Imaging with Kantorovich-Rubinstein discrepancy

Jan Lellmann

Cambridge Image Analysis
Department for Applied Mathematics and Theoretical Physics
University of Cambridge

D. A. Lorenz (TU Braunschweig)
C.-B. Schönlieb (U Cambridge)
T. Valkonen (EPN Quito/U Cambridge)

ICMS Workshop on Gradient Flows, April 2015
updated 22/04/2015

Problem

Rudin-Osher-Fatemi [Rudin, Osher, Fatemi'92] [Ring'00] [Chambolle'04] [Caselles, Chambolle, Novaga'09]

Given image data $u^0 : \Omega \to \mathbb{R}$, find

$$\inf_{u \in \mathrm{BV}(\Omega) \cap L^2(\Omega)} \left\{ \lambda \| u - u^0 \|_2^2 + TV(u) \right\}$$

Problem

Problem

Problem

$TV - L^1$ [Alliney'92] [Nikolova'04] [Chan, Esedoglu'05] [Duval, Aujol, Gousseau'09]

Given image data $u^0 : \Omega \to \mathbb{R}$, find

$$\inf_{u \in \mathrm{BV}(\Omega)} \left\{ \lambda \| u - u^0 \|_1 + TV(u) \right\}$$

Problem

Wasserstein $-TV$

Given image data $u^0 : \Omega \to \mathbb{R}$, find

$$\inf_{u \in \mathrm{BV}(\Omega)} \left\{ W_1(u\mathfrak{L}^n, u^0\mathfrak{L}^n) + TV(u) \right\}$$

Problem

Kantorovich-Rubinstein $-TV$

Given image data $u^0 : \Omega \to \mathbb{R}$, find

$$\inf_{u \in \mathrm{BV}(\Omega)} \left\{ \|u - u^0\|_{\mathrm{KR},(\lambda_1,\lambda_2)} + TV(u) \right\}$$

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

J. Lellmann, D. A. Lorenz, C. Schönlieb, T. Valkonen: *Imaging with Kantorovich-Rubinstein Discrepancy*, SIAM J. Imaging Sci., 7(4), 2014

Problem

Kantorovich-Rubinstein $-TV$

Given image data $u^0 : \Omega \to \mathbb{R}$, find

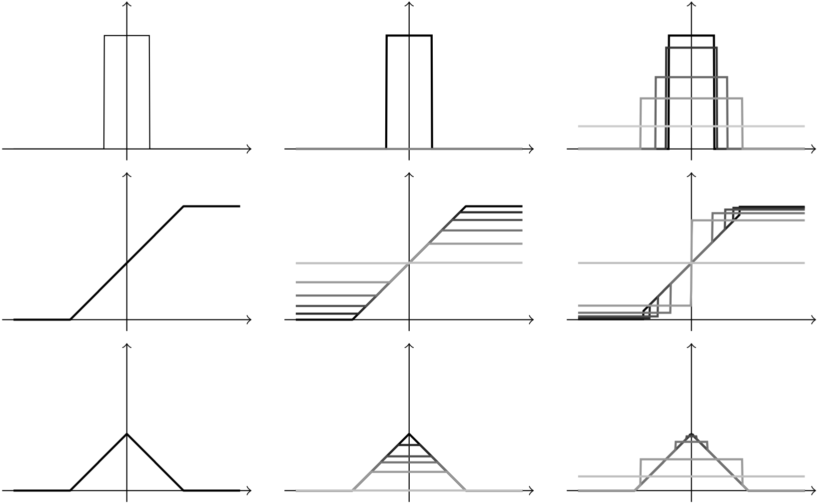
$$\inf_{u \in \mathrm{BV}(\Omega)} \left\{ \|u - u^0\|_{\mathrm{KR},(\lambda_1,\lambda_2)} + TV(u) \right\}$$

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

J. Lellmann, D. A. Lorenz, C. Schönlieb, T. Valkonen: *Imaging with Kantorovich-Rubinstein Discrepancy,* SIAM J. Imaging Sci., 7(4), 2014

optimal transport [Monge 1781] [Kantorovich 1942] [Jordan, Kinderlehrer, Otto'98] [Rachev, Rüschendorf'98] [Evans, Gangbo'99] [Ambrosio'03] [Villani'08], **registration** [Haker et al.'04], **segmentation, shape matching** [Chan, Esedoglu'09] [Swoboda, Schnörr'13] [Schmitzer, Schnörr'13], **image and shape retrieval** [Rubner, Tomasi, Guibas'00–02] [Rabin, Peyré'10] , **shape barycenter** [Rabin et al.'12], **color transfer** [Rabin, Peyré'11], **BRDF interpolation** [Bonneel et al.'11], **displacement interpolation, gradient flows** [McCann'97] [Benamou, Brenier'00+] [Ambrosio, Gigli, Savaré'06] [Burger, Carillo, Wolfram'10] [Düring, Mattes, Milisic'10] [Papadakis, Peyré, Oudet'14], **smoothing** and **decomposition** [Burger, Franek, Schönlieb'11]

Solution behavior, mass conservation and jump formation

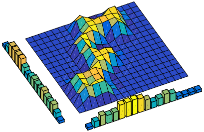


input	$L^1 - TV$	$\mathrm{KR} - TV$
	light $= \lambda_1$ smaller	$\lambda_1 \approx +\infty$
	$\lambda_2 \approx +\infty$	light $= \lambda_2$ smaller

Wasserstein distances

- $\Omega \subseteq \mathbb{R}^n$ image domain, μ, ν probability measures on Ω, then

$$W_p(\mu,\nu)^p := \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\Omega\times\Omega} |x-y|^p d\gamma(x,y)$$



is the **Wasserstein distance of order** p (here $p=1$).

- Π are the **couplings** of μ and ν: Probability measures on the product space $\Omega \times \Omega$ with **marginals** μ and ν.

$$\gamma \in \Gamma(\mu,\nu) \iff \gamma \in \mathfrak{M}(\Omega\times\Omega),\ \operatorname{proj}_1 \gamma = \mu,\ \operatorname{proj}_2 \gamma = \nu.$$

- Extends to non-negative measures with the same mass: $W_p(\mu,\nu) = +\infty$ if $|\mu| \neq |\nu|$.

Wasserstein distances

- For $p = 1$:
$$W_p(\mu,\nu) = \|\mu - \nu\|_{\mathrm{Lip}^*}$$
with the **dual Lipschitz norm**
$$\|\pi\|_{\mathrm{Lip}^*} := \sup\left\{ \int_\Omega f \, d\pi : \mathrm{Lip}(f) \leq 1 \right\}.$$
This requires $\pi(\Omega) = 0$, i.e., **zero mean**.
- Extension to bounded π:
$$\|\pi\|_{\mathrm{KR},1} := \sup\left\{ \int_\Omega f \, d\pi : |f| \leq 1, \mathrm{Lip}(f) \leq 1 \right\}.$$
  - sometimes *(dual) bounded Lipschitz norm* or variants: $f(x_0) = 0$, $\|f\|_1 \leq 1$, $\|f\|_\infty + Lip(f) \leq 1, \ldots$
  - other strategies: partial transport [Caffarelli, McCann'10] [Figalli'10], generalized Wasserstein distance [Piccoli, Rossi'14]

Wasserstein distances

- For $p = 1$:
$$W_p(\mu,\nu) = \|\mu - \nu\|_{\mathrm{Lip}^*}$$
with the **dual Lipschitz norm**
$$\|\pi\|_{\mathrm{Lip}^*} := \sup\left\{ \int_\Omega f \, d\pi : \mathrm{Lip}(f) \leq 1 \right\}.$$
This requires $\pi(\Omega) = 0$, i.e., **zero mean**.
- Extension to bounded π:
$$\|\pi\|_{\mathrm{KR},\lambda} := \sup\left\{ \int_\Omega f \, d\pi : \color{red}{|f| \leq \lambda_1,} \mathrm{Lip}(f) \leq \lambda_2 \right\}.$$
  - sometimes *(dual) bounded Lipschitz norm* or variants: $f(x_0) = 0$, $\|f\|_1 \leq 1$, $\|f\|_\infty + Lip(f) \leq 1, \ldots$
  - other strategies: partial transport [Caffarelli, McCann'10] [Figalli'10], generalized Wasserstein distance [Piccoli, Rossi'14]
- **"Kantorovich-Rubinstein norm"**
  - $\lambda_1 = +\infty$: dual Lipschitz norm $\|\cdot\|_{\mathrm{Lip}^*}$, $\lambda_2 = +\infty$: Radon/L^1 norm, $\|\cdot\|_{\mathfrak{M}} \approx \|\cdot\|_1$. $\lambda_1, \lambda_2 < +\infty$: $\|\delta_x - \delta_y\|_{\mathrm{KR},\lambda} \to 0$ as $x \to y$.
  - On bounded, convex, open domains: $\iff$ $\|\nabla f\|_\infty \leq \lambda_2$, "flat norm"

Relation to the Wasserstein distance

- Can show:

$$W_1(\mu,\nu) = \inf_{\gamma\geq 0}\left\{\int_{\Omega\times\Omega}|x-y|\,d\gamma : \operatorname{proj}_1\gamma - \operatorname{proj}_2\gamma = \mu-\nu\right\}.$$

- Our norm:

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

- **"Soft constrained" version** of dual Lipschitz norm/Wasserstein distance

Cascading formulation

Lemma (Cascading formulation – L., Lorenz, Schönlieb, Valkonen '14)

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

$$\|\mu\|_{\mathsf{KR},\lambda} = \min_{\vec{\nu} \in \mathfrak{M}(\overline{\Omega}, \mathbb{R}^n)} \{\lambda_1 \|\mu - \operatorname{div} \vec{\nu}\|_{\mathfrak{M}} + \lambda_2 \||\vec{\nu}|\|_{\mathfrak{M}}\}. \tag{1}$$

- $|\nu|$ analogous to the "transport density" for the Wasserstein distance [Evans, Gangbo'99]
- Proof:
  - Fenchel-Rockafellar duality theorem with Attouch/Brezis constraint qualification for strong duality
  - "min" from relatively weak* compactness of sets with bounded Radon measure.
- What if we know that $\mu \in L^1$?

Cascading formulation

Theorem (Data in L^1 – L., Lorenz, Schönlieb, Valkonen '14)

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

$$\|\mu\|_{\mathrm{KR},\lambda} = \min_{\nu \in W^{1,1}(\Omega;\mathrm{div})} \left\{ \lambda_1 \|\mu - \mathrm{div}\,\nu\|_{L^1(\Omega;\mathbb{R}^n)} + \lambda_2 \|\nu\|_{L^1(\Omega)} \right\}. \quad (2)$$

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

- Similar estimates exist for the pure Wasserstein case – estimates on the transport density [Pascale, Pratelli'02] [Santambrogio'09]
- Proof:
  - Approximate using finite sums of Dirac measures $\rightarrow$ ν consists of transport "rays"
  - Construct optimal ν from weakly* converging subsequence and show that convergence is strict
  - Show that $\mathrm{div}\,\nu \in L^1$ and $\nu \in L^1$: bound density of ν with respect to $\mathfrak{L}^n$ using the fact that transport rays intersecting a ball must be *approximately parallel*

KR-TV

Primal formulation

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

Dual/cascading formulation

$$\inf_{u,\nu} \left\{ \lambda_1 \|u - u^0 - \operatorname{div} \nu\|_{L^1} + \lambda_2 \||\nu|\|_{L^1} + TV(u) \right\}$$

Primal-dual/saddle-point formulation

$$\inf_{u} \sup_{\substack{|f| \leq \lambda_1 \\ |\nabla f| \leq \lambda_2}} \left\{ \int_\Omega (u - u^0) f \, dx + TV(u) \right\}$$

KR-TV

Primal formulation

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

Dual/cascading formulation

$$\inf_{u,\nu} \left\{ \lambda_1 \|u - u^0 - \operatorname{div} \nu\|_{L^1} + \lambda_2 \| |\nu| \|_{L^1} + TV(u) \right\}$$

Primal-dual/saddle-point formulation – $L^1 - TV$

$$\inf_{u} \sup_{\substack{|f| \le \lambda_1 \\ |\nabla f| \le +\infty}} \left\{ \int_\Omega (u - u^0) f \, dx + TV(u) \right\}$$

KR-TV

Primal formulation

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

Dual/cascading formulation

$$\inf_{u,\nu} \left\{ \lambda_1 \|u - u^0 - \operatorname{div} \nu\|_{L^1} + \lambda_2 \| |\nu| \|_{L^1} + TV(u) \right\}$$

Primal-dual/saddle-point formulation – mass conservation

$$\inf_{u} \sup_{\substack{|f| \leq +\infty \\ |\nabla f| \leq \lambda_2}} \left\{ \int_\Omega (u - u^0) f \, dx + TV(u) \right\}$$

Properties of the minimizer

- Recovery of the input data for large discrepancy weight:

Theorem (L., Lorenz, Schönlieb, Valkonen '14)

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 exist thresholds λ_1^* and λ_2^* such that for $\lambda_1 > \lambda_1^*$ and $\lambda_2 > \lambda_2^*$, the unique minimizer of KR-TV is u_0.

- Constant minimizer for small discrepancy weight:

Theorem (L., Lorenz, Schönlieb, Valkonen '14)

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 KR-TV is a constant whenever $1/C > \lambda_1$.

Properties of the minimizer

- **Weak** mass conservation:

Theorem (L., Lorenz, Schönlieb, Valkonen '14)

If $\frac{\lambda_2}{\lambda_1} \leq \frac{2}{\mathrm{diam}(\Omega)}$, *then there exists a minimizer* $\bar{u}$ *such that*

$$\int_\Omega \bar{u}(x)\,dx = \int_\Omega u^0(x)\,dx.$$

- Different to $TV - L^1$: Mass conservation even in the range of parameters where noise is removed
- Proof:
  - Consider saddle-point problem for $\lambda_1 = +\infty$ ($\rightarrow$ mass conservation)
  - modify duals to accomodate $\lambda_1 < +\infty$ but large enough

Properties of the minimizer

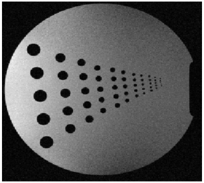
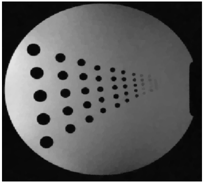
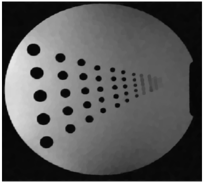
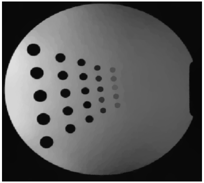
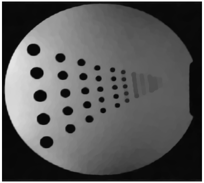
- **Weak** nonnegativity and boundedness:

Theorem (L., Lorenz, Schönlieb, Valkonen '14)

Let $u^0 \geq 0$. Then there exists a minimizer $\bar{u}$ that also fulfills $\bar{u} \geq 0$.

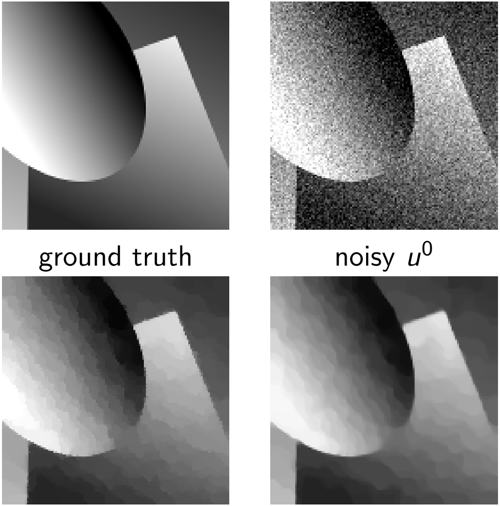
- Proof: Show that if $\bar{u}$ is a solution, then $\bar{u}^+$ is a solution using primal-dual optimality conditions
- Corollary (weak boundedness): Can find solution $\bar{u}$ with $\|\bar{u}\|_\infty \leq \|u^0\|_\infty$ if $u^0 \in L^\infty$.

Denoising

	L^1-TV	KR-TV
		
	$\lambda_1 = 0.6$	$\lambda_2 = 0.004$
		
	$\lambda_1 = 0.3$	$\lambda_2 = 0.002$

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

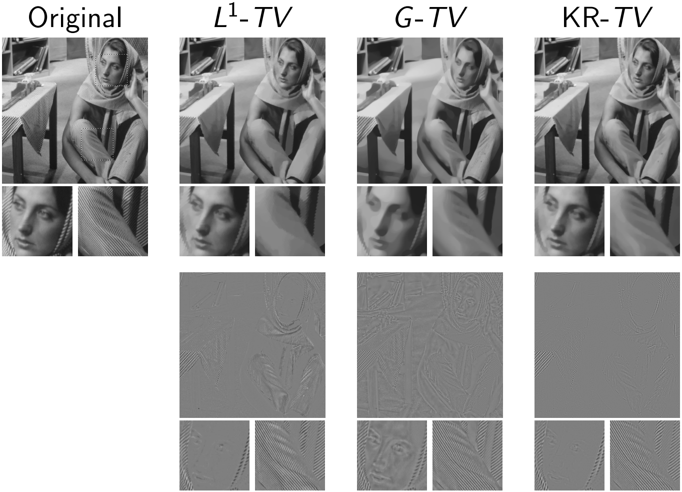
Denoising



ground truth | noisy u^0

L^1-TV $\|u - u^\dagger\|_{L^1} = 295.7$ | KR-TV $\|u - u^\dagger\|_{L^1} = 253.7$

Cartoon-texture decomposition



- **Top:** original and cartoon. **Bottom:** Texture.
- Parameters chosen so that the cartoon parts have the same TV.

Summary

- **Kantorovich-Rubinstein data term**
  - Extension of 1-Wasserstein distance
  - “Soft-constrained” dual Lipschitz norm
  - Cascading formulation
- **Analytical properties**
  - Recovery of input and constant solution
  - Mass conservation and weak non-negativity/boundedness
- **Experimental properties**
  - Structures “spread out”
  - Jump formation
  - Application in cartoon-texture decomposition

J. Lellmann, D. A. Lorenz, C. Schönlieb, T. Valkonen: *Imaging with Kantorovich-Rubinstein Discrepancy,* SIAM J. Imaging Sci., 7(4), 2014

Imaging with Kantorovich-Rubinstein discrepancy

Jan Lellmann

Cambridge Image Analysis
Department for Applied Mathematics and Theoretical Physics
University of Cambridge

D. A. Lorenz (TU Braunschweig)
C.-B. Schönlieb (U Cambridge)
T. Valkonen (EPN Quito/U Cambridge)

ICMS Workshop on Gradient Flows, April 2015
updated 22/04/2015

Relation to other norms

- Kantorovich-Rubinstein:

$$\|u\|_{\mathrm{KR},\lambda} = \inf_{\nu \in W^{1,1}(\Omega;\mathrm{div})} \left\{ \lambda_1 \|u - \operatorname{div} \nu\|_{L^1(\Omega;\mathbb{R}^n)} + \lambda_2 \|\nu\|_{L^1(\Omega)} \right\}.$$

- Meyer's G-Norm:

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

- Total Generalized Variation:

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