Collaborative Regularization Models for Color Imaging Problems

SIAM IS, Bologna, June 2018

Joan Duran*

*Dept. Mathematics and Computer Science, IAC3

Universitat de les Illes Balears, Mallorca, Spain

joint work with

Michael Moeller‡, Daniel Cremers§ and Catalina Sbert*

‡Inst. Vision and Graphics
University of Siegen, Germany

§Dept. Computer Science
Technical University of Munich, Germany

Universitat de les Illes Balears

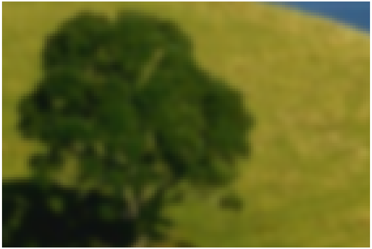
Ill-Posed Inverse Problems in Image Processing

- The classical inverse problem in imaging writes as $f = Au + \eta$.

Ill-Posed Inverse Problems in Image Processing

- The classical inverse problem in imaging writes as $f = Au + \eta$.
- Most of them are (highly) ill-posed.

u



f

Ill-Posed Inverse Problems in Image Processing

- The classical inverse problem in imaging writes as $f = Au + \eta$.
- Most of them are (highly) ill-posed.

u

$A^{-1}(f - \eta)$

Ill-Posed Inverse Problems in Image Processing

- The classical inverse problem in imaging writes as $f = Au + \eta$.
- Most of them are (highly) ill-posed.

u

$A^{-1}(f - \eta)$

- Regularization methods handles ill-posedness by introducing prior knowledge on u, usually assuming smooth solutions.

Ill-Posed Inverse Problems in Image Processing

- The classical inverse problem in imaging writes as $f = Au + \eta$.
- Most of them are (highly) ill-posed.

u

$A^{-1}(f - \eta)$

- Regularization methods handles ill-posedness by introducing prior knowledge on u, usually assuming smooth solutions.
- In the variational framework the regularized solution is computed as

$$\hat{u} = \arg\min_u R(u) + \lambda G(u; f).$$

Regularization Methods
Total Variation

- Consider the inverse problem

$$\min_{u \in \mathrm{BV}(\Omega, R)} R(u) + \frac{\lambda}{2} \|Au - f\|_2^2 ,$$

with $\Omega \subset \boldsymbol{R}^M$, $f \in L^2(\Omega, \boldsymbol{R})$ and a linear operator $A : L^2(\Omega) \to L^2(\Omega)$.

Regularization Methods
Total Variation

- Consider the inverse problem

$$\min_{u \in \mathrm{BV}(\Omega, \boldsymbol{R})} R(u) + \frac{\lambda}{2} \|Au - f\|_2^2 ,$$

with $\Omega \subset \boldsymbol{R}^M$, $f \in L^2(\Omega, \boldsymbol{R})$ and a linear operator $A : L^2(\Omega) \to L^2(\Omega)$.

- A popular regularizer is the total variation [Rudin, Osher, Fatemi '92]:

$$R(u) = \mathrm{TV}(u) = \underbrace{\int_\Omega \|\nabla u(x)\|_2 \, dx}_{u \in \mathcal{C}^1(\Omega, R)} = \underbrace{\sup_{\boldsymbol{\xi} \in \Xi} \left\{ \int_\Omega u \, \mathrm{div}\, \boldsymbol{\xi} \, dx \right\}}_{u \in L^1_{\mathrm{loc}}(\Omega, R)},$$

where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega, \boldsymbol{R}^M) \,:\, \|\boldsymbol{\xi}(x)\|_2 \leq 1, \ \forall x \in \Omega \right\}$.

Regularization Methods
Total Variation

- Consider the inverse problem

$$\min_{u \in \mathrm{BV}(\Omega, \boldsymbol{R})} R(u) + \frac{\lambda}{2} \|Au - f\|_2^2 ,$$

with $\Omega \subset \boldsymbol{R}^M$, $f \in L^2(\Omega, \boldsymbol{R})$ and a linear operator $A : L^2(\Omega) \to L^2(\Omega)$.

- A popular regularizer is the total variation [Rudin, Osher, Fatemi '92]:

$$R(u) = \mathrm{TV}(u) = \underbrace{\int_\Omega \|\nabla u(x)\|_2 \, dx}_{u \in \mathcal{C}^1(\Omega, R)} = \underbrace{\sup_{\boldsymbol{\xi} \in \Xi} \left\{ \int_\Omega u \operatorname{div} \boldsymbol{\xi} \, dx \right\}}_{u \in L^1_{\mathrm{loc}}(\Omega, R)},$$

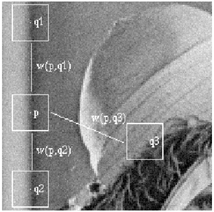
where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega, \boldsymbol{R}^M) \, : \, \|\boldsymbol{\xi}(x)\|_2 \leq 1, \ \forall x \in \Omega \right\}$.

- TV regularizes the image without smoothing the boundaries of the objects, but fails to recover fine structures and texture.

Regularization Methods

Nonlocal Total Variation

- Nonlocal means denoising algorithm [Buades, Coll, Morel '05]:



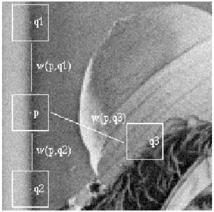
$$NL[u](x) = \frac{1}{\int_\Omega \omega_f(x,y)\,dy} \int_\Omega \omega_f(x,y) u(y)\,dy$$

$$\omega_f(x,y) = \exp\left(-\frac{\|f(P_x), f(P_y)\|}{h^2}\right)$$

Regularization Methods

Nonlocal Total Variation

- Nonlocal means denoising algorithm [Buades, Coll, Morel '05]:



$$NL[u](x) = \frac{1}{\int_\Omega \omega_f(x,y)\,dy} \int_\Omega \omega_f(x,y) u(y)\,dy$$

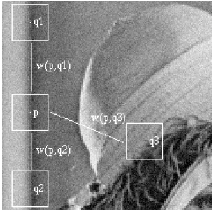
$$\omega_f(x,y) = \exp\left(-\frac{\|f(P_x), f(P_y)\|}{h^2} \right)$$

Natural images are self-similar!

Regularization Methods

Nonlocal Total Variation

- Nonlocal means denoising algorithm [Buades, Coll, Morel '05]:



$$NL[u](x) = \frac{1}{\int_\Omega \omega_f(x,y)\,dy} \int_\Omega \omega_f(x,y)u(y)\,dy$$

$$\omega_f(x,y) = \exp\left(-\frac{\|f(P_x), f(P_y)\|}{h^2}\right)$$

Natural images are self-similar!

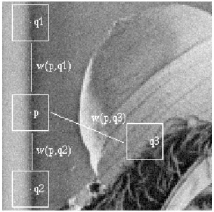
- Neighborhood filters as nonlocal regularization [Gilboa, Osher '08]:

$$\nabla_\omega u(x,y) = (u(y)-u(x))\sqrt{\omega_f(x,y)} \;\rightarrow\; R(u) = \int_\Omega \int_\Omega |\nabla_\omega u(x,y)|^2\,dy\,dx.$$

Regularization Methods

Nonlocal Total Variation

- Nonlocal means denoising algorithm [Buades, Coll, Morel '05]:



$$NL[u](x) = \frac{1}{\int_\Omega \omega_f(x,y)\,dy} \int_\Omega \omega_f(x,y) u(y)\,dy$$

$$\omega_f(x,y) = \exp\left(-\frac{\|f(P_x), f(P_y)\|}{h^2}\right)$$

Natural images are self-similar!

- Neighborhood filters as nonlocal regularization [Gilboa, Osher '08]:

$$\nabla_\omega u(x,y) = (u(y) - u(x))\sqrt{\omega_f(x,y)} \;\rightarrow\; R(u) = \int_\Omega \int_\Omega |\nabla_\omega u(x,y)|^2\,dy\,dx.$$

- The nonlocal total variation is defined as

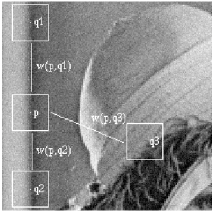
$$R(u) = \mathrm{NLTV}(u) = \int_\Omega \|\nabla_\omega u(x,\cdot)\|_2\,dx = \sup_{\boldsymbol{\xi}\in\Xi}\left\{\int_\Omega u\,\mathrm{div}_\omega\,\boldsymbol{\xi}\,dx\right\},$$

where $\Xi = \left\{\boldsymbol{\xi} \in C_c^1(\Omega\times\Omega, \boldsymbol{R}^M) \;:\; \|\boldsymbol{\xi}(x,\cdot)\|_2 \le 1, \forall x \in \Omega\right\}$.

Regularization Methods

Nonlocal Total Variation

- Nonlocal means denoising algorithm [Buades, Coll, Morel '05]:



$$NL[u](x) = \frac{1}{\int_\Omega \omega_f(x,y)\,dy} \int_\Omega \omega_f(x,y) u(y)\,dy$$

$$\omega_f(x,y) = \exp\left(- \frac{\|f(P_x), f(P_y)\|}{h^2} \right)$$

Natural images are self-similar!

- Neighborhood filters as nonlocal regularization [Gilboa, Osher '08]:

$$\nabla_\omega u(x,y) = (u(y) - u(x)) \sqrt{\omega_f(x,y)} \;\; \rightarrow \;\; R(u) = \int_\Omega \int_\Omega |\nabla_\omega u(x,y)|^2 \, dy\, dx.$$

- The nonlocal total variation is defined as

$$R(u) = \mathrm{NLTV}(u) = \int_\Omega \|\nabla_\omega u(x,\cdot)\|_2 \, dx = \sup_{\boldsymbol{\xi} \in \Xi} \left\{ \int_\Omega u \, \mathrm{div}_\omega \, \boldsymbol{\xi} \, dx \right\},$$

where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega \times \Omega, \boldsymbol{R}^M) \, : \, \|\boldsymbol{\xi}(x,\cdot)\|_2 \leq 1, \forall x \in \Omega \right\}$.

How can we generalize TV and NLTV to color images?

Vectorial Total Variation
Classical approaches

Consider a vector-valued image $\mathbf{u} : \Omega \to \boldsymbol{R}^c$ with C spectral channels.

Vectorial Total Variation

Classical approaches

Consider a vector-valued image $\mathbf{u} : \Omega \to \boldsymbol{R}^c$ with C spectral channels.

- Channel-wise summation [Blomgren, Chan '98]:

$$\mathrm{VTV}(\mathbf{u}) = \sum_{k=1}^{C} \mathrm{TV}(u_k) = \sup_{\boldsymbol{\xi} \in \Xi} \left\{ \sum_{k=1}^{C} \int_{\Omega} u_k \, \mathrm{div}\, \boldsymbol{\xi}_k \, dx \right\},$$

where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega, \boldsymbol{R}^M \times \cdots \times \boldsymbol{R}^M) \, : \, \|\boldsymbol{\xi}_k(x)\| \leq 1, \, \forall x \in \Omega, \, \forall k = 1, \ldots, C \right\}$.

Vectorial Total Variation

Classical approaches

Consider a vector-valued image $\mathbf{u} : \Omega \to \boldsymbol{R}^c$ with C spectral channels.

- Channel-wise summation [Blomgren, Chan '98]:

$$\mathrm{VTV}(\mathbf{u}) = \sum_{k=1}^{C} \mathrm{TV}(u_k) = \sup_{\boldsymbol{\xi} \in \Xi} \left\{ \sum_{k=1}^{C} \int_{\Omega} u_k \,\mathrm{div}\, \boldsymbol{\xi}_k \, dx \right\},$$

where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega, \boldsymbol{R}^M \times \cdots \times \boldsymbol{R}^M) \; : \; \|\boldsymbol{\xi}_k(x)\| \leq 1, \, \forall x \in \Omega, \, \forall k = 1, \ldots, C \right\}$.

- Global channel coupling [Sapiro, Ringach '96]:

$$\mathrm{VTV}(\mathbf{u}) = \int_{\Omega} \|\nabla \mathbf{u}\|_F \, dx = \sup_{\boldsymbol{\xi} \in \Xi} \left\{ \sum_{k=1}^{C} \int_{\Omega} u_k \,\mathrm{div}\, \boldsymbol{\xi}_k \, dx \right\},$$

where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega, \boldsymbol{R}^{M \cdot C}) \; : \; \|\boldsymbol{\xi}(x)\| \leq 1, \, \forall x \in \Omega \right\}$.

Vectorial Total Variation

Classical approaches

Consider a vector-valued image $\mathbf{u} : \Omega \to \boldsymbol{R}^c$ with C spectral channels.

- Channel-wise summation [Blomgren, Chan '98]:

$$\mathrm{VTV}(\mathbf{u}) = \sum_{k=1}^{C} \mathrm{TV}(u_k) = \sup_{\boldsymbol{\xi} \in \Xi} \left\{ \sum_{k=1}^{C} \int_{\Omega} u_k \operatorname{div} \boldsymbol{\xi}_k \, dx \right\},$$

where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega, \boldsymbol{R}^M \times \cdots \times \boldsymbol{R}^M) \, : \, \|\boldsymbol{\xi}_k(x)\| \leq 1, \, \forall x \in \Omega, \, \forall k = 1, \ldots, C \right\}$.

- Global channel coupling [Sapiro, Ringach '96]:

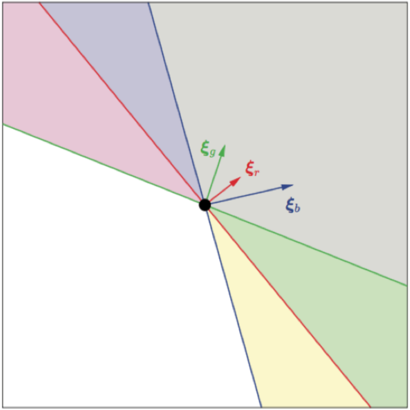
$$\mathrm{VTV}(\mathbf{u}) = \int_{\Omega} \|\nabla \mathbf{u}\|_F \, dx = \sup_{\boldsymbol{\xi} \in \Xi} \left\{ \sum_{k=1}^{C} \int_{\Omega} u_k \operatorname{div} \boldsymbol{\xi}_k \, dx \right\},$$

where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega, \boldsymbol{R}^{M \cdot C}) \, : \, \|\boldsymbol{\xi}(x)\| \leq 1, \, \forall x \in \Omega \right\}$.

- Spectral norm coupling [Goldluecke, Strekalovskiy, Cremers '12]:

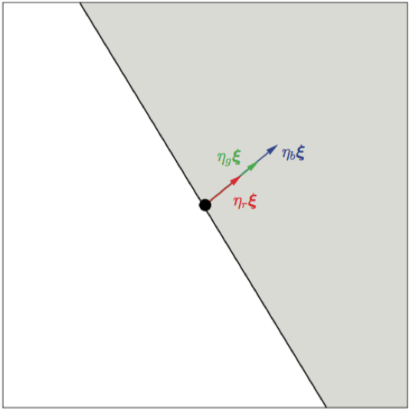
$$\mathrm{VTV}(\mathbf{u}) = \int_{\Omega} \|\nabla \mathbf{u}\|_{\sigma_1} \, dx = \sup_{(\boldsymbol{\xi}, \boldsymbol{\eta}) \in \Xi} \left\{ \sum_{k=1}^{C} \int_{\Omega} u_k \operatorname{div} (\boldsymbol{\eta}_k \boldsymbol{\xi}) \, dx \right\},$$

where $\Xi = \left\{ \boldsymbol{\xi} \in C_c^1(\Omega, \boldsymbol{R}^M), \boldsymbol{\eta} \in C_c^1(\Omega, \boldsymbol{R}^C) \, : \, \|\boldsymbol{\xi}(x)\| \leq 1, \|\boldsymbol{\eta}(x)\| \leq 1, \, \forall x \in \Omega \right\}$.



$$\int_\Omega \|\nabla \mathbf{u}\|_F \, dx$$

Channel coupling
Different edge direction



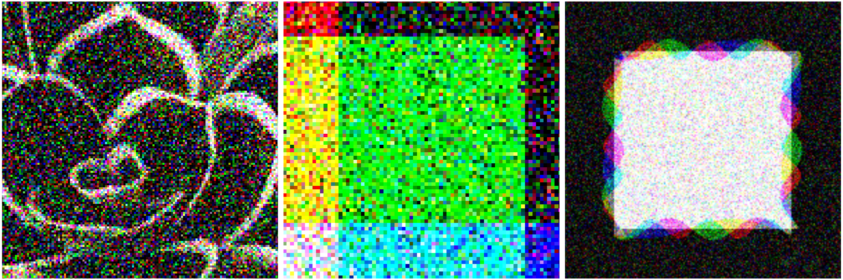
$$\int_\Omega \|\nabla \mathbf{u}\|_{\sigma_1} \, dx$$

Channel coupling
Common edge direction

Image from [Goldluecke, Strekalovskiy, Cremers '12]

Vectorial Total Variation

Which is the best VTV for color images?



Vectorial Total Variation
- coupling spatial derivatives
  - isotropic diffusion
  - anisotropic diffusion
- coupling color channels
  - ℓ^p – type coupling
  - Spectral coupling

Collaborative Total Variation for Multi-Channel Images

Proposed framework

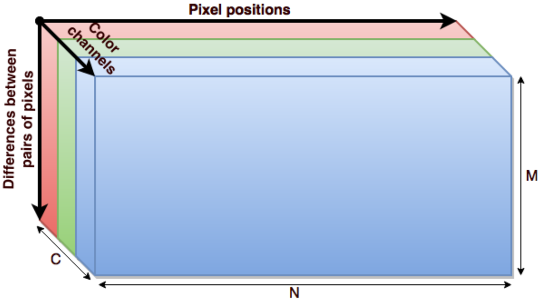
- Represent an image $\mathbf{u}$ with N pixels and C spectral channels by the matrix

$$\mathbf{u} = (u_1, \ldots, u_C) \in \boldsymbol{R}^{N \times C} \text{ s.t. } u_k \in \boldsymbol{R}^N, \forall k \in \{1, \ldots, C\}.$$

Collaborative Total Variation for Multi-Channel Images

Proposed framework

- Represent an image **u** with N pixels and C spectral channels by the matrix
$$\mathbf{u} = (u_1, \ldots, u_C) \in \boldsymbol{R}^{N \times C} \ \text{ s.t. } \ u_k \in \boldsymbol{R}^N, \forall k \in \{1, \ldots, C\}.$$
- The Jacobi matrix at each pixel defines a 3D tensor given by
$$D\mathbf{u} \equiv (Du)_{i,j,k} \in \boldsymbol{R}^{N \times M \times C},$$
with M directional derivatives.



Collaborative Total Variation for Multi-Channel Images

Proposed framework

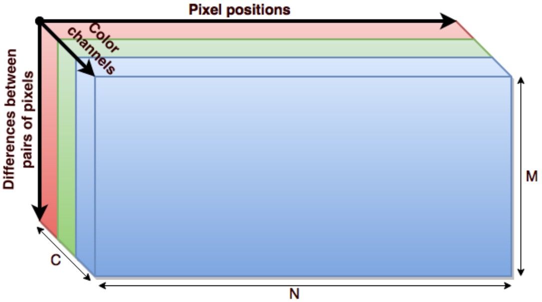
- Represent an image $\mathbf{u}$ with N pixels and C spectral channels by the matrix

$$\mathbf{u} = (u_1, \ldots, u_C) \in \boldsymbol{R}^{N \times C} \text{ s.t. } u_k \in \boldsymbol{R}^N, \forall k \in \{1, \ldots, C\}.$$

- The Jacobi matrix at each pixel defines a 3D tensor given by

$$D\mathbf{u} \equiv (Du)_{i,j,k} \in \boldsymbol{R}^{N \times M \times C},$$

with M directional derivatives.



- For local operators $D\mathbf{u}$ is of size $N \times 2 \times C$, while for nonlocal operators is of size $N \times N_\omega \times C$ with $N_\omega \ll N$ since few nonzero weights are considered.

Collaborative Total Variation for Multi-Channel Images

Collaborative sparsity enforcing norms

- Regularize $D\mathbf{u}$ by penalizing each dimension with a different norm.

Collaborative Total Variation for Multi-Channel Images

Collaborative sparsity enforcing norms

- Regularize $D\mathbf{u}$ by penalizing each dimension with a different norm.

Definition

Let $\|\cdot\|_a : \boldsymbol{R}^N \to \boldsymbol{R}$ be any vector norm and $\|\cdot\|_{\vec{b}} : \boldsymbol{R}^{M\times C} \to \boldsymbol{R}$ any matrix norm. Then, the collaborative norm of $A \in \boldsymbol{R}^{N\times M\times C}$ is defined as

$$\|A\|_{\vec{b},a} = \|v\|_a, \quad \text{with} \quad v_i = \|A_{i,:,:}\|_{\vec{b}}, \ \forall i \in \{1,\dots,N\},$$

where $A_{i,:,:}$ is the submatrix obtained by stacking the second and third dimensions of A at ith position.

Collaborative Total Variation for Multi-Channel Images

Collaborative sparsity enforcing norms

- Regularize $D\mathbf{u}$ by penalizing each dimension with a different norm.

Definition

Let $\|\cdot\|_a : \mathbb{R}^N \to \mathbb{R}$ be any vector norm and $\|\cdot\|_{\vec{b}} : \mathbb{R}^{M\times C} \to \mathbb{R}$ any matrix norm. Then, the collaborative norm of $A \in \mathbb{R}^{N\times M\times C}$ is defined as

$$\|A\|_{\vec{b},a} = \|v\|_a, \quad \text{with} \quad v_i = \|A_{i,:,:}\|_{\vec{b}}, \ \forall i \in \{1,\dots,N\},$$

where $A_{i,:,:}$ is the submatrix obtained by stacking the second and third dimensions of A at ith position.

Example ($\ell^{p,q,r}$ norms)

Let $A \in \mathbb{R}^{N\times M\times C}$ and consider $\|\cdot\|_{\vec{b}} = \ell^{p,q}$ and $\|\cdot\|_a = \ell^r$. Then, the $\ell^{p,q,r}$ norm is

$$\|A\|_{p,q,r} = \left(\sum_{i=1}^{N} \left(\sum_{j=1}^{M} \left(\sum_{k=1}^{C} |A_{i,j,k}|^p \right)^{q/p} \right)^{r/q} \right)^{1/r}.$$

Example ((S^p, ℓ^q) norm)

Let $A \in \mathbf{R}^{N\times M\times C}$ and consider $\|\cdot\|_{\vec{b}} = S^p$ and $\|\cdot\|_a = \ell^q$. Then the (S^p, ℓ^q) norm is

$$(S^p,\ell^q)(A) = \left(\sum_{i=1}^{N} \left\| \begin{pmatrix} A_{i,1,1} & \cdots & A_{i,1,C} \\ \vdots & \ddots & \vdots \\ A_{i,M,1} & \cdots & A_{i,M,C} \end{pmatrix} \right\|_{S^p}^{q} \right)^{1/q}.$$

- Schatten $p-$norms:
  - Fix a pixel location and consider the submatrix obtained by looking at the channel and derivative dimensions.
  - Compute SVD and penalize the singular values with an ℓ^p-norm:
    - $p = 1 \rightarrow$ nuclear norm, a convex relaxation of rank minimization,
    - $p = 2 \rightarrow$ Frobenius norm,
    - $p = \infty \rightarrow$ penalizing the largest singular value.

Example ((S^p, ℓ^q) norm)

Let $A \in \mathbf{R}^{N\times M\times C}$ and consider $\|\cdot\|_{\vec{b}} = S^p$ and $\|\cdot\|_a = \ell^q$. Then the (S^p, ℓ^q) norm is

$$(S^p, \ell^q)(A) = \left(\sum_{i=1}^{N} \left\| \begin{pmatrix} A_{i,1,1} & \cdots & A_{i,1,C} \\ \vdots & \ddots & \vdots \\ A_{i,M,1} & \cdots & A_{i,M,C} \end{pmatrix} \right\|_{S^p}^{q} \right)^{1/q}.$$

- Schatten $p-$norms:
  - Fix a pixel location and consider the submatrix obtained by looking at the channel and derivative dimensions.
  - Compute SVD and penalize the singular values with an ℓ^p-norm:
    - $p = 1 \rightarrow$ nuclear norm, a convex relaxation of rank minimization,
    - $p = 2 \rightarrow$ Frobenius norm,
    - $p = \infty \rightarrow$ penalizing the largest singular value.
- CTV norms are non invariant to permutations of the dimensions:

$$\ell^{p,q,r}(col, der, pix) \text{ and } (S^p(col, der), \ell^q(pix)).$$

Example ((S^p, ℓ^q) norm)

Let $A \in \mathbb{R}^{N\times M\times C}$ and consider $\|\cdot\|_{\vec{b}} = S^p$ and $\|\cdot\|_a = \ell^q$. Then the (S^p, ℓ^q) norm is

$$(S^p, \ell^q)(A) = \left(\sum_{i=1}^{N} \left\| \begin{pmatrix} A_{i,1,1} & \cdots & A_{i,1,C} \\ \vdots & \ddots & \vdots \\ A_{i,M,1} & \cdots & A_{i,M,C} \end{pmatrix} \right\|_{S^p}^{q} \right)^{1/q} .$$

- Schatten $p-$norms:
  - Fix a pixel location and consider the submatrix obtained by looking at the channel and derivative dimensions.
  - Compute SVD and penalize the singular values with an ℓ^p-norm:
    - $p = 1 \rightarrow$ nuclear norm, a convex relaxation of rank minimization,
    - $p = 2 \rightarrow$ Frobenius norm,
    - $p = \infty \rightarrow$ penalizing the largest singular value.
- CTV norms are non invariant to permutations of the dimensions:

$$\ell^{p,q,r}(col, der, pix) \ \text{ and } \ (S^p(col, der), \ell^q(pix)) .$$

- Any transform along each of the dimensions, in particular, color space transforms, can be applied before CTV.

Collaborative Total Variation for Multi-Channel Images

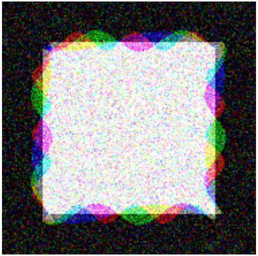
A unified framework for VTV

Continuous Formulation	Our Framework
$\int_\Omega \sum_{k=1}^{C} \sqrt{(\partial_{x_1} u_k(x))^2 + (\partial_{x_2} u_k(x))^2} dx$	$\ell^{2,1,1}(der, col, pix)$
$\int_\Omega \sum_{k=1}^{C} \left(\lvert\partial_{x_1} u_k(x)\rvert + \lvert\partial_{x_2} u_k(x)\rvert\right) dx$	$\ell^{1,1,1}(der, col, pix)$
$\sqrt{\sum_{k=1}^{C} \left(\int_\Omega \sqrt{(\partial_{x_1} u_k(x))^2 + (\partial_{x_2} u_k(x))^2} dx\right)^2}$	$\ell^{2,1,2}(der, pix, col)$
$\sqrt{\sum_{k=1}^{C} \left(\int_\Omega \left(\lvert\partial_{x_1} u_k(x)\rvert + \lvert\partial_{x_2} u_k(x)\rvert\right) dx\right)^2}$	$\ell^{1,1,2}(der, pix, col)$
$\int_\Omega \sqrt{\sum_{k=1}^{C} (\partial_{x_1} u_k(x))^2 + \sum_{k=1}^{C} (\partial_{x_2} u_k(x))^2} dx$	$\ell^{2,2,1}(col, der, pix)$
$\int_\Omega \sqrt{\sum_{k=1}^{C} \left(\lvert\partial_{x_1} u_k(x)\rvert + \lvert\partial_{x_2} u_k(x)\rvert\right)^2} dx$	$\ell^{1,2,1}(der, col, pix)$
$\int_\Omega \left(\sqrt{\sum_{k=1}^{C} (\partial_{x_1} u_k(x))^2} + \sqrt{\sum_{k=1}^{C} (\partial_{x_2} u_k(x))^2}\right) dx$	$\ell^{2,1,1}(col, der, pix)$
$\int_\Omega \left(\max_{1\le k\le C} \lvert\partial_{x_1} u_k(x)\rvert + \max_{1\le k\le C} \lvert\partial_{x_2} u_k(x)\rvert\right) dx$	$\ell^{\infty,1,1}(col, der, pix)$

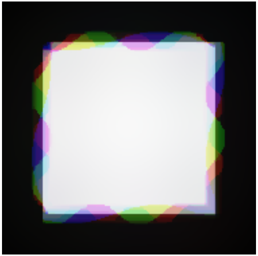
Continuous Formulation	Our Framework
$\int_\Omega \sqrt{\left(\max_{1\le k\le C} \lvert\partial_{x_1} u_k(x)\rvert\right)^2 + \left(\max_{1\le k\le C} \lvert\partial_{x_2} u_k(x)\rvert\right)^2} dx$	$\ell^{\infty,2,1}(col, der, pix)$
$\int_\Omega \max_{1\le k\le C} \sqrt{\left(\partial_{x_1} u_k(x)\right)^2 + \left(\partial_{x_2} u_k(x)\right)^2} dx$	$\ell^{2,\infty,1}(der, col, pix)$
$\int_\Omega \max\left\{\max_{1\le k\le C} \lvert\partial_{x_1} u_k(x)\rvert, \max_{1\le k\le C} \lvert\partial_{x_2} u_k(x)\rvert\right\} dx$	$\ell^{\infty,\infty,1}(col, der, pix)$
$\int_\Omega \left(\sqrt{\lambda^+(x)} + \sqrt{\lambda^-(x)}\right) dx$	$(S^1(col, der), \ell^1(pix))$
$\int_\Omega \sqrt{\lambda^+(x)}\, dx$	$(S^\infty(col, der), \ell^1(pix))$
$\int_\Omega \left(\sum_{k=1}^C \sqrt{\int_\Omega (u_k(y) - u_k(x))^2\, \omega(x,y)\, dy}\right) dx$	$\ell^{2,1,1}_\omega(der, col, pix)$
$\int_\Omega \left(\sum_{k=1}^C \int_\Omega \lvert u(y) - u(x)\rvert \sqrt{\omega(x,y)}\, dy\right) dx$	$\ell^{1,1,1}_\omega(der, col, pix)$
$\sqrt{\sum_{k=1}^C \left(\int_\Omega \sqrt{\int_\Omega (u_k(y) - u_k(x))^2\, \omega(x,y)\, dy\, dx}\right)^2}$	$\ell^{2,1,2}_\omega(der, pix, col)$
$\int_\Omega \int_\Omega \sqrt{\sum_{k=1}^C (u_k(y) - u_k(x))^2\, \omega(x,y)}\, dy\, dx$	$\ell^{2,1,1}_\omega(col, der, pix)$
$\int_\Omega \sqrt{\int_\Omega \sum_{k=1}^C (u_k(y) - u_k(x))^2\, \omega(x,y)\, dy\, dx}$	$\ell^{2,2,1}_\omega(col, der, pix)$
$\int_\Omega \int_\Omega \max_{1\le k\le C} \left((u_k(y) - u_k(x))^2\, \omega(x,y)\right) dy\, dx$	$\ell^{\infty,1,1}_\omega(col, der, pix)$

Which is the Best Channel Coupling?

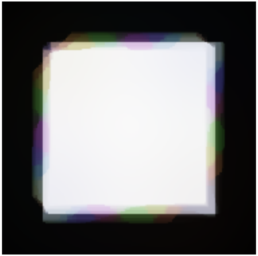
Inter-channel correlation



Noisy



ℓ^1 coupling



ℓ^2 coupling

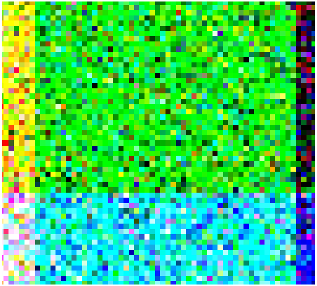
ℓ^∞ coupling

Noisy

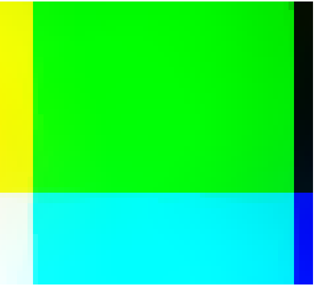
ℓ^1 coupling

ℓ^2 coupling

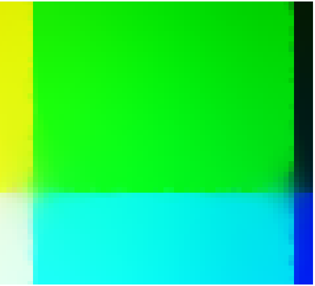
ℓ^∞ coupling



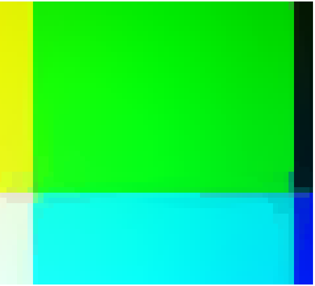
Noisy



ℓ^1 coupling



ℓ^2 coupling



ℓ^∞ coupling

Which is the Best Channel Coupling?

Singular vector analysis

Definition

Let F be a convex regularization s.t. $\partial F(\mathbf{u}) \neq \emptyset$ at any $\mathbf{u} \in \operatorname{dom} F$. Then, every function $\mathbf{u}_\lambda$ s.t. $\|\mathbf{u}_\lambda\| = 1$ and $\lambda \mathbf{u}_\lambda \in \partial F(\mathbf{u}_\lambda)$ is called a singular vector of F with singular value λ.

Which is the Best Channel Coupling?

Singular vector analysis

Definition

Let F be a convex regularization s.t. $\partial F(\mathbf{u}) \neq \emptyset$ at any $\mathbf{u} \in \operatorname{dom} F$. Then, every function $\mathbf{u}_\lambda$ s.t. $\|\mathbf{u}_\lambda\| = 1$ and $\lambda \mathbf{u}_\lambda \in \partial F(\mathbf{u}_\lambda)$ is called a singular vector of F with singular value λ.

- A signal can be restored well if it is a singular vector of F [Benning, Burger '13].

Which is the Best Channel Coupling?

Singular vector analysis

Definition

Let F be a convex regularization s.t. $\partial F(\mathbf{u}) \neq \emptyset$ at any $\mathbf{u} \in \operatorname{dom} F$. Then, every function $\mathbf{u}_\lambda$ s.t. $\|\mathbf{u}_\lambda\| = 1$ and $\lambda \mathbf{u}_\lambda \in \partial F(\mathbf{u}_\lambda)$ is called a singular vector of F with singular value λ.

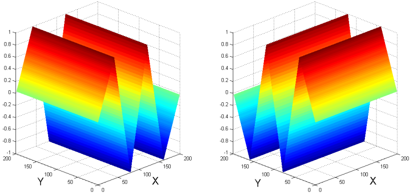
- A signal can be restored well if it is a singular vector of F [Benning, Burger '13].
- Singular vectors of CTV:

$$\mathbf{u} \in \partial \|D\mathbf{u}\|_{\vec{b},a} \Leftrightarrow \mathbf{u} = D^\top \mathbf{z}, \text{ with } \mathbf{z} \in \partial_{D\mathbf{u}}\left(\|D\mathbf{u}\|_{\vec{b},a}\right).$$

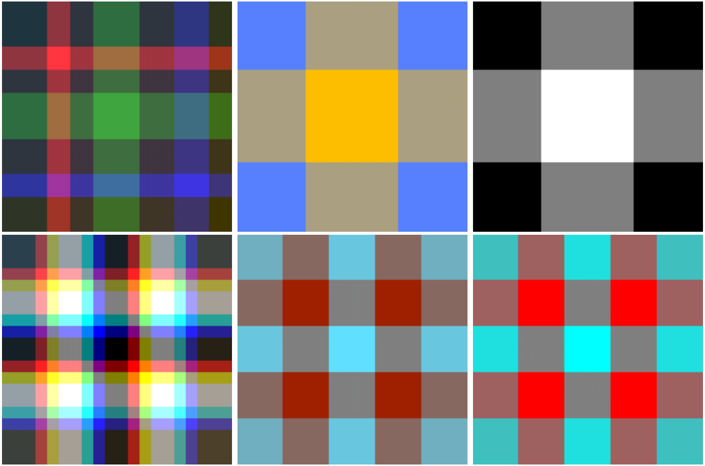
The functions whose divergence generates singular vectors reduce to

$$z_k^1(x_1, x_2) = c_k^1 l_k^1(x_1) \text{ and } z_k^2(x_1, x_2) = c_k^2 l_k^2(x_2),$$

where $c_k^r \in \boldsymbol{R}$, $|l_k^r(x)| \leq 1$, l_k^r piecewise linear and linearity changes iff $|l_k^r(x)| = 1$.

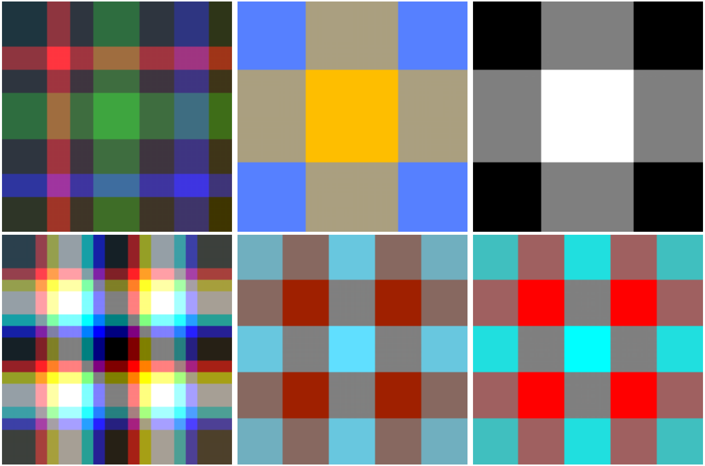


CTV	Singular Vectors	Properties
$\ell^{1,1,1}$	$u_k(x_1,x_2) = -c_k^1 D_1 l_k^1(x_1) - c_k^2 D_2 l_k^2(x_2)$	l_k^r depend on k and $c_k^r \in \{0, \pm 1\}$
$\ell^{2,1,1}$	$u_k(x_1,x_2) = -c_k^1 D_1 l^1(x_1) - c_k^2 D_2 l^2(x_2)$	l^r do not depend on k and $\lVert c^r \rVert_2 = 1$
$\ell^{\infty,1,1}$	$u_k(x_1,x_2) = -c_k^1 D_1 l^1(x_1) - c_k^2 D_2 l^2(x_2)$	l^r do not depend on k and $c_k^r \in \{0, \pm 1\}$



$\ell^{1,1,1}$ $\ell^{2,1,1}$ $\ell^{\infty,1,1}$

CTV	Singular Vectors	Properties
$\ell^{1,1,1}$	$u_k(x_1,x_2) = -c_k^1 D_1 l_k^1(x_1) - c_k^2 D_2 l_k^2(x_2)$	l_k^r depend on k and $c_k^r \in \{0, \pm 1\}$
$\ell^{2,1,1}$	$u_k(x_1,x_2) = -c_k^1 D_1 l^1(x_1) - c_k^2 D_2 l^2(x_2)$	l^r do not depend on k and $\|c^r\|_2 = 1$
$\ell^{\infty,1,1}$	$u_k(x_1,x_2) = -c_k^1 D_1 l^1(x_1) - c_k^2 D_2 l^2(x_2)$	l^r do not depend on k and $c_k^r \in \{0, \pm 1\}$



$\ell^{1,1,1}$ $\ell^{2,1,1}$ $\ell^{\infty,1,1}$

The ℓ^{∞} norm introduces the strongest channel coupling!

Primal-Dual Minimization

- Primal formulation:

$$\min_{\mathbf{u}\in R^{N\times C}} F(\mathbf{u}) + G(\mathbf{u}) = \|D\mathbf{u}\|_{\vec{b},a} + G(\mathbf{u}).$$

Primal-Dual Minimization

- Primal formulation:

$$\min_{\mathbf{u}\in R^{N\times C}} F\left(\mathbf{u}\right) + G\left(\mathbf{u}\right) = \|D\mathbf{u}\|_{\vec{b},a} + G\left(\mathbf{u}\right).$$

- Since F is closed and l.s.c., then

$$F(D\mathbf{u}) = F^{**}(D\mathbf{u}) = \sup_{\mathbf{p}\in R^{N\times M\times C}} \langle D\mathbf{u}, \mathbf{p}\rangle - F^{*}\left(\mathbf{p}\right).$$

Primal-Dual Minimization

- Primal formulation:

$$\min_{\mathbf{u}\in R^{N\times C}} F\left(\mathbf{u}\right) + G\left(\mathbf{u}\right) = \|D\mathbf{u}\|_{\vec{b},a} + G\left(\mathbf{u}\right).$$

- Since F is closed and l.s.c., then

$$F(D\mathbf{u}) = F^{**}(D\mathbf{u}) = \sup_{\mathbf{p}\in R^{N\times M\times C}} \langle D\mathbf{u}, \mathbf{p}\rangle - F^{*}\left(\mathbf{p}\right).$$

- If $F = \|\cdot\|$, then its Legendre-Fenchel transform is

$$F^{*}\left(\mathbf{p}\right) = \left\{ \begin{array}{ll} 0 & \text{if } \|\mathbf{p}\|_{*} \leq 1 \\ +\infty & \text{otherwise} \end{array} \right\} = \mathcal{X}_{\|\cdot\|_{*}\leq 1}\left(\mathbf{p}\right).$$

Primal-Dual Minimization

- Primal formulation:

$$\min_{\mathbf{u}\in R^{N\times C}} F(\mathbf{u}) + G(\mathbf{u}) = \|D\mathbf{u}\|_{\vec{b},a} + G(\mathbf{u}).$$

- Since F is closed and l.s.c., then

$$F(D\mathbf{u}) = F^{**}(D\mathbf{u}) = \sup_{\mathbf{p}\in R^{N\times M\times C}} \langle D\mathbf{u}, \mathbf{p}\rangle - F^{*}(\mathbf{p}).$$

- If $F = \|\cdot\|$, then its Legendre-Fenchel transform is

$$F^{*}(\mathbf{p}) = \left\{ \begin{array}{ll} 0 & \text{if } \|\mathbf{p}\|_{*} \leq 1 \\ +\infty & \text{otherwise} \end{array} \right\} = \mathcal{X}_{\|\cdot\|_{*}\leq 1}(\mathbf{p}).$$

Theorem

Let $\|\cdot\|_{\vec{b}^*}$ and $\|\cdot\|_{a^*}$ be the dual norms to $\|\cdot\|_{\vec{b}}$ and $\|\cdot\|_{a}$, respectively. Consider $A \in \mathbf{R}^{N\times M\times C}$ and define $v \in \mathbf{R}^{N}$ such that $v_i = \|A_{i,:,:}\|_{\vec{b}^*}$ for each $i \in \{1,\ldots,N\}$. If $\|v\|_{a^*}$ only depends on the absolute values of v_i's, then the dual norm to $\|\cdot\|_{\vec{b},a}$ is

$$\|A\|_{\vec{b}^*,a^*} = \|v\|_{a^*}, \quad \textit{with} \quad v_i = \|A_{i,:,:}\|_{\vec{b}^*}, \ \forall i \in \{1,\ldots,N\}.$$

- Saddle-point formulation:

$$\min_{\mathbf{u}\in R^{N\times C}} \max_{\mathbf{p}\in R^{N\times M\times C}} \langle D\mathbf{u}, \mathbf{p}\rangle - F^*(\mathbf{p}) + G(\mathbf{u}),$$

with optimality conditions

$$0 \in \partial G(\widehat{\mathbf{u}}) + D^\top \widehat{\mathbf{p}} \;\text{ and }\; 0 \in \partial F^*(\widehat{\mathbf{p}}) - D\widehat{\mathbf{u}}.$$

- Saddle-point formulation:

$$\min_{\mathbf{u}\in R^{N\times C}} \max_{\mathbf{p}\in R^{N\times M\times C}} \langle D\mathbf{u}, \mathbf{p}\rangle - F^*(\mathbf{p}) + G(\mathbf{u}),$$

with optimality conditions

$$0 \in \partial G(\widehat{\mathbf{u}}) + D^\top \widehat{\mathbf{p}} \;\text{ and }\; 0 \in \partial F^*(\widehat{\mathbf{p}}) - D\widehat{\mathbf{u}}.$$

- Primal-Dual algorithm [Chambolle, Pock '11]:

$$\begin{aligned}
\mathbf{u}^{n+1} &= \text{prox}_{\tau_n G}\left(\mathbf{u}^n - \tau_n D^\top \mathbf{p}^n\right) && \leftarrow \quad \text{Gradient descent step in } \mathbf{u} \\
\bar{\mathbf{u}}^{n+1} &= \mathbf{u}^{n+1} + \left(\mathbf{u}^{n+1} - \mathbf{u}^n\right), && \leftarrow \quad \text{Over-relaxation step in } \mathbf{u} \\
\mathbf{p}^{n+1} &= \text{prox}_{\sigma_n F^*}\left(\mathbf{p}^n + \sigma_n D\bar{\mathbf{u}}^{n+1}\right) && \leftarrow \quad \text{Gradient ascent step in } \mathbf{p}
\end{aligned}$$

where $\tau_n, \sigma_n > 0$ are adaptive step-size parameters and

$$\text{prox}_{\alpha f}(x) = \arg\min_y \left\{ \frac{1}{2\alpha}\|y - x\|_2^2 + f(y) \right\}.$$

- Saddle-point formulation:

$$\min_{\mathbf{u}\in R^{N\times C}} \max_{\mathbf{p}\in R^{N\times M\times C}} \langle D\mathbf{u},\mathbf{p}\rangle - F^*(\mathbf{p}) + G(\mathbf{u}),$$

with optimality conditions

$$0\in\partial G(\widehat{\mathbf{u}}) + D^\top\widehat{\mathbf{p}} \text{ and } 0\in\partial F^*(\widehat{\mathbf{p}}) - D\widehat{\mathbf{u}}.$$

- Primal-Dual algorithm [Chambolle, Pock '11]:

$$\begin{aligned}
\mathbf{u}^{n+1} &= \text{prox}_{\tau_n G}\left(\mathbf{u}^n - \tau_n D^\top\mathbf{p}^n\right) && \leftarrow \quad \text{Gradient descent step in } \mathbf{u}\\
\bar{\mathbf{u}}^{n+1} &= \mathbf{u}^{n+1} + \left(\mathbf{u}^{n+1} - \mathbf{u}^n\right), && \leftarrow \quad \text{Over-relaxation step in } \mathbf{u}\\
\mathbf{p}^{n+1} &= \text{prox}_{\sigma_n F^*}\left(\mathbf{p}^n + \sigma_n D\bar{\mathbf{u}}^{n+1}\right) && \leftarrow \quad \text{Gradient ascent step in } \mathbf{p}
\end{aligned}$$

where $\tau_n,\sigma_n > 0$ are adaptive step-size parameters and

$$\text{prox}_{\alpha f}(x) = \arg\min_y\left\{\frac{1}{2\alpha}\|y-x\|_2^2 + f(y)\right\}.$$

- The proximity operator of $F^* = \mathcal{X}_{\|\cdot\|_{\vec{b}*,a*}\leq 1}$ is

$$\text{prox}_{\sigma F^*}(\mathbf{p}) = \text{proj}_{\|\cdot\|_{\vec{b}*,a*}\leq 1}(\mathbf{p}).$$

Experimental Results

Image denoising

Noisy ($\sigma = 30$)

$\ell^{1,1,1}(col, der, pix)$

Experimental Results

Image denoising

Noisy ($\sigma = 30$)

$(S^{\infty}(col, der), \ell^{1}(pix))$

Experimental Results
Image denoising

Noisy ($\sigma = 30$)

$\ell^{2,1,1}(col, der, pix)$

Experimental Results

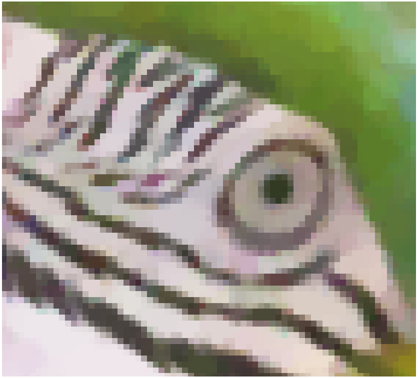
Image denoising

Noisy ($\sigma = 30$)

$\ell^{2,\infty,1}(der, col, pix)$

Experimental Results
Image denoising

Noisy ($\sigma = 30$)



$(S^1(col, der), \ell^1(pix))$

Experimental Results

Image denoising

Noisy ($\sigma = 30$)

$\ell^{\infty,1,1}(col, der, pix)$

Behaviour of CTV methods w.r.t. changing regularization parameter

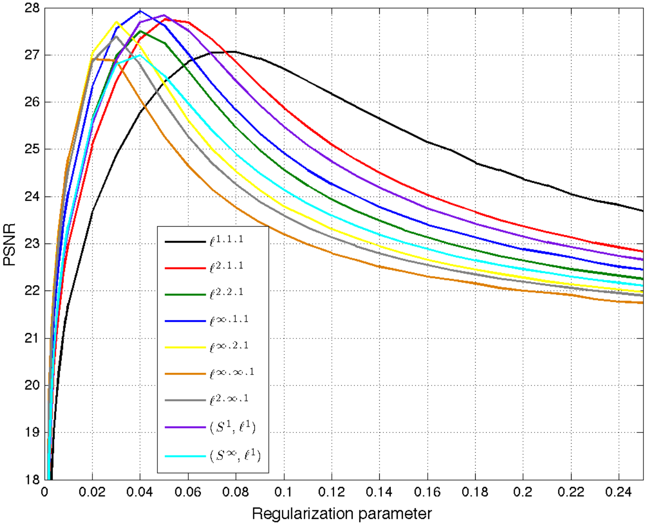


Image denoising on Kodak dataset

	Noisy	$\ell^{1,1,1}$	$\ell^{2,1,1}$	$\ell^{2,2,1}$	$\ell^{\infty,1,1}$	$\ell^{\infty,2,1}$	$\ell^{\infty,\infty,1}$	$\ell^{2,\infty,1}$	S^1, ℓ^1	S^∞, ℓ^1
1	24.78	28.14	29.07	28.51	**29.90**	29.19	28.60	29.07	29.20	27.96
2	24.76	28.54	29.48	29.22	**30.18**	29.87	29.36	29.66	29.83	28.62
3	24.80	29.20	30.15	29.81	**30.85**	30.51	29.84	30.25	30.33	29.24
4	24.68	30.92	32.22	31.80	**32.73**	32.71	31.54	32.13	32.32	31.01
5	24.71	31.50	32.75	32.41	33.13	**33.30**	32.10	32.64	32.81	31.65
6	24.72	27.36	28.19	27.98	**29.01**	28.64	28.29	28.52	28.59	27.47
7	24.71	29.46	30.39	30.12	**30.86**	30.71	29.99	30.35	30.57	29.53
8	24.96	31.08	32.10	31.84	**32.41**	32.40	31.62	32.02	32.20	31.22
9	25.68	30.92	31.74	31.54	**32.10**	32.00	31.49	31.78	31.85	31.11
10	24.66	29.75	30.81	30.49	**31.48**	31.29	30.52	30.94	31.05	29.84
11	24.66	30.14	31.10	30.84	**31.49**	31.46	30.68	31.07	31.22	30.25
12	24.71	31.85	33.15	32.84	33.45	**33.69**	32.47	33.03	33.25	32.05
	24.82	29.91	30.93	30.62	**31.47**	31.31	30.54	30.96	31.10	30.00

Image denoising on McMaster dataset

	Noisy	$\ell^{1,1,1}$	$\ell^{2,1,1}$	$\ell^{2,2,1}$	$\ell^{\infty,1,1}$	$\ell^{\infty,2,1}$	$\ell^{\infty,\infty,1}$	$\ell^{2,\infty,1}$	S^1, ℓ^1	S^∞, ℓ^1
1	25.32	29.29	29.83	29.64	29.74	29.52	28.97	29.25	**29.98**	29.16
2	24.90	27.80	28.41	28.26	28.43	28.32	27.80	28.02	**28.60**	27.75
3	25.46	30.44	30.96	30.84	30.78	30.66	30.16	30.39	**31.17**	30.33
4	25.14	29.26	29.91	29.75	29.95	29.82	29.30	29.54	**30.13**	29.22
5	25.62	31.11	31.46	31.40	30.97	30.84	30.33	30.55	**31.64**	30.89
6	25.01	29.83	30.49	30.32	30.34	30.13	29.55	29.84	**30.74**	29.68
7	25.21	30.96	31.63	31.48	31.41	31.21	30.66	30.98	**31.80**	30.87
8	25.34	31.98	32.72	32.60	32.50	32.30	31.78	32.15	**32.88**	31.99
9	25.21	32.54	33.36	33.32	33.08	32.93	32.50	32.85	**33.53**	32.70
10	24.69	32.26	33.06	33.02	32.70	32.54	32.10	32.49	**33.20**	32.37
11	25.55	30.21	30.85	30.75	30.87	30.73	30.35	30.59	**30.98**	30.29
12	25.21	30.58	31.18	30.99	31.11	30.87	30.36	30.69	**31.30**	30.50
	25.22	30.52	31.16	31.03	30.99	30.82	30.32	30.61	**31.33**	30.48

Experimental Results

Image denoising: local vs nonlocal CTV

Noisy

$\ell^{1,1,1} - \text{TV}$, PSNR $= 33.60$

Experimental Results

Image denoising: local vs nonlocal CTV

Noisy

$\ell^{1,1,1}$ – NLTV, PSNR = 35.41

Experimental Results

Image denoising: local vs nonlocal CTV

Noisy

$\ell^{\infty,1,1}$ – TV, PSNR = 34.88

Experimental Results

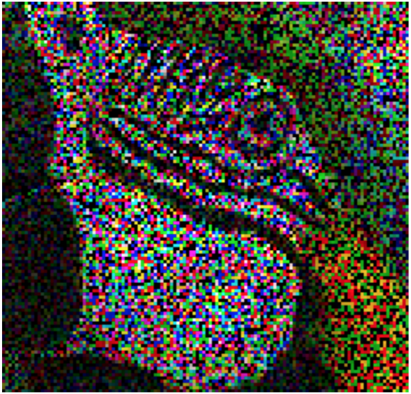
Image denoising: local vs nonlocal CTV

Noisy

$\ell^{\infty,1,1}$ – NLTV, PSNR = 35.65

Experimental Results

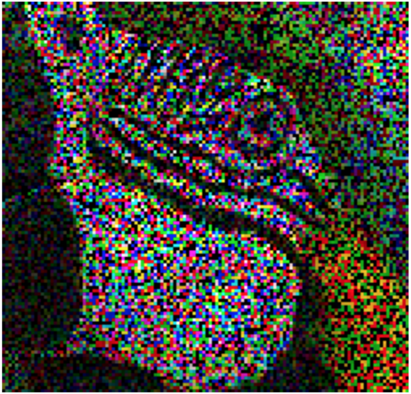
Image inpainting



Noisy ($\sigma = 30$)

$\ell^{1,1,1}(col, der, pix)$

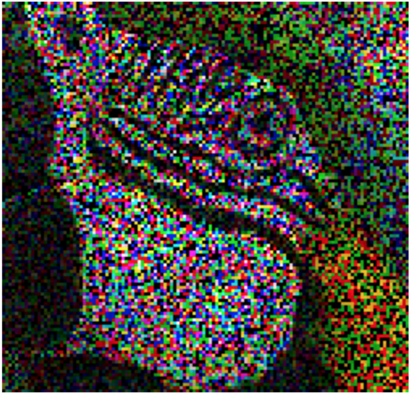
Experimental Results
Image inpainting



Noisy ($\sigma = 30$)

$(S^{\infty}(col, der), \ell^{1}(pix))$

Experimental Results
Image inpainting

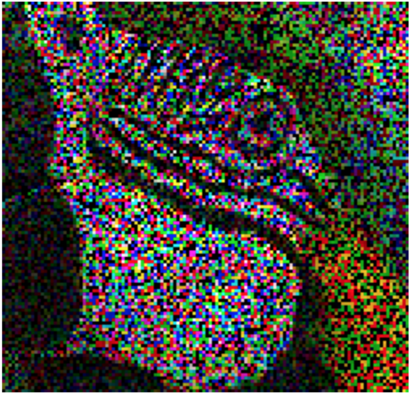


Noisy ($\sigma = 30$)

$\ell^{2,\infty,1}(der, col, pix)$

Experimental Results

Image inpainting

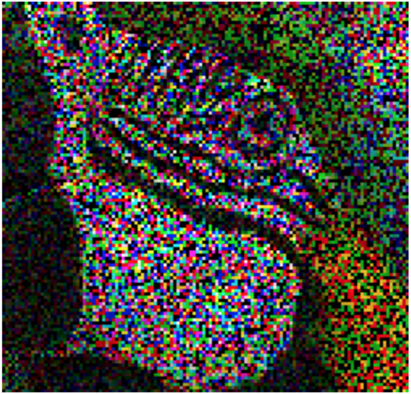


Noisy ($\sigma = 30$)

$\ell^{2,1,1}(col, der, pix)$

Experimental Results

Image inpainting

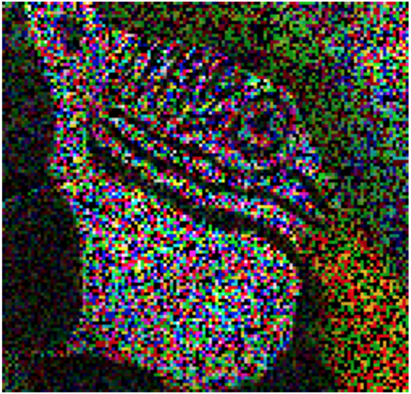


Noisy ($\sigma = 30$)

$(S^1(col, der), \ell^1(pix))$

Experimental Results

Image inpainting



Noisy ($\sigma = 30$)

$\ell^{\infty,1,1}(col, der, pix)$

Experimental Results
On Line Demo

IPOL
IMAGE PROCESSING ON LINE

Conclusions

- We introduced a unified framework for VTV based on collaborative enforcing norms.
- Depending on the inter-channel correlation, different CTV regularizations are suited.
- $\ell^{\infty,1,1}$ and (S^1, ℓ^1) best exploit inter-channel correlations.
- We introduced respective Nonlocal CTV regularizations.
- We proposed the primal-dual algorithm to solve the minimization problem.

Conclusions

- We introduced a unified framework for VTV based on collaborative enforcing norms.
- Depending on the inter-channel correlation, different CTV regularizations are suited.
- $\ell^{\infty,1,1}$ and $\left(S^1, \ell^1\right)$ best exploit inter-channel correlations.
- We introduced respective Nonlocal CTV regularizations.
- We proposed the primal-dual algorithm to solve the minimization problem.

References

- D., Moeller, Sbert, Cremers, "Collaborative Total Variation: A General Framework for Vectorial TV Models", SIIMS, 2016.
- D., Moeller, Sbert, Cremers, "On the Implementation of Collaborative TV Regularization: Application to Cartoon + Texture Decomposition", IPOL, 2016.
- D., Moeller, Sbert, Cremers, "A Novel Framework for Nonlocal Vectorial Total Variation Based on $\ell^{p,q,r}$-Norms", Proc. EMMCVPR, 2015.

Acknowledgements

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD

AGENCIA ESTATAL DE INVESTIGACIÓN

Collaborative Regularization Models for Color Imaging Problems

SIAM IS, Bologna, June 2018

Joan Duran*

*Dept. Mathematics and Computer Science, IAC3

Universitat de les Illes Balears, Mallorca, Spain

joint work with

Michael Moeller‡, Daniel Cremers§ and Catalina Sbert*

‡Inst. Vision and Graphics
University of Siegen, Germany

§Dept. Computer Science
Technical University of Munich, Germany

Universitat de les Illes Balears

Proximity operator of $\ell^{p,q,r}$ norms

- $\ell^{1,1,1}-$norm:

$$\left(\text{prox}_{\frac{1}{\sigma}\|\cdot\|_{1,1,1}}(A)\right)_{i,j,k} = \max\left(|A_{i,j,k}| - \frac{1}{\sigma}, 0\right)\text{sign}\left(A_{i,j,k}\right).$$

- $\ell^{2,1,1}-$norm:

$$\left(\text{prox}_{\frac{1}{\sigma}\|\cdot\|_{2,1,1}}(A)\right)_{i,j,k} = \max\left(\|A_{i,j,:}\|_2 - \frac{1}{\sigma}, 0\right)\frac{A_{i,j,k}}{\|A_{i,j,:}\|_2}.$$

- $\ell^{2,2,1}-$norm:

$$\left(\text{prox}_{\frac{1}{\sigma}\|\cdot\|_{2,2,1}}(A)\right)_{i,j,k} = \max\left(\|A_{i,:,:}\|_{2,2} - \frac{1}{\sigma}, 0\right)\frac{A_{i,j,k}}{\|A_{i,:,:}\|_{2,2}}.$$

- $\ell^{\infty,1,1}-$norm decouples at each j and k so we are left with an ℓ^{∞} problem computed by means of the projection onto unit ℓ^1 dual ball:

$$\left(\text{prox}_{\frac{1}{\sigma}\|\cdot\|_{\infty,1,1}}(A)\right)_{i,j,k} = A_{i,j,k} - \frac{1}{\sigma}\text{sign}\left(A_{i,j,k}\right)\left(\text{proj}_{\|\cdot\|_1\leq 1}\left(\sigma|A_{i,j,:}|\right)\right)_{i,j,k},$$

where $A_{i,j,:}$ denotes the vector obtained by staking third dimension.

- $\ell^{\infty,\infty,1}-$norm:

$$\left(\text{prox}_{\frac{1}{\sigma}\|\cdot\|_{\infty,\infty,1}}(A)\right)_{i,j,k} = A_{i,j,k} - \frac{1}{\sigma}\text{sign}\left(A_{i,j,k}\right)\left(\text{proj}_{\|\cdot\|_{1,1}\leq 1}\left(\sigma|A_{i,:,:}|\right)\right)_{i,j,k},$$

with $A_{i,:,:}$ being the vector obtained by stacking second and third dimensions.

- $\ell^{\infty,2,1}$−norm:

$$\left(\text{prox}_{\frac{1}{\sigma}\|\cdot\|_{\infty,2,1}}(A)\right)_{i,j,k} = A_{i,j,k} - \frac{1}{\sigma}\text{sign}(A_{i,j,k}) \left(\text{proj}_{\|\cdot\|_{1,2}\leq 1}(\sigma|A_{i,:,:}|)\right)_{i,j,k},$$

where $\text{proj}_{\|\cdot\|_{1,2}\leq 1}$ denotes the projection onto unit $\ell^{1,2}$−norm ball.

- $\ell^{2,\infty,1}$−norm:

$$\left(\text{prox}_{\frac{1}{\sigma}\|\cdot\|_{2,\infty,1}}(A)\right)_{i,j,k} = \frac{A_{i,j,k}}{\|A_{i,j,:}\|_2} \max\left(\|A_{i,j,:}\|_2 - \frac{1}{\sigma}v_{i,j}, 0\right),$$

where $v_{i,j} = \left(\text{prox}_{\|\cdot\|_1\leq 1}\left(\sigma\left(\|A_{i,j,:}\|_2\right)_j\right)\right)_{i,j}$, and $\left(\|A_{i,j,:}\|_2\right)_j$ denotes the vector obtained by stacking $\|A_{i,j,:}\|_2$ for all j.

Theorem

Let $f : \mathbf{R}^{n\times m} \to \mathbf{R}^n$ *be* $f_i(u) := \sqrt{\sum_{j=1}^m u_{i,j}^2} = \|u_{i,:}\|_2$, *and let* $g : \mathbf{R}^n \to \mathbf{R}$ *be proper convex function being nondecreasing in each argument. Then*

$$\left(prox_{\tau(g\circ f)}(u)\right)_{i,j} = \frac{u_{i,j}}{\|u_{i,:}\|_2} \max\left(\|u_{i,:}\|_2 - \tau v_i, 0\right),$$

where the v_i*'s are the components of the vector* $v \in \mathbf{R}^n$ *that solves*

$$v = \arg\min_{w\in R^n} \frac{1}{2}\left\|w - \frac{1}{\tau}f(u)\right\|^2 + \frac{1}{\tau}g^*(w).$$

Proximity operators of (S^p, ℓ^q) norms

If $q = 1$, the proximity operator decouples at each pixel:

- Define $M \times C$ submatrix $B_i := (A_{i,j,k})_{j=1,\dots,M;\, k=1,\dots,C}$.
- Let $B = B_i^T$, we need to solve at each pixel

$$\min_{D \in R^{M \times C}} \frac{1}{2}\|D - B\|_F^2 + \frac{1}{\sigma}\|D\|_{S^p}.$$

- Computing SVD of $B = U\Sigma_0 V^T$ and $\Sigma = U^T D V$, the problem is equivalent to

$$\min_{D \in R^{M \times C}} \frac{1}{2}\|U^T D V - \Sigma_0\|_F^2 + \frac{1}{\sigma}\|U^T D V\|_{S^p} \Leftrightarrow \min_{\Sigma \in R^{r \times r}} \frac{1}{2}\|\Sigma - \Sigma_0\|_F^2 + \frac{1}{\sigma}\|\Sigma\|_{S^p}.$$

- For diagonal matrices $S^p(\Sigma) = \ell^p\big(\mathrm{diag}(\Sigma)\big)$, so that we finally solve

$$\min_{s \in R^r} \frac{1}{2}\|s - s_0\|_2^2 + \frac{1}{\sigma}\|s\|_p,$$

  where $s_0 = \mathrm{diag}(\Sigma_0)$ and $s = \mathrm{diag}(\Sigma)$.

- Only need to compute eigenvalues, Σ_0, and eigenvectors, V, of $B_i^T B_i$.

- Let $\widehat{\Sigma}$ s.t. $\text{diag}(\widehat{\Sigma}) = \arg\min_s \frac{1}{2}\|s - s_0\|_2^2 + \frac{1}{\sigma}\|s\|_p$
- The proximity operator $\widehat{D} = \arg\min_D \frac{1}{2}\|D - B\|_2^F + \frac{1}{\sigma}\|D\|_{S^p}$ is $\widehat{D} = U\widehat{\Sigma}V^T$.
- Due to $B = U\Sigma_0 V^T$, communitation of diagonal matrices, and $\widehat{\Sigma}\Sigma_0\Sigma_0^\dagger = \widehat{\Sigma}$ – since $\widehat{\Sigma}$ has at most as many nonzero diagonal entries as Σ_0 –, one has

$$\begin{aligned} BV = U\Sigma_0 &\Rightarrow BV\widehat{\Sigma} = U\Sigma_0\widehat{\Sigma} = U\widehat{\Sigma}\Sigma_0 \\ &\Rightarrow BV\widehat{\Sigma}\Sigma_0^\dagger = U\widehat{\Sigma} \Rightarrow BV\widehat{\Sigma}\Sigma_0^\dagger V^T = U\widehat{\Sigma}V^T = \widehat{D}, \end{aligned}$$

where $\Sigma_0^\dagger$ denotes the pseudo-inverse matrix of Σ_0, i.e.

$$\left(\Sigma_0^\dagger\right)_{i,j} = \begin{cases} \dfrac{1}{(\Sigma_0)_{i,i}} & \text{if } i = j \text{ and } (s_0)_{i,i} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

- Therefore, the proximity operator is

$$\widehat{D} = BV\widehat{\Sigma}\Sigma^\dagger{}_0 V^T,$$

where

- $\text{diag}(\Sigma_0)$ consists of the square root of the eigenvalues of B^TB.
- $\text{col}(V)$ are the eigenvectors of B^TB.

Image Denoising

$$\min_{u \in R^{N\times C}} \|Ku\|_{b,a} + \frac{\lambda}{2}\|u - f\|_F^2,$$

where $f \in \mathbb{R}^{N\times C}$ is the noisy image, $\lambda > 0$ the regularization parameter, and $\|\cdot\|_{b,a}$ denotes either an $\ell^{p,q,r}$ norm or a Schatten (S^p, ℓ^q) norm.

The **proximity operator** of $G(u) = \frac{\lambda}{2}\|u - f\|_F^2$ is

$$\operatorname{prox}_{\tau G}(u) = \arg\min_{v \in X} \left\{ \frac{1}{2}\|v - u\|_F^2 + \tau\frac{\lambda}{2}\|v - f\|_F^2 \right\} \Leftrightarrow \operatorname{prox}_{\tau G}(u) = \frac{u + \tau\lambda f}{1 + \tau\lambda}.$$

Therefore, the solution of $u^{n+1} = \operatorname{prox}_{\tau_n G}\left(u^n - \tau_n K^T z^n\right)$ is given by

$$u^{n+1} = \frac{u^n + \tau_n\left(-K^T z^n + \lambda f\right)}{1 + \tau_n\lambda},$$

where $-K^T = \operatorname{div}$ is defined as $\langle -\operatorname{div} z, u\rangle_X = \langle z, Ku\rangle_Y$.

Image Deconvolution

$$\min_{u \in R^{N\times C}} \|Ku\|_{b,a} + \frac{\lambda}{2}\|Au - f\|_F^2,$$

with A being the linear operator modelling the convolution of u with a Gaussian kernel.s

The **proximity operator** of $G(u) = \frac{\lambda}{2}\|Au - f\|_F^2$ is

$$\widehat{u} = \arg\min_{v \in X} \left\{ \frac{1}{2}\|v - u\|_F^2 + \tau\frac{\lambda}{2}\|Av - f\|_F^2 \right\} \Leftrightarrow \widehat{u} = (I + \tau\lambda A^* A)^{-1}(u + \tau\lambda A^* f).$$

Computing $(I + \tau\lambda A^* A)^{-1}$ is huge time consuming in the spatial domain. On the contrary, using FFT, the solution can be efficiently computed as

$$\widehat{u} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(u) + \tau\lambda\mathcal{F}(A)\mathcal{F}(f)}{1 + \tau\lambda\mathcal{F}(A)^2}\right).$$