Multi–Frames in Thresholding Iterations for Nonlinear Operator Equations with Mixed Sparsity Constraints*

Gerd Teschke

June 29, 2005

Abstract

This paper is concerened with nonlinear inverse problems where the solution is assumed to have a sparse expansion with respect to several preassigned bases or frames. We develop a scheme which allows to minimize a Tikhonov functional where the usual quadratic regularization term is replaced by one–homogeneous (typically weighted ℓ_p, $1 \leq p \leq 2$) penalties on the coefficients (or isometrically transformed coefficients) of such multi–frame expansions. The computation of the solution amounts in this setting to a system of Landweber–fixed–point iterations with thresholding applied in each fixed–point iteration step.

1 Scope of the problem

We consider the computation of an approximation to a solution of a nonlinear operator equation

$$T(x) = y \ , \tag{1.1}$$

where $T : X \to Y$ is an operator between Hilbert spaces X, Y. In case of having only noisy data y^δ with $\|y^\delta - y\| \leq \delta$ available, there might be the problem of ill-posedness (in the sense of a discontinuous dependency of the solution on the data). Thus problem (1.1) has to be stabilized by regularization methods. In recent years, many of the well known methods for linear inverse problems have been generalized to nonlinear operator equations. But so far all the proposed schemes for nonlinear problems incorporate at most quadratic regularization whereas in many applications the solution is assumed to have sparse expansion which immediately leads to the involvement of nonquadratic penalties, e.g. ℓ_p norms with $p < 2$. In linear lore, this problem is still solved, see [2]. In nonlinear inverse problems there is an approach, see [5], which solves nonlinear operator equations with sparsity constraints. However, recent developments indicate that (higly) redundant systems, such as frames or systems of frames may yield a gain in this context (optimal representation/decomposition of the solution to be reconstructed). When dealing with dictionaries of frame systems, there exist certain methods, e.g. such as basis

*The author is with the Department of Mathematics, University of Bremen, Germany. G. T. was partially supported by Deutsche Forschungsgemeinschaft Grants TE 354/1-2, TE 354/3-1.

pursuit [1], that allow a decomposition of signals/functions into an optimal superposition of dictionary elements, where optimal means having smallest ℓ_1 norm of coefficients among all such decompositions. In [7], we have presented a method which combines an iterated thresholding scheme for solving linear inverse problems while requiring that the solution is assumed to have a sparse expansion in a multi–frame dictionary. In this paper, we also assume that the solution has a sparse expansion in a multi–frame dictionary but we aim now to extend the theory to nonlinear inverse problems with mixed multi–sparsity constraints. Thus the main result of this paper, coming out by combing results and technologies elaborated in [3, 4], [6, 5], and [7], is the development of a new method which is sort of thresholding Landweber iteration for solving a system of fixed point equations. This scheme is numerically illustrated by solving a few image processing task, but we also provide a regularization result which shows that this method is also well suited for ill-posed problems.

As in [7], let us assume we are given a finite family of preassigned frames $\{\phi^i_\lambda\}_{\lambda\in\Lambda_i, i\in\mathcal{I}} \subset X$, $n = \mathrm{card}(\mathcal{I})$, for which we have associated frame operators

$$F_i : X \to \ell_2 \text{ via } F_i x = \{\langle x, \phi^i_\lambda\rangle\}_{\lambda\in\Lambda_i} \text{ with } A_i \cdot I \le F_i^* F_i \le B_i \cdot I \ .$$

The variational formulation of the nonlinear inverse problem in a multi–frame setting with so–called multi–sparsity, or more general, multi–one–homogeneous constraints can be now casted as follows: find sequences of coefficients $\boldsymbol{g} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n) \in (\ell_2)^n$ such that

$$J_\alpha(\boldsymbol{g}) = \|y^\delta - T(K\boldsymbol{g})\|^2_Y + 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}) \tag{1.2}$$

is minimized, where $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\Psi_{\boldsymbol{L}}(\boldsymbol{g}) = (\Psi_1(\boldsymbol{L}_1\boldsymbol{g}_1), \ldots, \Psi_n(\boldsymbol{L}_n\boldsymbol{g}_n))$. In our case, $K\boldsymbol{g} = K(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n) = \sum_{i\in\mathcal{I}} F_i^*\boldsymbol{g}_i$, but one could also involve, as in [7], additional linear and bounded operators E_i, i.e. $K_E(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n) = \sum_{i\in\mathcal{I}} E_i F_i^* \boldsymbol{g}_i$. Moreover, the Ψ_i stand for positive, one–homogeneuos, lower semi–continuous and convex penalties (which are usually some weighted ℓ_p norms of the frame coefficients), and the infinite matrices $\boldsymbol{L}_i$ are restricted to be isometric mappings. In particular, we also need to require,

$$\|\boldsymbol{g}\|_{(\ell_2)^n} \le \|\Psi_{\boldsymbol{L}}(\boldsymbol{g})\|_{\ell_1} \ . \tag{1.3}$$

The strategies for nonlinear cases suggested in [6, 5], seem to be also adequate when dealing with multi–sparsity, or more general, with multi–one–homogeneous constraints. Before sketching the idea, we need to clarify the $(\ell_2)^n$–framework. First, for sake of simplicity, we restrict ourselves to $E_i = I$, for all i. Note that the suggested theory applies without any changes also to $E_i \neq I$. For the preassigned frame operators $F_i : X \to \ell_2$,

$$K : \ell_2 \times \ldots \times \ell_2 \to X \text{ via } (\ell_2)^n \ni \boldsymbol{g} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n) \mapsto \sum_{i\in\mathcal{I}} F_i^* \boldsymbol{g}_i \ ,$$

where the Hilbert space $(\ell_2)^n$ is endowed with the scalar $\langle \boldsymbol{g}, \boldsymbol{h}\rangle_{(\ell_2)^n} = \langle \boldsymbol{g}_1, \boldsymbol{h}_1\rangle_{\ell_2} + \ldots + \langle \boldsymbol{g}_n, \boldsymbol{h}_n\rangle_{\ell_2}$ and thus the associated norm is given by $\|\boldsymbol{g}\|^2_{(\ell_2)^n} = \|\boldsymbol{g}_1\|^2_{\ell_2} + \ldots + \|\boldsymbol{g}_n\|^2_{\ell_2}$. Moreover,

$$\langle K\boldsymbol{f}, h\rangle_X = \langle \boldsymbol{f}, (F_1 h, \ldots, F_n h)\rangle_{(\ell_2)^n} = \langle \boldsymbol{f}, K^* h\rangle_{(\ell_2)^n} \ ,$$

and thus,

$$\|K\| \le \sqrt{B_1 + \ldots + B_n} =: B \ .$$

The general idea for solving the nonlinear inverse problem in a multi–frame setting goes now as follows: we replace (1.2) by a sequence of functionals from which we hope that they are easier to treat and that the sequence of minimizers converge in some sense to, at least, a critical point of (1.2). To be more concrete, for $\boldsymbol{g} \in (\ell_2)^n$ and some auxiliary $\boldsymbol{a} \in (\ell_2)^n$, we introduce

$$J_\alpha^s(\boldsymbol{g}, \boldsymbol{a}) := J_\alpha(\boldsymbol{g}) + C\|\boldsymbol{g} - \boldsymbol{a}\|^2_{(\ell_2)^n} - \|T(K\boldsymbol{g}) - T(K\boldsymbol{a})\|^2_Y \tag{1.4}$$

and create an iteration process by:

1. Pick $\boldsymbol{g}_0 \in (\ell_2)^n$ and some proper constant $C > 0$
2. Derive a sequence $\{\boldsymbol{g}_k\}_{k=0,1,\dots}$ by the iteration:

$$\boldsymbol{g}_{k+1} = \arg \min_{\boldsymbol{g}_k \in (\ell_2)^n} J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k) \qquad k = 0, 1, 2, \dots$$

In order to avoid ambiguity, we will always denote $(\boldsymbol{g}_i)_k \in \ell_2$ as the k–th iterate of the i–th component of $\boldsymbol{g}_k \in (\ell_2)^n$, i.e. $(\ell_2)^n \ni \boldsymbol{g}_k = ((\boldsymbol{g}_1)_k, \dots, (\boldsymbol{g}_n)_k)$, and a particular coefficient of the k–th iterate with respect to some index $\lambda \in \Lambda_i$ is then denoted by $(\boldsymbol{g}_{\lambda,i})_k$; in its full glory we may thus write the k–th iterate $\boldsymbol{g}_k = (\{(\boldsymbol{g}_{\lambda,1})_k\}_{\lambda\in\Lambda_1}, \dots, \{(\boldsymbol{g}_{\lambda,n})_k\}_{\lambda\in\Lambda_n}) \in (\ell_2)^n$. As we shall see later on, in order to prove norm convergence of the iterates $\boldsymbol{g}_k$ towards a critical point of J_α, we have to restrict ourselves to a class of nonlinear problems for which all of the following three requirements hold true,

$$\begin{aligned} &\boldsymbol{g}_k \xrightarrow{w} \boldsymbol{g} \Longrightarrow T(K\boldsymbol{g}_k) \to T(K\boldsymbol{g}) \quad , \\ &F_j T'(K\boldsymbol{g}_k)^* z \to F_j T'(K\boldsymbol{g})^* z \quad \text{, for all } z \text{ and } j \quad , \\ &\|T'(K\boldsymbol{g}) - T'(K\boldsymbol{g}')\| \le LB\|\boldsymbol{g} - \boldsymbol{g}'\|_{(\ell_2)^n} \quad . \end{aligned} \tag{1.5}$$

It may happen that T already meets these conditions as an operator from $X \to Y$. If not, this can be achieved by assuming more regularity of the solution, i.e. changing the domain of T a little. To this end, we assume that there exists a function space X^s, and a compact embedding operator $i^s : X^s \to X$. Then we can consider $\tilde{T} = T \circ i^s : X^s \longrightarrow Y$. Lipschitz regularity is preserved. Moreover, if now $\boldsymbol{g}_k \xrightarrow{w} \boldsymbol{g}$ in X^s, then $\boldsymbol{g}_k \to \boldsymbol{g}$ in X and, moreover, $\tilde{T}'(K\boldsymbol{g}_k) \to \tilde{T}'(K\boldsymbol{g})$ in the operator norm. This argument applies to arbitrary nonlinear continuous and Fréchet differentiable operators $T : X \to Y$ with continuous Lipschitz derivative as long as a function space X^s with compact embedding i^s into X is available.

The remaining paper is organized as follows: In Section 2, we explain how the replacement functionals are constructed and discuss the well–posedness of the resulting problem. In Section 3, we derive conditions on the minimizing elements. The main result of the paper is presented in Section 4: strong convergence of the iterates towards a critical point. Moreover, in Section 5 we state conditions for which we may ensure that the scheme indeed has regularization properties. We end this paper with Section 6 in which we demonstrate the capabilities of the proposed scheme by solving nonlinear image processing tasks.

2 On the proper definition of the replacement functional

By the definition of J_α^s in (1.4) it is not clear whether the functional is positive definite or even bounded from below. This will be clarified in this section, i.e. we will show that this is the case provided the constant C is chosen properly.

For given multi–parameter $\alpha \in \mathbb{R}_+^n$ and $\boldsymbol{g}_0 \in (\ell_2)^n$ we may define a ball

$$K_r := \{\boldsymbol{g} \in (\ell_2)^n : \|\Psi_{\boldsymbol{L}}(\boldsymbol{g})\|_{\ell_1} \leq r\} \ ,$$

where the radius r is given by

$$r := J_\alpha(\boldsymbol{g}_0)/(2\min\{\alpha_i\}). \tag{2.1}$$

This obviously ensures, $\boldsymbol{g}_0 \in K_r$. Furthermore, we define the constant C by

$$C := 2B^2 \max\left\{ \left(\sup_{\boldsymbol{g}\in K_r} \|T'(K\boldsymbol{g})\| \right)^2, L\sqrt{\|y^\delta - T(K\boldsymbol{g}_0)\|^2 + 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_0)} \right\}, \tag{2.2}$$

where L is the Lipschitz constant of the Fréchet derivative of T. We assume that $\boldsymbol{g}_0$ was chosen such that $r < \infty$ and $C < \infty$.

Lemma 1 *Let r and C be chosen by* (2.1), (2.2). *Then, for all $\boldsymbol{g} \in K_r$,*

$$C\|\boldsymbol{g} - \boldsymbol{g}_0\|^2_{(\ell_2)^n} - \|T(K\boldsymbol{g}) - T(K\boldsymbol{g}_0)\|_Y^2 \geq 0 \tag{2.3}$$

and thus, $J_\alpha(\boldsymbol{g}) \leq J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_0)$.

Proof. By Taylors expansion we have

$$T(K\boldsymbol{g} + K\boldsymbol{h}) = T(K\boldsymbol{g}) + \int_0^1 T'(K\boldsymbol{g} + \tau K\boldsymbol{h})K\boldsymbol{h}\, d\tau$$

and thus we get with $\boldsymbol{h} = \boldsymbol{g}_0 - \boldsymbol{g}$

$$\begin{aligned} \|T(K\boldsymbol{g}) - T(K\boldsymbol{g}_0)\|_Y &\leq \int_0^1 \|T'(K\boldsymbol{g} + \tau K(\boldsymbol{g}_0 - \boldsymbol{g}))\| \|K(\boldsymbol{g}_0 - \boldsymbol{g})\|_X d\tau \\ &\leq \sup_{\boldsymbol{g}\in K_r} \|T'(K\boldsymbol{g})\| \|K(\boldsymbol{g}_0 - \boldsymbol{g})\|_X \\ &\leq \sup_{\boldsymbol{g}\in K_r} \|T'(K\boldsymbol{g})\| B \|\boldsymbol{g}_0 - \boldsymbol{g}\|_{(\ell_2)^n} \end{aligned}$$

Consequently, we get for all $\boldsymbol{g} \in K_r$

$$\begin{aligned} & C\|\boldsymbol{g} - \boldsymbol{g}_0\|^2_{(\ell_2)^n} - \|T(K\boldsymbol{g}) - T(K\boldsymbol{g}_0)\|_Y^2 \geq \\ & \qquad C\|\boldsymbol{g} - \boldsymbol{g}_0\|^2_{(\ell_2)^n} - B^2 \left(\sup_{\boldsymbol{g}\in K_r} \|T'(K\boldsymbol{g})\| \|\boldsymbol{g} - \boldsymbol{g}_0\|_{(\ell_2)^n} \right)^2 \\ = & \ \frac{C}{2}\|\boldsymbol{g} - \boldsymbol{g}_0\|^2_{(\ell_2)^n} \geq 0, \end{aligned}$$

and the functional $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_0)$ is non–negative for all $\boldsymbol{g} \in K_r$. □

Next, we show that this carries over to all of the iterates:

Proposition 2 *Let $\boldsymbol{g}_0, \alpha$ be given and r, C be defined by (2.1), (2.2). Then the functionals $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k)$ are bounded from below for all $k \in \mathbb{N}$ and have thus minimizers. For the minimizer $\boldsymbol{g}_{k+1}$ of $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k)$ holds $\boldsymbol{g}_{k+1} \in K_r$.*

Proof. The proof will be done by induction. For $k = 1$, we show in a first step that $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_0)$ is bounded from below. We have

$$\|y^\delta - T(K\boldsymbol{g})\|_Y^2 = \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 + \|T(K\boldsymbol{g}_0) - T(K\boldsymbol{g})\|_Y^2 + 2\langle y^\delta - T(K\boldsymbol{g}_0), T(K\boldsymbol{g}_0) - T(K\boldsymbol{g})\rangle_Y. \tag{2.4}$$

Thus,

$$\begin{aligned} J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_0) - 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}) &= \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 + 2\langle y^\delta - T(K\boldsymbol{g}_0), T(K\boldsymbol{g}_0) - T(K\boldsymbol{g})\rangle_Y \\ &\quad + C\|\boldsymbol{g} - \boldsymbol{g}_0\|_{(\ell_2)^n}^2 \qquad (2.5) \\ &\geq \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 - 2\|y^\delta - T(K\boldsymbol{g}_0)\|_Y \|T(K\boldsymbol{g}_0) - T(K\boldsymbol{g})\|_Y \\ &\quad + C\|\boldsymbol{g} - \boldsymbol{g}_0\|_{(\ell_2)^n}^2. \end{aligned} \tag{2.6}$$

Again by Taylor expansion,

$$\|T(K\boldsymbol{g}_0) - T(K\boldsymbol{g})\|_Y \leq B\|T'(K\boldsymbol{g}_0)\|\|\boldsymbol{g}_0 - \boldsymbol{g}\|_{(\ell_2)^n} + \frac{B^2 L}{2}\|\boldsymbol{g}_0 - \boldsymbol{g}\|_{(\ell_2)^n}^2. \tag{2.7}$$

Now let us assume that $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_0)$ is not bounded from below, e.g. there exists a sequence $\boldsymbol{g}_l$ such that $J_\alpha^s(\boldsymbol{g}_l, \boldsymbol{g}_0) \to -\infty$. This can only hold if $\|T(K\boldsymbol{g}_0) - T(K\boldsymbol{g}_l)\| \to \infty$, and because of (2.7) follows $\|\boldsymbol{g}_l\|_{(\ell_2)^n} \to \infty$ as well. In particular, for l large enough, we derive from (2.7)

$$\|T(K\boldsymbol{g}_0) - T(K\boldsymbol{g}_l)\|_Y \leq B^2 L\|\boldsymbol{g}_0 - \boldsymbol{g}_l\|_{(\ell_2)^n}^2,$$

and combining this estimate with (2.6) yields

$$J_\alpha^s(\boldsymbol{g}_l, \boldsymbol{g}_0) - 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_l) \geq \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 - 2B^2 L\|y^\delta - T(K\boldsymbol{g}_0)\|_Y \|\boldsymbol{g}_l - \boldsymbol{g}_0\|_{(\ell_2)^n}^2 + C\|\boldsymbol{g}_l - \boldsymbol{g}_0\|_{(\ell_2)^n}^2.$$

From the definition of C in (2.2) follows $2B^2 L\|y^\delta - T(K\boldsymbol{g}_0)\|_Y \leq C$ and thus

$$J_\alpha^s(\boldsymbol{g}_l, \boldsymbol{g}_0) - 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_l) \geq \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 \geq 0,$$

in contradiction to our assumption $J_\alpha^s(\boldsymbol{g}_l, \boldsymbol{g}_0) \to -\infty$, and thus $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_0)$ is bounded from below. By the same argument, we find $J_\alpha^s(\boldsymbol{g}_l, \boldsymbol{g}_0) \geq 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_l)$ for any sequence $\boldsymbol{g}_l$ with $\|\boldsymbol{g}_l\|_{(\ell_2)^n} \to \infty$, and by (1.3) we conclude $J_\alpha^s(\boldsymbol{g}_l, \boldsymbol{g}_0) \to \infty$, i.e. the functional is coercive and has a minimizer $\boldsymbol{g}_1$.

As in (2.6), we get by using (2.7),

$$\begin{aligned} J_\alpha^s(\boldsymbol{g}_1, \boldsymbol{g}_0) - 2\alpha\Psi(\boldsymbol{L}\boldsymbol{g}_1) &\geq \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 - 2B\|y^\delta - T(K\boldsymbol{g}_0)\|_Y \|T'(K\boldsymbol{g}_0)\|\|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{(\ell_2)^n} \\ &\quad - B^2 L\|y^\delta - T(K\boldsymbol{g}_0)\|_Y \|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{(\ell_2)^n}^2 + C\|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{(\ell_2)^n}^2 . \end{aligned}$$

By (2.2), $C/2 \geq B^2 L\|y^\delta - T(K\boldsymbol{g}_0)\|_Y$, and thus,

$$\begin{aligned} J_\alpha^s(\boldsymbol{g}_1, \boldsymbol{g}_0) - 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_1) &\geq \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 - 2B\|y^\delta - T(K\boldsymbol{g}_0)\|_Y \|T'(K\boldsymbol{g}_0)\|\|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{(\ell_2)^n} \\ &\quad + \frac{C}{2}\|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{(\ell_2)^n}^2. \end{aligned}$$

As $\boldsymbol{g}_0 \in K_r$, it follows from (2.2) that $B\|T'(K\boldsymbol{g}_0)\| \le \sqrt{C/2}$ holds, and consequently,

$$
\begin{aligned}
J_\alpha^s(\boldsymbol{g}_1, \boldsymbol{g}_0) - 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_1) &\ge \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 - 2\frac{\sqrt{C}}{\sqrt{2}}\|y^\delta - T(K\boldsymbol{g}_0)\|_Y\|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{(\ell_2)^n} \\
&\quad + \frac{C}{2}\|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{(\ell_2)^n}^2 \\
&= \left(\|y^\delta - T(K\boldsymbol{g}_0)\|_Y - \frac{\sqrt{C}}{\sqrt{2}}\|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{(\ell_2)^n}\right)^2 \ge 0.
\end{aligned}
$$

In particular,

$$
2\min\{\alpha_i\}\|\Psi_{\boldsymbol{L}}(\boldsymbol{g}_1)\|_{\ell_1} \le 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_1) \le J_\alpha^s(\boldsymbol{g}_1, \boldsymbol{g}_0) = \min_{\boldsymbol{g}} J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_0) \le J_\alpha^s(\boldsymbol{g}_0, \boldsymbol{g}_0) = J_\alpha(\boldsymbol{g}_0) \ ,
$$

i.e. $\|\Psi_{\boldsymbol{L}}(\boldsymbol{g}_1)\|_{\ell_1} \le J_\alpha(\boldsymbol{g}_0)/(2\min\{\alpha_i\}) = r$, and thus, $\boldsymbol{g}_1 \in K_r$. Next, thanks to Lemma 1,

$$
C\|\boldsymbol{g}_1 - \boldsymbol{g}_0\|_{\ell_2}^2 - \|T(K\boldsymbol{g}_1) - T(K\boldsymbol{g}_0)\|_Y^2 \ge 0 \quad \text{and } J_\alpha(\boldsymbol{g}_1) \le J_\alpha^s(\boldsymbol{g}_1, \boldsymbol{g}_0) \ ,
$$

and thus,

$$
\|y^\delta - T(K\boldsymbol{g}_1)\|_Y^2 \le J_\alpha(\boldsymbol{g}_1) \le J_\alpha^s(\boldsymbol{g}_1, \boldsymbol{g}_0) \le J_\alpha^s(\boldsymbol{g}_0, \boldsymbol{g}_0) \le \|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 + 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_0),
$$

and combining this estimate with the definition of C in (2.2) yields

$$
2B^2L\|y^\delta - T(K\boldsymbol{g}_1)\|_Y \le 2B^2L\sqrt{\|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 + 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_0)} \le C. \tag{2.8}
$$

For applying the induction step, assume that for all $i = 1, \cdots k-1$, the following properties hold true:

$$
\begin{aligned}
\boldsymbol{g}_i &\in K_r & (2.9)\\
C\|\boldsymbol{g}_i - \boldsymbol{g}_{i-1}\|_{(\ell_2)^n}^2 - \|T(K\boldsymbol{g}_i) - T(K\boldsymbol{g}_{i-1})\|_Y^2 &\ge 0 & (2.10)\\
2B^2L\|y^\delta - T(K\boldsymbol{g}_i)\|_Y &\le C, & (2.11)
\end{aligned}
$$

where $\boldsymbol{g}_i$ denotes a minimizer of the functional $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_{i-1})$. For $i = 1$, these properties have already been shown. As for the case $i = 1$, we have to show that the functional $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_{k-1})$ has a minimizer. First, we show that it is bounded from below: As in (2.6),

$$
\begin{aligned}
J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_{k-1}) - 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}) &\ge \|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y^2 \\
&\quad -2\|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y\|T(K\boldsymbol{g}_{k-1}) - T(K\boldsymbol{g})\|_Y + C\|\boldsymbol{g} - \boldsymbol{g}_{k-1}\|_{(\ell_2)^n}^2 \ .
\end{aligned}
$$

By Taylor expansion,

$$
\|T(K\boldsymbol{g}_{k-1}) - T(K\boldsymbol{g})\|_Y \le \|T'(K\boldsymbol{g}_{k-1})\|_Y\|\boldsymbol{g}_{k-1} - \boldsymbol{g}\|_{(\ell_2)^n} + \frac{B^2L}{2}\|\boldsymbol{g}_{k-1} - \boldsymbol{g}\|_{(\ell_2)^n}^2. \tag{2.12}
$$

Let us now assume that $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_{k-1})$ is not bounded from below. As in the case $k = 1$, there exists a sequence $\{\boldsymbol{g}_l\}_{l\in\mathbb{N}}$ with $\|\boldsymbol{g}_l\|_{(\ell_2)^n} \to \infty$ and $J_\alpha^s(\boldsymbol{g}_l, \boldsymbol{g}_{k-1}) \to -\infty$. In particular, for l large enough, follows from (2.12)

$$
\|T(K\boldsymbol{g}_{k-1}) - T(K\boldsymbol{g}_l)\|_Y \le B^2L\|\boldsymbol{g}_{k-1} - \boldsymbol{g}_l\|_{(\ell_2)^n}^2,
$$

and combining this estimate with (2.12) yields

$$\begin{aligned} J^s_\alpha(\boldsymbol{g}_l,\boldsymbol{g}_{k-1}) - 2\alpha\cdot\Psi_{\boldsymbol{L}}(\boldsymbol{g}_l) \;\ge\; & \|y^\delta - T(K\boldsymbol{g}_{k-1})\|^2_Y \\ & -2BL\|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y\|\boldsymbol{g}_l-\boldsymbol{g}_{k-1}\|^2_{(\ell_2)^n} + C\|\boldsymbol{g}_l-\boldsymbol{g}_{k-1}\|^2_{(\ell_2)^n}. \end{aligned}$$

By (2.11), $2B^2L\|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y \le C$ and thus

$$J^s_\alpha(\boldsymbol{g}_l,\boldsymbol{g}_{k-1}) - 2\alpha\cdot\Psi_{\boldsymbol{L}}(\boldsymbol{g}_l) \ge \|y^\delta - T(K\boldsymbol{g}_{k-1})\|^2 \ge 0,$$

in contradiction to our assumption $J^s_\alpha(\boldsymbol{g}_l,\boldsymbol{g}_{k-1}) \to -\infty$, and thus $J^s_\alpha(\boldsymbol{g},\boldsymbol{g}_{k-1})$ is bounded from below. By the same argument, we find $J^s_\alpha(\boldsymbol{g}_l,\boldsymbol{g}_{k-1}) \ge 2\alpha\cdot\Psi_{\boldsymbol{L}}(\boldsymbol{g}_l) \to \infty$ for any sequence $\boldsymbol{g}_l$ with $\|\boldsymbol{g}_l\|_{(\ell_2)^n} \to \infty$ and thus the functional is coercive and has a minimizer $\boldsymbol{g}_k$.
As in (2.12), we obtain

$$\begin{aligned} J^s_\alpha(\boldsymbol{g}_k,\boldsymbol{g}_{k-1}) - 2\alpha\cdot\Psi_{\boldsymbol{L}}(\boldsymbol{g}_k) \;\ge\; & \|y^\delta - T(K\boldsymbol{g}_{k-1})\|^2_Y \\ & -2B\|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y\|T'(K\boldsymbol{g}_{k-1})\|\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|_{(\ell_2)^n} \\ & -B^2L\|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|^2_{(\ell_2)^n} + C\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|^2_{(\ell_2)^n}. \end{aligned}$$

By (2.2) and assumption (2.11) we have $C/2 \ge B^2L\|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y$, and thus

$$\begin{aligned} J^s_\alpha(\boldsymbol{g}_k,\boldsymbol{g}_{k-1}) - 2\alpha\cdot\Psi_{\boldsymbol{L}}(\boldsymbol{g}_k) \;\ge\; & \|y^\delta - T(K\boldsymbol{g}_{k-1})\|^2_Y \\ & -2B\|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y\|T'(K\boldsymbol{g}_{k-1})\|\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|_{(\ell_2)^n} \\ & +\frac{C}{2}\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|^2_{(\ell_2)^n}. \end{aligned}$$

As $\boldsymbol{g}_{k-1}\in K_r$, it follows from (2.2) that $B\|T'(K\boldsymbol{g}_{k-1})\| \le \sqrt{C/2}$, and we consequently have

$$J^s_\alpha(\boldsymbol{g}_k,\boldsymbol{g}_{k-1}) - 2\alpha\cdot\Psi_{\boldsymbol{L}}(\boldsymbol{g}_k) \;\ge\; \left(\|y^\delta - T(K\boldsymbol{g}_{k-1})\|_Y - \frac{\sqrt{C}}{\sqrt{2}}\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|_{(\ell_2)^n}\right)^2 \ge 0.$$

In particular, it follows by (2.10),

$$\begin{aligned} 2\min\{\alpha_i\}\|\Psi_{\boldsymbol{L}}(\boldsymbol{g}_k)\|_{\ell_1} \;&\le\; 2\alpha\cdot\Psi_{\boldsymbol{L}}(\boldsymbol{g}_k) \le J^s_\alpha(\boldsymbol{g}_k,\boldsymbol{g}_{k-1}) = \min_{\boldsymbol{g}} J^s_\alpha(\boldsymbol{g},\boldsymbol{g}_{k-1}) \le J^s_\alpha(\boldsymbol{g}_{k-1},\boldsymbol{g}_{k-1}) \\ &=\; J^s_\alpha(\boldsymbol{g}_{k-1},\boldsymbol{g}_{k-2}) \le J^s_\alpha(\boldsymbol{g}_{k-2},\boldsymbol{g}_{k-2}) \le \cdots \le J^s_\alpha(\boldsymbol{g}_0,\boldsymbol{g}_0) \end{aligned}$$

i.e. $\|\Psi_{\boldsymbol{L}}(\boldsymbol{g}_k)\|_{\ell_1} \le J_\alpha(\boldsymbol{g}_0)/(2\min\{\alpha_i\}) = r$, and thus, $\boldsymbol{g}_k\in K_r$. As in Lemma 1, it follows

$$C\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|^2_{(\ell_2)^n} - \|T(K\boldsymbol{g}_k) - T(K\boldsymbol{g}_{k-1})\|^2_Y \ge 0$$

and

$$J_\alpha(\boldsymbol{g}) \le J^s_\alpha(\boldsymbol{g},\boldsymbol{g}_{k-1}),$$

and we obtain

$$\begin{aligned} \|y^\delta - T(K\boldsymbol{g}_k)\|^2_Y \;&\le\; J_\alpha(\boldsymbol{g}_k) \le J^s_\alpha(\boldsymbol{g}_k,\boldsymbol{g}_{k-1}) \le J^s_\alpha(\boldsymbol{g}_{k-1},\boldsymbol{g}_{k-1}) \le \cdots \le J^s_\alpha(\boldsymbol{g}_0,\boldsymbol{g}_0) \\ &=\; \|y^\delta - T(K\boldsymbol{g}_0)\|^2_Y + 2\alpha\cdot\Psi_{\boldsymbol{L}}(\boldsymbol{g}_0), \end{aligned} \tag{2.13}$$

and combining this estimate with the definition of C (2.2) yields

$$2B^2L\|y^\delta - T(K\boldsymbol{g}_k)\|_Y \le 2B^2L\sqrt{\|y^\delta - T(K\boldsymbol{g}_0)\|_Y^2 + 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}_0)} \le C, \tag{2.14}$$

i.e. we have shown that the assumptions (2.9)-(2.11) hold also for $i = k$. □

As an immediate consequence out of the latter proof we have

Corollary 3 *The sequences of functionals* $\{J_\alpha(\boldsymbol{g}_k)\}_{k=0,1,2,\ldots}$ *and* $\{J_\alpha^s(\boldsymbol{g}_{k+1}, \boldsymbol{g}_k)\}_{k=0,1,2,\ldots}$ *are non-increasing.*

3 Minimization of the replacement functional

In this section, we elaborate necessary conditions for a minimizer of the functional $J_\alpha^s(\boldsymbol{g}, \boldsymbol{a})$.

Lemma 4 *The necessary condition for a minimum of* $J_\alpha^s(\boldsymbol{g}, \boldsymbol{a})$ *is given by*

$$0 \in -F_jT'(K\boldsymbol{g})^*(y^\delta - T(K\boldsymbol{a})) + C\boldsymbol{g}_j - C\boldsymbol{a}_j + \alpha_j\boldsymbol{L}_j^*\partial\Psi_j(\boldsymbol{L}_j\boldsymbol{g}_j) \quad ,\text{for all } j = 1, \ldots, n \ . \tag{3.1}$$

Proof. In the notion of subgradients (which is allowed, see later on for a convexity result), we have for $j = 1, \ldots, n$,

$$\partial_j J_\alpha^s(\boldsymbol{g}, \boldsymbol{a}) = -2F_jT'(K\boldsymbol{g})^*(y^\delta - T(K\boldsymbol{a})) + 2C\boldsymbol{g}_j - 2C\boldsymbol{a}_j + 2\alpha_j\partial\Theta_j(\boldsymbol{g}_j) \ .$$

Consequently, through $\boldsymbol{v} \in \partial\Theta_j(\boldsymbol{g}_j) \Leftrightarrow \boldsymbol{L}_j\boldsymbol{v} \in \partial\Psi_j(\boldsymbol{L}_j\boldsymbol{g}_j)$, the necessary condition (3.1) follows immediately. □

Before giving an equivalent condition, we will have a closer look to the relation between the functionals Ψ_j and associated closed convex sets $\mathcal{C}_j$. We may consider the Fenchel or so–called dual functional of Ψ_j, which we will denote by Ψ_j^*. Since we have assumed Ψ_j to be a positive and one homogeneous functional, there exists a convex set $\mathcal{C}_j$ such that Ψ_j^* is equal to the indicator function $\chi_{\mathcal{C}_j}$ over $\mathcal{C}_j$. Moreover, in Hilbert space lore, we have total duality between convex sets and positive and one homogeneous functionals, i.e. $\Psi_j = (\chi_{\mathcal{C}_j})^*$.

Lemma 5 *Let* $M_j(\boldsymbol{g}, \boldsymbol{a}) := F_jT'(F^*\boldsymbol{g})^*(y^\delta - T(F^*\boldsymbol{a}))/C + \boldsymbol{a}_j$*, then the necessary conditions* (3.1) *can be casted as*

$$\boldsymbol{g}_j = \frac{\alpha_j}{C}\boldsymbol{L}_j^*\left(I - P_{\mathcal{C}_j}\right)\left(\frac{C}{\alpha_j}\boldsymbol{L}_jM_j(\boldsymbol{g}, \boldsymbol{a})\right) \quad , \qquad j = 1, \ldots, n \ . \tag{3.2}$$

where $P_{\mathcal{C}_j}$ *is the orthogonal projection onto the convex set* $\mathcal{C}_j$.

Proof. With the shorthand $M_j(\boldsymbol{g}, \boldsymbol{a})$ we may rewrite (3.1) for each j,

$$\boldsymbol{L}_j\frac{M_j(\boldsymbol{g}, \boldsymbol{a}) - \boldsymbol{g}_j}{\frac{\alpha_j}{C}} \in \partial\Psi_j(\boldsymbol{L}_j\boldsymbol{g}_j) \ ,$$

and thus, by standard arguments in convex analysis,

$$\frac{C}{\alpha_j}\boldsymbol{L}_j\boldsymbol{g}_j \in \frac{C}{\alpha_j}\partial\Psi_j^*\left(\boldsymbol{L}_j\frac{M_j(\boldsymbol{g}, \boldsymbol{a}) - \boldsymbol{g}_j}{\frac{\alpha_j}{C}}\right) \ .$$

In order to have an expression by means of projections (or generalized shrinkage operations), we expand the latter formula as follows,

$$\begin{aligned}
\boldsymbol{L}_j \frac{M_j(\boldsymbol{g},\boldsymbol{a})}{\frac{\alpha_j}{C}} &\in \boldsymbol{L}_j \frac{M_j(\boldsymbol{g},\boldsymbol{a}) - \boldsymbol{g}_j}{\frac{\alpha_j}{C}} + \frac{C}{\alpha_j}\partial\Psi_j^* \left(\boldsymbol{L}_j \frac{M_j(\boldsymbol{g},\boldsymbol{a}) - \boldsymbol{g}_j}{\frac{\alpha_j}{C}} \right) \\
&= \left(I + \frac{C}{\alpha_j}\partial\Psi_j^* \right) \left(\boldsymbol{L}_j \frac{M_j(\boldsymbol{g},\boldsymbol{a}) - \boldsymbol{g}_j}{\frac{\alpha_j}{C}} \right) ,
\end{aligned}$$

which is equivalent to

$$\left(I + \frac{C}{\alpha_j}\partial\Psi_j^* \right)^{-1} \left(\boldsymbol{L}_j \frac{M_j(\boldsymbol{g},\boldsymbol{a})}{\frac{\alpha_j}{C}} \right) = \boldsymbol{L}_j \frac{M_j(\boldsymbol{g},\boldsymbol{a}) - \boldsymbol{g}_j}{\frac{\alpha_j}{C}} .$$

Again, by standard results in convex analysis, it is known that $\left(I + \frac{C}{\alpha_j}\partial\Psi^* \right)^{-1}$ is nothing than the orthogonal projection onto the associated convex set $\mathcal{C}_j$, and hence the assertion follows,

$$\boldsymbol{g}_j = \frac{\alpha_j}{C}\boldsymbol{L}_j^*(I - P_{\mathcal{C}_j}) \left(\boldsymbol{L}_j \frac{M_j(\boldsymbol{g},\boldsymbol{a})}{\frac{\alpha_j}{C}} \right) .$$

□

The latter lemma states that for minimizing (1.4) we need to solve a system of n fixed point equations (3.2), which are nonlinearly coupled via the $P_{\mathcal{C}_j}$. To condense the notation a little, we introduce nonlinear operators (and call them generalized shinkage operators)

$$\mathbf{S}_j := \mathbf{S}_{\alpha_j,\boldsymbol{L}_j,\mathcal{C}_j} = \frac{\alpha_j}{C}\boldsymbol{L}_j^*(I - P_{\mathcal{C}_j})\boldsymbol{L}_j \frac{C}{\alpha_j} .$$

Thus, we may write

$$\boldsymbol{g} = (\mathbf{S}_1(M_1(\boldsymbol{g},\boldsymbol{a})), \ldots, \mathbf{S}_n(M_n(\boldsymbol{g},\boldsymbol{a}))) .$$

Let us now consider the associated fixed point map

$$\Phi(\boldsymbol{g},\boldsymbol{a}) = (\mathbf{S}_1(M_1(\boldsymbol{g},\boldsymbol{a})), \ldots, \mathbf{S}_n(M_n(\boldsymbol{g},\boldsymbol{a}))) .$$

Lemma 6 *For some generic $\boldsymbol{a}$, the operator $\Phi(\cdot,\boldsymbol{a})$ is a contraction if $B^2L/C\sqrt{J_\alpha(\boldsymbol{a})} < 1$, i.e.*

$$\|\Phi(\boldsymbol{g},\boldsymbol{a}) - \Phi(\tilde{\boldsymbol{g}},\boldsymbol{a})\|_{(\ell_2)^n} \leq q\|\boldsymbol{g} - \tilde{\boldsymbol{g}}\|_{(\ell_2)^n} \quad \textit{if} \quad q := \frac{B^2L}{C}\sqrt{J_\alpha(\boldsymbol{a})} < 1 .$$

Before proving this lemma, we need a result on projections onto convex sets.

Lemma 7 *Let K be a closed and convex set, then the mapping $I - P_K$ is non–expansive.*

This Lemma can be deduced by the following two standard properties of convex sets.

Lemma 8 *Let K be a closed and convex set in some Hilbert space H, then for all $u \in H$ and all $k \in K$ the inequality $\langle u - P_K u, k - P_K u\rangle \leq 0$ holds true.*

Proof. For all $\lambda \in [0,1]$ one has

$$\|u - ((1-\lambda)P_K u + \lambda k)\|^2 \geq \|u - P_K u\|^2 \ .$$

Thus, for all $\lambda \in [0,1]$

$$-2\lambda\langle u - P_K u, k - P_K u\rangle + \lambda^2\|k - P_K u\|^2 \geq 0,$$

and therewith we have $\langle u - P_K u, k - P_K u\rangle \leq 0$. $\square$

Lemma 9 *Let K be a closed and convex set, then for all $u, v \in H$ the inequality*

$$\|u - v - (P_K u - P_K v)\| \leq \|u - v\|$$

holds true.

Proof. We need to prove

$$-2\langle u - v, P_K u - P_K v\rangle + \|P_K u - P_K v\|^2 \leq 0 \ .$$

By Lemma 8 we have $\langle u - P_K u, P_K v - P_K u\rangle \leq 0$, or equivalently

$$-\langle u, P_K u\rangle + \langle u, P_K v\rangle + \|P_K u\|^2 - \langle P_K u, P_K v\rangle \leq 0 \ .$$

By symmetry we have

$$-\langle v, P_K v\rangle + \langle v, P_K u\rangle + \|P_K v\|^2 - \langle P_K v, P_K u\rangle \leq 0 \ .$$

Summing the two inequalities leads to

$$-\langle u - v, P_K u - P_K v\rangle + \|P_K u - P_K v\|^2 \leq 0 \ ,$$

and thus

$$-2\langle u - v, P_K u - P_K v\rangle + \|P_K u - P_K v\|^2 \leq -\|P_K u - P_K v\|^2 \leq 0 \ .$$

$\square$

Thanks to Lemma 9, we still have assured Lemma 7, and with Lemma 7 at hand, we are able to prove Lemma 6.

Proof. We have by Lemma 7 and the Lipschitz–continuity of T',

$$\begin{aligned}
\|\Phi(\boldsymbol{g},\boldsymbol{a}) - \Phi(\tilde{\boldsymbol{g}},\boldsymbol{a})\|^2_{(\ell_2)^n} &= \sum_{j=1}^n \|\mathbf{S}_j(\boldsymbol{g},\boldsymbol{a}) - \mathbf{S}_j(\boldsymbol{g},\boldsymbol{a})\|^2_{\ell_2} \\
&= \sum_{j=1}^n \frac{\alpha_j}{C}\left\|(I - P_{\mathcal{C}_j})\left(\boldsymbol{L}_j\frac{M_j(\boldsymbol{g},\boldsymbol{a})}{\frac{\alpha_j j}{C}}\right) - (I - P_{\mathcal{C}_j})\left(\boldsymbol{L}_j\frac{M(\tilde{\boldsymbol{g}},\boldsymbol{a})}{\frac{\alpha_j}{C}}\right)\right\|^2_{\ell_2} \\
&\leq \sum_{j=1}^n \|M_j(\boldsymbol{g},\boldsymbol{a}) - M_j(\tilde{\boldsymbol{g}},\boldsymbol{a})\|^2_{\ell_2} \\
&\leq \sum_{j=1}^n \frac{B_j}{C^2}\|T'(K\boldsymbol{g}) - T'(K\tilde{\boldsymbol{g}})\|^2\|y^\delta - T(K\boldsymbol{a})\|^2_Y \\
&\leq \sum_{j=1}^n \frac{B_j L^2}{C^2}\left(\sum_{i=1}^n B_i^{1/2}\|\boldsymbol{g}_i - \tilde{\boldsymbol{g}}_i\|_{\ell_2}\right)^2 J_\alpha(\boldsymbol{a}) \leq \frac{B^4 L^2}{C^2}\|\boldsymbol{g} - \tilde{\boldsymbol{g}}\|^2_{(\ell_2)^n} J_\alpha(\boldsymbol{a})
\end{aligned}$$

and the assertion follows. $\square$

Proposition 10 *The fixed point map $\Phi(\boldsymbol{g}, \boldsymbol{g}_k)$ is for all $k = 0, 1, 2, \dots$ a contraction.*

Proof. By the definition of C in (2.2) and Lemma 6 (setting $\boldsymbol{a} = \boldsymbol{g}_0$), we deduce that $\Phi(\boldsymbol{g}, \boldsymbol{g}_0)$ is a contraction with

$$q = \frac{B^2 L}{C}\sqrt{J_\alpha(\boldsymbol{g}_0)} \le \frac{1}{2} < 1.$$

With the help of Corollary 3, we complete the proof

$$\begin{aligned} \|\Phi(\boldsymbol{g}, \boldsymbol{g}_k) - \Phi(\tilde{\boldsymbol{g}}, \boldsymbol{g}_k)\|_{(\ell_2)^n} &\le \frac{B^2 L}{C}\sqrt{J_\alpha(\boldsymbol{g}_k)}\|\boldsymbol{g} - \tilde{\boldsymbol{g}}\|_{(\ell_2)^n} \\ &\le \dots \le \frac{B^2 L}{C}\sqrt{J_\alpha(\boldsymbol{g}_0)}\|\boldsymbol{g} - \tilde{\boldsymbol{g}}\|_{(\ell_2)^n} \le \frac{1}{2}\|\boldsymbol{g} - \tilde{\boldsymbol{g}}\|_{\ell_2}. \end{aligned}$$

$\square$

Up to here, we do know whether our fixed point iteration converges towards a critical point of $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k)$.

Proposition 11 *The necessary equation* (3.2) *for a minimum of the functional $J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k)$ has a unique fixed point, and the fixed point iteration converges towards the minimizer.*

Proof. To verify this assertion, we have to investigate the Taylor expansion of J_α^s more closely. By Taylor's expansion for T and the Lipschitz–continuity of T' we get

$$T(K\boldsymbol{g} + K\boldsymbol{h}) = T(K\boldsymbol{g}) + T'(K\boldsymbol{g})K\boldsymbol{h} + R(K\boldsymbol{g}, K\boldsymbol{h}) \tag{3.3}$$

with

$$\|R(K\boldsymbol{g}, K\boldsymbol{h})\|_Y \le \frac{B^2 L}{2}\|\boldsymbol{h}\|^2_{(\ell_2)^n}\ . \tag{3.4}$$

Denoting with ∇ the multi–valued (sub)gradient (still having in mind that the subgradient is set–valued) and with $\boldsymbol{g}_k$ the k–th iterate ($\boldsymbol{g}_j$ indicates the j–th component of $\boldsymbol{g}$),

$$\begin{aligned} J_\alpha^s(\boldsymbol{g} + \boldsymbol{h}, \boldsymbol{g}_k) - J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k) &= \nabla J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k)\cdot \boldsymbol{h} + C\|\boldsymbol{h}\|^2_{(\ell_2)^n} - 2\langle y^\delta - T(K\boldsymbol{g}_k), R(K\boldsymbol{g}, K\boldsymbol{h})\rangle_Y \\ &\quad + 2\sum_{j=1}^n \alpha_j\{\Theta_j(\boldsymbol{g}_j + \boldsymbol{h}_j) - \Theta(\boldsymbol{g}_j) - \partial\Theta_j(\boldsymbol{g}_j)\boldsymbol{h}_j\} \\ &\ge \nabla J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k)\cdot \boldsymbol{h} + C\|\boldsymbol{h}\|^2_{(\ell_2)^n} - 2\|y^\delta - T(K\boldsymbol{g}_k)\|_{\ell_2}\frac{B^2 L}{2}\|\boldsymbol{h}\|^2_{\ell_2} \\ &\quad + 2\sum_{j=1}^n \alpha_j\{\Theta_j(\boldsymbol{g}_j + \boldsymbol{h}_j) - \Theta(\boldsymbol{g}_j) - \partial\Theta_j(\boldsymbol{g}_j)\boldsymbol{h}_j\} \\ &\ge \nabla J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k)\cdot \boldsymbol{h} + \frac{C}{2}\|\boldsymbol{h}\|^2_{(\ell_2)^n} \\ &\quad + 2\sum_{j=1}^n \alpha_j\{\Theta_j(\boldsymbol{g}_j + \boldsymbol{h}_j) - \Theta(\boldsymbol{g}_j) - \partial\Theta_j(\boldsymbol{g}_j)\boldsymbol{h}_j\}. \end{aligned}$$

Assuming $\boldsymbol{g}$ is a critical point, i.e. $\nabla J_\alpha^s(\boldsymbol{g},\boldsymbol{g}_k)\cdot\boldsymbol{h}=0$ for all $\boldsymbol{h}$, we have

$$J_\alpha^s(\boldsymbol{g}+\boldsymbol{h},\boldsymbol{g}_k)-J_\alpha^s(\boldsymbol{g},\boldsymbol{g}_k)\geq\frac{C}{2}\|\boldsymbol{h}\|^2_{(\ell_2)^n}+2\sum_{j=1}^n\alpha_j\{\Theta_j(\boldsymbol{g}_j+\boldsymbol{h}_j)-\Theta(\boldsymbol{g}_j)-\partial\Theta_j(\boldsymbol{g}_j)\boldsymbol{h}_j\}\ .$$

By the definition of subgradients (for each individual j): an element $\boldsymbol{v}\in\ell_2$ belongs to $\partial\Theta_j(\boldsymbol{g}_j)$ if and only if for all $\boldsymbol{x}\in\ell_2$,

$$\Theta_j(\boldsymbol{g}_j)+\langle\boldsymbol{v},\boldsymbol{x}-\boldsymbol{g}_j\rangle_{\ell_2}\leq\Theta_j(\boldsymbol{x})\ ,$$

and, in particular for $\boldsymbol{x}=\boldsymbol{g}_j+\boldsymbol{h}_j$, this yields for all $\boldsymbol{v}\in\partial\Theta_j(\boldsymbol{g}_j)$ and all $\boldsymbol{h}_j\in\ell_2$,

$$\Theta_j(\boldsymbol{g}_j)+\langle\boldsymbol{v},\boldsymbol{h}_j\rangle_{\ell_2}\leq\Theta_j(\boldsymbol{g}_j+\boldsymbol{h}_j)\ \text{ or, equivalently, }\ 0\leq\Theta_j(\boldsymbol{g}_j+\boldsymbol{h}_j)-\Theta_j(\boldsymbol{g}_j)-\partial\Theta_j(\boldsymbol{g}_j)\boldsymbol{h}_j\ .$$

Consequently,

$$J_\alpha^s(\boldsymbol{g}+\boldsymbol{h},\boldsymbol{g}_k)-J_\alpha^s(\boldsymbol{g},\boldsymbol{g}_k)\geq\frac{C}{2}\|\boldsymbol{h}\|^2_{(\ell_2)^n}\ ,$$

and thus every critical point is a global minimizer of $J_\alpha^s(\boldsymbol{g},\boldsymbol{g}_k)$, and, again by the latter inequality, there exists only one global minimizer. □

By assuming more regularity on T, the latter statement can be improved:

Proposition 12 *Let T be a twice continuously differentiable operator. Then the functional $J_\alpha^s(\boldsymbol{g},\boldsymbol{g}_k)$ is strictly convex.*

Proof. Since the non–convex part of J_α^s is the discrepancy $\|y^\delta-T(K\boldsymbol{g})\|_Y^2$, it remains to show that

$$J^d(\boldsymbol{g}):=\|y^\delta-T(K\boldsymbol{g})\|_Y^2+C\|\boldsymbol{g}-\boldsymbol{g}_k\|^2_{\ell_2}-\|T(K\boldsymbol{g})-T(K\boldsymbol{g}_k)\|_Y^2 \tag{3.5}$$

is strictly convex in $\boldsymbol{g}$, i.e. we have to show that

$$J^d((1-\lambda)\boldsymbol{g}_1+\lambda\boldsymbol{g}_2)<(1-\lambda)J^d(\boldsymbol{g}_1)+\lambda J^d(\boldsymbol{g}_2)$$

holds for $\lambda\in(0,1)$ and arbitrary $\boldsymbol{g}_1,\boldsymbol{g}_2\in(\ell_2)^n$. At first, we express J^d by its Taylor expansion,

$$J^d(\boldsymbol{g}+\boldsymbol{h})=J^d(\boldsymbol{g})+\nabla J^d(\boldsymbol{g})\cdot\boldsymbol{h}+r(\boldsymbol{g},\boldsymbol{h})\ , \tag{3.6}$$

where

$$r(\boldsymbol{g},\boldsymbol{h}):=-2\langle y^\delta-T(K\boldsymbol{g}_k),R(K\boldsymbol{g},K\boldsymbol{h})\rangle_Y+C\|\boldsymbol{h}\|^2_{(\ell_2)^n}\ . \tag{3.7}$$

We have

$$\begin{aligned}J^d((1-\lambda)\boldsymbol{g}_1+\lambda\boldsymbol{g}_2)) &= J^d(\boldsymbol{g}_1+\lambda(\boldsymbol{g}_2-\boldsymbol{g}_1))=J^d(\boldsymbol{g}_2+(1-\lambda)(\boldsymbol{g}_1-\boldsymbol{g}_2))\\ &= (1-\lambda)J^d(\boldsymbol{g}_1+\lambda(\boldsymbol{g}_2-\boldsymbol{g}_1))+\lambda J^d(\boldsymbol{g}_2+(1-\lambda)(\boldsymbol{g}_1-\boldsymbol{g}_2))\end{aligned} \tag{3.8}$$

and with

$$\begin{aligned}J^d(\boldsymbol{g}_1+\lambda(\boldsymbol{g}_2-\boldsymbol{g}_1)) &= J^d(\boldsymbol{g}_1)+\lambda\nabla J^d(\boldsymbol{g}_1)\cdot(\boldsymbol{g}_2-\boldsymbol{g}_1)+r(\boldsymbol{g}_1,\lambda(\boldsymbol{g}_2-\boldsymbol{g}_1))\\ J^d(\boldsymbol{g}_2+(1-\lambda)(\boldsymbol{g}_1-\boldsymbol{g}_2)) &= J^d(\boldsymbol{g}_2)+(1-\lambda)\nabla J^d(\boldsymbol{g}_2)\cdot(\boldsymbol{g}_1-\boldsymbol{g}_2)+r(\boldsymbol{g}_2,(1-\lambda)(\boldsymbol{g}_1-\boldsymbol{g}_2))\end{aligned}$$

we obtain

$$\begin{aligned} J^d((1-\lambda)\boldsymbol{g}_1+\lambda\boldsymbol{g}_2)) &= (1-\lambda)J^d(\boldsymbol{g}_1)+\lambda J^d(\boldsymbol{g}_2)+\lambda(1-\lambda)\left[\nabla J^d(\boldsymbol{g}_1)-\nabla J^d(\boldsymbol{g}_2)\right]\cdot(\boldsymbol{g}_2-\boldsymbol{g}_1) \\ &\quad +(1-\lambda)r(\boldsymbol{g}_1,\lambda(\boldsymbol{g}_2-\boldsymbol{g}_1))+\lambda r(\boldsymbol{g}_2,(1-\lambda)(\boldsymbol{g}_1-\boldsymbol{g}_2))\ . \end{aligned}$$

Thus, J^s_α is strictly convex if for all $\lambda\in(0,1)$,

$$\begin{aligned} D(\boldsymbol{g}_1,\boldsymbol{g}_2,\lambda) &:= \lambda(1-\lambda)\left[\nabla J^d(\boldsymbol{g}_1)-\nabla J^d(\boldsymbol{g}_2)\right]\cdot(\boldsymbol{g}_2-\boldsymbol{g}_1) \\ &\quad +(1-\lambda)r(\boldsymbol{g}_1,\lambda(\boldsymbol{g}_2-\boldsymbol{g}_1))+\lambda r(\boldsymbol{g}_2,(1-\lambda)(\boldsymbol{g}_1-\boldsymbol{g}_2))<0\ . \end{aligned}$$

We have

$$\begin{aligned} \left[\nabla J^d(\boldsymbol{g}_1)-\nabla J^d(\boldsymbol{g}_2)\right]\cdot(\boldsymbol{g}_2-\boldsymbol{g}_1) &= -2C\|\boldsymbol{g}_2-\boldsymbol{g}_1\|^2_{(\ell_2)^n} \\ &\quad -2\langle y^\delta-T(K\boldsymbol{g}_k),(T'(K\boldsymbol{g}_1)-T'(K\boldsymbol{g}_2))K(\boldsymbol{g}_2-\boldsymbol{g}_1)\rangle_Y. \end{aligned}$$

As T is twice continuously Fréchet differentiable, it is

$$T'(K\boldsymbol{g}_1)=T'(K\boldsymbol{g}_2)+\int_0^1 T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2),\cdot)\,d\tau$$

and thus,

$$\begin{aligned} &\left[\nabla J^d(\boldsymbol{g}_1)-\nabla J^d(\boldsymbol{g}_2)\right]\cdot(\boldsymbol{g}_2-\boldsymbol{g}_1)= \\ &\quad -2C\|\boldsymbol{g}_2-\boldsymbol{g}_1\|^2_{(\ell_2)^n}+2\langle y^\delta-T(K\boldsymbol{g}_k),\int_0^1 T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2))^2d\tau\rangle, \end{aligned} \tag{3.9}$$

where we have used the shorthand $T''(\cdot)(\cdot,\cdot)=T''(\cdot)(\cdot)^2$. Again, as T is twice continuously Fréchet-differentiable, the function $R(K\boldsymbol{g},K\boldsymbol{h})$ in (3.7) is given by

$$R(K\boldsymbol{g},K\boldsymbol{h})=\int_0^1(1-\tau)T''(K\boldsymbol{g}+\tau K\boldsymbol{h})(K\boldsymbol{h})^2\,d\tau\ ,$$

and thus we obtain

$$\begin{aligned} R(K\boldsymbol{g}_1,\lambda K(\boldsymbol{g}_2-\boldsymbol{g}_1)) &= \lambda^2\int_0^1(1-\tau)T''(K\boldsymbol{g}_1+\tau\lambda K(\boldsymbol{g}_2-\boldsymbol{g}_1))(K(\boldsymbol{g}_2-\boldsymbol{g}_1))^2\,d\tau \\ &= \int_{1-\lambda}^1(\tau-(1-\lambda))T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2))^2\,d\tau \end{aligned} \tag{3.10}$$

and in the same way

$$R(K\boldsymbol{g}_2,(1-\lambda)K(\boldsymbol{g}_1-\boldsymbol{g}_2)) = \int\limits_0^{1-\lambda}(1-\lambda-\tau)T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2)^2\,d\tau\ . \quad (3.11)$$

Combining definition (3.7) and equations (3.9), (3.10) and (3.11) yields

$$D(\boldsymbol{g}_1,\boldsymbol{g}_2,\lambda) = -\lambda(1-\lambda)C\|\boldsymbol{g}_1-\boldsymbol{g}_2\|^2_{(\ell_2)^n} + 2\langle y^\delta - T(K\boldsymbol{g}_k), f(\boldsymbol{g}_1,\boldsymbol{g}_2,\lambda)\rangle_Y\ , \qquad (3.12)$$

where

$$\begin{aligned} f(\boldsymbol{g}_1,\boldsymbol{g}_2,\lambda) \;&:=\; \lambda(1-\lambda)\int\limits_0^1 T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2))^2\,d\tau \\ &\quad -(1-\lambda)\int\limits_{1-\lambda}^1(\tau-(1-\lambda))T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2))^2\,d\tau \\ &\quad -\lambda\int\limits_0^{1-\lambda}(1-\lambda-\tau)T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2)^2\,d\tau\ . \end{aligned}$$

The functional $f(\boldsymbol{g}_1,\boldsymbol{g}_2,\lambda)$ can now be recasted as follows

$$\begin{aligned} f(x_1,x_2,\lambda) \;&=\; \lambda\int\limits_0^{1-\lambda}\tau T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2))^2\,d\tau \\ &\qquad +(1-\lambda)\int\limits_{1-\lambda}^1(1-\tau)T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2))^2\,d\tau. \end{aligned}$$

In order to estimate $\|f(\boldsymbol{g}_1,\boldsymbol{g}_2,\lambda)\|_Y$ it is necessary to estimate the integrals separately. Due to the Lipschitz–continuity of the first derivative, the second derivative can be globally estimated by L, and it follows,

$$\begin{aligned} \lambda\left\|\int\limits_0^{1-\lambda}\tau T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2))^2\,d\tau\right\|_Y \;&\le\; \lambda\frac{(1-\lambda)^2}{2}B^2L\|\boldsymbol{g}_1-\boldsymbol{g}_2\|^2_{(\ell_2)^n} \\ (1-\lambda)\left\|\int\limits_{1-\lambda}^1(1-\tau)T''(K\boldsymbol{g}_2+\tau K(\boldsymbol{g}_1-\boldsymbol{g}_2))(K(\boldsymbol{g}_1-\boldsymbol{g}_2))^2\,d\tau\right\|_Y \;&\le\; (1-\lambda)\frac{\lambda^2}{2}B^2L\|\boldsymbol{g}_1-\boldsymbol{g}_2\|^2_{(\ell_2)^n} \end{aligned}$$

and thus

$$\|f(\boldsymbol{g}_1,\boldsymbol{g}_2,\lambda)\|_Y \le \frac{\lambda(1-\lambda)}{2}B^2L\|\boldsymbol{g}_1-\boldsymbol{g}_2\|^2_{(\ell_2)^n}\ . \qquad (3.13)$$

Combining (3.12) and (3.13) yields for $\lambda \in (0,1)$

$$\begin{aligned}
D(\boldsymbol{g}_1, \boldsymbol{g}_2, \lambda) &\leq -\lambda(1-\lambda)C\|\boldsymbol{g}_1 - \boldsymbol{g}_2\|^2_{(\ell_2)^n} + 2\|y^\delta - T(K\boldsymbol{g}_k)\|_Y \|f(\boldsymbol{g}_1, \boldsymbol{g}_2, \lambda)\|_Y \\
&\leq -\lambda(1-\lambda)C\|\boldsymbol{g}_1 - \boldsymbol{g}_2\|^2_{(\ell_2)^n} + \frac{\lambda(1-\lambda)}{2} 2B^2L\|y^\delta - T(K\boldsymbol{g}_k)\|_Y \|\boldsymbol{g}_1 - \boldsymbol{g}_2\|^2_{(\ell_2)^n} \\
&\overset{(2.14)}{\leq} -\lambda(1-\lambda)\frac{C}{2}\|\boldsymbol{g}_1 - \boldsymbol{g}_2\|^2_{(\ell_2)^n} < 0 \ ,
\end{aligned}$$

and thus the functional is strictly convex. □

4 Convergence properties of the iteration

Within this section we discuss convergence properties of the proposed scheme, i.e. we aim to show that the sequence of iterates $\{\boldsymbol{g}_k\}$ converges strongly towards a critical point of J_α, at least.

Lemma 13 *The sequence of iterates $\{\boldsymbol{g}_k\}_{k=0,1,2,\ldots}$ has a weakly convergent subsequence.*

Proof. This is an immediate consequence of Proposition 2, in which we have shown that for $k = 0,1,2,\ldots$ the iterates $\boldsymbol{g}_k$ are contained in K_r, and by requirement (1.3), $\|\boldsymbol{g}_k\|_{(\ell_2)^n} \leq r$. Since the iterates are uniformly bounded, we deduce that there exists at least one accumulation point $\boldsymbol{g}^\star_\alpha$ with $\boldsymbol{g}_{k,l} \stackrel{w}{\longrightarrow} \boldsymbol{g}^\star_\alpha$, where $\boldsymbol{g}_{k,l}$ denotes a subsequence of $\boldsymbol{g}_k$. □

Lemma 14 *For the iterates $\boldsymbol{g}_k$ holds $\lim_{k\to\infty} \|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|_{(\ell_2)^n} = 0$.*

Proof. With the help of Corollary 3, we observe that

$$\begin{aligned}
0 &\leq \sum_{k=0}^{N} \left\{ C\|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|^2_{(\ell_2)^n} - \|T(K\boldsymbol{g}_{k+1}) - T(K\boldsymbol{g}_k)\|^2_Y \right\} \\
&\leq \sum_{k=0}^{N} \left\{ J^s_\alpha(\boldsymbol{g}_{k+1}, \boldsymbol{g}_k) - J_\alpha(\boldsymbol{g}_{k+1}) \right\} \leq \sum_{k=0}^{N} \left\{ J_\alpha(\boldsymbol{g}_k) - J_\alpha(\boldsymbol{g}_{k+1}) \right\} \\
&= J_\alpha(\boldsymbol{g}_0) - J_\alpha(\boldsymbol{g}_{N+1}) \leq J_\alpha(\boldsymbol{g}_0) \ ,
\end{aligned}$$

i.e. the finite sums are uniformly bounded (independent on N). Now, by the Taylor expansion of T, we have

$$\|T(K\boldsymbol{g}_{k+1}) - T(K\boldsymbol{g}_k)\|^2_Y \leq \frac{C}{2}\|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|^2_{(\ell_2)^n} \ ,$$

and thus

$$0 \leq \frac{C}{2}\|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|^2_{(\ell_2)^n} \leq C\|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|^2_{(\ell_2)^n} - \|T(K\boldsymbol{g}_{k+1}) - T(K\boldsymbol{g}_k)\|^2_Y \longrightarrow 0$$

as $k \to \infty$ and the assertion follows. □

Lemma 15 *Every subsequence of $\boldsymbol{g}_k$ has a convergent subsequence $\boldsymbol{g}_{k,l}$ that converges strongly towards a function $\boldsymbol{g}_\alpha^\star$, and $\boldsymbol{g}_\alpha^\star$ satisfies the necessary condition for a minimizer of J_α:*

$$F_jT'(K\boldsymbol{g}_\alpha^\star)^*(y^\delta - T(K\boldsymbol{g}_\alpha^\star)) \in \alpha_j\partial\Theta_j((\boldsymbol{g}_j)_\alpha^\star)\ , \quad j = 1,\ldots,n\ . \tag{4.1}$$

Proof. According to Lemma 4, the minimizer $\boldsymbol{g}_{k+1}$ of $J_\alpha^s(\boldsymbol{g},\boldsymbol{g}_k)$ fulfills

$$0 \in F_jT'(K\boldsymbol{g}_{k+1})^*(y^\delta - T(K\boldsymbol{g}_k)) - C(\boldsymbol{g}_j)_{k+1} + C(\boldsymbol{g}_j)_k - \alpha_j\partial\Theta_j((\boldsymbol{g}_j)_{k+1}).$$

Thus, for all $j = 1,\ldots,n$,

$$\begin{aligned}(\boldsymbol{g}_j)_{k+1} - (\boldsymbol{g}_j)_k \ \in\ & \frac{1}{C}\big(F_jT'(K\boldsymbol{g}_{k+1})^*(y^\delta - T(K\boldsymbol{g}_{k+1})) - \alpha_j\partial_j\Theta_j((\boldsymbol{g}_j)_{k+1}) \\ & \quad +F_jT'(K\boldsymbol{g}_{k+1})^*(T(K\boldsymbol{g}_{k+1}) - T(K\boldsymbol{g}_k))\big)\end{aligned} \tag{4.2}$$

and, moreover, by Lemma 14,

$$\|F_jT'(K\boldsymbol{g}_{k+1})^*(T(K\boldsymbol{g}_{k+1}) - T(K\boldsymbol{g}_k))\|_Y \le \frac{CB_j^{1/2}}{2B}\|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|_{(\ell_2)^n} \to 0\ .$$

Passing to the limit $k \to \infty$ in (4.2),

$$0 \in \lim_{k\to\infty}\big(F_jT'(K\boldsymbol{g}_{k+1})^*(y^\delta - T(K\boldsymbol{g}_{k+1})) - \alpha_j\partial\Theta_j((\boldsymbol{g}_j)_{k+1})\big)\ . \tag{4.3}$$

Since $\boldsymbol{g}_k$ is bounded, every subsequence has a weakly convergent subsequence. Let $\boldsymbol{g}_{k,l}$ denote such a weakly convergent subsequence with weak limit $\boldsymbol{g}_\alpha^\star$ (for simplicity, we will denote this sequence by $\boldsymbol{g}_k$, too). Since

$$\begin{aligned}&F_jT'(K\boldsymbol{g}_{k+1})^*(y^\delta - T(K\boldsymbol{g}_{k+1})) = \\ &\quad F_jT'(K\boldsymbol{g}_{k+1})^*(y^\delta - T(K\boldsymbol{g}_\alpha^\star)) + F_jT'(K\boldsymbol{g}_{k+1})^*(T(K\boldsymbol{g}_\alpha^\star) - T(K\boldsymbol{g}_{k+1}))\ ,\end{aligned}$$

and because of

$$\|F_jT'(K\boldsymbol{g}_{k+1})^*(T(K\boldsymbol{g}_\alpha^\star - T(K\boldsymbol{g}_{k+1}))\|_{\ell_2} \le \frac{\sqrt{CB_j}}{\sqrt{2}B}\|T(K\boldsymbol{g}_\alpha^\star) - T(K\boldsymbol{g}_{k+1})\| \to 0$$

and by assumption (1.5), i.e.

$$F_jT'(K\boldsymbol{g}_{k+1})^*(y^\delta - T(K\boldsymbol{g}_\alpha^\star)) \to F_jT'(K\boldsymbol{g}_\alpha^\star)^*(y^\delta - T(K\boldsymbol{g}_\alpha^\star)),$$

we consequently obtain

$$\lim_{k\to\infty} F_jT'(K\boldsymbol{g}_{k+1})^*(y^\delta - T(K\boldsymbol{g}_{k+1})) = F_jT'(K\boldsymbol{g}_\alpha^\star)^*(y^\delta - T(K\boldsymbol{g}_\alpha^\star))\ . \tag{4.4}$$

Next, we have to consider $\lim_{k\to\infty}\partial\Theta_j((\boldsymbol{g}_j)_k)$. By an elementwise consideration we have, $\boldsymbol{v} \in \partial\Theta_j((\boldsymbol{g}_j)_k)$ if and only if for all $\boldsymbol{x} \in \ell_2$ the inequality $\Theta_j(\boldsymbol{x}) \ge \Theta_j((\boldsymbol{g}_j)_k) + \langle \boldsymbol{v}, \boldsymbol{x} - (\boldsymbol{g}_j)_k\rangle_{\ell_2}$ holds true. The assumption that Θ_j is lower semi-continuous and convex implies weak lower semi-continuity of all the Θ_j, i.e. $\Theta_j((\boldsymbol{g}_j)_\alpha^\star) \le \lim_{k\to\infty}\inf\Theta_j((\boldsymbol{g}_j)_k) \le \lim_{k\to\infty}\Theta_j((\boldsymbol{g}_j)_k)$. The same holds true for the ℓ_2–inner product. Thus, we deduce that for all $\boldsymbol{v} \in \lim_{k\to\infty}\partial\Theta_j((\boldsymbol{g}_j)_k)$

we have $\boldsymbol{v} \in \partial\Theta_j((\boldsymbol{g}_j)^\star_\alpha)$, i.e. $\lim_{k\to\infty} \partial\Theta_j((\boldsymbol{g}_j)_k) \subseteq \partial\Theta_j((\boldsymbol{g}_j)^\star_\alpha)$. Combining (4.4) with (4.2) proves that $\boldsymbol{g}_{k,l}$ converges, and as $\boldsymbol{g}^\star_\alpha$ is the weak limit of the sequence, $\boldsymbol{g}_{k,l} \to \boldsymbol{g}^\star_\alpha$. Equations (4.1) follow by passing to the limit in (4.3). $\square$

In principle, the limits of different convergent subsequences of $\boldsymbol{g}_k$ may differ. Let $\boldsymbol{g}_{k,l} \to \boldsymbol{g}^\star_\alpha$ be a subsequence of $\boldsymbol{g}_k$, and let $\tilde{\boldsymbol{g}}_{k,l}$ the predecessor of $\boldsymbol{g}_{k,l}$ in $\boldsymbol{g}_k$, i.e. $\boldsymbol{g}_{k,l} = \boldsymbol{g}_i$ and $\tilde{\boldsymbol{g}}_{k,l} = \boldsymbol{g}_{i-1}$. Then we observe, $J^s_\alpha(\boldsymbol{g}_{k,l}, \tilde{\boldsymbol{g}}_{k,l}) \to J_\alpha(\boldsymbol{g}^\star_\alpha)$. Moreover, as we have $J^s_\alpha(\boldsymbol{g}_{k+1}, \boldsymbol{g}_k) \le J^s_\alpha(\boldsymbol{g}_k, \boldsymbol{g}_{k-1})$ for all k, it turns out that the value of the Tikhonov functional for every limit $\boldsymbol{g}^\star_\alpha$ of a convergent subsequence remains the same, i.e. $J_\alpha(\boldsymbol{g}^\star_\alpha) = const$.

We may now summarize our findings and give a simple criterion that ensures strong convergence of the whole sequence $\{\boldsymbol{g}_k\}$.

Theorem 16 *Assume that there exists at least one isolated limit $\boldsymbol{g}^\star_\alpha$ of a subsequence $\boldsymbol{g}_{k,l}$ of $\boldsymbol{g}_k$. Then $\boldsymbol{g}_k \to \boldsymbol{g}^\star_\alpha$ as $k \to \infty$. The accumulation point $\boldsymbol{g}^\star_\alpha$ is a minimizer for the functional $J^s_\alpha(\boldsymbol{g}, \boldsymbol{g}^\star_\alpha)$.*

Proof. As in the proof of Proposition 11 we obtain, $J^s_\alpha(x^\star_\alpha + h, x^\star_\alpha) \ge J^s_\alpha(x^\star_\alpha, x^\star_\alpha) + \frac{C}{2}\|h\|^2$. The remaining proof of norm convergence can be directly taken from [6]. $\square$

5 A Regularization result

After stating norm convergence results for the proposed multi–frame approach for solving nonlinear operator equations, we now focus on how to optimally choose the parameter vector α. In a typical situation of an inverse problem, i.e. considering noisy data or equivalently the ill–posed case, the vector α plays the most important role in computing stabilized solutions. In this case, if the 'error' $e = y^\delta - T(K\boldsymbol{g}^\dagger)$ tends to zero, we wish our estimate for the solution of the inverse problem tend to $\boldsymbol{g}^\dagger$, since the minimizer of $J_\alpha(\boldsymbol{g})$ differs from $\boldsymbol{g}^\dagger$ for $\alpha \ne 0$. In inverse problems lore, this means to identify a functional relation between α and the noise floor δ, i.e. $\alpha = \alpha(\delta)$ with $\alpha(\delta) \to 0$ and $\|\boldsymbol{g}^\star_{\alpha(\delta)} - \boldsymbol{g}^\dagger\| \to 0$ as $\alpha \to 0$. If we find a parameter rule achieving this, then the suggested iteration scheme would regularize the ill–posed problem. However, in our context we have to face the fact that $\mathcal{N}(K)$ is nontrivial as long as we deal with frames, i.e. even if the inverse problem would have a unique solution, the corresponding vector of frame sequences to represent this solution will never have. Thus it is only reasonable to show that we approach one solution $\boldsymbol{g}^\dagger$ when passing to the limit $\delta \to 0$.

The next theorem provides conditions on the functional relation $\alpha(\delta)$ for which the constructed Landweber fixed point iteration with projections in each step is a regularization scheme (up to uniqueness).

Theorem 17 *Let $y^\delta \in Y$ with $\|y^\delta - y\|_Y \le \delta$, $\alpha_{\min}(\delta) = \min_j\{\alpha_j(\delta)\}$, $\alpha_{\max}(\delta) = \max_j\{\alpha_j(\delta)\}$, and assume $\alpha(\delta) = (\alpha_1(\delta), \ldots, \alpha_n(\delta))$ is chosen such that*

$$\alpha(\delta) \xrightarrow{\delta\to 0} 0\ , \quad \delta^2/\alpha_{\min}(\delta) \xrightarrow{\delta\to 0} 0\ , \quad \alpha_{\max}(\delta)/\alpha_{\min}(\delta) \xrightarrow{\delta\to 0} 1\ .$$

Then every sequence $\{\boldsymbol{g}^\star_{\alpha(\delta)}\}$ of minimizers of the functional $J_\alpha(\boldsymbol{g})$ where $\delta \to 0$ and $\alpha = \alpha(\delta)$ has a convergent subsequence. The limit of every convergent subsequence is a solution of $T(K\boldsymbol{g}) = y$ with minimal values of $\Psi(\boldsymbol{L}_j \boldsymbol{g}_j)$, $j = 1, \ldots, n$.

Proof. As $\boldsymbol{g}^\star_{\alpha(\delta)} = ((\boldsymbol{g}^\star_{\alpha(\delta)})_1, \ldots, (\boldsymbol{g}^\star_{\alpha(\delta)})_n)$ is a minimizer of J_α, we have

$$\|y^\delta - T(K\boldsymbol{g}^\star_{\alpha(\delta)})\|^2_Y + 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}^\star_{\alpha(\delta)}) \le \delta^2 + 2\alpha \cdot \Psi_{\boldsymbol{L}}(\boldsymbol{g}^\dagger) \ . \tag{5.1}$$

Thus, by the made assumptions on $\alpha(\delta)$, we achieve

$$\lim_{\delta\to 0} T(K\boldsymbol{g}^\star_{\alpha(\delta)}) = y \ .$$

Again by (5.1),

$$\|\Psi_{\boldsymbol{L}}(\boldsymbol{g}^\star_{\alpha(\delta)})\|_{\ell_1} \le \frac{\delta^2}{2\alpha_{\min}(\delta)} + \frac{\alpha_{\max}(\delta)}{\alpha_{\min}(\delta)} \|\Psi_{\boldsymbol{L}}(\boldsymbol{g}^\dagger)\|_{\ell_1}$$

implying,

$$\limsup_{\delta\to 0} \|\boldsymbol{g}^\star_{\alpha(\delta)}\|_{(\ell_2)^n} \le \limsup_{\delta\to 0} \|\Psi_{\boldsymbol{L}}(\boldsymbol{g}^\star_{\alpha(\delta)})\|_{\ell_1} \le \|\Psi_{\boldsymbol{L}}(\boldsymbol{g}^\dagger)\|_{\ell_1} \ ,$$

i.e. $\|\boldsymbol{g}^\star_{\alpha(\delta)}\|_{(\ell_2)^n}$ are uniformly bounded. Consequently, the sequence has a weakly convergent subsequence (again denoted by $\{\boldsymbol{g}^\star_{\alpha(\delta)}\}$) with weak limit $\boldsymbol{g}^\circ$,

$$\boldsymbol{g}^\circ = w - \lim_{\delta\to 0} \boldsymbol{g}^\star_{\alpha(\delta)} \ .$$

Since T is strongly continuous,

$$y = \lim_{\delta\to 0} T(K\boldsymbol{g}^\star_{\alpha(\delta)}) = T(K\boldsymbol{g}^\circ) \ ,$$

i.e. $\boldsymbol{g}^\circ$ is a solution of $T(K\boldsymbol{g}) = y$. Assume now $\boldsymbol{g}^\dagger$ is a solution of the inverse problem with minimal values of $\Psi_j(\boldsymbol{L}_j \ \cdot \ )$. Then, since all the Ψ_j are weak semi–continuous, we deduce

$$\Psi_j(\boldsymbol{L}_j \boldsymbol{g}^\circ_j) \le \limsup_{\delta\to 0} \Psi_j(\boldsymbol{L}_j (\boldsymbol{g}^\star_{\alpha(\delta)})_j) \le \Psi_j(\boldsymbol{L}_j \boldsymbol{g}^\dagger_j) \le \Psi_j(\boldsymbol{L}_j \boldsymbol{g}^\circ_j) \text{ for } j = 1, \ldots, n.$$

Hence $\boldsymbol{g}^\circ$ is also a solution with minimal values of $\Psi_j(\boldsymbol{L}_j \ \cdot \ )$. □

Resulting (regularization) iteration:

We may now summarize our findings and suggest the following regularization method. Assume that all the conditions we have imposed in the previous sections apply to our problem and, moreover, assume we have a parameter rule at hand that fulfills the conditions of Theorem 17. Then the algorithm goes as follows:

- Define a sequence $\{\alpha_n\}$ satifying the condition of Theorem 17, and pick $r \ge 1$, $\boldsymbol{g}_0$
- **while** $\|y^\delta - T(K\boldsymbol{g}^\star_{\alpha(\delta)})\| > r \cdot \delta$

- $\alpha = \alpha_n$
- pick an admissible C
- $[\boldsymbol{g}_\alpha^\star]$ = Iteration(T, y^δ, C, α, $\boldsymbol{g}_0$):

  $\boldsymbol{g}_{k+1} = \arg\min_{\boldsymbol{g}} J_\alpha^s(\boldsymbol{g}, \boldsymbol{g}_k)$ (solved by a projected fixed point iteration)

  $\boldsymbol{g}_\alpha^\star = \lim_{k\to\infty} \boldsymbol{g}_k$
- $\boldsymbol{g}_0 = \boldsymbol{g}_\alpha^\star$

end

In practice (treatment of limits), we have to incorporate stopping rules that will slightly modify this scheme:

- Define a sequence $\{\alpha_n\}$ satifying the condition of Theorem 17, and pick $r \geq 1$, $\boldsymbol{g}_0$, and additionally two tolerances τ_1, τ_2
- **while** $\|y^\delta - T(K\boldsymbol{g}^\star_{\alpha(\delta)})\| > r \cdot \delta$
  - $\alpha = \alpha_n$
  - pick an admissible C
  - $[\boldsymbol{g}_\alpha^\star]$ = Iteration(T, y^δ, C, α, τ_1, τ_2)

```
k = 0
while ||g_{k+1} - g_k||_{ℓ2} > τ1
      l = 0,  g_{k,0} = g_k
      while ||g_{k,l} - g_{k,l+1}||_{ℓ2} > τ2
          l = l + 1
          g_{k,l} = Φ_{α,C}(g_{k,l-1}, g_k)
          end
      g_{k+1} = g_{k,l}
      k = k + 1
end
g*_α = g_k
```

end

6 A numerical illustration

In this section, we apply the iterative machinery for solving nonlinear problems in a multi frame setting. For illustration purposes we focus on a sequence of synthetic nonlinear problems in the field of signal and image processing.

The first example is devoted to nonlinear image deformation. As the synthetic nonlinear operator we consider

$$T(x) = \cos(x) \ .$$

Assuming our image is given by some $x \in L_2(\Omega)$, where $\Omega = [0,1]^2$, then T is applied to each value $x(k,l)$, for all $(k,l) \in \Omega$,

$$T(x(k,l)) = \cos(x(k,l)).$$

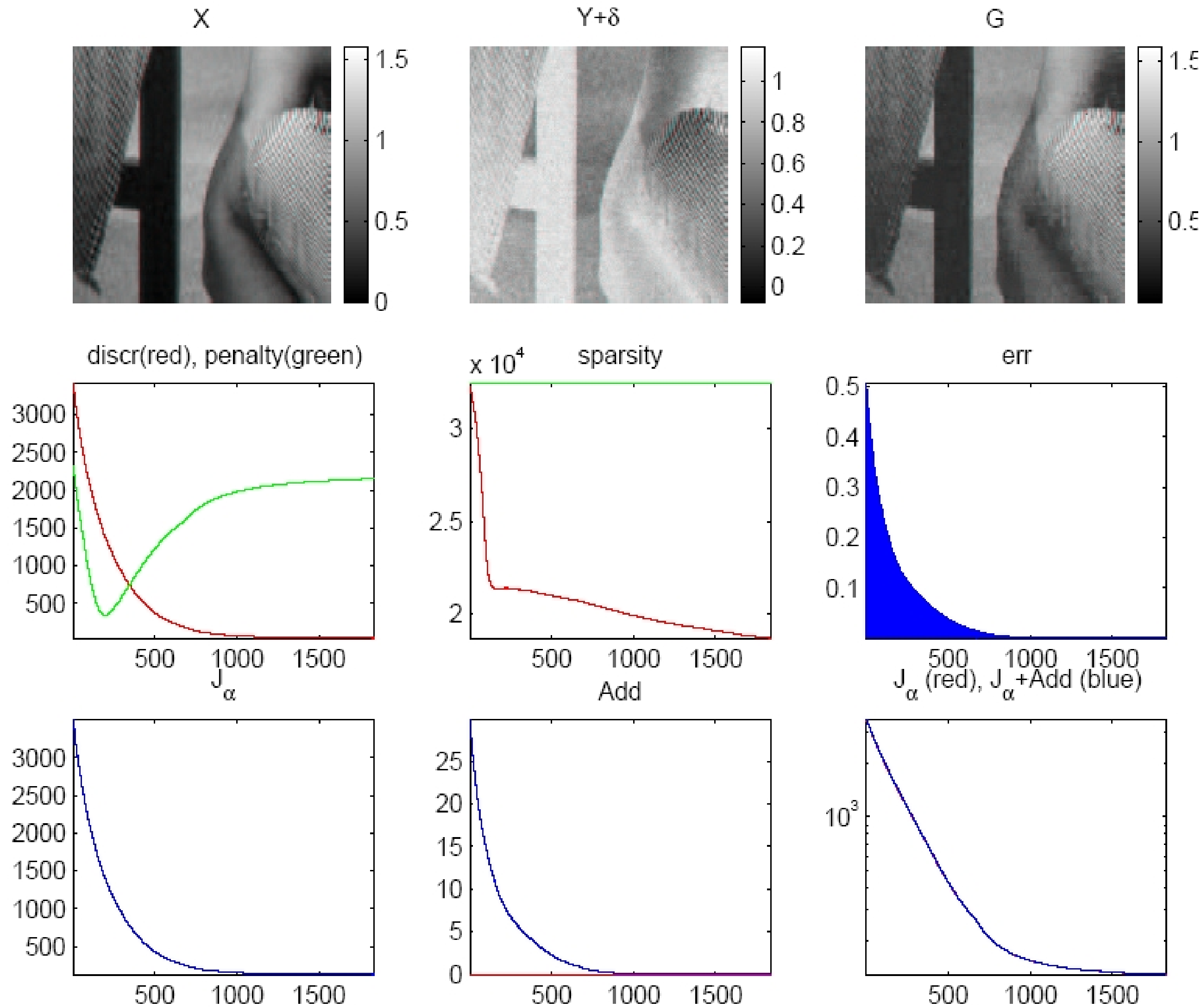

Figure 1: Thresholding Landweber fixed point iteration for the pixel basis and orthogonal Haar wavelet basis $\boldsymbol{L} = W$ and sparsity parameter $\alpha = 0.02$. From top left to up right: original image x; $T(x) + \delta = y^\delta$; final reconstruction of the solution; values of $\|y^\delta - T(F^*\boldsymbol{g})\|^2_{L_2(\Omega)}$ (red) and $|W\boldsymbol{g}|_{\ell_1}$ (green) during the whole iteration process; sparsity history (red, green indicates the reference to original total number of coefficients); error plot; J_α; Gaussian surrogate term; J_α (red) and $J^s_\alpha = J_\alpha$+'Gaussian surrogate term' (blue).

As the frame under consideration we chose the pixel basis with frame operator F and $x = F^*\boldsymbol{g}$ for some $\boldsymbol{g} \in \ell_2$. Moreover, we aim to reconstruct an image while requiring sparsity. Sparsity can be achieved when setting $p = 1$. However, we still know that sparsity cannot be well achieved when dealing with a pixel frame. Hence, it would be more feasible to switch to a wavelet frame (basis) when penalizing the approximation, i.e. we set $\boldsymbol{L} = W$ and W denoting

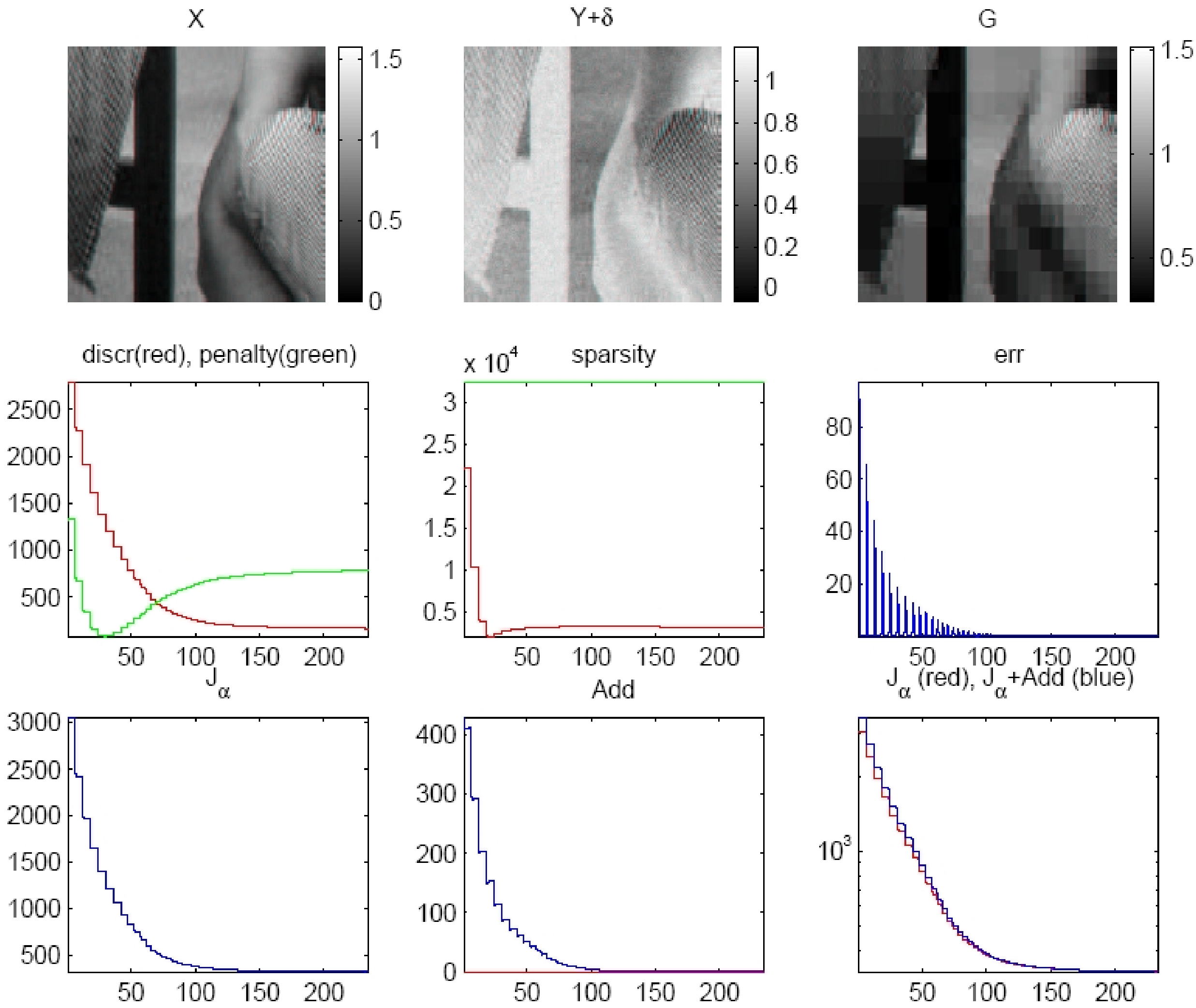

Figure 2: Thresholding Landweber fixed point iteration for the pixel basis and orthogonal Haar wavelet basis $\boldsymbol{L} = W$ and sparsity parameter $\alpha = 0.1$. From top left to up right: original image x; $T(x) + \delta = y^\delta$; final reconstruction of the solution; values of $\|y^\delta - T(F^*\boldsymbol{g})\|^2_{L_2(\Omega)}$ (red) and $|W\boldsymbol{g}|_{\ell_1}$ (green) during the whole iteration process; sparsity history (red, green indicates the reference to original total number of coefficients); error plot; J_α; Gaussian surrogate term; J_α (red) and $J^s_\alpha = J_\alpha$+'Gaussian surrogate term' (blue).

the orthogonal wavelet transform. Consequently, we may cast the problem as follows,

$$J_\alpha(\boldsymbol{g}) = \|y^\delta - T(F^*\boldsymbol{g})\|^2_{L_2(\Omega)} + 2\alpha|W\boldsymbol{g}|_{\ell_1} \ .$$

The resulting Landweber iteration is then based on solving the following fixed point equation in each step,

$$\boldsymbol{g}_{k+1} = \frac{\alpha}{C} W^*(I - P_{\mathcal{C}})(\frac{C}{\alpha} W M(\boldsymbol{g}_{k+1}, \boldsymbol{g}_k)) = \mathbf{S}_{\alpha,W,\mathcal{C}}(M(\boldsymbol{g}_{k+1}, \boldsymbol{g}_k)) \ ,$$

i.e. for each Landweber iteration we have to perform a fixed point iteration with a generalized shrinkage projection applied in each step

$$\boldsymbol{g}_{k+1,l+1} = \mathbf{S}_{\alpha,W,\mathcal{C}}(M(\boldsymbol{g}_{k+1,l}, \boldsymbol{g}_k)) \ .$$

We finally need to derive the generalized shrinkage operator $\mathbf{S}_{\alpha,\boldsymbol{L},\mathcal{C}}$. Since $p = 1$,

$$\Psi(W\boldsymbol{g}) = |W\boldsymbol{g}|_{\ell_1} = \sum_{\lambda\in\Lambda} |(W\boldsymbol{g})_\lambda| \ ,$$

the related convex set is then nothing else than

$$\mathcal{C} = \{W\boldsymbol{g} \in \ell_2 : \ \sup_{\lambda\in\Lambda} |(W\boldsymbol{g})_\lambda| \leq 1\} \ .$$

This yields the componentwise acting projection $P_{\mathcal{C}}(W\boldsymbol{g}) = \{P_{\mathcal{C}}((W\boldsymbol{g})_\lambda)\}_{\lambda\in\Lambda}$ with

$$P_{\mathcal{C}}((W\boldsymbol{g})_\lambda) = \begin{cases} (W\boldsymbol{g})_\lambda & \text{if } \ |(W\boldsymbol{g})_\lambda| \leq 1 \\ \operatorname{sgn}(W\boldsymbol{g})_\lambda & \text{if } \ |(W\boldsymbol{g})_\lambda| > 1 \end{cases} ,$$

where $\operatorname{sgn}(0) \in [-1, 1]$ and consequently,

$$(I - P_{\mathcal{C}})((W\boldsymbol{g})_\lambda) = \begin{cases} 0 & \text{if } \ |(W\boldsymbol{g})_\lambda| \leq 1 \\ \operatorname{sgn}(W\boldsymbol{g})_\lambda(|(W\boldsymbol{g})_\lambda| - 1) & \text{if } \ |(W\boldsymbol{g})_\lambda| > 1 \end{cases} .$$

This is the well–known soft shrinkage operation with threshold 1, which we denote here by S_1. Thus,

$$\boldsymbol{g}_{k+1,l+1} = S_{\alpha/C}(M(\boldsymbol{g}_{k+1,l}, \boldsymbol{g}_k)) \ .$$

The numerical results for two different parameters α are shown in Figures 1 and 2. In Figure 1 we have chosen $\alpha = 0.02$, in Figure 2, $\alpha = 0.1$. We may clearly observe that we achieve much better sparsity in the second case whereas the approximation quality is much higher in the first example, and that the number of iterations becomes less when α increases.

In the second illustration, we really compute a reconstruction when dealing with multi frames. For computational reasons we consider a one dimensional synthetic data set, see top left diagram in Figure 3. The nonlinearity comes into play by setting

$$y = T(x) = e^{-x} \ .$$

Our frame dictionary consists now of two different bases: Daubechies wavelet bases of order one (Haar wavelet basis) and ten. We denote the corresponding frame operators by F_1 and F_2, then

$$x = K\boldsymbol{g} = K(\boldsymbol{g}_1, \boldsymbol{g}_2) = F_1^*\boldsymbol{g}_1 + F_2^*\boldsymbol{g}_2 \ .$$

Moreover, we again aim to reconstruct a sparse solution of the inverse deformation problem. Since we still deal with a wavelet based dictionary it is customary to set $\boldsymbol{L}_1 = \boldsymbol{L}_2 = I$. The variational problem to be minimized reads then as

$$J_\alpha(\boldsymbol{g}) = J_\alpha(\boldsymbol{g}_1, \boldsymbol{g}_2) = \|y^\delta - T(K(\boldsymbol{g}_1, \boldsymbol{g}_2))\|^2_{L_2(\Omega)} + 2\alpha_1|\boldsymbol{g}_1|_{\ell_1} + 2\alpha_2|\boldsymbol{g}_2|_{\ell_1} \ .$$

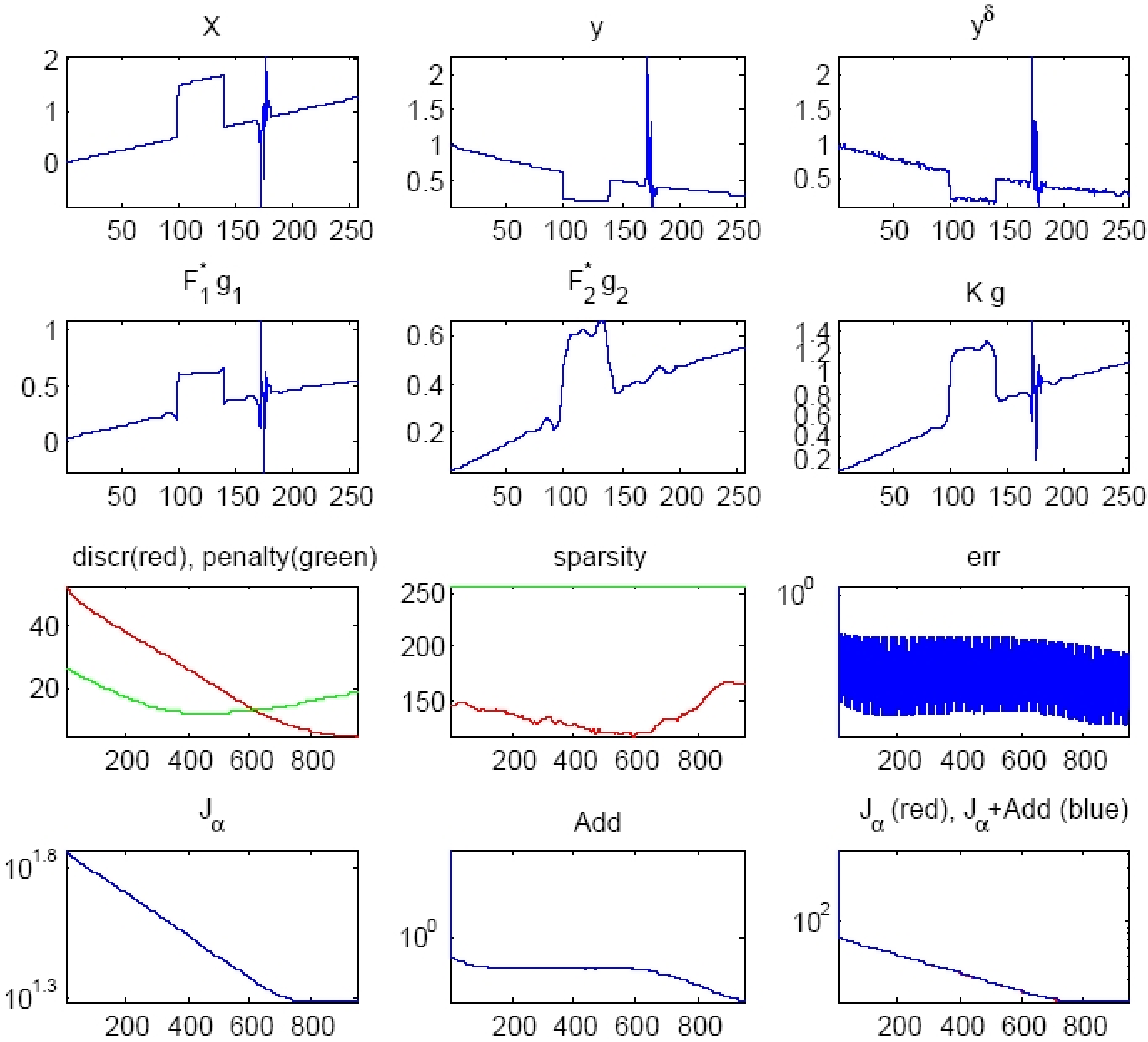

Figure 3: Thresholding Landweber fixed point iteration for a wavelet based dictionary ($F_1 \sim$ Haar system, $F_2 \sim$ Daubechies wavelet basis of order ten) and sparsity parameters $\alpha = (0.2, 0.5)$. From top left to up right: original data x; $T(x) = y$; $T(x) + \delta = y^\delta$; final Haar reconstruction; final Db10 reconstruction; final overall reconstruction; values of $\|y^\delta - T(F^*\boldsymbol{g})\|^2_{L_2(\Omega)}$ (red) and $|\boldsymbol{g}_1|_{\ell_1} + |\boldsymbol{g}_2|_{\ell_1}$ (green) during the whole iteration process; sparsity history (red, green indicates the reference to original total number of coefficients); error plot; J_α; Gaussian surrogate term; J_α (red) and $J^s_\alpha = J_\alpha$+'Gaussian surrogate term' (blue).

Hence, the resulting system of fixed point equations (to be solved in the same manner as before) is given by

$$
\begin{aligned}
(\boldsymbol{g}_1)_{k+1} &= S_{\alpha_1/C}(M_1(\boldsymbol{g}_{k+1}, \boldsymbol{g}_k)) \\
(\boldsymbol{g}_2)_{k+1} &= S_{\alpha_2/C}(M_2(\boldsymbol{g}_{k+1}, \boldsymbol{g}_k)) \quad .
\end{aligned}
$$

The results are visualized in Figure 3. The main observation is that we may indeed reconstruct with the proposed scheme an approximation of x. Moreover, we see that the different components of x are at most complementary covered by the two different frames: the Haar system essentially grabs the non–smooth part whereas the Db10 family describes smoother components of x. Of course, we must admit that the information is not completely split, i.e. there is still some redundant information in $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$. However, the reconstructed approximation of x requires even by using two bases much less coefficients (approx. 160 coefficients) than the original data set (256 coefficients).

References

[1] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *Report Stanford University*, 1995.

[2] I. Daubechies, M. Defrise, and C. DeMol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57:1413–1541, 2004.

[3] I. Daubechies and G. Teschke. Wavelet–based image decomposition by variational functionals. *Proc. SPIE Vol. 5266, p. 94-105, Wavelet Applications in Industrial Processing; Frederic Truchetet; Ed.*, Feb. 2004.

[4] I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising. *Applied and Computational Harmonic Analysis*, (19(1)):1–16, 2005.

[5] R. Ramlau and G. Teschke. A Thresholding Iteration for Nonlinear Operator Equations with Sparsity Constraints. *DFG-SPP-1114 preprint (submitted for journal publication)*, 2005.

[6] R. Ramlau and G. Teschke. Tikhonov Replacement Functionals for Iteratively Solving Nonlinear Operator Equations. *DFG-SPP-1114 preprint 85 (to appear in Inverse Problems)*, 2005.

[7] G. Teschke. Multi-Frame Representations in Linear Inverse Problems with Mixed Multi-Constraints. *DFG-SPP-1114 preprint 90 (submitted for journal publication)*, 2005.