SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Adaptive Regularization for Parseval Frames in Image Processing

M. Hintermüller A. Langer

SFB-Report No. 2014-014 October 2014

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

Der Wissenschaftsfonds.

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Adaptive Regularization for Parseval Frames in Image Processing

Michael Hintermüller* and Andreas Langer†

***Abstract*—For image data related to Parseval frames and Gaussian white noise, distributed (or locally adaptive) data fidelity weights for preserving image details while still considerably removing noise from homogeneous image regions in image reconstruction are computed automatically. While the underlying variational model is related to the Rudin-Osher-Fatemi (or total variation) model, the determination of the fidelity weights utilizes the Gumbel statistic for the maximum of a finite number of random variables associated with localized image residuals. An algorithm, which uses hierarchical image decompositions to speed up the iterative parameter adjustment process, is presented and tested numerically for reconstruction from partial Fourier data and for wavelet inpainting, respectively.**

***Index Terms*—locally adaptive regularization, Parseval frames, partial Fourier data, semismooth Newton, wavelet inpainting.**

I. Introduction

IMAGE restoration is one of the fundamental tasks in image processing. Its goal is to reconstruct an image from contaminated data f, which result from a deterioration due to some (linear) transformation operator K (such as convolution, wavelet or Fourier transform) applied to the original image $\hat{u}$ and a subsequent infliction of noise η, i.e., $u_d = K\hat{u} + \eta$. A popular approach to this restoration task rests on variational methods, i.e., the characterization of the reconstructed image u as the solution of a minimization problem of the type

$$\min_u \quad D(u; f) + \alpha R(u), \tag{1}$$

where $D(\cdot; f)$ represents a data fidelity term, $R(\cdot)$ an appropriate filter and $\alpha > 0$ a regularization parameter, which balances data fidelity and filtering. The choice of D is tyically dictated by the type of noise contamination, and R follows some prior information.

Here, we focus on noise η with zero mean and quadratic deviation from the mean equal to $\sigma^2 \geq 0$ Further, we aim at preserving edges in images well. These two pieces of a priori information lead to

$$D(u; f) := \frac{1}{2}\|Ku - f\|^2, \quad \text{and} \quad R(u) = |Du|(\Omega),$$

where the latter represents the total variation of a function u (see (7) below for its definition) and Ω denotes the image domain. Typical choices for the norm in D are the $L^2(\Omega)$-norm or the Euclidean norm on $\mathbb{R}^N$, where the latter is, e.g., relevant when K is a (finite) wavelet or Fourier transform. The resulting model (1) is the well-known Rudin-Osher-Fatemi model [28] which has been studied intensively in the literature; see, e.g., [7], [8], [9], [16], [22], [23], [26], [30], [31] as well as the monograph [34] and the many references therein.

The proper choice of α is delicate. In fact, large α, which would typically be favorable for large homogeneous image regions, not only removes noise, but also details in images. Small α, on the other hand, might be advantagous in regions with image details, but it adversely retains noise in homogeneous image regions. In order to address this issue, in [30] the notion of a scale of an image feature and its influence on α is studied. The subsequent work [29] proposes an update scheme for α within a deterministic regime. We note that instead of considering (1) one may equivalently study $\lambda D(u; f) + R(u)$ with $\lambda = 1/\alpha$. Based on this view and considering a piecewise constant function λ over the image domain, where the partitioning of the image into pieces is due to a pre-segmentation, in [4] scalars λ_i, $i = 1, \ldots$ #pieces, for each segment is computed by an augmented Lagrangian type algorithm. While still remaining in a deterministic regime, interestingly [4] uses a distributed (more precisely a piecewise constant) parameter λ.

Later its was noticed that stable choices of λ resp. α have to incorporate statistical properties of the noise. In this vein, in [2], [17] automated update rules for λ based on statistics of local constraints were proposed. For statistical multiscale methods we refer to [19], [20], [24]. While the methods in [2], [17] are highly competitive in practice, the adjustment of λ relies on the output of K being a deteriorated image again. This, however, limits the applicability of these approaches in situations where K transfers an image into a different type of range space. Particular examples of such operators are (discrete) wavelet or Fourier transforms. It is therefore the goal of this paper to study the approach of [17] in the context of Parseval frames. The latter include wavelet and Fourier transform, thus extending the approach of [17].

The rest of the paper is organized as follows. In the next section we introduce the concept of frames. Algorithmic appraches to adaptive regularization are the subject of section III. Numerical experiments are reported on in section IV.

II. Frames

Let $\mathcal{H}$ be a (separable) Hilbert space with the inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Then the development of this paper rests on the following definitions; cf. [10].

Manuscript received Month Day, Year; revised Month Day, Year; accepted Month Day, Year. Date of publication Month Day, Year; date of current version October 21, 2014. This work was support by ...

*Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany (email: hint@math.hu-berlin.de) (Corresponding author).

†University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany (e-mail: andreas.langer@mathematik.uni-stuttgart.de).

Definition II.1. *A sequence $(h_k)_{k\in\mathbb{N}}$ in $\mathcal{H}$ is called a* Bessel sequence *if there exists a constant $\overline{C} > 0$ such that*

$$\sum_k |\langle h, h_k\rangle|^2 \le \overline{C}\|h\|_{L^2}^2 \quad \textit{for all } h \in \mathcal{H}.$$

Every number $\overline{C}$ satisfying the above inequality is called a *Bessel bound* for $(h_k)_{k\in\mathbb{N}}$.

Definition II.2. *The sequence $(h_k)_{k\in\mathbb{N}}$ of elements in $\mathcal{H}$ is a* frame *for $\mathcal{H}$ if there exist $0 < \underline{C} \le \overline{C} < \infty$ such that for every $g \in \mathcal{H}$*

$$\underline{C}\|g\|^2 \le \sum_k |\langle g, h_k\rangle|^2 \le \overline{C}\|g\|^2. \tag{2}$$

The numbers $\underline{C}$, $\overline{C}$ are called frame bounds*. If $\underline{C} = \overline{C}$ then we call the frame a* tight frame*.*

Note that the frame bounds are not unique. Actually, the optimal upper frame bound is the infimum over all upper frame bounds, and the optimal lower frame bound is the supremum over all lower frame bounds.

A tight frame with $\underline{C} = \overline{C} = 1$ is sometimes called *Parseval frame*, as for example in [3], [25]. However, note that if the sequence $(h_k)_{k\in\mathbb{N}}$ is a tight frame, then after re-normalization of the elements h_k, $k \in \mathbb{N}$, one obtains

$$\sum_k |\langle g, h_k\rangle|^2 = \|g\|^2$$

for all $g \in \mathcal{H}$. This condition is equivalent to the *perfect reconstruction property*

$$g = \sum_k \langle g, h_k\rangle h_k \tag{3}$$

for all $g \in \mathcal{H}$. By Definition II.2 it is clear that an orthonormal basis is a tight frame, see e.g. [10, Theorem 3.4.2] or [36, Satz V.4.9], as well as [37].

Let the sequence $(h_k)_{k\in\mathbb{N}}$ be a frame for $\mathcal{H}$ with frame bounds $\underline{C}$ and $\overline{C}$. Then we define the *analysis operator*

$$T : \mathcal{H} \to \ell^2(\mathbb{N}), \qquad Tg = (\langle g, h_k\rangle)_{k\in\mathbb{N}}$$

and its adjoint T^*, called *synthesis operator*, by

$$T^* : \ell^2(\mathbb{N}) \to \mathcal{H}, \qquad T^*(c_k)_k = \sum_k c_k h_k.$$

With these definitions, property (2) can be written equivalently as $\underline{C}\|g\|^2 \le \|Tg\|^2 \le \overline{C}\|g\|^2$. Hence T is linear and bounded and thus $\|T\| = \|T^*\|$. Consequently by Definition II.1 we have that $\|T^*Tg\|^2 \le \overline{C}\|Tg\|^2$ for any $g \in \mathcal{H}$.

In this paper we consider operators K of the form $S \circ T$, where T is an analysis operator of a Parseval frame and S is a self-adjoint operator with the property $S^* \circ S = S$. Here we have in mind that S is a subsampling projection operator from the whole domain Ω onto a known subdomain $\mathcal{D} \subset \Omega$. Then we obtain that $\|K^*(Ku-f)\|^2 = \|Ku-f\|^2$, where we used that $Sf = f$, since f is the observed signal and therefore unknown in $\Omega \setminus \mathcal{D}$.

A. Fourier Frame

Note that the complex exponential functions $(e_k(x))_k := (\frac{1}{\sqrt{2\pi}}e^{ikx})_k$, $k \in \mathbb{N}$, constitute an orthonormal bases for $\mathcal{H} = L^2(0, 2\pi)$ and hence a Parseval frame for $L^2(0, 2\pi)$. Such a frame is called *Fourier frame* or *frame of exponentials* [10]. By the perfect reconstruction formula (3) there exists, for every $f \in L^2(0, 2\pi)$, an expansion

$$f(x) = \sum_k c_k e_k(x), \tag{4}$$

called *Fourier series*, with *Fourier coefficients* $c_k = \langle f, e_k\rangle$. Actually, if a function $f \in L^2(\mathbb{R})$ is *bandlimited*, i.e., its Fourier transform

$$\mathcal{F}f := \hat{f} := \frac{1}{\sqrt{2\pi}} \int e^{-ikx} f(x)dx = \langle f, e_k\rangle$$

has compact support, then $\hat{f}$ can be represented by its Fourier series. However, for any $f \in L^1(\mathbb{R})$ we have by the Parseval Theorem that $\|f\| = \|\mathcal{F}f\|$.

As an application, we are interested in the reconstruction of medical images from partial Fourier data. In this case the linear operator K equals $S \circ \mathcal{F}$, where S is a downsampling operator which selects only a few frequencies as output.

B. Wavelet Frames

Set $\mathcal{H} = L^2(\Omega)$ with $\Omega \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, and let $\Psi \subset L^2(\Omega)$ be a countable set. The corresponding *wavelet system* (or affine system) $X(\Psi)$ generated by the *mother wavelets* Ψ is defined by

$$X(\Psi) := \{\psi_{n,k} : \ n \in \mathbb{Z}, k \in \mathbb{Z}^d\},$$

where $\psi_{n,k} : x \mapsto 2^{nd/2}\psi(2^n x - k)$ with $\psi \in \Psi$. The associated analysis operator is given by $T : g \mapsto (\langle g, h\rangle)_{h\in X(\Psi)}$. Conversely, the function $g \in L^2(\Omega)$ can be well approximated by

$$g \approx T^*c = \sum_{h\in X(\Psi)} c_h h = \sum_{n\in\mathbb{Z}, k\in\mathbb{Z}^d} c_{n,k}\psi_{n,k}, \tag{5}$$

where $c = (c_h)_{h\in X(\Psi)}$ denote the wavelet coefficients of g.

If the functions in Ψ satisfy some mild smoothing conditions and if each of the mother wavelets has at least one vanishing moment, i.e., $\widehat{\psi}(0) = 0$, for all $\psi \in \Psi$, then $X(\Psi)$ is a Bessel system, see for example [11], [14]. We call the wavelet system $X(\Psi)$ a *wavelet frame*, if $X(\Psi)$ is a frame. If the system $X(\Psi)$ forms an orthogonal bases, then the coefficients $c_{n,k}$ in (5) are given by $c_{n,k} = \langle g, \psi_{n,k}\rangle$, while in general we have that $c_{n,k} = \langle g, \tilde{\psi}_{n,k}\rangle$, where $\tilde{\psi}_{n,k}$ is a dual bases of $\psi_{n,k}$, see [12]. The, perhaps, most traditional example of a function ψ for which $X(\Psi)$ constitutes an orthonormal basis for $L^2(\mathbb{R})$ and therefore forms a tight frame (and after renormalization even a Parseval frame) is the Haar function, see [12] for more details. Indeed, in our experiments we concentrate on orthogonal wavelets based on a multiresolution analysis. For more information on how to obtain such wavelets see [12].

Our motivating application is filling in missing wavelet-coefficients, which may have got lost during communication. Hence we have $K = P_I \circ T$, where T is the above defined analysis operator and P_I a sampling operator from the whole domain onto the known index set I.

III. Adaptive Regularization Approach

In this paper we focus on image restoration problems. Using the edge-preserving total variation regularization in a variational approach to image restoration leads to solving the constrained minimization problem

$$\min_u |Du|(\Omega) \quad \text{subject to (s.t.)} \int_\Omega Kudx = \int_\Omega f dx, \quad \int_\Omega (Ku-f)^2 dx = \sigma^2|\Omega|, \tag{6}$$

where $\Omega \subset \mathbb{R}^2$ is a simply connected domain with Lipschitz boundary $\partial\Omega$ and $|\Omega|$ its volume, K is a linear bounded operator, $f = K\hat{u} + \eta$ the observed data with $\hat{u}$ the unknown true signal, and η represents white Gaussian noise with zero mean and standard deviation σ. These statistical properties of η motivate the constraints in (6); see, e.g., [8]. Moreover, $|Du|(\Omega)$ denotes the total variation of $u \in L^1(\Omega)$ in Ω, defined by

$$|Du|(\Omega) = \sup\{\int_\Omega u \ \mathrm{div}\, \vec{p}\, dx : \vec{p} \in C_0^1(\Omega),\ \|\vec{p}\|_{L^\infty(\Omega)} \le 1\}. \tag{7}$$

Here and below, $L^q(\Omega)$, with $q \in [1,\infty]$, denotes the usual Lebesgue space [1], and $C_0^l(\Omega)$, $l \in \mathbb{N}$, is the space of l-times continuously differentiable functions with compact support in Ω. Usually (6) is addressed via the following unconstrained optimization problem:

$$\min_u |Du|(\Omega) + \frac{\lambda}{2}\int_\Omega |Ku-f|^2 dx \tag{8}$$

for a given constant $\lambda > 0$. For our purposes we modify the objective in (8) in order to handle the presence of the operator K. Hence, instead of tackling (8) directly we introduce a so-called *surrogate functional* [13] for $a \in L^2(\Omega)$, which is defined as

$$\begin{aligned} S(u,a) :&= |Du|(\Omega) + \frac{\lambda}{2}\Big(\|Ku-f\|^2_{L^2(\Omega)} + \delta\|u-a\|^2_{L^2(\Omega)} \\ &\quad - \|K(u-a)\|^2_{L^2(\Omega)}\Big) \\ &= |Du|(\Omega) + \frac{\lambda\delta}{2}\|u - f_K(a)\|^2_{L^2(\Omega)} + r(a,K,f,\lambda), \end{aligned} \tag{9}$$

with $f_K(a) := a - \frac{1}{\delta}K^*(Ka-f) \in L^2(\Omega)$, and where we assume $\delta > \|K\|^2$. Moreover, r is a function independent of u. We observe in (9) that the variable u is not longer directly affected by the action of K. Rather, minimizing $S(u,a)$ for fixed a resembles a typical image denoising problem. In order to approach a solution of (8), we consider the following iteration.

Surrogate Iteration: Choose $u^{(0)} \in L^2(\Omega)$ and compute for $k = 0, 1, 2, \ldots$

$$u^{(k+1)} = \arg\min_u |Du|(\Omega) + \frac{\lambda\delta}{2}\int_\Omega |u - f_K^{(k)}|^2 dx. \tag{10}$$

with $f_K^{(k)} := f_K(u^{(k)})$.

It can be shown that the iterative procedure (10) generates a sequence $(u^k)_{k\in\mathbb{N}}$ which converges to a minimizer of (8); see [13], [15]. It is well known that the minimization problem in (10) is strictly convex and can be efficiently solved by standard algorithms such as the primal-dual first-order algorithm [7], the split Bregman method [21], or the primal-dual semismooth Newton algorithm [23]. We note that the latter algorithm is able to solve the optimization problem in (8) directly.

Now we consider the sequence of problems in (10) and make the parameter λ locally dependent, i.e., $\lambda : \Omega \to \mathbb{R}_+$ is now a function. This leads to the following iterative procedure

$$u^{(k+1)} = \arg\min_u J_\lambda^{(k)}(u) := \frac{\delta}{2}\int_\Omega \lambda(x)(|u - f_K^{(k)}|^2)(x)dx + |Du|(\Omega), \quad \text{for } k = 0,1,2\ldots. \tag{11}$$

Problem (11) is related to the constrained minimization problem

$$\begin{aligned} &\min_u |Du|(\Omega) \\ &\text{s.t. } S(u) - A \le 0 \quad \text{almost everywhere (a.e.) in } \Omega, \end{aligned} \tag{12}$$

where $S(u)(x) = \int_\Omega w(x,y)(u - f_K^{(k)})^2(y)dy$ with ω being some localization filter and a constant $A \in \mathbb{R}_+$ depending on σ and K. For $u = u^{(k+1)}$, the constraint function in (12) reads

$$S(u^{(k+1)}) = \int_\Omega w(x,y)(u^{(k+1)} - u^{(k)} + \frac{1}{\delta}K^*(Ku^{(k)} - f))^2(y)dy. \tag{13}$$

Given the convergence result for scalar λ alluded to in connection with (10), for $k \to \infty$ one may expect that the term $u^{(k+1)} - u^{(k)}$ vanishes. Then it is enough to require that $\int_\Omega w(x,y)(\frac{1}{\delta}K^*(Ku^{(k)} - f))^2(y)dy \le A$. This motivates the locally, i.e., point-wise constrained minimization problem

$$\begin{aligned} &\min_u |Du|(\Omega) \\ &\text{s.t. } \int_\Omega w(\cdot,y)(\frac{1}{\delta}K^*(Ku-f))^2(y)dy \le A \quad \text{a.e. in } \Omega. \end{aligned} \tag{14}$$

Next we discuss the choice of A. We have that $\eta = K\hat{u} - f$ and thus $K^*(K\hat{u} - f) = K^*\eta$. Now we recall that η is assumed to be normally distributed with mean zero and standard deviation σ. But in general it is not obvious how $K^*\eta$ is distributed. For bounded linear operators $K : L^2(\Omega) \to L^2(\Omega)$ we have that there exists a constant $C > 0$ such that $\|Ku\| \le C\|u\|$ for any $u \in L^2(\Omega)$. With $\|K\| = \|K^*\|$ it follows that $\|K^*(Ku-f)\| \le C\|Ku - f\|$ and consequently the value A in (14) can be estimated from above by $A \le \frac{C\sigma^2}{\delta^2}$. If K is an analysis operator of a Parseval frame, for example $K = \mathcal{F}$, then we have that $A = \frac{\sigma^2}{\delta^2}$. Since we are interested in operators $K = S \circ T$, where S is a subsampling operator and T either the Fourier-Transform or an analysis operator of an orthogonal wavelet frame, the value A is given by $A = \frac{\sigma^2}{\delta^2}$. Thus, we can construct an updating scheme for λ as it was done in [17].

For this purpose, we recall the idea behind the λ-update rule in [17]: We assume that Ω_m is a discrete domain containing $m \times m$ pixels, $m \in \mathbb{N}$. For ease of notation we set $\Omega := \Omega_m$, and we note that the extension to rectangular pixel domains

is obvious. By $r(u) := f_K(u) - u$ we denote the discrete residual image with $r(u), f_K(u), u \in \mathbb{R}^{m^2}$ and reshape these quantities as $m \times m$ matrices. Note that in our setting $r(u)$ corresponds to K^* applied to $Ku - f$. For $\omega \in \mathbb{N}$ we define

$$\Omega_{i,j}^{\omega} = \{(s+i, t+j) : -\frac{\omega-1}{2} \leq s, t \leq \frac{\omega-1}{2}\},$$

a set of pixel-coordinates in a ω-by-ω window centered at (i, j) (with a symmetric extension at the boundary). In the discrete setting we regard η as an array of independent normally distributed random variables with zero mean and variance σ^2. Then $K^*\eta$ is also an array of independent normally distributed random variables with zero mean and variance A [18], [27]. The random variable

$$T_{i,j}^{\omega} = \frac{1}{A\delta^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (K^*\eta)_{s,t}^2$$

has then the χ^2-distribution with ω^2 degrees of freedom, i.e., $T_{i,j} \sim \chi^2_{\omega^2}$. If $u = \hat{u}$ satisfies $\eta = f_K(\hat{u}) - \hat{u}$, then, by specifically choosing the mean filter for w and applying it to the residual, we obtain

$$\begin{aligned} S_{i,j}^{\omega} &:= \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (f_K(\hat{u})_{s,t} - \hat{u}_{s,t})^2 \\ &= \frac{1}{\omega^2\delta^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} ((K^*\eta)_{s,t})^2 = \frac{A}{\omega^2} T_{i,j}^{\omega}. \end{aligned}$$

If u is an over-smoothed restored image, then the residual $f_K(u) - u$ contains details and we expect

$$S_{i,j}^{\omega} > \frac{A}{\omega^2} T_{i,j}^{\omega}.$$

Therefore we are interested in a bound B such that $S_{i,j}^{\omega} > B$ implies that some details in the neighborhood of the pixel (i, j) are left in the residual image $r(u)$. Utilizing distributions of extremal values (Gumbel distribution) and relaxing statistical dependencies, in [17] it was found that a good bound is

$$B := \frac{A}{\omega^2}(\mathbb{E}(T_{\max}) + \mathfrak{d}(T_{\max}))$$

where $T_{\max}$ is the maximum value of m^2 observations distributed along the χ^2-distribution with ω^2-degrees of freedom, $\mathbb{E}(\cdot)$ is the expected value and $\mathfrak{d}(\cdot)$ the standard deviation of a random variable. We note, however, that in our numerical experiments we found out that this bound is slightly too pessimistic for our applications and rather a lower bound should be used. In particular, since we are looking for a solution u^* such that $\sum_{(i,j)\in\Omega}(r(u^*)_{i,j})^2 = A|\Omega|$, we expect that if $S_{i,j}^{\omega} > A$, then u is an over-smoothed restoration and the residual $r(u)$ contains details. Hence, in our algorithm we choose $B := A$.

If now $S_{i,j}^{\omega} \in [0, B)$, then the residual should ideally only contain noise, while otherwise we suppose there are image details contained in the residual image in $\Omega_{i,j}^{\omega}$. Hence one defines the following modified local variance estimator

$$\tilde{S}_{i,j}^{\omega} := \begin{cases} S_{i,j}^{\omega} & \text{if } S_{i,j}^{\omega} \geq B, \\ \sigma^2 & \text{otherwise.} \end{cases}$$

This leads to the following update rule of λ, where we initially choose λ to be a small positive constant, which yields an over-smoothed restoration, i.e. most of the details remain in the residual. The parameter (function or vector, in the discrete setting) λ is subsequentially updated as follows: Let $\lambda_n \in \mathbb{R}^{m^2}$ be given. Then set

$$\lambda_{n+1} = \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} \left((\lambda_n)_{s,t} + \rho_n \left(\sqrt{(\tilde{S}_n^{\omega})_{s,t}} - \sigma \right) \right), \tag{15}$$

where $\rho_n = \frac{\|\lambda_n\|}{\sigma}$ denotes a positive scaling parameter.

The overall restoration algorithm with spatially adapted regularization is specified next. In its statement we consider $J_{\lambda,m}^k$ to be a suitable discrete version of J_{λ}^k. Moreover, for the ease of notation we omit the subscript m, which represents discrete quantities such as K_m, f_m, u_m or λ_m etc., respectively.

Adaptive Algorithm: Choose $\lambda_0 \in \mathbb{R}^{m^2}$, with $\lambda_0 > 0$, and $u_0 \in \mathbb{R}^{m^2}$, set $n := 0$, and iterate until a convergence criterion is satisfied:

1) Set $u^{(0)} := u_n$. For $k = 0, 1, 2, \ldots$ set $f_K(u^{(k)}) := u^{(k)} - \frac{1}{\delta}K^*(Ku^{(k)} - f)$ and compute
$$u^{(k+1)} = \arg\min_u J_{\lambda_n}^{(k)}$$
Let $u_{(n+1)}$ denote the outcome of this iteration.
2) Update λ_n to obtain λ_{n+1} according to (15); set $n := n + 1$.

While this iteration is functional in its own right and yields a convergent scheme, its convergence speed may still be improved. Indeed, following [17] we accelerate the above adaptive algorithm by employing a hierarchical decomposition of the image into scales. This idea, introduced by Tadmor, Nezzar and Vese in [32], [33], utilizes concepts from interpolation theory to represent a noisy image as the sum of "atoms" $u_{(\ell)}$, where every $u_{(\ell)}$ extracts features at a scale finer than the one of the previous $u_{(\ell-1)}$. Considering dyadic scales with a scalar λ, for example, the hierarchical decomposition operates as follows:

1) Choose $0 < \lambda_0 \in \mathbb{R}^{m^2}$, $u^{(0)} \in \mathbb{R}^{m^2}$ and iterate for $k = 0, 1, \ldots, K_0 \in \mathbb{N}$
$$u_0^{(k+1)} = \arg\min_u J_{\lambda_0}^{(k)}(u). \tag{16}$$
2) For $\ell = 0, 1, \ldots$ set $\lambda_{\ell+1} := 2^{\ell+1}\lambda_0$ and $v_\ell := f - Ku_\ell$. Then compute for $k = 0, 1, \ldots, K_{\ell+1} \in \mathbb{N}$
$$\hat{u}_\ell^{(k+1)} = \arg\min_u J_{\lambda_{\ell+1}}^{(k)}(u)$$
where $f_K(u_\ell^{(k)}) = u_\ell^{(k)} - \frac{1}{\delta}K^*(Ku_\ell^{(k)} - v_\ell)$ in $J_{\lambda_{\ell+1}}^{(k)}$. Set $u_{\ell+1} := u_\ell + \hat{u}_\ell^{(K_{\ell+1}+1)}$.

In order to extract image features at a finer scale in the spirit of the above adaptive version of the technique by Tadmor, Vese and Nezzar, we correspondingly modify the iterative adaption

of λ in (15) as suggested in [17] by setting

$$\begin{aligned}(\tilde{\lambda}_n)_{s,t} &:= (\lambda_n)_{s,t} + \rho_n \left(\sqrt{(\tilde{S}_n^{\omega})_{s,t}} - \sigma \right) \\ \lambda_{n+1} &= \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} \zeta \min\left((\tilde{\lambda}_n)_{s,t}, L \right),\end{aligned} \tag{17}$$

where $\zeta \geq 1$ and L is a large positive constant to ensure uniform boundedness of $(\lambda_k)_{k\in\mathbb{N}}$. Then the hierarchical spatial adaptiv algorithm can be written as follows.

sSA-TV Algorithm: Choose $u_0 \in \mathbb{R}^{m^2}$, $\lambda_0 \in \mathbb{R}_+^{m^2}$ and set $\zeta = 2$ and $n := 0$

1) If $n = 0$ solve (16) with $u^{(0)} := u_0$; else compute $v_n = f - Ku_n$. Set $\hat{u}_n^{(0)} := 0$ and compute for $k = 0, 1, 2, \ldots, K_n \in \mathbb{N}$
$$\hat{u}_n^{(k+1)} = \arg\min_u J_{\lambda_n}^{(k)}$$
with $f_K(\hat{u}_n^{(k)}) = \hat{u}_n^{(k)} - \frac{1}{\delta} K^*(K\hat{u}_n^{(k)} - v_n)$ in $J_{\lambda_n}^{(k)}$. Let $\hat{u}_n^{(K_n+1)}$ denote the corresponding solution.
2) Update $u_{n+1} := u_n + \hat{u}_n^{(K_n+1)}$.
3) Compute residual $r(u_{n+1})$ and stop if $\zeta < 2$ and $A > \|r(u_{n+1})\|$; otherwise do: If $\|r(u_{n+1})\| > A$ goto step 4); if $A > \|r(u_{n+1})\| > 0.9A$ stop; if $0.9A > \|r(u_{n+1})\|$ set $u_{n+1} := u_n$, $\lambda_n := \lambda_{n-1}$ (unless $n = 0$), and $\zeta = 1.7$ and continue with step 4)
4) Update λ_{n+1} based on u_{n+1} and (17).
5) Set $n := n + 1$ and return to step 1).

Note that we expect noise in the reconstruction u_n if the residual $\|u_n - f_K(u_n)\|^2$ is much smaller than A. Therefore we introduce the lower bound $0.9A$, i.e., 90 percent of the value A, which should ideally bound the residual form below. Hence, the sSA-TV algorithm is terminated as soon as the residual falls into the interval $[0.9A, A]$. If the norm of the residual $r(u_n)$ drops below the lower bound, then we reset $\zeta = 1.7$ to obtain a residual with $\|r(u_n)\| \leq \|r(u_{n+1})\|$ in the next iteration, and we stop the algorithm if $\|r(u_{n+1})\| \leq A$. For solving the minimization problems in the proposed algorithm we use the primal-dual Newton method suggested in [23]. The parameters in the primal-dual Newton algorithm are chosen as $\mu = 10^{-6}$, $\gamma = 10^{-4}$ and iteration is terminated as soon as its residual is smaller than 10^{-4}; see [23] for a detailed explanation of the algorithm. In order to obtain a sufficiently good approximation of the minimizer in each iteration n in step 2), we set K_n adaptively such that $\|\hat{u}_n^{(K_n+1)} - \hat{u}_n^{(K_n)}\| \leq 10^{-6}$.

IV. Numerical Experiments

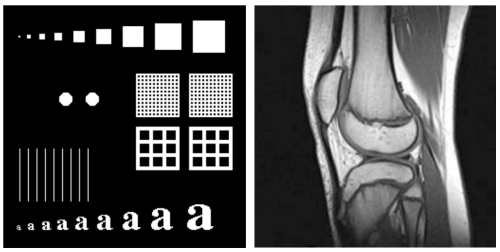
In the following we present numerical experiments for studying the behavior of the sSA-TV algorithm with respect to its image restoration capabilities and its stability concerning the choice of λ_0 and the window-size ω. We also compare the results obtained by the sSA-TV algorithm with the ones due to the primal-dual Newton algorithm for the experimentally best scalar choice of λ. The primal-dual Newton algorithm is terminated as soon as the norm of the residual is smaller than 10^{-4} and its parameters are chosen as before, i.e., $\mu = 10^{-6}$, $\gamma = 10^{-4}$. The performance of these methods is compared quantitatively by means of the peak signal-to-noise-ratio (PSNR) [5], which is widely used as an image quality assessment measure, and the structural similarity measure (MSSIM) [35], which relates to preceived visual quality better than PSNR. In general, when comparing PSNR- and MSSIM-values, large values indicate better reconstruction than smaller values. We also emphasize that for scalar λ other solvers (than primal-dual Newton) may be utilized without changing the conclusions of our findings. This is due to the fact that all of these methods aim at solving the problem (8).

In our experiments we concentrate on two application: (i) The reconstruction of partial Fourier-data which is a medical image processing task, typically related to magnetic resonance imaging (MRI); (ii) Filling-in of missing or damaged wavelet coefficients, which is sometimes also called *wavelet inpainting* [6], [38].

In all examples in this paper the image intensity range is scaled to $[0, 1]$. For both applications we have that $\|K\| = 1$, and since δ has influence on the convergence speed to the minimizer, i.e., the larger δ the slower is the convergence of the surrogate iteration, we set $\delta = 1.1$ in the sSA-TV Algorithm.

A. Reconstruction of Partial Fourier-Data

In magnetic resonance imaging one wishes to reconstruct an image which is only given by partial Fourier data and additionally deteriorated by some additive Gaussian noise with zero mean and standard deviation σ. Hence, the linear bounded operator is $K = S \circ \mathcal{F}$, where $\mathcal{F}$ is the $2D$ Fourier matrix and S is a downsampling operator which selects only a few output frequencies.



(a) (b)

Fig. 1. *Images obtained by full Fourier-data (a) "phantom" (b) "knee"*

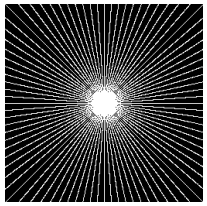
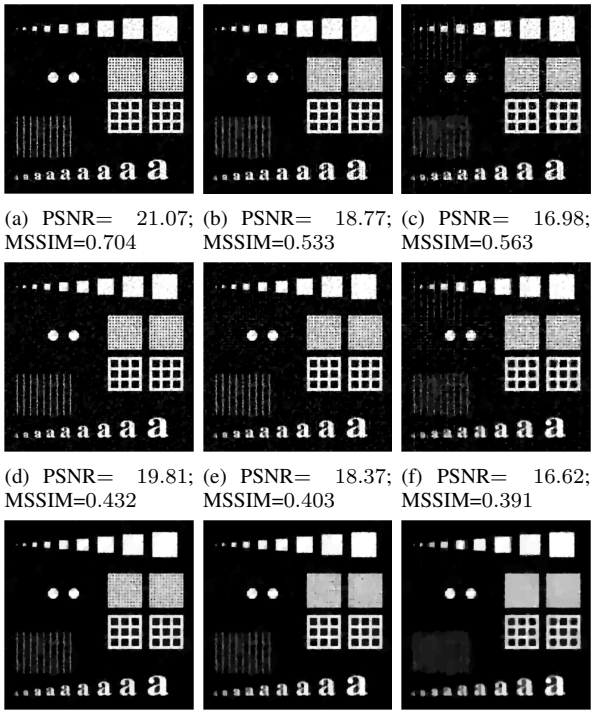


Fig. 2. *Sampling domain in the frequency plane, i.e., sampling operator S.*

The frequencies are usually sampled along radial lines in the frequency domain, in particular in our experiments along 44, 88, 132, and 176 radial lines, as visualized in Figure 2, and some additive Gaussian white noise with zero mean and standard deviation σ is added.

In our first example we consider a phantom-image of size 256×256 pixels, see Figure 1(a), transformed to its Fourier frequencies. As mentioned before, we sample the frequencies along radial lines and add some additive Gaussian noise with zero mean and standard deviation σ. In particular, we reconstruct the image via the sSA-TV algorithm by setting $\lambda_0 = 0.2$ and $\omega = 11$ and test its performance for different noise-levels, i.e., $\sigma \in \{0.3, 0.1, 0.05\}$. The obtained results are compared with the ones of the primal-dual Newton method with a scalar λ, see Table I. For the latter method we perform many experiments with different scalar λ-values and we present here the results with the largest PSNR- and MSSIM-values, which are in general not achieved by the same scalar λ, i.e., the λ, which yields the largest PSNR, need not necessarily also yield the largest MSSIM. The best values of these two quality measures are summarized in Table I. The sSA-TV method typically outperforms the primal-dual Newton method with respect to PSNR and MSSIM. Only when little information is available (strong subsampling) for the reconstruction, then the sSA-TV method may not outperform the best experimental scalar λ. Note that a sampling along 44, 88, 132, and 176 radial lines means that 21%, 39%, 55%, and 69% of the frequency data are available, respectively. Of course, if only very little data is available most details are lost, which explains why our algorithm may not help to improve the reconstruction. However, when we look at the reconstruction for $\sigma = 0.3$ and a sampling along 132, 88, and 44 radial lines, we see in Figure 3 that the sSA-TV improves the restoration in all cases. In particular, in Figure 3 we plot the results of the sSA-TV and the "best" results, in the sense of PSNR and MSSIM, of the primal-dual Newton method. The "best" PSNR-result of the primal dual Newton method seems to regularize only a little, which preserves features and therefore gives a high PSNR value, but the restoration contains more noise than the one generated by the sSA-TV. On the other hand the "best" MSSIM-result over-smoothes the reconstruction and hence details, like the dots in the squares, are lost in the solution. From this example we observe that the sSA-TV generates a reconstruction which removes noise in particular in the uniform parts and leaves details in the reconstruction at the same time. This observation somehow shows that one may not rely on only one of these quality measures but may consider both or a combination of both as a trustful reference, i.e., a reconstruction should have large values in both measures.

Next we consider an MRI-image of size 200×200 pixels of a human knee, see Figure 1(b). Similarly as above, the sSA-TV method outperforms the primal-dual Newton method with respect to PSNR and MSSIM when the number of samples is large enough, i.e., when a reasonable amount of frequencies are considered to reconstruct the image, see Table II. Now a sampling along 44, 88, 132, and 176 radial lines means that 26%, 48%, 66%, and 81% of the frequency data are available, respectively. In Figure 4 we show the reconstructions



(a) PSNR= 21.07; MSSIM=0.704 (b) PSNR= 18.77; MSSIM=0.533 (c) PSNR= 16.98; MSSIM=0.563

(d) PSNR= 19.81; MSSIM=0.432 (e) PSNR= 18.37; MSSIM=0.403 (f) PSNR= 16.62; MSSIM=0.391

(g) PSNR= 18.19; MSSIM=0.679 (h) PSNR= 16.86; MSSIM=0.650 (i) PSNR= 15.86; MSSIM=0.607

Fig. 3. *Results obtained by the sSA-TV (first row) and primal-dual Newton (second row - best PSNR-results; third row - best MSSIM-results) for* $\sigma = 0.3$ *and* 132 *(left column),* 88 *(mid column), and* 44 *(right column) radial lines.*

of the two methods for a sampling along 88 radial lines and with additive Gaussian noise with zero mean and $\sigma = 0.1$. We observe that the sSA-TV method best removes noise in uniform parts while preserving the edges in the image.

Independently of the noise level and the number of radial sampling lines, the final adaptive parameter λ obtained by our parameter choice rule is indeed selected according to the distribution of features in the image. In Figure 5 we show the final λ, where light gray indicates large values and dark gray refers to small values. We see that in detail regions λ is large in order to preserve details, while in uniform parts it is small to remove noise considerably.

1) Dependence on the initial choice λ_0: Next we investigate our algorithm concerning its stability with respect to the initial λ_0. Therefore we test the sSA-TV method for $\lambda_0 \in \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 1.5, 2, 2.5, 3\}$ and plot the PSNR- and MSSIM-values of the obtained reconstructions in Figure 6. Since the parameter λ controls the trade-off between a good data fit and the regularization coming from the TV-term, it has an effect on the variance of the residual $\|u^{(k)} - z(u^{(k)})\|_2$. This can be seen from the plots in Figure 6 where we also specify the number of updates on λ for each λ_0. From our tests we conclude that the sSA-TV method generates reconstructions with large PSNR and MSSIM when the initial λ_0 is sufficiently small, while for large λ_0 there is no guarantee for a good restoration. This can be attributed to the fact that our λ-update operates by increasing λ from an initial guess.

		primal dual Newton		sSA-TV method	
σ	# rays	PSNR	MSSIM	PSNR	MSSIM
0.3	176	21.19 ($\lambda = 7$)	0.710 ($\lambda = 5$)	**22.55**	**0.865**
0.3	132	19.81 ($\lambda = 10$)	0.679 ($\lambda = 5$)	**21.24**	**0.727**
0.3	88	**18.37** ($\lambda = 12$)	0.650 ($\lambda = 5$)	18.19	**0.706**
0.3	44	16.62 ($\lambda = 15$)	**0.607** ($\lambda = 5$)	**16.85**	0.571
0.1	176	29.84 ($\lambda = 25$)	0.943 ($\lambda = 12$)	**30.64**	**0.960**
0.1	132	**28.32** ($\lambda = 30$)	0.932 ($\lambda = 12$)	28.19	**0.960**
0.1	88	26.22 ($\lambda = 40$)	0.905 ($\lambda = 15$)	**26.75**	**0.941**
0.1	44	**20.81** ($\lambda = 40$)	**0.775** ($\lambda = 15$)	20.40	0.721
0.05	176	35.61 ($\lambda = 50$)	0.981 ($\lambda = 25$)	**35.80**	**0.983**
0.05	132	33.97 ($\lambda = 50$)	0.979 ($\lambda = 25$)	**34.92**	**0.980**
0.05	88	**31.80** ($\lambda = 75$)	0.970 ($\lambda = 25$)	31.12	**0.972**
0.05	44	22.56 ($\lambda = 50$)	**0.863** ($\lambda = 30$)	**23.39**	0.764

TABLE I

For the "phantom"-image we test for different standard deviations σ and different ratios of known Fourier coefficients the solution of the total variation minimization with a constant regularization parameter and compare it with the solution of the adaptive total variation approach with $\lambda_0 = 0.2$ and $\omega = 7$ in the sense of the quality measures PSNR and MSSIM.

		primal dual Newton		sSA-TV method	
σ	# rays	PSNR	MSSIM	PSNR	MSSIM
0.3	176	24.46 ($\lambda = 5$)	0.780 ($\lambda = 5$)	**25.21**	**0.802**
0.3	132	24.44 ($\lambda = 5$)	0.785 ($\lambda = 5$)	**25.07**	**0.804**
0.3	88	24.40 ($\lambda = 7$)	0.782 ($\lambda = 5$)	**24.69**	**0.793**
0.3	44	**23.83** ($\lambda = 7$)	0.759 ($\lambda = 5$)	23.51	**0.762**
0.1	176	29.81 ($\lambda = 20$)	0.896 ($\lambda = 15$)	**30.47**	**0.909**
0.1	132	29.64 ($\lambda = 20$)	0.894 ($\lambda = 20$)	**30.32**	**0.908**
0.1	88	29.05 ($\lambda = 25$)	0.886 ($\lambda = 20$)	**29.42**	**0.892**
0.1	44	27.06 ($\lambda = 30$)	0.845 ($\lambda = 25$)	**27.25**	**0.853**
0.05	176	33.54 ($\lambda = 20$)	0.942 ($\lambda = 40$)	**33.89**	**0.944**
0.05	132	33.08 ($\lambda = 50$)	**0.940** ($\lambda = 40$)	**33.21**	0.938
0.05	88	31.80 ($\lambda = 50$)	0.927 ($\lambda = 50$)	**32.49**	**0.935**
0.05	44	**28.71** ($\lambda = 75$)	0.879 ($\lambda = 50$)	28.70	**0.883**

TABLE II

For the "knee"-image we test for different standard deviations σ and different ratios of known fourier coefficients the solution of the total variation minimization with a constant regularization parameter and compare it with the solution of the adaptive total variation approach with $\lambda_0 = 0.3$ and $\omega = 7$ in the sense of the quality measures PSNR and MSSIM.

Thus, in case the latter is already too large for producing good reconstructions, the method could only recover if reductions in λ would be allowed as well.

2) Dependence on window size ω: Next we test our algorithm for different values of the window size ω varying from 3 to 23. Figure 7 shows the PSNR- and MSSIM-values of the restoration of the phantom image (degraded by Gaussian noise with $\sigma = 0.1$) via the sSA-TV method with $\lambda_0 = 0.2$. We observe that the PSNR and MSSIM are varying only slightly with respect to changing window-size. In particular, we observe that whenever the algorithm needs one or more additional steps to terminate, the PSNR exhibits an upwards jump, while the MSSIM seems to behave exactly in the opposite way. This behavior may be attributed to the observation that when the final residual is close to the lower bound $0.9A$, larger PSNR and lower MSSIM values are expected than when the residual is close to the upper bound A. Moreover, large window-sizes, i.e., $\omega > 11$, reduce the MSSIM considerably. These tests recommend to use a rather small window-size as done in our previous experiments, where we set $\omega = 7$.

B. Wavelet Inpainting

Wavelet inpainting is the problem of filling in missing data in the wavelet domain. This data may be lost by conventional communication techniques, which cannot provide an error free transmission. Hence, firstly a certain amount of wavelet coefficients is missing and secondly there might be some noise added to the data. In our experiments we consider random loss of data, which we specify with the amount of data available by its ratio. For example, if ratio $= 0.8$, then 80% of the data is available or in other words, 20% of the original data got lost. Additionally, the observed data is contaminated by Gaussian noise with zero mean and standard deviation σ.

As above we study the performance of the sSA-TV method and compare its results with the restorations of the primal-dual Newton method. For the latter method we again perform many experiments with different scalar λ-values and present here the results with the largest PSNR- and MSSIM-values.

For our investigation we consider the two images shown in Figure 8 both of size 256×256 pixels. We test our sSA-TV method for different noise-levels, i.e., $\sigma \in \{0.3, 0.1\}$, and different ratios, i.e., $0.9, 0.8, 0.7, 0.6$, and 0.5. The obtained PSNR- and MSSIM-values of the solution generated by the sSA-TV method with $\omega = 11$ and $\lambda_0 = 0.3$ and $\lambda_0 = 1$, respectively, are summarized in Table III and Table IV.

		primal dual Newton		sSA-TV method	
σ	ratio	PSNR	MSSIM	PSNR	MSSIM
0.3	0.9	19.11 ($\lambda = 5$)	0.671 ($\lambda = 1$)	**19.97**	**0.855**
0.3	0.8	17.91 ($\lambda = 5$)	0.645 ($\lambda = 1$)	**18.54**	**0.823**
0.3	0.7	17.36 ($\lambda = 5$)	0.643 ($\lambda = 5$)	**18.18**	**0.749**
0.3	0.6	16.16 ($\lambda = 5$)	0.642 ($\lambda = 5$)	**16.89**	**0.733**
0.3	0.5	14.72 ($\lambda = 5$)	0.626 ($\lambda = 5$)	**15.79**	**0.669**
0.1	0.9	26.17 ($\lambda = 20$)	0.963 ($\lambda = 10$)	**27.38**	**0.965**
0.1	0.8	24.82 ($\lambda = 20$)	0.951 ($\lambda = 10$)	**24, 97**	**0.952**
0.1	0.7	**23.18** ($\lambda = 20$)	0.929 ($\lambda = 10$)	22.87	**0.930**
0.1	0.6	**21.31** ($\lambda = 20$)	0.873 ($\lambda = 10$)	20.59	**0.894**
0.1	0.5	18.54 ($\lambda = 20$)	0.787 ($\lambda = 20$)	**18.75**	**0.843**

TABLE III
For the "phantom"-image we test for different standard deviations σ and different ratios of known wavelet coefficients the solution of the total variation minimization with a constant regularization parameter and compare it with the solution of the adaptive total variation approach with $\lambda_0 = 0.3$ and $\omega = 11$ in the sense of the quality measures PSNR and MSSIM.

		primal dual Newton		sSA-TV method	
σ	ratio	PSNR	MSSIM	PSNR	MSSIM
0.3	0.9	19.72 ($\lambda = 5$)	0.464 ($\lambda = 5$)	**20.34**	**0.522**
0.3	0.8	19.57 ($\lambda = 5$)	0.462 ($\lambda = 5$)	**19.95**	**0.501**
0.3	0.7	19.17 ($\lambda = 5$)	0.453 ($\lambda = 5$)	**19.32**	**0.477**
0.3	0.6	**18.80** ($\lambda = 5$)	0.442 ($\lambda = 5$)	**18.80**	**0.453**
0.3	0.5	15.75 ($\lambda = 10$)	0.381 ($\lambda = 5$)	**17.52**	**0.387**
0.1	0.9	23.51 ($\lambda = 20$)	0.719 ($\lambda = \frac{100}{7}$)	**24.03**	**0.756**
0.1	0.8	22.63 ($\lambda = 20$)	0.686 ($\lambda = \frac{100}{7}$)	**22.93**	**0.718**
0.1	0.7	21.62 ($\lambda = 20$)	0.645 ($\lambda = 20$)	**21.64**	**0.667**
0.1	0.6	**20.78** ($\lambda = 20$)	0.610 ($\lambda = 20$)	20.61	**0.614**
0.1	0.5	**19.86** ($\lambda = \frac{100}{7}$)	0.559 ($\lambda = \frac{100}{7}$)	19.61	**0.571**

TABLE IV
For the "barbara"-image we test for different standard deviations σ and different ratios of known wavelet coefficients the solution of the total variation minimization with a constant regularization parameter and compare it with the solution of the adaptive total variation approach with $\lambda_0 = 1$ and $\omega = 11$ in the sense of the quality measures PSNR and MSSIM.

From these results we observe that the sSA-TV method outperforms the primal-dual Newton method with respect to MSSIM. With respect to PSNR, the sSA-TV method also performs better than the primal-dual Newton method in most of the cases and especially when the ratio is large, i.e., when a lot of data is available. For example, when we consider a random loss of 20% of the wavelet coefficients (i.e., ratio $= 0.8$) and $\sigma = 0.05$, we see in Figure 9 that the sSA-TV method preserves image features (see the pattern on the scarf in Figure 9), while removing noise in the uniform parts. This is due to the fact, that the adaptively computed function λ is much higher in regions where details (e.g., see the pattern on the scarf) have to be reconstructed, see Figure 9 (c). The primal-dual Newton method leaves noise in the image while destroying image features, see Figure 9 (d)-(e) and (g)-(h). A similar behaviour is observed for the example in Figure 10 where $\sigma = 0.1$ and ratio$= 0.5$. Here we clearly see, that the primal-dual Newton method with a scalar regularization parameter is not able to preserve the intensities in the image, e.g., the light gray of the squares.

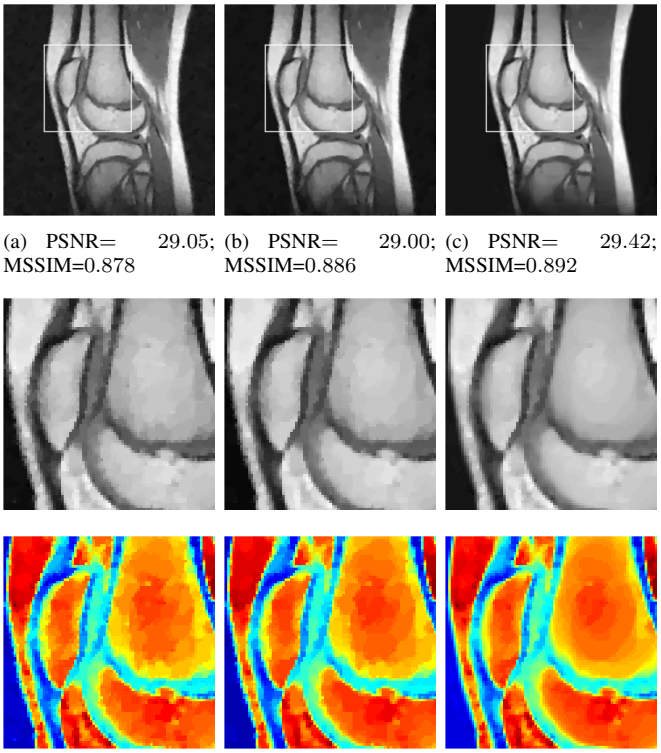
As above for the reconstruction of partial Fourier-data the final adaptive parameter λ obtained by our parameter choice rule is indeed chosen according to the features in the image, see Figure 9(c) and Figure 10(c). We see that in detail regions λ is large in order to preserve details, while in uniform parts it is small to remove noise considerably.

1) Dependence on the initial choice λ_0: Now we investigate our algorithm concerning its stability with respect to λ_0 for the application of wavelet inpainting. We test the sSA-TV method for $\lambda_0 \in \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$ and plot the PSNR- and MSSIM-values obtained for the reconstruction in Figure 11. In these plots we also specify the number of updates on λ for each λ_0. We observe a similar behaviour as in the case of MRI. Although now the reconstructions seem to be much more stable with respect to λ_0.

2) Dependence on the window size ω: As for the reconstruction from partial Fourier-data we test our algorithm again for different values of the window size ω varying from 3 to 23. In Figure 12 and Figure 13 we plot the PSNR- and MSSIM-values of the restoration of the image from Figure 10 and Figure 9 obtained by the sSA-TV method with $\lambda_0 = 0.3$ and $\lambda_0 = 1$, respectively. We observe that the reconstruction is remarkably stable with respect to the window size for all considered random losses of data.

V. Conclusion

In this work it has been shown that spatially adaptive data fidelity weights help to improve the quality of restored images with respect to PSNR and MSSIM. For undersampled data, this, of course, depends on the sampling rate, i.e. in case the sampling operation induces a loss of details in



(a) PSNR= 29.05; MSSIM=0.878 (b) PSNR= 29.00; MSSIM=0.886 (c) PSNR= 29.42; MSSIM=0.892

Fig. 4. *Reconstruction of the "knee"-image sampled in the Fourier domain along 88 radial lines (corresponds to* 48% *of the data) which is deteriorated by Gaussian noise with zero mean and standard deviation* $\sigma = 0.1$. *We show the results of the primal dual TV method with the best PSNR (left column), and the best MSSIM (mid column) and the result of the sSA-TV method (right column). In the second row we zoomed in on the area highlighted in the first row. In order to visualize the differences in the reconstruction we color the* zoomed area *in the third row.*

(a) $\sigma = 0.3$; # rays $= 132$ (b) $\sigma = 0.1$; # rays $= 132$ (c) $\sigma = 0.05$; # rays $= 132$

Fig. 5. *Final values of* λ *by our adaptive parameter choice rule.*

the data, then these details cannot be recovered properly by adjusting fidelity weights only. On the other hand, as long as the sampling allows to detect details in images, adaptive fidelity weights outperform best experimental scalar choices with respect to the aforementioned quality measures. In our context, the automated adjustment of the local weights is based on the Gumbel statistic and provides a rather robust update tool related to localized image residuals and confidence regions. It has the advantage of avoiding over/under-fitting due to properly chosen bounds in confidence tests. Algorithmically, the parameter adjustment scheme can be sped up by employing hierarchical decompositions, which aim at decomposing an image into so-called "atoms" at different scales, where a scale is induced by the magnitude of the fidelity weight. The

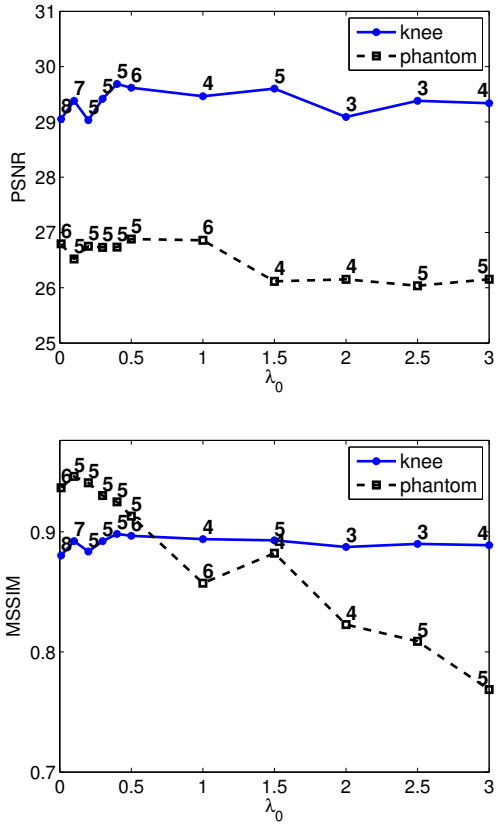


Fig. 6. *PSNR and MSSIM for the "phantom"- and "knee"-image restored from partial Fourier data (sampled along* 88 *radial lines), which is distorted by additive Gaussian noise with zero mean and standard deviation* $\sigma = 0.1$ *by our method for different* λ_0.

framework of the paper is suitable for data transfer operators K, which relate to Parseval frames, or when K encodes a blurring type operation. This is due to the structure of the output space and the properties of the adjoint operator K^*. In the context of Parseval frames, wavelet inpainting or recovery from partial Fourier data represent two specific applications, which can be treated by the framework developed in this paper.

Allowing for more general operators K and, hence, perhaps transferring from spatially adaptive data fidelity weights to distributed regularization weights remains a future challenge.

ACKNOWLEDGMENT

The authors would like to thank the Austrian Science Fund for supporting this research through sub-project P04 "FREELEVEL" of the Collaborative Research Center SFB F32 "Mathematical Optimization and Its Applications in Biomedical Sciences".

REFERENCES

[1] R.A. Adams and J.F. Fournier, *Sobolev spaces*. Pure and Applied Mathematics (Amsterdam), 140, 2nd ed., Elsevier/Academic Press, Amsterdam, 2003.

[2] A. Almansa, C. Ballester, V. Caselles, and G. Haro, *A TV based restoration model with local constraints*. J. Sci. Comput. 34 No. 3 (2008), 209–236.

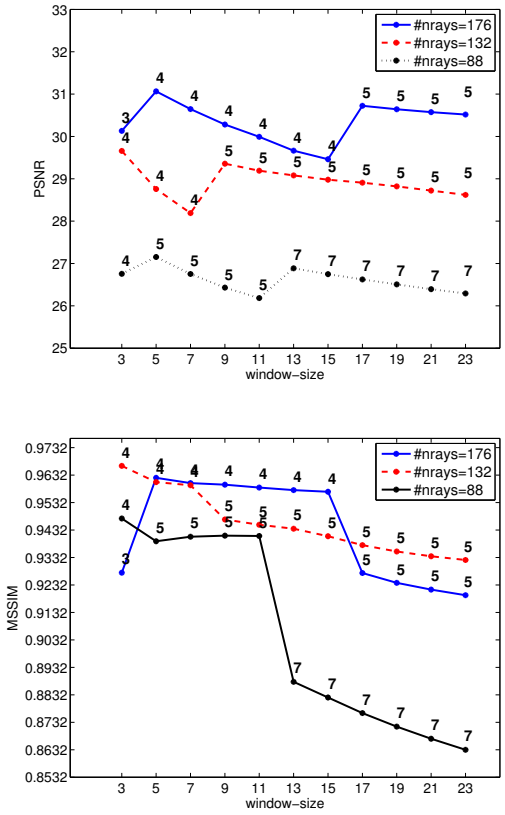


Fig. 7. *Restoration of the phantom-image with $\sigma = 0.1$ via the sSA-TV method for different ω.*

Fig. 8. *Images obtained by full Wavelet-coefficients (a) "phantom" (b) "barbara".*

Fig. 9. *Random loss of 20% of wavelet coefficients with Gaussian noise $\sigma = 0.1$ (a) Original image before information loss (b) Received image back-projected (c) Final values of λ by our adaptive parameter choice rule (d) Best PSNR-reconstruction by the primal-dual Newton algorithm ($\lambda = 20$, PSNR=22.63, MSSIM=0.676) (e) Best MSSIM-reconstruction by the primal-dual Newton algorithm ($\lambda = \frac{100}{7}$, PSNR=22.58, MSSIM=0.686) (f) Reconstruction via sSA-TV method (PSNR=22.93, MSSIM=0.718). In (g) - (i) we zoomed on the scarf.*

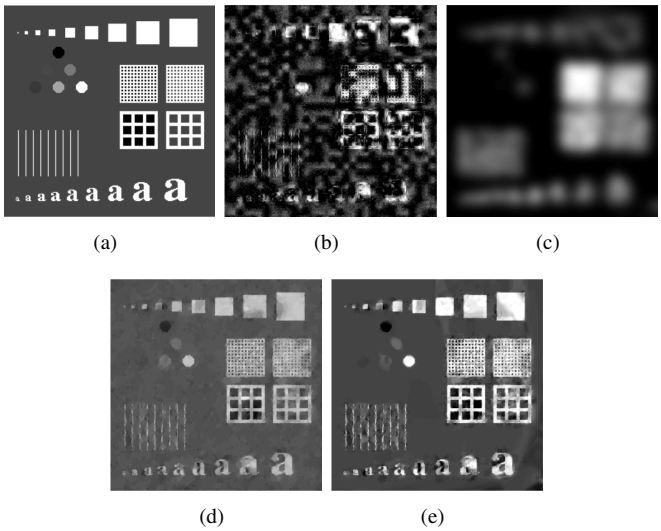


Fig. 10. *Random loss of 50% of wavelet coefficients with Gaussian noise $\sigma = 0.1$ (a) Original image before information loss (b) Received image back-projected (c) Final values of λ by our adaptive parameter choice rule (d) Best reconstruction by the primal-dual Newton algorithm ($\lambda = 20$, PSNR=18.54, MSSIM=0.787) (e) Reconstruction via sSA-TV method (PSNR=18.75, MSSIM=0.843).*

[3] J.J. Benedetto and E.J. King, *Smooth functions associated with wavelet sets on $\mathbb{R}^d$, $d > 1$, and frame bound gaps*, Acta. Appl. Math. 107 (2009), 121–142.

[4] M. Bertalmio, V. Caselles, B. Rougé, and A. Solé, *TV based image restoration with local constraints*. J. Sci. Comput. 19 (2003), 95–122.

[5] A. Bovik, *Handbook of Image and Video Processing*. Academic Press, San Diego, 2000.

[6] T. F. Chan, J. Shen, and H-M Zhou, *Total variation wavelet inpainting*, J. Math. Imaging Vision, 25:107–125, 2006.

[7] A. Chambolle, *An algorithm for total variation minimization and applications*. J. Math. Imaging Vision, 20 (2004), 89–97.

[8] A. Chambolle, and P.-L. Lions, *Image recovery via total variation minimization and related problems*. Numer. Math. 76 (1997), 167–188.

[9] Q. Chang, and I.-L. Chern, *Acceleration methods for total variation based image denoising*. SIAM J. Appl. Math. 25 (2003), 982–994.

[10] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Bosten, 2003.

[11] C.K. Chui and X. Shi, *Bessel sequences and affine frames*, Appl. Comp. Harmonic Anal. 1 (1993), 29–49.

[12] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, 1992.

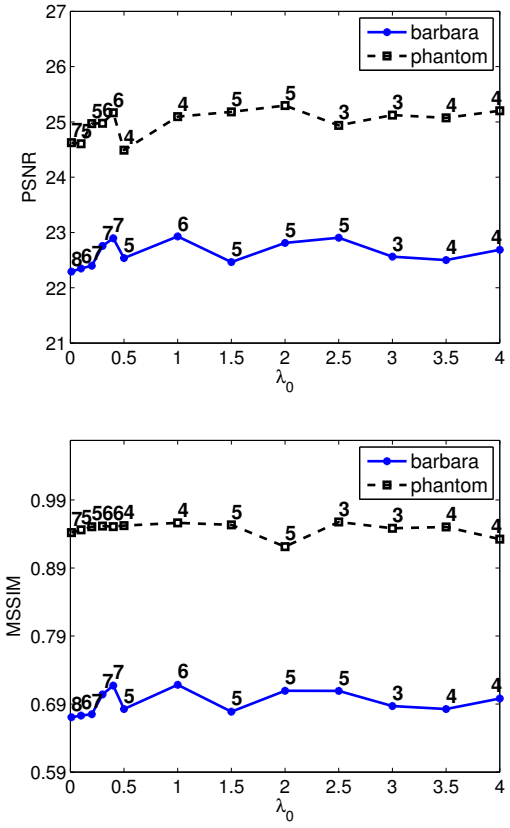


Fig. 11. *PSNR and MSSIM for the "phantom"- and "barbara"-image restored from partial wavelet coefficients (ratio* $= 0.8$*), which is distorted by additive Gaussian noise with zero mean and standard deviation* $\sigma = 0.1$*, by our method for different* λ_0.

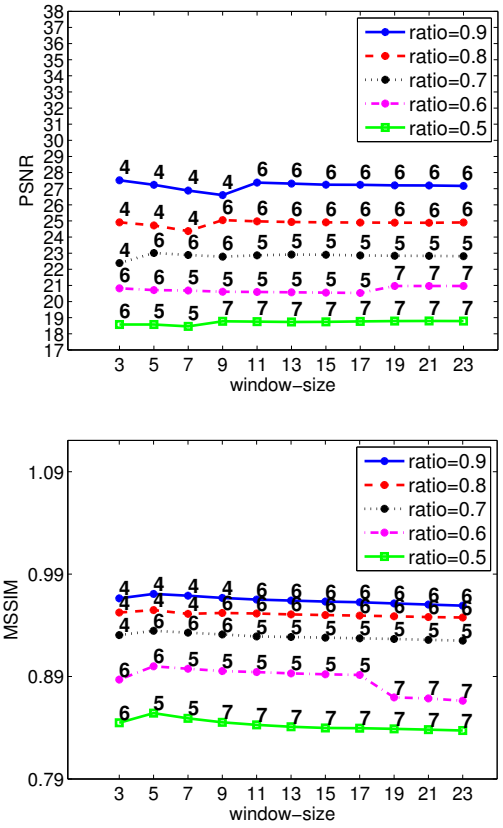


Fig. 12. *Restoration of the image from Figure 10 with* $\sigma = 0.1$ *via the sSA-TV method for different* ω.

[13] I. Daubechies, M. Defrise, Ch. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on Pure and Applied Mathematics, **57**(11), 2004, 1413–1457.

[14] I. Daubechies, B. Han, A. Ron, Z. Shen, *Framelets: MRA-based constructions of wavelet frames*, Applied and Computational Harmonic Analysis, 14 (2003), 1–46.

[15] I. Daubechies, G. Teschke, L. Vese, *Iteratively Solving Linear Inverse Problems Under General Convex Constraints*, Inverse Problems and Imaging, Vol 1, No 1 (2007), 29–46.

[16] D.C. Dobson, and C.R. Vogel, *Convergence of an iterative method for total variation denoising*. SIAM J. Numer. Anal. 34 (1997), 177991.

[17] Y. Dong, M. Hintermüller, and M. Rincon-Camacho, *Automated Regularization Parameter Selection in Multi-Scale Total Variation Models for Image Restoration*, J. Math. Imaging Vis., **40** (2011), 82–104.

[18] R. A. Fisher, "Tests of significance in harmonic analysis," *Proc. Roy. Soc. Ser. A*, vol. 125, pp. 54–59, June 1929.

[19] K. Frick and P. Marnitz, *Statistical multiresolution dantzig estimation in imaging: Fundamental concepts and algorithmic framework*. Electronic Journal of Statistics, 6 (2012), 231–268.

[20] K. Frick, P. Marnitz, and A. Munk, *Statistical multiresolution estimation for variational imaging: with an application in Poisson-biophotonics*. J. Math. Imaging Vision, 46, No. 3 (2013), 370–387.

[21] T. Goldstein and S. Osher, *The split bregman method for* ℓ_1 *regularized problems*, SIAM J. Imaging Sci., 2 (2009), 1311–1333.

[22] M. Hintermüller, and K. Kunisch, *Total bounded variation regularization as bilaterally constrained optimization problem*. SIAM J. Appl. Math. 64 (2004), 1311–1333.

[23] M. Hintermüller and G. Stadler, *An Infeasible Primal-Dual Algorithm for Total Bounded Variation-Based Inf-Convolution-Type Image Restoration*, SIAM J. Sci. Comput., 28 (2006), pp. 1–23.

[24] T. Hotz, P. Marnitz, R. Stichtenoth, L. Davies, Z. Kabluchko, and A. Munk, *Locally adaptive image denoising by a statistical multiresolution criterion*. Comput. Stat. Data Anal., 56, No. 33 (2012), 543–558.

[25] J. Krommweh and G. Plonka, *Directional Haar wavelet frames on triangles*, Appl. Comput. Harmon. Anal. **27**(2), 2009, 215–234.

[26] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul. 4 (2005), 460–489.

[27] M. Peligrad and W. B. Wu, " Central limit theorem for Fourier transforms of stationary processes," *The Annals of Probability*, vol. 38, no. 5, pp. 2009–2022, 2010.

[28] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*. Physica D 60 (1992), 2598.

[29] D. Strong, J.-F. Aujol, and T. Chan, *Scale recognition, regularization parameter selection, and Meyerł G norm in total variation regularization*. Technical report, UCLA, 2005.

[30] D. Strong, and T. Chan, *Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image processing*. Technical report, UCLA, 1996.

[31] D. Strong, and T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*. Inverse Probl. 19 (2003), 165–187.

[32] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical (BV,* L^2*) decompositions*. Multiscale Moel. Simul. 2, 554–579 (2004).

[33] E. Tadmor, S. Nezzar, and L. Vese, *Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation*. Commun. Math. Sci. (2008), 1–26.

[34] C.R. Vogel, *Computational Methods for Inverse Problems. Frontiers Appl. Math.*, vol. 23. SIAM, Philadelphia, 2002.

[35] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, *Image quality assessment: From error to structural similarity*. IEEE Tran. Image Process. **13**, 600–612, 2004.

[36] D. Werner, *Funktionalanalysis*, Springer, Berlin, 2007.

[37] K. Yosida, *Functional Analysis*, Springer, Berlin, 1995.

[38] X. Zhang and T. F. Chan, *Wavelet Inpainting by Nonlocal Total Variation*, Inverse Problems and Imaging, **4**(1), 2010, 191–210.

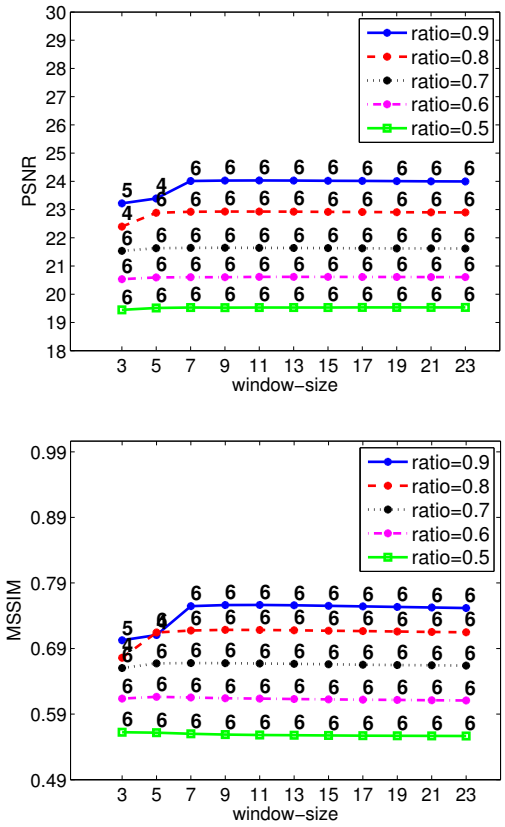


Fig. 13. *Restoration of the image from Figure 9 with* $\sigma = 0.1$ *via the sSA-TV method for different* ω.