Smoothed-NUV Priors for Imaging

Boxiao Ma, Nour Zalmai, and Hans-Andrea Loeliger

Abstract—Variations of $L1$-regularization including, in particular, total variation regularization, have hugely improved computational imaging. However, sharper edges and fewer staircase artifacts can be achieved with convex-concave regularizers. We present a new class of such regularizers using normal priors with unknown variance (NUV), which include smoothed versions of the logarithm function and smoothed versions of Lp norms with $p \leq 1$.

All NUV priors allow variational representations that lead to efficient algorithms for image reconstruction by iterative reweighted descent. A preferred such algorithm is iterative reweighted coordinate descent, which has no parameters (in particular, no step size to control) and is empirically robust and efficient.

The proposed priors and algorithms are demonstrated with applications to tomography. We also note that the proposed priors come with built-in edge detection, which is demonstrated by an application to image segmentation.

Index Terms—Image reconstruction, NUV, sparsity, iterative reweighted descent, expectation maximization, tomography, edge detection, image segmentation.

I. INTRODUCTION

Many problems in imaging, such as tomographic reconstruction, denoising, deblurring and so on, boil down to guessing an image from imperfect measurements. Such problems are often formulated in a statistical setting as follows. A grayscale image is represented by a vector $\mathbf{X} = (X_1, \ldots, X_L)^\mathsf{T} \in \mathbb{R}^L$ of pixel values. This image and the available observations $\mathbf{Y} \in \mathbb{R}^N$ are related by

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{Z}, \tag{1}$$

with $\mathbf{A} \in \mathbb{R}^{N \times L}$ and where $\mathbf{Z}$ is white Gaussian noise, i.e., $\mathbf{Z} \sim \mathcal{N}(0, \sigma_Z^2 \mathbf{I})$. In this paper, the matrix $\mathbf{A}$ is assumed to be known. From a specific observation $\mathbf{Y} = \mathbf{y}$, an estimate $\hat{x}$ of $\mathbf{X}$ has to be formed. This estimate is generally of the form

$$\hat{x} = \underset{\mathbf{x}}{\arg\min} \left(\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + 2\sigma_Z^2 \kappa(\mathbf{x}) \right), \tag{2}$$

with a regularization term $2\sigma_Z^2 \kappa(\mathbf{x})$. In statistical terms, the function

$$\rho(\mathbf{x}) \triangleq e^{-\kappa(\mathbf{x})} \tag{3}$$

may be viewed as a (possibly improper) prior on $\mathbf{X}$, in which case (2) is a maximum-a-posteriori (MAP) estimate.

The most popular priors/regularizers, including, in particular, total-variation (TV) regularization, are based on the $L1$ norm [1], [2] (or equivalently, on a Laplace prior), which preserves convexity in (2) while promoting sparsity (in some domain) of the solution. Many efficient algorithms have been proposed to solve the resulting minimization problem (2), cf.

[3], [4]. In consequence, the standard TV regularization and some of its variations have hugely improved computational imaging.

However, TV regularization is not perfect: it favors piecewise constant images ("under-smoothing") and produces staircase artifacts. In addition, it introduces bias on estimates of large coefficients.

Numerous methods have been proposed to improve TV regularization in different ways. For example, total generalized variation (TGV) [5], [6] uses higher-order derivatives to encourage piecewise linear (instead of piecewise constant) image reconstructions. Structure tensor total variation (STV) penalizes the image variation at each point (pixel) by taking into account the information in some local neighborhood. A combination of $L2$ and TV regularization is proposed in [7] to reconstruct piecewise smooth signals. More direct modifications of TV regularizations are proposed in [8], [9].

Giving up convexity opens additional options: nonconvex regularizers can reduce the bias on large coefficients and yield sparser solutions than TV regularization, thereby enabling reconstruction from fewer measurements [10]. Such nonconvex regularizers include the log-sum penalty [11], [12], capped $L1$ [13], minimax concave penalties (MCP) [14], Lp norms with $0 < p < 1$ [15]–[17], and others [18]–[21].

However, as pointed out in [22], many nonconvex regularizers on image gradients actually promote staircase artifacts. This problem can be largely avoided by amending a concave regularizer with a convex patch around the origin (cf. Fig. 4) [23]–[29].

In fact, we will confirm what has been noted before, that the best results are obtained with regularizers comprising a convex part (for small pixel differences) with a concave part (for large pixel difference) [22], [26], [30]. Such convex-concave regularizers (as exemplified by Fig. 4) promote smooth areas with crisp edges: the convex part takes care of the smooth areas while the concave part encourages sharp edges.

With nonconvex priors, the minimization in (2) is nontrivial. However, it is generally not necessary to find the global minimum of (2): a "good" local minimum may do. Proposed algorithmic approaches include graduated nonconvexity (GNC) [22], [31]–[33], half quadratic minimization [30], [34]–[37], Convex-ConCave optimization [38], the alternating direction method of multipliers (ADMM) [39], split Bregman iteration (SBI) [40], General Iterative Shrinkage and Thresholding (GIST) [41], and others.

In this paper, we consider a class of priors based on modeling the difference of neighboring pixels as normal random variables with individual unknown variances (NUV[1]). NUV priors are a key idea of sparse Bayesian learning [42]–[44], but

B. Ma and H.-A. Loeliger are, and N. Zalmai was, with the Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich, Switzerland. E-mail: ma@isi.ee.ethz.ch.

[1]NUV stands for Normal with Unknown Variance.

they are also closely related to variational representations of regularizers as detailed in [43], [45]–[47]. The explicit use of convex-concave smoothed NUV (SNUV[2]) priors for imaging was apparently introduced in [48] and has since been used in [49], [50].

The primary such prior is the plain-SNUV prior of [48]–[50]. Other versions of such priors include Lp-based regularizers with $p \le 1$ with a quadratic patch around the origin; for $p = 1$, we thus obtain the convex Huber function [47].

We will compare all these priors and find plain SNUV to be on a par with the best other convex-concave regularizers, and to outperform all other low-level[3] regularizers. However, the smoothed Lp norm with $p \lesssim 0.5$ performs nearly as well. Moreover, the Huber function performs best among the convex regularizers.

We will also propose a new generalization of these SNUV priors for images with vector-valued pixels such as color images. In addition, we note that all SNUV priors come with built-in edge detection based on the sparsity of the unknown variances.

Moreover, all SNUV priors have variational representations that make it easy to derive practical iterative algorithms for image reconstruction. Such algorithms include iterative reweighted least squares [51]–[54], iterative reweighted gradient descent (IRGD), and iterative reweighted coordinate descent (IRCD) [47]. These algorithms are similar in spirit to half quadratic minimization as in [30], [34]–[37].

A second group of reconstruction algorithms does not attempt to carry out the minimization in (2), but estimates the unknown variances of the NUV prior using traditional estimation techniques [55] such as expectation maximization [42], [56] or some approximation thereof. This second approach was used in [48]–[50]. In this paper, however, we find IRCD to be more reliable.

Related prior work also includes using approximate message passing (AMP) for imaging problems [57], [58]. However, AMP may fail to converge [59], especially when the matrix $\mathbf{A}$ in (1) has near-identical columns, which is typical for high-resolution imaging problems. The convergence problem of AMP is resolved by vector AMP as in [60], but vector AMP requires the singular-value decomposition of the matrix $\mathbf{A}$ to be computed, which is infeasible in many applications. An interesting new development is conjugate-gradient based AMP [61], the viability of which for high-resolution imaging remains to be assessed.

A first set of experimental results with the priors and algorithms of this paper is given in Figures 1 and 2, which show an example from computed tomography. All images in Fig. 1 are grayscale (from 0 to 1) with 256×256 ($= 65536$) pixels. The actual object (the ground truth, shown in Fig. 1.a) is a modified version of the Shepp–Logan phantom [62]. (The unmodified Shepp–Logan phantom is piecewise constant; the modification consists of adding a smooth gradient to the gray areas.) All reconstructions in Fig. 1 are based on 12 simulated projections onto a line detector with 450 individual pixels/detectors, amounting to 5400 scalar observations in total. More details will be given in Section V-A.

The first four rows in Fig. 1 show reconstructions with priors from the literature; the bottom row in Fig. 1 shows reconstructions with SNUV priors. For each reconstruction in Fig. 1, the root-mean-squared error (RMSE) is reported. Fig. 1.b results from filtered back projection (FBP); Fig. 1.c and Fig. 1.d result from $L2$ regularization on pixel values and neighboring pixel differences, respectively.

Fig. 1.e results from standard TV regularization [3]. Fig. 1.f results from $L1$ regularization on neighboring pixel differences (anisotropic TV). Figs. 1.g and 1.h are two versions of STV regularization (structure tensor total variation) as in [63]. Figs. 1.i–1.t all penalize the magnitude of neighboring pixel differences: in Fig. 1.i with the minimax concave penalty (MCP) from [14], in Fig. 1.j with the log-sum penalty, in Fig. 1.k with the $L0.5$-norm. Fig. 1.l results from the Mumford-Shah functional [29], which can be viewed as a truncated $L2$ regularizer applied to pixel differences and is therefore convex-concave. Figs. 1.m–1.p result from penalty functions from [26]. Figs. 1.q–1.t are obtained using the SNUV priors described in this paper.

The regularizers in Figs. 1.c–1.h and in Figs. 1.m and 1.q are convex; the other regularizers are not convex. The regularizers in Figs. 1.l–1.t are convex-concave, which clearly produce better reconstructions in this example.

Some more details of these same examples are shown in Fig. 2, which zooms into a small area inside the image.

It is obvious in this example that the plain-SNUV prior yields sharper edges and less staircase artifacts than most other priors, on a par with the prior in Fig. 1.p, which has nearly the same functional form. Moreover, IRCD marginally improves over approximate expectation maximization (EM).

In summary, in this paper, we describe, demonstrate, and discuss SNUV representations of priors for imaging and pertinent reconstruction algorithms in much more detail than in [48]–[50]. The explicit SNUV representation of smoothed Lp norms appears to be new, and so is the vector version of the SNUV priors. We also reconfirm the superiority of convex-concave priors over $L1$-based priors, and we point out that SNUV priors form a natural bridge between these two approaches.

Due to space constraints, we will not, in this paper, elaborate on the multi-resolution reconstruction proposed in [50]. Finally, we mention here that the plain-SNUV prior has successfully been used for blind image deblurring [64], [65].

The paper is structured as follows. The SNUV priors are introduced in Section II and further discussed in Section III. The minimization of the total cost function (2) is addressed in Section IV. Additional details on Fig. 1 and further examples from computed tomography are given in Section V. The generalization of the SNUV priors to color images and other vector-valued images is addressed in Section VI. SNUV priors come with built-in edge detection, which is used in Section VII for color image segmentation.

---

[2]SNUV stands for Smoothed NUV.

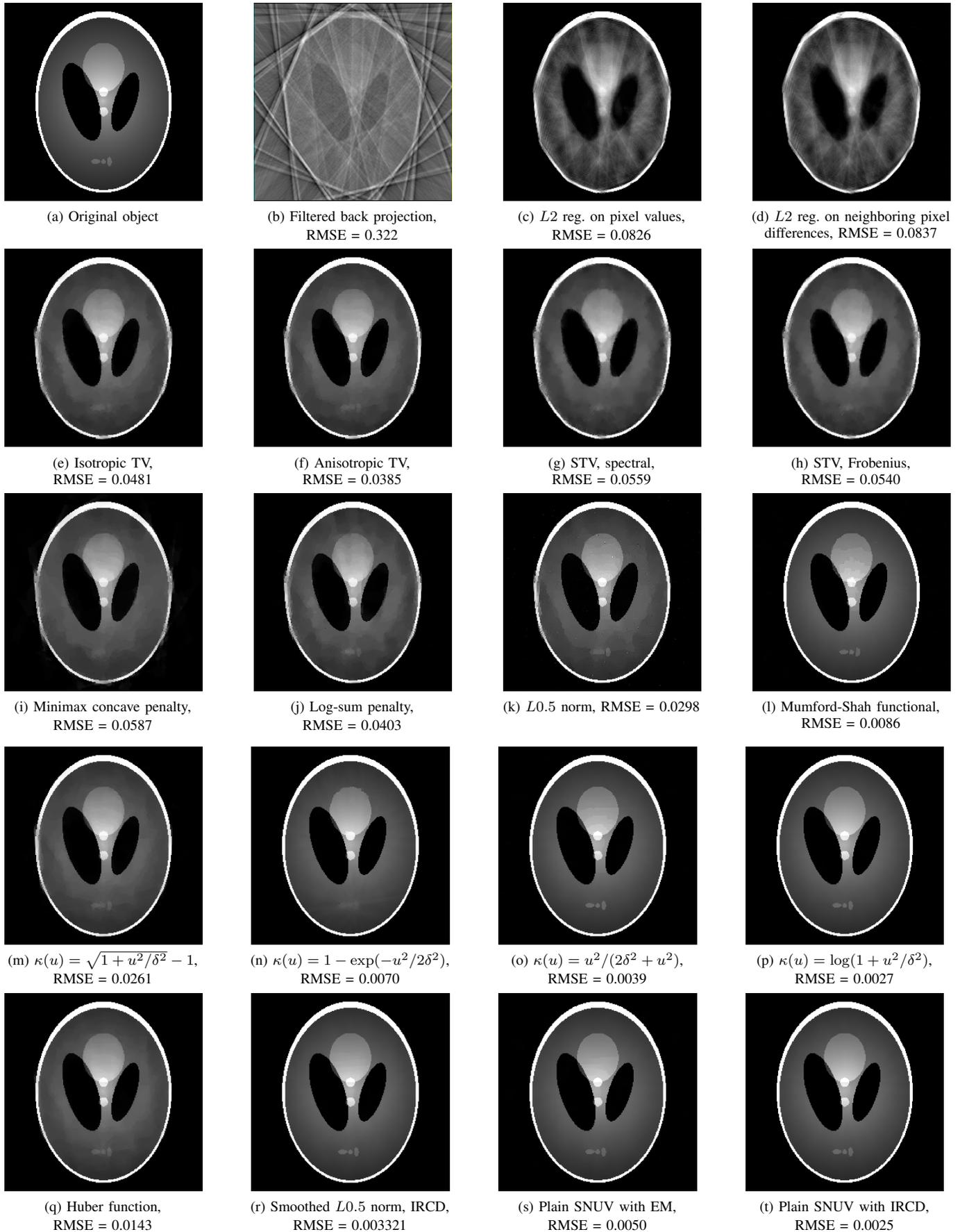[3]This paper is not concerned with neural-network priors.

Fig. 1. Tomographic reconstruction with different priors, using a modified Shepp–Logan phantom (a). The smoothed-NUV priors are in the bottom row.

(a) Original object

(b) Filtered back projection, RMSE = 0.322

(c) $L2$ reg. on pixel values, RMSE = 0.0826

(d) $L2$ reg. on neighboring pixel differences, RMSE = 0.0837

(e) Isotropic TV, RMSE = 0.0481

(f) Anisotropic TV, RMSE = 0.0385

(g) STV, spectral, RMSE = 0.0559

(h) STV, Frobenius, RMSE = 0.0540

(i) Minimax concave penalty, RMSE = 0.0587

(j) Log-sum penalty, RMSE = 0.0403

(k) $L0.5$ norm, RMSE = 0.0298

(l) Mumford-Shah functional, RMSE = 0.0086

(m) $\kappa(u) = \sqrt{1 + u^2/\delta^2} - 1$, RMSE = 0.0261

(n) $\kappa(u) = 1 - \exp(-u^2/2\delta^2)$, RMSE = 0.0070

(o) $\kappa(u) = u^2/(2\delta^2 + u^2)$, RMSE = 0.0039

(p) $\kappa(u) = \log(1 + u^2/\delta^2)$, RMSE = 0.0027

(q) Huber function, RMSE = 0.0143

(r) Smoothed $L0.5$ norm, IRCD, RMSE = 0.003321

(s) Plain SNUV with EM, RMSE = 0.0050

(t) Plain SNUV with IRCD, RMSE = 0.0025

Fig. 2. Zooming into a central part of Fig. 1.

Original

Isotropic TV, RMSE = 0.0261

Anisotropic TV ($L1$), RMSE = 0.0202

STV, Frobenius, RMSE = 0.0352

MCP RMSE = 0.0254

Log Sum penalty, RMSE = 0.0247

$L0.5$ norm, RMSE = 0.0209

Mumford-Shah RMSE = 0.0249

$\sqrt{1 + u^2/\delta^2} - 1$ RMSE = 0.0149

$1 - \exp(-u^2/2\delta^2)$ RMSE = 0.0154

$u^2/(2\delta^2 + u^2)$ RMSE = 0.0085

$\log(1 + u^2/\delta^2)$ RMSE = 0.0041

Huber function, RMSE = 0.0121

Smoothed $L0.5$ norm RMSE = 0.0054

Plain SNUV, IRCD, RMSE = 0.0040

II. THE REGULARIZERS / PRIORS

Throughout this paper, we assume the pixels or voxels $\mathbf{X} = (X_1, \ldots, X_L)$ to be arranged in a rectangular grid (in two or more dimensions), but the indices $1, \ldots, L$ need not be systematically related to this geometric arrangement. All we need of the geometry is the set $\Delta \subset \{1, \ldots, L\}^2$ of neighbors: $(k, \ell) \in \Delta$ if and only if $k < \ell$ and the pixels (or voxels) X_k and X_ℓ are immediate neighbors in some row or column. i.e., the Manhattan distance between these pixels (or voxels) equals one.

A. General Form

In this paper, the function $\kappa(\mathbf{x})$ in (2) has the form[4]

$$\kappa(\mathbf{x}) = \sum_{(k,\ell)\in\Delta} \kappa(x_k - x_\ell), \quad (4)$$

or, equivalently, the prior in (3) has the form[4]

$$\rho(\mathbf{x}) = \prod_{(k,\ell)\in\Delta} \rho(x_k - x_\ell). \quad (5)$$

In the special case where $\kappa(x_k - x_\ell) = |x_k - x_\ell|$, we thus obtain anisotropic TV regularization [66].

In this paper, we consider functions κ (with scalar argument) of the special form

$$\kappa(u_{k,\ell}) = \min_{\sigma_{k,\ell} \geq 0} \left(\frac{u_{k,\ell}^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)} - \ln \tilde{g}(\sigma_0, \sigma_{k,\ell}) \right) \quad (6)$$

with $u_{k,\ell} = x_k - x_\ell$, with parameter σ_0, and $\tilde{g}(\sigma_0, \sigma_{k,\ell})$ as discussed below. Equivalently, the prior ρ (with scalar argument) in (5) can be written as

$$\rho(u_{k,\ell}) = \max_{\sigma_{k,\ell} \geq 0} \left(\frac{\exp\left(\frac{-u_{k,\ell}^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)}\right)}{\sqrt{2\pi(\sigma_0^2 + \sigma_{k,\ell}^2)}} g(\sigma_0, \sigma_{k,\ell}) \right), \quad (7)$$

where

$$g(\sigma_0, \sigma_{k,\ell}) \triangleq \tilde{g}(\sigma_0, \sigma_{k,\ell})\sqrt{2\pi(\sigma_0^2 + \sigma_{k,\ell}^2)} \quad (8)$$

may be viewed as a prior—often an improper prior—on $\sigma_{k,\ell}$.

The function in the big parentheses in (7) is illustrated by the factor graph in Fig. 3 (cf. [47], [67], [68]). The dashed box in Fig. 3 represents the smoothed-NUV (SNUV) function[5]

$$\frac{1}{\sqrt{2\pi(\sigma_0^2 + \sigma_{k,\ell}^2)}} \exp\left(\frac{-u_{k,\ell}^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)} \right)$$

$$= \int_{-\infty}^{\infty} \frac{\exp\left(\frac{-\breve{u}_{k,\ell}^2}{2\sigma_{k,\ell}^2}\right)}{\sqrt{2\pi}\sigma_{k,\ell}} \cdot \frac{\exp\left(\frac{-(u_{k,\ell}-\breve{u}_{k,\ell})^2}{2\sigma_0^2}\right)}{\sqrt{2\pi}\sigma_0} d\breve{u}_{k,\ell}, \quad (9)$$

which is the sum of two zero-mean Gaussians with fixed variance σ_0^2 and unknown variance $\sigma_{k,\ell}^2$, respectively. The first

[4]In (4) and henceforth, with a slight abuse of notation, we use the same symbol κ both for the function $\kappa(\mathbf{x})$ (with a vector argument) and for the function $\kappa(u_{k,\ell}) = \kappa(x_k - x_\ell)$ (with a scalar argument). Likewise, we use ρ in (5) and henceforth both with a vector argument and with a scalar argument.

[5]The qualifier "smoothed" refers to the effect of $\sigma_0^2 > 0$ on the cost function around zero, as discussed later in this section.
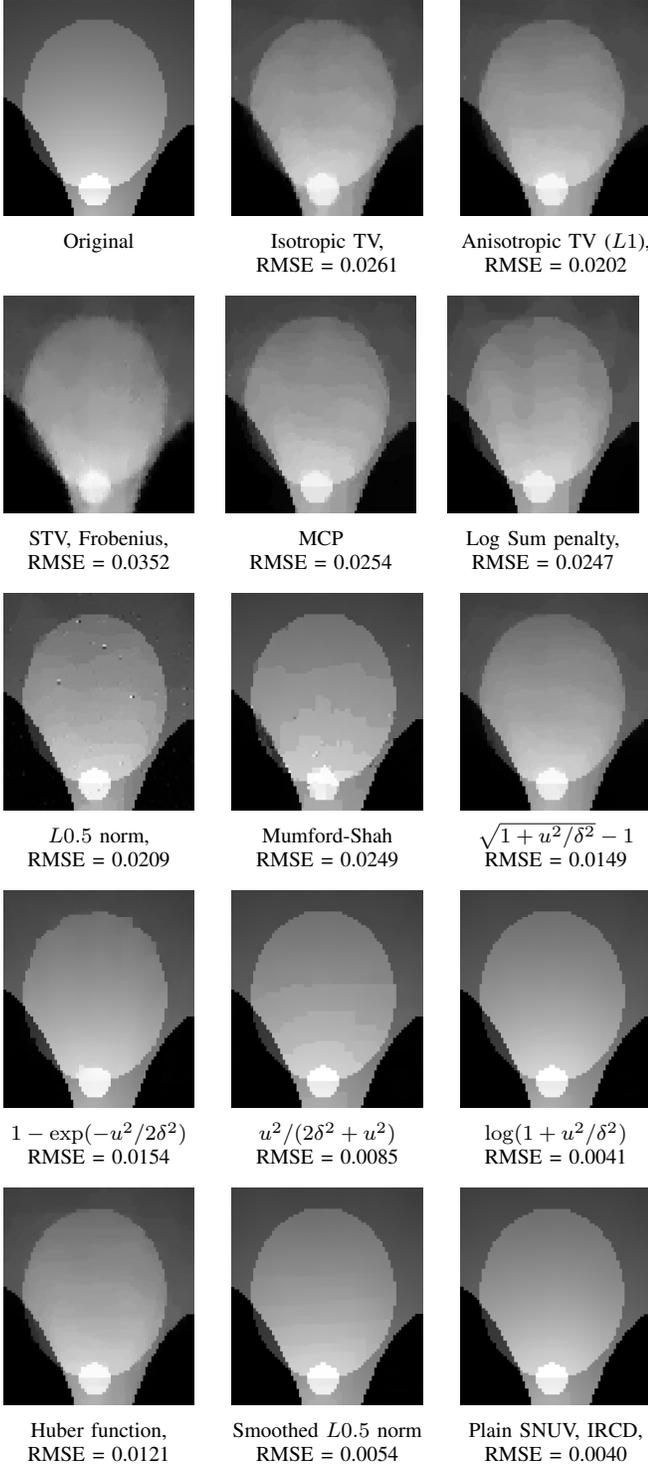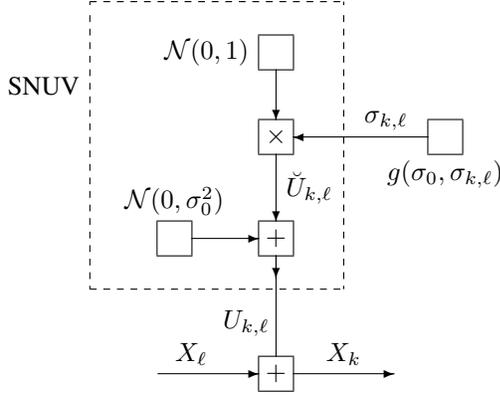
Fig. 3. Factor graph of SNUV regularization. The dashed box represents the SNUV function (9). $\mathcal{N}(m, \sigma^2)$ denotes a normal distribution with mean m and variance σ^2. The variable $U_{k,\ell} = X_k - X_\ell$ is the difference between two neighboring pixels or voxels X_k and X_ℓ. Maximization over $\sigma_{k,\ell}$ yields (7).

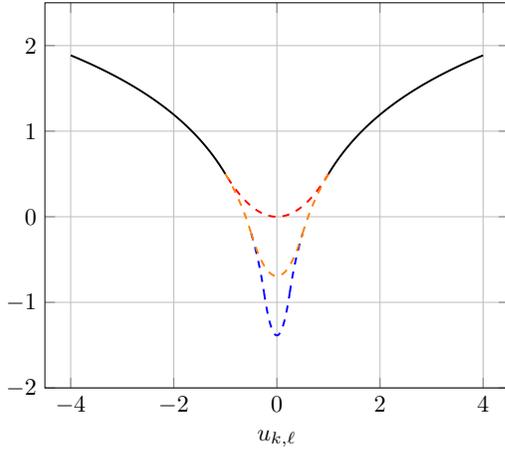Fig. 5. Solid and dashed blue: Huber function (17) for $\beta = 1$ and $\sigma_0^2 = 1$. Dotted: absolute-value function.

Fig. 4. Plain SNUV cost function (13). Dashed red: $\sigma_0^2 = 1$. Dashed orange: $\sigma_0^2 = 1/4$. Dashed blue: $\sigma_0^2 = 1/16$.

Fig. 6. Smoothed Lp norm (21): black: $p = 0.8$, $\sigma_0^2 = 5/8$, $\beta = (8/5)^{5/6}$; red: $p = 0.5$, $\sigma_0^2 = \beta = 1$; blue: $p = 0.25$, $\sigma_0^2 = 2$, $\beta = (1/2)^{4/7}$.

Gaussian accounts for the smooth parts of the image while the second Gaussian accounts for the edges, as will become clear later on.

With specific choices of $g(\sigma_0, \sigma_{k,\ell})$ (and $\tilde{g}(\sigma_0, \sigma_{k,\ell})$ by (8)), we obtain the plain-SNUV function (Fig. 4), the Huber function (Fig. 5), and smoothed versions of Lp norms (Fig. 6), as will be detailed below. The former yields the best empirical results, as exemplified by Figures 1 and 2.

However, there is more to (6) and (7): the specific form of these regularizers (or priors) is the key to the algorithms that will be discussed in Section IV. Moreover, these priors lead to sparse estimates of the parameters $\sigma_{k,\ell}$, which can be used for edge detection and more, as will be discussed in Sections III-B and VII.

B. The Plain SNUV Cost Function

The plain SNUV cost function (13) results from choosing $g(\sigma_0, \sigma_{k,\ell})$ to be some constant, i.e., using (7) without a prior on $\sigma_{k,\ell}$. In order for (13) to look nice, we choose this constant to be
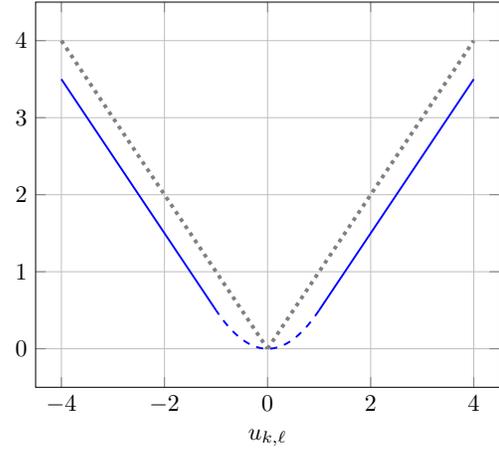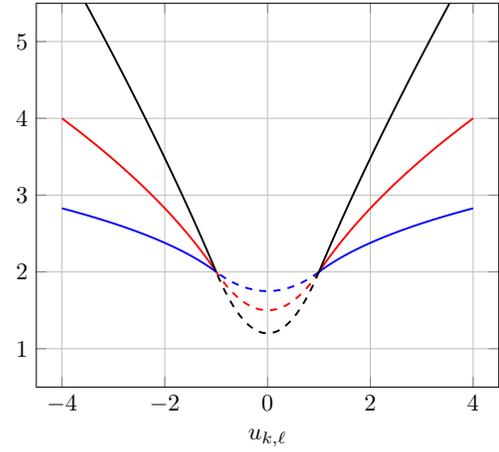
$$g(\sigma_0, \sigma_{k,\ell}) = \sqrt{2\pi}. \tag{10}$$

The minimizing $\sigma_{k,\ell}$ in (6) (= the maximizing $\sigma_{k,\ell}$ in (7)) is easily determined to be

$$\hat{\sigma}_{k,\ell}^2 = \underset{\sigma_{k,\ell}^2}{\operatorname{argmin}} \left(\frac{u_{k,\ell}^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)} + \frac{1}{2} \ln(\sigma_0^2 + \sigma_{k,\ell}^2) \right) \tag{11}$$

$$= \begin{cases} u_{k,\ell}^2 - \sigma_0^2, & \text{if } u_{k,\ell}^2 > \sigma_0^2 \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

resulting in

$$\kappa(u_{k,\ell}) = \begin{cases} \ln|u_{k,\ell}| + \frac{1}{2}, & \text{if } u_{k,\ell}^2 > \sigma_0^2 \\ \dfrac{u_{k,\ell}^2}{2\sigma_0^2} + \ln|\sigma_0|, & \text{otherwise.} \end{cases} \tag{13}$$

For $\sigma_0^2 \neq 0$, the function (13) is continuous and everywhere differentiable (even at the points $u_{k,\ell} = \pm\sigma_0$), cf. Fig. 4. Note that the two cases in (13) correspond to the concave and convex parts of κ.

C. The Huber Function

A SNUV representation of the Huber function (17) with parameters σ_0^2 and $\beta > 0$ is obtained by choosing

$$g(\sigma_0, \sigma_{k,\ell}) = \sqrt{2\pi(\sigma_0^2 + \sigma_{k,\ell}^2)} \exp\left(\frac{-\beta^2 \sigma_{k,\ell}^2}{2}\right). \quad (14)$$

(For $\sigma_0 = 0$, this happens to be a Rayleigh distribution in $\sigma_{k,\ell} \geq 0$, up to a scale factor. However, the function g does not seem to have a natural interpretation or justification other than its producing a useful function κ.) The minimizing $\sigma_{k,\ell}$ in (6) is easily determined to be

$$\hat{\sigma}_{k,\ell}^2 = \underset{\sigma_{k,\ell}^2}{\operatorname{argmin}} \left(\frac{u_{k,\ell}^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)} + \frac{\beta^2 \sigma_{k,\ell}^2}{2}\right) \quad (15)$$

$$= \begin{cases} |u_{k,\ell}|/\beta - \sigma_0^2, & \text{if } |u_{k,\ell}| > \beta\sigma_0^2 \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

resulting in

$$\kappa(u_{k,\ell}) = \begin{cases} \beta|u_{k,\ell}| - \frac{1}{2}\beta^2\sigma_0^2, & \text{if } |u_{k,\ell}| > \beta\sigma_0^2, \\ \dfrac{u_{k,\ell}^2}{2\sigma_0^2}, & \text{otherwise,} \end{cases} \quad (17)$$

which is the Huber function of robust statistics [69]. In the special case $\sigma_0^2 = 0$ and $\beta = 1$, (17) reduces to the absolute-value function. For $\sigma_0^2 \neq 0$, (17) is strictly convex and everywhere differentiable. The two cases in (17) correspond to the linear and quadratic parts of κ.

As it turns out, the parameter β is often redundant. Indeed, using the extended notation $\kappa(u_{k,\ell}, \sigma_0^2, \beta)$ for the function (17), we have

$$\kappa(u_{k,\ell}, \sigma_0^2, \beta) = \beta\kappa(u_{k,\ell}, \beta\sigma_0^2, 1). \quad (18)$$

Therefore, the effect of the parameter β on (2) is like a scale factor in σ_Z^2. In consequence, we can set $\beta = 1$ without loss of generality.

D. Smoothed Lp Norms

Eq. (14)–(17) can be generalized to smoothed Lp norms for all p with $0 < p < 2$ by choosing

$$g(\sigma_0, \sigma_{k,\ell}) = \sqrt{2\pi(\sigma_0^2 + \sigma_{k,\ell}^2)}$$
$$\cdot \exp\left(\frac{-\beta^2(2-p)(\sigma_0^2 + \sigma_{k,\ell}^2)^{\frac{p}{2-p}}}{2p}\right). \quad (19)$$

The minimizing $\sigma_{k,\ell}$ in (6) is given by

$$\hat{\sigma}_{k,\ell}^2 = \max\left\{0, \left|\frac{u_{k,\ell}}{\beta}\right|^{2-p} - \sigma_0^2\right\}, \quad (20)$$

resulting in

$$\kappa(u_{k,\ell}) = \begin{cases} \dfrac{\beta^2}{p}\left|\dfrac{u_{k,\ell}}{\beta}\right|^p, & \text{if } \left|\dfrac{u_{k,\ell}}{\beta}\right|^{2-p} > \sigma_0^2 \\ \dfrac{u_{k,\ell}^2}{2\sigma_0^2} + \beta^2\dfrac{2-p}{2p}(\sigma_0^2)^{\frac{p}{2-p}}, & \text{otherwise.} \end{cases} \quad (21)$$

For $\sigma_0^2 \neq 0$, the function (21) is continuous and everywhere differentiable. For $p = 1$, we obtain the Huber function of
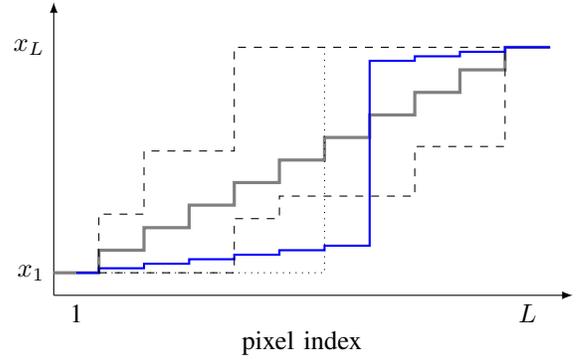

Fig. 7. One-dimensional interpolation between fixed x_1 and fixed x_L as discussed in Section III-A. A strictly convex cost function yields a regular staircase (solid gray) while a strictly concave cost function yields a single step at an arbitrary position (dotted). The absolute-value cost function assigns the same cost to every monotonous staircase (dashed). A convex-concave cost functions such as the plain SNUV yields a regular staircase (solid gray) if $|x_L - x_1|$ is small; otherwise, it yields a single large step flanked by flat regular ramps (solid blue).

Section II-C, except for a constant offset between (21) and (17). (Of course, shifting the offset between the two cases in (21) does not change the regularization properties.) In generalization of (18), we have

$$\kappa(u_{k,\ell}, \sigma_0^2, \beta) = \beta^{2-p}\kappa(u_{k,\ell}, \beta^{2-p}\sigma_0^2, 1). \quad (22)$$

For $\beta = 1$ and in the limit $p \to 0$, $\kappa(u_{k,\ell})$ as in (21) has the same derivative as (13). In consequence, the effect of the regularizer is virtually identical.

III. DISCUSSION AND EXTENSIONS

In the following, we assume all cost functions $\kappa(u_{k,\ell})$ to be monotonically increasing in $|u_{k,\ell}|$. A strictly convex cost function such as $\kappa(u_{k,\ell}) = |u_{k,\ell}|^p$ with $p > 1$ encourages smooth transitions, discourages sharp edges, and does not promote sparsity. By contrast, a strictly concave cost function such as $\kappa(u_{k,\ell}) = |u_{k,\ell}|^p$ with $0 < p < 1$ promotes sparsity and sharp edges, but creates staircase artifacts in smoothly varying areas. Convex-concave regularizers such as the plain SNUV or a smoothed Lp norm with $p \leq 1$ combine the advantages of convex and concave regularizers: smoothly varying areas of the image are addressed by the convex part while the concave part promotes sharp edges. The Huber function stands out as the only convex such regularizer, but the best empirical results are obtained with plain SNUV or smoothed Lp norms with $p < 1$.

A. Analytical 1D Interpolation

These general effects are illustrated in Fig. 7, where we consider interpolation in a one-dimensional pixel array $X_1, \ldots, X_L$ with fixed $X_1 = x_1$ and fixed $X_L = x_L$, and $x_2, \ldots, x_{L-1}$ to be interpolated by minimizing (4) with $\Delta = \{(k, k+1) : k = 1, \ldots, L-1\}$. Any strictly convex cost function yields the (unique) straight-line interpolation with $u_{1,2} = u_{2,3} = \ldots = u_{L-1,L}$. Any strictly concave cost function yields a single-step interpolation with $u_{k,k+1} = 0$ for all k except one, at an arbitrary position. The absolute-value

cost function[6] is minimized by all monotonous interpolations, most of which are irregular multi-step staircases.

By contrast, a convex-concave cost function such as the plain SNUV yields a straight line (i.e., a regular staircase) if $|x_L - x_1|$ is small, or else it yields a single big step flanked by flat regular ramps. For most imaging applications, the combination of these two modes is more satisfactory than either of them individually. The transition between these two modes is worth analyzing in detail.

Theorem. Using the plain SNUV prior[7] for interpolation with $L \geq 3$ (as defined above), the total cost (4) can have one or two local minima: the straight-line solution with $u_{1,2} = \ldots = u_{L-1,L}$ is a local (or global) minimum if and only if

$$|x_L - x_1| \leq \sigma_0(L-1). \tag{23}$$

The second local (or global) minimum exists if and only if

$$|x_L - x_1| \geq 2\sigma_0\sqrt{L-2}, \tag{24}$$

and it results in a single big step of size

$$|u_{k,k+1}| = \frac{|x_L - x_1|}{2}\left(1 + \sqrt{1 - 4(L-2)\frac{\sigma_0^2}{|x_L - x_1|^2}}\right) \tag{25}$$

for some arbitrary (but single) k. $\qquad\square$

The proof is given in Appendix A. Clearly, for

$$|x_L - x_1| \gg 2\sigma_0\sqrt{L-2}, \tag{26}$$

the single step (25) takes almost all of $|x_L - x_1|$. The intersection between (23) and (24) is never empty, and it may be large. Within this intersection, we do not bother to determine the global minimum since all algorithms considered in this paper will normally just converge to some local minimum.

B. Sparsity and Edge Detection

No matter what algorithm is used for the minimization in (27), the final estimates (12) or (16) or (20) of the variances $\sigma_{k,\ell}^2$ are normally sparse, with $\sigma_{k,\ell}^2 = 0$ indicating a smooth transition between X_k and X_ℓ and $\sigma_{k,\ell}^2 > 0$ indicating an edge. The threshold between these two cases is the transition point between the quadratic part and the concave part of the regularizer.

In other words, SNUV priors come with edge detection built in. An exemplary application will be described in Section VII.

IV. MINIMIZING THE TOTAL COST

With regularizers as in (4) and (6), the estimate (2) can be written as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\arg\min}\left(\frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2}{2\sigma_Z^2} + \sum_{(k,\ell)\in\Delta}\kappa(x_k - x_\ell)\right) \tag{27}$$

$$= \underset{\mathbf{x}}{\arg\min}\min_{\boldsymbol{\sigma}} f(\mathbf{x}, \boldsymbol{\sigma}), \tag{28}$$

---

[6]In one-dimensional "images", TV regularization amounts to an L1 penalty on the difference between neighboring pixels.

[7]The theorem does not apply to estimation by EM as in [48]–[50]

---

where $\boldsymbol{\sigma} \in \mathbb{R}^{|\Delta|}$ is the vector of all $\sigma_{k,\ell}$ with $(k,\ell) \in \Delta$ (in some arbitrary order), and

$$f(\mathbf{x}, \boldsymbol{\sigma}) \triangleq \frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2}{2\sigma_Z^2} + \sum_{(k,\ell)\in\Delta}\frac{(x_k - x_\ell)^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)}$$
$$- \sum_{(k,\ell)\in\Delta}\ln\tilde{g}(\sigma_0, \sigma_{k,\ell}). \tag{29}$$

In the special case where both $\sigma_0 = 0$ and $\sigma_{k,\ell} = 0$, $\kappa(x_k - x_\ell)$ enforces the constraint $x_k = x_\ell$, see also Fig. 3. However, in this paper, we henceforth assume $\sigma_0 > 0$.

We also note that both σ_Z^2 and σ_0^2 are effectively parameters of the proposed method. In this paper, these parameters are set (and optimized) manually, cf. Table III.

A. Iterative Reweighted Descent

The double minimization in (28) suggests to compute $\hat{\mathbf{x}}$ by algorithms[8] that iterate the following two steps:

1) A descent step in $\mathbf{x}$ with fixed $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{(\nu-1)}$.
2) Minimization over $\boldsymbol{\sigma}$ with fixed $\mathbf{x} = \mathbf{x}^{(\nu)}$:

$$\boldsymbol{\sigma}^{(\nu)} = \underset{\boldsymbol{\sigma}}{\arg\min} f(\mathbf{x}^{(\nu)}, \boldsymbol{\sigma}). \tag{30}$$

Beginning with initial values $\mathbf{x}^{(0)}$ and $\boldsymbol{\sigma}^{(0)}$, such an algorithm computes a sequence $\mathbf{x}^{(1)}, \boldsymbol{\sigma}^{(1)}, \mathbf{x}^{(2)}, \boldsymbol{\sigma}^{(2)}, \ldots$ with the idea that $\mathbf{x}^{(\nu)}$ converges to $\hat{\mathbf{x}}$ as in (28). Both steps are easy, as we are now going to discuss.

Concerning Step 2, it is obvious from (29) that (30) decomposes into scalar minimizations

$$\sigma_{k,\ell}^{(\nu)} = \underset{\sigma_{k,\ell}\geq 0}{\arg\min}\left(\frac{(x_k^{(\nu)} - x_\ell^{(\nu)})^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)} - \ln\tilde{g}(\sigma_0, \sigma_{k,\ell})\right), \tag{31}$$

which in turn have the closed-form solutions (12), (16), (20) with $u_{k,\ell} = x_k^{(\nu)} - x_\ell^{(\nu)}$.

Concerning Step 1, we first note that, with $\boldsymbol{\sigma}$ fixed, $f(\mathbf{x}, \boldsymbol{\sigma})$ is a quadratic form in $\mathbf{x}$. In other words, computing

$$\underset{\mathbf{x}}{\arg\min} f(\mathbf{x}, \boldsymbol{\sigma}^{(\nu-1)}) \tag{32}$$

is a least-squares problem, and using (32) in Step 1 results in an iterative reweighted least-squares algorithm in the style of [51]–[53]. However, computing the actual minimum (32) in Step 1 may be inefficient: making a "reasonable" step $\mathbf{x}^{(\nu-1)} \mapsto \mathbf{x}^{(\nu)}$ towards (32) is enough. For example, implementing Step 1 by a steepest-descent step will do, resulting in an iterative reweighted gradient descent (IRGD) algorithm [47].

An especially simple such method is coordinate descent, resulting in iterative reweighted coordinate descent (IRCD) as described below. This algorithm has no parameters and (in our experience) works well in practice.

Concerning the convergence of these algorithms, we note that $f(\mathbf{x}, \boldsymbol{\sigma})$ achieves its minimum at finite arguments and is continuous and continuously differentiable in all arguments. In consequence, the sequence $f(\mathbf{x}^{(\nu)}, \boldsymbol{\sigma}^{(\nu)})$, $\nu = 1, 2, \ldots$,

---

[8]The basic idea of this algorithm is far from new. For example, it includes half quadratic minimization as in [30], [34]–[37].

is guaranteed to converge. Moreover, both IRGD (with any reasonable step size control) and IRCD will neither escape to infinity nor get stuck except at a stationary point. While this does not suffice to actually prove convergence of the sequence $\mathbf{x}^{(\nu)}$, $\nu = 1, 2, \ldots$, it leaves nonconvergence only as a theoretical possibility under very special conditions. In practice, we have never observed any convergence problems.

B. Iterative Reweighted Coordinate Descent (IRCD)

In this algorithm, Step 1 of Iterative Reweighted Descent is implemented by a round of coordinate descent as follows. The first component of $\mathbf{x}^{(\nu)}$ is

$$x_1^{(\nu)} = \underset{x_1}{\arg\min} f\Big(\big(x_1, x_2^{(\nu-1)}, \ldots, x_L^{(\nu-1)}\big), \boldsymbol{\sigma}^{(\nu-1)}\Big), \quad (33)$$

the second component of $\mathbf{x}^{(\nu)}$ is

$$x_2^{(\nu)} = \underset{x_2}{\arg\min} f\Big(\big(x_1^{(\nu)}, x_2, x_3^{(\nu-1)}, \ldots, x_L^{(\nu-1)}\big), \boldsymbol{\sigma}^{(\nu-1)}\Big), \tag{34}$$

and so on. Each of these scalar minimizations is a least-squares problem with closed-form solution

$$x_k^{(\nu)} = \gamma_k^{-1} \left(\sigma_Z^{-2} \mathbf{A}_{.,k}^{\mathsf{T}}(\mathbf{y} - \breve{\mathbf{y}}_k) + \sum_{\ell:\,\ell \sim k} \frac{\tilde{x}_\ell}{\sigma_0^2 + \sigma_{k,\ell}^2} \right) \quad (35)$$

where $\mathbf{A}_{,k}$ denotes column k of $\mathbf{A}$, $\ell \sim k$ denotes the condition $(k, \ell) \in \Delta$ or $(\ell, k) \in \Delta$, $\sigma_{k,\ell}^2 \triangleq \sigma_{\ell,k}^2$ for $k > \ell$, and with the definitions

$$\breve{\mathbf{y}}_k \triangleq \mathbf{A}\big(x_1^{(\nu)}, \ldots, x_{k-1}^{(\nu)}, 0, x_{k+1}^{(\nu-1)}, \ldots, x_L^{(\nu-1)}\big)^{\mathsf{T}}, \quad (36)$$

$$\gamma_k \triangleq \sigma_Z^{-2} \|\mathbf{A}_{.,k}\|^2 + \sum_{\ell:\,\ell \sim k} (\sigma_0^2 + \sigma_{k,\ell}^2)^{-1}, \tag{37}$$

and

$$\tilde{x}_\ell \triangleq \begin{cases} x_\ell^{(\nu)} & \text{if } \ell < k \\ x_\ell^{(\nu-1)} & \text{if } \ell > k. \end{cases} \tag{38}$$

A proof of (35) is given in Appendix B.

Note that (36) need not be computed from scratch for every k, but can be updated according to

$$\breve{\mathbf{y}}_k = \breve{\mathbf{y}}_{k-1} + \mathbf{A}_{.,k-1} x_{k-1}^{(\nu)} - \mathbf{A}_{.,k} x_k^{(\nu-1)} \tag{39}$$

for $k > 1$.

In summary, IRCD repeats the following two steps, for $\nu = 1, 2, \ldots$ until convergence:

1) Compute (35) for $k = 1, \ldots, L$ using (39).
2) Update $\sigma_{k,\ell}^{(\nu)}$ for all $(k, \ell) \in \Delta$ using (12) or (16) (with $u_{k,\ell} = x_k^{(\nu)} - x_\ell^{(\nu)}$).

One round of IRCD requires about the same number of computations as computing the gradient in IRGD. We find IRCD to work very well in practice (but many iterations may be needed). In particular, we find IRCD to be at least as fast as IRGD and to give marginally better results, cf. also the remarks in Section V-D.

C. Orthogonal Columns and IRCD with Parallel Updates

The observation matrix $\mathbf{A}$ in (1) does not normally have orthogonal columns. In fact, in most high-resolution imaging problems, any two columns of $\mathbf{A}$ belonging to neighboring pixels (or voxels) are nearly identical. Nonetheless, there may be situations where $\mathbf{A}$ has orthogonal columns; for example, in Section VII below, $\mathbf{A}$ is an identity matrix.

If the columns of $\mathbf{A}$ are orthogonal, then it is easily seen from (36) that (35) simplifies to

$$x_k^{(\nu)} = \gamma_k^{-1} \left(\sigma_Z^{-2} \mathbf{A}_{.,k}^{\mathsf{T}} \mathbf{y} + \sum_{\ell:\,\ell \sim k} \frac{\tilde{x}_\ell}{\sigma_0^2 + \sigma_{k,\ell}^2} \right). \tag{40}$$

Note that the update rule (40) can be executed in parallel for any nonneighboring pixels/voxels.

In the special case where $\mathbf{A}$ is an identity matrix (cf. Section VII), (40) becomes

$$x_k^{(\nu)} = \gamma_k^{-1} \left(\frac{y_k}{\sigma_Z^2} + \sum_{\ell:\,\ell \sim k} \frac{\tilde{x}_\ell}{\sigma_0^2 + \sigma_{k,\ell}^2} \right), \tag{41}$$

where y_k is component k of $\mathbf{y}$ and

$$\gamma_k = \sigma_Z^{-2} + \sum_{\ell:\,\ell \sim k} (\sigma_0^2 + \sigma_{k,\ell}^2)^{-1}. \tag{42}$$

We conjecture that parallel updates of (35) (for nonneighboring pixels/voxels) work practically also if the pertinent columns of $\mathbf{A}$ are only approximately orthogonal, but we have not explored this systematically.

V. APPLICATION EXAMPLE I: COMPUTED TOMOGRAPHY

In tomography, the observations $\mathbf{y}$ in (2) are a collection of many noisy projections of the object, from many different angles, onto some detector array [70], as illustrated in Fig. 8. In two-dimensional (2D) tomography, the aim is to image a planar slice of the object from projections through the corresponding plane; in 3D tomography, the aim is to image an object in three dimensions. The projection matrix $\mathbf{A}$ may be obtained by methods such as the distance-driven projection [71]. Empty space is (traditionally and naturally) rendered as black (zero) while fully absorbing material is rendered white (one), as exemplified in Fig. 1.

A. 2D Modified Shepp–Logan Phantom (Fig. 1)

A first series of experimental results is shown in Figures 1 and 2, which were discussed in Section I. We now complete the discussion of Fig. 1 with a number of pertinent details. The setting is as in Fig. 8 with numerical values as in Table I, and with a line detector with 450 sensor pixels. The number of image pixels vs. the number of measurements is summarized in Table II.

The results in Fig. 1 used up to 8000 iterations of IRCD for the SNUV priors. For the other methods and regularizations in Fig. 1, every method is run until convergence and we optimized their parameters to get the best (in terms of RMSE) reconstruction possible.

The main parameters of plain SNUV with IRCD are σ_0^2 and σ_Z^2. An idea of the sensitivity to these parameters may be obtained from Table III.
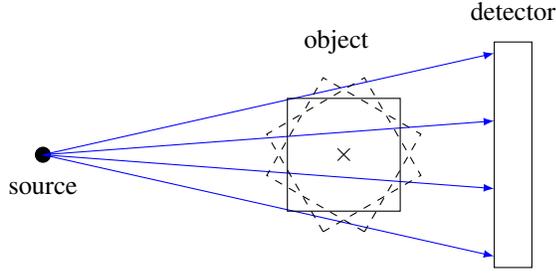
Fig. 8. Tomography with a point source. For the different projections, the object is rotated about the pivot (marked by "×") at its center.

TABLE I
DETAILS OF 2D EXAMPLES IN FIG. 1 AND FIG. 10.

distance from source to detector	800 units
distance from source to object center (pivot)	400 units
detector size	500 units
image size	200×200 units

TABLE II
NUMBER OF RECONSTRUCTED PIXELS (OR VOXELS) L AND NUMBER OF MEASUREMENTS N IN THE EXPERIMENTS OF FIG. 1 AND FIG. 11.

	L	N	N/L	RMSE
Fig. 1	256×256	12×450	8.2%	0.0025
Fig. 11	$128\times128\times128$	$10\times120\times120$	6.9%	0.0027

TABLE III
RMSE OF THE PLAIN-SNUV RECONSTRUCTION AS IN FIG. 1.T FOR DIFFERENT VALUES OF THE PARAMETERS σ_Z^2 AND σ_0^2. THE BEST RESULT IS DISPLAYED IN BOLD TYPE.

σ_Z^2	σ_0^2				
	$2\cdot10^{-5}$	$5\cdot10^{-5}$	$1\cdot10^{-4}$	$2\cdot10^{-4}$	$4\cdot10^{-4}$
$1\cdot10^{-4}$	0.1113	0.0893	0.0720	0.0532	0.0569
$2\cdot10^{-4}$	0.1065	0.0942	0.0056	0.0082	0.0158
$5\cdot10^{-4}$	0.1178	0.0863	**0.0025**	0.0028	0.0040
$1\cdot10^{-3}$	0.1312	0.0989	0.0026	0.0028	0.0037
$2\cdot10^{-3}$	0.1400	0.1050	0.0104	0.0086	0.0054

TABLE IV
DETAILS OF 3D EXAMPLE IN FIG. 11.

number of detector pixels	120×120
distance from source to detector	800 units
distance from source to object center (pivot)	400 units
detector size	450×450 units
object size	$200 \times 200 \times 200$ units

Since plain SNUV is not convex, initialization might matter. However, somewhat surprisingly, the initializations of the pixel values and the unknown variances do not seem to matter much. In our experiments, all pixels x_ℓ are initialized to some constant value (usually 0.5), and all variances $\sigma_{k,\ell}^2$ are initialized to some constant value between 10^{-5} and 10^{-3}.

B. Squares (Fig. 10)

Another series of experiments, using the phantom of Fig. 9, is summarized in Fig. 10. All images in Figures 9 and 10 have 128×128 pixels, and the smallest white squares in Fig. 9
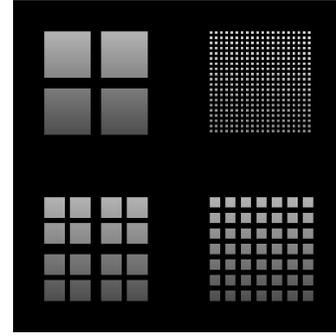
Fig. 9. 2D phantom for the experiments in Fig. 10.

consist of a single pixel. The reconstructions are based on P simulated projections onto a line detector with 115 sensor pixels (amounting to $115P$ scalar observations in total). The number P of projections is varied from 18 to 36.

The results in Fig. 10 used up to 8000 iterations of IRCD for the SNUV priors and up to 10'000 iterations of ASDPOCS for TV/L1 [3].

Again, it is obvious that plain SNUV beats the Huber function, which in turn beats the standard TV/L1 regularization.

C. 3D Modified Shepp–Logan Phantom (Fig. 11)

3D tomographic reconstruction with plain-SNUV regularization and IRCD is demonstrated in Fig. 11. Note that, in this case, $(k,\ell) \in \Delta$ if and only if $k < \ell$ and the voxels X_k and X_ℓ are immediate neighbors in any of the three coordinate directions.

The setting is as in Fig. 8 with numerical values as in Table IV. The parameters for the reconstruction are $\sigma_0^2 = 3\cdot10^{-4}$ and $\sigma_Z^2 = 3 \cdot 10^{-4}$, and up to 10'000 iterations were used. The number of voxels vs. the number of measurements is summarized in Table II.

D. Notes on the Algorithms

For the examples in this section (and similar examples not reported here), we have experimented with IRCD, IRGD, approximate EM and ASDPOCS (for standard TV). The experience with these different algorithms can be summarized as follows:

- IRCD is not slower than IRGD or approximate EM.
- IRGD is not slower than ASDPOCS.
- For plain SNUV regularization, IRCD yields marginally better results than IRGD and approximate EM.
- We normally used IRCD with a fixed (row by row) update order. Randomizing the update order does not significantly speed up the convergence and causes fluctuations in the quality of the results.
- IRCD requires $\sigma_0^2 > 0$. For concave (unsmoothed) Lp regularizers, IRGD can be used.

In summary, the simplicity, robustness, and efficiency of IRCD make it attractive for smoothed-NUV imaging.

Finally, we note here that the computational burden and the memory requirements may be reduced by adopting the multiresolution approach described in [50].
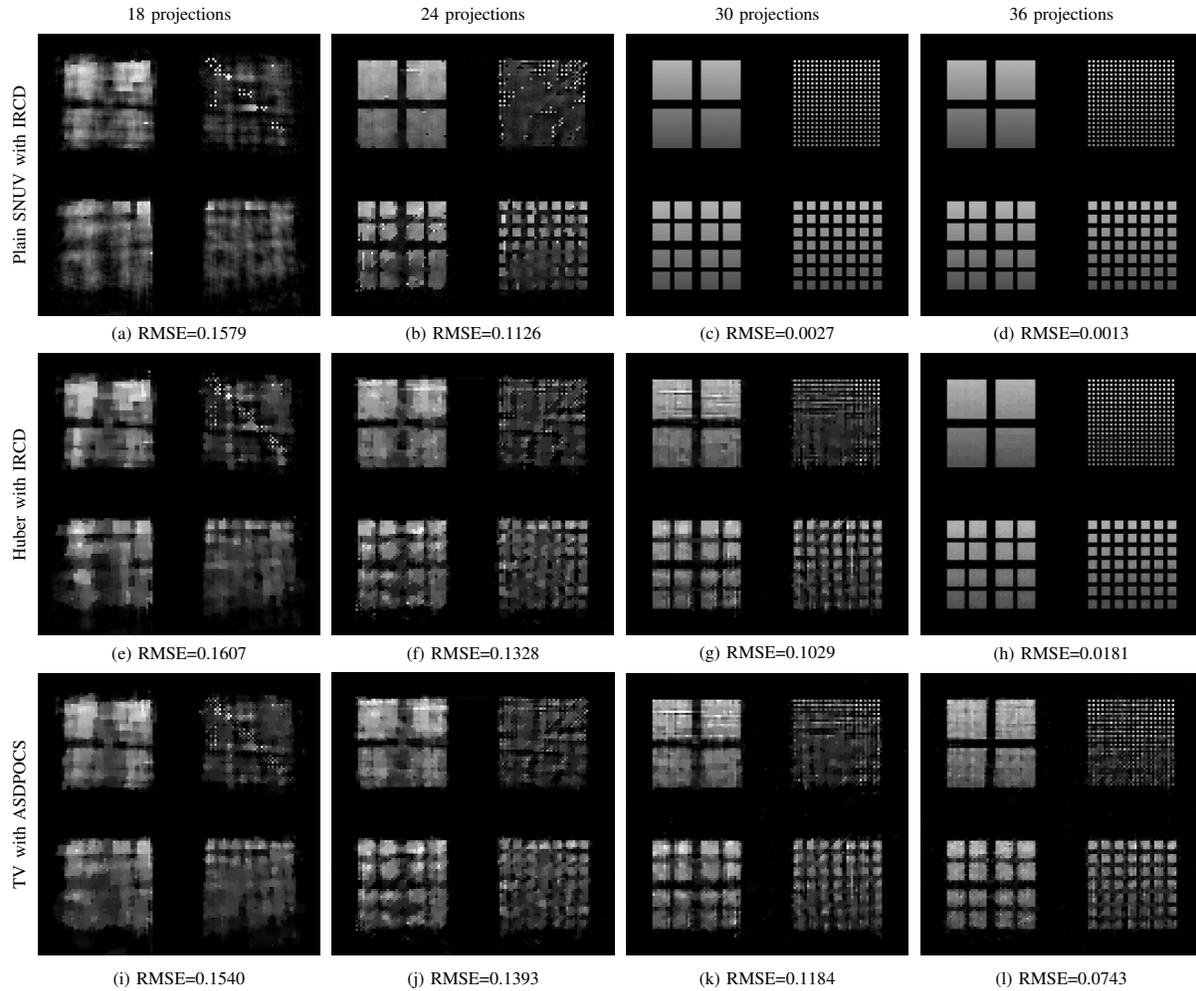
Fig. 10. Tomographic reconstructions of the phantom in Fig. 9 with different priors and different numbers of projections. Top row: plain SNUV with IRCD. Middle row: Huber function with IRCD. Bottom row: standard TV with ASDPOCS [3].

Slices of the original.

Slices of the reconstruction from 10 projections, RMSE = 0.0027.

Fig. 11. 3D tomographic reconstruction using a modified Shepp–Logan phantom and plain SNUV prior with IRCD.

VI. EXTENSION TO MULTI-CHANNEL IMAGES

So far in this paper, we have only considered grayscale images. We now address the extension of the SNUV priors to color images with pixel values $x_k \in \mathbb{R}^3$ or other images with pixel values $x_k \in \mathbb{R}^n$.
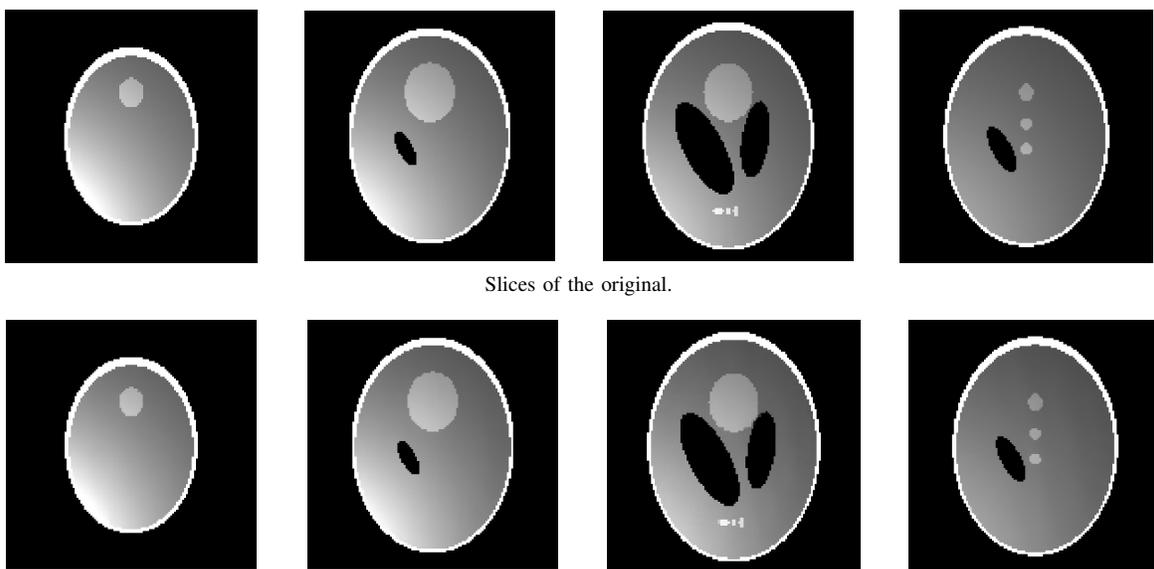
For color images, it is advantageous not to use the standard RGB representation, but the CIELAB color space [72], which is designed to match human visual perception with Euclidean distance. An example will be given in Section VII.

For pixel differences $u_{k,\ell} \in \mathbb{R}^n$, (6) is replaced by

$$\kappa(\|u_{k,\ell}\|) = \min_{\sigma_{k,\ell} \geq 0} \left(\frac{\|u_{k,\ell}\|^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)} - \ln \tilde{h}(\sigma_0, \sigma_{k,\ell}) \right) \quad (43)$$

with $\|u_{k,\ell}\|$ being the standard Euclidean norm. Note that σ_0 and $\sigma_{k,\ell}$ are still scalar. Correspondingly, the prior ρ in (7) is now

$$\rho(\|u_{k,\ell}\|) = \max_{\sigma_{k,\ell} \geq 0} \left(\frac{\exp\left(\frac{-\|u_{k,\ell}\|^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)} \right)}{(2\pi(\sigma_0^2 + \sigma_{k,\ell}^2))^{n/2}} h(\sigma_0, \sigma_{k,\ell}) \right), \quad (44)$$

with

$$h(\sigma_0, \sigma_{k,\ell}) \triangleq \tilde{h}(\sigma_0, \sigma_{k,\ell})(2\pi(\sigma_0^2 + \sigma_{k,\ell}^2))^{n/2}. \quad (45)$$

The choice of $\tilde{h}(\sigma_0, \sigma_{k,\ell})$ and the minimizing $\hat{\sigma}_{k,\ell}^2$ are very similar to the scalar case discussed in Section II.

A. Vector Plain SNUV

Choosing h as

$$h(\sigma_0, \sigma_{k,\ell}) = (2\pi)^{n/2}(\sigma_0^2 + \sigma_{k,\ell}^2)^{\mu/2}, \quad (46)$$

the minimizing $\sigma_{k,\ell}$ in (43) (= the maximizing $\sigma_{k,\ell}$ in (44)) is easily determined to be

$$\hat{\sigma}_{k,\ell}^2 = \operatorname*{argmin}_{\sigma_{k,\ell}^2} \left(\frac{\|u_{k,\ell}\|^2}{2(\sigma_0^2 + \sigma_{k,\ell}^2)} + \frac{\nu}{2}\ln(\sigma_0^2 + \sigma_{k,\ell}^2) \right) \quad (47)$$

$$= \begin{cases} \|u_{k,\ell}\|^2/\nu - \sigma_0^2, & \text{if } \|u_{k,\ell}\|^2 > \nu\sigma_0^2 \\ 0, & \text{otherwise} \end{cases} \quad (48)$$

with $\nu \triangleq n - \mu > 0$. This gives

$$\kappa(\|u_{k,\ell}\|) = \begin{cases} \nu \ln\|u_{k,\ell}\| + \frac{\nu(1-\ln\nu)}{2}, & \text{if } \|u_{k,\ell}\|^2 > \nu\sigma_0^2 \\ \frac{\|u_{k,\ell}\|^2}{2\sigma_0^2} + \nu\ln|\sigma_0|, & \text{otherwise.} \end{cases} \quad (49)$$

Note that μ (and thus ν) can be chosen freely, subject to $\nu > 0$. For $\mu = 0$ and $n = \nu = 1$, (48) and (49) reduce to (12) and (13), respectively.

B. Smoothed Powers of the Euclidean Norm

The generalization of the SNUV representation of scalar Lp norms is as follows. With

$$h(\sigma_0, \sigma_{k,\ell}) = \left(2\pi(\sigma_0^2 + \sigma_{k,\ell}^2)\right)^{n/2}$$
$$\cdot \exp\left(\frac{-\beta^2(2-p)(\sigma_0^2 + \sigma_{k,\ell}^2)^{\frac{p}{2-p}}}{2p} \right), \quad (50)$$

the minimizing $\sigma_{k,\ell}$ in (43) is given by

$$\hat{\sigma}_{k,\ell}^2 = \max\left\{ 0, \left| \frac{\|u_{k,\ell}\|}{\beta} \right|^{2-p} - \sigma_0^2 \right\}, \quad (51)$$

resulting in

$$\kappa(u_{k,\ell}) = \begin{cases} \frac{\beta^2}{p} \left| \frac{\|u_{k,\ell}\|}{\beta} \right|^p, & \left| \frac{\|u_{k,\ell}\|}{\beta} \right|^{2-p} > \sigma_0^2 \\ \frac{\|u_{k,\ell}\|^2}{2\sigma_0^2} + \beta^2 \frac{2-p}{2p}(\sigma_0^2)^{\frac{p}{2-p}}, & \text{otherwise.} \end{cases}$$
$$(52)$$

For $n = 1$, (52) reduces to the scalar smoothed Lp norm (21).

VII. APPLICATION EXAMPLE II: COLOR IMAGE SEGMENTATION BY ITERATIVE EDGE CUTTING

Image segmentation is the process of partitioning an image into non-overlapping segments so that pixels belonging to the same segment share certain properties. A great variety of segmentation algorithms exists, based on techniques such as thresholding, clustering, non-linear transformations, edge detection, and many others [73]–[77]. In addition, neural-networks have been very successful for segmentation tasks [78], [79], but such networks need to be trained with large numbers of correctly segmented images, and the results may be biased by the database used for training.

We here discuss a segmentation method, first proposed in [49], that is based on edge detection as in Sections III-B or VI. If the image is already given, we choose $\mathbf{A} = \mathbf{I}$ (an identity matrix) and $\mathbf{y}$ is the given image. However, the method can also be integrated into any imaging problem of the form (1), including tomography as in Section V. The preferred prior/regularizer is vector plain SNUV with $\mu = 0$ and $\nu = n$ as described in Section VI-A.

The algorithm iterates the following two steps:
1) Run a few iterations of Iterative Reweighted Descent as in Section IV-A.
2) Cut all detected edges.

In Step 2, an edge is detected as in Section III-B. Cutting an edge (k, ℓ) means to remove (k, ℓ) from Δ or, equivalently, to fix $\sigma_{k,\ell} = \infty$ for all subsequent iterations. The algorithm terminates when no new edges are detected and thus is guaranteed to terminate.

It turns out that this algorithm works quite well; in particular, it yields satisfactory closed segments.

The original algorithm proposed in [49] used approximate expectation maximization rather than IRCD in Step 1. In this paper, we report (new) results with IRCD, which works rather better and requires only few rounds of edge cutting. Moreover, we now use the vector plain SNUV from Section VI (instead of detecting edges in each color channel separately as in [49]).

Figures 12 and 13 show some experimental results using two test images from the Berkeley Segmentation Dataset (BSDS300) [80]. (These same examples were also used in [49], but with approximate EM instead of IRCD.) Note that, in these examples, the proposed method creates some single-pixel segments (which, of course, can easily be removed, if desired).
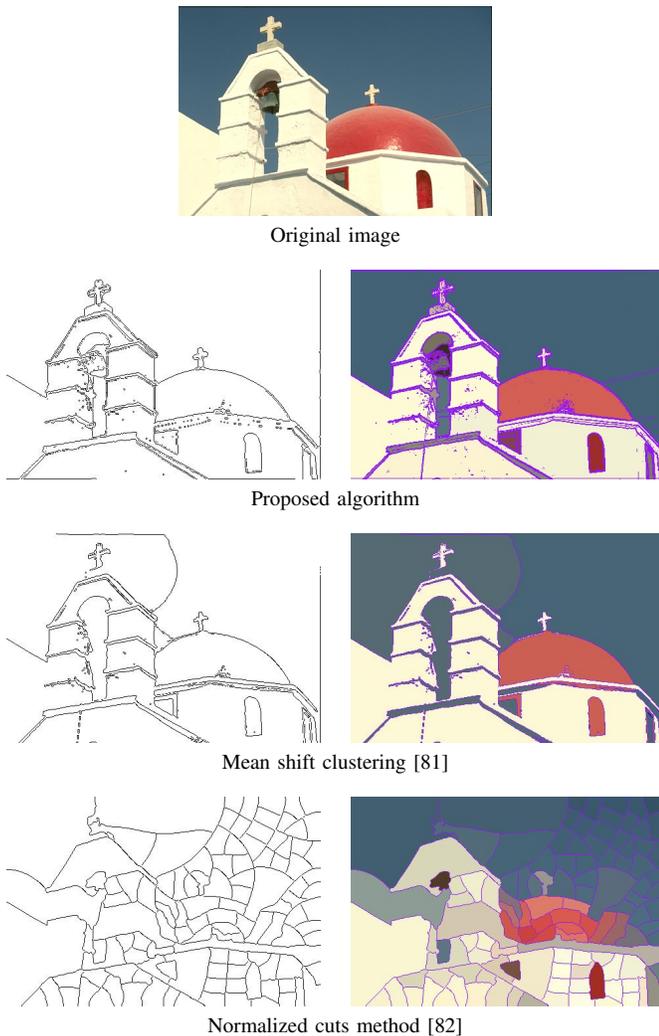
Fig. 12. A segmentation example. Left column: segment boundaries. Right column: segments colored with intrasegment average color, boundaries marked in purple.

Fig. 13. Another segmentation example. Left column: segment boundaries. Right column: segments colored with intrasegment average color.

The proposed algorithm was run with the following parameters (for both images): $\mathbf{A} = \mathbf{I}$, $\sigma_0^2 = 10^{-3}$, and $\sigma_Z^2 = 10^{-1}$, and 5 iterations of IRCD were used between subsequent rounds of edge cutting.

For comparison, Figures 12 and 13 show also the results of two other segmentation methods: mean shift [81] and normalized cuts [82]. Concerning the mean shift algorithm, note the crack in the sky behind the church, and the merging of the red parts of the horizontal ribbons with the background in the second image.

We also experimented with other standard algorithms including K-means [83] and the watershed transformation [84], but the results of these methods are inferior to all the methods in Figures 12 and 13. From these figures and from many other examples (not reported here), we conclude that the proposed algorithm is not inferior to, and arguably better than, the mean shift algorithm, and clearly outperforms the other methods.

VIII. CONCLUSION

While TV regularization prefers piecewise constant images, a preference for piecewise smooth images with crisp edges can be expressed by convex-concave regularizers/priors. We described and discussed a new class of such priors using normal priors with unknown variance (NUV).

Smoothed-NUV (SNUV) priors include smoothed versions of Lp norms with $p \leq 1$. However, the best empirical results are obtained with the plain SNUV prior, which performs on a par with the best prior convex-concave regularizers and outperforms all other low-level regularizers.

All SNUV priors have variational representations that allow practical iterative algorithms for image reconstruction. One such algorithm is iterative reweighted coordinate descent (IRCD), which has no parameters (in particular, no step size to control) and is more reliable than approximate expectation maximization (EM) that was used in prior work.

We also described a new generalization of the SNUV priors for images with vector-valued pixels such as color images. Moreover, SNUV priors come with built-in edge detection

based on the sparsity of the unknown variances.

The proposed priors and algorithms were demonstrated with applications to tomography and to color image segmentation based on built-in edge detection.

APPENDIX A
PROOF OF THE THEOREM IN SECTION III-A

Recall the notation $u_{k,k+1} \triangleq x_{k+1} - x_k$. The total cost function (4) is

$$\kappa(\mathbf{x}) \triangleq \sum_{k=1}^{L-1} \kappa(u_{k,k+1}) \tag{53}$$

with $\kappa(u_{k,k+1})$ as in (13). Instead of minimizing (53) over $\mathbf{x}$, we can equivalently minimize over $u_{1,2}, \ldots, u_{L-1,L}$ subject to the constraint

$$\sum_{k=1}^{L-1} u_{k,k+1} = x_L - x_1. \tag{54}$$

Recall that (13) is convex for $u_{k,k+1}^2 \leq \sigma_0^2$ and concave for $u_{k,k+1}^2 \geq \sigma_0^2$. It follows that, at any local minimum of (53), $u_{k,k+1}^2 > \sigma_0^2$ for at most one index k. Let ℓ be that single index. We thus have

$$u_{k,k+1}^2 \leq \sigma_0^2 \quad \text{for } k \neq \ell. \tag{55}$$

By concavity, we further have

$$u_{1,2} = \ldots = u_{L-1,L} \quad \text{except for } u_{\ell,\ell+1} \tag{56}$$

at any local minimum of (53).

We can thus reduce the minimization of (53) to a scalar minimization over $u_{\ell,\ell+1}$ according to the following lemma.

Lemma. The local minima of (53) are given by the local minima of the function

$$\tilde{\kappa}(u_{\ell,\ell+1}) \triangleq \kappa(u_{\ell,\ell+1}) + \frac{(x_L - x_1 - u_{\ell,\ell+1})^2}{2(L-2)\sigma_0^2} \tag{57}$$

together with

$$u_{k,k+1} = \frac{x_L - x_1 - u_{\ell,\ell+1}}{L-2} \quad \text{for } k \neq \ell. \tag{58}$$

$\square$

Proof of the Lemma. Using (55), the local minima of (53) coincide with the local minima of

$$\kappa(u_{\ell,\ell+1}) + \sum_{k \in \{1,\ldots,L-1\} \setminus \{\ell\}} \left(\frac{u_{k,k+1}^2}{2\sigma_0^2} + \ln |\sigma_0| \right). \tag{59}$$

From (56) and (54), we have

$$u_{k,k+1} = \frac{x_L - x_1 - u_{\ell,\ell+1}}{L-2} \quad \text{for } k \neq \ell. \tag{60}$$

Inserting (60) into (59) yields (57), up to an irrelevant constant. $\square$

From now on, we will use the shorthand $u_\ell \triangleq u_{\ell,\ell+1}$. The derivative of (57) is

$$\frac{d}{u_\ell} \tilde{\kappa}(u_\ell) = \begin{cases} \dfrac{1}{u_\ell} - \dfrac{x_L - x_1 - u_\ell}{(L-2)\sigma_0^2} & \text{if } u_\ell^2 > \sigma_0^2 \\[2ex] \dfrac{u_\ell}{\sigma_0^2} - \dfrac{x_L - x_1 - u_\ell}{(L-2)\sigma_0^2} & \text{if } u_\ell^2 \leq \sigma_0^2. \end{cases} \tag{61}$$

Note that (61) is continuous even at $|u_\ell| = \sigma_0$. Since (57) is continuous and grows to infinity for $|u_\ell| \to \infty$, there must be at least one local minimum, and all extrema of (57) can be found by setting (61) to zero.

For $u_\ell^2 \leq \sigma_0^2$, the derivative (61) is zero if and only if

$$u_\ell = \frac{x_L - x_1}{L-1} \tag{62}$$

which yields the straight-line solution

$$u_{1,2} = \ldots = u_{L-1,L}. \tag{63}$$

This solution exists if and only if $u_\ell^2 \leq \sigma_0^2$, i.e., if and only if

$$\frac{|x_L - x_1|}{L-1} \leq \sigma_0, \tag{64}$$

which is (23).

For $u_\ell^2 > \sigma_0^2$, the derivative (61) is zero if and only if

$$u_\ell(x_L - x_1 - u_\ell) - (L-2)\sigma_0^2 = 0. \tag{65}$$

This is a quadratic equation in u_ℓ with solutions

$$u_\ell = \frac{1}{2} \left(x_L - x_1 \pm \sqrt{(x_L - x_1)^2 - 4(L-2)\sigma_0^2} \right) \tag{66}$$

$$= \frac{x_L - x_1}{2} \left(1 \pm \sqrt{1 - \frac{4(L-2)\sigma_0^2}{(x_L - x_1)^2}} \right). \tag{67}$$

These solutions exist if

$$|x_L - x_1| \geq 2\sigma_0 \sqrt{L-2} \tag{68}$$

which is (24). Of the two solutions (67), the one closer to the origin is either a local maximum or it fails to satisfy $u_\ell^2 > \sigma_0^2$. The other solution

$$u_\ell = \frac{x_L - x_1}{2} \left(1 + \sqrt{1 - \frac{4(L-2)\sigma_0^2}{(x_L - x_1)^2}} \right) \tag{69}$$

is always a local minimum and satisfies

$$|u_\ell| > \frac{|x_L - x_1|}{2} \tag{70}$$

since $L \geq 3$. By (68), we then conclude that (69) is consistent with the condition $u_\ell^2 > \sigma_0^2$.

APPENDIX B
DERIVATION OF (35) BY GAUSSIAN MESSAGE PASSING

The solution (35) of the least-squares problems (33), (34), etc., can be derived by many standard methods. However, if the reader is familiar with factor graphs as in [67], [68], he may prefer to derive (35) by message passing in the factor graph of Fig. 14. In this factor graph, the pixel X_k is assumed to have the four neighbors $X_{\ell_1}, \ldots, X_{\ell_4}$. Note that this factor graph represents a Gaussian distribution with linear constraints that is equivalent to the least-squares problem at hand. Note also that this factor graph has no cycles, so that the exact (MAP/MMSE/LMMSE) solution can be determined by Gaussian message passing.

The momentarily fixed pixel values will be denoted by $\tilde{x}_\ell$ as in (38). Likewise, the momentarily fixed variances will be denoted by $\tilde{\sigma}_{k,\ell}$.

Fig. 14. Factor graph for the update of x_k in IRCD. The small filled boxes denote variables that are fixed and known (at this stage of the IRCD algorithm). The neighbor pixels of X_ℓ are denoted $X_{\ell_1}, \ldots, X_{\ell_4}$, with present values $\tilde{x}_{\ell_1}, \ldots, \tilde{x}_{\ell_4}$, respectively. The boxes labeled "$\mathcal{S}$" are SNUV factors as in Fig. 3. The blue arrows indicate the messages $\vec{\mu}_{X_k}$ and $\overleftarrow{\mu}_{X_k}$.

Using the computation tables for Gaussian messages in [44], all the following quantities can easily be read from Fig. 14. First, the backward message $\overleftarrow{\mu}_{X_k}$ along the edge X_k (i.e., the likelihood function) is a scalar Gaussian density, up to a scale factor, with precision

$$\overleftarrow{w}_{X_k} = \sigma_Z^{-2}(\mathbf{A}_{.,k})^\mathsf{T}\mathbf{A}_{.,k} \tag{71}$$

$$= \sigma_Z^{-2}\|\mathbf{A}_{.,k}\|^2 \tag{72}$$

and mean $\overleftarrow{m}_{X_k}$ determined by

$$\overleftarrow{w}_{X_k}\overleftarrow{m}_{X_k} = \sigma_Z^{-2}(\mathbf{A}_{.,k})^\mathsf{T}\left(\mathbf{y} - \sum_{\ell \in \{1,\ldots,L\}\setminus\{k\}} \mathbf{A}_{.,\ell}\tilde{x}_\ell\right) \tag{73}$$

$$= \sigma_Z^{-2}(\mathbf{A}_{.,k})^\mathsf{T}(\mathbf{y} - \breve{\mathbf{y}}_k) \tag{74}$$

with $\breve{\mathbf{y}}_k$ as in (36).

Second, the forward message $\vec{\mu}_{X_k}$ along the edge X_k (i.e., the prior) is a scalar density Gaussian with precision

$$\vec{w}_{X_k} = \sum_{\ell:\,\ell\sim k} (\sigma_0^2 + \tilde{\sigma}_{k,\ell}^2)^{-1} \tag{75}$$

and mean $\vec{m}_{X_k}$ determined by

$$\vec{w}_{X_k}\vec{m}_{X_k} = \sum_{\ell:\,\ell\sim k} \frac{\tilde{x}_\ell}{\sigma_0^2 + \tilde{\sigma}_{k,\ell}^2} \tag{76}$$

(In Fig. 14, the sums in (75) and (76) run over $\ell_1, \ldots, \ell_4$.)

Finally, the (posterior) marginal distribution of X_k is the product of $\vec{\mu}_{X_k}$ and $\overleftarrow{\mu}_{X_k}$, up to a scale factor, which is Gaussian with inverse variance given by

$$w_{X_k} = \vec{w}_{X_k} + \overleftarrow{w}_{X_k} \tag{77}$$

$$= \gamma_k \tag{78}$$

with γ_k as in (37), and with mean m_k given by

$$w_{X_k}m_k = \vec{w}_{X_k}\vec{m}_{X_k} + \overleftarrow{w}_{X_k}\overleftarrow{m}_{X_k}, \tag{79}$$

which is (35).

REFERENCES

[1] E. J. Candès and F. Guo, "New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction," *Signal Processing*, vol. 82, no. 11, pp. 1519–1543, 2002.

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[3] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, no. 17, p. 4777, 2008.

[4] M. Yan and L. A. Vese, "Expectation maximization and total variation-based model for computed tomography reconstruction from undersampled data," in *Proc. SPIE 7961, Medical Imaging*. Int. Society for Optics and Photonics, 2011.

[5] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.

[6] K. Bredies, "Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty," in *Efficient Algorithms for Global Optimization Methods in Computer Vision*. Springer, 2014, pp. 44–77.

[7] A. Gholami and S. M. Hosseini, "A balanced combination of Tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals," *Signal Processing*, vol. 93, no. 7, pp. 1945–1960, 2013.

[8] W. Stefan, R. A. Renaut, and A. Gelb, "Improved total variation-type regularization using higher order edge detectors," *SIAM Journal on Imaging Sciences*, vol. 3, no. 2, pp. 232–251, 2010.

[9] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling & Simulation*, vol. 7, no. 3, pp. 1005–1028, 2009.

[10] E. Y. Sidky, R. Chartrand, J. M. Boone, and X. Pan, "Constrained TpV minimization for enhanced exploitation of gradient sparsity: Application to CT image reconstruction," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 2, pp. 1–18, 2014.

[11] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[12] D. Tuia, R. Flamary, and M. Barlaud, "Nonconvex regularization in remote sensing," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6470–6480, 2016.

[13] D. Peleg and R. Meir, "A bilinear formulation for vector sparsity optimization," *Signal Processing*, vol. 88, no. 2, pp. 375–389, 2008.

[14] C.-H. Zhang *et al.*, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of statistics*, vol. 38, no. 2, pp. 894–942, 2010.

[15] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.

[16] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," *Advances in Neural Information Processing Systems*, vol. 22, pp. 1033–1041, 2009.

[17] M. Lai, Y. Xu, and W. Yin, "Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization," *SIAM Journal on Numerical Analysis*, vol. 51, no. 2, pp. 927–957, 2013.

[18] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ_0 norm," *IEEE Trans. on Signal Processing*, vol. 57, no. 1, pp. 289–301, 2008.

[19] J. Portilla, A. Tristán-Vega, and I. W. Selesnick, "Efficient and robust image restoration using multiple-feature $L2$-relaxed sparse analysis priors," *IEEE Trans. on Image Processing*, vol. 24, no. 12, pp. 5046–5059, 2015.

[20] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1607–1639, 2015.

[21] Y. Sun, X. Tan, X. Li, L. Lei, and G. Kuang, "Sparse optimization problem with s-difference regularization," *Signal Processing*, vol. 168, p. 107369, 2020.

[22] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *Multiscale Modeling and Simulation*, vol. 4, no. 3, pp. 960–991, 2005.

[23] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. on Medical Imaging*, vol. 8, no. 2, pp. 194–202, 1989.

[24] Y. Leclerc, "Constructing simple stable descriptions for image partitioning," *Int. Journal of Computer Vision*, vol. 3, no. 1, pp. 73–102, 1989.

[25] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on PAMI*, vol. 12, no. 7, pp. 629–639, 1990.

[26] E. Chouzenoux, J. Idier, and S. Moussaoui, "A majorize–minimize strategy for subspace optimization applied to image restoration," *IEEE Trans. on Image Processing*, vol. 20, no. 6, pp. 1517–1528, 2010.

[27] S. V. Venkatakrishnan, L. F. Drummy, M. Jackson, M. De Graef, J. Simmons, and C. A. Bouman, "Model-based iterative reconstruction for bright-field electron tomography," *IEEE Trans. on Computational Imaging*, vol. 1, no. 1, pp. 1–15, 2014.

[28] M. Foare, N. Pustelnik, and L. Condat, "Semi-linearized proximal alternating minimization for a discrete Mumford–Shah model," *IEEE Trans. on Image Processing*, vol. 29, pp. 2176–2189, 2019.

[29] K. Hohm, M. Storath, and A. Weinmann, "An algorithmic framework for Mumford–Shah regularization of inverse problems in imaging," *Inverse Problems*, vol. 31, no. 11, p. 115011, 2015.

[30] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. on Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.

[31] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT press, 1987.

[32] M. Nikolova, "Markovian reconstruction using a GNC approach," *IEEE Trans. on Image Processing*, vol. 8, no. 9, pp. 1204–1220, 1999.

[33] ——, "Weakly constrained minimization: application to the estimation of images and signals involving constant regions," *Journal of Mathematical Imaging and Vision*, vol. 21, no. 2, pp. 155–175, 2004.

[34] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[35] M. Nikolova and M. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM Journal on Scientific computing*, vol. 27, no. 3, pp. 937–966, 2005.

[36] R. He, W. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. on PAMI*, vol. 36, no. 2, pp. 261–275, 2013.

[37] M. Nikolova and R. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. on Image Processing*, vol. 16, no. 6, pp. 1623–1627, 2007.

[38] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural computation*, vol. 15, no. 4, pp. 915–936, 2003.

[39] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[40] J. Zhang, C. Zhao, D. Zhao, and W. Gao, "Image compressive sensing recovery using adaptively learned sparsifying basis via L0 minimization," *Signal Processing*, vol. 103, pp. 114–126, 2014.

[41] P. Gong, C. Zhang, Z. Lu, J. Huang, and J. Ye, "A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems," in *Proc. of the 30th Int. Conf. on Machine Learning*. PMLR, 2013, pp. 37–45.

[42] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[43] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[44] H.-A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmai, "On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing," *Information Theory and Applications Workshop (ITA)*, pp. 1–10, Feb 2016.

[45] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *Proc. 12th European Conf. on Computer Vision (ECCV)*. Springer, 2012, pp. 341–355.

[46] H. Zhang, D. P. Wipf, and Y. Zhang, "Multi-image blind deblurring using a coupled adaptive sparse prior," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2013, pp. 1051–1058.

[47] H.-A. Loeliger, B. Ma, H. Malmberg, and F. Wadehn, "Factor graphs with NUV priors and iteratively reweighted descent for sparse least squares and more," *Int. Symp. on Turbo Codes and Iterative Information Processing (ISTC)*, pp. 1–5, 2018.

[48] N. Zalmai, C. Luneau, C. Stritt, and H.-A. Loeliger, "Tomographic reconstruction using a new voxel-domain prior and Gaussian message passing," in *Proc. 24th European Signal Processing Conf. (EUSIPCO)*, 2016, pp. 2295–2299.

[49] B. Ma, N. Zalmai, R. Torfason, C. Stritt, and H.-A. Loeliger, "Color image segmentation using iterative edge cutting, NUV-EM, and Gaussian message passing," in *Proc. 5th IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, 2017, pp. 161–165.

[50] B. Ma, N. Zalmai, and H.-A. Loeliger, "A multi-resolution approach to complexity reduction in tomographic reconstruction," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6518–6522.

[51] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. on Graphics*, vol. 26, no. 3, p. 70, 2007.

[52] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

[53] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Foundations and Trends in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.

[54] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 331–372, 2015.

[55] S. S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields with applications to Bayesian tomography," *IEEE Trans. on Image Processing*, vol. 7, no. 7, pp. 1029–1044, 1998.

[56] D. P. Wipf, B. D. Rao, and S. S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Trans. on Information Theory*, vol. 57, no. 9, pp. 6236–6255, 2011.

[57] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a Markov-tree prior," *IEEE Trans. on Signal Processing*, vol. 60, no. 7, pp. 3439–3448, 2012.

[58] M. Borgerding, P. Schniter, J. Vila, and S. Rangan, "Generalized approximate message passing for cosparse analysis compressive sensing," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 3756–3760.

[59] S. Rangan, P. Schniter, A. K. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," *IEEE Trans. on Information Theory*, vol. 65, pp. 5339–5351, 2019.

[60] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. on Information Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.

[61] N. Skuratovs and M. Davies, "Upscaling vector approximate message passing," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 4757–4761.

[62] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. on Nuclear Science*, vol. 21, no. 3, pp. 21–43, 1974.

[63] S. Lefkimmiatis, A. Roussos, P. Maragos, and M. Unser, "Structure tensor total variation," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 1090–1122, 2015.

[64] B. Ma, J. Trisovic, and H.-A. Loeliger, "Multi-image blind deblurring using a smoothed NUV prior and iteratively reweighted coordinate descent," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2020, pp. 973–977.

[65] G. Marti, B. Ma, and H.-A. Loeliger, "Why maximum-a-posteriori blind image deblurring works after all," 29th European Signal Processing Conf. (EUSIPCO), Dublin, Ireland, Aug. 2021.

[66] R. Choksi, Y. van Gennip, and A. Oberman, "Anisotropic total variation regularized L^1-approximation and denoising/deblurring of 2D bar codes," *arXiv preprint arXiv:1007.1035*, 2010.

[67] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 28–41, 2004.

[68] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. of the IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.

[69] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in Statistics*. Springer, 1992, pp. 492–518.

[70] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. Springer-Verlag, 2009.

[71] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Physics in Medicine and Biology*, vol. 49, no. 11, p. 2463, 2004.

[72] J. Schanda, *Colorimetry: Understanding the CIE System*. John Wiley & Sons, 2007.

[73] K. K. Singh and A. Singh, "A study of image segmentation algorithms for different types of images," *Int. Journal of Computer Science*, vol. 7, no. 5, pp. 414–417, 2010.

[74] N. Senthilkumaran and R. Rajesh, "Edge detection techniques for image segmentation–a survey of soft computing approaches," *Int. Journal of Recent Trends in Engineering*, vol. 1, no. 2, pp. 250–254, 2009.

[75] A. Buluç, H. Meyerhenke, I. Safro, P. Sanders, and C. Schulz, "Recent advances in graph partitioning," in *Algorithm Engineering*. Springer, 2016, pp. 117–158.

[76] R. F. Moghaddam and M. Cheriet, "Adotsu: An adaptive and parameter-less generalization of Otsu's method for document image binarization," *Pattern Recognition*, vol. 45, no. 6, pp. 2419–2431, 2012.

[77] K. Parvati, P. Rao, and M. Mariya Das, "Image segmentation using gray-scale morphology and marker-controlled watershed transformation," *Discrete Dynamics in Nature and Society*, 2008.

[78] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. on PAMI*, vol. 40, no. 4, pp. 834–848, 2017.

[79] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Convolutional oriented boundaries: From image segmentation to high-level tasks," *IEEE Trans. on PAMI*, vol. 40, no. 4, pp. 819–833, 2017.

[80] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int. Conf. Computer Vision*, vol. 2, July 2001, pp. 416–423.

[81] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. on PAMI*, vol. 24, no. 5, pp. 603–619, 2002.

[82] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. on PAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[83] M. C. Chandhok, S. Chaturvedi, and A. Khurshid, "An approach to image segmentation using K-means clustering algorithm," *Int. Journal of Information Technology*, vol. 1, no. 1, pp. 11–17, 2012.

[84] A. Bala, "An improved watershed image segmentation technique using MATLAB," *Int. Journal of Scientific & Engineering Research*, vol. 3, no. 6, pp. 1–4, 2012.