Title
A Sparse Optimization Framework for the Numerical Solution of PDEs

Permalink
https://escholarship.org/uc/item/9rx1g1j8

Author
Tekin, Omer Faruk

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Sparse Optimization Framework for the Numerical Solution of PDEs

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mathematics

by

Ömer Faruk Tekin

2016

ABSTRACT OF THE DISSERTATION

A Sparse Optimization Framework for the Numerical Solution of PDEs

by

Ömer Faruk Tekin

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2016

Professor Russel E. Caflisch, Chair

This dissertation studies the extension of sparse optimization techniques to the numerical solution of partial differential equations for applications in scientific computing, in particular many-particle systems that are governed by a differential equation. Sparse optimization techniques have attracted much attention due to their substantial computational efficiency and feasibility for large-scale problems such as image processing, compressed sensing, and machine learning. In this dissertation, ℓ^1-minimization scheme has been studied for the solutions of elliptic and parabolic differential equations. Theoretical considerations for the effectiveness of the scheme, such as the sparsity properties, completeness, consistency, and the asymptotic behavior of the solutions are analyzed.

The dissertation of Ömer Faruk Tekin is approved.

Christina Kim

Vidvuds Ozolins

Stanley J. Osher

Russel E. Caflisch, Committee Chair

University of California, Los Angeles

2016

To My Love

Table of Contents

Acknowledgments

I would like to thank my advisor Russel Caflisch for his supervision and constant support. This dissertation would not have been possible without his encouragement. I am indebted to Stanley Osher for helpful insights and guidance. Many parts of this dissertation has been originated from useful discussions with him.

I would like to thank Vidvuds Ozolins for his initiative and pioneering approach on the problem formulation. I would like to thank my other collaborators: Farzin Barekat, Ke Yin, and Jonathan Siegel. The valuable information exchanges with my collaborators made it possible to search for new ideas.

I am glad I was able to attend lectures from outstanding mathematicians at our department. I thank Inwon Kim for her lectures in partial differential equations, which provided an excellent intuition in this area. I also would like to thank Mario Bonk and Marek Biskup for their classes and discussions that helped me to reinforce mathematical and analytical skills.

Vita

2008   Gold medal in the 49th International Mathematical Olympiad

2010-2011   E.A.P. Fellow at Mathematics Department, UCLA

2012   B.S. in Mathematics, Bilkent University, Ankara, Turkey

2013-2015   Teaching Assistant, Mathematics Department, UCLA

2015-2016   Teaching Assistant, Anderson School of Management, UCLA

2014-2016   Research Assistant, Mathematics Department, UCLA

PUBLICATIONS

Tekin, Ömer Faruk, Yin, Ke and Barekat, Farzin. *Spectral results for perturbed variational eigenvalue problems and their applications to compressed PDEs* Communications in mathematical sciences 14, no. 3 (2016): 777–797.

Siegel, Jonathan and Tekin, Ömer Faruk. *Compact support of L^1 penalized variational problems* Submitted to SIAM journal on applied mathematics.

CHAPTER 1

Introduction

It is very well known in many applications in machine learning such as Lasso regression [43], compressed sensing [19], matrix rank minimization [31, 36] and principal component analysis [12] that addition of ℓ^1 (or related quantity) to the optimization problems promotes a sparse structure for the solutions [13, 14]. Such problems are generally referred as ℓ^1-minimization problems, and the ℓ^1 term appears either as a constraint, or as a penalty term to the objective functional.

Until very recently, the applications of ℓ^1-minimization techniques were limited to discrete problems and the theoretical foundations are treated mostly from a combinatorial perspective. In [39], Schaeffer et al. introduced ℓ^1-minimization for applications in continuous systems such as PDEs, whereas the underlying ℓ^1-minimization is carried out in the spectral coefficients forming a discrete set. On the other hand, Ozolins et al. [33, 34] introduced ℓ^1-minimization on the spatial variable with applications in variational problems in mathematical physics such as the Schrödinger's equation in quantum mechanics, as well as to obtain wavelets [18] adapted to differential operators. The concept of spatial localization for basis functions have been introduced [46], and are called Wannier functions. The fundamental idea in [33, 34] is to promote spatial localization in variational problems by modifying the objective functional by a quantity related to the L^1 norm. Here, L^1 norm is the continuous analogue of ℓ^1 norm, whereas the spatial localization and compact support is regarded as the continuous analogue of sparsity. This approach introduced many interesting problems for both theoretical and practical considerations. Properties of the solution to the modified continuous problem can be studied in the mathematical analysis and partial differential equations context, whereas the numerical computations can be carried out via

efficient algorithms designed for sparse structures in the convex optimization, and machine learning context. This thesis consists of the study of theoretical aspects of the solutions to such continuous problems, and the design of algorithms to numerically compute these solutions. It is organized as follows.

In Chapter 2, motivated from the variational formulations of the eigenvalues of elliptic operators, the spectral properties of the solutions of L^1-penalized variational problems for elliptic operators are studied. The results are shown to be applicable for Compressed Modes (CM) and Compressed Plane Waves (CPW), which are devised to obtain a spatially localized basis for elliptic differential operators with multi-resolution capabilities. In particular, it is shown in Chapter 2 that the CMs and CPWs have proper approximation properties that make them viable for basis functions.

Chapter 3 focuses on the quantitative and asymptotic analysis of the energy of the solutions of L^1-penalized variational problems for elliptic operators. The analysis leads to an analogue of the Weyl's Law, as well as the stability of the energy quantity in terms of the L^1-penalization parameter.

The aim of Chapter 4 is to explore the support and regularity properties by studying the associated Euler-Lagrange equations. The results are first demonstrated in a general setting, and narrowed down for applications to CMs and CPWs.

Finally in Chapter 5, I consider parabolic problems with weighted L^1-penalization terms. In particular, the compact support phenomenon in both time and space variables are verified. Furthermore, numerical schemes are provided for solving such problems.

CHAPTER 2

Completeness Results

2.1 Introduction

The purpose of this section is to provide a framework for the verification of the completeness properties of the Compressed Modes (CM) and Compressed Plane Waves (CPW) introduced in [33, 34]. We consider the problem in a general functional analytic setting, and apply the corresponding results for CMs and CPWs. This chapter is taken with slight modification from [42].

We begin with the observation that the definition of CM and CPW is a variant of the Courant variational method (see e.g. [9, 20]). Namely, the Courant variational method characterizes the Dirichlet eigenvalues and eigenvectors of a second-order symmetric elliptic operator L defined on a bounded domain $\Omega \subset \mathbb{R}^d$

$$Lu = \lambda u \text{ in } \Omega,$$

$$u = 0 \text{ on } \partial\Omega.$$

via a hierarchical variational procedure involving the minimization of the objective functional $B[u, u] = \langle Lu, u \rangle_{L^2}$. As for the CMs and CPWs, the main difference is that the objective functional is modified by an L^1 regularization term. We treat this modification term as an arbitrary non-negative functional that satisfies certain boundedness criteria, and it turns out that the L^1 term has no discernible effect on completeness phenomenon. It is important to note that a perturbation on the objective functional may not result in a unique differential operator, for which the solutions are eigenfunctions. Therefore, the classical results in perturbation theory regarding the study of the spectral properties [29, 37] are not applicable for CMs and CPWs.

To put it in a general functional analytic setting, we would like to prove an analogue of the spectral theorem for a self-adjoint operator T when the modes (generalized eigenvectors) are computed as the minimizers of an objective functional of the following form

$$\mathcal{J}(u) := P(u) + \langle Tu, u \rangle,$$

where P is a non-negative penalty term. To tackle the technicalities associated with the above objective functional, we briefly discuss basic functional analytic facts in Section 2.2.

We prove that an analogue of the spectral theorem holds provided that the eigenvalues of the operator grows sufficiently fast. In particular, the growth condition holds when T is the weak realization of an elliptic operator on a bounded domain Ω lying inside $\mathbb{R}$ or $\mathbb{R}^2$.

The results presented in this chapter are further refinements of the results in [48, 4], where the authors provide error estimates for CMs. A closely related study on spectral properties for problems with L^1 type terms is [24], where the spectral theory for evolution equations with the total variation (TV) flow is developed for applications in image filtering where traditional filtering approaches are not applicable due to the geometry of the images.

2.2 Perturbed variational problems associated to linear operators

The proof of the spectral theorem for elliptic operators relies on the fact that the "inverse" of the elliptic differential operator is compact. The spectral theorem for symmetric compact operators in stated as follows.

Theorem 2.2.1 (Spectral Theorem for Compact Operators)**.** *Let $\mathcal{H}$ be a Hilbert space, and $K : \mathcal{H} \to \mathcal{H}$ be a linear symmetric compact operator. Then,*

1. *K has real eigenvalues ν_k, and $\nu_k \to 0$ as $k \to \infty$,*

2. *The (normalized) eigenvectors $\{\phi_k\}_{k=1}^{\infty}$, with $K\phi_k = \nu_k\phi_k$, form a complete orthonormal system in $\mathcal{H}$.*

As a consequence of the above spectral theorem, the inverse T, of a positive compact operator K satisfies the following spectral theorem.

Theorem 2.2.2 (Spectral Theorem for Inverse-Compact Operators). *Let $\mathcal{H}$ be a Hilbert space, and $K : \mathcal{H} \to \mathcal{H}$ be a linear bijective symmetric compact operator, that is also bounded below. Then, $T = K^{-1}$ satisfies the following properties*

1. *T has real eigenvalues λ_k, with $\{\lambda_k\}_{k=1}^{\infty}$ in increasing order, and $\lambda_k \to +\infty$ as $k \to \infty$,*

2. *The (normalized) eigenvectors $\{\phi_k\}_{k=1}^{\infty}$, with $T\phi_k = \lambda_k \phi_k$, form a complete orthonormal system.*

Remark 2.2.3. Notice that the operator T defined in Theorem 2.2.2 is unbounded, hence T must have a domain of definition, $\mathcal{D}(T)$, for which it is self-adjoint. We consider the following natural choice of the domain of definition,

$$\mathcal{D}(T) = \{\alpha = \sum_{n\in\mathbb{N}} \hat{\alpha}_n \phi_n \in \mathcal{H} | \sum_{n\in\mathbb{N}} \lambda_n \hat{\alpha}_n \phi_n \in \mathcal{H}, \text{ or equivalently, } \sum_{n\in\mathbb{N}} \lambda_n^2 |\hat{\alpha}_n|^2 < \infty\},$$

in which there is no ambiguity of the definition of T.

We now work with the operators T that can be represented as the inverse of some bijective symmetric compact operator. The eigenvalues and eigenvectors of T can be characterized via the following Courant-Fisher variational formulae (see e.g. [9])

$$
\begin{aligned}
\lambda_1 &= \min_{\substack{u\in\mathcal{D}(T)\\ ||u||=1}} \langle Tu, u \rangle, \\
\phi_1 &= \operatorname*{argmin}_{\substack{u\in\mathcal{D}(T)\\ ||u||=1}} \langle Tu, u \rangle, \\
\lambda_k &= \min_{\substack{u\in\mathcal{D}(T)\\ u\in\{\phi_1,...,\phi_{k-1}\}^{\perp}\\ ||u||=1}} \langle Tu, u \rangle, \\
\phi_k &= \operatorname*{argmin}_{\substack{u\in\mathcal{D}(T)\\ u\in\{\phi_1,...,\phi_{k-1}\}^{\perp}\\ ||u||=1}} \langle Tu, u \rangle.
\end{aligned}
$$

We consider a similar variational problem, where we perturb the objective functional $\langle Tu, u \rangle$. Strictly speaking, we define

$$J[u] = \langle Tu, u \rangle + P(u),$$

5

where $P : \mathcal{H} \to \mathbb{R}$ is a non-negative penalty term. We view the term $J[u]$ as the "energy" of the element u, and run a progressive energy-minimization procedure as in the Courant-Fisher formulae. In other words, we define

$$
\begin{aligned}
\zeta_1 &= \operatorname*{argmin}_{\substack{u \in \mathcal{D}(T) \\ ||u||=1}} J[u], \\[1em]
\zeta_k &= \operatorname*{argmin}_{\substack{u \in \mathcal{D}(T) \\ u \in \{\zeta_1,\dots,\zeta_{k-1}\}^\perp \\ ||u||=1}} J[u].
\end{aligned}
\tag{2.2.1}
$$

In case of a non-uniqueness in the minimization above, we define ζ_k to be one of the solutions to the corresponding minimization problem. To ensure the existence of ζ_k's, we impose that P is bounded and lower semi-continuous with respect to the norm-convergence, in the sense that

$$
\begin{aligned}
P(u) &\leq C||u||, \\[1em]
||u_n - u|| \to 0 \implies P(u) &\leq \liminf P(u_n).
\end{aligned}
\tag{2.2.2}
$$

The smallest constant C satisfying the boundedness of P is the functional norm of P, and is denoted by $||P||$.

We require that the eigenvectors of T, i.e. $\{\phi_n\}_{n\in\mathbb{N}}$, form a complete orthonormal system in $\mathcal{H}$. We conjecture that as long as the perturbation satisfies the existence criteria (2.2.2), such a spectral result still holds

Conjecture. The set $\{\zeta_n\}_{n\in\mathbb{N}}$ obtained via the variational procedure (2.2.1) forms a complete orthonormal system in $\mathcal{H}$.

This section mainly focuses on verifying this conjecture under certain growth assumptions on the eigenvalues λ_n. In order to verify this conjecture, one needs to show that

$$
\phi_k \in \overline{\operatorname{span}\{\zeta_n\}_{n\in\mathbb{N}}} \quad \forall k \in \mathbb{N},
\tag{2.2.3}
$$

where $\operatorname{span} E$ denotes the space consisting of the finite linear combinations of the elements in E. The following Hilbert theoretic result quantifies the relation (2.2.3).

6

Lemma 2.2.4. *Let* $\{e_n\}_{n\in\mathbb{N}}$ *be a maximal orthonormal system in a Hilbert space* $\mathcal{H}$. *Let* $\{f_n\}_{n\in\mathbb{N}}$ *be an orthonormal system in* $\mathcal{H}$. *Assume that each* f_n *has the expansion*

$$f_n = \sum_{k\in\mathbb{N}} a_{n,k} e_k, \ a_{n,k} \in \mathbb{C}.$$

Then, for each $k \in \mathbb{N}$,

$$d(e_k, \mathrm{span}\{f_n\})^2 = 1 - \sum_{n\in\mathbb{N}} |a_{n,k}|^2,$$

where $d(e, M)$ *denotes the distance between an element* $e \in \mathcal{H}$, *and a linear subspace* M *of* $\mathcal{H}$.

Proof. Let w be the projection of e_k onto $\mathrm{span}\{f_n\}_{n\in\mathbb{N}}$. Then, since $\{f_n\}_{n\in\mathbb{N}}$ is an orthonormal system, w is given by

$$\begin{aligned} w &= \sum_{n\in\mathbb{N}} \langle e_k, f_n \rangle f_n \\ &= \sum_{n\in\mathbb{N}} \langle e_k, \sum_{j} a_{n,j} e_j \rangle f_n \\ &= \sum_{n\in\mathbb{N}} \overline{a_{n,k}} f_n. \end{aligned}$$

Hence, we can compute the size of w

$$||w||^2 = \sum_{n\in\mathbb{N}} |a_{n,k}|^2. \tag{2.2.4}$$

Note by the property of the projection that $e_k - w \perp w$, therefore, by the Pythagorean identity, we have

$$||e_k||^2 = ||e_k - w||^2 + ||w||^2. \tag{2.2.5}$$

Notice also that w, being the projection of e_k onto $\mathrm{span}\{f_n\}_{n\in\mathbb{N}}$, is the closest point to e_k inside $\mathrm{span}\{f_n\}_{n\in\mathbb{N}}$, so that

$$d(e_k, \mathrm{span}\{f_n\})^2 = ||e_k - w||^2. \tag{2.2.6}$$

Combining (2.2.4)-(2.2.6), we get

$$d(e_k, \mathrm{span}\{f_n\})^2 = ||e_k||^2 - ||w||^2 = 1 - \sum_{n\in\mathbb{N}} |a_{n,k}|^2,$$

as desired. $\qquad\square$

Corollary 2.2.5. *Let $\{e_n\}_{n\in\mathbb{N}}$, $\{f_n\}_{n\in\mathbb{N}}$, $a_{n,k}$ be as in Lemma 2.2.4. Then, for each $k \in \mathbb{N}$,*

$$\sum_{n\in\mathbb{N}} |a_{n,k}|^2 \leq 1, \tag{2.2.7}$$

and

$$e_k \in \overline{\mathrm{span}\{f_n\}} \iff \sum_{n\in\mathbb{N}} |a_{n,k}|^2 = 1.$$

The following lemma yields an estimate for the elements that are lying inside the orthogonal complement of any arbitrary orthonormal system, in terms of the deviation of their functional values from the sum of the eigenvalues corresponding to the true eigenstates. This deviation is denoted by $R(N)$ (see the inequality (2.2.8).

Lemma 2.2.6. *Let ϕ_n be the eigenvectors of the operator T, with the corresponding eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ being in increasing order. Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal system satisfying*

$$\sum_{n=1}^{N} \langle Te_n, e_n \rangle \leq R(N) + \sum_{n=1}^{N} \lambda_n \quad \forall N \in \mathbb{N}. \tag{2.2.8}$$

Suppose that there exists $f \in \{e_n\}_{n\in\mathbb{N}}^{\perp}$, $\|f\| = 1$, with the expansion

$$f = \sum_{n\in\mathbb{N}} f_n \phi_n, \ f_n \in \mathbb{C}.$$

Then, we have

$$\sum_{n=1}^{N} |f_n|^2 (\lambda_{N+1} - \lambda_n) \leq R(N), \quad \forall N \in \mathbb{N}. \tag{2.2.9}$$

Proof. Let $\{a_{n,k}\}_{n,k\in\mathbb{N}}$ denote the coefficients when e_n expanded in the basis $\{\phi_k\}_{k\in\mathbb{N}}$, i.e.

$$e_n = \sum_{k\in\mathbb{N}} a_{n,k} \phi_k.$$

Applying the result (2.2.7) of Lemma 2.2.4 to the orthonormal systems $\{e_n\}_{n\in\mathbb{N}} \cup \{f\}$, and $\{\phi_n\}_{n\in\mathbb{N}}$, for each $k \in \mathbb{N}$, we get

$$\sum_{n\in\mathbb{N}} |a_{n,k}|^2 \leq 1 - |f_k|^2. \tag{2.2.10}$$

By the bilinearity of the inner product,

$$\langle Te_n, e_n \rangle = \sum_{k=1}^{\infty} |a_{n,k}|^2 \lambda_k.$$

8

Hence,

$$\sum_{n=1}^{N} \langle Te_n, e_n \rangle = \sum_{k=1}^{\infty} \left(\sum_{n=1}^{N} |a_{n,k}|^2 \right) \lambda_k. \tag{2.2.11}$$

Since e_n's have norm 1, in the expression (2.2.11), the coefficients of λ_k summed over k equals N. Given that the λ_k's are in the increasing order, the expression (2.2.11) is minimized when the coefficients of λ_k are maximized for small k. Having the constraint (2.2.10), we get

$$\sum_{n=1}^{N} \langle Te_n, e_n \rangle = \sum_{k=1}^{\infty} \left(\sum_{n=1}^{N} |a_{n,k}|^2 \right) \lambda_k \geq \sum_{k=1}^{N} \left(1 - |f_k|^2 \right) \lambda_k + \lambda_{N+1} \sum_{k=1}^{N} |f_k|^2.$$

Combining this, with the inequality (2.2.8) we get

$$\sum_{n=1}^{N} \lambda_n + R(N) \geq \sum_{n=1}^{N} \langle Te_n, e_n \rangle \geq \sum_{n=1}^{N} \left(1 - |f_n|^2 \right) \lambda_n + \lambda_{N+1} \sum_{n=1}^{N} |f_n|^2,$$

which implies

$$\sum_{n=1}^{N} |f_n|^2 (\lambda_{N+1} - \lambda_n) \leq R(N),$$

as desired. $\qquad \square$

Lemma 2.2.6 will be essential for proving the completeness result. The estimate (2.2.9) provides us an understanding of the elements lying inside the orthogonal complement in terms of $R(N)$, and the eigenvalues of T. If the estimate (2.2.9) is incompatible with the growth of eigenvalues, then we deduce that the orthogonal complement is empty, and hence the orthonormal system is maximal.

The next lemma provides an estimate for the deviation of the functional values of ζ_n, from the eigenvalues corresponding to the true eigenstates, hence Lemma 2.2.6 is applicable.

Lemma 2.2.7. *Let $\{\zeta_n\}_{n \in \mathbb{N}}$ be the solutions to the variational procedure (2.2.1). Let $\{\lambda_n\}_{n \in \mathbb{N}}$ be the eigenvalues of T, in increasing order. Then,*

$$J[\zeta_n] \leq \lambda_n + ||P||.$$

In particular, we have

$$\sum_{n=1}^{N} \langle T\zeta_n, \zeta_n \rangle \leq \sum_{n=1}^{N} J[\zeta_n] \leq \sum_{n=1}^{N} \lambda_n + ||P||N. \tag{2.2.12}$$

9

Proof. Let $\{a_{n,k}\}_{n,k \in \mathbb{N}}$ denote the coefficients when ζ_n expanded in the basis $\{\phi_k\}_{k \in \mathbb{N}}$

$$\zeta_n = \sum_{k \in \mathbb{N}} a_{n,k} \phi_k.$$

Let $n \in \mathbb{N}$ be fixed. For integers j with $1 \leq j \leq n-1$, define

$$\eta_j = \sum_{k=1}^{n} a_{j,k} \phi_k.$$

Clearly, $\{\eta_1, \eta_2, \ldots, \eta_{n-1}\} \subset \operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\}$. Furthermore,

$$\dim \operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\} = n,$$

i.e. the space $\operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\}$ has dimension larger than the cardinality of $\{\eta_1, \eta_2, \ldots, \eta_{n-1}\}$, so that we can find $\eta \in \operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\}$, $\eta \neq 0$, such that

$$\eta \perp \eta_j, \quad \forall j : 1 \leq j \leq n-1.$$

Now, since η_j's and η lie inside $\operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\}$, we get

$$\langle \eta, \zeta_j \rangle = \langle \eta, \eta_j \rangle = 0, \quad \forall j : 1 \leq j \leq n-1,$$

so that

$$\eta \perp \zeta_j, \quad \forall j : 1 \leq j \leq n-1.$$

By rescaling, we may assume $||\eta|| = 1$, so that η lies precisely in the class of functions where we look for a minimizer to determine ζ_n. The function η is in the solution set of the variational problem (2.2.1) at the nth step, hence

$$J[\zeta_n] \leq J[\eta] = \langle T\eta, \eta \rangle + P(\eta). \tag{2.2.13}$$

Since P is a bounded operator, we have

$$P(\eta) \leq ||P|| \, ||\eta|| = ||P||. \tag{2.2.14}$$

Suppose η has the expansion

$$\eta = \sum_{k=1}^{n} b_k \phi_k.$$

10

Since, $||\eta|| = 1$, we have $\sum_{k=1}^{n} |b_k|^2 = 1$. Furthermore, by the bilinearity of the inner product,

$$\langle T\eta, \eta \rangle = \sum_{k=1}^{n} |b_k|^2 \lambda_k.$$

We also have that λ_k's are in increasing order, so that

$$\langle T\eta, \eta \rangle = \sum_{k=1}^{n} |b_k|^2 \lambda_k \leq \lambda_n \sum_{k=1}^{n} |b_k|^2 = \lambda_n. \tag{2.2.15}$$

Combining (2.2.13)-(2.2.15), we obtain

$$J[\zeta_n] \leq \lambda_n + ||P||. \tag{2.2.16}$$

Summing up the inequality (2.2.16) for $n = 1, 2, \ldots, N$, and combining with the non-negativity of P, we verify (2.2.12). □

The following theorem provides the completeness of the orthonormal system $\{\zeta_n\}_{n \in \mathbb{N}}$, provided that the eigenvalues satisfy the super-linear growth.

Theorem 2.2.8. *Suppose the eigenvalues of T satisfy*

$$\lim_{n \to \infty} \frac{\lambda_n}{n} = \infty. \tag{2.2.17}$$

Then, $\{\zeta_n\}_{n \in \mathbb{N}}$, which are defined by the variational procedure (2.2.1), forms a complete orthonormal system in $\mathcal{H}$.

Proof. Lemma 2.2.7 implies

$$\sum_{n=1}^{N} \langle T\zeta_n, \zeta_n \rangle \leq \sum_{n=1}^{N} \lambda_n + ||P||N,$$

so that Lemma 2.2.6 is applicable to the orthonormal system $\{\zeta_n\}_{n \in \mathbb{N}}$ with the function $R(N) = ||P||N$. That is, assuming the existence of an $f \in \{\zeta_n\}_{n \in \mathbb{N}}^{\perp}$, $||f|| = 1$ with the expansion

$$f = \sum f_n \phi_n, \quad f_n \in \mathbb{C},$$

we obtain the estimate

$$\sum_{n=1}^{N} |f_n|^2 (\lambda_{N+1} - \lambda_n) \leq ||P||N, \quad \forall N \in \mathbb{N}. \tag{2.2.18}$$

11

This last inequality implies (assuming $f_n \neq 0$)

$$\lambda_{N+1} - \lambda_n \leq \frac{||P||N}{|f_n|^2}, \quad \forall N \in \mathbb{N},$$

which yields a contradiction by violating the growth condition (2.2.17) on λ_{N+1}, as we take limit as $N \to \infty$. Therefore, there is no non-zero function $f \in \{\zeta_n\}_{n\in\mathbb{N}}^{\perp}$, implying that the orthonormal system $\{\zeta_n\}_{n\in\mathbb{N}}$ is complete, as desired. $\qquad \square$

By trading the magnitude of the penalty term with the growth of λ_n, we can generalize the Theorem 2.2.8 so that it holds under a weaker growth condition on λ_n.

Theorem 2.2.9. *Suppose that the eigenvalues λ_n grows linearly in the sense that they satisfy*

$$\lambda_n = Mn + o(n), \ \ as \ n \to \infty \tag{2.2.19}$$

for some constant M. Suppose also that the penalty term P satisfies the following bound

$$||P|| < M. \tag{2.2.20}$$

Then, $\{\zeta_n\}_{n\in\mathbb{N}}$ forms a complete orthonormal system in $\mathcal{H}$.

Proof. Proceeding similarly as in the proof of Theorem 2.2.8, we get the inequality (2.2.18). Namely, if f is a function with unit norm, lying in the orthogonal complement of $\{\zeta_n\}_{n\in\mathbb{N}}$, then we have

$$\sum_{n=1}^{N} |f_n|^2 (\lambda_{N+1} - \lambda_n) \leq ||P||N, \quad \forall N \in \mathbb{N}. \tag{2.2.21}$$

Now, for $m < N$, by the monotonicity of λ_k, we can lower-bound the LHS of (2.2.21) by

$$(\lambda_{N+1} - \lambda_m) \sum_{n=1}^{m} |f_n|^2,$$

so that

$$\frac{\lambda_{N+1} - \lambda_m}{N} \sum_{n=1}^{m} |f_n|^2 \leq ||P||, \ \forall N \in \mathbb{N}, \forall m : 0 < m < N. \tag{2.2.22}$$

Taking the limit of (2.2.22) as $N \to \infty$, with the aid of the growth condition (2.2.19), we obtain

$$M \sum_{n=1}^{m} |f_n|^2 \leq ||P||, \ \forall m \in \mathbb{N}.$$

Now, taking the limit as $m \to \infty$, and by the fact that $||f|| = 1$, we obtain

$$M \leq ||P||,$$

contradicting (2.2.20).　　　　　　　　　　　　　　　　　　　　　　　　　　$\square$

The following theorem establishes how close the functions ζ_n approximate the subspaces generated by the first few eigenvectors of the operator T.

Theorem 2.2.10. *Let V_m be the subspace generated by the functions $\{\zeta_1, \zeta_2, \ldots, \zeta_m\}$. Then, for any $n \leq m$, we have*

$$\sum_{k=1}^{n} d(\phi_k, V_m)^2 \leq \frac{m||P||}{\lambda_{m+1} - \lambda_n},$$

provided $\lambda_{m+1} \neq \lambda_n$.

Proof. Recall from Lemma 2.2.7 that

$$\sum_{j=1}^{m} \langle T\zeta_j, \zeta_j \rangle \leq \sum_{j=1}^{m} J[\zeta_j] \leq m||P|| + \sum_{j=1}^{m} \lambda_j. \tag{2.2.23}$$

Recall by the bilinearity of the inner product that

$$\langle T\zeta_j, \zeta_j \rangle = \sum_{k=1}^{\infty} |a_{j,k}|^2 \lambda_k. \tag{2.2.24}$$

Combining (2.2.24) and (2.2.23), yields that

$$\sum_{j=1}^{m} \sum_{k=1}^{\infty} |a_{j,k}|^2 \lambda_k \leq m||P|| + \sum_{j=1}^{m} \lambda_j. \tag{2.2.25}$$

Rearranging (2.2.25), we obtain

$$\sum_{k=m+1}^{\infty} \sum_{j=1}^{m} |a_{j,k}|^2 \lambda_k - \sum_{k=1}^{m} \left(1 - \sum_{j=1}^{m} |a_{j,k}|^2 \right) \lambda_k \leq m||P||.$$

Lower-bounding the terms λ_k with $k > m$, by λ_{m+1} in the last expression, we obtain

$$\lambda_{m+1} \sum_{k=m+1}^{\infty} \sum_{j=1}^{m} |a_{j,k}|^2 - \sum_{k=1}^{m} \left(1 - \sum_{j=1}^{m} |a_{j,k}|^2 \right) \lambda_k \leq m||P||. \tag{2.2.26}$$

Since $\sum_{k=1}^{\infty} |a_{j,k}|^2 = 1$ for $j = 1, 2, \ldots, m$, we conclude that

$$\sum_{k=m+1}^{\infty} \sum_{j=1}^{m} |a_{j,k}|^2 = \sum_{k=1}^{m} \left(1 - \sum_{j=1}^{m} |a_{j,k}|^2 \right).$$

Substituting this into the inequality (2.2.26), and rearranging further we obtain

$$\sum_{k=1}^{m} \left(1 - \sum_{j=1}^{m} |a_{j,k}|^2 \right) (\lambda_{m+1} - \lambda_k) \le m||P||. \tag{2.2.27}$$

Notice by Lemma 2.2.4 that the coefficient in front of $\lambda_{m+1} - \lambda_k$ in (2.2.27) is equal to $d(\phi_k, V_m)^2$, so that (2.2.27) becomes

$$\sum_{k=1}^{m} d(\phi_k, V_m)^2 (\lambda_{m+1} - \lambda_k) \le m||P||.$$

Exploiting the monotonicity of λ_k once more, we obtain

$$\sum_{k=1}^{n} d(\phi_k, V_m)^2 (\lambda_{m+1} - \lambda_n) \le \sum_{k=1}^{m} d(\phi_k, V_m)^2 (\lambda_{m+1} - \lambda_k) \le m||P||,$$

implying

$$\sum_{k=1}^{n} d(\phi_k, V_m)^2 \le \frac{m||P||}{\lambda_{m+1} - \lambda_n},$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

2.3 Perturbed variational problems associated to elliptic operators

The results of Section 2.2 can now be applied to second-order linear symmetric elliptic operators. Let L be a second-order linear symmetric elliptic operator defined on a bounded domain Ω in $\mathbb{R}^d$. For simplicity, we consider symmetric elliptic operators with principal parts $-\Delta$, i.e. the Schrödinger's operator given by

$$Lu = -\Delta u + cu,$$

where $c : \Omega \to \mathbb{R}$ is a bounded measurable function.

As noted in the beginning of Section 2.2, according to the spectral theorem for second-order linear symmetric elliptic operators, L satisfies the following properties

1. L has real (Dirichlet) eigenvalues, λ_k, with $\{\lambda_k\}_{k=1}^\infty$ in increasing order, and $\lambda_k \to +\infty$ as $k \to \infty$,

2. The (normalized) eigenfunctions $\{\phi_k\}_{k=1}^\infty$, with $L\phi_k = \lambda_k \phi_k$, form a complete orthonormal system in $L^2(\Omega)$.

Furthermore, the Courant-Fisher principle applies to L, hence the eigenvalues and eigenfunctions of L is given by the following variational formulae

$$
\begin{aligned}
\lambda_1 &= \min_{\substack{u\in H_0^1(\Omega) \\ ||u||_2=1}} B[u,u], \\
\phi_1 &= \operatorname*{argmin}_{\substack{u\in H_0^1(\Omega) \\ ||u||_2=1}} B[u,u], \\
\lambda_k &= \min_{\substack{u\in H_0^1(\Omega) \\ u\in\{\phi_1,...,\phi_{k-1}\}^\perp \\ ||u||_2=1}} B[u,u], \\
\phi_k &= \operatorname*{argmin}_{\substack{u\in H_0^1(\Omega) \\ u\in\{\phi_1,...,\phi_{k-1}\}^\perp \\ ||u||_2=1}} B[u,u],
\end{aligned}
$$

where

$$
B[u,v] = \langle Lu, v \rangle = \int_\Omega \nabla u \cdot \overline{\nabla v} + c u \overline{v} \, d\mathbf{x}
$$

is the bilinear form associated to L.

We proceed similarly as in Section 2.2, where we perturb the functional $B[u,u]$. This time, we restrict ourselves to the penalty terms given by a constant multiple of the L^1 norm. Namely, we consider the energy functional

$$
J_\mu[u] = B[u,u] + \frac{1}{\mu}||u||_{L^1},
$$

and analogously define the functions $\{\psi_k\}_{k\in\mathbb{N}}$ via

$$
\begin{aligned}
\psi_1 &= \operatorname*{argmin}_{\substack{u\in H_0^1(\Omega) \\ ||u||_2=1}} J_\mu[u], \\
\psi_k &= \operatorname*{argmin}_{\substack{u\in H_0^1(\Omega) \\ u\in\{\psi_1,...,\psi_{k-1}\}^\perp \\ ||u||_2=1}} J_\mu[u].
\end{aligned}
\tag{2.3.1}
$$

15

We call these functions $\{\psi_k\}_{k\in\mathbb{N}}$ "Compressed Modes of second type" (CM-II) by analogy to the "Compressed Modes" (CM) defined in [33]. The main difference between CM and CM-II is that CMs are defined as the minimizers of a joint energy sum under orthogonality constraints, whereas CM-IIs marginally minimize the energy functional under a progressive orthogonality constraint. Nevertheless, CM-IIs $\{\psi_1, \psi_2, \ldots, \psi_m\}$, being an orthonormal sequence, are in the solution set of the associated variational problem for CMs. Therefore, the theory for CM-IIs can be used to prove the analogous results for CMs.

The following lemma establishes the existence of $\{\psi_k\}_{k\in\mathbb{N}}$ by verifying the existence criteria (2.2.2) for the L^1 penalty term in the definition of J_μ.

Lemma 2.3.1. *Let* $P : L^2(\Omega) \to \mathbb{R}$ *be defined by* $P(u) = \frac{1}{\mu}||u||_{L^1}$. *Then, P satisfies the criteria (2.2.2). Furthermore,*

$$||P|| = \frac{|\Omega|^{\frac{1}{2}}}{\mu}.$$

Proof. By Cauchy-Schwarz inequality,

$$\frac{1}{\mu}||u||_{L^1} \leq \frac{1}{\mu}||u||_{L^2}||\chi_\Omega||_{L^2} = \frac{|\Omega|^{\frac{1}{2}}}{\mu}||u||_{L^2}. \tag{2.3.2}$$

Here, χ_Ω denotes the characteristic function of the domain Ω. The equality in (2.3.2) holds when u is a non-zero constant function. Therefore, P is bounded with functional norm $\frac{|\Omega|^{\frac{1}{2}}}{\mu}$. As for the lower semi-continuity, consider a sequence $u_n \in L^2(\Omega)$ converging to some $u \in L^2(\Omega)$ in L^2. The inequality 2.3.2 implies that u_n converges to u also in L^1, so that $P(u) = \lim P(u_n)$, as desired. $\square$

Now, the Theorems 2.2.8, 2.2.9, 2.2.10, can be derived for $\{\psi_k\}_{k\in\mathbb{N}}$. Theorems 2.2.8, and 2.2.9 holds true when the eigenvalues grow super-linearly, and linearly, respectively. Weyl's law yields the exact asymptotic behavior of the eigenvalues of a second-order linear symmetric elliptic operator.

Theorem 2.3.2 (Weyl's Law (see e.g. [20]))**.** *Let L be a second-order linear elliptic operator on a bounded domain $\Omega \subset \mathbb{R}^d$, of the form*

$$Lu = -\Delta u + cu,$$

where $c : \Omega \to \mathbb{R}$ is a bounded measurable function. Let $\{\lambda_n\}_{n \in \mathbb{N}}$ be the eigenvalues of L, in increasing order. Then,

$$\lambda_n = \frac{(2\pi)^d}{\omega_d |\Omega|} n^{2/d} + o(n^{2/d}), \ \ as \ n \to \infty,$$

where ω_d denotes the volume of the unit ball in $\mathbb{R}^d$.

Therefore, we can deduce from the Weyl's law that super-linear and linear growth conditions on eigenvalues holds true precisely in dimensions 1, and 2, so that we have the following corollaries as direct consequences of the Theorems 2.2.8, 2.2.9, and 2.2.10.

Corollary 2.3.3 (Corollary to Theorem 2.2.8). *Let $L = -\Delta + c(x)$ be defined on a bounded open interval in $\mathbb{R}$, where c is a bounded measurable function. Then, the associated Compressed Modes of second type (CM-II), $\{\psi_n\}_{n \in \mathbb{N}}$, which are defined by the variational procedure (2.3.1), forms a complete orthonormal system in $L^2(\Omega)$.*

Corollary 2.3.4 (Corollary to Theorem 2.2.9). *Let $L = -\Delta + c(x)$ be defined on a bounded rectangular domain in $\mathbb{R}^2$, where c is a bounded measurable function. Suppose also that the penalty parameter μ satisfies the following bound*

$$\mu > \frac{|\Omega|^{\frac{3}{2}}}{4\pi}.$$

Then, the associated Compressed Modes of second type (CM-II), $\{\psi_n\}_{n \in \mathbb{N}}$, which are defined by the variational procedure (2.3.1), forms a complete orthonormal system in $L^2(\Omega)$.

Corollary 2.3.5 (Corollary to Theorem 2.2.10). *Let $\{\phi_k\}_{k \in \mathbb{N}}$ be the (Dirichlet) eigenfunctions of the operator $L = -\Delta + c(x)$, defined on a bounded domain $\Omega \subset \mathbb{R}^n$. Let $\{\lambda_k\}_{k \in \mathbb{N}}$ be the associated eigenvalues. Let $\{\psi_n\}_{n \in \mathbb{N}}$ be the functions defined by the variational procedure (2.3.1), and V_m be the subspace generated by the functions $\{\psi_1, \psi_2, \ldots, \psi_m\}$. Then, for any $n \le m$, we have*

$$\sum_{k=1}^{n} d(\phi_k, V_m)^2 \le \frac{m |\Omega|^{\frac{1}{2}}}{\mu(\lambda_{m+1} - \lambda_n)}, \tag{2.3.3}$$

provided $\lambda_{m+1} \ne \lambda_n$.

Remark 2.3.6. The inequality (2.3.3) in Corollary (2.3.5) quantifies how close the function spaces V_m approximate the true eigenfunctions in terms of the distribution of the eigenvalues. The distribution of eigenvalues has been widely studied for the Laplace equation in the context of universal inequalities (see for instance [26]). In Appendix B, we discuss how the RHS of the inequality (2.3.3) can be further refined to a more explicit form for specific domains Ω.

Remark 2.3.7. The Corollaries 2.3.3- 2.3.5 holds true (with modified inequalities) for second-order linear symmetric elliptic operators with appropriate coercivity properties, so that the eigenvalues grows at the same order as $-\Delta$.

The following theorem establishes that the elements that lie in the orthogonal complement of $\{\psi_n\}_{n\in\mathbb{N}}$ cannot lie inside the space $H_0^1(\Omega)$. In other words, the orthogonal complement consists of highly irregular functions.

Theorem 2.3.8. *Let $\Omega \subset \mathbb{R}^d$, and $\{\psi_n\}_{n\in\mathbb{N}}$ be the solutions to the variational procedure (2.2.1). Then,*

$$\{\psi_n\}_{n\in\mathbb{N}}^{\perp} \cap H_0^1(\Omega) = \{0\}.$$

Proof. Assume to the contrary that there is a non-zero $f \in \{\psi_n\}^{\perp} \cap H_0^1(\Omega)$. We may normalize f such that $||f||_2 = 1$. Since $f \in \{\psi_1,\ldots,\psi_{n-1}\}^{\perp}$ for all n, f is in the class of functions where we look for a minimizer to obtain ψ_n, hence it is in the solution set of the variational problem (2.2.1) at n^{th} step. As ψ_n is the actual solution to the corresponding minimization problem, we have

$$J_\mu[\psi_n] \leq J_\mu[f]. \tag{2.3.4}$$

We now prove that

$$\lim_{n\to\infty} J_\mu[\psi_n] = \infty, \tag{2.3.5}$$

which together with (2.3.4) implies

$$J_\mu[f] = \infty. \tag{2.3.6}$$

Recall that

$$\sum_{n=1}^{N} B[\psi_n, \psi_n] = \sum_{k=1}^{\infty} \left(\sum_{n=1}^{N} |a_{n,k}|^2 \right) \lambda_k, \tag{2.3.7}$$

18

where the coefficients in front of λ_k have magnitude less than or equal to 1 for each k, and their sum over k is N. Since λ_k's are in increasing order, the expression (2.3.7) is minimized when the coefficients of λ_k are maximized for small k. Therefore,

$$\sum_{n=1}^{N} B[\psi_n, \psi_n] \geq \sum_{k=1}^{N} \lambda_k,$$

and since P is non-negative

$$\sum_{n=1}^{N} J_\mu[\psi_n] = \sum_{n=1}^{N} B[\psi_n, \psi_n] + P(\psi_n) \geq \sum_{n=1}^{N} B[\psi_n, \psi_n] \geq \sum_{k=1}^{N} \lambda_k. \qquad (2.3.8)$$

We know that $\lim_{n \to \infty} \lambda_n = \infty$, and both λ_n's and $J_\mu[\psi_n]$'s are in increasing order. Therefore, the inequality (2.3.8) can hold only if (2.3.5) holds. Hence, we verify (2.3.6), i.e.

$$J_\mu[f] = B[f, f] + P(f) = \infty. \qquad (2.3.9)$$

Now, since P is bounded, the expression (2.3.9) yields

$$B[f, f] = \infty. \qquad (2.3.10)$$

On the other hand, as B is the bilinear form associated to a second order elliptic operator, it is bounded in the sense that

$$|B[u, v]| \leq C||u||_{H^1(\Omega)}||v||_{H^1(\Omega)}. \qquad (2.3.11)$$

Combining (2.3.10), and (2.3.11) applied to $u = v = f$, we get

$$||f||_{H^1(\Omega)} = \infty,$$

i.e. $f \notin H^1(\Omega)$, contradicting the assumption that $f \in \{\psi_n\}^\perp \cap H_0^1(\Omega)$. □

2.3.1 Scaling Properties

In this subsection, our aim is to provide the scaling invariance between μ and Ω. Let $\kappa\Omega$ denote the usual scaling of the domain Ω by a positive real number κ. Let $\tilde{L}$ on $\kappa\Omega$ be given by

$$\tilde{L}u = -\Delta u + \tilde{c}u = -\Delta u + \kappa^{-2}c\left(\frac{x}{\kappa}\right)u,$$

19

If the operator L on Ω has orthonormal eigenpairs $\{(\phi_n(x), \lambda_n)\}_{n=1}^{\infty}$, then

$$\{(\kappa^{-d/2}\phi_n\left(\frac{x}{\kappa}\right), \kappa^{-2}\lambda_n)\}_{n=1}^{\infty}$$

forms the orthonormal set of eigenpairs for $\tilde{L}$. Let $\{\psi_1, \psi_2, \dots, \}$ be the compressed modes of second type (CM-II), corresponding to the operator L. For each $n \in \mathbb{N}$, define $\tilde{\psi}_n(x) = \kappa^{-d/2}\psi_n(\frac{x}{\kappa})$. Then, clearly $\{\tilde{\psi}_1, \tilde{\psi}_2, \dots, \}$ forms an orthonormal system. Moreover,

$$\begin{aligned}
\tilde{B}[\tilde{\psi}_n, \tilde{\psi}_n] = \langle \tilde{L}\tilde{\psi}_n, \tilde{\psi}_n \rangle &= \int_{\kappa\Omega} |\nabla \tilde{\psi}_n|^2 dx + \int_{\kappa\Omega} \tilde{c}\tilde{\psi}_n^2 dx \\
&= \kappa^{-2} \int_{\Omega} |\nabla \psi_n|^2 dx + \kappa^{-2} \int_{\Omega} c\psi_n^2 dx \\
&= \kappa^{-2} B[\psi_n, \psi_n],
\end{aligned}$$

and

$$\|\tilde{\psi}_n\|_1 = \int_{\kappa\Omega} |\tilde{\psi}_n| dx = \kappa^{d/2} \int_{\kappa\Omega} |\psi_n| dx = \kappa^{d/2} \|\psi_n\|_1.$$

Hence, $\{\tilde{\psi}_1, \tilde{\psi}_2, \dots, \}$ is the corresponding CM-II for the operator $\tilde{L}$ on $\kappa\Omega$, with respect to the energy functional

$$\tilde{J}_{\tilde{\mu}}[u] = \tilde{B}[u, u] + \frac{1}{\tilde{\mu}}\|u\|_{L^1},$$

where

$$\tilde{\mu} = \frac{\mu}{\kappa^{2+d/2}}.$$

Notice that this last scaling relation is consistent with the scaling properties of the inequalities in Corollary 2.3.4 and Corollary 2.3.5.

2.4 Applications

We now establish the analogues of the Theorems 2.2.8-2.2.10 for the Compressed Modes and the Compressed Plane Waves. We first provide the precise definitions of Compressed Modes (CM) and Compressed Plane Waves (CPW) as introduced in [33, 34], and establish their connections to the theory we developed in Section 2.2, and then proceed with the verification of the analogous theorems.

2.4.1 Compressed Modes

Compressed Modes are defined via the following minimization procedure

$$\Psi^{(m)} = \{\psi_1^{(m)}, \ldots, \psi_m^{(m)}\} = \operatorname*{argmin}_{h_1, h_2, \ldots, h_m} \sum_{i=1}^{m} J_\mu[h_i] \quad \text{s.t.} \quad \langle h_j, h_k \rangle = \delta_{jk}, \tag{2.4.1}$$

where

$$J_\mu[u] = \frac{1}{\mu}||u||_{L^1} + \left\langle u, \left(-\frac{1}{2}\Delta + V\right)u \right\rangle = \frac{1}{\mu}||u||_{L^1} + \frac{1}{2}||\nabla u||_{L^2}^2 + \int_\Omega V u^2 d\mathbf{x}, \tag{2.4.2}$$

where V is a bounded measurable real-valued function defined on Ω. Here, the quantity $\left\langle u, \left(-\frac{1}{2}\Delta + V\right)u \right\rangle$ corresponds to the bilinear form associated to the elliptic operator $-\frac{1}{2}\Delta + V$. We denote the eigenvalues and eigenfunctions of $-\frac{1}{2}\Delta + V$ by λ_n and ϕ_n, with λ_n's being in increasing order, as usual.

As noted earlier, Compressed Modes of second type $\{\psi_1, \psi_2, \ldots, \psi_m\}$, defined by the variational procedure (2.3.1), being an orthonormal sequence, is in the solution set of the minimization problem (2.4.1), so that

$$\sum_{i=1}^{m} J_\mu[\psi_i^{(m)}] \le \sum_{i=1}^{m} J_\mu[\psi_i].$$

Combining this with the estimate (2.2.12), we obtain

$$\sum_{i=1}^{m} J_\mu[\psi_i^{(m)}] \le m\frac{|\Omega|^{\frac{1}{2}}}{\mu} + \sum_{j=1}^{m} \lambda_j. \tag{2.4.3}$$

The proof of Theorem 2.2.10 relies essentially on the estimation (2.2.23), and the orthonormality of the sequence $\{\psi_1, \ldots, \psi_m\}$. We still have the orthonormality, and the estimation (2.4.3) analogous to (2.2.23). Hence, the following corollary holds.

Corollary 2.4.1 (Corollary to Theorem 2.2.10). *Let V_m be the subspace generated by the Compressed Modes $\Psi = \{\psi_1^{(m)}, \psi_2^{(m)}, \ldots, \psi_m^{(m)}\}$. Then, for any $n \le m$, we have*

$$\sum_{k=1}^{n} d(\phi_k, V_m)^2 \le \frac{m|\Omega|^{\frac{1}{2}}}{\mu(\lambda_{m+1} - \lambda_n)},$$

provided $\lambda_{m+1} \ne \lambda_n$.

From Corollary 2.4.1, we deduce the following approximation result.

Corollary 2.4.2. *Let $\{\phi_k\}_{k \in \mathbb{N}}$ be the (Dirichlet) eigenfunctions of a second-order linear symmetric elliptic operator L, defined on a bounded domain $\Omega \subset \mathbb{R}$. Given any fixed parameter μ, the first m Compressed Modes up to a linear transformation, denoted by $\{\xi_1^{(m)}, \ldots, \xi_m^{(m)}\}$, satisfy*

$$\lim_{m \to \infty} \|\phi_k - \xi_k^{(m)}\|_2 = 0, \ \forall k \in \mathbb{N}.$$

Proof. Let $\xi_k^{(m)}$ denote the projection of ϕ_k onto the linear subspace spanned by $\{\psi_1^{(m)}, \ldots, \psi_m^{(m)}\}$, which we denote by V_m. Then, clearly, $\xi_k^{(m)}$ is a linear combination of $\{\psi_1^{(m)}, \ldots, \psi_m^{(m)}\}$. Furthermore, as a property of the projection, we have

$$d(\phi_k, V_m) = \|\phi_k - \xi_k^{(m)}\|_2,$$

so that Corollary 2.4.1 implies

$$\sum_{k=1}^{n} \|\phi_k - \xi_k^{(m)}\|_2^2 \leq \frac{m|\Omega|^{\frac{1}{2}}}{\mu(\lambda_{m+1} - \lambda_n)}. \tag{2.4.4}$$

As Ω is a bounded domain inside $\mathbb{R}$, By Weyl's law, we know that λ_m grows quadratically in m. Hence, taking the limit of (2.4.4) as $m \to \infty$, we conclude that the summands in the LHS of (2.4.4) decays to zero, i.e.

$$\lim_{m \to \infty} \|\phi_k - \xi_k^{(m)}\|_2 = 0, \tag{2.4.5}$$

as desired. $\square$

Corollary 2.4.2 can be viewed as a completeness result, since (2.4.5) yields that any eigenfunction ϕ_k is well approximated by its projection $\xi_k^{(m)}$ onto V_m. In this sense, V_m's trace the full space as $m \to \infty$.

2.4.2   Compressed Plane Waves

The construction of Compressed Plane Waves is closely related to that of Compressed Modes, where both involve minimizing a certain functional. The difference is that Compressed Plane

Waves have multi-resolution capabilities, which is achieved by adding the shift-orthogonality constraints. Let $\mathbf{w} = (\mathbf{w_1}, \ldots, \mathbf{w_d}) \in \mathbb{R}_+^{\mathbf{d}}$ be a basis of a d-dimensional lattice and let Ω be a rectangular box with

$$\Omega = [0, n_1 w_1] \times \cdots \times [0, n_d w_d], \quad (n_1, \ldots, n_d) \in \mathbb{N}^d.$$

Define the lattice

$$\Gamma_{\mathbf{w}} = \{\mathbf{jw} := (\mathbf{j_1 w_1}, \ldots, \mathbf{j_d w_d}) | \mathbf{0} \le \mathbf{j_1} < \mathbf{n_1}, \ldots, \mathbf{0} \le \mathbf{j_d} < \mathbf{n_d}\}.$$

The first n Basic Compressed Plane Waves (BCPWs) $\{\psi^k\}_{k=1}^n$, are defined via

$$\psi^1 = \underset{\psi}{\mathrm{argmin}} \; J_\mu[\psi] \;\; \text{s.t.} \;\; \langle \psi(\mathbf{x}), \psi(\mathbf{x} - \mathbf{jw}) \rangle = \delta_{\mathbf{j}, \mathbf{0}} \; \forall \mathbf{jw} \in \Gamma_{\mathbf{w}};$$

$$\psi^k = \underset{\psi}{\mathrm{argmin}} \; J_\mu[\psi] \;\; \text{s.t.} \;\; \begin{cases} \langle \psi(\mathbf{x}), \psi(\mathbf{x} - \mathbf{jw}) \rangle = \delta_{\mathbf{j}, \mathbf{0}} \; \forall \mathbf{jw} \in \Gamma_{\mathbf{w}} \\ \langle \psi(\mathbf{x}), \psi^{\mathbf{i}}(\mathbf{x} - \mathbf{jw}) \rangle = \mathbf{0} \quad \forall \mathbf{i} : \mathbf{0} < \mathbf{i} < \mathbf{k}, \end{cases}$$

where the functional J_μ is defined by

$$J_\mu[u] = \frac{1}{\mu}||u||_{L^1} + \langle u, -\frac{1}{2}\Delta u \rangle = \frac{1}{\mu}||u||_{L^1} + \frac{1}{2}||\nabla u||_{L^2}^2.$$

Notice that this functional is a special case of the functional (2.4.2), with $V \equiv 0$.

The translations of the BCPWs on the lattice $\Gamma_{\mathbf{w}}$ produce all CPWs. Unlike Compressed Modes that are solved in a single minimization problem, the Compressed Plane Waves are constructed hierarchically. This is similar to the shift-orthogonal wavelets [45], but a distinction of CPWs is that it is adapted to the Laplace operator.

Existence of CPW's essentially follows from the observation that shift orthogonality (i.e the constraints in the definition of BCPWs) is preserved under L^2-limits, so that any minimizing sequence has a subsequential limit, which still satisfies the shift orthogonality properties. The following theorem (see for example [6]) characterizes any orthonormal sequence of shift orthogonal functions.

Theorem 2.4.3. *Let $\Omega \subset \mathbb{R}^d$, and the lattice $\Gamma_{\mathbf{w}}$ be defined as above. Let $\{\xi^k\}_{k=1}^\infty$ be an orthonormal sequence of shift orthogonal functions. Then, the (Hilbert) space $\mathcal{H} = L^2(\Omega)$*

can be written as a direct sum

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \ldots \oplus \mathcal{H}_N,$$

where each $\mathcal{H}_k$ is the Hilbert space spanned by some eigenfunctions for the Laplace equation in the rectangular box Ω, with the property that if ξ^k has the decomposition

$$\sqrt{N}\xi^k = e_1^k + e_2^k + \cdots + e_N^k, \quad e_j^k \in \mathcal{H}_j,$$

then $E = \{e_j^k | j = 1, 2, \ldots, N; k \in \mathbb{N}\}$ forms an orthonormal system in $\mathcal{H}$. Furthermore, $N = n_1 n_2 \cdots n_d = |\Gamma_{\mathbf{w}}|$.

A detailed discussion of Theorem 2.4.3, with a characterization of the Hilbert spaces $\mathcal{H}_k$ is given in the Appendix A.

Remark 2.4.4. In Theorem 2.4.3, for a fixed k, both $\{\xi_{\mathbf{j}}^k | \mathbf{j}\mathbf{w} \in \Gamma_{\mathbf{w}}\}$, and $\{e_j^k | j = 1, 2, \ldots, N\}$ form an orthonormal system, and have the same cardinality, hence their linear span agree. Therefore,

$$\text{span}\{\xi_{\mathbf{j}}^k | \mathbf{j}\mathbf{w} \in \Gamma_{\mathbf{w}}; \mathbf{k} = 1, \ldots, \mathbf{M}\} = \text{span}\{\mathbf{e}_{\mathbf{j}}^{\mathbf{k}} | \mathbf{j} = 1, \ldots, \mathbf{N}; \mathbf{k} = 1, \ldots, \mathbf{M}\}$$

for any $M \in \mathbb{N} \cup \{\infty\}$.

With this remark and Theorem 2.4.3, instead of working with the CPW's $\{\psi_{\mathbf{j}}^k\}$, it is natural to switch to $\{e_j^k\}$ for completeness results. Let's define

$$J_\infty[u] = \frac{1}{2}||\nabla u||_{L^2(\Omega)}^2.$$

Then, we can write

$$J_\mu[u] = J_\infty[u] + \frac{1}{\mu}||u||_{L^1}.$$

As the Hilbert spaces $\mathcal{H}_j$ are the span of some eigenfunctions of the Laplacian, the functional J_∞ satisfies the following linearity property

$$J_\infty[e_1 + e_2 + \cdots + e_N] = J_\infty[e_1] + J_\infty[e_2] + \ldots + J_\infty[e_N], \quad e_j \in \mathcal{H}_j, \ j = 1, 2, \ldots, N.$$

Now, the functions $\{e_j^k\}$ can be regarded as the solutions to the following problems

$$\{e_1^1, e_2^1, \ldots, e_N^1\} = \operatorname*{argmin}_{\substack{f_j \in \mathcal{H}_j \\ ||f_j||_2 = 1}} J_\infty[f_1] + \ldots + J_\infty[f_N] + \frac{1}{\mu}||f_1 + \cdots + f_N||_1,$$

$$\{e_1^k, e_2^k, \ldots, e_N^k\} = \operatorname*{argmin}_{\substack{f_j \in \{e_j^1, \ldots, e_j^{k-1}\}^\perp \cap \mathcal{H}_j \\ ||f_j||_2 = 1}} J_\infty[f_1] + \ldots + J_\infty[f_N] + \frac{1}{\mu}||f_1 + \cdots + f_N||_1.$$

This is analogous to the variational procedure (2.2.1), except that at each step in the minimization, we obtain multiple functions. Nevertheless, we might regard one particular e_j^k, say e_1^k for simplicity, as the solution to the following minimization problem over $\mathcal{H}_1$

$$e_1^k = \operatorname*{argmin}_{\substack{f \in \{e_1^1, \ldots, e_1^{k-1}\}^\perp \cap \mathcal{H}_1 \\ ||f||_2 = 1}} J_\infty[f] + J_\infty[e_2^k] + \ldots + J_\infty[e_N^k] + \frac{1}{\mu}||f + e_2^k + \cdots + e_N^k||_1.$$

We still have the boundedness of the penalty term $P(f) = \frac{1}{\mu}||f + e_2^k + \cdots + e_N^k||_1$, as

$$\frac{1}{\mu}||f + e_2^k + \cdots + e_N^k||_1 \leq \frac{1}{\mu}\left(||f||_1 + ||e_2^k||_1 \ldots + ||e_N^k||_1\right)$$

$$(\text{Cauchy-Schwarz}) \leq \frac{|\Omega|^{\frac{1}{2}}}{\mu}\left(||f||_2 + ||e_2^k||_2 \ldots + ||e_N^k||_2\right)$$

$$= \frac{N|\Omega|^{\frac{1}{2}}}{\mu},$$

i.e.

$$||P|| \leq \frac{N|\Omega|^{\frac{1}{2}}}{\mu}. \tag{2.4.6}$$

Therefore, $\{e_j^k\}_{k \in \mathbb{N}}$ could be viewed as the solutions to an analogue of the variational procedure (2.2.1), in the Hilbert space $\mathcal{H}_j$, with the linear functional being the restriction of $-\frac{1}{2}\Delta$ on $\mathcal{H}_j$.

Next, let's enumerate the eigenfunctions forming each Hilbert space $\mathcal{H}_j$ as follows

$$\mathcal{H}_j = \operatorname{span}\{\phi_j^k | k = 1, 2, \ldots\},$$

$$-\frac{1}{2}\Delta\phi_j^k = \lambda_j^k \phi_j^k \tag{2.4.7}$$

$$\lambda_j^1 \leq \lambda_j^2 \leq \ldots.$$

The following theorem establishes an analogue of the Weyl's Law.

Theorem 2.4.5. *Let λ_j^k be defined as in (2.4.7), then*

$$\lambda_j^k = \frac{N(2\pi)^d}{2\omega_d|\Omega|}k^{2/d} + o(k^{2/d}), \ \ as \ k \to \infty,$$

where ω_d denotes the volume of the unit ball in $\mathbb{R}^d$.

A discussion of Theorem 2.4.5 can be found in Appendix A.

Overall, we verified that the functions $\{e_j^k\}$ are obtained via a variational procedure analogous to (2.2.1). We also noted in Remark 2.4.4 that the spans of $\{e_j^k\}$, and CPWs agree. Therefore, the theory developed in Section 2.2 applies to CPWs, so that we obtain the following corollaries as direct consequences of the Theorems 2.2.8, 2.2.9, and 2.2.10.

Corollary 2.4.6 (Corollary to Theorem 2.2.8). *Let Ω be a bounded interval in $\mathbb{R}$, that is an integer multiple of some lattice $\Gamma_{\mathbf{w}}$. Then, for any parameter μ, the set of Compressed Plane Waves $\{\psi_{\mathbf{j}}^k\}$ defined on Ω forms a complete orthonormal system in $L^2(\Omega)$.*

Proof. Notice that $\{e_j^k\}_{k\in\mathbb{N}} \subset \mathcal{H}_j$ are obtained as the solutions to a variational problem in $\mathcal{H}_j$, analogous to the variational procedure (2.2.1). Furthermore, since Ω lies inside $\mathbb{R}$, by Theorem 2.4.5, the corresponding eigenvalues grow super-linearly. Therefore, by Theorem 2.2.8, $\{e_j^k\}_{k\in\mathbb{N}}$ forms a complete orthonormal system in $\mathcal{H}_j$, for each $j = 1, 2, \ldots, N$. Finally, by Remark 2.4.4, $\{\psi_{\mathbf{j}}^k\}$ is a complete orthonormal system in $\mathcal{H}$. $\square$

Corollary 2.4.7 (Corollary to Theorem 2.2.9). *Let Ω be a rectangular domain inside $\mathbb{R}^2$, that is an integer multiple of some lattice $\Gamma_{\mathbf{w}}$. Then, for any parameter μ satisfying*

$$\mu > \frac{|\Omega|^{\frac{3}{2}}}{2\pi}, \tag{2.4.8}$$

the set of Compressed Plane Waves $\{\psi_{\mathbf{j}}^k\}$ defined on Ω forms a complete orthonormal system in $L^2(\Omega)$.

Proof. Since Ω lies inside $\mathbb{R}^2$, notice by Theorem 2.4.5 that the corresponding eigenvalues grow linearly, with the linearity constant $\frac{2\pi N}{|\Omega|}$. The proof proceeds analogous to the proof of Corollary 2.4.6, except that we rather apply Theorem 2.2.9. Note that the bound for the underlying penalty term is provided in (2.4.6). Hence, whenever the inequality (2.4.8) holds,

26

the assumption (2.2.20) in Theorem 2.2.9 is satisfied. Therefore, Theorem 2.2.9 yields the completeness of $\{\psi_{\mathbf{j}}^k\}$, as desired. $\qquad\square$

Corollary 2.4.8 (Corollary to Theorem 2.2.10)**.** *Let V_j^m be the subspace generated by the functions $\{e_j^1, e_j^2, \ldots, e_j^m\}$. Then, for any $n \leq m$, we have*

$$\sum_{k=1}^{n} d(\phi_j^k, V_j^m)^2 \leq \frac{mN|\Omega|^{\frac{1}{2}}}{\mu(\lambda_j^{m+1} - \lambda_j^n)},$$

provided $\lambda_j^{m+1} \neq \lambda_j^n$. Defining V^m via

$$V^m = \operatorname{span}\{\psi_{\mathbf{j}}^k | \mathbf{j} \in \mathbb{Z}^{\mathbf{d}}, \mathbf{k} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{m}\},$$

we further have

$$\sum_{k \leq m, j \leq N} d(\phi_j^k, V^m)^2 = \frac{mN|\Omega|^{\frac{1}{2}}}{\mu} \sum_{j \leq N} \frac{1}{\lambda_j^{m+1} - \lambda_j^n}.$$

2.5   Conclusions

In this chapter, we established a functional analytic framework to study the completeness properties of solutions of optimization problems that are formulated as a perturbation of the Courant-Fischer variational problem. This framework enabled us to verify the completeness properties of CM-I, CM-II, and CPW. In particular, for dimension $d = 1$, the completeness is proved unconditionally, and for dimension $d = 2$, it is shown that CM-II, and CPW form a complete basis provided that the underlying penalization term is sufficiently small. These results confirmed the conjecture posed in [33, 34] for dimensions $d = 1, 2$, whereas for higher dimensions the problem still remains open.

CHAPTER 3

Quantitative Analysis of Energy

3.1 An Analogue of Weyl's Law for Compressed Modes of second type

In this section, we would like to prove an analogue of the Weyl's law for the Compressed Modes of second type (CM-II) defined in Section 2.3. Recall that Weyl's law provides the asymptotic behavior of the (Dirichlet) eigenvalues of a second order symmetric elliptic linear operator. The eigenvalues of a second order symmetric elliptic linear operator, T, also corresponds to the values of the bilinear form $\langle Tu, u \rangle$ evaluated at the (normalized) eigenvectors

$$\lambda_n = \langle T\phi_n, \phi_n \rangle$$

In a similar construction, we consider the "energy" associated with each of the CM-II ζ_n defined by

$$\kappa_n = \langle T\zeta_n, \zeta_n \rangle + \frac{1}{\mu}\|\zeta_n\|_1$$

and consider the asymptotic behavior of the sequence κ_n. The following theorem provides the precise asymptotic behavior of κ_n by relating it to the associated Weyl's Law.

Theorem 3.1.1. *Let $\{\lambda_n\}_{n\in\mathbb{N}}$ be the (Dirichlet) eigenvalues of a second-order linear symmetric elliptic operator T on a bounded domain $\Omega \subset \mathbb{R}^d$. Suppose that C_Ω is the implicit constant in the Weyl's law for $\{\lambda_n\}_{n\in\mathbb{N}}$, so that*

$$\lambda_n = C_\Omega n^{\frac{2}{d}} + O(n^{\frac{1}{d}}) \ \text{as } n \to \infty.$$

Let $\{\kappa_n\}_{n\in\mathbb{N}}$ be the associated energies corresponding to the CM-II, in the sense that

$$\kappa_n = J[\zeta_n] := \langle T\zeta_n, \zeta_n \rangle + \frac{1}{\mu}\|\zeta_n\|_1.$$

Then, $\{\kappa_n\}_{n\in\mathbb{N}}$ *satisfies the following analogue of the Weyl's law*

$$\kappa_n = C_\Omega n^{\frac{2}{d}} + O(n^{\frac{3}{2d}}) \ as \ n \to \infty.$$

For the proof of Theorem 3.1.1, we analyze the cumulative sum of κ_n's, i.e. the function $F : \mathbb{R}^+ \to \mathbb{R}$ defined on integers via

$$F(n) = \sum_{m\leq n} \kappa_m, \forall n \in \mathbb{N},$$

and extended to the positive real numbers by interpolation. Notice that the above cumulative sum of energies are also considered in Lemma 2.2.7 for the completeness results. We also make use of those earlier estimates for F. Furthermore, due to the positivity and the increasing nature of the energies κ_n, the function F is a convex function. The following lemma provides an estimate for the derivative of convex functions with certain properties, and will be useful for the proof of Theorem 3.1.1.

Lemma 3.1.2. *Let* $F : \mathbb{R}^+ \to \mathbb{R}$ *be a convex function satisfying*

$$F(t) = t^\alpha + O(t^\beta), \ \ as \ t \to \infty,$$

where the real numbers α, *and* β *satisfies* $\alpha > 1$, *and* $\alpha > \beta > 0$. *Suppose further that* F *is differentiable on* $\mathbb{R}^+ \setminus E$, *where* E *is a discrete set. Then,*

$$F'(t) = \alpha t^{\alpha-1} + O(t^{\frac{\alpha+\beta}{2}-1}), \ \ as \ t \in \mathbb{R}^+ \setminus E \ and \ t \to \infty.$$

The proof of Lemma 3.1.2 consists of standard asymptotic analysis arguments, and is given in Appendix C. Now, we are ready to present the proof of Theorem 3.1.1.

Proof of Theorem 3.1.1. As noted earlier, we consider the function $F : \mathbb{R}^+ \to \mathbb{R}$ defined via

$$F(n) = \sum_{m\leq n} \kappa_m, \forall n \in \mathbb{N}$$

on integers, and extended to the whole positive real axis by linear interpolation. F is a convex function, and satisfies

$$\lim_{x\uparrow n} F'(x) = \kappa_n$$

29

From Lemma 2.2.7, we have

$$\sum_{m \leq n} \lambda_m \leq F(n) \leq \sum_{m \leq n} \lambda_m + n\|P\|, \tag{3.1.1}$$

where P is the underlying penalization operator (i.e. the L^1 norm).

The Weyl's law

$$\lambda_n = C_\Omega n^{\frac{2}{d}} + O(n^{\frac{1}{d}})$$

can be summed into

$$\sum_{m \leq n} \lambda_m = C_\Omega \sum_{m \leq n} m^{\frac{2}{d}} + O(n^{\frac{d+1}{d}}), \tag{3.1.2}$$

and the resulting quantity can be approximated by the standard integral test as

$$\frac{d}{d+2} n^{\frac{d+2}{d}} = \int_0^n x^{\frac{2}{d}} dx \leq \sum_{m \leq n} m^{\frac{2}{d}} \leq \int_0^{n+1} x^{\frac{2}{d}} dx = \frac{d}{d+2}(n+1)^{\frac{d+2}{d}},$$

so that

$$\sum_{m \leq n} m^{\frac{2}{d}} = \frac{d}{d+2} n^{\frac{d+2}{d}} + O(n^{\frac{2}{d}}).$$

Finally, since the remainder term in the last asymptotic relation is dominated by the remainder in (3.1.2), we obtain

$$\sum_{m \leq n} \lambda_m = \frac{d}{d+2} C_\Omega n^{\frac{d+2}{d}} + O(n^{\frac{d+1}{d}}). \tag{3.1.3}$$

Next, let's consider (3.1.1). By the asymptotic formula (3.1.3) for the cumulative sum of eigenvalues, both sides of (3.1.1) are controlled by the same quantity, so that we obtain

$$F(n) = C_\Omega n^{\frac{d+2}{d}} + O(n^{\frac{d+1}{d}}). \tag{3.1.4}$$

Now, applying Lemma 3.1.2 to F in the light of (3.1.4), we obtain

$$\kappa_n = \lim_{x \uparrow n} F'(x) = C_\Omega n^{\frac{2}{d}} + O(n^{\frac{3}{2d}}),$$

as desired. $\qquad \square$

Notice that the proof of Theorem 3.1.1 only requires the boundedness of the penalty term P, hence it holds for a general class of variational problems as long as the existence criteria (2.2.2) given in Chapter 2 is satisfied.

3.2 Consistency of Eigenvalues

As CM-II are defined by a modification of the variational principle for the eigenvalues of a second order symmetric elliptic linear operator, it is a question of interest to figure out whether the energies of the CM-II converges to the true eigenvalues as the L^1 regularization term approaches to zero. Specifically, for a second order symmetric elliptic linear operator T, we consider the energy associated to the CM-II $\zeta_n = \zeta_n(\mu)$ as a function of the regularization parameter μ

$$\kappa_n(\mu) = \langle T\zeta_n, \zeta_n \rangle + \frac{1}{\mu}\|\zeta_n\|_1. \tag{3.2.1}$$

The following theorem provides the rate of convergence for $\kappa_n(\mu)$ to the true eigenvalues λ_n.

Theorem 3.2.1. *Let $\{\lambda_n\}_{n\in\mathbb{N}}$ be the (Dirichlet) eigenvalues of a second order linear symmetric elliptic operator on a bounded domain $\Omega \subset \mathbb{R}^d$. Let $\{\kappa_n(\mu)\}_{n\in\mathbb{N}}$ be the energies corresponding to the CM-II defined as in (3.2.1). Then,*

$$\lambda_n - (2^{n-1} - 1)\frac{\|P\|}{\mu} \leq \kappa_n(\mu) \leq \lambda_n + 2^{n-1}\frac{\|P\|}{\mu} \quad \forall n \in \mathbb{N}, \tag{3.2.2}$$

where $\|P\|$ denotes the operator norm of the L^1 norm.

Proof. The analysis is based on the following cumulative energy sum, which has also been considered in Section 3.1

$$F(n) = \sum_{m \leq n} \kappa_m(\mu), \forall n \in \mathbb{N}$$

From Lemma 2.2.7, we have

$$\sum_{m \leq n} \lambda_m \leq F(n) \leq \sum_{m \leq n} \lambda_m + n\frac{\|P\|}{\mu} \tag{3.2.3}$$

We prove (3.2.2) by a generalized induction argument. For $n = 1$, the induction hypothesis follows simply by (3.2.3). Now, assume that (3.2.2) holds for $n = 1, 2, \ldots, r - 1$. Then,

$F(r-1)$ satisfies the following upper and lower bounds

$$
\begin{aligned}
F(r-1) &\leq \sum_{m \leq r-1} (\lambda_m + 2^{m-1} \frac{\|P\|}{\mu}) \\
&= \sum_{m \leq r-1} \lambda_m + \frac{\|P\|}{\mu} \sum_{m \leq r-1} 2^{m-1} \\
&= \sum_{m \leq r-1} \lambda_m + (2^{r-1} - 1) \frac{\|P\|}{\mu}
\end{aligned}
\tag{3.2.4}
$$

$$
\begin{aligned}
F(r-1) &\geq \sum_{m \leq r-1} (\lambda_m - (2^{m-1} - 1) \frac{\|P\|}{\mu}) \\
&= \sum_{m \leq r-1} \lambda_m - \frac{\|P\|}{\mu} \sum_{m \leq r-1} 2^{m-1} - 1 \\
&= \sum_{m \leq r-1} \lambda_m - (2^{r-1} - r) \frac{\|P\|}{\mu}
\end{aligned}
\tag{3.2.5}
$$

By (3.2.3), we have

$$
\sum_{m \leq r} \lambda_m - F(r-1) \leq \kappa_r(\mu) \leq \sum_{m \leq r} \lambda_m - F(r-1) + r \frac{\|P\|}{\mu}.
$$

Now, substituting the upper and lower bounds (3.2.4)-(3.2.5) in the above inequality for the respective sides, we obtain

$$
\lambda_r - (2^r - 1) \frac{\|P\|}{\mu} \leq \kappa_r(\mu) \leq \lambda_r - 2^r \frac{\|P\|}{\mu},
$$

as desired. □

By taking limit as $\mu \to \infty$ in (3.2.2), one obtains the following corollary.

Corollary 3.2.2. *Let $\{\lambda_n\}_{n \in \mathbb{N}}$ be the (Dirichlet) eigenvalues of a second-order linear symmetric elliptic operator on a bounded domain $\Omega \subset \mathbb{R}^d$. Let $\{\kappa_n(\mu)\}_{n \in \mathbb{N}}$ be the energies corresponding to the CM-II defined as in (3.2.1). Then,*

$$
\lim_{\mu \to \infty} \kappa_n(\mu) = \lambda_n.
$$

Remark 3.2.3. The convergence rate (3.2.3) is exponentially slow in terms of n, the order of the CM-II, hence is practical for only the lower order CM-II. On the other hand, the Weyl's law and Theorem 3.1.1 implies that

$$\lambda_n = C_\Omega n^{\frac{2}{d}} + O(n^{\frac{1}{d}}) \text{ as } n \to \infty,$$

$$\kappa_n(\mu) = C_\Omega n^{\frac{2}{d}} + O(n^{\frac{3}{2d}}) \text{ as } n \to \infty,$$

so that

$$|\kappa_n(\mu) - \lambda_n| = O(n^{\frac{3}{2d}}) \text{ as } n \to \infty,$$

hence suggesting that the convergence rate (3.2.3) would be polynomial for large values of n.

3.3   Conclusions

The analysis in this chapter has established the asymptotic behavior of the energies of CM-II as the depth of the functions approaches to infinity, or the penalization term approaches to zero. In particular, the asymptotic behavior as depth tends to infinity agrees with the associated Weyl's Law, hence implying the consistency with the unperturbed eigenvalue problem. As the penalization term vanishes, the resulting limit is found to be the corresponding eigenvalue, hence establishing the stability of the energies.

CHAPTER 4

Analysis of Euler Lagrange Equation

In this chapter, we consider the Euler-Lagrange equations satisfied by the Compressed Modes and Compressed Plane Waves, and derive certain properties of them by analyzing the corresponding differential equation. The study of Euler-Lagrange equations will be useful for the study of the asymptotic behavior of the support of the Compressed Modes, as well as the regularity properties of the Compressed Modes and Compressed Plane Waves.

4.1   Derivation of the Euler-Lagrange Equations

As the associated variational problems for the Compressed Modes and Compressed Plane Waves contain the L^1 regularization term, the Euler-Lagrange equations satisfied by these functions involve the subgradient of the absolute value function $f(x) = |x|$. We denote this subgradient term by p, where

$$p(u) = \begin{cases} -1, \text{ if } u < 0 \\ \in [-1, 1] \text{ if } u = 0 \\ 1, \text{ if } u > 0. \end{cases}$$

4.1.1   Euler-Lagrange Equation for Compressed Modes of Type One

Recall that the first N of the Compressed Modes for the second order linear symmetric elliptic differential operator $T = -\Delta + V(x)$, is calculated by the following variational problem

$$\{\psi_1, \psi_2, \ldots, \psi_N\} = \operatorname*{argmin}_{\tilde{\psi}_1, \tilde{\psi}_2, \ldots, \tilde{\psi}_N} \sum_{i=1}^{N} \int_{\mathbb{R}^d} \left(\frac{1}{\mu} |\tilde{\psi}_i| + \frac{1}{2} |\nabla \tilde{\psi}_i|^2 + V(x) \tilde{\psi}_i^2 \right) dx \text{ s.t. } \int_{\mathbb{R}^d} \tilde{\psi}_j \tilde{\psi}_k dx = \delta_{jk}$$

Now, incorporating the Lagrange multipliers associated to the above normalization and orthogonality constraints, we obtain the following Euler-Lagrange Equation for $\psi_i, i = 1, \ldots, N$

$$-\Delta\psi_i + 2V(x)\psi_i - 2\lambda_i\psi_i + \frac{1}{\mu}p(\psi_i) = \sum_{j \neq i} \lambda_{ij}\psi_j \qquad (4.1.1)$$

4.1.2 Euler-Lagrange Equation for Compressed Modes of Type Two

Similarly, the ith CM-II, ζ_i, is given by

$$\zeta_i = \underset{\zeta}{\operatorname{argmin}} \int_{\mathbb{R}^d} \left(\frac{1}{\mu}|\zeta| + \frac{1}{2}|\nabla\zeta|^2 + V(x)\zeta^2 \right) dx \text{ subject to } \begin{cases} \langle\zeta,\zeta\rangle = 1 \\ \langle\zeta,\zeta_m\rangle = 0 \text{ for } m < i \end{cases}$$

Now, since the ith CM-II only depends on the modes that are in the lower level, the following Euler-Lagrange Equation is satisfied by ζ_i

$$-\Delta\zeta_i + 2V(x)\zeta_i - 2\lambda_i\zeta_i + \frac{1}{\mu}p(\zeta_i) = \sum_{j<i} \lambda_{ij}\zeta_j \qquad (4.1.2)$$

Notice that the Euler-Lagrange equations for CM-I and CM-II are almost equivalent, except that the Lagrange multipliers λ_{ij} vanishes for CM-II whenever $i < j$.

4.1.3 Euler-Lagrange Equation for Compressed Plane Waves

The definition of CPWs are similar to CMs, where, in addition to the orthogonality constraints, the constraints imposed by the shift-orthogonality properties are also required. Furthermore, the differential operator is assumed to be $-\frac{1}{2}\Delta$. Namely,

$$\psi^n = \underset{\psi}{\operatorname{argmin}}\|\nabla\psi\|_2^2 + \frac{1}{\mu}\|\psi\|_1 \text{ subject to } \begin{cases} \langle\psi,\psi\rangle = 1 \\ \langle\psi,\psi_k\rangle = 0 \text{ for } 0 < k < N \\ \langle\psi,\psi_k^m\rangle = 0 \text{ for } m < n, 0 \leq k < N \end{cases}$$

For CPWs, the Lagrange multipliers are included both for the orthogonality constraints against the previously obtained functions, as well as the shifts of the function itself. Therefore, the Euler-Lagrange equation is given by

$$-\Delta\psi^n + \frac{1}{\mu}p(\psi^n) = \sum_{k=0}^{N-1} \lambda_k\psi_k + \sum_{m<n}\sum_{k=0}^{N-1} \lambda_k^m\psi_k^m. \qquad (4.1.3)$$

35

4.2   Regularity Results

The regularity properties of minimizers have been studied in the context of calculus of variations for linear and non-linear problems [2, 3, 21, 32]. The main method to obtain regularity of the solutions is the study of the associated Euler-Lagrange equation. The variational formulations for Compressed Modes and Compressed Plane Waves have solutions in the Sobolev Space $H^1(\Omega)$, hence the associated Euler-Lagrange Equations (4.1.1)-(4.1.3) are satisfied in the weak sense. In this section, we would like to verify higher order regularity such as classical differentiability of the solutions. We carry out this by the elliptic regularity theorems. One obstacle while using the elliptic regularity theorems would be the irregularities associated with the subdifferential term p. Therefore, the analysis of the term p constitutes a significant part of this section. We begin with stating the relevant theorems regarding the elliptic regularity, which will be used throughout this section. The following theorem yields the regularity of the solutions of the elliptic PDE's in the Sobolev sense.

Theorem 4.2.1 (Elliptic Regularity Theorem, see e.g. Section 6.3 in [22]). *Let L be an elliptic operator with C^∞ coefficients. Let $f \in H^m(U)$, where U is a bounded domain in $\mathbb{R}^d$, with C^1 boundary. Then, the solution $u \in H^1(U)$ to the PDE*

$$Lu = f \ in \ U$$

satisfies $u \in H^{m+2}_{loc}(U)$.

The following theorem converts the Sobolev regularity into classical regularity.

Theorem 4.2.2 (Sobolev Embedding Theorem, see e.g. Section 6.3 in [1]). *Let $u \in H^k(U)$, where U is a bounded domain in $\mathbb{R}^d$, with C^1 boundary. If*

$$k > \frac{d}{2},$$

then,

$$u \in C^{k - \lfloor \frac{d}{2} \rfloor - 1}(\bar{U}).$$

4.2.1 Problem Statement

The regularity treatment will be a rather generalized one, where we consider a generalized version of the Euler-Lagrange Equations (4.1.1)-(4.1.3). In its most general formulation, the Euler-Lagrange Equations (4.1.1)-(4.1.3) take the following form

$$Lu + \frac{1}{\mu}p(u) = Tu + F, \tag{4.2.1}$$

where L is a second order elliptic operator, p is the subdifferential of the absolute value function, F is a function possessing similar regularity properties as u (possibly a linear combination), and finally Tu is some (finite) linear combination of translates of u, which can be written as

$$Tu(x) = \sum_{i=1}^{k} c_i u(x - w_i), \tag{4.2.2}$$

for constants $c_1, \ldots, c_k$, and non-zero shift parameters $w_1, \ldots, w_k$.

For the Euler-Lagrange equations (4.1.1)-(4.1.3), the operator L is the original differential operator minus the Lagrange multiplier associated to the equation. For Compressed Modes, the term $Tu \equiv 0$, and F stands for the linear combination of the modes obtained at the previous levels. Whereas for the Compressed Plane Waves, Tu denotes the linear combinations of the shifts of the function, and F denotes the Compressed Plane Waves belonging to the previous levels. Throughout the regularity analysis of the differential equation (4.2.1), the main focus will be the regularity properties of the $p(u)$ term. By the following lemma, we show that the regularity analysis would be straightforward without the $p(u)$ term in the Euler-Lagrange equation.

Lemma 4.2.3. *Suppose that $u \in H^1(\Omega)$ solves the following differential equation*

$$Lu = Tu + F, \tag{4.2.3}$$

where Tu is a finite linear combination of translations of u as given by (4.2.2) for rectangular domains Ω, and $Tu \equiv 0$ otherwise. Suppose further that F possesses the same regularity properties as u, in the sense that

$$u \in H_{loc}^k(\Omega) \implies F \in H_{loc}^k(\Omega).$$

Then, $u \in C^\infty(\Omega)$.

Proof. The proof is a standard application of Elliptic Regularity Theorem (Theorem 4.2.1) to bootstrap the regularity of the solution u. Since the translation operator is a simple change of coordinates, it has no effect on regularity,

$$u \in H^k_{loc}(\Omega) \implies Tu \in H^k_{loc}(\Omega),$$

so that the equation (4.2.3) can be rewritten as

$$Lu = g,$$

where

$$u \in H^k_{loc}(\Omega) \implies g \in H^k_{loc}(\Omega).$$

In particular, $g \in H^1(\Omega)$, and hence by Theorem 4.2.1, $u \in H^3_{loc}(\Omega)$. Proceeding similarly, we obtain

$$u \in H^{2k+1}_{loc}(\Omega) \quad \forall k \in \mathbb{N},$$

or equivalently

$$u \in H^\infty_{loc}(\Omega).$$

Finally, by Sobolev Embedding Theorem (Theorem 4.2.2), we convert this last regularity result into

$$u \in C^\infty(\Omega),$$

as desired. □

For Compressed Modes, when there is no L^1 regularization term, the solutions are simply the eigenvectors of the underlying operator, and hence the $C^\infty(\Omega)$ regularity property is immediate. Similarly, if there is no L^1 regularization term in the definition of the Compressed Plane Waves, we obtain the Shift-Orthogonal Plane Waves (SOPW), which are represented as the finite linear combinations of Fourier modes [6], hence belong to the class $C^\infty(\Omega)$. Nevertheless, the proof of Lemma 4.2.3 is particularly important for the development of the rest of this section. Now, if we rewrite the equation (4.2.1) as

$$Lu = Tu + F - \frac{1}{\mu}p(u), \tag{4.2.4}$$

38

we can utilize the elliptic regularity theorem (Theorem 4.2.1) similarly as in the proof of Lemma 4.2.3 to bootstrap the regularity of u. The main difficulty is the irregularity of the term $p(u)$. Hence the zeros of u will have a particular importance. At this point, it is worth to mention that as an element of $H^1(\Omega)$, u does not necessarily assume well-defined pointwise values. Nevertheless, for $d \leq 3$, the solutions to the PDE (4.2.1) are continuous functions, as verified by the following lemma.

Lemma 4.2.4. *Let $\Omega \subset \mathbb{R}^d$ for $d \leq 3$, and suppose that u is a weak solution to the PDE (4.2.1). Then, $u \in C(\Omega)$.*

Proof. The term $p(u)$ is bounded, hence is square integrable. As a weak solution $u \in H^1(\Omega)$, therefore the RHS of (4.2.4) is also square integrable. Hence, by Theorem 4.2.1, $u \in H^2_{loc}(\Omega)$. For $d \leq 3$ by Theorem 4.2.2, the H^2-regularity of u can be converted into the classical regularity to obtain $u \in C(\Omega)$ ☐

For the case $d = 1$, Lemma 4.2.4 can be improved to provide higher order regularity.

Lemma 4.2.5. *Let $\Omega \subset \mathbb{R}$, and suppose u is a weak solution to the PDE (4.2.1). Then,*

(i) $u \in C^1(\Omega)$,

(ii) $u \in C^3(\Omega \setminus \{u = 0\})$.

Proof. By the proof of Lemma 4.2.4, $u \in H^2_{loc}(\Omega)$. Therefore, Theorem 4.2.2 applied with $d = 1$ yields (i). By continuity of u, $p(u) \in C^\infty_{loc}(\Omega \setminus \{u = 0\})$, and hence the RHS of (4.2.4) belongs to the class $H^2_{loc}(\Omega \setminus \{u = 0\})$. Now, by Theorem 4.2.1, $u \in H^4_{loc}(\Omega \setminus \{u = 0\})$. Finally, applying Theorem 4.2.2 with $k = 4$, and $d = 1$, we obtain $u \in C^3(\Omega \setminus \{u = 0\})$, proving (ii). ☐

The rest of this section is devoted to the the special case where $L = -\Delta$, and $d = 1$. In this case, the PDE (4.2.1) reads

$$-\partial_{xx}u + \frac{1}{\mu}p(u) = Tu + F. \tag{4.2.5}$$

The following regularity result is the main result for the rest of this section.

Theorem 4.2.6. *Let $u \in H^1(\Omega)$ be a solution to the PDE (4.2.1), where Ω is a subset of $\mathbb{R}$. Then, $u \in C^\infty$ off a finite set in Ω.*

Note by Lemma 4.2.5 that u is continuously differentiable up to third order away from the zeros of u, thus the equation (4.2.5) holds pointwise in the classical sense, i.e.

$$-\partial_{xx}u + \frac{1}{\mu}p(u) = Tu + F \quad \forall x \text{ s.t. } u(x) \neq 0. \tag{4.2.6}$$

Due to the potential discontinuities introduced by the term $p(u)$, the regularity analysis entails the study of the zeros of the solution. We now introduce the following definition to classify the zeros of u

Definition 4.2.7. Let $u(x) = 0$. Then, x is called a critical point for u, if u assumes both positive and negative values at every ball centered at x.

We proceed as follows. First, we prove that a solution to equation (4.2.6) admits finitely many critical zeros (Lemma 4.2.8). Then, we prove that non-critical zeros do not constitute any irregularities to the solution (Lemma 4.2.9).

Lemma 4.2.8. *Let a continuous function u satisfy (4.2.1). Then, u has finitely many critical points.*

Proof. Assume to the contrary that u has infinitely many critical points denoted by $x_1, x_2, \ldots$. Without loss of generality, we may assume that x_n's accumulate at some $x \in \Omega$. At every neighborhood of each x_n, u assumes both negative and positive values. Therefore, x is an accumulation point of both negative local minima, and positive local maxima. Let y_n, and z_n be those local minima and maxima accumulating at x, respectively. Since $u(y_n) < 0 < u(z_n)$, (4.2.6) implies

$$
\begin{aligned}
-\partial_{xx}u(y_n) - \frac{1}{\mu} &= Tu(y_n) + F(y_n) \\
-\partial_{xx}u(z_n) + \frac{1}{\mu} &= Tu(z_n) + F(z_n).
\end{aligned}
\tag{4.2.7}
$$

40

Furthermore, as y_n, and z_n are local minima and maxima for u, the second derivative of u satisfy $\partial_{xx}u(z_n) \leq 0 \leq \partial_{xx}u(y_n)$, which together with (4.2.7) implies

$$
\begin{aligned}
Tu(y_n) + F(y_n) &\leq -\frac{1}{\mu} \\
Tu(z_n) + F(z_n) &\geq \frac{1}{\mu}.
\end{aligned}
\tag{4.2.8}
$$

However, y_n, and z_n both accumulate at x. Therefore, the inequalities in (4.2.8) are incompatible with the continuity of u and f as $y_n \to x$ and $z_n \to x$, hence yielding a contradiction. $\qquad\square$

Now, we present the following lemma, which upgrades the differentiability of nonnegative functions up to the set where the function is equal to zero.

Lemma 4.2.9. *Let $I \in \mathbb{R}$ be an open interval. For $u : I \to \mathbb{R}^+ \cup \{0\}$, let K denote the collection of points where u is zero, i.e.*

$$
K = \{x | u(x) = 0\}.
$$

Suppose that K has empty interior, and $u \in C^1(I)$, and $u \in C^2(I \setminus K)$. Assume further that the second derivative of u on $I \setminus K$ can be extended continuously to I. Then, $u \in C^2(I)$.

The proof of Lemma 4.2.9 is discussed in the Appendix D. Now, we are ready to present the proof of Theorem 4.2.6.

Proof of Theorem 4.2.6. First, we verify that the assumptions of Lemma 4.2.9 are satisfied on an interval I where u is nonnegative. Lemma 4.2.5 implies $u \in C^1(I)$. Furthermore, since the RHS of (4.2.6) is continuous, u has a second derivative away from its zeros that can be extended continuously throughout I. Finally, if u is identically equal to zero on any subinterval $J \subset I$, then C^∞ property of u is immediate on J. Therefore, we may assume that the set $\{u = 0\}$ has empty interior. Hence, Lemma 4.2.9 is applicable for u. Now, the equation (4.2.6) is satisfied on I with $p(u) = 1$. We can repeat the same argument for the nonpositive values of u. Thus, we can differentiate (4.2.6) as many times as we wish to obtain the infinite differentiability of u away from the critical zeros and their shifts, which is a finite set by Lemma 4.2.8. $\qquad\square$

4.3 Support of Compressed Modes

In this section, we consider the asymptotic behavior of the support of the Compressed Modes, as the regularization term μ approaches to zero and infinity. Parabolic and Elliptic differential equations with compactly supported solutions have been considered in [7, 9, 8, 10]. In particular, in [8], the authors consider nonlinear elliptic equations involving maximum monotone graphs (such as the subgradient of L^1-norm as in our case). In [10], estimates for the volume of the support in terms of the time variable is provided for the parabolic equations. Our analysis differs from these studies in that the estimates are carried out in terms of the coefficients in the differential equation. The support of Compressed Modes are considered in [5], where the authors provide an asymptotic upper bound for the volume of the support as $\mu \to 0$. The results in this section are built upon the ideas in [5], and refines those results by providing an exact order of magnitude for the support as $\mu \to 0$. We also analyze the components of the energy associated to the Compressed Modes, and show the equipartition of the energy in the sense that the L^1-term and the gradient term of the energy grow in the same order. We then consider the other extreme case for μ, namely $\mu \to \infty$, and provide upper and lower bounds for the size of the support and the contribution of the L^1-term to the energy. The results in this section are given only in terms of CM-I, however they are also applicable to CM-II as the associated Euler-Lagrange equations are quite similar.

Notation. For two functions f, and g, we denote

$$f \lesssim_{c_1,c_2,\ldots,c_n} g \text{ as } \mu \to \mu_0,$$

if there exists a constant $C = C(c_1, \ldots, c_n)$ such that

$$f(\mu) \leq C(c_1, \ldots, c_n) g(\mu)$$

for the values of μ that are sufficiently close to μ_0. In case there is no ambiguity for the implicit constants $c_1, c_2, \ldots, c_n$, we omit them and simply write

$$f \lesssim g \text{ as } \mu \to \mu_0.$$

The expression $f \gtrsim g$ is defined analogously. If $g \lesssim f$, and $f \lesssim g$, then we denote $f \sim g$.

In [5], the following theorem is proven by studying the Euler-Lagrange equation (4.1.1) satisfied by Compressed Modes.

Theorem 4.3.1 (Theorem 4.1 in [5]). *For the Compressed Modes $\{\psi_i\}_{i=1}^N$ obtained from the regularization parameter μ, we have*

$$|\operatorname{supp}\psi_i| \lesssim_{N,d,\|V\|_\infty} \mu^{2d/(4+d)} \text{ for } j = 1,\ldots N \text{ as } \mu \to 0. \tag{4.3.1}$$

The proof mainly relies on the estimation of the Lagrange multipliers λ_i, λ_{ij} by multiplying the Euler-Lagrange equation (4.1.1) by ψ_i, and then integrating, to obtain the following expressions

$$\lambda_i = \frac{1}{2\mu}\int_{\mathbb{R}^d}|\psi_i|\,dx + \int_{\mathbb{R}^d}V(x)\psi_i^2\,dx + \frac{1}{2}\int_{\mathbb{R}^d}|\nabla\psi_i|^2\,dx, \tag{4.3.2a}$$

$$\lambda_{ij} = \frac{1}{\mu}\int_{\mathbb{R}^d}p(\psi_i)\psi_j\,dx + 2\int_{\mathbb{R}^d}V(x)\psi_i\psi_j\,dx + \int_{\mathbb{R}^d}\nabla\psi_i\cdot\nabla\psi_j\,dx. \tag{4.3.2b}$$

The total energy, E, of the first N Compressed Modes are defined by

$$E = \sum_{i=1}^N\int_{\mathbb{R}^d}(\frac{1}{\mu}|\psi_i| + \frac{1}{2}|\nabla\psi_i|^2 + V(x)\psi_i^2)\,dx.$$

For the proof of Theorem 4.3.1, the authors obtain the following results as $\mu \to 0$

$$E \sim \mu^{-4/(4+d)}, \tag{4.3.3}$$

$$\frac{1}{\mu}|\operatorname{supp}(\psi_i)| \leq (2\lambda_i + 2\|V\|_\infty)\int_{\mathbb{R}^d}|\psi_i|\,dx + \sum_{j\neq i}\lambda_{ij}\int_{\mathbb{R}^d}|\psi_j|\,dx, \tag{4.3.4}$$

The following theorem is the main result of this section, which refines the result of Theorem 4.3.1. Strictly speaking, it establishes the precise asymptotic behavior of the support as the regularization parameter $\mu \to 0$.

Theorem 4.3.2. *Let $\{\psi_i\}_{i=1}^N$ be the first N Compressed Modes. Then, as $\mu \to 0$,*

$$|\operatorname{supp}\psi_i| \sim \mu^{2d/(d+4)}. \tag{4.3.5}$$

As a straightforward consequence of Theorem 4.3.2, we obtain the following energy equipartition result.

Theorem 4.3.3. *Let $\{\psi_i\}_{i=1}^N$ be the first N Compressed Modes. Suppose the energy E is partitioned into $E = E_1 + E_2$, where*

$$E_1 = \sum_{i=1}^N \int_{\mathbb{R}^d} \frac{1}{\mu} |\psi_i| dx,$$

$$E_2 = \sum_{i=1}^N \int_{\mathbb{R}^d} \frac{1}{2} (|\nabla \psi_i|^2 + V(x)\psi_i^2)\, dx.$$

Then,

$$E_1 \sim \mu^{-4/(4+d)}, \quad E_2 \sim \mu^{-4/(4+d)} \ as \ \mu \to 0.$$

For the proof of Theorem 4.3.2, we use the classical Rayleigh-Faber-Krahn and the Poincare's inequalities. The Rayleigh-Faber-Krahn inequality provides a lower bound for the first (Dirichlet) eigenvalue of the Laplace operator in terms of the volume of the domain.

Theorem 4.3.4 (Rayleigh-Faber-Krahn inequality, see e.g. Section 3.2 in [26]). . *Let $\lambda_1(\Omega)$ denote the first (Dirichlet) eigenvalue of Laplace operator on a bounded domain $\Omega \subset \mathbb{R}^d$. Then,*

$$\lambda_1(\Omega) \gtrsim |\Omega|^{-2/d}, \tag{4.3.6}$$

where the implicit constant depends only on the dimension d.

On the other hand, the Poincare's inequality provides a bound for the L^2 norm of any function in $H_0^1(\Omega)$, in terms of the L^2 norm of its gradient.

Theorem 4.3.5 (Poincare's inequality, see e.g. Section 10.2 in [28]). *For $u \in H_0^1(\Omega)$,*

$$\|u\|_2^2 \leq \frac{\|\nabla u\|_2^2}{\lambda_1(\Omega)}. \tag{4.3.7}$$

Notice that in general, the Compressed Modes are elements of $H^1(\Omega)$. However, the Poincare's inequality holds true for functions in $H_0^1(\Omega)$, the space obtained by taking the H^1-norm closure of the $C_0^\infty(\Omega)$ functions. Nevertheless, since the Compressed Modes are

compactly supported as $\mu \to 0$, the Poincare's inequality is applicable for any domain that contains the support of the Compressed Modes.

The following lemma is obtained by a mere combination of the Rayleigh-Faber-Krahn and the Poincare's inequalities.

Lemma 4.3.6. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, and suppose that $u \in H^1(\Omega)$ has compact support inside Ω. Then,*

$$\|u\|_2 \lesssim |\operatorname{supp} u|^{1/d} \|\nabla u\|_2. \tag{4.3.8}$$

Proof. Construct the domains $\{V_k\}_{k=1}^{\infty}$ such that $\operatorname{supp} u \subset V_k$ and

$$\lim_{k \to \infty} |V_k| = |\operatorname{supp} u|.$$

Then, $u \in H_0^1(V_k)$, so that the Poincare's inequality (4.3.7) yields

$$\|u\|_2 \leq \frac{\|\nabla u\|_2}{\lambda_1(V_k)^{1/2}}. \tag{4.3.9}$$

On the other hand, the Rayleigh-Faber-Krahn inequality (4.3.6) provides a lower bound for $\lambda_1(V_k)$ as

$$\lambda_1(V_k) \gtrsim |V_k|^{-2/d}. \tag{4.3.10}$$

Substituting (4.3.10) in (4.3.9), we obtain

$$\|u\|_2 \lesssim |V_k|^{1/d} \|\nabla u\|_2.$$

Taking limit as $k \to \infty$, we obtain (4.3.8), as desired. $\qquad\qquad\square$

Now, we are ready to prove Theorem 4.3.2.

Proof of Theorem 4.3.2. Lets first estimate λ_i, and λ_{ij}. Note from the expression (4.3.2a) for λ_i that

$$\lambda_i \leq \frac{1}{2\mu} \int_{\mathbb{R}^d} |\psi_i| \, dx + \|V\|_\infty + \frac{1}{2} \int_{\mathbb{R}^d} |\nabla \psi_i|^2 \, dx \lesssim E.$$

Similarly from (4.3.2b),

$$
\begin{aligned}
|\lambda_{ij}| &\le \left| \frac{1}{\mu} \int_{\mathbb{R}^d} p(\psi_i)\psi_j \, dx + 2 \int_{\mathbb{R}^d} V(x)\psi_i\psi_j \, dx + \int_{\mathbb{R}^d} \nabla\psi_i \cdot \nabla\psi_j \, dx \right| \\
&\le \frac{1}{\mu} \int_{\mathbb{R}^d} |\psi_j| \, dx + 2 \int_{\mathbb{R}^d} |V(x)\psi_i\psi_j| \, dx + \int_{\mathbb{R}^d} \nabla|\psi_i \cdot \nabla\psi_j| \, dx \\
&\le \frac{1}{\mu} \int_{\mathbb{R}^d} |\psi_j| \, dx + \int_{\mathbb{R}^d} |V(x)|(\psi_i^2 + \psi_j^2) \, dx + \frac{1}{2} \int_{\mathbb{R}^d} |\nabla\psi_i|^2 + |\nabla\psi_j|^2 \, dx \\
&\lesssim E.
\end{aligned}
$$

Plugging the above upper bounds for λ_i, and λ_{ij} in inequality (4.3.4), we obtain the following upper bound for the support

$$
\frac{1}{\mu}|\operatorname{supp} \psi_i| \lesssim E \sum_{i=1}^{N} \int_{\mathbb{R}^d} |\psi_i| \, dx = \mu E E_1. \tag{4.3.11}
$$

Recall that the total energy of the system, E, grows at order $\mu^{4/(4+d)}$, so that

$$
E_1 \lesssim E \sim \mu^{4/(4+d)}.
$$

Combining this with the above inequality (4.3.11), we obtain

$$
|\operatorname{supp} \psi_i| \lesssim \mu^2 E E_1 \lesssim \mu^{2d/(d+4)}. \tag{4.3.12}
$$

Now, applying Lemma 4.3.6 to ψ_i, we get

$$
\begin{aligned}
\|\psi_i\|_2 &\lesssim |\operatorname{supp} \psi_i|^{1/d} \|\nabla\psi_i\|_2 \\
&\lesssim |\operatorname{supp} \psi_i|^{1/d} \sqrt{E},
\end{aligned} \tag{4.3.13}
$$

where the last inequality holds true since the term $\|\nabla\psi_i\|_2^2$ is contained in the total energy E. Since $E \sim \mu^{-4/(4+d)}$, the last inequality becomes

$$
\|\psi_i\|_2 \lesssim |\operatorname{supp} \psi_i|^{1/d} \mu^{-2/(4+d)},
$$

or equivalently,

$$
|\operatorname{supp} \psi_i| \gtrsim \mu^{2d/(d+4)},
$$

proving the upper bound for the support. The lower bound is already verified in (4.3.12), hence proving (4.3.5). $\qquad\square$

Proof of Theorem 4.3.3. The proof is mainly a repetition of the inequalities in the proof of Theorem 4.3.2 with the knowledge of the precise asymptotic behavior of the size of the support of $\psi_i|$ is given by

$$|\operatorname{supp} \psi_i| \sim \mu^{2d/(d+4)}.$$

First, we consider the inequality (4.3.11), which now becomes

$$\mu^{(d-4)/(d+4)} \sim \frac{1}{\mu}|\operatorname{supp} \psi_i| \lesssim \mu E E_1 \sim \mu^{d/(d+4)} E_1,$$

hence providing the following lower bound for E_1

$$E_1 \gtrsim \mu^{-4/(4+d)}.$$

We already have the upper bound $E_1 \leq E \sim \mu^{-4/(4+d)}$. Hence,

$$E_1 \sim \mu^{-4/(4+d)}.$$

Next, the inequality (4.3.13) becomes

$$1 \lesssim |\operatorname{supp} \psi_i|^{1/d}||\nabla\psi_i||_2 \sim \mu^{2/(4+d)}||\nabla\psi_i||_2.$$

Hence, we get

$$||\nabla\psi_i||_2 \gtrsim \mu^{-2/(d+4)}.$$

Now, we have

$$E_2 = \sum_{i=1}^{N}\int_{\mathbb{R}^d}\frac{1}{2}(|\nabla\psi_i|^2 + V(x)\psi_i^2)\,dx = \frac{1}{2}\sum_{i=1}^{N}||\nabla\psi_i||_2^2 - N||V||_\infty \gtrsim \mu^{-4/(d+4)}.$$

which proves $E_2 \sim \mu^{-4/(d+4)}$, since the lower bound is immediate. $\qquad\square$

So far, the asymptotic analysis of support and the energy partition is considered for the case $\mu \to 0$. Next, we analyze the case $\mu \to \infty$. In this case, since the L^1 term converges to zero, the energy of the Compressed Modes corresponding to a particular $\mu = \mu_0$ yields an upper bound for the energies corresponding to μ with $\mu \geq \mu_0$. Therefore, E stays bounded as $\mu \to \infty$, which can be written in the following notation

$$E \lesssim 1$$

47

The following theorem gives bounds for the support of the Compressed Modes, and their energy partition.

Theorem 4.3.7. *Let $\{\psi_i\}_{i=1}^N$ be the first N Compressed Modes. Then, as $\mu \to \infty$,*

$$1 \lesssim |\operatorname{supp} \psi_i| \lesssim \mu^2,$$
$$\mu^{-2} \lesssim E_1 \lesssim 1.$$

Proof. Similarly as in the proof of the previous theorem, we may estimate λ_i, and λ_{ij}'s by E, hence they are bounded, too. That is

$$\lambda_i \lesssim 1,$$
$$\lambda_{ij} \lesssim 1.$$

Now, the inequality (4.3.4) becomes

$$\frac{1}{\mu} |\operatorname{supp}(\phi_i)| \lesssim \mu E_1. \tag{4.3.14}$$

On the other hand, the inequality (4.3.8) yields

$$1 = ||u||_2 \lesssim |\operatorname{supp}(\psi_i)|^{1/d} ||\nabla \psi_i||_2.$$

Since $||\nabla \psi_i||_2 \lesssim \sqrt{E} \lesssim 1$, the above inequality implies $|\operatorname{supp}(\psi_i)| \gtrsim 1$. This, combined with (4.3.14) yields the result. $\qquad \square$

4.4 Conclusions

In this chapter, the size of the support for CM-I, and the regularity properties for CM-I, CM-II, and CPW have been studied by analyzing the associated Euler-Lagrange equations. The precise asymptotic behavior for the size of the support has been shown to agree with the upper bound given in [5]. A bootstrap argument based on the Elliptic Regularity Theorems has enabled us to conclude that for dimension $d = 1$, CM-I, CM-II, and CPW are infinitely differentiable on their domain of definition outside of finite set of points.

CHAPTER 5

Solutions to Variational Problems with Generalized L^1 Terms

This chapter presents the applications of the sparsity promoting techniques to different type of differential equations. There has been significant progress towards understanding the nature of broader family of differential equations involving the L^1 term, and developing numerical schemes to compute the solutions. Formulation of some non-linear elliptic and parabolic problems such as the Signum-Gordon and the divisible sandpile equations are considered in [11], where the authors provide an Alternating Direction Implicit (Douglas-Rachford) scheme to efficiently compute the solutions. A further reference on solving nonlinear multivalued evolution equations can be found in [30]. In [44], a general class of obstacle problems are formulated in terms of a variational form involving the L^1 norm, and a Split-Bregman scheme is developed for numerical solutions. It is later shown in [49] that these type of optimization problems can be efficiently solved via the primal-dual methods for convex problems (see [15]). Applications of L^1 type penalization to Hamilton-Jacobi equations are introduced in [17], where the authors provide a framework to efficiently solve high dimensional problems via the variational Hopf formula [27]. This chapter is taken with slight modification from Section 4 of [40], which studies the compact support properties of elliptic and parabolic problems with weighted L^1 terms.

In this chapter, we consider parabolic problems, in particular, the heat equation on an infinite domain that follows the gradient flow associated to an energy involving a weighted L^1 term and study the support of the solution. Our work is an extension of [10], where the authors study the parabolic variational inequalities, and discuss the connections to the free-boundary problems. The "free-boundary" essentially denotes the boundary of the "active"

region where the system is still governed by the underlying differential equation, hence evolves over time. Free-boundary problems have many application areas such as optimal control, fluid dynamics, and financial mathematics [35, 38]. A summary of the variational formulations of the free boundary problems can be found in [23].

Compressed Modes (CM) and Compressed Plane Waves (CPW) are designed to obtain spatially localized functions for applications in solid state physics. Spatial localization enables capturing the short-ranged interactions, and disregarding long-ranged interactions that are often beyond physical intuition. One of the very attractive features of CMs and CPWs is that the distinction between the short-ranged and long-ranged interactions are resolved by the variational formulation, hence no explicit thresholding is needed for such distinction. On the other hand, the scheme involves a parameter that controls the spatial localization at the expense of the accuracy of the solution. While the errors are inevitable as the true solutions are not fully spatially localized, errors in certain regions can be minimized by modifying the L^1 penalization term in a priori region. Namely, the extension of the original scheme for a spatially weighted L^1 penalization would enable such modification.

5.1   Problem Statement

The goal in this section is to analyze the following equation on $\mathbb{R}^n \times [0, T]$

$$u_t - \Delta u + \rho(x)p(u) = f$$
$$u(x, 0) = u_0(x)$$

(5.1.1)

and, in particular, classify the weight functions $\rho : \mathbb{R}^n \to \mathbb{R}^+ \cup \{0\}$, which would result in compactly supported functions in time and space domains. In [10], the parabolic variational inequality is formulated as

$$(u_t - \Delta u)(v - u) \geq f(v - u) \text{ a.e. for } x \in \mathbb{R}^n, 0 < t < T,$$
$$u \geq 0 \text{ for } x \in \mathbb{R}^n, 0 < t < T,$$
$$u(x, 0) = u_0(x).$$

(5.1.2)

for any non-negative measurable function v. Note that the problem of interest in this section, (5.1.1) can be seen as a special case of the problem (5.1.2) treated in [10]. Strictly speaking, if we denote the positive and negative parts of the solution of (5.1.1), u, by u_+, and u_-, respectively, then they satisfy the inequalities

$$(\partial_t u_+ - \Delta u_+)(v - u_+) \geq (f - \rho)(v - u_+)$$

$$(\partial_t u_- - \Delta u_-)(v - u_-) \geq (f + \rho)(v - u_-)$$

We generalize the compact support results in [10] to allow for a broader family of forcing terms f, so that it is applicable to our motivating problem (5.1.1).

The existence and uniqueness for the problem (5.1.2) is given in the above mentioned article. Furthermore, they prove theorems regarding the compact support of the solution (Theorem 3.1. and Theorem 3.2. in [10]) under the uniform negativity constraint on f, namely that there exist a positive real number ν, such that

$$f \leq -\nu. \tag{5.1.3}$$

For sufficient regularity assumptions, they also require

$$\begin{aligned} f &\in L^\infty(\mathbb{R}^n \times (0, T)), \\ f_t &\in L^\infty(\mathbb{R}^n \times (0, T)). \end{aligned} \tag{5.1.4}$$

We now quote the Theorems 3.1-3.2 from [10].

Theorem 5.1.1 (Theorem 3.1 in [10]). *Suppose that u is a solution to the parabolic variational inequality (5.1.2) with the assumptions (5.1.3) and (5.1.4) on f. Then, there is a positive number T_0 such that $u(x, t) \equiv 0$ for $t \geq T_0$.*

Theorem 5.1.2 (Theorem 3.2 in [10]). *Suppose that u is a solution to the parabolic variational inequality (5.1.2) with the assumptions (5.1.3) and (5.1.4) on f. Suppose further that u_0 has compact support. Then, there is a positive number R_0 such that $u(x, t) = 0$ when $|x| > R_0$.*

We show that we can relax the condition (5.1.3) so that it only holds away from a ball centered at the origin. Namely, we only require

$$f(x, t) \leq -\nu \text{ for } |x| > K, \tag{5.1.5}$$

along with non-strict negativity condition on f

$$f(x, t) \leq 0. \tag{5.1.6}$$

Theorem 5.1.3. *Suppose that u is a solution to the parabolic variational inequality (5.1.2) with the assumptions (5.1.4), (5.1.5), and (5.1.6) on f. Then, there is a positive number T_0 such that $u(x, t) \equiv 0$ for $t \geq T_0$.*

Theorem 5.1.4. *Suppose that u is a solution to the parabolic variational inequality (5.1.2) with the assumptions (5.1.4), (5.1.5), and (5.1.6) on f. Suppose further that u_0 has compact support. Then, there is a positive number R_0 such that $u(x, t) = 0$ if $|x| > R_0$.*

Corollary 5.1.5. *Let u satisfy the following PDE with a compactly supported initial condition u_0,*

$$u_t - \Delta u + \rho(x) p(u) = f,$$

where p is the sub-differential of the absolute value function, and $\rho : \mathbb{R}^n \to \mathbb{R}^+ \cup \{0\}$ is a weight function. Suppose further that

$$\lim_{(x,t) \to \infty} f(x, t) = 0$$

$$\liminf_{x \to \infty} \rho(x) > 0$$

Then, u is compactly supported on the (x, t)-space.

The proofs of Theorems 5.1.3 and 5.1.4 rely on the maximum principle applied to the family of functions β_ϵ, $u_{R,\epsilon}$ defined in [10]. We merely restate the definitions of these functions here. β_ϵ is a $C^\infty(\mathbb{R})$ function satisfying

$$\beta_\epsilon(x) = 0 \text{ for } x > 0,$$

$$\lim_{\epsilon \to 0} \beta_\epsilon(x) = -\infty \text{ for } x < 0,$$

$$\beta_\epsilon'(x) > 0 \text{ for } x < 0.$$

52

Then, for a given initial data u_0 and a source term f, the functions $u_{R,\epsilon}$ are defined to be the solution to the following problem

$$u_t - \Delta u + \beta_\epsilon(u) = f, \text{ for } |x| < R, 0 < t < T,$$

$$u(x,0) = u_0(x), \text{ for } |x| < R,$$

$$u(x,t) = 0, \text{ for } |x| = R, t > 0.$$

Proof of Theorem 5.1.3. We follow a similar construction as in [10]. From Theorem 2.1 in [10], there exists $M > 0$ such that $u_{R,\epsilon}(x,1) \leq M$. Let

$$v(x,t) = \begin{cases} M - \nu(t-1) \text{ for } |x| > K, \\ M - \nu(t-1) + \nu(K^2 - |x|^2)/2d \text{ for } |x| \leq K \end{cases}$$

and let $w = \max(0,v)$. Then, $w(x,t) = 0$ for $t > T_0 := 1 + M/\nu + K^2/2d$. Furthermore,

$$w_t - \Delta w = \begin{cases} 0 \text{ if } |x| < K \\ -\nu \text{ if } |x| > K, 1 \leq t \leq T_0 \\ 0 \text{ if } |x| > K, t > T_0 \end{cases}$$

In particular, w satisfies

$$w_t - \Delta w + \beta_\epsilon(w) \geq f.$$

Therefore, by the maximum principle applied on $w - u_{R,\epsilon}$, we conclude that $u_{R,\epsilon}(x,T_0) \leq 0$. Letting, $R \to \infty$, and $\epsilon \to 0$, we obtain $u(x,T_0) = 0$. Hence, $u \equiv 0$ for $t \geq T_0$. $\square$

Proof of Theorem 5.1.4. Let ρ denote the radius of the support of u_0, as in the original proof. The only difference is that, we proceed with $\tilde{\rho}$ such that $\tilde{\rho} > \max(\rho, K)$, and construct the comparison function in maximum principle argument in a slightly different way.

From Theorem 2.1 in [10], we know the existence of $N > 0$ such that

$$|u_{R,\epsilon}(t,x)| \leq N \text{ for } x \in \mathbb{R}^n, \tilde{\rho} \leq |x| \leq R, 0 < t < T_0.$$

For arbitrary positive constants μ, R_0 and for $r = |x|$, let w solve the heat equation for $|x| < \tilde{\rho}$ with source term f

$$w_t - \Delta w = f \text{ for } |x| < \tilde{\rho}$$

$$w(x, t) = \mu(R_0 - \tilde{\rho})^2 \text{ for } |x| = \tilde{\rho}, t > 0$$

$$w(x, 0) = \mu(R_0 - r)^2 \text{ for } |x| < \tilde{\rho}.$$

We choose parameters μ, R_0 such that $2\mu \leq \nu$, $\mu(R_0 - \tilde{\rho})^2 \geq N$, so that the following inequalities are satisfied

$$w_t - \Delta w + \beta_\epsilon(w) \geq -\nu \text{ if } |x| > \tilde{\rho}$$

$$w \geq N \text{ if } |x| = \tilde{\rho}.$$

Now, applying the maximum principle to $w - u_{R,\epsilon}$, we conclude that $w - u_{R,\epsilon} \geq 0$ if $\tilde{\rho} < |x| < R, 0 < t < T_0$. Therefore,

$$u_{R,\epsilon}(x, t) = 0 \text{ if } R_0 \leq |x| \leq R, 0 < t < T_0.$$

Letting $R \to \infty$, we obtain the spatial compactness of u, as desired. $\square$

Proof of Corollary 5.1.5. Observe that u_+ is a solution to the variational inequality (5.1.2) when f is replaced by $f - \rho(x)\mu$, so that the RHS of the variational inequality is strictly negative for large values of x and t. Now, by Theorem 5.1.4, u_+ is compactly supported in the space variable x. Let

$$\epsilon = \liminf_{x \to \infty} \rho(x).$$

Suppose $|f| < \frac{\epsilon}{2}$ for $t > T$. Then, Theorem 5.1.3 is applicable for u_+ provided that we replace the initial time with $t = T$ instead of $t = 0$, so that u_+ has compact support in t.

Repeating the above arguments for u_-, we conclude that u is compactly supported in time variable t, as desired. $\square$

5.2 Numerical Results

5.2.1 Heat Equation with L^1 minimization

For numerical results, we consider the initial value problem on $\mathbb{R}^n \times [0, T]$ without the forcing term, i.e.

$$u_t - \Delta u + \rho(x)p(u) = 0$$

$$u(x, 0) = u_0(x)$$

$$(5.2.1)$$

In order to numerically compute the solutions to the equation (5.2.1), we first discretize the equation in time via the implicit Euler scheme, then discretize the arising equations in space and carry out the computations via FFT solvers [16, 41]. In particular, the implicit Euler scheme for the problem (5.2.1) is given by

$$\frac{u^{n+1} - u^n}{\tau} = \Delta u^{n+1} - \rho(x)p(u^{n+1}). \tag{5.2.2}$$

Here, we denote $u^n = u^n(x, y) = u(n\tau, x, y)$ for the values of u at discrete time instances. Note that the spatial variable is not discretized in the equation (5.2.2). Furthermore, given u^n, equation (5.2.2) is a non-linear equation for u^{n+1}. Nevertheless, we can convert this equation into a variational form involving the (weighted) L^1 norm. Namely, the solution u^{n+1} to the equation (5.2.2) is also a solution to the following variational problem

$$\min_{u} \frac{1}{2}\|\nabla u\|_2^2 + \frac{1}{2\tau}\|u - u^n\|_2^2 + \|\rho(x)u\|_1. \tag{5.2.3}$$

Now, the above formulation is a convex minimization problem, which can be solved via convex optimization methods such as the Split-Bregman (ADMM) method [25, 30]. In order to implement convex optimization methods, we need to discretize the problem (5.2.3). First, we truncate the infinite domain of space variable into a sufficiently large finite rectangular domain, and enforce zero boundary conditions for the feasible set of solutions. Next, the Split-Bregman scheme is framed as follows. The associated Lagrangian is given by

$$\mathcal{L}(u, v, c) = \frac{1}{2}\|\nabla u\|_2^2 + \frac{1}{2\tau}\|u - u^n\|_2^2 + \|\rho(x)v\|_1 + \frac{\lambda}{2}\|u - v\|_2^2 + \lambda\langle c, u - v\rangle.$$

Here $\lambda > 0$ is the step-size parameter that controls the speed of convergence. The method consists of solving the following subproblems starting from an initial guess for the auxiliary variables v and c, until a desired level of convergence is obtained.

$$u_{k+1} = \underset{u}{\arg\min}\, \mathcal{L}(u, v_k, c_k) = \underset{u}{\arg\min}\, \frac{1}{2}\|\nabla u\|_2^2 + \frac{1}{2\tau}\|u - u^n\|_2^2 + \frac{\lambda}{2}\|u - v_k\|_2^2 + \lambda\langle c, u\rangle$$

$$\text{(P.1.1)}$$

$$v_{k+1} = \underset{v}{\arg\min}\, \mathcal{L}(u_{k+1}, v, c_k) = \underset{u}{\arg\min}\, \frac{\lambda}{2}\|v - u_{k+1}\|_2^2 + \|\rho(x)v\|_1 - \lambda\langle c, v\rangle \qquad \text{(P.1.2)}$$

$$c_{k+1} = c_k + u_{k+1} - v_{k+1} \qquad \text{(P.1.3)}$$

Without discretization, by Euler-Lagrange equations, the solution to the problem (P.1.1) also satisfies the following Poisson's equation

$$-\Delta u_{k+1} + \left(\frac{1}{\tau} + \lambda\right) u_{k+1} = \frac{1}{\tau}u^n + \lambda(v_k - c_k).$$

Hence, we consider the discretized version of the above problem, which can be solved via the discrete sine transform (DST). DST also ensures that the Dirichlet boundary conditions are met as well. The solution to subproblem (P.1.2) is given simply by a (weighted) soft-thresholding as

$$v_{k+1} = S\left(u_{k+1} + c_k, \frac{\rho}{\lambda}\right),$$

where S is the soft-thresholding operator applied coordinate-wise on the arguments. The soft-thresholding operator on scalars is given by

$$S(x, \alpha) = \text{sign}(x)\max(|x| - \alpha, 0).$$

The algorithm is summarized in Algorithm 1.

We contrast the results for the cases where there is no sub-differential term (Figure 5.1), with uniform sub-differential term (Figure 5.2), and with weighted sub-differential term whose weight is given by the characteristic function of the complement of a finite rectangular region (Figure 5.3).

We consider the problem with the two-dimensional space variable where the initial value function is taken to be an instance of the two-dimensional heat kernel. Namely, for Figure 5.1,

Algorithm 1 Split-Bregman Scheme (ADMM) for solving the heat equation with weighted subdifferential term.

Input: $U^0, \rho, \lambda, timesteps$, and convergence criteria

Output: U^n for $n = 1, 2, \ldots, timesteps$

   for $n = 0, 1, \ldots, timesteps - 1$ **do**

      while "not converged" **do**

        $u_{k+1}^{n+1} = \left(\left(\frac{1}{\tau} + \lambda \right) \mathbf{I} - \boldsymbol{\Delta} \right)^{-1} \left(\frac{1}{\tau} U^n + \lambda \left(v_k^{n+1} - c_k^{n+1} \right) \right)$

        $v_{k+1}^{n+1} = S \left(u_{k+1}^{n+1} + c_k^{n+1}, \frac{\rho}{\lambda} \right)$

        $c_{k+1}^{n+1} = c_k^{n+1} + u_{k+1}^{n+1} - v_{k+1}^{n+1}$

      end while

      $U^{n+1} = u_{k+1}^{n+1}$

   end for

---

we consider the standard heat equation

$$u_t - \Delta u = 0$$
$$u(x, y, 0) = u_0(x, y) = \frac{1}{4\pi} e^{-\frac{x^2 + y^2}{4}}, \tag{5.2.4}$$

whose solution is given analytically by the heat kernel as

$$u(x, y, t) = \frac{1}{4\pi(t+1)} e^{-\frac{x^2 + y^2}{4(t+1)}}.$$

Next, we modify the heat equation via the subdifferential term as

$$u_t - \Delta u + \gamma p(u) = 0$$
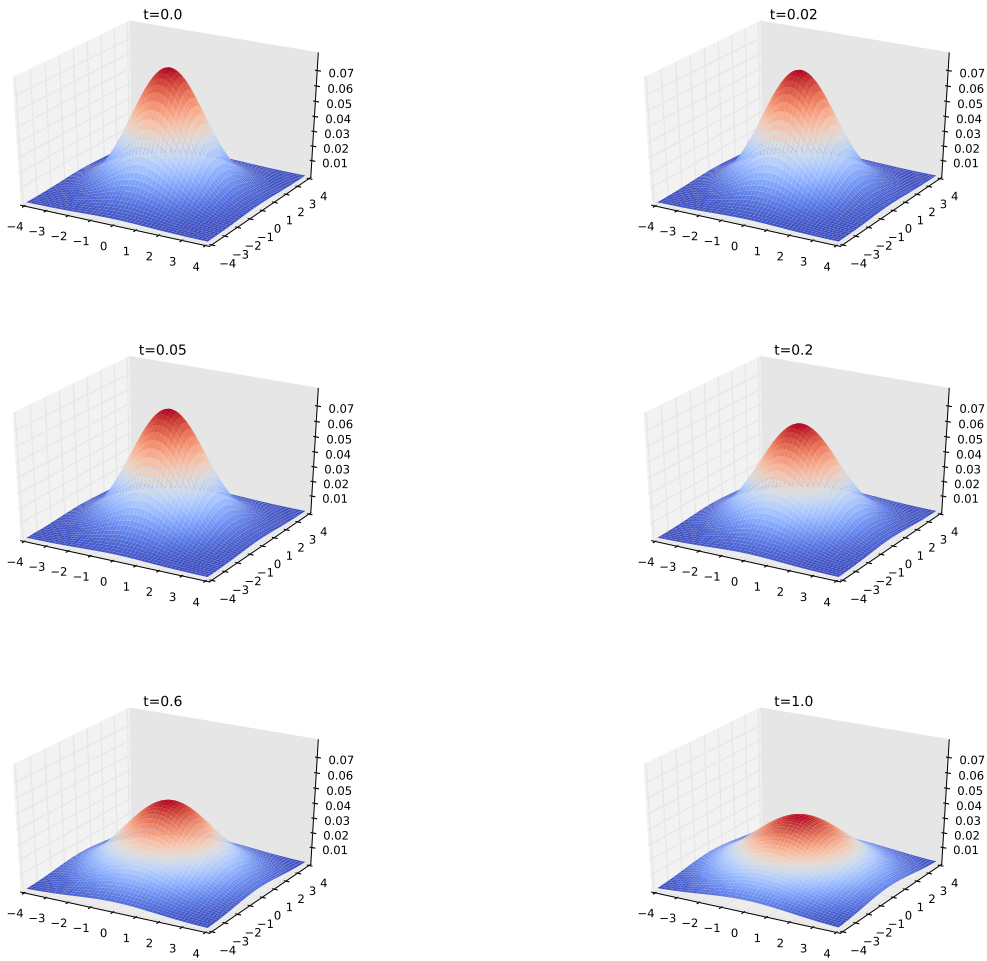$$u(x, y, 0) = u_0(x, y) = \frac{1}{4\pi} e^{-\frac{x^2 + y^2}{4}}, \tag{5.2.5}$$

with parameter

$$\gamma = 0.1. \tag{5.2.6}$$

Finally, the setup for the equation with weighted subdifferential term is given by

$$u_t - \Delta u + \gamma \chi_{R^c}(x) p(u) = 0$$
$$u(x, y, 0) = u_0(x, y) = \frac{1}{4\pi} e^{-\frac{x^2 + y^2}{4}}, \tag{5.2.7}$$

Figure 5.1: Analytical solution to the equation (5.2.4) at various t values.

with the following parameters.

$$\gamma = 0.1$$
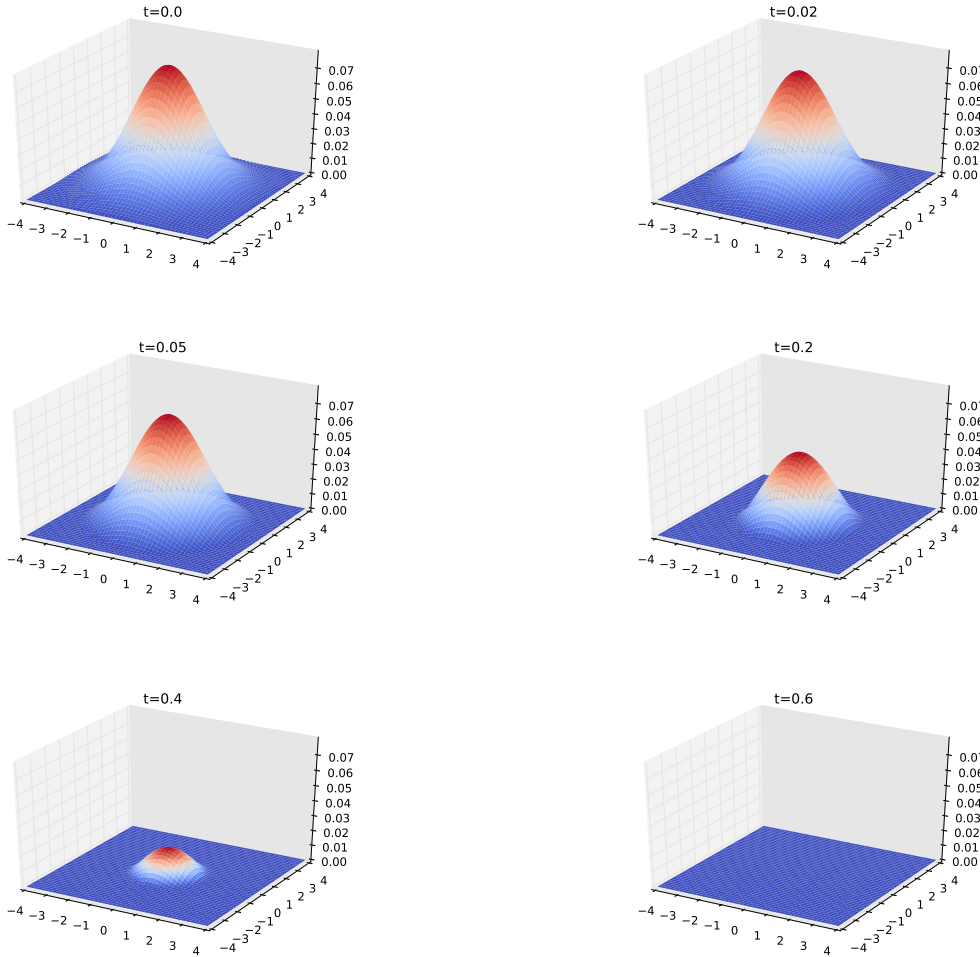$$R = [-2, 2] \times [-2, 2] \subset \mathbb{R}^2.$$

(5.2.8)

Notice that the weight term

$$\rho(x) = \gamma \chi_{R^c}(x)$$

satisfies

$$\liminf_{x \to \infty} \rho(x) = \lim_{x \to \infty} \rho(x) = \gamma,$$

hence Corollary 5.1.5 is applicable.

Figure 5.2: Solution to the equation (5.2.5) with the setup given by (5.2.6) at various t values.

Without the subdifferential term p, the solution spreads to infinity as $t \to \infty$ as shown in Figure 5.1. Whereas with the (uniform) subdifferential term, the solutions have compact support with sizes shrinking in time (see Figure 5.2). Hence, the solutions no longer spreads to infinity. On the other hand, when the subdifferential term is activated only outside of a finite rectangle, the solutions exhibit a similar behavior to the uniform subdifferential case for small values of t. However, for the large values of t, the support of the solutions tend to stay within the rectangle as there is no subdifferential term inside the rectangle as illustrated in Figure 5.3.

Figure 5.3: Solution to the equation (5.2.7) with the setup given by (5.2.8) at various t values.
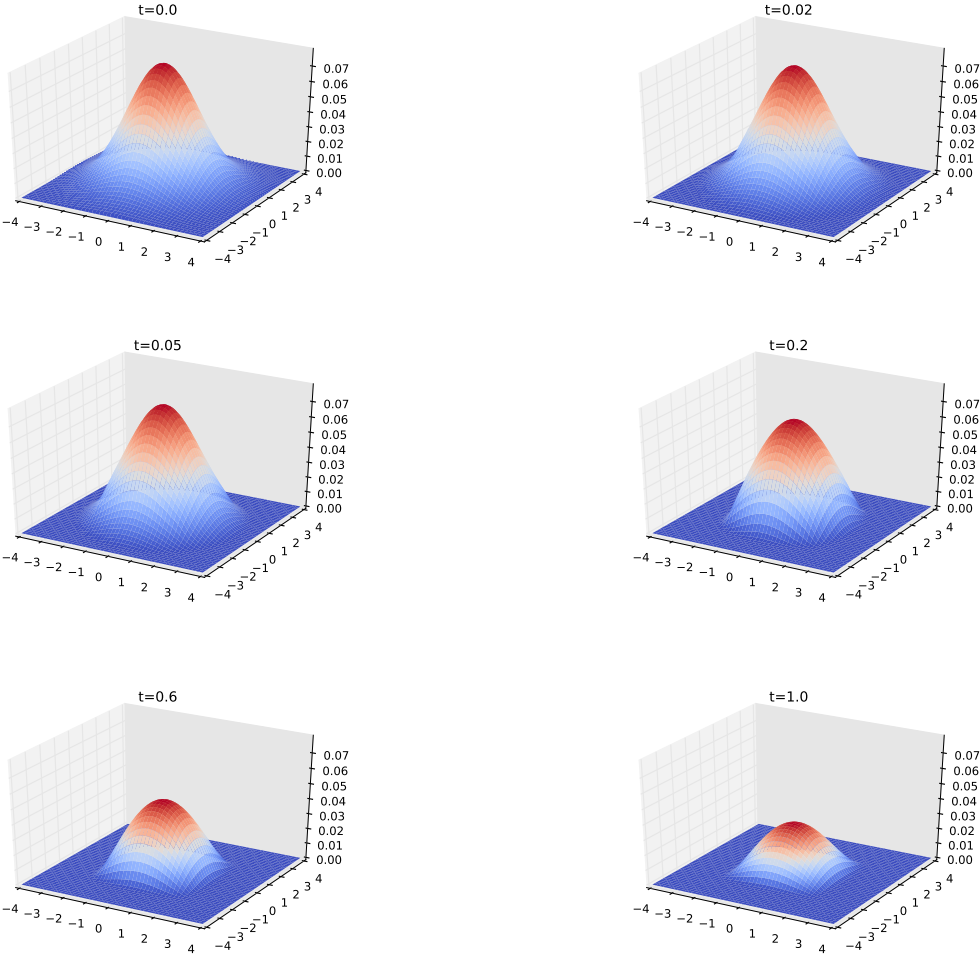
5.2.2   Heat Equation with Obstacle

Next, we consider the heat equation inside the cylinder $U \times [0, T]$ with an obstacle that is activated outside of a particular region $Q \subset \mathbb{R}^n$. More precisely, we consider the following

problem

$$u(x, t) \geq \phi(x) \text{ if } x \notin Q$$

$$u_t - \Delta u = f \text{ if } u(x, t) > \phi(x) \text{ or } x \in Q$$

$$u(x, t) = 0 \text{ for } x \in \partial U \tag{5.2.9}$$

$$u(x, 0) = u_0(x)$$

We employ a similar discretization scheme as in Section 5.2.1, so that

$$u^{n+1}(x, t) \geq \phi(x) \text{ if } x \notin Q$$

$$\frac{u^{n+1} - u^n}{\tau} = \Delta u^{n+1} + f \text{ if } x \in Q$$

$$\frac{u^{n+1} - u^n}{\tau} = \Delta u^{n+1} + f \text{ if } u^{n+1}(x, t) > \phi(x) \text{ and } x \notin Q$$

The above equation for u^{n+1} is an elliptic obstacle problem with zero (Dirichlet) boundary condition. A reformulation of elliptic obstacle problems as optimization problems are given in [44]. We follow their formulation with the modification that the penalization term is activated only outside of the region Q. Hence, the problem becomes

$$\min_u \frac{1}{2}\|\nabla u\|_2^2 + \frac{1}{2\tau}\|u - u^n\|_2^2 - \frac{1}{\tau}\langle u, f \rangle + \gamma \int_{Q^c} (\phi - u)^+ dx, \tag{5.2.10}$$

for some $\gamma > 0$ large enough. Following the standard iteration scheme described in Section 5 of [44] with the auxiliary variable $v = \phi - u$, the solution is obtained by solving the following problems until a desired level of convergence is obtained

$$u_{k+1} = \operatorname*{argmin}_u \frac{1}{2}\|\nabla u\|_2^2 + \frac{1}{2\tau}\|u - u^n\|_2^2 - \frac{1}{\tau}\langle u, f \rangle + \frac{\lambda}{2}\|u - (\phi - v_k)\|_2^2 + \lambda\langle c, u \rangle \tag{P.2.1}$$

$$v_{k+1} = \operatorname*{argmin}_u \frac{\lambda}{2}\|v - \phi + u_{k+1}\|_2^2 + \gamma \int_{Q^c} v^+ dx - \lambda\langle c, v \rangle \tag{P.2.2}$$

$$c_{k+1} = c_k + u_{k+1} + v_{k+1} - \phi \tag{P.2.3}$$

The solution to (P.2.1) can be computed by solving the elliptic equation

$$-\Delta u_{k+1} + \left(\frac{1}{\tau} + \lambda\right) u_{k+1} = \frac{1}{\tau}u^n + \frac{1}{\tau}f + \lambda(\phi - v_k - c_k),$$

whereas the solution to (P.2.2) is given by

$$v_{k+1}(x) = S_{Q^c}^+ \left(\phi(x) - u_{k+1}(x) + c_k(x), x, \frac{\gamma}{\lambda}\right)$$

61

where $S_{Q^c}^+$ is the soft-thresholding operator applied outside of Q and only to the positive values. Hence $S_{Q^c}^+$ is given by

$$S_{Q^c}^+(y, x, \alpha) = \begin{cases} y - \alpha \text{ if } x \notin Q, y \geq \alpha, \\ 0 \text{ if } x \notin Q, 0 < y < \alpha, \\ y \text{ otherwise.} \end{cases}$$

The algorithm is summarized in Algoritm 2.

Algorithm 2 Split-Bregman Scheme (ADMM) for solving the heat equation with obstacle that is activated outside of a given region

---

Input: $U^0, f, \phi, Q, \gamma, \lambda, timesteps$, and convergence criteria

Output: U^n for $n = 1, 2, \ldots, timesteps$

   for $n = 0, 1, \ldots, timesteps - 1$ **do**

      while "not converged" **do**

         $u_{k+1}^{n+1} = \left(\left(\frac{1}{\tau} + \lambda \right) \mathbf{I} - \mathbf{\Delta} \right)^{-1} \left(\frac{1}{\tau} U^n + \frac{1}{\tau} f + \lambda \left(\phi - v_k^{n+1} - c_k^{n+1} \right) \right)$

         $v_{k+1}^{n+1} = S_{Q^c}^+ \left(\phi - u_{k+1}^{n+1} + c_k^{n+1}, \frac{\gamma}{\lambda} \right)$

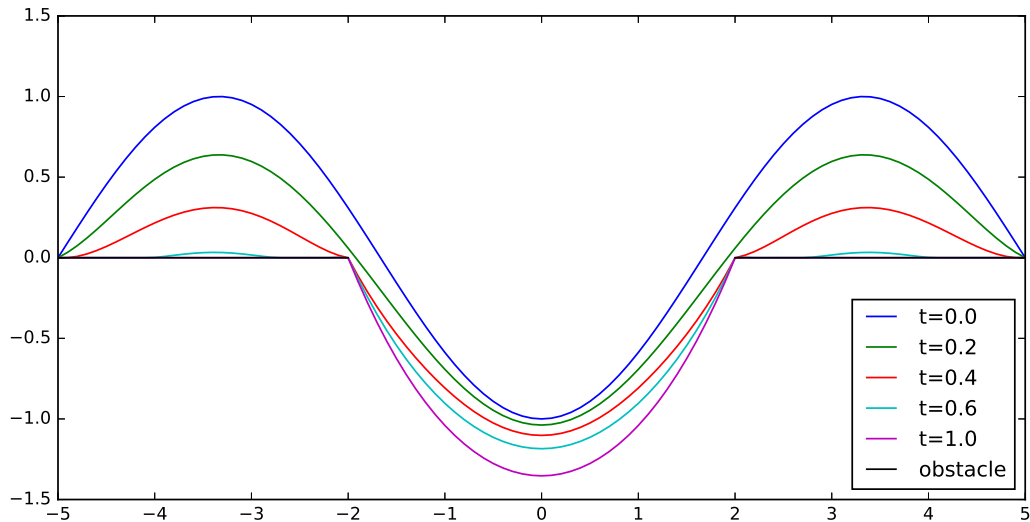 $c_{k+1}^{n+1} = c_k^{n+1} + u_{k+1}^{n+1} + v_{k+1}^{n+1} - \phi$

      end while

      $U^{n+1} = u_{k+1}^{n+1}$

   end for

Figure 5.4 illustrates the numerical solution of (5.2.9) with one-dimensional spatial variable along with the following setup

$$U = [-5, 5],$$
$$f(x, t) = -1,$$
$$\phi(x) = 0, \tag{5.2.11}$$
$$Q = [-2, 2],$$
$$u_0(x) = -\cos \frac{3\pi}{10} x$$

Figure 5.4: Solution to the equation (5.2.9) with the setup given by (5.2.11) at various t values

5.3 Conclusions

Using a weighted L^1 penalization in the context of parabolic problems, we have demonstrated that the deviation from the true solution can be reduced in pre-determined regions, yet still having compactly supported solutions. The numerical scheme is chosen appropriately so that the algorithm is applicable to a wide range of possibly discontinuous weights such as the characteristic functions of a sets. Numerical results suggest that the notion of weighted L^1 penalization can be applied to the parabolic obstacle problems, where the obstacle is not active at certain regions.

APPENDIX A

Shift Orthogonality

A.1 Characterization of Shift Orthogonality

We provide an explicit characterization of the Hilbert spaces $\mathcal{H}_j$ in Theorem 2.4.3. The eigenfunctions of the Laplace operator in a rectangular domain $\Omega = [0, n_1 w_1] \times \cdots \times [0, n_d w_d]$ is given by

$$\phi_{m_1,\ldots,m_d}(\mathbf{x}) = \mathbf{e}^{2\pi i \left(\frac{m_1 x_1}{n_1 w_1} + \frac{m_2 x_2}{n_2 w_2} + \cdots + \frac{m_d x_d}{n_d w_d} \right)}.$$

where $(m_1, m_2, \ldots, m_d) \in \mathbb{Z}^d$. Hence, if we form the lattice

$$\Pi_{\mathbf{w}} = \left\{ \left(\frac{m_1}{n_1 w_1}, \frac{m_2}{n_2 w_2}, \ldots, \frac{m_d}{n_d w_d} \right) \middle| (m_1, m_2, \ldots, m_d) \in \mathbb{Z}^d \right\},$$

then each of the eigenfunctions of the Laplace operator in the domain Ω can be represented as

$$\phi_v(\mathbf{x}) = \mathbf{e}^{2\pi i v \cdot \mathbf{x}}, \ v \in \Pi_{\mathbf{w}},$$

with the corresponding eigenvalue $\lambda_v = 4\pi^2 |v|^2$. Now, we define

$$\Lambda_{\mathbf{w}} = \left\{ \left(\frac{m_1}{n_1 w_1}, \frac{m_2}{n_2 w_2}, \ldots, \frac{m_d}{n_d w_d} \right) \middle| 0 \le m_1 < n_1, \ldots, 0 \le m_d < n_d \right\}.$$

Each $\rho \in \Lambda_{\mathbf{w}}$ has a natural periodic extension in $\Pi_{\mathbf{w}}$ with respect to $\Gamma_{\mathbf{w}}$. For each $\rho \in \Lambda_{\mathbf{w}}$, we denote such extension by Σ_ρ. Moreover, the family of Hilbert spaces $\mathcal{H}_j$ in Theorem 2.4.3 consists of the Hilbert spaces

$$\mathcal{H}_\rho = \text{span}\{\phi_v | v \in \Sigma_\rho\}.$$

Since each of $\Lambda_{\mathbf{w}}$ and $\Gamma_{\mathbf{w}}$ has cardinality $n_1 n_2 \ldots n_d$; the cardinality of the family of Hilbert spaces $\{\mathcal{H}_j\}_{j=1}^N$, satisfies $N = n_1 n_2 \ldots n_d = |\Gamma_{\mathbf{w}}|$, as asserted in Theorem 2.4.3.

We have already observed that the eigenvalue corresponding to ϕ_v is $\lambda_v = 4\pi^2 |v|^2$. Weyl's law in the rectangular domain case can be viewed as the growth of the size of the distance between lattice points and the origin. Therefore, with all these lattice characterization of the eigenfunctions, it is not hard to see that the growth of the eigenvalues corresponding to the eigenfunctions in each of the Hilbert spaces $\mathcal{H}_j$ are given precisely as in Theorem 2.4.5.

As an illustration, let's consider $\Omega = [0,2] \times [0,3] \subset \mathbb{R}^2$, and $\mathbf{w} = (\mathbf{1}, \mathbf{1})$. Then, $\Gamma_{\mathbf{w}}$ becomes

$$\Gamma_{\mathbf{w}} = \{(0,0), (0,1), (0,2), (1,0), (1,1), (1,2)\}.$$

The eigenfunctions for the Laplace equation in Ω are given by

$$\phi_{m,n}(x,y) = e^{2\pi i \left(\frac{mx}{2} + \frac{ny}{3}\right)},$$

so that

$$\Pi_{\mathbf{w}} = \left\{ \left(\frac{m}{2}, \frac{n}{3}\right) \Big| m, n \in \mathbb{Z} \right\}.$$

Now, the finite lattice $\Lambda_{\mathbf{w}}$ becomes

$$\Lambda_{\mathbf{w}} = \left\{ (0,0), \left(0, \frac{1}{3}\right), \left(0, \frac{2}{3}\right), \left(\frac{1}{2}, 0\right), \left(\frac{1}{2}, \frac{1}{3}\right), \left(\frac{1}{2}, \frac{2}{3}\right) \right\}$$

Finally, the decomposition given in Theorem 2.4.3 becomes

$$L^2(\Omega) = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_3 \oplus \mathcal{H}_4 \oplus \mathcal{H}_{\cdot 5} \oplus \mathcal{H}_6,$$

where

$$\mathcal{H}_1 = \mathrm{span}\{\phi_{2k,3l}\}_{k,l \in \mathbb{Z}}, \qquad \mathcal{H}_2 = \mathrm{span}\{\phi_{2k,3l+1}\}_{k,l \in \mathbb{Z}}, \qquad \mathcal{H}_3 = \mathrm{span}\{\phi_{2k,3l+2}\}_{k,l \in \mathbb{Z}},$$

$$\mathcal{H}_4 = \mathrm{span}\{\phi_{2k+1,3l}\}_{k,l \in \mathbb{Z}}, \qquad \mathcal{H}_5 = \mathrm{span}\{\phi_{2k+1,3l+1}\}_{k,l \in \mathbb{Z}}, \qquad \mathcal{H}_6 = \mathrm{span}\{\phi_{2k+1,3l+2}\}_{k,l \in \mathbb{Z}}.$$

A.2 An Elementary Proof of Shift Orthogonality Characterization

An elementary proof of Theorem 2.4.3 can be presented as follows. For simplicity, we consider the one dimensional case, where N is a positive integer, and work with the space $L^2([0,N])$,

where the shifts are simply the integer shifts. We denote the shift-orthogonal decomposition of a function $f \in L^2([0, N])$ by

$$f(x) = \sum_{k=1}^{N} f_k(x)e^{2\pi i \frac{k}{N} x}, \tag{A.2.1}$$

where each f_k is a periodic function with period 1. Existence and uniqueness of such a decomposition follows from expanding f in the Fourier basis and grouping the terms with the same frequencies modulo integers. Next, we consider the inner product of a function f with its shifts. If f has a decomposition as given in (A.2.1), then

$$f(x+m) = \sum_{k=1}^{N} f_k(x)e^{2\pi i \frac{k}{N}(x+m)} = \sum_{k=1}^{N} f_k(x)e^{\frac{2\pi i k m}{N}} e^{2\pi i \frac{k}{N} x}. \tag{A.2.2}$$

Setting $\xi = e^{\frac{2\pi i}{N}}$ i.e. as the N-th root of unity, we can rewrite (A.2.2) as

$$f(x+m) = \sum_{k=1}^{N} \xi^{km} f_k(x)e^{2\pi i \frac{k}{N} x}.$$

First, we note that the Fourier coefficients of a 1-periodic function is supported only on integers divisible by N. That is, for a 1-periodic function $g : [0, N] \to \mathbb{C}$, whenever n is not divisible by N

$$\int_0^N g(x)e^{2\pi i \frac{n}{N} x} dx = 0. \tag{A.2.3}$$

This last identity can be verified by the following direct computation using the 1-periodicity of g

$$\int_0^N g(x)e^{2\pi i \frac{n}{N} x} dx = \sum_{k=0}^{N-1} \int_k^{k+1} g(x)e^{2\pi i \frac{n}{N} x} dx = \int_0^1 g(x)e^{2\pi i \frac{n}{N} x} dx \sum_{k=0}^{N-1} \xi^{nk}.$$

As ξ is an N-th root of unity, and n is not divisible by N, we have

$$\sum_{k=0}^{N-1} \xi^{nk}$$

yielding (A.2.3).

Now, lets compute the inner product

$$\langle f(x), f(x+m) \rangle = \int_0^N f(x)\overline{f(x+m)}dx,$$

66

using the shift-orthogonal decomposition (A.2.1). By (A.2.2),

$$\langle f(x), f(x+m)\rangle = \int_0^N \left(\sum_{k=1}^N f_k(x)e^{2\pi i \frac{k}{N}x}\right)\overline{\left(\sum_{k=1}^N \xi^{km} f_k(x)e^{2\pi i \frac{k}{N}x}\right)} dx$$

$$= \int_0^N \left(\sum_{k=1}^N f_k(x)e^{2\pi i \frac{k}{N}x}\right)\left(\sum_{k=1}^N \xi^{-km}\overline{f_k(x)}e^{-2\pi i \frac{k}{N}x}\right) dx$$

$$= \int_0^N \left(\sum_{k=1}^N \sum_{\ell=1}^N \xi^{-\ell m} f_k(x)\overline{f_\ell(x)}e^{2\pi i \frac{k-\ell}{N}x}\right) dx.$$

Since $f_k(x)\overline{f_l(x)}$ is 1-periodic, we can use (A.2.3) to conclude that the summands in the last line above is non-zero only when $k = \ell$, yielding the following simplified expression

$$\langle f(x), f(x+m)\rangle = \int_0^N \sum_{k=1}^N \xi^{-km}|f_k(x)|^2 dx. \tag{A.2.4}$$

Denoting

$$w_k = \int_0^N |f_k(x)|^2 dx,$$

the identity (A.2.4) can be written in the following matrix form

$$\begin{bmatrix} \langle f(x), f(x+1)\rangle \\ \langle f(x), f(x+2)\rangle \\ \vdots \\ \langle f(x), f(x+N-1)\rangle \\ \langle f(x), f(x+N)\rangle \end{bmatrix} = \begin{bmatrix} \xi^{-1} & \xi^{-2} & \cdots & \xi^{-N} \\ \xi^{-2} & \xi^{-4} & \cdots & \xi^{-2N} \\ \vdots & \vdots & \ddots & \vdots \\ \xi^{-(N-1)} & \xi^{-2(N-1)} & \cdots & \xi^{-N(N-1)} \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n-1} \\ w_n \end{bmatrix}$$

or more compactly as follows

$$\mathbf{b} = \mathbf{A}\mathbf{w}. \tag{A.2.5}$$

This matrix equation can be solved directly since $\frac{1}{\sqrt{N}}A$ is an orthogonal matrix, i.e. $A^{-1} = \frac{1}{N}A^*$. The shift-orthogonality constraints amounts to

$$\mathbf{b} = [\mathbf{0}, \mathbf{0}, \dots, \mathbf{0}, \mathbf{1}]^{\mathbf{t}},$$

so that the solution to the equation (A.2.5) is given by

$$\mathbf{w} = \left[\frac{\mathbf{1}}{\mathbf{L}}, \frac{\mathbf{1}}{\mathbf{L}}, \dots, \frac{\mathbf{1}}{\mathbf{L}}, \frac{\mathbf{1}}{\mathbf{L}}\right]^{\mathbf{t}}.$$

Therefore, a shift orthogonal function f whose decomposition is given by (A.2.1) satisfies

$$\|f_k\|_2^2 = \int_0^N |f_k(x)|^2 dx = \frac{1}{L}.$$

APPENDIX B

Distribution of Eigenvalues for the Laplace Operator

We briefly discuss the properties of the distribution of the eigenvalues of the Laplace operator on rectangular domains to provide a further refinement of the approximation result (2.3.3) in Corollary 2.3.5. The idea is to obtain a uniform approximation results for dimensions $d \geq 3$, where there is no completeness result yet. The inequality (2.3.3) involves the difference between eigenvalues, $\lambda_m - \lambda_n$ in its RHS. First, we provide an estimate for the differences between eigenvalues. Recall that the eigenfunctions of the Laplace operator in a rectangular domain $\Omega = [0, w_1] \times, \ldots, \times [0, w_d]$ is given by

$$\phi_{m_1, \ldots, m_d}(\mathbf{x}) = \mathbf{e}^{2\pi \mathbf{i} \left(\frac{\mathbf{m_1 x_1}}{\mathbf{w_1}} + \frac{\mathbf{m_2 x_2}}{\mathbf{w_2}} + \cdots + \frac{\mathbf{m_d x_d}}{\mathbf{w_d}} \right)}.$$

where $(m_1, m_2, ..., m_d) \in \mathbb{Z}^d \setminus \{0\}$. Or alternatively, if we form the lattice

$$\Pi_{\mathbf{w}} = \left\{ \frac{m_1}{w_1}, \frac{m_2}{w_2}, \ldots, \frac{m_d}{w_d} \middle| (m_1, m_2, \ldots, m_d) \in \mathbb{Z}^d \setminus \{0\} \right\}$$

then each of the eigenfunctions of the Laplace operator in the domain Ω can be represented as

$$\phi_v(x) = e^{2\pi i v \cdot \mathbf{x}}, \quad v \in \Pi_{\mathbf{w}},$$

with the corresponding eigenvalue $\lambda_v = 4\pi^2 |\nu|^2$. Now, let's define the eigenvalue counting function $\mathcal{N} : \mathbb{R}^+ \to \mathbb{N}$ via

$$\mathcal{N}(x) = \# \{\lambda | \lambda \leq x, \lambda \text{ is an eigenvalue}\}.$$

From the observation above, we obtain

$$\mathcal{N}(x) = \# \left\{ v \in \Pi_{\mathbf{w}} \middle| |v| \leq \frac{\sqrt{x}}{2\pi} \right\}.$$

In other words, $\mathcal{N}$ counts the lattice points which lie on a ball with $\frac{\sqrt{x}}{2\pi}$. Notice that the number of lattice points inside a domain can be used as an estimate of the volume of the domain. In case of a ball, this relation can be converted into the following upper bound for $\mathcal{N}$

$$\mathcal{N}(x) \leq \frac{\left| B\left(\frac{\sqrt{x}}{2\pi}\right) \right|}{w_1 w_2 \dots w_d} \tag{B.0.1}$$

where $B\left(\frac{\sqrt{x}}{2\pi}\right)$ denotes the ball centered at the origin with radius $\frac{\sqrt{x}}{2\pi}$. On the other hand, if we consider another ball centered at the origin, whose radius is shorter than the previous one by the diagonal length ℓ of each cells of the lattice, we obtain a lower bound, namely

$$\mathcal{N}(x) \geq \frac{\left| B\left(\frac{\sqrt{x}}{2\pi} - \ell\right) \right|}{w_1 w_2 \cdots w_d} \tag{B.0.2}$$

Now, the relations (B.0.1)-(B.0.2) above become

$$\omega_d \frac{\left(\frac{\sqrt{x}}{2\pi} - l\right)^d}{w_1 w_2 \dots w_d} \leq \mathcal{N}(x) \leq \omega_d \frac{\left(\frac{\sqrt{x}}{2\pi}\right)^d}{w_1 w_2 \dots w_d}$$

where ω_d is the volume of the unit ball in $\mathbb{R}^d$. Inverting this last relation, we obtain

$$\mathcal{N}\left(4\pi^2 \left(\frac{w_1 w_2 \cdots w_d}{\omega_d}\right)^{2/d} n^{2/d}\right) \leq n \leq \mathcal{N}\left(4\pi^2 \left(\left(\frac{w_1 w_2 \cdots w_d}{\omega_d}\right)^{1/d} n^{1/d} + \ell\right)^2\right)$$

so that

$$4\pi^2 \left(\frac{w_1 w_2 \cdots w_d}{\omega_d}\right)^{2/d} n^{2/d} \leq \lambda_n \leq 4\pi^2 \left(\left(\frac{w_1 w_2 \cdots w_d}{\omega_d}\right)^{1/d} n^{1/d} + \ell\right)^2. \tag{B.0.3}$$

Let's define the constant C and the aspect ratio Υ_Ω of the domain Ω via

$$C = 4\pi^2 \left(\frac{w_1 w_2 \cdots w_d}{\omega_d}\right)^{2/d}, \quad \Upsilon_\Omega = \frac{\ell}{\left(\frac{w_1 w_2 \cdots w_d}{\omega_d}\right)^{\frac{1}{d}}}.$$

Then, the relation (B.0.3) becomes

$$Cn^{2/d} \leq \lambda_n \leq C\left(n^{\frac{1}{d}} + \Upsilon_\Omega\right)^2 \tag{B.0.4}$$

For $m \geq n$, by (B.0.4), we obtain the following separation result

$$\lambda_m - \lambda_n \geq Cm^{2/d} - C\left(n^{\frac{1}{d}} + \Upsilon_\Omega\right)^2.$$

We now have a lower bound for the quantity $\lambda_m - \lambda_n$. Recall that the inequality (2.3.3) is given by

$$\sum_{k=1}^{n} d(\phi_k, V_m)^2 \leq \frac{m|\Omega|^{\frac{1}{2}}}{\mu(\lambda_{m+1} - \lambda_n)}$$

Collecting the constants into the $C_{\Omega,\mu}$ term, the above inequality becomes

$$\sum_{k=1}^{n} d(\phi_k, V_m)^2 \leq C_{\Omega,\mu} \frac{m}{m^{\frac{2}{d}} - \left(n^{\frac{1}{d}} + \Upsilon_\Omega\right)^2}.$$

APPENDIX C

An Asymptotic Formula for Derivative of Convex Functions

We present the proof of the following lemma introduced in Chapter 3.

Lemma C.0.1. *Let $F : \mathbb{R}^+ \to \mathbb{R}$ be a convex function satisfying*

$$F(t) = t^\alpha + O(t^\beta), \ \ as \ t \to \infty,$$

where the real numbers α, and β satisfies $\alpha > 1$, and $\alpha > \beta > 0$. Suppose further that F is differentiable on $\mathbb{R}^+ \setminus E$, where E is a discrete set. Then,

$$F'(t) = \alpha t^{\alpha - 1} + O(t^{\frac{\alpha + \beta}{2} - 1}), \ \ as \ t \in \mathbb{R}^+ \setminus E \ and \ t \to \infty.$$

Proof of Lemma 3.1.2. Proof of Lemma 2.1. Let $C > 0$ be such that

$$t^\alpha - Ct^\beta \le F(t) \le t^\alpha + Ct^\beta, \quad \forall t > 0. \tag{C.0.1}$$

For any real number a with $a < 1$, we have, by convexity

$$\frac{F(t) - F(t - t^a)}{t^a} \le F'(t) \le \frac{F(t + t^a) - F(t)}{t^a} \tag{C.0.2}$$

which combined with (C.0.1) yields

$$\frac{t^\alpha - (t - t^a)^\alpha}{t^a} - C\frac{t^\beta + (t - t^a)^\beta}{t^a} \le F'(t) \le \frac{(t + t^a)^\alpha - t^\alpha}{t^a} + C\frac{(t + t^a)^\beta + t^\beta}{t^a} \tag{C.0.3}$$

Let's define

$$f_1(t) = \frac{t^\alpha - (t - t^a)^\alpha}{t^a},$$

$$g_1(t) = \frac{t^\beta + (t - t^a)^\beta}{t^a},$$

$$f_2(t) = \frac{(t + t^a)^\alpha - t^\alpha}{t^a},$$

$$g_2(t) = \frac{(t + t^a)^\beta + t^\beta}{t^a}.$$

Then, the inequality (C.0.3) becomes

$$f_1(t) - Cg_1(t) \le F'(t) \le f_2(t) + Cg_2(t). \tag{C.0.4}$$

We now analyze the growth of f_1, f_2, g_1 and g_2, as $t \to \infty$. Note that

$$f_1(t) = t^{\alpha-a} \left(1 - (1 - t^{a-1})^\alpha\right),$$

$$g_1(t) = t^{\beta-a} \left(1 + (1 - t^{a-1})^\beta\right),$$

$$f_2(t) = t^{\alpha-a} \left((1 + t^{a-1})^\alpha - 1\right),$$

$$g_2(t) = t^{\beta-a} \left(1 + (1 + t^{a-1})^\beta\right).$$

Taylor expanding the above expressions with the aid of

$$(1 + x)^\gamma = 1 + \gamma x + \frac{\gamma(\gamma - 1)}{2}x^2 + o(x^2) \text{ as } x \to 0, \tag{C.0.5}$$

we obtain

$$f_1(t) = t^{\alpha-a} \left(\alpha t^{a-1} - \frac{\alpha(\alpha - 1)}{2}t^{2a-2} + o(t^{2a-2})\right),$$

$$= \alpha t^{\alpha-1} - \frac{\alpha(\alpha - 1)}{2}t^{\alpha+a-2} + o(t^{\alpha+a-2})$$

$$g_1(t) = t^{\beta-a} \left(2 - \beta t^{a-1} + \frac{\beta(\beta - 1)}{2}t^{2a-2} + o(t^{2a-2})\right),$$

$$= 2t^{\beta-a} - \beta t^{\beta-1} - \frac{\beta(\beta - 1)}{2}t^{\beta+a-2} + o(t^{\beta+a-2})$$

and similarly

$$f_2(t) = \alpha t^{\alpha-1} + \frac{\alpha(\alpha - 1)}{2}t^{\alpha+a-2} + o(t^{\alpha+a-2})$$

$$g_2(t) = 2t^{\beta-a} + \beta t^{\beta-1} - \frac{\beta(\beta - 1)}{2}t^{\beta+a-2} + o(t^{\beta+a-2}).$$

72

Setting $a = \frac{\alpha - \beta}{2} - 1$, we obtain

$$f_1(t) - Cg_1(t) = \alpha t^{\alpha - 1} + O(t^{\frac{\alpha + \beta}{2} - 1}),$$
$$f_2(t) + Cg_2(t) = \alpha t^{\alpha - 1} + O(t^{\frac{\alpha + \beta}{2} - 1}),$$

which, together with (C.0.4) implies

$$F'(t) = \alpha t^{\alpha - 1} + O(t^{\frac{\alpha + \beta}{2} - 1}), \text{ as } t \to \infty, \tag{C.0.6}$$

as desired. $\qquad\qquad\square$

APPENDIX D

Improving Differentiability of Nonnegative Functions up to Their Zeros

This section is devoted to the proof of Lemma 4.2.9, which is stated as follows.

Lemma D.0.1. *Let $I \in \mathbb{R}$ be an open interval. For $u : I \to \mathbb{R}^+ \cup \{0\}$, let K denote the collection of points where u is zero, i.e.*

$$K = \{x | u(x) = 0\}.$$

Suppose that K has empty interior, and $u \in C^1(I)$, and $u \in C^2(I \setminus K)$. Assume further that the second derivative of u on $I \setminus K$ can be extended continuously to I. Then, $u \in C^2(I)$.

The main difficulty in the proof of Lemma 4.2.9 is that the zero set of nonnegative C^∞ functions can form arbitrary closed sets [47]. In particular, the set K can be a nowhere dense positive measure sets such as Smith-Volterra-Cantor set. Therefore, a careful topological analysis of zeros is required. Before proceeding with the proof of Lemma 4.2.9, we make the following definition to classify the zeros of a continuous function.

Definition D.0.1. Let f be a continuous function and let x be a point with $f(x) = 0$. Then, x is called an

  (i) *isolated zero* if there exists a neighborhood N around x with $f(y) \neq 0$ for $y \in N \setminus \{x\}$,

  (ii) *essential zero* if x is an accumulation points of other zeros.

Note that the sets of isolated zeros and essential zeros are mutually exclusive sets. Moreover, any particular zero of a function falls into precisely one of these categories.

Now, we are ready to present the proof.

Proof of Lemma 4.2.9. We denote $I = (a, b)$, and without loss of generality, assume $u(a) = u(b) = 0$. This is because the C^2 property holds when the function is nonzero, hence we can get rid of left and right segments of I where u is nonzero.

Let N denote the set of points of positivity of u. Now, we exploit the topological properties of $\mathbb{R}$, namely that any open set inside $\mathbb{R}$, hence N, can be represented as a union of countable open intervals as

$$N = \{x | u(x) > 0\} = \bigcup_{k \in \mathbb{N}} I_k = \bigcup_{k \in \mathbb{N}} (a_k, b_k).$$

Moreover, since K has no interior, we have $K = \partial N$.

Next, we prove that u_x is equal to zero on K except possibly at a, and b. Indeed, if x is a point with $u(x) = 0$, and $u_x(x) > 0$, then u attains negative values to the right of x, hence contradicting the nonnegativity of u. Similarly, $u_x(x) < 0$ is not feasible, implying $u_x(x) = 0$, as desired.

Let's denote the isolated and essential zeros of u by K_1 and K_2, respectively. Let f be the continuous extension of the second derivative of u. Since K_2 contains points that are accumulation points of zeros of u, where u_x is also equal to zero, the points in K_2 are accumulation points of the zeros of u_x. Hence, f vanishes on K_2. On the other hand, the isolated zeros, K_1, can form at most a countable set, so that $f = 0$ a.e. on K.

Now, let v be the solution to the Poisson problem

$$v_{xx}(x) = f(x) \text{ on } I$$

$$v(a) = v(b) = 0.$$

Notice that
$$\int_{a_k}^{b_k} f(x)dx = \int_{a_k}^{b_k} u_{xx}(x)dx = u_x(b_k) - u_x(a_k) = 0 \ \ \forall k \in \mathbb{N}.$$

Let y be point with $u(y) \neq 0$, i.e. $y \in N$. Then, $y \in (a_n, b_n)$ for some $n \in \mathbb{N}$. Since f vanishes a.e. on K, we have

$$\int_a^y f(x)dx = \int_{(a,y) \cap \bigcup_{k \in \mathbb{N}} (a_k, b_k)} f(x)dx = \int_{a_n}^y f(x)dx.$$

75

Hence,

$$v_x(y) - v_x(a) = \int_a^y v_{xx}(x)dx = \int_a^y f(x)dx = \int_{a_n}^y f(x)dx = u_x(y) - u_x(a_n) = u_x(y),$$

so that $v_x - u_x = v_x(a)$ on N. Since K is the boundary of N, by continuity of $v_x - u_x$, we conclude that $v_x - u_x$ is constant throughout I. Since v_x is continuously differentiable, u_x is continuously differentiable, too. Hence, $u \in C^2(I)$, as desired. $\qquad\square$

References

[1] Robert A Adams and John JF Fournier. *Sobolev spaces*, volume 140. Academic press, 2003.

[2] Frederick J Almgren. *Existence and regularity almost everywhere of solutions to elliptic variational problems with constraints*, volume 165. American Mathematical Soc., 1976.

[3] John M Ball. Strict convexity, strong ellipticity, and regularity in the calculus of variations. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 87(3), pages 501–513. Cambridge Univ Press, 1980.

[4] Farzin Barekat. On the consistency of compressed modes for variational problems associated with the schrödinger operator. *SIAM Journal of Mathematical Analysis*, 46(5):3568–3577, 2014.

[5] Farzin Barekat, Russel Caflisch, and Stanley Osher. On the support of compressed modes. *UCLA CAM Reports 14-14*, 2014.

[6] Farzin Barekat, Ke Yin, Russel E Caflisch, Stanley J Osher, Rongjie Lai, and Vidvuds Ozolins. Compressed Wannier modes found from an l_1 regularized energy functional. *arXiv preprint arXiv:1403.6883*, 2014.

[7] Philippe Benilan, Haim Brezis, and Michael G Crandall. A semilinear equation in $L^1(\mathbb{R}^N)$. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 2(4):523–555, 1975.

[8] Haim Brézis. Solutions with compact support of variational inequalities. *Russian Mathematical Surveys*, 29(2):103–108, 1974.

[9] Haim Brézis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer, 2011.

[10] Haim Brézis, Avner Friedman, et al. Estimates on the support of solutions of parabolic variational inequalities. *Illinois Journal of Mathematics*, 20(1):82–97, 1976.

[11] Russel E Caflisch, Stanley J Osher, Hayden Schaeffer, and Giang Tran. Pdes with compressed solutions. *Communications in Mathematical Sciences*, 13(8), 2015.

[12] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[13] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[14] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[15] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[16] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[17] Jérôme Darbon and Stanley Osher. Algorithms for overcoming the curse of dimensionality for certain hamilton-jacobi equations arising in control theory and elsewhere. *arXiv preprint arXiv:1605.01799*, 2016.

[18] Ingrid Daubechies et al. *Ten lectures on wavelets*, volume 61. SIAM, 1992.

[19] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[20] Lawrance Evans. *Partial Differential Equations*. American Mathematical Soc., 2012.

[21] Lawrence C Evans. Quasiconvexity and partial regularity in the calculus of variations. *Archive for Rational Mechanics and Analysis*, 95(3):227–252, 1986.

[22] Lawrence C Evans. Partial differential equations and monge-kantorovich mass transfer. *Current developments in mathematics*, pages 65–126, 1997.

[23] Avner Friedman. *Variational principles and free-boundary problems*. Courier Corporation, 2010.

[24] Guy Gilboa. A total variation spectral framework for scale and texture analysis. *SIAM Journal on Imaging Sciences*, 7(4):1937–1961, 2014.

[25] Tom Goldstein and Stanley Osher. The split bregman method for l^1-regularized problems. *SIAM journal on imaging sciences*, 2(2):323–343, 2009.

[26] Antoine Henrot. *Extremum problems for eigenvalues of elliptic operators*. Springer Science & Business Media, 2006.

[27] Eberhard Hopf. Generalized solutions of non-linear equations of first order(first order nonlinear partial differential equation discussing global locally-lipschitzian solutions via jacoby theorem extension). *Journal of Mathematics and Mechanics*, 14:951–973, 1965.

[28] Jürgen Jost. *Partial Differential Equations*. Springer Science & Business Media, 2013.

[29] Tosio Kato. *Perturbation theory for linear operators*, volume 132. Springer Science & Business Media, 2013.

[30] Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[31] Shiqian Ma, Donald Goldfarb, and Lifeng Chen. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1-2):321–353, 2011.

[32] Paolo Marcellini. Regularity of minimizers of integrals of the calculus of variations with non standard growth conditions. *Archive for Rational Mechanics and Analysis*, 105(3):267–284, 1989.

[33] Vidvuds Ozoliņš, Rongjie Lai, Russel Caflisch, and Stanley Osher. Compressed modes for variational problems in mathematics and physics. *Proceedings of the National Academy of Sciences*, 110(46):18368–18373, 2013.

[34] Vidvuds Ozoliņš, Rongjie Lai, Russel Caflisch, and Stanley Osher. Compressed plane waves yield a compactly supported multiresolution basis for the Laplace operator. *Proceedings of the National Academy of Sciences*, 111(5):1691–1696, 2014.

[35] Arshak Petrosyan and Henrik Shahgholian. Parabolic obstacle problems applied to finance. *Recent developments in nonlinear partial differential equations*, 439:117–133, 2007.

[36] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[37] Franz Rellich. *Perturbation theory of eigenvalue problems*. CRC Press, 1969.

[38] J-F Rodrigues. *Obstacle problems in mathematical physics*, volume 134. Elsevier, 1987.

[39] Hayden Schaeffer, Russel Caflisch, Cory D Hauck, and Stanley Osher. Sparse dynamics for partial differential equations. *Proceedings of the National Academy of Sciences*, 110(17):6634–6639, 2013.

[40] Jonathan Siegel and Omer Tekin. Compact support of l^1 penalized variational problems. *UCLA CAM Reports 15-56*, 2015.

[41] Paul N Swarztrauber. The methods of cyclic reduction, fourier analysis and the facr algorithm for the discrete solution of poisson's equation on a rectangle. *Siam Review*, 19(3):490–501, 1977.

[42] Ömer Faruk Tekin, Ke Yin, and Farzin Barekat. Spectral results for perturbed variational eigenvalue problems and their applications to compressed pdes. *Communications in Mathematical Sciences*, 14(3), 2016.

[43] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[44] Giang Tran, Hayden Schaeffer, William M Feldman, and Stanley J Osher. An l^1 penalty method for general obstacle problems. *SIAM Journal on Applied Mathematics*, 75(4):1424–1444, 2015.

[45] Michael Unser, Philippe Thevenaz, and Akram Aldroubi. Shift-orthogonal wavelet bases. *IEEE Transactions on Signal Processing*, 46(7):1827–1836, 1998.

[46] Gregory H Wannier. The structure of electronic excitation levels in insulating crystals. *Physical Review*, 52(3):191, 1937.

[47] Hassler Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934.

[48] Ke Yin and Stanley Osher. On the completeness of the compressed modes in the eigenspace. *UCLA CAM Reports 13-62*, 2013.

[49] Dominique Zosso, Braxton Osting, Mandy Xia, and Stanley J Osher. A fast primal-dual method for the obstacle problem. *UCLA CAM Reports 15-48*, 2015.