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 generalised 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 \cite{22,50}, texture synthesis and texture mixing \cite{52}, and surface mapping \cite{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 \cite{11,12,26,31,48,52,54,64}. In most cases, the 2-Wasserstein distance \cite{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 one 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 \cite{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 $\Omega$ and two constants $\lambda_1, \lambda_2 \geq 0$ we consider

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

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

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

The Kantorovich-Rubinstein norm \cite{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 \cite{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 \cite{7}. Finally, the KR-norm discrepancy shares properties with Meyer’s $G$-norm model \cite{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.

2
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 $\mathcal{M}(\Omega, \mathbb{R}^n)$ 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 $\vec{u}$ but sometime we omit the emphasis. The dual pairing between $\mathcal{M}(\Omega, \mathbb{R}^n)$ 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 $\mathcal{M}(\Omega, \mathbb{R}^n)$ is $\|\vec{u}\|_{\mathcal{M}} = \sup_{|f| \leq 1} \int \vec{f} \cdot d\vec{\mu}$ and is called the Radon norm. We identify $\mu$ and is called the Radon norm. We identify $u \in \mathcal{M}(\Omega, \mathbb{R}^n)$ with the corresponding measure $\mu$. Consequently, the norm on $\mathcal{M}(\Omega, \mathbb{R}^n)$ is $\{\mu \in \mathcal{M}(\Omega, \mathbb{R}^n) : \mu(\Omega) = 0\}$ for $\mu$ and $\nu$ are not probability 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 $\mu$ and $\nu$ define the $p$-Wasserstein distance

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

Note that this metric also makes sense if $\mu$ and $\nu$ are not probability measures but still non-negative and have equal mass, i.e., $\int_{\Omega} d\mu = \int_{\Omega} d\nu$. However, if the mass is not equal, no $\gamma$ with $\mu$ and $\nu$ as marginals would exist.
The celebrated Kantorovich duality \cite{28,63} states that, in the case of non-negative measures with equal mass, the Wasserstein metric can be equivalently expressed as

\[
W_p(\mu, \nu) = \left( \sup \left\{ \int \phi \, d\mu + \int \psi \, d\nu : \phi, \psi \in C_b(\Omega), \phi(x) + \psi(y) \leq |x-y|^p \right\} \right)^{1/p}.
\]

A particular special case is \( p = 1 \), and here, the Kantorovich-Rubinstein duality \cite{29,63} states that

\[
W_1(\mu, \nu) = \sup \left\{ \int f \, d(\mu - \nu) : \text{Lip}(f) \leq 1 \right\}.
\]

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

\[
\|\mu\|_{\text{Lip}^*} = \sup \left\{ \int f \, d\mu : \text{Lip}(f) \leq 1 \right\}
\]

one obtains the so-called dual Lipschitz norm on the space of measures with zero mean and finite first moments (cf. \cite{5,§8.10(viii)} where it is called modified Kantorovich-Rubinstein norm). Note that the supremum is unbounded if \( \mu \) 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

\[
\sup \left\{ \int f \, d\mu : |f| \leq 1, \text{Lip}(f) \leq 1 \right\},
\]

which is called Kantorovich-Rubinstein norm in \cite{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

\[
\|\mu\|_{\text{KR},\lambda} = \sup \left\{ \int f \, d\mu : |f| \leq \lambda_1, \text{Lip}(f) \leq \lambda_2 \right\}.
\] (2)

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

\[
\|\mu\|_{\text{KR},(\infty,1)} = \|\mu\|_{\text{Lip}^*},
\]

\[
\|\mu\|_{\text{KR},(1,\infty)} = \|\mu\|_{\text{MR}}.
\] (3)

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

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

\[
\|\mu\|_{\text{KR},\lambda} \leq \lambda_1 \|\mu\|_{\text{MR}}.
\]
If $\mu$ is non-negative it holds that

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

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

$$\|\mu\|_{KR,\lambda} \leq \lambda_1 \frac{\text{diam}(\Omega)}{2} \|\mu\|_{\mathcal{M}}.$$  

**Proof.** The first inequality follows directly from the definition of $\|\mu\|_{KR,\lambda}$ by dropping the constraint $|\nabla f| \leq \lambda_2$ 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 $\|f\|_{\infty} \leq \lambda_2$. However, since $\Omega$ has bounded diameter and $\mu$ has mean value zero, the constraint $\|\nabla f\|_{\infty} \leq \lambda_2 \text{diam}(\Omega)/2$ (indeed, $\lambda_2 \text{diam}(\Omega)$ is a bound on the value $\max f - \min f$, however, since $\int_\Omega d\mu = 0$, we may add a constant to $f$ without altering the outer supremum). We obtain

$$\|\mu\|_{KR,\lambda_1,\lambda_2} \leq \sup_{\|f\|_{\infty} \leq \lambda_2 \text{diam}(\Omega)/2} \int f \, d\mu \leq \lambda_2 \text{diam}(\Omega) \|\mu\|_{\mathcal{M}}/2.$$  

**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 $\|\mu\|_{\mathcal{M}} = 2$ while $\|\mu\|_{KR,\lambda} \to 0$ 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 $\text{Lip}(f) \leq \lambda_2$ by a pointwise constraint of the form $|f(x) - f(y)| \leq \lambda_2 |x - y|$, i.e., we have

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

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

$$\|\mu\|_{KR,\lambda} = \inf_{\gamma \geq 0} \left[ \lambda_1 \int_{\Omega} d|\mu - \text{proj}_1 \gamma + \text{proj}_2 \gamma| + \lambda_2 \int_{\Omega \times \Omega} |x - y| \, d\gamma \right]. \quad (4)$$  

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

$$\|\mu\|_{KR} = \inf_{\gamma \geq 0} \left\{ \int_{\Omega \times \Omega} |x - y| \, d\gamma : \text{proj}_1 \gamma - \text{proj}_2 \gamma = \mu \right\},$$  

5
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 \(\text{proj}_1 \gamma = \mu^+, \text{proj}_2 \gamma = \mu^-\) 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 \(\text{Lip}(f) \leq \lambda_2\) 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 \(\Omega\), it is indeed the case that \(\|\nabla f\|_\infty = \text{Lip}(f)\) (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

\[
\|\mu\|_{KR, \lambda} = \min_{\tilde{\nu} \in \mathcal{M}(\Omega, \mathbb{R}^n)} \lambda_1 \|\mu - \text{div} \tilde{\nu}\|_{\mathcal{M}} + \lambda_2 \||\tilde{\nu}||_{\mathcal{M}} \quad (5)
\]

where \(\text{div} \tilde{\nu}\) is understood to be taken in \(\Omega\) or, equivalently, in any open set \(U\) containing \(\Omega\).

**Proof.** Using indicator functions, we have

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

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

\[
A = \{f \in X : \sup_{x \in U} |f(x)| \leq \lambda_1\},
\]
\[
B = \{\tilde{g} \in Y : \sup_{x \in U} |\tilde{g}(x)| \leq \lambda_2\}.
\]

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

\[
F(f) = -\int_{\Omega} f \, d\nu + I_A(f), \quad G(\tilde{g}) = I_B(\tilde{g})
\]

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

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

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

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

Hence, we have

\[
\sup_{f \in X} -F(f) - G(Kf) = \inf_{\nu \in Y^*} F^*(-K^* \nu) + G^*(\nu).
\]
We have \( X^* = M(U) \) and \( Y^* = M(U, \mathbb{R}^n) \) and the conjugate functions of \( F \) and \( G \) are expressed with the help of the sets
\[
C = \{ \eta \in M(U) : |\eta|(U \setminus \overline{\Omega}) = 0 \} \\
D = \{ \tilde{\nu} \in M(U, \mathbb{R}^n) : |\tilde{\nu}|(U \setminus \overline{\Omega}) = 0 \}
\]
as
\[
F^*(\eta) = \lambda_1 ||\mu + \eta||_{\mathcal{M}(\overline{\Omega})} + I_C(\eta), \quad G^*(\tilde{\nu}) = \lambda_2 ||\tilde{\nu}||_{\mathcal{M}(\overline{\Omega})} + I_D(\tilde{\nu}).
\]
Since by the Kirszbraun theorem every \( f \) that is Lipschitz continuous on \( \Omega \) can be extended to \( U \) (with preservation of the Lipschitz constant) it follows with \( K^* = -\text{div} : Y^* \to X^* \) that
\[
\|\mu\|_{KR, \lambda} = \inf_{\tilde{\nu} \in Y^*} F^*(-K^*\tilde{\nu}) + G^*(\tilde{\nu}) \\
= \inf_{\nu \in M(U, \mathbb{R}^n)} \lambda_1 \|\mu - \text{div} \nu\|_{\mathcal{M}(\overline{\Omega})} + \lambda_2 \|\nu\|_{\mathcal{M}(U)} + I_C(\text{div} \nu) + I_D(\nu).
\]
Since bounded sets in \( M(U, \mathbb{R}^n) \) are relatively weakly* compact, we can replace \( M(U, \mathbb{R}^n) \) by \( M(\overline{\Omega}, \mathbb{R}^n) \) and drop the constraints \( C \) and \( D \) and arrive at
\[
\|\mu\|_{KR, \lambda} = \min_{\nu \in \mathcal{M}(\overline{\Omega}, \mathbb{R}^n)} \lambda_1 \|\mu - \text{div} \nu\|_{\mathcal{M}(\overline{\Omega})} + \lambda_2 \|\nu\|_{\mathcal{M}(\overline{\Omega})}
\]
as desired. \( \square \)

In Theorem 3.4 below we will prove that actually we can take \( \tilde{\nu} \) as an \( L^1 \) vector field with \( L^1 \) divergence in \( \overline{\Omega} \). Namely \( \tilde{\nu} \in W^{1,1}(\Omega; \text{div}) \), where for \( \Omega \subset \mathbb{R}^n \) an open domain, we define
\[
W^{1,1}(\Omega; \text{div}) := \{ \tilde{\nu} \in L^1(\Omega; \mathbb{R}^n) : \text{div} \tilde{\nu} \in L^1(\Omega) \}
\]
As such, our result is closely related to the work in [7], where this \( L^1 \) property is proved for the transport density \( |\tilde{\nu}| \). 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 \( \nu \) 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 \( a_j, b_j, a_k, b_k \in B(x, 2c\rho) + \mathbb{R}z \) for some unit vector \( z \).

**Proof.** The claim that \( \nu \) has the form \( \nu = \sum_{j=1}^M \beta_j [a_j, b_j] \) is trivial, as the problem in (5) with discrete \( \mu \) is a simple combinatorial problem.
Likewise and the vector giving the minimum distance between the lines
a constant distance \( \parallel \) of generality assume that \( v \) We may then find a plane \( P \) are large enough that \( \kappa \) segments \([\bar{a}, j] \) such that both segments lie in the cylinder \( B \) and that both segments have to cross. That is \([\bar{a}, j] \cap [\bar{b}, k] = \emptyset \) and \( \kappa > 2 \) is possible with \( \gamma \) of the points closer towards \( x \) depicted in Figure 1a or b. The segments \([\bar{a}, j] \) when looking from the direction \( v \) \( c > 1 \) as the segments \([\bar{a}, j] \) and \([\bar{b}, k] \) pass through \( B(\bar{x}, \gamma n) \), and so we may split each segment into three parts – two outside \( B(\bar{x}, \gamma n) \), and one inside.

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

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 \( B(x, \gamma n) \), then \( \|b_j - b_k\| < 2\kappa \rho \). This is because both segments \([\bar{a}, j] \) and \([\bar{b}, k] \) 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 \( B(\bar{x}, 2\kappa \rho) + \bar{z} \). Otherwise, for large enough \( c \), both \( |a_j - a_k| > \kappa \rho \) as well as \( |b_j - b_k| > \kappa \rho \). Since \( d \leq 2 \rho < \gamma \), i.e., some midpoints of the segments are closer than the end points, we observe that the two segments have to cross. That is \([\bar{a}, j] \cap [\bar{b}, k] = \emptyset \) for some \( \bar{q} \). If \( c \) and \( \kappa \) 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 \( \bar{a} - \bar{q} - \bar{b} \), compare Figure 1d it now follows that

\[
|\bar{a} - \bar{b}| \leq \sqrt{|\bar{a} - \bar{q}|^2 - (\kappa - 2)^2 \rho^2 + |\bar{b} - \bar{q}|^2 - (\kappa - 2)^2 \rho^2}.
\]

Likewise

\[
|\bar{a} - \bar{b}| \leq \sqrt{|\bar{a} - \bar{q}|^2 - (\kappa - 2)^2 \rho^2 + |\bar{b} - \bar{q}|^2 - (\kappa - 2)^2 \rho^2}.
\]

8
If \( n = 2 \), or more generally \( d = 0 \), it trivially follows that
\[
|a_j - b_k| + |a_k - b_j| < |a_j - q| + |b_k - q| + |a_k - q| + |b_j - q|
= |a_j - b_j| + |a_k - b_k|.
\]

Otherwise, minding that \( |d| \leq 2\rho \) and \( \kappa > 2 \), we calculate
\[
|a_j - b_k| + |a_j - b_k| = \sqrt{|\vec{a}_j - \vec{b}_k|^2 + |d|^2} + \sqrt{|\vec{a}_k - \vec{b}_j|^2 + |d|^2}
\leq \sqrt{(|\vec{a}_j - \vec{q}| + |\vec{b}_k - \vec{q}|)^2 - 2(\kappa - 2)^2\rho^2 + d^2}
\quad + \sqrt{(|\vec{a}_k - \vec{q}| + |\vec{b}_j - \vec{q}|)^2 - 2(\kappa - 2)^2\rho^2 + d^2}
< |a_j - q| + |b_k - q| + |a_k - q| + |b_j - q|
= |a_j - b_j| + |a_k - b_k|.
\]

This provides a contradiction to the optimality of the transport rays \([a_j, b_j]\) and \([a_k, b_k]\), and shows the claim. \( \square \)

**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
\[
\|\mu\|_{\text{KR},\lambda_1,\lambda_2} = \min_{\nu \in W^{1,1}(\Omega,\text{div})} \lambda_1 \|\mu - \text{div} \nu\|_{L^1(\Omega;\mathbb{R}^n)} + \lambda_2 \|\nu\|_{L^1(\Omega)}. \tag{6}
\]
Moreover the minimum is reached by \( \nu \) satisfying \( \int_{\Omega} \text{div} \nu \, d\Omega^n = 0 \).
Proof. We assume first that \( \mu \in L^\infty(\Omega) \). By Lemma 3.1 we have (5). To replace \( \Omega \) by \( \Omega \), we just have to show that that \( |\nu|(\partial\Omega) = 0 \) for any \( \nu \) reaching the minimum in (5). This follows if \( \nu \ll \mathcal{L}^n \). Hence it suffices to show that actually \( \nu \) and \( \text{div} \nu \) are also absolutely continuous with respect to \( \mathcal{L}^n \). This is where we need the convexity of \( \Omega \) and the absolute continuity of \( \mu \).

Clearly by (5) we have
\[
\|\mu\|_{KR, \lambda_1, \lambda_2} \leq \min_{\nu \in W^{1,1}(\Omega;\text{div})} \lambda_1 \|\mu - \text{div} \nu\|_{L^1(\Omega;\mathbb{R}^n)} + \lambda_2 \|\nu\|_{L^1(\Omega)},
\]
so it remains to show the opposite inequality. We approximate \( \mu \) 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 \( \alpha_{i,j} = \int_{V_{i,j}} \mu(x) \, dx \), as well as
\[
V_{i,j} \cap V_{i,k} = \emptyset, \quad (i \neq k), \quad \text{and} \quad \text{supp} \mu \subset \bigcup_{j=1}^{N_i} V_{i,j}, \quad (i = 1, \ldots, N_i). \quad (7)
\]
Then (5) is a finite-dimensional discrete/combinatorial problem, and we easily discover an optimal solution \( \nu^i \). Because transporting mass outside \( \Omega \) incurs a cost on \( \partial\Omega \), we see that
\[
\nu^i = \sum_{j=1}^{N_i} \beta_{i,j} \| a_{i,j}, b_{i,j} \|,
\]
for some \( \beta_{i,j} > 0 \) and \( a_{i,j}, b_{i,j} \in \{x_{i,1}, \ldots, x_{i,N_i}\} \). We calculate
\[
\text{div} [a, b] = \delta_b - \delta_a.
\]
Moreover
\[
\text{div} \nu^i(\Omega) = \text{div} \nu^i(\Omega) = 0, \quad \text{and} \quad \text{div} \nu^i \ll |\mu^i|.
\]
As minimisers, we have
\[
||\nu^i||_{\mathcal{M}(\Omega;\mathbb{R}^n)} \leq \frac{\lambda_1}{\lambda_2} \|\mu^i||_{\mathcal{M}(\Omega)} \leq \frac{\lambda_1}{\lambda_2} \|\mu||_{\mathcal{M}(\Omega)}.
\]
Therefore, after possibly moving to a subsequence, unlabelled, we may assume that \( \nu^i \rightharpoonup \nu \) for some \( \nu \in \mathcal{M}(\Omega;\mathbb{R}^n) \). But by (8) we may also assume that \( \text{div} \nu^i \rightharpoonup \lambda \in \mathcal{M}(\Omega) \), where \( \lambda \ll |\mu| \). From this absolute continuity it follows that \( \lambda(\partial\Omega) = 0 \). (A priori it might be that \( \lambda(\partial\Omega) \neq 0 \).) Necessarily \( \lambda = \text{div} \nu \), so that in particular \( \text{div} \nu \ll \mathcal{L}^n \). Because \( \partial\Omega \) is \( \mathcal{L}^n \)-negligible, it follows that \( \text{div} \nu(\Omega) = 0 \).

We want to show that \( \nu \) is an optimal solution to (5) for \( \mu \). 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 \( \mu \) within the cell \( V_{i,j} \) to the cell centre \( \delta_{x_{i,j}} \), or the other way around. That is
\[
\text{div} \nu_{i,j} = \mu \chi_{V_{i,j}} - \alpha_{i,j} \delta_{x_{i,j}}.
\]
with
\[ \|\nu_{i,j}\| \leq \epsilon_i \int_{\Omega} |\mu(x)| \, dx. \]

It follows that
\[ \sum_{j=1}^{N_i} \|\nu_{i,j}\| \leq \epsilon_i \|\mu\|. \]

If now \( \nu^* \) is an optimal solution to (5) for \( \mu \), defining
\[ \nu_0^i := \nu^* - \sum_{j=1}^{N_i} \nu_{i,j}, \]
we see that
\[ \|\nu_0^i\|_{\mathcal{M}(\Omega)} \leq \|\nu^*\|_{\mathcal{M}(\Omega)} + C\epsilon_i \]
and
\[ \text{div} \nu_0^i = \text{div} \nu^* - \mu + \mu^i. \]

Thus
\[ \lambda_1 \|\mu^i - \text{div} \nu^i\|_{\mathcal{M}(\Omega)} + \lambda_2 \|\nu^i\|_{\mathcal{M}(\Omega; \mathbb{R}^n)} \]
\[ \leq \lambda_1 \|\mu^i - \text{div} \nu_0^i\|_{\mathcal{M}(\Omega)} + \lambda_2 \|\nu_0^i\|_{\mathcal{M}(\Omega; \mathbb{R}^n)} \]
\[ \leq \lambda_1 \|\mu - \text{div} \nu^*\|_{\mathcal{M}(\Omega)} + \lambda_2 \|\nu^*\|_{\mathcal{M}(\Omega; \mathbb{R}^n)} + C\epsilon_i \]

By weak* lower semicontinuity
\[ \lambda_1 \|\mu - \text{div} \nu\|_{L^1(\Omega)} + \lambda_2 \|\nu\|_{\mathcal{M}(\Omega; \mathbb{R}^n)} \]
\[ \leq \liminf_{i \to \infty} \left( \lambda_1 \|\mu^i - \text{div} \nu^i\|_{\mathcal{M}(\Omega)} + \lambda_2 \|\nu^i\|_{\mathcal{M}(\Omega; \mathbb{R}^n)} \right) \]
\[ \leq \liminf_{i \to \infty} \left( \lambda_1 \|\mu - \text{div} \nu^*\|_{\mathcal{M}(\Omega)} + \lambda_2 \|\nu^*\|_{\mathcal{M}(\Omega; \mathbb{R}^n)} + C\epsilon_i \right) \]
\[ = \lambda_1 \|\mu - \text{div} \nu^*\|_{\mathcal{M}(\Omega)} + \lambda_2 \|\nu^*\|_{\mathcal{M}(\Omega; \mathbb{R}^n)}. \]

Thus \( \nu \) is an optimal solution to (5) for \( \mu \). Exploiting lower semicontinuity of both of the terms, we moreover see that \( \lim_{i \to \infty} \|\nu^i\|_{\mathcal{M}(\Omega; \mathbb{R}^n)} = \|\nu\|_{\mathcal{M}(\Omega; \mathbb{R}^n)} \).

Thus \( \{\nu^i\}_{i=1}^{\infty} \) converge to \( \nu \) strictly in \( \mathcal{M}(\Omega; \mathbb{R}^n) \). Likewise \( \{\mu^i - \text{div} \nu^i\}_{i=1}^{\infty} \) converge to \( \mu - \text{div} \nu \) strictly in \( \mathcal{M}(\Omega) \). But \( \{\mu^i\}_{i=1}^{\infty} \) were already constructed to converge strictly to \( \mu \), and we have above seen that \( |\text{div} \nu^i| \leq |\mu^i| \). Therefore also \( \{\text{div} \nu^i\}_{i=1}^{\infty} \) converge to \( \text{div} \nu \) strictly in \( \mathcal{M}(\Omega) \).

It remains to show that \( \nu \in W^{1,1}(\Omega; \text{div}) \). We have already shown \( \text{div} \nu \ll L^1(\Omega) \), so that \( \nu \ll L^1(\Omega) \). We just have to show that \( \nu \ll L^1(\Omega; \mathbb{R}^n) \) to show that \( \nu \in L^1(\Omega; \mathbb{R}^n) \). We do this by bounding the \( n \)-dimensional density of \( \nu \) 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

\[
\max_{\|z\|=1} \sum_{a_{i,j},b_{i,j} \in (B(x,\kappa\rho)+\mathbb{R}z) \cap \Omega} \beta_{i,j} \mathcal{H}^1(B(x,\rho) \cap [a_{i,j}, b_{i,j}])
\]
\[
\leq \max_{\|z\|=1} \sum_{a_{i,j},b_{i,j} \in (B(x,\kappa\rho)+\mathbb{R}z) \cap \Omega} \beta_{i,j} 2\rho\]
\[
\leq \max_{\|z\|=1} \sum_{x_{i,j} \in (B(x,\kappa\rho)+\mathbb{R}z) \cap \Omega} |a_{i,j}| 2\rho
\]
\[
\leq 2\rho \max_{\|z\|=1} \sum_{x_{i,j} \in (B(x,\kappa\rho)+\mathbb{R}z) \cap \Omega} \int_{V_{i,j}} |\mu(y)| \, dy
\]
\[
\leq 2\rho \max_{\|z\|=1} \int_{B(x,\kappa\rho+\epsilon_i)+z\mathbb{R}} |\mu(y)| \, dy\]
\[
\leq 2\rho (\kappa\rho + \epsilon_i)^{n-1} \text{diam}(\Omega) M.
\]

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

\[
\sum_{a_{i,j} \in B(x,c\rho)} \beta_{i,j} \mathcal{H}^1(B(x,\rho) \cap [a_{i,j}, b_{i,j}]) + \sum_{b_{i,j} \in B(x,c\rho)} \beta_{i,j} \mathcal{H}^1(B(x,\rho) \cap [a_{i,j}, b_{i,j}])
\]
\[
\leq \sum_{x_{i,j} \in B(x,c\rho)} 4\alpha_{i,j} \rho
\]
\[
= \sum_{x_{i,j} \in B(x,c\rho)} 4\rho \int_{V_{i,j}} |\mu(y)| \, dy
\]
\[
\leq 4\rho \int_{B(x,c\rho+\epsilon_i)} |\mu(y)| \, dy.
\]

It now follows that

\[
|\nu|(B(x,\rho)) \leq 4\rho \int_{B(x,c\rho+\epsilon_i)} |\mu(y)| \, dy + 2\rho (2\kappa\rho + \epsilon_i)^{n-1} \text{diam}(\Omega) M
\]

Letting \(i \to \infty\), we get by lower semicontinuity

\[
|\nu|(B(x,\rho)) \leq 4\rho \int_{B(x,c\rho)} |\mu(y)| \, dy + 2\rho (2\kappa\rho)^{n-1} \rho^n \text{diam}(\Omega) M
\]

Thus

\[
\lim_{\rho \to 0} \frac{|\nu|(B(x,\rho))}{\mathcal{L}^n(B(x,\rho))} \leq 0 + 2^n \kappa^{n-1} \text{diam}(\Omega) M
\]

It follows (see [35, Theorem 2.12]) that \(\nu \ll \mathcal{L}_0, \Omega\) with

\[
\|\nu\|_{L^1(\mathbb{R}_+ \Omega)} \leq 2^n \kappa^{n-1} \text{diam}(\Omega) M \mathcal{L}^n(\Omega).
\]

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

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

12
Then $\mu^\pm_M \leq \mu^\pm$. Applying the point-mass approximation above to both $\mu^k$ and $\mu$, we can take $(\mu^k_M)^\pm \leq (\mu^\pm)^\pm$. Then by a simple argument we also have $|\nu^i_M| \leq |\nu|$ for each $i, k = 1, 2, 3, \ldots$; compare [17] Proposition 4.3. Indeed, let $\tilde{\mu}^i_M := \text{div} \nu^i_M$. Clearly

$$(\tilde{\mu}^i_M)^\pm \leq (\mu^i_M)^\pm \leq (\mu^\pm)^\pm.$$  

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

$$\|\text{div} \nu - \text{div} \nu_M\|_{\mathcal{B}(\Omega)} \leq \|\mu - \mu_M\|_{\mathcal{B}(\Omega)}.$$  

(This can be verified by the point-mass approximation.) It follows that $\text{div} \nu_M \rightarrow \text{div} \nu$ strongly. In particular $\text{div} \nu_M - \mu_M \rightarrow \text{div} \nu - \mu$ strongly. By lower semi-continuity of $\|\cdot\|_{\text{KR}, \lambda_1, \lambda_2}$ we therefore deduce that $\lim \inf_{M \rightarrow \infty} |\nu_M|(|\Omega| \geq |\nu|(|\Omega|). Since $|\nu_M| \leq |\nu|$, it follows that $\nu_M \rightarrow \nu$ strongly in $\mathcal{M}(\Omega; \mathbb{R}^n)$. But the above paragraphs say that $\nu_M \in L^\infty(\Omega)$. Thus necessarily $\nu_M \in L^1(\Omega)$. \hfill \square

4 Kantorovich-Rubininstein-TV denoising

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

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

for some $u^0 \in L^1(\Omega)$ and $\lambda = (\lambda_1, \lambda_2) \geq 0$. We call this Kantorovich-Rubininstein-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 $\text{Lip}(f) \leq \lambda_2$ with the help of the distributional gradient as $|\nabla f| \leq \lambda_2$. Then problem (9) has the form

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

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

$$\min_{u, \bar{v}} \lambda_1\|u - u^0 - \text{div} \bar{v}\|_{L^1} + \lambda_2\|\bar{v}\|_{L^1} + \text{TV}(u). \tag{11}$$

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

$$H_{\lambda}(u, f) = \begin{cases} \int f(u - u_0) + \text{TV}(u), & \text{if } |f| \leq \lambda_1, |\nabla f| \leq \lambda_2 \\ -\infty, & \text{otherwise.} \end{cases}$$  \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 $\|\mu\|_{KR,(\lambda_1, \infty)} = \lambda_1 \|\mu\|_M$ and for $u \in L^1(\Omega)$ it holds that $\|u\|_M = \|u\|_{L^1}$. Hence, KR-TV is a generalization of the successful $L^1$-TV denoising [15]:

$$\min_u \|u - u_0\|_{KR,(\lambda_1, \infty)} + \text{TV}(u) = \min_u \|u - u_0\|_{L^1} + \frac{1}{\lambda_1} \text{TV}(u).$$  \tag{13}$$

We will study the influence of the additional parameter $\lambda_2$ 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 $\lambda_2$ 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

$$\min_u \max_{|f| \leq \lambda_1} \int f(u - u_0) + \text{TV}(u).$$

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

$$\min_u \max_{\text{Lip}(f) \leq \lambda_2} \int \Omega f(u - u_0) + \text{TV}(u)$$

for any $\lambda_2 \geq L$ 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

$$(\text{div } v(x))_i = \sum_{j=1}^n \frac{\partial v_{ij}}{\partial x_j}, \quad \text{div}^2 v(x) = \sum_{i,j=1}^n \frac{\partial^2 v_{ij}}{\partial x_j \partial x_i}.$$
The total generalized variation of order two for a parameter \( \alpha = (\alpha_1, \alpha_2) \) is

\[
\text{TGV}^2_\alpha(u) = \sup \left\{ \int_{\Omega} u \, \text{div}^2 v \, dx : \ v \in C^2_c(\Omega, S^{n \times n}), \right. \\
|v(x)| \leq \alpha_1, \ |\text{div} v(x)| \leq \alpha_2 \right\}
\]

The TGV term has an equivalent reformulation as follows: denote by 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 an \( S^{n \times n} \)-valued Radon measure. Then it holds that

\[
\text{TGV}^2_\alpha(u) = \inf_{\vec{w} \in \text{BD}(\Omega)} \alpha_1 \| |\mathcal{E}\vec{w}| |_{3\mathbb{R}} + \alpha_2 \| |\nabla u - \vec{w}| |_{3\mathbb{R}}
\]

(cf. \cite{8,9}), leading to the \( L^1 \)-TGV\(^2 \) denoising problem

\[
\min_{u \in L^1(\Omega), \ w \in \text{BD}(\Omega, \mathbb{R}^n)} \| u - u^0 \|_{L^1} + \alpha_1 \| |\mathcal{E}\vec{w}| |_{3\mathbb{R}} + \alpha_2 \| |\nabla u - \vec{w}| |_{3\mathbb{R}}.
\]

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

\[
\| \mu \|_{\text{KR}, \lambda} = \min_{\vec{v} \in \text{BD}(\Omega, \mathbb{R}^n)} \lambda_1 \| \mu - \text{div} \vec{v} \|_{2\mathbb{R}} + \lambda_2 \| |\vec{v}| |_{2\mathbb{R}}.
\]

We obtain a new (semi-)norm by “cascading” a higher order term in a new minimization problem. In the TV case we go from \( \text{TV}(u) = \| |\nabla u| |_{2\mathbb{R}} \) to \( \text{TGV}^2_\alpha \) by cascading with a vector field and penalizing the symmetrized gradient of this vector field. In the KR case, however, we go from \( \| u \|_{L^1} = \| u \|_{2\mathbb{R}} \) to \( \| \cdot \|_{\text{KR}, \lambda} \) by cascading with the divergence of a vector field and penalizing with the Radon norm of that vector field. One may say, that \( \text{TGV}^2_\alpha \) 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 \cite{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

\[
\| u \|_G = \inf \{ \| |\vec{g}| | \|_\infty : \text{div} \vec{g} = u, \ g \in L^\infty \}.
\]

Meyer proposed the following G-TV minimization problem

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

This differs from problem \cite{11} in two aspects: First, \( |\vec{g}| \) is penalized in the \( \infty \)-norm instead of the 1-like Radon norm and second, the equality \( \text{div} \vec{g} = u - u_0 \) is enforced exactly, while in \cite{11} a mismatch is allowed. The Meyer model has also been treated in numerous other papers, e.g. \cite{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 $\lambda_1$ and $\lambda_2$ such that the minimizer of (9) is $u_0$ (if $u_0$ is regular enough in some sense) if $\lambda_1$ and $\lambda_2$ 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. $|\vec{\phi}| \leq 1$
2. $\int u_0 \text{div} \vec{\phi} = TV(u_0)$.

Then there exists thresholds $\lambda_1^*$ and $\lambda_2^*$ 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

$$
\|u - u_0\|_{KR,\lambda_1,\lambda_2} + TV(u) \geq \int u \text{div} \vec{\phi} + \left[ \min_{\vec{\nu}} \lambda_1 \|u - u_0 - \text{div} \vec{\nu}\|_{\mathcal{M}} + \lambda_2 \|\vec{\nu}\|_{\mathcal{M}} \right]
$$

$$
= \int u_0 \text{div} \vec{\phi} + \min_{\vec{\nu}} \left[ \lambda_1 \|u - u_0 - \text{div} \vec{\nu}\|_{\mathcal{M}} + \lambda_2 \|\vec{\nu}\|_{\mathcal{M}} \right]
$$

$$
+ \int (u - u_0 - \text{div} \vec{\nu}) \text{div} \vec{\phi} + \int \text{div} \vec{\nu} \text{div} \vec{\phi}
$$

$$
\geq TV(u_0) + \min_{\vec{\nu}} \left[ (\lambda_1 - \|\text{div} \vec{\phi}\|_{\infty}) \|u - u_0 - \text{div} \vec{\nu}\|_{\mathcal{M}} + (\lambda_2 - \|\nabla \text{div} \vec{\phi}\|_{\infty}) \|\vec{\nu}\|_{\mathcal{M}} \right]
$$

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

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 $\check{u}$ to be the constant function that equals the mean value of $\bar{u}$ over $\Omega$, i.e.

$$
\check{u} \equiv \int_{\Omega} \bar{u}(x) \, dx.
$$

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

$$
H_\lambda(\check{u}, \check{f}) \geq H_\lambda(\bar{u}, f).
$$
In other words, using $TV(\bar{u}) = 0$, writing out $H_\lambda$, and rearranging terms

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

But, by the choice of $f$, we have

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

Therefore

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

An application of Poincaré’s inequality yields

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

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

The second of the above two theorems shows that for small $\lambda_1$ one recovers a constant solution. In fact, this has to be $\int_\Omega u^0 \, dx$. The first of the above two theorems shows that for parameters $\lambda_1$ and $\lambda_2$ 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 $\frac{\lambda_2}{\lambda_1} \leq \frac{2}{\text{diam}(\Omega)}$, then

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

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

**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 $\lambda_1$. Hence we start by denoting with $(\bar{u}, \bar{f})$ a solution of the following saddle-point problem:

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

With the notation [12], [14] reads as $\min_u \max_f H_{\infty, \lambda_1}(u, f)$.

It holds that $\int_\Omega \bar{u} \, dx = \int_\Omega u^0 \, dx$, because otherwise, the max would be $\infty$. In other words: with $\lambda_1 = \infty$ we have mass preservation.
Now let $\lambda_2 \leq \frac{2}{\text{diam}(\Omega)}$. We aim to show that there is constant $c$ such that $(\bar{u}, \bar{f} + c)$ is a solution of

$$\min_u \max_{|f| \leq \lambda_1, |\nabla f| \leq \lambda_2} \int f(u - u^0) \, dx + \text{TV}(u).$$

(15)

Since $\bar{f}$ is Lipschitz with constant $\lambda_2$, we get that $\bar{f}(x) - \bar{f}(y) \leq \lambda_2 |x - y|$, and hence, $\max \bar{f} - \min \bar{f} \leq \lambda_2 \text{diam}(\Omega)$. Consequently, there is a constant $c$ such that $|\bar{f} + c| \leq \lambda_2 \text{diam}(\Omega)$ in other words: $\bar{f} + c$ is feasible for (15). Since $\int \bar{u} = \int u^0$ we also have

$$H_{\lambda_1, \lambda_2}(\bar{u}, \bar{f} + c) = \int \bar{f}(u - u^0) \, dx + c \int (\bar{u} - u^0) \, dx + \text{TV}(u) = H_{\infty, \lambda_2}(\bar{u}, \bar{f}).$$

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

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

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

$$H_{\lambda_1, \lambda_2}(\bar{u}, \bar{f} + c) \leq H_{\infty, \lambda_2}(u + d, \bar{f}) = \int \bar{f}(u - u^0) \, dx + d \int \bar{f} \, dx + \text{TV}(u)$$

$$= \int (\bar{f} + c)(u - u^0) \, dx + \text{TV}(u) = H_{\lambda_1, \lambda_2}(u, \bar{f} + c).$$

(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).

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 $u \in BV(\Omega)$. Then $\partial TV(u) \subset \partial TV(u^+)$ and $\partial TV(u) \subset \partial TV(u^-)$.

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

\[ TV(u) = L(u). \]

This follows from applying the definition of the subdifferential

\[ TV(v) - TV(u) \geq L(v - u), \quad \text{for all } v \in BV(\Omega), \]  

(18)

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

\[ TV(u^-) \geq |L(u^-)|. \]

(19)

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

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

Referring to (19) we deduce $L \in \partial TV(u^+)$. \hfill \Box

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]

\[ 0 \in f + \partial TV(\bar{u}), \quad \text{and} \]

\[ \bar{u} - u_0 \in N_{C_1}(f) + N_{C_2}(f), \]  

(20)

(21)

where the constraint sets are

\[ C_1 := \{ f \in \text{Lip}(\Omega) \mid -\lambda_1 \leq f(x) \leq \lambda_1 \text{ for all } x \in \Omega \}, \quad \text{and} \]

\[ C_2 := \{ f \in \text{Lip}(\Omega) \mid \|\nabla f(x)\| \leq \lambda_2 \text{ for all } x \in \Omega \}. \]

(22)

(23)

Application of Lemma 4.4 shows that

\[ 0 \in f + \partial TV(\bar{u}^+), \]

(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 $\Sigma^n$-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

\[ \max_{f \in C_1 \cap C_2} \int_{\Omega} f(\bar{u} - u_0) \, dx, \]

19


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,

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

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

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

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

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

**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. \quad $\square$

**Corollary 4.7** (Non negative solutions if mass is preserved). If $u^0 \geq 0$ and $\frac{\lambda_2}{\lambda_1} \leq \frac{diam(\Omega)}{2}$ 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. $\int_{\Omega} \bar{u}^- dx > 0$) then $\bar{u}$ and $\bar{u}^+$ would have a different mean value which would contradict Theorem 1.3. \quad $\square$

## 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:

$$\min_{u} \max_{f, \phi} \int_{\Omega} f(u - u^0) dx + \int_{\Omega} \nabla u \cdot \phi dx - I_{\|\cdot\|_{\infty} \leq 1}(\phi)$$

$$- I_{\|\cdot\|_{\infty} \leq \lambda_1(f)} - I_{\|\cdot\|_{\infty} \leq \lambda_2(\nabla f)} \quad (26)$$
By dualizing the term $I_{\|\cdot\|_\infty \leq \lambda_2} (\nabla f)$ we obtain another primal variable $q$ and end up with

$$
\min_{u,q} \max_{f,\phi} \int_{\Omega} f(u - u^0) \, dx + \int_{\Omega} \nabla u \cdot \phi \, dx
- \lambda_1 \|q\|_{|\cdot|} - \int_{\Omega} q \cdot \nabla f.
$$

(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

$$
G(u, q) = \lambda_2 \|q\|_{|\cdot|}
$$

$$
F(f, \phi) = I_{\|\cdot\|_\infty \leq 1} (\phi) - I_{\|\cdot\|_\infty \leq \lambda_1} (f) + \int_{\Omega} f u^0 \, dx
$$

$$
K \begin{bmatrix} u \\ q \end{bmatrix} = \begin{bmatrix} \text{id} & \text{div} \\ \nabla & 0 \end{bmatrix} \begin{bmatrix} u \\ q \end{bmatrix} = \begin{bmatrix} u + \text{div} q \\ \nabla u \end{bmatrix}
$$

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

$$
\min_{u,\nu} \max_{\phi} \lambda_1 \|u - u_0 - \text{div} \nu\|_{|\cdot|} + \lambda_2 \|\nu\|_{|\cdot|} + \text{TV}(u)
$$

$$
= \min_{u,\nu} \max_{\phi} \lambda_1 \|u - u_0 - \text{div} \nu\|_{|\cdot|} + \lambda_2 \|\nu\|_{|\cdot|} + \int_{\Omega} \nabla u \cdot \phi \, dx
- \lambda_1 \|q\|_{|\cdot|}
$$

$$
= \min_{u,\nu} \max_{f,\phi} \lambda_1 \|u - u_0 - \text{div} \nu\|_{|\cdot|} + \int_{\Omega} \nabla u \cdot \phi \, dx
- \lambda_1 \|q\|_{|\cdot|}
$$

However, using $- \int_{\Omega} \text{div} \nu f \, dx = \int_{\Omega} \nu \cdot \nabla f \, dx$ we arrive back at precisely the same formulation as (27) (with $\nu$ 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 $\|K\| \leq \sqrt{\|\nabla\|^2 + 2}$. In our experiments we used the inertial forward-backward primal-dual method from [34] with a constant inertial
parameter $\alpha$. The iteration reads as

\begin{align*}
\bar{u}^k &= u^k + \alpha (u^k - u^{k-1}) \\
\bar{\nu}^k &= \nu^k + \alpha (\nu^k - \nu^{k-1}) \\
\bar{\sigma}^k &= \sigma^k + \alpha (\sigma^k - \sigma^{k-1}) \\
\bar{f}^k &= f^k + \alpha (f^k - f^{k-1}) \\
u^{k+1} &= u^k + \tau (-\text{div} \bar{\sigma}^k + \bar{f}^k) \\
\nu^{k+1} &= \text{prox}_{\tau \lambda_2 \|\cdot\|_{\infty}} (\bar{\nu}^k - \tau \nabla \bar{f}^k) \\
\sigma^{k+1} &= \text{proj}_{\|\cdot\|_{\infty} \leq \lambda_1} (\bar{\sigma}^k + \sigma (2u^{k+1} - \bar{u}^k)) \\
f^{k+1} &= \text{proj}_{\|\cdot\|_{\infty} \leq \lambda_1} (\bar{f}^k + \sigma (2u^{k+1} - \bar{u}^k - \text{div}(2\nu^{k+1} - \bar{\nu}^k) - u_0))
\end{align*}

with $\sigma$ and $\tau$ such that $\sigma \tau \leq \|K\|^{-2}$ 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 \cite{cvx1,cvx2} with the interior point solver from MOSEK \cite{mosek}.

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 $\lambda_1$ and $\lambda_2$ 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 $\lambda_1$ large enough) or totally disappears (for $\lambda_1$ 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 $u_0^0$ 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 $\lambda_1$. 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.

\texttt{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.

![One-dimensional illustrations for KR-TV denoising with varying parameters.](image)

Figure 2: One-dimensional illustrations for KR-TV denoising with varying parameters. Left: Original functions $u^0$. Middle: Corresponding $L^1$-TV minimizers with $\lambda_1$ decreasing (lighter gray corresponds to smaller $\lambda_1$); $\lambda_2$ is so large, that the respective constraint is inactive throughout. Right: Corresponding KR-TV minimizers with decreasing $\lambda_2$ (lighter gray corresponds to smaller $\lambda_2$); $\lambda_1$ 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 $u^1$. 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 $\lambda_1$ 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 $\lambda_1$ 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 $\lambda_1$ was set to $\infty$ while $\lambda_2$ 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 $\lambda_1$ is used), middle: KR-TV denoised by using the value $\lambda_2$ only ($\lambda_1$ so large, that the bound is inactive), right: $L^1$-TGV denoised. The respective values $\lambda_1$, $\lambda_2$ and $\alpha_1$, $\alpha_2$ 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 conservation 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, 55, 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


