Inverse problems: from regularized methods to learning

Nelly Pustelnik
CNRS, Laboratoire de Physique de l'ENS de Lyon

Workshop on Machine-Learning Assisted Image Formation
July, 10 2019

Sparsity

- Signal processing: decompose complex signals using elementary functions which are then easier to manipulate.

$$x(t) = \sum_{i=-\infty}^{+\infty} \alpha_i \varphi_i(t)$$

Sparsity

- Signal processing: decompose complex signals using elementary functions which are then easier to manipulate.

$$x(t) = \sum_{i=-\infty}^{+\infty} \alpha_i \varphi_i(t)$$

  - Fourier transform (1768)
    $\rightarrow$ Uncertainty principle: the energy spread of a function and its Fourier transform cannot be simultaneously arbitrarily small.
    $\rightarrow$ DFT and FFT (Gauss 1805, Cooley-Tukey 1965).

  - Wavelets transform: multiresolution
    $\rightarrow$ I. Daubechies: Compact support wavelet (1988).
    $\rightarrow$ DWT and Mallat recursive algorithm (1989).

Sparsity

- Signal processing: decompose complex signals using elementary functions which are then easier to manipulate.

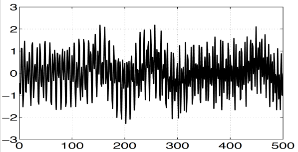
$$x(t) = \sum_{i=-\infty}^{+\infty} \alpha_i \varphi_i(t) \quad \Rightarrow \boxed{\text{Sparse} = \text{Few non-zero } \alpha_i}$$

- Fourier transform (1768)
  $\rightarrow$ Uncertainty principle: the energy spread of a function and its Fourier transform cannot be simultaneously arbitrarily small.
  $\rightarrow$ DFT and FFT (Gauss 1805, Cooley-Tukey 1965).

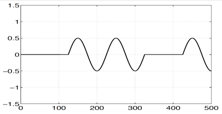
- Wavelets transform: multiresolution
  $\rightarrow$ I. Daubechies: Compact support wavelet (1988).
  $\rightarrow$ DWT and Mallat recursive algorithm (1989).

Sparsity

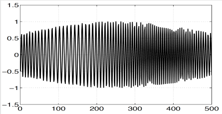
Example: Time-Frequency representation



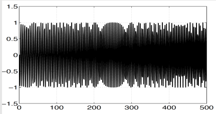
$=$



$+$

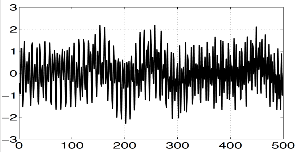


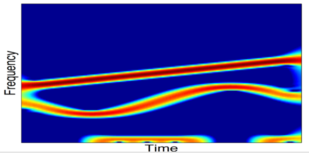
$+$



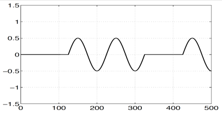
Sparsity

Example: Time-Frequency representation

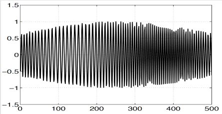




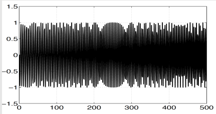
=



+



+

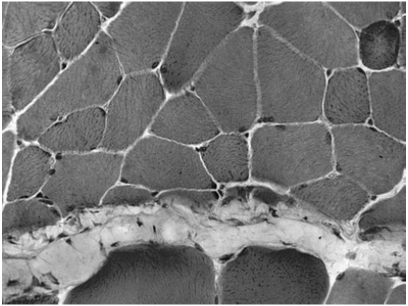


Wavelet transform

- Discrete setting: images on a grid $\Omega = \{1, \ldots, N_1\} \times \{1, \ldots, N_2\}$

$$\mathbf{x} = (x_{n_1,n_2})_{(n_1,n_2)\in\Omega}$$

$\rightarrow$ Vectorized representation denoted $x \in \mathbb{R}^N$ with $N = N_1 N_2$.



$\mathbf{x}$

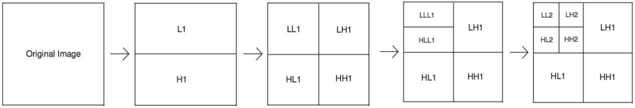
Wavelet transform

- Discrete setting: images on a grid $\Omega = \{1, \ldots, N_1\} \times \{1, \ldots, N_2\}$

$$\mathbf{x} = (x_{n_1,n_2})_{(n_1,n_2)\in\Omega}$$

$\rightarrow$ Vectorized representation denoted $x \in \mathbb{R}^N$ with $N = N_1 N_2$.

- Wavelets:
    - sparse representation of most natural signals/images.
    - DWT, denoted $F \in \mathbb{R}^{N\times N}$
      $\rightarrow$ orthonormal transform: $FF^* = F^*F = I$.
      $\rightarrow$ filterbank implementation:



Wavelet transform

- Discrete setting: images on a grid $\Omega = \{1, \dots, N_1\} \times \{1, \dots, N_2\}$

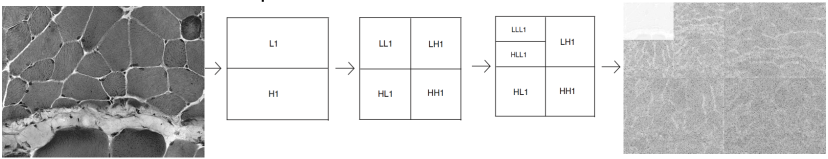
$$\mathbf{x} = (x_{n_1,n_2})_{(n_1,n_2)\in\Omega}$$

$\rightarrow$ Vectorized representation denoted $x \in \mathbb{R}^N$ with $N = N_1 N_2$.

- Wavelets:
  - sparse representation of most natural signals/images.
  - DWT, denoted $F \in \mathbb{R}^{N\times N}$

    $\rightarrow$ orthonormal transform: $FF^* = F^*F = I$.

    $\rightarrow$ filterbank implementation:

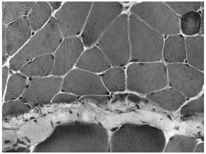


$\mathbf{x}$ $\boldsymbol{\alpha}$

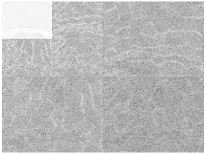
$(\alpha = Fx)$

Wavelet shrinkage

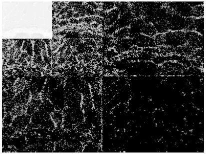
(Donoho-Jonhstone, 1992)



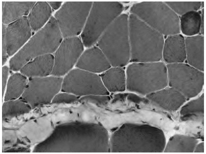
x



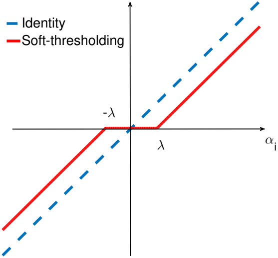
$\alpha = Fx$



$\mathrm{soft}_\lambda(Fx)$

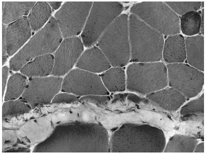


$\widehat{u} = F^*\mathrm{soft}_\lambda(Fx)$

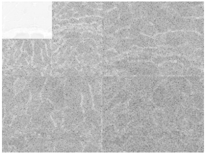


Wavelet shrinkage

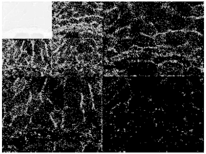
(Donoho-Jonhstone, 1992)



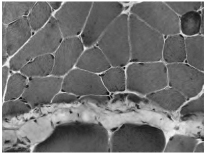
x



$\alpha = Fx$



$\mathrm{soft}_\lambda(Fx)$

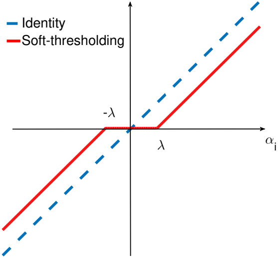


$\hat{u} = F^*\mathrm{soft}_\lambda(Fx)$

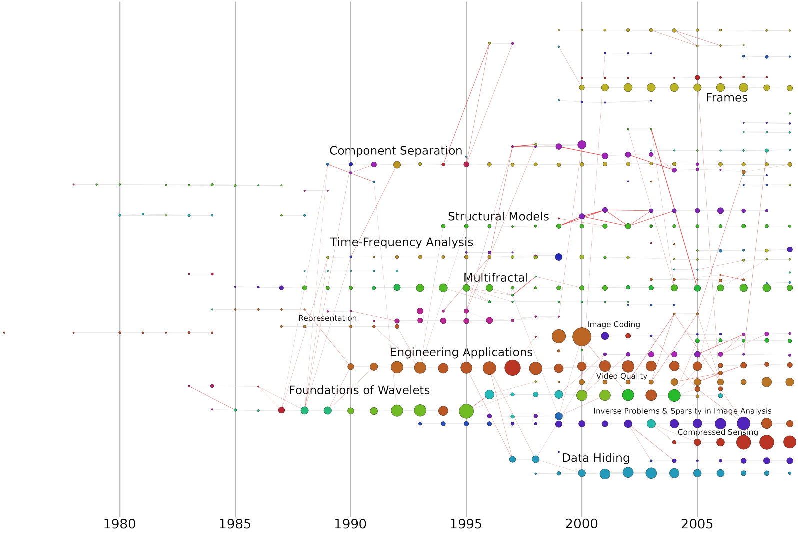
$$\mathrm{soft}_\lambda(\alpha) = \Big(\max\{|\alpha_{\underline{i}}| - \lambda, 0\}\mathrm{sign}(\alpha_{\underline{i}})\Big)_{\underline{i}\in\Omega}$$

$$= \underset{\nu\in\mathbb{R}^N}{\arg\min}\ \frac{1}{2}\|\nu - \alpha\|_2^2 + \lambda \underbrace{\sum_{\underline{i}} |\nu_{\underline{i}}|}_{\|\boldsymbol{\nu}\|_1}$$

$$\hat{u} = \underset{u\in\mathbb{R}^N}{\arg\min}\ \frac{1}{2}\|u - x\|_2^2 + \lambda\|Fu\|_1$$

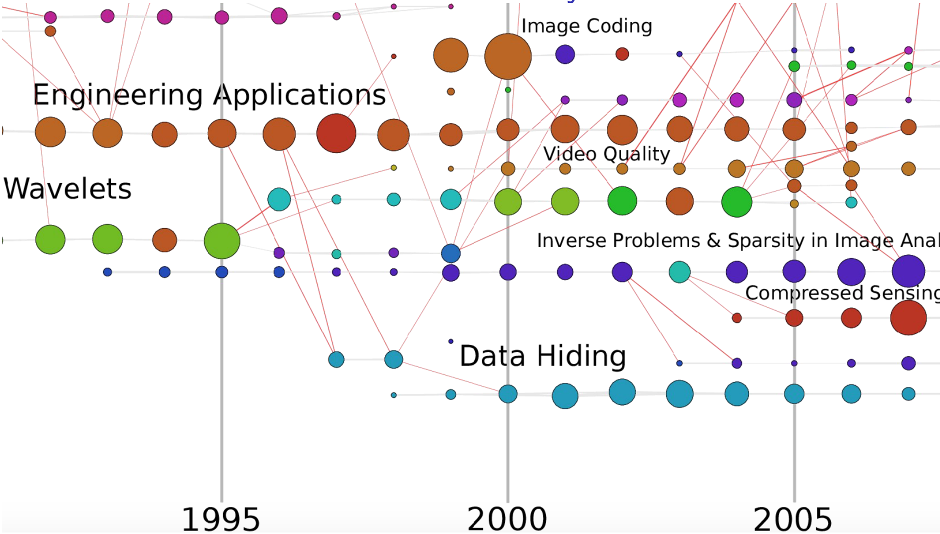


Wavelet: history



(extracted from M. Morini, P. Flandrin, E. Fleury, T. Venturini, P. Jensen, "Revealing evolutions in dynamical networks," 2018, arXiv:1707.02114)

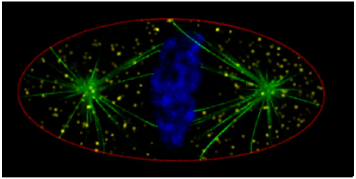
Wavelet: history



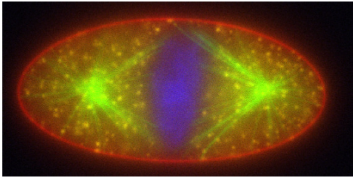
(extracted from M. Morini, P. Flandrin, E. Fleury, T. Venturini, P. Jensen, "Revealing evolutions in dynamical networks," 2018, arXiv:1707.02114)

Inverse problems

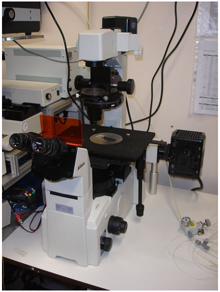
[Microscopy, ISBI Challenge 2013, F. Soulez]



Original image

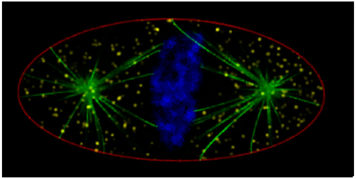


Degraded image



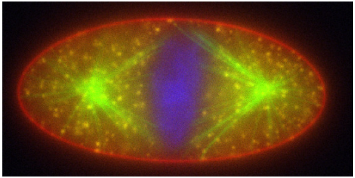
Inverse problems

[Microscopy, ISBI Challenge 2013, F. Soulez]



Original image

$\overline{x} \in \mathbb{R}^N$



Degraded image

$z = \mathcal{P}_\alpha(H\overline{x}) \in \mathbb{R}^M$

- $H \in \mathbb{R}^{M \times N}$: matrix associated with the degradation operator.
- $\mathcal{P}_\alpha \colon \mathbb{R}^M \to \mathbb{R}^M$: noise degradation with parameter α (e.g. Poisson noise).

Inverse problem: Find an estimate $\widehat{x}$ close to $\overline{x}$ from the observations z.

Inverse problems

Inverse problem: Find an estimate $\widehat{x}$ close to $\overline{x}$ from the observations z.

- Inverse filtering (if $M = N$ and H is invertible)

$$\begin{aligned}\widehat{x} &= H^{-1}z \\ &= H^{-1}(H\overline{x} + b) \quad \leftarrow \quad \text{if } b \in \mathbb{R}^M \text{ is an additive noise} \\ &= \overline{x} + H^{-1}b\end{aligned}$$

$\rightarrow$ Closed form expression, but amplification of the noise if H is ill-conditioned (*ill-posed problem*).

Inverse problems

Inverse problem: Find an estimate $\widehat{x}$ close to $\overline{x}$ from the observations z.

- Inverse filtering
- Variational approach: $\widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \mathcal{L}(Hx, z) + \lambda \mathcal{R}(Fx)$
  - $\Gamma_0(\mathcal{H})$: class of convex, lower semi-continuous, proper functions from $\mathbb{R}^N$ to $]-\infty, +\infty]$.
  - $\mathcal{L}(Hx, z)$: data fidelity term (in $\Gamma_0(\mathbb{R}^M)$),
  - $\mathcal{R}(Fx)$: regularization term (in $\Gamma_0(\mathbb{R}^N)$),
  - $\lambda > 0$: regularization parameter.

Inverse problems

Inverse problem: Find an estimate $\widehat{x}$ close to $\overline{x}$ from the observations z.

- Inverse filtering
- Variational approach: $\widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \mathcal{L}(Hx, z) + \lambda \mathcal{R}(Fx)$
  - $\Gamma_0(\mathcal{H})$: class of convex, lower semi-continuous, proper functions from $\mathbb{R}^N$ to $]-\infty, +\infty]$.
  - $\mathcal{L}(Hx, z)$: data fidelity term (in $\Gamma_0(\mathbb{R}^M)$),
  - $\mathcal{R}(Fx)$: regularization term (in $\Gamma_0(\mathbb{R}^N)$),
  - $\lambda > 0$: regularization parameter.
  - Example: ℓ_1-norm to deal with sparsity

$$\widehat{x} \in \underset{x \in \mathbb{R}}{\operatorname{Argmin}} \frac{1}{2} \|Hx - z\|_2^2 + \lambda \|Fx\|_1$$

$\rightarrow$ Soft-thresholding : $H = \mathrm{Id}$ (closed-form expression)

Bayesian interpretation

- $v = Hx = (v_{\underline{n}})_{\underline{n}\in\Omega}$: realization of a random vector V.
- z: realization of a random vector Z.
- $\alpha = Fx = (\alpha_{\underline{i}})_{\underline{i}\in\Upsilon}$: realization of a random vector $A = (A_{\underline{i}})_{\underline{i}\in\Upsilon}$ having independent components.

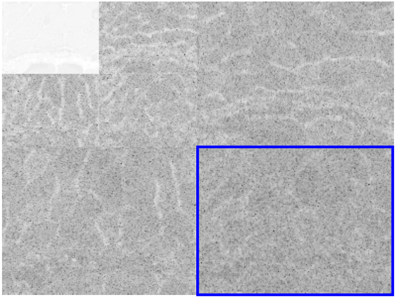
MAP estimator (Maximum A Posteriori)

$$\underset{x}{\text{maximize}} \;\; \mathsf{P}(V = Hx \mid Z = z)$$

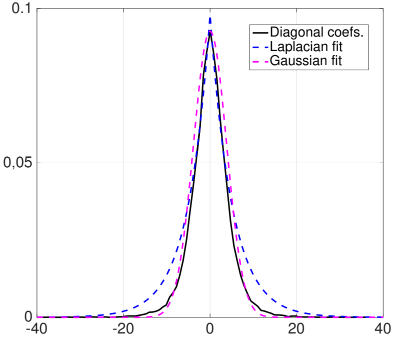
$$\underset{\boldsymbol{\alpha}}{\text{maximize}} \;\; \mathsf{P}(Z = z \mid V = HF^*\boldsymbol{\alpha}) \cdot \mathsf{P}(A = \alpha)$$

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \;\; \underbrace{-\ln \mathsf{P}(Z = z \mid V = HF^*\alpha)}_{\text{Data fidelity}} \underbrace{- \sum_{\underline{i}\in\Upsilon} \ln \mathsf{p}_{A_{\underline{i}}}(\alpha_{\underline{i}})}_{\text{A priori}}$$

Bayesian interpretation



Wavelet coefficients



Probability density function

Bayesian interpretation

MAP estimator (Maximum A Posteriori)

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \quad \underbrace{-\ln \mathrm{P}(X = x \mid V = HF^*\alpha)}_{\text{Data fidelity}} \underbrace{- \sum_{\underline{i} \in \Upsilon} \ln \mathrm{p}_{A_{\underline{i}}}(\alpha_{\underline{i}})}_{\text{A priori}}$$

where

$$\mathrm{P}(Z = z \mid V = HF^*\alpha) = \frac{1}{(2\pi\sigma^2)^{|\Omega|/2}} \exp\left\{ -\frac{\|HF^*\alpha - z\|_2^2}{2\sigma^2} \right\}$$

and

$$\mathrm{p}_{A_{\underline{i}}}(\alpha_{\underline{i}}) = \frac{1}{C_{\underline{i}}} \exp\{-\lambda_{\underline{i}} |\alpha_{\underline{i}}|\}$$

$$\underset{\alpha}{\text{minimize}} \quad \frac{1}{2\sigma^2} \|HF^*\alpha - z\|_2^2 + \sum_{\underline{i} \in \Upsilon} \lambda_{\underline{i}} |\alpha_{\underline{i}}|$$

$\Rightarrow$ "there can be other admissible Bayesian interpretations" (Gribonval, 2011)

Choice of $\mathcal{L}$, $\mathcal{R}$ and F

Synthesis formulation

$$\widehat{x} = F^*\widehat{\alpha} \quad \text{with} \quad \widehat{\alpha} \in \underset{\alpha}{\mathrm{Argmin}}\ \mathcal{L}(HF^*\alpha, z) + \lambda\mathcal{R}(\alpha) \qquad \lambda > 0$$

Analysis formulation $\quad \widehat{x} \in \underset{x}{\mathrm{Argmin}}\ \mathcal{L}(Hx, z) + \lambda\mathcal{R}(Fx) \qquad \lambda > 0$

- Analysis versus Synthesis
  - Equivalence for F orthonormal basis.
  - The analysis formulation is a particular case of the synthesis formulation.
- Few numerical comparisons.

(Elad, Milanfar, Ron, 2007) (Chaari, Pustelnik, Chaux, Pesquet, 2009) (Selesnick, Figueiredo, 2009), (Carlavan, Weiss, Blanc-Féraud, 2010) (Pustelnik, Benazza-Benhayia, Zheng, Pesquet, 2010)

Choice of $\mathcal{L}$, $\mathcal{R}$ and F

Observations	25.90	23.46	21.23	19.71	18.49
TV	27.10	26.33	25.38	24.77	24.53
DTCW (R)	27.50	26.70	25.77	**25.25**	**25.16**
Curvelets (R)	27.40	26.58	25.49	25.02	24.87
RDWT (R)	27.69	26.47	25.79	24.78	24.45
RDWT + Curvelets (R)	27.58	26.65	25.63	25.02	24.78
DTCW + Curvelets (R)	27.44	26.65	25.71	25.21	25.12
RDWT + DTCW (R)	27.77	26.70	25.72	25.09	24.86
DTCW (P)	27.73	26.78	**25.83**	25.24	25.15
Curvelets (P)	27.50	26.55	25.47	24.95	24.78
RDWT (P)	27.60	26.20	25.09	24.33	23.91
RDWT + Curvelets (P)	27.66	26.56	25.43	24.80	24.50
DTCW + Curvelets (P)	27.77	26.81	25.74	25.14	24.96
RDWT + DTCW (P)	**27.97**	**26.84**	25.58	24.75	24.33

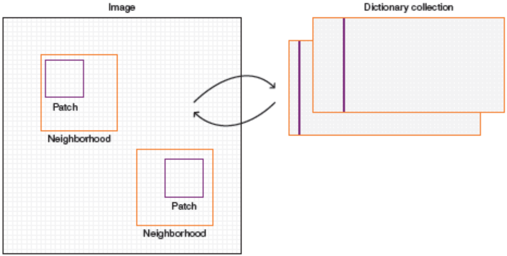
Tableau 1. *PSNR en dB des différentes régularisations utilisées sur l'image Barbara. (P) désigne un a priori de parcimonie tandis que (R) désigne un a priori de régularité.*

(extracted M. Carlavan, P. Weiss, L. Blanc-Féraud "Régularité et parcimonie pour les problèmes inverses en imagerie : algorithmes et comparaisons", Traitement du Signal, sept. 2010.) (P) Synthesis, (R) Analysis

Choice of $\mathcal{L}$, $\mathcal{R}$ and F

Choice for F (analysis):

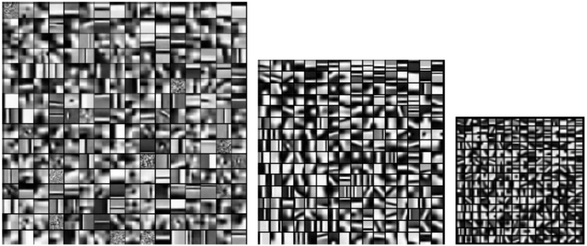
- Total variation: horizontal/vertical gradient.
- Hessian operator: second order derivative along horizontal, diagonal and vertical direction.
- Nonlocal total variation: weighted nonlocal gradients (Gilboa, Osher, 2008)(Bougleux, Peyré, Cohen, 2011)
- Local dictionaries of patches (Boulanger, Pustelnik, Condat, Piolot, Sengmanivong, 2018)



Choice of $\mathcal{L}$, $\mathcal{R}$ and F

Choice for F (synthesis):

- X-lets (webpage L. Duval) (Jacques, Duval, Chaux, Peyré, 2011)
- Sparse coding: Dictionary of patches: set of elementary signals (Aharon, Elad, Bruckstein, 2006)



(a) $s = 0$ (b) $s = 1$ (c) $s = 2$

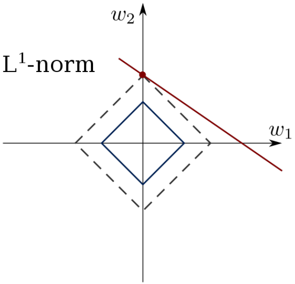
FIG. 6.1. *A learned 3-scales global dictionary, which has been trained over a large database of natural images.*

(extracted from Mairal, Sapiro, Elad, learning multiscale sparse representations for image and video restoration, 2007)

Choice of $\mathcal{L}$, $\mathcal{R}$ and F

Choice for $\mathcal{R}$:

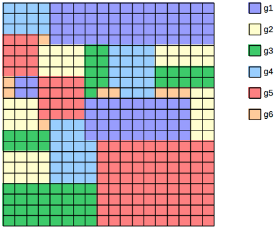
- ℓ_1-norm: $\mathcal{R} = \|\cdot\|_1$

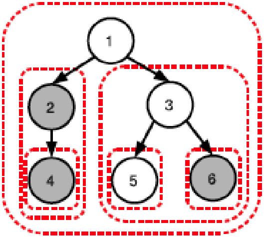


Choice of $\mathcal{L}$, $\mathcal{R}$ and F

Choice for $\mathcal{R}$:

- ℓ_1-norm: $\mathcal{R} = \|\cdot\|_1$
- Mixed-norm: $\mathcal{R} = \sum_{g \in \mathcal{G}} \|\theta_g\|_q$ with $q \geq 1$.
  - Non-overlapping groups: e.g. TV
  - Overlapping groups: Tree-structure (Zhao, Rocha, Yu, 2007), union of groups (Jacob, Obozinski, Vert, 2009).





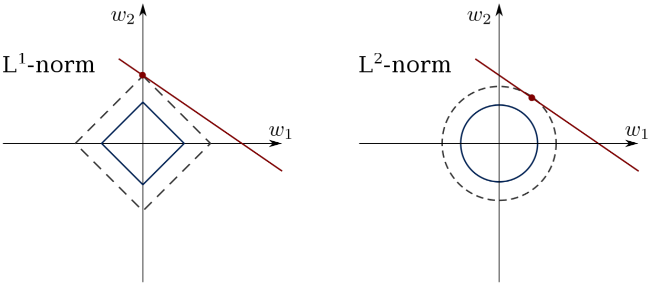
Choice of $\mathcal{L}$, $\mathcal{R}$ and F

Choice for $\mathcal{R}$:

- ℓ_1-norm: $\mathcal{R} = \| \cdot \|_1$
- Mixed-norm: $\mathcal{R} = \sum_{g \in \mathcal{G}} \|\theta_g\|_q$ with $q \geq 1$.
    - Non-overlapping groups: e.g. TV
    - Overlapping groups: Tree-structure (Zhao, Rocha, Yu, 2007), union of groups (Jacob, Obozinski, Vert, 2009).
- Schatten/Nuclear norm: $\mathcal{R} = \| \cdot \|_*$
- Non-convex:
    - $| \cdot |^q$ with $q \in]0, 1[$ (Frank, Friedman, 1993)
    - Log penalty: $\log(| \cdot | + \varepsilon)$ (Candès, Wakin, Boyd, 2008)
    - Several others (Nikolova, 2007)
    - Non-convex penalties leading to convex criterion (Parekh, Selesnick, 2015)

Choice of $\mathcal{L}$, $\mathcal{R}$ and F

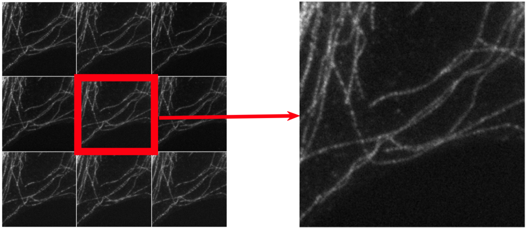
Choice for $\mathcal{R}$: $\min_{\eta_1 w_1 + \eta_2 w_2 = z} |w_1| + |w_2|$



Choice of $\mathcal{L}$ and $\mathcal{R}$

- Image degraded with Poisson noise

  e.g. tomography, microscopy



- Poisson likelihood

$$P(Z = z \mid V = HF^*\boldsymbol{\alpha}) = \prod_{\underline{n} \in |\Omega|} \frac{\exp(-\sigma(HF^*\boldsymbol{\alpha})_{\underline{n}}}{z_{\underline{n}}!} (\sigma(HF^*\boldsymbol{\alpha})_{\underline{n}})^{z_{\underline{n}}}$$

- Data fidelity term: Kullback-Leibler divergence

$$\mathcal{L}(HF^*\alpha, z) = \sum_{\underline{n} \in \widetilde{\Omega}} \psi_{\underline{n}}\big(\underbrace{(HF^*\alpha)_{\underline{n}}}_{v_{\underline{n}}}\big) \quad \text{where} \quad \psi_{\underline{n}}(v_{\underline{n}}) = -z_{\underline{n}} \ln \sigma v_{\underline{n}} + \sigma v_{\underline{n}}$$

(Combettes,Pesquet,2007) (Setzer, Steidl, Teuber, 2010) (Figueiredo, Bioucas-Dias, 2010) (Pustelnik, Chaux, Pesquet, 2011) (Antoine, Aujol, Boursier, Mélot, 2012)

Choice of $\mathcal{L}$, $\mathcal{R}$ and F

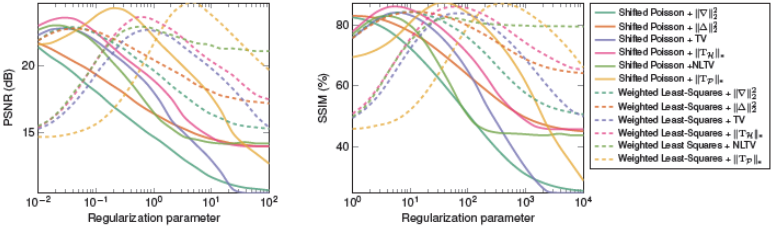


Figure 8. Evolution of the PSNR and the SSIM criterion as a function of the regularization parameters for the test image A.

Table 2. Best PSNR (dB) / SSIM (%) performance for both test images.

	$\|\nabla\|_2^2$	$\|\Delta\|_2^2$	TV	$\|T_{\mathcal{H}}\|_*$	NLTV	$\|T_{\mathcal{P}}\|_*$
Shifted Poisson	21.35 / 82.53	21.73 / 82.96	22.79 / 84.00	23.58 / 85.94	23.02 / 82.95	**24.28** / **87.18**
Weighted Least-Squares	22.88 / 84.20	22.83 / 84.30	22.66 / 83.84	23.61 / 85.94	22.94 / 82.36	**24.67** / **87.00**
	$\|\nabla\|_2^2$	$\|\Delta\|_2^2$	TV	$\|T_{\mathcal{H}}\|_*$	NLTV	$\|T_{\mathcal{P}}\|_*$
Shifted Poisson	26.70 / 85.42	26.86 / 85.89	27.99 / 87.59	28.38 / 88.08	27.86 / 86.52	**28.53** / 88.42
Weighted Least-Squares	27.82 / 86.86	27.79 / 86.97	27.68 / 87.81	28.03 / 88.51	27.12 / 85.57	28.16 / **88.61**

(extracted from Boulanger, Pustelnik, Condat, Piolot, Sengmanivong, Nonsmooth convex optimization for Structured Illumination Microscopy image reconstruction, 2018.)

Algorithmic strategy

Minimization problem

$$\text{Find} \qquad \widehat{y} \in \underset{y \in \mathcal{H}}{\text{Argmin}} \sum_{j=1}^{J} f_j(y)$$

where $(f_j)_{1 \leq j \leq J}$ belong to the class of convex functions, l.s.c., and proper from $\mathcal{H}$ to $]-\infty, +\infty]$. $\mathcal{H}$ finite dimensional Hilbert space.

- Example 1: $\widehat{u} \in \underset{\mathbf{u} \in \mathbb{R}^{|\Omega|}}{\text{Argmin}} \frac{1}{2}\|Hu - z\|_2^2 + \lambda \|Fu\|_1 + \iota_{\geq 0}(u)$
- Example 2: $\widehat{\alpha} \in \underset{\alpha \in \mathbb{R}^{|\Upsilon|}}{\text{Argmin}} \frac{1}{2}\|HF^*\alpha - z\|_2^2 + \lambda \|\alpha\|_1$
- Example 3: $\widehat{u} \in \underset{\mathbf{u} \in \mathbb{R}^{|\Omega|}}{\text{Argmin}} \sum_{\underline{n}} -z_{\underline{n}} \ln \sigma u_{\underline{n}} + \sigma u_{\underline{n}} + \lambda \sum_{g \in \mathcal{G}} \|(Fu)_g\|_2$

Algorithmic strategy

Minimization problem

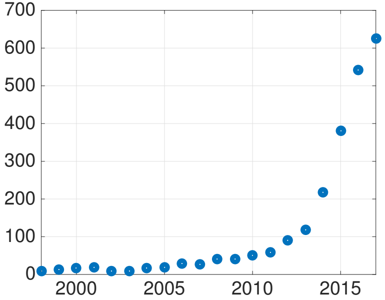
$$\text{Find} \qquad \widehat{y} \in \underset{y \in \mathcal{H}}{\operatorname{Argmin}} \sum_{j=1}^{J} f_j(y)$$

- Properties of the involved functions
  - smooth functions
    $\rightarrow$ gradient-based methods (Newton, Quasi-Newton, ...)
  - constraints
    $\rightarrow$ projection based methods (POCS, SIRT, ...)
  - non-smooth functions
    $\rightarrow$ proximal algorithms (FB, DR, PPXA, ADMM, Primal-Dual,...)
    $\rightarrow$ possible extension to infinite dimensional space.
    $\rightarrow$ flexibility in the design of objective functions.

Algorithmic strategy

Minimization problem

$$\text{Find} \qquad \widehat{y} \in \underset{y \in \mathcal{H}}{\operatorname{Argmin}} \sum_{j=1}^{J} f_j(y)$$



Number of articles per year on Google scholar containing "proximal algorithms" since 1997.

Proximity operator

Gradient descent

Solve $\widehat{y} \in \mathrm{Argmin}_y f(y)$ when $f \in \Gamma_0(\mathcal{H})$ with a Lipschitz gradient $\beta > 0$.

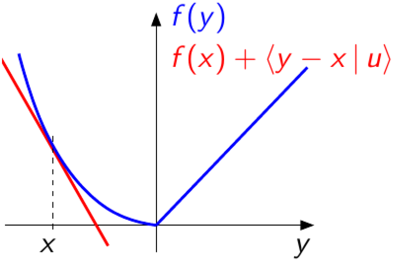
Set $\gamma_n \in]0, 2/\beta[$.
Set $y^{[0]} \in \mathcal{H}$.
For $k = 0, 1, \ldots$
$\lfloor$ $y^{[k+1]} = y^{[k]} - \gamma_k \nabla f(y^{[k]})$

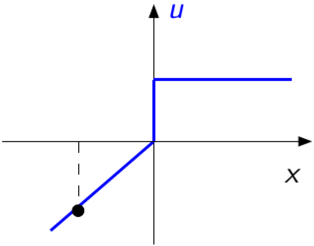
The sequence $(y^{[k]})_{k\in\mathbb{N}}$ converges to $\widehat{y}$.

Subdifferential of a convex function: properties

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\} \end{aligned}$$

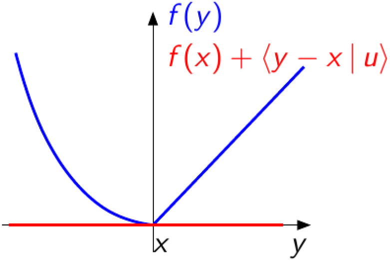


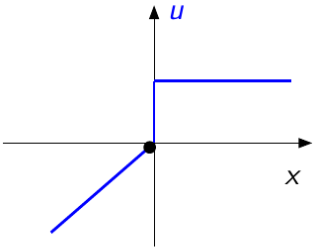


Subdifferential of a convex function: properties

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

$$\begin{aligned}\partial f : \mathcal{H} &\to 2^{\mathcal{H}}\\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \le f(y)\}\end{aligned}$$

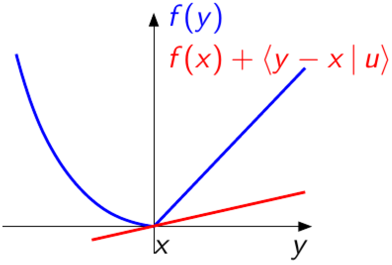


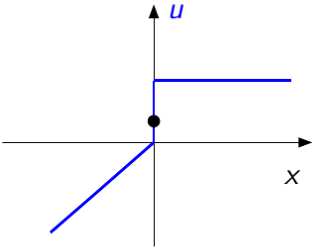


Subdifferential of a convex function: properties

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

$$\partial f : \mathcal{H} \to 2^{\mathcal{H}}$$
$$x \to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \leq f(y)\}$$

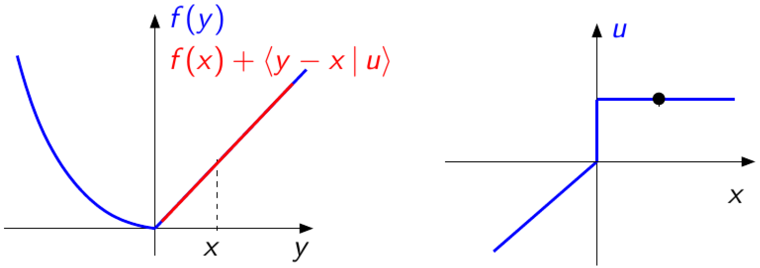




Subdifferential of a convex function: properties

Let $f : \mathcal{H} \to\ ]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

$$\partial f : \mathcal{H} \to 2^{\mathcal{H}}$$
$$x \to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \leq f(y)\}$$



Subdifferential of a convex function: properties

Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper function.
The (Moreau) subdifferential of f, denoted by ∂f, is such that

$$\begin{aligned} \partial f : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\to \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u\rangle + f(x) \leq f(y)\} \end{aligned}$$

- Fermat rule:

$$\begin{aligned} 0 \in \partial f(x) &\Leftrightarrow\ (\forall y \in \mathcal{H})\ \langle y - x \,|\, 0\rangle + f(x) \leq f(y) \\ &\Leftrightarrow x \in \mathrm{Argmin} f \end{aligned}$$

- If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$

Proximity operator

Subgradient descent (Shor, 1979)

Solve $\widehat{y} \in \mathrm{Argmin}_y f(y)$ when $f \in \Gamma_0(\mathcal{H})$ non-smooth.

$$\text{For } k = 0, 1, \ldots$$
$$\lfloor\; y^{[k+1]} = y^{[k]} - \gamma_k t^{[k]} \quad \text{with} \quad t^{[k]} \in \partial f(y^{[k]})$$

where $\partial f(y) = \{t \in \mathcal{H} \,|\, (\forall u \in \mathcal{H}) \;\; f(u) \geq f(y) + \langle t \,|\, u - y\rangle\}$.

Proximity operator

Subgradient descent (Shor, 1979)
Solve $\widehat{y} \in \mathrm{Argmin}_y f(y)$ when $f \in \Gamma_0(\mathcal{H})$ non-smooth.

$$\text{For } k = 0, 1, \ldots \\ \lfloor\; y^{[k+1]} = y^{[k]} - \gamma_k t^{[k]} \quad \text{with} \quad t^{[k]} \in \partial f(y^{[k]})$$

where $\partial f(y) = \{t \in \mathcal{H} \,|\, (\forall u \in \mathcal{H}) \;\; f(u) \geq f(y) + \langle t \,|\, u - y\rangle\}$.
Technical assumptions on γ_n to insure convergence:
$\Rightarrow$ decreasing step-size.

Proximity operator

Proximal point algorithm

Solve $\widehat{y} \in \operatorname{Argmin}_y f(y)$ when $f \in \Gamma_0(\mathcal{H})$ non-smooth

Set $\gamma_k > 0$ such that $\sum_{k=0}^{+\infty} \gamma_k^2 = \infty$.
Set $y^{[0]} \in \mathcal{H}$.
For $k = 0, 1, \ldots$
$\lfloor\ y^{[k+1]} = y^{[k]} - \gamma_k t^{[k]} \quad \text{with} \quad t^{[k]} \in \partial f(y^{[k+1]})$

where $\partial f(y) = \{t \in \mathcal{H} \mid (\forall u \in \mathcal{H})\ \ f(u) \geq f(y) + \langle t \mid u - y \rangle\}$.
The sequence $(y^{[k]})_{k \in \mathbb{N}}$ converges to $\widehat{y}$.

Proximity operator

Proximal point algorithm

Solve $\widehat{y} \in \text{Argmin}_y f(y)$ when $f \in \Gamma_0(\mathcal{H})$ non-smooth

Set $\gamma_k > 0$ such that $\sum_{k=0}^{+\infty} \gamma_k^2 = \infty$.
Set $y^{[0]} \in \mathcal{H}$.
For $k = 0, 1, \ldots$
$\lfloor\ y^{[k+1]} = y^{[k]} - \gamma_k t^{[k]} \quad \text{with} \quad t^{[k]} \in \partial f(y^{[k+1]})$

where $\partial f(y) = \{t \in \mathcal{H} \,|\, (\forall u \in \mathcal{H})\ f(u) \geq f(y) + \langle t \,|\, u - y\rangle\}$.
The sequence $(y^{[k]})_{k\in\mathbb{N}}$ converges to $\widehat{y}$. $\Rightarrow$ no decreasing step-size.

$$\Leftrightarrow \quad (\forall k \in \mathbb{N}) \quad y^{[k]} - y^{[n+1]} \in \gamma_k \partial f(y^{[k+1]})$$

$$\Leftrightarrow \quad (\forall k \in \mathbb{N}) \quad y^{[k+1]} = \text{prox}_{\gamma_k f}(y^{[k]})$$

Proximity operator

Definition (Moreau,1965) Let $f \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of f at point $u \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_f u$ such that

$$(\forall u \in \mathcal{H}) \qquad \mathrm{prox}_f u = \arg\min_{v \in \mathcal{H}} f(v) + \frac{1}{2} \|u - v\|^2$$

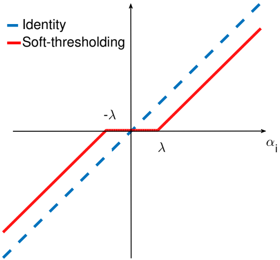
Proximity operator

Definition (Moreau,1965) Let $f \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of f at point $u \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_f u$ such that

$$(\forall u \in \mathcal{H}) \qquad \mathrm{prox}_f u = \arg\min_{v \in \mathcal{H}} f(v) + \frac{1}{2}\|u - v\|^2$$

Examples closed form expression

- $\mathrm{prox}_{\lambda\|\cdot\|_1}$: soft-thresholding with a fixed threshold $\lambda > 0$



Proximity operator

Definition (Moreau,1965) Let $f \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of f at point $u \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_f u$ such that

$$(\forall u \in \mathcal{H}) \qquad \mathrm{prox}_f u = \arg\min_{v \in \mathcal{H}} f(v) + \frac{1}{2}\|u - v\|^2$$

Examples closed form expression

- $\mathrm{prox}_{\lambda\|\cdot\|_1}$: soft-thresholding with a fixed threshold $\lambda > 0$
- $\mathrm{prox}_{\|\cdot\|_{1,2}}$(Peyré,Fadili,2011).
- $\mathrm{prox}_{\|\|_p^p}$ with $p = \{\frac{4}{3}, \frac{3}{2}, 2, 3, 4\}$(Chaux,Combettes,Pesquet,Wajs,2005).
- $\mathrm{prox}_{D_{KL}}$(Combettes,Pesquet,2007).
- $\mathrm{prox}_{\iota_C} = P_C$ projection onto the convex set C.

Proximity operator

Definition (Moreau,1965) Let $f \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of f at point $u \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_f u$ such that

$$(\forall u \in \mathcal{H}) \qquad \mathrm{prox}_f u = \arg\min_{v \in \mathcal{H}} f(v) + \frac{1}{2}\|u - v\|^2$$

Examples closed form expression

- $\mathrm{prox}_{\lambda\|\cdot\|_1}$: soft-thresholding with a fixed threshold $\lambda > 0$
- $\mathrm{prox}_{\|\cdot\|_{1,2}}$(Peyré,Fadili,2011).
- $\mathrm{prox}_{\|\|_p^p}$ with $p = \{\frac{4}{3}, \frac{3}{2}, 2, 3, 4\}$(Chaux,Combettes,Pesquet,Wajs,2005).
- $\mathrm{prox}_{D_{KL}}$(Combettes,Pesquet,2007).
- $\mathrm{prox}_{\iota_C} = P_C$: projection onto the convex set C.
  - $\rightarrow$ range constraint hypercube projection,
  - $\rightarrow$ $\ell_{1,p}$-ball constraint (Quattoni,Carreras,Collins,Darrell,2007) (Van Den Berg, Friedlander,2008)

Proximity operator

Definition (Moreau,1965) Let $f \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of f at point $u \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_f u$ such that

$$(\forall u \in \mathcal{H}) \qquad \mathrm{prox}_f u = \arg\min_{v \in \mathcal{H}} f(v) + \frac{1}{2}\|u - v\|^2$$

Examples closed form expression

- $\mathrm{prox}_{\lambda\|\cdot\|_1}$: soft-thresholding with a fixed threshold $\lambda > 0$
- $\mathrm{prox}_{\|\cdot\|_{1,2}}$(Peyré,Fadili,2011).
- $\mathrm{prox}_{\|\|_p^p}$ with $p = \{\frac{4}{3}, \frac{3}{2}, 2, 3, 4\}$(Chaux,Combettes,Pesquet,Wajs,2005).
- $\mathrm{prox}_{D_{KL}}$(Combettes,Pesquet,2007).
- $\mathrm{prox}_{\iota_C} = P_C$ projection onto the convex set C.
  - $\rightarrow$ range constraint hypercube projection,
  - $\rightarrow$ $\ell_{1,p}$-ball constraint (Quattoni,Carreras,Collins,Darrell,2007) (Van Den Berg, Friedlander,2008)
- $\mathrm{prox}_{\sum_{g\in\mathcal{G}}\|\cdot\|_q}$ with overlapping groups (Jenatton,Mairal,Obozinski,Bach, 2011)

Proximity operator

Definition (Moreau,1965) Let $f \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of f at point $u \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_f u$ such that

$$(\forall u \in \mathcal{H}) \qquad \mathrm{prox}_f u = \arg\min_{v \in \mathcal{H}} f(v) + \frac{1}{2}\|u - v\|^2$$

Examples closed form expression

- $\mathrm{prox}_{\lambda\|\cdot\|_1}$: soft-thresholding with a fixed threshold $\lambda > 0$
- $\mathrm{prox}_{\|\cdot\|_{1,2}}$(Peyré,Fadili,2011).
- $\mathrm{prox}_{\|\|_p^p}$ with $p = \{\frac{4}{3}, \frac{3}{2}, 2, 3, 4\}$(Chaux,Combettes,Pesquet,Wajs,2005).
- $\mathrm{prox}_{D_{KL}}$(Combettes,Pesquet,2007).
- $\mathrm{prox}_{\iota_C} = P_C$: projection onto the convex set C.
  - $\to$ range constraint hypercube projection,
  - $\to$ $\ell_{1,p}$-ball constraint (Quattoni,Carreras,Collins,Darrell,2007) (Van Den Berg, Friedlander,2008)
- $\mathrm{prox}_{\sum_{g \in \mathcal{G}} \|\cdot\|_q}$ with overlapping groups (Jenatton,Mairal,Obozinski,Bach, 2011)
- Composition with a linear operator: $\mathrm{prox}_{\varphi \circ L}$ closed form if $LL^* = \nu \mathrm{Id}$

Proximity operator

Definition (Moreau,1965) Let $f \in \Gamma_0(\mathcal{H})$ where $\mathcal{H}$ denotes a real Hilbert space. The proximity operator of f at point $u \in \mathcal{H}$ is the unique point denoted by $\mathrm{prox}_f u$ such that

$$(\forall u \in \mathcal{H}) \qquad \mathrm{prox}_f u = \arg\min_{v \in \mathcal{H}} f(v) + \frac{1}{2}\|u - v\|^2$$

Examples Proximity operator of a sum of two functions :

$$\mathrm{prox}_{f_1+f_2} = \mathrm{prox}_{f1} \circ \mathrm{prox}_{f2}?$$

- (Combettes-Pesquet, 2007) $N = 1$, $f_2 = \iota_C$ of a non-empty closed convex subset of C and f_1 is dierentiable at 0 with $h'(0) = 0$.
- (Chaux-Pesquet-Pustelnik,2009) C and f_2 are separable in the same basis.
- (Yu, 2013)(Shi et al., 2017) $\partial f_2(x) \subset \partial f_2(\mathrm{prox} f_1(x))$.
- Many recent results (Pustelnik, Condat, 2017)(Yukawa, Kagami, 2017)(del Aguila Pla, Jaldén, 2017)

Proximal algorithm: Forward-backward

Forward-backward algorithm

Solve $\widehat{y} \in \mathrm{Argmin}_y f_1(y) + f_2(y)$ with f_1 and f_2 in $\Gamma_0(\mathcal{H})$

Let $y^{[0]} \in \mathcal{H}$.
For $k = 0, 1, \dots$

$$\lfloor \; y^{[k+1]} = y^{[k]} + \tau_k \left(\mathrm{prox}_{\gamma_k f_1} (y^{[k]} - \gamma_k \nabla f_2 (y^{[k]})) - y^{[k]} \right)$$

Proximal algorithm: Forward-backward

Forward-backward algorithm

Solve $\widehat{y} \in \mathrm{Argmin}_y f_1(y) + f_2(y)$ with f_1 and f_2 in $\Gamma_0(\mathcal{H})$

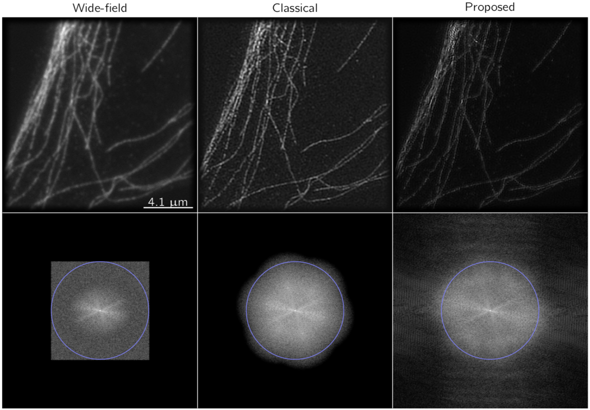
Let $y^{[0]} \in \mathcal{H}$.
For $k = 0, 1, \ldots$
$\lfloor\ y^{[k+1]} = y^{[k]} + \tau_k \left(\mathrm{prox}_{\gamma_k f_1} (y^{[k]} - \gamma_k \nabla f_2 (y^{[k]})) - y^{[k]} \right)$

Convergence (Combettes,Wajs,2005)

- f_2 is β-Lipschitz differentiable on $\mathcal{H}$ with $\beta > 0$
- $\gamma_k \in]0, 2/\beta[$: algorithm step-size
- $\tau_k \in]0, 1]$: relaxation parameter

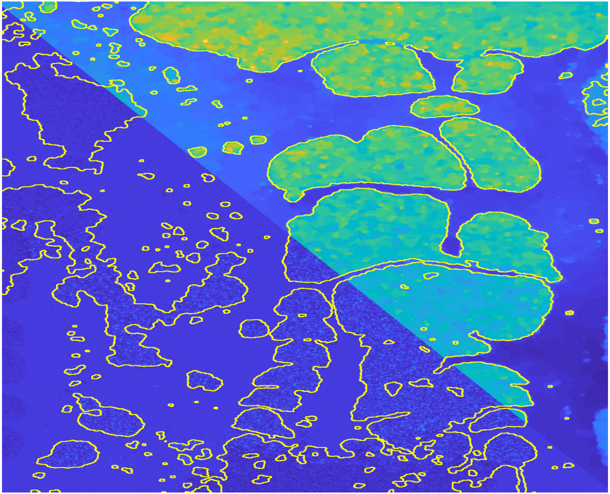
Under these assumptions, $(y^{[k]})_{k \in \mathbb{N}}$ converges to $\widehat{y}$.

Example: Structured illumination microscopy



*J. Boulanger, N. Pustelnik, L. Condat, T. Piolot, L. Sengmanivong, Nonsmooth convex optimization for Structured Illumination Microscopy image reconstruction, Inverse problems, vol. 34, no. 9, 22pp., July 2018.

Example: Multiphasic flow



*Collaboration LPENSL.

Example: Matrix completion

$$\underset{x}{\text{minimize}} \|x\|_* \quad \text{s.t.} \quad Hx = z$$

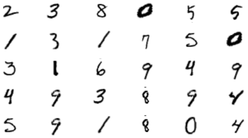
Example: Robust PCA

$$\underset{u,v}{\text{minimize}} \|u\|_* + \|v\|_1 \quad \text{s.t.} \quad z = u + v$$

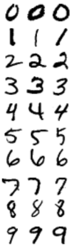
[From Goldfarb, Ma, Sheinberg, 2010]

Learning

Images to classify



Training set



Classification

- Training set of size L for K classes:

$$\mathcal{S} = \left\{ (u_\ell, z_\ell) \in \mathbb{R}^N \times \{1, \ldots, K\} \;\middle|\; \ell \in \{1, \ldots, L\} \right\}$$

examples: $u_\ell =$  and $z_\ell = 2$

$u_\ell =$  and $z_\ell = 9$

Learning: multiclass SVM

- $\phi(u)\colon \mathbb{R}^N \to \mathbb{R}^M$: mapping from the input space onto an arbitrary feature space with $M > N$

  $\Rightarrow$ linearization

  examples: convolution networks [Mirowski et al., 2008]

  scattering coefficients [Brunat,Mallat,2013]

- The predictor relies on K different discriminating functions $D_k\colon \mathbb{R}^N \to \mathbb{R}$:

$$D_k(u) = \phi(u)^\top x^{(k)} + b^{(k)}$$

- The predictor selects the class that best matches an observation

$$d(u) = \arg \max_{1 \le k \le K} D_k(u)$$

Learning: multiclass SVM

Objective of the learning stage: estimate x to correctly predict the input-output pair $(u_\ell, z_\ell) \in \mathcal{S}$ for every $\ell \in \{1, \ldots, L\}$,

$$z_\ell = \underset{1\leq k\leq K}{\arg\max}\ \varphi(u_\ell)^\top \mathrm{x}^{(k)}$$

$$\Leftrightarrow \quad \max_{k\neq z_\ell} \varphi(u_\ell)^\top(\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) < 0$$

[relax the strict ineqality with $\mu_\ell > 0$] $\Leftrightarrow \ \max_{k\neq z_\ell} \varphi(u_\ell)^\top(\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) \leq -\mu_\ell$

[deal with unfeasible constraints $\zeta^{(\ell)} \geq 0$] $\Leftrightarrow \ \max_{k\neq z_\ell} \varphi(u_\ell)^\top(\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) \leq \zeta^{(\ell)} - \mu_\ell$

$$\underset{(\mathrm{x},\xi)\in\mathbb{R}^{(M+1)K}\times\mathbb{R}^L}{\text{minimize}} \ \sum_{k=1}^{K} \|x^{(k)}\|_2^2 + \lambda \sum_{\ell=1}^{L} \xi^{(\ell)} \quad \text{subj. to}$$

$$\begin{cases} (\forall \ell \in \{1,\ldots,L\}) & \max_{k\neq z_\ell} \ \varphi(u_\ell)^\top(\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) \leq \xi^{(\ell)} - \mu_\ell \\ (\forall \ell \in \{1,\ldots,L\}) & \xi^{(\ell)} \geq 0, \end{cases}$$

Sparsity in learning

$$\underset{(\mathrm{x},\xi)\in\mathbb{R}^{(M+1)K}\times\mathbb{R}^L}{\text{minimize}} \sum_{k=1}^{K} \|x^{(k)}\|_1 + \lambda \sum_{\ell=1}^{L} \xi^{(\ell)} \quad \text{subj. to}$$

$$\begin{cases}(\forall \ell \in \{1,\dots,L\}) & \max_{k\neq z_\ell} \varphi(u_\ell)^\top(\mathrm{x}^{(k)} - \mathrm{x}^{(z_\ell)}) \leq \xi^{(\ell)} - \mu_\ell \\ (\forall \ell \in \{1,\dots,L\}) & \xi^{(\ell)} \geq 0,\end{cases}$$

- cf. work by F. Bach and reference therein.
- Possibility to learn quadratic interactions.

Image reconstruction with CNN

- Inverse problems : Tikhonov penalization

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\operatorname{Argmin}} \|Hx - z\|^2 + \lambda\|\Gamma x\|_2^2$$
$$\Leftrightarrow \widehat{x} = (H^*H + \lambda\Gamma^*\Gamma)^{-1}H^*z = Gz.$$

- Reformulation into a convolutional network using the kernel separability theorem relying on the existence of the decomposition $G = USV^\top$:

$$\widehat{x} = \sum_j s_j U_{j,\bullet}(V_{j,\bullet}^\top z).$$

where s_j denotes the j-th singular value, and $U_{j,\bullet}$ (resp. $V_{j,\bullet}$) denotes the j-th column of U (resp. V).
- 2D deconvolution can be reformulated as a weighted sum of separable 1D filters.
- $\widehat{x}$ can be well approximated by a small number of separable filters by dropping out kernel associated with very small s_j.

Image reconstruction with CNN: agnostic

- Image Deconvolution Convolutional Neural Networks (DCNN) [Xu et al, 2014] :

$$\begin{aligned}\widehat{x} &= f(z)\\ &= W_3\sigma(W_2\sigma(W_1 z + b_1) + b_2.\end{aligned}$$

  - W_3 denotes weights playing the same role than S,
  - W_2 and W_1: separable kernels acting horizontally or vertically,
  - σ denotes a nonlinear function.

- Goal: estimate $(W_i)_{i=1,2,3}$ and $(b_i)_{i=1,2}$ in order to minimize

$$\frac{1}{2|N|}\sum_{i\in N}\|f(z_\ell) - \overline{x}_\ell\|.$$

using training image pairs $\{\overline{x}_\ell, z_\ell\}_{\ell\in N}$.

Image reconstruction combining CNN and regularization techniques

Neumann Networks for Linear Inverse Problems in Imaging

Davis Gilton, Greg Ongie, Rebecca Willett*

June 5, 2019

Abstract

Many challenging image processing tasks can be described by an ill-posed linear inverse problem: deblurring, deconvolution, inpainting, compressed sensing, and superresolution all lie in this framework. Traditional inverse problem solvers minimize a cost function consisting of a data-fit term, which measures how well an image matches the observations, and a regularizer, which reflects prior knowledge and promotes images with desirable properties like smoothness. Recent advances in machine learning and image processing have illustrated that it is often possible to *learn* a regularizer from training data that can outperform more traditional regularizers. We present an end-to-end, data-driven method of solving inverse problems inspired by the Neumann series, which we call a Neumann network. Rather than unroll an iterative optimization algorithm, we truncate a Neumann series which directly solves the linear inverse problem with a data-driven nonlinear regularizer. The Neumann network architecture outperforms traditional inverse problem solution methods, model-free deep learning approaches, and state-of-the-art unrolled iterative methods on standard datasets. Finally, when the images belong to a union of subspaces and under appropriate assumptions on the forward model, we prove there exists a Neumann network configuration that well-approximates the optimal oracle estimator for the inverse problem and demonstrate empirically that the trained Neumann network has the form predicted by theory.

Image reconstruction combining CNN and regularization techniques

- Inverse problems : Tikhonov penalization

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\mathrm{Argmin}} \|Hx - z\|^2 + x^\top Rx \Leftrightarrow \widehat{x} = (H^*H + R)^{-1}H^*z$$

- Applying Neumann series expansion

  $+$ truncating the series

  $+$ $R = R_\theta$:

$$\widehat{x}(z,\theta) = \sum_{j=0}^{J} (I - \eta H^*H - \eta R_\theta)^j \eta H^* z$$

- Training from the dataset $(\overline{x}_\ell, z_\ell) \in \mathcal{S}$

$$\min_\theta \sum_{\ell=1}^{L} \|\widehat{x}(z_\ell,\theta) - \overline{x}_\ell\|_2^2$$

Image reconstruction combining CNN and regularization techniques

LEARNING THE INVISIBLE: A HYBRID DEEP LEARNING-SHEARLET FRAMEWORK FOR LIMITED ANGLE COMPUTED TOMOGRAPHY

TATIANA A. BUBBA, GITTA KUTYNIOK, MATTI LASSAS, MAXIMILIAN MÄRZ, WOJCIECH SAMEK, SAMULI SILTANEN, AND VIGNESH SRINIVASAN

ABSTRACT. The high complexity of various inverse problems poses a significant challenge to model-based reconstruction schemes, which in such situations often reach their limits. At the same time, we witness an exceptional success of data-based methodologies such as deep learning. However, in the context of inverse problems, deep neural networks mostly act as black box routines, used for instance for a somewhat unspecified removal of artifacts in classical image reconstructions. In this paper, we will focus on the severely ill-posed inverse problem of limited angle computed tomography, in which entire boundary sections are not captured in the measurements. We will develop a hybrid reconstruction framework that fuses model-based sparse regularization with data-driven deep learning. Our method is *reliable* in the sense that we only learn the part that can provably not be handled by model-based methods, while applying the theoretically controllable sparse regularization technique to the remaining parts. Such a decomposition into *visible* and *invisible* segments is achieved by means of the shearlet transform that allows to resolve wavefront sets in the phase space. Furthermore, this split enables us to assign the clear task of inferring unknown shearlet coefficients to the neural network and thereby offering an *interpretation* of its performance in the context of limited angle computed tomography. Our numerical experiments show that our algorithm significantly surpasses both pure model- and more data-based reconstruction methods.

[cs.CV] 12 Nov 2018

Image reconstruction combining CNN and regularization techniques

- Sparse regularization:

$$\widehat{x} \in \underset{x \in \mathbb{R}^N}{\mathrm{Argmin}} \|Hx - z\|^2 + \lambda \|Fx\|_1$$

- Train CNN to estimate the " invisible" from the visible:

$$\omega = \mathcal{N}\mathcal{N}_\theta(F\widehat{x})$$

- Combine the visible and the learned invisible coefficients:

$$\widehat{\widehat{x}} = F^*((F\widehat{x})_{\text{vis}} + \omega_{\text{inv}})$$

Image reconstruction combining CNN and regularization techniques

Method	RE	PSNR	SSIM	HaarPSI
$\boldsymbol{f}_{\texttt{FBP}}$	0.84	17.16	0.12	0.18
$\boldsymbol{f}^*$	0.22	28.76	0.94	0.47
$\boldsymbol{f}_{\texttt{TV}}$	0.21	29.54	0.95	0.54
$\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\texttt{FBP}})$	0.19	30.20	0.54	0.75
$\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\texttt{FBP}}))$	0.18	30.52	0.78	0.72
$\boldsymbol{f}_{\texttt{LtI}}$	**0.09**	**36.96**	**0.96**	**0.86**

Stability

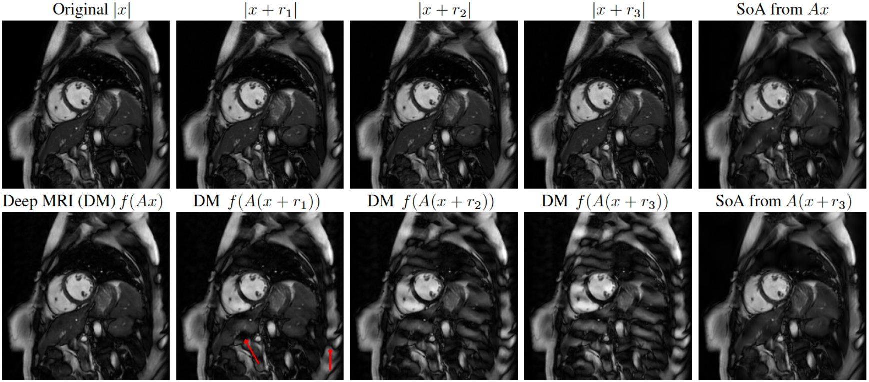
On instabilities of deep learning in image reconstruction - Does AI come at a cost?

Vegard Antun[1] Francesco Renna[2] Clarice Poon[3] Ben Adcock[4] Anders C. Hansen*[5,1]

Deep learning, due to its unprecedented success in tasks such as image classification, has emerged as a new tool in image reconstruction with potential to change the field. In this paper we demonstrate a crucial phenomenon: deep learning typically yields unstable methods for image reconstruction. The instabilities usually occur in several forms: (1) tiny, almost undetectable perturbations, both in the image and sampling domain, may result in severe artefacts in the reconstruction, (2) a small structural change, for example a tumour, may not be captured in the reconstructed image and (3) (a counterintuitive type of instability) more samples may yield poorer performance. Our new stability test with algorithms and easy to use software detects the instability phenomena. The test is aimed at researchers to test their networks for instabilities and for government agencies, such as the Food and Drug Administration (FDA), to secure safe use of deep learning methods.

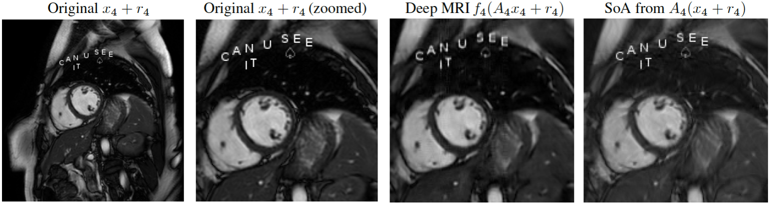
Stability

- tiny perturbation: incorrect representation.



Stability

- small structural changes: stable.



Conclusions and perspectives

- Deep learning allows us to obtain very good results for denoising task but general inverse problem still not solved.

  .

- From regularized methods to deep learning: model design + optimization.

- Design an objective function compatible with algorithmic strategies.

- Stronger guarantees in the non-convex setting (Mumford-Shah, deep learning).