DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Multilevel preconditioning for sparse optimization of functionals with nonconvex fidelity terms

S. Dahlke, M. Fornasier, U. Friedrich, T. Raasch

Preprint 155

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Multilevel preconditioning for sparse optimization of functionals with nonconvex fidelity terms

S. Dahlke, M. Fornasier, U. Friedrich, T. Raasch

Preprint 155

Multilevel preconditioning for sparse optimization of functionals with nonconvex fidelity terms*

S. Dahlke, M. Fornasier, U. Friedrich, T. Raasch

Abstract

This paper is concerned with the development of numerical schemes for the minimization of functionals involving sparsity constraints and nonconvex fidelity terms. These functionals appear in a natural way in the context of Tikhonov regularization of nonlinear inverse problems with ℓ_1 penalty terms. Our method of minimization is based on a generalized conditional gradient scheme. It is well–known that these algorithms might converge quite slowly in practice. Therefore, we propose an acceleration which is based on a decreasing thresholding strategy. Its efficiency relies on certain spectral properties of the problem at hand. We show that under certain boundedness and contraction conditions the resulting algorithm is linearly convergent to a *global* minimizer and that the iteration is monotone with respect to the functional. We study important classes of operator equations to which our analysis can be applied. Moreover, we introduce a certain multilevel preconditioning strategy which in practice promotes the aforementioned spectral properties for problems where the nonlinearity is a perturbation of a linear operator.

MSC 2010: 65K10, 65J15, 41A25, 65N12, 65T60, 47J06, 47J25.
Key Words: Conditional gradient method, non-convex optimization, sparse minimization, (nonlinear) operator equations, iterative thresholding, multilevel preconditioning, wavelets.

1 Introduction

The aim of this paper is to derive an efficient numerical algorithm for the global minimization of functionals of the form

$$\Gamma_{\boldsymbol{\alpha}}(\mathbf{u}) := \left\|K(\mathbf{u}) - y\right\|_Y^2 + 2\|\mathbf{u}\|_{\ell_{1,\boldsymbol{\alpha}}(\mathcal{J})}, \quad \mathbf{u} \in \ell_2(\mathcal{J}), \tag{1}$$

where $K : \ell_2(\mathcal{J}) \to Y$ is a nonlinear, continuously Fréchet differentiable operator acting between the sequence space $\ell_2(\mathcal{J})$ over the countable index

*December 13, 2013. This work has been supported by Deutsche Forschungsgemeinschaft (DFG), grant DA 360/12-2 and the LOEWE Center for Synthetic Microbiology (Synmikro), Marburg.

set $\mathcal{J}$ and a separable Hilbert space Y. Here $y \in Y$ is a given datum, and $\|\mathbf{u}\|_{\ell_{1,\boldsymbol{\alpha}}(\mathcal{J})} := \sum_{\mu \in \mathcal{J}} \alpha_\mu |u_\mu|$ denotes the weighted ℓ_1-norm of $\mathbf{u}$ with respect to a positive weight sequence $\boldsymbol{\alpha} \in \mathbb{R}_+^{\mathcal{J}}$. We shall assume that there exists $\alpha > 0$ such that $\alpha_\mu \geq \alpha$ for all $\mu \in \mathcal{J}$. Whenever the index set $\mathcal{J}$ is fixed and clear from the context, we will drop it in the notation and simply write ℓ_2 and $\ell_{1,\boldsymbol{\alpha}}$, respectively.

Typical examples where minimization problems of the form (1) arise are Tikhonov regularizations of nonlinear operator equations

$$\mathcal{K}(u) = y \tag{2}$$

when the forward operator $\mathcal{K} : X \to Y$ maps a separable Hilbert space X into Y. We refer, e.g., to [16] for a detailed discussion of Tikhonov regularization schemes. If the unknown solution is guaranteed to have a sparse expansion with respect to some suitable countable Riesz basis $\boldsymbol{\Psi} := \{\psi_\mu\}_{\mu \in \mathcal{J}}$ for X, it makes sense to utilize that the weighted ℓ_1-norm promotes sparse solutions. Denoting linear synthesis operator associated to $\boldsymbol{\Psi}$ with

$$u = \sum_{\mu \in \mathcal{J}} u_\mu \psi_\mu =: \mathcal{F}(\mathbf{u}), \quad \mathbf{u} \in \ell_2(\mathcal{J}),$$

and setting $K := \mathcal{K} \circ \mathcal{F}$, the minimization of (1) will produce a sparsely populated coefficient array $\mathbf{u}$ with $K(\mathbf{u}) \approx y$. The modeling motivation is the search of the "simplest" (in this case modeled by the "sparsest") explanation to the given datum y, resulting from the nonlinear process $\mathcal{K}$, in the spirit of the Occam's razor. Moreover, it is known that, under certain smoothness conditions, the *global minimizers* of (1) are regularizers for the problem.

By now there is a vast literature concerning sparse regularization of nonlinear inverse problems [2,3,19,23]. For most of the results in the literature related to minimizing algorithms for functionals of the type (1) usually only convergence to critical points is shown. Unfortunately, differently from global minimizers, nothing is really known concerning the regularization properties of critical points, significantly questioning the relevance of such convergence results.

The starting point of our present discussion is a generalized conditional gradient method which is known to guarantee the computation of subsequences converging to critical points of (1). The scope of this paper is to show under which *sufficient* conditions on $\mathcal{K}$ one may expect to have linear convergence of a suitable modification of this algorithm towards a *global minimizer*, hence guaranteeing regularization properties.

Several authors have independently proposed such an algorithm, see [14, 17, 20, 21] for the case of linear operators K and [2,3] for the generalization to the nonlinear case. The general setting can be described as follows. One introduces an auxiliary parameter $\lambda \in \mathbb{R}_+$, and considers the splitting

$$\Gamma_{\boldsymbol{\alpha}}(\mathbf{u}) = \underbrace{\|K(\mathbf{u}) - y\|_Y^2 - \lambda\|\mathbf{u}\|_{\ell_2}^2}_{=:\Gamma_\lambda^{(1)}(\mathbf{u})} + \underbrace{\lambda\|\mathbf{u}\|_{\ell_2}^2 + 2\|\mathbf{u}\|_{\ell_{1,\boldsymbol{\alpha}}}}_{=:\Gamma_{\lambda,\boldsymbol{\alpha}}^{(2)}(\mathbf{u})}. \tag{3}$$

Then $\Gamma_\lambda^{(1)}$ is continuously Fréchet differentiable and $\Gamma_{\lambda,\boldsymbol{\alpha}}^{(2)}$ is convex, lower semicontinuous, and coercive with respect to $\|\cdot\|_{\ell_2}$, so that all the necessary properties to set up a generalized conditional gradient method are satisfied. The algorithm is given by

Algorithm 1.1. 1. Choose $\mathbf{u}^{(0)} \in \ell_{1,\boldsymbol{\alpha}}$; $n := 0$;

2. Determine descent direction $\mathbf{v}^{(n)}$

$$\begin{aligned}\mathbf{v}^{(n)} \in \arg\min_{\mathbf{v}\in\ell_2} \Big(2\langle \big(K'(\mathbf{u}^{(n)})\big)^*\big(K(\mathbf{u}^{(n)}) - y\big) - \lambda\mathbf{u}^{(n)}, \mathbf{v}\rangle_{\ell_2} \\ + \lambda\|\mathbf{v}\|_{\ell_2}^2 + 2\|\mathbf{v}\|_{\ell_{1,\boldsymbol{\alpha}}}\Big);\end{aligned} \tag{4}$$

3. Determine step size $s^{(n)}$

$$s^{(n)} \in \arg\min_{s\in[0,1]} \Big(\Gamma_{\boldsymbol{\alpha}}(\mathbf{u}^{(n)} + s(\mathbf{v}^{(n)} - \mathbf{u}^{(n)}))\Big); \tag{5}$$

4. Set $\mathbf{u}^{(n+1)} := \mathbf{u}^{(n)} + s^{(n)}(\mathbf{v}^{(n)} - \mathbf{u}^{(n)})$; $n := n+1$; return to step 2.

Here $\big(K'(\mathbf{u}^{(n)})\big)^* \in \mathcal{L}(Y, \ell_2)$ denotes the adjoint mapping of $K'(\mathbf{u}^{(n)}) \in \mathcal{L}(\ell_2, Y)$. We refer to [2] for a detailed discussion and convergence analysis of Algorithm 1.1. If the parameter λ is chosen large enough, it is possible to choose $s^{(n)} = 1$ and to omit the third step of the algorithm, see Lemma 2.4 in [2]. Throughout this paper we always make this assumption, hence we focus on the minimization problem (4) in the following. Observe that by expanding the quadratic term below, (4) is equivalent to

$$\arg\min_{\mathbf{v}\in\ell_2} \Big(\|\mathbf{v} - \big(\mathbf{u}^{(n)} - \frac{1}{\lambda}\big(K'(\mathbf{u}^{(n)})\big)^*\big(K(\mathbf{u}^{(n)}) - y\big)\big)\|_{\ell_2}^2 + 2\|\mathbf{v}\|_{\ell_{1,\frac{\boldsymbol{\alpha}}{\lambda}}}\Big). \tag{6}$$

The minimizer of such a functional combining an ℓ_2-norm fidelity term and weighted ℓ_1-norm penalization can be directly computed using a soft thresholding operation, see [4, 13]. By defining

$$S_\alpha(x) := \begin{cases} x - \alpha, & x > \alpha, \\ 0, & |x| \le \alpha, \\ x + \alpha, & x < -\alpha, \end{cases}$$

and $\mathbb{S}_{\boldsymbol{\alpha}}(\mathbf{u})_\mu := S_{\alpha_\mu}(u_\mu)$ it holds that

$$\mathbb{S}_{\boldsymbol{\alpha}}(\mathbf{a}) = \arg\min_{\mathbf{v}\in\ell_2} \|\mathbf{v} - \mathbf{a}\|_{\ell_2}^2 + 2\|\mathbf{v}\|_{\ell_{1,\boldsymbol{\alpha}}}. \tag{7}$$

Consequently, through (6), we obtain that (4) is uniquely solved by

$$\mathbf{v}^{(n)} = \mathbb{S}_{\frac{\boldsymbol{\alpha}}{\lambda}}\Big(\mathbf{u}^{(n)} + \frac{1}{\lambda}\big(K'(\mathbf{u}^{(n)})\big)^*\big(y - K(\mathbf{u}^{(n)})\big)\Big). \tag{8}$$

This explains why Algorithm 1.1 is also known as the *iterated soft thresholding algorithm (ISTA)* or the *thresholded Landweber iteration.*

The convergence of Algorithm 1.1 for nonlinear operators K was studied in [3]. There it was shown that the sequence $\big(\mathbf{u}^{(n)}\big)_{n\in\mathbb{N}}$ has subsequences which are guaranteed to converge to a *stationary point* $\mathbf{u}^*$ of (1), i.e.,

$$\mathbf{u}^* \in \arg\min_{\mathbf{v}\in\ell_2}\Big(2\langle\big(K'(\mathbf{u}^*)\big)^*\big(K(\mathbf{u}^*) - y\big) - \lambda\mathbf{u}^*, \mathbf{v}\rangle_{\ell_2} + \lambda\|\mathbf{v}\|_{\ell_2}^2 + 2\|\mathbf{v}\|_{\ell_{1,\boldsymbol{\alpha}}}\Big),$$

or equivalently $\mathbf{0} \in \big(\partial\Gamma_{\boldsymbol{\alpha}}\big)(\mathbf{u}^*)$. However, it is known from the linear case, that the algorithm in its most basic form converges rather slowly. Strategies to accelerate the convergence of the method are necessary for its applicability. In [10] three of us considered the case of *linear* operators K and proposed to chose a decreasing thresholding strategy for the parameters $\boldsymbol{\alpha}^{(n)}$. In this $\mathbf{u}^* = \mathbf{u}^*_{\boldsymbol{\alpha}}$ is a global minimizer of (1). Moreover it has been possible to show that the resulting scheme is guaranteed to converge linearly, under spectral conditions of K, the so-called restricted isometry property, see (60) below. Furthermore this property is obtainable for certain classes of operators by means of multilevel preconditioners, we refer to [10] for details. This paper is concerned with the generalization of this strategy to nonlinear operators K. That is, we are interested in the convergence analysis of the iteration

$$\mathbf{u}^{(n+1)} := \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\Big(\mathbf{u}^{(n)} + \frac{1}{\lambda}\big(K'(\mathbf{u}^{(n)})\big)^*\big(y - K(\mathbf{u}^{(n)})\big)\Big), \tag{9}$$

where $\boldsymbol{\alpha}^{(n)} \in \mathbb{R}_+^{\mathcal{J}}$ is an entrywise decreasing sequence with $\lim_{n\to\infty}\boldsymbol{\alpha}^{(n)} = \boldsymbol{\alpha}$ and $\alpha_\mu^{(n)}, \alpha_\mu \geq \alpha \in \mathbb{R}_+, \mu \in \mathcal{J}$.

The basic convergence analysis is outlined in Section 2. Our analysis relies on two fundamental assumptions. We need that the operator K satisfies certain boundedness and Lipschitz continuity conditions, see (22). Moreover, we have to assume that the operator

$$T : \ell_2 \to \ell_2, \quad \mathbf{v} \mapsto T(\mathbf{v}) := \mathbf{v} + \frac{1}{\lambda}\big(K'(\mathbf{v})\big)^*\big(y - K(\mathbf{v})\big)$$

is a contraction on a sufficiently small ball around a critical point $\mathbf{u}^*$ of the functional $\Gamma_{\boldsymbol{\alpha}}$, which will turn out to be the unique global minimizer there. Then the iteration is linearly convergent, i.e.,

$$\|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2} \leq \gamma^n\|\mathbf{u}^*\|_{\ell_2}, \quad \text{for some } \gamma \leq 1.$$

Moreover, the iteration is monotone in the sense that

$$\Gamma_{\boldsymbol{\alpha}^{(n+1)}}(\mathbf{u}^{(n+1)}) < \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}), \tag{10}$$

provided that $\mathbf{u}^{(n)}$ is not a critical point of $\Gamma_{\boldsymbol{\alpha}^{(n)}}$. These properties are specified in Theorem 2.4 which is the main result of this paper.

The local contraction condition (42) on T may be hard to verify. In the Sections 2.2 and 2.3, we discuss in detail two classes of operators where it is satisfied. The first class consists of operators with bounded second derivatives and first derivative that satisfies the contraction property. The second class is given by nonlinear perturbations of linear operators satisfying the restricted isometry property (60). As already shown in [10] for large classes of linear operators $\mathcal{K}$ where (60) fails, it can actually be resumed by preconditioning. Details will be outlined for the case of a nonlinear $\mathcal{K}$ which is mild perturbation of a linear operator in Section 3.

The analysis in this paper is performed in an infinite dimensional setting. In this general setting, clearly the operator K and K' cannot be evaluated exactly. Therefore, in Section 4, we discuss strategies to solve the infinite dimensional problem by turning it into a finite dimensional one and using the expected sparsity of the minimizer. If implementable approximations of the actions of K and K' up to prescribed tolerances are applied, then the resulting inexact, but implementable, version of the algorithm will be again linearly convergent. If the underlying Riesz basis is of wavelet type, then the desired approximations are known in the literature for certain classes of nonlinearities [8, 12, 18].

2 Convergence analysis

In this section we analyze the convergence properties of the iteration (9). As a first step we show that under relatively mild assumptions $\Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)})$ decreases monotonically. It is known that in the case of constant thresholding parameters $\boldsymbol{\alpha}^n = \boldsymbol{\alpha}, n \in \mathbb{N}$, the sequence $\left(\mathbf{u}^{(n)}\right)_{n\in\mathbb{N}}$ has a convergent subsequence and every convergent subsequence converges to a stationary point of (1). However, we are in particular interested in the global minimizer of (1). Therefore, we prove that under more restrictive assumptions and for decreasing thresholing parameters $\boldsymbol{\alpha}^n$ the iterates converge linearly to the global minimizer of (1). In the remainder of this section we present examples of settings where our analysis can be applied. In Section 2.2, we describe how our assumptions can be fulfilled under smoothness conditions on the nonlinear operator K and its derivative. In Section 2.3 we present the important special case where K can be expressed as the sum of a linear operator satisfying the restricted isometry property, and a small nonlinear perturbation.

2.1 A general convergence result

We are particularly interested in computing approximations with the smallest possible number of nonzero entries to solutions of (2). As a benchmark, we recall that the most economical approximations of a given vector $\mathbf{v} \in \ell_2$ are provided by the *best N-term approximations* $\mathbf{v}_N$, defined by discarding in $\mathbf{v}$ all but the $N \in \mathbb{N}_0$ largest coefficients in absolute value. The error of best N-term approximation is defined as

$$\sigma_N(\mathbf{v}) := \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2}. \tag{11}$$

The subspace of all ℓ_2 vectors with best N-term approximation rate $s > 0$, i.e., $\sigma_N(\mathbf{v}) \lesssim N^{-s}$ for some decay rate $s > 0$, is commonly referred to as the *weak ℓ_τ space* $\ell_\tau^w(\mathcal{J})$, for $\tau = (s + \frac{1}{2})^{-1}$, which, endowed with

$$|\mathbf{v}|_{\ell_\tau^w(\mathcal{J})} := \sup_{N \in \mathbb{N}_0} (N+1)^s \sigma_N(\mathbf{v}), \tag{12}$$

becomes the quasi-Banach space $(\ell_\tau^w(\mathcal{J}), |\cdot|_{\ell_\tau^w(\mathcal{J})})$. Moreover, for any $0 < \epsilon \leq 2 - \tau$, we have the continuous embeddings $\ell_\tau(\mathcal{J}) \hookrightarrow \ell_\tau^w(\mathcal{J}) \hookrightarrow \ell_{\tau+\epsilon}(\mathcal{J})$, justifying why $\ell_\tau^w(\mathcal{J})$ is called weak $\ell_\tau(\mathcal{J})$. As before we omit the dependency on the index set $\mathcal{J}$ whenever it is clear from the context.

When it comes to the concrete computations of good approximations with a small number of active coefficients, one frequently utilizes certain thresholding procedures. Here small entries of a given vector are simply discarded, whereas the large entries may be slightly modified. In this paper, we will make use of *soft-thresholding* that we already introduced in (7). It is well-known that $\mathbb{S}_{\boldsymbol{\alpha}}$ is non-expansive for any $\boldsymbol{\alpha} \in \mathbb{R}_+^{\mathcal{J}}$,

Moreover, for any fixed $x \in \mathbb{R}$, the mapping $\beta \mapsto S_\beta(x)$ is Lipschitz continuous with

$$\big|S_\beta(x) - S_{\beta'}(x)\big| \leq |\beta - \beta'|, \quad \text{for all } \beta, \beta' \geq 0. \tag{13}$$

We readily infer the following technical estimate (for the proof we refer the reader to [10]).

Lemma 2.1. *Assume $\mathbf{v} \in \ell_2$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}_+^{\mathcal{J}}$ such that $0 < \alpha = \min\big(\inf_\mu \alpha_\mu, \inf_\mu \beta_\mu\big)$, and define $\Lambda_\alpha(\mathbf{v}) := \big\{\mu \in \mathcal{J} : |v_\mu| > \alpha\big\}$. Then*

$$\big\|\mathbb{S}_{\boldsymbol{\alpha}}(\mathbf{v}) - \mathbb{S}_{\boldsymbol{\beta}}(\mathbf{v})\big\|_{\ell_2} \leq \Big(\#\Lambda_\alpha(\mathbf{v})\Big)^{1/2} \max_{\mu \in \Lambda_\alpha(\mathbf{v})} |\alpha_\mu - \beta_\mu|. \tag{14}$$

In the sequel, we shall also use the following support size estimate, whose proof follows the lines of Lemma 5.1 in [6], more details are provided in [10].

Lemma 2.2. *Let $\mathbf{v} \in \ell_\tau^w$ and $\mathbf{w} \in \ell_2$ with $\|\mathbf{v} - \mathbf{w}\|_{\ell_2} \leq \epsilon$. Assume $\boldsymbol{\alpha} = (\alpha_\mu)_{\mu \in \mathcal{J}} \in \mathbb{R}_+^{\mathcal{J}}$ and $\inf_\mu \alpha_\mu = \alpha > 0$. Then it holds*

$$\#\operatorname{supp} \mathbb{S}_{\boldsymbol{\alpha}}(\mathbf{w}) \leq \#\Lambda_\alpha(\mathbf{w}) \leq \frac{4\epsilon^2}{\alpha^2} + 4C|\mathbf{v}|_{\ell_\tau^w}^\tau \alpha^{-\tau}, \tag{15}$$

where $C = C(\tau) > 0$. In particular if $\mathbf{v} \in \ell_0(\mathcal{J})$ then the estimate is refined

$$\#\operatorname{supp} \mathbb{S}_{\boldsymbol{\alpha}}(\mathbf{w}) \leq \#\Lambda_\alpha(\mathbf{w}) \leq \frac{4\epsilon^2}{\alpha^2} + \|\mathbf{v}\|_{\ell_0(\mathcal{J})}. \tag{16}$$

For the analysis of the iteration (9), we will always assume that the datum $y \in Y$ is fixed and contained in a bounded set, i.e.,

$$\|y\|_Y \leq C_Y < \infty. \tag{17}$$

In this setting, we define the operator

$$T : \ell_2 \to \ell_2, \quad \mathbf{v} \mapsto T(\mathbf{v}) := \mathbf{v} + \frac{1}{\lambda}\big(K'(\mathbf{v})\big)^*\big(y - K(\mathbf{v})\big). \tag{18}$$

In the following we want to show the convergence of the algorithm (9) to stationary points of $\Gamma_{\boldsymbol{\alpha}}$ and to estimate the rate of convergence. In order to do that we shall in particular show that, under certain *local* contraction properties of the operator T, the stationary point is actually unique in a predetermined ball around 0 and coincides with the global minimizer of $\Gamma_{\boldsymbol{\alpha}}$. First of all, we need to characterize the ball where the interesting stationary points should be searched.

To this end, we recall that for all $\boldsymbol{\alpha} \in \mathbb{R}_+^{\mathcal{J}}, \alpha_\mu \geq \alpha > 0, \mu \in \mathcal{J}$, the related energy functional $\Gamma_{\boldsymbol{\alpha}}$, defined by (1) is coercive, i.e., $\Gamma_{\boldsymbol{\alpha}}(\mathbf{v}) \to \infty$ as $\|\mathbf{v}\|_{\ell_2} \to \infty$. In particular this implies that

$$R := \sup\{\|\mathbf{v}\|_{\ell_2}, \Gamma_{\boldsymbol{\alpha}}(\mathbf{v}) \leq \Gamma_{\boldsymbol{\alpha}^{(0)}}(\mathbf{0})\} \tag{19}$$

is finite and we define

$$B(R) := \{\mathbf{u} \in \ell_2, \|\mathbf{u}\|_{\ell_2} \leq R\}. \tag{20}$$

Notice that for $\mathbf{v} \in \ell_2$ such that $\Gamma_{\boldsymbol{\alpha}}(\mathbf{v}) \leq \Gamma_{\boldsymbol{\alpha}^{(0)}}(\mathbf{0})$, we have

$$2\alpha\|\mathbf{v}\|_{\ell_2} \leq 2\|\mathbf{u}\|_{\ell_{1,\boldsymbol{\alpha}}(\mathcal{J})} \leq \Gamma_{\boldsymbol{\alpha}}(\mathbf{v}) \leq \Gamma_{\boldsymbol{\alpha}^{(0)}}(\mathbf{0}),$$

hence,

$$R \leq \frac{\Gamma_{\boldsymbol{\alpha}^{(0)}}(\mathbf{0})}{2\alpha}. \tag{21}$$

For the remainder of this section we will make the following additional standing hypothesis. The operators K and K' are Lipschitz continuous on closed bounded sets, i.e., for all closed and bounded $\mathcal{O} \subset \ell_2$ we assume

$$\begin{aligned} \|K(\mathbf{u}) - K(\mathbf{v})\|_Y &\leq C_K^{Lip}(\mathcal{O})\|\mathbf{u} - \mathbf{v}\|_{\ell_2}, \quad \mathbf{u}, \mathbf{v} \in \mathcal{O}, \\ \|K'(\mathbf{u}) - K'(\mathbf{v})\|_{\mathcal{L}(\ell_2, Y)} &\leq C_{K'}^{Lip}(\mathcal{O})\|\mathbf{u} - \mathbf{v}\|_{\ell_2}, \quad \mathbf{u}, \mathbf{v} \in \mathcal{O}. \end{aligned} \tag{22}$$

With a slight abuse of notation we denote the Lipschitz constants of K and K' on $B(R)$ by $C_K^{Lip}(R)$ and $C_{K'}^{Lip}(R)$, respectively. In particular (22) implies that K and K' are bounded on closed and bounded sets. Indeed, let $\mathcal{O} \subset \ell_2$ and $\mathbf{v}_0 \in \mathcal{O}$. Then we may bound K by estimating

$$\begin{aligned} \sup_{\mathbf{v} \in \mathcal{O}} \|K(\mathbf{v})\|_Y &\leq \sup_{\mathbf{v} \in \mathcal{O}} \|K(\mathbf{v}) - K(\mathbf{v}_0)\|_Y + \|K(\mathbf{v}_0)\|_Y \\ &\leq C_K^{Lip}(\mathcal{O}) \sup_{\mathbf{v} \in \mathcal{O}} \|\mathbf{v} - \mathbf{v}_0\|_{\ell_2} + \|K(\mathbf{v}_0)\|_Y < \infty, \end{aligned} \tag{23}$$

and K' by a similar estimate. We introduce the abbreviations

$$C_K^{bnd}(R) := \sup_{\mathbf{v}\in B(R)} \|K(\mathbf{v})\|_Y, \quad C_{K'}^{bnd}(R) := \sup_{\mathbf{v}\in B(R)} \|K'(\mathbf{v})\|_{\mathcal{L}(\ell_2,Y)}. \tag{24}$$

With these preliminaries, we can formulate the following proposition, which generalizes Lemma 2.4 in [2].

Proposition 2.3. *Suppose that* (17) *and* (22) *hold. For some* $\lambda_0 > 0$ *and* R *as in* (19) *we define*

$$R' := R + \frac{1}{\lambda_0} C_{K'}^{bnd}(R)(C_K^{bnd}(R) + C_Y). \tag{25}$$

Then $\|K(\cdot) - y\|_Y^2$ *is locally Lipschitz. We choose in* (3)

$$\lambda > \lambda_{\min} := \max\{\lambda_0, C_{K'}^{Lip}(R')(C_K^{bnd}(R') + C_Y) + C_{K'}^{bnd}(R')C_K^{Lip}(R')\} \tag{26}$$

and denote by $\left(\mathbf{u}^{(n)}\right)_{n\in\mathbb{N}}$ *the iterates of the decreasing thresholding iteration* (9) *starting from* $\mathbf{u}^{(\mathbf{0})} = \mathbf{0}$. *Then it holds*

$$\Gamma_{\boldsymbol{\alpha}^{(n+1)}}(\mathbf{u}^{(n+1)}) < \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}), \tag{27}$$

provided that $\mathbf{u}^{(n)}$ *is not yet a critical point of* $\Gamma_{\boldsymbol{\alpha}^{(n)}}$. *Furthermore the iterates fulfill the bound*

$$\|\mathbf{u}^{(n)}\|_{\ell_2} \le R, \quad n \in \mathbb{N}. \tag{28}$$

Proof. We shall prove by induction over n that

$$\|\mathbf{u}^{(n)}\|_{\ell_2} \le R \text{ and } \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}) \le \Gamma_{\boldsymbol{\alpha}^{(0)}}(\mathbf{0}). \tag{29}$$

We will show that if λ is chosen according to (26) and $\mathbf{u}^{(n)} \neq \mathbf{u}^{(n+1)}$, which is the case if $\mathbf{u}^{(n)}$ is no critical point of $\Gamma_{\boldsymbol{\alpha}^{(n)}}$, then this implies

$$\Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n+1)}) < \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}). \tag{30}$$

From the fact that $\boldsymbol{\alpha}^{(n)}$ decreases to $\boldsymbol{\alpha}$, together with (30) and (29) we would obtain

$$\Gamma_{\boldsymbol{\alpha}}(\mathbf{u}^{(n+1)}) \le \Gamma_{\boldsymbol{\alpha}^{(n+1)}}(\mathbf{u}^{(n+1)}) \le \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n+1)}) < \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}) \le \Gamma_{\boldsymbol{\alpha}^{(0)}}(\mathbf{0}).$$

By (19) this would also imply the validity of (29) for $n \to n+1$ and simultaneously of (27) and (28) for all $n \in \mathbb{N}$.

Notice that (29) in particular holds for $n = 0$. We begin by proving $\mathbf{u}^{(n+1)} \in B(R')$, where R' is defined in (25). We use the fact that soft shrinkage is nonexpansive, together with (24) and (17) to estimate

$$\begin{aligned} \|\mathbf{u}^{(n+1)}\|_{\ell_2} &\le \|\mathbf{u}^{(n)} + \frac{1}{\lambda}\big(K'(\mathbf{u}^{(n)})\big)^*\big(y - K(\mathbf{u}^{(n)})\big)\|_{\ell_2} \\ &\le \|\mathbf{u}^{(n)}\|_{\ell_2} + \frac{1}{\lambda} C_{K'}^{bnd}(R)(C_K^{bnd}(R) + C_Y) \le R'. \end{aligned} \tag{31}$$

Hence it follows that $\mathbf{u}^{(n)}, \mathbf{u}^{(n+1)} \in B(R')$. To prove (30) we shall use (4), reformulated for $\mathbf{u}^{(n)}$ and $\Gamma_{\boldsymbol{\alpha}}(\mathbf{u}^{(n)})$. In (3) we introduced the splitting $\Gamma_{\boldsymbol{\alpha}}(\mathbf{u}) = \Gamma_\lambda^{(1)}(\mathbf{u}) + \Gamma_{\lambda,\boldsymbol{\alpha}}^{(2)}(\mathbf{u})$, where $\Gamma_\lambda^{(1)}$ is continuously Fréchet differentiable. The derivative of $\Gamma_\lambda^{(1)}$ was already implicitly stated in (4) and may be reformulated by means of T as follows

$$\big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u})\mathbf{v} = 2\langle \big(K'(\mathbf{u})\big)^* \big(K(\mathbf{u}) - y\big) - \lambda \mathbf{u}, \mathbf{v}\rangle_{\ell_2} = -2\lambda \langle T(\mathbf{u}), \mathbf{v}\rangle_{\ell_2}. \quad (32)$$

Recall that by means of (7), (6), and (32), the definition of $\mathbf{u}^{(n+1)}$ in (9) can be reformulated as

$$\begin{aligned}
\mathbf{u}^{(n+1)} &= \arg\min_{\mathbf{v}\in\ell_2} \Big(\|\mathbf{v} - T(\mathbf{u}^{(n)})\|_{\ell_2}^2 + 2\|\mathbf{v}\|_{\ell_{1,\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}} \Big) \\
&= \arg\min_{\mathbf{v}\in\ell_2} \Big(-2\lambda \langle T(\mathbf{u}^{(n)}), \mathbf{v}\rangle_{\ell_2} + \lambda\|\mathbf{v}\|_{\ell_2}^2 + 2\|\mathbf{v}\|_{\ell_{1,\boldsymbol{\alpha}^{(n)}}} \Big) \\
&= \arg\min_{\mathbf{v}\in\ell_2} \Big(\big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)})\mathbf{v} + \Gamma_{\lambda,\boldsymbol{\alpha}^{(n)}}^{(2)}(\mathbf{v}) \Big).
\end{aligned}$$

In particular it follows that

$$\big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)})\mathbf{u}^{(n+1)} + \Gamma_{\lambda,\boldsymbol{\alpha}^{(n)}}^{(2)}(\mathbf{u}^{(n+1)}) \le \big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)})\mathbf{u}^{(n)} + \Gamma_{\lambda,\boldsymbol{\alpha}^{(n)}}^{(2)}(\mathbf{u}^{(n)})$$

holds, which is equivalent to

$$\big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)})(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}) \le \Gamma_{\lambda,\boldsymbol{\alpha}^{(n)}}^{(2)}(\mathbf{u}^{(n)}) - \Gamma_{\lambda,\boldsymbol{\alpha}^{(n)}}^{(2)}(\mathbf{u}^{(n+1)}). \quad (33)$$

Next, we apply the fundamental theorem of calculus to $\Gamma_\lambda^{(1)}$ and write

$$\begin{aligned}
&\Gamma_\lambda^{(1)}(\mathbf{u}^{(n+1)}) - \Gamma_\lambda^{(1)}(\mathbf{u}^{(n)}) \\
&\quad = \int_0^1 \big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)} + \tau(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}))(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})\mathrm{d}\tau \\
&\quad = \int_0^1 \Big(\big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)} + \tau(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})) - \big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)}) \Big)(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})\mathrm{d}\tau \\
&\qquad + \big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)})(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}).
\end{aligned}$$

This, together with (33) and (32) yields

$$\begin{aligned}
&\Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n+1)}) - \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}) \\
&\quad = \Big(\Gamma_\lambda^{(1)}(\mathbf{u}^{(n+1)}) + \Gamma_{\lambda,\boldsymbol{\alpha}^{(n)}}^{(2)}(\mathbf{u}^{(n+1)}) \Big) - \Big(\Gamma_\lambda^{(1)}(\mathbf{u}^{(n)}) + \Gamma_{\lambda,\boldsymbol{\alpha}^{(n)}}^{(2)}(\mathbf{u}^{(n)}) \Big) \\
&\quad \le \int_0^1 \Big(\big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)} + \tau(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})) - \big(\Gamma_\lambda^{(1)}\big)'(\mathbf{u}^{(n)}) \Big)(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})\mathrm{d}\tau \\
&\quad = \int_0^1 2\langle \big(K'(\mathbf{u}^{(n)} + \tau(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}))\big)^* \big(K(\mathbf{u}^{(n)} + \tau(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})) - y\big) \\
&\quad - \big(K'(\mathbf{u}^{(n)})\big)^* \big(K(\mathbf{u}^{(n)}) - y\big),\ (\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})\rangle_{\ell_2}\mathrm{d}\tau - \lambda\|\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}\|_{\ell_2}^2.
\end{aligned} \quad (34)$$

Moreover, by the assumptions (22) on K and K', we can estimate for $\mathbf{u}, \mathbf{v} \in B(R')$

$$\begin{aligned}
&\|\big(K'(\mathbf{u})\big)^*(K(\mathbf{u}) - y) - \big(K'(\mathbf{v})\big)^*(K(\mathbf{v}) - y)\|_{\ell_2} \\
&\quad = \|\Big(\big(K'(\mathbf{u})\big)^* - \big(K'(\mathbf{v})\big)^*\Big)(K(\mathbf{u}) - y) + \big(K'(\mathbf{v})\big)^*\big(K(\mathbf{u}) - K(\mathbf{v})\big)\|_{\ell_2} \\
&\quad \leq \lambda_{\min}\|\mathbf{u} - \mathbf{v}\|_{\ell_2}.
\end{aligned}$$

We apply this inequality for the special case $\mathbf{u} = \mathbf{u}^{(n)} + \tau(\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})$, $\mathbf{v} = \mathbf{u}^{(n)}$, to further estimate (34) as follows

$$\begin{aligned}
&\Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n+1)}) - \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}) \\
&\quad \leq \int_0^1 2\tau\lambda_{\min}\|\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}\|_{\ell_2}^2 \mathrm{d}\tau - \lambda\|\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}\|_{\ell_2}^2 \\
&\quad = (\lambda_{\min} - \lambda)\|\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)}\|_{\ell_2}^2.
\end{aligned} \tag{35}$$

Furthermore the right-hand side is negative if λ is chosen according to (26) and $\mathbf{u}^{(n)} \neq \mathbf{u}^{(n+1)}$, which is the case if $\mathbf{u}^{(n)}$ is no critical point of $\Gamma_{\boldsymbol{\alpha}^{(n)}}$, and this shows (30) and concludes the proof. $\square$

Notice that we decided to start our iteration from $\mathbf{u}^{(\mathbf{0})} = \mathbf{0}$. On the one hand, this choice is motivated by the fact that a priori we do not dispose of any information on potentially interesting stationary points and an arbitrary choice of the initial iteration has to be made. On the other hand, as we will show below, under certain assumptions, we will be able to identify in this way the unique global minimizer of the functional $\Gamma_{\boldsymbol{\alpha}}$. As we are seeking for stationary points which are limits of the sequence $\big(\mathbf{u}^{(n)}\big)_{n\in\mathbb{N}}$ of the iterates of the decreasing thresholding iteration (9) starting from $\mathbf{u}^{(\mathbf{0})} = \mathbf{0}$, in view of Proposition 2.3 we can assume without loss of generality that interesting stationary points $\mathbf{u}^*$ belongs to the ball $B(R)$. This assumption is not void, because a global minimizer $\mathbf{u}^\circ$ of $\Gamma_{\boldsymbol{\alpha}}$ necessarily has to lay in the ball $B(R)$, because $\Gamma_{\boldsymbol{\alpha}}(\mathbf{u}^\circ) < \Gamma_{\boldsymbol{\alpha}}(\mathbf{0})$. We shall also show below that all the iterates $\big(\mathbf{u}^{(n)}\big)_{n\in\mathbb{N}}$ are actually additionally confined within the ball

$$\mathcal{B} := \{\mathbf{v} \in \ell_2 : \|\mathbf{u}^* - \mathbf{v}\|_{\ell_2} \leq \|\mathbf{u}^*\|_{\ell_2}\}, \tag{36}$$

where $\mathbf{u}^*$ is an arbitrary stationary point of $\Gamma_{\boldsymbol{\alpha}}$ within $B(R)$. Hence, under the regularity assumption so far made for the operators K and K', the reference domain of the iterations of the algorithm is $\mathcal{B} \cap B(R)$. Within this setting we assume the following hypothesis: T satisfies the Lipschitz condition

$$\|T(\mathbf{u}^*) - T(\mathbf{v})\|_{\ell_2} \leq C_T^{Lip}\|\mathbf{u}^* - \mathbf{v}\|_{\ell_2}, \quad \mathbf{v} \in \mathcal{B} \cap B(R),\ y \in Y,\ \|y\|_Y \leq C_Y, \tag{37}$$

for any fixed stationary point $\mathbf{u}^*$. (As we shall see below, such a condition is not so strong as we shall apply it to only *one* stationary point.) Furthermore we define for some fixed $\lambda_0 > 0$ the analogue of (25) on $\mathcal{B} \cap B(R)$, that is

$$R'' := R + \frac{1}{\lambda_0} C_{K'}^{bnd}(\mathcal{B} \cap B(R))(C_K^{bnd}(\mathcal{B} \cap B(R)) + C_Y). \tag{38}$$

Then, the following convergence theorem holds.

Theorem 2.4. *Let $\mathbf{u}^*$ be a stationary point of (1) that satisfies $T(\mathbf{u}^*) \in \ell_\tau^w$ for some $0 < \tau < 2$. For some $\lambda_0 > 0$ and R'' as in (38) we choose*

$$\begin{aligned}\lambda > \max\{\lambda_0, \big(C_{K'}^{Lip}(\mathcal{B} \cap B(R'')(C_K^{bnd}(\mathcal{B} \cap B(R'')) + C_Y) \\ + C_{K'}^{bnd}(\mathcal{B} \cap B(R''))C_K^{Lip}(\mathcal{B} \cap B(R''))\big)\}.\end{aligned} \tag{39}$$

Furthermore let $\boldsymbol{\alpha}^{(n)}, \boldsymbol{\alpha} \in \mathbb{R}_+^{\mathcal{J}}$ with $\alpha_\mu^{(n)} \geq \alpha_\mu \geq \alpha \in \mathbb{R}_+, \mu \in \mathcal{J}$. We set

$$L := \frac{4(C_T^{Lip})^2 \|\mathbf{u}^*\|_{\ell_2}^2 \lambda^2}{\alpha^2} + 4C|T(\mathbf{u}^*)|_{\ell_\tau^w}^\tau \Big(\frac{\alpha}{\lambda}\Big)^{-\tau}, \tag{40}$$

with C is as in Lemma 2.2 and C_T^{Lip} as in (37). Moreover we define the set

$$\mathcal{B}_L := \{\mathbf{v} \in \ell_2 : \|\mathbf{u}^* - \mathbf{v}\|_{\ell_2} \leq \|\mathbf{u}^*\|_{\ell_2}, \# \operatorname{supp} \mathbf{v} \leq L\}. \tag{41}$$

Let us assume that there exists $0 < \gamma_0 < 1$, such that for all $\mathbf{v} \in \mathcal{B}_L$ and $\operatorname{supp}(\mathbf{v}) \subset \Lambda \subset \mathcal{J}$ *with $\#\Lambda \leq 2L$*

$$\|\big(T(\mathbf{u}^*) - T(\mathbf{v})\big)_{|S^* \cup \Lambda}\|_{\ell_2(S^* \cup \Lambda)} \leq \gamma_0 \|\mathbf{u}^* - \mathbf{v}\|_{\ell_2}, \tag{42}$$

where $S^* := \operatorname{supp} \mathbf{u}^*$. Then, for any $\gamma_0 < \gamma < 1$, the sequence $\big(\mathbf{u}^{(n)}\big)_{n \in \mathbb{N}}$ obtained by (9) fulfills

$$\big(\mathbf{u}^{(n)}\big)_{n \in \mathbb{N}} \subset \mathcal{B}_L \cap B(R) \tag{43}$$

and converges to $\mathbf{u}^*$ at a linear rate

$$\|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2} \leq \epsilon^{(n)} := \gamma^n \|\mathbf{u}^*\|_{\ell_2}, \tag{44}$$

whenever the $\boldsymbol{\alpha}^{(n)}$ are chosen according to

$$\max_{\mu \in \mathcal{J}} |\alpha_\mu^{(n)} - \alpha_\mu| \leq \lambda L^{-\frac{1}{2}}(\gamma - \gamma_0)\epsilon^{(n)}. \tag{45}$$

Moreover, the iteration is monotone

$$\Gamma_{\boldsymbol{\alpha}^{(n+1)}}(\mathbf{u}^{(n+1)}) < \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}),$$

provided that $\mathbf{u}^{(n)}$ is not yet a critical point of $\Gamma_{\boldsymbol{\alpha}^{(n)}}$.

Remark 2.5. Before proving this result, let us comment the following fundamental implication: under the assumptions of Lipschitzianity (37) and local contraction property (42), the iterations $\left(\mathbf{u}^{(n)}\right)_{n\in\mathbb{N}}$ of the algorithm starting from $\mathbf{u}^{(\mathbf{0})} = \mathbf{0}$ must converge to any stationary point $\mathbf{u}^* \in B(R)$, hence implying automatically its uniqueness! In fact, if there were another stationary point, it would also coincide with the limit of this sequence. In particular, the global minimizer $\mathbf{u}^\circ$ of $\Gamma_{\boldsymbol{\alpha}}$ necessarily lies in the ball $B(R)$ and is a stationary point of $\Gamma_{\boldsymbol{\alpha}}$, and we have linear convergence of the iterates to $\mathbf{u}^\circ$. Let us now address the proof of Theorem 2.4.

Proof. The proof is performed by induction over n. There is nothing to show for $n = 0$. The first step is to prove that $\mathbf{u}^{(n+1)}$ is indeed contained in $\mathcal{B}_L$. Let $\mathbf{u}^{(n)} \in \mathcal{B}_L \cap B(R)$, then since $\boldsymbol{\alpha}^{(n)}$ is decreasing to $\boldsymbol{\alpha}$ it holds that

$$\operatorname{supp} \mathbf{u}^{(n+1)} = \operatorname{supp} \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(T(\mathbf{u}^{(n)})\big) \subset \operatorname{supp} \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\mathbf{u}^{(n)})\big). \tag{46}$$

The Lipschitz property (37) implies that

$$\|T(\mathbf{u}^*) - T(\mathbf{u}^{(n)})\|_{\ell_2} \leq C_T^{Lip}\|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2}.$$

By Lemma 2.2 we can estimate

$$\begin{aligned} \# \operatorname{supp} \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\mathbf{u}^{(n)})\big) &\leq \Lambda_{\frac{\alpha}{\lambda}}\big(T(\mathbf{u}^{(n)})\big) \\ &\leq \frac{4(C_T^{Lip})^2\|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2}^2\lambda^2}{\alpha^2} + 4C|T(\mathbf{u}^*)|_{\ell_\tau^w}\left(\frac{\alpha}{\lambda}\right)^{-\tau} \leq L. \end{aligned} \tag{47}$$

We conclude that $\# \operatorname{supp} \mathbf{u}^{(n+1)} \leq L$. Let us denote $S^{(n)} = \operatorname{supp} \mathbf{u}^{(n)}$, $S^* = \operatorname{supp} \mathbf{u}^*$, and $\Lambda^{(n)} = S^* \cup S^{(n)} \cup S^{(n+1)}$. Notice that $\#S^{(n)} \cup S^{(n+1)} \leq 2L$. By the thresholding properties it is clear that after restriction to $\Lambda^{(n)}$

$$\mathbf{u}^*_{|\Lambda^{(n)}} = \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}(T(\mathbf{u}^*)_{|\Lambda^{(n)}}), \tag{48}$$

and

$$\mathbf{u}^{(n+1)}_{|\Lambda^{(n)}} = \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}(T(\mathbf{u}^{(n)})_{|\Lambda^{(n)}}) \tag{49}$$

hold. This, together with the nonexpansiveness of the soft-thresholding, Lemma 2.1 in the second inequality, and (42) together with $\# \operatorname{supp} \mathbf{u}^{(n+1)} \leq$

L in the third inequality yields

$$
\begin{aligned}
&\|\mathbf{u}^* - \mathbf{u}^{(n+1)}\|_{\ell_2} \\
&\quad = \|\mathbf{u}^*_{|\Lambda^{(n)}} - \mathbf{u}^{(n+1)}_{|\Lambda^{(n)}}\|_{\ell_2(\Lambda^{(n)})} \\
&\quad = \|\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}(T(\mathbf{u}^*)_{|\Lambda^{(n)}}) - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}(T(\mathbf{u}^{(n)})_{|\Lambda^{(n)}})\|_{\ell_2(\Lambda^{(n)})} \\
&\quad \leq \|\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}(T(\mathbf{u}^*)_{|\Lambda^{(n)}}) - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}(T(\mathbf{u}^{(n)})_{|\Lambda^{(n)}})\|_{\ell_2(\Lambda^{(n)})} \\
&\qquad\quad + \|\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}(T(\mathbf{u}^{(n)})_{|\Lambda^{(n)}}) - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}(T(\mathbf{u}^{(n)})_{|\Lambda^{(n)}})\|_{\ell_2(\Lambda^{(n)})} \\
&\quad \leq \|T(\mathbf{u}^*)_{|\Lambda^{(n)}} - T(\mathbf{u}^{(n)})_{|\Lambda^{(n)}}\|_{\ell_2(\Lambda^{(n)})} \\
&\qquad\quad + \frac{\big(\#\Lambda_{\frac{\alpha}{\lambda}}(T(\mathbf{u}^{(n)}))\big)^{1/2}}{\lambda}\big(\max_{\mu\in\Lambda_{\frac{\alpha}{\lambda}}(T(\mathbf{u}^{(n)}))} |\alpha_\mu - \alpha_\mu^{(n)}|\big) \\
&\quad \leq \gamma_0\|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2} + \frac{L^{1/2}}{\lambda}\big(\max_{\mu\in\Lambda_\alpha(T(\mathbf{u}^{(n)}))} |\alpha_\mu - \alpha_\mu^{(n)}|\big) \\
&\quad \leq \gamma_0\epsilon^{(n)} + (\gamma-\gamma_0)\epsilon^{(n)} = \gamma\epsilon^{(n)} = \epsilon^{(n+1)}.
\end{aligned}
$$

The last inequality is a consequence of induction hypothesis and (45). This proves $\mathbf{u}^{(n+1)} \in \mathcal{B}_L$. Obviously $\mathbf{u}^{(n+1)} \in B(R)$ because of the monotonicity of the iterations:

$$\Gamma_{\boldsymbol{\alpha}}(\mathbf{u}^{(n+1)}) \leq \Gamma_{\boldsymbol{\alpha}^{(n+1)}}(\mathbf{u}^{(n+1)}) \leq \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n+1)}) < \Gamma_{\boldsymbol{\alpha}^{(n)}}(\mathbf{u}^{(n)}) \leq \Gamma_{\boldsymbol{\alpha}^{(0)}}(\mathbf{0}).$$

□

2.2 Nonlinear operators with bounded second derivatives

In this section we state smoothness conditions on the nonlinear operator K which imply that the operator T defined in (18) fulfills (37) and (42). In the following we assume that S^* is the support of a global minimizer $\mathbf{u}^*$ of $\Gamma_{\boldsymbol{\alpha}}$ in $B(R)$. As discussed above, once we prove that T fulfills (37) and (42), then by Theorem 2.4 we automatically have that $\mathbf{u}^*$ is actually the unique stationary point of $\Gamma_{\boldsymbol{\alpha}}$ in $B(R)$.

Theorem 2.6. *Let the data fulfill assumption* (17)*. Assume that K is twice continuously differentiable on an open set that contains $\mathcal{B}$ and, together with its derivative K', is bounded on $\mathcal{B}$. Furthermore, assume that there exist $0 < \gamma_2 \leq \gamma_1 < 1$ such that for all $\Lambda \subset \mathcal{J}, \#\Lambda \leq 2L$ and $\boldsymbol{\zeta} \in \mathcal{B}, \operatorname{supp}\boldsymbol{\zeta} \subset S^* \cup \Lambda$, the following local contraction property*

$$\|\Big(\mathrm{Id} - \frac{1}{\lambda}\big(K'(\boldsymbol{\zeta})\big)^* K'(\boldsymbol{\zeta})\Big)_{|S^*\cup\Lambda\times S^*\cup\Lambda}\|_{\mathcal{L}(\ell_2(S^*\cup\Lambda),\ell_2(S^*\cup\Lambda))} \leq \gamma_2 \qquad (50)$$

holds. Moreover, let us assume that the uniform spectral gap condition

$$\|\Big(\frac{1}{\lambda}\big(K''(\boldsymbol{\zeta})(\cdot)\big)^*\big(y - K(\boldsymbol{\zeta})\big)\Big)_{|S^*\cup\Lambda\times S^*\cup\Lambda}\|_{\mathcal{L}(\ell_2(S^*\cup\Lambda),\ell_2(S^*\cup\Lambda))} \leq \gamma_1 - \gamma_2 \quad (51)$$

holds. Then T *defined in* (18) *fulfills* (37) *and* (42)*.*

Proof. The proof is an application of the mean value theorem. In order to compute the derivative of T we introduce the auxiliary operator

$$G : (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} + \frac{1}{\lambda}\big(K'(\mathbf{u})\big)^*\big(y - K(\mathbf{v})\big), \quad (\mathbf{u}, \mathbf{v}) \in \ell_2 \times \ell_2$$

and observe $T = G \circ (\mathrm{Id}, \mathrm{Id})^\intercal$. We compute

$$\begin{aligned} T'(\boldsymbol{\zeta})\,\mathbf{z} &= \Big(\frac{\partial G}{\partial \mathbf{u}}, \frac{\partial G}{\partial \mathbf{v}}\Big)\big((\boldsymbol{\zeta}, \boldsymbol{\zeta})\big) \circ \big(\,\mathrm{Id}, \mathrm{Id}\,\big)^\intercal \mathbf{z} \\ &= \Big(\,\mathrm{Id} + \frac{1}{\lambda}\big(K''(\boldsymbol{\zeta})(\cdot)\big)^*\big(y - K(\boldsymbol{\zeta})\big), -\frac{1}{\lambda}\big(K'(\boldsymbol{\zeta})\big)^* K'(\boldsymbol{\zeta})(\cdot)\Big) \circ \big(\,\mathrm{Id}, \mathrm{Id}\,\big)^\intercal \mathbf{z} \\ &= \mathbf{z} + \frac{1}{\lambda}\big(K''(\boldsymbol{\zeta})\,\mathbf{z}\big)^*\big(y - K(\boldsymbol{\zeta})\big) - \frac{1}{\lambda}\big(K'(\boldsymbol{\zeta})\big)^* K'(\boldsymbol{\zeta})\,\mathbf{z}. \end{aligned} \tag{52}$$

Observe that $K : \ell_2 \to Y$, $K' : \ell_2 \to \mathcal{L}(\ell_2, Y)$, and $K'' : \ell_2 \to \mathcal{L}(\ell_2, \mathcal{L}(\ell_2, Y))$. Therefore $K''(\boldsymbol{\zeta})\,\mathbf{z} \in \mathcal{L}(\ell_2, Y)$ holds. Consequently $\big(K''(\boldsymbol{\zeta})\,\mathbf{z}\big)^* \in \mathcal{L}(Y, \ell_2)$, so that the composition in (52) is well defined. By our assumptions K, K', and K'' are bounded on the bounded set $\mathcal{B}$. This, together with (17) implies $\sup_{\xi \in \mathcal{B}} \|T'(\xi)\|_{\mathcal{L}(\ell_2, \ell_2)} < \infty$. Since $\mathcal{B}$ is convex we can use the mean value theorem to conclude the Lipschitz property (37). In order to prove (42) let $\mathbf{v} \in \mathcal{B}_L$ and $\mathrm{supp}(\mathbf{v}) \subset \Lambda \subset \mathcal{J}$ with $\#\Lambda \leq 2L$. Then, by (52), (50), and (51) the estimate

$$\begin{aligned} &\|\big(T'(\boldsymbol{\zeta})\big)_{|S^* \cup \Lambda \times S^* \cup \Lambda}\|_{\mathcal{L}(\ell_2(S^* \cup \Lambda), \ell_2(S^* \cup \Lambda))} \\ &\quad \leq \|\Big(\,\mathrm{Id} - \frac{1}{\lambda}\big(K'(\boldsymbol{\zeta})\big)^* K'(\boldsymbol{\zeta})\Big)_{|S^* \cup \Lambda \times S^* \cup \Lambda}\|_{\mathcal{L}(\ell_2(S^* \cup \Lambda), \ell_2(S^* \cup \Lambda))} \\ &\quad + \|\Big(\frac{1}{\lambda}\big(K''(\boldsymbol{\zeta})(\cdot)\big)^*\big(y - K(\boldsymbol{\zeta})\big)\Big)_{|S^* \cup \Lambda \times S^* \cup \Lambda}\|_{\mathcal{L}(\ell_2(S^* \cup \Lambda), \ell_2(S^* \cup \Lambda))} \\ &\quad \leq \gamma_1 \end{aligned}$$

holds. The restriction $\mathcal{B}_{|S^* \cup \Lambda} := \{\mathbf{u}_{|S^* \cup \Lambda}, \mathbf{u} \in \mathcal{B}\}$ of $\mathcal{B}$ onto the index set $S^* \cup \Lambda$ is a convex set in $\ell_2(S^* \cup \Lambda)$. Hence, we can apply the mean value theorem again to finalize the proof as follows

$$\begin{aligned} &\|\big(T(\mathbf{u}^*) - T(\mathbf{v})\big)_{|S^* \cup \Lambda}\|_{\ell_2(S^* \cup \Lambda)} \\ &\leq \sup_{\boldsymbol{\zeta} \in \mathcal{B}_{|S^* \cup \Lambda}} \|\big(T'(\boldsymbol{\zeta})\big)_{|S^* \cup \Lambda \times S^* \cup \Lambda}\|_{\mathcal{L}(\ell_2(S^* \cup \Lambda), \ell_2(S^* \cup \Lambda))} \|\mathbf{u}^* - \mathbf{v}\|_{\ell_2} \\ &\leq \gamma_1 \|\mathbf{u}^* - \mathbf{v}\|_{\ell_2}. \end{aligned}$$

□

2.3 Nonlinear perturbation of linear operators

In this section we discuss the validity of (42) for the special case that K is given by the sum of a linear operator and a nonlinear perturbation. To be specific, we consider

$$K_\sigma = A + \sigma N, \tag{53}$$

where $\sigma \in \mathbb{R}_+$, $A \in \mathcal{L}(\ell_2, Y)$ and $N : \ell_2 \to Y$ is a nonlinear perturbation with the property that

$$\begin{aligned} N : \ell_2 &\to Y, & \quad\text{and}\quad N' : \ell_2 &\to \mathcal{L}(\ell_2, Y), \\ \mathbf{v} &\mapsto N(\mathbf{v}), & \mathbf{v} &\mapsto N'(\mathbf{v}) \end{aligned} \tag{54}$$

are Lipschitz continuous on closed bounded sets. Similarly to (22) and (24) we denote the respective Lipschitz constants and suprema on $B(R)$ with $C_N^{Lip}(R)$, $C_{N'}^{Lip}(R)$, $C_N^{bnd}(R)$, and $C_{N'}^{bnd}(R)$.

We begin by deriving uniform bounds for those constants. We denote

$$R(\sigma) = \sup\{\|\mathbf{v}\|_{\ell_2}, \Gamma_{\boldsymbol{\alpha},\sigma}(\mathbf{v}) \le \Gamma_{\boldsymbol{\alpha}^{(0)},\sigma}(\mathbf{0})\},$$

where $\Gamma_{\boldsymbol{\alpha},\sigma}$ is the functional $\Gamma_{\boldsymbol{\alpha}}$ for $K = K_\sigma$ depending on σ. Accordingly we denote

$$R'(\sigma) := R(\sigma) + \frac{1}{\lambda_0} C_{K'}^{bnd}(R(\sigma))(C_K^{bnd}(R(\sigma)) + C_Y).$$

We denote with $\mathcal{C}(\sigma)$ the set of critical points of $\Gamma_{\boldsymbol{\alpha},\sigma}$ in $B(R(\sigma))$. For any $\mathbf{u}^*(\sigma) \in \mathcal{C}(\sigma)$ we denote

$$L(\mathbf{u}^*(\sigma)) := \frac{4(C_T^{Lip})^2 \|\mathbf{u}^*(\sigma)\|_{\ell_2}^2 \lambda^2}{\alpha^2} + 4C|T(\mathbf{u}^*(\sigma))|_{\ell_\tau^w}^\tau \left(\frac{\alpha}{\lambda}\right)^{-\tau}. \tag{55}$$

Lemma 2.7. *Let the data fulfill assumption* (17). *Further let* $\sigma_0 \in \mathbb{R}_+$ *and* K_σ, $\sigma \in [0, \sigma_0]$, *be of the form* (53). *Suppose that the assumptions* (54) *hold. Then it holds that*

$$\begin{aligned} R_0 := \sup_{\sigma\in[0,\sigma_0]} R(\sigma) &< \infty, \\ R_0' := \sup_{\sigma\in[0,\sigma_0]} R'(\sigma) &< \infty, \\ \sup_{\sigma\in[0,\sigma_0]} C_{K_\sigma}^{Lip}(R'(\sigma)) &< \infty, \\ \sup_{\sigma\in[0,\sigma_0]} C_{K_\sigma'}^{Lip}(R'(\sigma)) &< \infty. \end{aligned} \tag{56}$$

Under the additional assumption

$$\sup_{\sigma\in[0,\sigma_0]} \sup_{\mathbf{u}^*(\sigma)\in\mathcal{C}(\sigma)} |T_\sigma(\mathbf{u}^*(\sigma))|_{\ell_\tau^w} < \infty \tag{57}$$

it further holds that

$$L_0 := \sup_{\sigma\in[0,\sigma_0]} \sup_{\mathbf{u}^*(\sigma)\in\mathcal{C}(\sigma)} L(\mathbf{u}^*(\sigma)) < \infty. \tag{58}$$

Remark 2.8. Before proving this lemma, let us comment on the condition (57), requiring to consider a supremum over the set $\mathcal{C}(\sigma)$, which a priori can be very large. As we will show later, the boundedness of the quantities of this lemma and an additional spectral properties on the operator A, the so-called *restricted isometry property*, see formula (60) below, will imply the operator T to fulfill (42). As already stated above, under this condition, the set $\mathcal{C}(\sigma)$ consists only of one point, i.e., the global minimizer of $\Gamma_{\boldsymbol{\alpha},\sigma}$. Hence the condition (57) will turn out to be much less restrictive as it seems at a first glance. Let us now prove the lemma.

Proof. It is immediate to see that (21) implies $R(\sigma) \leq \frac{\sup_{\sigma'\in[0,\sigma_0]}\left(\Gamma_{\boldsymbol{\alpha}^{(0)},\sigma'}(\mathbf{0})\right)}{2\alpha}$. Furthermore the term $\sup_{\sigma'\in[0,\sigma_0]}\left(\Gamma_{\boldsymbol{\alpha}^{(0)},\sigma'}(\mathbf{0})\right)$ is finite as

$$\sigma' \mapsto \Gamma_{\boldsymbol{\alpha}^{(0)},\sigma'}(\mathbf{0}) = \|\sigma' N(\mathbf{0}) - y\|_Y^2,$$

is continuous and bounded on $[0,\sigma_0]$, because of the assumptions (17) and (54). Hence, we conclude the boundedness of R_0 in (56). By the assumption (54) we may bound the Lipschitz constant of K_σ on $B(R_\sigma)$ as follows

$$C_{K_\sigma}^{Lip}(R(\sigma)) \leq \|A\|_{\mathcal{L}(\ell_2,Y)} + \sigma_0 C_N^{Lip}(R_0). \tag{59}$$

The constant $C_{K'_\sigma}^{Lip}(R(\sigma))$ may be bounded analogously. By the same reasoning as in (23) it follows that $C_{K_\sigma}^{bnd}(R(\sigma))$ and $C_{K'_\sigma}^{bnd}(R(\sigma))$ may be bounded independently of $\sigma \in [0,\sigma_0]$. This proves the existence of uniform bounds for $C_{K_\sigma}^{bnd}(R(\sigma))$ and $C_{K'_\sigma}^{bnd}(R(\sigma))$, and consequently of R'_0 in (56). Using assumption (54) on $B(R'_0)$ allows us to estimate similarly to (59) uniform bounds for $C_{K_\sigma}^{Lip}(R'(\sigma))$ and $C_{K'_\sigma}^{Lip}(R'(\sigma))$. It remains to prove the finiteness of L_0 in (58). To this end observe that the Lipschitz property of K and K' on $B(R_0)$ imply that T_σ, defined in (18), is Lipschitz on $B(R_0)$ and that the corresponding Lipschitz constants may be uniformly bounded in σ. The remaining terms in (55) are bounded uniformly in σ by assumption (57) and the estimate $\|\mathbf{u}^*(\sigma)\|_{\ell_2} \leq R(\sigma) \leq R_0$. □

We are now able to state conditions under which the fundamental contraction property (42) of the operators T_σ, defined by (18) can be ensured uniformly in σ for σ_0 sufficiently small.

Lemma 2.9. *Let the assumptions of Lemma 2.7 hold. Fix $\sigma_0 \in \mathbb{R}^+$. For all $\sigma \in [0,\sigma_0]$, we fix $\mathbf{u}^*(\sigma) \in \mathcal{C}(\sigma)$ and denote $S_\sigma^* := \operatorname{supp}\mathbf{u}^*(\sigma)$. We make the assumption that the linear part A of K_σ fulfills the restricted isometry property*

$$\|(\mathrm{Id} - A^*A)_{|\Lambda^\circ\times\Lambda^\circ}\|_{\mathcal{L}(\ell_2(\Lambda^\circ),\ell_2(\Lambda^\circ))} \leq \gamma_1 < 1, \tag{60}$$

for all $\Lambda^\circ \subset \mathcal{J}$ with $\#\Lambda^\circ \leq 3L_0$. By using the notations as in (56)*, the constant*

$$\begin{aligned} C :=& \|A\|_{\mathcal{L}(\ell_2,Y)} C_N^{Lip}(R_0) \\ &+ C_{N'}^{Lip}(R_0)\big(\|A\|_{\mathcal{L}(\ell_2,Y)} R_0 + \sigma_0 C_N^{bnd}(R_0) + C_Y\big) \\ &+ C_{N'}^{bnd}(R_0)\big(\|A\|_{\mathcal{L}(\ell_2,Y)} + \sigma_0 C_N^{Lip}(R_0)\big) \end{aligned} \tag{61}$$

is bounded and for all $\sigma \in [0, \min(\sigma_0, (1-\gamma_1)C^{-1}))$ the following holds: For all $\mathbf{v} \in B(R_0)$ with $\#\operatorname{supp}\mathbf{v} \leq L_0$, and $\operatorname{supp}(\mathbf{v}) \subset \Lambda \subset \mathcal{J}$ *with $\#\Lambda \leq 2L_0$, the contraction property*

$$\|\big(T_\sigma(\mathbf{u}^*(\sigma)) - T_\sigma(\mathbf{v})\big)_{|S_\sigma^*\cup\Lambda}\|_{\ell_2(S_\sigma^*\cup\Lambda)} \leq \gamma_0 \|\mathbf{u}^* - \mathbf{v}\|_{\ell_2}, \tag{62}$$

holds with $\gamma_0 := \gamma_1 + \sigma C < 1$.

Proof. We begin by proving that the constant C in (61) is bounded. To this end we apply Lemma 2.7 and observe that the Lipschitz property of N and N' on $B(R_0)$ implies similarly to (23) that N and N' are also bounded on $B(R_0)$.

Now fix $\sigma \in [0, \sigma_0]$ and let $\mathbf{v} \in B(R_0)$, $\#\operatorname{supp}\mathbf{v} \leq L_0$ and $\operatorname{supp}\mathbf{v} \subset \Lambda \subset \mathcal{J}$ with $\#\Lambda \leq 2L_0$ and denote $\Lambda^\circ := S_\sigma^* \cup \Lambda$.

We use the splitting

$$\begin{aligned} T_\sigma(\mathbf{v}) - T_\sigma(\mathbf{u}^*(\sigma)) = \mathbf{v} - \mathbf{u}^*(\sigma) &- A^*A(\mathbf{v} - \mathbf{u}^*(\sigma)) - \sigma A^*\big(N(\mathbf{v}) - N(\mathbf{u}^*(\sigma))\big) \\ &- \sigma\big(N'(\mathbf{v}) - N'(\mathbf{u}^*(\sigma))\big)^*\big((A+\sigma N)(\mathbf{v}) - y\big) \\ &- \sigma\big(N'(\mathbf{u}^*(\sigma))\big)^*\big((A+\sigma N)(\mathbf{v}) - (A+\sigma N)(\mathbf{u}^*(\sigma))\big) \end{aligned}$$

together with the assumption (60) to estimate

$$\begin{aligned} &\|\big(T_\sigma(\mathbf{v}) - T_\sigma(\mathbf{u}^*(\sigma))\big)_{|\Lambda^\circ}\|_{\ell_2(\Lambda^\circ)} \\ &\quad\leq \|\big((\mathrm{Id} - A^*A)(\mathbf{v} - \mathbf{u}^*(\sigma))\big)_{|\Lambda^\circ}\|_{\ell_2(\Lambda^\circ)} + \sigma\Big(\|\Big(A^*\big(N(\mathbf{v}) - N(\mathbf{u}^*(\sigma))\big)\Big)_{|\Lambda^\circ}\|_{\ell_2(\Lambda^\circ)} \\ &\quad+ \|\Big(\big(N'(\mathbf{v}) - N'(\mathbf{u}^*(\sigma))\big)^*\big((A+\sigma N)(\mathbf{v}) - y\big)\Big)_{|\Lambda^\circ}\|_{\ell_2(\Lambda^\circ)} \\ &\quad+ \|\Big(\big(N'(\mathbf{u}^*(\sigma))\big)^*\big((A+\sigma N)(\mathbf{v}) - (A+\sigma N)(\mathbf{u}^*(\sigma))\big)\Big)_{|\Lambda^\circ}\|_{\ell_2(\Lambda^\circ)}\Big) \\ &\quad\leq \Big(\gamma_1 + \sigma\big(\|A\|_{\mathcal{L}(\ell_2,Y)} C_N^{Lip}(R_0) \\ &\quad+ C_{N'}^{Lip}(R_0)\big(\|A\|_{\mathcal{L}(\ell_2,Y)} R_0 + \sigma_0 C_N^{bnd}(R_0) + C_Y\big) \\ &\quad+ C_{N'}^{bnd}(R_0)\big(\|A\|_{\mathcal{L}(\ell_2,Y)} + \sigma_0 C_N^{Lip}(R_0)\big)\big)\Big)\|\mathbf{v} - \mathbf{u}^*(\sigma)\|_{\ell_2} \\ &\quad= \Big(\gamma_1 + \sigma C\Big)\|\mathbf{v} - \mathbf{u}^*(\sigma)\|_{\ell_2}, \end{aligned}$$

which implies that the contraction property (62) holds for $\sigma < (1-\gamma_1)C^{-1}$.
□

The last lemma established the contraction property (62) uniformly in σ. Therefore, for the current choice of $K = K_\sigma$ as in (53), we are able to apply directly Theorem 2.4. Let us summarize the result as follows.

Theorem 2.10. *Let the assumptions of Lemma 2.9 hold. Then, for all $\sigma \in [0, \min(\sigma_0, (1-\gamma_1)C^{-1}))$ and $\gamma_0 < \gamma < 1$, if we choose $\left(\boldsymbol{\alpha}^{(n)}\right)_{n\in\mathbb{N}}$ according to (45) and λ such that*

$$\lambda > \max\{\lambda_0, C^{Lip}_{K'_\sigma}(R'_0)(C^{bnd}_{K_\sigma}(R'_0) + C_Y) + C^{bnd}_{K'_\sigma}(R'_0)C^{Lip}_{K_\sigma}(R'_0)\},$$

the sequence $\left(\mathbf{u}^{(n)}(\sigma)\right)_{n\in\mathbb{N}}$ defined by (9) with initial guess $\mathbf{u}^{(0)} = \mathbf{0}$ satisfies

$$\left(\mathbf{u}^{(n)}(\sigma)\right)_{n\in\mathbb{N}} \subset B(R_0). \tag{63}$$

Furthermore it converges to any $\mathbf{u}^*(\sigma) \in \mathcal{C}(\sigma)$ at a linear rate, i.e.,

$$\|\mathbf{u}^*(\sigma) - \mathbf{u}^{(n)}(\sigma)\|_{\ell_2} \le \gamma^n \|\mathbf{u}^*(\sigma) - \mathbf{u}^{(0)}(\sigma)\|_{\ell_2}, \tag{64}$$

and moreover

$$\Gamma_{\boldsymbol{\alpha}^{(n+1)},\sigma}(\mathbf{u}^{(n+1)}(\sigma)) < \Gamma_{\boldsymbol{\alpha}^{(n)},\sigma}(\mathbf{u}^{(n)}(\sigma)),$$

provided that $\mathbf{u}^{(n)}(\sigma)$ is not yet a critical point of $\Gamma_{\boldsymbol{\alpha}^{(n)},\sigma}$. In particular $\mathbf{u}^*(\sigma) \in \mathcal{C}(\sigma)$ has to be the only critical point of $\Gamma_{\boldsymbol{\alpha},\sigma}$ in $B(R_0)$ with $\#\operatorname{supp}\mathbf{u}^*(\sigma) \le L_0$, actually it is its unique global minimizer in $B(R_0)$.

3 Preconditioning

The convergence analysis in Section 2 for the iteration (9) relies on the contraction property (42) of the operator T defined in (18). This property also ensures that, despite the fact that $\Gamma_{\boldsymbol{\alpha}}$ is a nonconvex functional, it has nevertheless a unique global minimizer in a prescribed ball centered at $\mathbf{0}$ and that the iteration (9) is guaranteed to converge to it with linear rate. Unfortunately, we can not expect this powerful property to hold in general, even for the case where the underlying operator K is linear and compact. Therefore, in this section, we present how preconditioning can be applied to promote property (42) for K being a nonlinear perturbation of a linear operator. We have to imagine the action of this preconditioning as a sort of "stretching" of the functional $\Gamma_{\boldsymbol{\alpha}}$, so that no local minimizers or stationary points remain around $\mathbf{0}$ other than a unique global minimizer. Preconditioning also changes the topology of the minimization problem related to (1). Therefore, in Section 3.1, we begin by discussing the related topological issues. In Section 3.2 we present a preconditioning strategy and state conditions under which the restricted isometry property (60) will be satisfied. Finally in Section 3.3 we apply our findings to an interesting class of operators.

3.1 General setting

We shall consider the following modified functional

$$\big(\Gamma_{\boldsymbol{\alpha}} \circ D^{-1}\big)(\mathbf{z}) = \|(K \circ D^{-1})(\mathbf{z}) - y\|_Y^2 + 2\|D^{-1}\mathbf{z}\|_{\ell_{1,\boldsymbol{\alpha}}}, \quad \mathbf{z} \in \mathrm{Ran}(D), \tag{65}$$

where $D : \ell_2 \to \mathrm{Ran}(D)$ is a suitable preconditioning matrix with well defined formal inverse $D^{-1} : \mathrm{Ran}(D) \to \ell_2$. Moreover we assume that D maps *finitely supported vectors* on *finitely supported vectors* and that

$$\|D^{-1}\mathbf{z}\|_{\ell_{1,\boldsymbol{\alpha}}} \sim \|\operatorname{diag}(D^{-1})\mathbf{z}\|_{\ell_{1,\boldsymbol{\alpha}}}. \tag{66}$$

Note, that preconditioning of the energy functional (1) changes the topology of the associated minimization problem. Moreover, the preconditioning operator D may be *unbounded* in the topology of ℓ_2. However, as we will see below, this is not an issue here. Indeed, Theorem 4.3, which will be proved later in Section 4, enables us to reduce the setting to a *finite dimensional* one whenever needed, so that we can use the equivalence of norms on finite dimensional vector spaces.

To this end we begin with the observation that any stationary point $\mathbf{u}^*$ of (1) can be characterized by the subdifferential inclusion

$$0 \in \partial\big(\Gamma_{\boldsymbol{\alpha}}\big)(\mathbf{u}^*).$$

An analogous characterization holds for the stationary points $\mathbf{z}^*$ of (65). By the chain rule for subdifferentials, see [15, Proposition I.5.7], we have

$$0 \in \partial\big(\Gamma_{\boldsymbol{\alpha}} \circ D^{-1}\big)(\mathbf{z}^*) = \big(D^{-1}\big)^\star \partial\big(\Gamma_{\boldsymbol{\alpha}}\big)(D^{-1}\mathbf{z}^*),$$

where $\big(D^{-1}\big)^\star$ is the dual mapping of D^{-1}. In other words, there is a one-to-one relationship of the stationary points of (1) and (65). Moreover, by our assumptions on D, if $\mathbf{u}^*$ is a finitely supported stationary point of (1), the related stationary point $\mathbf{z}^* = D\mathbf{u}^*$ of (65) is also finitely supported.

We will use the assumption (66) to simplify the preconditioned energy functional $\Gamma_{\boldsymbol{\alpha}} \circ D^{-1}$. Indeed, motivated by the observation that $\|\operatorname{diag}(D^{-1})\mathbf{z}\|_{\ell_{1,\boldsymbol{\alpha}}} = \|\mathbf{z}\|_{\ell_{1,\operatorname{diag}(D^{-1})\boldsymbol{\alpha}}}$ and with a slight abuse of notation we will consider the modified energy functional

$$\Gamma^D_{\boldsymbol{\alpha}}(\mathbf{z}) := \|(K \circ D^{-1})(\mathbf{z}) - y\|_Y^2 + 2\|\mathbf{z}\|_{\ell_{1,\operatorname{diag}(D^{-1})\boldsymbol{\alpha}}}, \quad \mathbf{z} \in \mathrm{Ran}(D), \tag{67}$$

and the resulting minimization problem.

We avoid to deal with the topology of $\mathrm{Ran}(D)$ in the following way. Let $\mathbf{z}^*$ be a fixed stationary point of the preconditioned energy functional (67) of finite support and $\Lambda_0 \subset \mathcal{J}$ an arbitrary finite set such that $\operatorname{supp}\mathbf{z}^* \subset \Lambda_0$. The restriction of (67) onto Λ_0 is then given by

$$\Gamma^D_{\boldsymbol{\alpha},\Lambda_0}(\mathbf{z}) := \|(K \circ D^{-1})_{|\Lambda_0}(\mathbf{z}) - y\|_Y^2 + 2\|\mathbf{z}\|_{\ell_{1,(\operatorname{diag}(D^{-1})\boldsymbol{\alpha})_{|\Lambda_0}}(\Lambda_0)}, \quad \mathbf{z} \in \mathbb{R}^{\Lambda_0}. \tag{68}$$

The minimization problem can now be considered in $\mathbb{R}^{\Lambda_0}$ endowed with the Euclidean norm. We denote the restriction of $\mathbf{z}^*$ onto Λ_0 by $\mathbf{z}^*_{|\Lambda_0}$, by $E_{\Lambda_0} : \ell_2(\Lambda_0) \to \ell_2$ the trivial extension by 0, and by $\big(E_{\Lambda_0}\big)^\star$ its adjoint, being actually the restriction operator $\mathbf{z}^*_{|\Lambda_0} = \big(E_{\Lambda_0}\big)^\star \mathbf{z}^*$. Then it follows by the chain rule for subdifferentials that

$$\begin{aligned} 0 \in \big(E_{\Lambda_0}\big)^\star \partial\big(\Gamma^D_{\boldsymbol{\alpha}}\big)(\mathbf{z}^*) &= \big(E_{\Lambda_0}\big)^\star \partial\big(\Gamma^D_{\boldsymbol{\alpha}}\big)\big(E_{\Lambda_0}\mathbf{z}^*_{|\Lambda_0}\big) \\ &= \partial\big(\Gamma^D_{\boldsymbol{\alpha}} \circ E_{\Lambda_0}\big)\big(\mathbf{z}^*_{|\Lambda_0}\big) \\ &= \partial\big(\Gamma^D_{\boldsymbol{\alpha},\Lambda_0}\big)\big(\mathbf{z}^*_{|\Lambda_0}\big). \end{aligned}$$

Consequently $\mathbf{z}^*_{|\Lambda_0}$ is also a stationary point of the finite dimensional energy functional (68). Unfortunately, the vice versa is not valid, because $\big(E_{\Lambda_0}\big)^\star$ is not injective and a stationary point for $\Gamma^D_{\boldsymbol{\alpha},\Lambda_0}$ does not necessarily correspond a priori to the restriction to a finite dimensional set Λ_0 of a stationary point of $\Gamma^D_{\boldsymbol{\alpha}}$ in $\mathrm{Ran}(D)$.

Nevertheless, if one could assume that $\Gamma^D_{\boldsymbol{\alpha},\Lambda_0}$ has actually only one critical point in $\mathbb{R}^{\Lambda_0}$ for any choice of $\Lambda_0 \subset \mathcal{J}$ finite, then we can argue the uniqueness of the critical point of $\Gamma^D_{\boldsymbol{\alpha}}$ in $\mathrm{Ran}(D)$ as well. In fact, if there were two critical points $\mathbf{z}^*_1$ and $\mathbf{z}^*_2$ for $\Gamma^D_{\boldsymbol{\alpha}}$ in $\mathrm{Ran}(D)$, their support could be included in a finite set Λ'_0 of indexes. Without loss of generality this set could be assumed to be a subset of Λ_0 for the latter large enough. Hence, the assumed uniqueness of the critical point in $\mathbb{R}^{\Lambda_0}$ for $\Gamma^D_{\boldsymbol{\alpha},\Lambda_0}$ would immediately imply that $(\mathbf{z}^*_1)_{\Lambda_0} = (\mathbf{z}^*_2)_{\Lambda_0}$ or, equivalently, that $\mathbf{z}^*_1 = \mathbf{z}^*_2$. In turn this means that, in the situation of a unique critical point in finite dimensions, the minimization of the finite dimensional problem is actually equivalent to the minimization of the infinite dimensional one.

In Section 4 we will present an implementable numerical scheme, which solves the finite dimensional minimization problem related to (68). We shall also show that a priori knowledge of the set Λ_0 is not needed. In fact, it will be constructed on the fly by the presented adaptive scheme.

3.2 Multilevel preconditioning

In Section 2.3 we considered the case that K consists of a dominant linear part A and a nonlinear perturbation. In this setting, we were able to show that the contraction assumption (42) can be guaranteed if the linear part of the equation fulfills the restricted isometry property (60). In general this condition will fail to hold, even if A is a compact linear operator. Nevertheless, in this section, we show that this issue can be solved by a preconditioning strategy. To this end, we partly follow the lines of [10] and recall the corresponding results as far as they are needed for our purposes.

In the following we will assume that $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain and $\boldsymbol{\Psi} := \{\psi_\mu\}_{\mu\in\mathcal{J}}$ is a compactly supported wavelet basis or frame of wavelet type for $L_2(\Omega)$, see e.g. [5, §2.12]. Every $\mu \in \mathcal{J}$ is of the form $\mu = (j, k, e)$, where $j \in \mathbb{Z}$ is the *scale*, often denoted as $|\mu|$, $k \in \mathbb{Z}^d$ is the *spatial location* and e is the *type* of ψ_μ. We refer to [5, 7] for further details concerning this notation. We do not go into construction details concerning these bases or the alternative of wavelet frames. In fact, we simply assume the following properties for all $\mu \in \mathcal{J}$. Furthermore, for the ease of presentation, we formulate them for the case of an orthogonal wavelet basis on $\Omega = (0, 1)^d$:

(W_1) The support $\Omega_\mu := \operatorname{supp} \psi_\mu$ fulfills $|\Omega_\mu| \sim 2^{-|\mu|d}$. Furthermore there exists a suitable cube Q, centered at the origin, s.t., $\Omega_\mu \subset 2^{-|\mu|}k + 2^{-|\mu|}Q$, see [5, §2.12].

(W_2) The basis has the cancellation property $\int_\Omega \xi^{\boldsymbol{\beta}} \psi_\mu(\xi) \mathrm{d}\xi = 0$, $|\boldsymbol{\beta}| = 0, \ldots, d^* \in \mathbb{N}$.

(W_3) $\|\psi_\mu\|_{L_\infty(\Omega)} \leq C 2^{d/2|\mu|}$.

Examples of wavelet bases satisfying these conditions can be found in [11]. In this setting the *synthesis map* related to $\boldsymbol{\Psi}$ reads as

$$\mathcal{F} : \ell_2 \to L_2(\Omega), \quad \mathcal{F}(\mathbf{u}) := \sum_{\mu \in \mathcal{J}} u_\mu \psi_\mu, \quad \mathbf{u} \in \ell_2. \tag{69}$$

Its adjoint is given by

$$\mathcal{F}^* : L_2(\Omega) \to \ell_2, \quad \mathcal{F}^*(u) := (\langle u, \psi_\mu \rangle_{\ell_2})_{\mu \in \mathcal{J}}. \tag{70}$$

Let $\mathcal{A} \in \mathcal{L}(L_2(\Omega), Y)$ be a linear operator and consider its discretization $A := \mathcal{A}\mathcal{F}$. In this section we aim at stating conditions under which (60) can be ensured by means of a preconditioning strategy. We will make technical assumptions on the matrix $G = \big(G_{\mu,\nu}\big)_{\mu,\nu \in \mathcal{J}}$ given by

$$G := A^* A = \big(\langle \mathcal{A}^* \mathcal{A} \psi_\nu, \psi_\mu \rangle_{L_2(\Omega)}\big)_{\mu,\nu \in \mathcal{J}}. \tag{71}$$

To be specific, we will assume that there exist constants $c_1, c_2, c_3, s, \eta, r \in \mathbb{R}_+, r > d$, such that the following conditions hold for all $\mu = (j, k, e), \nu = (j', k', e') \in \mathcal{J}$:

- The entries of G satisfy the decay estimate

$$|G_{\mu,\nu}| \leq c_1 \frac{2^{-s||\mu|-|\nu||} 2^{-\eta \min(|\mu|,|\nu|)}}{\big(1 + 2^{\min(|\mu|,|\nu|)} \operatorname{dist}(\Omega_\mu, \Omega_\nu)\big)^r} \tag{72}$$

- On the diagonal, i.e., $\mu = \nu$, it holds that

$$|G_{\mu,\mu}| \geq c_2 2^{-\eta|\mu|}. \tag{73}$$

- For the same scale, i.e., $|\mu| = |\nu|$, the entries satisfy

$$|G_{\mu,\nu}| \le c_3 \frac{2^{-2\eta|\mu|}}{(1+|k-k'|)^r}. \tag{74}$$

Under these conditions the following holds.

Theorem 3.1 [10, Thm.4.6.]. *Suppose that G fulfills* (72)*,* (73)*, and* (74) *with $c_2 > c_3/(r-d)$. Let D be the block-diagonal matrix consisting of the diagonal level blocks of G, i.e.,*

$$D_{\mu,\nu} := \begin{cases} G_{\mu,\nu} & |\mu| = |\nu|, \\ 0 & \text{otherwise.} \end{cases} \tag{75}$$

Then there exists a constant $C = C(c_1, c_2, c_3, r, d)$ such that for each finite set $\Lambda \subset \mathcal{J}$ with $|\Lambda| \le 2^s C^{-1}$ the sub-matrix $(D^{-1/2}GD^{-1/2})_{|\Lambda\times\Lambda}$ satisfies

$$\|(\mathrm{Id} - D^{-1/2}GD^{-1/2})_{|\Lambda\times\Lambda}\| < C\, 2^{-(s-\frac{\eta}{2})}|\Lambda|$$

and

$$\kappa\big((D^{-1/2}GD^{-1/2})_{|\Lambda\times\Lambda}\big) \le \frac{1 + C\, 2^{-(s-\frac{\eta}{2})}|\Lambda|}{1 - C\, 2^{-(s-\frac{\eta}{2})}|\Lambda|}.$$

3.3 Integral operators with Schwartz kernels on disjoint domains

In this section we study a class of operators which fits into the setting of Section 2.3. Let $\Omega, \hat{\Omega} \subset \mathbb{R}^d$ be two bounded Lipschitz domains with $\mathrm{dist}(\Omega, \hat{\Omega}) = \delta > 0$. For fixed $t \in \mathbb{R}_+$ we consider

$$\mathcal{K} = \mathcal{A} + \sigma\mathcal{N} : L_2(\Omega) \to H^t(\hat{\Omega}),$$

where $\sigma \in \mathbb{R}_+$, $\mathcal{A} \in \mathcal{L}(L_2(\Omega), H^t(\hat{\Omega}))$ is linear, and $\mathcal{N} : L_2(\Omega) \to H^t(\hat{\Omega})$ is a nonlinear operator. Furthermore, we assume that the linear part $\mathcal{A}$ is an integral operator with a Schwartz kernel. To be specific, we assume that $\mathcal{A}$ is given by

$$v \mapsto \mathcal{A}v := \int_\Omega \Phi(\cdot, \xi) v(\xi) \mathrm{d}\xi, \tag{76}$$

where $\Phi : \hat{\Omega} \times \Omega \to \mathbb{R}$ is a kernel of Schwartz type, i.e.,

$$|\partial_x^{\boldsymbol{\alpha}} \partial_\xi^{\boldsymbol{\beta}} \Phi(x, \xi)| \le c_{\boldsymbol{\alpha},\boldsymbol{\beta}} |x - \xi|^{-(d+2t+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)}, \quad \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^d, \tag{77}$$

holds. Concerning the nonlinear perturbation $\mathcal{N}$, we assume that it is given by

$$v \mapsto \mathcal{N}(v) := \int_\Omega \tilde{\Phi}(\cdot, \xi) |v(\xi)|^2 \mathrm{d}\xi,$$

where $\tilde{\Phi}$ also fulfills (77). This condition implies that $\mathcal{A}$ and $\mathcal{N}$ are well defined as operators mapping into $H^t(\hat{\Omega})$. Moreover, the nonlinear perturbation $\mathcal{N}$ is twice continuously differentiable and consequently indeed $\mathcal{N}$ and $\mathcal{N}'$ are Lipschitz continuous on bounded closed sets: To see this, we write $\mathcal{N} = \mathcal{N}_1 \circ \mathcal{N}_2$ with

$$\begin{aligned} \mathcal{N}_1 : L_1(\Omega) &\to H^t(\hat{\Omega}), & \mathcal{N}_2 : L_2(\Omega) &\to L_1(\Omega), \\ v &\mapsto \int_\Omega \tilde{\Psi}(\cdot,\xi)v(\xi)\mathrm{d}\xi, & v &\mapsto |v|^2. \end{aligned}$$

Here the operator $\mathcal{N}_1$ as well as the derivative of $\mathcal{N}_2$, i.e.,

$$\mathcal{N}_2' : v \mapsto 2\langle v, \cdot\rangle_{L_2(\Omega)},$$

are linear. Recall that the synthesis map $\mathcal{F}$ associated to $\boldsymbol{\Psi}$ is given by (69). It is linear and hence Lipschitz. Together with the Lipschitz properties of $\mathcal{N}$, this implies that the discretized version of the nonlinear part, i.e., $N = \mathcal{N}\mathcal{F}$, fulfills (54).

Let us now assume that the linear term $A = \mathcal{A}\mathcal{F}$ of $K = A + \sigma N$ does not fulfill already (60). We want to show that setting

$$K \circ D = A \circ D + \sigma N \circ D,$$

for a suitable preconditioning matrix D, will allow us now to fulfill it for $A \circ D$. Moreover the new nonlinear perturbation $N \circ D$ will again satisfy the Lipschitz continuity conditions (54) as soon as we will remember that, eventually, the problem will be turned into a finite dimensional one. We shall construct the preconditioning matrix D by using the multilevel techniques presented in Section 3.2. To be specific, the remainder of this section is dedicated to the proof of property (72) of the matrix $G := \big(\mathcal{A}\mathcal{F}\big)^*\mathcal{A}\mathcal{F}$. (The other required properties (73) and (74) may be difficult to be shown, but they are often verified in practice.) To this end we follow the lines of [9]. To be explicit, with (71), the entries of G are given as

$$G_{\mu,\nu} = \langle \mathcal{A}^*\mathcal{A}\psi_\nu, \psi_\mu\rangle_{L_2(\Omega)} = \langle \mathcal{A}\psi_\nu, \mathcal{A}\psi_\mu\rangle_{H^t(\hat{\Omega})}. \tag{78}$$

We begin with the special case $t = 0$. and apply Taylor's formula to the kernel Φ around a point $\xi_0 \in \Omega_\mu$. For every $\xi \in \Omega_\mu$ there exists $\theta \in [0,1]$ such that

$$\Phi(x,\xi) = \sum_{|\boldsymbol{\beta}|\le d^*} \frac{\partial_\xi^{\boldsymbol{\beta}}\Phi(x,\xi_0)}{\boldsymbol{\beta}!}(\xi-\xi_0)^{\boldsymbol{\beta}} + \sum_{|\boldsymbol{\beta}|=d^*+1} \frac{\partial_\xi^{\boldsymbol{\beta}}\Phi(x,\xi+\theta(\xi-\xi_0))}{\boldsymbol{\beta}!}(\xi-\xi_0)^{\boldsymbol{\beta}}. \tag{79}$$

With (77) we can estimate

$$
\begin{aligned}
&|\sum_{|\boldsymbol{\beta}|=d^*+1} \frac{\partial_\xi^{\boldsymbol{\beta}}\Phi(x,\xi+\theta(\xi-\xi_0))}{\boldsymbol{\beta}!}(\xi-\xi_0)^{\boldsymbol{\beta}}| \\
&\qquad \leq \sum_{|\boldsymbol{\beta}|=d^*+1} \frac{1}{\boldsymbol{\beta}!}|(\xi-\xi_0)^{\boldsymbol{\beta}}| \sup_{\xi'\in\Omega_\mu} |\partial_\xi^{\boldsymbol{\beta}}\Phi(x,\xi')| \\
&\qquad \leq \sum_{|\boldsymbol{\beta}|=d^*+1} \frac{1}{\boldsymbol{\beta}!}|(\xi-\xi_0)^{\boldsymbol{\beta}}| c_{\mathbf{0},\boldsymbol{\beta}} \operatorname{dist}(x,\Omega_\mu)^{-(d+2t+d^*+1|)}.
\end{aligned}
\tag{80}
$$

The cancellation property (W_2) of $\psi_\mu \in \mathbf{\Psi}$, together with (79) and (80) yields

$$
\begin{aligned}
|\mathcal{A}\psi_\mu(x)| &= |\int_{\Omega_\mu} \Phi(x,\xi)\psi_\mu(\xi)\mathrm{d}\xi| \\
&\leq \sum_{|\boldsymbol{\beta}|=d^*+1} \frac{1}{\boldsymbol{\beta}!} c_{\mathbf{0},\boldsymbol{\beta}} \operatorname{dist}(x,\Omega_\mu)^{-(d+2t+d^*+1)} \int_{\Omega_\mu} |(\xi-\xi_0)^{\boldsymbol{\beta}}||\psi_\mu(\xi)|\mathrm{d}\xi.
\end{aligned}
\tag{81}
$$

By our assumptions (W_1) and (W_3) on the wavelets, i.e., $\Omega_\mu \subset 2^{-|\mu|}k + 2^{-|\mu|}Q$ and $\|\psi_\mu\|_{L_\infty(\Omega)} \leq C2^{d/2|\mu|}$, together with $\xi_0 \in \Omega_\mu$, it holds that

$$
\begin{aligned}
\int_{\Omega_\mu} |(\xi-\xi_0)^{\boldsymbol{\beta}}||\psi_\mu(\xi)|\mathrm{d}\xi &\leq C2^{\frac{d}{2}|\mu|} \int_{\Omega_\mu} |(\xi-\xi_0)^{\boldsymbol{\beta}}|\mathrm{d}\xi \\
&\leq C2^{\frac{d}{2}|\mu|} \int_Q |(2^{-|\mu|}(\xi'+k)-\xi_0)^{\boldsymbol{\beta}}| 2^{-d|\mu|}\mathrm{d}\xi' \\
&\leq C'2^{-|\mu|(\frac{d}{2}+|\boldsymbol{\beta}|)}.
\end{aligned}
$$

The combination of (81) and the last estimate implies

$$
|\mathcal{A}\psi_\mu(x)| \leq C_{d^*} \operatorname{dist}(x,\Omega_\mu)^{-(d+2t+d^*+1)} 2^{-|\mu|(\frac{d}{2}+d^*+1)}. \tag{82}
$$

Since we assumed that Ω and $\hat{\Omega}$ are disjoint domains, it holds for $\xi,\xi' \in \Omega$ with $\xi \neq \xi'$ and $\|\xi-x\|_2, \|\xi'-x\|_2 \geq \delta$ that $\frac{1}{|\xi-x||\xi'-x|} \leq C_{x,\delta}\frac{1}{|\xi-\xi'|}$. Further $C_{x,\delta}$ may be bounded independently of x. With (82) we prove immediately, for $\mu \neq \nu$ with $\operatorname{dist}(\Omega_\mu,\Omega_\nu) > 0$ the estimate

$$
\begin{aligned}
&|\langle \mathcal{A}\psi_\mu, \mathcal{A}\psi_\nu \rangle_{L_2(\hat{\Omega})}| \\
&\quad \leq (C_{d^*})^2 2^{-(|\mu|+|\nu|)(\frac{d}{2}+d^*+1)} \int_{\hat{\Omega}} \big(\operatorname{dist}(x,\Omega_\mu)\operatorname{dist}(x,\Omega_\nu)\big)^{-(d+2t+d^*+1)}\mathrm{d}x \\
&\quad \leq (C_{d^*})^2 |\hat{\Omega}| 2^{-(|\mu|+|\nu|)(\frac{d}{2}+d^*+1)} \operatorname{dist}(\Omega_\mu,\Omega_\nu)^{-(d+2t+d^*+1)} \\
&\quad = (C_{d^*})^2 |\hat{\Omega}| \frac{2^{-||\mu|-|\nu||(\frac{d}{2}+d^*+1)} 2^{-\min(|\mu|,|\nu|)(d^*+1-2t)}}{\big(2^{\min(|\mu|,|\nu|)}\operatorname{dist}(\Omega_\mu,\Omega_\nu)\big)^{d+2t+d^*+1}}.
\end{aligned}
\tag{83}
$$

For the case $\operatorname{dist}(\Omega_\mu, \Omega_\nu) = 0$ we use again (82) and apply $\operatorname{dist}(x, \Omega_\mu) \geq \delta$ directly to derive the simpler estimate

$$|\langle \mathcal{A}\psi_\mu, \mathcal{A}\psi_\nu \rangle_{L_2(\hat{\Omega})}| \leq (C_{d^*})^2 |\hat{\Omega}| 2^{-(|\mu|+|\nu|)(\frac{d}{2}+d^*+1)} \delta^{-2(d+2t+d^*+1)}. \tag{84}$$

Together (83) and (84) imply that $\big(\langle \mathcal{A}\psi_\mu, \mathcal{A}\psi_\nu \rangle_{L_2(\hat{\Omega})}\big)_{\mu,\nu \in \mathcal{J}}$ fulfills the assumption (72).

For the general case $t > 0$, we consider $\big(\langle \partial_x^{\boldsymbol{\alpha}}(\mathcal{A}\psi_\mu), \partial_x^{\boldsymbol{\alpha}}(\mathcal{A}\psi_\nu) \rangle_{L_2(\hat{\Omega})}\big)_{\mu,\nu \in \mathcal{J}}$, $\boldsymbol{\alpha} \in \mathbb{N}_0^d$. In this setting

$$\partial_x^{\boldsymbol{\alpha}}(\mathcal{A}\psi_\mu) = \int_\Omega \partial_x^{\boldsymbol{\alpha}} \Phi(\cdot, \xi) \psi_\mu(\xi) \mathrm{d}\xi$$

is again an integral with a Schwartz kernel. Indeed, an analogous argumentation as in the case $t = 0$ yields condition (72) for the case $t \in \mathbb{N}$, and consequently for $t \in \mathbb{R}_+$.

4 Equivalence to an inexact finite dimensional scheme

In practice, whenever we deal with infinite dimensional problems, the operators K and K' can not be evaluated exactly, and one has to replace their output by suitable numerical approximations. In this section we study the convergence behavior of the resulting inexact algorithm to solve the preconditioned minimization problem (67). Although the original problem is posed in general in infinite dimensions, adaptive approximations will allow us to show the confinement of the iteration within a well-determined finite dimensional space. In particular, in Theorem 4.3 below, we show that the global support of all iterates is contained in a finite set Λ_0. From a practical point of view, there would be no difference between the iterates produced by the adaptive scheme over the whole index set $\mathcal{J}$ or if we would restrict the set of possible indices to the (a priori unknown) set Λ_0. Therefore, by arguing as in Section 3.1, the combination of preconditioning and adaptive solvers yields an iterative scheme for the minimization of the unpreconditioned functional $\Gamma_{\boldsymbol{\alpha}}$.

We focus on the error introduced by the inexact evaluation of the nonlinear functional K and the linear operator $(K'(\cdot))^*$. To this end let us assume that for given tolerances $\rho, \delta > 0$, there exist approximation schemes which for every $\mathbf{v} \in \ell_2$ and pairs $(\mathbf{v}, w) \in \ell_2 \times Y$, respectively, compute finite dimensional approximations $[K(\mathbf{v})]_\rho$ and $[\big(K'(\mathbf{v})\big)^*(w)]_\delta$ such that

$$\begin{aligned} \|K(\mathbf{v}) - [K(\mathbf{v})]_\rho\|_Y &\leq \rho, \\ \|\big(K'(\mathbf{v})\big)^*(w) - [\big(K'(\mathbf{v})\big)^*(w)]_\delta\|_{\ell_2} &\leq \delta. \end{aligned} \tag{85}$$

This assumption is realistic, e.g., if the exact application of K and $(K'(\cdot))^*$ involves the solution of partial differential or integral equations and the numerical approximations can be computed by means of adaptive discretization schemes. Let us mention two prominent examples from the context of adaptive wavelet schemes of linear and nonlinear operator equations.

Example 4.1. Let $\mathbf{\Psi}_Y = \{\psi_{X,\mu}\}_{\mu\in\mathcal{J}_X}$ and $\mathbf{\Psi}_Y = \{\psi_{Y,\mu}\}_{\mu\in\mathcal{J}_Y}$ be wavelet Riesz bases for X and Y, respectively, such that the assumptions of Section 3.2 are satisfied. We denote the associated synthesis operators by $\mathcal{F}_X$ and $\mathcal{F}_Y$. Furthermore let $K = \mathcal{K}\circ\mathcal{F}_X : \ell_2 \to Y$ for some nonlinear operator $\mathcal{K} : X \to Y$.

(i) For the efficient approximate application of the linear operator $(K'(\mathbf{v}))^*$ to a given $w \in Y$, it is advantageous if the coefficient array $\mathbf{w} \in \ell_2(\mathcal{J}_Y)$ of $w = \mathcal{F}_Y(\mathbf{w})$, or at least good approximations of it, has a fast decay [7]. In that case, one may exploit the representation

$$(K'(\mathbf{v}))^*(w) = (\mathcal{F}_X^* \circ (\mathcal{K}'(\mathcal{F}_X(\mathbf{v}))^* \circ \mathcal{F}_Y)(\mathbf{w}) =: \mathbf{A_v w}$$

and the compressibility of the stiffness matrix $\mathbf{A_v} \in \mathcal{L}(\ell_2(\mathcal{J}_Y), \ell_2(\mathcal{J}_X))$. In fact, if $\mathbf{A_v} \in \mathcal{L}(\ell_\tau^w(\mathcal{J}_Y), \ell_\tau^w(\mathcal{J}_X))$ for all $0 < \tau_0 < \tau < 2$, then the second inequality in (85) can be ensured by suitable matrix compression techniques. In the special case of wavelet Riesz bases $\mathbf{\Psi}_X$, $\mathbf{\Psi}_Y$ and $\mathcal{K}'(\mathcal{F}_X(\mathbf{v}))$ being a differential operator or an integral operator with Schwartz kernel, e.g., we can expect that the stiffness matrix $\mathbf{A_v}$ is s^*-compressible, i.e., there exist biinfinite matrices $\mathbf{A}_{\mathbf{v},j}$ with at most a constant multiple of 2^j nontrivial entries per row and column, such that $\|\mathbf{A_v} - \mathbf{A}_{\mathbf{v},j}\|_2 \le C_s 2^{-js}$, $0 < s < s^*$. This property implies that $\mathbf{A_v}$ boundedly maps $\ell_\tau^w(\mathcal{J}_Y)$ into $\ell_\tau^w(\mathcal{J}_X)$. We refer to [7,22] and related works on the compressibility of operators in wavelet coordinates and the concrete realization of associated adaptive matrix-vector multiplications.

(ii) The approximate evaluation of the nonlinearity K itself at a given input $\mathbf{v} \in \ell_2$ is enabled under additional assumptions on the type of the nonlinearity. In the context of nonlinear operators, tree approximation techniques play an important role. Here a tree structure is imposed on the coefficient array of the output argument. For example, in the special case that X is a closed subspace of $H^s(\Omega)$, $s \ge 0$, $\Omega \subset \mathbb{R}^d$ a bounded domain, $Y = X'$ and $\mathcal{K}$ decomposes into $\mathcal{K} = \mathcal{A} + \mathcal{N}$ with a linear, boundedly invertible operator $\mathcal{A} : X \to X'$ and a Nemytskii-type nonlinearity

$$\mathcal{N} : X \to X', \quad (\mathcal{N}(v))(x) = f\big(\partial^{\beta_1} v(x), \ldots, \partial^{\beta_k} v(x)\big), \quad \beta_j \in \mathbb{N}_0^d,$$

adaptive wavelet tree approximation techniques have been developed and implemented in [1,8,12,18].

For simplicity, we will assume in the sequel that y is given exactly. For convenience, we define the analogue of (18) by

$$\tilde{T}_{\rho,\delta}(\mathbf{v}) := \mathbf{v} - \frac{1}{\lambda}[\big(K'(\mathbf{v})\big)^*([K(\mathbf{v})]_\rho - y)]_\delta. \tag{86}$$

An implementable version of ISTA with decreasing threshold parameters $\boldsymbol{\alpha}^{(n)}$, i.e., (9), is then given by

$$\tilde{\mathbf{u}}^{(n+1)} = \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(\tilde{T}_{\rho^{(n)},\delta^{(n)}}(\tilde{\mathbf{u}}^{(n)})\big). \tag{87}$$

The following theorem shows that if the parameters $\rho^{(n)}, \delta^{(n)}$ are suitably chosen, the overall algorithm is still linearly convergent.

Theorem 4.2. *Let $\mathbf{u}^*$ be a stationary point of (1) that satisfies $T(\mathbf{u}^*) \in \ell_\tau^w(\mathcal{J})$ for some $0 < \tau < 2$. Furthermore let $\boldsymbol{\alpha}^{(n)}, \boldsymbol{\alpha} \in \mathbb{R}_+^{\mathcal{J}}$ with $\alpha_\mu^{(n)} \geq \alpha_\mu \geq \alpha \in \mathbb{R}_+, \mu \in \mathcal{J}$. We set $\tilde{\mathbf{u}}^{(0)} = \mathbf{0}$ and assume that T fulfills condition (37). With C_T^{Lip} be as therein and C is as in Lemma 2.2 we set*

$$\tilde{L} := \frac{4\big((C_T^{Lip} + \tilde{\gamma} - \gamma)\|\mathbf{u}^*\|_{\ell_2}\big)^2\lambda^2}{\alpha^2} + 4C|T(\mathbf{u}^*)|_{\ell_\tau^w(\mathcal{J})}^\tau \Big(\frac{\alpha}{\lambda}\Big)^{-\tau},$$

and define $\mathcal{B}_{\tilde{L}}$ analogously to (41). Let us assume that there exists $0 < \gamma_0 < 1$, such that for all $\mathbf{v} \in \mathcal{B}_{\tilde{L}}$ and $\operatorname{supp}(\mathbf{v}) \subset \Lambda \subset \mathcal{J}$ with $\#\Lambda \leq 2\tilde{L}$

$$\|\big(T(\mathbf{u}^*) - T(\mathbf{v})\big)_{|S^*\cup\Lambda}\|_{\ell_2(S^*\cup\Lambda)} \leq \gamma_0\|\mathbf{u}^* - \mathbf{v}\|_{\ell_2}. \tag{88}$$

For the operator K' we assume

$$C_{K'}^{bnd}(\mathcal{B}_{\tilde{L}}) := \sup_{\mathbf{v}\in\mathcal{B}_{\tilde{L}}} \|K'(\mathbf{v})\|_{\mathcal{L}(\ell_2,Y)} < \infty. \tag{89}$$

Then, for any $\gamma_0 < \gamma < \tilde{\gamma} < 1$ the inexact thresholded iteration (87) fulfills

$$\big(\tilde{\mathbf{u}}^{(n)}\big)_{n\in\mathbb{N}} \subset \mathcal{B}_{\tilde{L}}$$

and converges to $\mathbf{u}^*$ at a linear rate

$$\|\mathbf{u}^* - \tilde{\mathbf{u}}^{(n)}\|_{\ell_2} \leq \tilde{\epsilon}^{(n)} := \tilde{\gamma}^n\|\mathbf{u}^*\|_{\ell_2}, \tag{90}$$

whenever the parameters and tolerances are chosen according to

$$\frac{1}{\lambda}\Big(C_{K'}^{bnd}(\mathcal{B}_{\tilde{L}})\rho^{(n)} + \delta^{(n)}\Big) \leq (\tilde{\gamma} - \gamma)\tilde{\epsilon}^{(n)}, \tag{91}$$

$$\max_{\mu\in\mathcal{J}} |\alpha_\mu^{(n)} - \alpha_\mu| \leq \lambda\tilde{L}^{-\frac{1}{2}}(\gamma - \gamma_0)\tilde{\epsilon}^{(n)}. \tag{92}$$

Proof. The proof is an induction over n. The case $n = 0$ is covered by the assumptions. Now let $\tilde{\mathbf{u}}^{(n)} \in \mathcal{B}_{\tilde{L}}$ and (90) hold for $n \in \mathbb{N}$. We begin by proving $\#\operatorname{supp} \tilde{\mathbf{u}}^{(n+1)} \leq \tilde{L}$. To this end we use the standing assumption (85) on the inexact operator evaluations, together with the assumption (91) to estimate for $\mathbf{v} \in \mathcal{B}_{\tilde{L}}$

$$\begin{aligned}
&\|T(\mathbf{v}) - \tilde{T}_{\rho,\delta}(\mathbf{v})\|_{\ell_2} \\
&\quad= \frac{1}{\lambda}\|\big(K'(\mathbf{v})\big)^*(K(\mathbf{v}) - y) - [\big(K'(\mathbf{v})\big)^*([K(\mathbf{v})]_\rho - y)]_\delta\|_{\ell_2} \\
&\quad\leq \frac{1}{\lambda}\Big(\|\big(K'(\mathbf{v})\big)^*(K(\mathbf{v}) - y) - \big(K'(\mathbf{v})\big)^*([K(\mathbf{v})]_\rho - y)\|_{\ell_2} \\
&\quad+ \|\big(K'(\mathbf{v})\big)^*([K(\mathbf{v})]_\rho - y) - [\big(K'(\mathbf{v})\big)^*([K(\mathbf{v})]_\rho - y)]_\delta\|_{\ell_2}\Big) \\
&\quad\leq \frac{1}{\lambda}\Big(\|\big(K'(\mathbf{v})\big)^*\|_{\mathcal{L}(Y,\ell_2)}\rho + \delta\Big) \\
&\quad\leq (\tilde{\gamma} - \gamma)\tilde{\epsilon}^{(n)}.
\end{aligned} \tag{93}$$

This inequality, applied for $\boldsymbol{v} = \tilde{\mathbf{u}}^{(n)} \in \mathcal{B}_{\tilde{L}}$, implies together with the Lipschitz continuity assumption (37) that

$$\begin{aligned}
&\|T(\mathbf{u}^*) - \tilde{T}_{\rho^{(n)},\delta^{(n)}}(\tilde{\mathbf{u}}^{(n)})\|_{\ell_2} \\
&\quad= \|T(\mathbf{u}^*) - T(\tilde{\mathbf{u}}^{(n)}) + T(\tilde{\mathbf{u}}^{(n)}) - \tilde{T}_{\rho^{(n)},\delta^{(n)}}(\tilde{\mathbf{u}}^{(n)})\|_{\ell_2} \\
&\quad\leq (C_T^{Lip} + \tilde{\gamma} - \gamma)\tilde{\epsilon}^{(n)},
\end{aligned}$$

By invoking Lemma 2.2 we can conclude

$$\#\operatorname{supp}(\tilde{\mathbf{u}}^{(n+1)}) = \#\operatorname{supp} \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(\tilde{T}_{\rho^{(n)},\delta^{(n)}}(\tilde{\mathbf{u}}^{(n)})\big) \leq \tilde{L}. \tag{94}$$

For the second part of the proof we set $\tilde{\Lambda}^{(n)} := S^* \cup \operatorname{supp} \tilde{\mathbf{u}}^{(n)} \cup \operatorname{supp} \tilde{\mathbf{u}}^{(n+1)}$. Notice that $\#\operatorname{supp} \tilde{\mathbf{u}}^{(n)} \cup \operatorname{supp} \tilde{\mathbf{u}}^{(n+1)} \leq 2\tilde{L}$. Because shrinkage is nonexpansive and by the assumption (88) we may estimate

$$\begin{aligned}
&\|\mathbf{u}^*_{|\tilde{\Lambda}^{(n)}} - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big)\|_{\ell_2(\tilde{\Lambda}^{(n)})} \\
&\quad= \|\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\mathbf{u}^*)_{|\tilde{\Lambda}^{(n)}}\big) - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big)\|_{\ell_2(\tilde{\Lambda}^{(n)})} \leq \gamma_0\tilde{\epsilon}^{(n)}.
\end{aligned} \tag{95}$$

Moreover, we may use the Lipschitz assumption (37) directly and invoke Lemma 2.2 directly to conclude

$$\begin{aligned}
\#\operatorname{supp} \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\tilde{\mathbf{u}}^{(n)})\big) &\leq \frac{4\|T(\mathbf{u}^*) - T(\tilde{\mathbf{u}}^{(n)})\|_{\ell_2}^2\lambda^2}{\alpha^2} + 4C|T(\mathbf{u}^*)|^\tau_{\ell^w_\tau(\mathcal{J})}\Big(\frac{\alpha}{\lambda}\Big)^{-\tau} \\
&\leq \frac{4(C_T^{Lip}\tilde{\epsilon}^{(n)})^2\lambda^2}{\alpha^2} + 4C|T(\mathbf{u}^*)|^\tau_{\ell^w_\tau(\mathcal{J})}\Big(\frac{\alpha}{\lambda}\Big)^{-\tau} \\
&\leq \tilde{L}
\end{aligned} \tag{96}$$

Since $\boldsymbol{\alpha}^{(n)}$ is decreasing to $\boldsymbol{\alpha}$ it holds that $\operatorname{supp} \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(T(\tilde{\mathbf{u}}^{(n)})\big) \subset \operatorname{supp} \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\tilde{\mathbf{u}}^{(n)})\big)$. This, together with (92) gives

$$\begin{aligned} &\|\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big) - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big)\|_{\ell_2(\tilde{\Lambda}^{(n)})} \\ &\quad\leq \frac{\big(\#\operatorname{supp} \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}(T(\tilde{\mathbf{u}}^{(n)}))\big)^{\frac{1}{2}}}{\lambda} \max_{\mu\in\mathcal{J}} |\alpha_\mu^{(n)} - \alpha_\mu| \\ &\quad\leq (\gamma-\gamma_0)\tilde{\epsilon}^{(n)}. \end{aligned} \tag{97}$$

Finally, we use that shrinkage in nonexpansive, together with (93) for $\tilde{\mathbf{u}}^{(n)}$ for the estimate

$$\|\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\Lambda^{(n)}}\big) - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(\tilde{T}_{\rho^{(n)},\delta^{(n)}}(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big)\|_{\ell_2(\tilde{\Lambda}^{(n)})} \leq (\tilde{\gamma}-\gamma)\tilde{\epsilon}^{(n)}. \tag{98}$$

The combination of (95), (97), and (98) finalizes the proof

$$\begin{aligned} \|\mathbf{u}^* - \tilde{\mathbf{u}}^{(n+1)}\|_{\ell_2} &\leq \|\mathbf{u}^*_{|\tilde{\Lambda}^{(n)}} - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big)\|_{\ell_2(\tilde{\Lambda}^{(n)})} \\ &\quad + \|\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big) - \mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big)\|_{\ell_2(\tilde{\Lambda}^{(n)})} \\ &\quad + \|\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big) - \tilde{\mathbf{u}}^{(n+1)}_{|\tilde{\Lambda}^{(n)}}\|_{\ell_2(\tilde{\Lambda}^{(n)})} \\ &\leq \gamma_0\tilde{\epsilon}^{(n)} + (\gamma-\gamma_0)\tilde{\epsilon}^{(n)} + (\tilde{\gamma}-\gamma)\tilde{\epsilon}^{(n)} = \tilde{\epsilon}^{(n+1)}. \qquad \square \end{aligned}$$

We have shown that the support size of each iterate $\tilde{\mathbf{u}}^{(n)}$ can be bounded by a uniform constant. As it turns out there also exists a bounded set $\Lambda_0 \subset \mathcal{J}$ that contains all those supports.

Theorem 4.3. *Let the assumptions of Theorem 4.2 hold. Let $N \in \mathbb{N}$ be large enough such that there exists $\delta > 0$ with*

$$\tilde{\epsilon}^{(N+1)} + \delta \leq \frac{1}{\lambda} \inf_{\mu\in\mathcal{J}} \alpha_\mu. \tag{99}$$

Then it holds that

$$\operatorname{supp}(\tilde{\mathbf{u}}^{(n)}) \subset \Lambda_\delta(T(\mathbf{u}^*)), \quad n \geq N,$$

and consequently

$$\operatorname{supp}(\tilde{\mathbf{u}}^{(n)}) \subset \Big(\bigcup_{j=0}^{N} \operatorname{supp}(\tilde{\mathbf{u}}^{(j)})\Big) \cup \Lambda_\delta(T(\mathbf{u}^*)) =: \Lambda_0, \quad n \in \mathbb{N}.$$

Proof. We prove that for any fixed $n \geq N$ all $\mu \in \mathcal{J} \setminus \Lambda_\delta(T(\mathbf{u}^*))$ it holds that $(\tilde{\mathbf{u}}^{(n+1)})_\mu = 0$. To this end let $0 < \gamma_0 < \gamma < \tilde{\gamma} < 1$ be as in Theorem 4.2 and denote $\tilde{\Lambda}^{(n)} := S^* \cup \operatorname{supp} \tilde{\mathbf{u}}^{(n)} \cup \tilde{\mathbf{u}}^{(n+1)}$. Recall that by estimating as in equation (93) it holds that

$$\|\big(T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}} - \tilde{T}_{\rho^{(n)},\delta^{(n)}}(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\big)\|_{\ell_2(\tilde{\Lambda}^{(n)})} \leq (\tilde{\gamma}-\gamma)\tilde{\epsilon}^{(n)}, \tag{100}$$

and further that, since $\#\operatorname{supp}\tilde{\mathbf{u}}^{(n)} \leq \tilde{L}$, we can use (42) to estimate

$$\|\big(T(\mathbf{u}^*)_{|\tilde{\Lambda}^{(n)}} - T(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\|_{\ell_2(\tilde{\Lambda}^{(n)})} \leq \gamma_0 \tilde{\epsilon}^{(n)}. \tag{101}$$

By definition $\mu \in \tilde{\Lambda}^{(n)} \setminus \Lambda_\delta(T(\mathbf{u}^*))$ implies that $|\big(T(\mathbf{u}^*)\big)_\mu| \leq \delta$. Therefore, for such μ we may use (100), (101), (99), and the fact that $\boldsymbol{\alpha}^{(n)}$ is decreasing to $\boldsymbol{\alpha}$ to estimate

$$\begin{aligned} |\big(\tilde{T}_{\rho^{(n)},\delta^{(n)}}(\tilde{\mathbf{u}}^{(n)})\big)_\mu| &\leq \|T(\mathbf{u}^*)_{|\tilde{\Lambda}^{(n)}} - \tilde{T}_{\rho^{(n)},\delta^{(n)}}(\tilde{\mathbf{u}}^{(n)})_{|\tilde{\Lambda}^{(n)}}\|_{\ell_2}(\tilde{\Lambda}^{(n)}) + |\big(T(\mathbf{u}^*)\big)_\mu| \\ &\leq \tilde{\epsilon}^{(N+1)} + \delta \leq \inf_{\nu \in \mathcal{J}} \alpha_\nu \leq \inf_{\nu \in \mathcal{J}} \alpha_\nu^{(n)}. \end{aligned}$$

Finally, by the definition of $\mathbb{S}_{\frac{1}{\lambda}\boldsymbol{\alpha}^{(n)}}$ it follows that $(\tilde{\mathbf{u}}^{(n+1)})_\mu = 0$. □

References

[1] A. Barinka, *Fast computation tools for adaptive wavelet schemes*, Ph.D. thesis, RWTH Aachen, 2005.

[2] T. Bonesky, K. Bredies, D.A. Lorenz, and P. Maass, *A generalized conditional gradient method for nonlinear operator equations with sparsity constraints*, Inverse Probl. **23** (2007), no. 5, 2041–2058.

[3] K. Bredies, D.A. Lorenz, and P. Maass, *A generalized conditional gradient method and its connection to an iterative shrinkage method*, Comput. Optim. Appl. **42** (2009), no. 2, 173–193.

[4] A. Chambolle, R.A. DeVore, N. Lee, and B.J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process. **7** (1998), no. 3, 319–335.

[5] A. Cohen, *Numerical analysis of wavelet methods*, Studies in Mathematics and its Applications, vol. 32, Elsevier, North-Holland, Amsterdam, 2003.

[6] A. Cohen, W. Dahmen, and R.A. DeVore, *Adaptive wavelet methods for elliptic operator equations — Convergence rates*, Math. Comput. **70** (2001), 27–75.

[7] ———, *Adaptive wavelet methods for elliptic operator equations: Convergence rates*, Math. Comput. **70** (2001), no. 233, 27–75.

[8] ———, *Adaptive wavelet schemes for nonlinear variational problems*, SIAM J. Numer. Anal. **41** (2003), no. 5, 1785–1823.

[9] S. Dahlke, M. Fornasier, and T. Raasch, *Multilevel preconditioning for adaptive sparse optimization*, Preprint 25, DFG-SPP 1324, August 2009.

[10] ______, *Multilevel preconditioning and adaptive sparse solution of inverse problems*, Math. Comput. **81** (2012), no. 277, 419–446.

[11] W. Dahmen and R. Schneider, *Wavelets with complementary boundary conditions – functions spaces on the cube*, Result. Math. **34** (1998), 255–293.

[12] W. Dahmen, R. Schneider, and Y. Xu, *Nonlinear functionals of wavelet expansions – Adaptive reconstruction and fast evaluation*, Numer. Math. **86** (2000), no. 1, 49–101.

[13] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pure Appl. Math. **57** (2004), no. 11, 1413–1457.

[14] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Ann. Stat. **32** (2004), no. 2, 407–499.

[15] I. Ekeland and R. Témam, *Convex analysis and variational problems. Unabridged, corrected republication of the 1976 english original*, Philadelphia, PA: Society for Industrial and Applied Mathematics, 1999.

[16] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Mathematics and its Applications, vol. 375, Kluwer Academic Publishers Group, Dordrecht, 1996.

[17] M. Figueiredo and R. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process. **12** (2003), no. 8, 906–916.

[18] J. Kappei, *Adaptive frame methods for nonlinear elliptic problems*, Appl. Anal. **90** (2011), no. 7–8, 1323–1353.

[19] R. Ramlau and G. Teschke, *A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints*, Numer. Math. **104** (2006), no. 2, 177–203.

[20] J.-L. Starck, D.L. Donoho, and E.J. Candés, *Astronomical image representation by the curvelet transform*, A&A **398** (2003), no. 2, 785–800.

[21] J.-L. Starck, M.K. Nguyen, and F. Murtagh, *Wavelets and curvelets for image deconvolution: a combined approach*, Signal Process. **83** (2003), no. 10, 2279–2283.

[22] R. Stevenson, *On the compressibility of operators in wavelet coordinates*, SIAM J. Math. Anal. **35** (2004), no. 5, 1110–1132.

[23] G. Teschke and C. Borries, *Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints*, Inverse Probl. **26** (2010), no. 2, 23.

Stephan Dahlke, Ulrich Friedrich
Philipps-Universität Marburg
FB Mathematik und Informatik, AG Numerik
Hans-Meerwein-Strasse
35032 Marburg, Germany
{dahlke, friedrich}@mathematik.uni-marburg.de

Massimo Fornasier
TU München
Fakultät für Mathematik, Einheit M15: Angewandte Numerische Analysis
Boltzmannstrasse 3
85748 Garching b. München, Germany
massimo.fornasier@ma.tum.de

Thorsten Raasch
Johannes Gutenberg-Universität Mainz
Institut für Mathematik, AG Numerische Mathematik
Staudingerweg 9
55099 Mainz, Germany
raasch@uni-mainz.de

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multi-level Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.

[92] G. Kutyniok. Data Separation by Sparse Representations. Preprint 92, DFG-SPP 1324, May 2011.

[93] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing. Preprint 93, DFG-SPP 1324, May 2011.

[94] H.-C. Kreusler and H. Yserentant. The Mixed Regularity of Electronic Wave Functions in Fractional Order and Weighted Sobolev Spaces. Preprint 94, DFG-SPP 1324, June 2011.

[95] E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. Preprint 95, DFG-SPP 1324, June 2011.

[96] S. Kunis and I. Melzer. On the Butterfly Sparse Fourier Transform. Preprint 96, DFG-SPP 1324, June 2011.

[97] T. Rohwedder. The Continuous Coupled Cluster Formulation for the Electronic Schrödinger Equation. Preprint 97, DFG-SPP 1324, June 2011.

[98] T. Rohwedder and R. Schneider. Error Estimates for the Coupled Cluster Method. Preprint 98, DFG-SPP 1324, June 2011.

[99] P. A. Cioica and S. Dahlke. Spatial Besov Regularity for Semilinear Stochastic Partial Differential Equations on Bounded Lipschitz Domains. Preprint 99, DFG-SPP 1324, July 2011.

[100] L. Grasedyck and W. Hackbusch. An Introduction to Hierarchical (H-) Rank and TT-Rank of Tensors with Examples. Preprint 100, DFG-SPP 1324, August 2011.

[101] N. Chegini, S. Dahlke, U. Friedrich, and R. Stevenson. Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates. Preprint 101, DFG-SPP 1324, September 2011.

[102] S. Dahlke, P. Oswald, and T. Raasch. A Note on Quarkonial Systems and Multilevel Partition of Unity Methods. Preprint 102, DFG-SPP 1324, September 2011.

[103] A. Uschmajew. Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation. Preprint 103, DFG-SPP 1324, September 2011.

[104] S. Kvaal. Multiconfigurational time-dependent Hartree method for describing particle loss due to absorbing boundary conditions. Preprint 104, DFG-SPP 1324, September 2011.

[105] M. Guillemard and A. Iske. On Groupoid C*-Algebras, Persistent Homology and Time-Frequency Analysis. Preprint 105, DFG-SPP 1324, September 2011.

[106] A. Hinrichs, E. Novak, and H. Woźniakowski. Discontinuous information in the worst case and randomized settings. Preprint 106, DFG-SPP 1324, September 2011.

[107] M. Espig, W. Hackbusch, A. Litvinenko, H. Matthies, and E. Zander. Efficient Analysis of High Dimensional Data in Tensor Formats. Preprint 107, DFG-SPP 1324, September 2011.

[108] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization Problems in Contracted Tensor Networks. Preprint 108, DFG-SPP 1324, October 2011.

[109] S. Dereich, T. Müller-Gronbach, and K. Ritter. On the Complexity of Computing Quadrature Formulas for SDEs. Preprint 109, DFG-SPP 1324, October 2011.

[110] D. Belomestny. Solving optimal stopping problems by empirical dual optimization and penalization. Preprint 110, DFG-SPP 1324, November 2011.

[111] D. Belomestny and J. Schoenmakers. Multilevel dual approach for pricing American style derivatives. Preprint 111, DFG-SPP 1324, November 2011.

[112] T. Rohwedder and A. Uschmajew. Local convergence of alternating schemes for optimization of convex problems in the TT format. Preprint 112, DFG-SPP 1324, December 2011.

[113] T. Görner, R. Hielscher, and S. Kunis. Efficient and accurate computation of spherical mean values at scattered center points. Preprint 113, DFG-SPP 1324, December 2011.

[114] Y. Dong, T. Görner, and S. Kunis. An iterative reconstruction scheme for photoacoustic imaging. Preprint 114, DFG-SPP 1324, December 2011.

[115] L. Kämmerer. Reconstructing hyperbolic cross trigonometric polynomials by sampling along generated sets. Preprint 115, DFG-SPP 1324, February 2012.

[116] H. Chen and R. Schneider. Numerical analysis of augmented plane waves methods for full-potential electronic structure calculations. Preprint 116, DFG-SPP 1324, February 2012.

[117] J. Ma, G. Plonka, and M.Y. Hussaini. Compressive Video Sampling with Approximate Message Passing Decoding. Preprint 117, DFG-SPP 1324, February 2012.

[118] D. Heinen and G. Plonka. Wavelet shrinkage on paths for scattered data denoising. Preprint 118, DFG-SPP 1324, February 2012.

[119] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. Preprint 119, DFG-SPP 1324, March 2012.

[120] C. Bender and J. Steiner. A-posteriori estimates for backward SDEs. Preprint 120, DFG-SPP 1324, April 2012.

[121] M. Espig, W. Hackbusch, A. Litvinenkoy, H.G. Matthiesy, and P. Wähnert. Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats. Preprint 121, DFG-SPP 1324, May 2012.

[122] O. Bokanowski, J. Garcke, M. Griebel, and I. Klompmaker. An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations. Preprint 122, DFG-SPP 1324, June 2012.

[123] A. Mugler and H.-J. Starkloff. On the convergence of the stochastic Galerkin method for random elliptic partial differential equations. Preprint 123, DFG-SPP 1324, June 2012.

[124] P.A. Cioica, S. Dahlke, N. Döhring, U. Friedrich, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R.L. Schilling. On the convergence analysis of Rothe's method. Preprint 124, DFG-SPP 1324, July 2012.

[125] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Classification Algorithms using Adaptive Partitioning. Preprint 125, DFG-SPP 1324, July 2012.

[126] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation of hierarchical Tucker and Tensor-Train tensors. Preprint 126, DFG-SPP 1324, July 2012.

[127] M. Kovács, S. Larsson, and K. Urban. On Wavelet-Galerkin methods for semilinear parabolic equations with additive noise. Preprint 127, DFG-SPP 1324, August 2012.

[128] M. Bachmayr, H. Chen, and R. Schneider. Numerical analysis of Gaussian approximations in quantum chemistry. Preprint 128, DFG-SPP 1324, August 2012.

[129] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. Preprint 129, DFG-SPP 1324, August 2012.

[130] P.A. Cioica, K.-H. Kim, K. Lee, and F. Lindner. On the $L_q(L_p)$-regularity and Besov smoothness of stochastic parabolic equations on bounded Lipschitz domains. Preprint 130, DFG-SPP 1324, December 2012.

[131] M. Hansen. $n-$term Approximation Rates and Besov Regularity for Elliptic PDEs on Polyhedral Domains. Preprint 131, DFG-SPP 1324, December 2012.

[132] R. E. Bank and H. Yserentant. On the H^1 -stability of the L_2 -projection onto finite element spaces. Preprint 132, DFG-SPP 1324, December 2012.

[133] M. Gnewuch, S. Mayer, and K. Ritter. On Weighted Hilbert Spaces and Integration of Functions of Infinitely Many Variables. Preprint 133, DFG-SPP 1324, December 2012.

[134] D. Crisan, J. Diehl, P.K. Friz, and H. Oberhauser. Robust Filtering: Correlated Noise and Multidimensional Observation. Preprint 134, DFG-SPP 1324, January 2013.

[135] Wolfgang Dahmen, Christian Plesken, and Gerrit Welper. Double Greedy Algorithms: Reduced Basis Methods for Transport Dominated Problems. Preprint 135, DFG-SPP 1324, February 2013.

[136] Aicke Hinrichs, Erich Novak, Mario Ullrich, and Henryk Wozniakowski. The Curse of Dimensionality for Numerical Integration of Smooth Functions. Preprint 136, DFG-SPP 1324, February 2013.

[137] Markus Bachmayr, Wolfgang Dahmen, Ronald DeVore, and Lars Grasedyck. Approximation of High-Dimensional Rank One Tensors. Preprint 137, DFG-SPP 1324, March 2013.

[138] Markus Bachmayr and Wolfgang Dahmen. Adaptive Near-Optimal Rank Tensor Approximation for High-Dimensional Operator Equations. Preprint 138, DFG-SPP 1324, April 2013.

[139] Felix Lindner. Singular Behavior of the Solution to the Stochastic Heat Equation on a Polygonal Domain. Preprint 139, DFG-SPP 1324, May 2013.

[140] Stephan Dahlke, Dominik Lellek, Shiu Hong Lui, and Rob Stevenson. Adaptive Wavelet Schwarz Methods for the Navier-Stokes Equation. Preprint 140, DFG-SPP 1324, May 2013.

[141] Jonas Ballani and Lars Grasedyck. Tree Adaptive Approximation in the Hierarchical Tensor Format. Preprint 141, DFG-SPP 1324, June 2013.

[142] Harry Yserentant. A short theory of the Rayleigh-Ritz method. Preprint 142, DFG-SPP 1324, July 2013.

[143] M. Hefter and K. Ritter. On Embeddings of Weighted Tensor Product Hilbert Spaces. Preprint 143, DFG-SPP 1324, August 2013.

[144] M. Altmayer and A. Neuenkirch. Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts. Preprint 144, DFG-SPP 1324, August 2013.

[145] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate functions by trigonometric polynomials based on rank-1 lattice sampling. Preprint 145, DFG-SPP 1324, September 2013.

[146] C. Bender, N. Schweizer, and J. Zhuo. A primal-dual algorithm for BSDEs. Preprint 146, DFG-SPP 1324, October 2013.

[147] D. Rudolf. Hit-and-run for numerical integration. Preprint 147, DFG-SPP 1324, October 2013.

[148] D. Rudolf and M. Ullrich. Positivity of hit-and-run and related algorithms. Preprint 148, DFG-SPP 1324, October 2013.

[149] L. Grasedyck, M. Kluge, and S. Krämer. Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors. Preprint 149, DFG-SPP 1324, October 2013.

[150] F. Filbir, S. Kunis, and R. Seyfried. Effective discretization of direct reconstruction schemes for photoacoustic imaging in spherical geometries. Preprint 150, DFG-SPP 1324, November 2013.

[151] E. Novak, M. Ullrich, and H. Woźniakowski. Complexity of Oscillatory Integration for Univariate Sobolev Spaces. Preprint 151, DFG-SPP 1324, November 2013.

[152] A. Hinrichs, E. Novak, and M. Ullrich. A Tractability Result for the Clenshaw Curtis Smolyak Algorithm. Preprint 152, DFG-SPP 1324, November 2013.

[153] M. Hein, S. Setzer, L. Jost, and S. Rangapuram. The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited. Preprint 153, DFG-SPP 1324, November 2013.

[154] M. Kovács, S. Larsson, and F. Lindgren. On the Backward Euler Approximation of the Stochastic Allen-Chan Equation. Preprint 154, DFG-SPP 1324, November 2013.

[155] S. Dahlke, M. Fornasier, U. Friedrich, and T. Raasch. Multilevel preconditioning for sparse optimization of functionals with nonconvex fidelity terms. Preprint 155, DFG-SPP 1324, December 2013.