Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Adaptive regularization for image reconstruction from subsampled data

Michael Hintermüller[1,2], Andreas Langer[3], Carlos N. Rautenberg[2], Tao Wu[4]

submitted: March 8, 2017

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin, Germany
E-Mail: michael.hintermueller@wias-berlin.de

[2] Department of Mathematics
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin, Germany
E-Mail: hint@math.hu-berlin.de
carlos.rautenberg@math.hu-berlin.de

[3] Universität Stuttgart
Pfaffenwaldring 57
70569 Stuttgart, Germany
E-Mail: andreas.langer@mathematik.uni-stuttgart.de

[4] Technische Universität München
Boltzmannstr. 3
85748 Garching, Germany
E-Mail: tao.wu@tum.de

No. 2379
Berlin 2017

2010 *Mathematics Subject Classification.* 68U10, 94A08, 49K20, 49K40, 49M37, 47A52, 65K10.

Key words and phrases. Image restoration, spatially adaptive regularization, partial Fourier data, wavelet inpainting.

This research was supported by the Austrian Science Fund (FWF) through START-Project Y305 “Interfaces and Free Boundariesänd SFB-Project F3204 “Mathematical Optimization and Applications in Biomedical Sciences", the German Research Foundation DFG through Project HI1466/7-1 “Free Boundary Problems and Level Set Methods”, as well as the Research Center MATHEON through Project C-SE15 “Optimal Network Sensor Placement for Energy Efficiency” supported by the Einstein Center for Mathematics Berlin.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Adaptive regularization for image reconstruction from subsampled data

Michael Hintermüller, Andreas Langer, Carlos N. Rautenberg, Tao Wu

Abstract

Choices of regularization parameters are central to variational methods for image restoration. In this paper, a spatially adaptive (or distributed) regularization scheme is developed based on localized residuals, which properly balances the regularization weight between regions containing image details and homogeneous regions. Surrogate iterative methods are employed to handle given subsampled data in transformed domains, such as Fourier or wavelet data. In this respect, this work extends the spatially variant regularization technique previously established in [15], which depends on the fact that the given data are degraded images only. Numerical experiments for the reconstruction from partial Fourier data and for wavelet inpainting prove the efficiency of the newly proposed approach.

1 Introduction

Image restoration is one of the fundamental tasks in image processing. The quality of the obtained reconstructions depends on several input factors: the quality of the given data, the choice of the regularization term or prior, the proper balance of data fidelity versus filtering, and perhaps several more. The goal of the present paper is to reconstruct an image, defined over the two-dimensional Lipschitz (image) domain Ω, from contaminated data f, defined over the data domain Λ. Given the original image $\widehat{u} : \Omega \to \mathbb{R}$, the data formation model is assumed to be

$$f = K\widehat{u} + \eta, \tag{1}$$

where $K\widehat{u}$ represents possibly subsampled data which results from a linear sampling strategy and η is related to white Gaussian noise (with zero mean). A more precise description of the data formation model is postponed until section 2.

A popular approach to image restoration rests on variational methods, i.e., the characterization of the reconstructed image u as the solution of a minimization problem of the type

$$\min_u \quad \Phi(u; f) + \alpha R(u), \tag{2}$$

where $\Phi(\cdot; f)$ represents a data fidelity term, $R(\cdot)$ an appropriate filter or prior, and $\alpha > 0$ a regularization parameter which balances data fidelity and filtering. The choice of Φ is typically dictated by the type of noise contamination. As long as Gaussian noise is concerned, following the maximum likelihood we choose

$$\Phi(u; f) = \frac{1}{2}\|Ku - f\|^2_{L^2(\Lambda)}.$$

On the other hand, R encodes prior information on the underlying image. For the sake of edge preservation, we choose

$$R(u) = |Du|(\Omega), \tag{3}$$

DOI 10.20347/WIAS.PREPRINT.2379 Berlin 2017

i.e., the total variation of a function u (see equation (5) below for its definition). Then the resulting model (2) becomes the well-known Rudin-Osher-Fatemi (ROF) model [31] which has been studied intensively in the literature; see, e.g., [6, 7, 8, 14, 21, 24, 29, 33, 34] as well as the monograph [38] and many references therein.

It is well known that the proper choice of α is delicate. A general guideline is the following one: Large α favorably removes noise in homogeneous image regions, but it also compromises image details in other regions; Small α, on the other hand, might be advantageous in regions with image details, but it adversely retains noise in homogeneous image regions. For an automated choice of α in (2) several methods have been devised; see for example [10, 18, 20, 32, 40] and the references therein, and see [22, 25] for the spatially distributed α methods. We note that instead of considering (2) one may equivalently study $\lambda\Phi(u; f) + R(u)$ 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 [2] a scalar λ_i, $i = 1, \ldots$ #pieces, for each segment is computed by an augmented-Lagrangian-type algorithm. While still operating in a deterministic regime, [2] interestingly uses a distributed (more precisely a piecewise constant) parameter function λ.

Later it was noticed that stable choices of λ (or respectively α) have to incorporate statistical properties of the noise. In this vein, in [1, 15] automated update rules for λ based on statistics of local constraints were proposed. For statistical multiscale methods we refer to [16, 17, 26]. A different approach has been proposed in [35] for image denoising only, where non-local means [4] has been used to create a non-local data fidelity term. While the methods in [1, 15, 23] are highly competitive in practice, the adjustment of λ requests the output of K to be a deteriorated image which is again defined over Ω. This, however, limits the applicability of these approaches in situations where K involves transformation of an image into a different type of data output space. Particular examples of such transformations include wavelet or Fourier transforms. It is therefore the goal of this paper to study the approach of [15] in the context of reconstructing from such non-image data, possibly coupled with subsampling for the sake of fast data acquisition.

Here we also mention other spatially weighted total variation methods from the existing literatures. Very often these methods, different from [15, 23] (and also the present paper), weight the total variation locally by certain edge indicators. In [9, 42, 43] the difference of the image curvature was used as an edge indicator, while alternatively the (modified) difference of eigenvalues of the image Hessian was considered by [41, 30]. Recently, the authors in [27, 28] used similar edge indicators to weight the total variation anisotropically under the framework of quasi-variational inequalities.

The rest of the paper is organized as follows. Section 2 describes in detail the problem settings and the notations. Our adaptive regularization approach is presented in section 3. Section 4 concludes the paper with numerical experiments on reconstruction of partial Fourier data and wavelet inpainting.

2 Problem Settings and Notations

In the data formation model (1), we shall consider the continuous linear operator K as a composition of two linear operators, i.e., $K = S \circ T$. More precisely, $T : L^2(\Omega) \to L^2(\Lambda)$ is a linear orthogonal transformation which preserves the inner product, i.e., $\langle u, v\rangle_{L^2(\Omega)} = \langle Tu, Tv\rangle_{L^2(\Lambda)}$ for any $u, v \in L^2(\Omega)$. Typical examples of T include Fourier and orthogonal wavelet transforms. Further, we denote the subsampling domain by $\tilde{\Lambda}$, which is assumed to be a (measurable) subset of Λ of finite positive measure, i.e., $0 < |\tilde{\Lambda}| < \infty$. Define $\mathbf{1}_{\tilde{\Lambda}}$ as the characteristic function on $\tilde{\Lambda}$, i.e., $\mathbf{1}_{\tilde{\Lambda}}$ equals 1 on $\tilde{\Lambda}$ and 0 elsewhere. Then the so-called subsampling operator $S : L^2(\Lambda) \to L^2(\Lambda)$ is defined by

$(Sf)(y) = \mathbf{1}_{\tilde{\Lambda}}(y)f(y)$ almost everywhere (a.e.) on Λ. It is worth mentioning that S is an orthogonal projection which satisfies idempotency, i.e., $S^2 = S$, and self-adjointness, i.e., $S^* = S$, and that the range of S, denoted by $\operatorname{Ran} S$, is a closed subspace of $L^2(\Lambda)$. In this setting, we consider the noise η as an arbitrary oscillatory function in $\operatorname{Ran} S$ with

$$\int_{\tilde{\Lambda}} \eta \, dy = 0, \quad \text{and} \int_{\tilde{\Lambda}} |\eta|^2 dy = \sigma^2 |\tilde{\Lambda}|, \tag{4}$$

for some $\sigma > 0$. As a direct consequence, the data f according to (1) also lies in $\operatorname{Ran} S$.

For $u \in L^1(\Omega)$, the total variation term $|Du|(\Omega)$ in (3) is defined as follows:

$$|Du|(\Omega) := \sup \Big\{ \int_\Omega u \ \operatorname{div} \vec{p} \, dx : \vec{p} \in C_0^1(\Omega; \mathbb{R}^2), \ \|\vec{p}\|_{L^\infty(\Omega)} \leq 1 \Big\}. \tag{5}$$

Here, $C_0^1(\Omega; \mathbb{R}^2)$ denotes the set of all $\mathbb{R}^2$-valued continuously differentiable functions on Ω with compact support.

3 Adaptive Regularization Approach

The focus of this paper is to reconstruct a high-quality image from subsampled data in a non-image data domain using an adaptive regularization approach. The present section is structured as follows. In section 3.1, we introduce the surrogate iteration method for solving the ROF-model [31]. Then in section 3.2 we incorporate spatially adaptive regularization into the surrogate iteration. We further accelerate the spatial adaptive algorithm by hierarchical decomposition.

3.1 ROF-Model and Surrogate Iteration

Our variational paradigm is chosen to follow Rudin, Osher and Fatemi [31], which allows to preserve edges in images. Further, due to the properties of the noise term η in (4), the ROF-model restores the image by solving the following constrained optimization problem:

$$\begin{aligned} &\text{minimize } (\min) \ |Du|(\Omega) \quad \text{over } u \\ &\text{subject to (s.t.)} \ \int_{\tilde{\Lambda}} Ku \, dy = \int_{\tilde{\Lambda}} f \, dy, \\ &\qquad\qquad \int_{\tilde{\Lambda}} |Ku - f|^2 dy = \sigma^2 |\tilde{\Lambda}|. \end{aligned} \tag{6}$$

Usually (6) is addressed via the following unconstrained optimization problem:

$$\min_u |Du|(\Omega) + \frac{\lambda}{2} \int_{\tilde{\Lambda}} |Ku - f|^2 \, dy \tag{7}$$

for a given constant $\lambda > 0$. Note that, since $Ku - f \in \operatorname{Ran} S$, the objective in (7) remains unchanged if the integration in the second term of the objective is performed over Λ rather than $\tilde{\Lambda}$. Assuming that K does not annihilate constant functions, one can show that there exists a constant $\lambda \geq 0$ such that the constrained problem (6) is equivalent to the unconstrained problem (7); see [7].

For our purposes we modify the objective in (7) in order to handle the presence of the operator K. Hence, instead of tackling (7) directly we introduce a so-called *surrogate functional* $\mathbb{S}$ [12]. In this vein, for given $a \in L^2(\Omega)$, $\mathbb{S}$ is defined as

$$\begin{aligned}\mathbb{S}(u,a) &:= |Du|(\Omega) + \frac{\lambda}{2}\Big(\|Ku-f\|^2_{L^2(\Lambda)} + \delta\|u-a\|^2_{L^2(\Omega)} - \|K(u-a)\|^2_{L^2(\Lambda)}\Big) \\ &= |Du|(\Omega) + \frac{\lambda\delta}{2}\|u - f_K(a)\|^2_{L^2(\Omega)} + \phi(a,K,f,\lambda),\end{aligned} \tag{8}$$

with

$$f_K(a) := a - \frac{1}{\delta}K^*(Ka-f) \in L^2(\Omega),$$

where we assume $\delta > 1$. Since $\|S^*\| = \|S\| \leq 1$ and $\|T^*\| = \|T\| = 1$, we have $\|K\| \leq 1 < \delta$. We note that here and below $\|\cdot\|$ denotes the operator norm $\|\cdot\|_{\mathcal{L}(L^2(\Omega))}$. We also emphasize that ϕ is a function independent of u. It is readily observed that minimization of $\mathbb{S}(u,a)$ over u is no longer affected by the action of K. Rather, minimizing $\mathbb{S}(u,a)$ for fixed a resembles a typical image denoising problem. In order to approach a solution of (7), we consider the following iteration.

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

$$u^{(k+1)} := \arg\min_u |Du|(\Omega) + \frac{\delta}{2}\int_\Omega \lambda|u - f_K^{(k)}|^2 dx. \tag{9}$$

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

It can be shown that the iteration (9) generates a sequence $(u^{(k)})_{k\in\mathbb{N}}$ which converges to a minimizer of (7); see [12, 13]. Moreover, the minimization problem in (9) is strictly convex and can be efficiently solved by standard algorithms such as the primal-dual first-order algorithm [6], the split Bregman method [19], or the primal-dual semismooth Newton algorithm [24].

3.2 Hierarchical Spatially Adaptive Algorithm

The problem in (9) is related to the globally constrained minimization problem

$$\min_u |Du|(\Omega) \quad \text{s.t.} \int_\Omega |u - f_K^{(k)}|^2 dx \leq A, \tag{10}$$

where $A > 0$ is a constant depending on σ and K; see [7]. In order to enhance image details while preserving homogeneous regions, we localize the constraint in (10), which leads to the modified variational model:

$$\min_u |Du|(\Omega) \quad \text{s.t. } \mathcal{S}(u) \leq A \quad \text{a.e. in } \Omega. \tag{11}$$

Here the local variance term $\mathcal{S}(u)(\cdot) := \int_\Omega w(\cdot,x)|u - f_K(u)|^2(x)dx$ is defined for some given localization filter w. Thus the constraint in (11) with $u = u^{(k+1)}$ reads

$$\mathcal{S}(u^{(k+1)})(\cdot) = \int_\Omega w(\cdot,x)\big|u^{(k+1)} - u^{(k)} + \frac{1}{\delta}K^*(Ku^{(k)} - f)\big|^2(x)dx \leq A. \tag{12}$$

Given the convergence result, as $k \to \infty$, for scalar λ alluded to in connection with (9), one expects the term $u^{(k+1)} - u^{(k)}$ to vanish. This indicates that $\int_\Omega w(\cdot,x)|\frac{1}{\delta}K^*(Ku^{(k)} - f)|^2(x)dx \leq A$ is

expected in the limit. This consideration leads to the following pointwisely constrained optimization problem:

$$\min_u |Du|(\Omega) \quad \text{s.t.} \int_\Omega w(\cdot,x)\Big|\frac{1}{\delta}K^*(Ku-f)\Big|^2(x)dx \le A \quad \text{a.e. in } \Omega. \tag{13}$$

Next we discuss the choice of A. In view of the (global) estimate for the *backprojected residual* $K^*(K\widehat{u}-f)$, i.e.,

$$\|K^*(K\widehat{u}-f)\|^2_{L^2(\Omega)} \le \|K^*\|^2\|K\widehat{u}-f\|^2_{L^2(\Lambda)} \le \sigma^2|\tilde{\Lambda}|,$$

we thus choose

$$A := \frac{\sigma^2|\tilde{\Lambda}|}{\delta^2}.$$

In deriving the above inequalities, we have used the facts that $\|K^*\| = \|K\| \le 1$ and $\|K\widehat{u}-f\|^2_{L^2(\Lambda)} = \sigma^2|\tilde{\Lambda}|$.

In a discrete setting, we now describe a strategy, based on a statistical local variance estimator, to adapt the spatially variant regularization parameter λ. For this purpose, consider a discrete image u defined over the discrete 2D index set Ω_h (of cardinality $|\Omega_h|$), whose nodes lie on a regular grid of uniform mesh size $h := \sqrt{1/|\Omega_h|}$. The total variation of a discrete image u is denoted by $|Du|(\Omega_h)$; see (15) below for a precise definition. We also define the residual image associated with $f_K(\cdot)$ by

$$r(u) := f_K(u) - u.$$

Concerning the filter w associated with $\mathcal{S}$ in (11), we exemplarily choose the mean filter pertinent to a square window centered at x. For this reason and in our discrete setting, we define the averaging window

$$\Omega^\omega_{i,j} := \left\{(i+hs, j+ht) : s,t \in \left[-\frac{\omega-1}{2}, \frac{\omega-1}{2}\right] \cap \mathbb{Z}\right\},$$

where $\omega > 1$ is an odd integer representing the window size, and then compute the estimated local variance at $(i,j) \in \Omega_h$ by

$$\mathcal{S}^\omega(u)_{i,j} := \frac{1}{\omega^2}\sum_{(\tilde{i},\tilde{j})\in\Omega^\omega_{i,j}} \left|r(u)_{\tilde{i},\tilde{j}}\right|^2.$$

Given the reconstruction u_n associated with λ_n, we use $\mathcal{S}^\omega(u_n)$ to check whether λ_n should be updated or it already yields a successful reconstruction u_n. In particular, motivated by [15], we intend to increase λ_n at the pixels where the corresponding local variance violates the upper estimate A. More specifically, we utilize the following update rule:

$$(\lambda_{n+1})_{i,j} = \frac{\zeta_n}{\omega^2}\sum_{(\tilde{i},\tilde{j})\in\Omega^\omega_{i,j}} \min\left\{\bar{\lambda}, \left((\lambda_n)_{\tilde{i},\tilde{j}} + \rho_n\|\lambda_n\|_{\ell^\infty}\left(\sqrt{\tilde{\mathcal{S}}^\omega(u_n)_{\tilde{i},\tilde{j}}/A}-1\right)\right)\right\}. \tag{14}$$

Here

$$\tilde{\mathcal{S}}^\omega(u)_{i,j} := \begin{cases}\mathcal{S}^\omega(u)_{i,j}, & \text{if } \mathcal{S}^\omega(u)_{i,j} > A,\\ A, & \text{otherwise,}\end{cases}$$

$\bar{\lambda} > 0$ is a prescribed upper bound, and $\|\lambda_n\|_{\ell^\infty}$ is a scaling factor suggested in [15]. Two step-size parameters, $\zeta_n > 1$ and $\rho_n > 0$, will allow a backtracking procedure should λ_{n+1} be overshot by (14), on which we refer to the HSA algorithm below for a more detailed account.

We are now ready to present our (basic) spatially adaptive (SA) image reconstruction algorithm.

SA Algorithm: Initialize $u_0 \in \mathbb{R}^{\Omega_h}$, $\lambda_1 \in \mathbb{R}^{\Omega_h}_+$, $n := 1$. Iterate as follows until a stopping criterion is satisfied:

1) Set $u_n^{(0)} := u_{n-1}$. For each $k = 0, 1, 2, ...$, compute $u_n^{(k+1)}$ according to

$$u_n^{(k+1)} := \arg\min_u |Du|(\Omega_h) + \frac{\delta h^2}{2} \sum_{(i,j)\in\Omega_h} (\lambda_n)_{i,j} \left|(u - f_n^{(k)})_{i,j}\right|^2,$$

with $f_n^{(k)} := u_n^{(k)} - \frac{1}{\delta} K^*(Ku_n^{(k)} - f)$. Let u_n be the outcome of this iteration.

2) Update λ_{n+1} according to (14). Set $n := n + 1$.

While the SA algorithm functions well in its own right, following [15] we further accelerate this algorithm by employing a hierarchical decomposition of the image into scales. This idea, introduced by Tadmor, Nezzar and Vese in [36, 37], utilizes concepts from interpolation theory to represent a noisy image as the sum of "atoms" $u_{(l)}$, where every $u_{(l)}$ extracts features at a scale finer than the one of the previous $u_{(l-1)}$. This method acts like an iterative regularization scheme, i.e., up to some iteration number $\bar{l}$ the method yields improvement on reconstruction results with a deterioration (due to noise influence and ill-conditioning) beyond $\bar{l}$.

Here we illustrate the basic workflow of hierarchical decomposition in a denoising problem (i.e., where K equals the identity). Given the exponential scales $\{\zeta^l \lambda_0 : l = 0, 1, 2, ...\}$ with $\lambda_0 \in \mathbb{R}^{\Omega_h}_+$ and $\zeta > 1$, the hierarchical decomposition operates as follows:

1 Initialize $u_0 \in \mathbb{R}^{\Omega_h}$ by

$$u_0 := \arg\min_u |Du|(\Omega_h) + \frac{h^2}{2} \sum_{(i,j)\in\Omega_h} (\lambda_0)_{i,j} |(u - f)_{i,j}|^2.$$

2 For $l = 0, 1, ...$, set $\lambda_{l+1} := \zeta\lambda_l$ and $v_l := f - u_l$. Then compute

$$d_l := \arg\min_u |Du|(\Omega_h) + \frac{h^2}{2} \sum_{(i,j)\in\Omega_h} (\lambda_{l+1})_{i,j} |(u - v_l)_{i,j}|^2,$$

and update $u_{l+1} := u_l + d_l$.

Now we incorporate such a hierarchical decomposition into the SA algorithm, which we shall refer to as the hierarchical spatially adaptive (HSA) algorithm. We note that all minimization (sub)problems in the HSA algorithm are solved by the primal-dual Newton method in [24]. There, the original ROF-model is approximated by a variational problem posed in $H_0^1(\Omega)$ via adding an additional regularization term $\frac{\mu}{2}\|\nabla u\|^2_{L^2(\Omega)}$, with $0 < \mu \ll 1/(\operatorname{ess\,sup} \lambda)$, to the objective and assuming, without loss of generality, homogeneous Dirichlet boundary conditions. In this case, the (discrete) total variation is given by

$$|Du|(\Omega_h) = h \sum_{(i,j)\in\Omega_h} \Big(|u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|\Big), \tag{15}$$

with $u_{i,j} = 0$ whenever $(i, j) \notin \Omega_h$. We refer to [24] for a detailed account of this algorithm.

We also remark that the initial $\lambda_1 \in \mathbb{R}^{\Omega_h}_+$ should be sufficiently small such that the resulting normalized data-fitting error θ_1 is much larger than 1. Then the HSA iterations are responsible for (monotonically)

lifting up λ_n in a spatially adaptive fashion as described earlier in this paper. Such a lifting is performed until the data-fitting error $\|Ku_n - f\|_{\ell^2}^2/|\tilde{\Lambda}_h|$ approaches the underlying noise level σ^2. If the data-fitting error drops too far below σ^2, then the algorithm may suffer from overfitting the noisy data. In this scenario, we backtrack on λ_n through potential reduction of ζ_n and ρ_n; see step 3 of the HSA algorithm.

HSA Algorithm: Input parameters $\delta > 1$, $\omega \in 2\mathbb{N}+1$. Initialize $u_0 \in \mathbb{R}^{\Omega_h}$, $\lambda_1 \in \mathbb{R}_+^{\Omega_h}$ (sufficiently small), $\zeta_0 > 1$, $\rho_0 > 0$.

1) Set $u_0^{(0)} := u_0$. For each $k = 0, 1, 2, ..., \kappa_0$, compute $u_0^{(k+1)}$ by

$$u_0^{(k+1)} := \arg\min_u |Du|(\Omega_h) + \frac{\delta h^2}{2} \sum_{(i,j)\in\Omega_h} (\lambda_1)_{i,j} \left|(u - f_0^{(k)})_{i,j}\right|^2,$$

with $f_0^{(k)} := u_0^{(k)} - \frac{1}{\delta}K^*(Ku_0^{(k)} - f)$. Let u_1 be the outcome of this iteration, and set $n := 1$.

2) Set $v_n := f - Ku_{n-1}$ and $d_n^{(0)} := 0$. For each $k = 0, 1, 2, ..., \kappa_n$, compute $d_n^{(k+1)}$ by

$$d_n^{(k+1)} := \arg\min_u |Du|(\Omega_h) + \frac{\delta h^2}{2} \sum_{(i,j)\in\Omega} (\lambda_n)_{i,j} \left|(u - f_n^{(k)})_{i,j}\right|^2,$$

with $f_n^{(k)} := d_n^{(k)} - \frac{1}{\delta}K^*(Kd_n^{(k)} - v_n)$. Let d_n be the outcome of this iteration, and update $u_n := u_{n-1} + d_n$.

3) Evaluate the (normalized) data-fitting error

$$\theta_n := \frac{\|Ku_n - f\|_{\ell^2}^2}{\sigma^2|\tilde{\Lambda}_h|}.$$

If $\theta_n > 1$, then set $\tilde{n} := n$, $\zeta_n := \zeta_{n-1}$, $\rho_n := \rho_{n-1}$, and continue with step 4;
If $0.8 \le \theta_n \le 1$, then return u_n, λ_n and stop;
If $\theta_n < 0.8$, then set $u_n := u_{\tilde{n}}$, $\lambda_n := \lambda_{\tilde{n}}$, $\zeta_n := \sqrt{\zeta_{n-1}}$, $\rho_n := \rho_{n-1}/2$, and continue with step 4.

4) Update λ_{n+1} according to formula (14). Set $n := n + 1$ and return to step 2.

4 Numerical Experiments

In this section, we present numerical results of the newly proposed HSA algorithm for two applications, namely reconstruction from partial Fourier data and wavelet inpainting. All experiments reported here were performed under Matlab. The image intensity is scaled to the interval $[0, 1]$ in advance of our computation. For the HSA algorithm, we always choose the following parameters: $\delta = 1.2$, $\omega = 11$, $\zeta_0 = 2$, $\rho_0 = 1$, $\bar{\lambda} = 10^6$, $u_0 = K^*f$. In the primal-dual Newton algorithm [24], we choose the H^1-regularization parameter $\mu = 10^{-4}$, the Huber smoothing parameter $\gamma = 10^{-3}$, and terminate the overall Newton iterations as soon as the initial residual norm is reduced by a factor of 10^{-4}. Besides, the maximum iteration numbers $\{\kappa_n\}$ for the surrogate iterations are adaptively chosen such that

$\|d_n^{(\kappa_n)} - d_n^{(\kappa_{n-1})}\|_{\ell^2} \leq 10^{-6}\sqrt{|\Omega_h|}$.

The images restored by HSA are compared, both visually and quantitatively, with the ones restored by the variational model in (7) with *scalar-valued* λ. For quantitative comparisons among restorations, we evaluate their peak signal-to-noise ratios (PSNR) [3] and also the structural similarity measures (SSIM) [39]; see Table 1. To optimize our choice for each scalar-valued λ, we adopt a bisection procedure, up to a relative error of 0.02, to maximize the following weighted sum of the PSNR- and SSIM-values of the resulting scalar-λ restoration

$$\frac{\mathsf{PSNR}(\lambda)}{\max\{\mathsf{PSNR}(\tilde{\lambda}) : \tilde{\lambda} \in I\}} + \frac{\mathsf{SSIM}(\lambda)}{\max\{\mathsf{SSIM}(\tilde{\lambda}) : \tilde{\lambda} \in I\}}$$

over the interval $I = [10^2, 10^5]$. The maximal PSNR and SSIM in the above formula are pre-computed up to a relative error of 0.001.

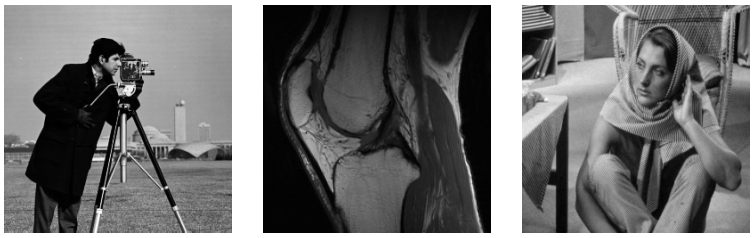

Figure 1: Test images (from left to right): “Cameraman", “Knee", and “Barbara".

4.1 Reconstruction of Partial Fourier Data

In magnetic resonance imaging, one aims to reconstruct an image which is only sampled by partial Fourier data and additionally distorted by additive white Gaussian noise of zero mean and standard deviation σ. Here the data-formation operator is given by $K = S \circ T$, where T is a 2D (discrete) Fourier transform and S represents a downsampling of Fourier data. In particular, we consider S which picks Fourier data along radial lines centered at zero frequency.

Our experiments are performed for the test images “Cameraman” and “Knee” with $\sigma \in \{0.05, 0.1\}$ and $\#\text{radials} \in \{75, 90, 105\}$ respectively. In these experiments, we have always initialized HSA with $\lambda_1 = 100$. The resulting restorations via the total-variation method with scalar-valued λ and via our HSA method are both displayed in Figures 2 and 3. We also show the ultimate spatially adapted λ from HSA in each test run, where the light regions in the λ-plot correspond to high values of λ and vice versa. It is observed that the values of λ in regions containing detailed features (e.g. the camera and the tripod in “Cameraman") typically outweigh its values in more homogeneous regions (e.g. the background sky in “Cameraman"). As a consequence, this favorably yields a sharper background-versus-detail contrast in the restored images via HSA. According to the quantitative comparisons reported in Table 1, HSA almost always outperforms scale-valued λ in terms of PSNR and SSIM. As a side remark, it is also observed that the spatially adapted λ via HSA is able to capture more features of the underlying image at a lower noise level.

To test the robustness of HSA, we perturb our choices of the window size ω and the initial choice of λ in our experiments. In Figure 4, we report the resulting PSNRs and SSIMs of such sensitivity tests on the particular Fourier-Cameraman example with $\sigma = 0.05$ and $\#\text{radials} = 90$. It is observed that HSA behaves relatively stable with different choices of ω. On the other hand, one should be cautioned

that the results of HSA deteriorate as the initial λ is chosen too large. Nevertheless, among all initial λ's smaller than a certain threshold (in this case 200), smaller choices do not always claim advantages over larger ones.

4.2 Wavelet Inpainting

Wavelet inpainting is about restoring missing wavelet coefficients due to lossy compression or error-prone data transmission; see, e.g., [5, 44]. Here we consider the scenario where a test image is compressed by storing the largest Daubechies-4 wavelet coefficients [11] in magnitude only up to a small sampling rate (s.r.), namely s.r. $\in \{2.5\%, 5\%, 10\%\}$. The compressed wavelet coefficients are further contaminated by additive white Gaussian noise of mean zero and standard deviation $\sigma \in \{0.05, 0.1\}$. For wavelet inpainting, we have initialized HSA with $\lambda_1 = 10$. The experiments are performed for the test images "Cameraman" and "Barbara", and the corresponding results, both restored images and the adapted λ's, are shown in Figures 5 and 6.

In the wavelet-Cameraman example, the results via scalar-valued λ's and HSA are almost identical to human eyes. Even though, HSA always outperforms the scale-valued λ in terms of PSNR, while the SSIM-comparison is somewhat even; see Table 1. Interestingly, the adapted λ's in this example exhibit patterns analogous to the ones in the Fourier-Cameraman example.

Our HSA method gains more advantages when it is applied to the "Barbara" image with a stronger cartoon-texture contrast than "Cameraman". In Figure 6, it is witnessed that the restored images via scalar-valued λ's suffer from undesirable staircase effects. In comparison, spatially adapted λ's yield significant improvements on the restorations, even in the cases where the pattern of λ is less transparent due to lack of data or strong noise. In Table 1, the PSNR- and SSIM-comparisons also dominantly favor the HSA method.

Fourier		Cameraman				Knee			
		scalar-valued λ		HSA		scalar-valued λ		HSA	
σ	#rad'l	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
0.05	75	26.7895	0.8051	**26.9559**	**0.8124**	30.4347	0.8247	**30.5399**	**0.8290**
0.05	90	**27.5399**	0.8215	27.5020	**0.8262**	30.8355	0.8337	**30.9442**	**0.8389**
0.05	105	28.1553	0.8307	**28.1667**	**0.8346**	31.1155	0.8402	**31.3328**	**0.8478**
0.1	75	24.9336	0.7576	**25.1809**	**0.7639**	28.2375	0.7570	**28.4896**	**0.7639**
0.1	90	25.2738	0.7666	**25.7072**	**0.7775**	28.4811	0.7627	**28.7140**	**0.7721**
0.1	105	25.6780	0.7740	**26.2317**	**0.7843**	28.5856	0.7662	**28.8373**	**0.7745**
Wavelet		Cameraman				Barbara			
		scalar-valued λ		HSA		scalar-valued λ		HSA	
σ	s.r.	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
0.05	2.5%	24.0319	0.7388	**24.5702**	**0.7436**	22.9489	0.6174	**24.4184**	**0.6777**
0.05	5%	26.5279	**0.7969**	**27.1539**	0.7964	24.6622	0.6922	**26.1698**	**0.7438**
0.05	10%	28.6248	**0.8374**	**29.5812**	0.8351	26.5317	0.7645	**27.8187**	**0.8083**
0.1	2.5%	23.7416	0.7301	**24.1510**	**0.7326**	22.7299	0.6067	**24.0291**	**0.6605**
0.1	5%	25.7625	0.7786	**26.5195**	**0.7791**	24.1307	0.6733	**25.2635**	**0.7095**
0.1	10%	27.3033	**0.8136**	**27.5671**	0.7937	25.4469	0.7410	**26.3245**	**0.7591**

Table 1: Comparisons with respect to PSNR and SSIM.

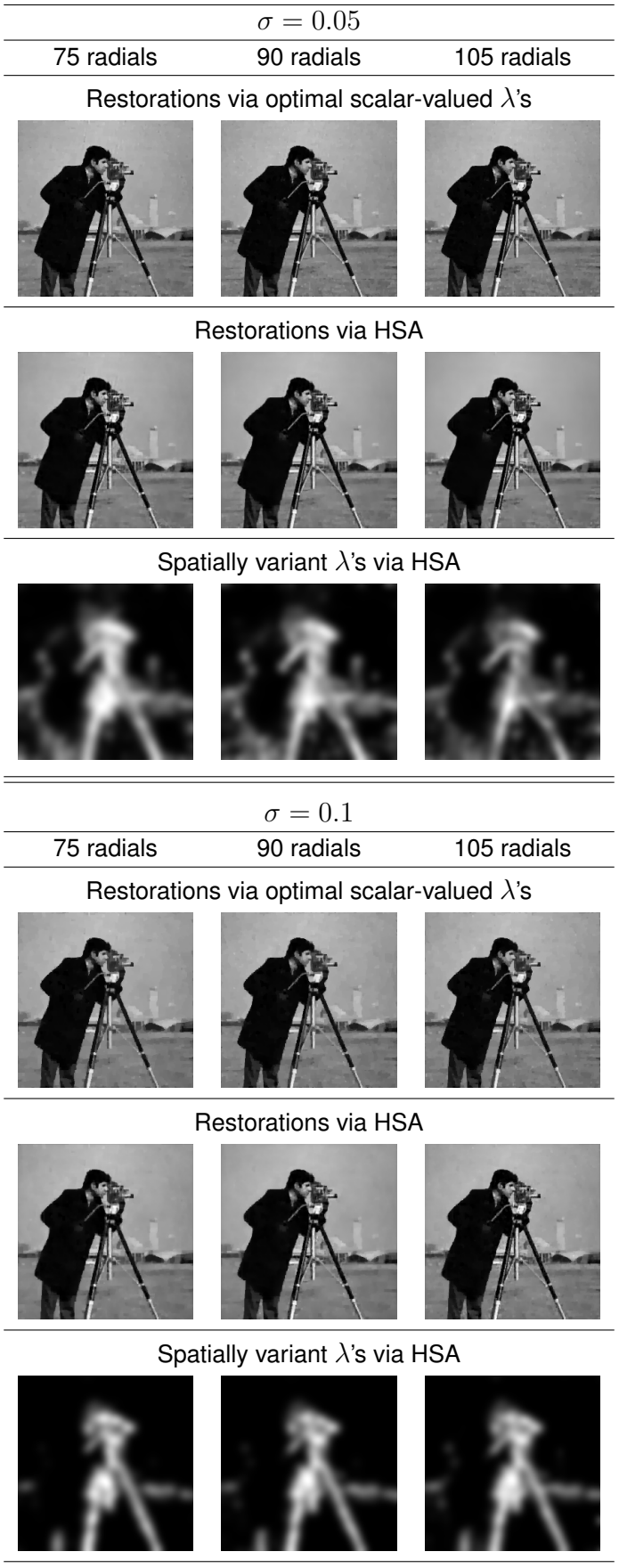

Figure 2: Reconstruction of partial Fourier data on “Cameraman".

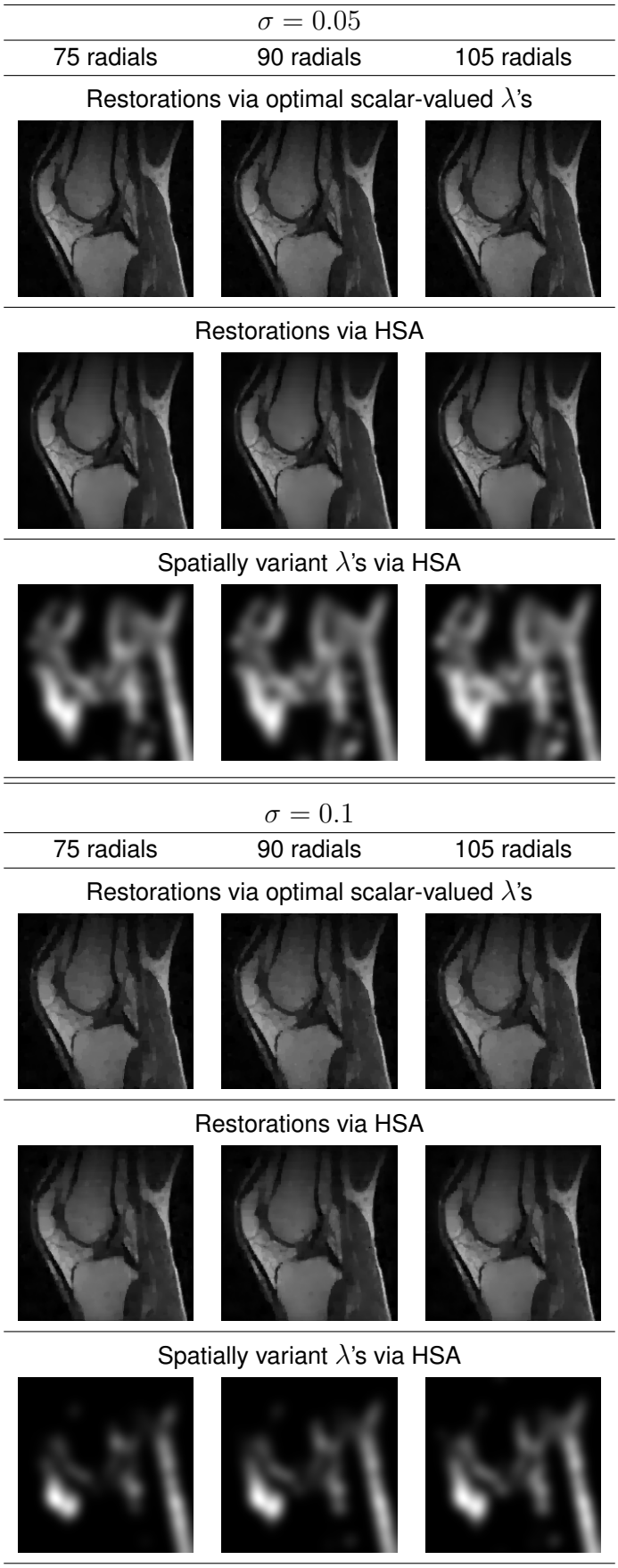

Figure 3: Reconstruction of partial Fourier data on “Knee".

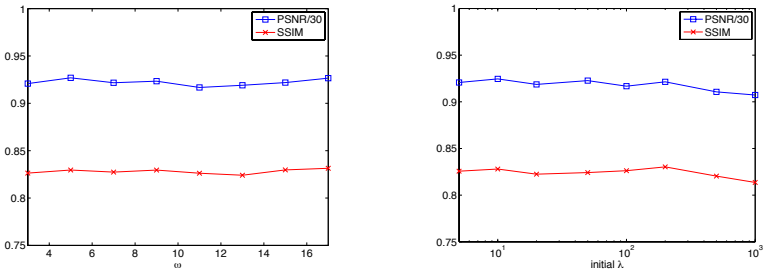

Figure 4: Sensitivity test: image = "Cameraman", $\sigma = 0.05$, #radials = 90.

5 Conclusion

In this work, it has been shown that spatially adapted data fidelity weights help to improve the quality of restored images. The automated adjustment of the local weights is developed based on the localized image residuals. Such a parameter adjustment scheme can be further accelerated by employing hierarchical decompositions, which aim at decomposing an image into so-called atoms at different scales. The framework of the paper is suitable for subsampled data in non-image domain, in particular incomplete coefficients from orthogonal Fourier- and wavelet transforms as illustrated in the numerical experiments.

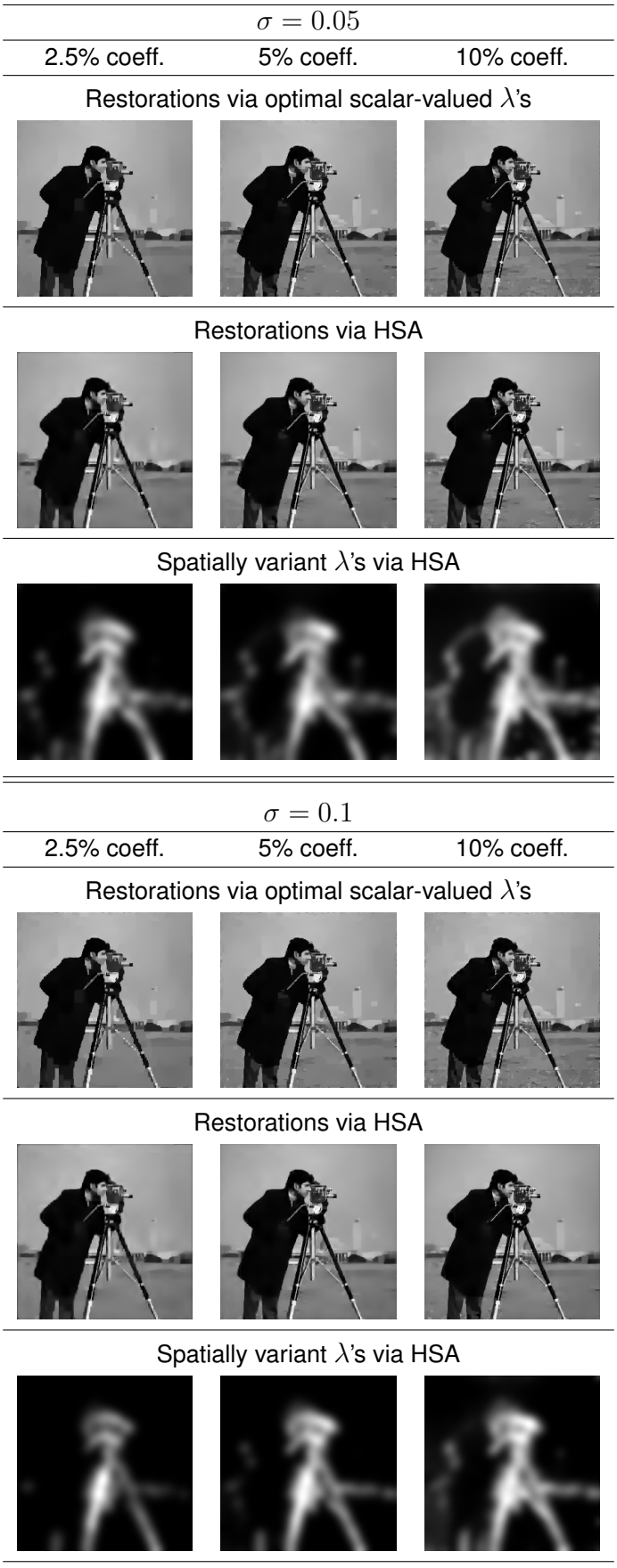

Figure 5: Wavelet inpainting on “Cameraman".

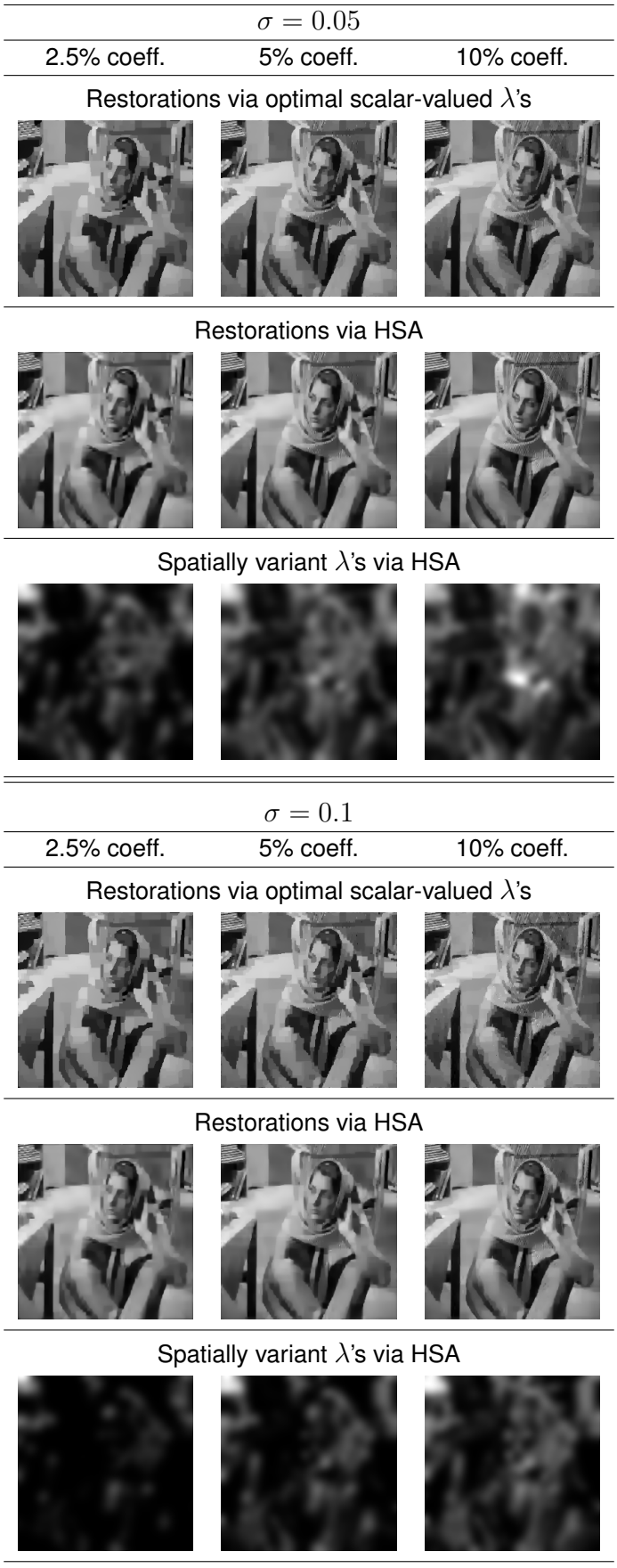

Figure 6: Wavelet inpainting on “Barbara".

References

[1] A. Almansa, C. Ballester, V. Caselles, and G. Haro: A TV based restoration model with local constraints, J. Sci. Comput., **34**, 209–236, (2008).

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

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

[4] A. Buades, B. Coll, and J.-M. Morel: A review of image denoising algorithms, with a new one, Multiscale Modeling and Simulation, **4**, 490–530, (2005).

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

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

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

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

[9] Q. Chen, P. Montesinos, Q.S. Sun, P.A. Heng, and D.S. Xia: Adaptive total variation denoising based on difference curvature, Image Vis. Comput. **28**, 298–306, (2010).

[10] K. Chen, E. Loli Piccolomini, and F. Zama: An automatic regularization parameter selection algorithm in the total variation model for image deblurring, Numer. Algor., **67**, 73–92, (2014).

[11] I. Daubechies, Ten lectures on wavelets, SIAM, Philadelphia (1992).

[12] I. Daubechies, M. Defrise, Ch. De Mol: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure Appl. Math., **57**, 1413–1457, (2004).

[13] I. Daubechies, G. Teschke, L. Vese: Iteratively Solving Linear Inverse Problems Under General Convex Constraints, Inverse Problems and Imaging, **1**, 29–46, (2007).

[14] D.C. Dobson, and C.R. Vogel: Convergence of an iterative method for total variation denoising, SIAM J. Numer. Anal., **34**, 1779–1791, (1997).

[15] 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**, 82–104, (2011).

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

[17] K. Frick, P. Marnitz, and A. Munk: Statistical multiresolution estimation for variational imaging: with an application in Poisson-biophotonics, J. Math. Imaging Vision, **46**, 370–387, (2013).

[18] R. Giryes, M. Elad, and Y. C. Eldar: The projected GSURE for automatic parameter tuning in iterative shrinkage methods, Appl. Comput. Harmon. Anal., **30**, 407–422, (2011).

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

[20] Ch. He, Ch. Hu, W. Zhang, and B. Shi: A fast adaptive parameter estimation for total variation image restoration, IEEE Transaction on Image Processing, **23**, 4954–4967, (2014).

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

[22] M. Hintermüller and C. N. Rautenberg: Optimal selection of the regularization function in a generalized total variation model. Part I: Modelling and theory, WIAS Preprint No. 2235, (2016) .

[23] M. Hintermüller and M. M. Rincon-Camacho: Expected absolute value estimators for a spatially adapted regularization parameter choice rule in L^1-TV-based image restoration, Inverse Problems, **26**, 085005, (2010).

[24] 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**, 1–23, (2006).

[25] M. Hintermüller, C. N. Rautenberg, T. Wu, and A. Langer: Optimal selection of the regularization function in a generalized total variation model. Part II: Algorithm, its analysis and numerical tests, WIAS Preprint No. 2236, (2016).

[26] 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, 543–558, (2012).

[27] F. Lenzen, F. Becker, J. Lellmann, S. Petra, and C. Schnörr: A class of quasi-variational inequalities for adaptive image denoising and decomposition, Comput. Optim. Appl. **54**, 371–398, (2013).

[28] F. Lenzen, J. Lellmann, F. Becker, and C. Schnörr: Solving quasi-variational inequalities for image restoration with adaptive constraint sets, SIAM J. Imaging Sci. **7**, 2139–2174, (2014).

[29] 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**, 460–489, (2005).

[30] Y. Ruan, H. Fang, and Q. Chen: Semiblind image deconvolution with spatially adaptive total variation regularization, Mathematical Problems in Engineering, 606170 (8 pages), (2014).

[31] L.I. Rudin, S. Osher, and E. Fatemi: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268, (1992).

[32] D. Strong, J.-F. Aujol, and T. Chan: Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization. Technical report, UCLA (2005).

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

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

[35] C. Sutour, Ch.-A. Deledalle, and J.-F. Aujol: Adaptive regularization of the NL-means: application to image and video denoising, IEEE Trans. Image Process. **23**, 3506–3521, (2014).

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

[37] E. Tadmor, S. Nezzar, and L. Vese: Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation. Commun. Math. Sci. **6**, 1–26, (2008).

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

[39] 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.

[40] Y.-W. Wen and R. H. Chan: Parameter selection for total -variation-based image restoration using discrepancy principle, IEEE Tran. Image Process. **21**, 1770–1781, (2012).

[41] L. Yan, H. Fang, and S. Zhong: Blind image deconvolution with spatially adaptive total variation regularization, Opt. Lett. **37**, 2778–2780, (2012).

[42] Q. Yuan, L. Zhang, and H. Shen: Multiframe super-resolution employing a spatially weighted total variation model, IEEE Trans. Circuits Syst. Video Technol. **22**, 379–392, (2012).

[43] Q. Yuan, L. Zhang, and H. Shen: Regional spatially adaptive total variation super-resolution with spatial information filtering and clustering, IEEE Trans. Image Process. **22**, 2327–2342, (2013).

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