Jump-Sparse and Sparse Recovery Using Potts Functionals

Martin Storath, *Member, IEEE*, Andreas Weinmann, and Laurent Demaret

***Abstract*—We recover jump-sparse and sparse signals from blurred incomplete data corrupted by (possibly non-Gaussian) noise using inverse Potts energy functionals. We obtain analytical results (existence of minimizers, complexity) on inverse Potts functionals and provide relations to sparsity problems. We then propose a new optimization method for these functionals which is based on dynamic programming and the alternating direction method of multipliers (ADMM). A series of experiments shows that the proposed method yields very satisfactory jump-sparse and sparse reconstructions, respectively. We highlight the capability of the method by comparing it with classical and recent approaches such as TV minimization (jump-sparse signals), orthogonal matching pursuit, iterative hard thresholding, and iteratively reweighted ℓ^1 minimization (sparse signals).**

***Index Terms*—ADMM, deconvolution, denoising, incomplete data, inverse Potts functional, jump-sparsity, piecewise constant signal, segmentation, sparsity.**

I. Introduction

In this article we aim at reconstructing jump-sparse (and sparse) signals $\overline{x} \in \mathbb{R}^n$ from linear noisy measurements $b \in \mathbb{R}^m$ (or $\mathbb{C}^m$) given by

$$b = A\overline{x} + \text{noise},$$

where A is a (general) $m \times n$ matrix. The reader may think of A being a Toeplitz matrix modeling blur or a Fourier matrix, or a combination of both. In particular, we deal with incomplete data meaning that the number of measurements m is significantly smaller than the size n of the original signal. Since this reconstruction problem is in general ill-posed it requires regularization. This is usually achieved by minimizing a suitable energy functional which expresses a tradeoff between data-fidelity and regularity. In view of the jump-sparsity of the underlying signal, the number of jumps $\|\nabla x\|_0 = |\{i : x_i \neq x_{i+1}\}|$ is a natural and powerful regularizing term [1]–[5]. The corresponding minimization problem, called *inverse Potts problem (iPotts),* reads

$$P_\gamma(x) = \gamma \|\nabla x\|_0 + \|Ax - b\|_p^p \to \min. \tag{1}$$

Here the parameter $\gamma > 0$ controls the tradeoff between jump-sparsity and data fidelity which is measured by some ℓ^p norm, $p \geq 1$. If the noise is Gaussian then $p = 2$ is the natural choice whereas $p = 1$ is the better choice for Laplacian or impulsive noise. (We use the notation $F(x) \to \min$ to denote the minimization problem for the functional F.)

The inverse Potts functional is not convex. To avoid the resulting difficulties, frequently the total variation (TV) penalty $\|\nabla x\|_1 = \sum_i |x_{i+1} - x_i|$ is used instead for piecewise constant signal restoration [6]–[12]. The TV problem can be solved using convex optimization and the algorithms converge to a global minimum [13]–[15]. However, the minimizers of the TV problem in general differ from those of the inverse Potts problem. It turns out that minimization of the Potts functional yields genuine jump-sparse signals whereas TV minimization does so only approximately, see for instance Fig. 1.

In this work, we are first concerned with the question of existence of minimizers which is more involved than it seems at first glance. In fact, we will see that the finite dimensional inverse Potts problem (1) has a minimizer whereas its continuous time counterpart in general need not have a minimizer at all. We further show that the inverse Potts problem is NP-hard; thus exact minimizers cannot be computed efficiently. Accepting this fact, we develop an ADMM optimization strategy which shows very good recovery performance in practice. Furthermore, we shed light on the relation between the jump-sparsity problem (1) and the sparse recovery problem. Let us be more precise.

A. Proposed ADMM Approach to the Inverse Potts Problem

We approach the inverse Potts problem (1) using the *alternating direction method of multipliers (ADMM).* ADMM strategies have recently become very popular in convex optimization especially TV minimization [16]–[20]. They have also shown their usefulness in non-convex optimization [18], [21]. We propose the iteration

$$\begin{cases} u^{k+1} \in \arg\min_u \gamma \|\nabla u\|_0 + \frac{\mu_k}{2} \left\| u - \left(v^k - \frac{\lambda^k}{\mu_k}\right)\right\|_2^2, \\ v^{k+1} = \arg\min_v \|Av - b\|_p^p + \frac{\mu_k}{2} \left\| v - \left(u^{k+1} + \frac{\lambda^k}{\mu_k}\right)\right\|_2^2, \\ \lambda^{k+1} = \lambda^k + \mu_k(u^{k+1} - v^{k+1}), \end{cases}$$

where the parameter μ_k is updated by $\mu_{k+1} = \tau\mu_k$ with fixed $\tau > 1$. The key point is that each subproblem of this *iPotts-*

Manuscript received October 09, 2013; revised March 04, 2014; accepted May 28, 2014. Date of publication June 05, 2014; date of current version June 25, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Emmanuel Candes. The research leading to these results was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant 267439 and the German Federal Ministry for Education and Research under SysTec Grant 0315508. The work of A. Weinmann was supported by the Helmholtz Association within the Young Investigator Group VH-NG-526.

M. Storath is with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: martin.storath@epfl.ch).

A. Weinmann and L. Demaret are with the Department of Mathematics, Technische Universität München, and also with the Institute of Computational Biology, Helmholtz Zentrum München, München 85764, Germany (e-mail: andreas.weinmann@helmholtz-muenchen.de; laurent.demaret@helmholtz-muenchen.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2329263

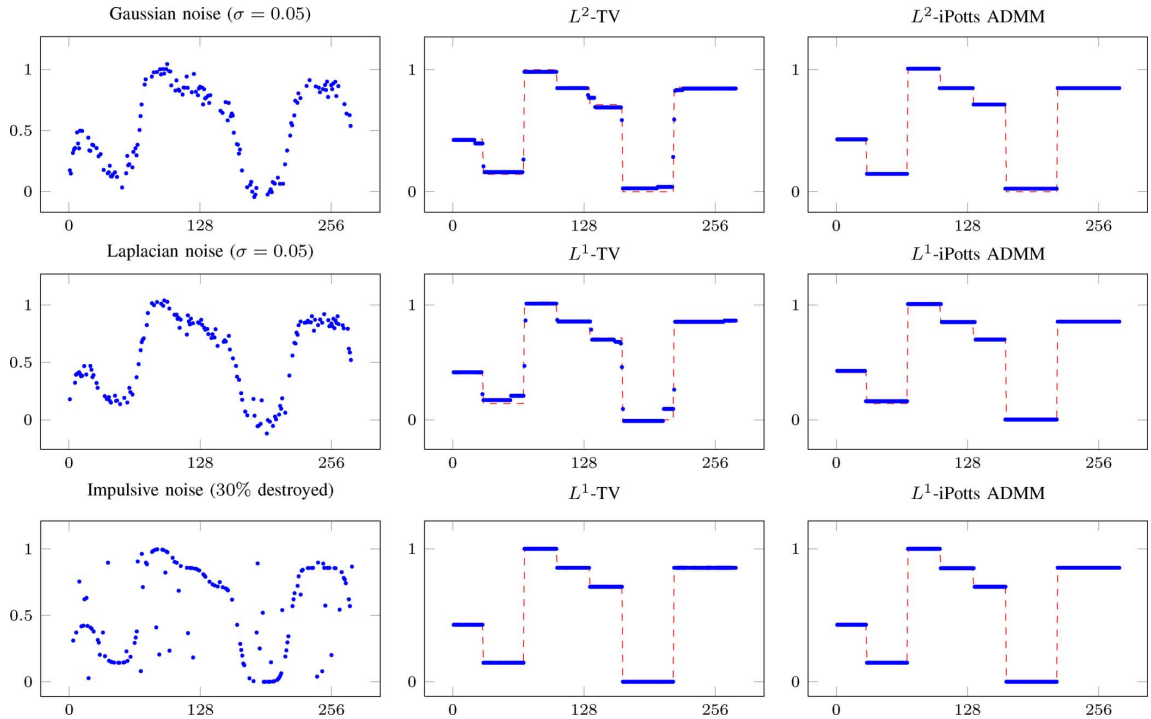


Fig. 1. The original signal (dashed line) is convolved by a Gaussian kernel of standard deviation 6 which $m = 138$ measurements are randomly selected from. The resulting data is corrupted by different types of noise (left). The total variation method (TV) mainly reconstructs the constant parts but adds transitional points in between the plateaus for Gaussian and Laplacian noise. The iPotts-ADMM recovers the true signal almost perfectly; in particular, the correct number of jumps. For impulsive noise, the iPotts-ADMM and the TV method perform equally well.

ADMM algorithm is numerically tractable. The first one is a classical Potts problem ((1) with $A = \mathrm{id}$) which can be solved fast and exactly in the univariate case. For multivariate data, such as images, we use the strategy of [2], [3]. The second subproblem consists of minimizing a classical Tikhonov functional. When $p = 2$, we solve a normal equation and, for $p = 1$, we use a fast semismooth Newton method [22]. We further show that our algorithm converges. Since the inverse Potts problem is NP-hard, we cannot expect that it converges to a global minimizer of (1) in general, but the numerical results are very satisfactory.

B. Inverse Potts Problems and Sparsity

The inverse Potts problem is closely connected to the "Lagrangian formulation" of the sparse recovery problem

$$S_\gamma(x) = \gamma \|x\|_0 + \|Ax - b\|_p^p \to \min. \tag{2}$$

The formulation (2) has been considered in [23], [24], for instance. General references concerning sparsity are the books [25]–[27] where also a variety of applications may be found in.

As with the inverse Potts problem and TV minimization, one can replace the number of non-zero entries $\|x\|_0$ by the absolute sum $\sum_i |x_i|$ to obtain a convex relaxation of the sparsity problem (2) called basis pursuit denoising (BPDN) or ℓ^1-minimization. It is one topic of compressed sensing [28]–[30] to clarify under which conditions a minimizer of the ℓ^1-functional minimizes the sparsity problem (2). Positive answers (with a high probability) are obtained under quite restrictive assumptions on the matrix A such as the restricted isometry property [31]. If such conditions are not met the solutions of BPDN are in general not minimizers of (2). Further related work replaces the jump-penalty $\|x\|_0$ by the non-convex functionals $\|x\|_q^q$ with $0 < q < 1$ [24], [32], [33].

In this work, instead of using relaxations, we transform the sparsity problem (2) to an inverse Potts problem of the form (1). We show that this can be done for all data fidelity terms based on the p-norm with $p \geq 1$. Thus we may approach the sparsity problem (2) using the proposed iPotts-ADMM algorithm.

An approach based on a transformation which is in a certain sense converse to ours is the one in [34]. There, Blake-Zisserman problems (which are certain discrete Mumford-Shah problems) with ℓ^2 data terms are transformed into separable sparsity type problems which are then approached by iterative thresholding algorithms.

C. Applications and Numerical Experiments

We apply the proposed iPotts-ADMM algorithm to reconstruct jump-sparse signals, which arise in various applications such as stepping rotations of bacterial flagella [35], the cross-hybridization of DNA [36]–[38], single-molecule fluorescence resonance energy transfer [39], and MALDI imaging [40]. Here, we recover jump-sparse signals from indirect measurements, for example from blurred data or Fourier data. The measurements are incomplete and corrupted with noise. The noise in our examples is Gaussian noise, Laplacian noise, or impulsive noise. The iPotts-ADMM algorithm is capable of recovering jump sparse signals almost perfectly from a reasonable level of noise, and gives in average higher reconstruction qualities than TV minimization.

We further apply the iPotts-ADMM based method to sparse recovery problems, which for example appear in source localization [41] or neuroimaging [42]. As for jump sparse signals

we consider blurred data under different types of noise. In our numerical experiments, we achieve similarly good results as for jump-sparse signals. We highlight the capability of our method by comparing it with orthogonal matching pursuit [43]–[45], basis pursuit denoising [46], iterative hard thresholding [23] and iteratively reweighted ℓ^1 minimization [5], which are the state-of-the-art approaches to sparse recovery.

In order to guarantee reproduciblity an implementation of our algorithms is freely available at http://pottslab.de.

D. Outline of the Paper

We start out to formulate our theoretical results on the inverse Potts problem in Section II. In Section III, we derive an ADMM algorithm for the inverse Potts problem. In Sections IV and V, we provide numerical experiments; Section IV deals with jump-sparse signals whereas, in Section V, we consider sparse signals. Finally, we supply the proofs in Section VI.

II. Inverse Potts Problems and Their Relation to Sparsity

We start our analysis of the inverse Potts problem by considering the question of existence of minimizers. It is remarkable that there is a significant difference between the finite dimensional discrete time case and its infinite dimensional continuous time counterpart. More precisely, we obtain a positive answer for the discrete time problem (1) but a negative answer for the corresponding continuous time problem.

Theorem 1: The inverse Potts problem (1) *has a minimizer.*

The proof of Theorem 1 is given in Section VI-A. It uses the compactness of the closed unit ball and the lower boundedness of an injective linear mapping which are features of finite dimensional spaces. Thus it does not carry over to the infinite dimensional continuous time case. We note that the existence of minimizers for Blake-Zisserman functionals with ℓ^2 data term has been shown in [34]. For ℓ^2 data terms, modifications of the proofs of [34] would also apply to our setting. However, for general ℓ^p data terms, the approach of [34] does not carry over.

The next theorem states that the continuous time counterpart of Theorem 1 is false in general. The continuous time counterpart of (1) is obtained by replacing the finite dimensional signal and data spaces by L^p function spaces and the matrix A by a bounded operator A between those function spaces.

Theorem 2: There are linear operators A and data b in L^p, $1 \leq p < \infty$, such that the continuous time inverse Potts problem with respect to A and b does not have a minimizer.

The proof of Theorem 2 is given in Section VI-A. The explicit counter-examples we give are convolution operators which are in fact important from a practical point of view.

The next natural step after showing the existence of minimizers (in the discrete case) is to clarify the complexity of computing such a minimizer. We obtain the following result.

Theorem 3: The inverse Potts problem (1) *is NP hard.*

As a consequence, a fast exact algorithm is not available (unless $P = NP$) and one has to resort to approximative strategies (see Section III). The proof of Theorem 3 is given in Section VI-B.

Finally, we are interested in the relations between sparsity problems and univariate inverse Potts problems. We first consider the sparsity problem (2). We find a corresponding univariate inverse Potts problem whose minimizers are directly related to the minimizers of the initial sparsity problem. We use this relation in Section V to apply our algorithm to sparsity problems.

Theorem 4: Let $x^* \in \mathbb{R}^{n+1}$ be a minimizer of the inverse Potts functional associated with the matrix $B = A\nabla$, i.e.,

$$x^* \in \arg\min_{x \in \mathbb{R}^{n+1}} \gamma \|\nabla x\|_0 + \|Bx - b\|_p^p. \tag{3}$$

Then $u^* = \nabla x^*$ minimizes the sparsity problem (2) *related to the matrix A and data b.*

We obtain a converse result for $p = 2$ (still for the univariate setting). The relations between the matrices A and B and between the data can be given explicitly but are not as simple as above. A similar relation has been used in [34] in the context of Blake-Zisserman functionals. The construction does not work for general $p \neq 2$ and it is not clear to us how to get a converse result when $p \neq 2$.

Theorem 5: For the inverse Potts problem (1) *associated with the matrix A and data b we consider the sparsity problem associated with the matrix $B = A'\nabla^+$ and data b'. Here ∇^+ is the pseudo-inverse of the discrete difference operator given by* (24)*. The modified data A' and b' are given in terms of A and b by* (29) *and* (31)*, respectively. Let u^* be a minimizer of the sparsity problem with respect to B, b',i. e.,*

$$u^* \in \arg\min_{u \in \mathbb{R}^{n-1}} \gamma \|u\|_0 + \|Bu - b'\|_2^2. \tag{4}$$

Then $x^* = \nabla^+ u^* + \mu(\nabla^+ u^*)e$ (with μ given by (27)*) is a solution of the inverse Potts problem* (1) *associated with A, b.*

The proofs of Theorem 4 and Theorem 5 are given in Section VI-B.

III. Minimization of the Potts Functional Using the Alternating Direction Method of Multipliers

In this section, we present our iterative approach to the inverse Potts problem (1).

A. A New ADMM Algorithm for the Inverse Potts Problem

The inverse Potts problem is equivalent to the bivariate constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & \gamma \|\nabla u\|_0 + \|Av - b\|_p^p \\ \text{subject to} \quad & u - v = 0. \end{aligned} \tag{5}$$

We incorporate the constraint $u - v$ into the target functional to obtain the unconstrained problem

$$\begin{aligned} L_\mu(u, v, \lambda) = {} & \gamma \|\nabla u\|_0 + \langle \lambda, u - v \rangle \\ & + \frac{\mu}{2} \|u - v\|_2^2 + \|Av - b\|_p^p \to \min. \end{aligned} \tag{6}$$

The parameter $\mu > 0$ regulates the coupling of u and v. The dual variable λ is an n-dimensional vector of Lagrange multipliers. Equation (6) is called the *augmented Lagrangian* of (5).

Completing the square in the second and third term of (6) yields

$$L_\mu(u,v,\lambda) = \gamma\|\nabla u\|_0 - \frac{\mu}{2}\left\|\frac{\lambda}{\mu}\right\|_2^2 + \frac{\mu}{2}\left\|u - v + \frac{\lambda}{\mu}\right\|_2^2 + \|Av-b\|_p^p. \quad (7)$$

In order to minimize the augmented Lagrangian (7) we use the *alternating direction method of multipliers (ADMM)*, see e.g., [18]. In the ADMM iteration we first fix v and λ and minimize $L_\mu(u,v,\lambda)$ with respect to u. Then we minimize $L_\mu(u,v,\lambda)$ with respect to v, keeping u and λ fixed. The third step is the update of the dual variable λ. Thus, the alternating direction method of multipliers for the inverse Potts problem (1) reads

$$\begin{cases} u^{k+1} \in \arg\min_u \gamma\|\nabla u\|_0 + \frac{\mu}{2}\left\|u - \left(v^k - \frac{\lambda^k}{\mu}\right)\right\|_2^2, \\ v^{k+1} = \arg\min_v \|Av-b\|_p^p + \frac{\mu}{2}\left\|v - \left(u^{k+1} + \frac{\lambda^k}{\mu}\right)\right\|_2^2, \\ \lambda^{k+1} = \lambda^k + \mu(u^{k+1} - v^{k+1}). \end{cases} \quad (8)$$

The crucial point is that both subproblems appearing in the first and the second line of (8) are computationally tractable (for $p \in [1,\infty]$). The first subproblem is the minimization of a classical Potts problem which we elaborate on in Section III-B. The second subproblem is the minimization of a classical Tikhonov-type problem which we explain in Section III-C.

We initialize the iteration with a small positive coupling parameter $\mu_0 > 0$ and increase it during the iteration by a factor $\tau > 1$. Hence, μ is given by the geometric progression

$$\mu = \mu_k = \tau^k \cdot \mu_0.$$

This assures that u and v can evolve quite independently at the beginning and that they are close to each other at the end of the iteration. We stop the iteration when the norm of $u - v$ falls below some tolerance. Our approach to the inverse Potts problem is summed up in Algorithm 1.

Algorithm 1: iPotts-ADMM

Input: Data $b \in \mathbb{R}^m$, model parameter $\gamma > 0$, measurement matrix $A \in \mathbb{R}^{m\times n}$

Output: Computed result $u \in \mathbb{R}^n$ to the inverse Potts problem **(1)**

begin

$v \leftarrow A^*b$; $\mu \leftarrow \mu_0$; $\lambda \leftarrow 0$;

repeat

$u \leftarrow$ Minimizer of classical L^2-Potts functional **(9)** with data $d = v - \frac{\lambda}{\mu}$ and parameter $\delta = \frac{2\gamma}{\mu}$;

$v \leftarrow$ Solution of Tikhonov problem **(11)** with data b, offset vector $w = u + \frac{\lambda}{\mu}$ and parameter $\frac{\mu}{2}$;

$\lambda \leftarrow \lambda + \mu(u - v)$;

$\mu \leftarrow \mu \cdot \tau$;

until $\|u - v\|_2^2 < \mathrm{TOL}$

end

We have the following convergence result, whose proof is given in Section VI-C.

Theorem 6: The ADMM iteration (8)*, and thus Algorithm 1, converges.*

Although we cannot expect convergence to a global minimum for the NP-hard inverse Potts problem, we see in the experimental section that Algorithm 1 gives very satisfactory reconstruction results.

In our experiments, reasonable numerical values for the parameters in Algorithm 1 are $\mu_0 = \gamma \cdot 10^{-6}$ as initial coupling, $\tau = 1.05$ for the increment of the coupling, and $\mathrm{TOL} = 10^{-6}$ for the stopping tolerance.

B. Minimization of the Classical Potts Subproblem

The first subproblem of the ADMM iteration (8) is a classical L^2-Potts problem of the form

$$P'_\delta(u) = \delta \cdot \|\nabla u\|_0 + \|u - f\|_2^2 \to \min \quad (9)$$

for parameter $\delta = \frac{2\gamma}{\mu_k}$ and data $f = v^k - \frac{\lambda^k}{\mu_k}$.

For univariate data this problem can be solved fast and exactly using dynamic programming [47]–[50]. The basic idea is that a minimizer of the Potts functional for data $(f_1,\ldots,f_r)$ can be computed in polynomial time provided that minimizers of the partial data $(f_1), (f_1,f_2), \ldots, (f_1,\ldots,f_{r-1})$ are known. The corresponding procedure works as follows. We denote the respective minimizers for the partial data by $u^1, u^2, \ldots, u^{r-1}$. In order to compute a minimizer for data $(f_1,\ldots,f_r)$, we create a set of r minimizer candidates $v^1,\ldots,v^r$, each of length r. These minimizer candidates are given by

$$v^\ell = (u^{\ell-1}, \underbrace{\mu_{[\ell,r]},\ldots,\mu_{[\ell,r]}}_{\text{Length } r-\ell+1}), \quad (10)$$

where u^0 is the empty vector and $\mu_{[\ell,r]}$ denotes the mean value of data $f_{[\ell,r]} = (f_\ell,\ldots,f_r)$. Among these candidates v^ℓ, one with the least Potts functional value is a minimizer for the data $f_{[1,r]}$. The dynamic program for the classical Potts problem (i.e., the recursive computation of u^n using (10)) can be performed in $O(n^2)$ time and $O(n)$ space complexity [49]. There are strategies to prune the search space which speed up the algorithm in practice [51], [52].

For multivariate data, we cannot solve the first subproblem of our ADMM algorithm exactly in reasonable time because the classical Potts problem (9) is NP-hard in two dimensions [3]. However, there exist well-working practical approaches based on graph cuts. We here use the max-flow/min-cut based algorithm of the library GCOptimization 3.0 [3], [53], [54].

C. Minimization of the Tikhonov Subproblem

The second subproblem of the ADMM iteration (8) is a classical Tikhonov problem with L^p data fitting of the form

$$\frac{\mu_k}{2}\|v - w\|_2^2 + \|Av - b\|_p^p \to \min, \quad (11)$$

where the offset vector w is given by $w = u^{k+1} + \frac{\lambda^k}{\mu_k}$. The problem is convex for all $p \in [1,\infty]$. Thus it can be solved efficiently using convex optimization. We briefly describe minimization strategies for the most relevant cases $p = 1$ and $p = 2$.

For $p = 2$, the solution is explicitly given by the solution of the normal equation

$$\left(A^*A + \frac{\mu_k}{2}\,\mathrm{id}\right) v = \frac{\mu_k}{2}u^{k+1} + \frac{1}{2}\lambda^k + A^*b. \qquad (12)$$

Here id denotes the identity matrix and A^* denotes the transposed of the conjugate. As the time complexity of solving (12) is $O(n^3)$ in general, the solution of (12) is the most expensive step in the ADMM iteration since the classical univariate Potts problem is in $O(n^2)$. However, if A^*A is a bandmatrix or if A^*A can be diagonalized efficiently then the system (12) can be solved fast and we are thus able to deal with large data sizes. For instance, if Ax describes the (circular) convolution of x with some vector h, i.e., $Ax = h * x$ then the solution of the normal equation is given by

$$w = \mathcal{F}^{-1}\left(\frac{\widehat{r}}{|\widehat{h}|^2 + \frac{\mu_k}{2}}\right)$$

where r denotes the right hand side of (12).

For $p = 1$ the minimization of the Tikhonov problem (11) is more challenging because the L^1 data term is not differentiable. Nevertheless, the problem can be treated by convex optimization. We use the approach proposed in [22]. There, the dual problem of (11) is solved iteratively by a semismooth Newton method, which converges superlinearly. The time complexity of every iteration depends on the number of measurements since an $m \times m$ linear system is solved in each iteration.

IV. Applications to Jump-Sparse Recovery and Numerical Experiments

In this section, we apply the inverse Potts ADMM (Algorithm 1) to the reconstruction of jump-sparse signals from blurred, noisy data. We consider both reconstruction from Fourier data and deconvolution under Gaussian, Laplacian or impulsive noise. (We refer to the Appendix for a formal description of the noise models.) We compare the results with the minimizers of the *total variation (TV)* problem given by

$$\gamma\|\nabla u\|_1 + \|Au - f\|_p^p \to \min. \qquad (13)$$

For the solution of this convex problem, we use the primal-dual method of [14] with 10 000 iterations.

The experiments were conducted on an Apple MacBook Pro, with Intel Core 2 Duo 2.66 GHz and 8 GB RAM. Typical runtimes are between 1 and 5 seconds for the one-dimensional experiments, and between 5 and 10 minutes for two dimensions.

A. Deconvolution of Blurred Incomplete Data Contaminated by Gaussian and Non-Gaussian Noise

Here, the measurement matrix A models the convolution with some kernel $h = (h_{-r}, \ldots, h_0, \ldots, h_r)$ of non-vanishing mean. We assume that only m measurements $\{j_1, \ldots, j_m\}$, $m < n$, are given. Hence, A is a reduced $m \times n$ Toeplitz matrix of the form

$$A_{j,k} = \begin{cases} h_{k-j}, & \text{if } |k-j| \le r \\ 0, & \text{else.} \end{cases} \qquad (14)$$

where $j = j_1, \ldots, j_m$, and $k = 1, \ldots, n$. In our experiments, h is a Gaussian convolution kernel of standard deviation 6.

In Fig. 1, data $b = A\overline{x}$ is corrupted by Gaussian, Laplacian and impulsive noise (from top to bottom) and $m = \frac{n}{2}$ random measurements are available. The noise variance is $\sigma = 0.05$ for Gaussian and Laplacian noise; in the impulsive noise case, 30% of the convolved signal is set to a random value between 0 and 1 (uniformly distributed). For data contaminated by Gaussian noise we use the L^2 data term, and for the other cases the L^1 data term. In the experiment (Fig. 1) we observe that the inverse Potts ADMM algorithm performs as well as the total variation for impulsive noise. For Gaussian and Laplacian noise, the minimizers of the total variation problem have additional plateaus as well as transitional points between the plateaus. In contrast, the iPotts-ADMM algorithm almost perfectly recovers the jump-sparse signal, and, in particular, the correct number of jumps.

B. Reconstruction of Jump-Sparse Signal From Noisy and Incomplete Fourier Spectrum

We measure an incomplete set of m frequency components of a jump-sparse signal $\overline{x} \in \mathbb{R}^n$. Hence, our measurement matrix is a reduced $(m \times n)$ Fourier matrix of the form

$$A_{j,k} = \frac{1}{\sqrt{n}} e^{-2\pi i j k/n}$$

where $k = 1, \ldots, n$ and j belongs to a set of m indices between 1 and n. Such reconstruction problems have been considered for example in [55]–[57]. Here, we measure every second frequency component, i.e., $j = 2, 4, \ldots, n$. We further assume that the complex valued Fourier data is corrupted by additive noise, i.e.,

$$b = A\overline{x} + \eta_\sigma + i\eta'_\sigma$$

where $\eta_\sigma, \eta'_\sigma$ are m-dimensional vectors of i.i.d. Gaussian random variables of variance σ.

In Fig. 2, we compare the performance of the inverse Potts algorithm (Algorithm 1) with that of TV minimization (13). We see that our method yields significantly higher peak signal-to-noise-ratios (PSNR) than minimizers of the total variation problem. The PSNR is given by

$$\mathrm{PSNR}(x) = 10\log_{10}\left(n\frac{\|\overline{x}\|_\infty^2}{\|\overline{x} - x\|_2^2}\right) \qquad (15)$$

where $\overline{x}$ denotes the groundtruth. We further observe that minimizers of the total variation problem have small variations within the plateaus and underestimate the jump heights ("contrast reduction"). The proposed inverse Potts ADMM algorithm reconstructs the original signal almost perfectly.

C. Reconstruction and Segmentation of Blurred Images

We use the inverse Potts functional in two-dimensions for the reconstruction of cartoon-like, i.e., piecewise constant, images. Such images serve as models in many applications, for instance in computed tomography [58]. In Fig. 3, we reconstruct a cartoon-like image from blurred and noisy data. Our approach recovers the piecewise constant image up to rounding off the corners.

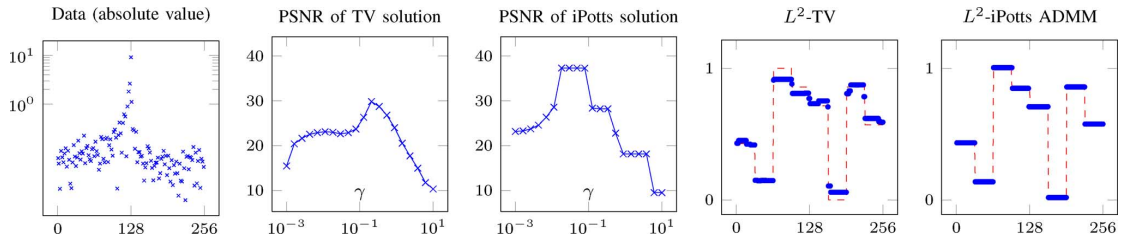


Fig. 2. Reconstruction of a jump-sparse signal using only every second frequency of the Fourier spectrum. Data is corrupted by Gaussian noise ($\sigma = 0.05$). The peak signal-to-noise-ratio of inverse Potts reconstructions are significantly higher than those of the TV reconstructions. The two plots on the righthand side show the reconstruction results corresponding to the optimal regularization parameter with respect to the PSNR ($\gamma = 0.21$ for TV and $\gamma = 0.02$ for iPotts).

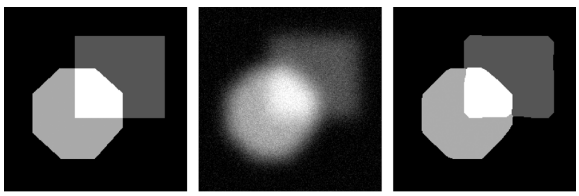


Fig. 3. Deconvolution of a geometric image (256 × 256 pixels), convolved by a Gaussian kernel of standard deviation 12 and corrupted by Gaussian noise ($\sigma = 0.05$). Our method nicely removes the blur. The result is in particular piecewise constant as the original image. (a) Original. (b) Data. (c) L^2-iPotts ADMM.

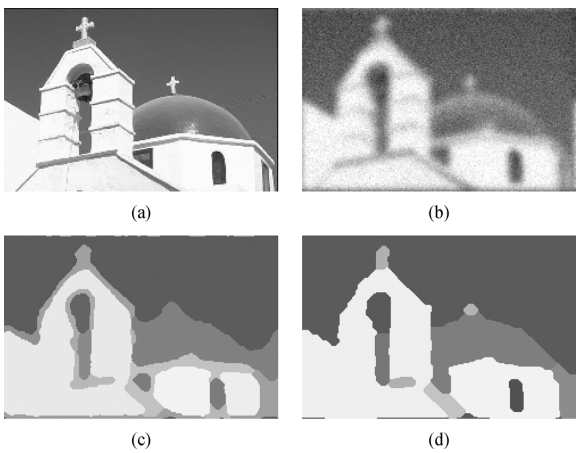


Fig. 4. Segmentation using the classical (9) and the inverse Potts functional (1) of a blurred and noisy natural image (size 241 × 161, image source [59]). Due to the blurring, the classical Potts segmentation exhibits additional segments at the boundaries, whereas segmentation with our inverse Potts ADMM detects sharp boundaries. (a) Original. (b) Data. (c) Classical L^2-Potts. (d) L^2-iPotts ADMM.

For natural images, the Potts functional is classically used for (multi-label) segmentation [2], [3]. (The Potts problem is sometimes called the *piecewise constant Mumford-Shah problem*.) We see in Fig. 4 that the inverse Potts functional (1), which incorporates the blurring operator A, performs better than the classical Potts functional (9) for this task. Here, we segment a blurred and noisy image using the inverse and the classical Potts functional. Due to the blurring, the segmentation using the classical Potts model introduces extra segments at the boundaries. Minimizing the inverse Potts problem, in contrast, detects sharp boundaries without producing additional boundary segments.

V. Applications to Sparse Recovery and Numerical Experiments

Theorem 4 asserts that solutions of the inverse Potts problem associated with $A\nabla$ yield solutions of the sparsity problem

$$S_\gamma(x) = \gamma\|x\|_0 + \|Ax - b\|_p^p \to \min.$$

Thus, we may apply the inverse Potts ADMM (Algorithm 1) to the sparsity problem. The corresponding method is depicted in Algorithm 2.

Algorithm 2: iPotts-ADMM for the sparsity problem

Input: Data $b \in \mathbb{R}^m$, model parameter $\gamma > 0$, measurement matrix $A \in \mathbb{R}^{m\times n}$

Output: Computed result $x \in \mathbb{R}^n$ of the sparsity problem **(2)**

begin

$y \leftarrow$ Solution of iPotts-ADMM (algorithm 1) with matrix $A\nabla$, data f, and model parameter γ;

$x \leftarrow \nabla y$;

end

We compare our method (Algorithm 2) with the following approaches to sparse recovery problems, which include the state-of-the-art methods.

- *Basis pursuit denoising (BPDN)* is the convex optimization problem

$$\gamma\,\|x\|_1 + \|Ax - b\|_p^p \to \min.$$

For the experiments, we use the toolbox YALL1 [46].

- *Iteratively reweighted ℓ^1 minimization (IRL1)* [5] solves a sequence of constrained optimization problems

$$\|x\|_{1,w} \to \min, \quad \text{s.t. } \|Ax - b\|_2 \le \delta, \tag{16}$$

where $\|x\|_{1,w} = \sum_i w_i|x_i|$ is a weighted ℓ^1 norm. The weights are initialized by $w_i = 1$ and are updated depending on the solution of the previous iteration by $w_i = \frac{1}{\epsilon + x_i}$. We perform five iterations and choose $\epsilon = 10^{-3}$. We use the toolbox YALL1 [46] for the minimization of (16).

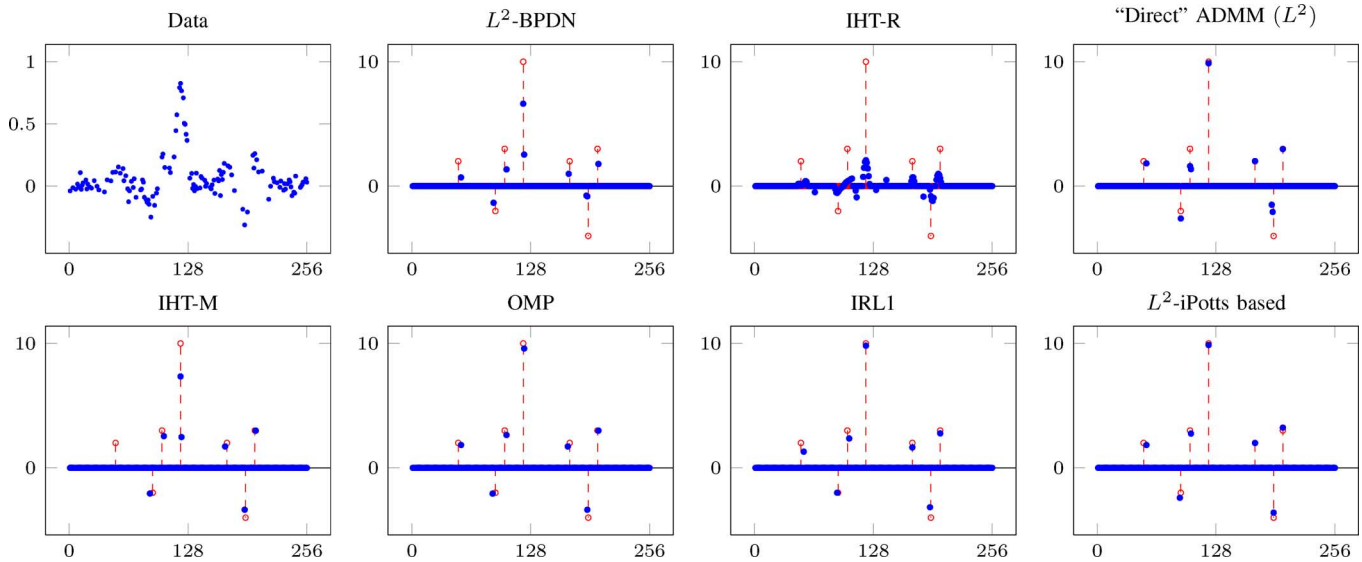


Fig. 5. The original signal (dashed stem plot) is blurred by a Gaussian kernel of standard deviation 5 and corrupted by Gaussian noise of standard deviation $\sigma = 0.05$. We took $m = \frac{n}{2} = 128$ measurements. Orthogonal matching pursuit (OMP), iteratively reweighted ℓ^1 minimization (IRL1) and our iPotts based approach have the best reconstruction quality (with respect to visual inspection), followed by iterative hard thresholding (IHT-M) and the "direct" splitting (cf. Section V-B). Basis pursuit denoising (BPDN) underestimates the heights of the spikes and the Lagrangian variant of iterative hard thresholding (IHT-R) reconstructs too many non-zero entries.

- *Orthogonal matching pursuit (OMP)* [45] greedily searches for minimizers of the constrained formulation of the L^2 sparsity problem

$$\min \|Ax - b\|_2^2, \quad \text{s.t. } \|x\|_0 \leq k.$$

  We use the implementation OMP.m of Stephen Becker available at Matlab's file exchange.
- *Iterative hard thresholding* [23] uses surrogate functionals (forward backward splitting) for the sparsity problem. We here use the two variants hard_l0_reg.m *(IHT-R)* and hard_l0_Mterm.m *(IHT-M)* of the toolbox sparsify 0.5.
- An ADMM method based on a "direct" splitting of (2) which we explain in Section V-B.

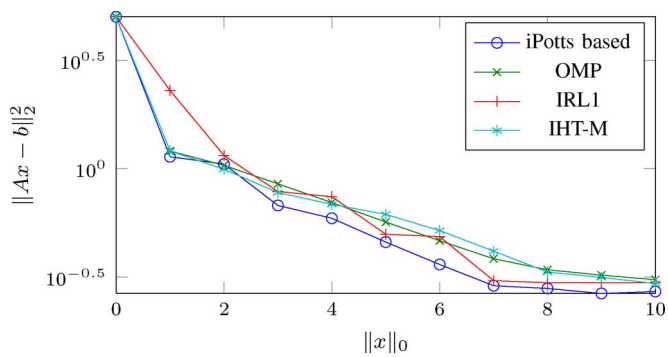


Fig. 6. Approximation error in dependence of the number of non-zero entries of solutions computed by state-of-the-art algorithms. Our iPotts-based method (Algorithm 2) yields lower approximation errors for any number of jumps. Here, the average values of 100 experiments are depicted.

A. Reconstruction of Noisy and Blurred Sparse Signals

Our goal is to reconstruct sparse signals from noisy, blurred and incomplete measurements. We model this reconstruction task by (2) where A is a reduced Toeplitz matrix. In our experiments, data is blurred by a Gaussian kernel and $m = \frac{n}{2}$ measurements are taken. Thus, we are in the setup of Section IV-A except that now the underlying signal is sparse instead of jump-sparse.

Our first example is the reconstruction of blurred and incomplete data under Gaussian noise (Fig. 5). The noise distribution suggests to employ the L^2 data penalty. In the experiment, basis pursuit denoising (BPDN) underestimates the height of the spikes, the Lagrangian variant of iterative hard thresholding (IHT-R) reconstructs too many non-zero entries and the "direct" splitting (Section V-B) has to many additional non-zero entries. Orthogonal matching pursuit (OMP), iteratively reweighted ℓ^1 minimization (IRL1), hard thresholding (IHT-M) and the proposed iPotts-ADMM based approach approximate the original signal quite well; in particular, they reconstruct the precise number of non-zero entries. Towards a deeper comparison of these four algorithms we quantify the reconstruction quality by looking at the average approximation error $\|Ax - f\|_2^2$ in dependence on the number of non-zero entries $\|x\|_0$ of a solution x; cf. Fig. 6. Here, the average values of a series of 100 runs is depicted where we used the setup of the experiment in Fig. 5. We observe that the iPotts based solutions have the least approximation errors in average.

In Fig. 7, we drive the same experiment as in Fig. 5 replacing Gaussian noise by impulsive noise. Due to this noise model, we employ the L^1 data term for our iPotts-based algorithm. For the other methods we also use the L^1 variant whenever it is available; to the best of our knowledge, this is the case for basis pursuit denoising and the direct splitting (Section V-B). We observe

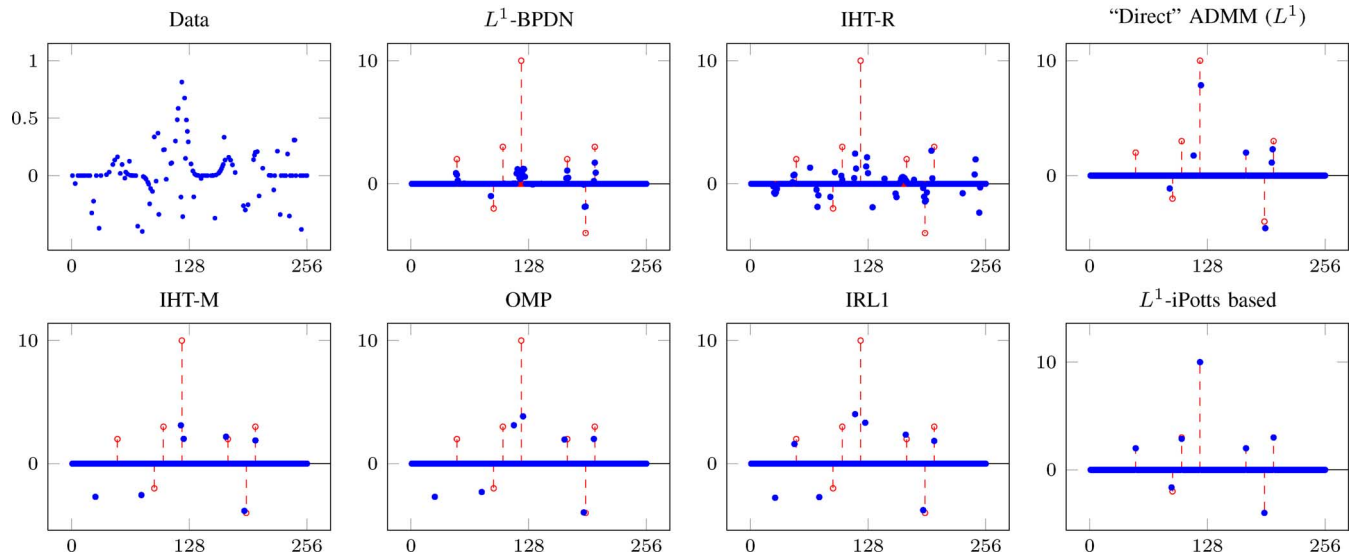


Fig. 7. The same setup as in Fig. 5 replacing Gaussian noise by impulsive noise (25% of data were set to a random value between -0.5 and 0.5). For the direct splitting (cf. Section V-B), basis pursuit (BPDN), and the iPotts based method we use L^1 data terms. We see that the proposed L^1-iPotts based algorithm performs significantly better than the other methods. It is able to recover the original signal almost perfectly.

that the proposed algorithm yields an almost perfect reconstruction also in presence of impulsive noise and that it performs significantly better than the other methods in this case.

B. Comparison With a "Direct" ADMM Approach to the Sparsity Problem

In analogy to (5), we consider the consensus form of the sparsity problem

$$\gamma\|u\|_0 + \|Av - b\|_p^p \to \min, \quad \text{s.t. } u - v = 0. \tag{17}$$

This leads to the augmented Lagrangian

$$\gamma\|u\|_0 + \langle \lambda, u - v\rangle + \frac{\mu}{2}\|u - v\|_2^2 + \|Av - b\|_p^p \to \min. \tag{18}$$

Proceeding as in Section III we obtain a "direct" ADMM algorithm for the sparsity problem. This algorithm is given by replacing $\|\nabla u\|_0$ by $\|u\|_0$ in the first line of (8). This leads to alternately solving a hard thresholding problem (instead of a Potts problem) and a classical Tikhonov problem associated with matrix A.

The difference between the "direct" ADMM approach and our iPotts-ADMM based method (Algorithm 2) is that they are based on different augmented Lagrangians. Indeed, when applying the iPotts-ADMM to the sparsity problem, we consider the inverse Potts problem associated with $A\nabla$ instead of A. Then, the augmented Lagrangian of the corresponding problem is obtained by replacing A by $A\nabla$ in (6). With the substitutions $\nabla u = u'$ and $\nabla v = v'$, (6) reads

$$\gamma\|u'\|_0 + \langle \lambda, \nabla^+(u' - v')\rangle + \frac{\mu}{2}\|\nabla^+(u' - v')\|_2^2 + \|Av' - b\|_p^p \to \min. \tag{19}$$

Comparing (18) and (19), we see that the direct method couples u and v directly whereas the iPotts based method involves the antiderivatives of u and v.

From the experiments (Figs. 5, 7, 8) we conclude that the iPotts-based method (Algorithm 2) is advantageous over the direct ADMM. In particular, the solutions of the iPotts-based method have lower energy than the "direct" method for the whole range of parameters γ; cf. Fig. 8.

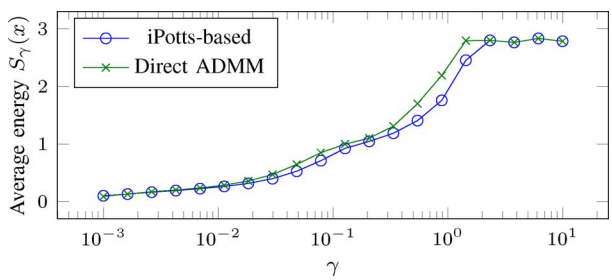


Fig. 8. Final total energy $S_\gamma(x)$ of the sparsity problem using the iPotts based method (Algorithm 2) and the "direct" ADMM method (Section V-B) for different parameters γ. The iPotts-ADMM algorithm reaches lower energies over the whole parameter range. This indicates the superiority of the first method. Computed values are averages over 50 experiments.

C. Sparse Image Recovery

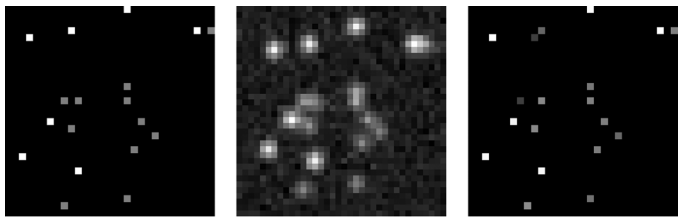
We also use our method to reconstruct sparse images. One may think of an image of small particles or of an astronomic image. We apply our procedure to images by reshaping the image to a vector and adapting the matrix A accordingly. Fig. 9 shows the deconvolution of a sparse image using our iPotts-ADMM based method (Algorithm 2). In the experiment we see that almost all spikes are recovered while only few false positives are reconstructed.

VI. Proofs

Here we provide the proofs of the theorems stated in the course of this paper.

A. Existence of Minimizers

We start out showing Theorem 1 which asserts that the inverse Potts problem (1) has a minimizer.



(a) (b) (c)

Fig. 9. Sparse image of size 30×30 blurred by a 7×7 Gaussian kernel (standard deviation 1) and corrupted by Gaussian noise ($\sigma = 0.05$). Data consists of 50% randomly selected pixels of the middle image. The iPotts based method recovers almost all spikes correctly. (For visualization purposes, the contrast of the middle image was increased.). (a) Original. (b) Blurred, noisy data. (c) iPotts based reconstruction.

Proof of Theorem 1:

In order to deal with the general case of a (possibly) singular matrix A we decompose the domain into $\ker A$ and a corresponding algebraic complement U. This means that $U + \ker A = \mathbb{R}^n$ (or $\mathbb{C}^n$) and $U \cap \ker A = \{0\}$. (In the following we proceed without drawing attention to $\mathbb{C}^n$ when writing $\mathbb{R}^n$, but the arguments work for the complex case as well.) For $x \in \mathbb{R}^n$ we frequently use the decomposition $x = u + v$, where $u = Q_U x$ is the projection Q_U of x to U, and v is the corresponding projection onto $\ker A$.

The matrix A restricted to the subspace U is invertible, and since we are in finite dimensional space, there is a positive constant c such that

$$\|Au\| \geq c\|u\| \quad \text{for any} \quad u \in U.$$

(Due to the finite dimension all norms are equivalent and the above inequality holds for any norm.) As a consequence, whenever, for a sequence u_k in U, the norm $\|u_k\|$ tends to ∞, the inverse Potts functional $P_\gamma(u_k)$ defined by (1) tends to ∞ as well. Therefore, for any sequence of vectors x_k in $\mathbb{R}^n$ (not only in U,) we obtain the implication:

$$P_\gamma(x_k) \text{ is bounded} \\ \Longrightarrow u_k = Q_U x_k \text{ has a converging subsequence.} \quad (20)$$

This is a consequence of $Ax_k = AQ_U x_k$.

Our next preparatory step introduces the mapping s on U which assigns to each $u \in U$ the minimal number of jumps of all vectors in $u + \ker A$, i.e.,

$$s(u) = \min_{v \in u + \ker A} \|\nabla v\|_0.$$

We show that this mapping s is lower semicontinuous which, in our context, means that the preimages of the sets $\{0, \ldots, k\}$ are closed for all $k \in \mathbb{N}$. To see this, we first observe that the set M_k of all vectors in $\mathbb{R}^n$ with at most k jumps is structurally a finite union of vector spaces (of dimension $k+1$.) More precisely,

$$M_k = \{x \in \mathbb{R}^n : \|\nabla x\|_0 \leq k\} = \bigcup_{J \subset \{1,\ldots,n-1\}, |J| = k} X_J,$$

where X_J are those vectors whose jump sets are contained in $J \subset \{1, \ldots, n-1\}$. Furthermore, a vector $u \in U$ has the property $s(u) \leq k$ if and only if there is a vector $x \in M_k$ (i.e., with at most k jumps) such that $Q_U x = u$. Summing up,

$$s^{-1}(\{0, \ldots, k\}) = Q_U(M_k) = \bigcup_{J \subset \{1,\ldots,n-1\}, |J| = k} Q_U(X_J). \quad (21)$$

We discuss the right hand side of (21) to see the lower semicontinuity of s. Each $Q_U(X_J)$ is a finite dimensional linear subspace and thus closed; so as a finite union of closed sets the right hand side of (21) is closed. Therefore the left hand side of (21) is closed which by definition implies the lower semicontinuity of s.

Now we can show the assertion of the theorem. We consider a sequence x_k such that the values $P_\gamma(x_k)$ of the inverse Potts functional P_γ tend to an infimum, i.e.,

$$\lim_{k \to \infty} P_\gamma(x_k) = \inf_{x \in \mathbb{R}^n} P_\gamma(x).$$

For every member of the sequence, we write $x_k = u_k + v_k$ with $u_k \in U$ and $v_k \in \ker A$. By (20) we find a subsequence x_{k_l} such that $u_{k_l} = Q_U x_{k_l}$ converges to some $u \in U$. Since $P_\gamma(x_{k_l})$ converges and $Au_{k_l} = Ax_{k_l}$ we have that

$$|\|\nabla x_{k_l}\|_0 - \|\nabla x_{k_r}\|_0| \leq |P_\gamma(x_{k_l}) - P_\gamma(x_{k_r})| \\ + |\|Au_{k_l} - b\|_p^p - \|Au_{k_r} - b\|_p^p| \to 0 \text{ as } l, r \to \infty.$$

This means that, for sufficiently large l, the number of jumps $\|\nabla x_{k_l}\|_0$ becomes constant; let us denote this constant by j. As a consequence, $s(u_{k_l}) \leq \|\nabla x_{k_l}\|_0 = j$, and thus, by the lower semicontinuity of s, $s(u) \leq j$. Hence, by the definition of s, there is a vector $x^* \in u + \ker A$ such that the number of jumps of x^* is smaller than or equal to j. Then,

$$P_\gamma(x^*) = \gamma\|\nabla x^*\|_0 + \|Ax^* - b\|_p^p = \gamma\|\nabla x^*\|_0 + \|Au - b\|_p^p \\ \leq j + \lim_l \|Au_{k_l} - b\|_p^p = \lim_l P_\gamma(x_{k_l})$$

which shows that x^* is a minimizer as desired. ■

Next we show Theorem 2 which states that the continuous-time analogue of Theorem 1 is wrong. We give counterexamples, i.e., we find bounded operators A and data f such that the continuous-time inverse Potts functional

$$P_\gamma(u) = \gamma \cdot \|\nabla u\|_0 + \|Au - f\|_p^p,$$

if u is a piecewise constant function on the interval $[0, 1]$, and $P_\gamma(u) = \infty$ otherwise, has no minimizer.

Proof of Theorem 2: We consider a positive function $g \in L^p[0, 1]$ with total mass 1 which is supported in the interval $\left[\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon\right]$ with positive $\varepsilon < \frac{1}{8}$. We use the symbol $\widetilde{g}$ for its left-shift by $\frac{1}{2}$. Our counterexamples are the (cyclic) convolution operators with functions $\widetilde{g}$ as above, i.e., operators A defined by $Au = \widetilde{g} * u$, and the data given by $f = g$.

We claim that, for Potts parameter γ with $\gamma < \gamma_0$ (defined in (23) below,)

$$\inf_v P_\gamma(v) = 2\gamma < P_\gamma(u) \quad \text{for all } u. \quad (22)$$

This means that there is no minimizer in that case and thus shows the assertion of the theorem. In order to show the equality in (22), we consider the sequence of characteristic functions $u_n = \frac{n}{2} 1_{\left[\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}\right]}$. We have that $\|\nabla u_n\|_0 = 2$ and $\|Au_n -$

$f\| \to 0$. Thus, $P_\gamma(u_n) \to 2\gamma$. This yields $\inf_v P_\gamma(v) \leq 2\gamma$. It remains to show the inequality in (22) (which in turn implies the equality in (22).) To this end, we have to consider the set of functions u with at most one jump and find $\gamma > 0$ such that $d(u) = \|Au - f\|_p^p > 2\gamma$ for all such u. The set $B = \{x \in [0,1] : f(x) \geq 2\}$ has positive Lebesgue measure $\lambda(B)$ since f has total mass 1 and is supported on an interval of length bounded by $\frac{1}{4}$. If $u < 1$, then $d(u) \geq \lambda(B)$. So in order to obtain $d(u) < \lambda(B)$, we need that $u \geq 1$ either to the left or to the right of its (sole) jump location. Then $Au \geq 1$ on at least one of the intervals $\left[\varepsilon, \frac{1}{2} - \varepsilon\right]$ and $\left[\frac{1}{2} + \varepsilon, 1 - \varepsilon\right]$. Both of these intervals have length $\frac{1}{2} - 2\varepsilon$, and, on both intervals, $f = 0$. Therefore, if $d(u) < \lambda(B)$, we necessarily have $d(u) \geq \frac{1}{2} - 2\varepsilon > \frac{1}{4}$. Then, for any u with at most one jump,

$$d(u) > \min\left(\lambda(B), \frac{1}{4}\right) =: 2\gamma_0. \tag{23}$$

If u has two or more jumps then trivially $P_\gamma(u) > 2\gamma$. Together, this implies that, for any γ with $\gamma < \gamma_0$, the inverse Potts functional P_γ fulfills $P_\gamma(u) > 2\gamma$ for all $u \in L^p[0,1]$. This shows (22) which completes the proof. ■

B. Relations to Sparsity

We first prove Theorem 4 which shows how to transform a sparsity problem into a jump-sparsity problem.

Proof of Theorem 4: For x^* satisfying (3), we define $u^* = \nabla x^*$. Towards a contradiction we assume that there is $u \in \mathbb{R}^n$ such that $\gamma\|u\|_0 + \|Au - b\|_p^p < \gamma\|u^*\|_0 + \|Au^* - b\|_p^p$. Then, for u, there is $x \in \mathbb{R}^{n+1}$ such that $u = \nabla x$. Then,

$$\begin{aligned}\gamma\|\nabla x\|_0 + \|A\nabla x - b\|_p^p &< \gamma\|u^*\|_0 + \|Au^* - b\|_p^p \\ &= \gamma\|\nabla x^*\|_0 + \|A\nabla x^* - b\|_p^p.\end{aligned}$$

which is a contradiction. ■

For $p = 2$ we show a converse statement. It is formulated as Theorem 5 and proved next. In its proof we make use of the decomposition of $\mathbb{R}^n$ into the orthogonal direct sum $\mathbb{R}^n = V \oplus \mathbb{R}e$, where e denotes the constant vector $(1, \ldots, 1)^T$ and V is the linear space of vectors with zero mean. Observing that the linear operator ∇ is bijective from the linear space V to $\mathbb{R}^{n-1}$, we use the symbol ∇^+ for the mapping $\mathbb{R}^{n-1} \to V$,

$$\nabla^+ = (\nabla|_V)^{-1}, \tag{24}$$

for the inverse of the mapping ∇ restricted to the subspace V.

Proof of Theorem 5: We consider the inverse Potts functional given by (1) for $p = 2$. We decompose $x \in \mathbb{R}^n$ according to $x = x_0 + \overline{x}$, with $x_0 \in V, \overline{x} \in \mathbb{R}e$. Applying this decomposition to (1) yields

$$\begin{aligned}P_\gamma(x) &= \gamma\|\nabla(x_0 + \overline{x})\|_0 + \|Ax_0 + A\overline{x} - b\|_2^2 \\ &= \gamma\|\nabla x_0\|_0 + \|Ax_0 + A\overline{x} - b\|_2^2.\end{aligned}$$

We write $\overline{x} = \mu e$ to obtain

$$P_\gamma(x) = \gamma\|\nabla x_0\|_0 + \|Ax_0 + \mu Ae - b\|_2^2. \tag{25}$$

Let us fix x_0 for the moment and let us look for $\mu = \mu(x_0)$ which minimizes the function $\mu \to P_\gamma(x_0 + \mu e)$. Since $\|\nabla(x_0 + \mu e)\|_0 = \|\nabla(x_0 + \mu' e)\|_0$ for all μ, μ' we have to minimize (w.r.t. μ)

$$\sum_{i=1}^m \left(\mu \sum_{j=1}^n A_{ij} + \sum_{j=1}^n A_{ij} x_{0,j} - b_i\right)^2 \to \min. \tag{26}$$

The corresponding minimizer $\mu(x_0)$ can be computed explicitly (e.g., by derivating). It is given by

$$\mu(x_0) = \frac{\sum_{i=1}^m \widetilde{A_i} b_i - \sum_{i=1}^m \widetilde{A_i} \sum_{j=1}^n A_{ij} x_{0,j}}{\sum_{i=1}^m \widetilde{A_i}^2}, \tag{27}$$

where $\widetilde{A_i}$ is the sum of the i^{th} row of the matrix A given by (30). In particular, $\mu(x_0)$ depends affine linearly on x_0, i.e., is of the form $d - Ex_0$ where d is a constant and E is a row vector of length n, both not depending on x_0. Plugging the expression (27) for $\mu(x_0)$ into (25), we obtain a minimization problem in x_0. It is given by

$$\gamma\|\nabla x_0\|_0 + \|A' x_0 - b'\|_2^2 \to \min, \tag{28}$$

where A' is the matrix given by

$$A'_{kj} = A_{kj} - \frac{\widetilde{A_k} \sum_{i=1}^m \widetilde{A_i} A_{ij}}{\sum_{i=1}^m \widetilde{A_i}^2}, \tag{29}$$

with

$$\widetilde{A_i} := \sum_{j=1}^n A_{ij} \tag{30}$$

and b' is the vector given by

$$b'_k = b_k - \frac{\widetilde{A_k} \sum_{i=1}^m \widetilde{A_i} b_i}{\sum_{i=1}^m \widetilde{A_i}^2}. \tag{31}$$

After these preparations we show the theorem; we consider a minimizer u^* of the sparsity problem (4) w.r.t. the matrix $B = A'\nabla$ and data b'. The crucial point is that ∇ is an isomorphism from V onto $\mathbb{R}^{n-1}$ which implies the equivalence

$$u^* \text{ minimizes } (4) \;\Leftrightarrow x_0^* = \nabla^+ u^* \text{ minimizes } (28).$$

Applying this equivalence, $x_0^* = \nabla^+ u^*$ is a minimizer of (28), and, using (27), the vector $x^* = x_0^* + \mu(x_0^*)$ is a minimizer of the original Potts problem (1) for A, b. ■

Using the relation between inverse Potts and sparsity problems we are now able to show the complexity statement Theorem 3 which asserts NP-hardness of the inverse Potts problem.

Proof of Theorem 3: The sparsity problem (2) is NP-hard by [32, Theorem 3] ($p \geq 1, \gamma > 0$.) According to Theorem 4 each instance of the sparsity problem (2) defines an instance of the inverse Potts problem (1). In particular, for any NP-hard instance of the sparsity problem (with matrix A and data b) there is a corresponding inverse Potts problem (with matrix $A\nabla$ and

data b.) The transformation of the functionals and the transformation of the corresponding minimizers given by Theorem 4 can obviously be done in polynomial time. Therefore the Potts problem is NP-hard. ∎

C. Convergence

In our presentation we have assumed that the sequence μ_k is a geometric progression. What we actually need is that μ_k is a non-decreasing sequence fulfilling

$$\sum_k \frac{1}{\sqrt{\mu_k}} < \infty \tag{32}$$

which is obviously satisfied for geometric progressions. So we show Theorem 6 assuming (32) instead.

Proof of Theorem 6: We consider the Potts ADMM iteration for u^k, v^k and λ^k given by (8). We show that

$$(u^k, v^k) \to (u^*, v^*) \text{ with } u^* = v^*, \quad \text{and} \quad \frac{\lambda^k}{\mu_k} \to 0, \tag{33}$$

which is a qualitative version of the assertion of the theorem.

We denote the functional occurring in the first line of (8) by F_k, i.e.,

$$F_k(u) = \gamma\|\nabla u\|_0 + \frac{\mu_k}{2}\left\|u - \left(v^k - \frac{\lambda^k}{\mu_k}\right)\right\|_2^2.$$

Using this notation, the first line of (8) reads $u^{k+1} \in \arg\min_u F_k(u)$. In order to estimate $\left\|u^{k+1} - \left(v^k - \frac{\lambda^k}{\mu_k}\right)\right\|_2$ we observe that $F_k(u^{k+1}) \le F_k\left(v^k - \frac{\lambda^k}{\mu_k}\right)$ which is a consequence of the minimality of u^{k+1}. Using the definition of F_k yields

$$\gamma\|\nabla u^{k+1}\|_0 + \frac{\mu_k}{2}\left\|u^{k+1} - \left(v^k - \frac{\lambda^k}{\mu_k}\right)\right\|_2^2 \le \gamma\left\|\nabla\left(v^k - \frac{\lambda^k}{\mu_k}\right)\right\|_0 \le \gamma n,$$

where n is the length of v_k. Since the first summand on the left hand side is non-negative we get that

$$\left\|u^{k+1} - \left(v^k - \frac{\lambda^k}{\mu_k}\right)\right\|_2^2 \le \frac{\gamma n}{\mu_k}. \tag{34}$$

In particular,

$$\lim_{k\to\infty} u^{k+1} - \left(v^k - \frac{\lambda^k}{\mu_k}\right) = 0. \tag{35}$$

Now we draw our attention to the second line of (8). We denote the corresponding functional by

$$G_k(v) = \|Av - b\|_p^p + \frac{\mu_k}{2}\left\|v - \left(u^{k+1} + \frac{\lambda^k}{\mu_k}\right)\right\|_2^2.$$

The minimality of v^{k+1} implies $G_k(v^{k+1}) \le G_k\left(u^{k+1} + \frac{\lambda^k}{\mu_k}\right)$. We apply the definition of G_k and estimate

$$\begin{aligned}\|Av^{k+1} - b\|_p^p + \frac{\mu_k}{2}\left\|v^{k+1} - \left(u^{k+1} + \frac{\lambda^k}{\mu_k}\right)\right\|_2^2 &\le \left\|A\left(u^{k+1} + \frac{\lambda^k}{\mu_k}\right) - b\right\|_p^p \\ &\le \left\|A\left(u^{k+1} + \frac{\lambda^k}{\mu_k} - v^k\right) + Av^k - b\right\|_p^p \\ &\le \left(\|A\|\left\|u^{k+1} + \frac{\lambda^k}{\mu_k} - v^k\right\|_2 + \|Av^k - b\|_p\right)^p. \end{aligned} \tag{36}$$

Here $\|A\|$ is the norm of A viewed as an operator from ℓ^2 to ℓ^p. We combine the inequalities (36) and (34) in order to obtain that

$$\|Av^{k+1} - b\|_p \le \frac{\|A\|\gamma n}{\mu_k} + \|Av^k - b\|_p.$$

Solving this recursion yields

$$\|Av^{k+1} - b\|_p \le \|A\|\gamma n \sum_{j=1}^{k} \frac{1}{\mu_j} + \|Av^0 - b\|_p,$$

which shows that the sequence $(\|Av^{k+1} - b\|_p)_{k\in\mathbb{N}}$ is bounded. Together with (36) this implies

$$\frac{\mu_k}{2}\left\|v^{k+1} - \left(u^{k+1} + \frac{\lambda^k}{\mu_k}\right)\right\|_2^2 \le \left(\|A\|\left\|u^{k+1} + \frac{\lambda^k}{\mu_k} - v^k\right\|_2 + C\right)^p,$$

where C is a positive constant independent of k. Using (35) we get that

$$\mu_k\left\|v^{k+1} - \left(u^{k+1} + \frac{\lambda^k}{\mu_k}\right)\right\|_2^2 \text{ is bounded.} \tag{37}$$

We show the convergence of the sequence v^k by showing that it is a Cauchy sequence. To this end we estimate

$$\|v^{k+1} - v^k\| \le \left\|v^{k+1} - u^{k+1} - \frac{\lambda^k}{\mu_k}\right\| + \left\|u^{k+1} + \frac{\lambda^k}{\mu_k} - v^k\right\|.$$

Now we apply (34) and (37) which yield

$$\|v^{k+1} - v^k\| \le \frac{C}{\sqrt{\mu_k}}$$

for some constant $C > 0$ which is independent of k. Assumption (32) on μ_k guarantees that v^k is a Cauchy sequence and hence that v^k converges to some v^*.

We use the third line of (8) to obtain the equality

$$\frac{\lambda^{k+1}}{\mu_{k+1}} = \frac{\mu_k}{\mu_{k+1}}\left(\left(\frac{\lambda^k}{\mu_k} + u^{k+1} - v^k\right) + (v^k - v^{k+1})\right). \tag{38}$$

By (35) and (37) each term in parenthesis converges to 0. Since μ_k is non-decreasing, we have that $\mu_k/\mu_{k+1} \leq 1$ and, therefore, (38) implies that

$$\lim_{k\to\infty} \frac{\lambda^k}{\mu_k} = 0, \quad \text{and} \quad \lim_{k\to\infty} \frac{\lambda^{k+1}}{\mu_k} = 0.$$

We rewrite the third line of (8) as $u^{k+1} - v^{k+1} = (\lambda^{k+1} - \lambda^k)/\mu_k$ to obtain the inequality

$$\|u^{k+1} - v^{k+1}\| \leq \frac{\|\lambda^{k+1}\|}{\mu_k} + \frac{\|\lambda^k\|}{\mu_k} \to 0.$$

This means that $u^k - v^k \to 0$ and, since v_k converges, also u_k converges and the corresponding limit u^* equals v^*. This shows (33) and completes the proof. ■

VII. Conclusion and Outlook

We have shown that the inverse Potts problem has a minimizer in the discrete setting but that the time continuous counterpart does not have minimizers in general. We further have shown that the computation of minimizers is an NP-hard problem. Having accepted that the computation of exact solutions are unfeasible, we have proposed a new approach to the inverse Potts problem based on the alternating direction method of multipliers. In our experiments we have compared the iPotts-ADMM algorithm with total variation minimization for jump-sparse reconstruction. We have observed that our method often performs better than but at least as well as TV minimization. We further have shown that the sparsity problem can be reduced to an inverse Potts problem for $p \geq 1$. The experiments indicate that the iPotts-based approach to the sparsity problem performs as least as well as the state-of-the-art algorithms in presence of Gaussian noise and significantly better in presence of impulsive noise.

Future research aims at faster algorithms for the multivariate inverse Potts problem and at Potts problems with manifold valued data.

Appendix

We consider Gaussian, Laplacian, and impulsive noise. The first two types of noise are additive. Thus the measurement is given by

$$b = A\overline{x} + \eta_\sigma,$$

where η_σ is a m-dimensional vector of i.i.d. random variables of standard deviation σ. In case of Gaussian noise, the probability density function is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

In the case of Laplacian noise, the density is defined by

$$p(x) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}}{\sigma}|x|}.$$

In the case of impulsive noise, we randomly choose a prescribed percentage of indices I between 1 and n and set each data point belonging to that index set to a random number, i.e.,

$$b_i = \begin{cases} (A\overline{x})_i, & \text{if } i \notin I, \\ \xi, & \text{else.} \end{cases}$$

Here, ξ is a random variable which is uniformly distributed in the interval $[0, 1]$ for the jump-sparsity experiments and in the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$ for the sparsity experiments.

References

[1] R. Potts, "Some generalized order-disorder transformations," in *Proc. Math. Cambridge Philos. Soc.*, 1952, vol. 48, pp. 106–109, 01.

[2] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.

[3] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[4] G. Winkler and V. Liebscher, "Smoothers for discontinuous signals," *J. Nonparametric Statist.*, vol. 14, no. 1-2, pp. 203–222, 2002.

[5] E. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ^1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

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

[7] C. Vogel and M. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.

[8] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.

[9] M. Unser and P. Tafti, "Stochastic models for sparse and piecewise-smooth signals," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 989–1006, Mar. 2011.

[10] F. Karahanoglu, I. Bayram, and D. Van De Ville, "A signal processing approach to generalized 1-D total variation," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5265–5274, Nov. 2011.

[11] D. Needell and R. Ward, "Near-optimal compressed sensing guarantees for total variation minimization," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3941–3949, Oct. 2013.

[12] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM J. Image Sci.*, vol. 6, no. 2, pp. 1035–1058, 2013.

[13] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Image Vis.*, vol. 20, no. 1, pp. 89–97, 2004.

[14] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Image Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[15] C. Clason, B. Jin, and K. Kunisch, "A duality-based splitting method for ℓ^1-TV image restoration with automatic regularization parameter choice," *SIAM J. Scientific Comput.*, pp. 1484–1505, 2009.

[16] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[17] M. K. Ng, P. Weiss, and X. Yuan, "Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods," *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2710–2736, 2010.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[19] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Berlin, Germany: Springer-Verlag, 2011, Springer Optimization and Its Applications, pp. 185–212.

[20] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM algorithm for a class of total variation regularized estimation problems," in *Proc. 16th IFAC Symp. Syst. Identification*, 2012, pp. 83–88.

[21] R. Chartrand and B. Wohlberg, "A nonconvex ADMM algorithm for group sparsity with sparse groups," presented at the IEEE Int. Conf. Acoust., Speech, an Signal Process. (ICASSP), Vancouver, BC, Canada, May 2013.

[22] C. Clason, B. Jin, and K. Kunisch, "A semismooth Newton method for l^1 data fitting with automatic choice of regularization parameters and noise calibration," *SIAM J. Image Sci.*, vol. 3, no. 2, pp. 199–231, 2010.

[23] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmonic Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[24] M. Lai and J. Wang, "An unconstrained ℓ_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems," *SIAM J. Optimiz.*, vol. 21, no. 1, pp. 82–101, 2011.
[25] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Berlin, Germany: Springer-Verlag, 2010.
[26] J. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[27] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. New York, NY, USA: Academic, 2008.
[28] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Aug. 2007.
[29] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
[30] E. Candès, "Compressive sampling," in *Proc. on the Int. Congress of Mathematicians*, 2006, pp. 1433–1452.
[31] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[32] X. Chen, D. Ge, Z. Wang, and Y. Ye, "Complexity of unconstrained L^2-L^p minimization," *Math. Programm.*, pp. 1–13, 2012.
[33] X. Chen, "Smoothing methods for nonsmooth, nonconvex minimization," *Math. Programm.*, pp. 1–29, 2012.
[34] M. Fornasier and R. Ward, "Iterative thresholding meets free-discontinuity problems," *Foundat. Comput. Mathemat.*, vol. 10, no. 5, pp. 527–567, 2010.
[35] Y. Sowa, A. Rowe, M. Leake, T. Yakushi, M. Homma, A. Ishijima, and R. Berry, "Direct observation of steps in rotation of the bacterial flagellar motor," *Nature*, vol. 437, no. 7060, pp. 916–919, 2005.
[36] A. Snijders *et al.*, "Assembly of microarrays for genome-wide measurement of DNA copy number by CGH," *Nature Genet.*, vol. 29, pp. 263–264, 2001.
[37] A. L. Drobyshev, C. Machka, M. Horsch, M. Seltmann, V. Liebscher, M. Hrabé de Angelis, and J. Beckers, "Specificity assessment from fractionation experiments (SAFE): A novel method to evaluate microarray probe specificity based on hybridisation stringencies," *Nucleic Acids Res.*, vol. 31, no. 2, p. e1, 2003.
[38] P. Hupé, N. Stransky, J. Thiery, F. Radvanyi, and E. Barillot, "Analysis of array CGH data: From signal ratio to gain and loss of DNA regions," *Bioinformatics*, vol. 20, no. 18, pp. 3413–3422, 2004.
[39] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha, "Advances in single-molecule fluorescence methods for molecular biology," *Annu. Rev. Biochem.*, vol. 77, pp. 51–76, 2008.
[40] R. Schönmeyer, G. Schmidt, S. Meding, A. Walch, and G. Binnig, "Automated co-analysis of MALDI and H&E images of retinal tissue for an improved spatial MALDI resolution," in *Bildverarbeitung für die Medizin (BVM)*, H.-P. Meinzer, T. Deserno, H. Hels, and T. Tolxdorff, Eds. Berlin, Germany: Springer-Verlag, 2013, pp. 217–222.
[41] D. Malioutov, M. Çetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
[42] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.
[43] A. Petukhov, "Fast implementation of orthogonal greedy algorithm for tight wavelet frames," *Signal Process.*, vol. 86, no. 3, pp. 471–479, 2006.
[44] V. Temlyakov, "Nonlinear methods of approximation," *Foundations Comput. Math.*, vol. 3, no. 1, pp. 33–107, 2003.
[45] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[46] J. Yang and Y. Zhang, "Alternating direction algorithms for ℓ^1-problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.
[47] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1985, vol. 17, pp. 137–154.
[48] A. Chambolle, "Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations," *SIAM J. Appl. Math.*, vol. 55, no. 3, pp. 827–863, 1995.
[49] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler, "Complexity penalized M-estimation," *J. Comput. Graphical Statist.*, vol. 17, no. 1, pp. 201–224, 2008.
[50] A. Weinmann, M. Storath, and L. Demaret, "The L^1-Potts Functional for Robust Jump-Sparse Reconstruction," ArXiv, 2013 [Online]. Available: arXiv:1207.4642, submitted for publication
[51] G. Rigaill, "Pruned Dynamic Programming for Optimal Multiple Change-Point Detection," ArXiv, 2010 [Online]. Available: arXiv:1004.0887, to be published
[52] R. Killick, P. Fearnhead, and I. Eckley, "Optimal detection of changepoints with a linear computational cost," *J. Amer. Statist. Assoc.*, vol. 107, no. 500, pp. 1590–1598, 2012.
[53] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
[54] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?," *IEEE Trans. Pattern Anal. Mach. ntell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.
[55] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[56] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
[57] J. Yang, Y. Zhang, and W. Yin, "A fast alternating direction method for TV L1-L2 signal reconstruction from partial Fourier data," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 288–297, Apr.. 2010.
[58] R. Ramlau and W. Ring, "A Mumford-Shah level-set approach for the inversion and segmentation of X-ray tomography data," *J. Comput. Phys.*, vol. 221, no. 2, pp. 539–557, 2007.
[59] 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. Comput. Vi.*, Jul. 2001, vol. 2, pp. 416–423.

Martin Storath (M'10) received his Diploma degree in mathematics in 2008, his Honours degree in technology management in 2009, and his Ph.D. degree *(summa cum laude)* in 2013, all from Technische Universität München.

From 2010 to 2013 he was a research associate at Helmholtz Center Munich. Currently he works as post-doctoral researcher in the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, Switzerland. His research interests include mathematical signal and image processing, especially variational recovery methods and complex wavelet based analysis.

Andreas Weinmann received his Diploma degree in mathematics and computer science from Technische Universität München in 2006 and his Ph.D. degree from Technische Universität Graz in 2010 (both with highest distinction).

Currently he is a research associate at Helmholtz Center Munich and is associated with the mathematics department at TUM. His research interests are applied analysis, approximation theory as well as signal and image processing.

Laurent Demaret received his Diploma degree in Fundamental Mathematics and Applications in 1998 and his Ph.D. degree in Signal Processing in 2002, both from Université de Rennes I. From 2002 to 2004 he was an associate scientist at Technische Universität München.

He is currently team leader of the Microscopic Image Analysis Group at the Helmholtz Center Munich, which he joined in 2004. His research interests include approximation theory, signal processing, and geometrical image representations.