Regularization and hyperparameter estimation

Lotfi CHAARI

February 14, 2011

Inverse problem

Observation model

$$\mathbf{y} = \mathbf{S}\mathbf{z} + \mathbf{n}$$

Maximum Likelihood estimation

$$\hat{\mathbf{z}}_{\mathrm{ML}} = [\mathbf{S}^{\mathsf{H}}\boldsymbol{\Psi}^{-1}\mathbf{S}]^{-1}\mathbf{S}^{\mathsf{H}}\boldsymbol{\Psi}^{-1}\mathbf{y}$$

Ill-posed inverse problem

$\Downarrow$

regularization

Regularization

Bayesian framework

Likelihood
Prior

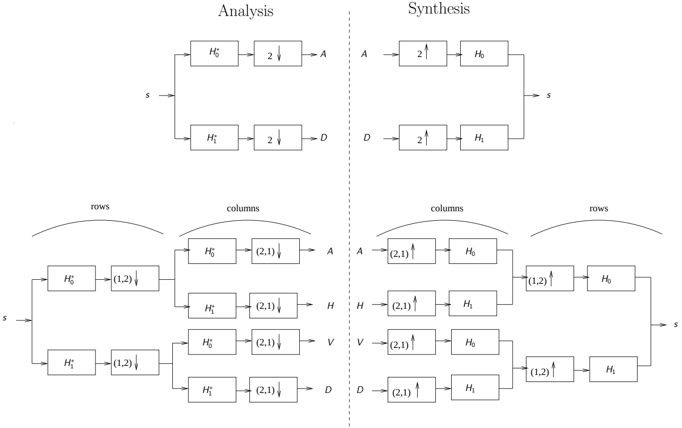
<table>
<tr><th>Original domain</th><th>Transform domain</th></tr>
<tr><td>- Tikhonov regularization
- Total Variation regularization</td><td>- Wavelet regularization</td></tr>
<tr><td colspan="2">Hybrid regularization</td></tr>
</table>

- **Wavelets ?**

Original space $\Longrightarrow$ Wavelet space

1D

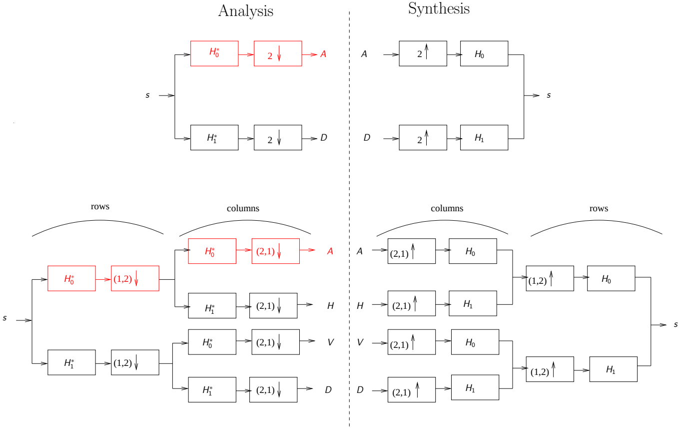
2D



3D ...

- **Wavelets ?**

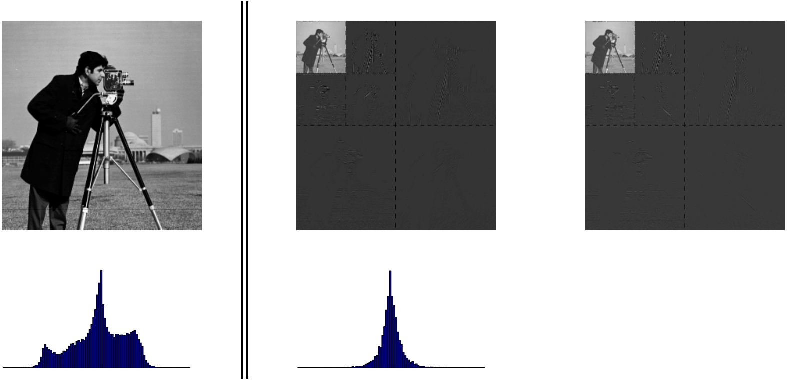
Original space $\Longrightarrow$ Wavelet space



1D

2D

3D ...



- Better spatial/frequential localization of artifacts
- More reliable/simple statistical model

Regularization

Regularization

MAP estimator

$$\mathbf{y} = \mathbf{S}(F^*\mathbf{x}) + \mathbf{n}$$

$$\hat{\mathbf{z}}_{\text{MAP}} = F^*\left\{\arg\min_{\mathbf{x}} \underbrace{D(\mathbf{y}, \mathbf{S}(F^*\mathbf{x}))}_{\textit{Likelihood}} + \underbrace{h_{\theta_1}(\mathbf{x}) + g_{\theta_2}(F^*\mathbf{x})}_{\textit{Prior}}\right\}$$

Frame analysis operator

$F : \mathbb{R}^L \to \mathbb{R}^K$
$\mathbf{z} \mapsto (\langle \mathbf{z} \mid \mathbf{e}_k \rangle)_{1 \le k \le K}$
$(K \ge L)$

Frame synthesis operator

$F^* : \mathbb{R}^K \to \mathbb{R}^L$
$(\xi_k)_{1 \le k \le K} \mapsto \sum_{k=1}^K \xi_k \mathbf{e}_k$

Problem

estimate $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2) \Longrightarrow$ estimate $\hat{\mathbf{z}}$

Regularization

MAP estimator

$$\mathbf{y} = \mathbf{S}(F^*\mathbf{x}) + \mathbf{n}$$

$$\hat{\mathbf{z}}_{\mathrm{MAP}} = F^*\left\{\arg\min_{\mathbf{x}} \underbrace{D(\mathbf{y}, \mathbf{S}(F^*\mathbf{x}))}_{\textit{Likelihood}} + \underbrace{h_{\theta_1}(\mathbf{x}) + g_{\theta_2}(F^*\mathbf{x})}_{\textit{Prior}}\right\}$$

$$\hat{\mathbf{z}}_{\mathrm{MAP}} = F^*\left\{\arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{S}(F^*\mathbf{x})\|^2 + \sum_{k\in\mathbb{K}} \frac{|x_k|^{p_k}}{\gamma_k} + \lambda\|F^*\mathbf{x}\|_{\mathrm{TV}}\right\}$$

$$\text{avec } \mathbb{K} = \{1, \ldots, K\}$$

Regularization

MAP estimator

$$\mathbf{y} = \mathbf{S}(F^*\mathbf{x}) + \mathbf{n}$$

$$\hat{\mathbf{z}}_{\text{MAP}} = F^*\left\{\arg\min_{\mathbf{x}} \underbrace{D(\mathbf{y}, \mathbf{S}(F^*\mathbf{x}))}_{\textit{Likelihood}} + \underbrace{h_{\theta_1}(\mathbf{x}) + g_{\theta_2}(F^*\mathbf{x})}_{\textit{Prior}}\right\}$$

$$\hat{\mathbf{z}}_{\text{MAP}} = F^*\left\{\arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{S}(F^*\mathbf{x})\|^2 + \sum_{g\in\mathbb{G}}\sum_{k\in S_g}\frac{|x_k|^{p_g}}{\gamma_g} + \lambda\|F^*\mathbf{x}\|_{\text{TV}}\right\}$$

$$\text{avec } \mathbb{G} = \{1, \ldots, G\}$$

$$\theta = \left((p_g, \gamma_g)_{g\in\mathbb{G}}, \lambda\right)$$

Optimization

- Optimality criterion

$$\mathcal{J} = D + h_{\theta_1} + g_{\theta_2}$$

  - D : convex, differentiable & even of β-Lipschitz gradient
  - h_{θ_1} and g_{θ_2} : convex but not necessarely differentiable

$\Longrightarrow$ Using standard optimization algorithms : complicated

- **Goal** : iteratively calculate $\widehat{\mathbf{x}}$ which minimizes $\mathcal{J}$

$$\Rightarrow \widehat{\mathbf{z}} = F^* \widehat{\mathbf{x}}$$

- **Tools** : convex optimization tools relying on proximity operators [Daubechies 04] [Combettes 05,08] [Chaux 07]

Optimization

- The proximity operator

Definition : [Moreau (65)]

* $\Gamma_0(\mathcal{H})$: lower semicontinuous convex functions from a separable real Hilbert space χ to $]-\infty, +\infty[$
* $\varphi \in \Gamma_0(\mathcal{H})$, $\forall x \in \mathcal{H}$, the function $\varphi(.) + \|. - x\|^2/2$ achieves its infimum at a unique point denoted by $\mathrm{prox}_\varphi x$

Optimization

- The proximity operator

Definition : [Moreau (65)]
* $\Gamma_0(\mathcal{H})$: lower semicontinuous convex functions from a separable real Hilbert space χ to $]-\infty,+\infty[$
* $\varphi \in \Gamma_0(\mathcal{H})$, $\forall x \in \mathcal{H}$, the function $\varphi(.) + \|. - x\|^2/2$ achieves its infimum at a unique point denoted by $\mathrm{prox}_\varphi x$

Example :

$$\begin{array}{ccc} \varphi : \mathbb{R} \to \mathbb{R} & \Longrightarrow & \mathrm{prox}_\varphi : \mathbb{R} \to \mathbb{R} \\ \xi \mapsto \omega|\xi|^p & & \xi \mapsto \mathrm{sign}(\xi)\eta \end{array}$$

where η is the unique solution in $[0,+\infty[$ to $\eta + p\eta^{p-1}\alpha = |\xi|$ with $p \in [1,\infty[$ and $\alpha \in \mathbb{R}_+^*$

complex-valued signals

Parallel ProXimal Algorithm (PPXA) [Combettes and Pesquet 08]

Let $f_1 = D$, $f_2 = h$ and $f_3 = g$.

- Let $\gamma = 200$ and $\lambda = 1.99$ ($\gamma \in]0, +\infty[$ and $\lambda \in]0, 2[$)
- Let $(\omega_\ell)_{1\leq\ell\leq 3} \in]0,1]^3$ so that $\sum_{\ell=1}^{3} \omega_\ell = 1$
- Let $(\mathbf{x}_\ell^{(0)})_{1\leq\ell\leq 3} \in (\mathbb{C}^K)^3$ and $\mathbf{x}^{(0)} = \sum_{\ell=1}^{3} \omega_\ell \mathbf{x}_\ell^{(0)}$

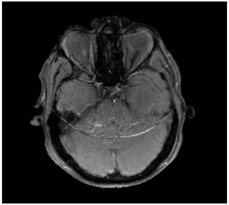
for $n = 0, 1, ...$ **do**

- for $\ell \in \{1, \dots, 3\}$, calculate $\mathbf{p}_\ell^{(n)} = \mathrm{prox}_{\gamma/\omega_\ell f_\ell} \mathbf{x}_\ell^{(n)}$
- $\mathbf{p}^{(n)} = \sum_{\ell=1}^{3} \omega_\ell \mathbf{p}_\ell^{(n)}$
- for $\ell \in \{1, \dots, 3\}$, calculate $\mathbf{x}_\ell^{n+1} = \mathbf{x}_\ell^{(n)} + \lambda \left(2\,\mathbf{p}^{(n)} - \mathbf{x}_n - \mathbf{p}_\ell^{(n)}\right)$
- $\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \lambda(\mathbf{p}^{(n)} - \mathbf{x}^{(n)})$

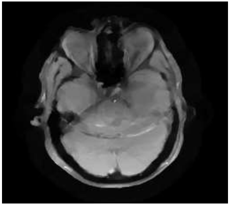
end for

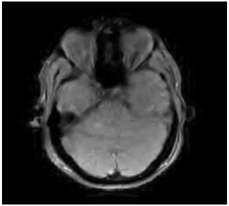
- After convergence, take $\widehat{\mathbf{z}} = F^* \mathbf{x}^{(n+1)}$

SENSE $(\mathrm{SNR} = 13.73\ \mathrm{dB})$

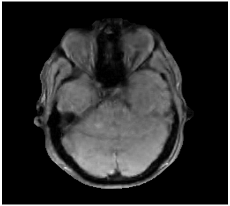


TV-SENSE $(\mathrm{SNR} = 13.39\ \mathrm{dB})$



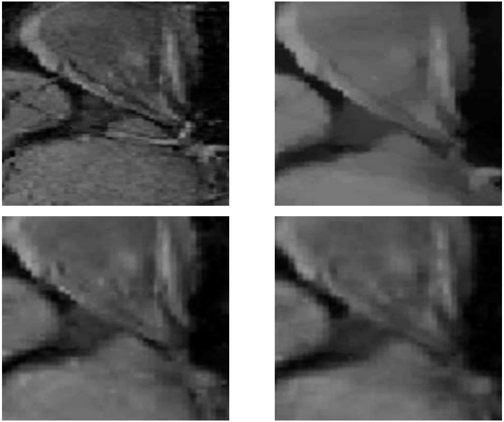


W-SENSE $(\mathrm{SNR} = 16.04\ \mathrm{dB})$



W-TV-SENSE $(\mathrm{SNR} = 16.37\ \mathrm{dB})$

SENSE (SNR $= 13.73$ dB) TV-SENSE (SNR $= 13.39$ dB)



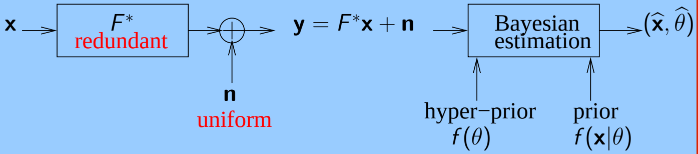
W-SENSE (SNR $= 16.04$ dB) W-TV-SENSE (SNR $= 16.37$ dB)

Hyperparameters estimation

Hyperparameters estimation

BAYES

- Reference image $\Rightarrow$ very low noise level
- Noisy observation with uniform noise



Measurement error : $C_\delta = \{\mathbf{x} \in \mathbb{R}^K \mid \|\mathbf{y} - F^*\mathbf{x}\| \leq \delta\}$

BAYES

Hierarchical Bayesian Model

Likelihood

$$f(\mathbf{y}|\mathbf{x}) \sim \mathcal{U}_{D_\delta}$$

with $D_\delta = \{\mathbf{y} \in \mathbb{R}^L \mid \|\mathbf{y} - F^*\mathbf{x}\| \leq \delta\}$

Frame coefficients prior

$$f(\mathbf{x}|\theta) = \frac{e^{-\lambda\|F^*\mathbf{x}\|_{\mathrm{TV}}}}{C(\theta)} \prod_{g\in\mathbb{G}} \frac{e^{\frac{-\sum_{k\in S_g}|\mathrm{x}_k|^{p_g}}{\gamma_g}}}{\gamma_g^{n_g/p_g}}$$

with $\theta = (\lambda, (\gamma_g, p_g)_{g\in\mathbb{G}})$

Hyperparameters prior

$$f(\theta) \propto C(\theta) 1_{[0,\lambda_{\max}]}(\lambda) \prod_{g\in\mathbb{G}} \left[\frac{1}{\gamma_g} 1_{\mathbb{R}^+}(\gamma_g) 1_{[0,p_{\max}]}(p_g)\right]$$

Hierarchical Bayesian Model

Posterior distribution

$$f(\mathbf{x},\theta|\mathbf{z}) = 1_{C_\delta}(\mathbf{x})\exp\left(-\lambda\|F^*\mathbf{x}\|_{\mathrm{TV}}\right)1_{[0,\lambda_{\max}]}(\lambda)$$

$$\prod_{g=1}^{G}\left[\left(\frac{1}{\gamma_g^{1/p_g}}\right)^{n_g}\exp\left(-\frac{1}{\gamma_g}\sum_{k\in S_g}|\mathbf{x}_k|^{p_g}\right)\left(\frac{1}{\gamma_g}1_{\mathbb{R}^+}(\gamma_g)1_{[0,p_{\max}]}(p_g)\right)\right]$$

- **Maximum a Posteriori ?**
- **Posterior Mean ?**

Posterior distribution

Bayesian estimators do not have simple closed form

$$\Downarrow$$

stochastic sampling techniques

Frame coefficients

Problem

Large dimension of F^* $\Longrightarrow$ direct sampling is time-consuming

Solution

Metropolis-Hastings sampler + appropriate proposal distribution q

$$\text{which : } \begin{cases} \text{guarantees that } \mathbf{x} \in C_\delta \\ \text{gives a tractable expression of the acceptance ratio} \end{cases}$$

Frame coefficients

Idea

$$\mathbb{R}^K = \mathrm{Ran}(F) \oplus \mathrm{Null}(F^*)$$
$$\mathbf{x} = \mathbf{x}_V + \mathbf{x}_{V^\perp}$$

$\hookrightarrow$ sample $\zeta \in C_\delta \Leftrightarrow$ sample $(\mathbf{x}_V, \mathbf{x}_{V^\perp})$
* <u>Sampling ζ_V :</u> $\mathbf{x}_V = F\nu \Leftarrow \nu = (F^*F)^{-1}\mathbf{u} \in \overline{C}_\delta \Leftarrow \mathbf{u} \sim U_{B_{\mathbf{y},\delta}}$
since $\mathbf{x} \in C_\delta \Leftrightarrow \nu \in \overline{C}_\delta = \{\nu | \, N(\mathbf{y} - F^*F\nu) \le \delta\}$
* <u>Sampling $\mathbf{x}_{V^\perp}$:</u> $\mathbf{x}_{V^\perp} = \Pi_{V^\perp}\mathbf{y} \Leftarrow \mathbf{z} \sim \mathcal{N}(\mathbf{x}^{(i-1)}, \sigma_{\mathbf{x}}^2 \mathbf{I})$

Hyperparameters

Idea

a) p_g : MH sampler with a truncated Gaussian distribution on $[0, p_{\max}]$ as proposal

b) γ_g : Sample directly according to $\mathcal{IG}\left(\frac{n_g}{p_g}, \sum_{k\in S_g} |\zeta_k|^{p_g}\right)$

c) λ : MH sampler with a truncated Gaussian distribution on $[0, \lambda_{\max}]$ as proposal

Without TV

take $\lambda = 0$

Results

Precision of the hyperparameters estimation

T : $\mathcal{B}_1 \bigcup \mathcal{B}_2$ - Daubechies 8 et 4
N : $\|\cdot\|_\infty$

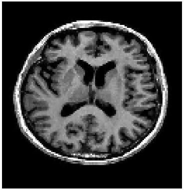
1) Generate the hyperparameters according to their prior
2) Generate the frame coefficients ($\delta = 10^{-4}$)

$\mathcal{B}_1$	MMSE		MAP	
	p	γ	p	γ
$h_{1,1}$	0.012	0.030	0.013	0.039
$v_{1,1}$	0.022	0.026	0.034	0.051
$d_{1,1}$	0.011	0.044	0.025	0.051
$h_{1,2}$	0.021	0.026	0.033	0.037
$v_{1,2}$	0.020	0.019	0.031	0.022
$d_{1,2}$	0.023	0.041	0.024	0.038
a_1	0.039	0.023	0.052	0.034

$\mathcal{B}_2$	MMSE		MAP	
	p	γ	p	γ
$h_{2,1}$	0.015	0.025	0.019	0.038
$v_{2,1}$	0.025	0.031	0.034	0.042
$d_{2,1}$	0.029	0.023	0.037	0.035
$h_{2,2}$	0.016	0.034	0.024	0.041
$v_{2,2}$	0.013	0.022	0.019	0.030
$d_{2,2}$	0.011	0.040	0.019	0.041
a_2	0.010	0.028	0.017	0.032

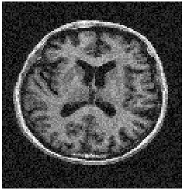
Application to image denoising

Original

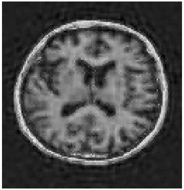


SNR=13.85 dB

Noisy



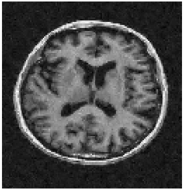
Variational



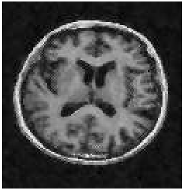
SNR=14.89 dB

uniform noise
on $[-\delta, \delta]^{128 \times 128}$
$\delta = 30$

SNR=15.69 dB



Wavelets



Wavelets + TV

SNR=**15.91** dB

Conclusion

- Regularization issues
- Hyperparameters and frame coefficient estimation for signals/images based on a noisy observation.

Thank you !