A Statistical Multiresolution Strategy for Image Reconstruction

Klaus Frick and Philipp Marnitz

Institute for Mathematical Stochastics
University of Göttingen
Goldschmidtstrasse 7, 37077 Göttingen
{frick,marnitz}@math.uni-goettingen.de
http://www.stochastik.math.uni-goettingen.de

Abstract. In this paper we present a fully data-driven and locally-adaptive method for image reconstruction that is based on the concept of *statistical multiresolution estimation* as introduced in [1]. It constitutes a statistical regularization technique that uses a ℓ_∞-type distance measure as data fidelity combined with a convex cost functional. The resulting convex optimization problem is approached by a combination of an inexact augmented Lagrangian method and Dykstra's projection algorithm.

Keywords: statistical multiresolution, extreme-value statistics, total-variation regularization, statistical inverse problems, statistical imaging.

1 Introduction

In this paper we are concerned with the reconstruction of an unknown gray-valued image $u^0 \in \mathrm{L}^2(\Omega)$ with $\Omega = [0,1]^2$ given the data

$$Y_{ij} = (Ku^0)_{ij} + \varepsilon_{ij}, \quad 1 \le i \le m, 1 \le j \le n. \tag{1}$$

We assume that ε_{ij} are independent and identically distributed Gaussian random variables with $\mathbf{E}(\varepsilon_{11}) = 0$ and $\mathbf{E}(\varepsilon_{11}^2) = \sigma^2 > 0$ and that $K : \mathrm{L}^2(\Omega) \to \mathbb{R}^{m\times n}$ is a linear and bounded operator. K is assumed to model image acquisition and sampling at the same time, i.e. $(Ku)_{ij}$ is assumed to be a sample at the pixel $(i/m, j/n)$ of a smoothed version of u.

Numerous methods for reconstructing the image u^0 from the data Y in the recent literature are covered by a common variational idea: an estimator $\hat{u}$ of u^0 is computed as the solution of the optimization problem

$$J(u) \to \inf \quad \text{s.t.} \quad \sup_{S\in\mathcal{S}} c_S^{-1} \sum_{(i,j)\in S} |(Ku)_{ij} - Y_{ij}|^2 \le 1, \tag{2}$$

where $J : \mathrm{L}^2(\Omega) \to \overline{\mathbb{R}}$ is a convex and lower-semicontinuous regularization functional. Moreover $\mathcal{S}$ denotes a system of subsets of the grid $G = \{1,\dots,m\} \times$

A.M. Bruckstein et al. (Eds.): SSVM 2011, LNCS 6667, pp. 74–85, 2012.

$\{1,\dots,n\}$ and $\{c_S \,:\, S\in\mathcal{S}\}$ is a set of positive regularization parameters that govern the trade-off between data-fit and regularity. Solutions of (2) are a special case of *statistical multiresolution estimators (SMRE)* as studied in [1]. In this context the statistic $T:\mathbb{R}^{m\times n}\to\mathbb{R}$ defined by

$$T(v)=\sup_{S\in\mathcal{S}} c_S^{-1}\sum_{(i,j)\in S}|v_{ij}|^2\,,\quad v\in\mathbb{R}^{m\times n} \tag{3}$$

is referred to as *multiresolution (MR) statistic*. Summarizing, we find the estimator $\hat{u}$ of u^0 such that $J(u)$ is minimal und the condition $T(Ku-Y)\le 1$.

The most popular instance of (2) is obtained by choosing $S=\{G\}$. Then, the MR-statistic coincides with the *quadratic fidelity* and problem (2) can be rewritten into

$$\hat{u}(\lambda)\in\operatorname*{argmin}_{u\in \mathrm{L}^2(\Omega)}\frac{\lambda}{2}\sum_{(i,j)\in G}|(Ku)_{ij}-Y_{ij}|^2+J(u) \tag{4}$$

for a suitable multiplier $\lambda>0$. In the seminal work [2], for example, the authors proposed the *total variation semi-norm*

$$J(u)=\begin{cases}|\mathrm{D}u|\,(\Omega) & \text{if } u\in \mathrm{BV}(\Omega)\\ +\infty & \text{else}\end{cases} \tag{5}$$

as penalization functional which has been a widely used model in imaging ever since. Here, $|\mathrm{D}u|\,(\Omega)$ denotes the total variation of the (measure-valued) gradient of u which coincides with $\int_\Omega|\nabla u|$ if u is smooth. Numerous efficient solution methods for (2) [3–5] and various modifications have been suggested so far (cf. [6–9] to name but a few).

However, the quadratic fidelity has an essential drawback: the information in the residual is incorporated *globally*, that is each pixel value $(Ku)_{ij}-Y_{ij}$ contributes equally to the statistic T *independent of its spatial position*. In practical situations this is clearly undesirable: images usually contain features of different scales and modality, i.e. constant and smooth portions as well as oscillating patterns both of different sizes. A solution $\hat{u}$ of (2) with a global fidelity T is hence likely to exhibit under- and oversmoothed regions at the same time.

Recently, also non-trivial choices of $\mathcal{S}$ that result in *locally adaptive* fidelity measures were considered. In [10] $\mathcal{S}$ is chosen to consist of a partition of G which is obtained beforehand by a Mumford-Shah segmentation. In [11, 12], a subset $S\subset G$ is fixed and afterwards $\mathcal{S}$ is defined as the collection of all translates of S. Both approaches allow for an approximate solution of (2) by means of an analogon of (4) with locally varying regularization parameter, i.e.

$$\hat{u}\in\operatorname*{argmin}_{u\in \mathrm{L}^2(\Omega)}\frac{1}{2}\sum_{(i,j)\in G}\lambda_{ij}\,|(Ku)_{ij}-Y_{ij}|^2+J(u). \tag{6}$$

In this work we amend this paradigm and present a numerical framework that is capable of directly solving (2) without any restrictions to $\mathcal{S}$. To this end we

extend the algorithmic ideas in [1] and propose a combination of an inexact augmented Lagrangian method [9, 13] with Dykstra's projection algorithm [14]. We also propose a novel a priori parameter choice rule for the constants c_S that allows for a statistical interpretation of the latter. We illustrate the capability of our approach by numerical examples, focusing on total variation regularization.

In the following we denote by $|S|$ the cardinality of $S \in \mathcal{S}$. We often refer to $|S|$ as the *scale* of S. We assume that $m, n \in \mathbb{N}$ are fixed and denote by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ the Euclidean inner-product and norm on $\mathbb{R}^{m\times n}$ and by $\|u\|_{\mathrm{L}^2}$ the L^2-norm of u.

2 Statistical Multiresolution Estimation

In this section we review sufficient conditions that guarantee existence of SMRE, that is of a solution of (2). Moreover we propose a statistically sound parameter choice model for the constants c_S and discuss how to choose the system $\mathcal{S}$.

2.1 Existence of SMRE

For the time being, let $\{c_S \ : \ S \in \mathcal{S}\}$ be a set of positive real numbers. We rewrite (2) to an equality constrained problem by introducing the slack variable $v \in \mathbb{R}^{m\times n}$. To be more precise, we aim for the solution of

$$J(u) + H(v) \to \inf \quad \text{s.t.} \quad Ku + v = Y \tag{7}$$

where H denotes the indicator function on the feasible set $\mathcal{C}$ of (2), i.e.

$$\mathcal{C} = \left\{ v \in \mathbb{R}^{m\times n} \ : \ T(v) \le 1 \right\} \quad \text{and} \quad H(v) = \begin{cases} 0 & \text{if } v \in \mathcal{C} \\ \infty & \text{else} \end{cases}. \tag{8}$$

Problems of type (7) are studied e.g. in [15, Chap. III]. There, Lagrangian multiplier methods are employed to solve (7). Recall the definition of the *augmented Lagrangian* of (7):

$$L_\lambda(u, v; p) = \frac{1}{2\lambda} \|Ku + v - Y\|^2 + J(u) + H(v) - \langle p, Ku + v - Y \rangle, \quad \lambda > 0. \tag{9}$$

Here $p \in \mathbb{R}^{m\times n}$ denotes the Lagrange multiplier for the linear constraint in (7). It is well known that existence of a saddle point of L_λ follows from certain constraint qualifications of the MR-statistic T. One typical example is given in Proposition 1 (see [1, Thm. 2.1] for a proof).

Proposition 1. *Assume that*

1. *there exists $\bar{u} \in L^2(\Omega)$ such that $J(\bar{u}) < \infty$ and $T(K\bar{u} - Y) < 1$ and that*
2. *for all $c \in \mathbb{R}$, the following sets are bounded:*

$$\left\{ u \in L^2(\Omega) \ : \ \sup_{S\in\mathcal{S}} \sum_{(i,j)\in S} |(Ku)_{ij} - Y_{ij}|^2 + J(u) \le c \right\}. \tag{10}$$

Then, there exist $\hat{u} \in L^2(\Omega)$ *and* $\hat{v}, \hat{p} \in \mathbb{R}^{m\times n}$ *such that*

$$L_\lambda(\hat{u}, \hat{v}; p) \le L_\lambda(\hat{u}, \hat{v}; \hat{p}) \le L_\lambda(u, v; \hat{p}), \quad \forall \left(u \in L^2(\Omega), v, p \in \mathbb{R}^{m\times n}\right).$$

Remark 1. 1. If $\hat{u} \in \mathrm{L}^2(\Omega)$ and $\hat{v}, \hat{p} \in \mathbb{R}^{m\times n}$ are as in Proposition 1, then $\hat{u}$ and $\hat{v}$ solve (7) and hence $\hat{u}$ is an SMR estimator.

2. Assumption 1) in Proposition 1 is called *Slater's constraint qualification*. It is for instance satisfied if the set $\left\{Ku \ : \ u \in \mathrm{L}^2(\Omega) \text{ and } J(u) < \infty\right\}$ is dense in $\mathbb{R}^{m\times n}$.
3. If J is chosen as the total variation semi-norm (5), then a sufficient condition for assumption (10) will be that there exists $(i, j) \in G$ such that $(K\mathbf{1})_{ij} \neq 0$, where $\mathbf{1} \in \mathrm{L}^2(\Omega)$ is the constant 1-function. This is immediate from Poincaré's inequality for functions in $\mathrm{BV}(\Omega)$ (cf. [16, Thm.5.11.1]).

2.2 An a Priori Parameter Selection Method

The choice of the *regularization parameters* c_S in (2) is of utmost importance for they determine the trade-off between smoothing and data-fit. We propose an a priori parameter choice method that is based on quantile values of extremes of transformed χ^2 distributions.

To this end, observe that for $S \in \mathcal{S}$ the random variable

$$t_S(\varepsilon) = \sigma^{-2} \sum_{(i,j)\in S} \varepsilon_{ij}^2$$

is χ^2-distributed with $|S|$ degrees of freedom (d.o.f.). We first aim for transforming $t_S(\varepsilon)$ to normality. It was shown in [17] that the *fourth root transform* $\sqrt[4]{t_S(\varepsilon)}$ is approximately normal with mean and variance

$$\mu_S = \sqrt[4]{|S| - 0.5} \quad \text{and} \quad \sigma_S^2 = \left(8\sqrt{|S|}\right)^{-1},$$

respectively. The fourth root transform outperforms other power transforms in the sense that the Kullback-Leibler distance to the normal distribution is minimized, see [17]. In particular, the approximation works well for small d.o.f.

Next, we consider the extreme value statistic

$$\sup_{S\in\mathcal{S}} \frac{\sqrt[4]{t_S(\varepsilon)} - \mu_S}{\sigma_S}. \tag{11}$$

We note that due to the transformation of the random variable $t_S(\varepsilon)$ to normality each scale contributes equally to the supremum in (11). Hence a parameter choice strategy based on the statistic (11) - like the one suggested in Proposition 2 below - is likely to balance the different scales occurring in $\mathcal{S}$.

It is important to note that the random variable $t_S(\varepsilon)$ and $t_{S'}(\varepsilon)$ are independent if and only if $S \cap S' = \emptyset$. As we do not assume that $\mathcal{S}$ consists of pairwise disjoint sets, (11) constitutes an extreme value statistic of *dependent* random variables. Except for special cases, little is known about the distribution of such statistics as a consequence of which the empirical distribution of (11) is considered in practice.

Proposition 2. *For $\alpha \in (0,1)$ and $S \in \mathcal{S}$ let q_α be the α-quantile of the statistic (11) and set $c_S = (q_\alpha \sigma_S + \mu_S)^4$. Then we get for each solution of (2):*

$$\mathbb{P}(J(\hat{u}) \leq J(u^0)) \geq \alpha. \tag{12}$$

Proof. From (1) and monotonicity of the fourth root transform it follows that

$$\begin{aligned}\mathbb{P}\left(T(Ku^0 - Y) \leq 1\right) &= \mathbb{P}\left(t_S(\varepsilon) \leq c_S \ \forall S \in \mathcal{S}\right) \\ &= \mathbb{P}\left(\sqrt[4]{t_S(\varepsilon)} \leq q_\alpha \sigma_S + \mu_S \ \forall S \in \mathcal{S}\right) \\ &= \mathbb{P}\left(\sup_{S \in \mathcal{S}} \frac{\sqrt[4]{t_S(\varepsilon)} - \mu_S}{\sigma_S} \leq q_\alpha\right) = \alpha.\end{aligned}$$

In other words, the constants c_S are chosen such that the true signal u^0 satisfies the constraints with probability α. By the fact that $\hat{u}$ is a solution of (2) it follows that $\mathbb{P}(T(Ku^0 - Y) \leq 1) \leq \mathbb{P}(J(\hat{u}) \leq J(u^0))$.

Remark 2. By the rule $c_S = (q_\alpha \sigma_S + \mu_S)^4$ in Proposition 2 the problem of selecting the *set* of regularization parameters c_S is reduced to the question on how to choose the *single* value $\alpha \in (0,1)$. The probability α plays the role of a regularization parameter and allows for a precise statistical interpretation: it constitutes a lower bound on the probability that the SMRE $\hat{u}$ is more regular than the true object u^0.

2.3 On the Choice of $\mathcal{S}$

In the previous section we addressed the question on how to select the regularization parameters $\{c_S\}_{S \in \mathcal{S}}$ for a given system of subsets $\mathcal{S}$ of the grid G. We will now comment on the choice of $\mathcal{S}$.

On the one hand, $\mathcal{S}$ should be chosen rich enough to resolve local features of the image sufficiently well. On the other hand, it is desirable to keep the cardinality of $\mathcal{S}$ small such that the optimization problem in (2) remains solvable within reasonable time. We suggest two different choices of $\mathcal{S}$, namely the set $\mathcal{S}_0$ of *all discrete squares in* G and the set $\mathcal{S}_2$ of *dyadic partitions of* G. The latter is obtained by recursively splitting the grid into four equal subsets until the lowest level of single pixels is reached. For the case $m = n$ it can be formally defined as

$$\mathcal{S}_2 = \bigcup_{l=1}^{\lfloor \log_2(n) \rfloor} \left\{ \left\{ k2^l, \dots, (k+1)2^l \right\}^2 \ : \ k = 0, \dots, 2^{\lfloor \log_2(n) \rfloor} \right\}.$$

Obviously, $\mathcal{S}_0$ contains much more elements than $\mathcal{S}_2$ and is hence likely to achieve a higher resolution. We indicate this behaviour in Figure 1.

Here, a solution $\bar{u}$ of (4) of a natural image from perturbed data is depicted (first row). Since this reconstruction method does not adapt the amount of regularization to the local image features, the reconstruction exhibits both over-

Fig. 1. True signal u^0, data Y with $\sigma = 0.1$ and solution of (4) $\bar{u}$ with $\lambda = 0.75$ (upper row). Oversmoothed regions identified on the scales $|S| = 4, 8$ and 16 (from left to right) for the system $\mathcal{S}_0$ (middle row) and $\mathcal{S}_2$ (lower row).

and undersmoothed regions. The oversmoothed regions can be identified via the MR-statistic T in (3) by marking those sets S in $\mathcal{S}$ for which

$$c_S^{-1} \sum_{(i,j)\in S} |Y_{ij} - (K\bar{u})_{ij}|^2 > 1.$$

The union of these sets for the systems S_0 (second row) and S_2 (third row) are highlighted in Figure 1 where we examine the scales $|S| = 4, 8, 16$ (from left to right). The parameters c_S are chosen according to Section 2.2 with $\alpha = 0.9$.

3 Algorithmic Methodology

In what follows, we present an algorithmic approach to the numerical computation of SMRE in practice that extends the methodology in [1]. We use an *inexact Uzawa-type algorithm* which decomposes the original problem into a series of subproblems which are substantially easier to solve.

3.1 Inexact Uzawa Algorithm

In order to compute the desired saddle point of the augmented Lagrangian function L_λ in (9), we use a modified version of the Uzawa-Algorithm (see e.g. [15,

Chap. III]). Starting with some initial $p_0 \in \mathbb{R}^{m\times n}$, the original algorithm consists in iteratively computing

1. $(u_k, v_k) \in \operatorname{argmin}_{u\in \mathrm{L}^2(\Omega), v\in\mathbb{R}^{m\times n}} L_\lambda(u, v; p_{k-1})$
2. $p_k = p_{k-1} - \lambda(Ku_k + v_k - Y)$.

Item 1. amounts to an implicit minimization step w.r.t. to the variabels u and v whereas 2. constitutes an explicit maximization step for the Lagrange multiplier p. The algorithm is usually stopped once the constraint in (7) is fulfilled up to a certain tolerance (e.g. with respect to the L^2-norm as described in Algorithm 1).

Rather than applying this algorithm in a straightforward manner, however, we carry out two modifications. Firstly, we add in the k-th step the following additional term to L_λ:

$$\frac{1}{2}\left(M\left\|u-u_{k-1}\right\|_{\mathrm{L}^2}^2 - \left\|K(u-u_{k-1})\right\|^2\right)^2 + \frac{\beta}{2}\left\|v-v_{k-1}\right\|^2. \tag{13}$$

Here M is chosen such that $M \geq \|K\|^2$ and $\beta \geq 0$. By adding (13) to L_λ the distance to the previous iterate is additionally penalized. As a result, we won't have to evaluate K repeatedly within an iterative minimization scheme, but only once at u_{k-1} as we will see when our algorithmic methodology will be addressed in the following subsection. Secondly, we perform successive minimization w.r.t. u and v instead of minimizing simultaneously. The resulting two subproblems can be tackled much more efficiently than the original problem. For details, we again refer to the next subsection .

After some rearrangements of the terms in L_λ and (13) and by exploiting the fact that H is the indicator function of the convex set $\mathcal{C}$, the modified Uzawa algorithm with successive minimization can be summarized as in Algorithm 1. In practice, Algorithm 1 is very stable and straightforward to implement, provided that efficient methods to solve (14) and (15) are at hand. However, a sound convergence analysis for Algorithm 1 in the present general setting is not available so far (see e.g. [18] for the linear case and [1, Thm. 2.2] for the case when the additional term in (13) is skipped).

3.2 Subproblems

Closer inspection of Algorithm 1 reveals that the original problem - computing a saddle point of L_λ - has been replaced by an iterative series of subproblems (14) and (15). We will now examine these two subproblems and propose methods that are suited to solve them. Here we proceed as in [1].

We focus on (15) first. Note that the problem given there amounts to computing the L_2-projection of $v_k := Y + \alpha p_{k-1} - Ku_{k-1}$ onto the feasible region $\mathcal{C}$ as defined in (8). Due to the supremum taken in the definition (3) of the statistic T, we can decompose $\mathcal{C}$ into $\mathcal{C} = \bigcap_{S\in\mathcal{S}} \mathcal{C}_S$ where

$$\mathcal{C}_S = \left\{ v \in \mathbb{R}^{m\times n} \;:\; c_S^{-1} \sum_{(i,j)\in S} |v_{ij}|^2 \leq 1 \right\}, \tag{17}$$

Algorithm 1. Inexact Uzawa Algorithm

Require: $Y \in \mathbb{R}^{m\times n}$ (data), $\lambda > 0$ (step size), $\tau \geq 0$ (tolerance).
Ensure: $(u[\tau], v[\tau])$ is an approximate solution of (7) computed in $k[\tau]$ iteration steps.

$u_0 \leftarrow \mathbf{0}_{\mathrm{L}^2}$ and $v_0 = p_0 \leftarrow 0$.
$r \leftarrow \|Ku_0 + v_0 - Y\|$ and $k \leftarrow 0$.
while $r > \tau$ **do**
$k \leftarrow k + 1$.
Minimize $L_\lambda(\cdot, v_{k-1}; p_{k-1}) + \frac{1}{2}\left(M\,\|\cdot - u_{k-1}\|_{\mathrm{L}^2}^2 - \|K(\cdot - u_{k-1})\|^2\right)$:

$$u_k \leftarrow \operatorname*{argmin}_{u\in \mathrm{L}^2(\Omega)} \frac{1}{2}\,\|u - (u_{k-1} - K^*(Ku_{k-1} + v_{k-1} - (Y + \lambda p_{k-1}))\|_{\mathrm{L}^2}^2 + \frac{\lambda}{M} J(u). \tag{14}$$

Minimize $L_\lambda(u_k, \cdot; p_{k-1}) + \frac{\beta}{2}\,\|\cdot - v_{k-1}\|^2$:

$$v_k \leftarrow \operatorname*{proj}_{\mathcal{C}} \left(\frac{Y + \lambda p_{k-1} + \beta v_{k-1} - Ku_k}{1+\beta}\right). \tag{15}$$

Update dual variable:

$$p_k \leftarrow p_{k-1} - \lambda^{-1}(Ku_k + v_k - Y). \tag{16}$$

$r \leftarrow \max(\|Ku_k + v_k - Y\|, \|K(u_k - u_{k-1})\|)$.
end while
$u[\tau] \leftarrow u_k$ and $v[\tau] \leftarrow v_k$ and $k[\tau] \leftarrow k$.

i.e. each $\mathcal{C}_S$ refers to the feasible region that would result if $\mathcal{S}$ contained S only. Note that all $\mathcal{C}_S$ are closed and convex sets. If we fix a $\mathcal{C}_S$ and consider some $v \notin \mathcal{C}_S$, the projection from v onto $\mathcal{C}_S$ can be stated explicitly as

$$(P_{\mathcal{C}_S}(v))_{i,j} = \begin{cases} v_{i,j} & \text{if} \quad (i,j) \notin S \\ v_{i,j}\left(1 + \sqrt{c_S / \textstyle\sum_{(k,l)\in S} |v_{k,l}|^2}\right) & \text{if} \quad (i,j) \in S. \end{cases} \tag{18}$$

This insight leads us to the conclusion that any method which computes the projection onto the intersection of closed and convex sets by projecting on the individual sets only would be feasible to solve (15). As it turns out, Dykstra's Algorithm [14] works exactly in this way and is hence our method of choice to solve (15). For a detailed statement of the algorithm and how the total number of sets that enter it may be decreased to speed up runtimes, see [1, Sec. 2.3].

We now turn our attention to (14). In contrast to the standard version of the Uzawa algorithm as stated in [15], this second subproblem in Algorithm 1 does not involve the inversion of the operator K, at least as long as a suitable constant M is chosen in (13). For this reason, (14) here simply amounts to solving an unconstrained denoising problem with a least-squares data-fit. Numerous methods for a wide range of different choices of J are available in order to cope with this problem. If J is chosen as the total variation seminorm, for example, the methods introduced in [3–5] will be suited (we will use the one in [3]).

4 Numerical Results

We conclude this paper by demonstrating the performance of SMRE as computed by our methodology introduced in Section 3. We will show SMRE computed for the *denoising* problem in Paragraph 4.1 as well as for *deconvolution* and *inpainting* problems in Paragraph 4.2. When it comes down to computation, we think of an image u as an $m \times n$ array of pixels rather than an element in $\mathrm{L}^2(\Omega)$. Accordingly, the operator K is realized as a $mn \times mn$ matrix.

4.1 Denoising

In this paragraph we consider data Y given by (1) when K is the identity matrix and u^0 is the test image in Figure 1 ($m = 341$ and $n = 512$). We compute SMRE based on the systems S_0 and S_2 as introduced in Paragraph 2.3 where we fixed $\alpha = 0.9$. To this end we utilize Algorithm 1 with $M = 1$ and $\beta = 0$, i.e. the standard Uzawa Algorithm.

We compare our estimators to the global estimators $\hat{u}(\lambda)$ ($\lambda > 0$) as defined in (4). We choose $\lambda = \lambda_2$ and $\lambda = \lambda_{\mathrm{B}}$ such that the mean squared distance and the mean symmetric Bregman distance to the true signal u^0 is minimized, respectively. To be more precise, we set

$$\lambda_2 = \mathbf{E}\left(\operatorname*{argmin}_{\lambda>0} \left\|u^0 - \hat{u}(\lambda)\right\|^2\right) \quad \text{and} \quad \lambda_{\mathrm{B}} = \mathbf{E}\left(\operatorname*{argmin}_{\lambda>0} D_J^{\mathrm{sym}}(u^0, \hat{u}(\lambda))\right), \quad (19)$$

where the symmetric Bregman distance for J as in (5) reads as

$$D_J^{\mathrm{sym}}(u, v) = \int_\Omega \left(\frac{\nabla u}{|\nabla u|} - \frac{\nabla v}{|\nabla v|}\right) \cdot (\nabla u - \nabla v)\, \mathrm{d}x.$$

Since the parameters λ_2 and λ_{B} are not accessible in practice as u^0 is unknown, we refer to $\hat{u}(\lambda_2)$ and $\hat{u}(\lambda_{\mathrm{B}})$ as L^2- and *Bregman-oracle*, respectively. In addition, we compare our approach to the *spatially adaptive TV (SA-TV)* method as introduced in [11]. The SA-TV algorithm approximates solutions of (2) for the case where $\mathcal{S}$ constitutes the set of all translates of a fixed window $S \subset G$ by computing a solution of (6) with a suitable spatially dependent regularization parameter λ. Starting from a (constant) initial parameter $\lambda = \lambda_0$ the SA-TV algorithm iteratively adjusts λ by increasing it in regions which were poorly reconstructed before according to the MR statistic.

For our numerical comparisons, we used the SA-TV-Algorithm as formulated in [11], considering square windows with side lengths 5 and 9, respectively. All parameters involved in the algorithm were chosen as suggested in [11]. As a breaking condition, we used the discrepancy principle which ended the reconstruction process after exactly three iteration steps in all of our experiments.

The reconstructions are displayed in Figure 2. By visual inspection, we find that the oracles are globally under- (L^2) and over-regularized (Bregman),

Fig. 2. Upper row: L^2 - and Bregman oracles. Middle row: SA-TV reconstruction with window size 5 and 9. Lower row: SMRE w.r.t. $\mathcal{S}_2$ and S_0 with $\alpha = 0.9$.

respectively. While the scalar parameter λ was chosen optimally w.r.t. the different distance measures, it still cannot cope with the spatially varying smoothness of the true object u^0.

In contrast, SMRE and SA-TV reconstructions exhibit the desired locally adaptive behaviour. Still the SMRE as formulated in this paper has the advantage that multiple scales are taken into account *at once*, while SA-TV only adapts the parameter on a single given scale. As a result, SA-TV reconstructions are of varying quality for finer and coarser features of the object, while the SMRE is capable of reconstructing such features equally well.

4.2 Deconvolution and Inpainting

We finally investigate the performance of our approach if the operator K in (1) is non-trivial. To be exact, we consider *inpainting* and *deconvolution* problems. For

the first we consider an inpainting domain that occludes 15% of the image with noise level $\sigma = 0.1$ (upper left panel in Figure 3) and for the latter a Gaussian convolution kernel with variance 2 and noise level $\sigma = 0.02$ (lower left panel in Figure 3). For all experiments we use the dyadic system $\mathcal{S}_2$ and $\alpha = 0.9$.

Fig. 3. Inpainting (upper row): data Y with $\sigma = 0.1$ (left) and SMRE (right). Deconvolution (lower row): data Y with $\sigma = 0.02$ (left) and SMRE (right).

Note that in both cases we have $K = K^*$ and $\|K\| = 1$; we therefore set $M = 1.01$ and $\beta = M - 1$ in (14) and (15), respectively. We use $\tau = 10^{-3}$ as breaking tolerance which results in both cases in $k[\tau] \sim 30$ iterations in Algorithm 1 and a total computation time of less than 4 min. The results are depicted in the upper right and lower right images of Figure 3, respectively.

Again, the results indicate that a reasonable trade-off between data fit and smoothing is found by the proposed a priori parameter choice rule and that the amount of smoothing is adapted according to the image features.

5 Conclusion

In this paper we showed how statistical multiresolution estimators, that is solutions of (2), can be employed for image reconstruction. We stressed that our method, combined with an a priori parameter selection rule, locally adapts the amount of regularization according to the image geometry. For the solution of the optimization problem (2) we suggested an inexact Uzawa algorithm. The performance of our method was illustrated for standard problems in imaging.

Acknowledgments. K.F. is supported by the DFG-SNF Research Group FOR916 *Statistical Regularization* (Z-Project). P.M is supported by the BMBF project 03MUPAH6 *INVERS* and by the SFB755 *Photonic Imaging on the Nanoscale*.

References

1. Frick, K., Marnitz, P., Munk, A.: Statistical multiresolution estimation in imaging: Fundamental concepts and algorithmic approach (2011), `http://arxiv.org/abs/1101.4373v1`
2. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D 60, 259–268 (1992)
3. Dobson, D.C., Vogel, C.R.: Convergence of an iterative method for total variation denoising. SIAM J. Numer. Anal. 34, 1779–1791 (1997)
4. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vision 20, 89–97 (2004)
5. Hintermüller, M., Kunisch, K.: Total bounded variation regularization as a bilaterally constrained optimization problem. SIAM J. Appl. Math. 64, 1311–1333 (2004)
6. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problmes. Numer. Math. 76, 167–188 (1997)
7. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. Multiscale Model. Simul. 4, 460–489 (2005) (electronic)
8. Frick, K., Scherzer, O.: Convex inverse scale spaces. In: Sgallari, F., Murli, A., Paragios, N. (eds.) SSVM 2007. LNCS, vol. 4485, pp. 313–325. Springer, Heidelberg (2007)
9. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM Journal on Imaging Sciences 3, 253–276 (2010)
10. Bertalmio, M., Caselles, V., Rougé, B., Solé, A.: TV based image restoration with local constraints. J. Sci. Comput. 19, 95–122 (2003)
11. Dong, Y., Hintermüller, M., Rincon-Camacho, M.: Automated regularization parameter selection in a multi-scale total variation model for image restoration. Technical report, Inst. of Mathematics and Scientific Computing, IFB Report 22 (2008)
12. Dong, Y., Hintermüller, M.: Multi-scale total variation with automated regularization parameter selection for color image restoration. In: Tai, X.C., Mørken, K., Lysaker, M., Lie, K.A. (eds.) SSVM 2009. LNCS, vol. 5567, pp. 271–281. Springer, Heidelberg (2009)
13. Elman, H.C., Golub, G.H.: Inexact and preconditioned Uzawa algorithms for saddle point problems. SIAM Journal on Numerical Analysis 31, 1645–1661 (1994)
14. Boyle, J.P., Dykstra, R.L.: A method for finding projections onto the intersection of convex sets in Hilbert spaces. In: Advances in Order Restricted Statistical Inference (Iowa City, Iowa, 1985). LNS, vol. 37, pp. 28–47. Springer, Berlin (1986)
15. Fortin, M., Glowinski, R.: Augmented Lagrangian methods. Studies in Mathematics and its Applications, vol. 15. North-Holland Publishing Co., Amsterdam (1983); Applications to the numerical solution of boundary value problems
16. Ziemer, W.P.: Weakly differentiable functions. Springer, New York (1989)
17. Hawkins, D.M., Wixley, R.: A note on the transformation of chi-squared variables to normality. Amer.Statist. 40, 296–298 (1986)
18. Zulehner, W.: Analysis of iterative methods for saddle point problems: a unified approach. Math. Comput. 71, 479–505 (2002)