Multi–frame representations in linear inverse problems with mixed multi–constraints

Gerd Teschke*

May 18, 2006

Abstract

This paper is concerned with linear inverse problems where the solution is assumed to have a sparse expansion with respect to several bases or frames. We were mainly motivated by the following two different approaches: 1st) B. Torrésani et.al. (2004, 2005) has suggested to represent audio signals by means of at least a wavelet for transient and a local cosine dictionary for tonal components. The suggested technology produces sparse representations of audio signals that are very efficient in audio coding. 2nd) Also quite recently, I. Daubechies et. al. (2003) has developed an iterative method for linear inverse problems that promote a sparse representation for the solution to be reconstructed. Here in this paper, we bring both ideas together and construct schemes for linear inverse problems where the solution might then have a sparse representation (we also allow smoothness constraints) with respect to several bases or frames. By a few numerical examples in the field of audio and image processing we show that the resulting method works quite nicely.

Keywords: Multi-frames, linear operator equations, shrinkage, sparsity, audio coding, image restoration

1 Introduction

Recent studies in the field of signal processing have shown the importance of sparse representations for various tasks, such as signal compression, denoising etc. Typically, such sparse representations are achieved by using a suitable orthonormal basis in the underlying function space. However, recent developments also indicate that redundant systems, such as frames, or dictionaries of 'waveform' systems may yield a gain in this context.

When dealing with dictionaries of 'waveform' systems, there exist several methods, e.g. best orthogonal basis, matching pursuit, basis pursuit etc., see, e.g., [2], that allow a decomposition of a signal into an 'optimal' superposition of dictionary elements, where optimal means having the smallest ℓ_1 norm of coefficients among all such decompositions. At least basis pursuit in highly overcomplete dictionaries leads to very large scale optimization problems (but can be attacked by linear programming, i.e. by interior-point methods).

*The author is with the Konrad–Zuse–Institute Berlin, Germany, teschke@zib.de.

In this paper we develop a new iterative method for finding the ℓ_p–optimal decomposition ($1 \leq p \leq 2$) of a given signal into dictionary building blocks. The skeletal structure of this scheme was originally discovered for solving linear inverse problems with sparsity constraints, see [3]. But instead of preselecting one orthonormal basis or frame only, we typically assume that the signal might be a superposition n different components and thus, we pick a dictionary consisting of a family of n frames. Moreover, we combine this with an inverse problem, namely assuming that we have not observed the signal directly, but only other quantities that are linearly related to the signal.

The advantage of the proposed method is that for achieving convergence of the iteration process, we do not need to require any further assumptions on the preselected family of frames (e.g. such as incoherence). Moreover, since each individual frame is separately penalized we may mix the constraints, i.e. the penalties may vary from ℓ_1–sparsity to ℓ_2–quadratic smoothness constraints. A similar attempt where mixed (sparsity and smoothness) constraints were used was made, e.g., in [4, 5, 7], but these approaches involve one single basis/frame only.

There exists are variety of papers that follow similar ideas as proposed here, e.g. [1, 8, 12]. However, this paper was mainly encouraged by several developments in the field of audio processing/coding, see e.g. [10, 9, 6]. In general, the claim in these papers is to use for different components (such as tonal and transient components) different 'waveform' atoms. Based on these principles, we develop a machinery which is, of course, useful in a wider range of applications. However, we limit our numerical experiments to the field of audio coding and image restoration/compression.

The remaining organization of the paper is as follows: in the following two subsections we collect some material on frames and inverse problems. In Section 2, we introduce our so–called multi–frame concept and cast the resulting inverse problem in its variational form. In Section 3, we explicitly compute our iterative scheme for which we show in Section 4 convergence in norm topology. In the last Section, we present our numerical experiments in audio and image processing.

1.1 Frames

A frame $\{\phi_\lambda\}_{\lambda\in\Lambda}$ in a Hilbert space $\mathcal{H}$ is a set of vectors for which there exists constants $A, B > 0$ such that, for all $v \in \mathcal{H}$,

$$A\|v\|^2_{\mathcal{H}} \leq \sum_{\lambda\in\Lambda} |\langle v, \phi_\lambda\rangle_{\mathcal{H}}|^2 \leq B\|v\|^2_{\mathcal{H}} \ .$$

Frames are typically 'overcomplete', i.e. for a given vector $v \in \mathcal{H}$, one can find many different sequences $g \in \ell_2$ of coefficients so that

$$v = \sum_{\lambda\in\Lambda} g_\lambda \phi_\lambda. \tag{1}$$

A few of them have special properties for which they are preferred, e.g. a sequence with minimal ℓ_2 norm. The problem of finding sequences g can be considered as an inverse problem. To this end, let us consider the operator F (often called the frame operator) that maps a function

$v \in \mathcal{H}$ to the element Fv of ℓ_2 by $Fv = \{\langle v, \phi_\lambda \rangle_{\mathcal{H}}\}_{\lambda \in \Lambda}$. The adjoint F^* maps a sequence $g \in \ell_2$ to the element $F^* g$ of $\mathcal{H}$ via $F^* g = \sum_{\lambda \in \Lambda} g_\lambda \phi_\lambda$, i.e. solving (1) amounts to solving $F^* g = v$. In order to show how to solve the last equation and to highlight the relation to standard frame lore, we observe that, for $v \in \mathcal{H}$, one has

$$F^* F v = \sum_{\lambda \in \Lambda} \langle v, \phi_\lambda \rangle_{\mathcal{H}} \, \phi_\lambda \ ;$$

for $g \in \ell_2$, the sequence $FF^* g$ is given by

$$(FF^* g)_\eta = \sum_{\lambda \in \Lambda} g_\lambda \, \langle \phi_\lambda, \phi_\eta \rangle_{\mathcal{H}} \ .$$

In this context, the sequence g of minimum ℓ_2-norm satisfying (1) is given by $g^\dagger = (F^*)^\dagger v$. Standard frame concepts suggest $g^\dagger = F(F^*F)^{-1} v$, so that $(F^*)^\dagger = F(F^*F)^{-1}$ in this case. The latter equation holds true since this inverse problem is well-posed: even though $\mathcal{N}(F^*) \neq \{0\}$, the operator F^*F has its spectrum completely within the interval $[A, B]$. The spectrum of the operator FF^* has a gap between the eigenvalue 0 and the remainder of the spectrum, which is contained in $[A, B]$ (the operator FF^* becomes only invertible if the frame satisfies special properties, e.g. if $\{\phi_\lambda\}_{\lambda \in \Lambda}$ forms a Riesz basis for $\mathcal{H}$). The relation between F^*F and FF^* is now as follows: since F^*F is invertible, every $v \in \mathcal{H}$ has expansions

$$v = \sum_{\lambda \in \Lambda} \left\langle v, (F^*F)^{-1} \phi_\lambda \right\rangle_{\mathcal{H}} \phi_\lambda = \sum_{\lambda \in \Lambda} \langle v, \phi_\lambda \rangle_{\mathcal{H}} \, (F^*F)^{-1} \phi_\lambda .$$

These expansions are only useful if it is possible to calculate $(F^*F)^{-1}\phi_\lambda$ (the so-called 'canonical' dual frame). Often it is convenient (and also more efficient) to employ an iterative reconstruction method, which is usually called the frame algorithm: given a relaxation parameter $0 < \gamma < 2/B$, set $\delta = \max\{|1 - \gamma A|, |1 - \gamma B|\} < 1$. Let $v_0 = 0$ and define the iteration

$$v_{m+1} = v_m + \gamma F^* F (v - v_m) \ ,$$

for which $\|v - v_m\|_{\mathcal{H}} \leq \delta^m \|v\|_{\mathcal{H}}$. This algorithm acts in function space. In order to express each v_m, we turn to the associated sequences g_m and assume $v_m = \sum_{\lambda \in \Lambda} (g_m)_\lambda \phi_\lambda = F^* g_m$. This implies

$$v_{m+1} \quad = \quad F^*(g_m + \gamma(Fv - FF^* g_m)) = F^*(g_m + \gamma F(v - F^* g_m))$$

inducing with $v_{m+1} = F^* g_{m+1}$ an iteration in sequence space which is now based on the Gram matrix F^*F,

$$g_{m+1} = g_m + \gamma F(v - F^* g_m) \qquad (2)$$

Scheme (2) is nothing than a Landweber iteration, which is a linear regularization scheme (assumed $0 < \gamma < 2/\|F^*\|^2$) and minimizes the discrepancy $\|v - F^* g\|^2_{\mathcal{H}}$. Consequently, the iterates of (2) approximate the minimum ℓ_2-norm sequence $\{\langle v, (F^*F)^{-1}\phi_\lambda \rangle_{\mathcal{H}}\}_{\lambda \in \Lambda}$.

It is now often of interest to find sequences that are sparser than the minimum ℓ_2-norm solution. For instance, one may know a priori that v is a noisy version of a linear combination of ϕ_λ with a coefficient sequence with small ℓ_p-norm ($p = 1$ or, more general, $1 \leq p \leq 2$). In this situation, it makes sense to compute some g that minimizes

$$\|v - F^* g\|_{\mathcal{H}}^2 + \alpha \|g\|_{\ell_p}^p \ .$$

For $p = 2$, a possible way to approach the minimizer for the last problem is given by damped Landweber iterations (for $\|F\|^2 < B'$)

$$g_{m+1} = \frac{1}{B' + \alpha} \left(B' g_m + F(v - F^* g_m) \right)$$

which converges because it is a contractive mapping and for $p = 1$, by Landweber iterations with shrinkage operation in each step

$$g_{m+1} = S_{\frac{\alpha}{2B'}} \left(g_m + (B')^{-1} F(v - F^* g_m) \right)$$

which converges too, see e.g. [3].

1.2 Linear inverse problems

As we have seen in the last section, the approximation of the dual frame $(F^*F)^{-1}\phi_\lambda$ or of the sequence $\{\langle v, (F^*F)^{-1}\phi_\lambda \rangle_{\mathcal{H}}\}_{\lambda \in \Lambda}$ can be directly related to solving a linear inverse problem in its variational form. However, in many applications, the features or signals of interest cannot be observed directly, but have to be inferred from other, observable quantities. Very often, there is a linear relationship between the feature modeled by a function v, and the derived quantities modeled by another function z, i.e. we can write the problem of inferring v from z as

$$Av = z \ .$$

This equation and the task of solving it makes only sense when everything is placed in an adequate setting. The observations (data), which we shall model by yet another function, f, are typically not exactly equal to $z = Av$, but a distortion of z. Often the distortion is modeled by an additive noise error term e (from which one typically assumes that it can be measured by its L_2-norm),

$$f = z + e = Av + e \ .$$

Therefore it is customary to take as the image space L_2; even if the true images z lie in a much smaller space. Thus, we shall always assume that A is a bounded operator from $\mathcal{H}$ to $\mathcal{H}'$ (think of $\mathcal{H}' = L_2$).

To find an estimate for v from observed f, one can minimize the discrepancy

$$\|f - Av\|_{\mathcal{H}'}^2 \ .$$

The minimizer of the discrepancy is called the pseudo-solution of the inverse problem. If A has a trivial null-space, the unique minimizer is given by $(A^*A)^{-1}A^*f$; if the null-space is non-trivial, one picks the unique element $z^\dagger$ of minimum norm. This function is called the generalized solution. Even when A^*A is not invertible, $z^\dagger = A^\dagger f$ is well-defined for all f with $A^*f \in \mathcal{R}(A^*A)$. But the generalized inverse may be unbounded, then the problem is ill-posed. In such cases, it has to be replaced by bounded approximants, so that numerically stable solutions can be used as meaningful approximant. This is the goal of regularization. A regularized version in its variational form, the so-called Tikhonov functional, is given by

$$\|f - Av\|^2_{\mathcal{H}'} + \alpha\|v\|^2_{\mathcal{H}} \ .$$

Let us now put the frame concept and regularization theory together, i.e. given an observation f, the task is to search for a sequence g of dual frame coefficients for our feature v. The variational problem of this inverse problems then takes the form

$$\Phi(g) = \|f - AF^*g\|^2_{\mathcal{H}'} + \alpha\|g\|^p_{\ell_p} \ , \tag{3}$$

where we allow $1 \leq p \leq 2$; the cases $p < 2$ promote sparse representations of v (typically one would pick $p = 1$). This kind of variational problem can be solved by applying the results and methods presented in [3].

2 Multi–frame setting and the variational problem

Instead of using one single frame only, we aim now to represent the function we are searching for by means of several frames. This makes sense since for certain classes of signals it often seems that one single frame is not always best suited (in the sense of locally best sparse approximation). Of course, a finite union of frames is again a frame. Thus, one might consider just one bigger frame. However, when putting mixed or different penalties on the different frames then a setup where each frame is treated individually is better suited. To this end we proceed this way and start with the development of a proper variational problem for this purpose.

Let $\mathcal{H}$ and $\mathcal{H}'$ as before. The suggested multi–frame setting requires the presence of a finite family of frames $\{\phi^i_\lambda\}_{\lambda\in\Lambda, i\in\mathcal{I}}$ where each individual collection $\{\phi^i_\lambda\}_{\lambda\in\Lambda}$ $(i = 1, 2, \ldots, n)$ is a frame for $\mathcal{H}$. For each frame we have the frame operator

$$F_i : \mathcal{H} \to \ell_2, \ \text{via} \ \ v \mapsto v^i := \{\langle v, \phi^i_\lambda \rangle\}_{\lambda\in\Lambda_i} \ . \tag{4}$$

All the frame operators may now be related by considering the following composition operator

$$K : (\ell_2)^n = \ell_2 \times \ldots \times \ell_2 \to \mathcal{H}, \ \text{via} \ \ (v^1, \ldots, v^n) \mapsto \sum_{i=1}^n F^*_i v^i \tag{5}$$

or more general, if we additionally involve a linear inverse problem by some bounded linear operator $A : \mathcal{H} \to \mathcal{H}'$

$$K_A : (\ell_2)^n = \ell_2 \times \ldots \times \ell_2 \to \mathcal{H}', \ \text{via} \ \ (v^1, \ldots, v^n) \mapsto \sum_{i=1}^n AF^*_i v^i \ . \tag{6}$$

For later use have to compute the adjoint and a bound for the operator norm of K_A.

Lemma 1 *The adjoint operator is given by*

$$K^*_A : \mathcal{H}' \mapsto (\ell_2)^n, \quad \text{via} \ \ g \mapsto K^*_A g = (F_1 A^* g, \ldots, F_n A^* g) \ . \tag{7}$$

Moreover, if we assume that $\|A\| < \tilde{C}$ *and that* B_i *denotes the upper frame bound for* F_i*, then*

$$\|K_A\| < \tilde{C}\sqrt{B_1 + \ldots + B_n}. \tag{8}$$

Proof. First, we note that the Hilbert space $(\ell_2)^n$ is endowed with the scalar

$$\langle g, h \rangle_{(\ell_2)^n} = \left\langle g^1, h^1 \right\rangle_{\ell_2} + \ldots + \langle g^n, h^n \rangle_{\ell_2}$$

and thus the associated norm is given by

$$\|g\|^2_{(\ell_2)^n} = \|g^1\|^2_{\ell_2} + \ldots + \|g^2\|^2_{\ell_2} \ .$$

For $f = (f^1, \ldots, f^n) \in (\ell_2)^n$ and $h \in \mathcal{H}'$ the adjoint operator (7) can now be easily derived:

$$\begin{aligned} \langle K_A f, h \rangle_{\mathcal{H}'} &= \sum_{i=1}^n \left\langle A F_i^* f^i, h \right\rangle_{\mathcal{H}'} = \sum_{i=1}^n \left\langle f^i, F_i A^* h \right\rangle_{\ell_2} \\ &= \langle f, (F_1 A^* h, \ldots, F_n A^* h) \rangle_{(\ell_2)^n} = \langle f, K_A^* h) \rangle_{(\ell_2)^n} \end{aligned}$$

and bound (8) follows then directly by

$$\begin{aligned} \|K_A^* f\|^2_{(\ell_2)^n} &= \|F_1 A^* f\|^2_{\ell_2} + \ldots + \|F_n A^* f\|^2_{\ell_2} \\ &\leq \|F_1\|^2 \|A\|^2 \|f\|^2_{\mathcal{H}} + \ldots + \|F_n\|^2 \|A\|^2 \|f\|^2_{\mathcal{H}} \ , \end{aligned}$$

i.e.

$$\|K_A\| < \tilde{C} \sqrt{B_1 + \ldots + B_n} \ .$$

□

With this specific operator K_A we may now formulate the following variational problem

$$\Phi(g) := \|f - K_A g\|^2_{\mathcal{H}'} + \alpha \cdot |||g||| \ , \tag{9}$$

where $g = (g^1, \ldots, g^n)$, $|||g||| := (|g^1|_{p_1,w_1}, \ldots, |g^n|_{p_n,w_n})$ with $|\cdot|_{p_i,w_i}$ denoting a weighted ℓ_{p_i} (semi)-norm, and $\alpha = (\alpha_1, \ldots, \alpha_n)$ represent n positive regularization parameters. In principle, we restrict ourselves also to $1 \leq p_i \leq 2$ with not necessarily requiring $p_i = p_j$. Thus Φ is in principle no longer homogeneous and this complicates the choice of α.

A strategy for solving this kind of problem for $n = 1$, $1 \leq p_1 \leq 2$ and $\{\phi_\lambda\}_{\lambda \in \Lambda}$ being a basis (also concepts for frames) is shown in [3], for $n = 2$, concepts are suggested in [4, 5, 7]. For our purposes, we will follow the techniques introduced there. The specialty here is that each component g^i of g is represented by another frame, i.e. we are searching for an approximation of v which is a composition of different frames (or different bases).

In what follows, we propose a strategy how to compute or to approximate the vector of sequences, g. First, we show that Φ is a convex functional. For general frame systems one cannot expect uniqueness (or strict convexity) since in principal we have $ker(F_i^*) \neq \{0\}$, i.e. even when $\mathcal{N}(A) = \{0\}$ we have no chance.

Lemma 2 *The functional* Φ *as defined in* (9) *is convex.*

Proof. Consider $g = (g^1, \ldots, g^n) \in (\ell_2)^n$ and $h = (h^1, \ldots, h^n) \in (\ell_2)^n$ and some $\tau \in (0,1)$. Then,

$$\begin{aligned}\Delta &= \Phi(\tau g + (1-\tau)h) - \tau\Phi(g) - (1-\tau)\Phi(h) \\ &= -\tau(1-\tau)\|K_A g - K_A h\|^2_{\mathcal{H}'} + \alpha \cdot (|||\tau g + (1-\tau)h||| - \tau|||g||| - (1-\tau)|||h|||) \ .\end{aligned}$$

The second term is non-positive since a Banach (semi) norm is convex, and the first term is also non-positive. Consequently, $\Delta \leq 0$ and Φ is convex. □

The next step is to construct a surrogate or so-called replacement functional for Φ from which we expect a simplification of the minimization process. The overall goal is to avoid the appearance of $\|K_A g\|^2_{\mathcal{H}'}$ which typically causes a non-linear coupling of all the frame coefficients we aim to compute. Defining a constant $C := \tilde{C}\sqrt{B_1 + \ldots + B_n}$ the standard Gaussian surrogate for the data discrepancy takes the following form

$$\Gamma^{sur}(g; a) \;=\; \|f - K_A g\|^2_{\mathcal{H}'} + C^2\|g - a\|^2_{(\ell_2)^n} - \|K_A g - K_A a\|^2_{\mathcal{H}'}$$

for some auxiliary element $a \in (\ell_2)^n$.

Lemma 3 *The functional $\Gamma^{sur}(g; a)$ is a proper surrogate for $\|f - K_A g\|^2_{\mathcal{H}'}$.*

Proof. 'Proper' in this context means that the problem remains convex and

$$\Gamma^{sur}(g; a) - \|f - K_A g\|^2_{\mathcal{H}'} \geq 0 \ .$$

To this end, consider

$$C^2\|g - a\|^2_{(\ell_2)^n} - \|K_A g - K_A a\|^2_{\mathcal{H}'} \;=\; \big\langle (C^2 - K_A^* K_A)(g - a), g - a \big\rangle_{(\ell_2)^n} \ . \tag{10}$$

Defining $L := \sqrt{C^2 - K_A^* K_A}$, we observe for $\tau \in (0,1)$ and $g, h \in (\ell_2)^n$

$$\begin{aligned}&\|L(\tau g + (1-\tau)h - a)\|^2_{(\ell_2)^n} - \tau\|L(g - a)\|^2_{(\ell_2)^n} - (1-\tau)\|L(h - a)\|^2_{(\ell_2)^n} \\ &\qquad = -\tau(1-\tau)\|L(g - h)\|^2_{(\ell_2)^n} \ .\end{aligned}$$

Since $C^2 - K_A^* K_A$ and therewith L are strictly positive operators, (10) is strictly convex as a function of g and positive for $g \neq a$. □

Now we are able to define the global surrogate for Φ:

$$\Phi^{sur}(g; a) := \Gamma^{sur}(g, a) + \alpha \cdot |||g||| \ , \tag{11}$$

satisfying

$$\Phi^{sur}(g; g) = \Phi(g) \ , \quad \Phi^{sur}(g; a) \geq \Phi(g) \quad \text{for all } a \in (\ell_2)^n \ . \tag{12}$$

The definition of our surrogate functional (11) suggests the following iteration in order to approach the minimizer $g^\star$ of the initial problem (9): starting from an arbitrarily chosen g_0, we determine the minimizer g_1 of (11) for $a = g_0$; each successive iterate g_m is then the minimizer for g of (11) anchored at the previous iterate $a = g_{m-1}$:

$$g_0 \text{ arbitrary} \ ; \quad g_{m+1} = \arg\min_g \Phi^{sur}(g; g_m) \qquad m = 0, 1, \ldots \tag{13}$$

3 Minimization of the surrogate functional

The general principles of minimizing (11) shown in [3] directly apply here. The difference is that instead with one single frame we deal with n frames and n mixed penalties. We limit the analysis to the case $p_j = 1$ for $j = 1, \ldots, n$. The other cases $1 < p_i \leq 2$ (not necessarily requiring $p_i = p_j$) cause no additional problems: each frame can be treated individually and in the same manner. The concept and the techniques can be retraced with the help of [3]. The goal here is to show the interplay when dealing with a family of frames. Note that all below appearing shrinkage operators will change when changing p_i. Nevertheless, all the lemmatas, propositions and theorems hold true. The case $0 < p_i < 1$ is not under consideration here since then the penalties become non–convex and a complete different theory is required.

First we discuss the minimization of (11) for some generic $a \in (\ell_2)^n$. The surrogate functional has the following form

$$\begin{aligned}
\Phi^{sur}(g; a) &= \|f\|^2_{\mathcal{H}'} - 2\,\langle g, K_A^* f\rangle_{(\ell_2)^n} + \alpha \cdot |||g||| \\
&\qquad C^2\|g - a\|^2_{(\ell_2)^n} + 2\,\langle g, K_A^* K_A a\rangle_{(\ell_2)^n} - \|K_A a\|^2_{\mathcal{H}'} \\
&= C^2\|g\|^2_{(\ell_2)^n} - 2\left\langle g, K_A^* f + C^2 a - K_A^* K_A a\right\rangle_{(\ell_2)^n} + \alpha \cdot |||g||| \\
&\qquad + \|f\|^2_{\mathcal{H}'} + C^2\|a\|^2_{(\ell_2)^n} - \|K_A a\|^2_{\mathcal{H}'} \\
&= \sum_{i=1}^{n} \sum_{\lambda \in \Lambda_i} \left(C^2 (g^i_\lambda)^2 - 2 g^i_\lambda \left[F_i A^* f + C^2 a^i - F_i A^* K_A a\right]_\lambda + \alpha_i |g^i_\lambda| \right) \\
&\qquad + \|f\|^2_{\mathcal{H}'} + C^2\|a\|^2_{(\ell_2)^n} - \|K_A a\|^2_{\mathcal{H}'}\ ,
\end{aligned} \tag{14}$$

where the g^i_λ denote the coefficients of the sequence g^i (and where implicitly assumed that we are dealing with real functions; otherwise one needs to parameterize by modulus and phase). The latter variational equation for the g^i_λ decouple. The summand is differentiable in g^i_λ except at $g^i_\lambda = 0$. To overcome this drawback we introduce set-valued derivatives, i.e. we allow $\mathrm{sgn}(0) \in [-1, 1]$. Then the minimization reduces to solving

$$g^i_\lambda + \frac{\alpha_i}{2C^2}\mathrm{sgn}(g^i_\lambda) \ni C^{-2}\left[F_i A^* f + C^2 a^i - F_i A^* K_A a\right]_\lambda\ . \tag{15}$$

Denoting the soft-shrinkage operator by S_t with shrinkage parameter t, we obtain an explicit expression for the coefficients

$$g^i_\lambda = S_{\frac{\alpha_i}{2C^2}}\left(C^{-2}\left[F_i A^* f + C^2 a^i - F_i A^* K_A a\right]_\lambda\right)\ . \tag{16}$$

Let us now introduce with a slight abuse of notation the soft-shrinkage operation for some $f \in \ell_2$ acting component-wise

$$S_t(f) = \{S_t(f_\lambda)\}_{\lambda \in \Lambda}\ .$$

With this shorthand we may introduce the combined shrinkage operator for some vector of sequences $(f^1,\dots,f^n)\in(\ell_2)^n$ and a multi parameter $t=(t_1,\dots,t_n)$

$$\mathbf{S}_t(f)=\left(S_{t_1}(f^1),\dots,S_{t_n}(f^n)\right)\ .$$

In this setting the minimizer g for (11) can be written in the much simpler form

$$g=\mathbf{S}_{\frac{\alpha}{2C^2}}\left(C^{-2}\left[K_A^*f+C^2a-K_A^*K_Aa\right]\right)\ . \tag{17}$$

We summarize our findings for the particular case $p_i=1$, $w_i=1$ for $i=1,\dots,n$:

Proposition 4 *Suppose the operator A maps a Hilbert space $\mathcal{H}$ to another Hilbert space $\mathcal{H}'$, with $\|A\|<\tilde{C}$, and suppose we are given n frames where the respective frame operators F_i map $\mathcal{H}$ to ℓ_2 with upper frame bounds B_i, and suppose f is an element of $\mathcal{H}'$. Pick $p_i=1$, $w_i=1$ for $i=1,\dots,n$, and $a\in(\ell_2)^n$. If $\Phi^{sur}(g;a)$ is defined as in (11) on $(\ell_2)^n$, then $\Phi^{sur}(g;a)$ has a unique minimizer in $(\ell_2)^n$. This minimizer is given by*

$$g=\mathbf{S}_{\frac{\alpha}{2C^2}}\left(C^{-2}\left[K_A^*f+C^2a-K_A^*K_Aa\right]\right)\ . \tag{18}$$

For all $h\in(\ell_2)^n$, one has

$$\Phi^{sur}(g+h;a)\geq\Phi^{sur}(g;a)+C^2\|h\|^2_{(\ell_2)^n}\ .$$

Proof. We observe that

$$\Phi^{sur}(g+h;a)-\Phi^{sur}(g;a)=C^2\|h\|^2_{(\ell_2)^n}+2\left\langle h,C^2g-C^2a-K_A^*(f-K_Aa)\right\rangle_{(\ell_2)^n}+\alpha\left(|||g+h|||-|||g|||\right).$$

Defining sets $\Lambda_i^0:=\{\lambda\in\Lambda_i|\ g_\lambda^i=0\}$, and $\Lambda_i^1:=\Lambda_i\setminus\Lambda_i^0$ and substituting (15) for g_λ^i, we recast the latter equation

$$\begin{aligned}\Phi^{sur}(g+h;a)-\Phi^{sur}(g;a)\ &=\ C^2\|h\|^2_{(\ell_2)^n}+\sum_{i=1}^n\sum_{\lambda\in\Lambda_i^0}\{\alpha_i|h_\lambda^i|-2h_\lambda^i\left[F_iA^*f+C^2a^i-F_iA^*K_Aa\right]_\lambda\}\\ &\quad+\sum_{i=1}^n\sum_{\lambda\in\Lambda_i^1}\alpha_i\{|g_\lambda^i+h_\lambda^i|-|g_\lambda^i|-h_\lambda^i\operatorname{sgn}(g_\lambda^i)\}.\end{aligned}$$

For $\lambda\in\Lambda_i^0$ we have $[F_iA^*f+C^2a^i-F_iA^*K_Aa]_\lambda\leq\alpha_i/2$, so that

$$\alpha_i|h_\lambda^i|-2h_\lambda^i[F_iA^*f+C^2a^i-F_iA^*K_Aa]_\lambda\geq 0.$$

For $\lambda\in\Lambda_i^1$, we consider two cases: if $g_\lambda^i>0$, then

$$|g_\lambda^i+h_\lambda^i|-|g_\lambda^i|-h_\lambda^i\operatorname{sgn}(g_\lambda^i)=|g_\lambda^i+h_\lambda^i|-(g_\lambda^i+h_\lambda^i)\geq 0;$$

if $g_\lambda^i<0$, then

$$|g_\lambda^i+h_\lambda^i|-|g_\lambda^i|-h_\lambda^i\operatorname{sgn}(g_\lambda^i)=|g_\lambda^i+h_\lambda^i|+(g_\lambda^i+h_\lambda^i)\geq 0,$$

which proves the assertion. ☐

Proposition 4 directly carries over to iteration (13):

Corollary 5 *Make the same assumptions as in Proposition 4. Pick $g_0\in(\ell_2)^n$, and define the functions g_m by the algorithm (13). Then*

$$g_{m+1}=\mathbf{S}_{\frac{\alpha}{2C^2}}\left(C^{-2}\left[K_A^*f+C^2g_m-K_A^*K_Ag_m\right]\right)\ . \tag{19}$$

4 Convergence Analysis

In this section we consider the convergence of the proposed iteration (13):

Theorem 6 *Suppose the operator A maps a Hilbert space $\mathcal{H}$ to another Hilbert space $\mathcal{H}'$, with $\|A\| < \tilde{C}$, and suppose we are given n frames where the respective frame operators F_i map $\mathcal{H}$ to ℓ_2 with upper frame bounds B_i, and suppose f is an element of $\mathcal{H}'$. Then the sequence of iterates*

$$g_{m+1} = \mathbf{S}_{\frac{\alpha}{2C^2}} \left(C^{-2} \left[K_A^* f + C^2 g_m - K_A^* K_A g_m \right] \right) \quad , \quad m = 1, 2, \ldots \quad ,$$

with g_0 arbitrarily chosen in $(\ell_2)^n$, converges in norm to a minimizer of the functional

$$\Phi(g) = \|f - K_A g\|_{\mathcal{H}'}^2 + \alpha \cdot |||g||| \ .$$

In order to prove this result, we follow the lines in [3]. The techniques and strategies developed in [3] can be applied with natural changes caused by our multi-dimensional setting. For readers convenience, we will show in detail how to achieve strong convergence: first, we prove weak convergence, and we show that the weak limit is a minimizer for Φ; and next, we show that the convergence holds also in norm.

With the following shorthand

$$\mathbf{T} g = \mathbf{S}_{\frac{\alpha}{2C^2}} \left(C^{-2} \left[K_A^* f + C^2 g - K_A^* K_A g \right] \right) \ ,$$

i.e. $g_m = \mathbf{T}^m g_0$, we may formulate the weak convergence result as follows:

Proposition 7 *The sequence $\mathbf{T}^m g_0$, $n = 1, 2, \ldots$ convergences weakly, and its limit is a fixed point for $\mathbf{T}$.*

This result can be achieved by applying Opial's Theorem, see [11]:

Theorem 8 (Opial) *Let the mapping $\mathbf{A}$ from $\mathcal{H}$ to $\mathcal{H}$ satisfy the following conditions:*
i) $\mathbf{A}$ is non-expansive, i.e. for all $v, w \in \mathcal{H}$, $\|\mathbf{A}v - \mathbf{A}w\| \leq \|v - w\|$,
ii) $\mathbf{A}$ is asymptotically regular: for all $v \in \mathcal{H}$, $\|\mathbf{A}^{n+1}v - \mathbf{A}^n v\| \stackrel{n\to\infty}{\longrightarrow} 0$,
iii) the set $\mathcal{F}$ of fixed points of $\mathbf{A}$ in $\mathcal{H}$ is not empty.
Then, for all $v \in \mathcal{H}$, the sequence $\{\mathbf{A}^n v\}_{n \in \mathbb{N}}$ converges weakly to a fixed point in $\mathcal{F}$.

In order to prove Proposition 7, we apply Theorem 8 to $\mathbf{T}$. To this end, we have to verify conditions $i), ii)$ and $iii)$. We do this by the following series of lemmas.

Lemma 9 *The operator $\mathbf{S}_t$ is non-expansive, i.e. for all $v, w \in (\ell_2)^n$,*

$$\|\mathbf{S}_t(v) - \mathbf{S}_t(w)\|_{(\ell_2)^n} \leq \|v - w\|_{(\ell_2)^n} \ .$$

Proof. The results is obtained by applying the fact that each single shrinkage operator is non-expansive, see, e.g., [3],

$$\begin{aligned}\|\mathbf{S}_t(v) - \mathbf{S}_t(w)\|^2_{(\ell_2)^n} &= \sum_{i=1}^n \|S_t(v^i) - S_t(w^i)\|^2_{\ell_2} = \sum_{i=1}^n \sum_{\lambda\in\Lambda_i} |S_t(v^i_\lambda) - S_t(w^i_\lambda)|^2 \\ &\leq \sum_{i=1}^n \sum_{\lambda\in\Lambda_i} |v^i_\lambda - w^i_\lambda|^2 = \|v-w\|_{(\ell_2)^n} \ .\end{aligned}$$

□

Lemma 10 *The mapping* $\mathbf{T}$ *is non-expansive, i.e. for all* $v, w \in (\ell_2)^n$

$$\|\mathbf{T}v - \mathbf{T}w\|_{(\ell_2)^n} \leq \|v-w\|_{(\ell_2)^n}.$$

Proof. By Lemma 9 we have

$$\begin{aligned}\|\mathbf{T}v - \mathbf{T}w\|_{(\ell_2)^n} &\leq \|(I - C^{-2}K_A^*K_A)(v-w)\|_{(\ell_2)^n} \\ &\leq \|(v-w)\|_{(\ell_2)^n}\end{aligned}$$

since have chosen C such that $\|K_A\| < C$. □

Hence, $\mathbf{T}$ satisfies condition $i)$ in Theorem 8. Next, we verify condition $ii)$:

Lemma 11 *The sequences* $\{\Phi(g_m)\}_{m\in\mathbb{N}}$ *and* $\{\Phi^{sur}(g_{m+1}; g_m)\}_{m\in\mathbb{N}}$ *are non-increasing.*

Proof. By the definition of L and (13) we have

$$\Phi(g_{m+1}) + \|L(g_{m+1} - g_m)\|^2_{(\ell_2)^n} = \Phi^{sur}(g_{m+1}; g_m) \leq \Phi^{sur}(g_m; g_m) = \Phi(g_m) \tag{20}$$

and

$$\Phi^{sur}(g_{m+2}; g_{m+1}) \leq \Phi(g_{m+1}) \leq \Phi(g_{m+1}) + \|L(g_{m+1} - g_m)\|^2_{(\ell_2)^n} = \Phi^{sur}(g_{m+1}; g_m).$$

□

Lemma 12 *The series* $\sum_{m=0}^\infty \|g_{m+1} - g_m\|^2_{(\ell_2)^n}$ *is convergent.*

Proof. Since L is a strictly positive operator, we have

$$\sum_{m=0}^N \|g_{m+1} - g_m\|^2_{(\ell_2)^n} \leq \frac{1}{M} \sum_{m=0}^N \|L(g_{m+1} - g_m)\|^2_{(\ell_2)^n}$$

where M is a strictly lower bound for L^*L. By Lemma 11 and (20),

$$\sum_{m=0}^N \|L(g_{m+1} - g_m)\|^2_{(\ell_2)^n} \leq \sum_{m=0}^N (\Phi(g_m) - \Phi(g_{m+1})) \leq \Phi(g_0),$$

regardless of the choice of $N \in \mathbb{N}$ and the infinite series converges. □

Consequently, we have that

Lemma 13 *The mapping* $\mathbf{T}$ *is asymptotically regular, i.e.*

$$\|\mathbf{T}^{m+1}g_0 - \mathbf{T}^m g_0\|_{(\ell_2)^n} \to 0 \quad \text{for} \quad n \to \infty \ .$$

We finalize the proof of Theorem 7 with verifying condition *iii*):

Lemma 14 *The* $\|g_m\|_{(\ell_2)^n}$ *are bounded uniformly in* n.

Proof. By Lemma 11 and since $\alpha_i > 0$ for $i = 1, \ldots, n$, we have

$$\alpha \cdot |||g_m||| \leq \Phi(g_m) \leq \Phi(g_0) \text{ , thus } \left| |||g_m||| \right|_{\ell_1} \leq \frac{1}{\min_i \alpha_i} \Phi(g_0).$$

Hence the g_m are uniformly bounded. Moreover, since

$$\|g_m\|^2_{(\ell_2)^n} \ = \ \left| |||g_m||| \right|^2_{\ell_1} \ ,$$

we also have a uniform bound on the $\|g_m\|^2_{(\ell_2)^n}$. □

As pointed out in [3], we still have

Lemma 15 *Suppose the mapping* $\mathbf{A}$ *from* $\mathcal{H}$ *to* $\mathcal{H}$ *satisfies the conditions i) and ii) in Theorem 8. Then, if a subsequence of* $\{\mathbf{A}^n v\}_{n\in\mathbb{N}}$ *converges weakly in* $\mathcal{H}$, *then its limit is a fixed point of* $\mathbf{A}$.

Lemma 16 *The set of fixed points of* $\mathbf{T}$ *is not empty.*

Proof. By Lemma 14, the $\mathbf{T}^m g_0$ are uniformly bounded in m. By the Banach-Alaoglu Theorem, the sequence has a weak accumulation point. By Lemma 15, this weak accumulation point is a fixed point for $\mathbf{T}$ and consequently, the set of fixed points of $\mathbf{T}$ is not empty. □

Finally, by Lemmas 10, 13, and 16, we have shown Proposition 7. Moreover, we can show that this fixed point is also a minimizer for Φ:

Proposition 17 *A fixed point for* $\mathbf{T}$ *is a minimizer for the functional* Φ.

Proof. If $g_\star = \mathbf{T}g_\star$, then by Proposition 4, we have that $g_\star$ is a minimizer for $\Phi^{sur}(g; g_\star)$. Moreover, for all $h \in (\ell_2)^n$,

$$\Phi^{sur}(g_\star + h; g_\star) \geq \Phi^{sur}(g_\star; g_\star) + C^2\|h\|^2_{(\ell_2)^n} \ .$$

With $\Phi^{sur}(g_\star; g_\star) = \Phi(g_\star)$ and $\Phi^{sur}(g_\star + h; g_\star) = \Phi(g_\star + h) + C^2\|h\|^2_{(\ell_2)^n} - \|K_A h\|^2_{\mathcal{H}'}$, we deduce that, for all $h \in (\ell_2)^n$,

$$\Phi(g_\star + h) \geq \Phi(g_\star) + \|K_A h\|^2_{\mathcal{H}'} \ ,$$

which proves that $g_\star$ is also a minimizer for Φ. □

Next, we shall prove that the convergence of $\{g_m\}_{m\in\mathbb{N}}$ holds also in the Hilbert space norm $\|\cdot\|_{(\ell_2)^n}$. Let us introduce the following shorthands

$$g_w = \text{weak} - \lim_{m\to\infty} g_m \ , \quad u_m = g_m - g_w \ , \quad h = g_w + C^{-2}K_A^*(f - K_A g_w) \ . \tag{21}$$

Lemma 18 *$\|K_A u_m\|_{(\ell_2)^n} \to 0$ for $m \to \infty$.*

Proof. First, observe that with

$$\begin{aligned} u_{m+1} &= g_{m+1} - g_w = \mathbf{S}_t\left(g_m + C^{-2}K_A^*(f - K_A g_m)\right) - \mathbf{S}_t(h) \\ &= \mathbf{S}_t\left(h + (I - C^{-2}K_A^*K_A)u_m\right) - \mathbf{S}_t(h) \end{aligned}$$

one has

$$u_{m+1} - u_m = \mathbf{S}_t\left(h + (I - C^{-2}K_A^*K_A)u_m\right) - \mathbf{S}_t(h) - u_m$$

and since $\|u_{m+1} - u_m\|_{(\ell_2)^n} = \|g_{m+1} - g_m\|_{(\ell_2)^n} \to 0$ for $n \to \infty$ (by Lemma 13), we have

$$\|\mathbf{S}_t\left(h + (I - C^{-2}K_A^*K_A)u_m\right) - \mathbf{S}_t(h) - u_m\|_{(\ell_2)^n} \to 0 \quad \text{for} \quad n \to \infty\ . \tag{22}$$

By triangle inequality,

$$\left|\|u_m\|_{(\ell_2)^n} - \|\mathbf{S}_t\left(h + (I - C^{-2}K_A^*K_A)u_m\right) - \mathbf{S}_t(h)\|_{(\ell_2)^n}\right| \to 0 \quad \text{for} \quad n \to \infty\ . \tag{23}$$

By Lemma 9, we have

$$\|\mathbf{S}_t\left(h + (I - C^{-2}K_A^*K_A)u_m\right) - \mathbf{S}_t(h)\|_{(\ell_2)^n} \le \|(I - C^{-2}K_A^*K_A)u_m\|_{(\ell_2)^n} \le \|u_m\|_{(\ell_2)^n}$$

and thus the modulus in (23) can be dropped, which implies

$$\|u_m\|_{(\ell_2)^n} - \|(I - C^{-2}K_A^*K_A)u_m\|_{(\ell_2)^n} \to 0 \quad \text{for} \quad n \to \infty\ . \tag{24}$$

Since $\|u_m\|_{(\ell_2)^n} + \|(I - C^{-2}K_A^*K_A)u_m\|_{(\ell_2)^n} \le 2\|g_m - g_w\|_{(\ell_2)^n} \le 2(\|g_w\|_{(\ell_2)^n} + \sup_m \|g_m\|_{(\ell_2)^n}) = \tau$ where τ is finite by Lemma 14, we obtain by (24)

$$\begin{aligned} 0 &\le \|u_m\|^2_{(\ell_2)^n} - \|(I - C^{-2}K_A^*K_A)u_m\|^2_{(\ell_2)^n} \\ &\le \tau\left(\|u_m\|_{(\ell_2)^n} - \|(I - C^{-2}K_A^*K_A)u_m\|_{(\ell_2)^n}\right) \to 0 \quad \text{for} \quad n \to \infty\ . \end{aligned}$$

The inequality

$$\begin{aligned} \|u_m\|^2_{(\ell_2)^n} - \|(I - C^{-2}K_A^*K_A)u_m\|^2_{(\ell_2)^n} &= 2C^{-2}\|K_A u_m\|^2_{(\ell_2)^n} - \|C^{-2}K_A^*K_A u_m\|^2_{(\ell_2)^n} \\ &\ge C^{-2}\|K_A u_m\|^2_{(\ell_2)^n} \end{aligned}$$

finally proves the assertion. □

Lemma 19 $\|\mathbf{S}_t(h + u_m) - \mathbf{S}_t(h) - u_m\|_{(\ell_2)^n} \to 0$ for $n \to \infty$.

Proof. By Lemma 9,

$$\begin{aligned} \|\mathbf{S}_t(h + u_m) - \mathbf{S}_t(h) - u_m\|_{(\ell_2)^n} &\le \|\mathbf{S}_t(h + (I - C^{-2}K_A^*K_A)u_m) - \mathbf{S}_t(h) - u_m\|_{(\ell_2)^n} \\ &\quad + \|\mathbf{S}_t(h + u_m) - \mathbf{S}_t(h + (I - C^{-2}K_A^*K_A)u_m)\|_{(\ell_2)^n} \\ &\le \|\mathbf{S}_t(h + (I - C^{-2}K_A^*K_A)u_m) - \mathbf{S}_t(h) - u_m\|_{(\ell_2)^n} \\ &\quad + \|C^{-2}K_A^*K_A u_m\|_{(\ell_2)^n}\ . \end{aligned}$$

The assertion follows because of Lemma 18 and (22). □

The next lemma establishes norm convergence.

Lemma 20 *If for some* $h \in (\ell_2)^n$ *and a sequence* $\{w_m\}_{m\in\mathbb{N}}$ *with* weak$-\lim_{m\to\infty} w_m = 0$ *one has* $\lim_{m\to\infty} \|\mathbf{S}_t(h+w_m) - \mathbf{S}_t(h) - w_m\|_{(\ell_2)^n} = 0$, *then* $\|w_m\|_{(\ell_2)^n} \to 0$ *for* $m \to \infty$.

Proof. First, note again that

$$\|w_m\|^2_{(\ell_2)^n} = \sum_{i=1}^n \sum_{\lambda\in\Lambda_i} |(w_m^i)_\lambda|^2 \ .$$

For each index i we define finite (possibly empty) sets $\Lambda_i^0 \subset \Lambda_i$, so that $\sum_{\lambda\in\Lambda_i\setminus\Lambda_i^0} |h_\lambda^i|^2 \le \left(\frac{\alpha_i}{4C^2}\right)^2$. Since each Λ_i^0 is finite, we have by the weak convergence of the w_m that

$$\sum_{\lambda\in\Lambda_i^0} |(w_m^i)_\lambda|^2 \to 0 \quad \text{for} \quad m\to\infty$$

holds for every i, thus

$$\sum_{i=1}^n \sum_{\lambda\in\Lambda_i^0} |(w_m^i)_\lambda|^2 \to 0 \quad \text{for} \quad m\to\infty \ .$$

Let us now focus on the remaining sums $\sum_{\lambda\in\Lambda_i\setminus\Lambda_i^0} |(w_m^i)_\lambda|^2$: for each i and each m we split $\Lambda_i^1 = \Lambda_i \setminus \Lambda_i^0$ into two subsets:

$$\Lambda_i^{1,m} = \{\lambda\in\Lambda_i^1 \ : \ |(w_m^i)_\lambda + h_\lambda^i| < \alpha_i/2C^2\} \ \text{ and } \ \tilde{\Lambda}_i^{1,m} = \Lambda_i^1 \setminus \Lambda_i^{1,m}.$$

If $\lambda \in \Lambda_i^{1,m}$, then $S_{\frac{\alpha_i}{2C^2}}((w_m^i)_\lambda + h_\lambda^i) = S_{\frac{\alpha_i}{2C^2}}(h_\lambda^i) = 0$ such that

$$|(w_m^i)_\lambda - S_{\frac{\alpha_i}{2C^2}}((w_m^i)_\lambda + h_\lambda^i) + S_{\frac{\alpha_i}{2C^2}}(h_\lambda^i)| = |(w_m^i)_\lambda|$$

and thus

$$\begin{aligned}
\sum_{i=1}^n \sum_{\lambda\in\Lambda_i^{1,m}} |(w_m^i)_\lambda|^2 &= \sum_{i=1}^n \sum_{\lambda\in\Lambda_i^{1,m}} |(w_m^i)_\lambda - S_{\frac{\alpha_i}{2C^2}}((w_m^i)_\lambda + h_\lambda^i) + S_{\frac{\alpha_i}{2C^2}}(h_\lambda^i)|^2 \\
&\le \sum_{i=1}^n \sum_{\lambda\in\Lambda_i} |(w_m^i)_\lambda - S_{\frac{\alpha_i}{2C^2}}((w_m^i)_\lambda + h_\lambda^i) + S_{\frac{\alpha_i}{2C^2}}(h_\lambda^i)|^2 \\
&= \|\mathbf{S}_t(h+u_m) - \mathbf{S}_t(h) - u_m\|^2_{(\ell_2)^n} \to 0 \quad \text{for} \quad m\to\infty \ . \qquad (25)
\end{aligned}$$

It remains to consider the case $\lambda \in \tilde{\Lambda}_i^{1,m}$. In this situation, $|(w_m^i)_\lambda + h_\lambda^i| \ge \alpha_i/2C^2$, and hence we have for all i and m

$$|(w_m^i)_\lambda| \ge |(w_m^i)_\lambda + h_\lambda^i| - |h_\lambda^i| \ge \alpha_i/2C^2 - \alpha_i/4C^2 = \alpha_i/4C^2 \ge |h_\lambda^i| \ ,$$

implying that $(w_m^i)_\lambda + h_\lambda^i$ and $(w_m^i)_\lambda$ have the same sign and it follows that for $\lambda \in \tilde{\Lambda}_i^{1,m}$

$$\begin{aligned}
|(w_m^i)_\lambda - S_{\frac{\alpha_i}{2C^2}}((w_m^i)_\lambda + h_\lambda^i) + S_{\frac{\alpha_i}{2C^2}}(h_\lambda^i)| &= |(w_m^i)_\lambda - S_{\frac{\alpha_i}{2C^2}}((w_m^i)_\lambda + h_\lambda^i)| \\
&= |(w_m^i)_\lambda - ((w_m^i)_\lambda + h_\lambda^i) + \alpha_i/2C^2 \mathrm{sgn}((w_m^i)_\lambda)| \\
&\ge \alpha_i/2C^2 - |h_\lambda^i| \ge \alpha_i/4C^2 \ .
\end{aligned}$$

Consequently, for each individual i and m we deduce,

$$\sum_{\lambda \in \tilde{\Lambda}_i^{1,m}} |(w_m^i)_\lambda - S_{\frac{\alpha_i}{2C^2}}((w_m^i)_\lambda + h_\lambda^i) + S_{\frac{\alpha_i}{2C^2}}(h_\lambda^i)|^2 \geq (\alpha_i/4C^2)^2 |\tilde{\Lambda}_i^{1,m}| \ ,$$

but since $\|\mathbf{S}_t(h+u_m) - \mathbf{S}_t(h) - u_m\|^2_{(\ell_2)^n} \to 0$ for $m \to \infty$, necessarily implying

$$\|S_{\frac{\alpha_i}{2C^2}}(h^i + u_m^i) - S_{\frac{\alpha_i}{2C^2}}(h^i) - u_m^i\|^2_{\ell_2} \to 0 \quad \text{for} \quad m \to \infty \ ,$$

there exists an index m^* uniform in i so that for $m > m^*$,

$$\sum_{\lambda \in \tilde{\Lambda}_i^{1,m}} |(w_m^i)_\lambda - S_{\frac{\alpha_i}{2C^2}}((w_m^i)_\lambda + h_\lambda^i) + S_{\frac{\alpha_i}{2C^2}}(h_\lambda^i)|^2 < (\alpha_i/4C^2)^2 \ ,$$

which implies that for all i the sets $\tilde{\Lambda}_i^{1,m}$ are empty when $m > m^*$. Consequently,

$$\sum_{i=1}^n \sum_{\lambda \in \tilde{\Lambda}_i^{1,m}} |(w_m^i)_\lambda|^2 = 0 \quad \text{for} \quad m > m^* \ ,$$

which completes the proof since we have

$$\|w_m\|^2_{(\ell_2)^n} = \sum_{i=1}^n \left(\sum_{\lambda \in \Lambda_i^0} |(w_m^i)_\lambda|^2 + \sum_{\lambda \in \Lambda_i^{1,m}} |(w_m^i)_\lambda|^2 + \sum_{\lambda \in \tilde{\Lambda}_i^{1,m}} |(w_m^i)_\lambda|^2 \right) \to 0 \quad \text{for} \quad m \to \infty \ .$$

□

Setting $w_m = u_m$ and h and u_m as in (21), we have shown that

$$\|g_m - g_w\|_{(\ell_2)^n} \to 0 \quad \text{for} \quad m \to \infty \ .$$

Moreover, setting $v_m = \sum_{i=1}^n \sum_{\lambda \in \Lambda_i} (g_m^i)_\lambda \phi_\lambda^i$ and $v_\star = \sum_{i=1}^n \sum_{\lambda \in \Lambda_i} (g_w^i)_\lambda \phi_\lambda^i$, the estimate

$$\|v_m - v_\star\|_{\mathcal{H}} \leq \left(n \max_i \{B_i\}\right)^{1/2} \|g_m - g_w\|_{(\ell_2)^n}$$

ensures that the convergence holds also in the $\mathcal{H}$-norm topology.

Remark: In the proposed scheme, the damping/threshold vector α must be chosen by the 'user'. In terms of compression, this vector controls the sparsity to be achieved. In the context of inverse problems, this parameter controls the regularization/stabilization of the inverse problem. Moreover, if the 'error' $e = f - AK_I g_o$ tends to zero, we wish our estimate for the solution of the inverse problem tend to g_o, since the minimizer of $\Phi(g)$ differs from g_o for $\alpha \neq 0$. Thus one typically aims to identify a functional relation between α and the noise floor ϵ, i.e. $\alpha = \alpha(\epsilon)$ with $\alpha(\epsilon) \to 0$ and $\|g_{w,\alpha(\epsilon)} - g_o\| \to 0$ as $\epsilon \to 0$. A regularization theory for the univariate case ($n = 1$) achieving this is provided in [3]. But for $n > 1$ and a non–homogeneous mixing of penalties no regularization theory is established so far.

5 Numerical Experiments

This section is devoted to show the usefulness of the proposed multi–frame scheme. We present a little number of numerical experiments from different perspectives: convergence, sparsity, approximation quality, and applicability to synthetic data, real audio data, and images.

The overall configuration of our algorithm is as follows: for sake of simplicity, we pick as our underlying frames a wavelet basis (Haar system) and a (non–local) Fourier basis only. Hence, $B_1 = B_2 = 1$. In the examples, we restrict ourselves to $A = I$. Consequently, the constant C in our Gaussian surrogate is not allowed to be equal or smaller than $\sqrt{2}$. We aim to achieve sparsity in both representations, i.e. we set $p_1 = p_2 = 1$; moreover, we do not involve additional penalty weight sequences, i.e. $w_1 = w_2 = 1$. The variational problem is thus simply given by

$$\Phi(g^1, g^2) = \|f - (F_1^* g^1 + F_2^* g^2)\|^2 + \alpha_1 \|g^1\|_{\ell_1} + \alpha_2 \|g^2\|_{\ell_1} \ ,$$

and the minimization by Gaussian surrogates yields the following iteration

$$\begin{pmatrix} (g^1)_{m+1} \\ \\ (g^2)_{m+1} \end{pmatrix} = \begin{pmatrix} S_{\alpha_1/2C^2} \left(C^{-2} \{ F_1 f + C^2 (g^1)_m - F_1 F_1^* (g^1)_m - F_1 F_2^* (g^2)_m \} \right) \\ \\ S_{\alpha_2/2C^2} \left(C^{-2} \{ F_2 f + C^2 (g^2)_m - F_2 F_1^* (g^1)_m - F_2 F_2^* (g^2)_m \} \right) \end{pmatrix} .$$

Since we deal with bases only, the application of $F_1 F_1^*$, $F_1 F_2^*$, $F_2 F_1^*$, and $F_2 F_2^*$ simplifies to the discrete decomposition and reconstruction schemes. If one really goes beyond bases, i.e. using frames, one indeed has to compute (approximate) all the (mixed) gram matrices. This might be of course costly but can be optimized by picking localized and reasonably incoherent frames. As an experimental observation, in case the frame generating analyzing atmos are not reasonably distinct, the scheme is not able to separate the signal components adequately, i.e. all the sequences g^i contain very similar informations.

5.1 Application to Audio Coding

Let us now denote by $\{\phi^1_\lambda\}$ the Fourier system and with $\{\phi^2_\lambda\}$ the wavelet system. Then we consider two frame operators $F_1 : v \mapsto \{\langle v, \phi^1_\lambda \rangle\} = g^1$ and $F_2 : v \mapsto \{\langle v, \phi^2_\lambda \rangle\} = g^2$.

A synthetic Example: In this example we have simulated a signal f that is a composition of two different components: a harmonic wave and noisy perturbation within the interval $[350, 400]$. As a sampled discrete vector it has a total number of 631 coefficients in the time–domain representation. This discrete vector is used as input for our algorithm. The results for $\alpha_1 = \alpha_2 = 0.2$ are visualized in Figure 1. We find that involving the Haar wavelet basis and the Fourier basis splits the signal in very sparse and well separated components. The sparseness evolution of the two individual components can be seen in the 'sparsity' plot in Figure 1 approving that the chosen frames meet quite nicely the signal structure.

Example "Glockenspiel": This data set represents a real audio signal consisting of tonal components and a sequence of (bell) attacks. We again try to apply Haar wavelet and Fourier splitting. For $\alpha_1 = 0.02$ and $\alpha_2 = 0.01$ the results are shown in Figure 2. As expected, the Haar system captures all the bell attacks very well, and, moreover, the Fourier system the tonal components. The sparsity evolution graph shows the rapid decay of the number of wavelet coefficients which can be explained by a fast "bell attacks" localization process through the iteration.

We summarize, whenever the dictionary consists of complementary frames, the proposed algorithm produces a sparse representation in which the individual components overlap inconsiderably. However, a different choice of penalty weights would of course imply a different splitting of the signal: if $\alpha_1 >> \alpha_2$, then almost everything of the signal would be captured by the wavelet system and vice versa.

5.2 Application to Image Restoration and Compression

In what follows we provide evidence that the machinery can naturally be applied for image restoration and compression tasks. We consider as test data "Part of woman image" and verify that non–optimally chosen families of frames may achieve the same reconstruction results but that the sparsity gets essentially worse (what is typically expected).

Example Wavelet–Wavelet: In this case, see Figure 3 ($\alpha_1 = 10$, $\alpha_2 = 20$), we have picked a Daubechies–6–wavelet and the Haar wavelet system. Both bases capture local structures but of different smoothness. In order to provide a comparison of sparsity with the next example, the parameters α_1 and α_2 are chosen such that similar SNR's and relative approximation errors are achieved.

Example Wavelet–Fourier: This example is the same as the latter one except that we have exchanged the Daubechies–6–wavelet system with the Fourier system (here for reasons just explained we have chosen $\alpha_1 = 15$ and $\alpha_2 = 13$). As we may observe in Figure 4 (and since the Haar and the Fourier system are complementary), we achieve much better sparsity as before. The complementary selectivity of wavelets and harmonics manifests here very visible when splitting into local jumps and oscillatory components. We finally conclude that even for image restoration/compression tasks (here only denoising is illustrated, but deblurring – or more generally: inverting operator equations – is by construction possible) the proposed method has demonstrated its capabilities.

6 Conclusion

In this paper we have proposed a Landweber–like iterative scheme that provides a decomposition of the solution of a linear inverse problem into several components. In the present setting, each component is assumed to have an expansion with respect to a preassigned frame. The algorithm allows to put mixed constraints (such as sparsity, smoothness etc.) on the individual frame coefficients. The applicability is numerically tested and approved in a simplified situation in which $A = I$. In this setting we have shown how synthetic, audio and image data can be decomposed into components that can be sparsely represented by means of wavelet and Fourier systems. The audio results can be downloaded from `http://www.zib.de/AG_InverseProblems/wav/`. The parameters α_i are chosen manually in order achieve as much sparsity as possible. This regrettably reflects the current infirmity of the proposed scheme which can hopefully be circumvented in one of the forthcoming papers.

Acknowledgment

The author was partially supported by Deutsche Forschungsgemeinschaft Grants TE 354/1-2, TE 354/3-1. He would like to thank Bruno Torrésani (LATP, CMI Marseille, France) for inspiring discussions and for graciously providing several sound tracks and material for performance comparison, and, moreover, Ingrid Daubechies (Princeton University, USA) and Patrick Wolfe (Havard University, Cambridge, USA) for the short dated three hours meeting where we had discussions on the topic. Finally, the author wishes to thank the referees for their very valuable and helpful comments and suggestions.

References

[1] S. Anthoine. Different Wavelet-based Approaches for the Separation of Noisy and Blurred Mixtures of Components. Application to Astrophysical Data. PhD thesis, Princeton University, `http://www.latp.univ-mrs.fr/ anthoine/`, 2005.

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

[3] 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.

[4] 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.

[5] 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.

[6] L. Daudet and B. Torrsani. Hybrid representations for audiophonic signal encoding. *Signal Processing, Special issue on Image and Video Coding Beyond Standards*, 82(11):1595–1617, 2002.

[7] M. Defrise and C. DeMol. Inverse imaging with mixed penalties. *Conference proceedings*, 2004.

[8] M. Elad, J.-L. Starck, D. Donoho, and P. Querre. Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *to appear in Applied and Computational Harmonic Analysis*, 2006.

[9] F. Jaillet and B. Torresani. Time–frequency jigsaw puzzle: adaptive multiwindow and multilayered Gabor expansions. *preprint*, 2005.

[10] S. Molla and B. Torresani. An hybrid audio scheme using hidden markov models of waveforms. *to appaer in Applied and Computational Harmonic Analysis*, 2005.

[11] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, (73):591–597, 1967.

[12] J.-L. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representation and a variational approach. *IEEE Transaction on Image Processing*, 14:10ff, 2005.

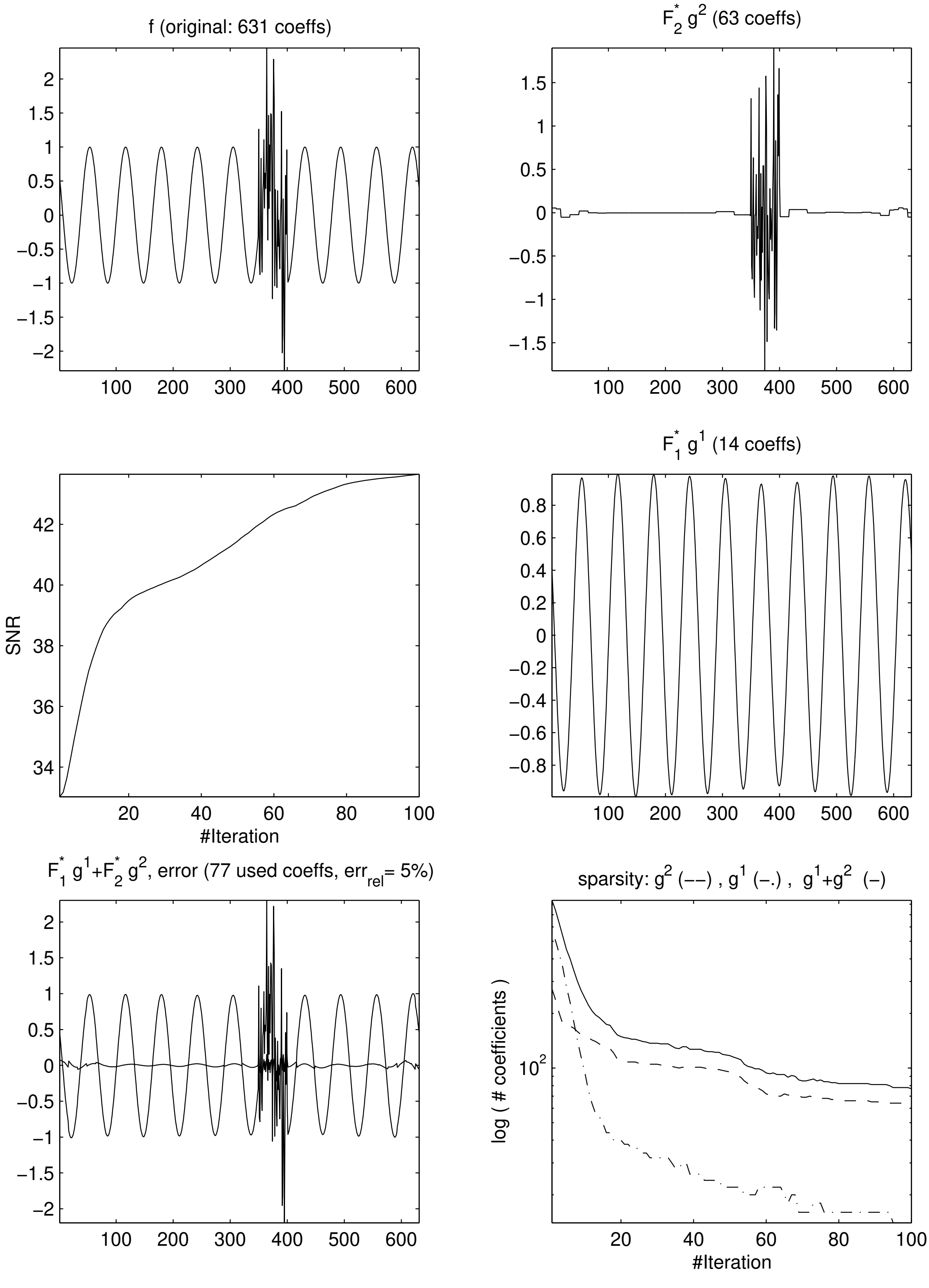

Figure 1: From top left to up right: synthetic data, Haar wavelet component (g^2) (in time domain) after 100 iterations, SNR evolution through the iteration process, Fourier component (g^1) (in time domain) after 100 iterations, reconstruction and error after 100 iterations, and sparsity evolution through the iteration process.

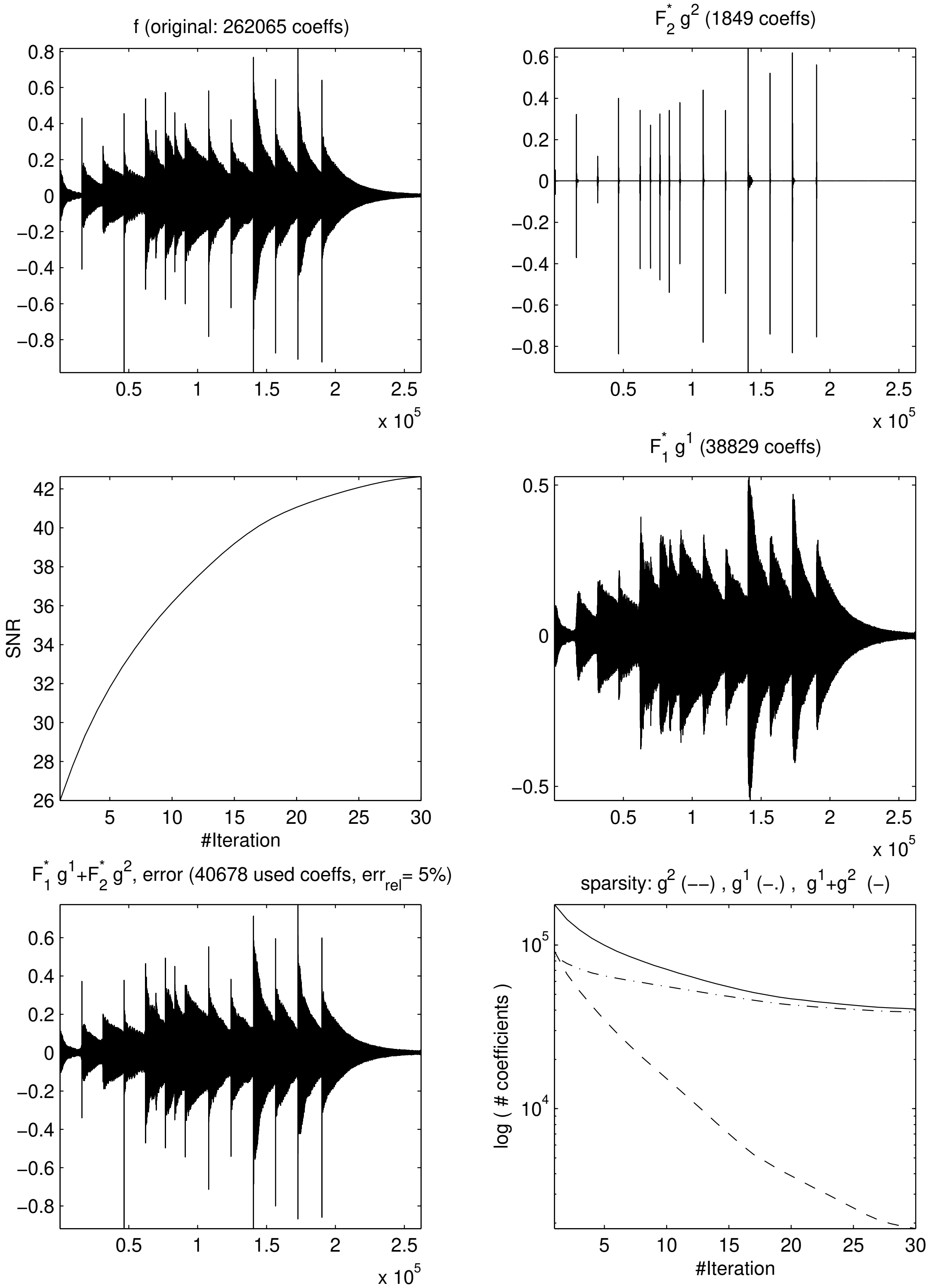

Figure 2: From top left to up right: "Glockenspiel" data, Haar wavelet component (g^2) (in time domain) after 30 iterations, SNR evolution through the iteration process, Fourier component (g^1) (in time domain) after 30 iterations, reconstruction after 30 iterations, and sparsity evolution through the iteration process.

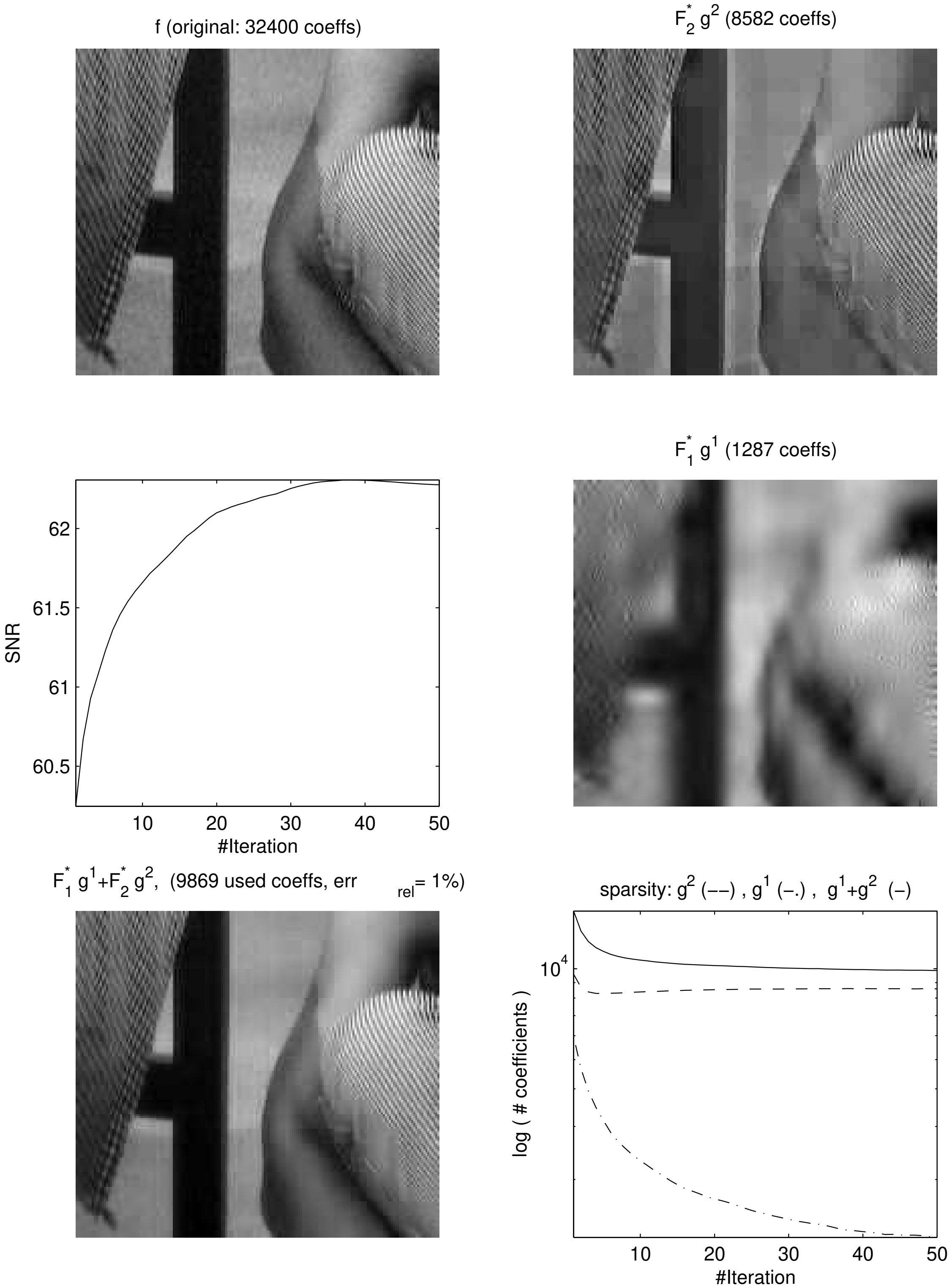

Figure 3: "Part of woman image", Haar wavelet component (g^2) (in time domain) after 50 iterations, SNR evolution through the iteration process, Daubechies–6–wavelet component (g^1) (in time domain) after 50 iterations, reconstruction after 50 iterations, and sparsity evolution through the iteration process.

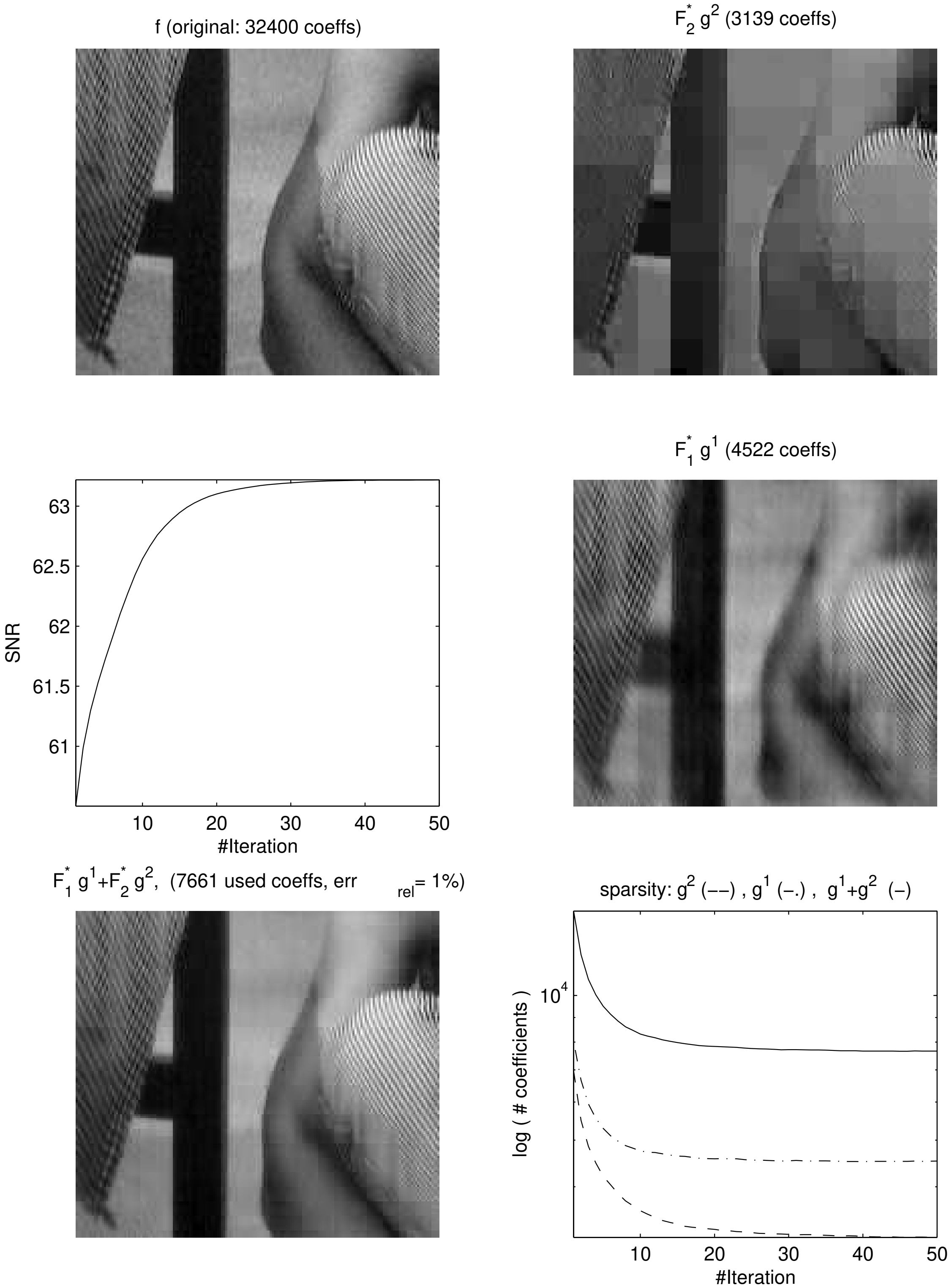

Figure 4: "Part of woman image", Haar wavelet component (g^2) (in time domain) after 50 iterations, SNR evolution through the iteration process, Fourier component (g^1) (in time domain) after 50 iterations, reconstruction after 50 iterations, and sparsity evolution through the iteration process.