

Chapter 5

Functional Lifting for Variational Problems with Higher-Order Regularization



Benedikt Loewenhausner and Jan Lellmann

Abstract Variational approaches are an established paradigm in the field of image processing. The non-convexity of the functional can be addressed by functional lifting and convex relaxation techniques, which aim to solve a convex approximation of the original energy on a larger space. However, so far these approaches have been limited to first-order, gradient-based regularizers such as the total variation. In this work, we propose a way to extend functional lifting to a second-order regularizer derived from the Laplacian. We prove that it can be represented efficiently and thus allows numerical optimization. We experimentally demonstrate the usefulness on a synthetic convex denoising problem and on synthetic as well as real-world image registration problems.

Introduction and Related Work

In this work, we consider variational energy minimization problems of the form

$$\inf_{u:\Omega\rightarrow\Gamma} \int_{\Omega} \rho(x, u(x)) dx + \lambda S(u), \quad (5.1)$$

for estimating some unknown data u defined on an open, bounded, connected—usually rectangular—image domain $\Omega \subseteq \mathbb{R}^d$ with values in $\Gamma \subseteq \mathbb{R}^n$. The data term in (5.1) is of integral form, with the integrand $\rho(x, u(x))$ typically depending

B. Loewenhausner
Institute of Mathematics and Image Computing (MIC), University of Lübeck, Lübeck, Germany
Technical University of Munich, Garching, Germany
e-mail: benedikt.loewenhausner@in.tum.de

J. Lellmann (✉)
Institute of Mathematics and Image Computing (MIC), University of Lübeck, Lübeck, Germany
e-mail: jan.lellmann@uni-luebeck.de

on some noisy, corrupted measurements. We are particularly interested in the case where ρ is *non-convex* in $u(x)$.

The regularizer S , weighted by a parameter $\lambda > 0$, encodes prior knowledge in order to account for randomness and is often used to resolve ambiguities and render the problem well-posed.

A classical convex example is the Rudin-Osher-Fatemi model with

$$\rho(x, u(x)) := \frac{1}{2}(u(x) - g(x))^2 \quad \text{and} \quad S(u) := \text{TV}(u), \quad (5.2)$$

which can be used to remove noise from a given image $g : \Omega \rightarrow \Gamma$ while preserving discontinuities [35]. The *total variation* $\text{TV}(u)$ is defined as the integral

$$\text{TV}(u) := \int_{\Omega} d\|Du\|, \quad (5.3)$$

where the vector-valued Radon measure Du is used to represent the distributional derivative of u in order to allow for discontinuities [1, 41]. For (weakly) differentiable u , the total variation assumes the simpler form

$$\text{TV}(u) = \int_{\Omega} \|\nabla u(x)\|_2 dx. \quad (5.4)$$

As we will be mainly focused on the discretized setting, we will restrict ourselves to the regular case and use the more suggestive notation (5.4).

In the ROF model, as ρ is convex, computing a global minimizer of (5.1) numerically is feasible even for large problems [4]. However, in many applications, one cannot assume convexity. As a prime example, consider the problem of *image registration* [24], also sometimes referred to as *large-displacement optical flow*: one starts with two images $R, T : \Omega \rightarrow \mathbb{R}$ and aims to find a *deformation*, also called *displacement*, $u : \Omega \rightarrow \mathbb{R}^d$ which is “sufficiently regular” and aligns R and T in the sense that

$$R(x) \approx T(x + u(x)) \quad (5.5)$$

for all $x \in \Omega$. A suitable energy is

$$\frac{1}{2} \int_{\Omega} (R(x) - T(x + u(x)))^2 dx + \lambda S(u). \quad (5.6)$$

This data term is also referred to as sum-of-squares distance (SSD) [25].

Numerically minimizing (5.6) is a challenging problem: not only is the data term generally non-convex, the degree of non-convexity is also completely determined by the data R and T , which are generally noisy and result in an energy landscape with many local minima.

Typical methods for minimizing (5.6) therefore rely on local solvers such as gradient-based and (Quasi-)Newton methods; see [30] for an algorithmic overview. In the context of *optical flow* [16], the classical, and still most common, method is to linearize T around a current estimate, which renders the problem convex. These approaches suffer from the typical issue of local non-convex optimization: the algorithm can get stuck in local minima and requires a good initial estimate. Much work has been dedicated to finding such a starting point, such as “warping” and coarse-to-fine strategies [3].

Non-convexity also appears in much simpler settings, such as q -(pseudo-)norm denoising with energies of the form

$$\int_{\Omega} |u(x) - g(x)|^q dx + \lambda S(u), \quad (5.7)$$

with $q < 1$. This choice of q makes the method more robust against outliers in the data g , as the influence of outliers diminishes as $q \rightarrow 0$. Choosing $q < 1$ also encourages the sparsity of the argument more than convex variants with $q \geq 1$; this is a particularly useful feature in the context of sparse representation [11]. See also [29] for an extensive analysis of non-convex regularization. However, it again renders the data term non-smooth and non-convex. A recent development is to modify methods for non-smooth convex optimization to the non-convex setting [26, 27], however these are again local and convergence results are currently very limited.

Computational and algorithmic advances have recently made another strategy viable: Instead of solving the non-convex problem directly, one aims to approximate it by a—usually much larger—*convex* one, which can be solved to a global optimum. In order to approximate the original problem well, one relies on *functional lifting*, i.e., embedding the original problem into a much larger space: Instead of solving

$$\inf_{u: \Omega \rightarrow \Gamma} f(u), \quad (5.8)$$

one solves the *lifted problem*

$$\inf_{\bar{u}: \Omega \rightarrow \mathcal{P}(\Gamma)} \bar{f}(\bar{u}), \quad (5.9)$$

where $\mathcal{P}(\Gamma)$ is the set of *probability measures* over the range Γ , and \bar{f} is a suitable extension of f on this larger function space in the following sense: with each $u : \Omega \rightarrow \Gamma$, one can associate a function $\bar{u} : \Omega \rightarrow \mathcal{P}(\Gamma)$ which is a Dirac measure at every point, $\bar{u}(x) := \delta_{u(x)}$, and require that $\bar{f}(\bar{u}) = f(u)$ for all u in the original space. On the other hand, if the solution of (5.9) is a Dirac measure $\delta_{u(x)}$ at every point and \bar{f} does not introduce artificial minimizers, then u will be a solution of the original problem (5.8), as each element of the feasible set of (5.8) has a corresponding element in the feasible set of (5.9).

This leaves the question of how to define \bar{f} on arguments \bar{u} that are not Dirac measures, but rather mixtures or even diffuse measures. There is a series of publications discussing different strategies for deriving “good” liftings, starting with image segmentation [20, 32, 38, 39], general convex first-order regularization [33] as a functional-analytic formulation of the classical paper [17], recently advancing the framework to manifold-valued problems [21] and more accurate discretizations [18, 28, 40].

However, all these works assume that the regularizer depends only on the first-order derivative ∇u or its distributional counterpart. For natural images, such first-order regularization is often sub-optimal, as it penalizes linear parts and, in the case of TV, prefers -wise constant solutions.

For natural images, regularizers that use second- and higher-order derivatives have been found to be much more suitable [2, 7, 22, 23, 31, 36]. Therefore one would like to use these more advanced regularizers in the functional lifting framework. However, so far there has been little progress in this direction. The reason is that the space of probability measures $\mathcal{P}(\Gamma)$ is usually discretized as a discrete probability measure on ℓ chosen points in the range Γ . If one follows the same strategy as for lifting TV, one ends up with a large number of constraints on the dual variables, which is at least cubic with respect to ℓ . This requires to choose ℓ very small, which corresponds to a very rough discretization of the range Γ of u and brings the accuracy below acceptable thresholds.

Contributions

In this work, we propose a method for approximating energies of the form (5.1) using functional lifting and convex relaxation, where ρ is a general non-convex data term and the regularizer S incorporates *second-order* information:

- We investigate the non-smooth “Absolute Laplacian” regularizer, which incorporates second-order derivatives and coincides with TV^2 on one-dimensional domains (section “[Lifting for Absolute Laplacian Regularization](#)”).
- After reviewing mathematical prerequisites (section “[Notation and Mathematical Preliminaries](#)”) and the discretized version of the problem, we discuss where the usual strategy for computing a convex extension of the regularizer fails for more involved regularizers (section “[Approximate Relaxation of the Absolute Laplacian](#)”).

We prove that by introducing an approximation step, the number of required constraints can be reduced from cubic (ℓ^3) to linear (ℓ) growth in the one-dimensional case (Theorem 5.1). We propose an extension to the case $d \geq 2$, which—although currently without theoretical guarantees—has been very successful in all of our experiments.

- In order to show that a non-convex data term combined with higher-order regularization has practical benefits, we illustrate the method on a synthetical

q -pseudo-norm denoising example as in (5.7) with second-order regularization (section “[Non-convex Denoising with Second-Order Regularity](#)”).

- We demonstrate the applicability to the non-convex problem of image registration as in (5.6) (section “[Image Registration Using the Absolute Laplacian](#)”).

We conclude with an outlook and notes on further open questions (section “[Conclusion and Outlook](#)”).

Lifting for Absolute Laplacian Regularization

In the following, we will consider a special case of second-order regularization: For $u = (u_1, \dots, u_n) : \Omega \rightarrow \mathbb{R}^n$, we define the *absolute Laplacian* regularizer

$$S_{AL}(u) := \int_{\Omega} \|\Delta u(x)\|_1 dx. \quad (5.10)$$

By convention the Laplacian $\Delta u := (\Delta u_1, \dots, \Delta u_n)^\top$ is vector-valued for $n > 1$.

Similar to the total variation, S_{AL} can be extended to functions with distributional Laplacians as well using a dual formulation; it can also be viewed as the set of functions with a gradient of bounded deformation [37]. Again we will focus on the discretized energy and therefore use the simplified notation (5.10).

The absolute Laplacian regularizer (5.10) has some drawbacks: most importantly, it is not isotropic in the sense that $S_{AL}(u) = S_{AL}(Ru)$ for some rotation matrix $R \in \mathbb{R}^{n \times n}$, and it has a large kernel that includes all harmonic functions. The latter issue was also discussed in detail in [14] for quadratic Laplacian regularization.

It is tempting to substitute a full Hessian regularization such as [9, 15, 23]

$$\int_{\Omega} \left(\sum_{i=1}^n \|\nabla^2 u_i(x)\|_2^2 \right)^{\frac{1}{2}} dx, \quad (5.11)$$

however this couples all components of u , which invalidates the argument used in the proof of Theorem 5.1 below. As of now, we have not found a way for efficiently computing a convex relaxation in the full Hessian-regularized case.

In contrast, the absolute Laplacian (5.10) decouples in the components u_i . Moreover, in the one-dimensional scalar case with $d = 1$ and $n = 1$, it is identical to the second-order total variation [9],

$$S_{AL}(u) = \int_{\Omega} |u''(x)| dx \quad (5.12)$$

or its distributional equivalent.

The absolute Laplacian is motivated by a regularizer that is—in a slightly loose interpretation of the term—known as “curvature” regularization in the medical

image registration community [10] and penalizes the squared Laplacians $\|\Delta u(x)\|_2^2$ instead of $\|\Delta u(x)\|_1$. However, as we will see in the next sections, the 1-homogeneous nature of S_{AL} is crucial in order to accurately lift the regularizer.

Notation and Mathematical Preliminaries

In the following, we detail the discretized lifting approach. We follow the notation in [28]. In order to discretize the probability measures $\mathcal{P}(\Gamma)$, we choose an n -dimensional regular grid of points $\{t_1, \dots, t_\ell\} \subseteq \Gamma$, which are referred to as *labels*. The number of labels in each dimension of the range Γ is denoted by l_k , $k = 1, \dots, n$, and the grid spacing h is assumed to be uniform and constant.

The space $\mathcal{P}(\Gamma)$ is discretized as the unit simplex in \mathbb{R}^ℓ ,

$$\Delta_\ell := \{\bar{p} \in \mathbb{R}^\ell \mid \bar{p} \geq 0, \sum_{i=1}^{\ell} \bar{p}_i = 1\}. \quad (5.13)$$

In a slight abuse of notation, we will from now on denote by \bar{u} a function mapping into the set of *discretized* probability measures, i.e., $\bar{u} : \Omega \rightarrow \Delta_\ell$. The i -th unit vector $e_i \in \Delta_\ell$, $i \in \{1, \dots, \ell\}$, is associated with the Dirac measure δ_{t_i} at label t_i . Rather than associating a general vector $\bar{u}(x) \in \Delta_\ell$ with a weighted sum of Dirac measures as is commonly done, we assign to each vector a *single* Dirac measure $\delta_{u(x)}$, where $u(x) \in \Gamma$ is obtained by linear weighting of the labels:

$$u(x) = \sum_{i=1}^{\ell} \bar{u}_i(x) t_i. \quad (5.14)$$

Whenever (5.14) holds, we refer to $\bar{u}(x) \in \mathbb{R}^\ell$ as a *lifted representation* of $u(x) \in \Gamma$. A function $\bar{u} : \Omega \rightarrow \Delta_\ell$ is called a *lifted representation* of the function $u : \Omega \rightarrow \Gamma$ if (5.14) holds point-wise for all $x \in \Omega$.

Approximate Relaxation of the Absolute Laplacian

In order to illustrate the basic process of constructing an energy function for the lifted representation, first consider the data term in integral form:

$$F(u) := \int_{\Omega} \rho(x, u(x)) dx. \quad (5.15)$$

We discretize $\mathcal{P}(\Gamma)$ as in the previous section, and seek a suitable convex extension of F ,

$$\bar{F}(\bar{u}) := \int_{\Omega} \bar{\rho}(x, \bar{u}(x)) dx, \quad (5.16)$$

to all $\bar{u} : \Omega \rightarrow \Delta_{\ell}$. A classical way [6] is to find the *largest convex* $\bar{\rho} : \Omega \times \Delta^{\ell} \rightarrow \mathbb{R}$ such that

$$\bar{\rho}(x, e_i) = \rho(x, t_i), \quad i = 1, \dots, \ell. \quad (5.17)$$

In order to do so for some fixed x , one first defines a function

$$\phi(p) := \begin{cases} \rho(x, t_i), & \text{if } p = e_i, \\ +\infty, & \text{otherwise,} \end{cases} \quad (5.18)$$

and sets $\bar{\rho}(x, p) := \phi^{**}(p)$, where ϕ^{**} is the Legendre-Fenchel biconjugate [34]. More precisely,

$$\phi^*(f) := \sup_p \{\langle p, f \rangle - \phi(p)\} = \max_{i \in \{1, \dots, \ell\}} \{\langle e_i, f \rangle - \rho(x, t^i)\}, \quad (5.19)$$

$$\phi^{**}(p) := \sup_f \{\langle p, f \rangle - \phi^*(f)\}. \quad (5.20)$$

As can be seen from (5.19), even for integrands ρ that depend only on a single value $u(x)$, the conjugate is generally composed of ℓ pieces. Using common first-order solvers, this incurs a cost of ℓ dual or auxiliary variables per point.

For the regularizer, this issue is much worse: Assume $\Omega \subseteq \mathbb{R}$, then the Laplacian of u at a point x is simply the second derivative and commonly discretized as

$$u''(x) \approx (u(x - \eta) - 2u(x) + u(x + \eta))/\eta^2, \quad (5.21)$$

which depends on *three* different values of u . A finite difference-based second-order regularizer will therefore be of the form

$$\int_{\Omega} \rho(u(x - \eta), u(x), u(x + \eta)) dx, \quad (5.22)$$

which results in three running indices in (5.19) and thus ℓ^3 terms in the maximum. Even for a very moderate choice of $\ell = 10$, this results in 1000 additional variables per point, which is impractical.

In this section, we therefore consider an approximation of this process for the special case of the absolute Laplacian regularizer (5.10), which only requires *linear*

complexity. We derive the model for the one-dimensional case $d = 1$ and $n = 1$,

$$\int_{\Omega} |u''(x)| dx, \quad (5.23)$$

and subsequently discuss how to generalize it to n -dimensional image domains and vector-valued u .

The first step is to separate computation of the second derivatives from the lifting process, i.e., we also apply the derivative operator to the *lifted* representation \bar{u} and seek a lifted regularizer

$$\int_{\Omega} \bar{\rho}(\bar{u}''(x)) dx \approx \int_{\Omega} \bar{\rho} \left((\bar{u}(x + \eta) - 2\bar{u}(x) + \bar{u}(x - \eta)) / \eta^2 \right) dx, \quad (5.24)$$

where $x \pm \eta$ are the neighboring points of x . For simplicity, we assume $\eta = 1$.

We apply the same process as in (5.18) to $\rho(z) = |z|$ and set

$$\phi(p) = \begin{cases} |\mu| \cdot |t_{i_1} - 2t_{i_0} + t_{i_2}|, & \text{if } p = \mu \cdot (e_{i_1} - 2e_{i_0} + e_{i_2}), \\ +\infty, & \text{otherwise,} \end{cases} \quad (5.25)$$

where $1 \leq i_0, i_1, i_2 \leq \ell$. The free variable $\mu \in \mathbb{R}$ is not required, but ensures that ϕ is positively homogeneous. This implies that the conjugate ϕ^* is an indicator function of some set, which simplifies the later optimization. Taking the convex conjugate, we obtain

$$\phi^*(f) = \delta_{K_{1D}}(f) := \begin{cases} 0, & f \in K_{1D}, \\ +\infty, & \text{otherwise,} \end{cases} \quad (5.26)$$

with the set

$$K_{1D} := \bigcap_{1 \leq i_0, i_1, i_2 \leq \ell} \{f \in \mathbb{R}^{\ell} : f_{i_1} - 2f_{i_0} + f_{i_2} \leq h |i_1 - 2i_0 + i_2|\}. \quad (5.27)$$

This is a straightforward computation following from the definition of the convex conjugate and making use of the 1-homogeneity of ϕ , and using the assumption that the labels t_i are uniformly spaced with distance h . The above formulation consists of ℓ^3 constraints, which would render the problem numerically intractable except for very small ℓ .

A main contribution of this work is the following theorem, which shows that the number of constraints can be reduced to linear order.

Theorem 5.1 *The set K_{1D} in (5.27) with ℓ^3 linear constraints can be equivalently represented by ℓ linear constraints:*

$$K_{1D} = \{f \in \mathbb{R}^\ell : f_2 - f_1 \leq h, \quad f_\ell - f_{\ell-1} \geq -h\} \cap \bigcap_{2 \leq i \leq \ell-1} \{f \in \mathbb{R}^\ell : f_{i-1} - 2f_i + f_{i+1} \leq 0\}. \quad (5.28)$$

Proof Denoting the right-hand side in (5.28) by K_{1D}^{red} , and using the definition of K_{1D} in (5.27), we have to show that $K_{1D} = K_{1D}^{red}$.

$$K_{1D} \subseteq K_{1D}^{red}$$

Assume $f \in K_{1D}$ as in (5.27), i.e., $f_{i_1} - 2f_{i_0} + f_{i_2} \leq h|i_1 - 2i_0 + i_2|$ holds for all triples $i_0, i_1, i_2 \in \{1, \dots, \ell\}$. Choose $i_1 = i_2 = 2$ and $i_0 = 1$, then the first inequality in (5.28) follows. Analogously we obtain the second inequality $f_\ell - f_{\ell-1} \geq -h$ by setting $i_1 = i_2 = \ell - 1$ and $i_0 = \ell$. All other inequalities in (5.28) follow by setting $i_1 = i - 1, i_0 = i, i_2 = i + 1$, therefore $f \in K_{1D}^{red}$.

$$K_{1D} \supseteq K_{1D}^{red}$$

Suppose $f \in K_{1D}^{red}$, i.e., the inequalities in (5.28) hold. We define the vector $a \in \mathbb{R}^{\ell-1}$, $a_i := f_{i+1} - f_i$ as the difference between two consecutive components of f . Using this notation, we reformulate the constraints (5.28) in terms of a :

$$a_1 \leq h, \quad (5.29)$$

$$a_{\ell-1} \geq -h, \quad (5.30)$$

$$a_{i-1} \geq a_i, \quad \forall i \in \{2, \dots, \ell - 1\}. \quad (5.31)$$

Thus the components of a_i form a finite, monotonously non-increasing sequence that is absolutely bounded by h , i.e., $a \in S := \{x \in [-h, +h]^{\ell-1} : x_i \geq x_{i+1}\}$.

If $i_0 = i_1 = i_2$, the inequality in (5.27) holds trivially. Otherwise, if two of the indices agree, then the inequality in (5.27) takes the form

$$f_j - f_k \leq h|j - k|. \quad (5.32)$$

Assuming without loss of generality that $j > k$, this inequality follows from

$$f_j - f_k = a_k + \dots + a_{j-1} \leq |a_k| + \dots + |a_{j-1}| \leq h|j - k| \quad (5.33)$$

due to the observation that all a_i are bounded by $\pm h$.

We are left with the last case of distinct i_0, i_1, i_2 . Without loss of generality assume $i_1 > i_2$, otherwise we swap the symbols.

As all inequalities are invariant with respect to the addition of a constant to f , it suffices to prove the claim for all f with f_1 fixed to some constant. Therefore we can assume $f_1 = 0$. Under this assumption, the linear map between vectors $f \in \mathbb{R}^\ell$

in K_{1D}^{red} and vectors $a \in \mathbb{R}^{\ell-1}$ satisfying (5.29)–(5.31) is bijective. As the vertices of the latter set consist of the vectors of the form $(h, \dots, h, -h, \dots, -h)$, from bijectivity we deduce that the vertices of the set $K_{1D}^{red} \cap \{f \mid f_1 = 0\}$ are the elements satisfying the equality $|f_{i+1} - f_i| = h$ and the inequality $f_{i-1} - 2f_i + f_{i+1} \leq 0$.

Showing that all f satisfying (5.28) are contained in the set in (5.27) is equivalent to showing

$$\max_{f \in K_{1D}^{red} \cap \{f \mid f_1 = 0\}} \{f_{i_1} - 2f_{i_0} + f_{i_2}\} \leq h|i_1 - 2i_0 + i_2|. \quad (5.34)$$

As the maximum problem is a linear program, it assumes its maximum on the set of vertices of $K_{1D}^{red} \cap \{f \mid f_1 = 0\}$. Therefore we only have to show that

$$f_{i_1} - 2f_{i_0} + f_{i_2} \leq h|i_1 - 2i_0 + i_2| \quad (5.35)$$

for all f in the finite set of vertices, i.e., satisfying $|f_{i+1} - f_i| = h$ and the inequality $f_{i+1} - 2f_i + f_{i-1} \leq 0$ (and still $f_1 = 0$). This can be argued case by case:

$$i_0 < i_2 < i_1$$

As the left-hand side in (5.35) can be written as $(f_{i_1} - f_{i_0}) + (f_{i_2} - f_{i_0})$ and due the observation (5.33), the maximum is assumed on the vertex f satisfying $f_{i+1} = f_i + h$ for all i , with maximum value

$$f_{i_1} - 2f_{i_0} + f_{i_2} = h(i_1 - i_0) + h(i_2 - i_0) = h(i_1 - 2i_0 + i_2) = h|i_1 - 2i_0 + i_2|, \quad (5.36)$$

which shows that the inequality in (5.27) holds for this case.

$$i_2 < i_0 < i_1$$

In this case the maximum is assumed if either $f_{i+1} = f_i + h$ or $f_{i+1} = f_i - h$ for all i , depending on which of $i_2 - i_0$ and $i_0 - i_1$ is larger. Therefore

$$f_{i_1} - 2f_{i_0} + f_{i_2} \leq \max\{\pm(h(i_0 - i_2) - h(i_1 - i_0))\} = h|i_1 - 2i_0 + i_2|. \quad (5.37)$$

$$i_2 < i_1 < i_0$$

Again with the observation (5.33), we see that in this case the maximum is attained for $f_{i+1} = f_i - h$ for all i , in which case

$$f_{i_1} - 2f_{i_0} + f_{i_2} = -(f_{i_0} - f_{i_1}) - (f_{i_0} - f_{i_2}) = h(-i_2 + 2i_0 - i_1) = h|i_1 - 2i_0 + i_2|. \quad (5.38)$$

This shows that (5.35) holds for all vertices in the set $K_{1D}^{red} \cap \{f \mid f_1 = 0\}$, and therefore for all points, which concludes the proof of the remaining inclusion $K_{1D}^{red} \subseteq K_{1D}$. \square

Interestingly, in the classical convex relaxation for the (first-order) total variation used in [19, 33], the dual constraint set is of the form

$$K_{\text{TV},1D} = \bigcap_{1 \leq i \leq \ell-1} \{f \in \mathbb{R}^\ell : |f_i - f_{i+1}| \leq h\}. \quad (5.39)$$

As the second intersection in (5.28) enforces $f_{i+1} - f_i \leq f_i - f_{i-1}$, we obtain

$$K_{1D} = K_{\text{TV},1D} \cap \bigcap_{2 \leq i \leq \ell-1} \{f \in \mathbb{R}^\ell : f_{i-1} - 2f_i + f_{i+1} \leq 0\}. \quad (5.40)$$

Thus, when moving from first- to second-order regularization in the proposed way, the only addition is an extra non-positivity constraint on the second derivative of the dual variable f .

So far we have only considered the case of a one-dimensional domain Ω . In order to generalize the construction in (5.25) to $d > 1$ dimensions, we replace the one-dimensional three-point stencil by the corresponding Laplacian stencil in higher dimensions:

$$\phi(p) = \begin{cases} |\mu| \cdot \left| \sum_{j=1}^d (i_{1,j} - 2i_0 + i_{2,j}) \right|, & \text{if } p = \mu \cdot \sum_{j=1}^d (e_{i_{1,j}} - 2e_{i_0} + e_{i_{2,j}}), \\ +\infty, & \text{otherwise,} \end{cases} \quad (5.41)$$

where $i_{1,j}$ and $i_{2,j}$ are the indices of the neighboring points of i_0 in the j -th spatial direction. The convex conjugate can be computed in a similar fashion as in the one-dimensional case:

$$\phi^*(f) = \delta_K(f) \quad (5.42)$$

with the set

$$K := \bigcap_{1 \leq i_0, i_{1,1}, i_{2,1}, \dots, \leq \ell} \left\{ f \in \mathbb{R}^\ell : \sum_{j=1}^d (f_{i_{1,j}} - 2f_{i_0} + f_{i_{2,j}}) \leq h \left| \sum_{k=1}^d (i_{1,k} - 2i_0 + i_{2,k}) \right| \right\}. \quad (5.43)$$

Taken all together, the lifted absolute Laplacian regularizer for scalar-valued images in a d -dimensional image domain becomes

$$\bar{S}_{AL,s}(\bar{u}) := \int_{\Omega} \sup_{f \in K} \langle \Delta \bar{u}(x), f \rangle dx. \quad (5.44)$$

In order to approximate the absolute Laplacian for lifted *vector*-valued functions $u = (u_1, \dots, u_n)$, we apply (5.44) to the marginal distributions $\bar{u}^{(k)}(x) :=$

$\Pi_k \bar{u}(x) \in \Delta^{l_k}$ separately in each component $k \in \{1, \dots, n\}$, where

$$\Pi_k := \underbrace{(1, \dots, 1)}_{l_1 \cdot l_2 \cdot \dots \cdot l_{k-1} \text{ ones}} \otimes \text{Id}_{l_k} \otimes \underbrace{(1, \dots, 1)}_{l_{k+1} \cdot l_{k+2} \cdot \dots \cdot l_n \text{ ones}} \in \mathbb{R}^{l_k \times \ell} \quad (5.45)$$

computes the k -th marginal distribution by summing the entries of \bar{u} over all dimensions of the range with the exception of the k -th dimension. As the absolute Laplacian regularizer decouples in the components of u , it can be approximated by summing the one-dimensional regularizer of the marginalized label distribution over the label dimensions:

$$\bar{S}_{AL}(\bar{u}) := \sum_{i=1}^n \bar{S}_{AL,s}(\Pi_i \bar{u}) = \sum_{i=1}^n \int_{\Omega} \sup_{f^i \in K_{l_i}} \langle \Delta \Pi_i \bar{u}(x), f^i(x) \rangle dx. \quad (5.46)$$

Here $K_{l_i} \subseteq \mathbb{R}^{l_i}$ denotes a set of the form (5.43) in l_i -dimensional space, which accounts for the fact that there may be a different number of labels in each dimension of the range.

After discretizing the image domain $\Omega \subseteq \mathbb{R}^d$ on a d -dimensional Cartesian grid Ω' , the full discretized problem can be formulated in saddle point form:

$$\inf_{\bar{u}: \Omega' \rightarrow \Delta^\ell} \sup_{f^i: \Omega' \rightarrow K_{l_i}, i=1, \dots, n} \sum_{x \in \Omega'} \bar{\rho}(x, \bar{u}(x)) + \lambda \sum_{x \in \Omega'} \sum_{i=1}^n \langle \Delta \Pi_i \bar{u}(x), f^i(x) \rangle. \quad (5.47)$$

This problem can be readily solved using any available primal-dual method for non-smooth convex optimization.

We do not know of a result similar to Theorem 5.1 yet in order to reduce the number of constraints for the sets K_{l_i} in a similar way as for K_{1D} . Therefore, we take a pragmatic approach: we approximate each of the sets K_{l_i} by the set K_{1D} in the corresponding dimension, which amounts to an outer approximation of K_{l_i} . We can then apply Theorem 5.1 to solve the problem using the reduced number of constraints.

Experimental Results

We evaluate the proposed strategy for higher-order relaxation of non-convex problems on two applications. Firstly, we consider a non-convex denoising problem, using the MATLAB extension CVX [12, 13] to solve the primal formulation of the saddle-point problem (5.47) on an Intel Core i7-4500U CPU with 8 GB of RAM.

Secondly, we examine a real-world image registration problem, using a CUDA 7.5.17 implementation¹ of a first order primal-dual algorithm with diagonal preconditioning [5] which runs on an Nvidia GeForce GTX 680 GPU with an Intel Core i7 960 CPU and 24 Gb RAM. The implementation uses a more recent “sublabel-accurate” approach for lifting the data term in order to reduce the required resolution for the data term [18, 28].

Non-convex Denoising with Second-Order Regularity

In order to illustrate that non-convexity can be beneficial when combined with second-order regularization, we consider the simple one-dimensional denoising problem

$$\inf_{u: \Omega \rightarrow \mathbb{R}} \int_{\Omega} |u(x) - g(x)|^q dx + \lambda \int_{\Omega} |u''(x)| dx \quad (5.48)$$

with $\Omega \subseteq \mathbb{R}$. For $q = 1$, one obtains a simple *convex* $\text{TV}^2 - L^1$ denoising model, while for $q < 1$, the energy is generally non-convex. We used the proposed method to approximate a global solution of (5.48).

The method was applied to a smooth input signal g distorted by heavy salt-and-pepper noise, with 80% of the values randomly set to 0 or 1. The locations of the outliers were unknown to the solver, and no additional preprocessing or outlier masking was performed.

As can be seen from Figs. 5.1 and 5.2, combining higher-order regularization with a non-convex data term allows to reconstruct the signal more faithfully. While both approaches prefer piecewise linear results as expected from the function-space formulation, in the convex approach with $q = 1$, input noise is carried over into the output before the structure is fully visible.

While convex methods relying on L^1 data terms are often—rightfully—referred to as “robust” methods in comparison to methods using smooth or quadratic data terms, the non-convex approach with $q = 0.1$ is even more robust against outliers and returns a decent reconstruction for a range of λ on this challenging problem. Run times were in the order of 0.3 s for a discretization of Ω using 120 grid points and $\ell = 63$ labels.

¹See [28] and <http://github.com/tum-vision/prost> for the most recent version.

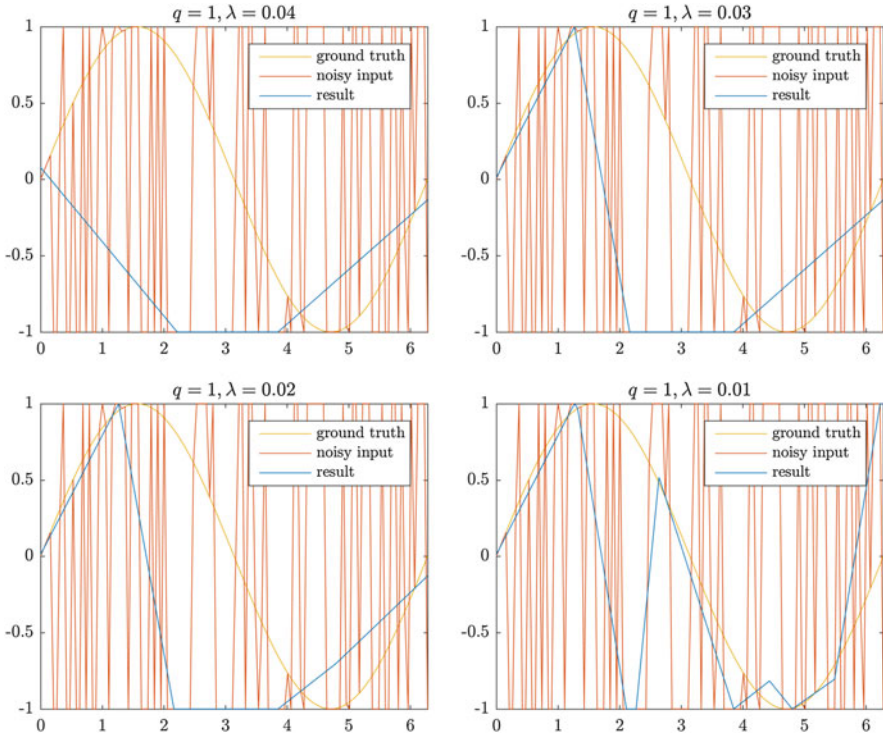


Fig. 5.1 Classical convex second-order ($TV^2 - L^1$) denoising of a smooth signal corrupted by 80% blind salt-and-pepper noise using varying regularization strength λ . The result is piecewise affine as expected from TV^2 regularization. Starting from large λ with heavy over-regularization and decreasing λ , noise is picked up early. There is no regimen where both noise is removed and the signal reconstructed faithfully

Image Registration Using the Absolute Laplacian

For a more challenging application, we apply the method to the image registration problem with SSD data term (5.6) and absolute Laplacian regularization.

Translation-Only Synthetic Image

We first apply the absolute Laplacian regularizer to a synthetic binary image registration problem. The input reference image R is a binary 64×64 image of two vertical boxes. The template image T is obtained by translating the input image by 12 pixels (Fig. 5.3, first row). Thus the ground truth is a uniform translation by 12 pixels and constitutes a global minimizer, as it has vanishing data term and the second-order regularizer does not penalize linear deformations. This configuration

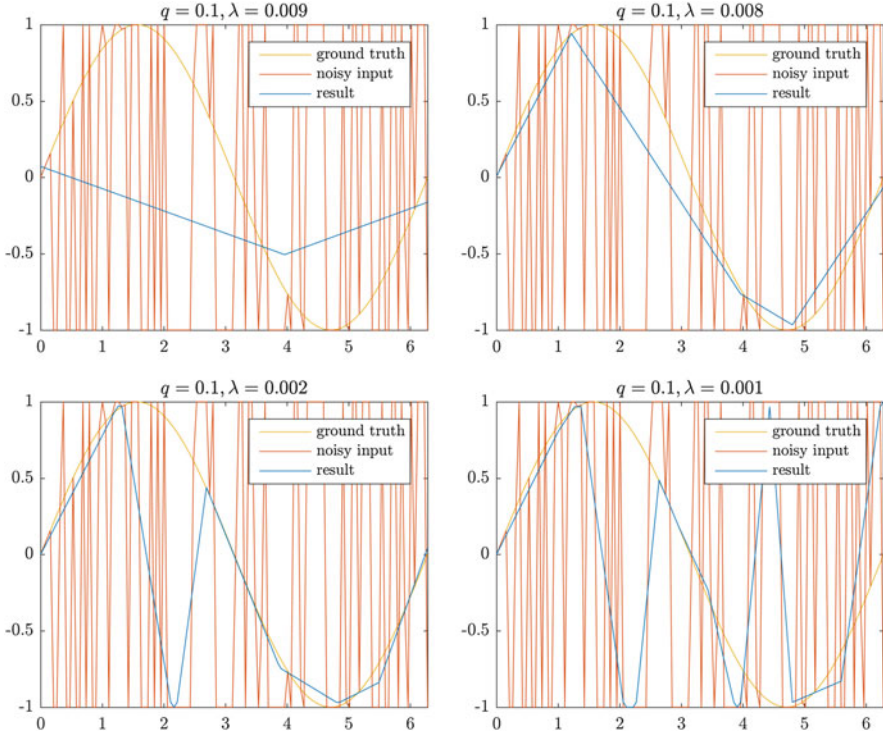


Fig. 5.2 *Non-convex* second-order ($\text{TV}^2 - L^q$ with $q = 0.1$) denoising of the signal in Fig. 5.1 using the proposed convex lifting and approximation with varying regularization strength λ . The proposed approach allows to approximate *global* minimizers of such higher-order regularized non-convex models. The additional non-convexity achieves a better reconstruction of the signal (**top right**) than in the convex case (Fig. 5.1) before giving in to noise (**bottom left**)

is challenging for methods based on local optimization, as there is a strong local minimum. Furthermore, as the images contain large constant regions, the energy landscape has extensive flat regions with zero gradient.

We compare our approach to a traditional curvature-regularized model solved using a single-resolution local minimization method implemented in the MATLAB extension FAIR [24, 25]. The regularization strength was manually set to $\lambda = 10$, however a wide range of values for λ produced the same qualitative behavior. The traditional approach leads to a solution that is not globally optimal (Fig. 5.3, second row). Using our approach, we retrieve the globally optimal ground truth with $\ell = 9$ labels in the label space $\Gamma = [-12, 12]^2$ and a run time of 85 s, without having to resort to approaches such as coarse-to-fine or affine pre-registration for initialization (Fig. 5.3, bottom row).

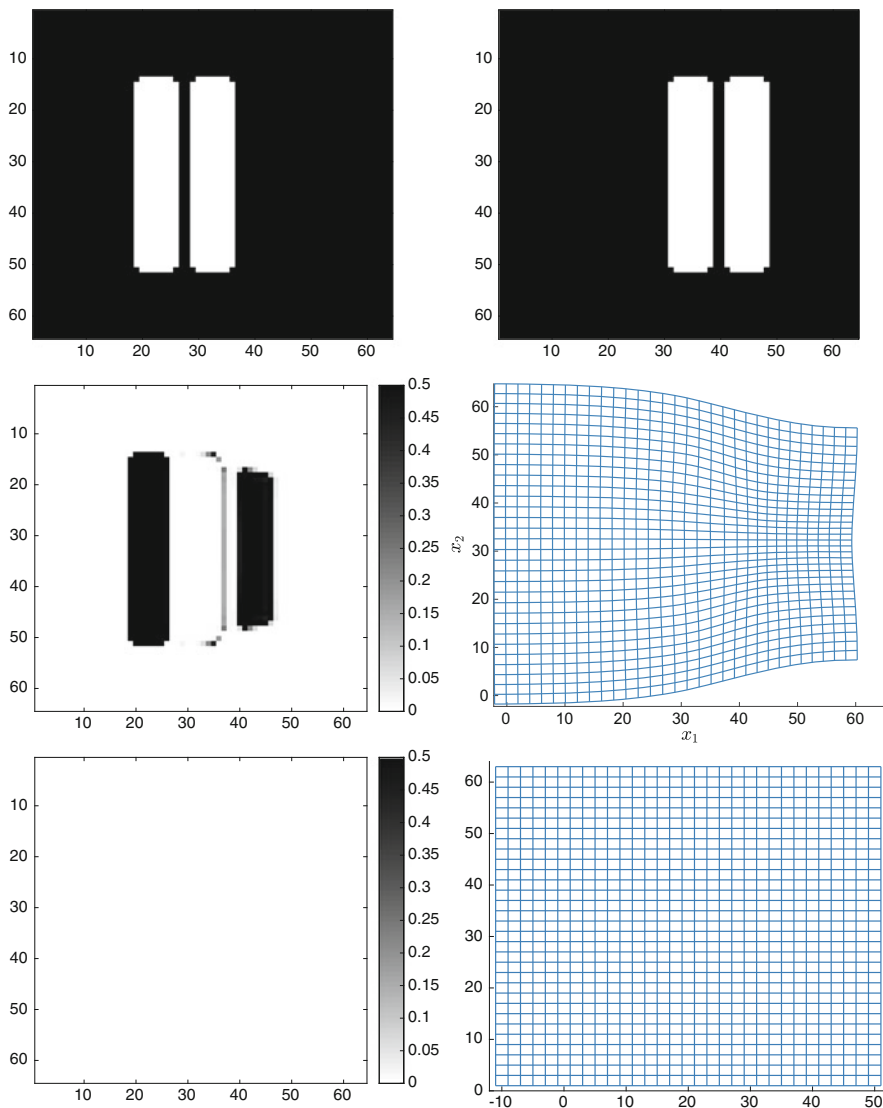


Fig. 5.3 Application of the proposed lifting for absolute Laplacian regularization to a synthetic image registration problem. A traditional curvature-regularized model solved using a local Gauss-Newton method serves as a baseline. The input reference image R (top left) and template image T (top right) differ by a ground truth translation of 12 pixels. The second and third row show the final difference images $\frac{1}{2}(R(x) - T(x + u(x)))^2$ (left) and obtained deformation u visualized as a deformation grid (right). The classical local optimization method (second row) converges to a local solution which is not globally optimal and yields a non-constant deformation with a mean displacement of 2.3 pixels. Using the proposed functional lifting for absolute Laplacian regularization (bottom row), the global optimum is retrieved accurately with an average displacement of 12.0002 pixels

Real-World Image Registration

As a real-world example, we employ the SSD energy with absolute Laplacian regularization to solve the image registration problem on a pair of X-ray images, and compare to the existing lifting approach [18] with total variation regularization. The regularization strength was manually set to $\lambda = 0.05$. Run times were 933 s for total variation, and 515 s for absolute Laplacian minimization.

As can be seen from the numerical results (Fig. 5.4), while the first-order total variation regularization achieves a very good data fit, it results in a physically implausible self-intersecting deformation grid (Fig. 5.4, second row). This behavior can be partly attributed to the well-known fact that total variation promotes piecewise constant solutions, also commonly referred to as stair-casing effect [8]. In the context of medical image registration, this is a highly undesired behavior, as jumps in the deformation map u correspond to infinite stretch or compression and often lead to self-intersections. In contrast, the proposed second-order regularizer (Fig. 5.4, bottom row) maintains a physically meaningful deformation, while still achieving an acceptable data fit.

Conclusion and Outlook

In this work, we have taken a first step towards extending the convex relaxation and functional lifting framework to second-order regularization. We showed how to solve the main issue of an exploding number of constraints for the absolute Laplacian regularization.

Experiments on a denoising problem showed that the combination of higher-order regularization and non-convex data terms can lead to better results than a convex model, and allows to recover highly corrupted data in a piecewise linear fashion. In the application of image registration, the absolute Laplacian faithfully retrieves simple translations and leads to a more realistic deformation grid than total variation regularization on a real-world problem.

While our relaxation allows to reduce the number of required constraints to linear complexity, it is an approximation, rather than a “tight” relaxation in the sense of an exact biconjugate, and the proof is still limited to one dimension. An open question is whether one can find a similar compact representation for the tight relaxation in more than one dimension.

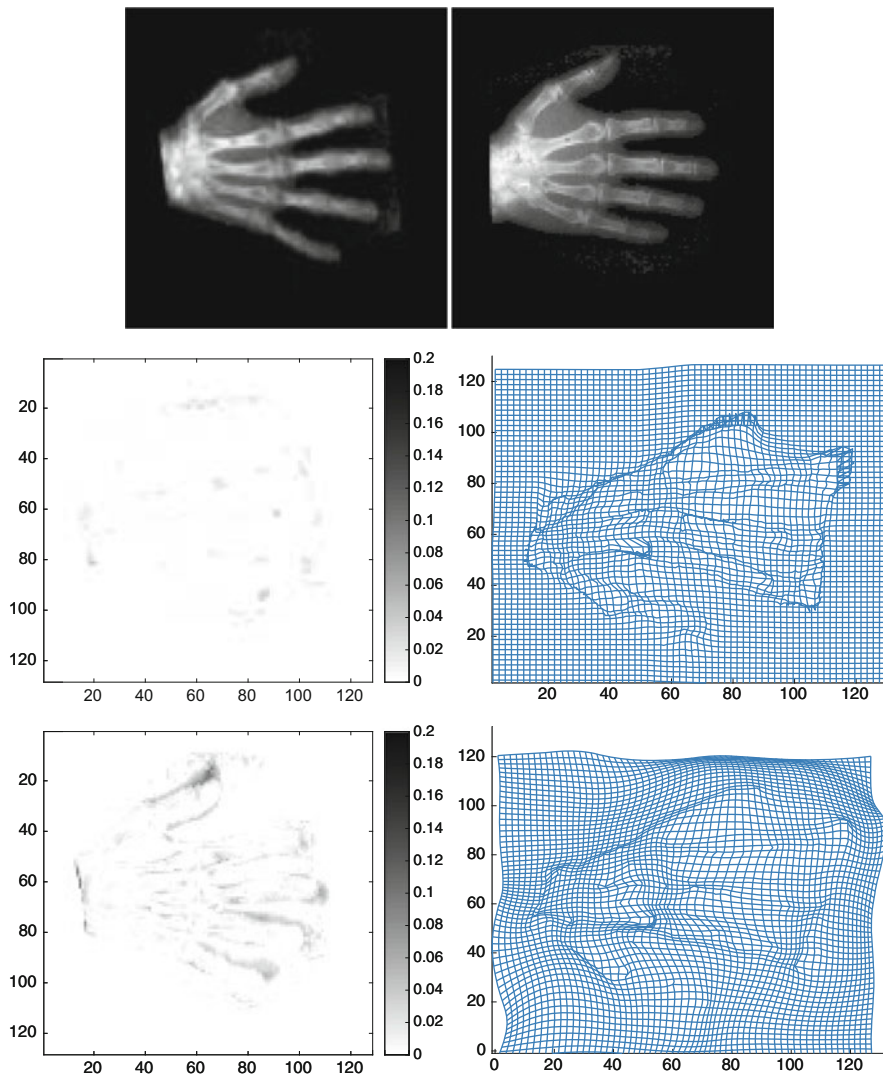


Fig. 5.4 Comparison between two global optimization methods for medical image registration: classical first-order total variation regularization [18] and the proposed second-order lifting approach. The input data consists of a pair of 128×128 grayscale X-ray images of two right hands (**top row**). Both approaches are evaluated using $\ell = 10^2 = 100$ labels, $\Gamma = [-12, 12]^2$, $\Omega = [0, 128]^2$, and a regularization strength of $\lambda = 0.05$. The classical first-order total variation regularization generates piecewise constant deformations and a physically implausible self-intersecting deformation grid (**second row**). The second-order regularizer avoids discontinuities and maintains a physically meaningful deformation grid (**bottom row**)

Finally, in this work we have constrained ourselves to the discretized setting. A functional-analytic discussion as well as an extension to the more recent manifold-valued and sublabel-accurate relaxations remain subject of future work.

Acknowledgements This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—394737018 “Functional Lifting 2.0: Efficient Convexifications for Imaging and Vision”. We would like to thank Emanuel Laude and Thomas Möllenhoff for providing their library `prost`, which was used to solve the saddle-point formulation of the problems.

References

1. L. Ambrosio, N. Fusco, D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems* (Clarendon Press, Oxford, 2000)
2. K. Bredies, K. Kunisch, T. Pock, Total generalized variation. *J. Imag. Sci.* **3**(3), 294–526 (2010)
3. T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping, in *European Conference on Computer Vision* (Springer, Berlin, 2004), pp. 25–36
4. A. Chambolle, An algorithm for total variation minimization and applications. *J. Math. Imag. Vis.* **20**, 89–97 (2004)
5. A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.* **40**(1), 120–145 (2011)
6. A. Chambolle, D. Cremers, T. Pock, A convex approach to minimal partitions. *J. Imag. Sci.* **5**(4), 1113–1158 (2012)
7. T. Chan, A. Marquina, P. Mulet, Higher-order total variation-based image restoration. *J. Sci. Comput.* **22**(2), 503–516 (2000)
8. T. Chan, S. Esedoğlu, F. Park, A. Yip, Total variation image restoration: overview and recent developments, in *The Handbook of Mathematical Models in Computer Vision* (Springer, Berlin, 2005)
9. F. Demengel, Fonctions à Hessien borné. *Ann. Inst. Fourier* **34**, 155–190 (1985)
10. B. Fischer, J. Modersitzki, Curvature based image registration. *J. Math. Imag. Vis.* **18**(1), 81–85 (2003)
11. W.J. Fu, Penalized regressions: the bridge versus the lasso. *J. Comput. Graph. Stat.* **7**(3), 397–416 (1998)
12. M. Grant, S. Boyd, Graph implementations for nonsmooth convex programs, in *Recent Advances in Learning and Control* (Springer, Berlin, 2008), pp. 95–110
13. M. Grant, S. Boyd, CVX: matlab software for disciplined convex programming (2014). <http://cvxr.com/cvx>
14. S. Henn, A full curvature based algorithm for image registration. *J. Math. Imag. Vis.* **24**(2), 195–208 (2006)
15. W. Hinterberger, O. Scherzer, Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing* **76**(1), 109–133 (2006)
16. B.K.P. Horn, B.G. Schunck, Determining optical flow. *Artif. Intell.* **17**, 185–203 (1981)
17. H. Ishikawa, Exact optimization for Markov random fields with convex priors. *Pattern. Anal. Mach. Intell.* **25**(10), 1333–1336 (2003)
18. E. Laude, T. Möllenhoff, M. Moeller, J. Lellmann, D. Cremers, Sublabel-accurate convex relaxation of vectorial multilabel energies, in *European Conference on Computer Vision* (Springer, Berlin, 2016), pp. 614–627
19. J. Lellmann, C. Schnörr, Continuous multiclass labeling approaches and algorithms. *SIAM J. Imag. Sci.* **4**(4), 1049–1096 (2011)

20. J. Lellmann, J. Kappes, J. Yuan, F. Becker, C. Schnörr, Convex multi-class image labeling by simplex-constrained total variation, in *Scale Space and Variational Methods in Computer Vision*. Lecture Notes in Computer Science, vol. 5567 (2009), pp. 150–162
21. J. Lellmann, E. Strekalovskiy, S. Koetter, D. Cremers, Total variation regularization for functions with values in a manifold, in *International Conference on Computer Vision* (2013), pp. 2944–2951
22. M. Lysaker, X.C. Tai, Iterative image restoration combining total variation minimization and a second-order functional. *Int. J. Comput. Vis.* **66**(1), 5–18 (2006)
23. M. Lysaker, A. Lundervold, X.C. Tai, Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Trans. Image Process.* **12**(12), 1579–1590 (2003)
24. J. Modersitzki, *Numerical Methods for Image Registration* (Oxford University Press on Demand, Oxford, 2004)
25. J. Modersitzki, *FAIR: Flexible Algorithms for Image Registration* (SIAM, Philadelphia, 2009)
26. T. Möllenhoff, E. Strekalovskiy, M. Moeller, D. Cremers, The primal-dual hybrid gradient method for semiconvex splittings. *SIAM J. Imag. Sci.* **8**(2), 827–857 (2015)
27. T. Möllenhoff, E. Strekalovskiy, M. Möller, D. Cremers, Low rank priors for color image regularization, in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (Springer, Berlin, 2015), pp. 126–140
28. T. Möllenhoff, E. Laude, M. Moeller, J. Lellmann, D. Cremers, Sublabel-accurate relaxation of nonconvex energies, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 3948–3956
29. M. Nikolova, Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *Multiscale Model. Simul.* **4**(3), 960–991 (2005)
30. J. Nocedal, S.J. Wright, *Numerical Optimization* (Springer, Berlin, 2006)
31. K. Papafitsoros, C.B. Schönlieb, A combined first and second order variational approach for image reconstruction. *J. Math. Imag. Vision* **48**(2), 308–338 (2014)
32. T. Pock, A. Chambolle, D. Cremers, H. Bischof, A convex relaxation approach for computing minimal partitions, in *Computer Vision and Pattern Recognition* (2009)
33. T. Pock, D. Cremers, H. Bischof, A. Chambolle, Global solutions of variational models with convex regularization. *J. Imag. Sci.* **3**(4), 1122–1145 (2010)
34. R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970)
35. L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268 (1992)
36. S. Setzer, G. Steidl, T. Teuber, Infimal convolution regularizations with discrete 11-type functionals. *Commun. Math. Sci.* **9**(3), 797–872 (2011)
37. P. Suquet, Existence et régularité des solutions des équations de la plasticité parfaite. *C. R. Acad. Sci. Paris, Ser. A* **286**, 1201–1204 (1978)
38. J. Yuan, E. Bae, X.C. Tai, Y. Boykov, A continuous max-flow approach to Potts model, in *European Conference on Computer Vision* (2010), pp. 379–392
39. C. Zach, D. Gallup, J.M. Frahm, M. Niethammer, Fast global labeling for real-time stereo using multiple plane sweeps, in *Vision, Modeling, and Visualization* (2008)
40. C. Zach, C. Häne, M. Pollefeys, What is optimized in convex relaxations for multilabel problems: connecting discrete and continuously inspired map inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(1), 157–170 (2014)
41. W.P. Ziemer, *Weakly Differentiable Functions* (Springer, Berlin, 1989)