Imaging with Kantorovich-Rubinstein discrepancy

Jan Lellmann* Dirk A. Lorenz† Carola Schönlieb‡
Tuomo Valkonen§

2nd October 2014

Abstract

We propose the use of the Kantorovich-Rubinstein norm from optimal transport in imaging problems. In particular, we discuss a variational regularisation model endowed with a Kantorovich-Rubinstein discrepancy term and total variation regularization in the context of image denoising and cartoon-texture decomposition. We point out connections of this approach to several other recently proposed methods such as total generalized variation and norms capturing oscillating patterns. We also show that the respective optimization problem can be turned into a convex-concave saddle point problem with simple constraints and hence, can be solved by standard tools. Numerical examples exhibit interesting features and favourable performance for denoising and cartoon-texture decomposition.

1 Introduction

In this paper we introduce a distance function from optimal transport to the field of mathematical imaging. Optimal transport is the theory that answers questions about how to transport a given initial mass distribution to a desired new distribution and do so in the most efficient way (according to some cost functions), see [63] for a recent review and further references. Distance functions related to ideas from optimal transport have appeared in various places in imaging problems in the last ten years. The main applications in this context are image and shape classification [36–40, 45, 51, 59], segmentation [16,44,48,56,57], registration and warping [27,46,66], image smoothing [11],

*Department for Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom, j.lellmann@damtp.cam.ac.uk

†Institute for Analysis and Algebra, TU Braunschweig, 38092 Braunschweig, Germany, d.lorenz@tu-braunschweig.de

‡Department for Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom

§Prometeo Fellow, Center for Mathematical Modeling (Modemat), EPN Quito, Ecuador

contrast and colour modification [22, 50], texture synthesis and texture mixing [52], and surface mapping [6,10,32,33]. Being a distance function applicable to very general densities (continuous and discrete (Dirac deltas) densities) the Wasserstein distance had an increasing impact on robust distance measures in imaging [11,12,26,31,48,52,54,64]. In most cases, the 2-Wasserstein distance [2] is used.

In this work we propose the use of the so-called Kantorovich-Rubinstein norm (KR-norm) in imaging. In combination with total variation (TV) denoising, we investigate the KR-TV denoising problem. Consider a given noisy image u^0 for which a denoised version u is sought. In variation denoising one formulates this problem as a minimization problem where on minimizes the sum of a *discrepancy term* which measures the distance from the given image u^0 to the image u, and a *penalty term* that penalizes images u that are not natural in some sense [55]. We investigate the case in which the discrepancy term is the KR-norm and the penalty term it the TV seminorm, i.e. for a given noisy image u^0 on a set Ω and two constants $\lambda_1, \lambda_2 \geq 0$ we consider

$$\min_u \|u - u^0\|_{\mathrm{KR},(\lambda_1,\lambda_2)} + \mathrm{TV}(u)$$

where the KR-norm is defined for a Radon measure μ (and hence, also for L^1-functions) on a set $\Omega \subset \mathbb{R}^n$ by

$$\|\mu\|_{\mathrm{KR},(\lambda_1,\lambda_2)} = \sup\{\int_\Omega f \,\mathrm{d}\mu \ : \ |f| \leq \lambda_1,\ \mathrm{Lip}(f) \leq \lambda_2\}.$$

The Kantorovich-Rubinstein norm [5, §8.3] is closely related to the 1-Wasserstein distance and hence, to optimal transport problems. It will turn out that this norm has interesting relations to other well known concepts in imaging: The KR-norm is a generalization of the L^1 norm, and hence, a KR-TV denoising model inherits and generalizes some of the favorable properties of the L^1-TV denoising [15]. The generalization of L^1-norm discrepancies to KR-norm discrepancies shares some similarities with the generalization from the TV penalty to the total generalized variation (TGV) penalty [7]. Finally, the KR-norm discrepancy shares properties with Meyer's G-norm model [41, 62] for oscillating patterns and for cartoon-texture decomposition. Also from the computational point of view, the KR-norm has favorable properties. It turns out that the KR-TV denoising problem has a formulation as a saddle-point problem that can be solved by means of several primal-dual methods. The computational cost per iteration as well as the needed storage requirements are almost as low as for similar algorithms for L^1-TV denoising.

The paper is organized as follows: After fixing the notation we introduce and recall transport metrics in Section 2. In Section 3 we derive two reformulations of the KR-norm that will be used to analyze and interpret the KR-TV denoising problem, which is the content of Section 4. In Section 5 we illustrate how the KR-TV denoising problem can be solved numerically by primal-dual methods. Finally, in Section 6 we present examples for KR-TV denoising and cartoon-texture decomposition and then finish the paper with a conclusion.

1.1 Notation

We work in a domain $\Omega \subset \mathbb{R}^n$ and use $|x|$ as the euclidean absolute value for $x \in \Omega$. We denote by $\mathfrak{M}(\Omega, \mathbb{R}^n)$ the space of $\mathbb{R}^n$-valued Radon measures, i.e. the dual space of $(C_0(\Omega, \mathbb{R}^n), \||\cdot|\|_\infty)$ of continuous functions that vanish "at infinity". If we want to emphasize that a function or a measure is vector valued we write $\vec{\nu}$ but sometime we omit the emphasis. The dual pairing between $\mathfrak{M}(\Omega, \mathbb{R}^n)$ and $C_0(\Omega, \mathbb{R}^n)$ (and any two other spaces in duality) will be denoted by $\langle \vec{f}, \vec{\mu} \rangle$. Consequently, the norm on $\mathfrak{M}(\Omega, \mathbb{R}^n)$ is $\|\vec{\mu}\|_{\mathfrak{M}} = \sup_{|\vec{f}| \leq 1} \int \vec{f} \cdot \mathrm{d}\vec{\mu}$ and is called the Radon norm. We identify $u \in L^1(\Omega, \mathbb{R}^n)$ with the corresponding measure $u \in \mathfrak{M}(\Omega, \mathbb{R}^n)$, i.e. we treat $L^1(\Omega, \mathbb{R}^n)$ embedded into $\mathfrak{M}(\Omega, \mathbb{R}^n)$. The n-dimensional Lebesgue measure is denoted by $\mathfrak{L}^n$ while the d-dimensional Hausdorff measure is $\mathfrak{H}^d$.

For a measure μ on Ω, another set Ω' and $F : \Omega \to \Omega'$ the push-forward of μ by F is $\mu \# F(A) = \mu(F^{-1}(A))$. On $\Omega \times \Omega$ we denote by $\mathrm{proj}_{1/2}$ the projections onto the first and second component, respectively. Having a measure γ on $\Omega \times \Omega$ we denote (with slight abuse of notation) by $\mathrm{proj}_{1/2} \gamma$ the push forward of γ by $\mathrm{proj}_{1/2}$, i.e. the marginals of γ. The restriction of some measure μ onto some set A is denoted by $\mu \llcorner A$. By $C_b(\Omega, \mathbb{R}^n)$ we denote the space of bounded and continuous functions on Ω. For $f : \Omega \to \mathbb{R}$ we denote by $\mathrm{Lip}(f) = \sup_{x \neq y} |f(x) - f(y)|/|x - y|$ the Lipschitz constant of f.

For two points $a, b \in \mathbb{R}^n$ we define the line interval $[a, b] = \{ta + (1 - t)b \mid t \in [0, 1]\}$ and the vector measure $[\![a, b]\!]$ to be

$$[\![a, b]\!] = \frac{b - a}{|b - a|} \mathfrak{H}^1 \llcorner [a, b].$$

By $\mathrm{diam}(\Omega) = \sup\{|x - y| \;:\; x, y \in \Omega\}$ we denote the diameter on Ω. For a set C we denote by I_C the indicator function, i.e. $I_C(u) = 0$ for $u \in C$ and $= \infty$ otherwise.

2 Transport metrics

A variety of different metrics exist on measure spaces. As the study of metrics on measure spaces has its origins in probability theory, most metrics are defined on the space of probability measures, i.e., non-negative measures with total mass equal to one. A popular class of such metrics is given by the Wasserstein metrics: For $p \geq 1$ and two probability measures μ and ν define the p-Wasserstein distance

$$W_p(\mu, \nu) = \Big(\inf\{ \int_{\Omega \times \Omega} |x - y|^p \,\mathrm{d}\gamma(x, y) \;:\; \mathrm{proj}_1 \gamma = \mu,\ \mathrm{proj}_2 \gamma = \nu\} \Big)^{1/p}. \quad (1)$$

Note that this metric also makes sense if μ and ν are not probability measures but still non-negative and have equal mass, i.e., $\int_\Omega \mathrm{d}\mu = \int_\Omega \mathrm{d}\nu$. However, if the mass is not equal, no γ with μ and ν as marginals would exist.

The celebrated Kantorovich duality [28, 63] states that, in the case of non-negative measures with equal mass, the Wasserstein metric can be equivalently expressed as

$$W_p(\mu,\nu) = \Big(\sup\{\int_\Omega \phi \,\mathrm{d}\mu + \int_\Omega \psi \,\mathrm{d}\nu \ :\ \phi,\psi \in C_b(\Omega),\ \phi(x)+\psi(y) \le |x-y|^p\}\Big)^{1/p}.$$

A particular special case is $p = 1$, and here, the Kantorovich-Rubinstein duality [29, 63] states that

$$W_1(\mu,\nu) = \sup\{\int_\Omega f \,\mathrm{d}(\mu-\nu) \ :\ \mathrm{Lip}(f) \le 1\}.$$

A particularly interesting fact is that this metric only depends on the difference $\mu - \nu$. In fact, by setting

$$\|\mu\|_{\mathrm{Lip}^*} = \sup\{\int_\Omega f \,\mathrm{d}\mu \ :\ \mathrm{Lip}(f) \le 1\}$$

one obtains the so-called dual Lipschitz norm on the space of measures with zero mean and finite first moments (cf. [5, §8.10(viii)] where it is called modified Kantorovich-Rubinstein norm). Note that the supremum is unbounded if μ has a nonzero mean. To prevent the norm from blowing up in this case, and hence, to obtain a norm on the space of all signed measures with finite first moments, one can add the constraint that the test functions f shall be bounded. This leads to the expression

$$\sup\{\int_\Omega f \,\mathrm{d}\mu \ :\ |f| \le 1,\ \mathrm{Lip}(f) \le 1\},$$

which is called Kantorovich-Rubinstein norm in [5, §8.3]. Since we would like the bound on the values of f and the bound on its Lipschitz constant to vary independently in the following, we introduce for $\lambda = (\lambda_1, \lambda_2)$ the norm

$$\|\mu\|_{\mathrm{KR},\lambda} = \sup\{\int_\Omega f \,\mathrm{d}\mu \ :\ |f| \le \lambda_1,\ \mathrm{Lip}(f) \le \lambda_2\}. \tag{2}$$

Note that in the extreme cases $\lambda_1 = \infty$ and $\lambda_2 = \infty$ we recover the dual Lipschitz and the Radon norm

$$\begin{aligned} \|\mu\|_{\mathrm{KR},(\infty,1)} &= \|\mu\|_{\mathrm{Lip}^*} \\ \|\mu\|_{\mathrm{KR},(1,\infty)} &= \|\mu\|_{\mathfrak{M}}. \end{aligned} \tag{3}$$

Note that the norm $\|\mu\|_{\mathrm{KR},(\lambda_1,\lambda_2)}$ with $\lambda_1, \lambda_2 > 0$ is equivalent to the bounded Lipschitz norm [63, §6] where one takes the supremum over all functions f such that $|f| + \mathrm{Lip}(f) \le 1$. In general we have the following simple estimates:

Lemma 2.1 (Estimates by the Radon norm)**.** *For any* $\lambda = (\lambda_1, \lambda_2) \ge 0$ *it holds that*

$$\|\mu\|_{\mathrm{KR},\lambda} \le \lambda_1 \|\mu\|_{\mathfrak{M}}.$$

If μ is non-negative it holds that

$$\|\mu\|_{\mathrm{KR},\lambda} = \lambda_1 \|\mu\|_{\mathfrak{M}}.$$

If Ω has finite diameter $\operatorname{diam}(\Omega)$, then it holds for any μ with $\int_\Omega \mathrm{d}\mu = 0$ that

$$\|\mu\|_{\mathrm{KR},\lambda} \le \lambda_2 \tfrac{\operatorname{diam}(\Omega)}{2} \|\mu\|_{\mathfrak{M}}.$$

Proof. The first inequality follows directly from the definition of $\|\mu\|_{\mathrm{KR},\lambda}$ by dropping the constraint $|\nabla f| \le \lambda_2$ and the second claim by observing that the supremum is attained at $f \equiv \lambda_1$.

For the last claim we estimate from above by dropping the constraint $\|f\|_\infty \le \lambda_1$. However, since Ω has bounded diameter and μ has mean value zero, the constraint $\||\nabla f|\|_\infty \le \lambda_2$ implies that one also has a bound $\|f\|_\infty \le \lambda_2 \operatorname{diam}(\Omega)/2$ (indeed, $\lambda_2 \operatorname{diam}(\Omega)$ is a bound on the value $\max f - \min f$, however, since $\int_\Omega \mathrm{d}\mu = 0$, we may add a constant to f without altering the outer supremum). We obtain

$$\|\mu\|_{\mathrm{KR},\lambda_1,\lambda_2} \le \sup_{\|f\|_\infty \le \lambda_2 \operatorname{diam}(\Omega)/2} \int f \, \mathrm{d}\mu \le \lambda_2 \operatorname{diam}(\Omega) \|\mu\|_{\mathfrak{M}}/2.$$

□

Remark 2.2. Note that the KR-norm may not be bounded from below by the Radon norm in general: For $\mu = \delta_{x_0} + \delta_{x_1}$ it holds that $\|\mu\|_{\mathfrak{M}} = 2$ while $\|\mu\|_{\mathrm{KR},\lambda} \to 0$ for $|x_0 - x_1| \to 0$.

3 Primal formulations of the KR-norm

We present two reformulations of the KR-norm. The first, only shown formally, is similar to the Kantorovich-Rubinstein duality and shows the relation to optimal transport.

The idea for the first reformulation is to replace the constraint $\operatorname{Lip}(f) \le \lambda_2$ by a pointwise constraint of the form $|f(x) - f(y)| \le \lambda_2 |x - y|$, i.e., we have

$$\|\mu\|_{\mathrm{KR},\lambda} = \sup\{\int f \, \mathrm{d}\mu \ : \ |f(x)| \le \lambda_1, \ |f(x) - f(y)| \le \lambda_2 |x - y|\}.$$

We express the pointwise constraints by $f(x) - \lambda_1 \le 0$, $-f(x) - \lambda_1 \le 0$, $f(x) - f(y) - \lambda_2|x - y| \le 0$ and $f(y) - f(x) - \lambda_2|x - y| \le 0$, introduce Lagrange multipliers and clean up the resulting expression and finally arrive at

$$\|\mu\|_{\mathrm{KR},\lambda} = \inf_{\gamma \ge 0} \left[\lambda_1 \int_\Omega \mathrm{d}|\mu - \operatorname{proj}_1 \gamma + \operatorname{proj}_2 \gamma| + \lambda_2 \int_{\Omega \times \Omega} |x - y| \, \mathrm{d}\gamma \right]. \tag{4}$$

This expression may be compared to the following variant from [53]

$$\|\mu\|_{\mathrm{KR}'} = \inf_{\gamma \ge 0} \{ \int_{\Omega \times \Omega} |x - y| \, \mathrm{d}\gamma \ : \ \operatorname{proj}_1 \gamma - \operatorname{proj}_2 \gamma = \mu \},$$

which is a "strict constraint" version of (4). Because we have a metric cost function $(x, y) \mapsto |x-y|$, this is the same as requiring $\operatorname{proj}_1 \gamma = \mu^+, \operatorname{proj}_2 \gamma = \mu^-$ and we recover the Wasserstein metric with $p = 1$ from (1).

We get another reformulation by dualizing the problem slightly differently. The idea is to reformulate the constraint $\operatorname{Lip}(f) \leq \lambda_2$ with the help of the distributional derivative of f as $\||\nabla f|\|_\infty \leq \lambda_2$. This is allowed since for bounded, convex and open domains Ω, it is indeed the case that $\||\nabla f|\|_\infty = \operatorname{Lip}(f)$ (cf. [1, Prop. 2.13]). Through this reformulation, the KR-norm can be seen to be equivalent to the flat norm in the theory of currents [21, 43].

Lemma 3.1. *Let $\Omega \subset \mathbb{R}^n$ be open, convex, and bounded, and let $\lambda = (\lambda_1, \lambda_2) \geq 0$. Then it holds that*

$$\|\mu\|_{\mathrm{KR},\lambda} = \min_{\vec{\nu} \in \mathfrak{M}(\overline{\Omega}, \mathbb{R}^n)} \lambda_1 \|\mu - \operatorname{div} \vec{\nu}\|_{\mathfrak{M}} + \lambda_2 \||\vec{\nu}|\|_{\mathfrak{M}} \tag{5}$$

where $\operatorname{div} \vec{\nu}$ is understood to be taken in $\overline{\Omega}$ or, equivalently, in any open set U containing $\overline{\Omega}$.

Proof. Using indicator functions, we have

$$\|\mu\|_{\mathrm{KR},\lambda} = \sup_f \int_\Omega f \,\mathrm{d}\mu - I_{\{\|\cdot\|_\infty \leq \lambda_1\}}(f) - I_{\{\||\cdot|\|_\infty \leq \lambda_2\}}(\nabla f).$$

Now let U be an open set containing $\overline{\Omega}$, define the Banach spaces $X = C_c^1(U)$ and $Y = C_0(U, \mathbb{R}^n)$, and the subsets

$$A = \{f \in X \ : \ \sup_{x \in \overline{\Omega}} |f(x)| \leq \lambda_1\}$$
$$B = \{\vec{g} \in Y \ : \ \sup_{x \in \overline{\Omega}} |\vec{g}(x)| \leq \lambda_2\}.$$

Further define functionals $F : X \to \mathbb{R} \cup \{\infty\}$ and $G : Y \to \mathbb{R} \cup \{\infty\}$ by

$$F(f) = -\int_\Omega f \,\mathrm{d}\nu + I_A(f), \qquad G(\vec{g}) = I_B(\vec{g})$$

as well as the linear operator $K = \nabla : X \to Y$. With this notation we have

$$\|\mu\|_{\mathrm{KR},\lambda} = \sup_{f \in X} -F(f) - G(Kf).$$

To use the Fenchel-Rockafellar duality [20] we use the constraint qualification from [3], i.e., that it holds that

$$\bigcup_{\alpha > 0} \alpha[\operatorname{dom}(G) - K \operatorname{dom}(F)] \supset \bigcup_{\alpha > 0} \alpha A = Y.$$

Hence, we have

$$\sup_{f \in X} -F(f) - G(Kf) = \inf_{\nu \in Y^*} F^*(-K^* \nu) + G^*(\nu).$$

We have $X^* = \mathfrak{M}(U)$ and $Y^* = \mathfrak{M}(U, \mathbb{R}^n)$ and the conjugate functions of F and G are expressed with the help of the sets

$$C = \{\eta \in \mathfrak{M}(U) \ : \ |\eta|(U \setminus \overline{\Omega}) = 0\}$$
$$D = \{\vec{\nu} \in \mathfrak{M}(U, \mathbb{R}^n) \ : \ |\vec{\nu}|(U \setminus \overline{\Omega}) = 0\}$$

as

$$F^*(\eta) = \lambda_1 \|\mu + \eta\|_{\mathfrak{M}(\overline{\Omega})} + I_C(\eta), \qquad G^*(\vec{\nu}) = \lambda_2 \||\vec{\nu}|\|_{\mathfrak{M}(\overline{\Omega})} + I_D(\vec{\nu}).$$

Since by the Kirszbraun theorem every f that is Lipschitz continuous on Ω can be extended to U (with preservation of the Lipschitz constant) it follows with $K^* = -\operatorname{div} : Y^* \to X^*$ that

$$\begin{aligned}\|\mu\|_{\mathrm{KR},\lambda} &= \inf_{\vec{\nu} \in Y^*} F^*(-K^* \vec{\nu}) + G^*(\vec{\nu}) \\ &= \inf_{\nu \in \mathfrak{M}(U, \mathbb{R}^n)} \lambda_1 \|\mu - \operatorname{div} \vec{\nu}\|_{\mathfrak{M}(U)} + \lambda_2 \||\vec{\nu}|\|_{\mathfrak{M}(U)} + I_C(\operatorname{div} \vec{\nu}) + I_D(\vec{\nu}).\end{aligned}$$

Since bounded sets in $\mathfrak{M}(U, \mathbb{R}^n)$ are relatively weakly* compact, we can replace the infimum by a minimum and since $\operatorname{supp} \vec{\nu} \subset \overline{\Omega}$ implies that $\operatorname{supp} \operatorname{div} \vec{\nu} \subset \overline{\Omega}$ we can replace $\mathfrak{M}(U, \mathbb{R}^n)$ by $\mathfrak{M}(\overline{\Omega}, \mathbb{R}^n)$ and drop the constraints C and D and arrive at

$$\|\mu\|_{\mathrm{KR},\lambda} = \min_{\nu \in \mathfrak{M}(\overline{\Omega}, \mathbb{R}^n)} \lambda_1 \|\mu - \operatorname{div} \vec{\nu}\|_{\mathfrak{M}(\overline{\Omega})} + \lambda_2 \||\vec{\nu}|\|_{\mathfrak{M}(\overline{\Omega})}$$

as desired. □

In Theorem 3.4 below we will prove that actually we can take $\vec{\nu}$ as an L^1 vector field with L^1 divergence in (5). Namely $\vec{\nu} \in W^{1,1}(\Omega; \operatorname{div})$, where for $\Omega \subset \mathbb{R}^n$ an open domain, we define

$$W^{1,1}(\Omega; \operatorname{div}) := \{\vec{\nu} \in L^1(\Omega; \mathbb{R}^n) \mid \operatorname{div} \vec{\nu} \in L^1(\Omega)\}.$$

As such, our result is closely related to the work in [17], where this L^1 property is proved for the transport density $|\vec{\nu}|$. Our proof is however different and shorter, based on the following simpler geometric estimate.

Lemma 3.2. *Let $\Omega \subset \mathbb{R}^n$ be convex, open and bounded, and $\mu = \sum_{i=1}^N \alpha_i \delta_{x_i}$. Then any optimal solution ν to* (5) *has the form $\nu = \sum_{j=1}^M \beta_j [\![a_j, b_j]\!]$, where $a_j, b_j = x_i$ for some i. Moreover, the transport rays $[a_j, b_j]$ are* approximately parallel *in the following sense: there exist constants $c = c(n)$ and $\kappa = \kappa(n)$ such that if $[a_j, b_j] \cap B(x, \rho) \neq \emptyset$ and $[a_k, b_k] \cap B(x, \rho) \neq \emptyset$ with $a_j, b_j, a_k, b_k \notin B(x, c\rho)$, then $[a_j, b_j]$ and $[a_k, b_k]$ satisfy $a_j, b_j, a_k, b_k \in B(x, 2\kappa\rho) + \mathbb{R}z$ for some unit vector z.*

Proof. The claim that ν has the form $\nu = \sum_{j=1}^M \beta_j [\![a_j, b_j]\!]$ is trivial, as the problem in (5) with discrete μ is a simple combinatorial problem.

Suppose $[a_j, b_j] \cap B(x, \rho) \neq \emptyset$ and $[a_k, b_k] \cap B(x, \rho) \neq \emptyset$, and that $a_j, b_j, a_k, b_k \notin B(x, c\rho)$, for c yet to be determined. If $n = 2$, let $\bar{a}_j := a_j$, $\bar{b}_j := b_j$, $\bar{a}_k := a_k$, and $\bar{b}_k := b_k$. Also set $d := 0$, and $v := 0$. Otherwise, if $n > 2$, let $v \in \mathbb{R}^n$ be the vector giving the minimum distance between the lines

$$L_j := a_j + \mathbb{R}(b_j - a_j), \quad \text{and} \quad L_k := a_k + \mathbb{R}(b_k - a_k).$$

We may then find a plane $P \subset \mathbb{R}^n$ orthogonal to v such that $L_j \subset P$ and $L_k \subset v + P$. After rotation and translation, if necessary, we may without loss of generality assume that $v = (0, d) \in \mathbb{R}^n$ for some $d \in \mathbb{R}^{n-2}$, and

$$a_j = (\bar{a}_j, 0),\ b_j = (\bar{b}_j, 0), \quad \text{and} \quad a_k = (\bar{a}_k, d),\ b_k = (\bar{b}_k, d).$$

We also denote $x = (\bar{x}, x_0)$. Since L_j and L_k lie on the planes P and $v + P$ at a constant distance $\|d\| \leq 2\rho$ apart, we find that $\bar{a}_j, \bar{b}_j, \bar{a}_k, \bar{b}_k \notin B(\bar{x}, \gamma_n c\rho)$, for some dimensional constant $\gamma_n \in (0, 1)$. In fact, we may assume by shifting all of the points closer towards x that

$$\bar{a}_j, \bar{b}_j, \bar{a}_k, \bar{b}_k \in \partial B(\bar{x}, \gamma_n c\rho),$$

This is possible with $c > 1$ as the segments $[\bar{a}_j, \bar{b}_j]$ and $[\bar{a}_k, \bar{b}_k]$ pass through $B(\bar{x}, \rho)$, and so we may split each segment into three parts – two outside $B(\bar{x}, \gamma_n c\rho)$, and one inside.

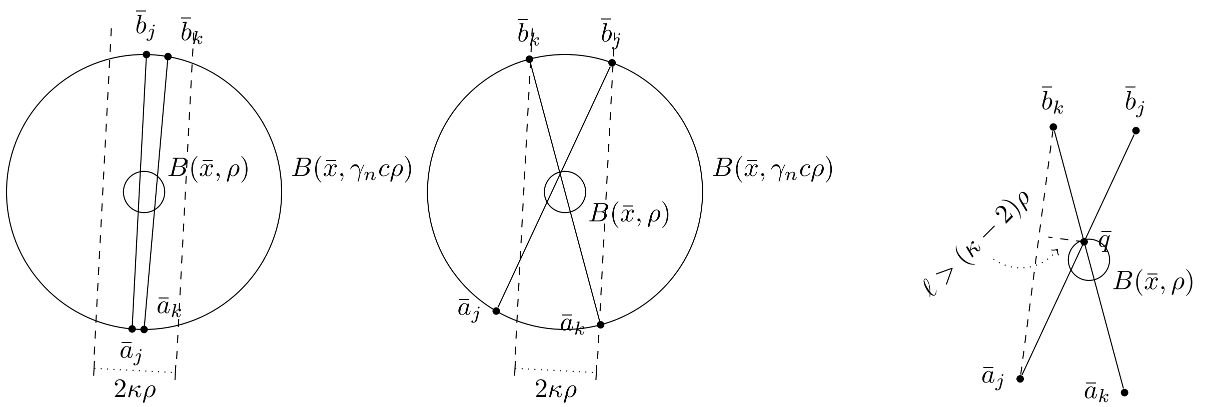
Let $\kappa > 2$. Observe now that in case $n = 2$ and generally for $n > 2$, when looking from the direction v, we have one of the two-dimensional situation depicted in Figure 1a or b. The segments $[\bar{a}_j, \bar{b}_j]$ and $[\bar{a}_k, \bar{b}_k]$, starting and ending on $\partial B(\bar{x}, \gamma_n c\rho)$, both pass through approximately ($c \gg 1$) in the middle of this sphere, through $\partial B(x, \rho)$. They are either within a cylinder of width $2\kappa\rho$, as in Figure 1b, or are not, as in Figure 1a.

If $\|a_j - a_k\| < \kappa\rho$ and c is large enough that $B(x, \rho)$ reduces to almost to a point in comparison to $B(x, \gamma_n c\rho)$, then $\|\bar{b}_j - \bar{b}_k\| < 2\kappa\rho$. This is because both segments $[\bar{a}_j, \bar{b}_j]$ and $[\bar{a}_k, \bar{b}_k]$ also pass through the ball $B(x, \rho)$ and so cannot diverge much on the opposite side of the ball. Trivially a unit vector z exists, such that both segments lie in the cylinder $B(x, 2\kappa\rho) + \mathbb{R}z$. Otherwise, for large enough c, both $|\bar{a}_j - \bar{a}_k| \geq \kappa\rho$ as well as $|\bar{b}_j - \bar{b}_k| > \kappa\rho$. Since $d \leq 2\rho < \kappa\rho$, i.e., some midpoints of the segments are closer than the end points, we observe that the two segments have to cross. That is $[\bar{a}_j, \bar{b}_j] \cap [\bar{a}_k, \bar{b}_k] = \bar{q}$ for some $\bar{q}$. If c and κ are large enough that $B(x, \rho)$ reduces to a point in comparison to everything else, we can make $\bar{q} \in B(\bar{x}, \rho)$. By simple geometrical reasoning, on the triangle $\bar{a}_j - \bar{q} - \bar{b}_k$, compare Figure 1c, it now follows that

$$|\bar{a}_j - \bar{b}_k| \leq \sqrt{|\bar{a}_j - \bar{q}|^2 - (\kappa - 2)^2\rho^2} + \sqrt{|\bar{b}_k - \bar{q}|^2 - (\kappa - 2)^2\rho^2}.$$

Likewise

$$|\bar{a}_k - \bar{b}_j| \leq \sqrt{|\bar{a}_k - \bar{q}|^2 - (\kappa - 2)^2\rho^2} + \sqrt{|\bar{b}_j - \bar{q}|^2 - (\kappa - 2)^2\rho^2}.$$



(a) Approximately parallel case (b) Intersecting case (c) Improvement of transport ray

Figure 1: Illustration of the two-dimensional projection in the proof of Lemma 3.2.

If $n = 2$, or more generally $d = 0$, it trivially follows that

$$\begin{aligned} |a_j - b_k| + |a_k - b_j| &< |a_j - q| + |b_k - q| + |a_k - q| + |b_j - q| \\ &= |a_j - b_j| + |a_k - b_k|. \end{aligned}$$

Otherwise, minding that $|d| \leq 2\rho$ and $\kappa > 2$, we calculate

$$\begin{aligned} |a_j - b_k| + |a_j - b_k| &= \sqrt{|\bar{a}_j - \bar{b}_k|^2 + |d|^2} + \sqrt{|\bar{a}_k - \bar{b}_j|^2 + |d|^2} \\ &\leq \sqrt{(|\bar{a}_j - \bar{q}| + |\bar{b}_k - \bar{q}|)^2 - 2(\kappa-2)^2\rho^2 + d^2} \\ &\quad + \sqrt{(|\bar{a}_k - \bar{q}| + |\bar{b}_j - \bar{q}|)^2 - 2(\kappa-2)^2\rho^2 + d^2} \\ &< |a_j - q| + |b_k - q| + |a_k - q| + |b_j - q| \\ &= |a_j - b_j| + |a_k - b_k|. \end{aligned}$$

This provides a contradicion to the optimality of the transport rays $[a_j, b_j]$ and $[a_k, b_k]$, and shows the claim. □

Remark 3.3. If $n = 2$, we can take $\kappa = 2$, and the argument is simplified considerably.

Theorem 3.4. *Suppose $\Omega \subset \mathbb{R}^n$ is convex, open, and bounded, and $\mu \in L^1(\Omega)$. Then*

$$\|\mu\|_{\mathrm{KR},\lambda_1,\lambda_2} = \min_{\nu \in W^{1,1}(\Omega;\mathrm{div})} \lambda_1 \|\mu - \operatorname{div}\nu\|_{L^1(\Omega;\mathbb{R}^n)} + \lambda_2 \|\nu\|_{L^1(\Omega)}. \tag{6}$$

Moreover the minimum is reached by ν satisfying $\int_\Omega \operatorname{div}\nu \, \mathrm{d}\mathfrak{L}^n = 0$.

Proof. We assume first that $\mu \in L^\infty(\Omega)$. By Lemma 3.1, we have (5). To replace $\overline{\Omega}$ by Ω, we just have to show that that $|\nu|(\partial\Omega) = 0$ for any ν reaching the minimum in (5). This follows if $\nu \ll \mathfrak{L}^n$. Hence it suffices to show that actually ν and $\operatorname{div}\nu$ are also absolutely continuous with respect to $\mathfrak{L}^n$. This is where we need the convexity of Ω and the absolute continuity of μ.

Clearly by (5) we have

$$\|\mu\|_{\mathrm{KR},\lambda_1,\lambda_2} \leq \min_{\nu \in W^{1,1}(\Omega;\operatorname{div})} \lambda_1 \|\mu - \operatorname{div}\nu\|_{L^1(\Omega;\mathbb{R}^n)} + \lambda_2 \|\nu\|_{L^1(\Omega)},$$

so it remains to show the opposite inequality. We approximate μ in terms of strict convergence of measures by $\{\mu^i\}_{i=1}^\infty$, where $\mu^i = \sum_{j=1}^{N_i} \alpha_{i,j}\delta_{x_{i,j}}$. We may clearly assume that $x_{i,j} \in \Omega$, because $|\mu|(\partial\Omega) = 0$ by absolutely continuity. Moreover, given a sequence $\epsilon_i \searrow 0$, we may assume that there exist Voronoi cells $V_{i,j} \subset B(x_{i,j}, \epsilon_i)$, such that $\alpha_{i,j} = \int_{V_{i,j}} \mu(x)\,\mathrm{d}x$, as well as

$$V_{i,j} \cap V_{i,k} = \emptyset, (i \neq k), \quad \text{and} \quad \operatorname{supp}\mu \subset \bigcup_{j=1}^{N_i} V_{i,j}, \quad (i = 1, \ldots, N_i). \tag{7}$$

Then (5) is a finite-dimensional discrete/combinatorial problem, and we easily discover an optimal solution ν^i. Because tranporting mass outside Ω incurs a cost on $\partial\Omega$, we see that

$$\nu^i = \sum_{j=1}^{M_i} \beta_{i,j} [\![a_{i,j}, b_{i,j}]\!],$$

for some $\beta_{i,j} > 0$ and $a_{i,j}, b_{i,j} \in \{x_{i,1}, \ldots, x_{i,N_i}\}$. We calculate

$$\operatorname{div}[\![a, b]\!] = \delta_b - \delta_a.$$

Moreover

$$\operatorname{div}\nu^i(\overline{\Omega}) = \operatorname{div}\nu^i(\Omega) = 0, \quad \text{and} \quad \operatorname{div}\nu^i \ll |\mu^i|. \tag{8}$$

As minimisers, we have

$$\|\nu^i\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)} \leq \frac{\lambda_1}{\lambda_2}\|\mu^i\|_{\mathfrak{M}(\overline{\Omega})} \leq \frac{\lambda_1}{\lambda_2}\|\mu\|_{\mathfrak{M}(\overline{\Omega})}.$$

Therefore, after possibly moving to a subsequence, unrelabelled, we may assume that $\nu^i \xrightarrow{*} \nu$ for some $\nu \in \mathfrak{M}(\overline{\Omega};\mathbb{R}^n)$. But by (8) we may also assume that $\operatorname{div}\nu^i \xrightarrow{*} \lambda \in \mathfrak{M}(\overline{\Omega})$, where $\lambda \ll |\mu|$. From this absolute continuity it follows that $\lambda(\overline{\Omega}) = 0$. (A priori it might be that $\lambda(\overline{\Omega}) \neq 0$.) Necessarily $\lambda = \operatorname{div}\nu$, so that in particular $\operatorname{div}\nu \ll \mathfrak{L}^n$. Because $\partial\Omega$ is $\mathfrak{L}^n$-negligible, it follows that $\operatorname{div}\nu(\Omega) = 0$.

We want to show that ν is an optimal solution to (5) for μ. We do this as follows. With i fixed, within each $V_{i,j}$, $(j = 1, \ldots, N_i)$, we may construct a map $\nu_{i,j}$ transporting the mass of μ within the cell $V_{i,j}$ to the cell centre $\delta_{x_{i,j}}$, or the other way around. That is

$$\operatorname{div}\nu_{i,j} = \mu\chi_{V_{i,j}} - \alpha_{i,j}\delta_{x_{i,j}}$$

with

$$\|\nu_{i,j}\| \leq \epsilon_i \int_{V_{i,j}} |\mu(x)| \, \mathrm{d}x.$$

It follows that

$$\sum_{j=1}^{N_i} \|\nu_{i,j}\| \leq \epsilon_i \|\mu\|.$$

If now ν^* is an optimal solution to (5) for μ, defining

$$\nu_0^i := \nu^* - \sum_{j=1}^{N_i} \nu_{i,j},$$

we see that

$$\|\nu_0^i\|_{\mathfrak{M}(\overline{\Omega})} \leq \|\nu^*\|_{\mathfrak{M}(\overline{\Omega})} + C\epsilon_i$$

and

$$\operatorname{div} \nu_0^i = \operatorname{div} \nu^* - \mu + \mu^i.$$

Thus

$$\begin{aligned}
\lambda_1 \|\mu^i - \operatorname{div} \nu^i\|_{\mathfrak{M}(\overline{\Omega})} + &\lambda_2 \|\nu^i\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)} \\
&\leq \lambda_1 \|\mu^i - \operatorname{div} \nu_0^i\|_{\mathfrak{M}(\overline{\Omega})} + \lambda_2 \|\nu_0^i\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)} \\
&\leq \lambda_1 \|\mu - \operatorname{div} \nu^*\|_{\mathfrak{M}(\overline{\Omega})} + \lambda_2 \|\nu^*\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)} + C\epsilon_i.
\end{aligned}$$

By weak* lower semicontinuity

$$\begin{aligned}
\lambda_1 \|\mu - \operatorname{div} \nu\|_{L^1(\Omega)} + &\lambda_2 \|\nu\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)} \\
&\leq \liminf_{i\to\infty} \big(\lambda_1 \|\mu^i - \operatorname{div} \nu^i\|_{\mathfrak{M}(\overline{\Omega})} + \lambda_2 \|\nu^i\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)}\big) \\
&\leq \liminf_{i\to\infty} \big(\lambda_1 \|\mu - \operatorname{div} \nu^*\|_{\mathfrak{M}(\overline{\Omega})} + \lambda_2 \|\nu^*\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)} + C\epsilon_i\big) \\
&= \lambda_1 \|\mu - \operatorname{div} \nu^*\|_{\mathfrak{M}(\overline{\Omega})} + \lambda_2 \|\nu^*\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)}.
\end{aligned}$$

Thus ν is an optimal solution to (5) for μ. Exploiting lower semicontinuity of both of the terms, we moreover see that $\lim_{i\to\infty} \|\nu^i\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)} = \|\nu\|_{\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)}$. Thus $\{\nu^i\}_{i=1}^\infty$ converge to ν strictly in $\mathfrak{M}(\overline{\Omega};\mathbb{R}^n)$. Likewise $\{\mu^i - \operatorname{div} \nu^i\}_{i=1}^\infty$ converge to $\mu - \operatorname{div} \nu$ strictly in $\mathfrak{M}(\overline{\Omega})$. But $\{\mu^i\}_{i=1}^\infty$ were already constructed to converge strictly to μ, and we have above seen that $(\operatorname{div} \nu^i)^\pm \leq (\mu^i)^\pm$. Therefore also $\{\operatorname{div} \nu^i\}_{i=1}^\infty$ converge to $\operatorname{div} \nu$ strictly in $\mathfrak{M}(\overline{\Omega})$.

It remains to show that $\nu \in W^{1,1}(\Omega; \operatorname{div})$. We have already shown $\operatorname{div} \nu \ll \mathfrak{L}^n \llcorner \Omega$, so that $\operatorname{div} \nu \in L^1(\Omega)$. We just have to show that $\nu \ll \mathfrak{L}^n \llcorner \Omega$ to show that $\nu \in L^1(\Omega; \mathbb{R}^n)$. We do this by bounding the n-dimensional density of ν at each point. Let $M := \|\mu\|_{L^\infty(\Omega)}$. We now refer to Lemma 3.2, and approximate the mass of the set of approximately parallel transport rays passing through

$B(x,\rho)$ by

$$\begin{aligned}
&\max_{\|z\|=1} \sum_{a_{i,j},b_{i,j}\in(B(x,\kappa\rho)+\mathbb{R}z)\cap\Omega} \beta_{i,j}\mathfrak{H}^1(B(x,\rho)\cap[a_{i,j},b_{i,j}]) \\
&\quad\le \max_{\|z\|=1} \sum_{a_{i,j},b_{i,j}\in(B(x,\kappa\rho)+\mathbb{R}z)\cap\Omega} \beta_{i,j}2\rho \\
&\quad\le \max_{\|z\|=1} \sum_{x_{i,j}\in(B(x,\kappa\rho)+\mathbb{R}z)\cap\Omega} |\alpha_{i,j}|2\rho \\
&\quad\le 2\rho \max_{\|z\|=1} \sum_{x_{i,j}\in(B(x,\kappa\rho)+\mathbb{R}z)\cap\Omega} \int_{V_{i,j}} |\mu(y)|\,\mathrm{d}y \\
&\quad\le 2\rho \max_{\|z\|=1} \int_{B(x,\kappa\rho+\epsilon_i)+z\mathbb{R}} |\mu(y)|\,\mathrm{d}y \\
&\quad\le 2\rho(\kappa\rho+\epsilon_i)^{n-1}\operatorname{diam}(\Omega)M.
\end{aligned}$$

Also the mass of the set of transport rays with start or end point in $B(x,c\rho)$ may be approximated by

$$\begin{aligned}
&\sum_{a_{i,j}\in B(x,c\rho))} \beta_{i,j}\mathfrak{H}^1(B(x,\rho)\cap[a_{i,j},b_{i,j}]) + \sum_{b_{i,j}\in B(x,c\rho))} \beta_{i,j}\mathfrak{H}^1(B(x,\rho)\cap[a_{i,j},b_{i,j}]) \\
&\quad\le \sum_{x_{i,j}\in B(x,c\rho))} 4\alpha_{i,j}\rho \\
&\quad= \sum_{x_{i,j}\in B(x,c\rho))} 4\rho\int_{V_{i,j}} |\mu(y)|\,\mathrm{d}y \\
&\quad\le 4\rho\int_{B(x,c\rho+\epsilon_i)} |\mu(y)|\,\mathrm{d}y.
\end{aligned}$$

It now follows that

$$|\nu^i|(B(x,\rho)) \le 4\rho\int_{B(x,c\rho+\epsilon_i)} |\mu(y)|\,\mathrm{d}y + 2\rho(2\kappa\rho+\epsilon_i)^{n-1}\operatorname{diam}(\Omega)M$$

Letting $i\to\infty$, we get by lower semicontinuity

$$|\nu|(B(x,\rho)) \le 4\rho\int_{B(x,c\rho)} |\mu(y)|\,\mathrm{d}y + 2^n\kappa^{n-1}\rho^n\operatorname{diam}(\Omega)M$$

Thus

$$\lim_{\rho\searrow 0}\frac{|\nu|(B(x,\rho))}{\mathfrak{L}^n(B(x,\rho))} \le 0 + 2^n\kappa^{n-1}\operatorname{diam}(\Omega)M$$

It follows (see [35, Theorem 2.12]) that $\nu \ll \mathfrak{L}^n\llcorner\Omega$ with

$$\|\nu\|_{L^1(\Omega;\mathbb{R}^n)} \le 2^n\kappa^{n-1}\operatorname{diam}(\Omega)M\mathfrak{L}^n(\Omega).$$

Finally, we consider the case of unbounded $\mu\in L^1(\Omega)$. We take

$$\mu_M(x) := \max\{-M,\min\{\mu(x),M\}\},\quad (M=1,2,3,\ldots).$$

Then $\mu_M^{\pm} \leq \mu^{\pm}$. Applying the point-mass approximation above to both μ^k and μ, we can take $(\mu_M^i)^{\pm} \leq (\mu^i)^{\pm}$. Then by a simple argument we also have $|\nu_M^i| \leq |\nu^i|$ for each $i, k = 1, 2, 3, \ldots$; compare [17, Proposition 4.3]. Indeed, let $\tilde{\mu}_M^i := \operatorname{div} \nu_M^i$. Clearly

$$(\tilde{\mu}_M^i)^{\pm} \leq (\mu_M^i)^{\pm} \leq (\mu^i)^{\pm}.$$

We can therefore find a measure $\tau_M^i \in \mathfrak{M}(\Omega; \mathbb{R}^n)$ with $|\tau_M^i| \leq |\nu^i|$ such that $\operatorname{div} \tau_M^i = \tilde{\mu}_M^i$. If τ_M^i is not optimal, then we find a contradiction to ν^i being optimal by replacing it with $\nu^i + \nu_M^i - \tau_M^i$. We may therefore assume that $\nu_M^i = \tau_M^i$. Consequently $|\nu_M^i| \leq |\nu^i|$. Similarly we prove that $|\nu_M^i| \leq |\nu_{M+1}^i|$. By the strict convergence of ν^i to ν, we now deduce that $|\nu_M| \leq |\nu|$ and $|\nu_M| \leq |\nu_{M+1}|$. By an analogous argument we prove that $(\operatorname{div} \nu_M^i)^{\pm} \leq (\operatorname{div} \nu^i)^{\pm}$, $(\operatorname{div} \nu_M^i)^{\pm} \leq (\operatorname{div} \nu_{M+1}^i)^{\pm}$, and consequently $(\operatorname{div} \nu_M)^{\pm} \leq (\operatorname{div} \nu)^{\pm}$ and $(\operatorname{div} \nu_M)^{\pm} \leq (\operatorname{div} \nu_{M+1})^{\pm}$. Also $|\operatorname{div} \nu_M|(\Omega) \to |\operatorname{div} \nu|(\Omega)$, because

$$\|\operatorname{div} \nu - \operatorname{div} \nu_M\|_{\mathfrak{M}(\Omega)} \leq \|\mu - \mu_M\|_{\mathfrak{M}(\Omega)}.$$

(This can be verified by the point-mass approximation.) It follows that $\operatorname{div} \nu_M \to \operatorname{div} \nu$ strongly. In particular $\operatorname{div} \nu_M - \mu_M \to \operatorname{div} \nu - \mu$ strongly. By lower semicontinuity of $\|\cdot\|_{\mathrm{KR},\lambda_1,\lambda_2}$ we therefore deduce that $\liminf_{M\to\infty} |\nu_M|(\Omega) \geq |\nu|(\Omega)$. Since $|\nu_M| \leq |\nu|$, it follows that $\nu_M \to \nu$ strongly in $\mathfrak{M}(\Omega; \mathbb{R}^n)$. But the above paragraphs say that $\nu_M \in L^\infty(\Omega)$. Thus necessarily $\nu_M \in L^1(\Omega)$. ☐

4 Kantorovich-Rubinstein-TV denoising

In this section we assume that Ω is a bounded, convex and open domain in $\mathbb{R}^n$ and study the minimization problem

$$\min_u \|u - u^0\|_{\mathrm{KR},\lambda} + \mathrm{TV}(u) \tag{9}$$

for some $u^0 \in L^1(\Omega)$ and $\lambda = (\lambda_1, \lambda_2) \geq 0$. We call this *Kantorovich-Rubinstein-*TV *denoising*, or short KR-TV denoising. Using the different forms of the KR-norm we have two different forms of the KR-TV denoising problem. The first uses the definition (2) but we replace the constraint $\mathrm{Lip}(f) \leq \lambda_2$ with the help of the distributional gradient as $|\nabla f| \leq \lambda_2$. Then problem (9) has the form

$$\min_u \max_{\substack{|f| \leq \lambda_1 \\ |\nabla f| \leq \lambda_2}} \int_\Omega f(u - u^0) + \mathrm{TV}(u). \tag{10}$$

We call this form, the *saddle point formulation.* Another formulation is obtained by using Theorem 3.4 to obtain

$$\min_{u,\vec{\nu}} \lambda_1 \|u - u^0 - \operatorname{div} \vec{\nu}\|_{L^1} + \lambda_2 \||\vec{\nu}|\|_{L^1} + \mathrm{TV}(u). \tag{11}$$

We call this the *cascading* or *dual formulation.*

Note that the optimal transport formulation (4) will not be used any further in this paper. The reason is, that this formulation does not seem to be suited for numerical purposes as it involves a measure on the domain $\Omega \times \Omega$ which leads, if discretized straightforwardly, to too large storage demands.

We denote

$$H_\lambda(u, f) = \begin{cases} \int f(u - u^0) + \mathrm{TV}(u), & \text{if } |f| \le \lambda_1,\ |\nabla f| \le \lambda_2 \\ -\infty, & \text{otherwise.} \end{cases} \tag{12}$$

Then, (10) reads as $\min_u \max_f H_{\lambda_1,\lambda_2}(u, f)$.

4.1 Relation to L^1-TV denoising

Similar to (3) one has $\|\mu\|_{\mathrm{KR},(\lambda_1,\infty)} = \lambda_1 \|\mu\|_{\mathfrak{M}}$ and for $u \in L^1(\Omega)$ it holds that $\|u\|_{\mathfrak{M}} = \|u\|_{L^1}$. Hence, KR-TV is a generalization of the successful L^1-TV denoising [15]:

$$\min_u \|u - u^0\|_{\mathrm{KR},(\lambda_1,\infty)} + TV(u) = \min_u \|u - u^0\|_{L^1} + \tfrac{1}{\lambda_1}\,\mathrm{TV}(u). \tag{13}$$

We will study the influence of the additional parameter λ_2 in Section 6.1 and 6.2 numerically. Note, however, that it is possible that the minimizer of (13) may also be a minimizer of (9) for λ_2 large enough but finite: To see this, we express L^1-TV as a saddle point problem by dualizing the L^1 norm to obtain

$$\min_u \max_{|f| \le \lambda_1} \int_\Omega f(u - u^0) + \mathrm{TV}(u).$$

We denote by $(\bar u, \bar f)$ a saddle point for this functional. If the function $\bar f$ is already Lipschitz continuous with constant L, then $(\bar u, \bar f)$ is also a solution of the saddle point problem

$$\min_u \max_{\substack{|f| \le \lambda_1 \\ \mathrm{Lip}(f) \le \lambda_2}} \int_\Omega f(u - u^0) + \mathrm{TV}(u)$$

for any $\lambda_2 \ge L$ and consequently, $\bar u$ is a solution of the KR-TV problem.

4.2 Relation to TGV denoising

The cascading formulation (11) reveals an interesting conceptional relation to the total generalized variation (TGV) model [7]. To define it, we introduce $S^{n\times n}$ as the set of symmetric $n \times n$ matrices and for a function v with values in $S^{n\times n}$ we set

$$(\operatorname{div} v(x))_i = \sum_{j=1}^n \frac{\partial v_{ij}}{\partial x_j}, \qquad \operatorname{div}^2 v(x) = \sum_{i,j=1}^n \frac{\partial^2 v_{ij}}{\partial x_j \partial x_i}.$$

The total generalized variation of order two for a parameter $\alpha = (\alpha_1, \alpha_2)$ is

$$\mathrm{TGV}^2_\alpha(u) = \sup\Big\{ \int_\Omega u \operatorname{div}^2 v \,\mathrm{d}x \;:\; v \in C^2_c(\Omega, S^{n\times n}), \\ |v(x)| \le \alpha_1,\ |\operatorname{div} v(x)| \le \alpha_2 \Big\}$$

The TGV term has an equivalent reformulation as follows: denote by $\mathrm{BD}(\Omega)$ the space of vector fields of bounded deformation, i.e. vector fields $\vec{w} \in L^1(\Omega, \mathbb{R}^n)$ such that the symmetrized distributional gradient $\mathcal{E}\vec{w} = \frac{1}{2}(\nabla\vec{w} + \nabla\vec{w}^T)$ is a $S^{n\times n}$-valued Radon measure. Then it holds that

$$\mathrm{TGV}^2_\alpha(u) = \inf_{\vec{w}\in\mathfrak{M}(\Omega,\mathbb{R}^n)} \alpha_1 \| |\mathcal{E}\vec{w}| \|_{\mathfrak{M}} + \alpha_2 \| |\nabla u - \vec{w}| \|_{\mathfrak{M}}$$

(cf. [8,9]), leading to the L^1-TGV^2 denoising problem

$$\min_{u\in L^1(\Omega),\ \vec{w}\in\mathfrak{M}(\Omega,\mathbb{R}^n)} \|u - u^0\|_{L^1} + \alpha_1 \| |\mathcal{E}\vec{w}| \|_{\mathfrak{M}} + \alpha_2 \| |\nabla u - \vec{w}| \|_{\mathfrak{M}}.$$

Note that this reformulation resembles the spirit of the reformulation of the Kantorovich-Rubinstein norm from Lemma 3.1

$$\|\mu\|_{\mathrm{KR},\lambda} = \min_{\vec{\nu}\in\mathfrak{M}(\overline{\Omega},\mathbb{R}^n)} \lambda_1 \|\mu - \operatorname{div}\vec{\nu}\|_{\mathfrak{M}} + \lambda_2 \| |\vec{\nu}| \|_{\mathfrak{M}}.$$

We obtain a new (semi-)norm by "cascading"a higher order term in a new minimization problem. In the TV case we go from $\mathrm{TV}(u) = \| |\nabla u| \|_{\mathfrak{M}}$ to TGV^2_α by cascading with a vector field and penalizing the symmetrized gradient of this vector field. In the KR case, however, we go from $\|u\|_{L^1} = \|u\|_{\mathfrak{M}}$ to $\|\cdot\|_{\mathrm{KR},\lambda}$ by cascading with the divergence of a vector field and penalizing with the Radon norm of that vector field. One may say, that TGV^2_α is a higher order generalization of the total variation while the KR-norm is a *lower order* generalization of the L^1 norm (or the Radon norm).

4.3 Relation to G-norm cartoon-texture decomposition

In [41] Meyer introduced the G-norm as a discrepancy term in denoising problems to allow for oscillating patterns in the denoised images. The G-norm is defined as

$$\|u\|_G = \inf\{\| |\vec{g}| \|_\infty \;:\; \operatorname{div}\vec{g} = u,\ g \in L^\infty\}.$$

Meyer proposed the following G-TV minimization problem

$$\min_u \lambda\|u - u_0\|_G + \mathrm{TV}(u) = \min_{u,\vec{g}} \lambda\| |\vec{g}| \|_\infty + \mathrm{TV}(u) + \delta_{\{0\}}(\operatorname{div}\vec{g} - (u - u_0)).$$

This differs from problem (11) in two aspects: First, $|\vec{g}|$ is penalized in the ∞-norm instead of the 1-like Radon norm and second, the equality $\operatorname{div}\vec{g} = u - u_0$ is enforced exactly, while in (11) a mismatch is allowed. The Meyer model has also been treated in numerous other papers, e.g. [4,19,30,65].

4.4 Properties of KR-TV denoising

Similar to the case of L^1-TV denoising (cf. [15, Lemma 5.5]) there exist thresholds for λ_1 and λ_2 such that the minimizer of (9) is u_0 (if u_0 is regular enough in some sense) if λ_1 and λ_2 are above the thresholds:

Theorem 4.1. *Let $u_0 \in BV(\Omega)$ and assume that there exists a continuously differentiable vector field $\vec{\phi}$ with compact support such that*

1. $|\vec{\phi}| \leq 1$ *and*
2. $\int u_0 \operatorname{div} \vec{\phi} = TV(u_0)$.

Then there exists thresholds λ_1^* and λ_2^* such that for $\lambda_1 > \lambda_1^*$ and $\lambda_2 > \lambda_2^*$, the unique minimizer of (9) *is u_0.*

Proof. For any $u \in BV$ we have

$$\begin{aligned}
\|u - u_0\|_{\mathrm{KR},\lambda_1,\lambda_2} + \mathrm{TV}(u) &\geq \int u \operatorname{div} \vec{\phi} + \Big[\min_{\vec{\nu}} \lambda_1 \|u - u_0 - \operatorname{div} \vec{\nu}\|_{\mathfrak{M}} + \lambda_2 \||\vec{\nu}|\|_{\mathfrak{M}}\Big] \\
&= \int u_0 \operatorname{div} \phi + \min_{\vec{\nu}} \Big[\lambda_1 \|u - u_0 - \operatorname{div} \vec{\nu}\|_{\mathfrak{M}} + \lambda_2 \||\vec{\nu}|\|_{\mathfrak{M}} \\
&\qquad + \int (u - u_0 - \operatorname{div} \vec{\nu}) \operatorname{div} \vec{\phi} + \int \operatorname{div} \vec{\nu} \operatorname{div} \vec{\phi}\Big] \\
&\geq \mathrm{TV}(u_0) + \min_{\vec{\nu}} \Big[(\lambda_1 - \|\operatorname{div} \vec{\phi}\|_\infty) \|u - u_0 - \operatorname{div} \vec{\nu}\|_{\mathfrak{M}} + \\
&\qquad (\lambda_2 - \||\nabla \operatorname{div} \vec{\phi}|\|_\infty) \||\vec{\nu}|\|_{\mathfrak{M}}\Big]
\end{aligned}$$

Hence, the values $\lambda_1^* = \|\operatorname{div} \vec{\phi}\|_\infty$ and $\lambda_2^* = \||\nabla \operatorname{div} \vec{\phi}|\|_\infty$ are valid thresholds as claimed. $\square$

Likewise there are thresholds in the opposite direction, again similarly to the L^1-TV case.

Theorem 4.2. *Let $\Omega \subset \mathbb{R}^n$ be a convex open domain with Lipschitz boundary. Then there exists a constant $C = C(\Omega)$ such that any solution $\bar{u}$ to* (9) *is a constant whenever $1/C > \lambda_1$.*

Proof. Let f maximize $H_\lambda(\bar{u}, \cdot)$. Define $\tilde{u}$ to be the constant function that equals the mean value of $\bar{u}$ over Ω, i.e.

$$\tilde{u} \equiv \fint_\Omega \bar{u}(x)\,\mathrm{d}x.$$

Let $\tilde{f}$ maximize $H_\lambda(\tilde{u}, \cdot)$. Since $\bar{u}$ solves (9), we have

$$H_\lambda(\tilde{u}, \tilde{f}) \geq H_\lambda(\bar{u}, f).$$

In other words, using $\mathrm{TV}(\tilde{u}) = 0$, writing out H_λ, and rearranging terms

$$\int_\Omega \tilde{f}(\bar{u} - u^0)\,\mathrm{d}x + \int_\Omega \tilde{f}(\tilde{u} - \bar{u})\,\mathrm{d}x \geq \int_\Omega f(\bar{u} - u^0)\,\mathrm{d}x + \mathrm{TV}(\bar{u}).$$

But, by the choice of f, we have

$$\int_\Omega \tilde{f}(\bar{u} - u^0)\,\mathrm{d}x \leq \int_\Omega f(\bar{u} - u^0)\,\mathrm{d}x.$$

Therefore

$$\mathrm{TV}(\bar{u}) \leq \int_\Omega \tilde{f}(\tilde{u} - \bar{u})\,\mathrm{d}x.$$

An application of Poincaré's inequality yields

$$\mathrm{TV}(\bar{u}) \leq \lambda_1 C\,\mathrm{TV}(\bar{u}).$$

This is a contradiction unless $1 < \lambda_1 C$ or $\mathrm{TV}(\bar{u}) = 0$, i.e., $\bar{u}$ is a constant. □

The second of the above two theorems shows that for small λ_1 one recovers a constant solution. In fact, this has to be $⨍_\Omega u^0\,\mathrm{d}x$. The first of the above two theorems shows that for parameters λ_1 and λ_2 large enough, one recovers the input u^0 from the KR-TV denoising problem. This behavior is similar to the L^1-TV denoising problem. If one leaves the regime of exact reconstruction one usually observes that for L^1-TV denoising mass disappears and also the phenomenon of "suddenly vanishing sets" (cf. [18]). In contrast, for the KR-TV denoising model, we have mass conservation of the minimizer even in the range of parameters, where exact reconstruction does not happen anymore and noise is being removed. The precise statement is given in the next theorem:

Theorem 4.3 (Mass preservation). *If $\frac{\lambda_2}{\lambda_1} \leq \frac{2}{\mathrm{diam}(\Omega)}$, then*

$$\min_u \|u - u^0\|_{\mathrm{KR},\lambda_1,\lambda_2} + \mathrm{TV}(u)$$

has a minimizer $\bar{u}$ such that $\int_\Omega \bar{u}(x)\,\mathrm{d}x = \int_\Omega u^0(x)\,\mathrm{d}x$.

Proof. The idea is, to prove that a minimizer of the KR-TV denoising problem with $\lambda_1 = \infty$ is also a minimizer of the problem with finite but large enough λ_1. Hence we start by denoting with $(\bar{u}, \bar{f})$ a solution of the following saddle-point problem:

$$\min_u \max_{|\nabla f| \leq \lambda_2} \int f(u - u^0)\,\mathrm{d}x + \mathrm{TV}(u) \tag{14}$$

With the notation (12), (14) reads as $\min_u \max_f H_{\infty,\lambda_2}(u, f)$.

It holds that $\int_\Omega \bar{u}\,\mathrm{d}x = \int_\Omega u^0\,\mathrm{d}x$, because otherwise, the max would be ∞. In other words: with $\lambda_1 = \infty$ we have mass preservation.

Now let $\frac{\lambda_2}{\lambda_1} \leq \frac{2}{\mathrm{diam}(\Omega)}$. We aim to show that there is constant c such that $(\bar{u}, \bar{f} + c)$ is a solution of

$$\min_u \max_{\substack{|f| \leq \lambda_1 \\ |\nabla f| \leq \lambda_2}} \int f(u - u^0)\,\mathrm{d}x + \mathrm{TV}(u). \tag{15}$$

Since $\bar{f}$ is Lipschitz with constant λ_2, we get that $\bar{f}(x) - \bar{f}(x) \leq \lambda_2 |x - y|$, and hence, $\max \bar{f} - \min \bar{f} \leq \lambda_2 \,\mathrm{diam}(\Omega)$. Consequently, there is a constant c such that

$$|\bar{f} + c| \leq \lambda_2 \frac{\mathrm{diam}(\Omega)}{2} \leq \lambda_1$$

in other words: $\bar{f} + c$ is feasible for (15). Since $\int \bar{u} = \int u^0$ we also have

$$H_{\lambda_1,\lambda_2}(\bar{u}, \bar{f} + c) = \int \bar{f}(\bar{u} - u^0)\,\mathrm{d}x + c \underbrace{\int (\bar{u} - u^0)\,\mathrm{d}x}_{=0} + TV(u) = H_{\infty,\lambda_2}(\bar{u}, \bar{f}).$$

Since all f that are feasible for (15) are also feasible for (14), we have for all these f that

$$H_{\lambda_1,\lambda_2}(\bar{u}, f) \leq H_{\infty,\lambda_2}(\bar{u}, f) \leq H_{\infty,\lambda_2}(\bar{u}, \bar{f}) = H_{\lambda_1,\lambda_2}(\bar{u}, \bar{f} + c). \tag{16}$$

Also we have by $(\bar{u}, \bar{f})$ being a saddle-point for all u that

$$H_{\lambda_1,\lambda_2}(\bar{u}, \bar{f} + c) = H_{\infty,\lambda_2}(\bar{u}, \bar{f}) \leq H_{\infty,\lambda_2}(u, \bar{f}).$$

But since $\mathrm{TV}(u) = \mathrm{TV}(u + d)$ for every constant d we also have with $d = c \int (u - u^0)\,\mathrm{d}x / \int \bar{f}\,\mathrm{d}x$ that

$$\begin{aligned} H_{\lambda_1,\lambda_2}(\bar{u}, \bar{f} + c) \leq H_{\infty,\lambda_2}(u + d, \bar{f}) &= \int \bar{f}(u - u^0)\,\mathrm{d}x + d \int \bar{f}\,\mathrm{d}x + \mathrm{TV}(u) \\ &= \int (\bar{f} + c)(u - u^0)\,\mathrm{d}x + \mathrm{TV}(u) = H_{\lambda_1,\lambda_2}(u, \bar{f} + c). \end{aligned} \tag{17}$$

Together, (16) and (17) show that for all f and u it holds that

$$H_{\lambda_1,\lambda_2}(\bar{u}, f) \leq H_{\lambda_1,\lambda 2}(\bar{u}, \bar{f} + c) \leq H_{\lambda_1,\lambda_2}(u, \bar{f} + c)$$

and this shows that $(\bar{u}, \bar{f} + c)$ is a solution of (15). $\square$

Note that the above theorem remains valid if we replace the TV penalty by any other penalty that is invariant under addition of constants such as Sobolev semi-norms.

We state a lemma on the subdifferential of the total variation of the positive and negative part of a function which we use in the following theorem.

Lemma 4.4. *Let $u \in \mathrm{BV}(\Omega)$. Then $\partial\,\mathrm{TV}(u) \subset \partial\,\mathrm{TV}(u^+)$ and $\partial\,\mathrm{TV}(u) \subset \partial\,\mathrm{TV}(u^-)$.*

Proof. It suffices to prove the inclusion $\partial\,\mathrm{TV}(u) \subset \partial\,\mathrm{TV}(u^+)$, the other inclusion being completely analogous. We begin by observing that if $L \in \partial\,\mathrm{TV}(u)$, as a linear functional $L^* \in [\mathrm{BV}(\Omega)]^*$, then

$$\mathrm{TV}(u) = L(u).$$

This follows from applying the definition of the subdifferential

$$\mathrm{TV}(v) - \mathrm{TV}(u) \geq L(v-u), \quad \text{for all } v \in \mathrm{BV}(\Omega), \tag{18}$$

to both $v = 0$ and $v = 2u$. If we now apply the definition to $v = u^-$, and also use the fact that $-L \in \partial\,\mathrm{TV}(-u)$, we deduce

$$\mathrm{TV}(u^-) \geq |L(u^-)|. \tag{19}$$

Using $\mathrm{TV}(u) = \mathrm{TV}(u^+) + \mathrm{TV}(u^-)$ to rearrange (18), we have

$$\mathrm{TV}(v) - \mathrm{TV}(u^+) \geq L(v-u^+) + \big(\mathrm{TV}(u^-) + L(u^-)\big), \quad \text{for all } v \in \mathrm{BV}(\Omega).$$

Referring to (19) we deduce $L \in \partial\,\mathrm{TV}(u^+)$. □

Theorem 4.5 (Weak maximum principle)**.** *Let $u^0 \geq 0$. Then there exists a minimizer $\bar{u}$ of* (9) *that also fulfills $\bar{u} \geq 0$.*

Proof. Writing the necessary and sufficient optimality conditions for the saddle point formulation (10) of (9), we have [20, Theorem 4.1 & Proposition 3.2, Chapter III]

$$0 \in f + \partial\,\mathrm{TV}(\bar{u}), \quad \text{and} \tag{20}$$

$$\bar{u} - u_0 \in N_{C_1}(f) + N_{C_2}(f), \tag{21}$$

where the constraint sets are

$$C_1 := \{f \in \mathrm{Lip}(\Omega) \mid -\lambda_1 \leq f(x) \leq \lambda_1 \text{ for all } x \in \Omega\}, \quad \text{and} \tag{22}$$

$$C_2 := \{f \in \mathrm{Lip}(\Omega) \mid \|\nabla f(x)\| \leq \lambda_2 \text{ for all } x \in \Omega\}. \tag{23}$$

Application of Lemma 4.4 shows that

$$0 \in f + \partial\,\mathrm{TV}(\bar{u}^+), \tag{24}$$

so that the first condition (20) is satisfied by $\bar{u}^+$ as well. Let us show that also (21) is satisfied by $\bar{u}^+$. To begin with we observe that at $\mathfrak{L}^n$-a.e. point x with $\bar{u}(x) < 0$, either C_1 or C_2 is active. Indeed, since $\bar{u}(x) - u_0(x) < 0$ at such point, in the problem

$$\max_{f \in C_1 \cap C_2} \int_\Omega f(\bar{u} - u_0)\,\mathrm{d}x,$$

the solution f should be as negative as possible within the constraints. If it is as negative as possible, C_1 is active, and

$$[N_{C_1}(f)](x) = (-\infty, 0].$$

Otherwise, C_2 has to be active, with f going as fast as possible to the least possible value it can achieve. In this case,

$$[N_{C_2}(f)](x) = [0, \infty) \operatorname{sign}[-\operatorname{div} \nabla f(x)].$$

If C_1 is not active, this has to be

$$[N_{C_2}(f)](x) = (-\infty, 0],$$

for $\bar{u}$ to satisfy (21). In either case, the right hand side of (21) is $(-\infty, 0]$. Therefore, trivially

$$\bar{u}^+ - u_0 \in N_{C_1}(f) + N_{C_2}(f) = (-\infty, 0]. \tag{25}$$

Combining (24) and (25) shows that $\bar{u}^+$ is a solution to (10). □

Corollary 4.6 (Weak boundedness)**.** *Let $u^0 \in L^\infty(\Omega)$. Then there exists a solution $\bar{u}$ of* (9) *fulfilling $\|\bar{u}\|_{L^\infty(\Omega)} \le \|u^0\|_{L^\infty(\Omega)}$.*

Proof. The problem (9) is affine-invariant, i.e., for data $au^0 + c$ for any constants $a, c \in \mathbb{R}$ we have $a\bar{u} + c$ as a solution. Setting $M := \|u^0\|_{L^\infty(\Omega)}$ and applying Theorem 4.5 to data $u^0 + M$ and $-u^0 + M$ proves the claim. □

Corollary 4.7 (Non negative solutions if mass is preserved)**.** *If $u^0 \ge 0$ and $\frac{\lambda_2}{\lambda_1} \le \frac{2}{\operatorname{diam}(\Omega)}$ then any minimizer of* (9) *is non-negative.*

Proof. The proof of Theorem 4.5 reveals that if $\bar{u}$ is a solution, then also $\bar{u}^+$ is a solution. However, if $\bar{u}$ would have a negative part (i.e. $\int_\Omega \bar{u}^- \, \mathrm{d}x > 0$) then $\bar{u}$ and $\bar{u}^+$ would have a different mean value which would contradict Theorem 4.3. □

5 Numerical solution

In this section we briefly sketch how one may solve the KR-TV denoising problem (9) numerically. Basically, we rely on methods to solve convex-concave saddle point problems, see, e.g. [14, 23, 34].

For the saddle point formulation (9) with Lipschitz constraint we reformulate as follows:

$$\begin{aligned}\min_u \max_{f,\phi} \int_\Omega f(u - u^0) \, \mathrm{d}x + \int_\Omega \nabla u \cdot \phi \, \mathrm{d}x - I_{\|\cdot\|_\infty \le 1}(|\phi|) \\ - I_{\|\cdot\|_\infty \le \lambda_1}(f) - I_{\||\cdot|\|_\infty \le \lambda_2}(\nabla f)\end{aligned} \tag{26}$$

By dualizing the term $I_{\||\cdot|\|_\infty\le\lambda_2}(\nabla f)$ we obtain another primal variable q and end up with

$$\begin{aligned}\min_{u,q}\max_{f,\phi}\int_\Omega f(u-u^0)\,\mathrm{d}x+\int_\Omega \nabla u\cdot\phi\,\mathrm{d}x\\ -I_{\|\cdot\|_\infty\le 1}(|\phi|)-I_{\|\cdot\|_\infty\le\lambda_1}(f)+\lambda_2\||q|\|_{\mathfrak{M}}-\int_\Omega q\cdot\nabla f.\end{aligned}\tag{27}$$

This is of the form

$$\min_{u,q}\max_{f,\phi} G(u,q)+\langle K(u,q),(f,\phi)\rangle-F(f,\phi)$$

with

$$\begin{aligned}G(u,q)&=\lambda_2\||q|\|_{\mathfrak{M}}\\ F(f,\phi)&=I_{\||\cdot|\|_\infty\le 1}(\phi)+I_{\|\cdot\|_\infty\le\lambda_1}(f)+\int_\Omega f\,u^0\,\mathrm{d}x\\ K\begin{bmatrix}u\\q\end{bmatrix}&=\begin{bmatrix}\mathrm{id}&\mathrm{div}\\ \nabla&0\end{bmatrix}\begin{bmatrix}u\\q\end{bmatrix}=\begin{bmatrix}u+\mathrm{div}\,q\\ \nabla u\end{bmatrix}\end{aligned}$$

Remark 5.1. We may also start from the cascading formulation (11) which is already almost in saddle-point form:

$$\begin{aligned}&\min_{u,\nu}\lambda_1\|u-u_0-\mathrm{div}\,\nu\|_{\mathfrak{M}}+\lambda_2\||\nu|\|_{\mathfrak{M}}+\mathrm{TV}(u)\\ &=\min_{u,\nu}\max_{\phi}\lambda_1\|u-u_0-\mathrm{div}\,\nu\|_{\mathfrak{M}}+\lambda_2\||\nu|\|_{\mathfrak{M}}+\int_\Omega\nabla u\cdot\phi\,\mathrm{d}x-I_{\|\cdot\|_\infty\le 1}(|\phi|)\\ &=\min_{u,\nu}\max_{f,\phi}\int_\Omega(u-u_0-\mathrm{div}\,\nu)\,f\,\mathrm{d}x+\lambda_2\||\nu|\|_{\mathfrak{M}}+\int_\Omega\nabla u\cdot\phi\,\mathrm{d}x\\ &\qquad -I_{\||\cdot|\|_\infty\le 1}(|\phi|)-I_{\|\cdot\|_\infty\le\lambda_1}(f).\end{aligned}$$

However, using $-\int_\Omega \mathrm{div}\,\nu\,f\,\mathrm{d}x=\int_\Omega\nu\cdot\nabla f\,\mathrm{d}x$ we arrive back at precisely the same formulation as (27) (with ν instead of q).

Note that both F and G admit simple proximity operators (both implementable in complexity proportional to the number of variables in F or G, respectively). Moreover, the operator K and its adjoint involve only one application of the gradient and the divergence (and some pointwise operations) and hence, can also be implemented in linear complexity. Hence, the application of general first order primal-dual methods leads to methods with very low complexity of the iterations and usually fast initial progress of the iterations. Moreover, note that the norm of K can be estimated with the help of the norm of the (discretized) gradient operator as $\|K\|\le\sqrt{\|\nabla\|^2+2}$. In our experiments we used the inertial forward-backward primal-dual method from [34] with a constant inertial parameter α. The iteration reads as

$$
\begin{aligned}
\bar{u}^k &= u^k + \alpha(u^k - u^{k-1}) \\
\bar{\nu}^k &= \nu^k + \alpha(\nu^k - \nu^{k-1}) \\
\bar{\phi}^k &= \phi^k + \alpha(\phi^k - \phi^{k-1}) \\
\bar{f}^k &= f^k + \alpha(f^k - f^{k-1}) \\
u^{k+1} &= \bar{u}^k - \tau(-\operatorname{div}\bar{\phi}^k + \bar{f}^k) \\
\nu^{k+1} &= \operatorname{prox}_{\tau\lambda_2 \||\cdot|\|_{\mathfrak{M}}}(\bar{\nu}^k - \tau\nabla\bar{f}^k) \\
\phi^{k+1} &= \operatorname{proj}_{\||\cdot|\|_\infty \le 1}(\bar{\phi}^k + \sigma\nabla(2u^{k+1} - \bar{u}^k)) \\
f^{k+1} &= \operatorname{proj}_{\|\cdot\|_\infty \le \lambda_1}(\bar{f}^k + \sigma(2u^{k+1} - \bar{u}^k - \operatorname{div}(2\nu^{k+1} - \bar{\nu}^k) - u^0))
\end{aligned}
$$

with σ and τ such that $\sigma\tau \le \|K\|^{-2}$ and some $\alpha \in [0, 1/3[$ (cf. [34, Remark 3]).

For our one-dimensional examples in Section 6.1 the total number of variables is small enough so that general purpose solvers for convex optimization can be applied. Here we used CVX [24,25] with the interior point solver from MOSEK.[1]

6 Experiments

In this section we present examples of minimizers of the KR-TV problem. In each subsection we do not have the aim to show that KR-TV outperforms any existing method but to point out additional features of this new approach. Hence, we do in general not compare the KR-TV functional against the most successful method for the respective task, but to the closest relative among the successful methods, i.e. to the L^1-TV method.

6.1 One-dimensional examples

Figure 2 shows the influence of the parameters λ_1 and λ_2 in three simple but instructive examples: a plateau, a ramp and a hat.

For the L^1-TV case the plateau either stays exact (for λ_1 large enough) or totally disappears (for λ_1 small enough). If the plateau would have been wide enough, then it would not disappear, but the minimizer would be constant 1 since the minimizer always approaches the constant median value for $\lambda_1 \to 0$. In the KR-TV case, however, the plateau gets wider and flatter while the total mass is preserved. In the limit $\lambda_2 \to 0$ the minimizer converges to a constant but still has the same mass than u^0 since for $\lambda_2 \to 0$ one approaches the constant mean value.

For the ramp, L^1-TV shows the known behavior that the ramp is getting flatter and flatter for decreasing λ_1. In the limit $\lambda_1 \to 0$ one obtains the constant median. For KR-TV, somewhat unexpectedly, the ramp not only gets flatter (it approaches the constant mean value, which equals the median here) but also forms new jumps. For some parameter value, the minimizer is even a pure jump.

[1] `http://mosek.com`

The observation for the hat is somehow similar to the ramp: L^1-TV just cuts off the hat-tip while KR-TV creates additional jumps.

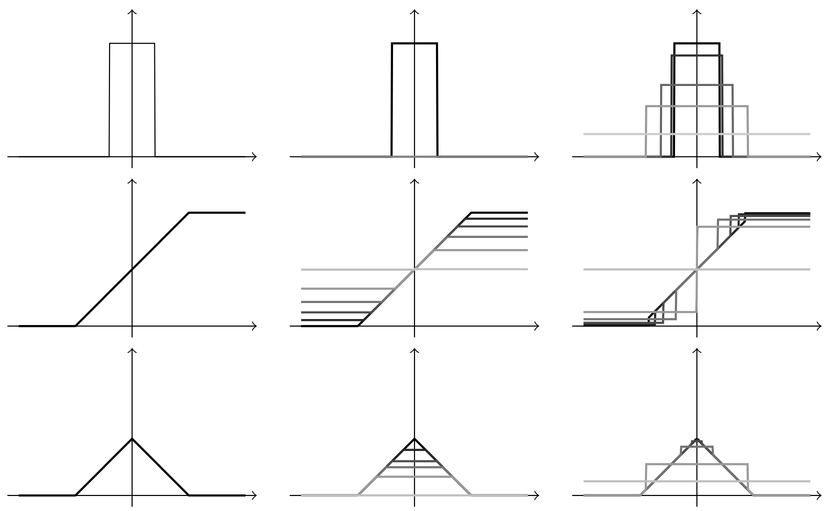


Figure 2: One-dimensional illustrations for KR-TV denoising with varying parameters. Left: Original functions u^0. Middle: Corresponding L^1-TV minimizers with λ_1 decreasing (lighter gray corresponds to smaller λ_1); λ_2 is so large, that the respective constraint is inactive throughout. Right: Corresponding KR-TV minimizers with decreasing λ_2 (lighter gray corresponds to smaller λ_2); λ_1 is so large, that the respective constraint is inactive throughout.

6.2 Two dimensional denoising with KR-TV

We illustrate the denoising capabilities of KR-TV in comparison with L^1-TV in Figures 3 and 4. Figure 3 shows effects similar to those shown in Figure 2 in one dimension. While both L^1-TV and KR-TV denoise the image well, L^1-TV tends to remove small structures completely while KR-TV mashes small structures together before they are merged with the background.

In Figure 4 we took a piecewise affine image, contaminated by noise and denoised it by L^1-TV, KR-TV and L^1-TGV. The parameters have been tuned by hand to give a minimal L^1-error to the ground truth, i.e. to the noise-free $u^\dagger$. Even though this choice seems to be perfectly suited for L^1-TV it turns out that KR-TV achieves a smaller error. One the other hand, the superiority of L^1-TGV shows that the choice of the regularizer has a far larger impact in this experiment. Also note that staircasing is slightly reduced by KR-TV in comparison to L^1-TV but also edges are a little more blurred for KR-TV. Since L^1-TGV is perfectly suited to this image (consisting of affine parts and jumps) it is no surprise that this produces by far the best results on this image.

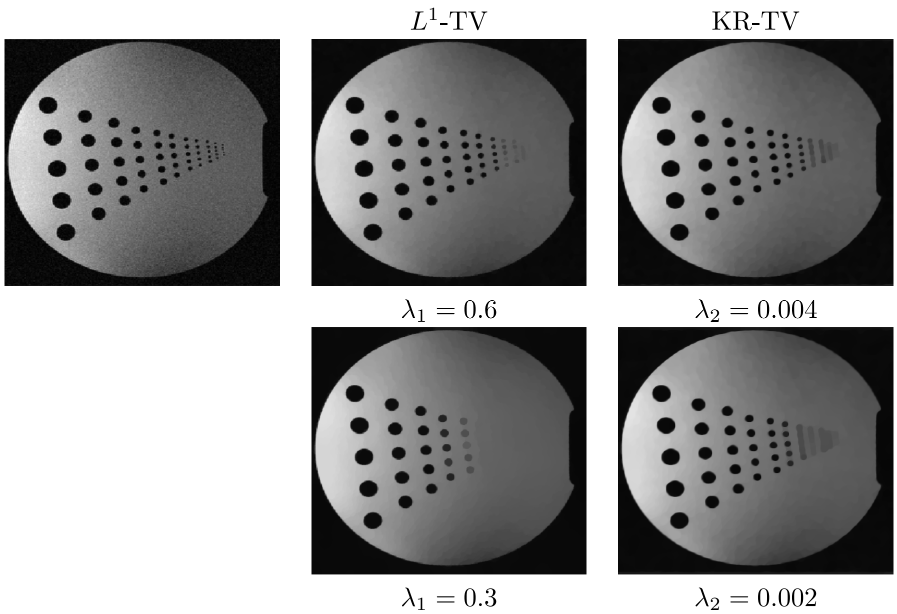


Figure 3: Denoising with KR-TV and L^1-TV. In the right images λ_1 is so large that the respective constraint is inactive.

6.3 Cartoon-Texture decomposition

We compare the KR-TV model for cartoon texture decomposition with L^1-TV and also with Meyer's G-TV (cf. Section 4.3). In Figure 5 we show decompositions of Barbara into its cartoon and texture part. The parameters have been chosen as follows: We started with the value λ_1 for the L^1-TV decomposition (i.e. $\lambda_2 = \infty$) and chose it such that most texture is in the texture component but also some structure is already visible. Then, for the G-TV the parameter was adjusted such that the cartoon part has the same total variation as the cartoon part from the L^1-TV decomposition. For the KR-TV decomposition, the value λ_1 was set to ∞ while λ_2 was again chosen such that the total variation of the cartoon part equals the total variation of the other cartoon parts. The rationale behind this choice is that the total variation is used as a prior for the cartoon part in all three models. We remark that choosing the parameters such that the L^1-discrepancy of the texture part is equal for all three decompositions leads to slightly different, but visually comparable results.

Note that, for these parameters the L^1-TV decomposition already has some structure in the texture part (parts of the face and of the bookshelf) and the G-TV decomposition has structure and texture severely mixed, while for KR-TV the texture component still mainly contains texture. Also note that KR-TV manages to keep the smooth structure of the clothes in the cartoon part (see e.g. the scarf and the trousers) while L^1-TV gives a more "piecewise constant"

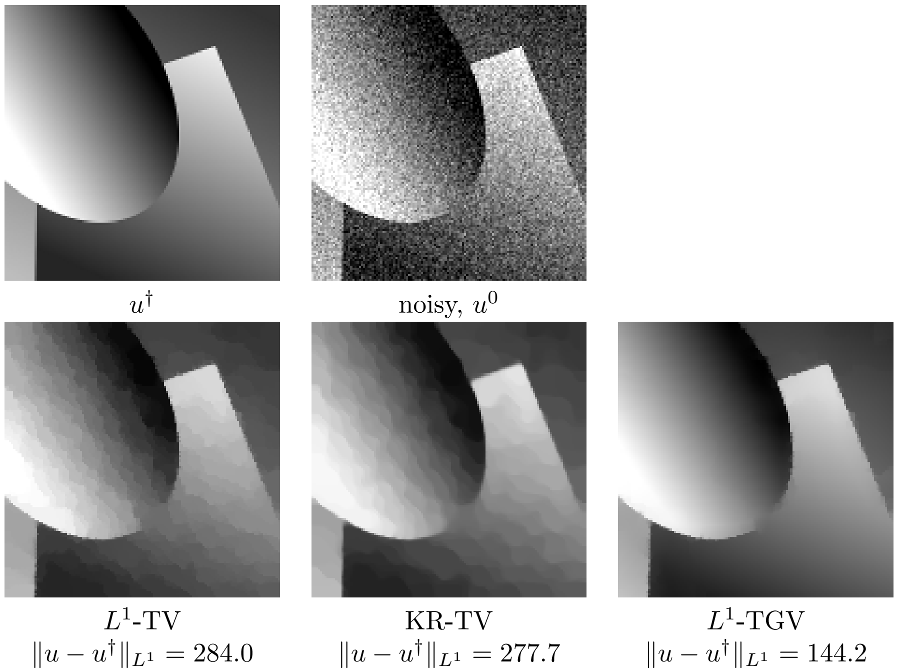


Figure 4: Denoising with KR-TV and L^1-TV. Left: L^1-TV denoising (i.e. only λ_1 is used), middle: KR-TV denoised by using the value λ_2 only (λ_1 so large, that the bound is inactive), right: L^1-TGV denoised. The respective values λ_1, λ_2 and α_1, α_2 have been optimized to result is the smallest L^1 error to the original noise-free image.

cartoon image.

7 Conclusion

In this paper we propose a new discrepancy term in a total variation regularisation approach for images that is motivated by optimal transport. The proposed discrepancy term is the Kantorovich-Rubinstein transport norm. We show relations of this norm to other standard discrepancy terms in the imaging literature and derive qualitative properties of minimizers of a total variation regularization model with a KR discrepancy. Indeed, we find that the KR discrepancy can be seen as a generalization of the dual Lipschitz norm and the L^1 norm, both of which can be derived from the Kantorovich-Rubinstein norm by letting one of the parameters go to infinity, respectively. Moreover, we show that this specialization is in fact crucial for obtaining a model in which the solution conserves mass and that the model has a solution which preserves positivity.

The paper is furnished with a discussion of experiments where we use the KR-TV regularisation approach in the context of image denoising and image

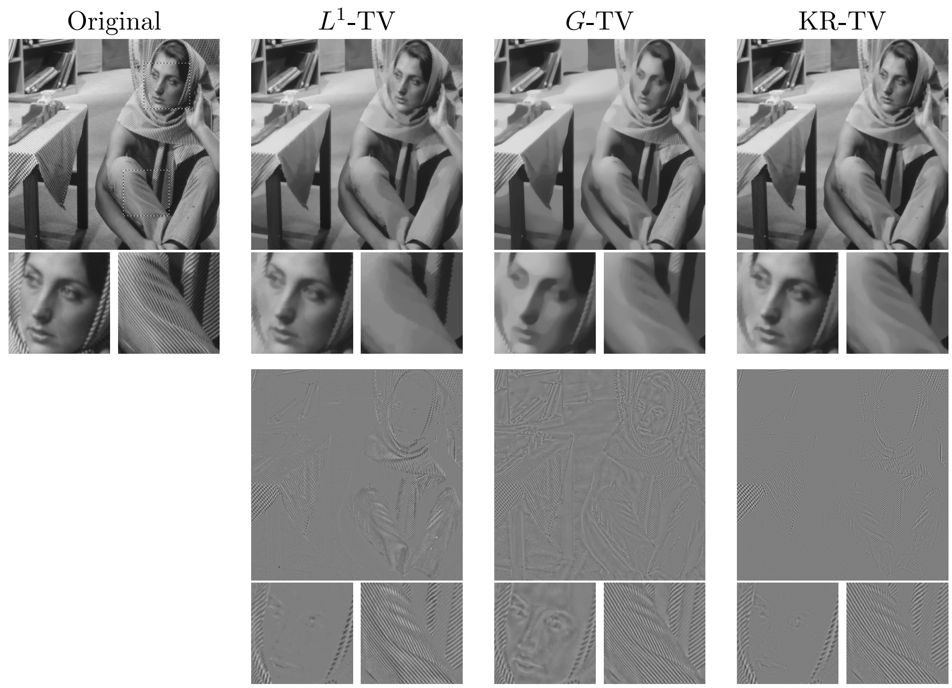


Figure 5: Cartoon-texture decomposition with L^1-TV, G-TV, and KR-TV. Top row: original and cartoon parts, bottom row: texture parts.

decomposition. Our numerical discussion suggests that the use of the KR norm can reduce the TV staircasing effect and performs better when decomposing an image into a cartoon-like and oscillatory component. Due to the mass conversation property we also expect that this approach is interesting in medical imaging, where images are usually indeed density functions of physical quantities, as well as in the context of density estimation where total variation approaches have been used before in the context of earthquakes and fires, see [42] for instance. The applicability of the KR discrepancy in other imaging problems such as optical flow, image sequence interpolation or stereo vision has to be investigated in future research.

While some analytical properties of the KR-TV method have been established (e.g. a weak maximum principle and a mass preservation property), a deeper understanding of the geometrical properties, as has been carried out for L^1-TV and L^2-TV, as well as to some extent for TGV (see, e.g., [8, 13, 18, 47, 49, 58, 60, 61]), would indeed be interesting. However, due to the non-locality of the KR discrepancy, the analysis may be more complicated.

Acknowledgement

This project has been financially supported by the King Abdullah University of Science and Technology (KAUST) Award No. KUK-I1-007-43, and the EPSRC first grant Nr. EP/J009539/1 "Sparse & Higher-order Image Restoration". T. Valkonen has further been supported by a Senescyt (Ecuadorian ministry of Education, Science, and Technology) Prometeo Fellowship. J. Lellmann has been supported by the Leverhulme Early Career Fellowship ECF-2013-436.

References

[1] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*, volume 254. Clarendon Press Oxford, 2000.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[3] Hedi Attouch and Haïm Brezis. Duality for the sum of convex functions in general Banach spaces. In Jorge Alberto Barroso, editor, *Aspects of Mathematics and its Applications*, volume 34 of *North-Holland Mathematical Library*, pages 125–133. Elsevier, 1986.

[4] Jean-François Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. Structure-texture image decomposition—modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.

[5] Vladimir. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[6] Doug M. Boyer, Yaron Lipman, Elizabeth St. Clair, Jesus Puente, Biren A. Patel, Thomas Funkhouser, Jukka Jernvall, and Ingrid Daubechies. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. *Proceedings of the National Academy of Sciences*, 108(45):18221–18226, 2011.

[7] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[8] Kristian Bredies, Karl Kunisch, and Tuomo Valkonen. Properties of L^1-TGV^2: The one-dimensional case. *Journal of Mathematical Analysis and Applications*, 398:438–454, 2013.

[9] Kristian Bredies and Tuomo Valkonen. Inverse problems with second-order total generalized variation constraints. In *Proceedings of the 9th International Conference on Sampling Theory and Applications (SampTA) 2011*, Singapore, 2011.

[10] Jonathan M. Bunn, Doug M. Boyer, Yaron Lipman, Elizabeth St. Clair, Jukka Jernvall, and Ingrid Daubechies. Comparing Dirichlet normal surface energy of tooth crowns, a new technique of molar shape quantification for dietary inference, with previous methods in isolation and in combination. *American Journal of Physical Anthropology*, 145(2):247–261, 2011.

[11] Martin Burger, Marzena Franek, and Carola-Bibiane Schönlieb. Regularized regression and density estimation based on optimal transport. *Applied Mathematics Research eXpress*, 2012(2):209–253, 2012.

[12] Giuseppe Buttazzo and Filippo Santambrogio. A model for the optimal planning of an urban area. *SIAM J. Math. Anal.*, 37(2):514–530, 2005.

[13] Vicent Caselles, Antonin Chambolle, and Matteo Novaga. The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale modeling & simulation*, 6(3):879–894, 2007.

[14] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[15] Tony F. Chan and Selim Esedoglu. Aspects of total variation regularized L^1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.

[16] Tony F. Chan, Selim Esedoglu, and Kangyu Ni. Histogram based segmentation using Wasserstein distances. In *Scale Space and Variational Methods in Computer Vision*, pages 697–708. Springer, 2007.

[17] Luigi De Pascale and Aldo Pratelli. Regularity properties for Monge transport density and for solutions of some shape optimization problem. *Calculus of Variations and Partial Differential Equations*, 14(3):249–274, 2002.

[18] Vincent Duval, Jean-François Aujol, and Yann Gousseau. The TVL1 model: a geometric point of view. *Multiscale Modeling & Simulation. A SIAM Interdisciplinary Journal*, 8(1):154–189, 2009.

[19] Vincent Duval, Jean-François Aujol, and LuminitaA. Vese. Mathematical modeling of textures: Application to color image decomposition with a projected gradient algorithm. *Journal of Mathematical Imaging and Vision*, 37(3):232–248, 2010.

[20] Ivar Ekeland and Roger Temam. *Convex analysis and variational problems.* SIAM, 1999.

[21] Herbert Federer. *Geometric Measure Theory.* Springer, 1969.

[22] Sira Ferradans, Nicolas Papadakis, Julien Rabin, Gabriel Peyré, and Jean-François Aujol. Regularized discrete optimal transport. In *Scale Space and Variational Methods in Computer Vision*, pages 428–439. Springer, 2013.

[23] Tom Goldstein, Ernie Esser, and Richard Baraniuk. Adaptive primal-dual hybrid gradient methods for saddle-point problems. arXiv preprint arXiv:1305.0546, 2013.

[24] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[25] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[26] Kristen Grauman and Trevor Darrell. Fast contour matching using approximate earth mover's distance. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–220. IEEE, 2004.

[27] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004.

[28] Leonid V. Kantorovič. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[29] Leonid V. Kantorovič and Gennadi Š. Rubinšteĭn. On a functional space and certain extremum problems. *Doklady Akademii Nauk SSSR*, 115:1058–1061, 1957.

[30] Stefan Kindermann, Stanley Osher, and Jinjun Xu. Denoising by BV-duality. *Journal of Scientific Computing*, 28(2-3):411–444, 2006.

[31] Haibin Ling and Kazunori Okada. An efficient earth mover's distance algorithm for robust histogram comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(5):840–853, 2007.

[32] Yaron Lipman and Ingrid Daubechies. Conformal Wasserstein distances: Comparing surfaces in polynomial time. *Advances in Mathematics*, 227(3):1047–1077, 2011.

[33] Yaron Lipman, Jesus Puente, and Ingrid Daubechies. Conformal Wasserstein distance: II. Computational aspects and extensions. *Math. Comput.*, 82(281):331–381, 2013.

[34] Dirk A. Lorenz and Thomas Pock. An inertial forward-backward method for monotone inclusions. To appear in *Journal of Mathematical Imaging and Vision*, 2014. [doi:10.1007/s10851-014-0523-2, arXiv:1403.3522].

[35] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces: Fractals and rectifiability.* Cambridge University Press, 1999.

[36] Facundo Mémoli. On the use of Gromov-Hausdorff distances for shape comparison. In *Eurographics symposium on point-based graphics*, pages 81–90. The Eurographics Association, 2007.

[37] Facundo Mémoli. Gromov-Hausdorff distances in euclidean spaces. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008.

[38] Facundo Mémoli. Spectral Gromov-Wasserstein distances for shape matching. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 256–263. IEEE, 2009.

[39] Facundo Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011.

[40] Facundo Mémoli. A spectral notion of Gromov-Wasserstein distance and related methods. *Applied and Computational Harmonic Analysis*, 30(3):363–401, 2011.

[41] Yves Meyer. *Oscillating patterns in image processing and nonlinear evolution equations.* American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

[42] George O. Mohler, Andrea L. Bertozzi, Thomas A. Goldstein, and Stanley J. Osher. Fast TV regularization for 2D maximum penalized likelihood estimation. *Journal of Computational and Graphical Statistics*, 20(2):479–491, 2011.

[43] Frank Morgan. *Geometric Measure Theory: A Beginner's Guide*. Academic Press, 1987.

[44] Kangyu Ni, Xavier Bresson, Tony F. Chan, and Selim Esedoglu. Local histogram based segmentation using the Wasserstein distance. *International Journal of Computer Vision*, 84(1):97–111, 2009.

[45] Laurent Oudre, Jérémie Jakubowicz, Pascal Bianchi, and Chantal Simon. Classification of periodic activities using the Wasserstein distance. *Biomedical Engineering, IEEE Transactions on*, 59(6):1610–1619, 2012.

[46] Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[47] Konstantinos Papafitsoros and Kristian Bredies. A study of the one dimensional total generalised variation regularisation problem. arXiv preprint arXiv:1309.5900, 2013.

[48] Gabriel Peyré, Jalal Fadili, and Julien Rabin. Wasserstein active contours. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2541–2544. IEEE, 2012.

[49] Christiane Pöschl and Otmar Scherzer. Exact solutions of one-dimensional TGV. arXiv preprint arXiv:1309.7152, 2013.

[50] Julien Rabin and Gabriel Peyré. Wasserstein regularization of imaging problems. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1541–1544. IEEE, 2011.

[51] Julien Rabin, Gabriel Peyré, and Laurent D Cohen. Geodesic shape retrieval via optimal mass transport. In *Computer Vision–ECCV 2010*, pages 771–784. Springer, 2010.

[52] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2012.

[53] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[54] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[55] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, and Frank Lenzen. *Variational methods in imaging*, volume 167. Springer, 2008.

[56] Bernhard Schmitzer and Christoph Schnörr. Modelling convex shape priors and matching based on the Gromov-Wasserstein distance. *Journal of Mathematical Imaging and Vision*, 46(1):143–159, 2013.

[57] Bernhard Schmitzer and Christoph Schnörr. Object segmentation by shape matching with Wasserstein modes. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 123–136. Springer, 2013.

[58] David Strong and Tony Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems*, 19(6):S165, 2003.

[59] Paul Swoboda and Christoph Schnörr. Convex variational image restoration with histogram priors. *SIAM Journal on Imaging Sciences*, 6(3):1719–1735, 2013.

[60] Tuomo Valkonen. The jump set under geometric regularisation. Part 1: Basic technique and first-order denoising. arXiv preprint arXiv:1407.1531, July 2014.

[61] Tuomo Valkonen. The jump set under geometric regularisation. Part 2: Higher-order approaches. arXiv preprint arXiv:1407.2334, July 2014.

[62] Luminita A Vese and Stanley J Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19(1-3):553–572, 2003.

[63] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[64] Wei Wang, John A. Ozolek, Dejan Slepcev, Ann B Lee, Cheng Chen, and Gustavo K. Rohde. An optimal transportation approach for nuclear structure-based pathology. *Medical Imaging, IEEE Transactions on*, 30(3):621–631, 2011.

[65] Wotao Yin, Donald Goldfarb, and Stanley Osher. A comparison of three total variation based texture extraction models. *Journal of Visual Communication and Image Representation*, 18(3):240–252, 2007.

[66] Lei Zhu, Yan Yang, Steven Haker, and Allen Tannenbaum. An image morphing technique based on optimal mass preserving mapping. *IEEE Transactions on Image Processing*, 16(6):1481–1495, 2007.