UDC 004.93'1:519.157

J. Lellmann, C. Schnörr

Regularizers for Vector-Valued Data and Labeling Problems in Image Processing

Дан обзор последних результатов в области регуляризаторов, основанных на полных вариациях, применительно к векторным данным. Результаты оказались полезными для хранения или улучшения мультимодальных данных и задач разметки на непрерывной области определения. Возможные регуляризаторы и их свойства рассматриваются в рамках единой модели.

The review of recent developments on total variation-based regularizers is given with the emphasis on vector-valued data. These have been proven to be useful for restoring or enhancing data with multiple channels, and find particular use in relaxation techniques for labeling problems on continuous domains. The possible regularizers and their properties are considered in a unified framework.

Наведено огляд останніх результатів у галузі регуляризаторів, що базуються на повних варіаціях, стосовно векторних даних. Результати виявилися корисними для зберігання та покращення мультимодальних даних і задач розмітки на неперервній області визначення. Можливі регуляризатори та їх властивості розглядаються в рамках єдиної моделі.

Abstract

We review recent developments on total variation-based regularizers, with emphasis on vector-valued data. These have been proven to be useful for restoring or enhancing data with multiple channels, and find particular use in recent advances on relaxation techniques for multiclass labeling problems on continuous domains. Many of the proposed approaches only differ in the norm that is used. We provide a review of the possible regularizers and their properties in a unified framework.

1. Introduction

Total variation-based variational approaches in an image processing exhibit some unique properties. Most prominently, they can be viewed as a number of «stacked» problems on the sublevelsets of a function, which allows to gain insight into the structure of their minimizers, and can be applied to finding global minimizers of segmentation/labeling problems formulated on continuous domains.

This work is intended as an introductory overview of the field. We first discuss the related functionals that are commonly used in image processing (Sect. 2). Although they were originally formulated for scalar-valued total variation regularizers, they can be readily extended to vector-valued data. This occurs in particular when dealing with convex relaxations of segmentation problems.

Recently, many different total variation-based regularizers have been proposed in different contexts. We try to give a systematic account of the existing approaches and their properties (Sect. 3). Although the intention of this work is to provide a broad overview on the subject, some technicalities cannot be avoided and will be defined in the following section.

Preliminaries

In the following, superscripts v^i denote a collection of vectors or matrix columns, while subscripts v_k denote vector components or matrix rows, i.e. we denote, for $A \in \mathbb{R}^{d \times l}$,

$$A = (a^1 | \ldots | a^l) = (a_1 | \ldots | a_d)^{\mathrm{T}}. \qquad (1)$$

We denote by $\Delta_l := \{x \in \mathbb{R}^l \mid x \geq 0, e^{\mathrm{T}} x = 1\}$ the unit simplex in $\mathbb{R}^l$, where $e := (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^l$. The i-th unit vector is denoted by e^i, I_n is the identity matrix in $\mathbb{R}^n$ and $\|\cdot\|_2$ the usual Euclidean norm for vectors resp. the Frobenius norm for matrices. Analogously, the standard inner product $\langle \cdot, \cdot \rangle$ extends to pairs of matrices as the sum over their element-wise product. $\mathcal{B}_r(x)$ denotes the ball of radius r in x, and S^{d-1} the set of $x \in \mathbb{R}^d$ with $\|x\| = 1$. The indicator function $1_{\mathcal{S}}(x)$ of a set $\mathcal{S}$ is defined as $1_{\mathcal{S}}(x) = 1$ iff $x \in \mathcal{S}$ and $1_{\mathcal{S}}(x) = 0$ otherwise. For a convex set $\mathcal{C}$, $\sigma_{\mathcal{C}}(u) := \sup_{v \in \mathcal{C}} \langle u, v \rangle$ is the support function from convex analysis.

For simplicity, we assume that the the image domain $\Omega \subseteq \mathbb{R}^d$ is the open unit box, $\Omega = (0,1)^d$. All results can equally be formulated for bounded open domains with compact Lipschitz boundary. The images or generally data with a spatial domain Ω are represented as vector-valued func-

Keywords: variational image segmentation, image labeling, convex relaxation, nonsmooth convex optimization.

tions $u:\Omega\to\mathbb{R}^l$ which are absolutely integrable, i.e. $u\in L^1(\Omega)^l$.

$C_c^k(\Omega)$ is the space of *k*-times continuously differentiable functions on Ω with compact support. As usual, $\mathcal{L}^d$ denotes the *d*-dimensional Lebesgue measure, while $\mathcal{H}^k$ denotes the *k*-dimensional Hausdorff measure.

We say that $u:\Omega\to\mathbb{R}^l$ is of *bounded variation*, i.e. $u\in BV(\Omega)^l$, if $u\in L^1(\Omega)^l$ and its distributional derivative corresponds to a finite Radon measure; i.e. $u_j\in L^1(\Omega)$ and there exist $\mathbb{R}^d$-valued measures $Du_j=(D_1u_j,\ldots,D_du_j)$ on the Borel subsets $\mathcal{B}(\Omega)$ of Ω such that

$$\sum_{j=1}^{l}\int_\Omega u_j \operatorname{div} v_j dx=-\sum_{j=1}^{l}\sum_{i=1}^{d}\int_\Omega v_j^i dD_i u_j \quad (2)$$
$$\forall v\in (C_c^\infty(\Omega))^{d\times l}.$$

The measures in (2) form the distributional gradient Du, which is again a measure that vanishes on any $\mathcal{H}^{(d-1)}$-negligible set.

The *total variation* of *u* is defined as the measure-theoretic total variation of its distributional gradient [AFP00, Prop. 3.6]: For some (possibly vector-valued) measure μ on the measure space $(X,\mathcal{E})$, the total variation measure of μ on a set $E\subset\mathcal{E}$ is defined as

$$|\mu|(E):=\sup\left\{\sum_{k=0}^{\infty}\mu\|(E_k)\|_2 \,\middle|(E_k)\subseteq \right.$$
$$\left.\subseteq\mathcal{E}\text{ pairwise disjoint, } E=\bigcup_{k=0}^{\infty}E_k\right\}. \quad (3)$$

Consequently, the total variation of $u\in L^1(\Omega)^l$ is defined as

$$TV(u):=|Du|(\Omega)=\int_\Omega 1d\,|Du|. \quad (4)$$

The total variation also has the dual representation

$$TV(u)=\sup_{v\in C_c^\infty(\Omega)^{d\times l},\|v(x)\|\le 1}\int_\Omega\langle u,\operatorname{Div} v\rangle dx, \quad (5)$$
$$\operatorname{Div} v:=\left(\operatorname{div} v_1,\ldots,\operatorname{div} v_l\right)^T.$$

By partial integration, this implies that for continuously differentiable $u\in C^1(\Omega)^l$,

$$TV(u)=\int_\Omega\|\nabla u\|_2 dx, \quad (6)$$

however the definition of the total variation allows *u* to be discontinuous or even piecewise constant. For more general functionals depending on Du, first note that Du can be represented in terms of its density with respect to $|Du|$, i.e. $Du = (Du/|Du|)|Du|$. Since $Du/|Du|$ is actually a $|Du|$-integrable function, the total variation is

$$TV(u)=\int_\Omega\left\|\frac{Du}{|Du|}\right\|_2 d|Du|. \quad (7)$$

Based on this principle it is possible to construct non-standard *TV*-based regularizers: For $\Psi:\mathbb{R}^{d\times l}\to\mathbb{R}_{\ge 0}$ continuous, positively homogeneous and convex, we define the measure

$$\Psi(Du):=\Psi\left(\frac{Du}{|Du|}\right)|Du|. \quad (8)$$

Essentially Ψ generates a new measure from Du by transforming its density with respect to $|Du|$ [AFP00, Thm. 2.38]. Note that by definition Ψ is a seminorm, and a norm if $\Psi(z)=0\Leftrightarrow z=0$.

2. *TV*-Based Functionals in Image Analysis

In this paper, we will be concerned with a particular class of variational problems used in image processing and analysis. We refer to [SGG+09] for a general overview. The output of a variational method is defined as the minimizer

$$u^*:=\arg\min_{u\in\mathcal{C}} f(u), \quad (9)$$

where $\mathcal{C}$ is some subset of a space of functions defined on Ω, and *f* a functional depending on the input data. The interpretation of *u* is governed by the problem to be solved: for the prototypical example of color denoising, $u:\Omega\to[0,1]^3$ could directly describe the colors of the output image on the image domain $\Omega\subseteq\mathbb{R}^d$; while for segmentation problems, $u:\Omega\to[0,1]$ could assign each point to the foreground ($u(x)=1$) or background ($u(x)=0$) class. We will in particular consider the case where u is vector-valued.

Usually the objective *f* is composed of a *data term* $H(u)$ and a *regularizer* $J(u)$,

$$f(u)=H(u)+J(u). \quad (10)$$

The data term strongly depends on the input data – such as color values of a recorded image, depth measurements, or other features – and promotes a good fit of the minimizer to the input data. However, due to the presence of noise in the input data, it is generally necessary to incorporate the additional prior knowledge about the «typical» appearance of the desired output, which is the purpose of the regularizer. We will now consider some important choices for data term and regularizer.

2.1. $L^2 - L^2$: Gaussian Denoising

The central question is how to construct the individual terms, and which combinations lead to good results. The most basic, classical example is Gaussian denoising, where the goal is to find, given an input image $I:\Omega \to \mathbb{R}^l$, some $u:\Omega \to \mathbb{R}^l$ minimizing

$$f(u)=\frac{1}{2}\int_\Omega \| u-I\|_2^2\,dx+\frac{\lambda}{2}\int_\Omega \| Du\|_2^2\,dx \quad (11)$$

for some weighting parameter $\lambda>0$. Usually u is required to be differentiable, nevertheless we use the distributive derivative Du to simplify comparison with the following variants. Note that both the data term and the regularizer exhibit quadratic growth.

The problem is convex, and after the discretization can be solved globally optimal as a linear equation system. While the approach removes Gaussian noise very well, it tends to smear hard edges in the image. This is caused by the quadratic growth of the regularizers, which makes it susceptible to «outliers» – i.e. high gradients – in the form of hard edges.

2.2. $L^2 - TV$: Rudin-Osher-Fatemi

Many approaches have been proposed to circumvent this problem. Most prominently, the anisotropic diffusion approach consists in solving (11) using gradient descent, at each step locally modifying the norm in the regularizer to reduce smoothing across directions where the current iterate has a large gradient, i.e. across potential edges. While this is widely used and gives good results in many cases, the output cannot be characterized in the variational way as the minimizer of a certain functional. A more one-step approach is Rudin-Osher-Fatemi (ROF) denoising [ROF92],

$$f(u)=\frac{1}{2}\int_\Omega \| u-I\|_2^2\,dx+\lambda\int_\Omega \| Du\|_2. \quad (12)$$

The right-hand side should be seen as a shorthand notation for the total variation of u. The scalar-valued case has been extensively studied and works well for removing Gaussian noise while preserving hard edges; also, the overall problem is still convex and therefore can be solved globally optimal. The key difference is that while the data term still has quadratic growth, the regularizer only grows linearly. In this paper, we will be concerned with regularizers that also have this property.

2.3. $L^1 - TV$

For non-Gaussian noise such as salt-and-pepper, (12) is suboptimal, as it is quite sensitive to outliers in the input image I. Also, ROF denoising invariably leads to a reduction in contrast. These drawbacks are addressed by the $L^1 - TV$ model (see [Nik01] for an overview),

$$f(u)=\int_\Omega \| u-I\|_1\,dx+\lambda\int_\Omega \| Du\|_2 dx. \quad (13)$$

Here, both the data term as well as the regularizer exhibit linear growth. Existence and well-posedness of (12) and (13) can be shown in a precise sense within the class of functions of bounded variation [AFP00, AMT91].

The functionals such as (13) are extremely tolerant to noise. The downside is that they also tend to generate a «staircasing» effect on smooth gradients, i.e. the solution tends to be piecewise constant. We refer to [DAG09] and the references therein for a detailed analysis. For image denoising this is certainly not desirable, therefore some effort has been put into reducing staircasing while preserving robustness, mostly by including the higher-order derivatives (see e.g. [CL97, Sch98, CMM00, LT06, BKP10]). It should also be noted that the robustness can be increased even more at the cost of convexity [Nik].

2.4. Linear– TV: Scalar Case

Upon closer inspection there are applications where staircasing – or more specifically piecewise constancy – is exactly what is desired for the output. This was noted in [CEN06] for the case of *geometry denoising*: Here the goal is to find, for some given *set* $T\subseteq\Omega$, a set $S\subseteq\Omega$ that minimizes

$$f(S):=|T\Delta S|+\lambda Per(S), \quad (14)$$

where $|T \Delta S| := |(T \setminus S) \cup (S \setminus T)|$ is the volume of the symmetric difference of the sets T and S, and $Per(S)$ is the *perimeter* of S, $Per(S) := TV(1_S)$, which coincides with the classical length resp. area of the boundary if S is sufficiently smooth. Setting $u := 1_S$ and relaxing $u(x)$ to the whole of R, we arrive at the problem of minimizing

$$f(u) = \int_\Omega |u - 1_T| \, dx + \lambda \int_\Omega \| Du \|_2, \qquad (15)$$

over $u \in BV(\Omega)$, which is a $L^1 - TV$ denoising problem as in (13) with $I = 1_T$. The important aspect is that we departed from the requirement that u must be an indicator function, allowing intermediate values. This leads to an overall convex problem (15) which can be solved globally optimal; however it introduces the possibility that we may obtain *non-discrete* solutions, which do not correspond to an indicator function.

In this case it is very desirable that the minimizer is piecewise constant. Ideally it should be a *discrete* solution, i.e. take only the values 0 or 1. In fact this is often observed in practice. Moreover, in [CEN06] it was noted that *any* non-discrete solution may be turned into a *discrete* one by a simple thresholding (i.e. rounding) of u. This allows to solve *combinatorial* problems globally optimal by solving *convex* problems and post-processing the solutions obtained.

This thresholding property applies to the much wider class «Linear-TV» problems

$$f(u) = \int_\Omega u(x) s(x) dx + \lambda \int_\Omega \| Du \|_2, \qquad (16)$$

with the constraint $u \in BV(\Omega)$. As shown in [CEN06], this allows to solve the Chan-Vese two-class image segmentation problem [CV01]. Here the desired image is piecewise constant, taking only one of two grey values, c_0 or c_1. Each point in the image should be assigned either to the foreground ($u(x) = 1$) or the background ($u(x) = 0$), such that

$$f(u) = \int_\Omega u \left(c_1 - I(x)\right)^2 dx + \\ + \int_\Omega (1-u)\left(c_0 - I(x)\right)^2 dx + \lambda \int_\Omega \| Du \|_2 \qquad (17)$$

is minimized over all indicator functions $u \in BV(\Omega)$. For fixed $c_1, c_2 \in \mathbb{R}$, this is equivalent to a problem of the form (16) for $s = (c_1 - I(x))^2 - (c_0 - I(x))^2$.

In a sense, the problem (16) extends the concept of graph cut-based segmentation – formulated on a finite set of points in the image domain, typically a uniform grid – to continuous image domains. Therefore it is also referred to as a *continuous cut* . For an extensive analysis of a closely related problem and its associated dual problem we refer to [Str83].

An important property is that solutions of the ROF and $L^1 - TV$ problems (12), (13) for scalar-valued u are intimately connected with solutions of associated continuous cut problems of the form (16). This comes from the fact that the scalar total variation can be written in terms of it level sets,

$$TV(u) = \int_{-\infty}^{\infty} TV(1_{\{x \in \Omega | u(x) > \alpha\}}) d\alpha. \qquad (18)$$

This *coarea formula* [FR60] intuitively connects the problem of finding $u : \Omega \to R$ with a series of problems of finding *indicator functions* for problems of the type (16), and can be exploited in two directions: On the one hand, it allows to solve a family of parametrized continuous cut problems by solving a single ROF-type problem and thresholding [CD09, Ber09, SKO09]. On the other hand, ROF- and $L^1 - TV$ type problems can be solved by solving a series of continuous cuts to find the super-level sets $\{x \in \Omega \,|\, u(x) > \alpha\}$ for all α, and from these reconstructing u [Hoc01, DS06a, DS06b, GY07, CD09, DAG09].

While in theory this would require to solve infinitely many problems, in practice the range of values of u is often quantized, such as {0, 1, …, 256} for greyscale images. Under a suitable discretization that respects the coarea formula [CD09], it is possible to apply the same results to find the quantized u – originally a combinatorial problem – by solving a finite number of continuous cuts, or even graph cuts [DS06a, DS06b, DAG09].

2.5. Linear – TV : Multiclass Labeling

The results from the previous section do not transfer to the case of vector-valued u, as there is no natural extension of the coarea formula (18) and of the thresholding process to more than two classes. However it is still interesting to consider the extension of (16) to vector-valued u,

$$f(u) = \int_\Omega \langle u(x), s(x) \rangle \quad dx + \lambda \int_\Omega \| Du \|_2. \qquad (19)$$

This problem class naturally occurs in relaxations of *multi-class image labeling problems* . Here the goal is to find, for each pixel x in the image domain $\Omega \subseteq \mathbb{R}^d$, a label $\ell(x) \in \{1,\ldots,l\}$ which assigns one of l class labels to x so that the labeling function ℓ adheres to some data fidelity and spatial coherency constraints. In contrast to the image restoration applications above, u must assume one of a finite number of values at each point.

Several authors [ZGFN08, CCP08, LKY+09] independently proposed linearization techniques for this problem; the basic idea is also equivalent to the basis-function technique in [LLT06], and is a continuous counterpart to solving combinatorial finite-dimensional problems using LP relaxation [KT99].

We follow the notation used in [ZGFN08, LKY++09]: Identify label i with the i-th unit vector $e^i \in \mathbb{R}^l$, set $E := \{e^1,\ldots,e^l\}$, and find $u : \Omega \to E$ minimizing

$$f(u) = \int_\Omega \langle u(x), s(x)\rangle \quad dx + \lambda \int_\Omega \| Du \|_2 . \quad (20)$$

The *data term* assigns to each label $u(x) = e^i$ a *local cost* $s_i(x)$. If one constrains $u \in BV(\Omega)^l$, the right-hand side is well-defined even though u is not necessarily differentiable, and represents the total weighted length of interfaces between regions of constant labeling. As the data term is linear, the local costs s may be arbitrarily complex without affecting the overall problem class.

To tackle the combinatorial nature of (20), the problem is *relaxed*,

$$\inf_{u \in \mathcal{C}} \left\{ \int_\Omega \langle u(x), s(x)\rangle dx + \lambda \int_\Omega \| Du \|_2 \right\}, \quad (21)$$

where $\mathcal{C} := \{u \in BV(\Omega)^l \mid u_i(x) \geqslant 0, \sum_{i=1}^{l} u_i(x) = 1\}$ is a *convex* set which constrains each $u(x)$ to the unit simplex, i.e. to the convex hull of E. Since the regularizer is still convex, the overall problem is as well. On the other hand, due to the relaxation artificial *non-discrete* solutions with $u(x) \notin E$ at some points may be introduced.

As the data term was linearized, the local costs s may be arbitrarily complex, possibly derived from a probabilistic model, without losing convexity. Thus the important question is what effects can be achieved by modifying the regularizer. In the scope of this paper, we will consider approaches where the regularizer is replaced by

$$J(u) := \int_\Omega \Psi(Du), \quad (22)$$

where $\Psi : \mathbb{R}^{l \times d} \to \mathbb{R}_{\geq 0}$ is continuous, convex, positively homogeneous, i.e. $\Psi(cz) = c\Psi(z)$ for $c > 0$, and satisfies $\rho_u \| z \| \geq \Psi(z) \geq \rho_l \| z \|_2$ for some $\rho_u \geqslant \rho_l > 0$. In the $Linear-TV$ case, existence of a solution for the problem

$$\inf_{u \in \mathcal{C}} \left\{ \int_\Omega \langle u(x), s(x)\rangle dx + \int_\Omega \Psi(Du) \right\}, \quad (23)$$

follows from [AFP00, LS10].

Although we mainly motivated the use of modified regularizer in the labeling framework, they are as well interesting for ROF- and $TV - L^1$ type problems (12), (13) applied to vector-valued data.

3. Total Variation-Based Regularizers

We will first review some basic properties of the custom regularizer (22). For the proofs and details we refer to [LS10].

3.1. Basic Properties

A structure of the regularizer. Consider the effect of the regularizer (22) on labeling functions u that takes only *two* discrete values, e^i and e^j for some $i, j \in \{1,\ldots,l\}$, i.e. $u = e^i 1_S + e^j 1_{\Omega \setminus S}$ for some $S \subseteq \Omega$ with $Per(S) < \infty$. Then by [AFP00, Thm. 3.84] (a variant of this was also shown in [LS10, Prop. 4]),

$$J(e^i 1_S + e^j 1_{\Omega \setminus S}) = \int_{\Omega \cap \mathcal{F}S} \Psi\left(\nu_S \left(e^i - e^j\right)^{\mathrm{T}}\right) d\mathcal{H}^{d-1}, \quad (24)$$

where $\mathcal{F}S \subseteq \Omega$ denotes the *reduced boundary* of S, and $\nu_S : \mathcal{F}S \to S^{d-1}$ is the *generalized inner normal* of S [AFP00, Def. 3.54].

Therefore the regularizer locally penalizes jumps along the (reduced) boundary of S depending on the normal at each point, and on the labels of the adjoining regions.

Isotropy

Often, the regularizer is assumed to be rotation invariant (*isotropic*):

$$\Psi(Rz) = \Psi(z) \quad \forall R \in SO(d) \quad (25)$$

This is obviously the case if and only if $\Psi(\nu x^{\mathrm{T}}) = \Psi(e^1 x^{\mathrm{T}})$ for any $\nu \in S^{d-1}$. If Ψ is isotropic, for the application of multiclass labeling we may define the *interaction potential*

$$d(i,j) := \Psi\left(e^1\left(e^i - e^j\right)^{\mathrm{T}}\right). \qquad (26)$$

Due to the isotropy, we have $\Psi\left(e^1(e^i - e^j)^{\mathrm{T}}\right) = \Psi\left(-e^1(e^i - e^j)^{\mathrm{T}}\right) = \Psi(e^1(e^j - e^i))$, therefore $d(i,j) = d(j,i)$. From convexity and positive homogeneity it follows that d must be subadditive, and from the lower bound ρ_l we get that $d(i,j) = 0 \Leftrightarrow i = j$.

Hence for isotropic regularizers, the interaction potential d must be a metric, and defines the behavior on label functions taking only discrete label values $(e^1, \ldots, e^l)$ in the spirit of (24) by

$$J(e^i 1_S + e^j 1_{\Omega \setminus S}) = \int_{\Omega \cap \mathcal{F}S} d(i,j) d\mathcal{H}^{d-1} = d(i,j)\mathrm{Per}(S), \qquad (27)$$

i.e. boundaries between regions of constant labeling are penalized by their length, weighted by $d(i,j)$ depending on the labels i, j of the adjoining regions.

Permutation Invariance. This refers to the invariance of the regularizer with respect to permutations of the elements of u, i.e. of the label set in multiclass labeling. In terms of Ψ, invariance is given if $\Psi(z) = \Psi(zP)$ for any permutation matrix $P \in \mathbb{R}^{l \times l}$.

Separability. If Ψ can be written as a sum of terms that depend only on individual components or directional derivatives of u, it is called *separable* in the components of u resp. in space. Separability usually simplifies optimization, as it reduces the coupling between variables.

Homogeneity. In (21) we assume spatial homogeneity (translation invariance), as Ψ does not depend on x. It is also possible to consider general non-homogeneous regularizers of the form

$$\int_\Omega \Psi_x(Du). \qquad (28)$$

This is a common practice for anisotropic Ψ in combination with a process where the anisotropy is controlled by local properties of the input or the current iterate. An even more general regularization approach is

$$\int_\Omega g(x,u,Du), \qquad (29)$$

however this considerably complicates the requirements for the existence of solutions [AFP00, Chap. 5], and is out of scope for this paper.

Dual Formulation. In analogy to the dual definition for the total variation, for any continuous, positively homogeneous and convex Ψ it is possible to give a dual definition: By Fenchel duality, Ψ is the support function of some closed convex set $\mathcal{D}_{loc} \subseteq \mathbb{R}^{d \times l}$ [RW04],

$$\Psi(z) = \sigma_{\mathcal{D}_{loc}}(z) = \sup_{y \in \mathcal{D}_{loc}} \langle z, y \rangle. \qquad (30)$$

Then, in analogy to the dual formulation of the total variation,

$$J(u) = \sup\{\int \langle u, Div v \rangle dx \mid v \in \mathcal{D}\}, \qquad (31)$$

$$\mathcal{D} := \{v \in C_c^\infty(\Omega)^{d \times l} \mid v(x) \in \mathcal{D}_{loc} \ \forall x \in \Omega\}. \qquad (32)$$

Thus Ψ can be defined implicitly by its *dual set* $\mathcal{D}_{loc}$. This representation is also convenient for dual or primal-dual optimization methods [Cha05, ZGFN08, CCP08, LBS10, LS10].

3.2. Isotropic Approaches

3.2.1. Frobenius Norm

The most classical choice for Ψ is the Frobenius norm,

$$\Psi_F(Du) := \left(\sum_{i,j} \left(D^i u_j\right)^2\right)^{\frac{1}{2}}. \qquad (33)$$

This definition is the basis for large parts of geometric measure theory and the theory of functions of bounded variation [AFP00], and is sometimes referred to as MTV in the context of denoising of vector-valued data [SR96, DAV08]. It is isotropic and permutation invariant, however it is neither separable in the components of u nor in space. The dual set $\mathcal{D}_{loc}^F$ is just the ℓ_2 unit ball in $\mathbb{R}^{d \times l}$. The associated potential is

$$\left(1/\sqrt{2}\right)\Psi\left(\nu\left(e^i - e^j\right)^{\mathrm{T}}\right) = 1_{i \neq j} =: d_u(i,j). \qquad (34)$$

The potential d_u on the right-hand side is known as *uniform*, *discrete*, or *Potts* metric, and widely used in approaches defined on finite grids [KT99, BVZ01, KT07]. Optimization approaches for more involved discretizations can be found in [LLT06, LKY+09].

3.2.2. Linear Transformations

A straightforward modification that can be applied to most Ψ is to introduce a matrix $A=\left(a^1\,|\ldots|\,a^l\right)\in\mathbb{R}^{k\times l}$ for some $k\leqslant l$, and define

$$\Psi_{F,A}(Du):=\Psi_F(D(Au))=\Psi_F\left((Du)A^{\mathrm{T}}\right). \quad (35)$$

This corresponds to substituting the Frobenius matrix norm on the distributional gradient with a linearly weighted variant. While we only consider the Frobenius norm here, the approach can in principle be used to augment all other norms. It retains isotropy of the underlying norm, but neither permutation invariance nor separability.

Applied to a jump from label i to label j, this results in the potential

$$\Psi_{F,A}\left(\nu\left(e^i-e^j\right)^{\mathrm{T}}\right)=\left\|a^i-a^j\right\|_2=:d_A(i,j). \quad (36)$$

As noted in [LBS09], metrics of the form d_A are known as *Euclidean* metrics. This class comprises some important special cases:

- the uniform metric, with $A=(1/\sqrt{2})I$;
- the *linear* (label) metric, $d(i,j)=c\,|\,i-j\,|$, with $A=(c,2c,\ldots,lc)$. This regularizer is suitable to problems where the labels can be naturally ordered, e.g. the depth from stereo or grayscale image denoising.
- More generally, if label i corresponds to a prototypical vector z^i in k-dimensional feature space, and the Euclidean norm is an appropriate metric on the features, it is natural to set $d(i,j)=\left\|z^i-z^j\right\|$, which is Euclidean by construction. This corresponds to a regularization in feature space, rather than in «label space».

Also, non-Euclidean metrics such as the *truncated linear metric*, $d(i,j)=\min\{2,|\,i-j\,|\}$, can be approximated by solving a (convex) SDP problem, cf. [LS10] and the references therein.

One major advantage of this kind of modification is that it is quite powerful, but retains a closed-form expression for the regularizer. The dual set is just $\mathcal{D}^A_{loc}=\mathcal{D}^F_{loc}\mathrm{A}^{\mathrm{T}}$. Alternatively, it is often easier to formally merge A into the linear gradient operator Du, which allows to keep the structure of the dual set and requires only few modifications to the optimization method [LKY+09].

3.2.3. Channel-By-Channel

Maybe the most straightforward approach to transfer from scalar total variation to the vector-valued case is to sum up the total variations of the components, i.e.

$$\Psi_1(Du):=\|Du_1\|_2+\ldots+\|Du_l\|_2. \quad (37)$$

In this formulation, the objective is separable in the components of u, which potentially simplifies numerical optimization [Blo98, ZGFN08]. For unconstrained u, a variant of the coarea formula (18) still holds, i.e.

$$\int_\Omega\Psi_1(Du)=\int_{-\infty}^{\infty}\left(\int_\Omega\Psi_1(Dw^\alpha)\right)d\alpha,\ \left(w^\alpha\right)_i:=1_{u_i>\alpha}. \quad (38)$$

However for labeling problems, where one is looking for $u:\Omega\to E=\{e^1,\ldots,e^l\}$, this is of very limited interest, since $w^\alpha(x)\notin E$ in general.

Similar to the Frobenius norm, Ψ_1 implements the uniform metric,

$$(1/2)\Psi_1\left(\nu\left(e^i-e^j\right)^{\mathrm{T}}\right)=d_u(i,j), \quad (39)$$

and is isotropic, with $\mathcal{D}^1_{loc}=\{(z^1\,|\ldots|\,z^l)\,|\,\|z^i\|_2\leqslant \leqslant 1\ \forall i\in\{1,\ldots,l\}\}$. As in the case of the Frobenius norm, linearly transformed variants of the form

$$\Psi_{1,A}(Du):=\Psi_1(D(Au)) \quad (40)$$

could be used. A straightforward transformation shows that $\Psi_{1,A}=\left\|a^i-a^j\right\|_1$. The class of metrics covered by this approach would thus again be those representable using a linear embedding into a space that is now endowed with the norm given by Ψ_1 instead of the Euclidean norm.

3.2.4. Convex Envelope Approach

Motivated by [ABDM01], Chambolle et al. [CCP08] proposed an approach for constructing regularizers for multiclass labeling problems, where the potential d is given in advance as $d(i,j)=\tau(|\,i-j\,|)$ for a positive concave function τ. While they used a different notation for the parametrization of the unit simplex through u, their parametrization is equivalent to (20) under a linear transformation of the components of u.

The approach is derived by setting

$$\Psi'(z):=\begin{cases}\tau(|\,i-j\,|), & z=\nu(e^i-e^j)^{\mathrm{T}},\\ +\infty, & \mathit{otherwise}.\end{cases} \quad (41)$$

Then Ψ is constructed as the convex envelope of Ψ', i.e. the largest convex function smaller or equal to Ψ'. This potentially generates Ψ that are as large as possible while still satisfying (27) for the given *d*, and thus we may hope that the relaxation to the unit simplex (21) does not generate too many artificial non-discrete solutions. This leads to

$$\mathcal{D}^{\tau}_{loc} := \bigcap_{i \neq j} \{v = \left(v^1, \ldots, v^l\right) \in \in \mathbb{R}^{d \times l} \mid \| v^i - v^j \|_2 \leq \tau(|i-j|), e^{\mathrm{T}} v = 0\}. \quad (42)$$

The approach can be extended to arbitrary metrics *d* by setting [LS10]

$$\mathcal{D}^{d}_{loc} := \bigcap_{i \neq j} \{v = \left(v^1, \ldots, v^l\right) \in \in \mathbb{R}^{d \times l} \mid \| v^i - v^j \|_2 \leq d(i,j), e^{\mathrm{T}} v = 0\} \quad (43)$$

for some given interface potential $d(i, j)$. By definition, $\mathcal{D}^{d}_{loc}$ and thus $\Psi_d = \sigma_{\mathcal{D}^{d}_{loc}}$ are isotropic. In [LS10] it was shown that, for metrics *d*, the resulting Ψ_d satisfies the desired

$$\Psi_d(\nu(e^i - e^j)) = d(i,j). \quad (44)$$

This provides an approach to construct regularizers with arbitrary (metric) prescribed interaction potentials *d*. The downside is that there is no simple closed expression for Ψ and thus for the regularizer, and the dual set can be quite involved, which potentially complicates optimization.

From (44) it can be seen that Ψ_d is permutation invariant only for $d = \lambda d_u, \lambda \geqslant 0$. Also note that in order to define Ψ_d, *d* does not have to be a metric. However (44) then only holds as an inequality, so *J* is not a true extension of the desired regularizer.

3.2.5. Eigenvalue-Based Norms

In the anisotropic diffusion community it is widely used practice ([SR96], see also [WS01] and the references therein) to employ weighted norms based on the eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_d \geqslant 0$ of the structure tensor

$$G(Du) = (Du)(Du)^{\mathrm{T}}. \quad (45)$$

Let $\lambda^2(Du) := (\lambda_1^2, \ldots, \lambda_d^2)$, i.e. the λ_i represent the magnitudes of the singular values of *Du*. While originally rooted in a diffusion framework, the approach can also be used to construct TV-like regularizers. It includes the Frobenius norm, since

$$\Psi_F(Du) = \sqrt{e^{\mathrm{T}} \lambda^2(Du)}. \quad (46)$$

In addition, for some rotation matrix $R \in R^{d \times d}$ and permutation matrix $P \in R^{l \times l}$,

$$G(R(Du)P) = R(Du)PP^{\mathrm{T}}(Du)^{\mathrm{T}} R^{\mathrm{T}} = = RG(Du)R^{\mathrm{T}}, \quad (47)$$

therefore $\lambda^2(Du) = \lambda^2(R(Du)P)$, i.e. all norms derived from these singular values are isotropic and permutation invariant.

The Frobenius norm approach was considered for color denoising in [Blo98]. They observed that, since

$$\Psi_F(Du) := \left(\sum_i \| Du_i \|^2 \right)^{1/2}, \quad (48)$$

the Frobenius norm prefers transitions with similar magnitude in all channels: The transition (0,0)→ →(1,1) is assigned a much lower penalty than the two consecutive transitions (0,0)→(1,0)→(0,1). This phenomenon does not occur with the Channel-by-Channel regularizer Ψ_1. For color denoising, this leads to a color smearing effect at edges and a color shift towards the greyscale image [Blo98]. A similar effect was observed in [CCP08] for multiclass segmentation, where the preference towards similar gradients leads to minimizers that assume non-discrete values more frequently than is the case for e.g. Ψ_d.

In [GC10a] it was proposed to employ

$$\Psi_2(Du) := \sqrt{\lambda_1^2(G)}, \quad (49)$$

which amounts to the standard ℓ_2 operator norm on *Du*. The corresponding dual set can be represented as

$$\mathcal{D}^2_{loc} = \{v x^{\mathrm{T}} \mid v \in R^d, x \in R^l, \|v\|_2 \leq 1, \|x\|_2 \leq 1\}. \quad (50)$$

While this more difficult to handle numerically than e.g. $\mathcal{D}^F_{loc}$, it can be dealt with reasonably well in primal-dual methods, and experimentally reduces color smearing and channel coupling in denoising, deblurring and superresolution applications [GC10a]. Regarding labeling approaches, we have

$$\left(1/\sqrt{2}\right) \Psi_2\left(v\left(e^i - e^j\right)^{\mathrm{T}}\right) = d_u(i,j), \quad (51)$$

i.e. Ψ_2 represents the uniform metric. However, since $\Psi_2 \leqslant \Psi_F$, for multiclass labeling the energy when using Ψ_2 potentially generates more undesired minima than the standard choice Ψ_F.

3.3. Anisotropic Approaches

In this section we will consider approaches that are not rotation invariant. Note that most of these have been developed for scalar-valued total variation; however they extend to the vector-valued case in a straightforward manner and could be coupled with any of the above approaches for weighting different labels.

3.3.1. Wulff Shapes

For scalar-valued u, the use of anisotropic variants of the total variation has been studied in [EO04] for the ROF model, where the authors characterize minimizers of such functionals. They base their analysis on the «Wulff shape» associated with Ψ, which is identical to $\mathcal{D}_{loc}$ for the scalar-valued case. As an example, consider setting $\mathcal{D}_{loc}^b := [0,1]^d$. Then, for (scalar!) $u \in BV(\Omega)$,

$$\Psi_b(Du) = \| Du \|_1 = | D^1 u | + \ldots + | D^d u |. \quad (52)$$

Essentially, the Wulff shape defines the norm Ψ via the unit ball of its dual norm. It can be shown that the structure of $\mathcal{D}_{loc}$ is reflected in the structure of the minimizer of the ROF functional (12) in the sense that it does not affect structures in the shape of $\mathcal{D}_{loc}$ itself: For ν small enough, the minimizer $u^* u^*$ of

$$f(u) = \frac{1}{2}\int_\Omega \left(u - 1_{\mathcal{D}_{loc}} \right)^2 + \nu \int_\Omega \Psi(Du) \quad (53)$$

is just a multiple of the input, i.e. $u^* = c 1_{\mathcal{D}_{loc}}$ [EO04, Thm. 4.1]. Applied to the above definition for Ψ_b, this means that the unit box may occur as the minimizer of the anisotropic ROF model, which cannot happen for the standard ROF model [Mey01].

Based on these results, it was shown in [ZNF09] that the thresholding property for isotropic continuous cuts can be transferred to their anisotropic counterparts, i.e. it is still possible to recover discrete solutions of the anisotropic continuous cut problems by thresholding. When combined with a spatially varying, edge-driven adaptation of the Wulff shape, they observed improved visual quality when applied to the reconstruction of depth maps and 3D structure.

These anisotropies can also be extended to the vector-valued case, e.g. by setting

$$\Psi(Du) = \left(\sum_i \Psi_b(Du_i) \right)^2, \quad (54)$$

or by replacing $\|\cdot\|_2$ by Ψ_b in (43). However as in the isotropic case this invalidates the thresholding property.

3.3.2. Anisotropy from Discretization: 4-neighborhoods

A large class of anisotropies that occurs in practice are actually induced by the discretization used to approximate the total variation on grids. A very common scheme is to add the total variation of the individual components,

$$\Psi_{a,\|\cdot\|}(Du) := \left\| D^1 u \right\| + \ldots + \left\| D^d u \right\|. \quad (55)$$

Here $\|\cdot\|$ refers to some norm on $\mathbb{R}^l$, notable cases include $\|\cdot\|_2$ and the completely separable case

$$\Psi_{a,1}(Du) := \Psi_{a,\|\cdot\|_1}(Du) := \| D^1 u \|_1 + \ldots + \| D^d u \|_1 = \sum_{i=1}^{d} \sum_{j=1}^{l} | D^i u_j |. \quad (56)$$

This is in fact the anisotropy that is implicitly assumed by many algorithms that use a grid-based representation with pairwise potentials [KT99, BVZ01, KT07]: Assume that $\Omega = (0,1)^2$, $N \in N$ and $h := 1/N$, and u is discretized by its values on the uniform grid $\{x^{i,j} = (ih, jh) \mid i, j \in \{0, \ldots, N\}\}$. The usual 4-neighborhood discretization then amounts to (with appropriate boundary conditions) approximating the directional derivatives via

$$D^1 u(x^{i,j}) \approx \frac{1}{h}\left(u\left(x^{i+1,j}\right) - u\left(x^{i,j}\right) \right), \quad D^2 u(x^{i,j}) \approx \frac{1}{h}\left(u\left(x^{i,j+1}\right) - u\left(x^{i,j}\right) \right). \quad (57)$$

In the case of scalar-valued u (or for the individual terms in $\Psi_{a,1}$), the total variation with Neumann boundary conditions is then discretized as

$$J(u) \approx \frac{1}{h} \sum_{i,j=0}^{N-1} \left| u\left(x^{i+1,j}\right) - u\left(x^{i,j}\right) \right| +$$

$$+\frac{1}{h}\sum_{i,j=0}^{N-1}\left|u\left(x^{i,j+1}\right)-u\left(x^{i,j}\right)\right|. \qquad (58)$$

This type of energy is tremendously popular as it is convex, contains only pairwise terms (i.e. terms depending on only two different variables) and is therefore easy to implement and analyze. Moreover, for two-class problems (resp. scalar-valued u), the energy has the thresholding property and is *submodular*, which allows to find a *discrete* solution in polynomial time, for example using graph-cut methods [BVZ01].

For infinitesimal grid size, it implements the $\Psi_{a,1}$ norm. The main drawback is that edges parallel to the coordinate axes are preferred to diagonal edges, which often leads to «zig-zag» artifacts on diagonal structures.

3.3.3. The anisotropy from Discretization: Larger Neighborhoods

To some extent, the discretization-induced anisotropy can be reduced by increasing the neighborhood, i.e. by increasing the number of pairwise terms and adding proper weighting factors. In [Boy03, KB05] it was shown that anisotropies formulated as a certain class of metrics can asymptotically be approximated arbitrarily well using pairwise terms. However this requires the grid spacing and the neighborhood size to approach zero and infinity, respectively: For discretizations with a fixed number of neighbors, true isotropy cannot be guaranteed even for an arbitrarily fine grid.

In practice, increasing the neighborhood size generally reduces artifacts but increases runtime. This effect is even more pronounced in higher-dimensional data [KSK+08].

3.4. Other Approaches

3.4.1. Color TV

For the restoration of multichannel data such as color images [Blo98] suggests to use

$$J(u)=\left(\sum_{i=1}^{l}TV(u^i)^2\right)^{1/2}. \qquad (59)$$

While this approach is fully isotropic and permutation invariant, and seems to improve ROF restoration of images with large intensity deviation between color channels, it cannot be represented in the form (22) using Ψ. Moreover, it has the distinct disadvantage that it lacks locality, completely coupling all points in the image. While this problem is less severe when using PDE-based schemes for optimization, it becomes a larger impediment when applying more advanced schemes that rely on some form of sparsity.

3.4.2. The lifting of nonconvex problems

In [ABDM01, CCP08, PCBC09] a technique was developed to minimize general variational functionals of the form

$$f'(u') := \int_{\Omega} h(x,u',\nabla u')dx, \qquad (60)$$

over $u'\in W^{1,1}(\Omega')$, where f is convex in $\nabla u'$, but not necessarily in u'. The approach relies on the same approach as applied in [Ish03] in the graph cut framework, essentially *lifting* the problem originally formulated on $\Omega'\subseteq R^d$ to a higher-dimensional domain $\Omega\subseteq R^{d+1}$. It was shown that

$$f'(u') = \int_{\Omega\times\mathbb{R}}\Psi_{h,x}\left(D1_{\{(x,t)\in\mathbb{R}^d\times\mathbb{R}|u'(x)\geq t\}}\right), \qquad (61)$$

where $\Psi_{h,x} := \sigma_{\mathcal{D}^{h,x}_{loc}}$ is defined implicitly via the Legendre-Fenchel conjugate of h with respect to the last argument:

$$\mathcal{D}^{h,x}_{loc} := \{(v,w)\in R^d\times R \mid w \geqslant h^*(x,t,v)\}. \qquad (62)$$

Essentially, this transforms the problem of finding the optimal u' into the problem of finding the set of points below its graph, which can be seen as a two-class segmentation problem in $\mathbb{R}^{d+1}$. This can be solved by a relaxation technique applied to (61). The problem can thus be treated as a highly anisotropic segmentation problem.

3.4.3. Partially separable norms

For linearizations of labeling problems that involve a large number of labels at each point, optimization can be made more efficient by exploiting separability in the regularizer. This occurs for example in optical flow estimation, where the two-dimensional flow vectors $u=(u_1,u_2)$ at each point are quantized using M^2 labels, which requires a prohibitively large amount of memory for fine quantizations.

If the regularizer decomposes with respect to u_1 and u_2, i.e. $J(u)=J_1(u_1)+J_2(u_2)$, it is possible to apply the relaxation technique in [GC10b], which only requires memory in the order of $O(2M)$ as opposed to $O(M^2)$.

4. Conclusion

In this paper, we tried to give an overview over recent variational methods that make use of the unique properties of total variation-based regularizers. These range from traditional restoration approaches to recent advanced in relaxation techniques for approximately solving combinatorial problems using convex optimization.

For all these approaches, it is essential to choose a suitable norm when constructing the regularizer. We hope that this overview will prove useful as a reference and for weighing the advantages and disadvantages of the numerous variants that have been proposed.

1. [ABDM01] *Alberti G.*, *Bouchitté G.*, *Dal Maso G.* The calibration method for the Mumford-Shah functional and free-discontinuity problems. Technical Report math/0105013v1, arXiv, 2001.
2. [AFP00] *Ambrosio L.*, *Fusco N.*, *Pallara D.* Functions of Bounded Variation and Free Discontinuity Problems. Clarendon Press, 2000.
3. [AMT91] *Ambrosio L.*, *Mortola S.*, *Tortorelli V.M.* Functionals with linear growth defined on vector valued BV functions // J. Math. pures et appl., 70:193–323, 1991.
4. [Ber09] *Berkels B.* An unconstrained multiphase thresholding approach for image segmentation // Scale Space and Var. Meth. in Comp. Vis., **5567** of Springer LNCS. – 2009. – P. 26–37.
5. [BKP10] *Bredies K.*, *Kunisch K.*, *Pock T.* Total generalized variation // SIAM J. Imaging Sci. – 2010. – N 3(3). – P. 294–526.
6. [Blo98] *Blomgren P.V.* Total Variation Methods for Restoration of Vector Valued Images. PhD thesis, UCLA, 1998.
7. [Boy03] *Boykov Y.* Computing geodesics and minimal surfaces via graph cuts // Int. Conf. Comp. Vis. – 2003. – P. 26–33.
8. [BVZ01] *Boykov Yu.*, *Veksler O.*, *Zabih R.* Fast approximate energy minimization via graph cuts // Patt. Anal. Mach. Intell. – 2001. – N 23(11). – P. 1222–1239.
9. [CCP08] *Chambolle A.*, *Cremers D.*, Pock T. A convex approach for computing minimal partitions. Tech. Rep. 649, Ecole Polytechnique CMAP, 2008.
10. [CD09] *Chambolle A.*, *Darbon J.* On total variation minimization and surface evolution using parametric maximum flows // Int. J. Comp. Vis. // – 2009. – **84**. – P. 288–307.
11. [CEN06] *Chan T.F.*, *Esedoḡlu S.*, *Nikolova M.* Algorithms for finding global minimizers of image segmentation and denoising models // J. Appl. Math. – 2006. – **66**(5). – P. 1632–1648.
12. [Cha05] *Chambolle A.* Total variation minimization and a class of binary MRF models // EMMCVPR. – 2005. – **3757**. – P. 136–152.
13. [CL97] *Chambolle A.*, *Lions P.-L.* Image recovery via total variation minimization and related problems // Numer. Math. – 1997. – **76**(2). – P. 167–188.
14. [CMM00] *Chan T.*, *Marquina A.*, *Mulet P.* Higher-order total variation-based image restoration // J. Sci. Comput. – 2000. – **22**(2). – P. 503–516.
15. [CV01] *Chan T.F.*, *Vese L.A.* Active contours without edges // IEEE Trans. Image Proc. – 2001. – **10**(2). – P. 266–277.
16. [DAG09] *Duval V.*, *Aujol J.-F.*, *Gousseau Y.* The TVL1 model: a geometric point of view // Technical Report 00380195, HAL, 2009.
17. [DAV08] *Duval V.*, *Aujol J.-F.*, *Vese L.* A projected gradient algorithm for color image decomposition // CMLA Preprint, (2008–21), 2008.
18. [DS06a] *Darbon J.*, *Sigelle M.* Image restoration with discrete constrained total variation part I: Fast and exact optimization // J. Math. Imaging Vis. – 2006. – **26**(3). – P. 261–276.
19. [DS06b] *Darbon J.*, *Sigelle M.* Image restoration with discrete constrained total variation part II: Levelable functions, convex priors and non-convex cases // Ibid. – P. 277–291.
20. [EO04] *Esedoḡlu S.*, *Osher S.J.* Decomposition of images by the anisotropic Rudin-Osher-Fatemi model // Comm. Pure Appl. Math. – 2004. – **57**. – P. 1609–1626.
21. [FR60] *Fleming W.H.*, *Rishel R.* An integral formula for total gradient variation // Archiv der Mathematik. – 1960. – **11**(1). – P. 218–222.
22. [GC10a] *Goldluecke B.*, *Cremers D.* An approach to vectorial total variation based on geometric measure theory // Conf. Comp. Vis. Patt. Recog., 2010.
23. [GC10b] *Goldluecke S.*, *Cremers D.* Convex relaxation for multilabel problems with product label spaces // Europ. Conf. Comp. Vis., 2010.
24. [GY07] *Goldfarb D.*, *Yin Y.* Parametric maximum flow algorithms for fast total variation minimization. CAAM tech. rep., Rice Univ., 2007.
25. [Hoc01] *Hochbaum D.* An efficient algorithm for image segmentation, Markov random fields and related problems // J. of the ACM. –2001. – 48.
26. [Ish03] *Ishikawa H.*. Exact optimization for Markov random fields with convex priors // Patt. Anal. Mach. Intell.. – 2003. – **25**(10). – P. 1333–1336.
27. [KB05] *Kolmogorov V.*, *Boykov Y.* What metrics can be approximated by geo-cuts, or global optimization of length/area and flux // Int. Conf. Comp. Vis. –2005. – N 1. – P. 564–571.
28. [KSK+08] *Klodt M.*, *Schoenemann T.*, *Kolev K.* et al. An experimental comparison of discrete and continu-

ous shape optimization methods // Europ. Conf. Comp. Vis., Marseille, France, Oct. 2008.
29. [KT99] *Kleinberg J.M.*, *Tardos E.* Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields // Found. Comp. Sci. –1999. – P. 14–23.
30. [KT07] *Komodakis N.*, *Tziritas G.* Approximate labeling via graph cuts based on linear programming // Patt. Anal. Mach. Intell. – 2007. – **29**(8). – P. 1436–1453.
31. [LBS09] *Lellmann J.*, *Becker F.*, *Schnörr C.* Convex optimization for multi-class image labeling with a novel family of total variation based regularizers // Int. Conf. Comp. Vis., 2009.
32. [LBS10] *Lellmann J.*, *Breitenreicher D.*, *Schnörr C.* Fast and exact primal-dual iterations for variational problems in computer vision // Europ. Conf. Comp. Vis. – 2010. – **6312** of LNCS. – P. 494–505.
33. [LKY+09] *Convex* multi-class image labeling by simplex-constrained total variation / J. Lellmann, J. Kappes, J. Yuanet al. // Scale Space and Var. Meth. – 2009. – **5567** of LNCS. – P. 150–162.
34. [LLT06] *Lie J.*, *Lysaker M.*, *Tai X.-C.* A variant of the level set method and applications to image segmentation // Math. Comp. – 2006. – **75**. – P. 1155–1174.
35. [LS10] *Lellmann J.*, *Schnörr C.* Continuous multiclass labeling approaches and algorithms // Tech. rep., Univ. of Heidelberg, Feb. 2010.
36. [LT06] *Lysaker M.*, *Tai X.-C.* Iterative image restoration combining total variation minimization and a second-order functional // Int. J. Comp. Vis. – 2006. – **66**(1). – P. 5–18.
37. [Mey01] *Meyer Y.* Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, 22 of Univ. Lect. Series. AMS, 2001.
38. [Nik] *Nikolova M.* Functionals for signal and image reconstruction: properties of their minimizers and applications. CMLA research report, ENS de Cachan.
39. [Nik01] *Nikolova M.* Smoothing of outliers in image restoration by minimizing regularized objective functions with nonsmooth data-fidelity terms // Int. Conf. Image Proc. – 2001. – **1**. – P. 233–236.
40. [PCBC09] *Global* solutions of variational models with convex regularization / T. Pock, D. Cremers, H. Bischof et al., Technical report, Graz Univ. of Tech., 2009. Preprint.
41. [ROF92] *Rudin L.*, *Osher S.*, *Fatemi E.* Nonlinear total variation based noise removal algorithms // Physica D. – 1992. – **60**. – P. 259–268.
42. [RW04] *Rockafellar R.T.*, *Wets R.J.-B.* Variational Analysis. Springer, 2004.
43. [Sch98] *Scherzer O.* Denoising with higher order derivatives of bounded variation and an application to parameter estimation // Computing. – 1998. – **60**. – P. 1–27.
44. [SGG+09] *Scherzer O.*, *Grasmair M.*, Grossauer H., *Haltmeier M.*, *Lenzen F.* Variational Methods in Imaging. – **167** of Applied Mathematical Sciences. Springer, 2009.
45. [SKO09] *Strandmark P.*, *Kahl F.*, *Overgaard N.C.* Optimizing parametric total variation models // Int. Conf. Comp. Vis. – 2009.
46. [SR96] *Sapiro G.*, *Ringach D.L.* Anisotropic diffusion of multivalued images with applications to color filtering // IEEE Trans. Image Proc. – 1996. – **5**(11). – P. 1582–1586.
47. [Str83] *Strang G.* Maximal flow through a domain // Math. Prog. – 1983. – **26**. – P. 123–143.
48. [WS01] *Weickert J.*, *Schnörr C.* A theoretical framework for convex regularizers in PDE-based computation of image motion // Int. J. Comp. Vis. – 2002. – **45**(3). – P. 245–264.
49. [ZGFN08] *Fast* global labeling for real-time stereo using multiple plane sweeps / C. Zach, D. Gallup, J. Frahm et al. // Vis. Mod. Vis, 2008.
50. [ZNF09] *Zach C.*, *Niethammer M.*, Frahm *J.-M.* Continuous maximal flows and Wulff shapes: Application to MRFs // Comp. Vis. Patt. Recogn. – 2009. – P. 1911–1918.

●