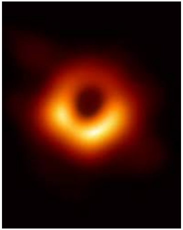
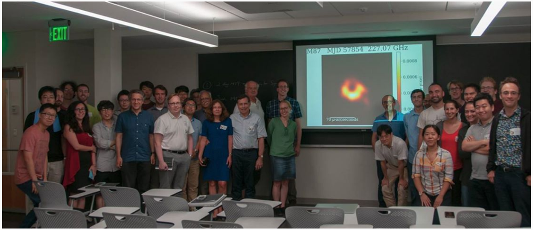
Optimization
– Inverse problems –

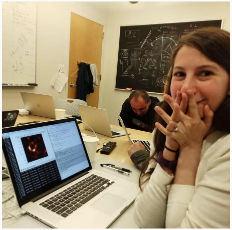
Nelly Pustelnik

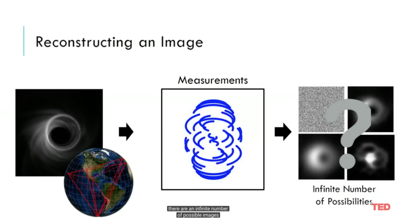
CNRS, Laboratoire de Physique de l'ENS de Lyon, France

Image analysis: serving other sciences



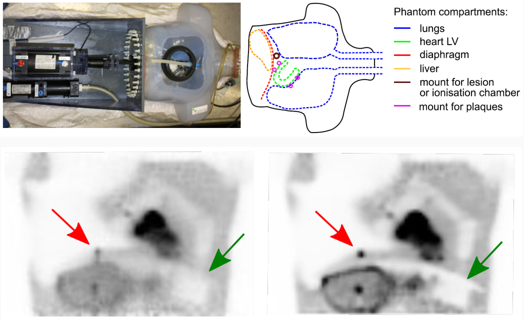






Black hole, Galaxy M87, Event Horizon Telescope (EHT)

Image analysis: serving other sciences



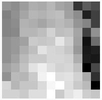
(source : F. Jolivet)

H2020 Nexis Project

Forward model

Notations

- Image $\mathrm{x} \in \mathbb{R}^{N_1 \times N_2}$



$$\mathrm{x} = \left(\mathrm{x}_{n_1,n_2}\right)_{1 \leq n_1 \leq N_1, 1 \leq n_2 \leq N_2}$$

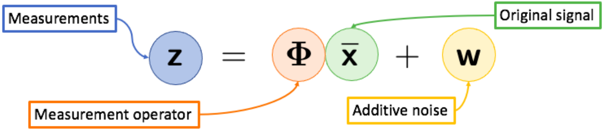
- Vector consisting of the values of the image of size $N = N_1 \times N_2$ arranged column-wise $\mathrm{x} \in \mathbb{R}^N$

  (with $N = N_1 \times N_2$)

$$\mathrm{x} = \left(\mathrm{x}_n\right)_{1 \leq n \leq N}$$

Direct model

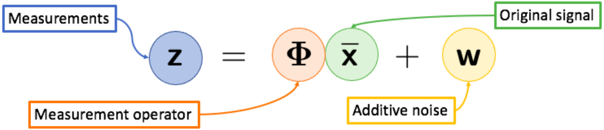
Observation model:



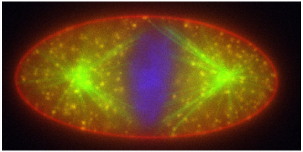
Objective: Find an estimate $\widehat{\mathrm{x}} \in \mathbb{R}^N$ of $\overline{\mathrm{x}}$ from $\mathrm{z} \in \mathbb{R}^M$.

Direct model

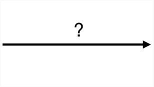
Observation model:

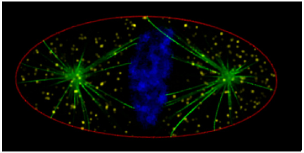


Objective: Find an estimate $\widehat{x} \in \mathbb{R}^N$ of $\overline{x}$ from $z \in \mathbb{R}^M$.



Degraded image z





Original image $\overline{x}$

Direct model: convolution

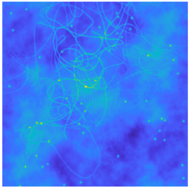
$$\mathrm{z} = \Phi\overline{\mathrm{X}} \quad \Leftrightarrow \quad \mathrm{z} = \phi * \overline{\mathrm{X}}$$

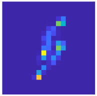
- $\{\phi * \overline{\mathrm{x}}\}$: convolution product with the Point Spread Function (PSF) ϕ of size $Q_1 \times Q_2$.
- Φ is a block-circulant matrix with circulant blocks related to ϕ, then $\Phi = \mathbf{F}^* \Lambda \mathbf{F}$ where
  - Λ: diagonal matrix,
  - $\mathbf{F}$: represents the discrete Fourier transform where $\cdot^*$ denotes here the transpose conjugate and $\mathbf{F}^* = \mathbf{F}^{-1}$.

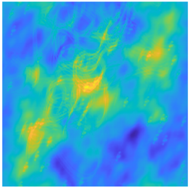
- Efficient computation of $\Phi\overline{\mathrm{x}}$ by means of its Fourier transform of $\overline{\mathrm{X}}$:

$$\begin{aligned} \Phi\overline{\mathrm{x}} &= \mathbf{F}^* \Lambda \mathbf{F} \overline{\mathrm{x}} \\ &= \mathbf{F}^* \Lambda \overline{\mathrm{X}}. \end{aligned}$$

Direct model: convolution

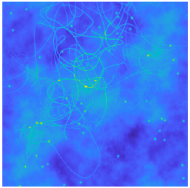




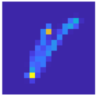


$$\bar{\mathrm{x}} * \phi = \mathrm{z}$$

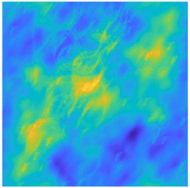
Direct model: convolution



$*$

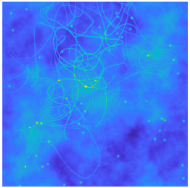


$=$



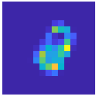
$\bar{x}$ ϕ z

Direct model: convolution



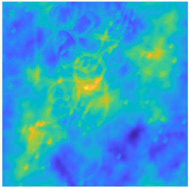
$\bar{x}$

$*$



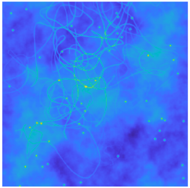
ϕ

$=$

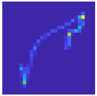


z

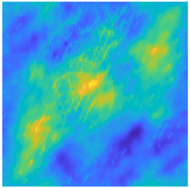
Direct model: convolution



$*$

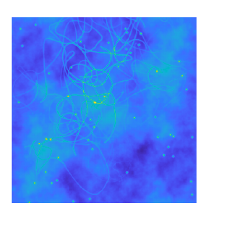


$=$

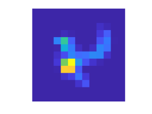


$\overline{\mathrm{x}}$ ϕ z

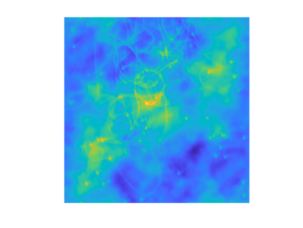
Direct model: convolution



$*$

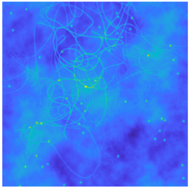


$=$

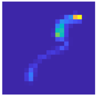


$\bar{\mathrm{x}}$ ϕ z

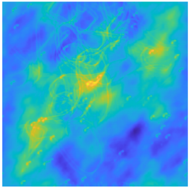
Direct model: convolution



$*$

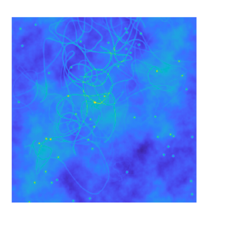


$=$

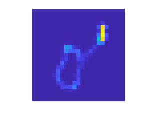


$\bar{\mathrm{x}}$ ϕ z

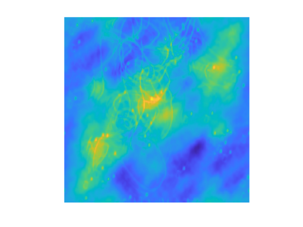
Direct model: convolution



$*$

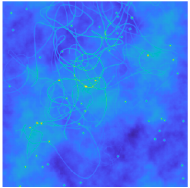
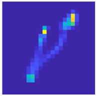
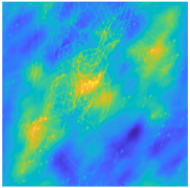


$=$



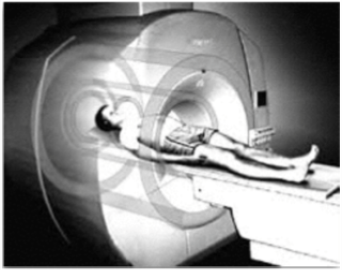
$\bar{\mathrm{x}}$ ϕ z

Direct model: convolution

 $*$  $=$ 

$\bar{\mathrm{x}}$ ϕ z

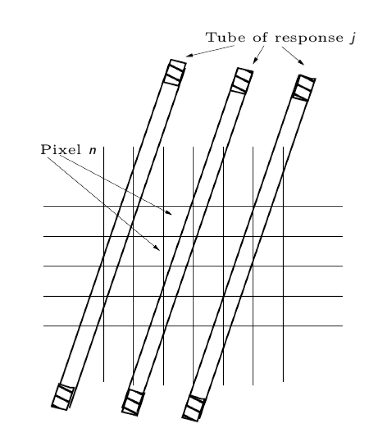
Direct model in medecine: MRI



$$z \simeq \Phi \overline{x}$$

- $\overline{x} \in \mathbb{R}^N$: vectorized original (unknown) image.
- $\Phi = \mathbf{F} \colon \mathbb{R}^N \to^M$: measurement operator selecting (mask : $^N \to^M$) Fourier coefficients (2D Fourier transform $\mathbf{F} \colon \mathbb{R}^N \to^N$).
- z: vector containing the observed values (undersampled Fourier coefficients).

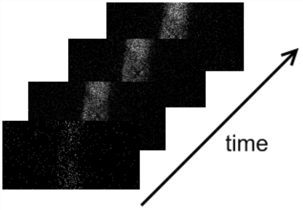
Direct model in medecine: Tomography



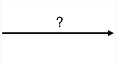
$$\mathrm{z} \simeq \Phi \overline{\mathrm{x}}$$

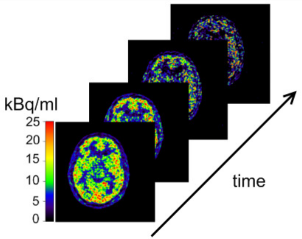
- $\overline{\mathrm{x}} = (\overline{\mathrm{x}}_n)_{1 \le n \le N} \in \mathbb{R}^N$: vector consisting of the (unknown) values of the original image of size $N = N_1 \times N_2$.
- $\Phi = (\Phi_{m,n})_{1 \le m \le M, 1 \le n \le N}$: probability to detect an event in the tube/line of response.
- $\mathrm{z} = (\mathrm{z}_m)_{1 \le m \le M} \in \mathbb{R}^M$: vector containing the observed values (sinogram).

Application examples in medecine: Tomography



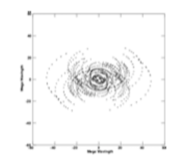
Degraded images





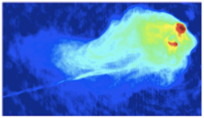
Reconstructed images

Direct model in Astronomy: Radio-interferometry



Fourier sampling

Very Large Array, New Mexico

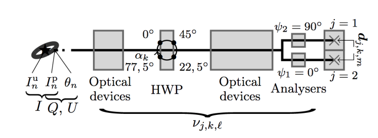


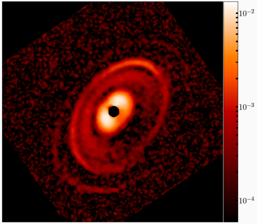
Cygnus A

$$z \simeq \Phi \overline{x}$$

- $\overline{x} \in \mathbb{R}^N$: vectorized original (unknown) 2D image.
- $\Phi = \mathbf{F} : \mathbb{R}^N \to^M$: measurement operator selecting Fourier coefficients.
    - $\mathbf{F} : \mathbb{R}^N \to^{\overline{N}}$: 2D Fourier transform (with zero-padding),
    - $\in^{M \times \overline{N}}$: (de)-gridding matrix modelling non-uniform (undersampled) Fourier transform, and direction (in)dependent effect (calibration artefacts).
- z: vector containing the observed values (undersampled Fourier coefficients)

Application examples in astronomy: High-contrast imagery





RXJ 1615 (Avenhaus et al. 2018)

$$\mathrm{z}_{j,k} \simeq \sum_{m=1}^{3} \nu_{j,k,m}\, \mathrm{T}_{j,k} \mathrm{A}\, \mathrm{S}_m$$

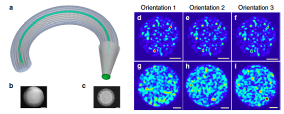
where

- $\mathrm{A} : \mathbb{R}^N \to \mathbb{R}^N$: invariant blur.
- $\mathrm{T}_{j,k} : \mathbb{R}^N \to \mathbb{R}^j$: geometric transform of the j-th polariser during the k-th acquisition.
- Stokes versus Jones formalisms:

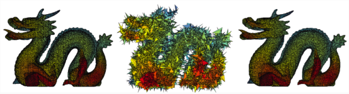
$$\mathrm{I}_{j,k}^{\mathrm{det}} = \frac{1}{2}\mathrm{I}_{\mathrm{u}} + \mathrm{I}_{\mathrm{p}} \cos^2(\theta - 2\alpha_k - \psi_j)$$

$$\Leftrightarrow \quad \mathrm{I}_{j,k}^{\mathrm{det}} = \nu_{j,k,1} + \nu_{j,k,2} + \nu_{j,k,3}$$

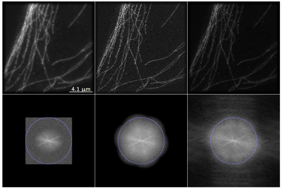
Many others!



Photon imaging



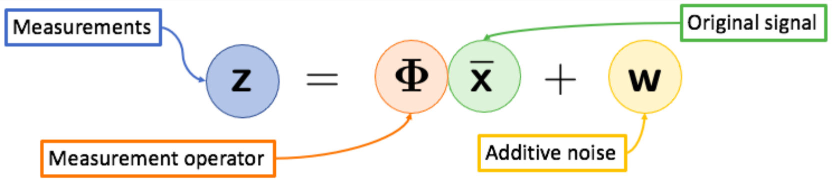
3D Mesh denoising



Structured Illumination Microscopy

Computer vision

Direct model



+ Φ is often known or can be approximated.
+ Φ is a sparse matrix.
− Φ is generally ill-conditioned.

Hadamard conditions (1902)

The problem $\mathrm{z} = \Phi\overline{\mathrm{x}}$ is said to be well-posed if it fulfills the **Hadamard conditions**:

1. **existence of a solution,**

   i.e. the range $\operatorname{ran}\Phi$ of Φ is equal to $\mathbb{R}^M$,
2. **uniqueness of the solution,**

   i.e. the nullspace $\ker\Phi$ of Φ is equal to $\{0\}$,
3. **stability of the solution $\widehat{\mathrm{x}}$ relatively to the observation,**

   i.e. $\big(\forall(\mathrm{z},\mathrm{z}') \in \big(\mathbb{R}^M\big)^2\big)$

$$\|\mathrm{z}-\mathrm{z}'\| \to 0 \quad \Rightarrow \quad \|\widehat{\mathrm{x}}(\mathrm{z}) - \widehat{\mathrm{x}}(\mathrm{z}')\| \to 0.$$

Hadamard conditions (1902)

The problem $\mathrm{z} = \Phi\overline{\mathrm{x}}$ is said to be well-posed if it fulfills the **Hadamard conditions**:

1. **existence of a solution**,

   i.e. every vector z in $\mathbb{R}^M$ is the image of a vector x in $\mathbb{R}^N$,
2. **uniqueness of the solution**,

   i.e. if $\widehat{\mathrm{x}}(\mathrm{z})$ and $\widehat{\mathrm{x}}'(\mathrm{z})$ are two solutions, then they are necessarily equal since $\widehat{\mathrm{x}}(\mathrm{z}) - \widehat{\mathrm{x}}'(\mathrm{z})$ belongs to $\ker \Phi$,
3. **stability of the solution** $\widehat{\mathrm{x}}$ relatively to the observation,

   i.e. ensure that a small perturbation of the observed image leads to a slight variation of the recovered image.

Inverse problem solving

Inversion

- **Inverse filtering** (if $M = N$ and Φ est inversible)

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

Remark :

$\rightarrow$ Closed form expression but noise amplification if Φ ill-conditioned (*ill-posed problem*).

Inversion

- **Inverse filtering** (if $M \geq N$ and rank of Φ is N)

$$\begin{aligned}\widehat{\mathrm{x}} &= (\Phi^*\Phi)^{-1}\Phi^\top \mathrm{z} \\ &= (\Phi^*\Phi)^{-1}\Phi^*(\Phi\overline{\mathrm{x}} + \mathrm{w}) \quad \text{if additive noise } \mathrm{w} \in \mathbb{R}^M \\ &= \overline{\mathrm{x}} + (\Phi^*\Phi)^{-1}\Phi^*\mathrm{w}\end{aligned}$$

Remark :

$\rightarrow$ Closed form expression but noise amplification if Φ ill-conditioned (*ill-posed problem*).

Regularization

- **Variational approach**: Restore the degraded image z i.e., find $\widehat{\mathrm{x}}$ close to $\bar{\mathrm{x}}$:

$$\widehat{\mathrm{x}} \in \operatorname*{Argmin}_{\mathrm{x} \in \mathbb{R}^N} \underbrace{\frac{1}{2}\|\Phi \mathrm{x} - \mathrm{z}\|_2^2}_{\text{Data-term}} + \quad \lambda \quad \underbrace{R(\mathrm{x})}_{\text{Penalization}}$$

Remarks

$\rightarrow$ λ: regularization parameter.

$\rightarrow$ If $\lambda = 0$: inverse filtering.

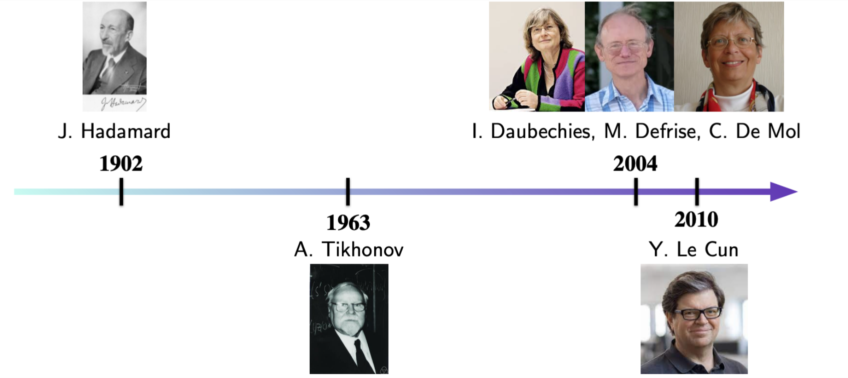
Regularization

- **Variational approach**: Restore the degraded image z i.e., find $\widehat{x}$ close to $\bar{x}$:

$$\widehat{x} \in \operatorname*{Argmin}_{x\in\mathbb{R}^N} \underbrace{\frac{1}{2}\|\Phi x - z\|_2^2}_{\text{Data-term}} + \lambda \underbrace{R(x)}_{\text{Penalization}}$$

(a) Degraded Uniform blur 9×9 Gaussian noise (b) Inverse filtering Quadratic regularisation (c) $\Lambda = \mathrm{Id}$ (d) Λ Laplacian (e) Total variation

Inverse problems: a brief story



Data-fidelity

Maximum A Posteriori (MAP)

Maximum A Posteriori (MAP)

Let x and z be random vector realizations X and Z.

$$\widehat{\mathrm{x}} \in \operatorname*{Argmax}_{\mathrm{x}\in\mathbb{R}^N} \underbrace{\mu_{X|Z=\mathrm{z}}(\mathrm{x})}_{\text{Posterior distribution}}$$

Maximum A Posteriori (MAP)

Maximum A Posteriori (MAP)

Let x and z be random vector realizations X and Z.

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}\in\mathbb{R}^N}{\operatorname{Argmax}} \quad \underbrace{\mu_{X|Z=\mathrm{z}}(\mathrm{x})}_{\text{Posterior distribution}}$$

Bayes rule:

$$\max_{\mathrm{x}\in\mathbb{R}^N} \mu_{X|Z=z}(\mathrm{x}) \Leftrightarrow \max_{\mathrm{x}\in\mathbb{R}^N} \mu_{Z|X=\mathrm{x}}(\mathrm{z}) \cdot \mu_X(\mathrm{x})$$

$$\Leftrightarrow \min_{\mathrm{x}\in\mathbb{R}^N} \Big\{ -\log(\mu_{Z|X=\mathrm{x}}(\mathrm{z})) - \log(\mu_X(\mathrm{x})) \Big\}$$

Maximum A Posteriori (MAP)

Maximum A Posteriori (MAP)

Let x and z be random vector realizations X and Z.

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}\in\mathbb{R}^N}{\operatorname{Argmax}} \quad \underbrace{\mu_{X|Z=\mathrm{z}}(\mathrm{x})}_{\text{Posterior distribution}}$$

Bayes rule:

$$\max_{\mathrm{x}\in\mathbb{R}^N} \mu_{X|Z=z}(\mathrm{x}) \Leftrightarrow \max_{\mathrm{x}\in\mathbb{R}^N} \mu_{Z|X=\mathrm{x}}(\mathrm{z}) \cdot \mu_X(\mathrm{x})$$

$$\Leftrightarrow \min_{\mathrm{x}\in\mathbb{R}^N} \Big\{ \underbrace{-\log(\mu_{Z|X=\mathrm{x}}(\mathrm{z}))}_{\text{Data-term}} \underbrace{-\log(\mu_X(\mathrm{x}))}_{\text{A priori}} \Big\}$$

$$\Leftrightarrow \min_{\mathrm{x}\in\mathbb{R}^N} L(\mathrm{x}) + R(\mathrm{x})$$

Data-term: Gaussian noise

$$(\forall \mathrm{x} \in \mathbb{R}^N) \quad L(\mathrm{x}) = -\log(\mu_{Z|X=\mathrm{x}}(\mathrm{z}))$$

- Let $\mathrm{z} = \Phi\overline{\mathrm{x}} + \mathrm{w}$ with $\mathrm{w} \sim \mathcal{N}(0, \alpha)$

- Gaussian likelihood:

$$\mu_{Z|X=\mathrm{x}}(\mathrm{z}) = \prod_{n=1}^{M} \frac{1}{\sqrt{2\pi\alpha}} \exp\left(\frac{((\Phi\mathrm{x})_n - \mathrm{z}_n)^2}{2\alpha}\right)$$

- Data-term:

$$L(\mathrm{x}) = \sum_{n=1}^{M} \frac{1}{2\alpha}((\Phi\mathrm{x})_n - \mathrm{z}_n)^2$$

Data-term: Poisson noise

$$(\forall \mathrm{x} \in \mathbb{R}^N) \quad L(\mathrm{x}) = -\log(\mu_{Z|X=\mathrm{x}}(\mathrm{z}))$$

- Let $\mathrm{z} = \mathcal{D}_\alpha(\Phi\overline{\mathrm{x}})$ where $\mathcal{D}_\alpha$ Poisson noise with parameter α.

- Poisson likelihood:

$$\mu_{Z|X=\mathrm{x}}(\mathrm{z}) = \prod_{n=1}^{M} \frac{\exp\big(-\alpha(\Phi\mathrm{x})_n\big)}{z_n!} \big(\alpha(\Phi\mathrm{x})_n\big)^{\mathrm{z}_n}$$

- Data-term: $L(\mathrm{x}) = \sum_{n=1}^{M} \Psi_i\big((\Phi\mathrm{x})_n\big)$

$$(\forall \upsilon \in \mathbb{R}) \quad \Psi_i(\upsilon) = \begin{cases} \alpha\upsilon - \mathrm{z}_n \ln(\alpha\upsilon) & \text{if } \mathrm{z}_n > 0 \text{ and } \upsilon > 0, \\ \alpha\upsilon & \text{si } \mathrm{z}_n = 0 \text{ and } \upsilon \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Blind deconvolution

$$(\widehat{x}, \widehat{\Phi}) \in \underset{x,\Phi}{\text{Argmin}} \frac{1}{2}\|\Phi x - z\|_2^2 + \lambda_1 R_1(x) + \lambda_2 R_2(\Phi)$$

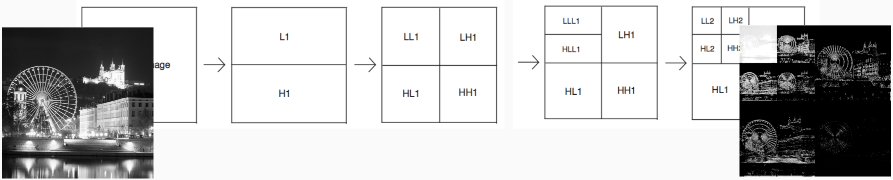
- **Typical choice for R_2:**
  - Φ sensitivity map: piecewise constant.
  - Φ associated with a blur:
    - sparsity,
    - nonnegativity,
    - bounds on vertical/horizontal variations of the blur.

- Examples in Part III.

Regularization

Wavelet denoising: $\mathbf{z} = \overline{\mathbf{x}} + \mathrm{w}$ with $\mathrm{w} = \mathcal{N}(0, \alpha \mathrm{Id})$

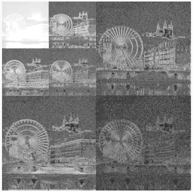
- Wavelets: sparse representation of most natural signals.
- Filterbank implementation of a dyadic wavelet transform: $\Psi \in \mathbb{R}^{N \times N}$.



$\mathbf{z} \in \mathbb{R}^N$ $\qquad$ $\boldsymbol{\zeta} = \Psi \mathbf{z}$

Wavelet denoising: $\mathbf{z} = \overline{\mathbf{x}} + \mathrm{w}$ with $\mathrm{w} = \mathcal{N}(0, \alpha \mathrm{Id})$

$\mathbf{z}$



$\boldsymbol{\zeta} = \Psi \mathbf{z}$

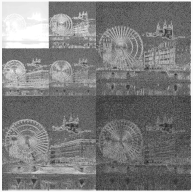
$\mathrm{soft}_\lambda(\Psi \mathbf{z})$

$\widehat{\mathrm{x}} = \Psi^* \mathrm{soft}_\lambda(\Psi \mathbf{z})$

$$
\begin{aligned}
\mathrm{soft}_\lambda(\boldsymbol{\zeta}) &= \big(\max\{|\zeta_{\underline{i}}| - \lambda, 0\} \mathrm{sign}(\zeta_{\underline{i}}) \big)_{\underline{i} \in \Omega} \\
&= \arg\min_{\boldsymbol{\nu}} \frac{1}{2} \|\boldsymbol{\nu} - \boldsymbol{\zeta}\|_2^2 + \lambda \|\boldsymbol{\nu}\|_1
\end{aligned}
$$

Wavelet denoising: $\mathbf{z} = \overline{\mathbf{x}} + \mathrm{w}$ with $\mathrm{w} = \mathcal{N}(0, \alpha \mathrm{Id})$

$\mathbf{z}$



$\boldsymbol{\zeta} = \Psi \mathbf{z}$

$\mathrm{soft}_\lambda(\Psi \mathbf{z})$

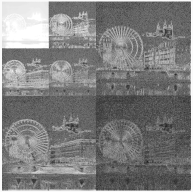
$\widehat{\mathrm{x}} = \Psi^* \mathrm{soft}_\lambda(\Psi \mathbf{z})$

$$\begin{aligned}
\mathrm{soft}_\lambda(\boldsymbol{\zeta}) &= \big(\max\{|\zeta_{\underline{i}}| - \lambda, 0\} \mathrm{sign}(\zeta_{\underline{i}}) \big)_{\underline{i} \in \Omega} \\
&= \arg\min_{\boldsymbol{\nu}} \frac{1}{2} \|\boldsymbol{\nu} - \boldsymbol{\zeta}\|_2^2 + \lambda \|\boldsymbol{\nu}\|_1 \\
&= \mathrm{prox}_{\lambda \|\cdot\|_1}(\boldsymbol{\zeta})
\end{aligned}$$

$\rightarrow$ **proximity operator**

Wavelet denoising: $\mathbf{z} = \overline{\mathrm{x}} + \mathrm{w}$ with $\mathrm{w} = \mathcal{N}(0, \alpha \mathrm{Id})$

$\mathbf{z}$



$\boldsymbol{\zeta} = \Psi \mathbf{z}$

$\mathrm{soft}_{\lambda}(\Psi \mathbf{z})$

$\widehat{\mathrm{x}} = \Psi^* \mathrm{soft}_{\lambda}(\Psi \mathbf{z})$

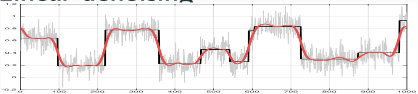
$$
\begin{aligned}
\mathrm{soft}_{\lambda}(\boldsymbol{\zeta}) &= \big(\max\{|\zeta_{\underline{i}}| - \lambda, 0\} \mathrm{sign}(\zeta_{\underline{i}})\big)_{\underline{i} \in \Omega} \\
&= \arg\min_{\boldsymbol{\nu}} \frac{1}{2} \|\boldsymbol{\nu} - \boldsymbol{\zeta}\|_2^2 + \lambda \|\boldsymbol{\nu}\|_1 \\
&= \mathrm{prox}_{\lambda \|\cdot\|_1}(\boldsymbol{\zeta}) \quad \rightarrow \textbf{proximity operator} \\
\widehat{\mathrm{x}} &= \mathrm{prox}_{\lambda \|\Psi \cdot\|_1}(\mathrm{z})
\end{aligned}
$$

Piecewise constant denoising: $\mathbf{z} = \overline{\mathbf{x}} + \mathrm{w}$ with $\mathrm{w} = \mathcal{N}(0, \sigma^2 \mathrm{Id})$

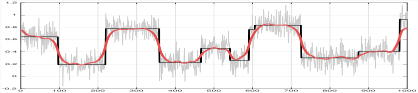
- **Minimization problem**:

$$\widehat{\mathrm{x}}(\mathrm{z}; \widehat{\lambda}) = \arg\min_{\mathrm{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathrm{x} - \mathrm{z}\|_2^2 + \lambda \|\mathrm{D}\mathrm{x}\|_{\bullet} \quad \text{where} \quad \begin{cases} \mathrm{D}\mathrm{x} = \psi * \mathrm{x} \\ \lambda > 0 \end{cases}$$

- **Linear denoising**



$\psi = [1 \quad -1]$; $\| \cdot \|_{\bullet} = \| \cdot \|_2^2$; Large λ



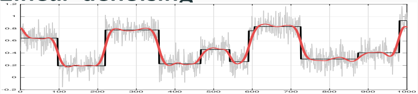
$\psi = [1 \quad -1]$; $\| \cdot \|_{\bullet} = \| \cdot \|_2^2$; Small λ

Piecewise constant denoising: $\mathbf{z} = \overline{\mathbf{x}} + \mathbf{w}$ with $\mathbf{w} = \mathcal{N}(0, \sigma^2 \mathrm{Id})$

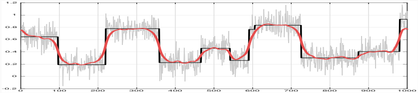
- **Minimization problem**:

$$\widehat{\mathrm{x}}(\mathrm{z}; \widehat{\lambda}) = \arg \min_{\mathrm{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathrm{x} - \mathrm{z}\|_2^2 + \lambda \|\mathrm{Dx}\|_\bullet \quad \text{where} \quad \begin{cases} \mathrm{Dx} = \psi * \mathrm{x} \\ \lambda > 0 \end{cases}$$

- **Linear denoising**

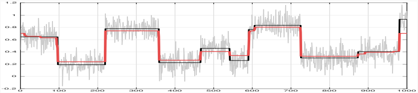


$\psi = [1 \quad -1]$; $\| \cdot \|_\bullet = \| \cdot \|_2^2$; Large λ

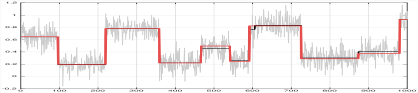


$\psi = [1 \quad -1]$; $\| \cdot \|_\bullet = \| \cdot \|_2^2$; Small λ

- **Non-linear denoising**.



$\psi = [1 \quad -1]$ and $\| \cdot \|_\bullet = \| \cdot \|_1$



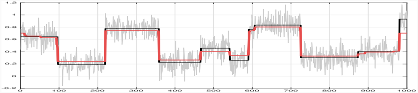
$\psi = [1 \quad -1]$ and $\| \cdot \|_\bullet = \| \cdot \|_0$

Piecewise linear denoising: $\mathbf{z} = \overline{\mathbf{x}} + \mathrm{w}$ with $\mathrm{w} = \mathcal{N}(0, \sigma^2 \mathrm{Id})$

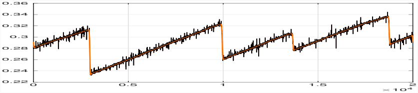
- **Minimization problem**:

$$\widehat{\mathrm{x}}(\mathrm{z}; \widehat{\lambda}) = \arg \min_{\mathrm{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathrm{x} - \mathrm{z}\|_2^2 + \lambda \|\mathrm{D}\mathrm{x}\|_\bullet \quad \text{where} \quad \begin{cases} \mathrm{D}\mathrm{x} = \psi * \mathrm{x} \\ \lambda > 0 \end{cases}$$

- **Non-linear denoising: piecewise constant/linear**



$\psi = \begin{bmatrix} 1 & -1 \end{bmatrix}$ and $\| \cdot \|_\bullet = \| \cdot \|_1$



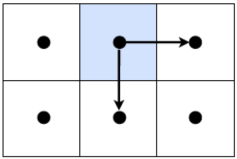
$\psi = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}$ and $\| \cdot \|_\bullet = \| \cdot \|_1$

2D-total variation

Anisotropic total variation (Rudin et al. 1992)

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - z\|_2^2 + \lambda \sum_{n_1=1}^{N-1} \sum_{n_2=1}^{N-1} \left(|x_{n_1+1,n_2} - x_{n_1,n_2}| + |x_{n_1,n_2+1} - x_{n_1,n_2}|\right)$$

- Horizontal and vertical difference filters:
  $\psi_1 = [1 - 1]$ and $\psi_2 = \psi_1^\top$
- Link between (ψ_1, ψ_2) and (D_1, D_2):
  cf. slides 7
- Sparse transform: $D = [D_1^\top, D_2^\top]^\top \in \mathbb{R}^{2N \times N}$
- Regularization: $R(x) = \|Dx\|_1 = \|D_1 x\|_1 + \|D_2 x\|_1$



2D-total variation

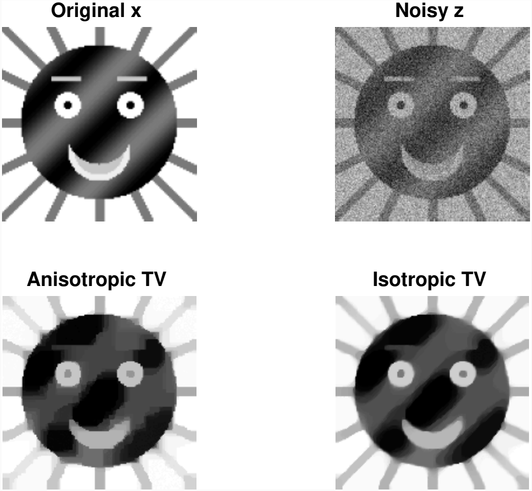
Isotropic total variation

$$\widehat{\mathrm{x}} = \underset{\mathrm{x}\in\mathbb{R}^N}{\arg\min}\ \frac{1}{2}\|\mathrm{x}-\mathrm{z}\|_2^2 + \lambda \sum_{n_1=1}^{N-1}\sum_{n_2=1}^{N-1}\sqrt{|x_{n_1+1,n_2}-x_{n_1,n_2}|^2+|x_{n_1,n_2+1}-x_{n_1,n_2}|^2}$$

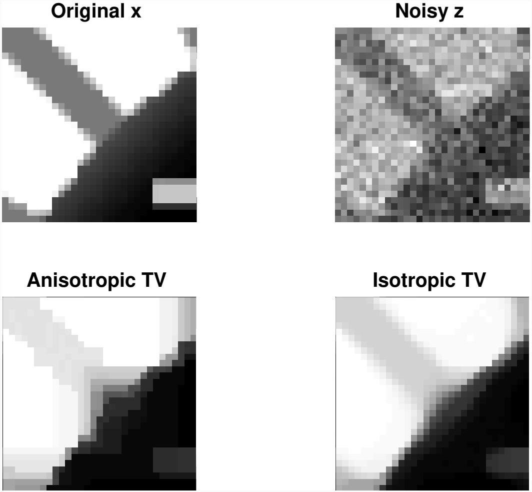
- Horizontal and vertical difference filters: $\psi_1 = [1 - 1]$ and $\psi_2 = \psi_1^\top$
- Link between (ψ_1, ψ_2) and $(\mathrm{D}_1, \mathrm{D}_2)$:

  cf. slides 7
- Sparse transform: $\mathrm{D} = [\mathrm{D}_1^\top, \mathrm{D}_2^\top]^\top \in \mathbb{R}^{2N\times N}$
- Regularization: $R(\mathrm{x}) = \|\mathrm{D}\mathrm{x}\|_{1,2} \rightarrow$ **coupling**

2D-total variation

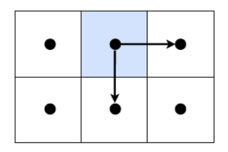


2D-total variation



Penalization choices

- Horizontal/vertical gradient, TV : $R(\mathrm{x}) = \|\mathrm{Dx}\|_1 = \|\mathrm{D}_1\mathrm{x}\|_1 + \|\mathrm{D}_2\mathrm{x}\|_1$



Penalization choices

- Horizontal/vertical gradient, TV
- Hessian operator: second order derivative along horizontal, diagonal and vertical direction. (Lefkimmiatis,Ward, Unser, 2013)

$$[\mathbf{T}_{\mathcal{H}}\mathbf{x}]_n = \begin{bmatrix} [\mathbf{D}^2_{11}\mathbf{x}]_n & [\mathbf{D}^2_{12}\mathbf{x}]_n \\ [\mathbf{D}^2_{12}\mathbf{x}]_n & [\mathbf{D}^2_{22}\mathbf{x}]_n \end{bmatrix} \quad \Rightarrow \quad R(\mathrm{x}) = \sum_n \|[\mathbf{T}_{\mathcal{H}}\mathbf{x}]_n\|_p$$

Penalization choices

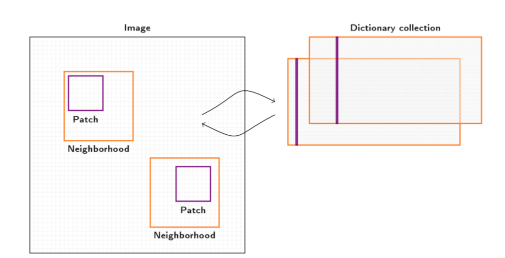
- Horizontal/vertical gradient, TV
- Hessian operator: second order derivative along horizontal, diagonal and vertical direction. (Lefkimmiatis,Ward, Unser, 2013)
- Weighted nonlocal gradients (Gilboa, Osher, 2008)(Bougleux, Peyré, Cohen, 2011)

$$[\mathbf{T}_{\mathrm{NL}}\mathbf{x}]_n = \begin{bmatrix} [\mathbf{W}_1(\mathbf{F}_1\mathbf{x}-\mathbf{x})]_n \\ \vdots \\ [\mathbf{W}_T(\mathbf{F}_T\mathbf{x}-\mathbf{x})]_n \end{bmatrix} \quad \Rightarrow \quad R(\mathrm{x}) = \sum_n \|[\mathbf{T}_{\mathrm{NL}}\mathbf{x}]_n\|_{1,2}$$

$\mathbf{W}_t = \mathrm{diag}\left(\exp\left(-\frac{1}{\eta}\mathbf{B}(\mathbf{F}_t\tilde{\mathbf{x}}-\tilde{\mathbf{x}})^2\right)\right)$: diagonal weight matrices,
$\mathbf{F}_t$: translation operator,
$\mathbf{B}$: lowpass filtering.

Penalization choices

- Horizontal/vertical gradient, TV
- Hessian operator: second order derivative along horizontal, diagonal and vertical direction. (Lefkimmiatis,Ward, Unser, 2013)
- Weighted nonlocal gradients (Gilboa, Osher, 2008)(Bougleux, Peyré, Cohen, 2011)
- Local dictionaries of patches + nuclear norm (i.e. l1) (Boulanger, Pustelnik, Condat, Piolot, Sengmanivong, 2018)



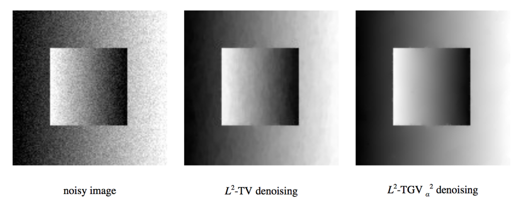
Penalization choices

- Horizontal/vertical gradient, TV
- Hessian operator: second order derivative along horizontal, diagonal and vertical direction. (Lefkimmiatis,Ward, Unser, 2013)
- Weighted nonlocal gradients (Gilboa, Osher, 2008)(Bougleux, Peyré, Cohen, 2011)
- Local dictionaries of patches + nuclear norm (i.e. l1) (Boulanger, Pustelnik, Condat, Piolot, Sengmanivong, 2018)
- TGV (Bredies,Kunisch,Pock,2010)

$$R(\mathbf{x}) = \min_{\mathbf{y}} \|\mathrm{Dx} - \mathbf{y}\|_1 + \gamma \|\tilde{\mathrm{D}}\mathbf{y}\|_1$$

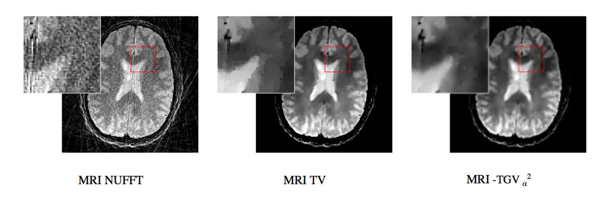
Penalization choices

- Horizontal/vertical gradient, TV
- Hessian operator: second order derivative along horizontal, diagonal and vertical direction. (Lefkimmiatis,Ward, Unser, 2013)
- Weighted nonlocal gradients (Gilboa, Osher, 2008)(Bougleux, Peyré, Cohen, 2011)
- Local dictionaries of patches + nuclear norm (i.e. l1) (Boulanger, Pustelnik, Condat, Piolot, Sengmanivong, 2018)
- TGV (Bredies,Kunisch,Pock,2010)



Penalization choices

- Horizontal/vertical gradient, TV
- Hessian operator: second order derivative along horizontal, diagonal and vertical direction. (Lefkimmiatis,Ward, Unser, 2013)
- Weighted nonlocal gradients (Gilboa, Osher, 2008)(Bougleux, Peyré, Cohen, 2011)
- Local dictionaries of patches + nuclear norm (i.e. l1) (Boulanger, Pustelnik, Condat, Piolot, Sengmanivong, 2018)
- TGV (Bredies,Kunisch,Pock,2010)



Penalization choices

- Horizontal/vertical gradient, TV
- Hessian operator: second order derivative along horizontal, diagonal and vertical direction. (Lefkimmiatis,Ward, Unser, 2013)
- Weighted nonlocal gradients (Gilboa, Osher, 2008)(Bougleux, Peyré, Cohen, 2011)
- Local dictionaries of patches + nuclear norm (i.e. l1) (Boulanger, Pustelnik, Condat, Piolot, Sengmanivong, 2018)
- TGV (Bredies,Kunisch,Pock,2010)
- 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)

Penalization choices

Synthesis formulation

$\widehat{\mathrm{x}} = \mathrm{D}^*\widehat{\zeta}$ with $\mathrm{D} \in \mathbb{R}^{P \times N}$

$$\widehat{\zeta} \in \underset{\zeta}{\mathrm{Argmin}} \frac{1}{2}\|\Phi \mathrm{D}^* \zeta - \mathrm{z}\|_2^2 + \lambda \|\zeta\|_\bullet$$

- X-lets
- Sparse coding

Analysis formulation

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}} \frac{1}{2}\|\Phi \mathrm{x} - \mathrm{z}\|_2^2 + \lambda \|\mathrm{D}\mathrm{x}\|_\bullet$$

- Horizontal/vertical gradients:
- Hessian operator
- NLTV
- ...

(webpage L. Duval)(Aharon, Elad, Bruckstein, 2006) (Mairal, Sapiro, Elad, 2007)(Gilboa, Osher, 2008)(K Bredies, K Kunisch, T Pock, 2010)(Jacques, Duval, Chaux, Peyré, 2011) (S Lefkimmiatis, A Bourquard, M Unser,2011) (Zoran, Weiss, 2011) (G Kutyniok, D Labate, 2012)(Chierchia et al.,2014)(Boulanger et al., 2018)...

Penalization choices

Synthesis formulation

$\widehat{\mathrm{x}} = \mathrm{D}^* \widehat{\zeta}$ with $\mathrm{D} \in \mathbb{R}^{P \times N}$

$$\widehat{\zeta} \in \underset{\zeta}{\operatorname{Argmin}} \frac{1}{2} \|\Phi \mathrm{D}^* \zeta - \mathrm{z}\|_2^2 + \lambda \|\zeta\|_\bullet$$

Analysis formulation

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\operatorname{Argmin}} \frac{1}{2} \|\Phi \mathrm{x} - \mathrm{z}\|_2^2 + \lambda \|\mathrm{D}\mathrm{x}\|_\bullet$$

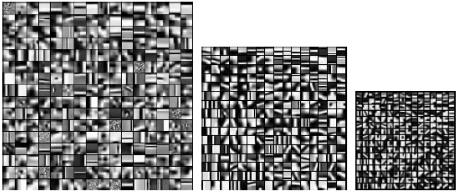
$\Rightarrow$ **Equivalence for D orthonormal basis.**

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

Penalization choices

Choice for D (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)

Penalization choices

(Pustelnik, Benazza-Benhayia, Zheng, Pesquet, 2010)

Property 4.2 *The analysis formulation (51) is a particular case of the synthesis formulation (52).*

Proof. By definition of a frame, F^* is surjective. Consequently, for every $y \in \mathbb{R}^N$, there exists an element x in $\mathbb{R}^K$ such that $y = F^* x$ and (51) can be rewritten as

$$\widehat{y} \in \underset{y=F^*x}{\operatorname{Argmin}} \sum_{i=1}^{I} f_i(F^*x) + \sum_{j=1}^{J} g_j(FF^*x), \tag{53}$$

that is

$$\widehat{y} = F^*\widehat{x} \quad \text{with} \quad \widehat{x} \in \underset{x}{\operatorname{Argmin}} \sum_{i=1}^{I} f_i(F^*x) + \sum_{j=1}^{J} h_j(x), \tag{54}$$

where, for every $j \in \{1, \ldots, J\}$, $h_j = g_j(FF^* \cdot)$. □

Property 4.3 *Let F be a non bijective tight frame analysis operator. If, for every $j \in \{1, \ldots, J\}$, g_j can be written as a sum of functions $h_{j,1} \colon \operatorname{Im} F \to]-\infty, +\infty]$ and $h_{j,2} \colon \ker F^* \to]-\infty, +\infty]$, i.e.,*

$$\big(\forall (y, x_\perp) \in \mathbb{R}^N \times \ker F^*\big) \qquad g_j(Fy + x_\perp) = h_{j,1}(Fy) + h_{j,2}(x_\perp) \tag{55}$$

where $\ker F^*$ is the nullspace of F^*, and if, for every $u \in \ker F^*$, $h_{j,2}(u) \geq h_{j,2}(0)$, then the analysis formulation (51) and the synthesis formulation (52) are equivalent.

Penalization choices

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.*

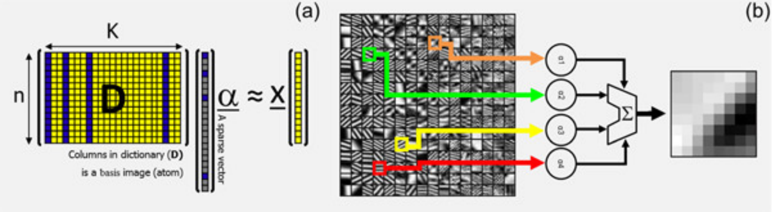
(P) Synthesis, (R) Analysis

(extracted M. Carlavan, P. Weiss, L. Blanc-Féraud, 2010.)

Dictionary learning

$$\widehat{\zeta} \in \operatorname*{Argmin}_{\zeta,\mathrm{D}} \frac{1}{2}\|\Phi \mathrm{D}\zeta - \mathrm{z}\|_2^2 + \lambda\|\zeta\|_1 \quad \text{s.t.} \quad \|\mathrm{D}_j\|_2 \le 1$$

- Overcomplete dictionaries for natural images
- Sparse decomposition
- (Olshausen and Field, 1997; Elad and Aharon, 2006; Raina et al., 2007)



[Source image : link]

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: $\widehat{\mathrm{x}} \in \underset{\mathrm{x}\in\mathbb{R}^N}{\text{Argmin}} \frac{1}{2}\|\Phi\mathrm{x}-\mathrm{z}\|_2^2 + \lambda\|\mathrm{D}\mathrm{x}\|_1 + \iota_{\geq 0}(\mathrm{x})$
- Example 2: $\widehat{\boldsymbol{\zeta}} \in \underset{\boldsymbol{\zeta}\in\mathbb{R}^K}{\text{Argmin}} \frac{1}{2}\|\Phi\mathrm{D}^*\boldsymbol{\zeta}-\mathrm{z}\|_2^2 + \lambda\|\boldsymbol{\zeta}\|_1$
- Example 3: $\widehat{\mathrm{x}} \in \underset{\mathrm{x}\in\mathbb{R}^N}{\text{Argmin}} \sum_{\underline{n}} -z_{\underline{n}} \ln \sigma\mathrm{x}_{\underline{n}} + \sigma\mathrm{x}_{\underline{n}} + \lambda \sum_{g\in\mathcal{G}} \|(\mathrm{D}\mathrm{x})_g\|_2$

Proximal algorithms

General objective function

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

where H_j linear operator from $\mathcal{H}$ to $\mathcal{G}_j$ and $(f_j)_{1 \leq j \leq J}$ belong to the class of convex functions, l.s.c., and proper from $\mathcal{G}_j$ to $]-\infty, +\infty]$.

Proximal algorithms

General objective function

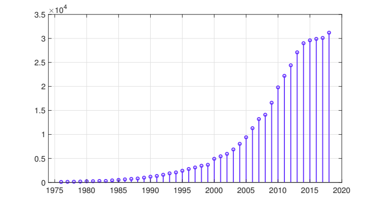
$$\text{Find} \qquad \widehat{y} \in \underset{y \in \mathcal{H}}{\text{Argmin}} \sum_{j=1}^{J} f_j(H_j y)$$

where H_j linear operator from $\mathcal{H}$ to $\mathcal{G}_j$ and $(f_j)_{1 \leq j \leq J}$ belong to the class of convex functions, l.s.c., and proper from $\mathcal{G}_j$ to $]-\infty, +\infty]$.

- Numerous proximal algorithms

  [Bauschke-Combettes, 2017]

  - Forward-Backward
  - Douglas-Rachford
  - ADMM
  - Primal-dual . . .



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