + \tilde{h}_0] = P_{\mathcal{X}}[p_0].
\end{aligned}$$

Induction step: if claim holds for $n \in \mathbb{N}_0$, then:

$$\begin{aligned}
g_{n+1} + h_n &= P_{\mathcal{X}}[\underbrace{\tilde{g}_n + \tilde{h}_n}_{=p_n}] + \underbrace{g_n + h_{n-1}}_{=q_n} - \underbrace{\tilde{g}_n}_{=P_{\mathcal{Y}}[q_n]} = q_{n+1};\\
\tilde{g}_{n+1} + \tilde{h}_{n+1} &= P_{\mathcal{Y}}[\underbrace{g_{n+1} + h_n}_{=q_{n+1}}] + \underbrace{\tilde{g}_n + \tilde{h}_n}_{=p_n} - \underbrace{g_{n+1}}_{=P_{\mathcal{X}}[p_n]} = p_{n+1};\\
g_{n+1} &= P_{\mathcal{X}}[\tilde{g}_n + \tilde{h}_n] = P_{\mathcal{X}}[p_n],
\end{aligned}$$

completing the proof.

DR. Let $g_0 \in \mathcal{X}$, let the sequences (g_n), (p_n), and (q_n) be generated as described in Table 9.1. We need to show that $p_n = 2P_{\mathcal{Y}}[g_{n-1}] - g_{n-1}$ and $q_n = g_n$ for all $n \in \mathbb{N}$. Base case: by construction,

$$\begin{aligned}
p_1 &= g_0 + 2\left(-\frac{g_0 + g_0}{2} + P_{\mathcal{Y}}[g_0]\right) = 2P_{\mathcal{Y}}[g_0] - g_0;\\
q_1 &= g_0 - \left(\frac{2P_{\mathcal{Y}}[g_0] - g_0 + g_0}{2} + P_{\mathcal{X}}[p_1]\right)\\
&= g_0 + P_{\mathcal{X}}[2P_{\mathcal{Y}}[g_0] - g_0] - P_{\mathcal{Y}}[g_0] = g_1.
\end{aligned}$$

Induction step: if the claim holds for $n \in \mathbb{N}$, then:

$$\begin{aligned}
p_{n+1} &= p_n + 2\left(-\frac{p_n + q_n}{2} + P_{\mathcal{Y}}[q_n]\right) = 2P_{\mathcal{Y}}[q_n] - q_n = 2P_{\mathcal{Y}}[g_n] - g_n;\\
q_{n+1} &= q_n - \left(\frac{2P_{\mathcal{Y}}[q_n] - q_n + q_n}{2} + P_{\mathcal{X}}[p_{n+1}]\right)\\
&= g_n + P_{\mathcal{X}}[2P_{\mathcal{Y}}[g_n] - g_n] - P_{\mathcal{Y}}[g_n] = g_{n+1}.
\end{aligned}$$

□

The reformulations of AP, Dykstra, and DR in terms of variables p, q (second row of Table 9.1) motivate corresponding evolution equations (row "2v-FPF" of Table 9.1). The structure of these evolution equations becomes more apparent, if one introduces the center–of–mass coordinates $s = \frac{p+q}{2}$ and $d = \frac{p-q}{2}$, see row "2vFPF in CM" in Table 9.1. The resulting variational form (row "$F[s, d]$ form" in Table 9.1) is connected to the flow through the following lemma.

LEMMA 9.2. *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be weakly closed, let $s, d \in \mathcal{H}$, let $E_\mathcal{X}$ be Fréchet-differentiable at $s + d$, let $E_\mathcal{Y}$ be Fréchet-differentiable at $s - d$. Let*

$$F[s, d] := \frac{1}{2} E_\mathcal{X}[s + d] + \frac{1}{2} E_\mathcal{Y}[s - d] - \frac{1}{2} \|d\|_2^2.$$

Then,

$$\begin{aligned} \frac{\delta}{\delta s} F[s, d] &= s - \frac{P_\mathcal{X}[s + d] + P_\mathcal{Y}[s - d]}{2}; \\ \frac{\delta}{\delta d} F[s, d] &= \quad - \frac{P_\mathcal{X}[s + d] - P_\mathcal{Y}[s - d]}{2}. \end{aligned}$$

Proof. By Remark 4.12,

$$\begin{aligned} \frac{\delta}{\delta s} \frac{1}{2} E_\mathcal{X}[s + d] &= \frac{s + d}{2} - \frac{P_\mathcal{X}[s + d]}{2}; \\ \frac{\delta}{\delta d} \frac{1}{2} E_\mathcal{Y}[s - d] &= \frac{s - d}{2} - \frac{P_\mathcal{Y}[s - d]}{2}, \end{aligned}$$

and by direct calculation, $\frac{\delta}{\delta d} \frac{1}{2} \|d\|_2^2 = d$. Combine these equations to obtain the desired result. □

Remark 9.3 (2v-FPF as Reflection-Reflection algorithm). Recall that the DR update can be written using reflection operators:

$$g_{n+1} = \frac{1}{2}(g + R_\mathcal{X} \circ R_\mathcal{Y}[g]),$$

and this update is also known as Relaxed-Reflect-Reflect (cf. the reformulation on p. 85).

The similarity between this update and rescaled 2v-FPF variant of DR is striking. Namely, by the previous proposition, DR corresponds to the algorithm with updates

$$\begin{cases} p_{n+1} = p_n + 2 \left(-\frac{p_n + q_n}{2} + P_\mathcal{Y}[q_n] \right), \\ q_{n+1} = q_n - 1 \left(\frac{p_{n+1} + q_n}{2} + P_\mathcal{X}[p_{n+1}] \right). \end{cases}$$

Rescaled to be symmetric, the corresponding equations state:

$$\begin{cases} \partial_t p = -\frac{p+q}{2} + P_\mathcal{Y}[q] = \frac{1}{2} \left(-p + R_\mathcal{Y}[q] \right); \\ \partial_t q = -\frac{p+q}{2} + P_\mathcal{X}[p] = \frac{1}{2} \left(-q + R_\mathcal{X}[p] \right). \end{cases} \tag{DRF}$$

Similarly to DR in Definition 5.53, this is a combination of two relaxed reflections. We call the equations above DRF (Douglas-Rachford Flow).

Remark 9.4. If (p,q) is a fixed point of DRF, then $s = \frac{p+q}{2} \in \mathcal{X} \cap \mathcal{Y}$. Indeed, if

$$0 = \partial_t p = -\frac{p+q}{2} + P_{\mathcal{Y}}[q], \quad \text{and}$$
$$0 = \partial_t q = -\frac{p+q}{2} + P_{\mathcal{X}}[p],$$

then

$$\frac{p+q}{2} = P_{\mathcal{Y}}[q] \in \mathcal{Y}, \quad \text{and} \quad \frac{p+q}{2} = P_{\mathcal{X}}[p] \in \mathcal{X}.$$

Therefore, one can use $E_{\mathcal{X}}[s] + E_{\mathcal{Y}}[s]$ as a criterion for DRF convergence. This is the same energy functional as the one we used to show energy dissipation for alternating projections (Proposition 6.4).

Table 9.1: Gradient system reformulations of AP, Dykstra, and DR algorithms

	AP (ER)	**Dykstra (BIO)**	**DR (HIO)**
Algorithm	Given $g_0 \in \mathcal{X}$, set $g_{n+1} = P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g_n].$	Given g_0, set $h_{-1} = \tilde{h}_0 = 0$, set $\tilde{g}_n = P_{\mathcal{Y}}[g_n + h_{n-1}],$ $h_n = g_n + h_{n-1} - \tilde{g}_n,$ $g_{n+1} = P_{\mathcal{X}}[\tilde{g}_n + \tilde{h}_n],$ $\tilde{h}_{n+1} = \tilde{g}_n + \tilde{h}_n - g_{n+1}.$	Given g_0, set $g_{n+1} = g_n + P_{\mathcal{X}}[2P_{\mathcal{Y}}[g_n] - g_n] - P_{\mathcal{X}}[g_n].$
Reformulation	Set $q_0 = g_0, p_1 = P_{\mathcal{Y}}[g_0]$, $q_1 = P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g_0]$, set $p_{n+1} = p_n + 2\left(-\frac{p_n + q_n}{2} + \frac{P_{\mathcal{X}}[p_n] + P_{\mathcal{Y}}[q_n]}{2}\right),$ $q_{n+1} = q_n + 2\left(-\frac{p_{n+1} + q_n}{2} + \frac{P_{\mathcal{X}}[p_{n+1}] + P_{\mathcal{Y}}[q_n]}{2}\right).$ Then, $q_n = g_n, p_n = P_{\mathcal{Y}}[g_{n-1}]$ for all $n \in \mathbb{N}$.	Set $p_0 = P_{\mathcal{Y}}[g_0]$, $q_0 = g_0$, set $q_{n+1} = q_n + 2\left(\frac{P_{\mathcal{X}}[p_n] - P_{\mathcal{Y}}[q_n]}{2}\right),$ $p_{n+1} = p_n - 2\left(\frac{P_{\mathcal{X}}[p_n] - P_{\mathcal{Y}}[q_{n+1}]}{2}\right).$ Then, $p_n = \tilde{g}_n + \tilde{h}_n,$ $q_n = g_n + h_{n-1},$ and $P_{\mathcal{X}}[p_n] = g_{n+1}$ for all $n \in \mathbb{N}_0$.	Set $p_0 = g_0$, $q_0 = g_0$, set $p_{n+1} = p_n + 2\left(-\frac{p_n + q_n}{2} + P_{\mathcal{Y}}[q_n]\right),$ $q_{n+1} = q_n - 1\left(\frac{p_{n+1} + q_n}{2} + P_{\mathcal{X}}[p_{n+1}]\right).$ Then, $p_n = 2P_{\mathcal{Y}}[g_{n-1}] - g_{n-1},$ $q_n = g_n$ for all $n \in \mathbb{N}$.

Continued on next page

Table 9.1 – *Continued from previous page*

	AP (ER)	**Dykstra (BIO)**	**DR (HIO)**
2v-FPF	Pass to equations. **For DR**: accelerate time in the equation for q by the factor of 2. Get: $\partial_t p = -\frac{p+q}{2} + \frac{P_{\mathcal{X}}[p] + P_{\mathcal{Y}}[q]}{2};$ $\partial_t q = -\frac{p+q}{2} + \frac{P_{\mathcal{X}}[p] + P_{\mathcal{Y}}[q]}{2}.$	$\partial_t q = +\frac{P_{\mathcal{X}}[p] - P_{\mathcal{Y}}[q]}{2};$ $\partial_t p = -\frac{P_{\mathcal{X}}[p] - P_{\mathcal{Y}}[q]}{2}.$	$\partial_t p = -\frac{p+q}{2} + P_{\mathcal{Y}}[q];$ $\partial_t q = -\frac{p+q}{2} + P_{\mathcal{X}}[p].$
2v-FPF in CM	Let $s := (p+q)/2$, $d := (p-q)/2$. Get: $\partial_t s = -s + \frac{P_{\mathcal{X}}[s+d] + P_{\mathcal{Y}}[s-d]}{2};$ $\partial_t d = 0.$	$\partial_t s = 0;$ $\partial_t d = -\frac{P_{\mathcal{X}}[s+d] - P_{\mathcal{Y}}[s-d]}{2}.$	$\partial_t s = -s + \frac{P_{\mathcal{X}}[s+d] + P_{\mathcal{Y}}[s-d]}{2};$ $\partial_t d = \quad -\frac{P_{\mathcal{X}}[s+d] - P_{\mathcal{Y}}[s-d]}{2}.$
$F[s,d]$ form	Let $F[s,d] := \frac{1}{2}E_{\mathcal{X}}[s+d] + \frac{1}{2}E_{\mathcal{Y}}[s-d] - \frac{1}{2}\\|d\\|_2^2$. Then: $\partial_t \begin{pmatrix} s \\ d \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{\delta}{\delta s} \\ \frac{\delta}{\delta d} \end{pmatrix} F[s,d]$	$\partial_t \begin{pmatrix} s \\ d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\delta}{\delta s} \\ \frac{\delta}{\delta d} \end{pmatrix} F[s,d]$	$\partial_t \begin{pmatrix} s \\ d \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\delta}{\delta s} \\ \frac{\delta}{\delta d} \end{pmatrix} F[s,d]$

9.3 DYNAMICS OF 2V-FPF: AN EXAMPLE

The goal of this section is to demonstrate a heuristic explanation of how DR can escape some stagnation points of AP. The following arguments are proposed to justify the explanation.

1) Remark 9.5 juxtaposes AP and Dykstra, arguing the difference in how they solve feasibility problems. Namely, in a (very loose) sense, AP seems to "close the distance" to $\mathcal{X}$ and $\mathcal{Y}$ (minimize $F[s,d]$ in s), while Dykstra "goes to infinity" (maximize $F[s,d]$ in d).

2) Proposition 9.7 illustrates why "going to infinity" can be a good tactic for solving a simple feasibility problem.

3) Example 9.10 illustrates the update rules of 2v-FPF.

4) Example 9.11 illustrates dynamics of 2v-FPF for three different feasibility problems (convex feasible, convex non-feasible, and non-convex feasible problems; see Figures 9.3, 9.4, 9.5, respectively).

Remark 9.5 (AP vs Dykstra). Assume that (s,d) solves 2v-FPF for AP in center-of-mass coordinates (Table 9.1) with initial values $s(0) = s_0 \in \mathcal{H}$, $d(0) = 0$. Then, $d(t) \equiv 0$, and $s(t)$ solves ERF. In particular,

$$F[s,d] = F[s,0] = E_{\mathcal{X}}[s] + E_{\mathcal{Y}}[s]$$

is non-increasing (Proposition 6.4) and $E_{\mathcal{X}}[s+d] - E_{\mathcal{Y}}[s-d]$ is decreasing or equals zero (Proposition 7.22). In this sense, $p = q = s$ do not increase their distance to $\mathcal{X}$ and $\mathcal{Y}$.

Assume that $(\tilde{s}, \tilde{d})$ solves 2v-FPF for Dykstra (Table 9.1) with initial values $s(0) = 0$, $d(0) = d_0 \in \mathcal{H}$. Then, $s(t) \equiv 0$, and, formally,

$$\begin{aligned}\frac{d}{dt}F[s(t),d(t)] &= \langle d - P_{\mathcal{X}}[d] + (-d - P_{\mathcal{Y}}[-d])(-1) - 2d, -\frac{P_{\mathcal{X}}[d] - P_{\mathcal{Y}}[-d]}{2}\rangle \\ &= \frac{1}{2}\|P_{\mathcal{X}}[d] - P_{\mathcal{Y}}[-d]\|_2^2,\end{aligned}$$

meaning that $F[0,d]$ is non-decreasing.

Further, for many cases of interest $F[0,d]$ may be unbounded as $\|d\|_2 \to \infty$. For example, assume that $\mathcal{X}$ and $\mathcal{Y}$ are bounded, such that $\mathcal{X}$ and $\mathcal{Y}$ are contained in $B_R(0)$ for some $R > 0$. Then,

$$\frac{\|d - P_{\mathcal{X}}[d]\|_2}{\|d\|_2} \geqslant \frac{\|d\|_2 - R}{\|d\|_2} \to 1$$

as $d \to \infty$, and likewise for $\mathcal{Y}$; therefore,

$$\frac{F[0,d]}{\|d\|_2} \geqslant \frac{\frac{1}{2}\|d\|_2 - \frac{1}{2}R + \frac{1}{2}\|d\|_2 - \frac{1}{2}R - \frac{1}{2}\|d\|_2}{\|d\|_2} \to \frac{1}{2},$$

showing $F[0, d] \to \infty$ as $d \to \infty$.

This indicates that in pursuit of $F[s, d]$ maximization in d, 2v-FPF for Dykstra can lead to $d \to \infty$.

Exploration of such unbounded regions may seem counterintuitive at first. One can put forth the following reasoning to why exploration of such regions may be benefitial.

Main tools for exploration of $\mathcal{H}$ (in search of $\mathcal{X} \cap \mathcal{Y}$) are projections $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$. At a point $p = s + d \in \mathcal{H}$, projection $P_{\mathcal{X}}[p]$ conceptually carries two pieces of information:

i) $P_{\mathcal{X}}[p]$ is the closest point to p in $\mathcal{X}$;

ii) no points of $\mathcal{X}$ lie in the interior of $B_{\|p-P_{\mathcal{X}}[p]\|_2}(p)$.

If the ball $B_{\|p-P_{\mathcal{X}}[p]\|_2}(p)$ is very large (as can be the case when $p \to \infty$), a very large region can be excluded from the search, which may lead to recovery of some solution $f \in \mathcal{X} \cap \mathcal{Y}$. This is illustrated further in Proposition 9.7 below.

Heuristically, one can argue that AP exploits information i), while Dykstra exploits information ii).

In Proposition 9.7, we use the notion of regularity that is integral to convex feasibility problems [BB96].

DEFINITION 9.6 (REGULARITY). *Let $M \in \mathbb{N}$, let $\mathcal{X}_1, \dots, \mathcal{X}_M \subset \mathcal{H}$ be weakly closed, assume that $\mathcal{X} := \cap_{i=1}^{M} \mathcal{X}_i \neq \varnothing$. The tuple $(X_1, \dots, X_M)$ is called regular, if $\forall \varepsilon > 0\, \exists \delta > 0$ such that for all $g \in \mathcal{H}$ with*

$$\max_{i \in \{1,\dots,M\}} \left\{ \|g - P_{\mathcal{X}_i}[g]\|_2 \right\} \leqslant \delta$$

holds $\inf_{f \in \mathcal{X}} \|g - f\|_2 \leqslant \varepsilon$.

To quote [BB96], “the geometric idea behind this definition is extremely simple: ’If you are close to all sets, then the intersection can not be too far away’.”

The following proposition exemplifies how “escaping to infinity” may be used to solve a feasibility problem when one of the sets is a half-space. A similar setting was used in [AABT16] to demonstrate global convergence of DR.

PROPOSITION 9.7 (HALF-SPACE FEASIBILITY PROBLEM). *Let $e \in \mathcal{H}$ be such that $\|e\|_2 = 1$, let $b \geqslant 0$, define the closed half-plane $\mathcal{X} = \{f \in \mathcal{H} \mid \langle f, e\rangle \geqslant b\}$. Let $\mathcal{Y} \subset \mathcal{H}$ be weakly closed, assume that $\mathcal{X} \cap \mathcal{Y} \neq \varnothing$.*

Let $g_n = nbe$ for $n \in \mathbb{N}$.

Then, $E_{\mathcal{X}}[P_{\mathcal{Y}}[g_n]] + E_{\mathcal{Y}}[P_{\mathcal{Y}}[g_n]] \to 0$ as $n \to \infty$.

In particular, if $(\mathcal{X}, \mathcal{Y})$ is regular, then the distance between $P_{\mathcal{Y}}[g_n]$ and $\mathcal{X} \cap \mathcal{Y}$ goes to zero, i. e.

$$\lim_{n \to \infty} \inf_{f \in \mathcal{X} \cap \mathcal{Y}} \|P_{\mathcal{Y}}[g_n] - f\|_2 = 0.$$

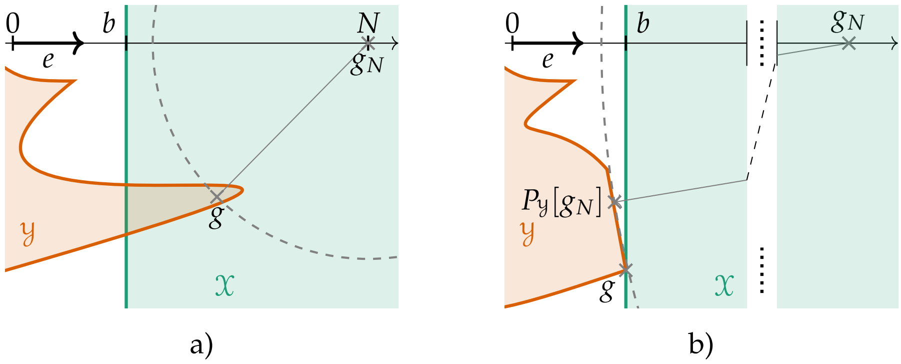

Figure 9.1: Illustration to Proposition 9.7 (half-space feasibility problem). The proposition states: assume $\mathcal{X} \cap \mathcal{Y} \neq \varnothing$, let $g_N = Ne$; then, $P_{\mathcal{Y}}[g_N]$ goes arbitrarily close to $\mathcal{X}$ as $N \to \infty$. The figures illustrate the claims made in the proof of Proposition 9.7.

a) *Claim 1* for $\varepsilon = 0$. If $\|g - g_N\|_2 < \|Ne - be\|_2 = N - b$, then $P_{\mathcal{Y}}[g_N]$ must belong to $\mathcal{X}$. Here, $N = 5$ and $g_N = 5e$. The sequence $P_{\mathcal{Y}}[g_N]$ arrives at a solution in finitely many steps.

b) *Claim 2*. For any solution g, the distance $\|g - g_N\|_2$ becomes less or equal than $N - b + \varepsilon$ for any $\varepsilon > 0$ as $N \to \infty$. In conjuncture with *Claim 1* this means that $P_{\mathcal{Y}}[g_N]$ becomes arbitrarily close to $\mathcal{X}$. Here, $N = 15$ and $g_N = 15e$. The sequence $P_{\mathcal{Y}}[g_N]$ converges to the solution g.

Proof. The proof can be split into two separate claims. First claim states that if the distance between g_n and any solution g is not too large, then $P_{\mathcal{Y}}[g_n]$ is not too far from the half-space $\mathcal{X}$. Second claim states that for large enough n, distance between g_n and g is not too large. In mathematical terms:

Claim 1. Let $\varepsilon > 0$, let $g \in \mathcal{X} \cap \mathcal{Y}$. If $\|g - g_N\|_2 \leqslant N - b + \varepsilon$ for some $N \in \mathbb{N}$, then $\|P_{\mathcal{Y}}[g_n] - P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g_n]\|_2 \leqslant \varepsilon$ for all $n \geqslant N$. In particular, if $\varepsilon = 0$, then $P_{\mathcal{Y}}[g_n] \in \mathcal{X}$.

Claim 2. Let $g \in \mathcal{X} \cap \mathcal{Y}$. Then, for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\|g - g_N\|_2 \leqslant N - b + \varepsilon$.

See Figure 9.1 for an illustration of these claims.

Proof of Claim 1. Let $\varepsilon > 0$, let $g \in \mathcal{X} \cap \mathcal{Y}$, assume that $\|g - g_N\|_2 \leqslant N - b + \varepsilon$ for some $N \in \mathbb{N}$. Since $g \in \mathcal{Y}$,

$$\|g_N - P_{\mathcal{Y}}[g_N]\|_2 \leqslant \|g_N - g\|_2 \leqslant N - b + \varepsilon.$$

Therefore, by triangle inequality have

$$\|P_{\mathcal{Y}}[g_N]\|_2 \geqslant \|g_N\|_2 - \|g_N - P_{\mathcal{Y}}[g_N]\|_2 \geqslant N - (N - b + \varepsilon) \geqslant b - \varepsilon.$$

Further,

$$\|g_N - P_{\mathcal{Y}}[g_N]\|_2^2 = \|g_N\|^2 - 2\langle g_N, P_{\mathcal{Y}}[g_N]\rangle + \|P_{\mathcal{Y}}[g_N]\|_2^2 \\ \geqslant N^2 - 2N\langle P_{\mathcal{Y}}[g_N], e\rangle + (b-\varepsilon)^2.$$

This inequality can be rewritten in the form

$$\begin{aligned} 2N\langle P_{\mathcal{Y}}[g_N], e\rangle &\geqslant N^2 + (b-\varepsilon)^2 - \|g_N - P_{\mathcal{Y}}[g_N]\|_2^2 \\ &\geqslant N^2 + (b-\varepsilon)^2 - (N-b+\varepsilon)^2 \\ &\geqslant 2N(b-\varepsilon); \end{aligned}$$

hence, $\langle P_{\mathcal{Y}}[g_N], e\rangle \geqslant b - \varepsilon$. This implies that $\langle P_{\mathcal{Y}}[g_N] + \varepsilon e, e\rangle \geqslant b$, meaning that $P_{\mathcal{Y}}[g_N] + \varepsilon e \in \mathcal{X}$ and that

$$\|P_{\mathcal{Y}}[g_N] - P_{\mathcal{X}} \circ P_{\mathcal{Y}}[g_N]\|_2 \leqslant \|P_{\mathcal{Y}}[g_N] - (P_{\mathcal{Y}}[g_N] + \varepsilon e)\|_2 \leqslant \varepsilon.$$

To extend this inequality to $n \geqslant N$, observe that if $\|g_n - g\|_2 \leqslant n - b + \varepsilon$ for some $n \geqslant N$, then

$$\|g_{n+1} - g\|_2 = \|g_n + e - g\|_2 \leqslant \|g_n - g\|_2 + 1 \leqslant (n+1) - b + \varepsilon.$$

Thus, *Claim 1* follows by induction in n.

Proof of Claim 2. Let $g \in \mathcal{X} \cap \mathcal{Y}$. Define the following vectors:

$$g_{\parallel} = \langle g, e\rangle e; \qquad g_{\perp} = g - g_{\parallel}.$$

Further, let

$$\begin{cases} \varepsilon \in (0, b) \text{ if } b > 0, \\ \varepsilon \in (0, \frac{\|g_{\parallel}\|_2}{2}) \text{ if } b = 0 \text{ and } \|g_{\parallel}\|_2 > 0, \\ \varepsilon > 0 \text{ if } b = 0 \text{ and } \|g_{\parallel}\|_2 = 0. \end{cases}$$

With these definitions, it is straightforward to verify that

$$c := \|g_{\parallel} - (b-\varepsilon)e\|_2 \geqslant \varepsilon. \tag{9.1}$$

Indeed, since $\|g_{\parallel}\|_2 \geqslant b$ due to $g \in \mathcal{X}$, for $b > 0$, $\varepsilon \in (0, b)$ have

$$\|g_{\parallel} - (b-\varepsilon)e\|_2 = \|g_{\parallel}\|_2 - b + \varepsilon \geqslant \varepsilon;$$

for $b = 0, \|g_{\parallel}\|_2 > 0, \varepsilon \in (0, \frac{\|g_{\parallel}\|_2}{2})$ have

$$\|g_{\parallel} - (b-\varepsilon)e\|_2 = \|g_{\parallel}\|_2 - \varepsilon \geqslant \|g_{\parallel}\|_2/2 \geqslant \varepsilon;$$

for $b = 0, \|g_{\parallel}\|_2 = 0, \varepsilon > 0$ have

$$\|g_{\parallel} - (b-\varepsilon)e\|_2 = \varepsilon.$$

Let $b_\varepsilon := b - \varepsilon$.

Let $N \in \mathbb{N}$ be larger that b_ε. Since $g_N - b_\varepsilon e = (N - b_\varepsilon)e$, we have the following cosine equality:

$$\frac{\langle g_N - b_\varepsilon e, g - b_\varepsilon e\rangle}{\|g_N - b_\varepsilon e\|_2 \|g - b_\varepsilon e\|_2} = \frac{\langle e, g - b_\varepsilon e\rangle}{\|g - b_\varepsilon e\|_2} = \frac{c}{\|g - b_\varepsilon e\|_2}.$$

Therefore, by the cosine theorem have

$$\begin{aligned}\|g_N - g\|_2^2 &= \|g - b_\varepsilon e\|_2^2 + \|g_N - b_\varepsilon e\|_2^2 - 2\|g - b_\varepsilon e\|_2 \|g_N - b_\varepsilon e\|_2 \frac{c}{\|g - b_\varepsilon e\|_2} \\ &= \|g - b_\varepsilon e\|_2^2 + \|g_N - b_\varepsilon e\|_2^2 - 2c\|g_N - b_\varepsilon e\|_2 \\ &= c^2 + \|g_\perp\|_2^2 + (N - b_\varepsilon)^2 - 2c(N - b_\varepsilon).\end{aligned}$$

The goal is to choose N large enough so that

$$\begin{aligned}\|g_N - g\|_2^2 &\overset{!}{\leqslant} (N - b + \varepsilon)^2 = (N - b_\varepsilon)^2 &\Leftrightarrow \\ c^2 + \|g_\perp\|_2^2 + (N - b_\varepsilon)^2 - 2c(N - b_\varepsilon) &\overset{!}{\leqslant} (N - b_\varepsilon)^2 &\Leftrightarrow \\ c^2 + \|g_\perp\|_2^2 &\overset{!}{\leqslant} 2c(N - b_\varepsilon) &\Leftrightarrow \\ N > b + \frac{c}{2} + \frac{1}{c}\|g_\perp\|_2^2.\end{aligned}$$

This latter condition is satisfied by Equation (9.1), if we choose

$$N > b + \frac{\varepsilon}{2} + \frac{1}{\varepsilon}\|g_\perp\|_2^2,$$

showing *Claim 2*.

Thus, *Claim 2* shows that *Claim 1* is applicable; thus, $E_{\mathcal{X}}[P_{\mathcal{Y}}[g_n]] \to 0$ as $n \to \infty$, and trivially $E_{\mathcal{Y}}[P_{\mathcal{Y}}[g_n]] = 0$.

Final statement of the proposition follows directly from the definition of regularity and definitions of $E_{\mathcal{X}}$ and $E_{\mathcal{Y}}$. □

Let us illustrate the dynamics of 2v-FPF using numerically generated examples. To this end, consider the following discretizations of 2v-FPF. Note that we use the coordinates (p, q) rather than (s, d), as discretized algorithms are more numerically stable in these coordinates.

DEFINITION 9.8 (DISCRETIZED 2V-FPF). *Let $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ be weakly closed with projections $P_{\mathcal{X}}, P_{\mathcal{Y}}$, let $p_0, q_0 \in \mathcal{H}$, let $\varepsilon > 0$. The sequences $(p_n^{(\varepsilon)})_{n\in\mathbb{N}}$, $(q_n^{(\varepsilon)})_{n\in\mathbb{N}}$ are generated by the d2v-FPF (discretized 2v-FPF) variant of AP, Dykstra, or DR, if $p_0^{(\varepsilon)} = p_0$, $q_0^{(\varepsilon)} = q_0$, and the following updates hold for all $n \in \mathbb{N}$:*

$$\begin{cases} p_{n+1} = p_n + \varepsilon\left(-\frac{p_n \quad + q_n}{2} + \frac{P_{\mathcal{X}}[p_n \quad] + P_{\mathcal{Y}}[q_n]}{2}\right) \\ q_{n+1} = q_n + \varepsilon\left(-\frac{p_{n+1} + q_n}{2} + \frac{P_{\mathcal{X}}[p_{n+1}] + P_{\mathcal{Y}}[q_n]}{2}\right) \end{cases}$$

for AP;

$$\begin{cases} q_{n+1} = q_n + \varepsilon \left(\frac{P_{\mathcal{X}}[p_n] - P_{\mathcal{Y}}[q_n]}{2} \right) \\ p_{n+1} = p_n - \varepsilon \left(\frac{P_{\mathcal{X}}[p_n] - P_{\mathcal{Y}}[q_{n+1}]}{2} \right) \end{cases}$$

for Dykstra;

$$\begin{cases} p_{n+1} = p_n + \varepsilon \left(-\frac{p_n + q_n}{2} + P_{\mathcal{Y}}[q_n] \right) \\ q_{n+1} = q_n - \varepsilon \left(\frac{p_{n+1} + q_n}{2} + P_{\mathcal{X}}[p_{n+1}] \right) \end{cases} \qquad \text{(dDRF)}$$

for DR.

Remark 9.9 (Sequential and parallel updates). We call update rules from Definition 9.8 *sequential*, meaning that variable instance that is updated in the first equation appears in the second. Thus, iterates p_n, q_n, p_{n+1} and q_{n+1} appear on right-hand sides of sequential updates. These updates are convenient for computational implementation: one always uses the last instance of iterates p and q.

If we replace the indices $n+1$ on the right-hand sides by n, we call the resulting updates *parallel*, meaning that the same instances of variables are used in the first and second equations. Only iterates p_n and q_n appear on right-hand sides of parallel updates. These updates are slightly less convenient for computational implementation: one has to keep a copy of both p_n and q_n to calculate new values of p and q. However, parallel updates exhibit more symmetry and are easier to visualize.

Example 9.10 (Update rules). Figure 9.2 illustrates 2v-FPF evolution equations (Table 9.1). Right-hand sides of these equations correspond one parallel (Remark 9.9) update for Definition 9.8 with $\varepsilon = 1$. The update is shown for a case when $\mathcal{X}$, $\mathcal{Y}$ are two-dimensional non-intersecting balls. Since the parallel update is defined solely by $p_n, q_n, P_{\mathcal{X}}[p_n]$ and $P_{\mathcal{Y}}[q_n]$, the illustration remains valid for other global geometries of $\mathcal{X}$ and $\mathcal{Y}$ as long as $P_{\mathcal{X}}[p_n]$ and $P_{\mathcal{Y}}[q_n]$ are not changed.

Example 9.11 (A two-dimensional example). Figures 9.3, 9.4, 9.5 show examples of 2v-FPF flow $q(t), p(t)$ with starting point $p(0) = q(0) = s_0$ for three feasibility problems.

Figure 9.3 demonstrates a feasible convex problem. The algorithms 2v-FPF AP and 2v-FPF Dykstra converge to $P_{\mathcal{X}\cap\mathcal{Y}}[s_0]$ while 2v-FPF DR converges to some other point in $\mathcal{X} \cap \mathcal{Y}$. This is consistent with the known convex optimization results on Halpern and Dykstra algorithms (cf. Section 6.3.4.1).

Figure 9.4 demonstrates a non-feasible convex problem.

- 2v-FPF AP converges to a minimum of $E_{\mathcal{X}}[s] + E_{\mathcal{Y}}[s]$.

- 2v-FPF Dykstra's p and q escape to infinity; $P_{\mathcal{X}}[p]$ converges to some point in $\mathcal{X}$.
- 2v-FPF DR exhibits combined features: s converges to a minimum of $E_{\mathcal{X}}[s] + E_{\mathcal{Y}}[s]$, and p and q escape to infinity.

Figure 9.5 demonstrates a feasible non-convex problem.

- 2v-FPF AP converges to a local minimum of $E_{\mathcal{X}}[s] + E_{\mathcal{Y}}[s]$ that is not a solution.
- 2v-FPF Dykstra's p and q escape to infinity. $P_{\mathcal{X}}[p]$ converges to some point in $\mathcal{X}$ that is not a solution. $P_{\mathcal{Y}}[q]$ (not shown on the picture) converges to a solution; the jump to the solution may be reckognized at the knick in $q(t)$.
- 2v-FPF DR's p and q converge to a solution. At first, the algorithm exhibits behavior similar to the feasible convex case from Figure 9.3; at a stagnation point, p and q move away from the point as in the non-feasible convex case from Figure 9.4; when p and q are far enough (from the 2v-FPF AP stagnation point), the algorithm escapes the stagnation region and converges to a solution.

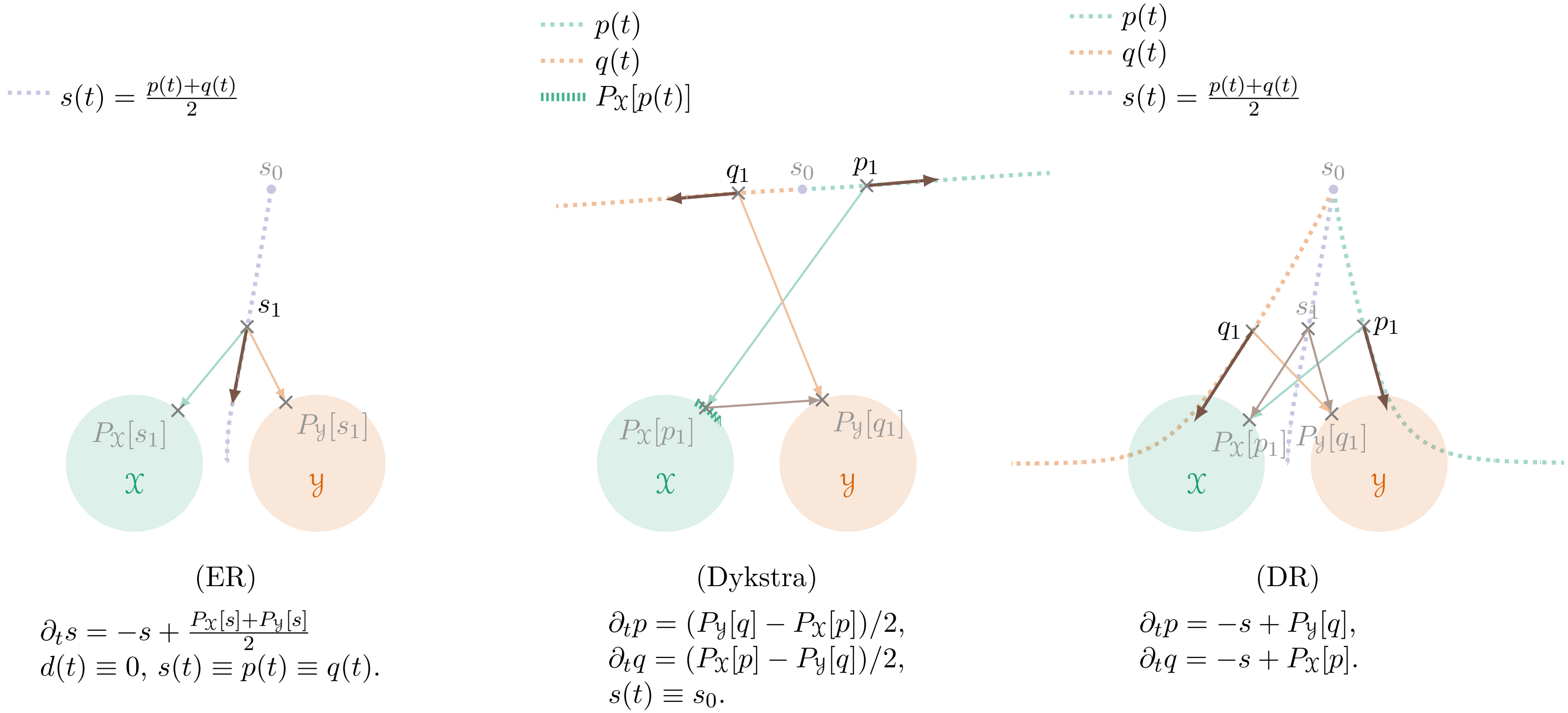

Figure 9.2: Updates of discretized 2v-FPF (Example 9.10)

Brown arrows illustate updates of algorithms at points $q_1 = q(1)$, $p_1 = p(1)$, $s_1 = (p_1 + q_1)/2$. Faded dotted lines illustrate the corresponding 2v-FPF $q(t)$, $p(t)$ with the starting point $p(0) = q(0) = s_0$. See Example 9.11 for details on $q(t)$, $p(t)$.

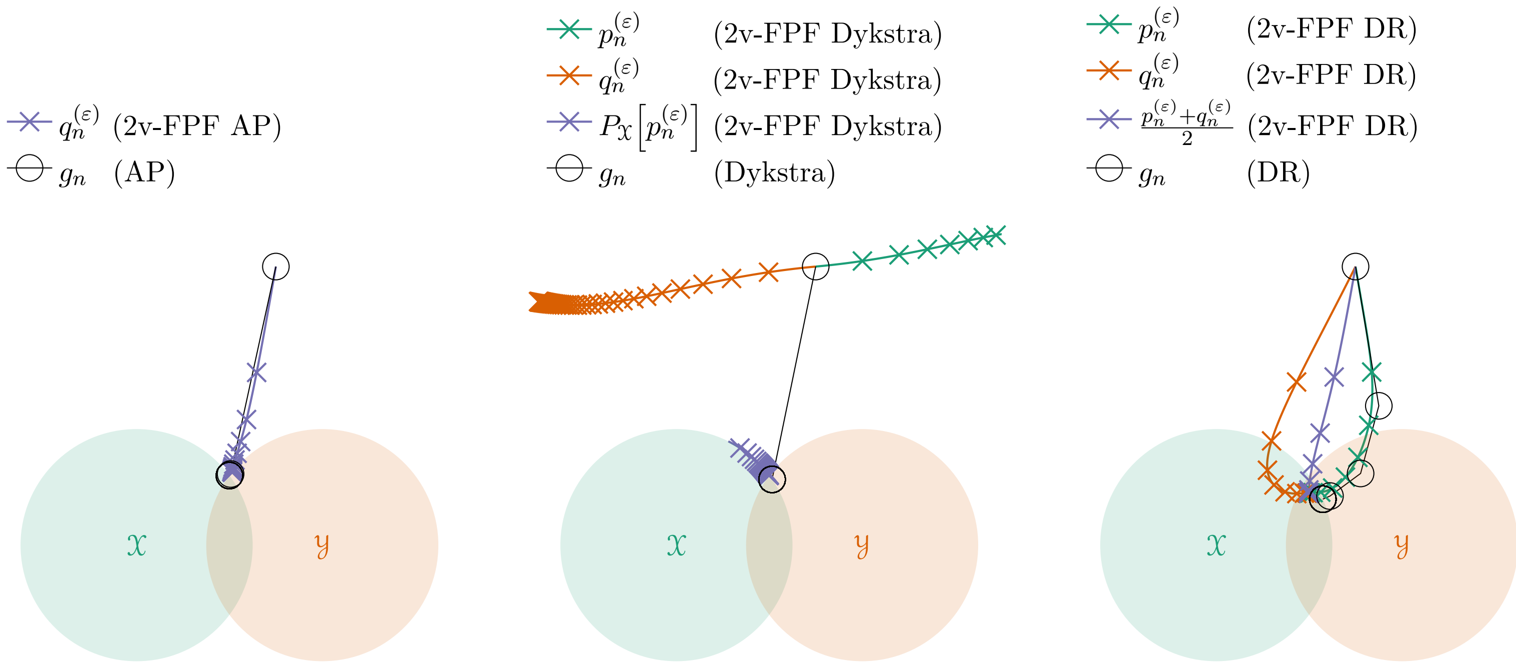

Figure 9.3: Dynamics of 2v-FPF AP, Dykstra and DR: feasible convex case

An example of 2v-FPF flow $q(t), p(t)$ with starting point $p(0) = q(0) = s_0$. The flow is generated using sequential 2v-FPF updates from Definition 9.8 using $\varepsilon = 1/20$; crosses mark points in time $t = (20\varepsilon)n = n$ for $n \in \mathbb{N}$. Classical AP, Dykstra, DR iterates are shown using black circles. See Example 9.11 for details.

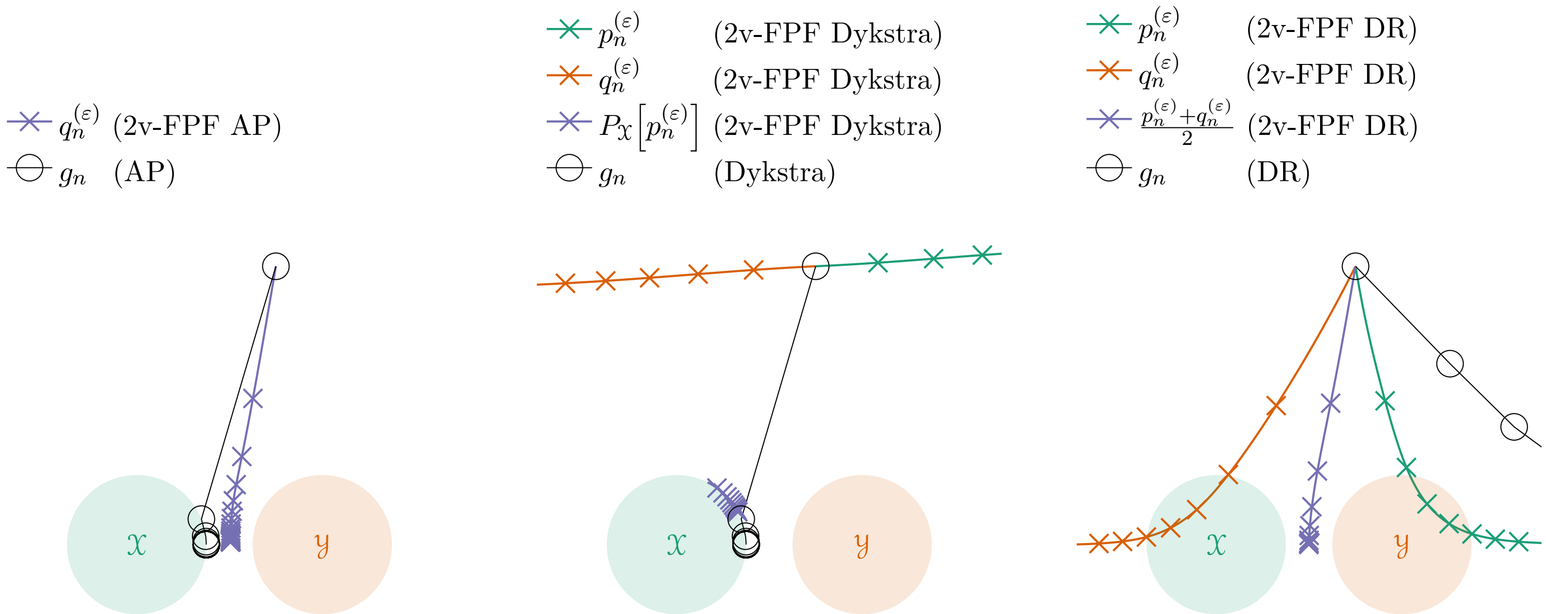

Figure 9.4: Dynamics of 2v-FPF AP, Dykstra and DR: non-feasible convex case

An example of 2v-FPF flow $q(t), p(t)$ with starting point $p(0) = q(0) = s_0$. The flow is generated using sequential 2v-FPF updates from Definition 9.8 using $\varepsilon = 1/20$; crosses mark points in time $t = (20\varepsilon)n = n$ for $n \in \mathbb{N}$. Classical AP, Dykstra, DR iterates are shown using black circles. See Example 9.11 for details.

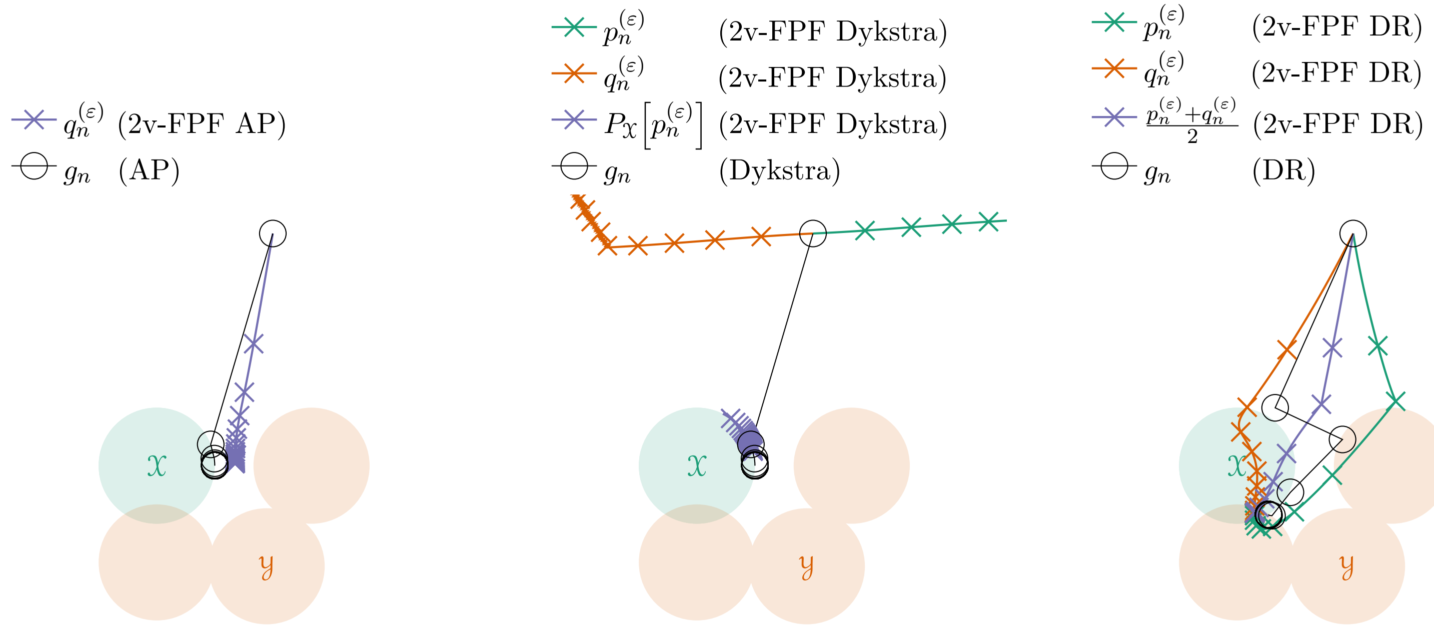

Figure 9.5: Dynamics of Dv-FPF AP, Dykstra and DR: feasible non-convex case

An example of 2v-FPF flow $q(t), p(t)$ with starting point $p(0) = q(0) = s_0$. The flow is generated using sequential 2v-FPF updates from Definition 9.8 using $\varepsilon = 1/20$; crosses mark points in time $t = (20\varepsilon)n = n$ for $n \in \mathbb{N}$. Classical AP, Dykstra, DR iterates are shown using black circles. See Example 9.11 for details.

9.4 DR/HIO FLOW

Specifically, in context of phase retrieval, consider the following set of equations we call the Douglas-Rachford/Hybrid-Input-Output Flow:

$$\begin{aligned} \partial_t p &= -\frac{p+q}{2} + P_{\mathcal{P}}[q] \\ \partial_t q &= -\frac{p+q}{2} + P_{\mathcal{M}}[p]. \end{aligned} \tag{DR/HIO-F}$$

This section demonstrates that DR/HIO-F admits weak solutions with

$$p, q \in L^\infty\big((0,T); L^2(\mathbb{T}^d)\big) \cap C\big([0,T]; L^2(\mathbb{T}^d)\big)$$

for some $T > 0$, and discusses some properties of the solutions.

The proofs are very similar to the ones presented in Chapter 7. A notable difference — that does not significantly alter the proofs — comes from the fact that we are not able to establish a bound on p, q that is uniform in time on $(0, \infty)$. The demonstrated bounds (Lemma 9.13) allow linear growth of $\|p\|_{L^2}$ and $\|q\|_{L^2}$ in time.

9.4.1 *Global weak solutions*

The following is an analogon of Definition 7.5.

Definition 9.12 (Approximation sequence). *Let $\sqrt{I} \in l^1(\mathbb{Z}^d)$ be even and non-negative, let $p_0, q_0 \in L^2(\mathbb{T}^d)$. The sequence $(p_n^{(\varepsilon)}, q_n^{(\varepsilon)})$ is called a d-DR/HIO-F sequence, if it is generated by the following parallel (Remark 9.9) dDRF updates:*

$$\begin{aligned} p_{n+1}^{(\varepsilon)} &= p_n^{(\varepsilon)} + \varepsilon \left(-\frac{p_n^{(\varepsilon)} + q_n^{(\varepsilon)}}{2} + P_{\mathcal{P}}[q_n^{(\varepsilon)}] \right) \\ q_{n+1}^{(\varepsilon)} &= q_n^{(\varepsilon)} + \varepsilon \left(-\frac{p_n^{(\varepsilon)} + q_n^{(\varepsilon)}}{2} + P_{\mathcal{M}}[p_n^{(\varepsilon)}] \right). \end{aligned} \tag{d-DR/HIO-F}$$

Define the piecewise constant interpolations

$$\begin{aligned} &\check{p}^{(\varepsilon)}, \check{q}^{(\varepsilon)} \colon [0, +\infty) \mapsto L^2(\mathbb{T}^d) \\ &\check{p}^{(\varepsilon)}(t) = p_n^{(\varepsilon)} \quad \textit{for } t \in \big[n\varepsilon, (n+1)\varepsilon\big), n \in \mathbb{N}_0; \\ &\check{q}^{(\varepsilon)}(t) = q_n^{(\varepsilon)} \quad \textit{for } t \in \big[n\varepsilon, (n+1)\varepsilon\big), n \in \mathbb{N}_0. \end{aligned}$$

Using $\tau = t - \lfloor t \rfloor_\varepsilon$ (cf. Notation 7.4), define the piecewise linear interpolations

$$\begin{aligned}
&\overline{p}^{(\varepsilon)}, \overline{q}^{(\varepsilon)} \colon [0, +\infty) \mapsto L^2(\mathbb{T}^d) \\
&\widehat{p}^{(\varepsilon)}(t) = \left(1 - \frac{\tau}{h}\right) p_n^{(\varepsilon)} + \frac{\tau}{h} p_{n+1}^{(\varepsilon)} \quad \text{for } t \in [nh, (n+1)\varepsilon), n \in \mathbb{N}_0, \\
&\widehat{q}^{(\varepsilon)}(t) = \left(1 - \frac{\tau}{h}\right) q_n^{(\varepsilon)} + \frac{\tau}{h} q_{n+1}^{(\varepsilon)} \quad \text{for } t \in [nh, (n+1)\varepsilon), n \in \mathbb{N}_0.
\end{aligned}$$

The following is an analogon of Corollary 7.6.

Lemma 9.13 (Interpolations are bounded). *Let $(p_n^{(\varepsilon)}, q_n^{(\varepsilon)})_{n \in \mathbb{N}_0}$ be the d-DR/HIO-F sequence with initial value (p_0, q_0) and step size $\varepsilon \in (0,1]$. Let $\mathcal{H} \in \{H^1(\mathbb{T}^d), L^2(\mathbb{T}^d)\}$, let $c_0 = \max\{\|p_0\|_{\mathcal{H}}, \|q_0\|_{\mathcal{H}}\}$.*

i) Then, $\|p_N^{(\varepsilon)}\|_{\mathcal{H}} \leqslant c_0 + N\varepsilon \|\widecheck{\sqrt{I}}\|_{\mathcal{H}}$, and the same holds for $\|q_N^{(\varepsilon)}\|_{\mathcal{H}}$.

ii) For $r \in [1, \infty]$, the interpolations $\widehat{p}^{(\varepsilon)}$, $\widehat{q}^{(\varepsilon)}$ and $\overline{p}^{(\varepsilon)}$, $\overline{q}^{(\varepsilon)}$ belong to $L^r(0, T; H^1(\mathbb{T}^d))$ with the following estimates:

$$\sup_{t \in (0,T)} \|\widehat{p}^{(\varepsilon)}(t)\|_{H^1} \leqslant c_0 + T \|\widecheck{\sqrt{I}}\|_{H^1}, \tag{9.2}$$

$$\sup_{t \in (0,T)} \|\overline{p}^{(\varepsilon)}(t)\|_{H^1} \leqslant c_0 + T \|\widecheck{\sqrt{I}}\|_{H^1}, \tag{9.3}$$

$$\left(\int_0^T \|\widehat{p}^{(\varepsilon)}(t)\|_{H^1}^r dt \right)^{1/r} \leqslant c_0 T^{1/r} + \frac{T^{2/r}}{2} \|\widecheck{\sqrt{I}}\|_{H^1}; \tag{9.4}$$

$$\left(\int_0^T \|\overline{p}^{(\varepsilon)}(t)\|_{H^1}^r dt \right)^{1/r} \leqslant c_0 T^{1/r} + \frac{T^{2/r}}{2} \|\widecheck{\sqrt{I}}\|_{H^1}; \tag{9.5}$$

and Equations (9.2) to (9.5) remain valid if $\widehat{p}^{(\varepsilon)}$ is replaced by $\widehat{q}^{(\varepsilon)}$, or if $\overline{p}^{(\varepsilon)}$ is replaced by $\overline{q}^{(\varepsilon)}$. Further,

$$\sup_{t \in (0,T)} \left\| \frac{d}{dt} \widehat{p}^{(\varepsilon)}(t) \right\|_{L^2} \leqslant c_0 + T \|\widecheck{\sqrt{I}}\|_{H^1}; \tag{9.6}$$

$$\sup_{t \in (0,T)} \left\| \frac{d}{dt} \widehat{q}^{(\varepsilon)}(t) \right\|_{L^2} \leqslant c_0 + (T+1) \|\widecheck{\sqrt{I}}\|_{H^1}; \tag{9.7}$$

$$\left(\int_0^T \left\| \frac{d}{dt} \widehat{p}^{(\varepsilon)}(t) \right\|_{L^2}^r dt \right)^{1/r} \leqslant T^{1/r} c_0 + \frac{T^{2/r}}{2} \|\widecheck{\sqrt{I}}\|_{H^1}; \tag{9.8}$$

$$\left(\int_0^T \left\| \frac{d}{dt} \widehat{q}^{(\varepsilon)}(t) \right\|_{L^2}^r dt \right)^{1/r} \leqslant T^{1/r} c_0 + \frac{(T+1)^{2/r}}{2} \|\widecheck{\sqrt{I}}\|_{H^1}. \tag{9.9}$$

Proof. i) We show the statement for $\mathcal{H} = H^1(\mathbb{T}^d)$; the other case follows analogously. The claim is shown by induction in N. The base case $N = 0$ follows by definition of c_0. If the claim is true for $N \in \mathbb{N}$, then

$$\|p_{N+1}^{(\varepsilon)}\|_{H^1} = \left(1 - \frac{\varepsilon}{2}\right) \left\|p_N^{(\varepsilon)}\right\|_{H^1} + \frac{\varepsilon}{2}\Big\|\underbrace{-q_{N+1}^{(\varepsilon)} + 2P_{\mathcal{P}}[q_{N+1}^{(\varepsilon)}]}_{=q_{N+1}^{(\varepsilon)}}\Big\|_{H^1}$$
$$\leqslant \max\left\{q_N^{(\varepsilon)}, p_N^{(\varepsilon)}\right\} \leqslant c_0 + N\varepsilon\|\check{\sqrt{I}}\|_{H^1} \leqslant c_0 + (N+1)\varepsilon\|\check{\sqrt{I}}\|_{H^1}.$$

Similarly,

$$\left\|q_{N+1}^{(\varepsilon)}\right\|_{H^1} = \left(1 - \frac{\varepsilon}{2}\right) \left\|q_N^{(\varepsilon)}\right\|_{H^1} + \frac{\varepsilon}{2}\left\|p_{N+1}^{(\varepsilon)}\right\|_{H^1} + \varepsilon\left\|P_{\mathcal{M}}[p_N^{(\varepsilon)}]\right\|_{H^1}$$
$$\leqslant c_0 + N\varepsilon\|\check{\sqrt{I}}\|_{H^1} + \varepsilon\|\check{\sqrt{I}}\|_{H^1} = c_0 + (N+1)\varepsilon\|\check{\sqrt{I}}\|_{H^1}.$$

ii) Equations (9.2) to (9.5) follow immediately from the definition of $\widetilde{p}^{(\varepsilon)}, \overline{\widetilde{p}}^{(\varepsilon)}$ and i). For Equation (9.6) observe that for almost all $t \in (0, T)$ there exists $n \in \mathbb{N}$ such that $n \leqslant N_\varepsilon(T)$ (with N_ε as in Notation 7.4), and such that

$$\left\|\frac{d}{dt}\widetilde{p}^{(\varepsilon)}(t)\right\|_{L^2} \leqslant \left\|-\frac{p_n^{(\varepsilon)} + q_n^{(\varepsilon)}}{2} + P_{\mathcal{P}}[q_n^{(\varepsilon)}]\right\|_{L^2}$$
$$\leqslant \frac{1}{2}\left\|-p_n^{(\varepsilon)} + |q_n^{(\varepsilon)}|\right\|_{L^2} \leqslant \frac{1}{2}\left(\left\|p_{N_\varepsilon(T)}^{(\varepsilon)}\right\|_{L^2} + \left\|q_{N_\varepsilon(T)}^{(\varepsilon)}\right\|_{L^2}\right)$$
$$\leqslant c_0 + N_\varepsilon(T)\varepsilon\|\check{\sqrt{I}}\|_{L^2} = c_0 + \lfloor T\rfloor_\varepsilon\|\check{\sqrt{I}}\|_{L^2} \leqslant c_0 + T\|\check{\sqrt{I}}\|_{L^2},$$

using the fact that $-q + 2P_{\mathcal{P}}[q] = |q|$ for all $q \in L^2(\mathbb{T}^d)$. Similarly,

$$\left\|\frac{d}{dt}\widetilde{q}^{(\varepsilon)}(t)\right\|_{L^2} \leqslant \left\|-\frac{p_n^{(\varepsilon)} + q_n^{(\varepsilon)}}{2} + P_{\mathcal{M}}[p_n^{(\varepsilon)}]\right\|_{L^2}$$
$$\leqslant \frac{1}{2}\left\|-p_n^{(\varepsilon)} + |q_n^{(\varepsilon)}|\right\|_{L^2} + \left\|\check{\sqrt{I}}\right\|_{L^2} \leqslant c_0 + N_\varepsilon(T)\varepsilon\|\check{\sqrt{I}}\|_{L^2} + \left\|\check{\sqrt{I}}\right\|_{L^2}$$
$$= c_0 + (\lfloor T\rfloor_\varepsilon + 1)\|\check{\sqrt{I}}\|_{L^2} \leqslant c_0 + (T+1)\|\check{\sqrt{I}}\|_{L^2}.$$

Equations (9.8) and (9.9) follow from Equations (9.6) and (9.7). □

This result can be extended to the support constraint $\mathcal{S}$ instead of positivity. For other constraints, the bounds must be established on a case-by-case basis.

The following is an analogon of Corollary 7.10.

Corollary 9.14 (Strongly convergent subsequence exists). *Let $\tilde{p}_0, \tilde{q}_0 \in H^1(\mathbb{T}^d)$, let $r \in [1, \infty]$, let $T \in (0, \infty)$. Let $J_1 \subset (0, 1]$ be the set of discretization step sizes with $0 \in \overline{J_1}$. Let $(\widetilde{p}^{(\varepsilon)}, \widetilde{q}^{(\varepsilon)})$ be as in Definition 9.12 with initial value $(\tilde{p}_0, \tilde{q}_0)$.*

Then, there exists a set $J_2 \subset J_1$ with $0 \in \overline{J_2}$ such that $\widetilde{p}^{(\varepsilon)}$ converges strongly to some $p \in L^p_{t,L^2} \cap C_{t,L^2}$ and $\widetilde{q}^{(\varepsilon)}$ converges strongly to some $q \in L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.

Proof. Follows from Lemma 9.13 and from the Aubin-Lions-Simon lemma with

$$X_s = H^1(\mathbb{T}^d) \times H^1(\mathbb{T}^d),\ X = L^2(\mathbb{T}^d) \times L^2(\mathbb{T}^d),\ X_w = L^2(\mathbb{T}^d) \times L^2(\mathbb{T}^d),$$

equipped with tensor product norms (inherited from Hilbert spaces H^1 and L^2).

Alternatively, one can apply the Aubin-Lions-Simon lemma with $X_s = H^1(\mathbb{T}^d)$, $X = L^2(\mathbb{T}^d)$, $X_w = L^2(\mathbb{T}^d)$ to construct a set $J_{1.5} \subset J_1$ with $0 \in \overline{J_{1.5}}$ such that $\widetilde{p}^{(\varepsilon)}$ converges strongly to some p; then apply the Aubin-Lions-Simon lemma with $X_s = H^1(\mathbb{T}^d)$, $X = L^2(\mathbb{T}^d)$, $X_w = L^2(\mathbb{T}^d)$ to construct a set $J_2 \subset J_{1.5}$ with $0 \in \overline{J_2}$ such that $\widetilde{q}^{(\varepsilon)}$ converges strongly to some q. □

The following is an analogon of Lemma 7.12.

Lemma 9.15 (Initial value of the limit). *Under assumptions of Corollary 9.14 holds* $\lim_{t\to 0} p(t) = \tilde{p}_0$ *and* $\lim_{t\to 0} q(t) = \tilde{q}_0$*; the limits are taken in* $\|\cdot\|_{L^2}$.

Proof. Throughout this proof, use $\|\cdot\|_2 = \|\cdot\|_{L^2}$. Let us show the claim for $\lim_{t\to 0} q(t) = \tilde{q}_0$.

As in the proof of Lemma 7.12, one has $\inf_{\varepsilon\in J_2} \|q(t) - \widetilde{q}^{(\varepsilon)}(t)\|_2 = 0$, and

$$\|q(t) - \tilde{q}_0\|_2 \leqslant \inf_{\varepsilon\in J_2} \|\widetilde{q}^{(\varepsilon)}(t) - \widetilde{q}^{(\varepsilon)}(0)\|_2.$$

Again, as in the proof of Lemma 7.12

$$\begin{aligned}
\|\widetilde{q}^{(\varepsilon)}(t) - \widetilde{q}^{(\varepsilon)}(0)\|_2 &\leqslant \sum_{n=0}^{N_\varepsilon(t)} \|\widetilde{q}^{(\varepsilon)}(\varepsilon(n+1)) - \widetilde{q}^{(\varepsilon)}(\varepsilon n)\|_2 \\
&\leqslant \sum_{n=0}^{N_\varepsilon(t)} \varepsilon \left\| -\frac{\widetilde{p}^{(\varepsilon)}(\varepsilon n) + \widetilde{q}^{(\varepsilon)}(\varepsilon n)}{2} + P_{\mathcal{M}}[\widetilde{p}^{(\varepsilon)}(\varepsilon n)]\right\|_2 \\
&\overset{(*)}{\leqslant} \sum_{n=0}^{N_\varepsilon(t)} \varepsilon(c_0 + t\|\sqrt{\check{I}}\|_2 + 1\|\sqrt{\check{I}}\|_2),
\end{aligned}$$

the last inequality being derived just as Equation (9.6) in Lemma 9.13. Continue with

$$\begin{aligned}
&\leqslant N_\varepsilon(t)\varepsilon(c_0 + (t+1)\|\sqrt{\check{I}}\|_2) \\
&\leqslant t(c_0 + (t+1)\|\sqrt{\check{I}}\|_2) \to 0
\end{aligned}$$

as $t \to 0$, proving the claim. The case $\lim_{t\to 0} p(t) = \tilde{p}_0$ is analogous, with the bound $c_0 + t\|\sqrt{\check{I}}\|_2$ instead of $c_0 + (t+1)\|\sqrt{\check{I}}\|_2$ in $(*)$. □

The following is an analogon of Lemma 7.13.

Lemma 9.16 (Convergence of $\widetilde{p}^{(\varepsilon)}$ implies convergence of $\breve{p}^{(\varepsilon)}$). *Let $p_0, q_0 \in H^1(\mathbb{T}^d)$, let $r \in [1,\infty]$, let $T \in (0,\infty)$. Let $(\widetilde{p}^{(\varepsilon)}, \widetilde{q}^{(\varepsilon)}), (\breve{p}^{(\varepsilon)}, \breve{q}^{(\varepsilon)})$ be as in Definition 7.5. Let J_2, p, q be as in Corollary 7.10, meaning that $\widetilde{p}^{(\varepsilon)} \to p$ and $\widetilde{q}^{(\varepsilon)} \to q$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.*

Then, $\breve{p}^{(\varepsilon)} \to p$ and $\breve{q}^{(\varepsilon)} \to q$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.

Proof. The proof is similar to the proof of Lemma 7.13. Let us show that $\breve{q}^{(\varepsilon)} \to q$; the case $\breve{p}^{(\varepsilon)} \to p$ follows analogously.

Let us check the convergence in L^p_{t,L^2} for $p \in [1,\infty)$. Proceeding as in Lemma 7.13, have

$$\begin{aligned}
&\int_0^T \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \|\widetilde{q}^{(\varepsilon)}(t) - \breve{q}^{(\varepsilon)}(t)\|^r_{L^2} dt \\
&= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon - 1} \int_0^\varepsilon \Big\| \left(1 - \frac{\tau}{\varepsilon}\right) q_n^{(\varepsilon)} \\
&\qquad + \frac{\tau}{\varepsilon}\Big(q_n^{(\varepsilon)} \underbrace{- \varepsilon \frac{p_n^{(\varepsilon)} + q_n^{(\varepsilon)}}{2} + \varepsilon P_{\mathcal{M}}[p_n^{(\varepsilon)}]}_{\text{do not cancel out}}\Big) - q_n^{(\varepsilon)} \Big\|^r_{L^2} d\tau \\
&= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=0}^{N_\varepsilon - 1} \int_0^\varepsilon \tau^{r/2} \left\| \frac{p_n^{(\varepsilon)} + q_n^{(\varepsilon)}}{2} + P_{\mathcal{M}}[p_n^{(\varepsilon)}] \right\|^r_{L^2} \\
&\leqslant \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \underbrace{\varepsilon^{r/2}}_{\to 0} \underbrace{N_\varepsilon(T)\varepsilon}_{\to T} \big(\underbrace{c_0 + (N_\varepsilon(T)\varepsilon + 1)\|\breve{\sqrt{I}}\|_{L^2}}_{\leqslant c_0 + (T+1)\|\breve{\sqrt{I}}\|_{L^2}}\big)^r = 0,
\end{aligned}$$

where in the last inequality we used Lemma 9.13 with $c_0 = \max\{p_0, q_0\}$.

The convergence in L^∞_{t,L^2} and C_{t,L^2} follows similarly, cf. proof of Lemma 7.13. □

The following is an analogon of Remark 7.14.

Remark 9.17 (Formal derivative of the solution candidate). Let $p_0, q_0 \in H^1(\mathbb{T}^d)$, let $r \in [1,\infty]$, let $T \in (0,\infty)$. Let $\widetilde{p}^{(\varepsilon)}, \breve{p}^{(\varepsilon)}$ and $\widetilde{q}^{(\varepsilon)}, \breve{q}^{(\varepsilon)}$ be as in Definition 9.12. Let J_2, p, q be as in Corollary 7.10, meaning that $\widetilde{p}^{(\varepsilon)} \to p$ and $\widetilde{q}^{(\varepsilon)} \to q$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.

Then, for a test function $\zeta \in C_0^\infty([0,T], L^2(\mathbb{T}^d))$ have — by the same argumentation as in Remark 7.14 —

$$\begin{aligned}
&\int_0^T \langle \partial_t q(t), \zeta(t) \rangle \, dt = \\
&= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \int_0^{\lfloor T \rfloor_\varepsilon} \left\langle -\frac{\breve{p}^{(\varepsilon)}(t) + \breve{q}^{(\varepsilon)}(t)}{2} + P_{\mathcal{M};0}[\breve{p}^{(\varepsilon)}(t)], \zeta(t) \right\rangle dt
\end{aligned}$$

the limit can be brought into the integral by the dominated convegence theorem, applicable due to Lemma 9.13

$$= \int_0^T \left\langle -\frac{1}{2} \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \breve{p}^{(\varepsilon)}(t) - \frac{1}{2} \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \breve{q}^{(\varepsilon)}(t) + \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\breve{p}^{(\varepsilon)}(t)], \zeta(t) \right\rangle dt$$

$$= \int_0^T \left\langle -\frac{p(t)+q(t)}{2} + \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\breve{p}^{(\varepsilon)}(t)], \zeta(t) \right\rangle dt. \tag{9.10}$$

Analogously,

$$\int_0^T \langle \partial_t q(t), \zeta(t)\rangle \, dt = \int_0^T \left\langle -\frac{p(t)+q(t)}{2} + P_{\mathcal{P}}[q(t)], \zeta(t) \right\rangle dt, \tag{9.11}$$

where we used sequential continuity of $P_{\mathcal{P}}$ (Lemma 6.21).

The derivation of Equation (9.11) is rigorous. To make Equation (9.10) rigorous, we need to establish the existence of $\lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} P_{\mathcal{M};0}[\breve{p}^{(\varepsilon)}(t)]$. To this end, we show that $t \mapsto q(t)$ is Lipshitz and therefore a. e. Fréchet-differentiable by the Rademacher theorem (Theorem 7.15).

The following is an analogon of Corollary 7.16.

Corollary 9.18. *Let $p_0, q_0 \in H^1(\mathbb{T}^d)$, let $r \in [1,\infty]$, let $T \in (0,\infty)$. Let $\widetilde{p}^{(\varepsilon)}, \breve{p}^{(\varepsilon)}$ and $\widetilde{q}^{(\varepsilon)}, \breve{q}^{(\varepsilon)}$ be as in Definition 9.12. Let J_2, g be as in Corollary 9.14, meaning that $\widetilde{p}^{(\varepsilon)} \to p$ in $L^p_{t,L^2} \cap C_{t,L^2}$ and $\widetilde{q}^{(\varepsilon)} \to q$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.*

Then, there exists a function $\partial_t q\colon (0,T) \to L^2(\mathbb{T}^d)$ such that

$$\lim_{\varepsilon\to 0} \frac{\|q(t+\varepsilon) - q(t) - \varepsilon\partial_t q(t)\|_{L^2}}{\varepsilon} = 0$$

for Lebesgue–almost all $t \in (0,T)$.

Proof. Proceed as in the proof of Corollary 7.16 show that q is Lipshitz. Let $s,t \in (0,T)$; let $t < s$.

Then,

$$\begin{aligned}
\|q(s) - q(t)\|_2 &= \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \|\breve{q}^{(\varepsilon)}(s) - \breve{q}^{(\varepsilon)}(t)\|_2 \\
&\leqslant \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \sum_{n=N_\varepsilon(t)}^{N_\varepsilon(s)+1} \varepsilon \left\| -\frac{p_n^{(\varepsilon)} + q_n^{(\varepsilon)}}{2} + P_{\mathcal{M};0}[p_n^{(\varepsilon)}] \right\| \\
&\leqslant \lim_{\substack{\varepsilon\to 0\\ \varepsilon\in J_2}} \underbrace{(N_\varepsilon(s) + 1 - N_\varepsilon(t))\varepsilon}_{=\lfloor s\rfloor_\varepsilon - \lfloor t\rfloor_\varepsilon + \varepsilon} (c_0 + (t+1)\|\widecheck{\sqrt{I}}\|_2) \\
&\leqslant (s-t)(c_0 + (t+1)\|\widecheck{\sqrt{I}}\|_2),
\end{aligned}$$

using Lemma 9.13 with $c_0 = \max\{\|p_0\|_{\mathcal{H}}, \|q_0\|_{\mathcal{H}}\}$. The claim follows by the Rademacher theorem (Theorem 7.15). □

The following is an analogon of Lemma 7.17.

Lemma 9.19 (Limit of $P_{\mathcal{M}}[\breve{\bar{p}}^{(\varepsilon)}]$ exists). *Let $p_0, q_0 \in H^1(\mathbb{T}^d)$, let $r \in [1, \infty]$, let $T \in (0, \infty)$. Let $\hat{\check{p}}^{(\varepsilon)}, \breve{\bar{p}}^{(\varepsilon)}$ and $\hat{\check{q}}^{(\varepsilon)}, \breve{\bar{q}}^{(\varepsilon)}$ be as in Definition 9.12. Let J_2, p, q be as in Corollary 9.14, meaning that $\hat{\check{p}}^{(\varepsilon)} \to p$ and $\hat{\check{q}}^{(\varepsilon)} \to q$ in $L^p_{t,L^2} \cap C_{t,L^2}$ as $\varepsilon \to 0$ for $\varepsilon \in J_2$.*

Then, there exists a function

$$\varphi\colon (0,T) \to \left\{ \tilde{\varphi}\colon \mathbb{Z}^d \to [0, 2\pi) \mid \sin \tilde{\varphi}(t) \text{ odd for a. a. } t \in (0,T) \right\}$$

such that

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} P_{\mathcal{M};0}[\breve{\bar{p}}^{(\varepsilon)}(t)] = P_{\mathcal{M};\varphi(t)}[p(t)]$$

in $L^2(\mathbb{T}^d)$ for Lebesgue–almost all $t \in (0,T)$.

Proof. By Corollary 9.18, $\partial_t q$ exists and belongs to $L^2(\mathbb{T}^d)$ for almost all $t \in (0,T)$. Define

$$m_t := \partial_t q(t) + \frac{p(t) + q(t)}{2}.$$

One can use substantially the same argumentation as in Lemma 7.17 to show that $\lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} P_{\mathcal{M};0}[\breve{\bar{p}}^{(\varepsilon)}(t)]$ exists and equals m_t in $L^2(\mathbb{T}^d)$ for Lebesgue–almost all $t \in (0,T)$, and to show that there exists the desired φ such that

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon \in J_2}} P_{\mathcal{M};0}[\breve{\bar{p}}^{(\varepsilon)}(t)] = P_{\mathcal{M};\varphi(t)}[p(t)]$$

in $L^2(\mathbb{T}^d)$. □

The following is an analogon of Theorem 7.18.

Theorem 9.20 (Global solutions exist). *Let $\sqrt{I} \in L^2(\mathbb{T}^d)$ be non-negative, let $\tilde{p}_0, \tilde{q}_0 \in L^2(\mathbb{T}^d)$, let $p \in (1, \infty)$, let $T > 0$.*

Then, there exist

$$\varphi\colon (0,T) \to \left\{ \tilde{\varphi}\colon \mathbb{Z}^d \to [0, 2\pi) \mid \sin \tilde{\varphi}(t) \text{ odd for a. a. } t \in (0,T) \right\}$$

and

$$p, q \in L^\infty\big((0,T); L^2(\mathbb{T}^d)\big) \cap C\big([0,T]; L^2(\mathbb{T}^d)\big)$$

such that p, q belong to $L^p\big((0,T); L^2(\mathbb{T}^d)\big)$ and are weak solutions of the equation

$$\begin{aligned} \partial_t p &= -\frac{p+q}{2} + P_{\mathcal{P}}[q] \\ \partial_t q &= -\frac{p+q}{2} + P_{\mathcal{M};\varphi}[p]. \end{aligned}$$

Proof. The substantial difference of this theorem to Theorem 7.18 lies in the fact that p and q do not belong to $L^\infty\big((0,\infty); L^2(\mathbb{T}^d)\big)$, as they may grow linearly by Lemma 9.13. Otherwise, the proof is essentially the same as the proof of Theorem 7.18. Existence of solution candidates p, q is guaranteed by Corollary 9.14. By Corollary 9.18, $\partial_t q$ exists, and by Lemma 9.19, there exists a function φ necessary to make the argument Remark 9.17 rigorous.

By Lemma 9.15, $p(0) = \tilde{p}_0$ and $q(0) = \tilde{q}_0$, concluding the proof. □

9.4.2 *Discussion of properties*

This section does not contain any rigorous mathematical results, but discusses certain aspects of boundedness and convergence of the DR/HIO-F solutions.

Remark 9.21 (Boundedness). In phase retrieval simulations, we observed discretized DR/HIO-F to remain bounded for $\varepsilon \in (0, 1]$, both for the non-negativity and support size additional constraints.

These observations indicate that it is reasonable to search for an improvement of the bound of Lemma 9.13.

Observe that for the update in q for all $k \in \mathbb{Z}^d$, $\varepsilon \in (0, 1]$, one has

$$\begin{aligned} |\hat{q}(k) + \varepsilon \partial_t \hat{q}(k)| &= \left| \left(1 - \frac{\varepsilon}{2}\right) \hat{q}(k) + \frac{\varepsilon}{2} |2\sqrt{I}(k) - |\hat{p}|(k)| \right| \\ &\leqslant \left(1 - \frac{\varepsilon}{2}\right) |\hat{q}(k)| + \frac{\varepsilon}{2} \max\left\{2\sqrt{I}(k), |\hat{p}|(k)\right\}. \end{aligned}$$

Unfortunately, this pointwise estimate can not be extended to $\|\cdot\|_2$, as, in general,

$$\left\|2\sqrt{I}(k) - |\hat{q}|(k)\right\|_2 \nleqslant \max\left\{\left\|2\sqrt{I}\right\|_2, \left\|\hat{q}\right\|_2.\right\}$$

(Note that for the finite-dimensional cases of a DR variant, certain boundedness conditions have been established, for example, in [LP16], cf. Remark 9.23 below.)

Remark 9.22 (Convergence of DR). Some recent results cover convergence of DR in a variety of settings (see [LS20] for a survey). To the best of our knowledge, first results on the local convergence in a nonconvex feasible setting were demonstrated in [HL13]. A global convergence result for a feasible intersection of semi-algebraic sets was demonstrated in [LP16]. For the non-feasible case, local convergence has recently been established under appropriate assuptions on the regularity of the sets, see [LM20].

The convergence analysis of DR/HIO-F equations lies beyond the scope of this work. One can observe the following formulation of convergence for DR/HIO-F.

As motivated in Remark 9.4, one can use $E_{\mathcal{A}}[s] + E_{\mathcal{M}}[s]$ as a DR/HIO-F convergence criterion. Let (s, d) be a solution of DR/HIO-F in the center-of-mass coordinates. Then, formally,

$$\begin{aligned} &\frac{d}{dt}\left(E_{\mathcal{M}}[s] + E_{\mathcal{A}}[s]\right) \\ &\qquad = \langle s - P_{\mathcal{M}}[s], -s + \frac{P_{\mathcal{A}}[s+d] + P_{\mathcal{M}}[s-d]}{2}\rangle \\ &\qquad +\langle s - P_{\mathcal{A}}[s], -s + \frac{P_{\mathcal{A}}[s+d] + P_{\mathcal{M}}[s-d]}{2}\rangle \\ &\qquad = -\frac{1}{2}\langle s - \frac{P_{\mathcal{A}}[s] + P_{\mathcal{M}}[s]}{2}, s - \frac{P_{\mathcal{A}}[s+d] + P_{\mathcal{M}}[s-d]}{2}\rangle. \end{aligned}$$

Thus, the convergence condition

$$\frac{d}{dt} E_{\mathcal{M}}[s] + E_{\mathcal{A}}[s] < 0$$

can be interpreted geometrically: the angle between vectors $\frac{P_{\mathcal{A}}[s]+P_{\mathcal{M}}[s]}{2} - s$ and $\frac{P_{\mathcal{A}}[s+d]+P_{\mathcal{M}}[s-d]}{2} - s$ must be acute.

One could ask the following: if $s(t_*)$ solves phase retrieval at some time $t_* \in \mathbb{R}$, does $s(t)$ remain a solution at later times t? Observe that if $s(t_*) \in \mathcal{A} \cap \mathcal{M}$ for some $t_* \in \mathbb{R}$, then

$$s(t_*) - \frac{P_{\mathcal{A}}[s(t_*)] + P_{\mathcal{M}}[s(t_*)]}{2} = 0;$$

therefore, $\frac{d}{dt} E_{\mathcal{M}}[s(t_*)] + E_{\mathcal{A}}[s(t_*)] = 0$. Unfortunately, this does not necessarily guarantee that $E_{\mathcal{M}}[s(t)] + E_{\mathcal{A}}[s(t)] = 0$ for all $t > t_*$.

Remark 9.23. The idea to provide an alternative formulation of the Douglas-Rachford algorithm using an appropriate functional has been explored in literature before. For example, results on boundedness of Douglas-Rachford for the non-convex case were established in [LP16] (and extended in other papers, see references in [LS20]) using a three-variables merit function that in our notation would have the form

$$F_{LP}[x, y, z] = \frac{1}{2} E_{\mathcal{X}}[y] + \frac{1}{2} E_{\mathcal{Y}}[z] - \frac{1}{2\eta}\|y - z\|_2^2 + \frac{1}{\eta}\langle x - y, z - y\rangle,$$

where $x, y, z \in \mathcal{H}$, and $\eta > 0$. This would correspond to the functional F introduced in the beginning of this chapter, if one were to set $\eta = 1$ and omit the last term in F_{LP}.

In [LP16], a relaxed version of the Douglas-Rachford algorithm is obtained from the merit function F_{LP} using a splitting technique.

10 NUMERICAL EXAMPLES

This chapter contains figures that illustrate the dynamics of ER, DR, ERF, and DR/HIO-F. All figures are constructed using the same problem, corresponding to the sum of three Gaussians, see Figure E.1; or to its symmetrized version (the sum of six Gaussians) for even-restricted problems. This problem was chosen as a "generic" instance of phase retrieval. It is kept deliberately simple, smooth and exponentially decaying. It is very different from the realistic cases described, for example, in [ELB18].

One reason for this simplicity stems from the fact that this thesis uses non-negativity rather than support size as the additional constraint. (Non-negativity alone is insufficient to reconstruct meaningful solutions of [ELB18] problems, as the algorithms quickly converge to some non-negative, but not sparse function.)

Another reason for this simplicity is the fact that — even for this toy problem — one can observe noteworthy dynamical features of ER and DR.

Feature 1. For the presented problem, performances of ER and DR do not seem to correlate with the underlying discretization dimension. This supports the conjecture that properties of feasibility sets — properties that govern the dynamics of ER and DR — carry over to the infinite-dimensional setting.

Feature 2. The numerical local convergence radius of ER — at least, for some initial values — seems to be conciderably larger than one can expect from theoretical results. This may be connected to the presence of numerous saddle points in the landscape of the energy functional $E_{\mathcal{M}} + E_{\mathcal{P}}$. These saddle points correspond to unstable fixed points of ERF.

Feature 3. Global convergence of DR is not guaranteed, but is observed for some inital values. The local convergence speed of DR is comparative to the local convergence speed of ER.

Note that in these features we wrote "convergence" for readability, while it is more accurate to write "energy decay".

For readability, figures are presented in three sections.

The first section presents figures that illustrate the behavior of the classical ER and DR algorithms for different initializations (global and local) and different dimensions (from 31×31 to 255×255).

The second section presents figures that compare the dynamics of the classical ER and DR algorithms to their discretized flow counterparts. It also shows that the angle between ERF iterates changes in a discontinuous way, which may explain why acceleration of ER may not be very effective.

The third section presents figures that demonstrate Hessians of ER fixed points:

— at an unstable even fixed point far from the solution;

— at a stable non-even fixed point far from the solution;

— at an unstable even fixed point near the (even) solution.

It also demonstrates correlations between support and phase of Hessian eigenvectors corresponding to the smallest negative eigenvalues.

Unless explicitely specified otherwise, all figures use the non-negativity constraint $\mathcal{P}$ and the modulus constraint $\mathcal{M}(\sqrt{I})$ with $\sqrt{I} := |\mathcal{F}(g_{3G;N})|$, $N \in \mathbb{N}$, from Figure E.1. For certain explicitely specified figures — those considering even-restricted versions of algorithms — we use the modulus constraint $\mathcal{M}(\sqrt{I})$ with $\sqrt{I} := |\mathcal{F}(g_{3G;N}^{(\mathrm{even})})|$, $N \in \mathbb{N}$, where

$$g_{3G;N}^{(\mathrm{even})} := \mathcal{F}^{-1}\big(\mathrm{Re}(\hat{g}_{3G;N})\big)$$

is the (real-valued) symmetrized version of $g_{3G;N}$.

For the energy estimation, we use the functional

$$E_{\mathcal{M}}[g_n] = E_{\mathcal{P}}[g_n] + E_{\mathcal{M}}[g_n]$$

for ER iterates g_n (as in [Fie82]), and the functional

$$E_{\mathcal{P}}\big[P_{\mathcal{M}} \circ R_{\mathcal{P}}[g_n]\big] = E_{\mathcal{P}}\big[P_{\mathcal{M}} \circ R_{\mathcal{P}}[g_n]\big] + E_{\mathcal{M}}\big[P_{\mathcal{M}} \circ R_{\mathcal{P}}[g_n]\big] \qquad (10.1)$$

for DR iterates g_n (as in [ELB18], cf. Remark 5.54).

For the discretized ERF estimates, we use the functional $E_{\mathcal{P}}[g_n] + E_{\mathcal{M}}[g_n]$. For the discretized DR/HIO-F estimates, we apply the functional from (10.1) to the iterate s_n in an attempt at maximal consistency. In our simulations this estimate was closely correlated with the quantity $E_{\mathcal{P}}[s_n] + E_{\mathcal{M}}[s_n]$.

10.1 DYNAMICS OF CLASSICAL ER AND DR

This section contains the following figures.

— Example 10.1: a sample ER and DR run, $N = 63$ (with $g \in \mathbb{R}^{N\times N}$), global initialization.

— Example 10.2: “average” ER and DR energy decay, $N = 63$, random global initialization, 20 runs.

— Example 10.3: “average” ER and DR energy decay, $N = 63$, random local initialization, 20 runs.

— Example 10.4: "average" ER and DR energy decay, $N \in \{31, 63, 127, 255\}$, random local initialization, 20 runs for every dimension.

— Example 10.5: ER and DR energy decay, $N \in \{31 \dots 40, 63 \dots 72, 127 \dots 136, 255 \dots 264\}$, zero phase initialization.

— Example 10.6: ER and DR: example of a non-converging run for $g_{3G;31}$.

Example 10.1 (ER and DR: an example run, $N = 63$). Figure 10.1 demonstrates an example of ER and DR dynamic. Initial value of the algorithms is $P_{\mathcal{M}}[g_0]$, where g_0 is uniformly sampled between 0 and 1 in each pixel.

For the plotted initialization, one can see that ER stagnates while DR finds the close vicinity of a solution.

Example 10.2 (ER and DR: global rand. init.; $N = 63$). Figure 10.2 demonstrates the energy of ER and DR runs. The algorithms are instantiated 20 times each using $P_{\mathcal{M}}[g_0]$, where g_0 is uniformly sampled between 0 and 1 in each pixel.

One can see that ER either stagnates or converges; DR consistently finds the close vicinity of a solution, potential convergence is comparable to the convergence of the succesful ER runs.

Example 10.3 (ER and DR: local phase flip init.; $N = 63$). Figure 10.3 demonstrates the ER and DR behavior near the solution. The algorithms are instantiated 20 times each using initial value g_0, where g_0 is obtained from $g_{3G;63}$ by flipping approximately 1 percent of all phases of $\hat{g}_{3G;63}$ by π. The precise prosedure is as follows. First, choose $g_* \in R^{63x63}$ by sampling every pixel uniformely between 0 and 1. Second, let $\varphi := \arg \hat{g}_*$, with the phase chosen such that $\varphi(k) \in [-\pi, \pi]$. Third, change the phase of $\mathcal{F}(g_{3G;63})$ by π at all indices k where $|\varphi(k)| > 0.99\pi$.

All instances of ER and DR yield $E[g_n] \to 0$. The energy decay speed is comparable for ER and DR.

Graphs of sample solutions and distances to $g_{3G;63}$ indicate that — even for problems such as $g_{3G;63}$ with an exponential decay — instances of ER and DR can converge to points that are non-trivially distinct from $g_{3G;63}$.

Example 10.4 (ER and DR: global rand. init.; $N \in \{31, 63, 127, 255\}$). Figure 10.4 illustrates the behavior of ER and DR for different dimensions, for the $g_{3G;N}$, $N \in \{31, 63, 127, 255\}$. For each dimension, the algorithms are instantiated 20 times each using $P_{\mathcal{M}}[g_0]$, where g_0 is uniformly sampled between 0 and 1 in each pixel.

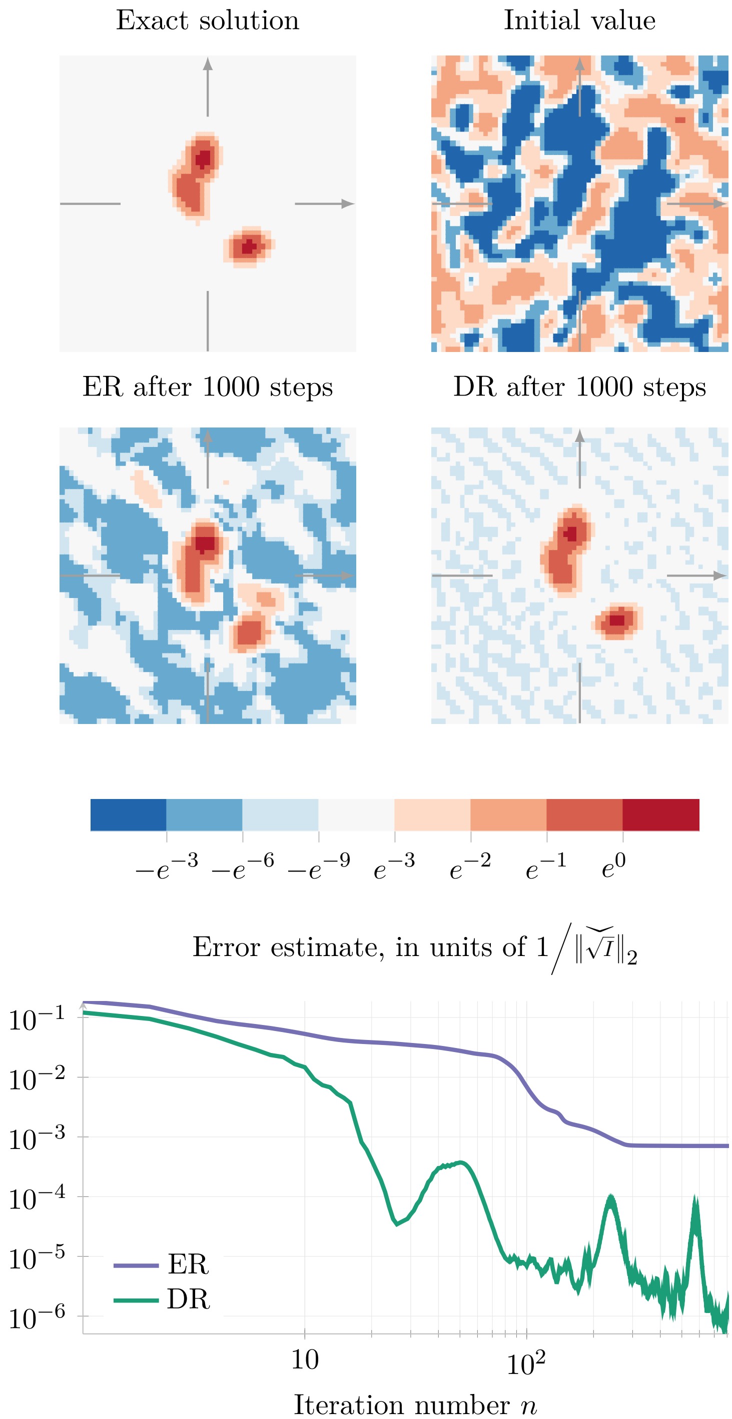

Figure 10.1: ER and DR: example run for $g_{3G;63}$ (Example 10.1). ER and DR for $g_{3G;63}$ with random phase initialization. The energy is plotted in units of $1/\|\check{\sqrt{I}}\|_2$.

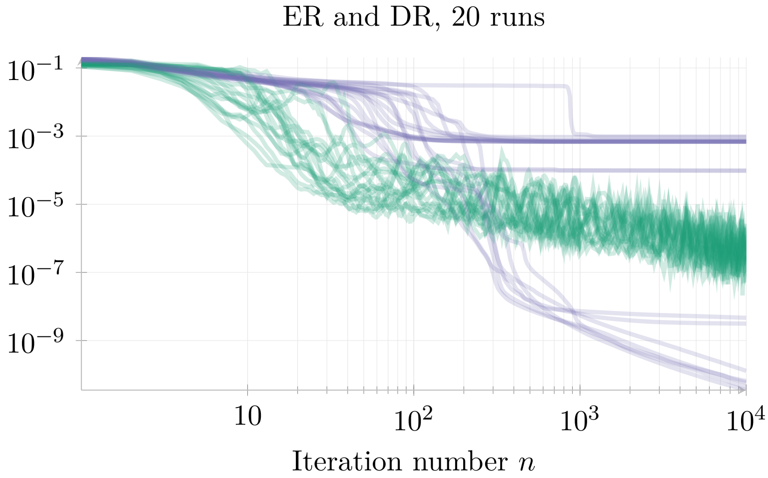

Figure 10.2: ER and DR: global rand. init.; $N = 63$ (Example 10.2). Energy for 20 runs of ER and DR for $g_{3G;63}$ with random initialization, cf. Example 10.2. The energy is plotted in units of $1 / \|\check{\sqrt{I}}\|_2$.

The behavior of the algorithms stays comparable across all depicted dimensions. This suggests that phase retrieval on an infinite-dimensional domain (such as $\mathbb{T}^d$) is a reasonable setting to investigate features of ER and DR.

Example 10.5 (ER and DR: global zero phase init.; different N). Figure 10.5 illustrates that ER and DR are sensitive not only to the initial value, but to change in dimension.

The behavior of ER and DR is shown for $g_{3G;N}$, $N \in \{31 \dots 40, 63 \dots 72, 127 \dots 136, 255 \dots 264\}$. The algorithms are instantiated in every dimension using $P_{\mathcal{M}}[g_0]$, where $g_0 \equiv 1$.

The behavior of the algorithms is highly sensitive to the change in dimension. The non-converging outlier in dimension 31×31 illustrates that there exist combinations of discretizations and starting values for which ER and DR yield poor results, see also Example 10.6.

We note that although the starting value is even, all instances — with the exception of the non-convergent outlier with $N = 31$ — escape the subspace of even functions through numeric perturbations. (This is not shown in the figure.)

Example 10.6 (ER and DR: a non-converging run). Figure 10.6 demonstrates an example of a non-converging run for ER and DR. The algorithms are instantiated using $P_{\mathcal{M}}[g_0]$, where $g_0 \equiv 1$.

For the plotted initialization, one can see that both algorithms can not find a solution after 10^5 iterations. In certain sense, it is natural that the algorithms do not converge, as both ER and DR preserve evenness of the iterate, and $P_{\mathcal{M}}[g_0]$ is even. Thus, all iterates g_n must

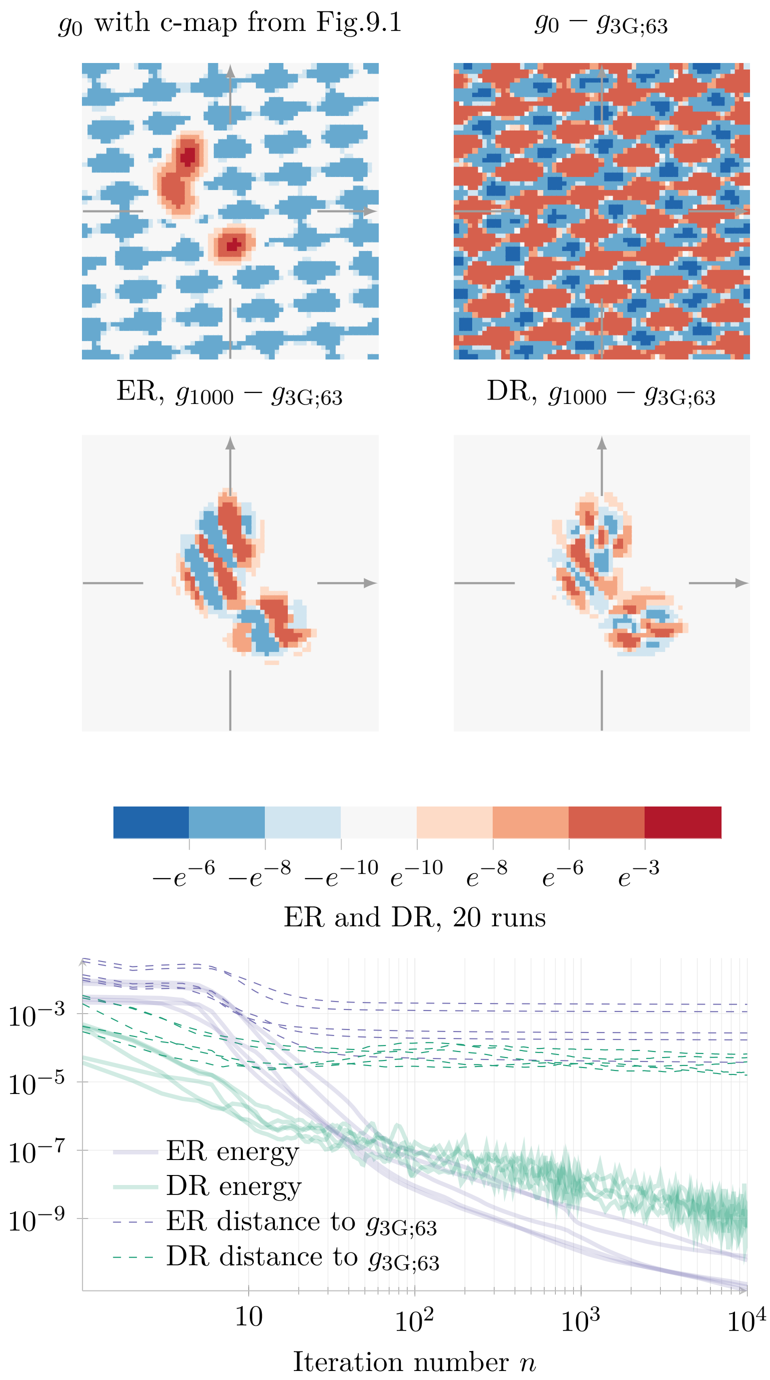

Figure 10.3: ER and DR: local phase flip init.; $N = 63$ (Example 10.3). ER and DR for $g_{3G;63}$ with 20 local (1 percent phase flip) initializations, described in Example 10.3. An example of such initialization is shown in the upper left corner; to ease the comparison, it is shown using the colormap of Figure 10.1. Resulting iterates can not be distinguished from the solution by the naked eye. Hence, other plots show differences between iterates and the solution; they use the colormap specified in this figure.
The bottom graph shows the corresponding energy, and the distances between the iterates and $g_{3G;63}$. The energies and distances are plotted in units of $1 \big/ \|\sqrt{\check{I}}\|_2$.

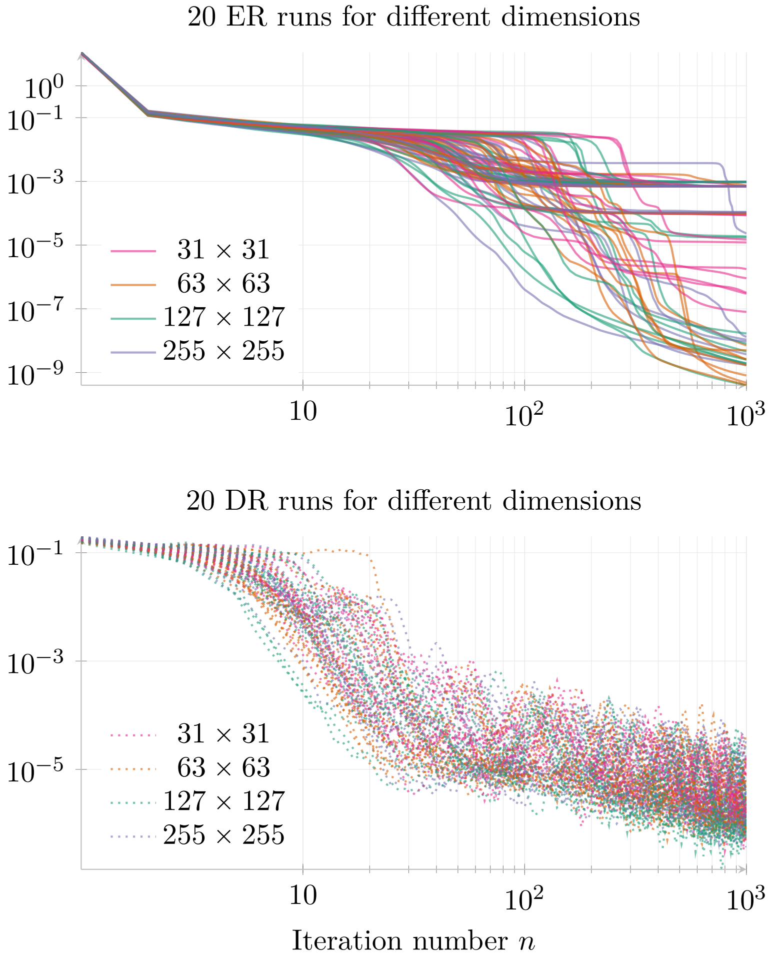

Figure 10.4: ER and DR: global rand. init.; $N \in \{31, 63, 127, 255\}$ (Example 10.4).

The energy of ER and DR for $g_{3G;N}$, $N \in \{31, 63, 127, 255\}$; 20 runs with random initialization. The algorithms are instantiated 20 times each using $P_{\mathcal{M}}[g_0]$, where g_0 is uniformly sampled between 0 and 1 in each pixel. The energy is plotted in units of $1 \big/ \|\check{\sqrt{I}}\|_2$.

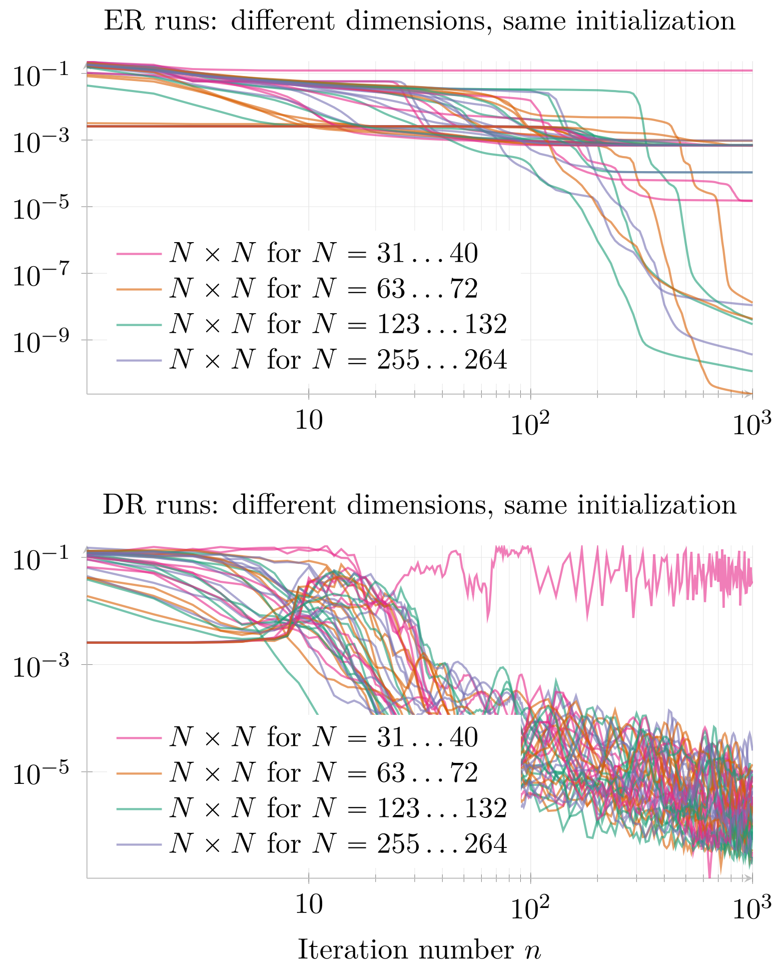

Figure 10.5: ER and DR: global zero phase init.; different N (Example 10.5). The energy of ER and DR for $g_{3G;N}$, $N \in \{31 \ldots 40,\ 63 \ldots 72,\ 127 \ldots 136,\ 255 \ldots 264\}$. The algorithms are instantiated using $P_{\mathcal{M}}[g_0]$, where $g_0 \equiv 1$. The energy is plotted in units of $1 \big/ \|\sqrt{\check{I}}\|_2$. The outlier for ER and DR — the case with highest energy at the end of the run — is the case $N = 31$. For this N, numerical errors fail to perturb evenness of the algorithms, and they fail to converge (cf. Example 10.6).

be even, both for ER and DR, and thus can not converge to the non-even solution $g_{3G;31}$. It is noteworthy that of all dimensions considered in Example 10.5, only $N = 31$, for $g_0 \equiv 1$ exibits this behavior.

10.2 CLASSICAL ALGORITHMS VS. THE DISCRETIZED FLOWS

This section contains the following figures.

— Figure 10.7: energy decay of ER and discretized ERF, for $N \in \{31, 127\}$, for different step sizes ε.

— Figure 10.7: energy decay of DR and discretized DR/HIO-F, for $N \in \{31, 127\}$, for different step sizes ε.

— Figure 10.9: change of direction for the discretized ERF.

Example 10.7 (Similar dynamics of ER and ERF). Figure 10.7 illustrates that the energy decay of ER and ERF is similar for different setups. One can observe a minor difference for the discretized ERF instance with $\varepsilon = 1.25$. This step size is too large for the energy dissipation to hold, cf. Proposition 6.4. As for the decay, this accelerated version is comparable to other discretizations with $\varepsilon \leqslant 1$.

The first row illustrates two runs with the initial value $g_0 = g_{9G;N}$, $N \in \{31, 127\}$, demonstrated in Figure E.2. It was picked as a non-even starting value that can be consistently chosen across different dimensions.

The second row illustrates two runs with the initial values $g_0 = g_{3G;N} + 10^{-4} g_{9G;N}$, $N \in \{31, 127\}$, where $g_{3G;N}$ is the solution we were trying to reconstruct.

The third row illustrates two even-restricted runs, meaning that symmetry of the approximations is enforced after every step by taking $\mathcal{F}^{-1}(\mathrm{Re}(\hat{g}_n^{(\varepsilon)}))$. The exact solution is given by the (real-valued) symmetrized version of $g_{3G;N}$:

$$g_{3G;N}^{(\text{even})} := \mathcal{F}^{-1}(\mathrm{Re}(\hat{g}_{3G;N})),$$

and $\sqrt{I} := |\mathcal{F}(g_{3G;N}^{(\text{even})})|$, $N \in \mathbb{N}$. The initial value is generated by flipping phases at the pixels k with smallest values of $\sqrt{I}(k)$. One can see that the accelerated discretized variant ($\varepsilon = 1$) was able to escape the fixed point in the case $N = 31$.

Overall, performances of ER and ERF remained comparable. This indicates that ERF — and the functional $E_{\mathcal{M}} + E_{\mathcal{P}}$ that generates ERF — can be used as a theoretical tool to study ER.

Example 10.8 (Similar dynamics of ER and ERF). Figure 10.8 illustrates that the energy decay of DR and DR/HIO-F is similar for different

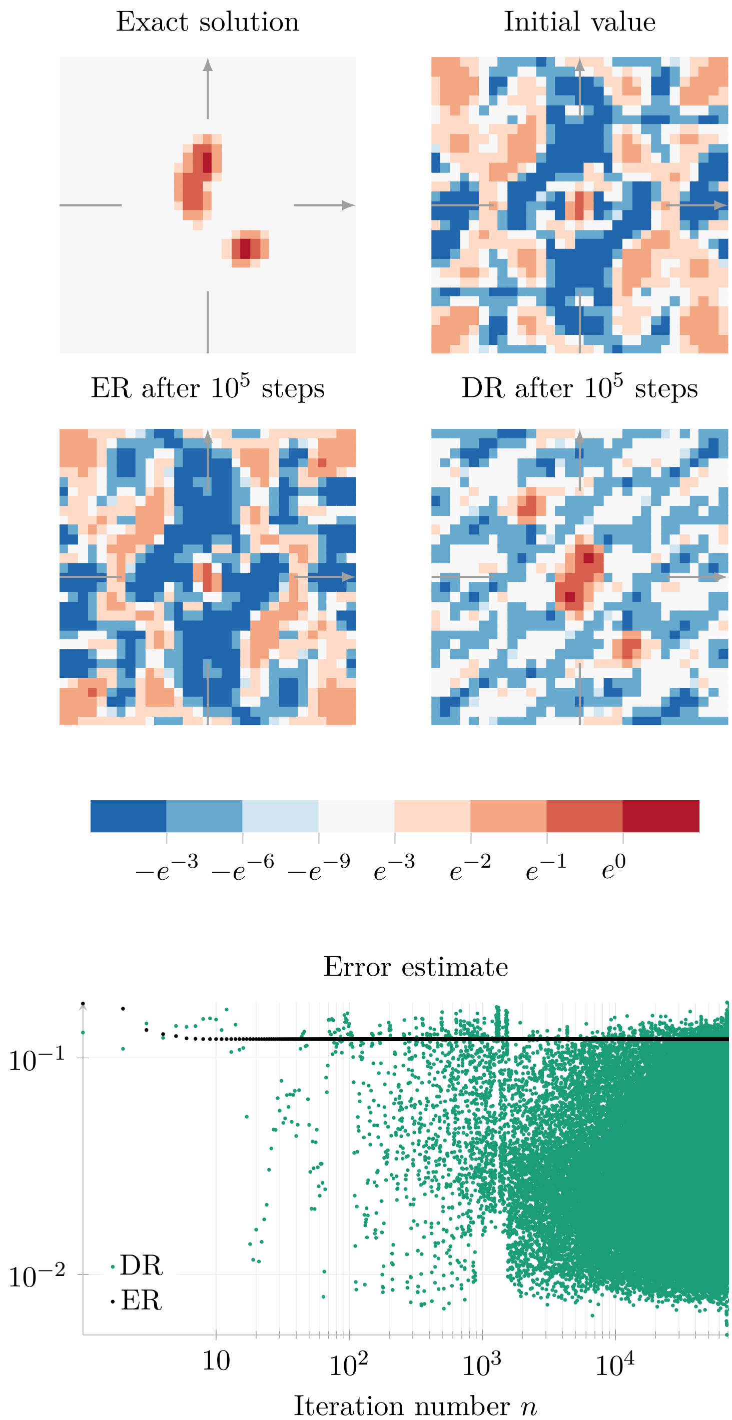

Figure 10.6: ER and DR: a non-converging run (Example 10.6). ER and DR for $g_{3G;31}$ with the zero phase initialization, cf. Example 10.6. The energy is plotted in units of $1 \big/ \|\check{\sqrt{I}}\|_2$. The algorithms do not converge after 10^5 iterations.

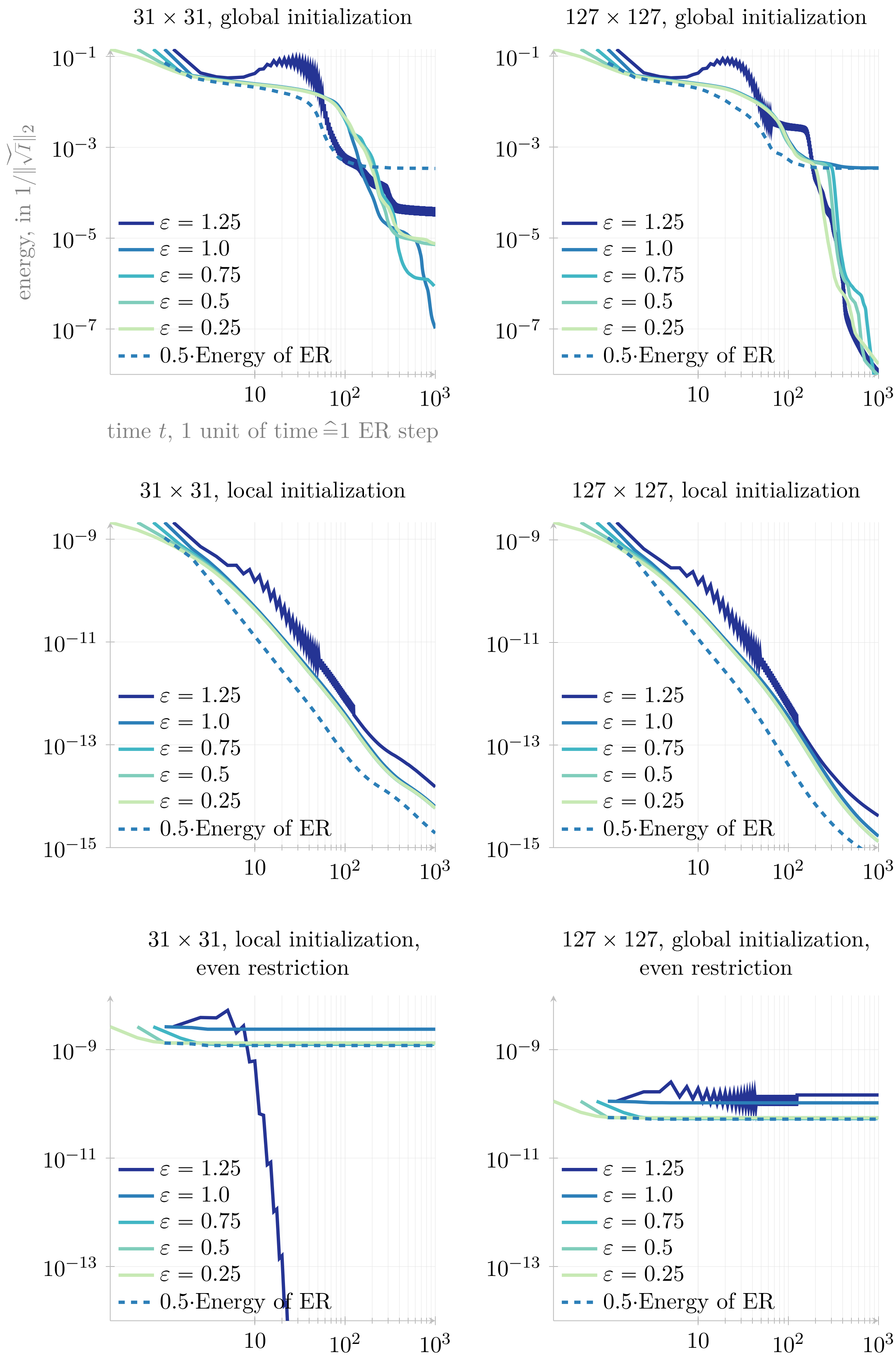

Figure 10.7: Similar dynamics of ER and ERF (see Example 10.7). Axis labels of the upper left graphs apply to all graphs of this figure.

setups. The initializations are the same as in Example 10.7 for comparisson of DR and DRF.

Overall, performances of DR and DR/HIO-F remained comparable. This indicates that DR/HIO-F can be used as a theoretical tool to study DR.

Example 10.9 (Direction change of ERF). Figure 10.9 illustrates that ERF — or, to be more precise, discretized ERF with step size $\varepsilon = 0.005$, — discontinuously changes direction. The plots depict the behavior of finely discretized ERF on the time that corresponds to a single ER step.

The figure plots the quantity $1 - \cos\alpha_n$, where

$$\cos\alpha_n = \frac{\langle g^{(\varepsilon)}_{n+1} - g^{(\varepsilon)}_n, g^{(\varepsilon)}_n - g^{(\varepsilon)}_{n-1}\rangle}{\|g^{(\varepsilon)}_{n+1} - g^{(\varepsilon)}_n\|_2 \, \|g^{(\varepsilon)}_n - g^{(\varepsilon)}_{n-1}\|_2}$$

is the cosine of the angle between two consecutive updates.

This quantity is plotted thrice.

In the top row the algorithm is initialized as follows. First, we run 200 steps of dERF using $\varepsilon = 0.5$ and the initial value $g_0 = g_{9G;127}$ as in Example 10.7. The resulting approximation is used as the initial value for the dERF with $\varepsilon = 0.005$ that is plotted in Figure 10.9. (This instantiation is performed so that the initial value is in a stagnation region.)

In the middle row the algorithm is initialized in a similar way, except for the very first initialization we use $g_0 = g_{3G;127} + 10^{-4} g_{9G;127}$ instead of $g_0 = g_{9G;127}$ (again, as as in Example 10.7).

In the bottom row the algorithm is initialized in a similar way, except we use the symmetrized problem with the solution $g^{(\text{even})}_{3G;N}$, the very first initialization is symmetrized: $g_0 = g^{(\text{even})}_{3G;127} + 10^{-4} g^{(\text{even})}_{9G;127}$, and the symmetry is enforced after each dERF step.

For reference, corresponding energy is plotted in orange for all three instances.

Overall, one can see that discontinuous jumps in the angle appear for all three instances, even for the even-restricted instance where the algorithm typically rapidly converges to a fixed point.

10.3 THE HESSIAN AT THE ERF FIXED POINTS

This section discusses the Hessian of $E = E_{\mathcal{P}} + E_{\mathcal{M}}$. Numerical examples indicate that for phase retrieval with positivity, for solutions with small support, unstable fixed points are prevalent.

Further, — and consistent with the construction made in the proof of Proposition 8.6, — numerical examples indicate that for unstable

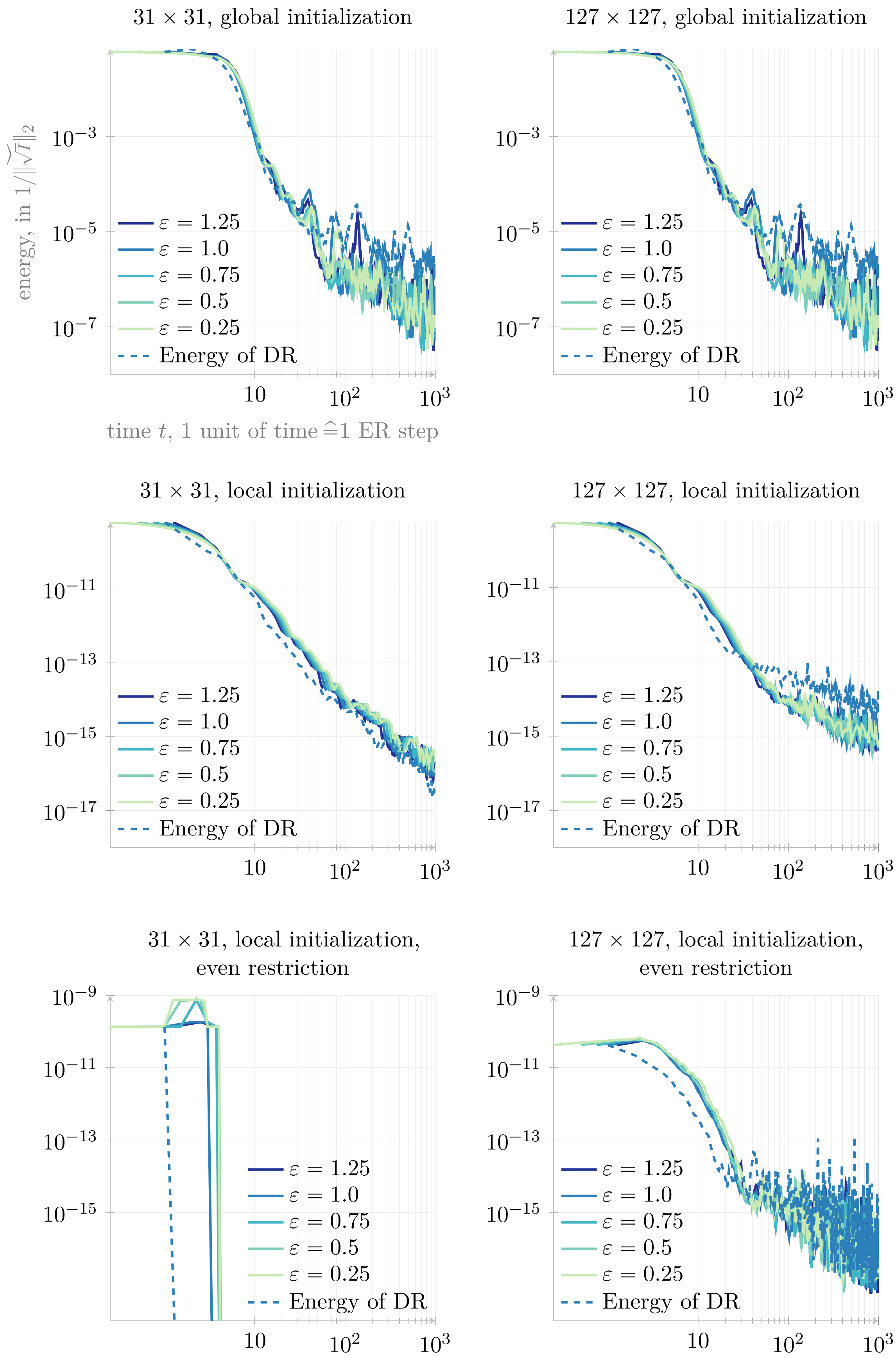

Figure 10.8: Similar dynamics of DR and DRF (see Example 10.8). Axis labels of the upper left graphs apply to all graphs of this figure.

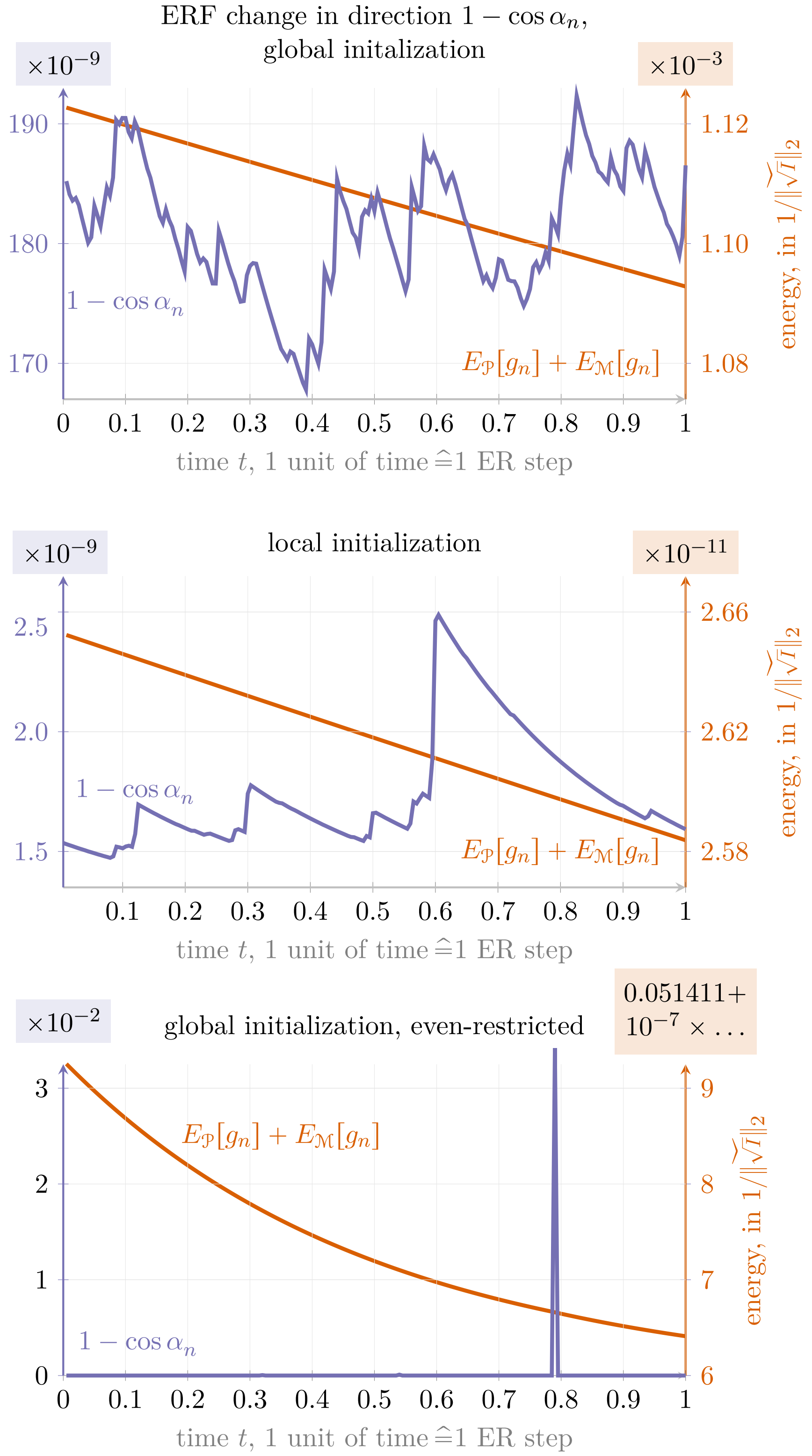

Figure 10.9: Direction change of ERF (see Example 10.9).
The plots depict the quantity $1-\cos\alpha_n$, where the cosine angle is the angle between two consecutive updates, see Example 10.9.
One can see that — also for the even-restricted case, where the algorithm converges to a fixed point (implied by the energy decay) — the algorithm changes the direction in a discontinuous manner.

fixed points, eigenvectors v_i that correspond to the lowest negative eigenvalues λ_i satisfy

$$\operatorname{Re}(\hat{v}_i(k)^* \hat{g}(k)) \approx 0;$$
$$\operatorname{supp} v_i \approx \{k \mid \sqrt{I}(k)/|\hat{g}(k)| > 1\}.$$

This section contains the following figures.

- Figure 10.10 illustrates the Hessian at an unstable even fixed point far from the solution.
- Figure 10.11 illustrates the Hessian at a stable fixed point far from the solution.
- Figure 10.12 illustrates the Hessian at an unstable fixed point near the solution.
- Figure 10.13 illustrates correlations between the support and Fourier phases of Hessian eigenvectors corresponding to the smallest negative eigenvalues.

Example 10.10 (The Hessian: unstable FP far from solution). Figure 10.10 shows Hessian at the even ERF fixed point $g = 0.5(P_{\mathcal{P}}[f] + P_{\mathcal{M}}[f])$, where f is the ER fixed point generated in Example 10.6. For this example, one has

$$\| \operatorname{Re}(\hat{v}_1(k)^* \hat{g}(k))\|_2^2 = 0, \quad \text{and} \|\mathbb{1}_{\{\sqrt{I}/|\hat{g}|.>1\}} v_1\|_2^2 \approx 0.97\|v_1\|_2^2.$$

(Here, exact equality is meant within machine precision.) Similar relationships hold for other eigenvalues, see Figure 10.13.

Example 10.11 (The Hessian: stable FP). Figure 10.11 shows the Hessian at a stable non-even ERF fixed point generated by a long ERF run with a random initialization. The support and Fourier phase eigenvalue correlations that are present for unstable fixed points do not appear in this case.

Example 10.12 (The Hessian: unstable FP near solution). Figure 10.12 shows the Hessian at an unstable even fixed point near the exact solution.

The exact solution is given by the (real-valued) symmetrized version of $g_{3G;N}$:

$$g_{3G;N}^{(\text{even})} := \mathcal{F}^{-1}(\operatorname{Re}(\hat{g}_{3G;N})),$$

and $\sqrt{I} := |\mathcal{F}(g_{3G;N}^{(\text{even})})|$, $N \in \mathbb{N}$. The initial value is generated by flipping the phases at the pixels k with smallest values of $\sqrt{I}(k)$. The fixed point is generated by running an even-restricted dERF with $\varepsilon = 0.5$.

Similarly to Example 10.10, one has

$$\| \operatorname{Re}(\hat{v}_1(k)^* \hat{g}(k))\|_2^2 = 0, \quad \text{and} \quad \|\mathbb{1}_{\{\sqrt{I}/|\hat{g}|>1\}} v_1\|_2^2 \approx 0.93\|v_1\|_2^2.$$

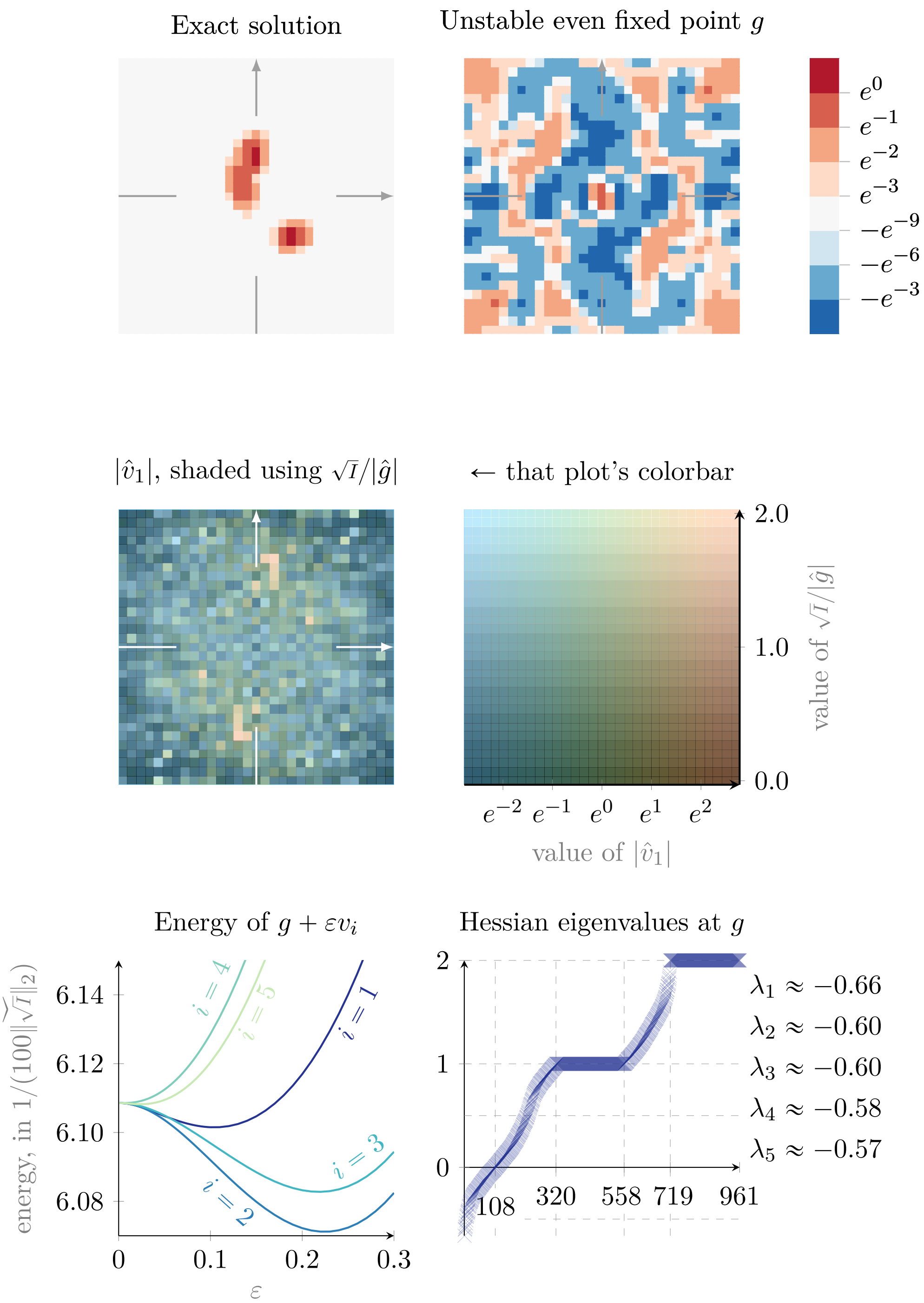

Figure 10.10: The Hessian: unstable FP far from solution (see Example 10.10).

For $g_{3G;31}$ (upper left) with zero phase initialization, Error-Reduction converges to the even fixed point g (upper right), cf. Example 10.6.

The numerical Hessian H at g, described in Section 8.3, shows that this fixed point is unstable. Eigenvalues of H are shown in the lower right corner. The lowest eigenvalue $\lambda_1 \approx -0.66$ is negative. For the corresponding eigenvector v_1, values of $|\hat{v}_1|$ are plotted in the center left. The eigenvector v_1 is odd, meaning that $\operatorname{Im} v_1(k)^* g(k) = 0$ for all k. The support of $|\hat{v}_1|$ is primarily contained in the region $\sqrt{I}/|\hat{g}| > 1.0$.

The bottom left graph shows energy along eigenvectors corresponding to the five lowest eigenvalues.

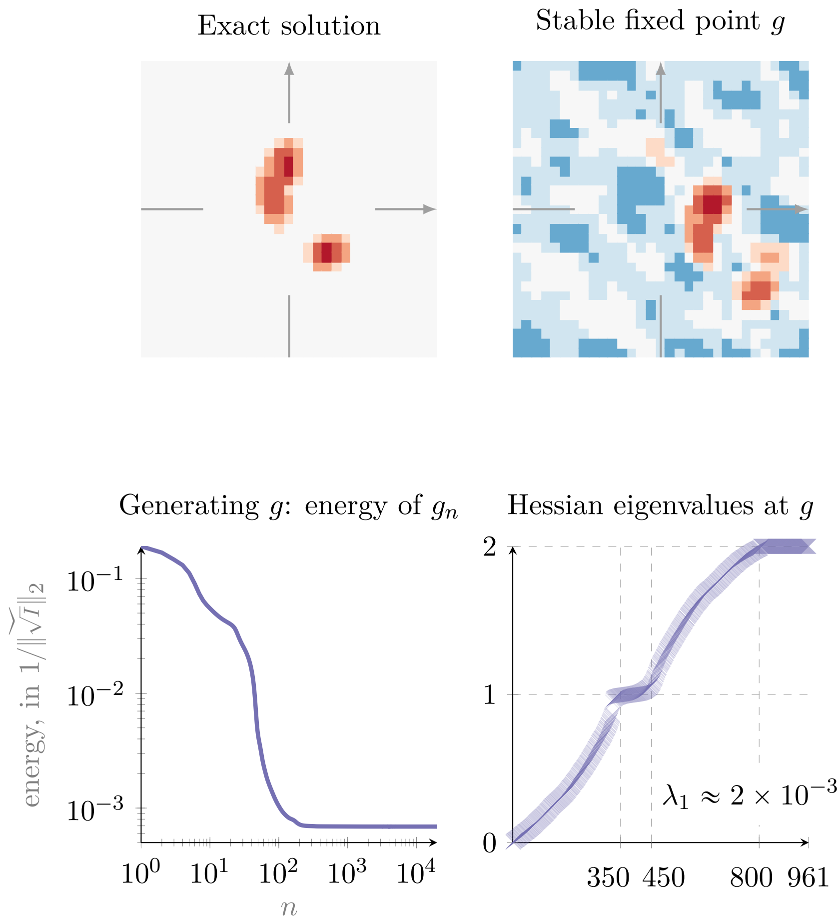

Figure 10.11: The Hessian: stable FP (see Example 10.11). For $g_{3G;31}$ (upper left) with random phase initialization, Error-Reduction converges to the stable fixed point g (in the upper right), the energy $E_{\mathcal{P}}[g] + E_{\mathcal{M}}[g]$ is plotted in the bottom left. The numerical Hessian H at g, described in Section 8.3, shows that this fixed point is stable. Eigenvalues of H are shown in the bottom right. The lowest eigenvalue $\lambda_1 \approx 2 \times 10^{-3}$ is positive. The correlations between λ_i, support of $|\hat{v}_i|$, and angle between $\hat{g}$ and $\hat{v}_i$ that can be observed for negative eigenvalues λ_i are not present here, see Figure 10.13.

(Here, exact equality is meant within machine precision.) Similar relationships hold for the other eigenvalues, see Figure 10.13.

Example 10.13. Figure 10.13 illustrates that at the unstable fixed points, the Hessian eigenvectors v_i that correspond to the lowest negative eigenvalues λ_i satisfy

$$\begin{gathered}\operatorname{Re}(\hat{v}_i(k)^* \hat{g}(k)) \approx 0; \\ \operatorname{supp} v_i \approx \{k \mid \sqrt{I}(k)/|\hat{g}(k)| > 1\}.\end{gathered}$$

This is consistent with the form of the Hessian.

This correlation is not present for the stable fixed points.

We are not able to verify these correlations for unstable non-even fixed points, as it is not clear how such points can be generated.

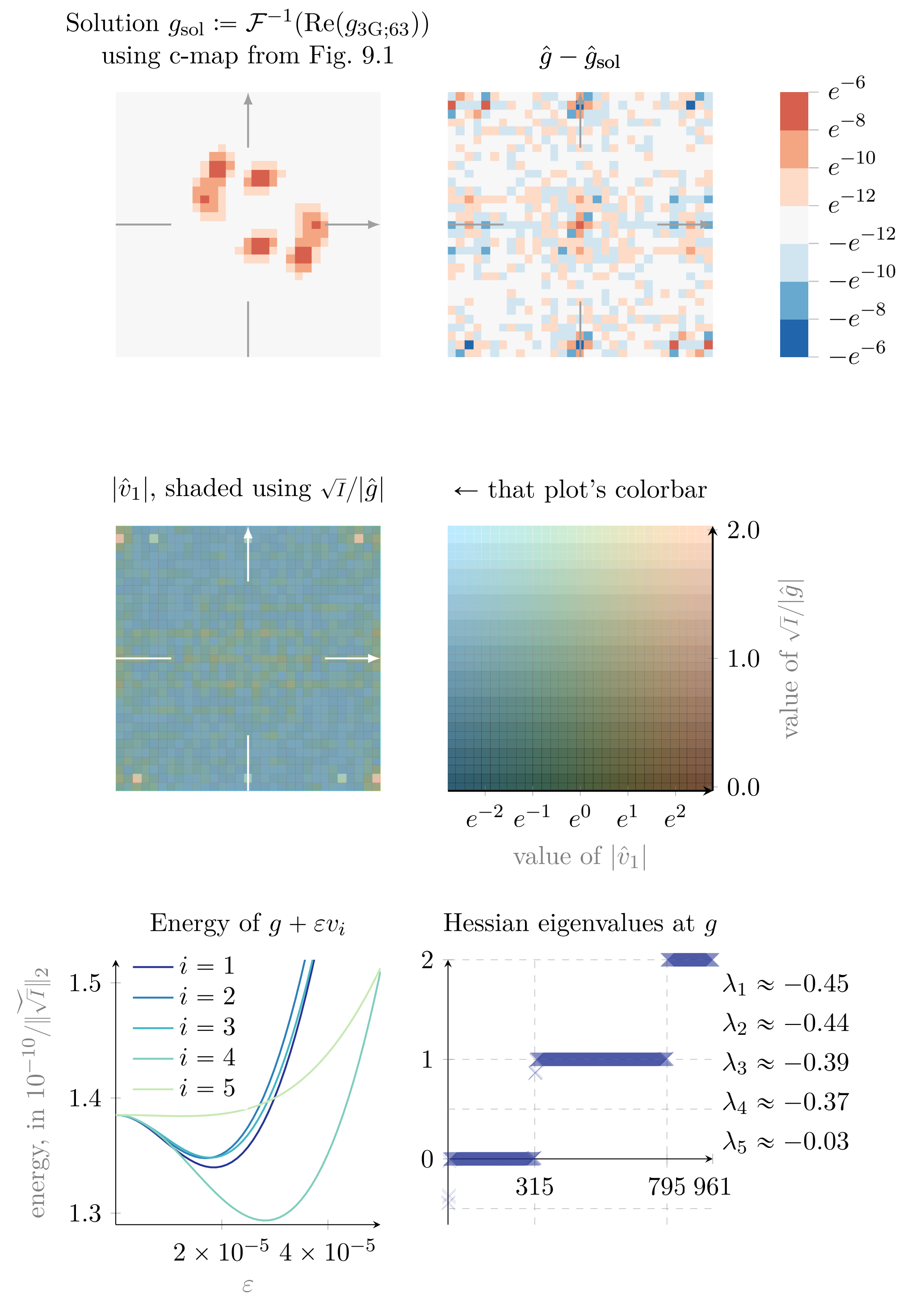

Figure 10.12: The Hessian: unstable FP near solution (see Example 10.12). For $g^{(\mathrm{even})}_{3G;31}$ (upper left) with local phase phase flip initialization, even-restricted ERF converges to the even fixed point g (upper right), undistinguishable from the solution by the naked eye. The numerical Hessian H at g, described in Section 8.3, shows that this fixed point is unstable. Eigenvalues of H are shown in the lower right corner. The lowest eigenvalue $\lambda_1 \approx -0.45$ is negative. For the corresponding eigenvector v_1, values of $|\hat{v}_1|$ are plotted in the center left. The support of $|\hat{v}_1|$ is primarily contained in the region $\sqrt{I}/|\hat{g}| > 1.0$.
The bottom left graph shows energy along eigenvectors corresponding to five lowest eigenvalues.

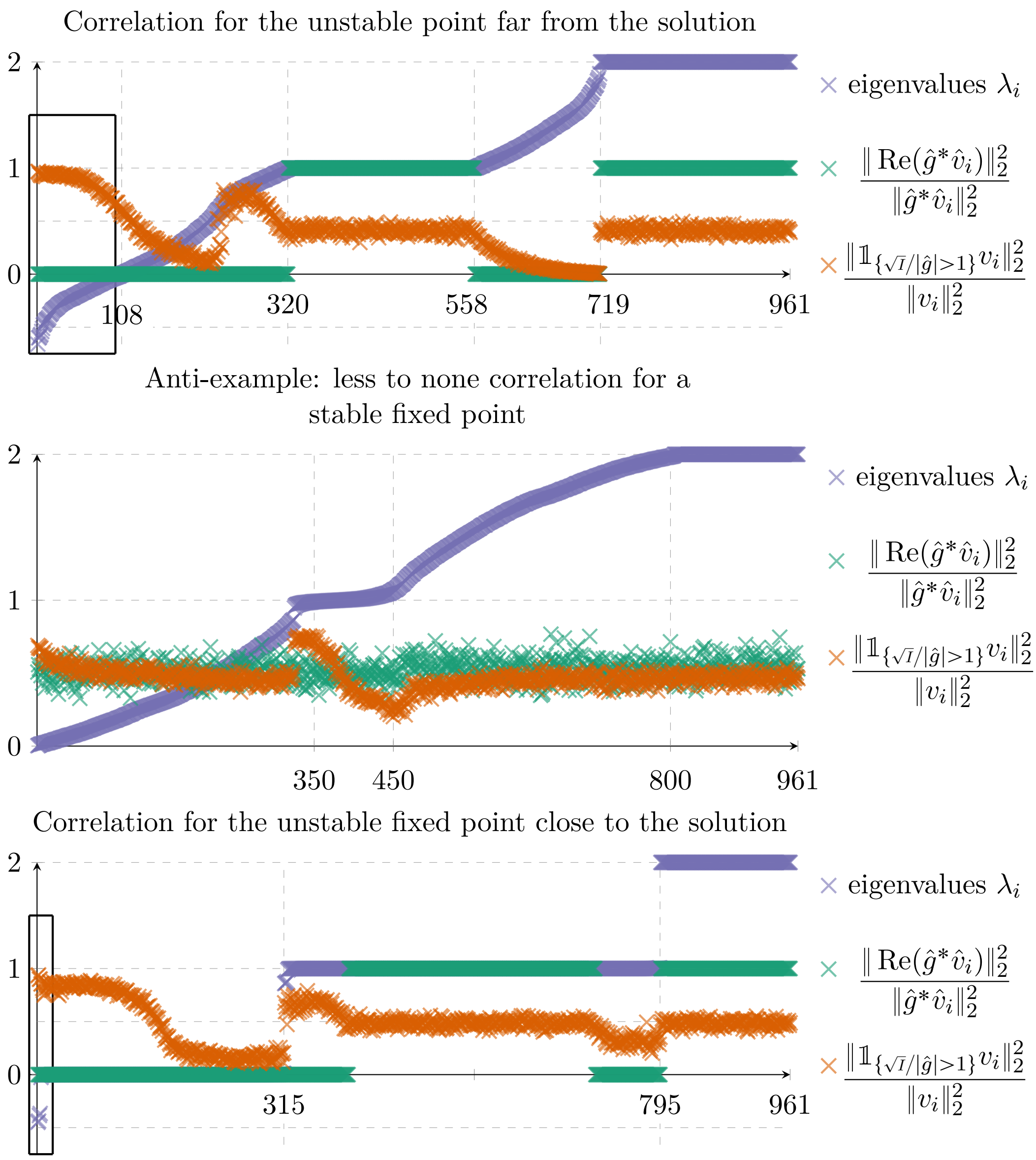

Figure 10.13: Eigenvector correlations for unstable fixed points (Example 10.13).

11 OUTLOOK

In conclusion, we briefly discuss some open questions connected to the results presented in this work.

11.1 WEAK CLOSEDNESS OF THE MODULUS SET AND SUBDIFFERENTIAL SELECTION

We have demonstrated that on bounded domains Ω, the set $\mathcal{M}$ is weakly closed (Section 3.3), and that for weakly closed sets $\mathcal{X}$

$$\begin{aligned} \overline{\partial} E_{\mathcal{X}}[g] &= \overline{\operatorname{conv} g - \Pi_{\mathcal{X}}[g]}^{*} \quad \text{(Clarke subdifferential), and} \\ \partial_{\mathrm{KM}} E_{\mathcal{X}}[g] &= g - \Pi_{\mathcal{X}}[g] \quad \text{(generalized subdifferential)} \end{aligned}$$

(Section 4.3).

While we use a selection of the generalized subdifferential to establish connection to existing algorithms, one could use selections of the Clarke subdifferential to explore variants of ER and DR where the projection $P_{\mathcal{M}}$ is replaced by another operator, for example, the operator

$$\hat{g} \mapsto \begin{cases} \sqrt{I}(k)\frac{\hat{g}(k)}{|\hat{g}(k)|} & \text{if } \hat{g}(k) \neq 0, \\ 0 & \text{else.} \end{cases}$$

In applications, such an operator could be translated to an appropriate regularized version of $P_{\mathcal{M}}$, potentially opening new ways to analyze phase retrieval.

11.2 EXISTENCE AND CONVERGENCE OF ERF

We have demonstrated that Error-Reduction Flow

$$\partial_t g = -(g - P_{\mathcal{M}}[g]) - (g - P_{\mathcal{A}}[g])$$

with non-negative additional constraint $\mathcal{A} = \mathcal{P}$ has global weak solutions that belong to

$$L^{\infty}\big((0,\infty); L^2(\Omega)\big) \cap C\big([0,\infty); L^2(\Omega)\big),$$

where Ω is bounded and has a continuous boundary. The following open questions are connected to this result.

Convergence of ERF. As discussed in Section 5.2, it is known that for non-convex sets, the Alternating Projections algorithm may fail to converge: the set of its fixed points may be a compact continuum. To the extent of our knowledge, this behaviour has not been observed for phase retrieval.

In context of this work, the corresponding questions are:

i) Does every solution of ERF converge to a fixed point (cf. Remark 7.25)?

ii) If Case i) is not true, can one establish convergence under stronger assumptions — for example, for regularized additional constraints, or for even solutions?

iii) If convergence can not be established, can one construct an example — akin to the one shown in [BN13] — for which the solution does not converge?

Existence of ERF for nonconvex additional constraints. As discussed in Remark 7.20, one can attempt to demonstrate existence results for other additional constraints, for example, for the support size constraint $\mathcal{A} = \mathcal{T}_s(\nu)$.

The underlying difficulty is that projections on non-convex constraint sets are not necessarily continuous. Therefore one must substantially modify the arguments of Section 7.2 (or find new arguments) to establish that the solution candidate — constructed by Aubin-Lions — is indeed a solution of ERF.

Existence of ERF on unbounded domains. As briefly elaborated in Remark 6.20, if $\|\sqrt{I}/\widehat{P_{\mathcal{P}}[g_n]}\|_\infty$ is not bounded from below for dERF iterates g_n, one can not expect $P_{\mathcal{M}}[g_n]$ to be sufficiently smooth in Fourier space, or to decay sufficiently fast in object space. Here, "sufficiently smooth" and "sufficiently fast" mean smoothness and decay that would guarantee relative compactness of $\{(g_n)_{n\in\mathbb{N}}\}$ (for example, using criteria of [Peg85]).

One can therefore ask whether it is possible to extend existence results to unbounded domains — possibly, by regularizing the constraints or altering them in an appropriate manner.

This question is particularly relevant for non-crystallographic applications, where the object is not embedded in a periodic structure and thus can not be adequately modeled on bounded domains.

11.3 STABILITY AND DIFFICULTY

We have demonstrated that fixed points f of ERF correspond — under necessary and sufficient conditions on $\|\sqrt{I}/\hat{f}\|_\infty$ — to fixed points of ER. Further, we have demonstrated that fixed points of ERF that do not correspond to ER are unstable (under mild additional assumptions), and that fixed points of even-restricted ER and ERF are likely to be stable (in a finite-dimensional setting, or along appropriate directions in an infinite-dimensional setting). Also, numerical evidence points towards the conjecture that — at least, for certain problems — unstable fixed points are prevalent in ER dynamics.

The following open questions are connected to these results.

Other stability criteria. Further developement of fixed point stability and instability criteria may be of significant importance for phase retrieval.

For example, one can attempt to extend existing criteria to other additional constraints such as sparsity. This case poses considerable challenges due to lack of regularity in the sparsity projection.

As for the modulus constraint, it may be benefitial to change coordinate systems or to develop appropriate restrictions of tangent spaces.

A very ambitious question one could ask is the following: provided a solution g, how great is the distance between

g or associated solution (like a translate of g)

and

the closest (non-trivially distinct from g) stable fixed point of $E_{\mathcal{A}} + E_{\mathcal{M}}$

for given $\sqrt{I}$ and $\mathcal{A}$? The answer to this question could explain why — for some phase retrieval problems — numerically observed convergence radius of ER greatly exceeds the theoretically accessible one.

Adaptive reconstruction algorithms. Ideas of adaptively restricting additional constraints are known in literature. For example, in [YAY14] random restriction of the support in object space is used to reconstruct sparse signals.

In context of our work, convergence of ERF in an energy landscape with many saddle points can justify analysis of adaptive algorithm variants.

For example, let $\sqrt{I}_\alpha := P_{\mathcal{J}_a(\alpha)}[\sqrt{I}]$ denote the data subjected to amplitude thresholding. Assume that the maximum of $\sqrt{I}$ is attained only at the origin (which is reasonable for $\sqrt{I}$ corresponding to non-negative problems). Then, $\sqrt{I}_{\|\sqrt{I}\|_\infty}(k) = \delta_0(k)\sqrt{I}(k)$ and $\sqrt{I}_0(k) = \sqrt{I}(k)$. Thus, one can run ER (or ERF) with data $\sqrt{I}_\alpha$, starting with $\alpha = \|\sqrt{I}\|_\infty$ and adaptively decreasing α to 0. Effectively, the algorithm first finds positions of waves that contribute to data the most, and then adaptively refines the features.

In simulations, we observed such algorithms to converge to a fixed saddle point. This approach has shown itself superior to classical ER, but inferiour to DR, and not directly compatible with sparsity constraints.

The setting of ERF could provide theoretical framework to investigate adaptive algorithms more closely. More refined results on fixed point stability can yield insight on a suitable design of adaptive algorithms.

Difficulty criteria for phase retrieval. While not directly addressed in this work, the question of gauging the difficulty of a phase retrieval instance is very intriguing. The question is: given data $\sqrt{I}$, how difficult is it to construct a solution?

The paper [ELB18] uses sparsity $\|\sqrt{I}\|_4/\|\sqrt{I}\|_2$ — the smaller this quantity is, the harder is phase retrieval — in conjunction with additional criteria (number of atoms in the molecule) to quantify difficulty of phase retrieval instances (assuming support size and atomicity constraints, the latter imposing a restriction on minimal distance between atoms).

One can ask whether these criteria can be reformulated to a more theoretically accessible form. It is obvious that sparsity of the data alone is not sufficient, and that additional criteria must play an important role. Indeed, considering finite spaces for simplicity, one can see that $\|\sqrt{I}\|_4/\|\sqrt{I}\|_2$ is smallest for $\sqrt{I} \equiv \text{const}$. However, this data corresponds to a delta peak object and is easily reconstructed by the Douglas-Rachford algoritm.

In context of our work, one can ask the following: is there a correlation between the difficulty of a phase retrieval instance and the energy functional $E_{\mathcal{P}} + E_{\mathcal{M}}$?

11.4 DOUGLAS-RACHFORD FLOW

As established in Chapter 9, Alternating Projections, Dykstra, and Douglas-Rachford algorithms can be connected through variation of the functional

$$F[s,d] := \frac{1}{2}E_{\mathcal{X}}[s+d] + \frac{1}{2}E_{\mathcal{Y}}[s-d] - \frac{1}{2}\|d\|_2^2,$$

and for the resulting Douglas-Rachford variant, one can use the energy $E_{\mathcal{X}}[s] + E_{\mathcal{Y}}[s]$ as a termination criterion.

Variation of $F[s,d]$ leads to different equations, such as the Douglas-Rachford Flow

$$\begin{aligned} \partial_t s &= -s + \frac{P_{\mathcal{A}}[s+d] + P_{\mathcal{M}}[s-d]}{2} \\ \partial_t d &= -\frac{P_{\mathcal{A}}[s+d] + P_{\mathcal{M}}[s-d]}{2} \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} \partial_t p &= -\frac{p+q}{2} + P_{\mathcal{A}}[p] \\ \partial_t q &= -\frac{p+q}{2} + P_{\mathcal{M}}[q], \end{aligned}$$

where $s = \frac{p+q}{2}$, $d = \frac{p-q}{2}$. This system of equations admits global weak solutions, albeit the solutions are allowed to grow in time (which is not observed in practice).

A natural question is whether one can establish boundedness for solutions of the Douglas-Rachford flow. More involved questions are, for example, i) whether one can show local convergence of the solutions, or ii) whether one can establish existence of periodic orbits of the flow.

APPENDIX

A

FOURIER ANALYSIS

This chapter briefly recalls some relevant results from Fourier analysis, which can be found in almost every book on the topic (e. g., [Fri07], [Str03] or [RS75]).

A.1 FOURIER TRANSFORM ON INFINITE-DIMENSIONAL DOMAINS

Throughout this section, $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d\}$, and $\Omega_F = \mathbb{R}^d$ if $\Omega = \mathbb{R}^d$, and $\Omega_F = \mathbb{Z}^d$ if $\Omega = \mathbb{T}^d$. We use on equal footing the notation $\mathcal{F}[\cdot] = \hat{\cdot}$ to denote the Fourier transform, and the notation $\mathcal{F}^{-1}[\cdot] = \check{\cdot}$ to denote the inverse Fourier transform.

Definition A.1 (Fourier transform on L^1). *For $f \in L^1(\Omega;\mathbb{C})$, its Fourier transform $\hat{f}\colon \Omega_F \to \mathbb{C}$ is defined as*

$$\hat{f}(k) = \int_\Omega f(x) e^{-ik\cdot x} dx,$$

The following lemma uses the convolution of two functions, defined as

$$\begin{aligned} *\colon L^1(\Omega;\mathbb{C}) \times L^1(\Omega;\mathbb{C}) &\to L^1(\Omega;\mathbb{C}) \\ f * g(y) &= \int f(x) g(y-x)\, dx \quad \text{for a. a. } y \in \Omega, \end{aligned}$$

where $f * g \in L^1(\Omega;\mathbb{C})$ by Fubini's theorem. We also use the notation $f_-(x) := f(-x)$.

Lemma A.2 (Basic properties). *Let $f, g \in L^1(\Omega;\mathbb{C})$. Then,*

i) $\mathcal{F}[fg] = (2\pi)^{-d}\mathcal{F}[f] * \mathcal{F}[g]$, and $\mathcal{F}[f * g] = (2\pi)^{-d}\mathcal{F}[f]\mathcal{F}[g]$;

ii) $\widehat{f_-^*} = (\hat{f})^*$.

iii) If f is continuously differentiable and $\partial_{x_j} f(x) \in L^1(\Omega;\mathbb{C})$, then $\widehat{\partial_{x_j} f}(k) = ik_j \hat{f}(k)$ for a. a.$k \in \Omega_F$.

iv) For any $a \in \Omega$, $\widehat{f(\cdot + a)}(k) = e^{ik\cdot a}\hat{f}(k)$ for a. a. $k \in \Omega_F$.

Theorem A.3 (Fourier transform on L^2, Plancherel). *There exists a unique continuous map $\mathcal{F}\colon L^2(\Omega;\mathbb{C}) \to L^2(\Omega_F;\mathbb{C})$ that agrees with the Fourier transform on $L^1(\Omega;\mathbb{C}) \cap L^2(\Omega;\mathbb{C})$.*

Moreover, the rescaled version $(2\pi)^{-d/2}\mathcal{F}$ is an isometric isomorphism, i. e. for all $f, g \in L^2(\Omega;\mathbb{C})$ holds

$$\langle 2\pi^{-d/2}\mathcal{F}[f], 2\pi^{-d/2}\mathcal{F}[g]\rangle_{L^2(\Omega_F)} = \langle f, g\rangle_{L^2(\Omega)}.$$

In particular, the Plancherel identity holds: for all $f \in L^2(\Omega;\mathbb{C})$,

$$\|f\|^2_{L^2(\Omega)} = \frac{1}{(2\pi)^d}\|\hat{f}\|^2_{L^2(\Omega_F)}.$$

The inverse operator $\mathcal{F}^{-1}\colon L^2(\Omega_F;\mathbb{C}) \to L^2(\Omega;\mathbb{C})$ is called the inverse Fourier transform on $L^2(\Omega_F;\mathbb{C})$ and satisfies

$$\mathcal{F}^{-1}[f](x) = \frac{1}{(2\pi)^d}\int_{\Omega_F} e^{ik\cdot x}\hat{f}(k)\,dk.$$

It is easy to verify that properties ii)–iv) of Lemma A.2 transfer to $f, g \in L^2(\Omega;\mathbb{C})$.

Corollary A.4. *A function $f \in L^2(\Omega;\mathbb{C})$ is real-valued if and only if $\hat{f}(-k) = \hat{f}^*(k)$ for a. a.$k \in \Omega_F$.*

Proof. "⇒" If f is real-valued,

$$\hat{f}(-k) = \int_\Omega f(x)e^{ik\cdot x}dx = \left(\int_\Omega f(x)e^{-ik\cdot x}dx\right)^* = \hat{f}^*(k)$$

for a. a. $k \in \Omega_F$.

"⇒" If $\hat{f}(-k) = \hat{f}^*(k)$ for a. a.$k \in \Omega_F$,

$$\begin{aligned} f(x) - f^*(x) &= \frac{1}{(2\pi)^d}\int \hat{f}(k)e^{ik\cdot x} - \frac{1}{(2\pi)^d}\int \hat{f}^*(k)e^{-ik\cdot x} \\ &= \frac{1}{(2\pi)^d}\int \hat{f}(k)e^{ik\cdot x} - \frac{1}{(2\pi)^d}\int \hat{f}(-k)e^{-ik\cdot x} \\ 0 \end{aligned}$$

using the substitution $\tilde{k} = -k$ in the second integral and exploiting the symmetry of the integration domain. □

In particular, if f is real-valued and $\hat{f} = \sqrt{I}e^{i\varphi}$ for some non-negative function $\sqrt{I} \in L^2(\Omega_F;\mathbb{R}_{\geqslant 0})$ and phase $\varphi\colon \Omega_F \to [0, 2\pi)$, then $\sqrt{I}$ is even, and $e^{i\varphi(-k)} = e^{-i\varphi(k)}$ for a. a.$k \in \Omega_F$. I. e., $\cos\varphi$ is even and $\sin\varphi$ is odd.

Also, if f is real-valued and even, then $\hat{f}$ is real-valued and even.

A.2 FINITE-DIMENSIONAL FOURIER TRANSFORM

For $\Omega = \mathbb{T}^d_N$ and $\Omega_F = \mathbb{T}^d_N$, $N = (N_1, \dots, N_d)$, the spaces $L^2(\Omega, \mathbb{C})$ and $L^2(\Omega_F, \mathbb{C})$ are isomorph to $\mathbb{C}^{N_1\times\dots\times N_d}$.

In this setting, it is common to define the Fourier transform as follows.

Definition A.5 (Finite-dimensional Fourier transform). *For $f \in \mathbb{C}^{N_1\times\cdots\times N_d}$, its Fourier transform $\hat{f} \in \mathbb{C}^{N_1\times\cdots\times N_d}$ is defined as*

$$\hat{f}(k_1, \dots k_d) = \sum_{x_1=1}^{N_1}\cdots\sum_{x_d=1}^{N_d} f(x_1, \dots, x_d)\prod_{j=1}^{d} e^{-2\pi i(k_j-1)(x_j-1)/N_j},$$

where x_j *and* k_j *belong to* $\{1, \ldots N_j\}$ *for* $j \in \{1, \ldots, d\}$.

In this setting, properties of Lemma A.2 remain valid with necessary modifications, and Plancherel's theorem reads

$$\sum_{x_1=1}^{N_1} \cdots \sum_{x_d=1}^{N_d} |f(x_1, \ldots, x_d)|^2 = \frac{1}{|N|} \sum_{k_1=1}^{N_1} \cdots \sum_{k_d=1}^{N_d} |\hat{f}(k_1, \ldots, k_d)|^2,$$

where $|N| = N_1 \cdot \ldots \cdot N_d$.

B A NON-UNIQUENESS EXAMPLE

Phase retrieval is notoriously ill-posed, especially in the infinite-dimensional setting; see [GKR20] for a recent review on the topic.

This chapter shows a well-known example from [BS79], further studied in [Hay82], of non-trivial ambiguities in phase retrieval. The example is based on the fact that for two non-negative functions $f, g \in \mathcal{H}$, the convolutions $f * g_-$ and $f * g$ are non-negative and have the same Fourier transform modulus $|\hat{f}\hat{g}|$.

Remark B.1 (Trivial ambiguities). It is straightforward to check that for any $g \in \mathcal{H}$, its translation $g(\cdot + x_0)$ for $x_0 \in \Omega$, and its reflection $g_-(x) := g(-x)$ for a. a. $x \in \Omega$ have the same Fourier modulus as g.

These ambiguities are commonly called trivial ambiguities of phase retrieval.

Remark B.2 (Local non-uniqueness). In certain settings phase retrieval may exhibit local non-uniqueness. For example, let $\Omega = \mathbb{T}$, let $\varphi \in [0, 2\pi)$, let $\alpha \in (0, 1)$, define $\sqrt{I} \colon \mathbb{Z}^d \to \mathbb{R}$ using

$$\sqrt{I}(k) = \delta_0(k) + \frac{\alpha}{2}\left(\delta_1(k)e^{i\varphi} + \delta_{-1}(k)e^{-i\varphi}\right),$$

δ being the Kronecker delta. Then,

$$\check{\sqrt{I}}(x) = \frac{1}{2\pi}\left(1 + \alpha\cos(x + \varphi)\right),$$

which is non-negative for any phase $\varphi \in [0, 2\pi)$.

In general, if g is a solution of some phase problem with non-negativity, with $g \geqslant \varepsilon$, then the phase of any Fourier pixel with $\sqrt{I} < \varepsilon/2$ is ambiguous.

One way to address this ambiguity is to restrict the support of the solution. For finite-dimensional phase problem, this corresponds to oversampling in Fourier space. A more detailed discussion on this matter can be found, e. g., in [Bar+20] and references therein.

Example B.3 (Non-local non-uniqueness). Let $f, g \in L^2(\Omega)$ be non-negative. Then, convolutions $f * g$ and $f * g_-$ are non-negative and have the same Fourier modulus. Indeed, by Fourier calculus (see Appendix A) have

$$|\mathcal{F}[f * g]| = |\hat{f}\hat{g}| = |\hat{f}\hat{g}^*| = |\mathcal{F}[f * g_-]|.$$

Using this observation, it is possible to construct distinct solutions to the same phase retrieval problem. Let $\zeta \in C_c^\infty(\mathbb{R})$, let

$$f(x) = g(x) = (2\delta_0 + \delta_1) * \zeta(x) = 2\zeta(x) - \zeta(x-1),$$

where δ is the Dirac delta distribution. Then,

$$\begin{aligned} f * g(x) &= \mathcal{F}^{-1}\left((2e^{ik0} + e^{ik1})(2e^{ik0} + e^{ik1})|\hat{\zeta}|^2\right) \\ &= \mathcal{F}^{-1}\left((4e^{ik0} + 4e^{ik1} + e^{ik2})|\hat{\zeta}|^2\right) \\ &= (4\delta_0 + 4\delta_1 + 1\delta_2)\mathcal{F}^{-1}\left(|\hat{\zeta}|^2\right). \end{aligned}$$

Analogously,

$$\begin{aligned} f * g_-(x) &= \mathcal{F}^{-1}\left((2e^{ik0} + e^{ik1})(e^{ik0} + 2e^{ik1})|\hat{\zeta}|^2\right) \\ &= \mathcal{F}^{-1}\left((2e^{ik0} + 5e^{ik1} + 2e^{ik2})|\hat{\zeta}|^2\right) \\ &= (2\delta_0 + 5\delta_1 + 2\delta_2)\mathcal{F}^{-1}\left(|\hat{\zeta}|^2\right) \neq f * g(x). \end{aligned}$$

Figure B.1 illustrates a variant of this example on a two-dimensional torus.

The presented non-uniqueness can be adapted to other domains, function spaces and constraints. For example, to illustrate non-uniqueness for the support size constraint, one can choose $\mathcal{F}^{-1}\left(|\hat{\zeta}|^2\right) = \mathbb{1}_{B_r(0)}$ for $r < \frac{1}{2}$. Then both $f * g$ and $f * g_-$ satisfy the support size constraint with support size $9\lambda(B_r(0))$.

Remark B.4. The previos non-local non-uniqueness example appears in many phase retrieval formulations. In the finite setting, it can be eliminated with sufficient oversampling [Hay82]. In the infinite-dimensional setting, it can be eliminated using generalized, rather than Fourier, measurements [Ala+19].

As argued in [Luk17], for feasibility formulations of phase retrieval — find an element in $\mathcal{M} \cap \mathcal{A}$ for some $\mathcal{A} \in \mathcal{H}$ — the question of uniqueness is far less relevant than the question whether $\mathcal{M} \cap \mathcal{A}$ is empty or not. This reasoning may be applied to the minimization of the energy $E_\mathcal{M} + E_\mathcal{A}$ as well.

In the energy minimization formulation, one could investigate not whether a global minimizer is unique, but whether it is possible to estimate the attraction region near global minimizers.

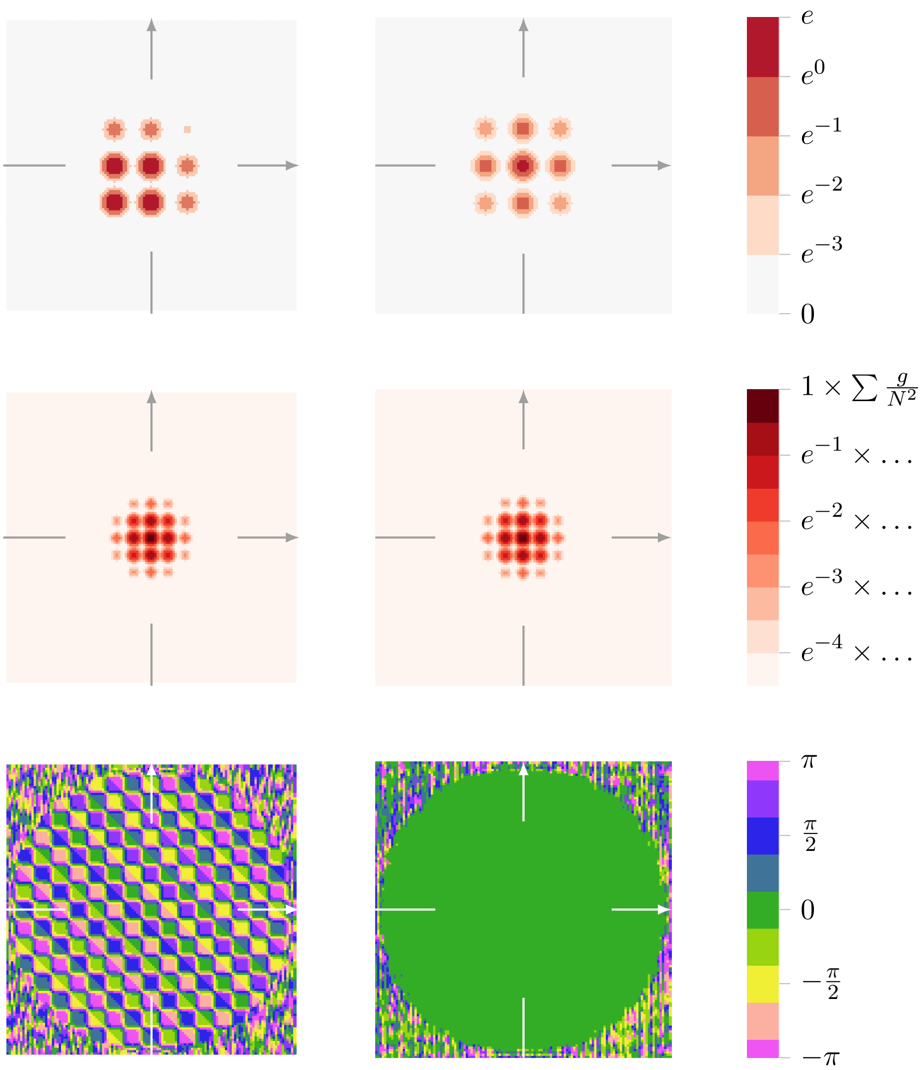

Figure B.1: Non-uniqueness of solutions (Example B.3)

Consider the domain $\mathbb{T}^2_{[-1,1];127}$, let $a = 1/2$ (amplitude factor), $b = 16/63$ (shift), $A = 125$ (gaussian width), let

$$f_1(x) = g_1(x) = a(2\delta_0(x) + \delta_b(x)) * \exp(-A|x|^2);$$
$$f(x,y) = g(x,y) = f_1(x)f_1(y).$$

Top left: $\delta_{(-b,-b)} * f * g$. Top right: $\delta_{(-b,-b)} * f * g_-$. Center left: $|\widehat{f * g}|$. Center right: $|\widehat{f * g_-}|$. Bottom left: $\arg(\widehat{f * g})$. Bottom right: $\arg(\widehat{f * g_-})$. Graphs of phases are modulated by $e^{ix_0 \cdot}$, where the shift x_0 is approximately equals to the barycenter of the corresponding function (to minimize visual cluttering due to phase oscillations).

The function $\delta_{(-b,-b)} * f * g_-$ is even; thus, its Fourier transform is real-valued. In the bottom row, at the edges $|\widehat{f * g}|$ and $|\widehat{f * g_-}|$ become comparable with machine precision and their phases become random.

C

SOME STANDARD RESULTS ON GENERALIZED DIFFERENTIATION

This chapter recalls the following standard results on Clarke and generalized subdifferentials that are used in the thesis:

- If a functional is Fréchet-differentiable, then Clarke and generalized subdifferentials contain only the corresponding Fréchet-derivative.
- For a sum of two functionals one of which is Fréchet-differentiable, subdifferentials of the sum equal to the sum of subdifferentials.

Corollary C.1 (Clark subdiff. of cts. Fréchet-diff.). *Let $E\colon \mathcal{H} \to \mathbb{R}$ be continuously Fréchet-differentiable at $g_* \in \mathcal{H}$ with Fréchet-derivative $\nabla E[g_*]$. Then, $\widehat{dE}[g_*][w] = \langle \nabla E[g_*], w \rangle$ for all $w \in \mathcal{H}$, and $\overline{\partial} E[g_*] = \{\nabla E[g_*]\}$.*

Proof. By definition,

$$\begin{aligned}
\widehat{dE}[g_*][w] &= \limsup_{g \to g_*, \varepsilon \searrow 0} \frac{E[g + \varepsilon w] - E[g]}{\varepsilon} \\
&= \limsup_{g \to g_*, \varepsilon \searrow 0} \frac{\varepsilon \langle \nabla E[g], w \rangle + o(\varepsilon)}{\varepsilon} \\
&= \limsup_{g \to g_*} \langle \nabla E[g], w \rangle = \langle \nabla E[g_*], w \rangle,
\end{aligned}$$

where we used that ∇E is continuous in the last step.

Therefore, for $v = \nabla E[g]$ have

$$\langle v, w \rangle = \langle \nabla E[g], w \rangle = \widehat{E}[g][w],$$

and $\nabla E[g] \in \overline{\partial} E[g]$.

Let $v \neq \nabla E[g]$. Then, for $w = v - \nabla E[g]$ have

$$\langle v - \nabla E[g], w \rangle = \| v - \nabla E[g] \|_2^2 > 0,$$

implying

$$\langle v, w \rangle \nleq \langle \nabla E[g], w \rangle \text{ for all } w \in \mathcal{H},$$

so that $\nabla E[g]$ is the only element contained in $\overline{E}[g]$. □

Lemma C.2 (Additivity if one summand is cts. Fréchet-diff.). *Let $E, F\colon \mathcal{H} \to \mathbb{R}$. Let E be Lipschitz continuous at g_*, let F be continuously Fréchet-differentiable at g_* for some $g_* \in \mathcal{H}$. Then,*

$$\overline{\partial}(E + F)[g_*] = \overline{\partial} E[g_*] + \overline{\partial} F[g_*] \tag{C.1}$$

Proof. First, observe that for all $w \in \mathcal{H}$

$$\widehat{d}(E+F)[g_*][w] = \widehat{d}E[g_*][w] + \widehat{d}F[g_*][w]. \tag{C.2}$$

Indeed, since

$$\lim_{g_n \to g, \varepsilon_n \searrow 0} \frac{F[g_n + \varepsilon_n w] - F[g_n]}{\varepsilon_n} = \langle \nabla F, w \rangle \text{ for all } w \in \mathcal{H}$$

and for all sequences $(g_n)_n, (\varepsilon_n)_n$ with $g_n \to g_*, \varepsilon_n \searrow 0$ as $n \to \infty$, it is also true for the sequence pair (g_n), (ε_n) that attains the supremum in $\widehat{d}E[g_*][w]$. Therefore, for such (g_n), (ε_n) have:

$$\begin{aligned} \widehat{d}(E+F)[g_*][w] &\geqslant \lim_{n\to\infty} \frac{E[g_n + \varepsilon w] - E[g_n] + F[g_n + \varepsilon w] - F[g_n]}{\varepsilon} \\ &= \limsup_{g \to g_*, \varepsilon \searrow 0} \frac{E[g+\varepsilon w] - E[g]}{\varepsilon} + \langle \nabla F[g_*], w \rangle \\ &= \widehat{d}E[g_*][w] + \widehat{d}F[g_*][w]. \end{aligned} \tag{C.3}$$

Further, subadditivity of lim sup immediatly implies

$$\widehat{d}(E+F)[g_*][w] \leqslant \widehat{d}E[g_*][w] + \widehat{d}F[g_*][w],$$

establishing (C.2).

Second, let us show the desired result (C.1). For "$\subseteq$", if

$$\langle v, w \rangle \leqslant \widehat{d}(E+F)[g_*][w] = \widehat{d}E[g_*][w] + \langle \nabla F[g_*], w \rangle \tag{C.4}$$

for all $w \in \mathcal{H}$, then for $v_E = v - \nabla F[g_*]$ and $v_F = \nabla F[g_*]$ have

$$\langle v_E, w \rangle \leqslant \widehat{d}E[g_*][w] \text{ and } \langle v_F, w \rangle \leqslant F[g_*][w] \tag{C.5}$$

For "$\supseteq$", the proof is similar and straightforward by (C.2). □

D

SUPPLEMENTAL CALCULATIONS

The following property of complex numbers was used in the thesis.

Lemma D.1. *Let* $a, b \in \mathbb{C}$*, let* $b \neq 0$*. Then,*

$$\left|a - \frac{b}{|b|}|a|\right| \leqslant 2|a-b|.$$

Proof. Distinguish between two cases. First, assume that $|a-b| < |a|$. If $|b| = |a|$, $\left|a - \frac{b}{|b|}|a|\right| = |a-b| \leqslant 2|a-b|$. If $|b| > |a|$, the angle formed by the points $a, \frac{b}{|b|}|a|, b$ is obtuse, and $\left|b - \frac{b}{|b|}|a|\right| \leqslant |a-b|$ by the cosine theorem. Therefore $\left|a - \frac{b}{|b|}|a|\right| \leqslant |a-b| + \left|b - \frac{b}{|b|}|a|\right| \leqslant 2|a-b|$. If $|b| < |a|$,

$$\begin{aligned}\left|a - \frac{b}{|b|}|a|\right| &\leqslant |a-b| + \left|b - \frac{b}{|b|}|a|\right| = |a-b| + \underbrace{\left|\frac{b}{|b|}|a|\right|}_{=|a|} - |b| \\ &\leqslant |a-b| + |a| - |b| \leqslant |a-b| + |a-b| = 2|a-b|.\end{aligned}$$

Second, assume that $|a-b| \geqslant |a|$. Then, $\left|a - \frac{b}{|b|}|a|\right| \leqslant 2|a| \leqslant 2|a-b|$. □

D.1 COMPACTNESS OF THE MODULUS CONSTRAINT

The modulus constraint set is compact on the torus. This can be demonstrated using a standard diagonalization argument.

Proposition D.2. *Let* $I \in \ell^1(\mathbb{Z}^d)$ *with* $I \geqslant 0$ *(so that* $\sqrt{I}$ *belongs to* $\ell^2(\mathbb{Z}^d)$*). Then the modulus set*

$$\mathcal{M}(\sqrt{I}) = \{f \in L^2(\mathbb{T}^d) \mid |\hat{f}| = \sqrt{I}\}$$

is strongly sequentially compact (and thus strongly compact).

Proof. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}(\sqrt{I})$.

For readability, we split the proof into the following steps:

1) For any sequence $(k_{\tilde{m}})_{\tilde{m}\in\mathbb{N}}$ that counts all elements of $\mathbb{Z}^d$; use Z_m to denote the set that contains first m^d elements of the sequence. (The choice of m^d first elements makes it easier to pick a sequence (k_m) for which Z_m contains growing d-dimensional cubes centered at the origin.)

2a) Use induction to extract subsequences $(\hat{f}_n)_{n\in M_m}$, $m \in \mathbb{N}$, from $(\hat{f}_n)_{n\in\mathbb{N}}$ that converge pointwise on Z_m. Specifically, for any $m \in \mathbb{N}$, extract subsequences such that

$$\mathbb{N} = M_0 \supset M_1 \supset M_2 \supset \ldots \supset M_m \supset \ldots,$$

such that $(\hat{f}_n)_{n\in M_m}$ converges to some $\hat{f}^{(m)}$ pointwise on Z_m, and such that $f^{(m)} \in \mathcal{M}(\sqrt{I})$.

2b) Show that $(f^{(m)})_{m\in\mathbb{N}}$ is Cauchy in $L^2(\mathbb{T}^d)$; follow that it converges to some $f \in L^2(\mathbb{T}^d)$.

3) Show that $f \in \mathcal{M}(\sqrt{I})$.

4) Use a diagonal argument to construct a subsequence of $(f_n)_n$ that converges to f, concluding the proof.

Let us demonstrate these steps. 1) Let $(k_m)_{m\in\mathbb{N}}$ be a sequence that counts all $k \in \mathbb{Z}^d$ (i.e. for any $k \in \mathbb{Z}^d$ there exists exactly one $m \in \mathbb{N}$ with $k_m = k$) such that

$$\begin{aligned}
Z_1 &:= \bigcup_{\tilde{m}=1}^{1^d} \{k_{\tilde{m}}\} &&= \bigtimes_{i=1}^{d} \{0\} \\
Z_3 &:= \bigcup_{\tilde{m}=1}^{3^d} \{k_{\tilde{m}}\} &&= \bigtimes_{i=1}^{d} \{-1, 0, 1\} \\
Z_5 &:= \bigcup_{\tilde{m}=1}^{5^d} \{k_{\tilde{m}}\} &&= \bigtimes_{i=1}^{d} \{-2, -1, 0, 1, 2\} \\
&\ldots &&\ldots \quad \ldots \\
Z_m &:= \bigcup_{\tilde{m}=1}^{m^d} \{k_{\tilde{m}}\} &&= \bigtimes_{i=1}^{d} \left\{-\frac{m-1}{2}, \ldots, \frac{m-1}{2}\right\}
\end{aligned}$$

for all odd $m \in \mathbb{N}$; informally, such a sequence consequently counts all the elements in each cube around the origin before proceeding to the next "layer". With this construction, $B_{\frac{N-1}{2}}(0) \subset Z_{N^d}$ for any odd $N \in \mathbb{N}$.

2a) Construct a sequence of functions $(f^{(m)})_{m\in\mathbb{N}}$ such that subsequences of $(f_n)_{n\in\mathbb{N}}$ converge pointwise to $f^{(m)} \in \mathcal{M}(\sqrt{I})$ on certain domains $Z_m \subset \mathbb{Z}^d$ that grow with m.

Induction statement: for all $m \in \mathbb{N}$, there exists an unbounded set $M_m \subset \mathbb{N}$ and a function $f^{(m)} \in \mathcal{M}(\sqrt{I})$ such that $\hat{f}_n(k)$ converges to $\hat{f}^{(m)}(k)$ for all $k \in Z_m$ as $n \to \infty$ in M_m, and $|\hat{f}^{(m)}(k)| = \sqrt{I}(k)$ for $k \in \mathbb{Z}^d \backslash Z_m$.

Base case: let $m = 1$, let $M_0 = \mathbb{N}$. Since $\{a \in \mathbb{C} \mid |a| = \sqrt{I}(k_1)\}$ is compact, the sequence $\left(\hat{f}_n(k_1)\right)_{n\in M_0}$ has a subsequence that converges to some $b \in \mathbb{C}$ with $|b| = \sqrt{I}(k_1)$. Define

$$\hat{f}^{(1)}\colon \mathbb{Z}^d \to \mathbb{C}; \qquad k \mapsto \begin{cases} b & \text{if } k = k_1; \\ \sqrt{I}(k) & \text{else.} \end{cases}$$

By construction, $f^{(1)} \in \mathcal{M}(\sqrt{I})$ is the desired function. In other words, there exists an unbounded set $M_1 \subset M_0$ and a function $f^{(1)} \in \mathcal{M}(\sqrt{I})$ such that $\hat{f}_n(k)$ converges to $f^{(1)}(k)$ for all $k \in Z_1 = \{k_1\}$ as $n \in M_1$ goes to infinity, while $\hat{f}^{(1)}(k) = \sqrt{I}(k)$ for $k \in \mathbb{Z}^d \backslash Z_1$.

Induction step: assume the statement holds for m, M_m and $\hat{f}^{(m)}$. Since $\{a \in \mathbb{C} \mid |a| = \sqrt{I}(k_{m+1})\}$ is compact, the sequence $\left(\hat{f}_n(k_{m+1})\right)_{n\in M_m}$ has a convergent subsequence, i.e. such that $\hat{f}_n(k_{m+1}) \to b \in \mathbb{C}$ — with $|b| = \sqrt{I}(k_{m+1})$ — as $n \to \infty$ for $n \in M_{m+1}$ for some unbounded $M_{m+1} \subset M_m$. Define

$$\hat{f}^{(m+1)}\colon \mathbb{Z}^d \to \mathbb{C}; \qquad k \mapsto \begin{cases} \hat{f}^{(m)}(k) & \text{if } k \in Z_m; \\ b & \text{if } k = k_{m+1}; \\ \sqrt{I}(k) & \text{else.} \end{cases}$$

Thus, for $m+1$, there there exists an unbounded sequence $M_{m+1} \subset \mathbb{N}$ and a function $f^{(m+1)} \in \mathcal{M}$ such that $\hat{f}_n(k)$ converges to $\hat{f}^{(m+1)}(k)$ for all $k \in Z_{m+1}$, and $\hat{f}^{(m+1)}(k) = \sqrt{I}(k)$ for $k \in \mathbb{Z}^d \backslash Z_{m+1}$.

2b) The sequence $(\hat{f}^{(m)})_{m\in\mathbb{N}}$ is Cauchy in $l^2(\mathbb{Z}^d)$. Indeed, for an odd $N^d \in \mathbb{N}$ and any $n_1, n_2 \geqslant N^d$ have

$$\|\hat{f}^{(n_1)} - \hat{f}^{(n_2)}\|^2_{l^2(\mathbb{Z}^d)} \leqslant 2 \sum_{k\in\mathbb{Z}^d\backslash Z_{N^d}} \sqrt{I}(k)^2 \leqslant 2 \sum_{k\in\mathbb{Z}^d\backslash B_{\frac{N-1}{2}}(0)} \sqrt{I}(k)^2 \to 0$$

as $N \to \infty$, since $\sum_{k\in\mathbb{Z}^d} \sqrt{I}(k)^2 < \infty$ by assumption. Therefore, and because $l^2(\mathbb{Z}^d)$ is complete, $\hat{f}^{(m)} \to \hat{f}$ for some $\hat{f} \in l^2(\mathbb{Z}^d)$ as $m \to \infty$.

3) ($f \in \mathcal{M}$.) Further, $|\hat{f}(k)| = \sqrt{I}(k)$ for all $k \in Z^d$. Indeed, if we assume that $\left||\hat{f}(k)| - \sqrt{I}(k)\right| = \varepsilon > 0$ for some $k \in \mathbb{Z}^d$, then

$$\inf_{m\in\mathbb{N}} \left||\hat{f}(k)| - |\hat{f}^{(m)}(k)|\right| \geqslant \varepsilon$$

and $f^{(m)}$ could not converge to f in $L^2(\mathbb{T}^d)$, leading to a contradiction.

4) (Diagonal argument.)

Let $M \subset \mathbb{N}$ be the set constructed using m-th element of M_m for $m \in \mathbb{N}$. Then, M is unbounded, and $(f_n)_{n\in M}$ converges pointwise to $\hat{f}$. By dominated convergence theorem with majorant $\sqrt{I}^2$, $\hat{f}_n$ converges to $\hat{f}$ in l^2 as $n \to \infty$ for $n \in M$.

Therefore,

$$\|f_n - f\|^2_{L^2(\mathbb{T}^d)} \leqslant (2\pi)^d \|f_n - f\|^2_\infty \leqslant (2\pi)^{2d} \|\hat{f}_n - \hat{f}\|_{l^2(\mathbb{Z}^d)^2} \to 0$$

as $n \to \infty$ for $n \in M$, so $\mathcal{M}$ is sequentially compact. □

Corollary D.3. *Let $\Omega \subset \mathbb{R}^d$ bounded and Lebesgue-measurable, and let $I \in L^1(\mathbb{R}^d)$ with $I \geqslant 0$ (so that $\sqrt{I}$ belongs to $L^2(\mathbb{R}^d)$).*

Then the modulus set $\mathcal{M}(\sqrt{I}) = \{f \in L^2(\Omega) \mid |\hat{f}| = \sqrt{I}\}$ is strongly compact (where $\hat{f}$ is the Fourier transform of the extension of f by zero to all of $\mathbb{R}^d$).

Proof. W.l.o.g. (see remark at the very end of the proof), assume that $\Omega \subset \mathbb{T}^d$. Let

$$\mathcal{X}_{\mathbb{T}^d} = \{f \in L^2(\mathbb{T}^d) \mid \operatorname{supp} f \subseteq \Omega\}, \tag{D.1}$$

$$\mathcal{X}_{\mathbb{R}^d} = \{f \in L^2(\mathbb{R}^d) \mid \operatorname{supp} f \subseteq \Omega\}. \tag{D.2}$$

It is straightforward to verify that the embedding, which is defined as

$$\iota\colon \mathcal{X}_{\mathbb{T}^d} \to \mathcal{X}_{\mathbb{R}^d}, \tag{D.3}$$

$$f(x) \mapsto \begin{cases} f(x) & \text{if } x \in \Omega, \\ 0 & \text{else} \end{cases} \tag{D.4}$$

for almost all $x \in \mathbb{R}^d$, is a homeomorphism between $\left(\mathcal{X}_{\mathbb{T}^d}, \|\cdot\|_{L^2(\mathbb{T}^d)}\right)$ and $\left(\mathcal{X}_{\mathbb{R}^d}, \|\cdot\|_{L^2(\mathbb{R}^d)}\right)$.

Thus, if a set $\mathcal{Y} \subseteq \mathcal{X}_{\mathbb{R}^d}$ is not compact, then $\iota(\mathcal{Y}) \subseteq \mathcal{X}_{\mathbb{T}^d}$ is not compact.

Assume that $\mathcal{M}(\sqrt{I}) \subseteq \mathcal{X}_{\mathbb{R}^d}$ is not compact. Then, $\iota(fM(\sqrt{I})) = \mathcal{M}(\widetilde{\sqrt{I}}) \subseteq \mathcal{X}_{\mathbb{T}^d}$ is not compact, where — by a straightforward calculation — $\widetilde{\sqrt{I}} \in \ell^2(\mathbb{Z}^d)$ with $\widetilde{\sqrt{I}}(k) = |\hat{f}|(k)$ for any $f \in \mathcal{M}(\sqrt{I})$, and the Fourier transform of f is taken in $L^2(\mathbb{R}^d)$. However, $\mathcal{M}(\widetilde{\sqrt{I}})$ being not compact contradicts Proposition D.2, meaning that $\mathcal{M}(\sqrt{I}) \subseteq \mathcal{X}_{\mathbb{R}^d}$ is compact.

In the beginning of the proof, we assume that $\Omega \subseteq \mathbb{T}^d$ w.l.o.g. Indeed, since Ω is bounded by assumption, have $c := \sup_{x \in \Omega} \|x\|_{\mathbb{R}^d} < \infty$, such that the coordinate scaling

$$\begin{aligned} L^2(\mathbb{R}^d) &\to L^2(\mathbb{R}^d) \\ f(x) &\mapsto f(x/c) \end{aligned}$$

for all $x \in \mathbb{R}^d$ is a homeomorphism that maps any subset of $L^2(\Omega)$ onto a subset of $L^2(\tilde{\Omega})$ such that $\tilde{\Omega} \subseteq \mathbb{T}^d$, and the topological properties of both subsets coincide. □

D.2 COMPACTNESS OF THE HILBERT CUBE

Similarly to Proposition D.2, one can demonstrate (strong) compactness of the Hilbert cube.

First, recall the following standard results on integral and series convergence.

LEMMA D.4. *Let $d \in \mathbb{N}$, let $\|\cdot\|_{2;\mathbb{R}^d}$ denote the Euclidean norm of a d-dimensional vector. For $r > d$, for $R > 0$ have*

$$\int_{|x|>R} \frac{C}{1+\|x\|^r_{2;\mathbb{R}^d}} dx \leqslant \frac{C_r}{R^{r-d}}$$

for $C_r := 2^d C\lambda(B_1^{d-1}(0)) < \infty$.

Also, $\|\frac{C}{1+\|\cdot\|^r_{2;\mathbb{R}^d}}\|_2 < \infty$.

Proof. By a straightforward calculation,

$$\begin{aligned}\int_{|x|>R} \frac{C}{1+\|x\|^r_{2;\mathbb{R}^d}} dx \leqslant \int_{|x|>R} \frac{C}{\|x\|^r_{2;\mathbb{R}^d}} dx \\ = 2^d C \int_{z=R}^{\infty} \frac{1}{z^r} \lambda(B_z^{d-1}(0))\, dz 2^d \lambda(B_1^{d-1}(0)) \int_{z=R}^{\infty} \frac{z^{d-1}}{z^r}\, dz \\ = 2^d C\lambda(B_1^{d-1}(0)) \frac{1}{r-d} \left.\frac{1}{z^{r-d}}\right|_{z=R}^{\infty} \leqslant \frac{C_r}{R^{r-d}}\end{aligned}$$

for $C_r = 2^d C\lambda(B_1^{d-1}(0))$.

Further,

$$\int_{\mathbb{R}^d} \frac{C^2}{(1+\|x\|^r_{2;\mathbb{R}^d})^2} dx \leqslant \int_{B_1^d(0)} C^2 dx + \int_{|x|>1} \frac{C^2}{(1+\|x\|^r_{2;\mathbb{R}^d})^2} dx \leqslant$$
$$C^2\lambda(B_1^d(0)) + CC_R < \infty,$$

since $(1+\|x\|^r_{2;\mathbb{R}^d})^2 \geqslant (1+\|x\|^r_{2;\mathbb{R}^d})$ for all $x \in \mathbb{R}^d$. □

LEMMA D.5. *Let $d \in \mathbb{N}$, let $\|\cdot\|_{2;\mathbb{R}^d}$ denote the Euclidean norm of a d-dimensional vector. For $r > d$,*

$$\sum_{k\in\mathbb{Z}^d} \frac{1}{1+\|k\|^r_{2;\mathbb{R}^d}} < \infty.$$

Proof. The sum can be bound by an integral, which can be evaluated explicitly.

Define the function

$$\begin{aligned} f\colon \mathbb{R}^d &\to \mathbb{R}^d \\ q = (q_1, \ldots, q_d) &\mapsto f(q) \text{ with } f(q)_j = \left\lfloor q_j + \frac{1}{2} \right\rfloor,\ j \in \{1, \ldots, d\},\end{aligned}$$

using the floor function $\lfloor\cdot\rfloor$. Then, for any $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$ holds

$$\int_{[k_1-\frac{1}{2};k_1+\frac{1}{2})\times\ldots\times[k_d-\frac{1}{2};k_d+\frac{1}{2})} \frac{1}{\|f(q)\|^r_{2;\mathbb{R}^d}} dq_1\cdots dq_d$$
$$= \int_{[k_1-\frac{1}{2};k_1+\frac{1}{2})\times\ldots\times[k_d-\frac{1}{2};k_d+\frac{1}{2})} \frac{1}{\|k\|^r_{2;\mathbb{R}^d}} dq_1\cdots dq_d = \frac{1}{\|k\|^r_{2;\mathbb{R}^d}}.$$

Therefore,

$$\sum_{k\in\mathbb{Z}^d\setminus[-1;1]^d} \frac{1}{\|k\|^r_{2;\mathbb{R}^d}} = \int_{\left((-\infty;-\frac{3}{2}]\cup[\frac{3}{2};\infty)\right)^d} \frac{1}{\|f(q)\|^r_{2;\mathbb{R}^d}} dq. \tag{D.5}$$

Since for any $a \in \mathbb{R}$ holds $\left|\lfloor a+\frac{1}{2}\rfloor\right| \geqslant |a| - \frac{1}{2}$, we have

$$\|f(q)\|_{2;\mathbb{R}^d} = \left(\sum_{j=1}^d \left\lfloor q_j + \frac{1}{2}\right\rfloor^2\right)^{\frac{r}{2}} \geqslant \left(\sum_{j=1}^d \left(|q_j| - \frac{1}{2}\right)^2\right)^{\frac{r}{2}},$$

which, inserted into Equation (D.5), yields

$$\begin{aligned}
&\int_{\left((-\infty;-\frac{3}{2}]\cup[\frac{3}{2};\infty)\right)^d} 1\Big/\|f(q)\|^r_{2;\mathbb{R}^d} \, dq \\
\leqslant\ &\int_{\left((-\infty;-\frac{3}{2}]\cup[\frac{3}{2};\infty)\right)^d} 1\Big/\left(\textstyle\sum_{j=1}^d \left(|q_j| - \frac{1}{2}\right)^2\right)^{\frac{r}{2}} dq \\
=\ &2^d \int_{[\frac{3}{2};\infty)^d} 1\Big/\left(\textstyle\sum_{j=1}^d \left(q_j - \frac{1}{2}\right)^2\right)^{\frac{r}{2}} dq \\
=\ &2^d \int_{[1;\infty)^d} 1\Big/\left(\textstyle\sum_{j=1}^d q_j^2\right)^{\frac{r}{2}} dq \\
=\ &2^d \int_{[1;\infty)^d} 1\Big/\|q\|^r_{2;\mathbb{R}^d} \, dq.
\end{aligned}$$

Further, for the ball $B_1(0)$ holds $(1;\infty)^d \subseteq \mathbb{R}^d\setminus B_1(0)$, since for any $q \in B_1(0)$ holds $q_j \leqslant 1$ for all $j \in \{1, \ldots, d\}$. Therefore,

$$\begin{aligned}
2^d \int_{[1;\infty)^d} \frac{1}{\|q\|^r_{2;\mathbb{R}^d}} dq &\leqslant 2^d \int_{\mathbb{R}^d\setminus B_1(0)} \frac{1}{\|q\|^r_{2;\mathbb{R}^d}} dq \\
&= 2^d \int_{z=1}^{\infty} \frac{1}{z^r} \lambda(B_z^{d-1}(0))\, dz = 2^d \lambda(B_1^{d-1}(0)) \int_{z=1}^{\infty} \frac{1}{z^r} z^{d-1}\, dz \\
&= 2^d \lambda(B_1^{d-1}(0)) \frac{1}{(r-d)} < \infty
\end{aligned}$$

for $r - d > 0$.

Thus,

$$\sum_{k\in\mathbb{Z}^d\setminus[-1;1]^d} \frac{1}{\|k\|^r_{2;\mathbb{R}^d}} < \infty$$

for $r > d$, so that

$$\sum_{k\in\mathbb{Z}^d} \frac{1}{1+\|k\|_{2;\mathbb{R}^d}^r} \leqslant \sum_{k\in\mathbb{Z}^d\cap[-1;1]^d} 1 + \sum_{k\in\mathbb{Z}^d\setminus[-1;1]^d} \frac{1}{\|k\|_{2;\mathbb{R}^d}^r}$$
$$\leqslant 3^d + \sum_{k\in\mathbb{Z}^d\setminus[-1;1]^d} \frac{1}{\|k\|_{2;\mathbb{R}^d}^r} < \infty. \qquad \square$$

Recall the definition of a Hilbert cube from Example 6.39: for $\rho \in l^1(\mathbb{Z}^d;\mathbb{R}_{\geqslant 0})$,

$$\mathcal{Q}_\rho^{\mathrm{F}} = \left\{ f \in L^1(\mathbb{T}^d) \mid |\hat{f}(k)| \leqslant \rho \text{ for all } k \in \mathbb{Z}^d \right\}.$$

PROPOSITION D.6 (HILBERT CUBE IS COMPACT). *Let $\rho \in l^1(\mathbb{Z}^d;\mathbb{R}_{\geqslant 0})$. Then, $\mathcal{Q}_\rho^{\mathrm{F}}$ is compact in $L^p(\mathbb{T}^d)$ for all $p \in [1,\infty]$.*

Proof. The proof is extremely similar to the proof of Proposition D.2.

One minor difference is that instead of using the compactness of the circumference $\{z \in \mathbb{C} \mid |z| = a\}$ for $a \geqslant 0$, it is necessary to use the compactness of the circle $\{z \in \mathbb{C} \mid |z| \leqslant a\}$ for $a \geqslant 0$.

Another minor difference is that when the dominated convergence theorem is used, instead of picking $\sqrt{I}^2 \in \ell^1$ as the majorant, one picks the majorant $\rho \in \ell^1$ to follow the convergence. $\square$

COROLLARY D.7 (A COMPACT HILBERT CUBE). *Let $C > 0$, $r > d$. Then, $\mathcal{Q}_{C;r}^{\mathrm{F}}$ is compact in $L^p(\mathbb{T}^d)$ for all $p \in [1,\infty]$.*

Proof. Follows from Proposition D.6 for $\rho(k) = C/(1+\|k\|_{2;\mathbb{R}^d}^r)$, which belongs to $l^1(\mathbb{Z}^d;\mathbb{R}_{\geqslant 0})$ by Lemma D.5. $\square$

D.3 FORMAL SECOND DERIVATIVES OF ENERGIES

Remark D.8 (Formal derivatives of $E_{\mathcal{P}}$ and $E_{\mathcal{M}}$). Let $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d, \mathbb{T}_N^d\}$, , for $g, h, \chi \in L^2(\Omega)$. Observe the following formal calculation.

$$\begin{aligned}
&\left.\frac{d}{d\varepsilon_h}\right|_{\varepsilon_h=0} \left.\frac{d}{d\varepsilon_\chi}\right|_{\varepsilon_\chi=0} E_{\mathcal{P}}[g+\varepsilon_h h+\varepsilon_\chi \chi] \\
&\quad= \left.\frac{d}{d\varepsilon_h}\right|_{\varepsilon_h=0} \int (g(x)+\varepsilon_h h(x) - P_{\mathcal{P}}[g+\varepsilon_h h](x))\chi(x)\,dx \\
&\quad= \frac{1}{2}\left.\frac{d}{d\varepsilon_h}\right|_{\varepsilon_h=0} \int (g(x)+\varepsilon_h h(x) - |g(x)+\varepsilon_h h(x)|)\chi(x)\,dx \\
&\quad= \frac{1}{2}\left.\frac{d}{d\varepsilon_h}\right|_{\varepsilon_h=0} \int (g(x)+\varepsilon_h h(x) - |g|(x) - \frac{\varepsilon_h g(x)h(x)}{|g(x)|} + O(\varepsilon_h{}^2))\chi(x)\,dx \\
&\quad= \int\int \mathbb{1}_{g<0}(x)\delta(x-y)\chi(x)h(y)\,dx\,dy
\end{aligned}$$

in object space, and

$$= \int\int \hat{\chi}^*(k)\widehat{\mathbb{1}}_{g<0}(k-q)\hat{h}(q)\,dq\,dk$$

in Fourier space. Analogously,

$$\begin{aligned}
&\left.\frac{d}{d\varepsilon_h}\right|_{\varepsilon_h=0} \left.\frac{d}{d\varepsilon_\chi}\right|_{\varepsilon_\chi=0} E_{\mathcal{M}}[g+\varepsilon_h h+\varepsilon_\chi \chi] \\
&\quad = \left.\frac{d}{d\varepsilon_h}\right|_{\varepsilon_h=0} C_{\mathcal{F}} \int \left(\hat{g}+\varepsilon_h\hat{h}-\sqrt{I}\frac{\hat{g}+\varepsilon_h\hat{h}}{|\hat{g}+\varepsilon_h\hat{h}|}\right)^* \hat{\chi} \\
&\quad = C_{\mathcal{F}} \int \left(\hat{h}-\sqrt{I}\frac{\hat{h}}{|\hat{g}|}+\sqrt{I}\frac{\hat{g}}{|\hat{g}|^2}\frac{\hat{g}^*\hat{h}+\hat{g}\hat{h}^*}{2|\hat{g}|}\right)^* \hat{\chi} \\
&\quad = C_{\mathcal{F}} \int \left(1-\frac{\sqrt{I}}{2|\hat{g}|}\right)\hat{h}^*\hat{\chi} + C_{\mathcal{F}}\int \frac{\sqrt{I}}{2|\hat{g}|}\frac{(\hat{g}^*)^2}{|\hat{g}|^2}\hat{h}\hat{\chi}
\end{aligned}$$

— so far the integration variable has been omitted for readability; now it is introduced in order to simplify the expression. Recall that, for real-valued functions h, $\hat{h}(k)=\hat{h}^*(-k)$, and continue with —

$$\begin{aligned}
&= C_{\mathcal{F}} \int \underbrace{\left(1-\frac{\sqrt{I(k)}}{2|\hat{g}(k)|}\right)}_{=:A_{\mathrm{F}}(k)} \hat{h}(k)^*\hat{\chi}(k)\,dk + C_{\mathcal{F}} \int \underbrace{\frac{\sqrt{I(k)}}{2|\hat{g}(k)|}\frac{(\hat{g}(k)^*)^2}{|\hat{g}(k)|^2}}_{=:B_{\mathrm{F}}(k)} \hat{h}(-k)^*\hat{\chi}(k)\,dk \\
&= C_{\mathcal{F}} \int\int \big(A_{\mathrm{F}}(k)\delta(k-q)+B_{\mathrm{F}}(k)\delta(k+q)\big)\hat{h}(q)^*\hat{\chi}(k)\,dk\,dq
\end{aligned}$$

in Fourier space, and can be rewritten as

$$\begin{aligned}
&= \int \chi(x)h(x)\,dx - \int \chi(x)\Bigg(\underbrace{\mathcal{F}^{-1}\left(\frac{\sqrt{I}}{2|\hat{g}|}\right)}_{=:A(x)} * h\Bigg)(x)\,dx \\
&\qquad\qquad + \int \chi(x)\Bigg(\underbrace{\mathcal{F}^{-1}\left(\frac{\sqrt{I}}{2|\hat{g}|}\frac{(\hat{g}^*)^2}{|\hat{g}|^2}\right)}_{=:B(x)} * h\Bigg)(-x)\,dx \\
&= \int \chi(x)h(x)\,dx - \int \chi(x)\int A(x-y)h(y)\,dy\,dx \\
&\qquad\qquad + \int \chi(x)\int B(x-y)h(-y)\,dy\,dx \\
&= \int\int \big(1-A(x-y)+B(x+y)\big)\chi(x)h(y)\,dx\,dy
\end{aligned}$$

in object space. The inverse Fourier transforms in A, B are formal. For example, they are well-defined as inverse Fourier transforms if the underlying space is finite and $k \mapsto \frac{\sqrt{I(k)}}{|\hat{g}(k)|}$ is bounded.

E

DETAILS OF NUMERICAL EXAMPLES

This chapter contains more details on the setting used in Chapter 10 to generate numerical examples. It describes the exact form of the solution $g_{3G;N}$, and contains figures that show $g_{3G;N}$ for $N \in \{31, 127\}$, the corresponding Fourier moduli and Fourier phases.

It also describes the exact form of the function $g_{9G;N}$ that was used as initalization in Example 10.7 and Example 10.8.

For reference, this chapter contains figures illustrating some other example solutions. For these solutions, the observed behavior of reconstruction algorithms was comparable to the results presented in Chapter 10.

The densities were generated using the space discretization

$$\mathbb{T}^2_{[-1,1];N} := \left\{ (x,y) \in [-1,1]^2 \;\middle|\; \begin{aligned} x &= -1 + \frac{2i_x}{N-1} \text{ for } i_x \in [N-1]_0 \\ y &= -1 + \frac{2i_y}{N-1} \text{ for } i_y \in [N-1]_0 \end{aligned} \right\},$$

where $[N-1]_0 = \{0, 1, \ldots, N-1\}$.

The used problem was generated as a Gaussian sum of the following form.

Definition E.1 (Gaussian sum). *Let $N \in \mathbb{N}$ denote the number of discretization points along each dimension, let $N_G \in \mathbb{N}$ denote the number of Gaussians. Let $a \in \mathbb{R}^{N_G}$ with $a_i \geqslant 0$ for $i \in [N_G]$ describe maximal amplitudes (of Gaussians). Let $m \in (\mathbb{R}^2)^{N_G}$ describe means (of Gaussians), let $A \in (\mathbb{R}^{2\times 2})^{N_G}$ describe scaled inverse covariance matrices (of Gaussians).*

Define the Gaussian sum corresponding to a, m, A as

$$g_{a,m,A}(x) = \sum_{i=1}^{N_G} a_i \exp\big((x - m_i)^\intercal A_i (x - m_i)\big).$$

In a mild deviation from standard definition of a multivariate normal distribution, the factor $\frac{1}{2}$ is not present in the exponent.

The values of a, m, and A used in $g_{3G;N}$ are collected in Table E.1; values of a, m, and A used in $g_{3G;N}$ are collected in Table E.2. these tables, the values of A are written row-wise, i. e. $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is writtend down as $[a_{11} a_{12}; a_{21} a_{22}]$.

$i \in [3]$	a	m	A
1	1.28	[0.02 -0.23]	[104.83 -24.10 -24.10 171.17]
2	1.40	[-0.29 0.38]	[169.98 -15.83 -15.83 90.02]
3	0.90	[-0.38 0.14]	[150.70 43.15 43.15 79.30]

Table E.1: 3 Gaussians.
The entries of a are sampled from a uniform distribution on $[0.75, 1.5]$. The entries of m are sampled from a uniform distribution on the ball $B_{0.5}(0)$. The matrices A_i are generated as follows: i) construct a diagonal matrix $D_i \in \mathbb{R}^{2\times 2}$ from samples of a uniform distribution on $[5, 205]$; ii) rotate it by an angle sampled from a uniform distribution on $[0, 2\pi)$.
For a, m, A as in this table, we write $g_{3G} := g_{a,m,A}$. Further, we write $g_{3G;N} := g_{3G}|_{T^2_{[-1,1];N}}$ when g_{3G} is restricted to the discretized torus $T^2_{[-1,1];N}$.

Table E.2: 9 Gaussians

$i \in [9]$	a	m	A
1	1.48	[-0.41 -0.14]	[198.75 11.78; 11.78 139.25]
2	1.05	[0.42 -0.27]	[140.30 -32.92;-32.92 49.70]
3	1.29	[0.13 -0.30]	[173.25 17.73; 17.73 150.75]
4	0.85	[-0.08 -0.03]	[151.50 24.43; 24.43 128.50]
5	1.44	[0.02 0.41]	[78.94 23.12; 23.12 163.06]
6	0.79	[-0.22 0.45]	[72.46 -13.49;-13.49 121.54]
7	1.02	[0.31 0.20]	[183.13 -0.99; -0.99 190.87]
8	1.11	[0.05 0.12]	[121.49 23.95; 23.95 124.51]
9	1.07	[0.28 0.07]	[139.09 -1.50; -1.50 162.91]

The entries of a are sampled from a uniform distribution on $[0.75, 1.5]$. The entries of m are sampled from a uniform distribution on the ball $B_{0.5}(0)$. The matrices A_i are generated as follows: i) construct a diagonal matrix $D_i \in \mathbb{R}^{2\times 2}$ from samples of a uniform distribution on $[5, 205]$; ii) rotate it by an angle sampled from a uniform distribution on $[0, 2\pi)$.
For a, m, A as in this table, we write $g_{9G} := g_{a,m,A}$. Further, we write $g_{9G;N} := g_{9G}|_{T^2_{[-1,1];N}}$ when g_{9G} is restricted to the discretized torus $T^2_{[-1,1];N}$.

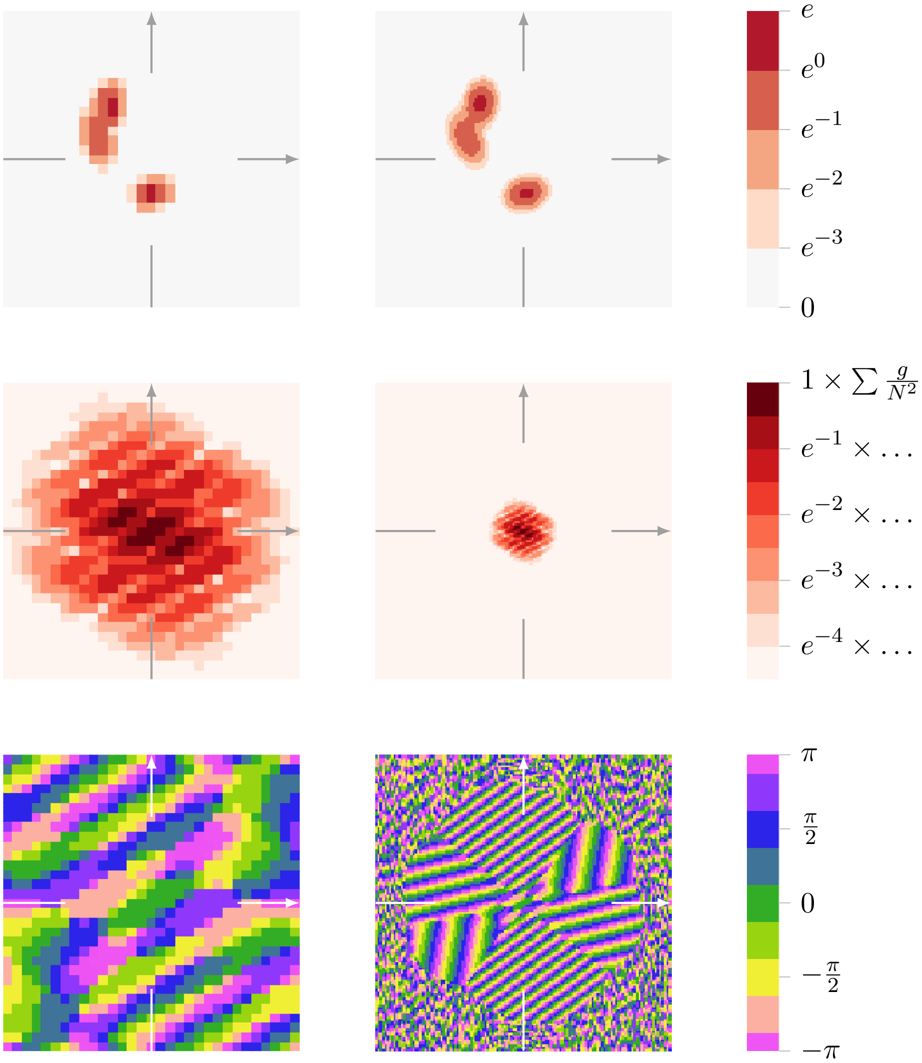

Figure E.1: Sum of 3 Gaussians from Table E.1
Top left: $g_{3G;31}$. Top right: $g_{3G;127}$. Center left: $|\hat{g}_{3G;31}|$. Center right: $|\hat{g}_{3G;127}|$. Bottom left: $\arg(\hat{g}_{3G;31})$. Bottom right: $\arg(\hat{g}_{3G;127})$. Graphs of phases are modulated by $e^{ix_0\cdot}$, where the shift x_0 is approximately equals to the barycenter of the corresponding function (to minimize visual cluttering due to phase oscillations).
The low-frequency content of $g_{3G;31}$ and $g_{3G;127}$ is similar; thus, center of $|\hat{g}_{3G;127}|$ looks very similar to $|\hat{g}_{3G;31}|$.
On the bottom right, at the edges values of $|\hat{g}_{3G;127}|$ are comparable with machine precision and $\arg(\hat{g}_{3G;127})$ looks like white noise.

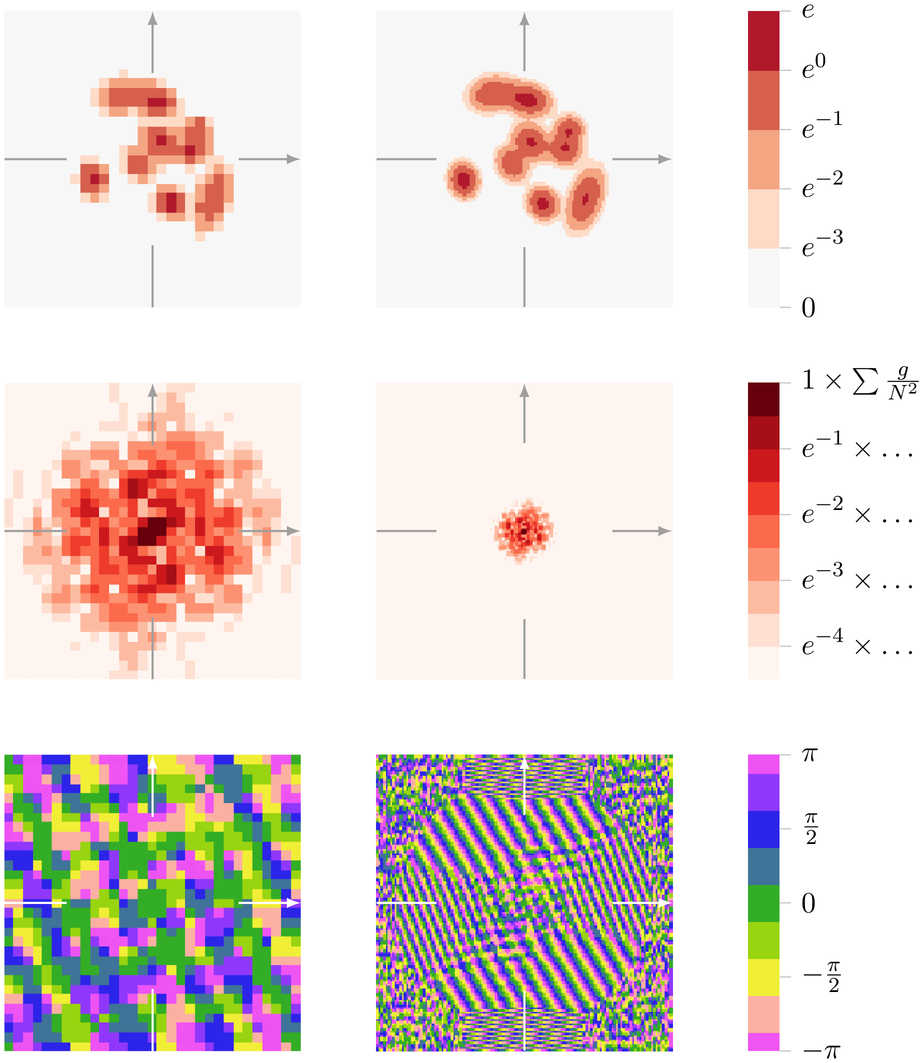

Figure E.2: Sum of 9 Gaussians from Table E.2

Top left: $g_{9G;31}$. Top right: $g_{9G;127}$. Center left: $|\hat{g}_{9G;31}|$. Center right: $|\hat{g}_{9G;127}|$. Bottom left: $\arg(\hat{g}_{9G;31})$. Bottom right: $\arg(\hat{g}_{9G;127})$. Graphs of phases are modulated by $e^{ix_0 \cdot}$, where the shift x_0 is approximately equals to the barycenter of the corresponding function (to minimize visual cluttering due to phase oscillations).

The low-frequency content of $g_{9G;31}$ and $g_{9G;127}$ is similar; thus, center of $|\hat{g}_{9G;127}|$ looks very similar to $|\hat{g}_{9G;31}|$.

On the bottom right, at the edges values of $|\hat{g}_{9G;127}|$ are comparable with machine precision and $\arg(\hat{g}_{9G;127})$ looks like white noise.

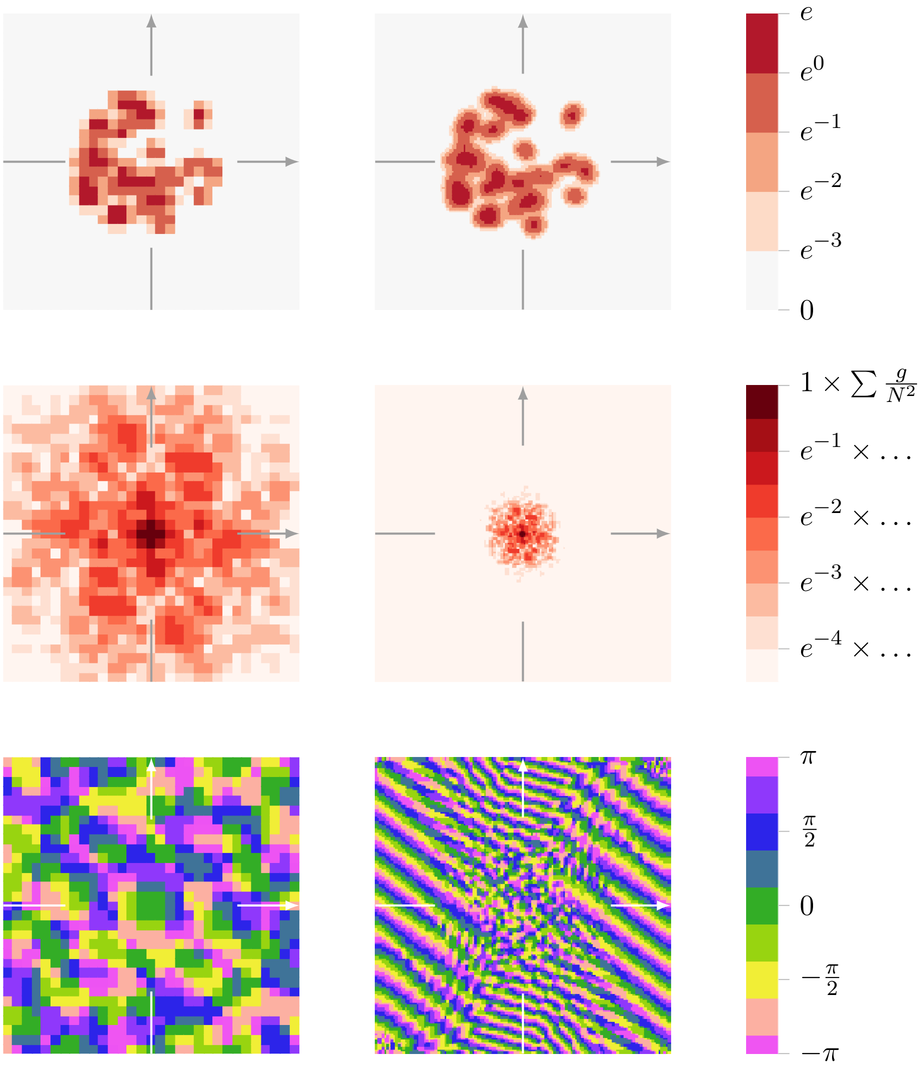

Figure E.3: Sum of 27 Gaussians

Top left: $g_{27G;31}$. Top right: $g_{27G;127}$. Center left: $|\hat{g}_{27G;31}|$. Center right: $|\hat{g}_{27G;127}|$. Bottom left: $\arg(\hat{g}_{27G;31})$. Bottom right: $\arg(\hat{g}_{27G;127})$. Graphs of phases are modulated by $e^{ix_0\cdot}$, where the shift x_0 is approximately equals to the barycenter of the corresponding function (to minimize visual cluttering due to phase oscillations).

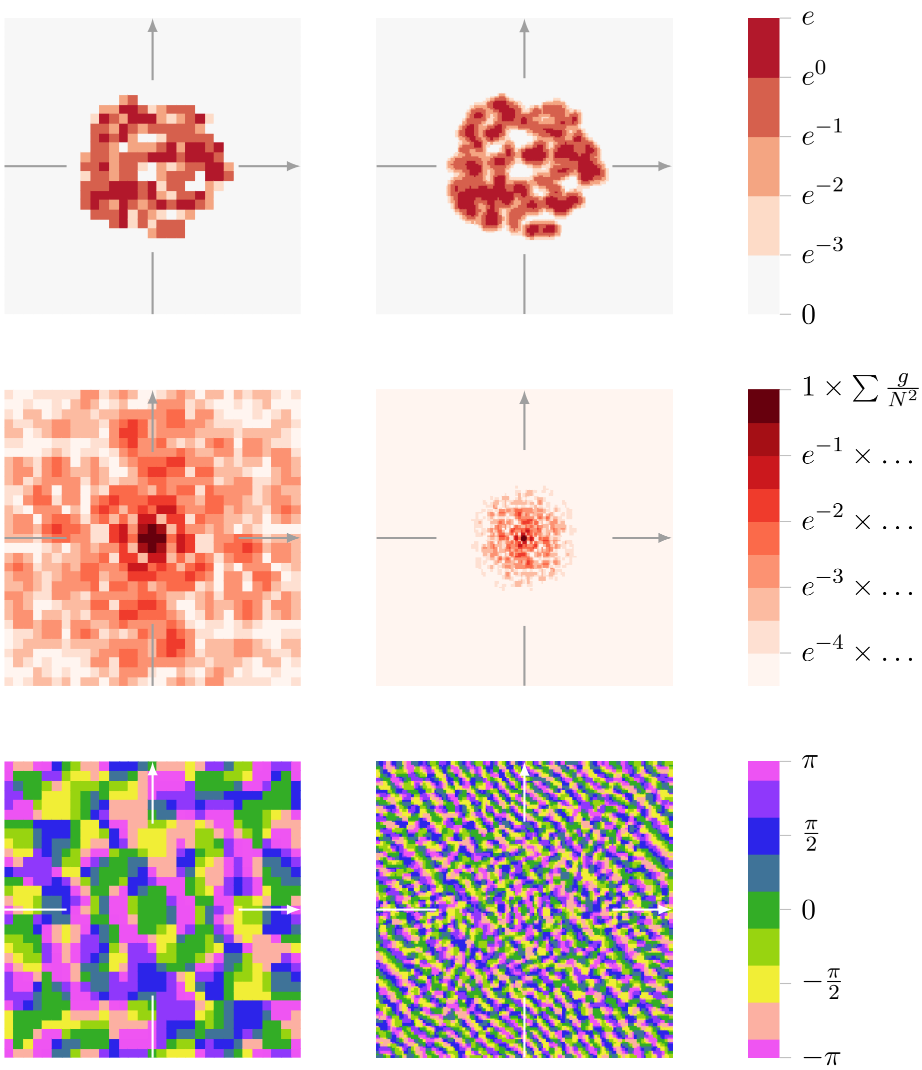

Figure E.4: Sum of 84 Gaussians
Top left: $g_{84G;31}$. Top right: $g_{84G;127}$. Center left: $|\hat{g}_{84G;31}|$. Center right: $|\hat{g}_{84G;127}|$. Bottom left: $\arg(\hat{g}_{84G;31})$. Bottom right: $\arg(\hat{g}_{84G;127})$. Graphs of phases are modulated by $e^{ix_0 \cdot}$, where the shift x_0 is approximately equals to the barycenter of the corresponding function (to minimize visual cluttering due to phase oscillations).

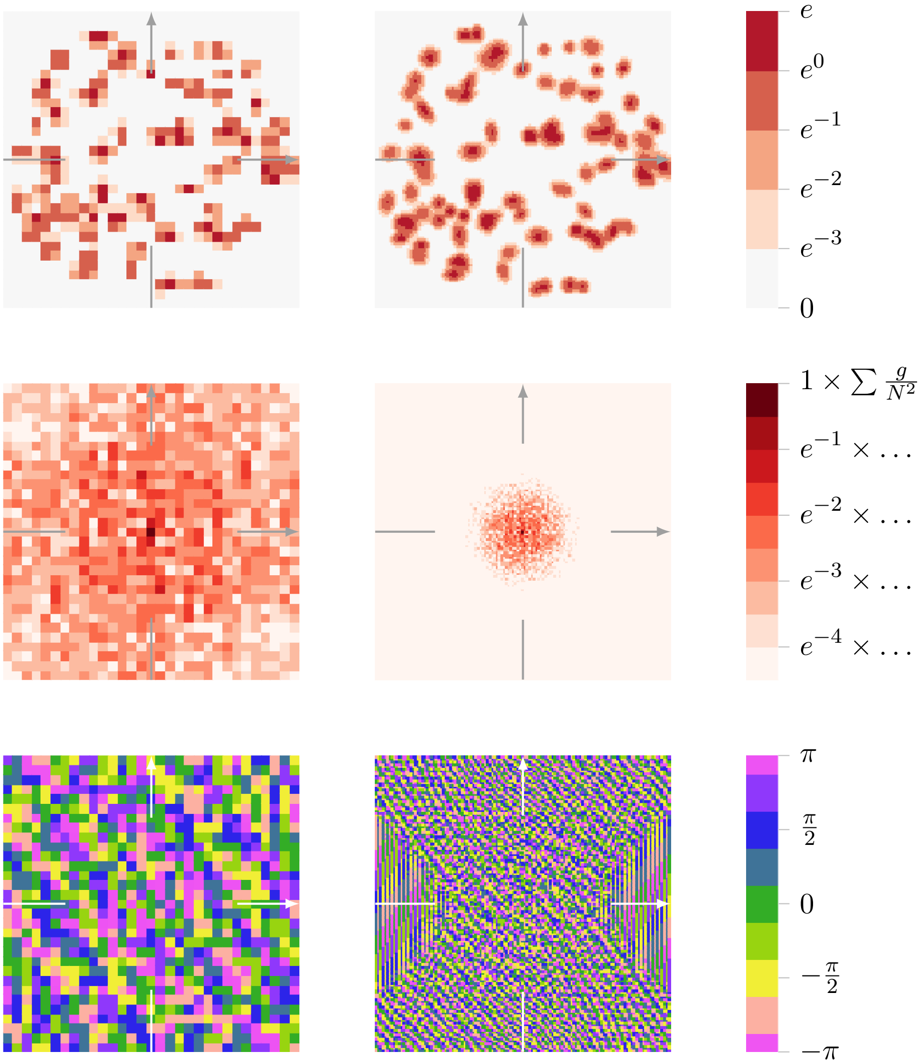

Figure E.5: Sum of 84 Gaussians, large support radius

Top left: $g_{84G;31}$. Top right: $g_{84G;127}$. Center left: $|\hat{g}_{84G;31}|$. Center right: $|\hat{g}_{84G;127}|$. Bottom left: $\arg(\hat{g}_{84G;31})$. Bottom right: $\arg(\hat{g}_{84G;127})$. Graphs of phases are modulated by $e^{ix_0 \cdot}$, where the shift x_0 is approximately equals to the barycenter of the corresponding function (to minimize visual cluttering due to phase oscillations).

In comparisson to concentrated Gaussians (Figure E.4), solution phase (in the bottom row) is notably less regular.

[Ala+19] Rima Alaifari, Ingrid Daubechies, Philipp Grohs, and Rujie Yin. "Stable Phase Retrieval in Infinite Dimensions." en. In: *Foundations of Computational Mathematics* 19.4 (Aug. 2019), pp. 869–900. ISSN: 1615-3383. DOI: 10.1007/s10208-018-9399-7.

[AGS01] L. Ambrosio, N. Gigli, and G. Savare. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhauser Verlag, 2001. ISBN: 3-7643-2428-7.

[AK00] L. Ambrosio and B. Kirchheim. "Rectifiable Sets in Metric and Banach Spaces." In: *Mathematische Annalen* 318 (Nov. 2000), pp. 527–555.

[AG13] Luigi Ambrosio and Nicola Gigli. "A User's Guide to Optimal Transport." en. In: *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle*. Ed. by Luigi Ambrosio, Alberto Bressan, Dirk Helbing, Axel Klar, and Enrique Zuazua. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, 2013, pp. 1–155. ISBN: 978-3-642-32160-3. DOI: 10.1007/978-3-642-32160-3_1.

[AABT16] Francisco J. Aragón Artacho, Jonathan M. Borwein, and Matthew K. Tam. "Global Behavior of the Douglas–Rachford Method for a Nonconvex Feasibility Problem." en. In: *Journal of Global Optimization* 65.2 (June 2016), pp. 309–327. ISSN: 1573-2916. DOI: 10.1007/s10898-015-0380-6.

[Aum65] Robert J Aumann. "Integrals of set-valued functions." In: *Journal of Mathematical Analysis and Applications* 12.1 (1965), pp. 1–12. ISSN: 0022-247X. DOI: https://doi.org/10.1016/0022-247X(65)90049-1. URL: https://www.sciencedirect.com/science/article/pii/0022247X65900491.

[Bar+20] Alexander Barnett, Charles L. Epstein, Leslie Greengard, and Jeremy Magland. "Geometry of the Phase Retrieval Problem." In: *arXiv:1808.10747 [math-ph]* (Apr. 2020). arXiv: 1808.10747 [math-ph].

[BC17] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer International Publishing, 2017. ISBN: 978-3-319-48310-8.

[BB93] H.H. Bauschke and J.M. Borwein. "On the Convergence of von Neumann's Alternating Projection Algorithm for Two Sets." In: *Set-Valued Analysis* 1.2 (1993).

[BB96] H.H. Bauschke and J.M. Borwein. "On Projection Algorithms for Solving Convex Feasibility Problems." In: *SIAM Review* 38.3 (1996).

[BCL02] Heinz H. Bauschke, Patrick L. Combettes, and D. Russell Luke. "Phase Retrieval, Error Reduction Algorithm, and Fienup Variants: A View from Convex Optimization." EN. In: *JOSA A* 19.7 (July 2002), pp. 1334–1345. ISSN: 1520-8532. DOI: 10.1364/JOSAA.19.001334.

[BCL03] Heinz H. Bauschke, Patrick L. Combettes, and D. Russell Luke. "Hybrid Projection–Reflection Method for Phase Retrieval." In: *J. Opt. Soc. Am. A* 20.6 (June 2003), pp. 1025–1034. DOI: 10.1364/JOSAA.20.001025.

[Bau+13] Heinz H. Bauschke, D. Russell Luke, Hung M. Phan, and Xianfu Wang. "Restricted Normal Cones and the Method of Alternating Projections: Applications." en. In: *Set-Valued and Variational Analysis* 21.3 (Sept. 2013), pp. 475–501. ISSN: 1877-0541. DOI: 10.1007/s11228-013-0238-3.

[BN13] Heinz Bauschke and Dominikus Noll. "On Cluster Points of Alternating Projections." In: (July 2013).

[Ber+00] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. "The Protein Data Bank." en. In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 235–242. ISSN: 0305-1048. DOI: 10.1093/nar/28.1.235.

[Bor07] Jonathan M. Borwein. "Proximality and Chebyshev Sets." en. In: *Optimization Letters* 1.1 (Jan. 2007), pp. 21–32. ISSN: 1862-4480. DOI: 10.1007/s11590-006-0014-5.

[BF13] Franck Boyer and Pierre Fabrie. *Mathematical Tools for the Study of the Incompressible Navier-Stokes Equations andRelated Models*. en. Applied Mathematical Sciences. New York: Springer-Verlag, 2013. ISBN: 978-1-4614-5974-3. DOI: 10.1007/978-1-4614-5975-0.

[Bre+13] Cynthia Brewer, Mark Harrower, Ben Sheesley, Andy Woodruff, and David Heyman. *ColorBrewer: Color Advice for Maps*. https://colorbrewer2.org/#. 2013.

[BP67] F. E Browder and W. V Petryshyn. "Construction of Fixed Points of Nonlinear Mappings in Hilbert Space." en. In: *Journal of Mathematical Analysis and Applications* 20.2 (Nov. 1967), pp. 197–228. ISSN: 0022-247X. DOI: 10.1016/0022-247X(67)90085-6.

[BS79] Yu M. Bruck and L. G. Sodin. "On the Ambiguity of the Image Reconstruction Problem." In: *Optics Communications* 30.3 (1979), pp. 304–308. ISSN: 0030-4018. DOI: 10.1016/0030-4018(79)90358-4.

[BL03] James V. Burke and D. Russel Luke. "Variational Analysis Applied to the Problem of Optical Phase Retrieval." In: *SIAM Journal on Control and Optimization* 42.2 (2003), pp. 576–595.

[CLS15] E. J. Candes, X. Li, and M. Soltanolkotabi. "Phase Retrieval via Wirtinger Flow: Theory and Algorithms." In: *IEEE Transactions on Information Theory* 61.4 (2015), pp. 1985–2007.

[CSV11] E. J. Candes, T. Strohmer, and V. Voroninski. "PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming." In: *ArXiv e-prints* (Sept. 2011).

[CCV52] W. Cochran, F. H. Crick, and V. Vand. "The Structure of Synthetic Polypeptides. I. The Transform of Atoms on a Helix." en. In: *Acta Crystallographica* 5.5 (Sept. 1952), pp. 581–586. ISSN: 0365-110X. DOI: 10.1107/S0365110X52001635.

[CT90] P. L. Combettes and H. J. Trussell. "Method of Successive Projections for Finding a Common Point of Sets in Metric Spaces." en. In: *Journal of Optimization Theory and Applications* 67.3 (Dec. 1990), pp. 487–507. ISSN: 1573-2878. DOI: 10.1007/BF00939646.

[CLW67] J.W. Cooley, P.A.W. Lewis, and P.D. Welch. "Historical Notes on the Fast Fourier Transform." In: *Proceedings of the IEEE* 55.10 (Oct. 1967), pp. 1675–1677. ISSN: 1558-2256. DOI: 10.1109/PROC.1967.5959.

[CFT90] T. R. Crimmins, J. R. Fienup, and B. J. Thelen. "Improved Bounds on Object Support from Autocorrelation Support and Application to Phase Retrieval." EN. In: *JOSA A* 7.1 (Jan. 1990), pp. 3–13. ISSN: 1520-8532. DOI: 10.1364/JOSAA.7.000003.

[DF87] J. Dainty and James Fienup. "Phase Retrieval and Image Reconstruction for Astronomy." In: *Image Recovery: Theory Appl* 13 (Jan. 1987).

[Deu01] Frank R. Deutsch. *Best Approximation in Inner Product Spaces*. en. CMS Books in Mathematics. New York: Springer-Verlag, 2001. ISBN: 978-0-387-95156-0. DOI: 10.1007/978-1-4684-9298-9.

[DR56] Jim Douglas and Jr H. H. Rachford. "On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables." In: 1956. DOI: 10.1090/S0002-9947-1956-0084194-4.

[Els03] V. Elser. "Phase Retrieval by Iterated Projections." In: *J. Opt. Soc. Am. A* 20.1 (2003), pp. 40–52.

[ELB18] V. Elser, T.-Y. Lan, and T. Bendory. “Benchmark Problems for Phase Retrieval.” In: *SIAM Journal on Imaging Sciences* 11.4 (2018), pp. 2429–2455.

[Els17] Veit Elser. “Matrix Product Constraints by Projection Methods.” en. In: *Journal of Global Optimization* 68.2 (June 2017), pp. 329–355. ISSN: 1573-2916. DOI: 10.1007/s10898-016-0466-9.

[ER11] René. Escalante and Marcos. Raydan. *Alternating Projection Methods*. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, Jan. 2011. ISBN: 978-1-61197-193-4. DOI: 10.1137/9781611971941.

[Eva10] Lawrence C. Evans. *Partial Differential Equations*. Providence, R.I.: American Mathematical Society, 2010. ISBN: 978-0-8218-4974-3 0-8218-4974-3.

[Fie82] J. R. Fienup. “Phase Retrieval Algorithms: A Comparison.” In: *Appl. Opt.* 21.15 (1982), pp. 2758–2769.

[Fie13] James R. Fienup. “Phase Retrieval Algorithms: A Personal Tour [Invited].” EN. In: *Applied Optics* 52.1 (Jan. 2013), pp. 45–56. ISSN: 2155-3165. DOI: 10.1364/AO.52.000045.

[FM14] James Fletcher and Warren Moors. “Chebyshev Sets.” In: *Journal of the Australian Mathematical Society* 98 (Apr. 2014), pp. 161–231. DOI: 10.1017/S1446788714000561.

[FD97] G. Friesecke and G. Doltzmann. “Implicit Time Discretization and Global Existence for a Quasi-Linear Evolution Equation with Nonconvex Energy.” In: *SIAM Journal on Mathematical Analysis* 2.28 (1997), pp. 363–380.

[Fri07] Gero Friesecke. *Lectures on Fourier Analysis*. Mathematics Institute, University of Warwick (Unpublished), May 2007.

[FJJ16] Gero Friesecke, Richard D. James, and Dominik Jüstel. “Twisted X-Rays: Incoming Waveforms Yielding Discrete Diffraction Patterns for Helical Structures.” In: *SIAM Journal on Applied Mathematics* 76.3 (Jan. 2016), pp. 1191–1218. ISSN: 0036-1399. DOI: 10.1137/15M1043418.

[GS72] R W Gerchberg and W O Saxton. “A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures.” en. In: 35 (1972), p. 6.

[Goo04] Joseph Goodman. *Introduction to Fourier Optics*. English. 3rd edition. Englewood, Colo: Roberts and Company Publishers, Dec. 2004. ISBN: 978-0-9747077-2-3.

[GKR20] Philipp Grohs, Sarah Koppensteiner, and Martin Rathmair. “Phase Retrieval: Uniqueness and Stability.” In: *arXiv:1901.07911 [math]* (Feb. 2020). arXiv: 1901.07911 [math].

[Hay82] M. Hayes. "The Reconstruction of a Multidimensional Sequence from the Phase or Magnitude of Its Fourier Transform." In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 30.2 (Apr. 1982), pp. 140–154. ISSN: 0096-3518. DOI: 10.1109/TASSP.1982.1163863.

[HL13] Robert Hesse and D. Russell Luke. "Nonconvex Notions of Regularity and Convergence of Fundamental Algorithms for Feasibility Problems." In: *arXiv:1212.3349 [math]* (June 2013). arXiv: 1212.3349 [math].

[JEH15] K. Jaganathan, Y. C. Eldar, and B. Hassibi. "Phase Retrieval: An Overview of Recent Developments." In: *ArXiv e-prints* (Oct. 2015).

[Jam99] Philippe Jaming. "Phase Retrieval Techniques for Radar Ambiguity Problems." In: *The Journal of Fourier Analysis and Applications* 5 (Jan. 1999), pp. 309–329. DOI: 10.1007/BF01259373.

[Kov15] Peter Kovesi. "Good Colour Maps: How to Design Them." In: *arXiv:1509.03700 [cs]* (Sept. 2015). arXiv: 1509.03700 [cs].

[KM80] Alexander Kruger and B.Sh Mordukhovich. "Extremal points and the Euler equation in nonsmooth optimization." In: *Doklady Akademii Nauk BSSR* 24 (Jan. 1980), pp. 684–687.

[Leo09] Giovanni Leoni. *A First Course in Sobolev Spaces*. eng. Providence, R.I: American Mathematical Society, 2009. ISBN: 978-0-8218-4768-8.

[LLM09] A. S. Lewis, D. R. Luke, and J. Malick. "Local Linear Convergence for Alternating and Averaged Nonconvex Projections." en. In: *Foundations of Computational Mathematics* 9.4 (Aug. 2009), pp. 485–513. ISSN: 1615-3383. DOI: 10.1007/s10208-008-9036-y.

[LP16] Guoyin Li and Ting Kei Pong. "Douglas–Rachford Splitting for Nonconvex Optimization with Application to Nonconvex Feasibility Problems." en. In: *Mathematical Programming* 159.1 (Sept. 2016), pp. 371–401. ISSN: 1436-4646. DOI: 10.1007/s10107-015-0963-5.

[LM13] Matthias Liero and Alexander Mielke. "Gradient Structures and Geodesic Convexity for Reaction–Diffusion Systems." In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.2005 (Dec. 2013), p. 20120346. DOI: 10.1098/rsta.2012.0346.

[LS19] Scott B. Lindstrom and Brailey Sims. "Survey: Sixty Years of Douglas–Rachford." In: *arXiv:1809.07181 [math]* (Feb. 2019). arXiv: 1809.07181 [math].

[LS20] Scott B. Lindstrom and Brailey Sims. "Survey: Sixty Years of Douglas–Rachford." In: *arXiv:1809.07181 [math]* (Jan. 2020). arXiv: 1809.07181 [math].

[LM79] P. L. Lions and B. Mercier. "Splitting Algorithms for the Sum of Two Nonlinear Operators." In: *SIAM Journal on Numerical Analysis* 16.6 (1979), pp. 964–979.

[LBL02] D. Russel Luke, James V. Burke, and Richard G. Lyon. "Optical Wavefront Reconstruction: Theory and Numerical Methods." In: *SIAM Review* 44.2 (2002), pp. 169–224.

[Luk05] D. Russell Luke. "Relaxed Averaged Alternating Reflections for Diffraction Imaging." In: *Inverse Problems* 21.1 (Feb. 2005), pp. 37–50. ISSN: 0266-5611, 1361-6420. DOI: 10.1088/0266-5611/21/1/004. arXiv: math/0405208.

[LM20] D. Russell Luke and Anna-Lena Martins. "Convergence Analysis of the Relaxed Douglas–Rachford Algorithm." In: *SIAM Journal on Optimization* 30.1 (Jan. 2020), pp. 542–584. ISSN: 1052-6234. DOI: 10.1137/18M1229638.

[Luk17] D. Luke. "Phase Retrieval, What's New?" In: *SIAM SIAG/OPT* 25 (Apr. 2017), p. 1.

[MP18] André Miede and Ivo Pletikosic. *Classicthesis Package*. https://bitbucket.org/ar 2018.

[MS96] B. S. Mordukhovich and Y. Shao. "Nonsmooth Sequential Analysis in Asplund Spaces." In: *Tras. Amer. Math. Soc.* 328.4 (1996), pp. 1235–1280.

[Mor18] Boris S. Mordukhovich. *Variational Analysis and Applications*. Springer, 2018. ISBN: 978-3-319-92773-2.

[Mou16] Walaa M. Moursi. "The Douglas–Rachford Operator in the Possibly Inconsistent Case : Static Properties and Dynamic Behaviour." eng. PhD thesis. University of British Columbia, 2016. DOI: 10.14288/1.0340501.

[MSS14] Pejman Mowlaee, Rahim Saeidi, and Yannis Stylianou. "INTERSPEECH 2014 Special Session: Phase Importance in Speech Processing Applications." In: 2014.

[NR16] D. Noll and A. Rondepierre. "On Local Convergence of the Method of Alternating Projections." In: *Foundations of Computational Mathematics* 16.2 (2016), pp. 425–455.

[Pau47] Wolfgang Pauli. *Die allgemeinen Prinzipien der Wellenmechanik*. de. J. W. Edwards, 1947.

[Pau+18] E. Pauwels, A. Beck, Y. C. Eldar, and S. Sabach. "On Fienup Methods for Sparse Phase Retrieval." In: *IEEE Transactions on Signal Processing* 66.4 (2018), pp. 982–991.

[Peg85] Robert L. Pego. "Compactness in L2 and the Fourier Transform." In: *Proceedings of the American Mathematical Society* 95.2 (1985), pp. 252–254.

[Pha15] Hung M. Phan. "Linear Convergence of the Douglas-Rachford Method for Two Closed Sets." In: *arXiv:1401.6509 [math]* (Feb. 2015). arXiv: 1401.6509 [math].

[RJ93] Lawrence Rabiner and Biing-Hwang Juang. *Fundamentals of Speech Recognition*. English. 1 edition. Englewood Cliffs, N.J: Pearson, Apr. 1993. ISBN: 978-0-13-015157-5.

[RS75] Michael Reed and Barry Simon. *II: Fourier Analysis, Self-Adjointness*. en. Elsevier, 1975. ISBN: 978-0-12-585002-5.

[RW09] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Springer-Verlag Berlin Heidelberg, 2009.

[Say52] D. Sayre. "The Squaring Method: A New Method for Phase Determination." en. In: *Acta Crystallographica* 5.1 (Jan. 1952), pp. 60–65. ISSN: 0365-110X. DOI: 10.1107/S0365110X52000137.

[Sch70] H.A. Schwarz. "Ueber Einen Grenzübergang Durch Alternirendes Verfahren." In: (1870), pp. 272–286.

[She+15] Yoav Shechtman, Yonina C. Eldar, Oren Cohen, Henry N. Chapman, Jianwei Miao, and Mordechai Segev. "Phase Retrieval with Application to Optical Imaging." In: *IEEE Signal Processing Magazine* 32.3 (May 2015), pp. 87–109. ISSN: 1053-5888.

[Str03] Robert S. Strichartz. *Guide to Distribution Theory and Fourier Transforms*. Englisch. Singapore ; River Edge, NJ: WSPC, July 2003. ISBN: 978-981-238-430-0.

[WC53] J. D. Watson and F. H. C. Crick. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." en. In: *Nature* 171.4356 (Apr. 1953), pp. 737–738. ISSN: 1476-4687. DOI: 10.1038/171737a0.

[Wen+12] Zaiwen Wen, Chao Yang, Xin Liu, and Stefano Marchesini. "Alternating Direction Methods for Classical and Ptychographic Phase Retrieval." en. In: *Inverse Problems* 28.11 (Oct. 2012), p. 115010. ISSN: 0266-5611. DOI: 10.1088/0266-5611/28/11/115010.

[YAY14] Y. Shechtman, A. Beck, and Y.C. Eldar. "GESPAR: Efficient Phase Retrieval of Sparse Signals." In: *IEEE Transactions on Signal Processing* 62.4 (2014), pp. 928–938.

LIST OF FIGURES

LIST OF TABLES

MATHEMATICAL NOTATION

$x, y \in \Omega$ variables in the object space domain
$k, q \in \Omega_F$ variables in the Fourier space domain
$t, \tau, T \in \mathbb{R}$ time variables
$\varepsilon \in \mathbb{R}_{\geqslant 0}$ step size in discretized algorithms
$(\cdot)^*$ complex conjugation
$C_{\mathcal{F}}$ normalization constant in Plancherel's theorem 16, 229
∂ depending on the context, this symbol denotes the boundary of a set, the subgradient of a convex functional ∂_c, the generalized (Kruger-Mordukhovich) subdifferential ∂_{KM}, or — used as ∂_t — the partial derivative of a function in time
$\overline{\partial}$ Clarke subdifferential
$B_R(f), \mathring{B}_R(f)$ closed and open balls around f (radius R; metric $\|\cdot\|_{\mathcal{H}}$)

$f, g, h, p, q, s, d \in \mathcal{H}$ functions in the object space
$g_n, p_n, q_n \in \mathcal{H}$ algorithm iterates 128, 194
$\widetilde{g}^{(\varepsilon)}, \widetilde{p}^{(\varepsilon)}, \widetilde{q}^{(\varepsilon)} \colon [0, T) \to \mathcal{H}$ linear interpolations in time between algorithm iterates 145, 194
$\breve{g}^{(\varepsilon)}, \breve{p}^{(\varepsilon)}, \breve{q}^{(\varepsilon)} \colon [0, T) \to \mathcal{H}$ pcw. constant interpolations in time between algorithm iterates 145, 194
$\varphi, \psi \colon \Omega_F \to \mathbb{C}$ phase of complex-valued functions 40, 42
$\sqrt{I}$ measurement data (square root of measured intensity) 19
$\mathbb{1}_{\{\text{condition}\}}(x)$ means $\mathbb{1}_{\{x \in \Omega | \text{condition at } x\}}(x)$ 33
A, B, T (possibly multivalued) operators acting on $\mathcal{H}$ 100
$A^{(\mathbf{loc})}$ local version of the operator A 44, 48, 80
$J_{\lambda A} = (I + A)^{-1}$ (possibly multivalued) resolvent of A 105

Ω physical space (domain of the object space); $\mathbb{R}^d$, $\mathbb{T}^d$, or $\mathbb{T}^d_N$ 15
Ω_F domain of the Fourier space; isomorph to $\mathbb{R}^d$, $\mathbb{Z}^d$, or $\mathbb{T}^d_N$ 16
$\mathcal{H}(\Omega)$ object space; typically, $L^2(\Omega)$ 15
$\widehat{\mathcal{H}}(\Omega_F)$ Fourier space; typically, $L^2(\Omega_F)$ 16

$\mathcal{D} \subset \mathcal{H}$, $\widehat{\mathcal{D}} \subset \widehat{\mathcal{H}}$ domains of operators and functionals
$\mathcal{X}$, $\mathcal{Y} \subset \mathcal{H}$ generic sets, typically proximal, non-empty 78
$\mathcal{C} \subset \mathcal{H}$ generic convex set, typically proximal, non-empty 97
$\mathcal{M}(\sqrt{I})$, $\mathcal{M}^{(\mathbf{i})}(\sqrt{I}; S_F) \subset \mathcal{H}$ modulus constraint set and incomplete modulus constraint set 19
$\mathcal{A} \subset \mathcal{H}$ additional constraint, typically proximal, non-empty 56
$\mathcal{P}$ non-negativity constraint 19
$\mathcal{S}(S)$ support constraint (of functions supported $S \subset \Omega$) 19
$\mathcal{T}_{\mathrm{a}}(\alpha)(\alpha)$ amplitude thresholding constraint 19
$\mathcal{T}_{\mathrm{s}}(\nu)(\nu)$ support size constraint 19
$\Pi_{\mathcal{X}}$ (set-valued) projection operator onto a proximal set $\mathcal{X}$ 24,

$P_{\mathcal{X}}$ a single-valued selection of the projection onto a proximal set $\mathcal{X}$ 24

$P_{\mathcal{M}}$ single-valued modulus projection selection (onto $\mathcal{M}$). Whenever $P_{\mathcal{M}}$ is used, it is implicitely understood that there exists a phase φ with $P_{\mathcal{M}} = P_{\mathcal{M};\varphi}$. 40, 130

$P_{\mathcal{P}}$ positivity projection (onto $\mathcal{P}$) 33, 130, 131

$R_{\mathcal{X}}$ reflection by a proximal set $\mathcal{X}$ 24, 115, 179

$E_{\mathcal{X}}$ projection energy functional onto a proximal set $\mathcal{X}$ 53, 120

$E_{\mathcal{M}}$ modulus energy functional 54, 58

$E_{\mathcal{P}}$ positivity energy functional 54, 58

H formal Hessian of $E_{\mathcal{M}} + E_{\mathcal{P}}$ 171, 245

INDEX

COLOPHON

The thesis was typeset using the package `classicthesis` [MP18], see

https://bitbucket.org/amiede/classicthesis/

Figures use color palettes from [Kov15] and [Bre+13], see

https://peterkovesi.com/projects/colourmaps/,
https://colorbrewer2.org/.

Final Version as of January 23, 2023 .