Weighted ℓ_1-Minimization for Parametric Operator Equations

Master's Thesis

Submitted to Lehrstuhl C für Mathematik (Analysis)
in Partial Fulfillment of the Requirements for the
Degree of

Master of Science

by
Benjamin Bykowski
Matriculation Number: 329364

Thesis Supervisor:
Prof. Dr. Holger Rauhut
and
Prof. Dr. Martin Grepl

Aachen, August 2015

Declaration

(Translation from German)

I hereby declare that I prepared this thesis entirely on my own and have not used outside sources without declaration in the text. Any concepts or quotations applicable to these sources are clearly attributed to them. This thesis has not been submitted in the same or substantially similar version, not even in part, to any other authority for grading and has not been published elsewhere.

Original Declaration Text in German:

Erklärung

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen worden ist. Alle Ausführungen, die wörtlich oder sinngemäß übernommen worden sind, sind als solche gekennzeichnet.

City, Date

Signature

Contents

1 Introduction and Overview

The number of models in science and engineering relying on Partial Differential Equations (PDEs) is abundant. Solutions to corresponding PDEs are however rarely analytically available, but have to be approximated numerically, often involving very time consuming computations. This might especially become a greater nuisance if the model – as certainly almost always is the case – depends on some parameters and one is interested in the solutions for many sets of parameters to determine certain properties of the system with respect to these influences. Even if the qualitative behaviour of a given model is clear and one is able to roughly estimate certain quantities after looking at solutions for only a few different parameters, it is usually hard to make these estimates precise.

1.1 An Example Problem

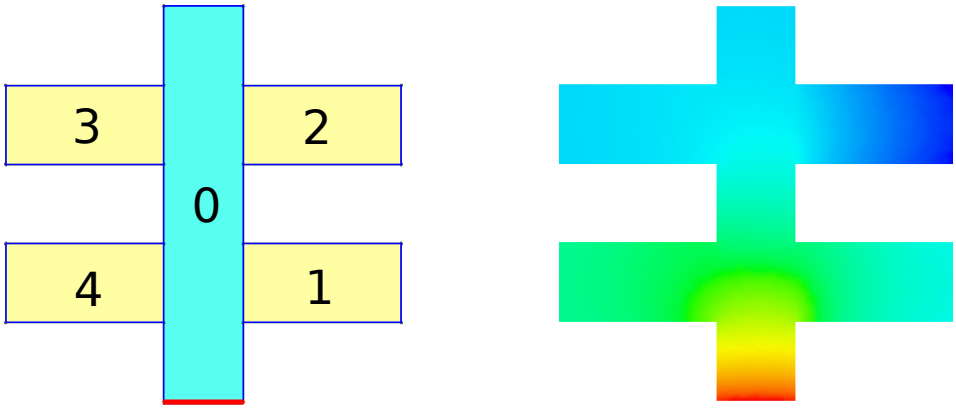
To have a concrete example in mind, one may consider a thermal fin, made of five different materials with different heat transfer coefficients cooling down some heat source at its root as illustrated in Figure 1.1 and assume one is interested in the average temperature of the fin at the heat source using different materials. The fin may be modelled using a simple diffusion equation

$$-\nabla \cdot (a(x,z)\nabla u(x,z)) = 0 \tag{1.1}$$

with parameter-dependent diffusion coefficient $a(\cdot, z)$ and parameter vector z. In the above example the coefficient simply varies spatially in the five regions, i.e.

$$a(x,z) := \sum_{j=1}^{5} z_j \Psi_j(x)$$

with $z_j \in [a,b]$ for some $0 < a < b$ and $\Psi_j(x)$ the characteristic function of the j-th region. Assuming this model is reasonable, the average temperature should continuously or even smoothly decrease with increasing heat transfer coefficient in any region. Given enough values for enough different sets of coefficients, it thus should be possible to estimate the temperature for any other set of parameters arbitrarily well, by, for example, simply linearly interpolating between the known values. This approach is however promptly subject to the *curse of dimensionality*. If, for example, only 10 samples are taken for every set of four fixed parameters, this results in a total number of 10^5 evaluations. Furthermore, although in this



(a) Schematic (b) Example solution

Figure 1.1: A thermal fin split into five regions having different heat transfer coefficients. The heat source is located at the red line in the schematic.

case linear interpolation should intuitively give a reasonable approximation, in general this might not be the case. Without sufficient intuition on the problem, one thus may not trust this method any longer for any moderate number of known values.

1.2 Uncertainty Quantification

In the setting of Uncertainty Quantification, where one is concerned with statistical properties of a *Quantity of Interest (QoI)* like the average temperature above, one almost always lacks intuition on the behaviour of the system in the parameters. Again using above example, the diffusion coefficient here may instead originate from a stochastic process via, for example, a Karhunen-Loève transformation

$$a(x,z) := \bar{a}(x) + \sum_{j\geq 1} z_j \Psi_j(x), \tag{1.2}$$

where $\Psi_j(x)$ now corresponds to the j-th eigenfunction of the process and z is an element of a potentially infinite-dimensional parameter space. The eigenfunctions often have to be approximated numerically themselves and all information available a-priori is a decay estimate on the $\Psi_j(x)$ with respect to a given norm.
Over the last 20 years, problems of these kind have received much well-deserved attention. The methods analysed and employed the most so far may be divided in *Monte Carlo* and *spectral* methods.
Monte Carlo methods essentially rely on nothing but the strong law of large num-

bers and are thus suited for a great class of problems (cf. [Nie92]). But in a sense these are thereby also bound to ignore a lot of information, such as that given QoI depends smoothly on the parameters and known convergence properties of Ψ_j.
The spectral methods make use of exactly this kind of additional knowledge. Put shortly, analogous to orthogonal decomposition of functions in a Fourier series, it is possible to expand a QoI, which is subject to a Gaussian stochastic process, into an exponentially L_2-convergent countable sequence of Hermite Polynomials and Gaussian random variables, analogous to the Karhunen-Loève expansion (1.2) above, by the Cameron-Martin Theorem (cf. [EMSU12, Section 2.1] [Xiu10]). This idea has successfully been applied to stochastic PDE via the so called *Stochastic Galerkin* method (cf. [GS91]). It has however also been observed that if the input to a given model is well-behaved in the sense of the Cameron-Martin Theorem, but not Gaussian, albeit an expansion via Hermite Polynomials is still possible, the exponential convergence rate is lost. As one might intuitively guesses, the exponential convergence can be restored by using polynomials other than the Hermite basis, that are specific to the given stochastic process at hand. These are explicitly known for processes based on standard probability distributions, such as the Poisson distribution. This approach is referred to as *generalized Polynomial Chaos (gPC)* (cf. [XK02b, Xiu10, EMSU12, Wiener-Askey Scheme]), making the spectral approach very promising for a large class of Uncertainty Quantification problems. Its efficiency has been demonstrated inter alia in [XS09, XK02a, WXGK05]. More importantly for the remainder of this thesis, an estimate for stochastic moments such as the expectation value and variance are easily obtained from the expansion, thanks to orthonormality of given gPC basis with respect to considered probability distribution. The expectation value, for example, simply corresponds to the coefficient of the constant basis function. Further generalizations to, e.g. non-smooth problems have also been investigated (cf. [MNGK04]).
The spectral methods may be further classified as *intrusive* and *non-intrusive* methods. The just mentioned Stochastic Galerkin method, for instance, belongs to the first kind. To project the solution in the given polynomial basis, the Galerkin method used to solve the deterministic PDE has to be modified. The implementation of this modification is of course an additional error source and in general also a highly non-trivial task. Non-intrusive or *stochastic collocation* methods on the other hand rely on nothing but solutions to given PDE for certain parameter values. Although this introduces an additional interpolation error, it has been observed that these methods can still perform very well [Elm11, BNTT11].
Finally stochastic collocation methods may again be split into the two subcategories of *interpolation* and *pseudospectral (discrete-projection)* methods. As suggested by the name, former methods try to interpolate a given QoI via some samples in a given gPC basis. An example for this kind of methods is given by random least squares projection as introduced in [CDL13, MNST14, MNvST13, Mig13]. Pseudospectral methods also interpolate the QoI in some gPC basis, but specifically do so by directly evaluating the orthogonal projections on the coefficients of the QoI in the gPC basis via quadrature (cf. [Xiu07] [Xiu10, Section 7.3]).

1.3 Aim of this Thesis

Recently, a new method based on results from Compressive Sensing has been introduced in [RS14], that by above terminology may be classified as stochastic collocation interpolation method. This so called *Compressive Sensing Petrov-Galerkin (CSPG)* method has been shown to potentially yield an efficient to obtain and evaluate interpolant in the Chebyshev basis to quantities, such as the average temperature at the root of the fin above, using only a relatively low number of solutions at randomly chosen parameters.
These findings rely themselves on two major recent results. First, the solutions to a large class of problems are indeed smooth, even analytic, in the parameters, as has been first shown for elliptic PDEs such as (1.1) in [CDS10a]. Second, it relies on the generalization of standard to Weighted Compressive Sensing, which has been introduced in [RW15], providing an improved framework for function interpolation in a polynomial basis such as the Chebyshev basis.

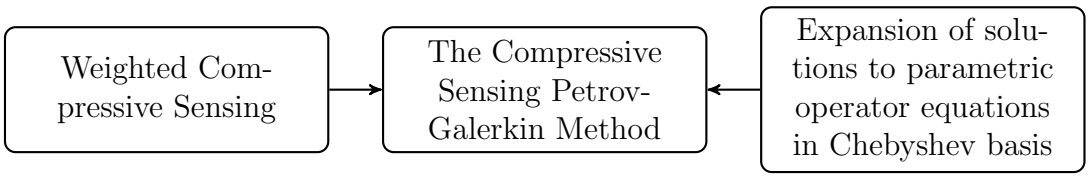


Figure 1.2: Overview

The *aim of this thesis* is to review and compile all these results and to give a first impression on its practicability and actual performance via numerical tests. The at first glance certainly major restriction to the Chebyshev basis is kept for the same technical reasons as in [RS14] throughout this thesis. As discussed later in Subsection 3.2.3, the CSPG method is also applicable using any other gPC. To solve occurring ℓ_1-minimization problems, algorithms well known from standard Compressive Sensing, such as *Hard Thresholding Pursuit* `(HTP)` and *Chambolle and Pock's method*, are translated to the weighted context. Furthermore, it is shown that the approach actually transfers to a much larger class of problems than elliptic PDEs employing the findings of [CCS14].
The potential strengths of CSPG may be summarized by the following properties:
It is non-intrusive, embarrassingly parallel and easy to implement overall. Nothing but a certain least number m of approximate evaluations of the corresponding function are required. Any method to solve the PDE in focus may be applied, as long as it is assured that the error of the evaluations lie within some prescribed bound. The values can simply be computed by executing the chosen method m times in parallel. The resulting weighted ℓ_1-minimization problems are efficiently solvable via various simple-to-implement algorithms.
The method is robust. The samples may be taken independently, i.e. non-adaptively. If samples obtained for certain parameter values turn out to be faulty,

it is possible to replace just these samples.
The interpolant is efficient to evaluate. As briefly mentioned in the previous section, thanks to the properties of the gPC basis, stochastic moments such as the mean value and variance of the interpolant with respect to corresponding probability measure are easily obtained from the computed coefficients (cf. Section 4.2). In case more involved properties of given function are of interest, by the very ansatz of CSPG the interpolant is a simple polynomial having only a low number of non-zero coefficients and is hence fast to evaluate.
A-priori error bounds are available. These allow especially to determine in which cases the CSPG method is a reasonable ansatz and potentially yields superior results when compared to, e.g. quasi-Monte Carlo methods (cf. Remark 4.3).

1.4 Related Work

Albeit the CSPG method is new, the idea to apply Compressive Sensing methods or weighted l_1-minimization to problems of these kind is not. Similar approaches, that can also be categorized as stochastic collocation interpolation methods, have been taken in [PHD14, DO11]. These however do not make use of the results established in [RW15], that link a-priori estimates of the coefficients to the solution in a gPC basis to a favourable ambient index set. A stochastic collocation pseudospectral approach has been taken in [Tan13] and shown to be effective, that is however of no further concern in this thesis.
Finally, the CSPG method may be interpreted as an extension to the random least squares projection method introduced in [CDL13, MNST14, MNvST13, Mig13]. It has been shown that simple least squares projection yields especially good results if the number of samples is proportional to the square of the number of basis functions N^2. Which basis functions to project on in the first place, is however not clear. It has been established that choosing a monotone and nested index set corresponding to the basis functions result in good convergence rates (cf. [Mig13, p. 51], [CCS14, Theorem 3.1]). In contrast to least-squares projection, one has not to be aware of precisely which basis functions are actually significant in the expansion of the solution. Instead, one only has to know that these s significant basis functions lie within some ambient space of N basis functions. Since the number of required samples $m = Cs\log^3(s)\log(N)$ scales only logarithmically in the ambient dimension N, the ambient space can be chosen quite conservatively and is not required to be monotone nor nested. Nevertheless, in [RS14], up to logarithmic factors, similar convergence rates as in [CCS14, Theorem 3.1] in the number of samples have been established. Thus, ℓ_1-minimization, so to speak, selects the best basis functions automatically. The difference to direct least-squares projection manifests particularly strikingly in the *Hard Thresholding Pursuit* algorithm and is discussed again at the end of Subsection 2.2.3.

1.5 Thesis Structure

The basic theory of *Weighted* Compressive Sensing is briefly reviewed in Chapter 2. Furthermore `Hard Thresholding Pursuit` (`HTP`), `Compressive Sensing Matching Pursuit` (`CoSaMP`) and `Chambolle and Pock's Primal Dual Algorithm` are translated to the weighted context. Reconstruction guarantees analogous to those known from standard Compressive Sensing in terms of the – also in this chapter introduced – weighted restricted isometry constant and quasi-best weighted approximation are given. In Chapter 3 the Holomorphy Assumption is introduced and shown that under this assumption a Chebyshev expansion of the solutions to a parametric operator equation with respect to the parameters is possible. It is furthermore demonstrated that the Holomorphy Assumption allows to obtain the expansion coefficients via weighted ℓ_1-minimization. Utilizing these results, the Compressive Sensing Petrov-Galerkin (CSPG) method is presented in Chapter 4. The algorithm and error estimates for the reconstructed function are given and the ingredients for its efficient application are analysed. Numerical tests of the CSPG method are then discussed in Chapter 5. Finally, the results are summarized, issues revised and directions for further research given in Chapter 6.

1.6 A Short Note on the Notation

Unless stated otherwise, operations on sequences are to be understood pointwise. For instance, given sequences $(\rho_j)_{j\in\mathbb{N}}$, $(b_j)_{j\in\mathbb{N}}$ the statement $\rho > 1$ is equivalent to $\rho_j > 1, \forall j \in \mathbb{N}$ and $\rho \cdot b$ is defined as the sequence $(\rho_j \cdot b_j)_{j\in\mathbb{N}}$.

2 Weighted Compressive Sensing

Starting with the seminal papers of Donoho, Candes, Romberg and Tao [Don06, CRT06], Compressive Sensing has been a popular subject in applied mathematics for almost a decade now. At the heart of Compressive Sensing lies the observation that a high-dimensional vector $x \in \mathbb{C}^N$ is recoverable from a vector $y \in \mathbb{C}^m$ with $m \ll N$ related by a linear system of equations $Ax = y$, given only a few entries of x are non-zero and the matrix $A \in \mathbb{C}^{m \times N}$ fulfills certain properties discussed later. This setting applies in many situations (cf. [FR13, Section 1.2]). The concrete application important in the remainder of this thesis is the reconstruction of some function f in the Chebyshev basis via just a relatively few known values $f(z_j) = y_j$ at points z_j. The vector x then corresponds to the Chebyshev coefficients, the vector y to the function values, the matrix A to the sample matrix $A_{kl} = T_l(z_k)$, where T_l is the l-th Chebyshev polynomial. Given the samples y are obtained by a numerical approximation of f, the condition $Ax = y$ is certainly impractical, since it is at best known, that the values lie within some range of reliability, i.e. satisfy $\|Ax - y\|_2 \leq \eta$ for some $\eta \geq 0$. Also the condition that x has some exact number of non-zero entries is very restricting in practice. One probably rather just knows that only a certain number of coefficients in x are significant, i.e. there exists a vector x_S having a small number s of entries that are non-zero such that $\|x - x_S\|_1 \leq \varepsilon$ for some $\varepsilon \geq 0$. Given the matrix A satisfies the so called *robust null space property*, it is possible to reconstruct (a good approximation to) x even under these weaker conditions by solving the efficiently treatable ℓ_1-optimization problem (cf. [FR13, Thm. 4.15])

$$\min_{z \in \mathbb{C}^N} \|z\|_1 , \quad \text{s.t. } \|Ax - y\|_2 \leq \eta. \tag{2.1}$$

The robust null space property is however often hard to check. To simplify this task, the so called *restricted isometry property* has been introduced, which implies the null space property. While still no specific matrix is known to fulfill the restricted isometry property either, it has nevertheless been proven, that *subgaussian random matrices*, where the entries are simply given by $m \cdot N$ independent draws of a subgaussian random process, satisfy the restricted isometry property with high probability, provided

$$m \geq Cs \log(N),$$

where C is an universal constant independent of s and N. Thus a vector x in a high dimensional space is recoverable from a relatively short measurement vector y,

whose size scales only linear in the number of actually significant entries s of x and *logarithmically* in the ambient dimension N of x. It has also been shown that any matrix of the form $A = (\Phi_j(z_i))_{ij}$, where $\{\Phi_j \mid j \in [N]\}$ is some orthonormal system, such that $\|\Phi_j\|_\infty \leq K$ for all j, and z_i are independent random draws with respect to the orthogonalization measure associated to the orthonormal system, satisfy the restricted isometry property with high probability, provided

$$m \geq CK^2 s \log^3(s) \log(N). \tag{2.2}$$

If a function f is described by an approximately sparse vector x in the Chebyshev basis, it is thus possible to infer x from a relatively small set of samples y, and thereby reconstruct the function f by only a relatively few evaluations of this function. Naive application of this method however yields in many cases two major problems. First, as it is the case with interpolation via, e.g. least-squares, the reconstructed function may exhibit a behaviour resembling the Runge phenomenon (cf. [RW15]). Second, the number of required samples grows quadratically with the L_∞-bound K of the orthonormal system. When interpolating using, for example, Legendre-polynomials L_j, where $\|L_j\|_\infty = \sqrt{2j+1}$, the logarithmic scaling $\log(N)$ of the number of measurements with the ambient dimension is therefore rendered obsolete, due to the quick growth of K with N. One faces the same problem, when interpolating in a multidimensional Chebyshev basis as shown later in Section 3.2.
Recently, the more general theory of Weighted Compressive Sensing has been introduced in [RW15], which inter alia addresses just mentioned issues. Assuming an expansion x of f in the orthornomal system $\{\Phi_j \mid j \in [N]\}$ lies in a weighted ℓ_1 space, i.e. there exists a weight vector ω satisfying $\omega_j \geq \|\Phi_j\|_\infty$ and

$$\sum_{j \geq 1} |x_j| \, \omega_j < \infty,$$

it has been established that

$$m \geq CK^2 s \log^3(s) \log(N)$$

samples are enough to reconstruct f, now assuming the expansion of f is so called ω-s-sparse. Briefly, Weighted Compressive Sensing methods thereby take into account information about the rate of convergence of the given function in an orthonormal system. Information of this kind is often available. It is, for example, a basic result in classical Fourier Analysis, that – roughly speaking – the smoother a function, the faster its Fourier series converges. More precisely, given a function f, such that for $r \in \mathbb{N}$ all derivatives $\partial^\alpha f$ with $|\alpha| \leq r$ exist, its Fourier coefficients

decay with order r, i.e.

$$\left|\hat{f}(m)\right| \leq c_{n,r} \frac{\max\left(\|f\|_{L_1}, \sup_{|\alpha|=r} \|\partial^\alpha f\|_{\dot{\Lambda}_\gamma}\right)}{(1+|m|)^r} \tag{2.3}$$

where $\|\cdot\|_{\dot{\Lambda}_\gamma}$ is the so called γ-Lipschitz-Norm (cf. [Gra08, Corollary 3.2.10.]). If f is smooth, one may thus expect these methods to perform better compared to methods known from standard Compressive Sensing.
As the name implies, Weighted Compressive Sensing generalizes the whole theory of standard Compressive Sensing by additionally considering some weight sequence associated to the vector x and potentially applies to a much wider range of problems than just function interpolation. This of course also implies, that all of the definitions, theorems and algorithms known from standard Compressive Sensing have to be modified appropriately. The aim of the first part of this section is to summarize all of these definitions and results, important for the remainder of the thesis. Setting the frequently recurring weight sequence ω as $\omega \equiv 1$ recovers the definitions and results known from standard Compressive Sensing. The second part then introduces the weighted counterparts to the `Hard Thresholding Pursuit` and `Compressive Sampling Matching Pursuit` algorithms found in standard Compressive Sensing and Chambolle and Pock's Primal Dual Algorithm adapted to weighted ℓ_1-minimization.

2.1 Theory

Now for the formal definition of already mentioned and for all further results fundamental weighted ℓ_p-spaces.

Definition 2.1 (Weighted ℓ_p-spaces)
Given $p \in [0,2]$ and an index set Λ, the weighted ℓ_p-space over Λ is defined as

$$\ell_{\omega,p} := \left\{ x = (x_j)_{j\in\Lambda} \mid \|x\|_{\omega,p} < \infty \right\},$$

where $\|x\|_{\omega,p} := \left(\sum_{j\in\Lambda} |x_j|^p \, \omega_j^{2-p}\right)^{\frac{1}{p}}$ and $\|x\|_{\omega,0} := \sum_{j\in\operatorname{supp} x} \omega_j^2$ with $\operatorname{supp} x := \left\{ j \in J \mid x_j \neq 0 \right\}$.

For $p \in [1,2]$, the map $\|\cdot\|_{\omega,p}$ is indeed a norm by a straightforward generalization of the standard result in unweighted ℓ_p-spaces.

Definition 2.2
A vector $x \in \mathbb{C}^N$ is called *weighted s-sparse* or *ω-s-sparse* if $\|x\|_{\omega,0} \leq s$.

The at first glance peculiar choice of summing x against ω_j^{2-p} instead of ω_j^p has been made for ease of notation later. While the best approximation error of a

vector x by an s-sparse vector, commonly notated as

$$\sigma_s(x)_p = \min_{\{z \,|\, \|z\|_0 \leq s\}} \|x - z\|_p \,,$$

is easy to compute, the weighted analogue

$$\sigma_s(x)_{\omega,p} = \min_{\{z \,|\, \|z\|_{\omega,0} \leq s\}} \|x - z\|_{\omega,p}$$

is in general hard to calculate. To simplify this task, one makes the following definitions.

Definition 2.3 (Quasi-Best Weighted s-term Approximation [RW15, p.19])
Given x in some $\ell_{\omega,p}(\Lambda)$, define the non-increasing rearrangement v by $v_k := \left(\left|x_{\pi(j)}\right|^p \omega_{\pi(j)}^{-p}\right)$, where π is the unique permutation, such that $v_1 \geq v_2 \geq \cdots \geq 0$ and $\pi(j) > \pi(k)$ if $j > k$ in case $v_j = v_k$. Denote by k the maximal integer satisfying $\sum_{j=1}^{k} \omega_{\pi(j)}^2 \leq s$ and by S the set $\{\pi(1), \ldots, \pi(k)\}$. Then x_S is called *quasi-best ω-s-term approximation to x.*

Remark 2.4
Notice that the index set S in Definition 2.3 is independent of p.

Definition 2.5
With notations as in Definition 2.3 the *approximation error of the quasi-best ω-s-term approximation to x* is defined by

$$\tilde{\sigma}_s(x)_{\omega,p} := \|x - x_S\|_{\omega,p} = \|x_{S^c}\|_{\omega,p} \,.$$

While $\sigma_s(x)_{\omega,p} \leq \tilde{\sigma}_s(x)_{\omega,p}$ by definition, there are cases, where the quasi-best and actual best weighted approximation to a vector x do not coincide. An example is the case $s = 3$, $\omega = [1, \sqrt{2}, \sqrt{3}]$ and $x = [9, 9, 10]$ due to Jason Jo [Jo13]. Here one has $\tilde{\sigma}_3(x)_{\omega,2} = 9\sqrt{2}$, where $\tilde{x} = [0, 0, 10]$ is the quasi-best weighted approximation, while $\sigma_3(\hat{x})_{\omega,2} = 10 < 9\sqrt{2}$ for the actual best 3-sparse weighted approximation $\hat{x} = [9, 9, 0]$. The quasi-best weighted approximation is still very useful, as the following Lemma shows.

Lemma 2.6 (Quasi-Best approximation Error Bound [RW15, Lemma 3.1])
If $s \geq \|\omega\|_\infty^2$, then, for x in some $\ell_{\omega,p}(\Lambda)$, it holds

$$\tilde{\sigma}_{3s}(x)_{\omega,p} \leq \sigma_s(x)_{\omega,p}.$$

As the easier computability of the quasi-bound might already indicate, this estimate will be very important in the next section on reconstruction algorithms. Using this inequality one is also able to prove a weighted analogue to the well-known Stechkin-inequality (cf. [FR13, Proposition 2.3]).

Theorem 2.7 (Weighted Stechkin [RW15, Theorem 3.2])
For $0 < p < q \leq 2$, x in some $\ell_{\omega,p}(\Lambda)$ and $s \geq \|\omega\|_\infty^2$ the following inequalities hold

$$\sigma_s(x)_{\omega,q} \leq \tilde{\sigma}_s(x)_{\omega,q} \leq \left(s - \|\omega\|_\infty^2\right)^{1/q-1/p} \|x\|_{\omega,p}\,.$$

The condition $s \geq \|\omega\|_\infty^2$ is natural, since a vector x with a non-zero entry x_j corresponding to a weight $\omega_j > s$ is not s-sparse by definition.

Recovery under the Weighted Null Space Property

As mentioned in the introduction, the so called robust null space property allows for the recovery of a sparse vector x via solving the ℓ_1-minimization problem (2.1). An analogous result holds for matrices A fulfilling the ω-NSP defined below.

Definition 2.8 (ω-NSP [RW15, Definition 4.1])
Given weights ω, a matrix $A \in \mathbb{C}^{m\times N}$ is said to satisfy the *weighted robust null space property* (ω-NSP) of order s with constants $\rho \in (0,1)$ and $\tau > 0$ if

$$\forall v \in \mathbb{C}^N, S \subseteq [N], \omega(S) \leq s : \|v_S\|_{\omega,1} \leq \frac{\rho}{\sqrt{s}} \left\|v_{\overline{S}}\right\|_{\omega,1} + \tau \|Av\|_2\,.$$

Error bounds for sparse recovery via weighted ℓ_1-minimization in the form given in the following theorem are used in all subsequent results based on Compressive Sensing in this thesis.

Theorem 2.9 (Weighted ℓ_1-Recovery Error Bounds [RW15, Corollary 4.3])
Let $A \in \mathbb{C}^{m\times N}$ satisfy the ω-NSP of order s with constants $\rho \in (0,1)$ and $\tau > 0$. Then for $x \in \mathbb{C}^N$, $y = Ax + \xi$ with $\|\xi\|_2 \leq \eta$ and the solution $x^\#$ to

$$\min_{z\in\mathbb{C}^N} \|z\|_{\omega,1} \text{ s.t. } \|Az - y\|_2 \leq \eta, \tag{2.4}$$

it holds

$$\left\|x - x^\#\right\|_{\omega,1} \leq c_1 \sigma_s(x)_{\omega,1} + d_1 \sqrt{s}\eta.$$

And if additionally $s \geq 2\|\omega\|_\infty^2$, it holds

$$\left\|x - x^\#\right\|_2 \leq c_2 \frac{\sigma_s(x)_{\omega,1}}{\sqrt{s}} + d_2\eta,$$

where $c_1, c_2, d_1, d_2 > 0$ depend only on ρ and τ.

Problem 2.4 is also called *quadratically constrained weighted basis pursuit*. The proof to Theorem 2.9 employs the following lemma of rather technical nature.

Lemma 2.10 (ω-NSP is Robust [RW15, Theorem 4.2])
Let $A \in \mathbb{C}^{m\times N}$ be a matrix satisfying the ω-NSP of order s with constants $\rho \in (0,1)$ and $\tau > 0$. Then one has the bound

$$\forall x, z \in \mathbb{C}^N : \|z - x\|_{\omega,1} \leq \frac{1+\rho}{1-\rho}\left(\|z\|_{\omega,1} - \|x\|_{\omega,1} + 2\sigma_s(x)_{\omega,1}\right) + \frac{2\tau\sqrt{s}}{1-\rho}\left\|A(z-x)\right\|_2 .$$

And if additionally $s \geq 2\|\omega\|_\infty^2$, it holds

$$\forall x, z \in \mathbb{C}^N : \|z - x\|_2 \leq \frac{C_1}{\sqrt{s}}\left(\|z\|_{\omega,1} - \|x\|_{\omega,1} + 2\sigma_s(x)_{\omega,1}\right) + C_2\left\|A(z-x)\right\|_2 ,$$

where C_1, C_2 are absolute constants only depending on τ, ρ, s and $\|\omega\|_\infty$.

Proof (Theorem 2.9) By definition of $x^\#$ one has $\left\|x^\#\right\|_{\omega,1} - \|x\|_{\omega,1} \leq 0$ and

$$\left\|A(x - x^\#)\right\|_2 \leq \|Ax - y\|_2 + \left\|Ax^\# - y\right\|_2 \leq 2\eta.$$

Setting $z = x^\#$ in Lemma 2.10 then directly implies the result. □

Recovery under the Weighted Restricted Isometry Property

The ω-NSP is a little unwieldy and does not always directly provide much insight into what matrices allow for reconstruction via weighted basis pursuit. Again in complete analogy to standard Compressive Sensing theory, the ω-RIP below provides some remedy.

Definition 2.11 (ω-RIP [RW15, Definition 1.3])
For $A \in \mathbb{C}^{m\times N}$, $s \geq 1$ and weights ω, A is said to fulfill the *weighted restricted isometry property* (ω-RIP) with *weighted restricted isometry constant* $\delta_{\omega,s}$ if

$$\left(1 - \delta_{\omega,s}\right)\|x\|_2^2 \leq \|Ax\|_2^2 \leq \left(1 + \delta_{\omega,s}\right)\|x\|_2^2 \tag{2.5}$$

for all $x \in \mathbb{C}^N$ with $\|x\|_{w,0} \leq s$.

The ω-RIP above is weaker than the standard RIP-condition. Following the argument of [Jo13] if one defines the largest number of non-zero entries with weighted cardinality less than s as

$$P_\omega(s) := \max_{\omega(I)\leq s} |I| ,$$

it is immediately evident, that if a matrix A satisfies the standard RIP of order $P_\omega(s)$ for some weights ω, it also satisfies the ω-RIP of order s, since $\|x\|_{\omega,0} \geq \|x\|_0$. Except for the case $\omega \equiv 1$, where $\|\cdot\|_{\omega,0} \equiv \|\cdot\|_0$ however, it is possible that $\|x\|_{\omega,0} > \|x\|_0$ and thus $\|x\|_0 \leq s$ does not imply $\|x\|_{\omega,0} \leq s$ and hence inequality (2.5) for $\|\cdot\|_0$ might not hold. Now for the promised main result of this subsection.

Theorem 2.12 (ω-RIP Implies ω-NSP [RW15, Theorem 4.5])
Given weights ω, $s \geq 2\|\omega\|_\infty^2$ and $A \in \mathbb{C}^{m\times N}$ satisfying the ω-RIP with

$$\delta_{\omega,3s} \leq \frac{1}{3}, \tag{2.6}$$

the matrix A satisfies the ω-NSP of order s with constants

$$\rho = \frac{2\delta_{\omega,3s}}{1-\delta_{\omega,3s}} \in (0,1) \quad \text{and} \quad \tau = \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,3s}} > 0.$$

Remark 2.13
By employing more involved methods, the constant in the above bound (2.6) may be improved [FR13, Theorem 6.12]. Also conditions on $\delta_{\omega,2s}$ instead of $\delta_{\omega,3s}$ are possible.

Remark 2.14
The ω-RIP constant can be rewritten as $\delta_{\omega,s} = \max_{S\subseteq[N],\,\omega(S)\leq s}\left\|A_S^* A_S - \mathrm{Id}\right\|_{2\to 2}$, where $\|\cdot\|_{2\to 2}$ is the usual spectral norm.

Weighted Reconstruction of a Function

As mentioned in the introduction, random sampling matrices coming from some orthonormal system satisfy the restricted isometry property with high probability. Its weighted generalization is stated in the following theorem.

Theorem 2.15 (ω-RIP for Matrix from ONS [RW15, Theorem 5.2])
Given an orthonormal system $(\Phi_j)_{j\in\Lambda}$ with $N := |\Lambda| < \infty$, weights $\omega_j \geq \left\|\Phi_j\right\|_\infty$, parameters $\delta, \gamma \in (0,1)$ and

$$m \geq C\delta^{-2} s \max\left\{\log^3(s)\log(N), \log\frac{1}{\gamma}\right\} \tag{2.7}$$

samples t_l drawn i.i.d. with respect to the orthogonalization measure associated to $(\Phi_j)_{j\in\Lambda}$, the normalized sampling matrix

$$\tilde{A}_{lk} := \frac{1}{\sqrt{m}}\Phi_k(t_l)$$

satisfies the ω-RIP of order s with $\delta_{\omega,s} \leq \delta$ with probability exceeding $1-\gamma$.

Remark 2.16
It is conjectured, that at least a factor $\log^2(s)$ in the inequality for the required number of samples (2.7) is superfluous.

Essentially an application of Theorem 2.12 and Theorem 2.9 to this result then yields the following theorem, granting interpolation of functions using weighted ℓ_1-minimization. Details follow in the proof below.

Theorem 2.17 (Weighted Reconstruction of a Function)
Let $(\Phi_j)_{j\in J}$ be an orthonormal system with a possibly countably infinite index set J, a sparsity parameter $s \geq 1$, weights ω, such that $\omega_j \geq \left\|\Phi_j\right\|_\infty$, and

$$J_0^s(\omega) := \left\{ j \in J \mid \omega_j^2 \leq \frac{s}{2} \right\} \tag{2.8}$$

with $|J_0^s| < \infty$. Let further $f(t) = \sum_{j\in J} x_j \Phi_j(t)$ be a fixed function with $\|f\|_{\omega,1} < \infty$ and $\eta_1 \geq 0$ such that $\left\|f - f_{J_0^s}\right\|_2 \leq \eta_1$, where $f_{J_0^s} := \sum_{j\in J_0^s} x_j \Phi_j(t)$. Then given

$$m \geq c_0 s \max\left\{ \log^3(s)\log(N), \log\frac{1}{\gamma} \right\} \tag{2.9}$$

sampling points t_l drawn i.i.d. with respect to the orthogonalization measure associated to $(\Phi_j)_{j\in J}$, corresponding samples $y_l = f(t_l) + \xi_l$, such that $\|\xi\|_2 \leq \eta_2$, the sampling matrix $A_{ij} := \Phi_j(t_i)$ and the solution $x^\#$ to

$$\min_z \|z\|_{\omega,1} \text{ s. t. } \|Az - y\|_2 \leq \eta := \eta_1 + \eta_2$$

the reconstructed function $f^\#(t) = \sum_{j\in J_0^s} x_j^\# \Phi_j(t)$ satisfies with probability at least $1-\gamma$ the inequalities

$$\left\|f - f^\#\right\|_\infty \leq \left\|f - f^\#\right\|_{\omega,1} \leq c_1 \sigma_s(f)_{\omega,1} + d_1 \sqrt{\frac{s}{m}}\eta, \tag{2.10}$$

$$\left\|f - f^\#\right\|_2 \leq c_2 \frac{\sigma_s(f)_{\omega,1}}{\sqrt{s}} + d_2 \sqrt{\frac{1}{m}}\eta, \tag{2.11}$$

where c_0, c_1, d_1, c_2, d_2 are absolute constants.

Proof (Theorem 2.17) The basic idea of the proof is to interpret the samples taken from f as noisy samples of $f_{J_0^s}$ via the decomposition $f = f_{J_0^s} + f_{J\setminus J_0^s}$, where analogously to the definition of $f_{J_0^s}$, $f_{J\setminus J_0^s} := \sum_{j\in J\setminus J_0^s} x_j \Phi_j(t)$. The part $f_{J_0^s}$ is now dealt with via Theorem 2.15, which implies that the resulting normalized matrix $\tilde{A}$ only depending on the finite index set J_0^s satisfies the ω-RIP of order s with $\delta_{\omega,s} \leq \frac{1}{3}$ and probability at least $1-\gamma$ using a number m of measurements as in (2.9). Then Theorem 2.12 implies that $\tilde{A}$ also satisfies the ω-NSP of order s, where it is used that $s \geq 2\|\omega\|_\infty^2$ by definition of J_0^s. Noting that $\|Az - y\| \leq \eta$ with $y = Ax + \hat{\eta}$ is equivalent to $\left\|\tilde{A}z - \tilde{y}\right\| \leq \frac{\eta}{\sqrt{m}}$ with $\tilde{y} = \tilde{A}x + \frac{\hat{\eta}}{\sqrt{m}}$, Theorem 2.9 implies the bounds on $\left\|f - f^\#\right\|_{\omega,1}$ and $\left\|f - f^\#\right\|_2$.

The bound on $\left\|f - f^{\#}\right\|_\infty$ is established via

$$\left\|f - f^{\#}\right\|_\infty \leq \sum_{j=1}^{N} \left|x_j - x_j^{\#}\right| \left\|\Phi_j\right\|_\infty \leq \left\|f - f^{\#}\right\|_{\omega,1}. \qquad \square$$

How good are these results? As a quick reminder: the best known bound on the number of required samples in standard Compressive Sensing is given by (2.2) as

$$m \geq CK^2 s \log^3(s) \log(N),$$

where K is some constant so that $\left\|\Phi_j\right\|_\infty \leq K$ for all $j \in \Lambda$. Comparing this to (2.7), the factor K^2 has been eliminated. But the sparsity s of a set S now depends on weights via $\omega(S) = \sum_{j \in S} \omega_j^2$. By the condition $\omega_j \geq \left\|\Phi_j\right\|_\infty$, a vector that has been 1-sparse in the standard sense might now be $\left\|\Phi_j\right\|_\infty^2$-sparse in the weighted sense. In the case $\left\|\Phi_j\right\|_\infty$ is constant, (2.2) and (2.7) are equivalent and nothing has been gained. In fact the bound is even worse due to the $\log(s)$ factors, making these very suspicious as stated in Remark 2.16. In some important cases however, the new bound (2.7) is an improvement. The advantage is particularly significant, when interpolating a smooth function f in some orthonormal system, where $\left\|\Phi_j\right\|_\infty$ grows with j, since by results like (2.3), the largest and thus most significant coefficients of a function are those corresponding to small j and hence small $\left\|\Phi_j\right\|_\infty$.
Also certainly standing out at first glance is the error bound (2.10), which grants an L_∞-bound that is not available in standard Compressive Sensing. Possibly also surprising is the dependence on the sample error η. One must not overlook that $\eta = \eta_1 + \eta_2$ might grows with the number of samples, as $\|\xi\|_2$ and thus η_2 might grows with m. A better intuition on these can be obtained by splitting the error η into η_1 and η_2 in the right hand sides of (2.10) and (2.11). Assuming $\left|y_l - f(t_l)\right| \leq \tau$ and setting $\eta_2 = \sqrt{m}\tau$ one then gets the following corollary:

Corollary 2.18
With notations as in Theorem 2.17 and given $\left|y_l - f(t_l)\right| \leq \tau$, it holds

$$\left\|f - f^{\#}\right\|_\infty \leq \left\|f - f^{\#}\right\|_{\omega,1} \leq c_1 \sigma_s(f)_{\omega,1} + d_1 \sqrt{\frac{s}{m}} \eta_1 + d_1 \sqrt{s} \tau,$$
$$\left\|f - f^{\#}\right\|_2 \leq c_2 \frac{\sigma_s(f)_{\omega,1}}{\sqrt{s}} + d_1 \frac{\eta_1}{\sqrt{m}} + d_2 \tau.$$

Now setting $\eta_1 = \left\|f - f_{J_0^s}\right\|_2$, η_1 decreases with increasing s. On the other hand τ remains constant. In the limit, the error $\left\|f - f^{\#}\right\|_2$ thus depends constantly on τ and the error $\left\|f - f^{\#}\right\|_\infty$ grows with $\sqrt{s}$.

2.2 Reconstruction Algorithms

In this section some of the well known reconstruction algorithms from standard Compressive Sensing are adapted to the weighted case.

2.2.1 Approximation Algorithms

The adaption and analysis of the algorithms are quite similar to each other. Except for two steps each, where the properties of the quasi-best hard thresholding operator (2.21) are used, the proofs carry over verbatim from the standard case as found in [FR13, Chapter 6].
The following two simple lemmas are of crucial importance to the analysis of any of the algorithms presented in this section.

Lemma 2.19 ([FR15, Lemma 3.6])
Given $u, v \in \mathbb{C}^N$ and an index set $S \subseteq [N]$, it holds

$$\begin{aligned} \left|\left\langle u, (\mathrm{Id} - A^*A)\, v\right\rangle\right| &\leq \delta_{\omega,t}\|u\|_2\|v\|_2\,, && \text{if } \omega(\mathrm{supp}(u) \cup \mathrm{supp}(v)) \leq t, \\ \left\|\left((\mathrm{Id} - A^*A)\, v\right)_S\right\|_2 &\leq \delta_{\omega,t}\|v\|_2\,, && \text{if } \omega(S \cup \mathrm{supp}(v)) \leq t. \end{aligned}$$

Lemma 2.20 ([FR15, Lemma 3.8])
Given $y \in \mathbb{C}^m$, $S \subseteq [N]$ and $\omega(S) \leq s$ one has the estimate

$$\left\|(A^*y)_S\right\|_2 \leq \sqrt{1+\delta_{\omega,s}}\|y\|_2\,.$$

Further motivated by the quasi-best ω-approximation in Definition 2.3 one defines the following surrogate to the *hard thresholding operator* in standard Compressive Sensing:

Definition 2.21
With notations as in Definition 2.3 denote the *quasi-best hard thresholding operator* by $\tilde{H}_s(x)_\omega := x_S$.

By Remark 2.4 the operator $\tilde{H}_s(x)_\omega$ is indeed well-defined. The standard hard thresholding operator H satisfies $\|u_S\|_p \leq \left\|H_s(u)\right\|_p$ given any index set S of cardinality at most s. An analogous result is valid for $\tilde{H}_s(u)_\omega$.

Lemma 2.22
Given u in some $\ell_{\omega,p}(\Lambda)$, where $s \geq \|\omega\|_\infty$, any index set S with $\omega(S) \leq s$ and $R := \mathrm{supp}\, \tilde{H}_{3s}(u)_\omega$, it holds

$$\|u_S\|_{\omega,p} \leq \|u_R\|_{\omega,p}\,.$$

PROOF Define T as the index set of the best $\|\cdot\|_{\omega,p}$-approximation to u satisfying $\omega(T) \leq s$, that is

$$\sigma_s(u)_{\omega,p} = \|u - u_T\|_{\omega,p} = \left\|u_{\overline{T}}\right\|_{\omega,p}.$$

By definition of T, it holds

$$\left\|u_{\overline{T}}\right\|_{\omega,p} \leq \left\|u_{\overline{S}}\right\|_{\omega,p} \text{ or equivalently } \|u_S\|_{\omega,p} \leq \|u_T\|_{\omega,p}.$$

Further Lemma 2.6 leads to the inequality

$$\left\|u_{\overline{R}}\right\|_{\omega,p} \leq \left\|u_{\overline{T}}\right\|_{\omega,p}, \text{ which is equivalent to } \|u_T\|_{\omega,p} \leq \|u_R\|_{\omega,p}.$$

Combining these estimates then implies claim. □

Weighted Hard Thresholding Pursuit and Iterative Hard Thresholding

At the heart of the algorithms presented in this paragraph lie the following observations: Given a matrix A satisfying the weighted restricted isometry property for some s and ω, it holds $(A^*A)_S \approx \mathrm{Id}$ for sets S with $\omega(S) \leq s$. Further, if the error in the samples y is small, one has $Ax \approx y$. Assuming x and x^k are s-sparse, it thus holds $A^*(y - Ax^k) = A^*A(x - x^k) \approx x - x^k$ or equivalently $x^k + A^*(y - Ax^k) \approx x$. Setting $x^0 := 0$ and $x^{k+1} := \tilde{H}_{3s}(x^k + A^*(y - Ax^k))_\omega$ then defines a sequence of ω-$3s$-sparse vectors, that are close to x under the conditions of Lemma 2.6. This directly motivates the weighted analogue to `IHT` given by Algorithm 1. Analysis for Weighted Iterative Hard Thresholding (`WIHT`) may be found in [FR15, Jo13].

Algorithm 1 Weighted Iterative Hard Thresholding [FR15, Algorithm 3.4]

Input:
Normalized measurement matrix $A \in \mathbb{C}^{m \times N}$
Measurement vector $y \in \mathbb{C}^m$
Weights $\omega \in \mathbb{R}^N_{\geq 1}$
Output: $3s$-sparse approximation to $x^\#$
1: Initialize target support set $S^0 := \emptyset$, vector $x^0 := 0$
2: **while** Stopping criterion is not met **do**
3: $\quad x^{k+1} := \tilde{H}_{3s}(x^k + A^*(y - Ax^k))_\omega$

When using standard Iterative Hard Thresholding with random matrices A, the resulting sequences x^k in practice sometimes diverge. This might be due to bad luck, i.e. A not actually satisfying the weighted restricted isometry property. Solving a least-squares problem in between the thresholding steps lead to a more stable behaviour. This algorithm is known as `Hard Thresholding Pursuit` [FR13, p.150]. Its weighted analogue is stated in Algorithm 2.
Reconstruction guarantees for these algorithms are given by the following lemma.

Algorithm 2 Weighted Hard Thresholding Pursuit (WHTP)

Input:
Normalized measurement matrix $A \in \mathbb{C}^{m\times N}$
Measurement vector $y \in \mathbb{C}^m$
Weights $\omega \in \mathbb{R}^N_{\geq 1}$
Output: $3s$-sparse approximation to $x^\#$
1: Initialize target support set $S^0 := \emptyset$, vector $x^0 := 0$
2: **while** Stopping criterion is not met **do**
3: $S^{k+1} := \operatorname{supp} \tilde{H}_{3s}(x^k + A^*(y - Ax^k))_\omega$
4: $x^{k+1} := \operatorname{argmin} \left\{ \|y - Av\|_2 \mid v \in \mathbb{C}^N, \operatorname{supp}(v) \subseteq S^{k+1} \right\}$

Lemma 2.23 (Reconstruction via WHTP and WIHT)
Suppose the weighted restricted isometry constant $\delta_{\omega,7s}$ of $A \in \mathbb{C}^{m\times N}$ satisfies

$$\delta_{\omega,7s} \leq \frac{1}{\sqrt{3}} \approx 0.5773 \tag{2.12}$$

and $x \in \mathbb{C}^N$ is ω-s-sparse with $s \geq \|\omega\|^2_\infty$. Then for $e \in \mathbb{C}^m$, $S \subseteq [N]$ with $\omega(S) \leq s$ and $y = Ax + e$ the sequence x^n defined by Algorithm 2 satisfies

$$\|x^n - x_S\|_2 \leq \rho^n \left\|x^0 - x_S\right\|_2 + \tau \left\|Ax_{\overline{S}} + e\right\|_2, \tag{2.13}$$

where the constants $\rho \in (0,1)$ and $\tau \in \mathbb{R}^+$ depend only on $\delta_{\omega,7s}$.

Notice that Lemma 2.23 only guarantees convergence of a subsequence to a cluster point.
The condition on the restricted isometry property of A is for both algorithms the same. The improvement of WHTP over WIHT is reflected by smaller constants τ and ρ in the above lemma.
By line 4, the vector x^{k+1} is determined by the index set S^{k+1} alone. All other quantities, such as y, s and ω are constant over all iterations. Since S^{k+1} depends in the same manner only on x^k, S^{k+1} actually depends only on S^k and a natural stopping criterion for the algorithm is given by $S^{n+1} = S^n$. In case $S^{n+1} \neq S^n$, but, e.g. $S^{n+2} = S^n$, the algorithm might not terminate using this criterion alone. One thus may also terminate the algorithm after a fixed number $\bar{n}$ of steps.

Proof (Lemma 2.23) As mentioned before, the proof very closely resembles the proof to Theorem 3.8 on the recovery properties of WIHT in [FR15].
The goal is to obtain an estimate of the form

$$\left\|x^{n+1} - x_S\right\|_2 \leq \rho \|x^n - x_S\|_2 + (1-\rho)\tau \left\|Ax_{\overline{S}} + e\right\|_2.$$

Noticing $\|\cdot\|_{\omega,2} = \|\cdot\|_2$, an application of Lemma 2.22 with $u := x^n + A^*(y - Ax^n)$,

$R := S^{n+1} = \operatorname{supp}(\tilde{H}_{3s}(u^n))$ and $S := S$ yields

$$\left\| \left(x^n + A^*(y - Ax^n)\right)_S \right\|_2^2 \leq \left\| \left(x^n + A^*(y - Ax^n)\right)_{S^{n+1}} \right\|_2^2. \tag{2.14}$$

Splitting left- and right-hand side as

$$\begin{aligned}
\left\| \left(x^n + A^*(y - Ax^n)\right)_S \right\|_2^2 &= \left\| \left(x^n + A^*(y - Ax^n)\right)_{S \setminus S^{n+1}} \right\|_2^2 + \left\| \left(x^n + A^*(y - Ax^n)\right)_{S \cap S^{n+1}} \right\|_2^2 \\
\left\| \left(x^n + A^*(y - Ax^n)\right)_{S^{n+1}} \right\|_2^2 &= \left\| \left(x^n + A^*(y - Ax^n)\right)_{S^{n+1} \setminus S} \right\|_2^2 + \left\| \left(x^n + A^*(y - Ax^n)\right)_{S \cap S^{n+1}} \right\|_2^2
\end{aligned}$$

substracting $\left\| \left(x^n + A^*(y - Ax^n)\right)_{S \cap S^{n+1}} \right\|_2^2$ from both sides of (2.14) and taking the square root then implies

$$\left\| \left(x^n + A^*(y - Ax^n)\right)_{S \setminus S^{n+1}} \right\|_2 \leq \left\| \left(x^n + A^*(y - Ax^n)\right)_{S^{n+1} \setminus S} \right\|_2.$$

Adding zeros to both sides one has

$$\left\| \left(x_S - x^{n+1} + x^n - x_S + A^*(y - Ax^n)\right)_{S \setminus S^{n+1}} \right\|_2 \leq \left\| \left(x^n - x_S + A^*(y - Ax^n)\right)_{S^{n+1} \setminus S} \right\|_2.$$

An application of the triangle inequality to the left-hand side yields

$$\begin{aligned}
&\left\| \left(x_S - x^{n+1}\right)_{S \setminus S^{n+1}} \right\|_2 - \left\| \left(x^n - x_S + A^*(y - Ax^n)\right)_{S \setminus S^{n+1}} \right\|_2 \\
\leq &\left\| \left(x_S - x^{n+1} + x^n - x_S + A^*(y - Ax^n)\right)_{S \setminus S^{n+1}} \right\|_2 \\
\leq &\left\| \left(x^n - x_S + A^*(y - Ax^n)\right)_{S^{n+1} \setminus S} \right\|_2
\end{aligned}$$

or rearranged

$$\begin{aligned}
\left\| \left(x_S - x^{n+1}\right)_{S \setminus S^{n+1}} \right\|_2 \leq &\left\| \left(x^n - x_S + A^*(y - Ax^n)\right)_{S^{n+1} \setminus S} \right\|_2 \\
&+ \left\| \left(x^n - x_S + A^*(y - Ax^n)\right)_{S \setminus S^{n+1}} \right\|_2.
\end{aligned}$$

Using the notation $S \Delta S^{n+1} := (S \setminus S^{n+1}) \cup (S^{n+1} \setminus S)$ one thus has

$$\left\| \left(x_S - x^{n+1}\right)_{S \setminus S^{n+1}} \right\|_2 \leq \sqrt{2} \left\| \left(x^n - x_S + A^*(y - Ax^n)\right)_{S \Delta S^{n+1}} \right\|_2. \tag{2.15}$$

Now by line 4 of the `WHTP`-algorithm $y - Ax^{n+1}$ is orthogonal to

$$\left\{ Az \in \mathbb{C}^n \mid \operatorname{supp}(z) \subseteq S^{n+1} \right\}$$

and thus characterized by

$$\left\langle y - Ax^{n+1}, Az \right\rangle = 0 \qquad \forall z, \operatorname{supp}(z) \subseteq S^{n+1}$$

or equivalently

$$\left(A^*(y - Ax^{n+1})\right)_{S^{n+1}} = 0.$$

From this and inequality (2.15), it follows with $e' := Ax_{\overline{S}} + e$

$$\begin{aligned}
&\left\|x^{n+1} - x_S\right\|_2^2 = \left\|\left(x^{n+1} - x_S\right)_{S^{n+1}}\right\|_2^2 + \left\|\left(x^{n+1} - x_S\right)_{S\setminus S^{n+1}}\right\|_2^2 \\
\leq &\left\|\left(x^{n+1} - x_S + A^*(y - Ax^{n+1})\right)_{S^{n+1}}\right\|_2^2 + 2\left\|\left(x^n - x_S + A^*(y - Ax^n)\right)_{S\Delta S^{n+1}}\right\|_2^2 \\
\leq &\left[\left\|\left((\mathrm{Id} - A^*A)(x^{n+1} - x_S)\right)_{S^{n+1}}\right\|_2 + \left\|\left(A^*e'\right)_{S^{n+1}}\right\|_2\right]^2 \\
&+ 2\left[\left\|\left((\mathrm{Id} - A^*A)(x^n - x_S)\right)_{S\Delta S^{n+1}}\right\|_2 + \left\|\left(A^*e'\right)_{S\Delta S^{n+1}}\right\|_2\right]^2.
\end{aligned}$$

Now by Lemma 2.19, Lemma 2.20 and the inequalities

$$\begin{aligned}
\omega(S^{n+1} \cup \operatorname{supp}(x^{n+1} - x_S)) &\leq \omega(S^{n+1}) + \omega(S) \leq 4s, \\
\omega(S\Delta S^{n+1} \cup \operatorname{supp}(x^n - x_S)) &\leq \omega(S^n) + \omega(S^{n+1}) + \omega(S) \leq 7s,
\end{aligned}$$

it holds

$$\begin{aligned}
\left\|x^{n+1} - x_S\right\|_2^2 \leq &\left[\delta_{\omega,4s}\left\|x^{n+1} - x_S\right\|_2 + \sqrt{1+\delta_{\omega,3s}}\left\|e'\right\|_2\right]^2 \\
&+ 2\left[\delta_{\omega,7s}\|x^n - x_S\|_2 - \sqrt{1+\delta_{\omega,4s}}\left\|e'\right\|_2\right]^2.
\end{aligned}$$

Setting

$$a := \left\|x^{n+1} - x_s\right\|_2, \quad b := \delta_{\omega,4s}, \quad c := \sqrt{1+\delta_{\omega,3s}}\left\|e'\right\|_2$$

and using the identity $a^2 - (ba + c)^2 = (1 - b^2)(a + c/(1 + b))(a - c/(1 - b))$ this is equivalent to

$$\begin{aligned}
&2\left[\delta_{\omega,7s}\|x^n - x_S\|_2 + \sqrt{1+\delta_{\omega,4s}}\left\|e'\right\|_2\right]^2 \geq \\
&(1 - \delta_{\omega,4s}^2)\left(\left\|x^{n+1} - x_S\right\|_2 + \frac{\sqrt{1+\delta_{\omega,3s}}}{1+\delta_{\omega,4s}}\left\|e'\right\|_2\right)\left(\left\|x^{n+1} - x_S\right\|_2 - \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,4s}}\left\|e'\right\|_2\right).
\end{aligned}$$

If it holds the assumption that

$$\left\|x^{n+1} - x_S\right\|_2 \geq \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,4s}}\left\|e'\right\|_2, \tag{2.16}$$

then using the inequality above one concludes

$$2\left[\delta_{\omega,7s}\|x^n - x_S\|_2 + \sqrt{1+\delta_{\omega,4s}}\|e'\|_2\right]^2 \geq (1-\delta_{\omega,4s}^2)\left(\left\|x^{n+1} - x_S\right\|_2 - \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,4s}}\|e'\|_2\right)^2$$

and thus after yet another rearrangement

$$\left\|x^{n+1} - x_S\right\|_2 \leq \frac{\sqrt{2}\delta_{\omega,7s}}{\sqrt{1-\delta_{\omega,4s}^2}}\left\|x^{n+1} - x_S\right\|_2 + \left(\frac{\sqrt{2}}{\sqrt{1-\delta_{\omega,4s}}} + \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,4s}}\right)\|e'\|_2 .$$

In case assumption (2.16) is not satisfied, this inequality still holds. Hence the claim holds with constants

$$\rho := \frac{\sqrt{2}\delta_{\omega,7s}}{\sqrt{1-\delta_{\omega,4s}^2}} \leq \frac{\sqrt{2}\delta_{\omega,7s}}{\sqrt{1-\delta_{\omega,7s}^2}} < 1 \text{ and } (1-\rho)\tau := \left(\frac{\sqrt{2}}{\sqrt{1-\delta_{\omega,4s}}} + \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,4s}}\right) \leq 5.15. \ \square$$

Weighted Compressive Sampling Matching Pursuit

Originally introduced in [NT10], Compressive Sampling Matching Pursuit (`CoSaMP`) is a greedy algorithm, improving on some of the shortcomings of the simpler Orthogonal Matching Pursuit (`OMP`), given as Algorithm 3 below. The fundamental idea however is the same and certainly easier illustrated by considering `OMP` first.

Algorithm 3 Orthogonal Matching Pursuit (`OMP`)

Input:
Normalized measurement matrix $A \in \mathbb{C}^{m\times N}$
Measurement vector $y \in \mathbb{C}^m$
Output: Approximation to $x^\#$

1: Initialize target support set $S^0 := \emptyset$, vector $x^0 := 0$
2: **while** Stopping criterion is not met **do**
3: $\quad j^{k+1} := \underset{j\in[N]}{\operatorname{argmax}}\left|(A^*(y - Ax^k))_j\right|$
4: $\quad S^{k+1} := S^k \cup \left\{j^{k+1}\right\}$
5: $\quad x^{k+1} := \operatorname{argmin}\left\{\|y - Av\|_2 \mid v \in \mathbb{C}^N, \operatorname{supp}(v) \subseteq S^{k+1}\right\}$

Starting with an empty support set S^0 and the zero vector x^0, `OMP` tries to incrementally approximate the solution $x^\#$ by choosing a new index to add to the support set in step 3 and finding the vector with support on the set S^k that matches the data the most via orthogonal projection in step 5. With a more precise illustration provided below, the effectiveness of `OMP` is then essentially established by the following observation (cf. [FR13, Proof to Lemma 3.3]):

$$\forall v \in \mathbb{C} : \min_{t\in C}\left\|y - A(v + te_j)\right\|_2^2 = \|y - Av\|_2^2 - \left|(A^*(y - Av))_j\right| .$$

Given some vector x^k supported on some set S^k, the largest decrease in the error $\left\|y - A(x^k + te_j)\right\|_2^2$ by adding an index j to S^k is thus achieved by selecting j as in step 3. The effort to reduce the local error $\left\|y - A(x^k + te_j)\right\|_2^2$ the most is also the reason why `OMP` is classified as a greedy algorithm. Selection of j in this manner albeit only guarantees

$$\|y - Aw\|_2^2 \leq \|y - Av\|_2^2 - \left|(A^*(y - Av))_j\right|^2$$

with $w := \operatorname{argmin}_{z \in \mathbb{C}^N} \left\{\|y - Az\|_2 \mid \operatorname{supp} z \subseteq S \cup \{j\}\right\}$. Thus given the solution $x^\#$ is s-sparse, `OMP` is not guaranteed to select the proper index set $S \supseteq \operatorname{supp} x^\#$ after, e.g. s steps, as shown in [FR13, Section 6.4]. This problem is addressed by `CoSaMP`. Assuming the solution $x^\#$ is s-sparse, instead of a single index, `CoSaMP` adds the indices of the $2s$ largest coefficients of $(A^*(y - Ax^k))$ to the potential index set S^k. `CoSaMP` then performs an orthogonal projection on this larger index set and thresholds the result to a s-sparse vector. Again replacing the hard thresholding operator by the quasi-best hard thresholding operator and s by $3s$ everywhere in the original version (cf. [FR13, p.164]), the weighted analogue to `CoSaMP` then reads as in Algorithm 4.

Algorithm 4 Weighted Compressive Sampling Matching Pursuit (`WCoSaMP`)

Input:
Normalized measurement matrix $A \in \mathbb{C}^{m \times N}$
Measurement vector $y \in \mathbb{C}^m$;
Weights $\omega \in \mathbb{R}^N_{\geq 1}$

Output: $3s$-sparse approximation to $x^\#$

1: Initialize target support set $S^0 := \emptyset$, vector $x^0 := 0$
2: **while** Stopping criterion is not met **do**
3: $\quad U^{k+1} := \operatorname{supp} x^k \cup \operatorname{supp} \tilde{H}_{4(3s)}(A^*(y - Ax^k))_\omega$
4: $\quad u^{k+1} := \operatorname{argmin} \left\{\|y - Av\|_2 \mid v \in \mathbb{C}^N, \operatorname{supp}(v) \subseteq U^{k+1}\right\}$
5: $\quad x^{k+1} := \tilde{H}_{3s}(u^{k+1})_\omega$

Analogously to [FR13, Theorem 6.27] one has in the weighted case the following recovery guarantees for `WCoSaMP`:

Lemma 2.24 (Reconstruction via `WCoSaMP`)
Suppose the restricted isometry constant $\delta_{\omega,16s}$ of $A \in \mathbb{C}^{m \times N}$ satisfies

$$\delta_{\omega,16s} \leq \frac{\sqrt{\sqrt{11/3} - 1}}{2} \approx 0.4782 \tag{2.17}$$

and $x \in \mathbb{C}^N$ is ω-s-sparse with $s \geq \|\omega\|_\infty^2$. Then for $e \in \mathbb{C}^m$, $y = Ax + e$ and $S \subseteq [N]$ with $\omega(S) \leq s$, the sequence $(x^n)_n$ defined by Algorithm 4 satisfies

$$\|x^n - x_S\|_2 \leq \rho^n \left\|x^0 - x_S\right\|_2 + \tau \left\|Ax_{\overline{S}} + e\right\|_2, \tag{2.18}$$

where the constants $\rho \in (0,1)$ and $\tau \in \mathbb{R}^+$ depend only on $\delta_{\omega,16s}$.

Lemma 2.24 again only guarantees convergence to some cluster point.
A closer looks at Algorithm 4 reveals that U^{k+1}, except for constant quantities such as y and ω, only depends on x^k. In the same manner u^{k+1} only depends on U^{k+1} and x^{k+1} on u^{k+1}. Thus U^{k+1} is determined by U^k alone and a natural stopping criterion for `WCoSaMP` is given by the condition $U^{k+1} = U^k$. As discussed for `WHTP` if, e.g. $U^{k+1} \neq U^k$, but $U^{k+2} = U^k$, the algorithm does not terminate using this criterion alone. One thus may add the condition of a maximal number $\bar{n}$ of iterations.

Proof (Lemma 2.24) The proof very closely follows the proof of Theorem 6.27 in [FR13]. Discarding the term $Ax_{\overline{S}} + e$ for the moment, it is first shown that the lines 3 to 5 of Algorithm 4 essentially lead to inequalites of the following forms

$$\left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2 \lesssim \|x^n - x_S\|_2 \tag{2.19}$$

$$\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 \lesssim \left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2 \tag{2.20}$$

$$\left\|x_S - x^{n+1}\right\|_2 \lesssim \left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2 + \left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2. \tag{2.21}$$

Combining the inequalities then gives an estimate of $\left\|x^{n+1} - x_S\right\|_2$ in terms of $\|x^n - x_S\|_2$. The inequalities are now established in reversed order.
Starting with line 5 of Algorithm 4, it holds $\left\|u^{n+1} - x^{n+1}\right\|_2 \leq \left\|u^{n+1} - x_{S\cap U^{n+1}}\right\|_2$ by Lemma 2.22. Also by lines 3 and 5 of Algorithm 4 it is $\operatorname{supp} x^{n+1} \subseteq U^{n+1}$. Therefore

$$\begin{aligned}\left\|(x_S - x^{n+1})_{U^{n+1}}\right\|_2 &= \left\|x_{S\cap U^{n+1}} - x^{n+1}\right\|_2 \leq \left\|u^{n+1} - x^{n+1}\right\|_2 + \left\|u^{n+1} - x_{S\cap U^{n+1}}\right\|_2 \\ &\leq 2 \cdot \left\|u^{n+1} - x_{S\cap U^{n+1}}\right\|_2 = 2 \cdot \left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2.\end{aligned}$$

Using $(x^{n+1})_{\overline{U^{n+1}}} = 0$ and $(u^{n+1})_{\overline{U^{n+1}}} = 0$ one gets the estimate of the desired form (2.21) via

$$\begin{aligned}\left\|x_S - x^{n+1}\right\|_2^2 &= \left\|(x_S - x^{n+1})_{\overline{U^{n+1}}}\right\|_2^2 + \left\|(x_S - x^{n+1})_{U^{n+1}}\right\|_2^2 \\ &\leq \left\|(x_S - x^{n+1})_{\overline{U^{n+1}}}\right\|_2^2 + 4 \cdot \left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2^2\end{aligned} \tag{2.22}$$

Now by line 4 holds for all $z \in \mathbb{C}^N$ with $\operatorname{supp}(z) \subseteq U^{n+1}$

$$\left\langle y - Au^{n+1}, Az \right\rangle = 0$$

or equivalently, since $\left\langle y - Au^{n+1}, Az \right\rangle = \left\langle A^*(y - Au^{n+1}), z \right\rangle$,

$$\left(A^*(y - Au^{n+1})\right)_{U^{n+1}} = 0.$$

By $y = Ax_S + e'$, where $e' := Ax_{\overline{S}} + e$, it immediately follows

$$(A^*A(x_S - u^{n+1}))_{U^{n+1}} = -(A^*e')_{U^{n+1}}$$

and thus via Lemma 2.19 by the inequality

$$\omega(\operatorname{supp}(x_S - u^{n+1}) \cup U^{n+1}) \leq \omega(S \cup U^{n+1}) \leq \omega(S) + \omega(U^{n+1}) \leq 16s,$$

it holds

$$\begin{aligned}\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 &\leq \left\|\left((\mathrm{Id} - A^*A)(x_S - u^{n+1})\right)_{U^{n+1}}\right\|_2 + \left\|(A^*e')_{U^{n+1}}\right\|_2 \\ &\leq \delta_{\omega,16s} \cdot \left\|x_S - u^{n+1}\right\|_2 + \left\|(A^*e')_{U^{n+1}}\right\|_2.\end{aligned}$$

Now it is assumed that $\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 > 1/(1-\delta_{\omega,16s})\left\|(A^*e')_{U^{n+1}}\right\|_2$ as otherwise (2.23) is immediate. Then it is

$$\begin{aligned}\left[\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 - \left\|(A^*e')_{U^{n+1}}\right\|_2\right]^2 \leq \delta^2_{\omega,16s}\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2^2 \\ + \delta^2_{\omega,16s}\left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2^2\end{aligned}$$

Using the identity $a^2 - b^2 = (a+b)(a-b)$, one has

$$\begin{aligned}\delta^2_{\omega,16s}\left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2^2 &\geq \left[\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 - \left\|(A^*e')_{U^{n+1}}\right\|_2\right]^2 \\ &\quad - \delta^2_{\omega,16s}\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2^2 \\ &= \left((1+\delta_{\omega,16s})\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 - \left\|(A^*e')_{U^{n+1}}\right\|_2\right) \\ &\quad \cdot \left((1-\delta_{\omega,16s})\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 - \left\|(A^*e')_{U^{n+1}}\right\|_2\right) \\ &= (1-\delta^2_{\omega,16s}) \\ &\quad \cdot \left(\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 - \frac{1}{1+\delta_{\omega,16s}}\left\|(A^*e')_{U^{n+1}}\right\|_2\right) \\ &\quad \cdot \left(\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 - \frac{1}{1-\delta_{\omega,16s}}\left\|(A^*e')_{U^{n+1}}\right\|_2\right).\end{aligned}$$

Since the middle term is greater than or equal to the bottom term, this follows

$$\frac{\delta^2_{\omega,16s}}{1-\delta^2_{\omega,16s}}\left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2^2 \geq \left(\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 - \frac{1}{1-\delta_{\omega,16s}}\left\|(A^*e')_{U^{n+1}}\right\|_2\right)^2.$$

Taking the square root and rearranging implies the inequality of form (2.20) by

$$\begin{aligned}\left\|(x_S - u^{n+1})_{U^{n+1}}\right\|_2 \leq & \frac{\delta_{\omega,16s}}{\sqrt{1-\delta_{\omega,16s}^2}}\left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2 \\ & + \frac{1}{1-\delta_{\omega,16s}}\left\|(A^*e')_{U^{n+1}}\right\|_2 .\end{aligned} \tag{2.23}$$

Defining $T^{n+1} := \operatorname{supp}(\tilde{H}_{4(3s)}(A^*(y - Ax^n))_\omega)$ and $S^n := \operatorname{supp} x^n$ it holds

$$\left\|(A^*(y - Ax^n))_{S\cup S^n}\right\|_2^2 \leq \left\|(A^*(y - Ax^n))_{T^{n+1}}\right\|_2^2$$

by Lemma 2.22. Substracting $\left\|(A^*(y - Ax^n))_{(S\cup S^n)\cap T^{n+1}}\right\|_2^2$ this follows

$$\left\|(A^*(y - Ax^n))_{(S\cup S^n)\setminus T^{n+1}}\right\|_2^2 \leq \left\|(A^*(y - Ax^n))_{T^{n+1}\setminus(S\cup S^n)}\right\|_2^2 .$$

Now, since $\operatorname{supp}(x_S - x^n) \subseteq S \cup S^n$, it holds for the right hand side

$$\left\|(A^*(y - Ax^n))_{T^{n+1}\setminus(S\cup S^n)}\right\|_2^2 = \left\|(x_S - x^n + A^*(y - Ax^n))_{T^{n+1}\setminus(S\cup S^n)}\right\|_2^2 .$$

Noting $(S \cup S^n) \setminus T^{n+1} \subseteq \overline{T^{n+1}}$, for the left hand side follows

$$\begin{aligned}\left\|(A^*(y - Ax^n))_{(S\cup S^n)\setminus T^{n+1}}\right\|_2^2 \geq & \left\|(x_S - x^n)_{\overline{T^{n+1}}}\right\|_2^2 \\ & - \left\|(x_S - x^n + A^*(y - Ax^n))_{T^{n+1}\setminus(S\cup S^n)}\right\|_2^2 .\end{aligned}$$

Using the notation $A\Delta B = (A \setminus B) \cup (B \setminus A)$, combining above estimates holds

$$\begin{aligned}\left\|(x_S - x^n)_{\overline{T^{n+1}}}\right\|_2^2 &\leq \left\|(A^*(y - Ax^n))_{T^{n+1}\setminus(S\cup S^n)}\right\|_2^2 \\ &\quad + \left\|(x^n - x_S + A^*(y - Ax^n))_{T^{n+1}\setminus(S\cup S^n)}\right\|_2^2 \\ &\leq \sqrt{2}\left\|(x^n - x_S + A^*(y - Ax^n))_{T^{n+1}\Delta(S\cup S^n)}\right\|_2^2 \\ &\leq \sqrt{2}\left\|\left((\mathrm{Id} - A^*A)(x^n - x_S)\right)_{T^{n+1}\Delta(S\cup S^n)}\right\|_2^2 + \left\|(A^*e')_{(S\cup S^n)\Delta T^{n+1}}\right\|_2^2 ,\end{aligned}$$

where $y = Ax_S + e'$. Again using $(x^{n+1})_{\overline{U^{n+1}}} = (u^{n+1})_{\overline{U^{n+1}}} = 0$ the left-hand side may be bounded via

$$\left\|(x_S - x^n)_{\overline{T^{n+1}}}\right\|_2 \geq \left\|(x_S - x^n)_{\overline{U^{n+1}}}\right\|_2 = \left\|(x_S)_{\overline{U^{n+1}}}\right\|_2 = \left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2 .$$

By Lemma 2.19 and since

$$\omega(T^{n+1}\Delta(S \cup S^n) \cup \operatorname{supp}(x^n - x_S)) \leq \omega(T^{n+1} \cup S \cup S^n) \leq 16s$$

it also holds an inequality of form (2.21), more precisely

$$\left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2 \leq \sqrt{2}\delta_{\omega,16s}\|x^n - x_S\|_2 + \sqrt{2}\left\|(A^*e')_{T^{n+1}\Delta(S\cup S^n)}\right\|_2 . \quad (2.24)$$

With (2.22)-(2.24) all ingredients for the proof have been collected. Now one first combines (2.22) and (2.23) to get the inequality

$$\begin{aligned}\left\|x_S - x^{n+1}\right\|_2^2 \leq& \left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2^2 \\ &+ 4\left(\frac{\delta_{\omega,16s}}{\sqrt{1-\delta_{\omega,16s}^2}}\left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2 + \frac{1}{1-\delta_{\omega,16s}}\left\|(A^*e')_{U^{n+1}}\right\|_2\right)^2 \\ &\leq \left(\sqrt{\frac{1+3\delta_{\omega,16s}^2}{1-\delta_{\omega,16s}^2}}\left\|(x_S - u^{n+1})_{\overline{U^{n+1}}}\right\|_2 + \frac{2}{1-\delta_{\omega,16s}}\left\|(A^*e')_{U^{n+1}}\right\|_2\right)^2\end{aligned}$$

where in the last line the inequality $a^2 + (b+c)^2 \leq (\sqrt{a^2+b^2}+c)^2$ has been used. This implies with inequality (2.24) the bound

$$\begin{aligned}\left\|x_S - x^{n+1}\right\|_2^2 \leq& \sqrt{\frac{2\delta_{\omega,16s}^2(1+3\delta_{\omega,16s}^2)}{1-\delta_{\omega,16s}^2}}\|x^n - x_S\|_2 \\ &+ \sqrt{\frac{2(1+3\delta_{\omega,16s}^2)}{1-\delta_{\omega,16s}^2}}\left\|(A^*e')_{(S\cup S^n)\Delta T^{n+1}}\right\|_2 + \frac{2}{1-\delta_{\omega,16s}}\left\|(A^*e')_{U^{n+1}}\right\|_2 .\end{aligned}$$

Inductively applying the previous inequality, the claim (2.18) follows via Lemma 2.20 with constants

$$\rho = \sqrt{\frac{2\delta_{\omega,16s}^2(1+3\delta_{\omega,16s}^2)}{1-\delta_{\omega,16s}^2}}\ ,\ (1-\rho)\tau = \sqrt{\frac{2(1+3\delta_{\omega,16s}^2)(1+\delta_{\omega,9s})}{1-\delta_{\omega,16s}^2}} + \frac{2\sqrt{1+\delta_{\omega,16s}}}{1-\delta_{\omega,16s}}$$

if $\rho \in (0,1)$. This is the case if $\delta_{\omega,16s}^2$ is smaller than the largest root of $6t^2+3t-1$, which in turn is the case if $\delta_{\omega,16s}^2 < (\sqrt{11/3}-1)/4$, i.e. if (2.17) holds. □

Reconstruction Guarantees in $\sigma_s(x)_{\omega,p}$

For the algorithms discussed above are also reconstruction guarantees in the form of (2.10), (2.11) in Theorem 2.17 available, as established by the theorem below.

Theorem 2.25 (Reconstruction Guarantees in $\sigma_s(x)_{\omega,p}$)
Given the ω-RIP condition (2.12) or (2.17) of order t, i.e. $\delta_{\omega,t} \leq c$ for any $A \in \mathbb{C}^{m\times N}$, such that $\delta_{\omega,2t} \leq c$ holds for the RIP-constant $\delta_{\omega,2t}$ of A, and $x \in \mathbb{C}^N$, $e \in \mathbb{C}^m$, executing the algorithm with an input sparsity of $2t$, $x^0 = 0$

and $y = Ax + e$ results in the sequence x^n satisfying

$$\|x - x^n\|_{\omega,p} \leq Ct^{\frac{1}{p}-1}\tilde{\sigma}_t(x)_{\omega,1} + Dt^{\frac{1}{p}-\frac{1}{2}}\|e\|_2 + 2\rho^n t^{\frac{1}{p}-\frac{1}{2}}\|x\|_2\,,$$

where $C, D > 0$ and $0 < \rho < 1$ depend only on $\delta_{\omega,2t}$. In particular, it holds

$$\begin{aligned}\|x - x^n\|_{\omega,1} &\leq C\tilde{\sigma}_t(x)_{\omega,1} + D\sqrt{t}\|e\|_2 + 2\sqrt{t}\rho^n\|x\|_2\,,\\ \|x - x^n\|_2 &\leq C/\sqrt{t}\cdot\tilde{\sigma}_t(x)_{\omega,1} + D\|e\|_2 + 2\rho^n\|x\|_2\,,\end{aligned}$$

and if $x^\# \in \mathbb{C}^N$ is a cluster point of the sequence

$$\begin{aligned}\left\|x - x^\#\right\|_{\omega,1} &\leq C\tilde{\sigma}_t(x)_{\omega,1} + D\sqrt{t}\|e\|_2\,,\\ \left\|x - x^\#\right\|_2 &\leq C/\sqrt{t}\cdot\tilde{\sigma}_s(x)_{\omega,1} + D\|e\|_2\,.\end{aligned}$$

The proof employs the following weighted analogue to [FR13, Lemma 6.23]:

Theorem 2.26 ([FR15, Theorem 3.9])
Let $A \in \mathbb{C}^{m\times N}$ be a matrix with ω-RIP constant $\delta_{\omega,s} < 1$ and $x, x' \in \mathbb{C}^N$, such that $\omega(x') \leq \kappa s$ for some $\kappa > 0$. Then if there exist $\tau > 0$, $\xi \geq 0$, such that

$$\left\|x_T - x'\right\|_2 \leq \tau\left\|Ax_{\overline{T}} + e\right\|_2 + \xi$$

with the support for the quasi-best ω-$2s$-approximation $T \subseteq [N]$ to x, it holds

$$\begin{aligned}\left\|x - x'\right\|_{\omega,p} \leq& s^{\frac{1}{p}-1}\left[2^{1-\frac{1}{p}} + 2\tau(1+\kappa)^{\frac{1}{p}-\frac{1}{2}}\sqrt{1+\delta_{\omega,s}}\right]\tilde{\sigma}_s(x)_{\omega,1}\\ &+ s^{\frac{1}{p}-\frac{1}{2}}(2+\kappa)^{\frac{1}{p}-\frac{1}{2}}\left(\tau\|e\|_2 + \xi\right).\end{aligned} \tag{2.25}$$

Proof (Theorem 2.25) By Lemma 2.23, Lemma 2.24 there exist $0 < \rho < 1$ and $\tau > 0$ depending only on $\delta_{\omega,2t}$, such that

$$\|x_T - x^n\|_2 \leq \tau\left\|Ax_{\overline{T}} + e\right\|_2 + \rho^n\|x_T\|_2$$

where T is the support of the quasi-best weighted $2s$-approximation to x. Then Theorem 2.26 with $x' = x^n$, $\xi = \rho^n\|x_T\|_2 \leq \rho^n\|x\|_2$, $\kappa = 2$ implies for any $1 \leq p \leq 2$ the bound

$$\begin{aligned}\|x - x^n\|_{\omega,p} &\leq t^{\frac{1}{p}-1}C\tilde{\sigma}_t(x)_{\omega,1} + t^{\frac{1}{p}-\frac{1}{2}}4^{\frac{1}{p}-\frac{1}{2}}\left(\tau\|e\|_2 + \rho^n\|x\|_2\right)\\ &= Ct^{\frac{1}{p}-1}\tilde{\sigma}_t(x)_{\omega,1} + Dt^{\frac{1}{p}-\frac{1}{2}}\|e\|_2 + 2\rho^n t^{\frac{1}{p}-\frac{1}{2}}\|x\|_2\end{aligned}$$

where C, D depend only on τ, which in turn only depends on $\delta_{\omega,2t}$. □

2.2.2 Chambolle and Pock's Preconditioned Primal-Dual Algorithm

The weighted ℓ_1-minimization problem presented in Theorem 2.17 is a convex program and one may directly apply any algorithm suited for this kind of problem. In case of unweighted ℓ_1-minimization there are however a number of algorithms specialized to the specific structure of the task as, e.g. presented in [FR13, Moe12, YMO11, DZ13]. One of these algorithms is the method of Chambolle and Pock (cf. [FR13, section 15.2], [CP11]) and its preconditioned version (cf. [PC11]), which actually apply to a more general class of optimization problems. In the following section the basic ingredients and properties of Chambolle and Pock's method are briefly presented. It is then shown how this method applies to the weighted basis pursuit problem (2.4).

General Algorithm

Chambolle and Pock's algorithm 5 is suited for any problem of the form

$$\min_{x\in\mathbb{R}^N} F(Ax) + G(x), \tag{2.26}$$

where $A \in \mathbb{R}^{m\times N}$ and $F : \mathbb{R}^m \to (-\infty, \infty]$, $G : \mathbb{R}^N \to (-\infty, \infty]$ are extended, real-valued, proper, convex functions. Central to the algorithm are the so called *proximal mappings*, defined as

$$\begin{aligned} P_G(T;\xi) &:= \operatorname*{argmin}_{z\in\mathbb{R}^N} \left(TG(z) + \frac{1}{2}\|z-\xi\|_2^2 \right), \\ P_{F^*}(\Sigma;x) &:= \operatorname*{argmin}_{z\in\mathbb{R}^m} \left(\Sigma F^*(z) + \frac{1}{2}\|z-x\|_2^2 \right), \end{aligned} \tag{2.27}$$

which regularize the functions F, G in the sense, that these are strictly convex and have several other beneficial properties in the context of optimization (cf. [FR13, pp. 553-555]).
The algorithm is indeed effective by the following theorem.

Theorem 2.27
Let $A \in \mathbb{R}^{m\times N}$ and $F : \mathbb{R}^m \to (-\infty, \infty]$, $G : \mathbb{R}^N \to (-\infty, \infty]$ be extended real-valued proper convex functions with $\operatorname{effdom}(F) = \mathbb{R}^m$ or $\operatorname{effdom}(G) = \mathbb{R}^N$, such that there exists $x \in \mathbb{R}^N$ with $Ax \in \operatorname{effdom}(F)$, attaining the optima

$$\min_{x\in\mathbb{R}^N} F(Ax) + G(x), \quad \max_{x\in\mathbb{R}^m} -F^*(\xi) - G^*(-A^*\xi).$$

Then for $\Theta = 1$ and positive definite symmetric matrices T, Σ satisfying

$$\left\| \Sigma^{\frac{1}{2}} A T^{\frac{1}{2}} \right\|_{\mathcal{L}(\mathbb{R}^N,\mathbb{R}^m)} < 1,$$

Algorithm 5 Preconditioned Primal-Dual Algorithm

Input:

Normalized measurement matrix $A \in \mathbb{R}^{m\times N}$; $\Theta \in [0,1]$;

T, Σ positive definite symmetric matrices satisfying $\left\|\Sigma^{\frac{1}{2}} A T^{\frac{1}{2}}\right\|_{\mathcal{L}(\mathbb{R}^N,\mathbb{R}^m)} < 1$

Output:

Approximation $\xi^{\#} = \xi^{\bar{n}}$ to dual problem;

Approximation $x^{\#} = x^{\bar{n}}$ to primal problem

1: Initialize $x^0 \in \mathbb{R}^N$, $\xi^0 \in \mathbb{R}^m$, $\bar{x}^0 := x^0$

2: **while** Stopping criterion is not met at $n = \bar{n}$ **do**

3: $\quad x^{n+1} := P_G(T; x^n - TA^*\xi^{n+1})$

4: $\quad \xi^{n+1} := P_{F^*}(\Sigma; \xi^n + \Sigma A(x^{n+1} + \Theta(x^{n+1} - x^n)))$

the sequence x^n defined by Algorithm 5 converges to an optimal solution of problem (2.26), i.e.

$$\lim_{n\to\infty} F(Ax^n) + G(x^n) = \min_{x\in\mathbb{R}^N} F(Ax) + G(x).$$

Matrices Σ and T as above always exist. Specific valid choices for Σ and T are

$$T := \operatorname{diag}(\tau_1, \dots, \tau_N), \qquad \Sigma := \operatorname{diag}(\sigma_1, \dots, \sigma_m),$$
$$\tau_j := \frac{1}{\sum_{i=1}^m |A_{ij}|}, \qquad \sigma_i := \frac{1}{\sum_{j=1}^N |A_{ij}|}.$$

Proof By [FR13, Theorem B.30] the problem (2.26) is equivalent to the saddle-point problem

$$\min_{x\in\mathbb{R}^N} \max_{\xi\in\mathbb{R}^m} \operatorname{Re}\langle Ax, \xi\rangle + G(x) - F^*(\xi).$$

Thus convergence of the sequence defined in Algorithm 5 is assured via [PC11, Theorem 1]. By [PC11, Lemma 2] T and Σ are indeed positive, symmetric maps satisfying

$$\left\|\Sigma^{\frac{1}{2}} A T^{\frac{1}{2}}\right\|_{\mathcal{L}(\mathbb{R}^N,\mathbb{R}^m)} < 1. \qquad \square$$

Weighted ℓ_1-Minimization

To cast the quadratically constrained weighted ℓ_1-minimization problem

$$\min \|z\|_{\omega,1} \ \text{ s.t } \|Az - y\|_2 \le \eta \tag{2.28}$$

in the form (2.26), one defines

$$F(z) := \chi_{B(y,\eta)}(z) := \begin{cases} 0 & \text{if } \|z-y\|_2 \leq \eta \\ \infty & \text{else} \end{cases} \quad \text{and} \quad G(z) := \|z\|_{\omega,1}\,.$$

By [FR13, example B.19(d)] the convex conjugate to F is then obtained as

$$F^*(\xi) = \sup_{\{z \,|\, \|z-y\|_2 \leq \eta\}} \operatorname{Re}\langle z, \xi\rangle = \operatorname{Re}\langle y, \xi\rangle + \eta\|\xi\|_2$$

and by [FR13, example B.19(c)] the convex conjugate to G as

$$G^*(\xi) = \begin{cases} 0 & \text{if } \left\|\xi\omega^{-1}\right\|_\infty \leq 1 \\ \infty & \text{else} \end{cases}.$$

Now F, G, F^* and G^* are all convex real-valued functions on a finite dimensional real vector space and thus obtain their minima. It is furthermore $\operatorname{effdom}(G) = \mathbb{R}^N$ and it also holds $Au \in \operatorname{effdom}(F)$ for $u := 0 \in \mathbb{R}^N$. Thus all conditions of Theorem 2.27 are satisfied and the sequence x^n given by Algorithm 5 converges to a solution of problem (2.28). The proximal mappings (2.27) evaluate to

$$P_{|\cdot|}(\tau; x) = \operatorname*{argmin}_{z\in\mathbb{R}} \left\{\frac{1}{2}|x-z|^2 + \tau|z|\right\} = \begin{cases} \operatorname{sgn}(x)(|x|-z) & \text{if } |z| \geq \tau \\ 0 & \text{else} \end{cases} =: S_\tau(x),$$

$$P_G(T; x_l) = S_{\tau_l\omega_l}(x_l), \quad l \in [N],$$

where $S_\cdot(\cdot)$ is also called *soft thresholding operator* and

$$P_{F^*}(\Sigma; \xi) = \begin{cases} 0 & \text{if } \left\|\Sigma^{-1}\xi - y\right\|_2 \leq \eta \\ \left(1 - \frac{\eta}{\left\|\Sigma^{-1}\xi - y\right\|_2}\right)(\xi - \Sigma y) & \text{else} \end{cases}.$$

Since above proximal mappings are given in closed form, Chambolle and Pock's algorithm is in this case easy to implement and solves the problem efficiently.
By simply setting $\eta = 0$ everywhere above, the method may also be employed for equality constrained weighted ℓ_1-minimization.

Stopping Criterions and Convergence Rate

A suitable stopping criterion is provided by the *primal dual gap*

$$E(x, \xi) := F(Ax) + G(x) - (-F^*(\xi) - G^*(-A^*\xi)) \geq 0$$

essentially measuring the distance between the primal and dual variables (cf. [FR13, p. 482]). By Theorem 2.27 it holds $E(x, \xi) = 0$ for a primal solution

x and dual solution ξ to problem (2.26). It has been shown that the gap $E(x^n, \xi^n)$ converges for x^n, ξ^n as in Algorithm 5 with rate $(O)(1/n)$ (cf. [FR13, Exercise 15.7], [CP11, Theorem 1], [PC11]). A stopping criterion for the general algorithm is hence given by the condition $E(x^n, \xi^n) \leq \eta$ for some $\eta > 0$. If only a specific amount of time for the execution of the algorithm is available, one may further set a maximal number $\bar{n}$ of iterations. For the specific case of quadratically constrained weighted ℓ_1-minimization, one could also simply terminate the algorithm as soon as the error $\left\|Ax^k - y\right\|_2$ falls below a certain threshold.

2.2.3 Comparison of the Algorithms

A major difference between algorithms directly attacking the quadratically constrained basis pursuit problem (2.4),

$$\min \|z\|_{\omega,1} \text{ s.t } \|Az - y\|_2 \leq \eta,$$

as the Chambolle-Pock algorithm, and approximation algorithms as given in Subsection 2.2.1, are the required input quantities.
As (2.4) depends on η, it has to be an input parameter to any algorithm directly solving this problem. This concretely manifests in the definition of F^* and P_{F^*} employed in Chambolle and Pock's algorithm. In the approximation algorithms, while the reconstruction guarantees depend on η, it is not used in the computations. However, using the quasi-best hard thresholding operator, here at least an estimate of the sparsity s of the vector to be reconstructed has to be passed. Chambolle and Pock's algorithm does not require a sparsity parameter.
Common to all algorithms is their simplicity. Chambolle and Pock's algorithm uses nothing but pointwise vector operations and multiplication with A and A^*. The approximation algorithms employ additionally the hard-thresholding operator, the application of which is however equivalent to a simple search and a pointwise vector multiplication. `WCoSaMP` and `WHTP` also compute a solution to a least-squares problem, which by the assumption on s is, after all, low dimensional. The most costly part of presented algorithms is thus the frequent multiplication with A and A^*. The efficiency of these algorithms hence centrally depend on the availability of fast transformations with these matrices.
Another difference between the algorithms is their execution speed. The approximation algorithms need only a few interations until a reasonable reconstruction is obtained. Chambolle and Pock's algorithm however only converges with rate $\mathcal{O}(1/n)$ in the number of steps. Albeit it should allow for a more precise solution compared to the approximation algorithms, in general a high number of steps is required until the precision of its solution actually surpasses that of the approximation algorithms.
Beside the methods introduced in this section, there are a larger number of interesting algorithms not discussed here, that should carry over to the weighted case likewise. Other greedy approximation algorithms are, e.g. Orthogonal Match-

ing Pursuit and Subspace Pursuit [DM09]. Further iterative thresholding algorithms are given by Renormalized Iterative Hard Thresholding (NIHT) [BD10] and Graded Iterative Hard Thresholding [BFH13]. As mentioned in the introduction to Subsection 2.2.2, one may also use any algorithm suited for convex programming to directly solve the weighted ℓ_1-minimization problem (2.4). Another accelerated ℓ_1-minimization method one should certainly mention here is the Homotopy Method [FR13, Section 15.1]. Furthermore, there are so called Split Bregman Methods [Moe12, YMO11, GO09] and methods based on Douglas-Rachford splitting [DZ13], that are closely related to Chambolle and Pock's method. Again common to all these algorithms is their demand for fast transformations with A and A^* for efficient execution.

Comparing furthermore any of these algorithms to *direct random least-squares projection* on some polynomial basis (cf. [CDL13, MNST14, MNvST13, Mig13]), the basis space has not to be chosen precisely beforehand. Although selection of the best basis functions seems like a very demanding combinatorial task, `WHTP`, for instance, employs nothing but a few additional matrix-vector, vector-vector operations and an application of the quasi-best hard thresholding operator $\tilde{H}_s(x)_\omega$, to extract the best s-sparse basis from a potentially much larger ambient space, thanks to the RIP-property of the resulting projection matrix A.

One may moreover wonder whether it is feasible to reduce the number of samples even further, by adaptively selecting the basis functions. By the findings of [AcCD11] however, no matter how sophisticated the adaption method may be, this at best cuts the number of required samples in half.

3 Chebyshev Expansion of Solutions to Parametric Operator Equations

Given a parameter dependent operator equation between Banach spaces X and Y

$$\mathcal{D} : X \times U \to Y, \qquad \mathcal{D}(u, z) = 0, \tag{3.1}$$

with parameter space $U := [-1, 1]^{\mathbb{N}}$, is it possible to expand the solutions u to $\mathcal{D}$ in a series of Chebyshev polynomials in the parameter? If X and Y were simply the complex plane $\mathbb{C}$ and U was just the one dimensional space $[-1, 1]$, then, given the derivative of $\mathcal{D}$ w.r.t. u does not vanish on U, u was a holomorphic function on U by the implicit function theorem and the expansion existed. An analogous result holds for arbitrary Banach spaces as well (cf. [KR97, Theorem 3.3.1], [Che13, Section 3.3, Theorem 3]). In [CCS14] it has recently been established, that under the so called *Holomorphy Assumption*, the expansion of the solutions to such problems exist and their coefficients lie in some ℓ_p space even if $U = [-1, 1]^{\mathbb{N}}$. Furthermore, the framework in [CCS14] includes problems of the form (1.2), where the parameter itself is expanded in a series

$$h : U \to L, \qquad h(z) := \sum_{j \geq 1} z_j \Psi_j(\cdot),$$

and L is again some Banach space. Problem (3.1) then takes the form

$$\mathcal{D}(u, z) = \mathcal{P}(u, h(z)) = \mathcal{P}(u, \sum_{j \geq 1} z_j \Psi_j(\cdot)) = 0 \text{ with } \mathcal{P} : X \times L \to Y. \tag{3.2}$$

In this chapter, the Holomorphy Assumption will be shortly reviewed and inter alia established for the special case of linear elliptic equations with Robin boundary equations. Using the results of [RS14] it is then shown that the coefficients to corresponding expansion even lie in a *weighted* ℓ_p space.

3.1 The Holomorphy Assumption

The goal of this and the next section is to determine sufficient conditions so that problem (3.2) above is well-defined for all $z \in U$ and a treatment via the Compressive Sensing method as presented in Chapter 4 is possible. This is essentially the case if the *Holomorphy Assumption* below is fulfilled.

Definition 3.1
Denote by $\mathcal{H}(A, B)$ the set of functions from A to B having a continuous Fréchet derivative.

Definition 3.2 (HA (p, δ) [CCS14, Definition 2.1])
Given $\delta > 0$ and $p \in (0, 1]$, an operator $\mathcal{D} : X \times U \to Y$ is said to fulfill the **HA** (p, δ)-assumption if

(i) $\forall z \in U \exists! u(z) \in X : \mathcal{D}(u, z) = 0$,

(ii) $\sup_{z \in U} \left\| u(z) \right\|_X \leq C_0$,

(iii) there exists $b \in \ell_p(\mathbb{N})$, such that $b > 0$, and $C_\delta > 0$ so that for all sequences ρ, such that $\rho > 1$ and $\sum_{j \geq 1} (\rho_j - 1) b_j \leq \delta$ holds $z \mapsto u(z) \in \mathcal{H}(\mathcal{O}_\rho, X)$ and $\sup_{z \in \mathcal{E}_\rho} \left\| u(z) \right\|_X \leq C_\delta$, where $\mathcal{O}_\rho := \bigotimes_{j \geq 1} \mathcal{O}_{\rho_j}$, $\mathcal{O}_{\rho_j}$ open, $\mathcal{E}_\rho := \bigotimes_{j \geq 1} \mathcal{E}_{\rho_j}$, $\mathcal{E}_{\rho_j} \subseteq \mathcal{O}_{\rho_j} \subseteq \mathbb{C}$, and

$$\mathcal{E}_\sigma := \left\{ \frac{w + w^{-1}}{2} \mid 1 \leq |w| \leq \sigma \right\} \subseteq \mathbb{C}$$

is the *Bernstein Ellipse*.

Remark 3.3
A sequence ρ as in (iii) always exists, since in particular $b \in \ell_1(\mathbb{N})$ and thus $\rho_j := 1 + \frac{\delta}{\|b\|_{\ell_1(\mathbb{N})}}$ is a valid choice.

Directly checking this assumption is certainly cumbersome in most cases. The following theorem reduces this task to a more tractable problem in many cases.

Theorem 3.4 (Class of $\mathcal{P}$ Fulfilling HA (p, δ) [CCS14, Theorem 4.3])
Assume

(i) $\exists p \in (0, 1] : \left(\left\| \Psi_j \right\|_L \right)_{j \geq 1} \in \ell_p(\mathbb{N})$

(ii) $\forall u \in X, h \in h(U) \exists! u \in X : \mathcal{P}(u, h) = 0$

(iii) $\mathcal{P}(\cdot, \cdot) \in \mathcal{H}(X \times L, W)$

(iv) $\forall h \in h(U) : \frac{\partial \mathcal{P}}{\partial u}(u(h), h) : X \to W$ is an isomorphism onto W

where $h(U) := \left\{ \sum_{j \geq 1} z_j \Psi_j(\cdot) \mid z \in U \right\}$.
Then there exists a $\delta > 0$, for which $\mathcal{P}$ satisfies **HA** (p, δ).

The theorem above only yields a qualitative result. Neither the constant δ itself nor an estimate are given, since the existence of $\delta > 0$ in the proof to Theorem 3.4 found in [CCS14] is provided by the implicit function theorem.
Next, Theorem 3.4 is simply applied to three examples, the first two of which will be considered in more depth later.

Example 3.5 (Elliptic Diffusion Equation [RS14, CCS14])
Consider the following elliptic partial differential equation with a spatially dependent *diffusion coefficient* $a(x, z)$, a source $f \in L_2(\Omega)$ and homogeneous Dirichlet boundary conditions on a sufficiently smooth region $\Omega \subseteq \mathbb{R}^n$

$$-\nabla \cdot (a(\cdot, z)\nabla u) - f = 0 \text{ in } \Omega, \quad u|_{\partial\Omega} = 0 \tag{3.3}$$

Assume the diffusion coefficient can be decomposed as

$$h(z) := a(\cdot, z) = \bar{a}(\cdot) + \sum_{j \geq 1} z_j \Psi_j(\cdot), \quad x \in \Omega, z \in U \tag{3.4}$$

with $\bar{a} \in L_\infty(\Omega)$ so that

$$b := \left(\left\| \Psi_j \right\|_{L_\infty(\Omega)} \right)_{j \in \mathbb{N}} \in \ell_p(\mathbb{N}), \tag{3.5}$$

$$R := \|\bar{a}\|_{L_\infty(\Omega)} + \|b\|_{\ell_1(\mathbb{N})}, \ \bar{a} - \|b\|_{\ell_1(\mathbb{N})} \geq r, \quad r, R \in \mathbb{R}^+. \tag{3.6}$$

Further define

$$A(h(z)) : H_0^1(\Omega) \to H^{-1}(\Omega), u \mapsto \left(v \mapsto \int_\Omega a(x, z) \nabla u \nabla v \, \mathrm{d}x \right),$$
$$l := \left(v \mapsto \int_\Omega f v \, \mathrm{d}x \right) \in H^{-1}(\Omega).$$

To fit this into the framework above, one finally defines the operator

$$\mathcal{P}(u, h) : H_0^1(\Omega) \times L_\infty(\Omega) \to H^{-1}(\Omega), (u, h) \mapsto -A(h)u - l.$$

Then (3.5) already establishes Theorem 3.4(i). Also $h(z) = a(\cdot, z)$ satisfies by (3.6) the *uniform ellipticity assumption*

$$0 < r \leq h(z) \leq R < \infty, \quad \forall z \in U, \tag{3.7}$$

which in turn via the Lax-Milgram implies that the problem (3.2) is well-posed and Theorem 3.4(ii) holds. Now $\mathcal{P}$ is well-defined and continuously differentiable from $H_0^1(\Omega) \times L_\infty(\Omega)$ to $H^{-1}(\Omega)$ so that Theorem 3.4(iv) holds. The operator $\mathcal{P}$ is furthermore by definition a linear operator in u. Thus

$$\frac{\partial \mathcal{P}}{\partial u}(u, h)(v) = \mathcal{P}(v, h)$$

and $\frac{\partial \mathcal{P}}{\partial u}(u(h(z)), h(z))$ is an isomorphism from $H_0^1(\Omega)$ onto $H^{-1}(\Omega)$ for every $z \in U$ by just established point (ii) of Theorem 3.4. Thus also Theorem 3.4(iii) is satisfied.

Now for a straightforward generalization of the previous example.

Example 3.6 (Linear Elliptic Equation with Robin Boundary Conditions) The linear elliptic equation with Robin boundary conditions is given by

$$\begin{aligned} -\nabla \cdot (a(x,z)\nabla u + l(x,z)u) + c(x,z)\nabla u + \lambda(x,z)u - f(x,z) = 0 \text{ in } \Omega \\ \alpha \frac{\partial u}{\partial n} + \beta u = g(x,z) \text{ on } \partial\Omega \end{aligned} \tag{3.8}$$

where $c(x,z)$ is called *convection coefficient*, $\lambda(x,z)$ *reaction coefficient*. Assume

$$\alpha \geq 0, \quad \beta > 0, \quad f, g \in H^{-1}(\Omega) \times L_\infty(U), \tag{3.9}$$

$$\exists R, r \in \mathbb{R} : R \geq a(x,z), \left(\lambda(x,z) - \frac{1}{2}\nabla(l(x,z) + c(x,z))\right) \geq r > 0, \tag{3.10}$$

$$\forall s \in \partial\Omega : (b(s,z) \cdot n) = 0 \tag{3.11}$$

and $\Omega \subset \mathbb{R}^n$ again sufficiently smooth. In case $\alpha = 0$, define $\alpha^{-1} = 0$ everywhere below.
Assume further that there exist expansions of a, b, c, λ, f, g in

$$\Psi^a, \Psi^b, \Psi^c, \Psi^\lambda, \Psi^f, \Psi^g \in \ell_p(\mathbb{N})$$

analogous to the expansion of the diffusion coefficient in (3.4). Zipping these into a single sequence

$$\Psi := (\Psi_1^a, \Psi_1^b, \Psi_1^c, \Psi_1^\lambda, \Psi_1^f, \Psi_1^g, \Psi_2^a, \Psi_2^b, \Psi_2^c, \Psi_2^\lambda, \Psi_2^f, \Psi_2^g, \dots),$$

it holds $\Psi \in \ell_p(\mathbb{N})$, i.e. Theorem 3.4(i).
Now cast problem (3.8) into its weak form. First, observe

$$\begin{aligned} &\int_\Omega \left[-\nabla \cdot (a(x,z)\nabla u + l(x,z)u) + c(x,z)\nabla u + \lambda(x,z)u\right] v \,\mathrm{d}x \\ &= \int_\Omega (a(x,z)\nabla u + l(x,z)u)\nabla v \,\mathrm{d}x + \int_\Omega (c(x,z)\nabla u + \lambda(x,z)u) v \,\mathrm{d}x \\ &\quad - \int_{\partial\Omega} \left[(a(s,z)\nabla u + b(s,z)u) \cdot n\right] v \,\mathrm{d}s \end{aligned}$$

and using the Robin condition, i.e. $\frac{\partial u}{\partial n} = \nabla u \cdot n = \alpha^{-1}(g - \beta u)$, one gets

$$\int_{\partial\Omega} \left[(a(s,z)\nabla u + b(s,z)u) \cdot n\right] v \,\mathrm{d}s = \int_{\partial\Omega} \left[a(s,z)\alpha^{-1}(g(s,z) - \beta u) + (b(s,z) \cdot n)u\right] v \,\mathrm{d}s.$$

Splitting this into right- and left-hand side, the weak formulation to above problem is hence obtained as

$$\begin{aligned} B(h,u,v) :=& \int_\Omega \left[\nabla \cdot (a(x,z)\nabla u + l(x,z)u) + c(x,z)\nabla u + \lambda(x,z)u\right] v \,\mathrm{d}x \\ &+ \int_{\partial\Omega} (a(s,z)\alpha^{-1}\beta - (b(s,z)\cdot n))uv \,\mathrm{d}s \\ =& \int_\Omega fv \,\mathrm{d}x + \int_{\partial\Omega} \alpha^{-1} a(s,z) g(s,z) v \,\mathrm{d}s := M(h,v) \end{aligned}$$

Lax-Milgram then again ensures the existence of the solution via (3.10), since

$$\begin{aligned} B(h,u,u) =& \int_\Omega a(x,z)|\nabla u|^2 + (l(x,z)+c(x,z))\cdot\nabla u u + \lambda(x,z)|u|^2 \,\mathrm{d}x \\ &+ \int_{\partial\Omega} (a(s,z)\alpha^{-1}\beta - (b(s,z)\cdot n))|u|^2 \,\mathrm{d}s \\ =& \int_\Omega a(x,z)|\nabla u|^2 + \left(\lambda(x,z) - \frac{1}{2}\nabla\cdot(l(x,z)+c(x,z))\right)|u|^2 \,\mathrm{d}x \\ &+ \int_{\partial\Omega} (a(s,z)\alpha^{-1}\beta - \underbrace{(b(s,z)\cdot n)}_{=0 \text{ by } 3.11})|u|^2 \,\mathrm{d}s \\ \geq& \int_\Omega a(x,z)|\nabla u|^2 + \left(\lambda(x,z) - \frac{1}{2}\nabla\cdot(l(x,z)+c(x,z))\right)|u|^2 \,\mathrm{d}x \\ \geq& C(h)\cdot\|u\|^2_{H^1(\Omega)} \end{aligned}$$

with some $C(h) > 0$ by the Poincare inequality.
Now define $h(z) := \bar{h}(\cdot) + \sum_{j\geq 1} z_j \Psi_j(\cdot)$, where z is split accordingly and $\bar{h}(\cdot) \in L_\infty(\Omega)$ is constant in $z \in U$ analogous to definition (3.4). Further define linear operators A and L via

$$\begin{aligned} A(h) : H^1(\Omega) \to H^{-1}(\Omega),\ u \mapsto B(h,u,\cdot), \\ L(h) := M(h,\cdot) \in H^{-1}(\Omega) \end{aligned}$$

and finally the operator $\mathcal{P}$ as required for the framework of Theorem 3.4 by

$$\mathcal{P}(u,h) : H^1(\Omega)\times L_\infty(\Omega) \to H^{-1}(\Omega),\ (u,h) \mapsto A(h)u - L(h).$$

As in the example above $\mathcal{P}$ is well-defined by Lax-Milgram and continuously differentiable from $H^1(\Omega)\times L_\infty(\Omega)$ to $H^{-1}(\Omega)$ so that Theorem 3.4(ii) and (iii) are satisfied. The derivative $\frac{\partial\mathcal{P}}{\partial u}(u(h(z)),h(z))(v)$ at v is again equal to $\mathcal{P}(v,h(z))$ by the linearity of $\mathcal{P}$ in u and thus an isomorphism from $H^1(\Omega)$ onto $H^{-1}(\Omega)$ for every $z \in U$ by the already established point (ii) of Theorem 3.4. So also Theorem 3.4(iv) holds.

The conditions (3.9), (3.10) and (3.11) in Example 3.6 have been chosen for simplicity of exposition. Improved conditions may be found in, e.g. [Lad68, Section3.6].
The framework of Theorem 3.4 also applies in case of time-dependent operator equations, as shortly illustrated in the next example.

Example 3.7 (Linear Parabolic Evolution Equation [CCS14])
Consider the linear parabolic evolution equation

$$\partial_t u - \nabla \cdot (a(\cdot, z)\nabla u) - f = 0 \text{ in } (0,T) \times \Omega \tag{3.12}$$

with spatially dependent diffusion coefficient $a(x,z)$, Gel'fand evolution triple $V \subseteq H \cong H' \subseteq V'$ and boundary conditions

$$u|_{\partial\Omega} = 0 \quad \forall t \in (0,T), \quad u|_{t=0} = u_0 \in H \quad \forall z \in U \tag{3.13}$$

With $V = H_0^1(\Omega)$ and $H = L_2(\Omega)$ a solution space is (cf. [CCS14])

$$Y := L_2(0,T; H_0^1(\Omega)) \times L_2(\Omega) \tag{3.14}$$

This is cast into the framework above via definition of $\mathcal{P}$ as

$$\mathcal{P}(u,h) : L_2(0,T;V) \cap H^1(0,T;V') \times L_\infty(\Omega) \to L_2(0,T;V') \times H, \tag{3.15}$$

$$(u,h) \mapsto \big(\partial_t u - \nabla \cdot (h\nabla u) - f, u(\cdot,0)\big) \tag{3.16}$$

Further examples may be found in [CCS14]. Those examples show that the framework is also compatible to, e.g. elliptic diffusion equations with certain non-affine diffusion coefficients, non-linear operator equations and operator equations with parameter-dependent domain.

3.2 Chebyshev Expansion of Solutions

The main aim of this section is to link the Holomorphy Assumption introduced above with the expansion of the solutions to $\mathcal{P}$ via Chebyshev polynomials on a parameter space of the form $U = [-1,1]^{\mathbb{N}}$. It also connects the results established here to the feasibility of the application of weighted ℓ_1-methods as presented in Chapter 2.

3.2.1 The Chebyshev Orthonormal Basis

Let us start with the definition of the Chebyshev polynomials in one dimension.

Definition 3.8 (1-Dimensional Chebyshev Basis)
The n-th *Chebyshev polynomial* is given by the mapping

$$T_n : [-1,1] \to \mathbb{R},\ t \mapsto \sqrt{2}\cos(n \arccos(t)),\ n \in \mathbb{N}, \quad T_0(t) \equiv 1$$

and their $L_\infty(U)$-norm by $\theta := \sqrt{2} = \|T_n\|_{L_\infty(U)}$ for $n \geq 1$. Further define the corresponding one-dimensional *Chebyshev probability measure* via

$$\mathrm{d}\Gamma(t) := \frac{1}{\pi\sqrt{1-t^2}}\mathrm{d}t.$$

Remark 3.9

The normalization factor $\sqrt{2}$ has been chosen, so that the Chebyshev polynomials are orthonormal with respect to the Chebyshev probability measure, i.e. satisfy

$$\int_{-1}^{1} T_j(z)T_k(z)\,\mathrm{d}\Gamma(z) = \delta_{jk}.$$

The measure $\mathrm{d}\Gamma$ is indeed a *probability measure*. Using the substitution $t := \cos(x)$ one has

$$\int_{-1}^{1} \frac{\mathrm{d}t}{\pi\sqrt{1-t^2}} = \int_{-\pi}^{0} \frac{-\sin(x)}{\pi\big|\sin(x)\big|}\,\mathrm{d}x = \int_{0}^{\pi} \frac{\sin(x)}{\pi\big|\sin(x)\big|}\,\mathrm{d}x = \frac{1}{\pi}\int_{0}^{\pi} \mathrm{d}x = 1.$$

These may be used to interpolate functions in one-dimensional space. For interpolation in N-dimensional space one can simply use products of the form

$$T_\nu(z) = \prod_{j=1}^{N} T_{\nu_j}(z_j)$$

where ν is a multi-index, basis. In infinite-dimensional space, i.e. if $U = [-1,1]^{\mathbb{N}}$, above product with $N = \infty$ however diverges, e.g. if $\nu \equiv 1$. One thus restricts the support sequences ν to sequences having only a finite number of non-zero elements. More precisely, one makes the following definitions:

Definition 3.10 (Chebyshev Polynomials for ∞-Dimensional Parameter Spaces)

Define the *space of positive integer sequences with finite support* as

$$\mathcal{F} := \big\{\nu = (\nu_1, \nu_2, \dots) \mid \nu_j \in \mathbb{N}_0 \wedge \operatorname{supp}\nu < \infty\big\}$$

and the corresponding tensorized Chebyshev polynomials via

$$T_\nu(z) := \prod_{j=1}^{\infty} T_{\nu_j}(z_j) = \prod_{j\in\operatorname{supp}\nu} T_{\nu_j}(z_j), \qquad z \in U.$$

Further define the infinite-dimensional *Chebyshev product measure* as

$$\mathrm{d}\Gamma := \bigotimes_{j\geq 1} \mathrm{d}\Gamma(z_j) = \bigotimes_{j\geq 1} \frac{\mathrm{d}z_j}{\pi\sqrt{1-z_j^2}}, \qquad z \in U.$$

Remark 3.11
These tensorized Chebyshev polynomials are again orthonormal with respect to the corresponding Chebyshev measure, i.e. satisfy

$$\int_U T_\nu(z)T_\mu(z)\,\mathrm{d}\Gamma(z) = \delta_{\nu\mu} \tag{3.17}$$

and form an orthonormal basis in $L_2(U, \mathrm{d}\Gamma)$.

3.2.2 Expansion of Solutions to Operator Equations

Employing Chebyshev polynomials, $u(z)$ may be formally expanded as

$$u(z) = \sum_{\nu\in\mathcal{F}} d_\nu T_\nu(z), \qquad d_\nu \in X. \tag{3.18}$$

It is however not known yet that this expansion is well-defined, i.e. lies in $\ell_1(\mathcal{F})$. A first step in this direction is taken by the following Theorem.

Theorem 3.12
If $\mathcal{D}$ fulfills some **HA** (p, δ), it holds with C_δ and ρ as in **HA** (p, δ)(iii)

$$\|d_\nu\|_X \leq \sqrt{2}C_\delta\rho^{-\nu} = \tilde{C}_\delta\rho^{-\nu}, \tag{3.19}$$

where $\rho^\nu := \prod_{k=1}^{\infty} \rho_k^{\nu_k}$, $\tilde{C}_\delta := \sqrt{2}C_\delta$.

Proof
Essentially the proof of [RS14, Proposition 4.1] applies with $(\delta\mu_0)^{-1}\|f\|_Y$ replaced by C_δ, if C_δ is chosen as the bound of G given by the holomorphic version of the implicit function theorem – although not explicitly – as in the proof to [CCS14, Theorem 4.3]. Since there are however slight differences in the setting and notation, it will be given here again in detail.
First, by orthonormality of the T_ν with respect to $\mathrm{d}\Gamma$, one has

$$d_\nu = \int_U u(z)T_\nu(z)\,\mathrm{d}\Gamma(z) \in X.$$

Setting $\nu = n \cdot e_1 = (n, 0, 0, \dots)$, $U =: [-1, 1] \times U'$ and $z =: (z_1, z')$, this implies

$$d_{ne_1} = \int_{z'\in U'} \int_{-1}^{1} T_n(t)u(t, z')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}}\,\mathrm{d}\Gamma(z').$$

Then substituting $t := \cos(\phi)$ and again $\xi := e^{i\phi}$, this follows

$$\begin{aligned}
&\int_{-1}^{1} T_n(t)u(t,z')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}} = \frac{\sqrt{2}}{\pi}\int_0^{\pi} u(\cos(\phi),z')\cos(n\phi)\,\mathrm{d}\phi \\
=&\frac{\sqrt{2}}{2\pi}\int_{-\pi}^{\pi} u(\cos(\phi),z')\cos(n\phi)\,\mathrm{d}\phi = \frac{\sqrt{2}}{2\pi i}\int_{|\xi|=1} u(\frac{\xi+\xi^{-1}}{2},z')\frac{\xi^n+\xi^{-n}}{2}\frac{\mathrm{d}\xi}{\xi} \\
=&\frac{\sqrt{2}}{4\pi i}\int_{|\xi|=1} u(\frac{\xi+\xi^{-1}}{2},z')\xi^{n-1}\,\mathrm{d}\xi + \frac{\sqrt{2}}{4\pi i}\int_{|\xi|=1} u(\frac{\xi+\xi^{-1}}{2},z')\xi^{-n-1}\,\mathrm{d}\xi.
\end{aligned}$$

With $S^1 := \left\{z \in \mathbb{C} \mid |z| = 1\right\}$ the map

$$J : \mathbb{C} \supseteq S^1 \to [-1,1],\ \xi \mapsto \frac{\xi+\xi^{-1}}{2}$$

is also known as the *Joukowsky map* and is a double cover of $[-1,1]$. More generally one has for $S_\sigma := \left\{z \in \mathbb{C} \mid |z| = \sigma\right\}$ the mapping

$$J_\sigma : \mathbb{C} \supseteq S_\sigma \to \mathcal{E}_\sigma,\ \xi \mapsto \frac{\xi+\xi^{-1}}{2},$$

where $\mathcal{E}_\sigma$ the Bernstein Ellipse as in Definition 3.2. Defining the annulus

$$A_\sigma := \left\{z \in \mathbb{C} \mid \sigma^{-1} \leq |z| \leq \sigma\right\},$$

it thus is $J_\sigma(A_\sigma) \subseteq \mathcal{E}_\sigma \subseteq \mathcal{E}_{\rho_1}$, since $\sigma < \rho_1$.
By **HA** (p,δ)(iii) the map $z_1 \mapsto u(z_1,z')$ is analytic on the set $\mathcal{O}_{\rho_1} \supseteq \mathcal{E}_{\rho_1}$. Hence

$$\xi \mapsto u(\frac{\xi+\xi^{-1}}{2},z')\xi^{n-1}, \quad \xi \mapsto u(\frac{\xi+\xi^{-1}}{2},z')\xi^{-n-1}$$

are analytic maps on A_{ρ_1} and by Cauchy's theorem for any $\sigma \in (1,\rho_1)$, it holds

$$\begin{aligned}
&\int_{-1}^{1} T_n(t)u(t,z')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}} \\
&= \frac{\sqrt{2}}{4\pi i}\int_{|\xi|=\sigma^{-1}} u(\frac{\xi+\xi^{-1}}{2},z')\xi^{n-1}\,\mathrm{d}\xi + \frac{\sqrt{2}}{4\pi i}\int_{|\xi|=\sigma} u(\frac{\xi+\xi^{-1}}{2},z')\xi^{-n-1}\,\mathrm{d}\xi.
\end{aligned}$$

Then employing $\sup_{z\in\mathcal{E}_\rho}\left\|u(z)\right\|_X \leq C_\delta$ by **HA** (p,δ)(iii) one has

$$\left\|\int_{-1}^{1} T_n(t)u(t,z')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}}\right\|_X \leq \frac{2\pi\sigma^{-1}}{\sqrt{2}\cdot 2\pi}C_\delta\sigma^{-n+1} + \frac{2\pi\sigma}{\sqrt{2}\cdot 2\pi}C_\delta\sigma^{-n-1} = \sqrt{2}C_\delta\sigma^{-n}.$$

Taking the limit $\sigma \to \rho_1$ and using the fact that Γ is a probability measure, i.e.

$$\int_{U'} C \, \mathrm{d}\Gamma(z') = C \text{ for any constant } C,$$

this yields

$$\|d_{ne_1}\|_X \leq \sqrt{2} C_\delta \rho_1^{-n}.$$

For general $\nu \in \mathcal{F}$ the argument may be repeated for every $j \in \mathrm{supp}(\nu)$, where, by definition, $\mathrm{supp}(\nu)$ is finite. □

By (3.19) the convergence properties of the coefficients d_ν in (3.18) are thus closely connected to sequences ρ as in **HA** (p, δ)(iii), i.e. $\rho > 1$ satisfying

$$\sum_{j \geq 1} (\rho_j - 1) b_j \leq \delta.$$

The convergence properties of d_ν are hence in turn apparently linked to the decay of the sequence b. If it is, e.g. known that

$$\sum_{j \geq 1} v_j^{\frac{2-p}{p}} b_j \leq \delta \tag{3.20}$$

for some sequence $v \geq 1$ and $p \in (0, 1]$, a valid sequence ρ as in **HA** (p, δ)(iii) may be defined via $\rho_j := v_j + 1$. Taking again a glance at (3.19),

$$\|d_\nu\|_X \leq \tilde{C}_\delta \rho^{-\nu},$$

the weight sequence v in (3.20) hence captures the convergence properties of $(\|d_\nu\|_X)_{\nu \in \mathcal{F}}$ in manner of Figure 3.1. This gives reason to associate weights of the following form to the tensorized Chebyshev polynomials.

Definition 3.13 (Tensorized Weights Associated to a Sequence v)
For $\theta \geq 1$ and some sequence $v \geq 1$ define $(\omega_\theta(v))_\nu := \theta^{\|\nu\|_0} v^\nu$ for $\nu \in \mathcal{F}$, where again $v^\nu := \prod_{k=1}^{\infty} v_k^{\nu_k}$.

The factor θ may appear puzzling at this point. It chiefly later provides a way to ensure the condition $\omega_\nu \geq \|\Phi_\nu\|_\infty$ in Theorem 2.17. Using the Chebyshev Polynomials, i.e. $\Phi_\nu = T_\nu$, one may set θ as defined in Definition 3.8. The condition $\omega_\nu \geq \|\Phi_\nu\|_\infty$, respectively the L_∞-properties of the Chebyshev basis functions, are also partly the reason why not, e.g. Legendre polynomials have been considered here. Details follow at the end of this section.
For ease of notation, make the following definition, before proceeding to the main result in this section.

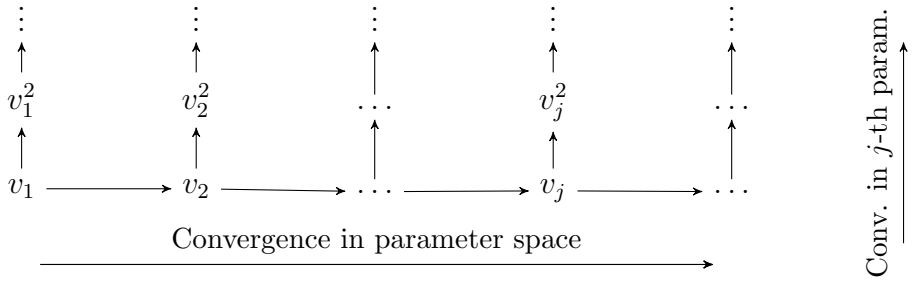


Figure 3.1: Relation of weights v to convergence of $(\|d_\nu\|_X)_{\nu\in\mathcal{F}}$.

Definition 3.14
Given v and p, define $\tilde{v} := v^{\frac{2-p}{p}}$.

Now finally for the main result of this section, establishing inter alia the well-definedness of (3.18).

Theorem 3.15
Given $\mathcal{D}$ satisfies some **HA** (p, δ), such that for a sequence $v \geq 1$ holds

$$\sum_{j\geq 1} b_j \tilde{v}_j < \delta, \qquad (3.21) \qquad\qquad b \in \ell_{v,p}(\mathbb{N}), \qquad (3.22)$$

with b as in **HA** (p, δ), one has $\left(\|d_\nu\|_X\right)_{\nu\in\mathcal{F}} \in \ell_{\omega_\theta(v),p}(\mathcal{F})$ for $\theta \geq 1$.

The proof is based on the following lemma.

Lemma 3.16 ($\ell_p(\mathcal{F})$ Summability of a Sequence [CDS10b, Theorem 7.2])
For $p \in (0, 1]$ and a sequence $(a_j)_{j\in\mathbb{N}}$ holds

$$\left(\frac{|\nu|!}{\nu!}a^\nu\right)_{\nu\in\mathcal{F}} \in \ell_p(\mathcal{F}) \quad \Longleftrightarrow \quad \sum_{j\geq 1} a_j < 1 \text{ and } (a_j)_{j\in\mathbb{N}} \in \ell_p(\mathbb{N}).$$

Proof (Theorem 3.15) The proof proceeds as in [RS14, Theorem 4.2].
The idea is to employ Theorem 3.12 via constructing a suitable sequence ρ, satisfying the conditions of **HA** (p, δ)(iii). Without loss of generality, assume $v > 1$. Else define v, e.g. by $v_j = 1 + \tau$ with $0 < \tau < \left(\frac{\delta}{\|b\|_{\ell_1(\mathbb{N})}}\right)^{\frac{p}{2-p}} - 1$.
For $\nu \in \mathcal{F}$ construct E, such that for $F := \mathbb{N} \setminus E$ and $0 < \varepsilon := \delta - \sum_{j\geq 1}(\tilde{v}_j - 1)b_j$

$$\sum_{j\in F} \tilde{v}_j b_j < \frac{\varepsilon}{e\tilde{\theta}}, \qquad (3.23)$$

where such E, F exist by (3.21). Further, define $\alpha > 1$, such that

$$(\alpha - 1)\sum_{j\in E} \tilde{v}_j b_j \leq \frac{\varepsilon}{2}.$$

Then define

$$\rho_j := \begin{cases} \alpha\tilde{v}_j & , j \in E \\ \max\left\{\tilde{v}_j, \frac{\varepsilon}{2|\nu_F| b_j}\right\} & , j \in F \end{cases}.$$

One has $\rho > 1$, since $\alpha, \tilde{v}_j > 1$, and ρ satisfies the conditions in **HA** (p,δ)(iii) by

$$\begin{aligned}\sum_{j\geq 1}(\rho_j - 1)b_j &= \sum_{j\in E}\alpha\tilde{v}_j b_j + \sum_{j\in F}\max\left\{\tilde{v}_j, \frac{\varepsilon}{2|\nu_F|\, b_j}\right\} b_j - \sum_{j\geq 1} b_j \\ &\leq (\alpha-1)\sum_{j\in E}\tilde{v}_j b_j + \sum_{j\in E}\tilde{v}_j b_j + \sum_{j\in F}\tilde{v}_j b_j + \frac{\varepsilon}{2} - \sum_{j\geq 1} b_j \\ &\leq \varepsilon + \sum_{j\geq 1}(\tilde{v}_j - 1)b_j = \delta.\end{aligned}$$

Therefore by Theorem 3.12, it holds

$$\|d_\nu\|_X \leq \tilde{C}_\delta \prod_{j\in E}(\alpha\tilde{v}_j)^{-\nu_j} \prod_{j\in F}\min\left\{\tilde{v}_j^{-\nu_j}, \left(\frac{|\nu_F|\, g_j}{\nu_j}\right)^{\nu_j}\right\}$$

with $g_j := \varepsilon^{-1} b_j$ and the convention that a factor equals 1 if $\nu_j = 0$. Defining $\mathcal{F}_E := \{\nu \in \mathcal{F} \mid \operatorname{supp}(\nu) \in E\}$ and $\mathcal{F}_F := \mathcal{F} \setminus \mathcal{F}_E$ then follows

$$\begin{aligned}\sum_{\nu\in\mathcal{F}} \omega_\nu^{2-p}\|d_\nu\|_X^p &= \sum_{\nu\in\mathcal{F}} \tilde{\omega}_\nu^p \|d_\nu\|_X^p \\ &\leq \tilde{C}_\delta^p \sum_{\nu\in\mathcal{F}} \tilde{\theta}^{p\|\nu\|_0} \left(\prod_{j\in E}\tilde{v}_j^{p\nu_j}(\alpha\tilde{v}_j)^{-p\nu_j}\right)\left(\prod_{j\in F}\tilde{v}_j^{p\nu_j}\min\left\{\tilde{v}_j^{-p\nu_j}, \left(\frac{|\nu_F|\, g_j}{\nu_j}\right)^{p\nu_j}\right\}\right) \\ &= \tilde{C}_\delta^p \left(\sum_{\nu\in\mathcal{F}_E}\tilde{\theta}^{p\|\nu\|_0}\alpha^{-p\nu}\right)\left(\sum_{\nu\in\mathcal{F}_F}\tilde{\theta}^{p\|\nu\|_0}\prod_{j\in F}\tilde{v}_j^{p\nu_j}\min\left\{\tilde{v}_j^{-p\nu_j}, \left(\frac{|\nu_F|\, g_j}{\nu_j}\right)^{p\nu_j}\right\}\right) \\ &\leq \tilde{C}_\delta^p \left(\sum_{\nu\in\mathcal{F}_E}\tilde{\theta}^{p\|\nu\|_0}\alpha^{-p\nu}\right)\left(\sum_{\nu\in\mathcal{F}_F}\tilde{\theta}^{p\|\nu\|_0}\prod_{j\in F}\left(\frac{|\nu_F|\,\tilde{v}_j^{p\nu_j} g_j}{\nu_j}\right)^{p\nu_j}\right).\end{aligned}$$

The first factor may be rewritten via

$$\sum_{\nu\in\mathcal{F}_E}\tilde{\theta}^{p\|\nu\|_0}\prod_{j\in E}\alpha^{-p\nu_j} = \prod_{j\in E}\left(1 + \tilde{\theta}^p\sum_{k=1}^{\infty}\alpha^{-pk}\right) = \left(1 + \frac{\tilde{\theta}^p\alpha^{-p}}{1-\alpha^{-p}}\right)^{|E|}$$

and thus is finite by $\alpha > 1$. For the second factor one employs Stirlings formula

$$\frac{n!e^n}{e\sqrt{n}} \leq n^n \leq \frac{n!e^n}{\sqrt{2\pi}\sqrt{n}}$$

and, after defining $h_j := e\tilde{\theta}\tilde{v}_j g_j$, obtains

$$\begin{aligned}\prod_{j\in F}\left(\frac{|\nu_F|\,\tilde{v}_j^{\nu_j} g_j}{\nu_j}\right)^{\nu_j} &= |\nu_F|^{|\nu_F|}\,\frac{1}{\nu_F^{\nu_F}}(\tilde{v}g)^{\nu_F} \leq \frac{|\nu_F|!e^{|\nu_F|}}{\sqrt{2\pi}\sqrt{|\nu_F|}}\,\frac{\prod_{j\in F}\max\left\{1, e\tilde{\theta}\sqrt{\nu_j}\right\}}{\nu_F!e^{|\nu_F|}}(\tilde{v}g)^{\nu_F}\\ &\leq \frac{|\nu_F|!}{\nu_F!}\prod_{j\in F}(\tilde{v}_j g_j)^{\nu_j}\max\left\{1, e\tilde{\theta}\sqrt{\nu_j}\right\} \leq \frac{|\nu_F|!}{\nu_F!}h^{\nu_F},\end{aligned}$$

where in the last line the crude estimate $\max\left\{1, \tilde{\theta}e\sqrt{\nu_j}\right\} \leq (\tilde{\theta}e)^{\nu_j}$ has been used. Now by (3.23) one obtains

$$\sum_{j\in F}\tilde{v}_j b_j < \frac{\varepsilon}{e\tilde{\theta}} \Leftrightarrow \sum_{j\in F} h_j = \sum_{j\in F} e\tilde{\theta}\tilde{v}_j\varepsilon^{-1}b_j < 1.$$

Condition (3.22) implies $h \in \ell_p(\mathbb{N})$ and by Lemma 3.16 thus $\left(\frac{|\nu_F|!}{\nu_F!}h^{\nu_F}\right)_\nu \in \ell_p(\mathcal{F}_F)$. Putting all together one has $\left(\|d_\nu\|_X\,\tilde{\omega}_\nu\right)_{\nu\in\mathcal{F}} \in \ell_p(\mathcal{F})$, which is equivalent to

$$\sum_{\nu\in\mathcal{F}}\|d_\nu\|_X^p\,\omega_\nu^{2-p} < \infty$$

or $\left(\|d_\nu\|_X\right)_{\nu\in\mathcal{F}} \in \ell_{\omega_\theta(v),p}(\mathcal{F})$. □

By **HA** (p,δ)(iii) one has $b \in \ell_p(\mathbb{N}) \subseteq \ell_1(\mathbb{N})$ and choosing $v \equiv 1$ the conditions (3.21), (3.22) are always satisfied. Setting $v \equiv 1$ and $\theta = 1$ in the theorem above then yields the well-definedness of the series in (3.18), as soon as $\mathcal{D}$ fulfills some Holomorphy Assumption **HA** (p,δ).
Applying a linear functional $G : X \to \mathbb{R}$ to the expansion then furthermore holds

$$F(z) := G(u(z)) = \sum_{\nu\in\mathcal{F}} g_\nu T_\nu(z), \qquad g_\nu := G(d_\nu) \in \mathbb{R},$$

so that also an expansion in Chebyshev polynomials of a linear functional is possible. From Theorem 3.15 one thus immediately obtains the following corollary.

Corollary 3.17
Under the same conditions as in Theorem 3.15 and with $G \in \mathcal{L}(X,\mathbb{R})$ the expansion coefficients $g = (g_\nu)_{\nu\in\mathcal{F}}$ of $F(z) = G(u(z))$ satisfy $g \in \ell_{\omega_\theta(v),p}(\mathcal{F})$.

Apart from that, Theorem 3.15 yields even more, namely a hint at the influence of the coefficients d_ν on the $\ell_{\omega_\theta(v),p}(\mathcal{F})$-norm of $u(z)$. Again peeking at Theorem 2.17

and, for example, taking $v \equiv 1$, θ as in Definition 3.8, it shows that Chebyshev basis functions T_ν non-constant in many variables contribute much more to the weighted sparsity of the expansion than those which are only non-constant in few variables. The most basis functions corresponding to indices in J_0^s as in Theorem 2.17 are thus non-constant in only a few variables.
As pointed out in the proof, as soon as $\|b\|_{\ell_1} < \delta$, it is always possible to define $v > 1$, such that (3.21) and (3.22) hold. The index set J_0^s is then finite for every s and the weights $\omega_\theta(v)$. Setting again θ as in Definition 3.8, all conditions of Theorem 2.17 are met and one may use weighted ℓ_1-minimization for reconstruction of some functional F. Combining these observations yields the following corollary.

Corollary 3.18 (Functional Reconstruction via Weighted ℓ_1-Minimization) Given $\mathcal{D}$ satisfying **HA** (p, δ) with a sequence b, such that $\|b\|_{\ell_1} < \delta$, there exists $v > 1$ and thus weights $\omega = \omega_\theta(v)$ for $\theta > 1$, such that all conditions of Theorem 2.17 are met and reconstruction via weighted ℓ_1-minimization is feasible.

3.2.3 Why Chebyshev Polynomials?

Ultimately the methods presented in the last chapter are suitable for reconstruction of some functional depending on a parametric operator equation in the Chebyshev basis. But why has one been considering a Chebyshev and not, e.g. a Legendre basis all the time? The main reason is the existence of the nice bound provided by Theorem 3.12. Defining the tensorized Legendre polynomials

$$L_\nu(z) := \prod_{j=1}^{\infty} L_{\nu_j}(z_j) = \prod_{j \in \operatorname{supp} \nu} L_{\nu_j}(z_j), \qquad z \in U,$$

an analogous bound is given by [CDS10a, Lemma 4.2] as

$$\left\| d_\nu^L \right\|_X \leq C_\delta \prod_{j \in \operatorname{supp} \nu} (2\nu_j + 1) \frac{\pi \rho_j}{2(\rho_j - 1)} \rho_j^{-\nu_j}, \tag{3.24}$$

where d_ν^L are now the coefficients in the Legendre expansion. Using this bound one may also prove a result alike to Theorem 3.15. Due to the growth of the L_∞-norm of the Legendre polynomials with their degree as $\|L_\nu\|_\infty = \prod_{j \in \operatorname{supp} \nu} \sqrt{2\nu_j + 1}$, this is bound to become more involved, without really giving any further insight. However, one may notice that for Theorem 2.17 to hold, the weights ω_ν have to grow also with each index ν_j and with them the number of samples. The number of required samples to obtain a certain set of expansion coefficients thus varies with employed orthonormal basis.
As suggested in [RS14, Remark 4.6], one might consider the preconditioned Leg-

endre polynomials

$$Q_j(z_j) := \sqrt{\frac{\pi}{2}}(1-z_j^2)^{\frac{1}{4}} L_j(z_j),$$

which again satisfy an uniform bound $\|Q_j\|_\infty \leq \sqrt{3}$, to improve on this situation. A minor drawback of these is that it does *not* hold $\|Q_0\|_\infty = 1$, but $\|Q_0\|_\infty = \sqrt{\frac{\pi}{2}}$, so that the infinite product (3.24) is not well defined with L replaced by Q. A possible workaround here is to truncate the parameter space to a finite dimension first. This is often introduced as the *finite dimensional noise assumption* (cf. [BTNT12, Assumption 2.2]). A greater drawback however is that the Q_j are no longer orthonormal with respect to the uniform probability measure. The corresponding expansion coefficients are therefore less suited to evaluate stochastic moments with respect to the uniform probability measure than those to L_j, as the identities discussed Section 4.2 do not apply any longer.

3.3 Operator Equations with Affine Dependence on a Finite Number of Parameters

The framework above clearly comes with at least two major shortcomings for practical purposes. First, it is not clear how to choose ρ such that the bound (3.19) in Theorem 3.12 is employed optimally or how to choose ρ at all. In the proof to Theorem 3.15 such a sequence is explicitly constructed given just weights v, capturing the convergence properties of b via (3.21), (3.22), and the L_∞-constant θ associated to the Chebyshev basis. But since the proof also depends on several crude estimates, this is probably not an optimal choice.
Second, also the constant C_δ in (3.19) is not known. Its existence has so far only been shown implicitly in a few examples by Theorem 3.4. Likewise the constants δ, p and a sequence b are not provided by Theorem 3.4.
Regarding the first problem, one may observe, that in case the sequence b as in some $\mathbf{HA}(p,\delta)$ is finite, i.e. it holds $b_j = 0$ for all $j > d$ and some $d \in \mathbb{N}$, and constant, i.e. $b_j = \beta$ for $j \in [d]$ and some $\beta > 0$, the bound on $\|d\|_{\ell^p_{\omega_\theta(v)}(\mathcal{F})}$ appears to be much more precise, since all the crude estimates are employed to derive a bound on the infinite set F. In this case one may choose $v \equiv w$ for some constant $w > 1$ so that

$$\sum_{j\geq 1} \tilde{v}_j b_j = \tilde{w}\beta d < \delta$$

and define ρ formally via

$$\rho_j := \begin{cases} \left(\alpha + \frac{1}{\tilde{v}_j}\right) \tilde{v}_j = \alpha\tilde{w} + 1 & \text{if } j \leq d \\ \infty & \text{else} \end{cases},$$

satisfying $\sum_{j\geq 1}(\rho_j - 1)b_j = \alpha\tilde{w}d\beta \leq \delta$ or equivalently $\alpha \leq \frac{\delta}{\tilde{w}\beta d}$. By the estimate provided in the proof to Theorem 3.15, it holds

$$\sum_{\nu\in\mathcal{F}} \omega_\nu^{2-p} \|d_\nu\|_X^p \leq \tilde{C}_\delta^p \sum_{\nu\in\mathcal{F}} \tilde{\theta}^{p|\nu\|_0} \left(\prod_{j\in E} \alpha^{-p\nu_j}\right) = \tilde{C}_\delta^p \left(1 + \frac{\tilde{\theta}^p\alpha^{-p}}{1-\alpha^{-p}}\right)^d$$

$$\Leftrightarrow \|d\|_{\ell_{\omega_\theta(v),p}(\mathcal{F})} \leq \tilde{C}_\delta \left(1 + \frac{\tilde{\theta}^p\alpha^{-p}}{1-\alpha^{-p}}\right)^{\frac{d}{p}} = \tilde{C}_\delta \left(1 + \frac{\theta^{2-p}}{\alpha^p - 1}\right)^{\frac{d}{p}}$$

and the optimal choice for α is given by $\alpha := \frac{\delta}{\tilde{w}\beta d} > 1$. At least in this case thus seemingly precise bounds are available.
The constant C_δ is nonetheless still missing. So are p, δ and b. These certainly depend on the concrete operator equation at hand. Luckily, Example 3.5 and Example 3.6 are operator equations with an affine parameter dependence, where an alternative framework used throughout [RS14] is available and provides upper bounds on these constants. It essentially holds the following Theorem.

Theorem 3.19 (HA (p, δ) for Affine Parametric Operator Equations)
Given linear operators A_j and the affine operator equation

$$\mathcal{D}(u, z) := A(z)u - f(z) = 0, \quad A(z) := A_0 + \sum_{j\geq 1} z_j A_j(z_j), \quad A(z) : X \to X',$$

where X' is the dual space to X, assume A_0 is boundedly invertible, i.e. there exists $\mu_0 > 0$ such that $\left\|A_0^{-1}\right\|_{\mathcal{L}(X',X)} \leq \mu_0$, and for b defined via $b_j := \left\|A_0^{-1}A_j\right\|_{\mathcal{L}(X,X)}$ holds $\|b\|_{\ell_p(\mathbb{N})} < 1$. Further assume the function f is holomorphic in every component on $U_\delta := [-1-\delta, 1+\delta]^{\mathbb{N}}$ and satisfies $\sup_{z\in U_\delta}\left\|f(z)\right\|_{X'} < \infty$ for some $\delta \in (0, 1 - \|b\|_{\ell_1(\mathbb{N})})$.
Then **HA** (p, δ) holds for $\mathcal{D}(u, z)$ with b, p, δ as above.

Proof First, establish **HA** (p, δ)(iii) for b. Given a sequence $\rho > 1$ such that

$$\sum_{j\geq 1}(\rho_j - 1)b_j \leq \delta, \tag{3.25}$$

define the polydisc $D_\rho := \prod_{j\in\mathbb{N}} D_{\rho_j}$ corresponding to ρ, where $D_r := \{z \in \mathbb{C} \mid |z| \leq r\}$ is the disk of radius r centered at the origin. Furthermore define $B_j(z_j) := A_0^{-1}A_j(z_j)$, take $z \in D_\rho$ and fix all components of z except for an arbitrary

$z_k \in \mathcal{D}_{\rho_k}$. Now the parametric solution

$$u(z) = (A(z))^{-1} f(z) = \left(\left(\mathrm{Id} + \sum_{j \neq k} z_j B_j(z_j) \right) + z_k B_k(z_k) \right)^{-1} A_0^{-1} f(z)$$

is holomorphic in z_k and well-defined, since (3.25) holds with $\delta \in (0, 1 - \|b\|_{\ell_1(\mathbb{N})})$ and thus $\left(\mathrm{Id} + \sum_{j \geq 1} z_j B_j(z_j)\right)$ is invertible for $z \in D_\rho$. Because z_k was chosen arbitrarily, the solution $u(z)$ is holomorphic in every $z_j \in D_{\rho_j}$. Moreover, it holds

$$\begin{aligned} \big\|u(z)\big\|_X &= \left\| \left(\mathrm{Id} + \sum_{j \geq 1} z_j B_j(z_j) \right)^{-1} A_0^{-1} f(z) \right\|_X \\ &\leq \left\| \left(\mathrm{Id} + \sum_{j \geq 1} z_j B_j(z_j) \right)^{-1} \right\|_{\mathcal{L}(X,X)} \big\|A_0^{-1} f(z)\big\|_X \\ &\leq \frac{1}{1-\delta} \frac{\big\|f(z)\big\|_{X'}}{\mu_0} \leq \frac{1}{1-\delta} \frac{\sup_{z \in U} \big\|f(z)\big\|_{X'}}{\mu_0} =: C_\delta \end{aligned} \tag{3.26}$$

and since also $\mathcal{E}_\rho \subseteq D_\rho$ hence **HA** (p, δ)(iii) is fulfilled with $\mathcal{O}_\rho := D_\rho$ and C_δ as above.
By Remark 3.3, a sequence ρ as above always exists and by $\rho > 1$ it is $D_\rho \supseteq U$. Consequently **HA** (p, δ)(i) holds by the existence of the solution $u(z)$ for $z \in D_\rho$. Finally, inequality (3.26) implies **HA** (p, δ)(ii) with $C_0 := C_\delta$. □

Remark 3.20

In case of affine parameter dependence, regardless of the concrete operator equation at hand, it holds $C_\delta \sim \frac{1}{1-\delta}$.

The restriction $\|b\|_{\ell_1} < 1$ is quite natural and indeed equivalent to the existence of a lower bound r in the uniform ellipticity condition (3.7) of Example 3.5, given, e.g. the basis functions $\Psi_j(\cdot)$ in (3.4) are constant.

4 The Compressive Sensing Petrov-Galerkin Method

In this chapter the Weighted Compressive Sensing methods introduced in Chapter 2 will be applied to the problem of recovering a linear functional $F(z) := G(u(z))$ depending on solutions $u(z)$ to a parametric operator equation $\mathcal{D}(u, z)$, as they have been introduced in Chapter 3.
In the first section, the Compressive Sensing Petrov Galerkin (CSPG) method is presented, that computes an approximation $F^\#$ to F such that $\left\|F - F^\#\right\| \leq \varepsilon$, where $\|\cdot\|$ is either $\|\cdot\|_2$ or $\|\cdot\|_{\omega,1}$, and its effectiveness proven. The CSPG method depends on a number of quantities, which are in practice certainly hard to estimate, such as a set J_0^s where $\left\|F - F_{J_0^s}\right\|_2 \leq \eta_1$ for some $\eta_1 > 0$. Nevertheless it gives an overview of the quantities to be considered, when applying weighted ℓ_1-minimization to such problems. Finally Corollary 4.2 provides asymptotic bounds of the error $\left\|F - F^\#\right\|$ in terms of the weighted convergence properties of F. These show especially that the approach may perform reasonably well in the important cases the Chebyshev expansion of F is known to depend only on a finite number of parameters or decays polynomially with the parameter dimension. Next, the use of CSPG in Uncertainty Quantification for the computation of stochastic moments is briefly demonstrated. In the final section a few notes on the practical implementation and performance of the algorithm are given.

4.1 The General Algorithm

The results established in Chapter 2 and Chapter 3 directly motivate Algorithm 6 below. Furthermore, these imply the following error bounds on the reconstructions.

Theorem 4.1 (Reconstruction Guarantees for a Linear Functional)
Using the notations of Corollary 3.17 and Theorem 2.17, the via Algorithm 6 reconstructed functional $F^\#$ satisfies

$$\begin{aligned} \left\|F - F^\#\right\|_\infty &\leq c_1 \sigma_s(F)_{\omega_\theta(v),1} + d_1 \sqrt{\frac{s}{m}} \eta, \\ \left\|F - F^\#\right\|_{L_2(U,\Gamma)} &\leq c_2 \frac{\sigma_s(F)_{\omega_\theta(v),1}}{\sqrt{s}} + d_2 \sqrt{\frac{1}{m}} \eta. \end{aligned} \tag{4.1}$$

If $\|g\|_{\omega,p} < \infty$ for some $p \in (0,1)$, the weighted Stechkin estimate Theorem 2.7 and Corollary 3.17 furthermore yield

$$\sigma_s(F)_{\omega(v),1} \leq s^{1-\frac{1}{p}} \cdot \theta^{\frac{2}{p}} \|g\|_{\ell_{\omega(v),p}(\mathcal{F})} \,. \tag{4.2}$$

Similarly to Corollary 2.18, one thus obtains from Theorem 4.1 the following bounds.

Corollary 4.2 (Asymptotic Reconstruction Guarantees for a Linear Functional) With the same notations as in Theorem 4.1, $p \in (0,1)$ and sampling errors $|\xi_j| \leq \tau$, it holds

$$\begin{aligned} \left\| F - F^{\#} \right\|_{\infty} &\leq C_1(g,p) \left(\frac{\log^3(m)\log(N)}{m} \right)^{\frac{1}{p}-\frac{1}{2}} + \frac{D_1}{\sqrt{\log(N)}} \left(\eta_1 + \sqrt{m}\tau \right), \\ \left\| F - F^{\#} \right\|_{L_2(U,\Gamma)} &\leq C_2(g,p) \left(\frac{\log^3(m)\log(N)}{m} \right)^{\frac{1}{p}-1} + D_2 \left(\frac{\eta_1}{\sqrt{m}} + \tau \right), \end{aligned} \tag{4.3}$$

where D_1, D_2 are universal constants and $C_1(g,p), C_2(g,p)$, except for other universal constants, only depend on p and $\|g\|_{\ell_{\omega(v),p}(\mathcal{F})}$.

As noted after Corollary 2.18, setting $\eta_1 = \left\| F - F_{J_0^s} \right\|_2$, η_1 decreases with increasing s. The error in the samples τ however does in general not decrease with the number of samples m. While its influence on the L_2-error stays constant with m, it grows in the L_∞-bound. Assuming τ is so small in all instances that it is neglectable anyway and looking at the asymptotic case of large s, one may consider the following cases.

Remark 4.3 ([RS14, Remark 5.5])

(a) **Constant Weights.** Given only a finite number d of entries in the sequence b in $\mathbf{HA}\,(p,\delta)$ are non-zero, one may define weights of the form

$$v_j = \begin{cases} w & , j \in [d] \\ \infty & , \text{else} \end{cases}$$

with some $w > 1$. For these the bound $m \geq Cs\log(s)^3\log(N)$ becomes

$$m \geq \begin{cases} C_w \cdot \log(d) \cdot s \cdot \log^4(s) & , s \leq (2w)^{2d} \\ C_w \cdot ds \cdot \log^3(s) \cdot \log(\log(s)) & , s > (2w)^{2d} \end{cases}$$

with some univeral constant C_w only depending on w and $\|g\|_{\ell_{\omega(v),p}(\mathcal{F})}$.

(b) **Polynomial Weights.** In case of polynomially growing weights $v_j = c \cdot j^\alpha$ for some $c > 1$, $\alpha > 0$, one requires with some universal constant $C_{\alpha,c}$ only

$$m \geq C_{\alpha,c} s \log^5(s)$$

samples of the functional, where $C_{\alpha,c}$ only depends on $\|g\|_{\ell_{\omega(v),p}(\mathcal{F})}$, α and c.

(c) **Exponential Weights.** Given weights of the form $v_j = c \cdot \alpha^j$, $c > 1$, by [RS14, Corollary 5.4(c)] it actually holds $N \in \mathcal{O}(s)$ and it is possible that $m > N$. Here weighted ℓ_1-minimization might yield inferior performance compared to, e.g. direct least squares projection.

Setting $\tau = 0$ in (4.3) one obtains from Remark 4.3 in the polynomial and constant case the easy to remember asymptotic error bounds

$$\left\|F - F^\#\right\|_{L_2(U,\Gamma)} \leq C \left(\frac{\log^r(m)}{m}\right)^{\frac{1}{p}-\frac{1}{2}}, \quad \left\|F - F^\#\right\|_{L_\infty(U,\Gamma)} \leq C \left(\frac{\log^r(m)}{m}\right)^{\frac{1}{p}-1} \tag{4.4}$$

with $r \leq 5$ and universal constant C. Neglecting logarithmic factors in these cases and given $p < \frac{1}{2}$, $d > 5$ the convergence rate in the number of samples m of above L_2-bound surpasses the best asymptotic bound known for quasi-Monte Carlo methods $\mathcal{O}(\frac{\log(m)^d}{m})$ (cf. [Nie92, Section 1.2]).
Let us now consider the parameters employed in Algorithm 6 in a little more detail. The algorithm mainly depends on five central quantities: the target accuracy ε, an accuracy η_2 and a sparsity s, such that the via weighted ℓ_1-minimization reconstructed functional $F^\#$ satisfies the estimate $\left\|F - F^\#\right\| \leq \varepsilon$ given samples of F with accuracy η_2, weights v reflecting the convergence properties of F in the Chebyshev basis and the samples of F within the accuracy η_2 themselves.
For a better overview, the quantities used in Algorithm 6 and their dependencies are again illustrated in Figure 4.1.
The parameter ε is given by the context in which Algorithm 6 is applied and thus is simply chosen by the user. Computing the samples within an accuracy η of F is a task completely orthogonal to finding the other quantities. This may be achieved using an adaptive Petrov-Galerkin method as discussed later in Section 4.3. The combination of a Petrov-Galerkin solver with Algorithm 6 then yields the *Compressive Sensing Petrov-Galerkin* method.
The constants η_1, η_2, the weights ω and the ω-sparsity s strongly depend on each other and it is not obvious how to get ahold of them. Since the most costly point is safely assumed to be the computation of the samples with accuracy η_2 and $m \geq c_0 s \log^3(s) \log(N)$ samples are required for the guarantees (4.1) to hold, where $N = \left|J_0^s(\omega)\right| = \left|\left\{\nu \in \mathcal{F} \mid \omega_\nu^2 \leq \frac{s}{2}\right\}\right|$, the best parameters are determined by the smallest s and largest ω, η_2, such that the conditions of Theorem 4.1 are satisfied. Given a function $\kappa(\eta_2, m)$ which measures the cost of obtaining the

Algorithm 6 Algorithm for the Approximation of a Functional $F(z) = G(u(z))$ Depending on a Parametric Operator Equation via Weighted ℓ_1-Minimization in the Chebyshev Basis

Input:

Desired accuracy ε;

Method to (approximately) solve a weighted ℓ_1-minimization problem;

Accuracies η_1, η_2, sparsity s, weights $v > 1$, so that $\left\|F - F_{\eta_2}\right\|_2 \leq \eta_2$ implies $\left\|F - F^{\#}\right\| \leq \varepsilon$ for employed (approximate) weighted ℓ_1-minimization algorithm;

Method to compute samples with accuracy η_2;

Output: Reconstructed functional $F(z) = G(u(z))$, such that $\left\|F - F^{\#}\right\| \leq \varepsilon$

1: Compute the finite index set $J_0^s = \left\{\nu \in \mathcal{F} \mid \theta^{2\cdot\|\nu\|_0} v^{2\cdot\nu} \leq \frac{s}{2}\right\}$, set $d := \dim(J_0^s) := \underset{j\in\mathbb{N}}{\operatorname{argmax}} \left\{\theta^2 v_j^2 \leq \frac{s}{2}\right\}$, $F_d(z) := F(z_1, z_2, \dots, z_d, 0, 0, \dots)$ and define U_d as the d-dimensional parameter space corresponding to J_0^s.

2: Compute $N := |J_0^s|$.

3: Compute $m := Cs\log^3(s)\log(N)$.

4: Choose m sample points $z_1, \dots, z_m \in U_d$ i.i.d with respect to the Chebyshev product measure $\mathrm{d}\Gamma$.

5: Compute m samples $y := (F_d^\eta(z_1), \dots, F_d^\eta(z_m))^T$ so that

$$\left|F_d(z_k) - F_d^\eta(z_k)\right| \leq \eta_2 \qquad \forall k \in [m].$$

6: Compute weights $\omega_\nu := \theta^{\|\nu\|_0} v^\nu$.

7: Apply the weighted ℓ_1-algorithm to samples y, weights $(\omega_\nu)_{\nu\in J_0^s}$ and operator

$$A_{l,\nu} = T_\nu(z_l),\ l \in [m],\ \nu \in J_0^s$$

to (approximately) solve the problem

$$\min_{z\in\mathbb{C}^N} \|z\|_{\omega,1} \text{ s.t. } \|Az - y\|_2 \leq \eta = \eta_1 + \eta_2$$

for (an approximation to) a coefficent vector g representing $F(z) = G(u(z))$ in the Chebyshev basis.

8: Assemble the solution $F^{\#}(z) = \sum_{\nu\in J_0^s} g_\nu T_\nu(z)$.

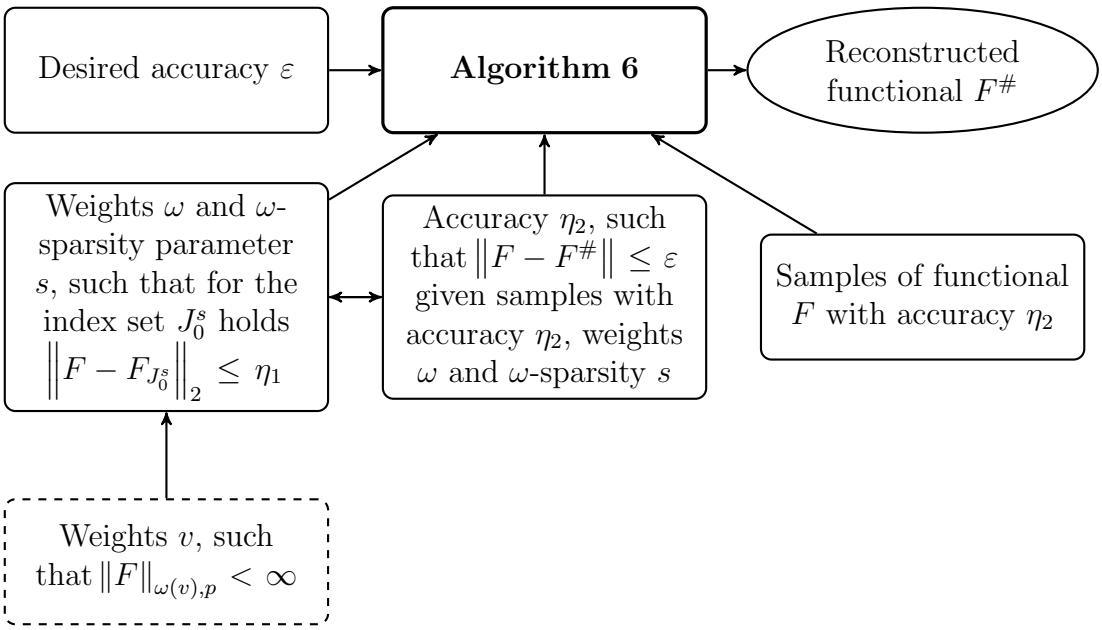


Figure 4.1: Illustration of input and output to Algorithm 6 and its dependencies. Dashed boxes correspond to quantities that are not strictly necessary for the reconstruction guarantees to hold, however are employed in Algorithm 6.

samples, one thus, e.g. for a given L_∞-bound ε, in theory needs to solve the minimization problem

$$\min_{\eta,m,s,\omega} \kappa(\eta_2, m) \quad \text{s.t .} \quad c_1\sigma_s(F)_{\omega,1} + d_1\sqrt{\frac{s}{m}}\eta \leq \varepsilon,$$

where the error bound in (4.1) has been used. In practice η_2 may just be fixed, for instance if a finite element method on a specific grid with specific elements is employed. Furthermore, assuming the cost of evaluating the operator equation is constant over all parameters z, instead of minimizing the cost function $\kappa(\eta_2, m)$, one only requires to minimize the number of samples m. By the bound given in Theorem 4.1 this is in turn equivalent to minimizing $c_0 s \log^3(s)\log(N)$.
Using the weights $\omega_\theta(v)$ the index set size N further only depends on s and a simpler one-dimensional sequence of weights v. It then remains the problem

$$\min_{s,v,\eta} s\log^3(s)\log(N(s,v)) \text{ s.t. } c_1\sigma_s(F)_{\omega_\theta(v),1} + d_1\left(c_0\log^3(s)\log(N(s,v))\right)^{-\frac{1}{2}}\eta \leq \varepsilon.$$

Again assuming the error η is very small, i.e. $\eta \approx 0$, this simplifies to

$$\min_{s,v} s\log^3(s)\log(N(s,v)) \text{ s.t. } c_1\sigma_s(F)_{\omega_\theta(v),1} \leq \varepsilon.$$

Now again applying the weighted Stechkin estimate Theorem 2.7 and Corollary 3.17, the bound given in Theorem 3.12 provides

$$\begin{aligned}\sigma_s(F)_{\omega_\theta(v),1} &\leq s^{1-\frac{1}{p}} \cdot \theta^{\frac{2}{p}} \|g\|_{\ell_{\omega_\theta(v),p}(\mathcal{F})} \leq s^{1-\frac{1}{p}} \cdot \theta^{\frac{2}{p}} \|G\|_{L_\infty(U)} \|d\|_{\ell_{\omega_\theta(v),p}\mathcal{F}} \\ &\leq s^{1-\frac{1}{p}} \cdot \theta^{\frac{2}{p}} \|G\|_{L_\infty(U)} \tilde{C}_\delta \cdot \left\|(\rho^{-\nu})_{\nu\in\mathcal{F}}\right\|_{\ell_{\omega_\theta(v),p}(\mathcal{F})}.\end{aligned}$$

It remains to estimate $\tilde{C}_\delta$ and $\left\|(\rho^{-\nu})_{\nu\in\mathcal{F}}\right\|_{\ell_{\omega_\theta(v),p}(\mathcal{F})}$. For $\left\|(\rho^{-\nu})_{\nu\in\mathcal{F}}\right\|_{\ell_{\omega_\theta(v),p}(\mathcal{F})}$ a general estimate is given by [CDS10b, Theorem 7.2], which is however quite crude and hence of little practical use. Also the constant $\tilde{C}_\delta$ depends strongly on the concrete operator equation that one is trying to approach. These issues have already been pointed out in Section 3.3.
But even in the presumably simple case of an affine operator equation depending only on a finite number d of parameters as in Section 3.3, this estimate appears to be of little practical use. Using the notations of Section 3.3 and noting that by Remark 3.20 it always holds $\tilde{C}_\delta = C/(1-\delta)$ with some fixed constant C, follows

$$\sigma_s(F)_{\omega_\theta(v),1} \leq s^{1-\frac{1}{p}} \cdot \theta^{\frac{2}{p}} \cdot \frac{C}{1-\tilde{w}\beta d\alpha} \cdot \left(1 + \frac{\theta^{2-p}}{\alpha^p - 1}\right)^{\frac{d}{p}},$$

where $w > 1$, $p \in (0,1)$ and $\alpha \in (1, \frac{1-\beta d}{\tilde{w}\beta d})$. Analogously bounding the L_2-error implies

$$\sigma_s(F)_{\omega_\theta(v),1} \leq s^{\frac{1}{2}-\frac{1}{p}} \cdot \theta^{\frac{2}{p}} \cdot \frac{C}{1-\tilde{w}\beta d\alpha} \cdot \left(1 + \frac{\theta^{2-p}}{\alpha^p - 1}\right)^{\frac{d}{p}}.$$

Given some desired accuracy ε and using above inequality, it has to hold

$$s \geq \left(\frac{\varepsilon}{\theta^{\frac{1}{p}} \cdot \frac{C}{1-\tilde{w}\beta d\alpha} \cdot \left(1 + \frac{\theta^{2-p}}{\alpha^p - 1}\right)^{\frac{d}{p}}}\right)^{\frac{1}{\frac{1}{2}-\frac{1}{p}}}.$$

Unsurprisingly, this inequality however does not yield practical results in general. Setting, for example,

$$\beta := \frac{1}{4.3},\ d := 4,\ \varepsilon := 10^{-3},\ C := 10,$$

using the heuristically near-optimal parameters $p := \frac{1}{7}$, $\alpha = \frac{1}{\tilde{w}}\left(\frac{1}{\beta d} - 1\right)$ and optimizing for δ still results in bounds larger than 10^4 for s given any admissible weight $w > 1$. As seen later in Chapter 5, this is certainly far too conservative in most cases.

4.2 Compressive Sensing Petrov-Galerkin and Uncertainty Quantification

While it might be possible to improve these a-priori estimates by considering more and more specific problems to get better error guarantees, in the context of Uncertainty Quantification however only certain statistical quantities depending on the functional, like the mean value or the variance of the functional values with respect to a certain probability distribution, are of interest. Here the representation of the reconstructed functional $F^\#$ in an orthonormal basis comes in very handy. By orthonormality of the Chebyshev polynomials (Remark 3.11) it holds for the mean value with respect to the Chebyshev measure Γ the identity

$$\mathbb{E}[F^\#] = \int_U F^\#(z)\,\mathrm{d}\Gamma(z) = \sum_{\nu\in J_0^s} g_\nu \int_{[-1,1]^{\mathbb{N}}} T_\nu(z)\,\mathrm{d}\Gamma(z) = g_0 \tag{4.5}$$

Again by orthonormality of the Chebyshev basis one obtains for the variance

$$\begin{aligned}
\mathbb{V}[F^\#] = \mathbb{E}[(F^\# - \mathbb{E}[F^\#])^2] &= \int_U (F^\#(z) - \mathbb{E}[F^\#])^2\,\mathrm{d}\Gamma(z) \\
&= \int_U \left(\sum_{\nu\in J_0^s\setminus\{0\}} T_\nu(z) g_\nu\right)^2 \mathrm{d}\Gamma(z) \\
&= \int_U \sum_{\nu\in J_0^s\setminus\{0\}} T_\nu(z)^2 g_\nu^2\,\mathrm{d}\Gamma(z) + \int_U \sum_{\nu\in J_0^s\setminus\{0\}} \sum_{\mu\in J_0^s\setminus\{0,\nu\}} T_\nu(z)T_\mu(z) g_\nu g_\mu\,\mathrm{d}\Gamma(z) \\
&= \sum_{\nu\in J_0^s\setminus\{0\}} g_\nu^2 \int_U T_\nu(z)^2\,\mathrm{d}\Gamma(z) + \sum_{\nu\in J_0^s\setminus\{0\}} \sum_{\mu\in J_0^s\setminus\{0,\nu\}} g_\nu g_\mu \int_U T_\nu(z)T_\mu(z)\,\mathrm{d}\Gamma(z) \\
&= \sum_{\nu\in J_0^s\setminus\{0\}} g_\nu^2.
\end{aligned} \tag{4.6}$$

Expectation and variance may be computed for any other probability distribution in the same manner by simply replacing T_ν with the corresponding associated orthonormal basis given by the so called Wiener-Askey Scheme (cf. [XK02b, Xiu10, EMSU12]). Statistical quantities depending on higher moments, like the skewness of $F^\#$, may be obtained in the same manner. Although one might not be able to estimate the error compared to $\mathbb{E}[F]$, $\mathbb{V}[F]$ etc. in these quantities precisely a-priori, the convergence rates of (4.4) still transfer. Hence one may pragmatically invoke Algorithm 6 for increasing values of s, until the variation of the respective estimates via $F^\#$ drops below a certain predetermined bound. As noted in Subsection 3.2.3, the number of required samples to reach a certain threshold in the error seems thereby to depend on the probability distribution.

4.3 Notes on the Implementation

In this section, Algorithm 6 is analysed line by line and the complexity of the required computations briefly discussed.
Computing the finite index set J_0^s is equivalent to enumerating the points in several d-dimensional simplices centered at the origin. More precisely, one may decompose J_0^s as in the proof to [RS14, Theorem 5.3] with $A := \log_2 s/2$, $T := 2\log_2(\theta)$, $a := 2\log_2 v_j$ and $d := \max\left\{j \in \mathbb{N} \mid a_j \leq A - T\right\}$ as

$$J_0^s = \{0\} \cup \bigcup_{k=1}^{d} \bigcup_{\substack{S \subseteq [d] \\ |S|=k}} \iota_d(a, A - Tk) \text{ and } \iota_d(a, L) := \left\{ \nu \in \mathbb{N}_+^d \mid \sum_{j=1}^{d} a_j \nu_j \leq L \right\}.$$

The union over k translates to a simple `for`-loop. Enumerating over all $S \subset [d]$ with size k is less straightforward, but is accomplishable efficiently. The elements of $\iota_d(a, L)$ correspond to the positive integer-valued vectors in a d-dimensional simplex at the origin, which are again known to be computable with linear complexity in the number of points in this simplex. Furthermore, cleverly summing over S, no empty $\iota_d(a, L)$ have to be evaluated and overall thus enumeration of J_0^s is in $\mathcal{O}(\#J_0^s|)$.
As a byproduct of the previous step, N in line 2 is also computable in linear time. Computing the number of samples m in line 3 given s and N is now a trivial task. The m sample points i.i.d. with respect to the Chebyshev product measure $\mathrm{d}\Gamma$ in U_d as in line 4 may be obtained via drawing m i.i.d. uniform samples $\tilde{z}_1, \ldots, \tilde{z}_m$ and the transformation $z_j := \cos(\pi \tilde{z}_j)$.
Computation of the samples $F(z_l)$ with some guaranteed accuracy η_2 as in line 5 is possible as mentioned before via, e.g. an adaptive Petrov-Galerkin method. However the dimension d of the parameters z_l might vary with the other input parameters. In this case one has to ensure the employed method supports parameters with any given dimension d. Evaluation of the samples furthermore may be sped up using, for instance, some reduced basis computed beforehand. It might also be accelerated using methods similar to those employed in solving time-dependent problems, treating the parameter z similar to the time t there.
Computing the weights in line 6 is again completely straightforward.
Now in line 7 one may apply any of the methods discussed in Section 2.2. As pointed out in Subsection 2.2.3, all of these are very simple and rely, besides the application of the transformations A and A^*, at most on pointwise vector operations, solutions to low-dimensional least-squares problems and the quasi-best hard thresholding operator $\tilde{H}$, all of which are efficiently computable. Multiplication with A and A^* is also efficiently possible via standard matrix multiplication given A and A^* in explicit form as long as m and N are not too large. If N is quite large however, explicit multiplication with A and A^* might become too costly. In the one-dimensional case, one may employ a Nonequidistant Fast

Fourier Transformation [PST98] to speed up the multiplications. Such algorithms are also available for the multidimensional case [Pot03]. Unfortunately, these scale poorly in the dimension d, which in the setting of high-dimensional operator equations is certainly a drawback. Fast multiplication with sampling matrices originating from truly random parameters z is in general presumably not possible. It might however be possible to impose additional structure on the sample points, allowing for a fast transformation. Better scaling in the dimension d could be achieved by, for example, choosing the sampling points along one-dimensional lattice $\Lambda(r, M) := \{r \cdot j \mid j \in [M]\}$ with $r \in \mathbb{R}^d \setminus \mathbb{Q}^d$ (cf. [KKP12]). But even whether the RIP-property holds for sampling matrices coming from such pseudo-random sequences is not clear and subject to further research.
In step 8 one finally has the coefficients g_ν at hand and with them an approximation $F^\#$ to given functional F. If not only certain statistical properties of $F^\#$ as in Section 4.2 are of interest, one probably wants to evaluate the reconstructed functional for certain parameters $(q_1, \dots, q_r) =: q$ at some point. Creating a matrix $A(q)$ from q just like the sampling matrix A in step 8 from z, this can be done by a matrix-vector multiplication of $A(q)$ with the normalized coefficients $g/\sqrt{m}$. Restricting g to non-zero entries only and adjusting $A(q)$ accordingly, the number of columns in $A(q)$ is significantly lower than the number of columns in A in step 8. Explicit multiplication with $A(q)$ is thus still an option, even if q is quiet large. For very large q one again faces problems similar to the step before.
In conclusion the algorithm is in all cases very easily stacked on top of legacy code providing solutions to some operator equation. Its performance mainly depends on efficient transformations with the sample matrices A and A^*. In case the parameter space dimension d is not too large and the size of A is moderate, Algorithm 6 may be efficiently executed using explicit matrix-vector multiplication.

5 Numerical Results

In this chapter numerical tests of the Compressive Sensing Petrov-Galerkin method as introduced in Chapter 4 are presented and discussed briefly. None of the examples below have been derived from a specifically stochastic problem. Especially the results to the examples considered in Section 5.2 and Section 5.3 are nevertheless hopefully much easier to interpret and evaluate than most problems having a particular stochastic background.
As has been pointed out in Section 4.1, finding a combination of parameters η_1, η_2, s and v, that guarantee the error of the reconstruction $F^\#$ to be bounded by a desired accuracy ε, is hard. Instead, the algorithm is simply executed for different values of η_2, s and v. The quality of the reconstructed solutions is then evaluated in three different manners via comparison to numerical values again obtained by an adaptive Petrov-Galerkin method also using the respective error bound η_2. The function corresponding to these numerical values is in the following called $\hat{F}$.
First, the reconstructions $F^\#$ for different s and weights v are compared to $\hat{F}$ over $N = 10^4$ randomly chosen parameter values via

$$H_{L_2}(z) := \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left|\hat{F}(z_j) - F^\#(z_j)\right|^2} \quad \text{and} \quad H_{L_\infty}(z) := \max_{j=1,\dots,N}\left|\hat{F}(z_j) - F^\#(z_j)\right|.$$

Next, the solutions $F^\#$ are compared pointwise to $\hat{F}$ for different sparsity values s in a number of different parameter (sub-)spaces.
Furthermore, the mean $\mathbb{E}[F]$ of F with respect to the Chebyshev probability measure of the solutions in the parameter space is approximated using the Monte Carlo method and the reconstructed functional. Here the following two observations are employed: For the expansion of $F^\#$ in the polynomials, it holds the identity (4.5),

$$\mathbb{E}[F^\#] = g_0.$$

Further, given m i.i.d. samples $\hat{F}(z_1), \dots, \hat{F}(z_m)$ of some functional $\hat{F}$, it holds by [BSZ11, Lemma 4.1] for the expectation value with respect to some probability measure η that

$$\left\|\mathbb{E}[\hat{F}] - E_m[\hat{F}]\right\|_{L_2(U;\eta)} \leq m^{-\frac{1}{2}}\left\|\hat{F}\right\|_{L_2(U;\eta)},$$

where $E_m[\hat{F}] := \frac{1}{m}\sum_{k=1}^{m}\hat{F}(z_k)$ is the empirical mean of the samples. The mean thus converges with rate $\mathcal{O}(\sqrt{m})$ in the number of samples. Using so called

low-discrepancy sequences instead of i.i.d. samples, the rate can be improved to $\mathcal{O}(\frac{\log(N)^d}{N})$, where d is again the parameter space dimension. The computation of $E_m[\hat{F}]$ over these sampling sequences is known as quasi-Monte Carlo method (cf. [Nie92, section 1.2]). The empirical means $E_m[\hat{F}]$ are in the following computed using the Monte Carlo method over the same samples that were employed to obtain the reconstruction $F^\#$ as well as the quasi-Monte Carlo method taking the respective *same number* of samples along the Halton sequence, and then compared to $\mathbb{E}[F^\#] = g_0$.
The parameter η_1 required to formulate the ℓ_1-minimization problem in step 7 of Algorithm 6 is unknown. Using an approximation algorithm, such as `WHTP` and `WCoSaMP` introduced in Section 2.2, knowledge of η_1 is however not required. Hence presented solutions to all problems have been obtained via `WHTP`.
Also the actual number of required samples m given some sparsity s and ambient dimension N is by Theorem 4.1 not known exactly, but only up to a constant factor C and possible logarithmic factors in s. Thus the optimistic value

$$m = 2s\log(N)$$

has been employed in view of Remark 2.16. `WHTP` performed well using these parameters in all upcoming examples.
All numerical tests have been implemented using the `Python` programming language. The numerical solutions to all problems have been obtained using the `AdaptiveLinearVariationalSolver` of the `FEniCS` numerics environment (cf. [LMW+12]), guaranteeing a respective error bound η_2 in the samples. The reconstructed functionals have been computed using the `NumPy` and `SciPy` packages (cf. [JOP+]). The elements of the Halton sequence have been computed using the `ghalton` package.

5.1 Diffusion Equation with Trigonometric Coefficient

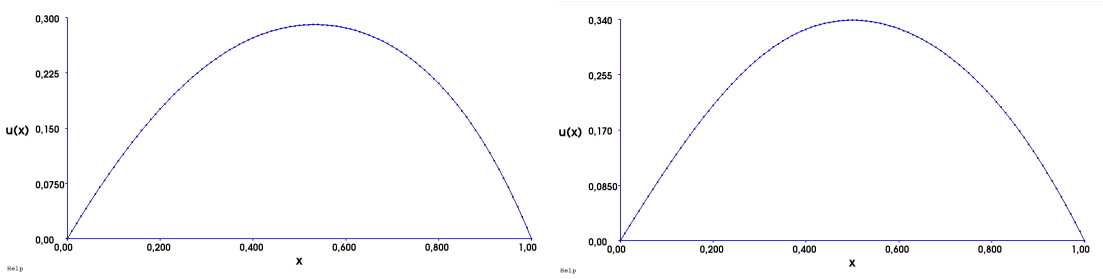


Figure 5.1: Two example solutions to (5.2), (5.1)

s	m	N	$H_{L_2}(z)$	$H_{L_\infty}(z)$
35	560	2945	1.956e-05	6.206e-05
50	881	6690	1.045e-05	4.916e-05
65	1212	11173	6.301e-06	1.843e-05
80	1564	17484	4.188e-06	1.765e-05
95	1920	24433	2.667e-06	9.246e-06
110	2273	30586	1.815e-06	5.826e-06
125	2645	39234	1.535e-06	7.368e-06
140	3008	46195	1.094e-06	4.615e-06
155	3391	56289	9.457e-07	2.652e-06

(a) $v \equiv 1.03$

s	m	N	$H_{L_2}(z)$	$H_{L_\infty}(z)$
35	442	549	3.065e-05	1.119e-04
50	696	1051	1.337e-05	4.995e-05
65	975	1804	9.388e-06	3.104e-05
80	1239	2306	6.390e-06	1.862e-05
95	1541	3315	4.309e-06	1.639e-05
110	1822	3949	3.470e-06	1.532e-05
125	2114	4685	2.429e-06	7.471e-06
140	2414	5536	1.883e-06	5.876e-06
155	2723	6516	1.506e-06	8.149e-06

(b) $v \equiv 1.06$

Table 5.1: Values of $H_{L_2}(z)$ and $H_{L_\infty}(z)$ for a number of parameter combinations

The first problem is a simple one-dimensional diffusion equation as in Example 3.5

$$-\nabla \cdot (a(\cdot, z)\nabla u) - 10 = 0 \text{ in } \Omega \tag{5.1}$$

$$u = 0 \text{ on } \partial\Omega \tag{5.2}$$

with $\Omega := [0, 1]$ and diffusion coefficient

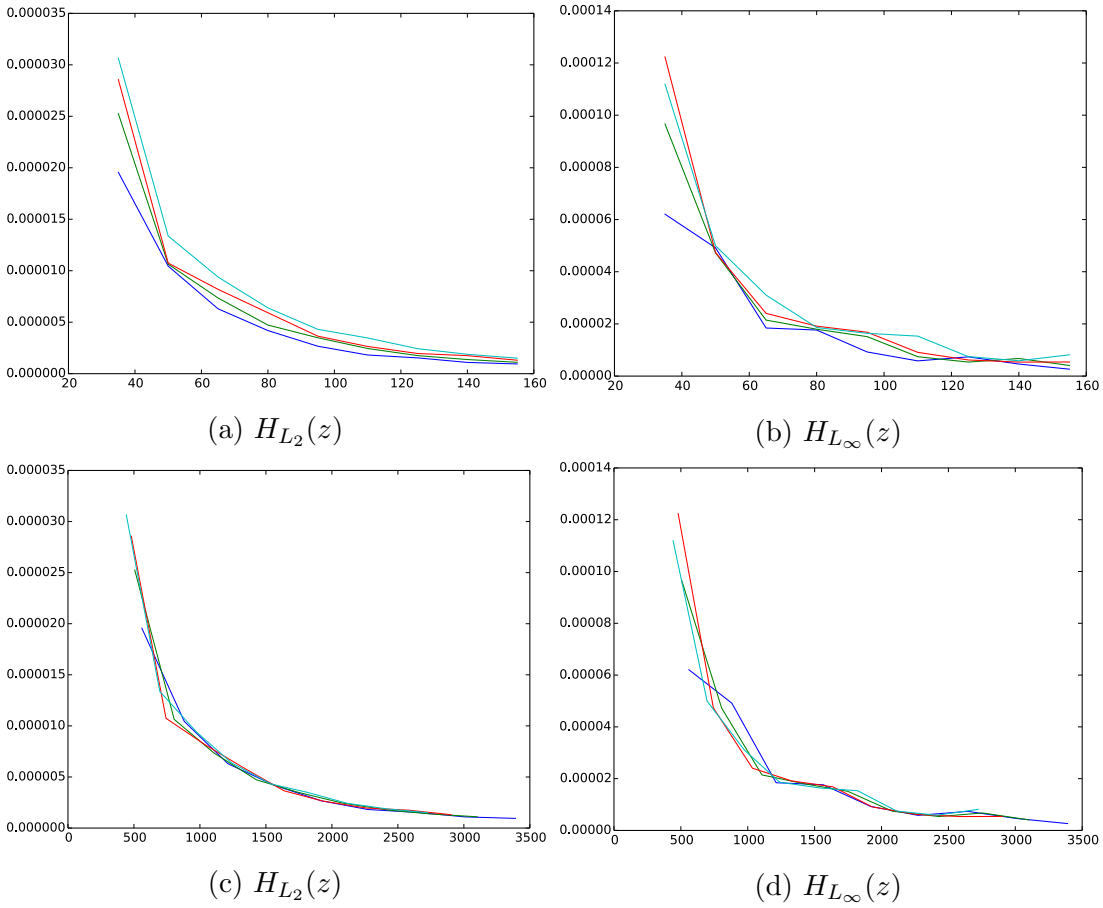
$$a(z, x) := a_0(x) + \sum_{j=1}^{4} z_j a_j(x) \tag{5.3}$$

$$a_0(x) := 4.3, \quad a_{2j-1}(x) := \cos(j\pi x), \quad a_{2j}(x) := \sin(j\pi x), \quad j = 1, 2. \tag{5.4}$$

The functional considered is the spatial average of the solution

$$F(z) := G(u(z)) := \int_\Omega u(z, x)\,\mathrm{d}x. \tag{5.5}$$

The error on the samples has been bounded by $\eta_2 = 10^{-8}$.
Figure 5.2 shows that the error in the parameter space with respect to the L_2-norm decreases steadily with an increasing number of samples, i.e. with increasing sparsity s and decreasing size of the weights v. Similar behaviour is observed in the L_∞-error. While it does not decrease strictly monotonically, it still converges to zero. The reconstructed functionals also converge pointwise against the original as can be observed in the plots Figure 5.3 and Figure 5.4.
As seen in Table 5.1b, increasing v has a substantial effect on the number of basis functions, while less so on the number of required samples m. Choosing v small thus might not have a major impact on the quality of the obtained reconstruction, but the ℓ_1-minimization quickly becomes intreatable due to its size.
By Figure 5.5 the approximated expectation value $\mathbb{E}[F^\#]$ of the functional appears to have converged to four significant places using only about $m = 500$ samples. The values obtained by the Monte Carlo methods agree with $\mathbb{E}[F^\#]$ to at least three significant places. Moreover, the variation of $\mathbb{E}[F^\#]$ with increasing m is much smaller compared to the values computed by the Monte Carlo methods.



(a) $H_{L_2}(z)$ (b) $H_{L_\infty}(z)$

(c) $H_{L_2}(z)$ (d) $H_{L_\infty}(z)$

Figure 5.2: Comparison of the reconstructed functionals $F^{\#}$ against $\hat{F}$ corresponding to the functional $F(z)$ as in (5.5) over 10^4 randomly chosen parameter values. The x-axis corresponds to the sparsity s in the first row and the number of samples m in the second row.
Blue: $v \equiv 1.03$, Green: $v \equiv 1.04$, Red: $v \equiv 1.05$, Turquoise: $v \equiv 1.06$.

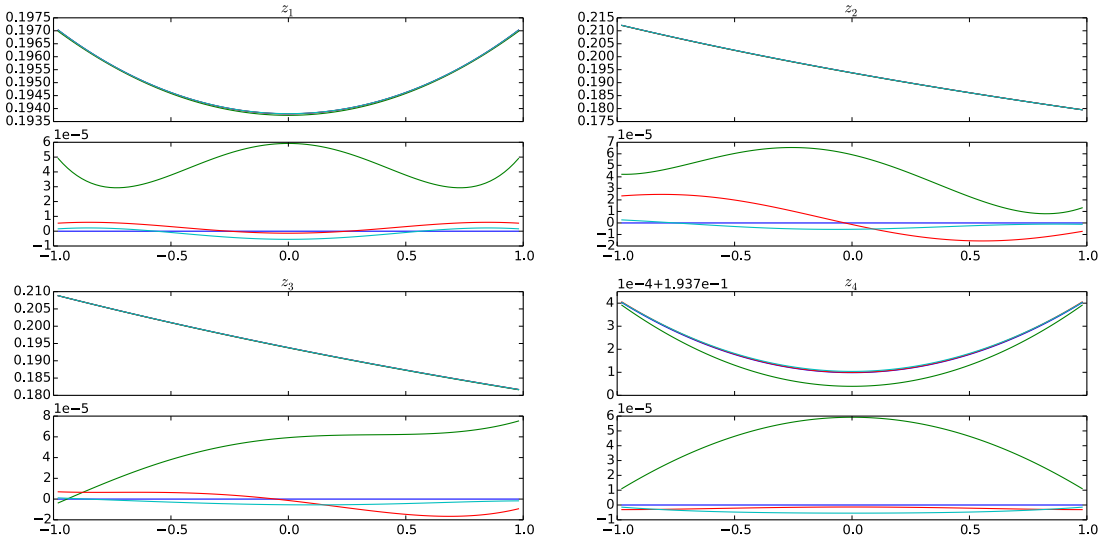
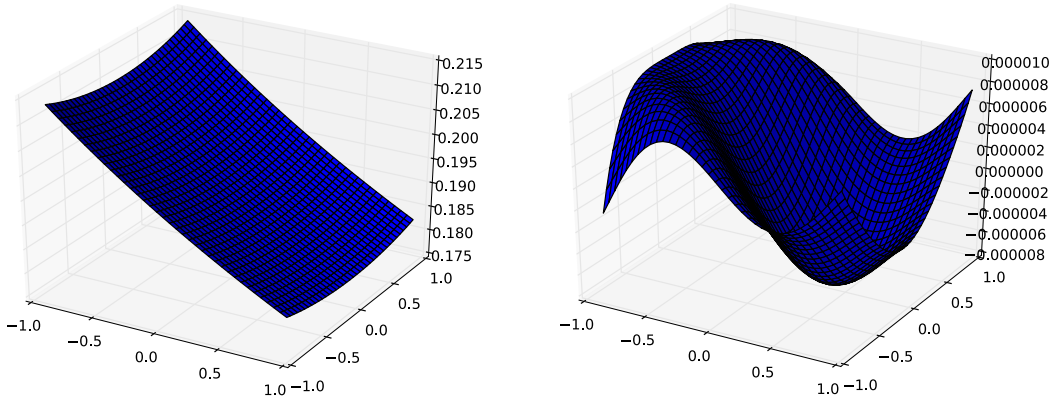


Figure 5.3: A pointwise comparison of $F^{\#}$ against $\hat{F}$ corresponding to the functional $F(z)$ as in (5.5) for different sparsity values s and weights $v \equiv 1.06$. In every plot, the parameter in the title is varied along the x-axis, while all other parameters are set identically zero. The upper subplots show the actual values of the functionals. The lower subplots show $\hat{F} - F^{\#}$.
Blue: $\hat{F}$, Green: $s = 35$, Red: $s = 65$, Turquoise: $s = 140$



(a) $F^{\#}(z)$ (b) $\hat{F}(z) - F^{\#}(z)$

Figure 5.4: A pointwise comparison of via $F^{\#}$ reconstructed values against $\hat{F}$ corresponding to the functional $F(z)$ as in (5.5) for sparsity value $s = 95$ and weights $v \equiv 1.06$. The z_1-parameter corresponds to the x-axis, the z_2-parameter to the y-axis.

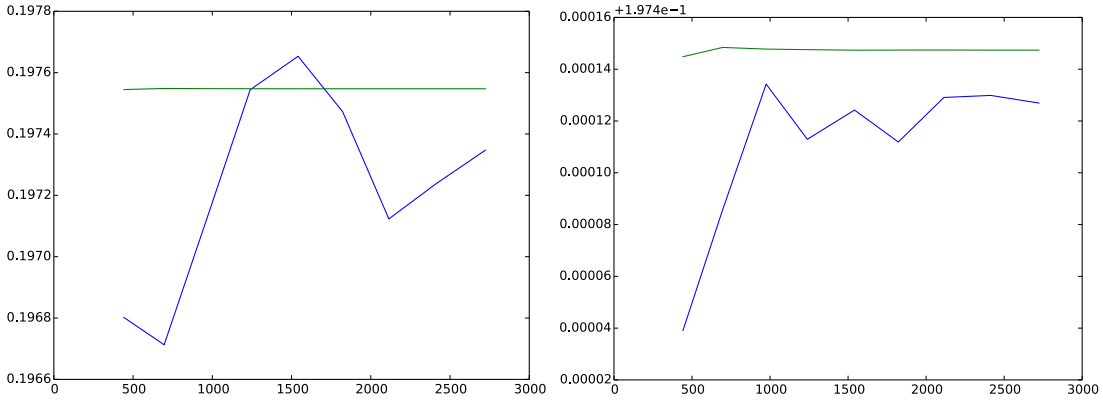
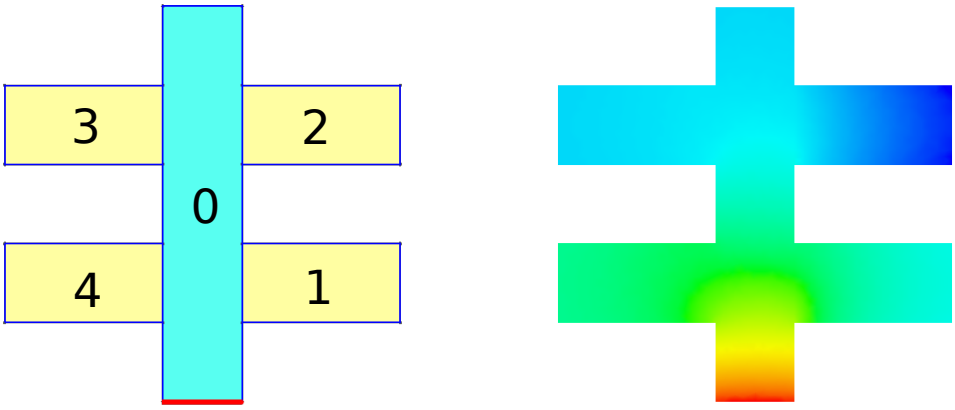


Figure 5.5: Plot of the average value $\mathbb{E}[F^{\#}]$ obtained via the reconstruction in green and the average value $E_m[\hat{F}]$ obtained via Monte Carlo approximation over the same samples in blue in (a) and $E_m[\hat{F}]$ obtained sampling along the Halton-Sequence in (b). The number of samples m corresponds to the x-axis. The weights have been chosen as $v \equiv 1.06$.

5.2 Thermal Fin



(a) Schematic (b) Example solution

Figure 5.6: The thermal fin is split into five regions and heated from below at Γ_{Root} marked red in the schematic.

Now finally back to the introductory example of Section 1.1. The heat transfer inside the fin is modelled using a diffusion equation and the interaction with the surrounding medium by Newton's Law of Cooling in the boundary conditions,

except for the root, where a unit heat inflow is assumed. This gives a linear elliptic equation with Robin boundary conditions, as introduced in Example 3.6,

$$-\nabla \cdot (a(\cdot, z)\nabla u) = 0 \text{ in } \Omega \tag{5.6}$$

$$\frac{\partial u}{\partial n} = 1 \text{ on } \Gamma_{\text{Root}}, \quad \frac{\partial u}{\partial n} = \beta u \text{ on } \partial\Omega \setminus \Gamma_{\text{Root}} \tag{5.7}$$

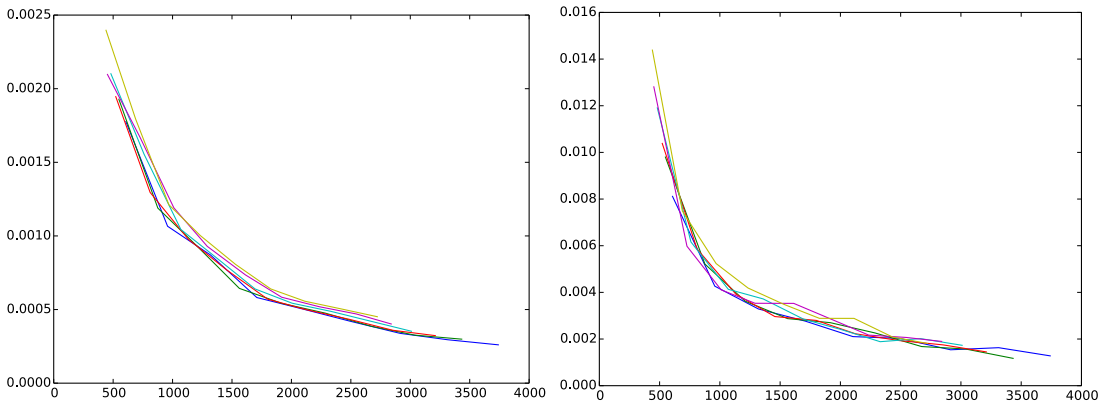
where n is the unit normal vector to the boundary and β is the so called Biot number. The Biot number describes how well the fin is isolated from the surrounding medium, where a small number effectively corresponds to thermal conductor. It thus has been chosen rather small as $\beta = 0.1$. The different materials of the fin in the respective regions are captured by the the parameter dependent diffusion coefficient

$$a(x, z) := \sum_{j=1}^{5} \left(1 + 9 \cdot \frac{z_j + 1}{2}\right) \Psi_j(x),$$

where $z_j \in [-1, 1]$ and $\Psi_j(x)$ is the characteristic function of the j-th region. The average temperature at the bottom of the fin for some parameter z equals to the value of the functional

$$F(z) := G(u(z)) := \int_{\Gamma_{\text{Root}}} u(z, s)\, \mathrm{d}s. \tag{5.8}$$

The samples have been obtained with a maximal error of $\eta_2 = 10^{-5}$.
The overall L_2- and L_∞-error qualitatively show the same behaviour as in the previous example of the diffusion equation with a trigonometric coefficient. The L_2-error decreases monotonically with the number of samples, the L_∞-error still appears to converge to zero. This behaviour is again reflected in plots Figure 5.8 and Figure 5.9, although this time the respective parameters have a greater influence on the values $F(z)$ and the error does not converge as fast as before. By Figure 5.7b a much better pointwise error is also not expected, since the error in the samples might be as large as $\eta_2 = 10^{-5}$.
Once more Table 5.2 illustrates that, while the weights v are not central to the quality of the solution, choosing these too small result in a very large ℓ_1-minimization problem. Comparing Table 5.2a and Table 5.1a also shows, how sensible N is to the parameter space dimension d. Going up from $d = 4$ to just $d = 5$ parameters, the number of basis functions almost triples compared to the previous example for the still very moderate sparsity value $s = 155$.
As seen in Figure 5.10, the expectation value $\mathbb{E}[F^\#]$ does not converge as fast as in the example before, but nevertheless still significantly outperforms the Monte Carlo methods. About 10^3 samples appear to be enough to approximate the expectation value of the functional to three significant figures.



(a) $H_{L_2}(z)$ (b) $H_{L_\infty}(z)$

Figure 5.7: Comparison of the reconstructed functionals $F^{\#}$ against $\hat{F}$ corresponding to the functional $F(z)$ as in (5.8) over 10^4 randomly chosen parameter values. The number of samples m corresponds to the x-axis. Blue: $v \equiv 1.03$, Green: $v \equiv 1.04$, Red: $v \equiv 1.05$, Turquoise: $v \equiv 1.06$, Magenta: $v \equiv 1.07$, Greenyellow: $v \equiv 1.08$.

s	m	N	$H_{L_2}(z)$	$H_{L_\infty}(z)$
35	607	5801	1.777e-03	8.116e-03
50	958	14426	1.066e-03	4.265e-03
65	1320	25546	8.699e-04	3.289e-03
80	1709	43356	5.815e-04	2.767e-03
95	2104	64372	5.019e-04	2.104e-03
110	2494	83692	4.207e-04	2.026e-03
125	2914	115218	3.383e-04	1.539e-03
140	3313	137398	2.951e-04	1.629e-03
155	3741	173668	2.603e-04	1.278e-03

(a) $v \equiv 1.03$

s	m	N	$H_{L_2}(z)$	$H_{L_\infty}(z)$
35	440	531	2.097e-03	1.281e-02
50	689	981	1.669e-03	5.978e-03
65	969	1716	1.191e-03	4.106e-03
80	1234	2231	9.292e-04	3.532e-03
95	1530	3126	7.354e-04	3.528e-03
110	1826	4016	5.837e-04	2.863e-03
125	2113	4681	5.191e-04	2.168e-03
140	2432	5917	4.701e-04	2.068e-03
155	2723	6517	4.019e-04	1.897e-03

(b) $v \equiv 1.08$

Table 5.2: Values of $H_{L_2}(z)$ and $H_{L_\infty}(z)$ for a number of parameter combinations

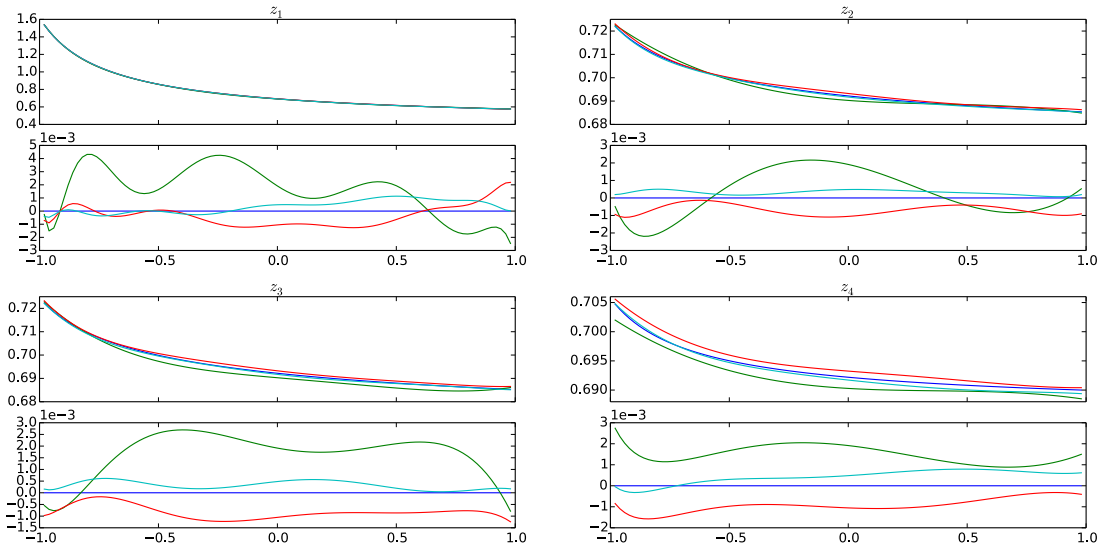
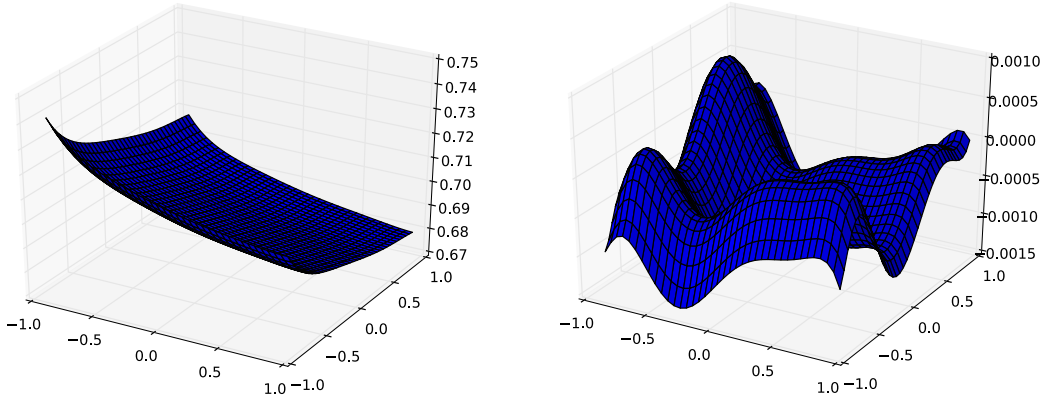


Figure 5.8: A pointwise comparison of $F^{\#}$ against $\hat{F}$ corresponding to the functional $F(z)$ as in (5.8) for different sparsity values s and weights $v \equiv 1.08$. In every plot, the parameter in the title is varied along the x-axis, while all other parameters are set to zero. The upper subplots show the actual values of the functional. The lower subplots show $\hat{F} - F^{\#}$.
Blue: $\hat{F}$, Green: $s = 35$, Red: $s = 95$, Turquoise: $s = 155$



(a) $F^{\#}(z)$ (b) $\hat{F}(z) - F^{\#}(z)$

Figure 5.9: A pointwise comparison of $F^{\#}$ against $\hat{F}$ corresponding to the functional $F(z)$ as in (5.8) for sparsity value $s = 95$ and weights $v \equiv 1.08$. The z_2-parameter corresponds to the x-axis, the z_3-parameter to the y-axis.

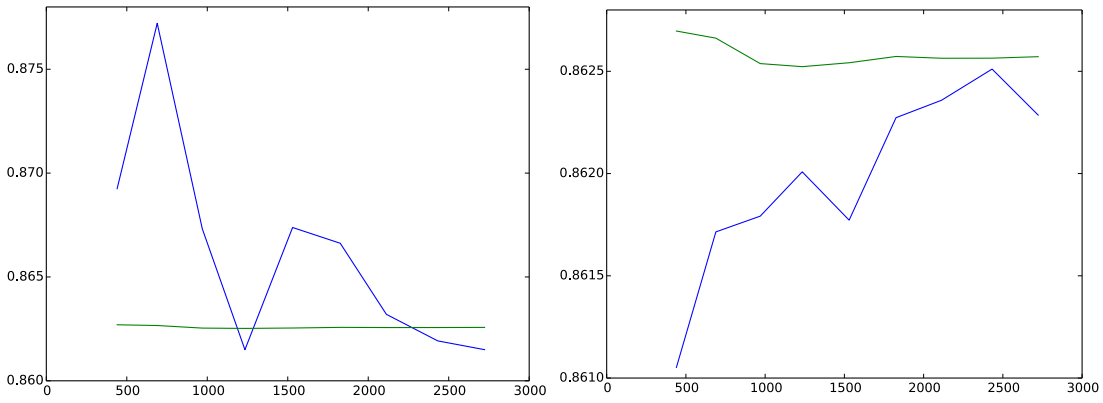
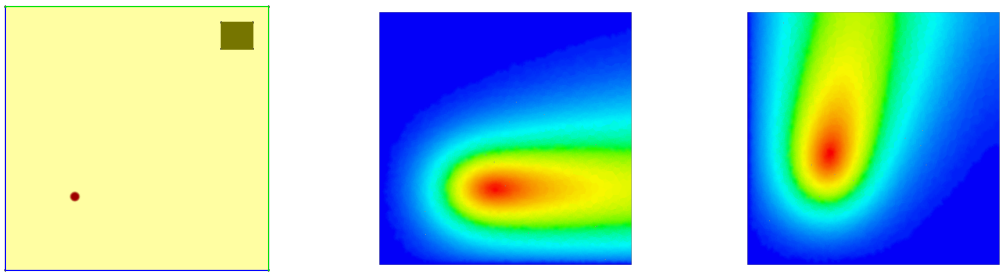


Figure 5.10: Plot of the average value $\mathbb{E}[F^{\#}]$ obtained via the reconstruction in green and the average value $E_m[\hat{F}]$ obtained via Monte Carlo approximation over the same samples in blue in (a) and $E_m[\hat{F}]$ obtained sampling along the Halton-Sequence in (b). The number of samples m corresponds to the x-axis. The weights have been chosen as $v \equiv 1.08$.

5.3 Dispersion of Pollutant



(a) Schematic (b) Example: $c = (7.5, 0)$ (c) Example: $c = (13.5, 8.6)$

Figure 5.11: Schematic and two example solutions for the pollutant problem. The blue boundary corresponds to a zero Dirichlet condition (an absorbing wall), the green boundary to a zero Neumann condition (surrounding air). The source is located at the bottom left dot. The concentration of the pollutant is measured inside the box in the top right corner.

The last example is concerned with the propagation of some pollutant emitted near a building subject to different wind speeds. This phenomenon may be modelled

by a convection-diffusion equation as introduced in Example 3.6 of the form

$$-\nabla \cdot (a(\cdot)\nabla u) + c(\cdot, z)\nabla u - f(\cdot) = 0 \text{ in } \Omega, \tag{5.9}$$

$$u = 0 \text{ on } \Gamma_1, \quad \frac{\partial u}{\partial n} = 0 \text{ on } \Gamma_2, \tag{5.10}$$

where the varying wind speed is represented by the parameter dependent convection coefficient $c(\cdot, z)$ and the source of pollution by some function $f(\cdot)$. The condition $u = 0$ on Γ_1 furthermore enforces the total pollution to be zero in Γ_1, so that Γ_1 corresponds to an inpenetrable absorbing surface, e.g. a wall or building. By $\frac{\partial u}{\partial n} = 0$ the flow direction of the pollutant does not change in Γ_2, i.e. the pollution just leaves the considered region Ω without facing any obstructions.
To keep things simple, assume a square region $\Omega := \left\{x \in \mathbb{R}^2 \mid 0 \leq x_1, x_2 \leq 5\right\}$ with Γ_1, Γ_2 as in Figure 5.11 and one is interested in the total concentration of the pollutant in the region $\Omega_{\text{Box}} := \left\{x \in \mathbb{R}^2 \mid 4.2 \leq x_1, x_2 \leq 4.7\right\}$, that is

$$F(z) := G(u(z)) := \int_{\Omega_{\text{Box}}} u(z, x)\, \mathrm{d}x. \tag{5.11}$$

Further, assume the wind speed varies between 0 and 15 units of speed along the x-axis and between 0 and 30 along the y-axis respectively with the parameters, but is spatially constant, i.e. set

$$c(x, (z_1, z_2)) := (15 \cdot \frac{z_1 + 1}{2}, 30 \cdot \frac{z_2 + 1}{2})$$

with $z_1, z_2 \in [-1, 1]$ and assume only minor diffusion given by

$$a(x) := 1.3.$$

Finally, model the emission of the pollutant by

$$f(x) := 20 \cdot e^{-2((x_1 - 1.5)^2 + (x_2 - 1.5)^2)}.$$

All samples considered have been obtained with a maximal error of $\eta_2 = 10^{-5}$.
Compared to the previous examples, Algorithm 6 appears to perform less well in this very low-dimensional case. As illustrated by Figure 5.14, the reconstruction for $s = 35$, $v \equiv 1.02$ using $m = 466$ samples can not even compete with the simple linear interpolation plotted in Figure 5.14a using only 400 samples.
Compared to the previous examples, also the expectation value $\mathbb{E}[F^{\#}]$ converges slowly. To obtain the correct value to only two significant places, one requires as many as 2000 samples, as shown by Figure 5.15. This is probably not due to the low dimensionality of the problem, but rather due to the more complex behaviour of the system in the parameters. Again, the variation of $\mathbb{E}[F^{\#}]$ with increasing m is smaller compared to the values computed by the Monte Carlo methods.

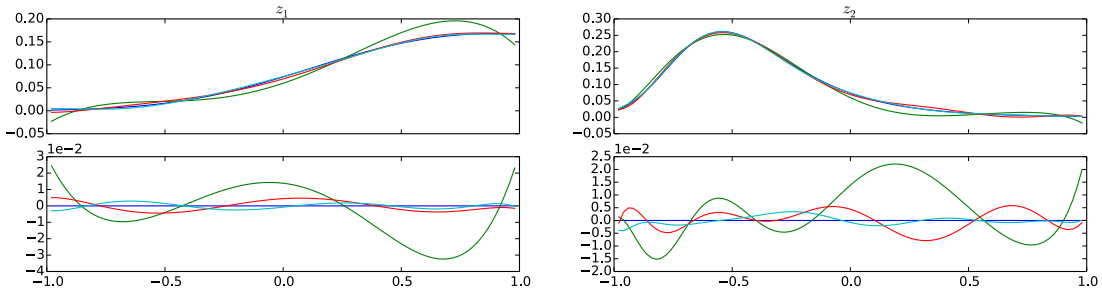
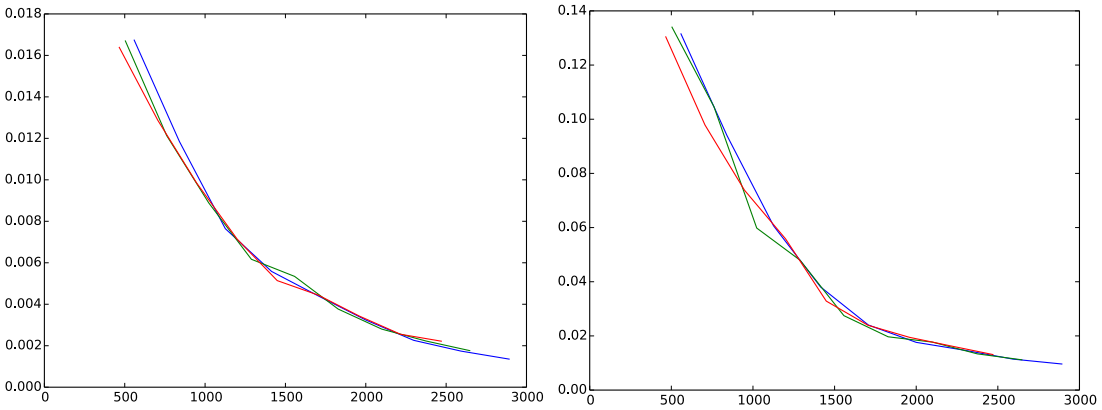


Figure 5.12: A pointwise comparison of $F^{\#}$ against $\hat{F}$ corresponding to the functional $F(z)$ as in (5.11) for different sparsity values s and weights $v \equiv 1.02$. In every plot, the parameter in the title is varied along the x-axis, while $z_2 = 0$ in the first and $z_1 = 0$ in the second plot. The upper subplots show the actual values of the functionals. The lower subplots show $\hat{F} - F^{\#}$.
Blue: $\hat{F}$, Green: $s = 35$, Red: $s = 95$, Turquoise: $s = 155$



(a) $H_{L_2}(z)$ (b) $H_{L_\infty}(z)$

Figure 5.13: Comparison of the reconstructed functionals $F^{\#}$ against $\hat{F}$ corresponding to the functional $F(z)$ as in (5.11) over 10^4 randomly chosen parameter values. The number of samples m corresponds to the x-axis.
Blue: $v \equiv 1.01$, Green: $v \equiv 1.015$, Red: $v \equiv 1.02$.

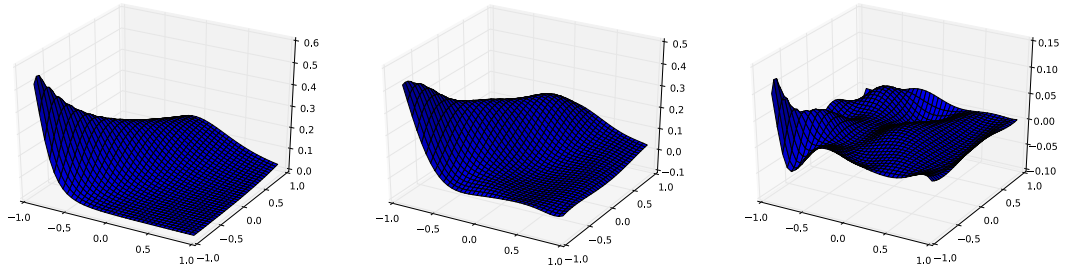
s	m	N	$H_{L_2}(z)$	$H_{L_\infty}(z)$
35	559	2918	1.658e-02	9.569e-02
50	840	4439	1.244e-02	5.979e-02
65	1126	5741	7.637e-03	4.682e-02
80	1414	6856	5.535e-03	3.003e-02
95	1707	7945	3.888e-03	2.470e-02
110	2000	8848	3.124e-03	1.653e-02
125	2298	9798	2.419e-03	1.227e-02
140	2593	10510	1.735e-03	1.093e-02
155	2892	11245	1.267e-03	9.956e-03

(a) $v \equiv 1.01$

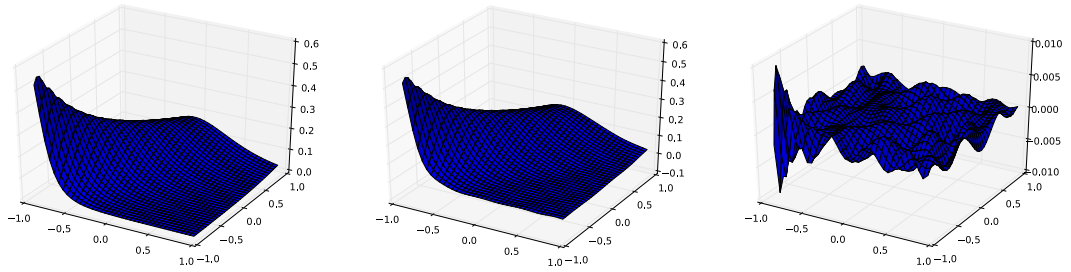
s	m	N	$H_{L_2}(z)$	$H_{L_\infty}(z)$
35	466	775	1.769e-02	1.170e-01
50	706	1162	1.224e-02	7.592e-02
65	948	1467	1.020e-02	5.270e-02
80	1200	1804	7.351e-03	4.263e-02
95	1449	2050	5.146e-03	2.957e-02
110	1705	2312	4.404e-03	2.441e-02
125	1958	2519	3.453e-03	1.785e-02
140	2216	2735	2.553e-03	1.463e-02
155	2470	2886	2.258e-03	1.160e-02

(b) $v \equiv 1.02$

Table 5.3: Values of $H_{L_2}(z)$ and $H_{L_\infty}(z)$ for a number of parameter combinations.



(a) $\hat{F}(z)$ (b) $F^{\#}(z)$ (c) $\hat{F}(z) - F^{\#}(z)$



(d) $\hat{F}(z)$ (e) $F^{\#}(z)$ (f) $\hat{F}(z) - F^{\#}(z)$

Figure 5.14: Pointwise comparison of true to reconstructed values of the functional $F(z)$ as in (5.11). The z_1-parameter corresponds to the x-axis, the z_2-parameter to the y-axis. The reconstructed functional $F^{\#}(z)$ in the top row has been obtained for $s = 35$, $v = 1.02$ and in the bottom row for $s = 155$, $v = 1.01$.

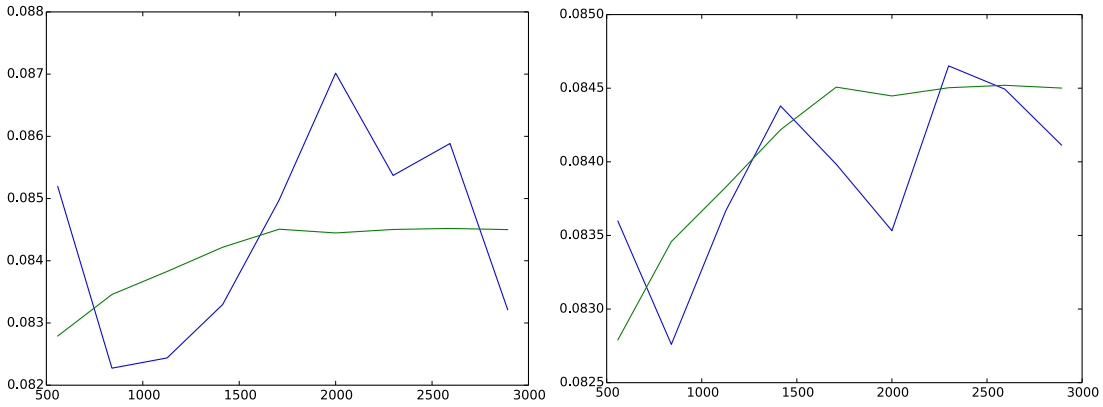


Figure 5.15: Plot of the average value $\mathbb{E}[F^{\#}]$ obtained via the reconstruction in green and the average value $E_m[\hat{F}]$ obtained via Monte Carlo approximation over the same samples in blue in (a) and $E_m[\hat{F}]$ obtained sampling along the Halton-Sequence in (b). The number of samples m corresponds to the x-axis. The weights have been chosen as $v \equiv 1.01$.

6 Summary and Future Work

6.1 Summary

The Compressive Sensing Petrov-Galerkin (CSPG) method introduced in [RS14] has been shown to be practical and efficient. It has furthermore been demonstrated to yield a new approach to evaluation of statistical quantities such as the mean value or variance, that is potentially more efficient than quasi-Monte Carlo methods by a series of numerical tests. The findings of Weighted Compressive Sensing (cf. [RW15]) central to CSPG have also been compiled and reviewed. To solve in the CSPG method arising weighted ℓ_1-minimization problems, new algorithms have been introduced and discussed. Using the results of [CCS14] it has furthermore been demonstrated that the CSPG method applies indeed to a large class of problems.

6.2 Future Work

While the findings so far strengthen confidence in the CSPG method, there is certainly still plenty of room for improvements.
As pointed out in Chapter 4 and Chapter 5, it is not clear at all how to choose the parameters to the CSPG method, especially s, η_1, J_0^s and v.
The properties of the input quantities, e.g. the convergence estimates of the diffusion coefficient in the thermal fin example, have been related to the weights v via a-priori bounds on the error in the reconstructed functional. As it has also been demonstrated in Chapter 4, it is however unclear of how much further practical use these estimates are. In [PHD14, Section 4.1.1] it has been suggested to instead obtain the weights numerically by one of the approaches taken in [BTNT12].
It is not evident whether the total degree form of the index set J_0^s is a good choice, although it appears to work quite well in practice (cf. [BTNT12, Section 3.2]).
Furthermore, with η_1 the quadratic constraint constant η in the ℓ_1-minimization problem is also unknown. This is not a problem in practice if one uses a greedy or iterative method like Weighted Hard Thresholding Pursuit, but it is not obvious how to apply, e.g. Chambolle and Pock's method. It has been suggested in [RS14] to replace the quadratic by an equality constraint, presumably yielding reconstruction errors that are comparable to the bounds in the original problem (cf. [FR13, Chapter 11]).
Finally, even given all these parameters, the random sampling matrix A still satisfies the RIP only probabilistically and hence is maybe not optimal in all

cases. Additionally, using any algorithm, the efficient solution of the weighted ℓ_1-minimization problem crucially depends on fast transformations with A and A^*, as discussed in Subsection 2.2.3 and Section 4.3. However, finding deterministic sampling matrices that are only guaranteed to satisfy some RIP is by itself an open research topic in Compressive Sensing.

One might also consider to extend the CSPG method.

For example, so far only parameter dependent *functionals* have been reconstructed. It is not clear, how to deal with the reconstruction of complete solutions, which in general lie in an infinite dimensional solution space. But when using a FEM, one might consider applying methods similar to low-rank techniques in Compressive Sensing on the finite dimensionals solution vectors provided.

Except for the convergence estimates of the functionals encoded by some weights v, no further information that may come with the given operator equation or method used to obtain the samples have been employed. It has, for instance, already been noted in the introductory example that the temperature at the root of the fin decreases *monotonically* with increasing heat transfer coefficients. Such information may be converted to additional constraints to the corresponding ℓ_1-minimization problem. Following an approach different from the one presented here, but nevertheless employing ℓ_1-methods, it has been shown in [Tan13] that taking derivatives to a solution into consideration might yield better results.

One may also postprocess the solution based on such knowledge. The reconstructed functionals in Section 5.3, for example, take negative values for some parameter values, although this would correspond to a negative concentration and does not make any sense in this context. The solution thus can already be improved by simply thresholding the reconstructed function to values larger than zero.

Finally, considering an actual real-world example would doubtlessly help to identify the in practice most important directions for future research.

Bibliography

[AcCD11] Ery Arias-castro, Emmanuel J. Candes, and Mark A. Davenport. On the fundamental limits of adaptive sensing. Technical report, Stanford, 2011.

[BD10] Thomas Blumensath and Michael E. Davies. Normalized iterative hard thresholding: Guaranteed stability and performance. *J. Sel. Topics Signal Processing*, 4(2):298–309, 2010.

[BFH13] Jean-Luc Bouchot, Simon Foucart, and Pawel Hitczenko. Hard thresholding pursuit algorithms: Number of iterations. submitted, 2013.

[BNTT11] Joakim Bäck, Fabio Nobile, Lorenzo Tamellini, and Raul Tempone. Stochastic spectral galerkin and collocation methods for PDEs with random coefficients: A numerical comparison. In Jan S. Hesthaven and Einar M. Rønquist, editors, *Spectral and High Order Methods for Partial Differential Equations*, volume 76 of *Lecture Notes in Computational Science and Engineering*, pages 43–62. Springer Berlin Heidelberg, 2011.

[BSZ11] Andrea Barth, Christoph Schwab, and Nathaniel Zollinger. Multi-level monte carlo finite element method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik*, 119(1):123–161, 2011.

[BTNT12] Joakim Beck, Raul Tempone, Fabio Nobile, and Lorenzo Tamellini. On the optimal polynomial approximation of stochastic PDEs by galerkin and collocation methods. *Mathematical Models and Methods in Applied Sciences*, 22(09):1250023, 2012.

[CCS14] Albert Cohen, Abdellah Chkifa, and Christoph Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. *Journ. Math. Pures et Appliquees*, 2014.

[CDL13] Albert Cohen, Mark A. Davenport, and Dany Leviatan. On the stability and accuracy of least squares approximations. *Foundations of Computational Mathematics*, 13(5):819–834, 2013.

[CDS10a] Albert Cohen, Ronald DeVore, and Christoph Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs. *Analysis and Applications*, 9(1):11–47, 2010.

[CDS10b] Albert Cohen, Ronald DeVore, and Christoph Schwab. Convergence rates of best N-term galerkin approximations for a class of elliptic sPDEs. *Journ. Found. Comp. Math.*, 10(6):615–646, 2010.

[Che13] W. Cheney. *Analysis for Applied Mathematics*. Graduate Texts in Mathematics. Springer New York, 2013.

[CP11] 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.

[CRT06] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theor.*, 52(2):489–509, February 2006.

[DM09] Wei Dai and Olgica Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5):2230–2249, 2009.

[DO11] Alireza Doostan and Houman Owhadi. A non-adapted sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys.*, 230(8):3015–3034, April 2011.

[Don06] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.

[DZ13] Laurent Demanet and Xiangxiong Zhang. Eventual linear convergence of the douglas rachford iteration for basis pursuit. *arXiv*, abs/1301.0542, 2013.

[Elm11] Elman, H.C. and Miller, C.W. and Phipps, E.T. and Tuminaro, R.S. Assessment of collocation and Galerkin approach to linear diffusion equations with random data. *Intern. J. for uncertainty quantification*, 1(1):19–34, 2011.

[EMSU12] Oliver G. Ernst, Antje Mugler, Hans-Jörg Starkloff, and Elisabeth Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46:317–339, 3 2012.

[FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer, Berlin Heidelberg, 2013.

[FR15] Jonathan Fell and Holger Rauhut. Sparse approximation from random fourier samples of wavelet expansions via weighted iterative hard thresholding (unreleased). 2015.

[GO09] Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[Gra08] Loukas Grafakos. *Classical Fourier Analysis*. Springer, New York; [London], 2008.

[GS91] Roger G. Ghanem and Pol D. Spanos. *Stochastic Finite Elements*. Springer, Berlin Heidelberg, 1991.

[Jo13] Jason Jo. Iterative hard thresholding for weighted sparse approximation. *arXiv*, abs/1312.3582, 2013.

[JOP$^+$] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed 2014-09-18].

[KKP12] Lutz Kämmerer, Stefan Kunis, and Daniel Potts. Interpolation lattices for hyperbolic cross trigonometric polynomials. *Journal of Complexity*, 28(1):76 – 92, 2012.

[KR97] R.V. Kadison and J.R. Ringrose. *Fundamentals of the Theory of Operator Algebras*. Number Vol. 1. American Mathematical Society, 1997.

[Lad68] Olga A. Ladyzhenskaya. *Linear and quasilinear elliptic equations*. Mathematics in Science and Engineering. Elsevier Science, 1968.

[LMW$^+$12] Anders Logg, Kent-Andre Mardal, Garth N. Wells, et al. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012.

[Mig13] Giovanni Migliorati. *Polynomial approximation by means of the random discrete L^2 projection and application to inverse problems for PDEs with stochastic data*. Theses, Ecole Polytechnique X, April 2013.

[MNGK04] O.P. Le Maitre, H.N. Najm, R.G. Ghanem, and O.M. Knio. Multi-resolution analysis of wiener-type uncertainty propagation schemes. *Journal of Computational Physics*, 197(2):502 – 531, 2004.

[MNST14] G. Migliorati, F. Nobile, E. Schwerin, and R. Tempone. Analysis of discrete l^2 projection on polynomial spaces with random evaluations. *Found. Comput. Math.*, 14(3):419–456, June 2014.

[MNvST13] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Approximation of quantities of interest in stochastic PDEs by the random discrete l^2 projection on polynomial spaces. *SIAM J. Scientific Computing*, 35(3), 2013.

[Moe12] Michael Moeller. *Multiscale Methods for (Generalized) Sparse Recovery and Applications in High Dimensional Imaging.* PhD thesis, Universität Münster, 2012.

[Nie92] Harald Niederreiter. *Random Number Generation and quasi-Monte Carlo Methods.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.

[NT10] Deanna Needell and Joel A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Commun. ACM*, 53(12):93–100, December 2010.

[PC11] Thomas Pock and Antonin Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *International Conference on Computer Vision (ICCV 2011)*, 2011. To Appear.

[PHD14] Ji Peng, Jerrad Hampton, and Alireza Doostan. A weighted ℓ_1-minimization approach for sparse polynomial chaos expansions. *Journal of Computational Physics*, 267(0):92 – 111, 2014.

[Pot03] Daniel Potts. Fast algorithms for discrete polynomial transforms on arbitrary grids. *Linear Algebra and its Applications*, 366(0):353 – 370, 2003.

[PST98] Daniel Potts, Gabriele Steidl, and Manfred Tasche. Fast algorithms for discrete polynomial transforms. *Math. Comput*, 67:1577–1590, 1998.

[RS14] Holger Rauhut and Christoph Schwab. Compressive sensing petrov-galerkin approximation of high-dimensional parametric operator equations. Technical Report 2014-29, Seminar for Applied Mathematics, ETH Zürich, 2014.

[RW15] Holger Rauhut and Rachel Ward. Interpolation via weighted ℓ_1 minimization. *Applied and Computational Harmonic Analysis*, 2015.

[Tan13] Gary Tang. *Methods for high dimensional uncertainty quantification.* PhD thesis, Stanford, 2013.

[WXGK05] Xiaoliang Wan, Dongbin Xiu, George, and Em Karniadakis. Stochastic solutions for the twodimensional advection-diffusion equation. *SIAM J. Sci. Computing*, 26:578–590, 2005.

[Xiu07] Dongbin Xiu. Efficient collocational approach for parametric uncertainty analysis. *Commun. Comput. Phys*, pages 293–309, 2007.

[Xiu10] Dongbin Xiu. *Numerical Methods for Stochastic Computations*. Princeton University Press, 2010.

[XK02a] Dongbin Xiu and George Em Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*, 191(43):4927 – 4948, 2002.

[XK02b] Dongbin Xiu and George Em Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644, February 2002.

[XS09] Dongbin Xiu and Jie Shen. Efficient stochastic galerkin methods for random diffusion equations. *J. Comput. Phys.*, 228(2):266–281, February 2009.

[YMO11] Y. Yang, M. Moeller, and S. Osher. A dual split bregman method for fast ℓ^1 minimization. CAM report 11-57, UCLA, 2011.