ML ESTIMATION OF WAVELET REGULARIZATION HYPERPARAMETERS IN INVERSE PROBLEMS

Roberto Cavicchioli[1], Caroline Chaux[2], Laure Blanc-Féraud[3] and Luca Zanni[1]

[1]University of Modena and Reggio Emilia, [2]Aix-Marseille Univ., [3] MORPHEME I3S/INRIA/IBV

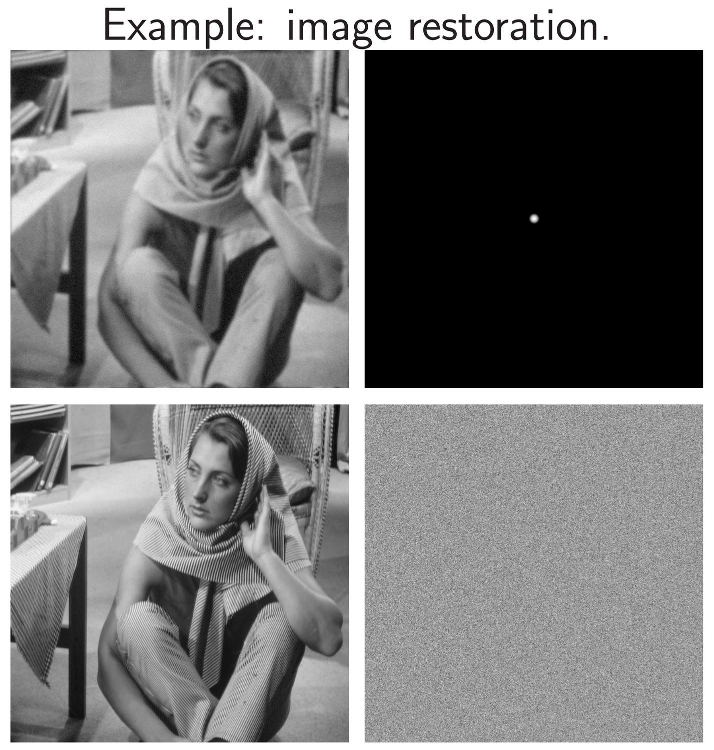
Maximum Likelihood estimation of the hyperparameters

Inverse Problem

$$g = Au + \eta$$

- A blurring linear operator
- η additive Gaussian white noise $\mathcal{N}(0, \sigma^2 I)$
- $u, g, \eta \in \mathbb{R}^k$ (k number of pixels)

Example: image restoration.



Regularization of *ill-posed* inverse problem

$$J(u) = \underbrace{\frac{\|g - Au\|^2}{2\sigma^2}}_{\text{likelihood term}} + \underbrace{\sum_{\mathbf{m}=1}^{M} \lambda_{\mathbf{m}} \phi_{\mathbf{m}}(F_{\mathbf{m}} u)}_{\text{regularization term}}. \qquad (1)$$

$\phi = \ell_1$-norm, F = m-band *Wavelet Transform*, $\boldsymbol{\lambda} = (\lambda_{\mathbf{m}})_{\mathbf{m}=1,\dots,M}$ hyperparameters

Estimating $\boldsymbol{\lambda} \Leftrightarrow$ maximizing $p_{\boldsymbol{\lambda}}(g)$ w.r.t $\boldsymbol{\lambda}$

$$p_{\boldsymbol{\lambda}}(g) = \int_u p_{\boldsymbol{\lambda}}(g, u) du = \int_u p(g|u) p_{\boldsymbol{\lambda}}(u) du.$$

$$p(g|u) = \frac{1}{(2\pi)^{k/2}\sigma^k} \exp\left(\frac{-\|g - Au\|^2}{2\sigma^2}\right)$$

$$p_{\boldsymbol{\lambda}}(u) = \frac{1}{\int_u \exp\left(-\sum_{\mathbf{m}=1}^{M} \lambda_{\mathbf{m}} \phi_{\mathbf{m}}(F_{\mathbf{m}} u)\right) du} \prod_{\mathbf{m}=1}^{M} \exp\left(-\lambda_{\mathbf{m}} \phi_{\mathbf{m}}(F_{\mathbf{m}} u)\right) \qquad (2)$$

The **maximization** can be performed using a **gradient method**.

$$\frac{\partial \log p_{\boldsymbol{\lambda}}(g)}{\partial \lambda_{\mathbf{m}}} = \mathrm{E}_{\boldsymbol{\lambda}}[\phi_{\mathbf{m}}(F_{\mathbf{m}} u)] - \mathrm{E}_{\sigma, \boldsymbol{\lambda}}[\phi_{\mathbf{m}}(F_{\mathbf{m}} u)]$$

$\mathrm{E}_{\boldsymbol{\lambda}}$ defined according to *a priori* law (2)

$\mathrm{E}_{\sigma,\boldsymbol{\lambda}}$ defined according to the *a posteriori* law:

$$p_{\sigma,\boldsymbol{\lambda}}(u|g) = \frac{\exp\left(\frac{-\|g-Au\|^2}{2\sigma^2} - \sum_{\mathbf{m}=1}^{M} \lambda_{\mathbf{m}} \phi_{\mathbf{m}}(F_{\mathbf{m}} u)\right)}{\int_u \exp\left(\frac{-\|g-Au\|^2}{2\sigma^2} - \sum_{\mathbf{m}=1}^{M} \lambda_{\mathbf{m}} \phi_{\mathbf{m}}(F_{\mathbf{m}} u)\right) du}.$$

Operators A and F must be split to be diagonalized in different spaces!

Equivalent criterion to (1):

$$J(u, w) = \frac{1}{2\sigma^2 \mu}\left((w - (\mathrm{I} + C)^{-1} u)^{\mathrm{T}} (\mathrm{I} + C)(w - (\mathrm{I} + C)^{-1} u)\right) + \frac{1}{2\sigma^2}\left(\|g - Au\|^2\right) + \sum_{\mathbf{m}=1}^{M} \lambda_{\mathbf{m}} \phi_{\mathbf{m}}(F_{\mathbf{m}} u)$$

with $(\mathrm{I} + C)^{-1} = \mathrm{I} - \mu A^* A$ (μ such that $\mu\|A^*A\| < 1$).

Minimizing $J(u)$ w.r.t u is *equivalent* to minimize $J(u, w)$ w.r.t (u, w).

Decoupling of the operators and Sampling

New criterion $\Rightarrow$ new $p_{\boldsymbol{\lambda}}(g)$

$$p_{\boldsymbol{\lambda}}(g) = \int_{u,w} p(g|u) p(w|u) p_{\boldsymbol{\lambda}}(u) du dw$$

$$p(w|u) = \frac{\exp\left(-\frac{(w-(\mathrm{I}+C)^{-1}u)^{\mathrm{T}}(\mathrm{I}+C)(w-(\mathrm{I}+C)^{-1}u)}{2\sigma^2\mu}\right)}{(2\pi\sigma^2\mu)^{k/2} (\det(\mathrm{I} + C))^{-1/2}}$$

Remark: $p(w|u)$ is a Gaussian law $\mathcal{N}\left((\mathrm{I} + C)^{-1} u, \sigma^2 \mu (\mathrm{I} + C)^{-1}\right)$

Compute gradient for standard **Gradient Ascent** (GA):

$$\frac{\partial \log p_{\boldsymbol{\lambda}}(g)}{\partial \lambda_{\mathbf{m}}} = \mathrm{E}_{\boldsymbol{\lambda}}[\phi_{\mathbf{m}}(F_{\mathbf{m}} u)] - \mathrm{E}_{\sigma,\boldsymbol{\lambda},\mu}[\phi_{\mathbf{m}}(F_{\mathbf{m}} u)].$$

$\mathrm{E}_{\boldsymbol{\lambda}}$ can be computed analytically.

$\mathrm{E}_{\sigma,\boldsymbol{\lambda},\mu}$ sample u according to *a posteriori* law $p_{\sigma,\boldsymbol{\lambda},\mu}(u, w|g)$.

Rewrite $J(u, w)$ decoupling u and w:

$$J(u, w) = \frac{1}{2\sigma^2\mu}\left(\|u - w\|^2 + \langle Cw, w\rangle\right) + \frac{1}{2\sigma^2}\left(\|g\|^2 - 2\langle u, A^* g\rangle\right) + \sum_{\mathbf{m}=1}^{M} \lambda_{\mathbf{m}} \phi_{\mathbf{m}}(F_{\mathbf{m}} u).$$

Both variables estimated in **decorrelated** spaces (*Wavelets, Fourier*).

Sampling now possible in *reasonable* computing time!

1. Generate w samples (**Gibbs sampler**) according to $p(w|u)$ (Gaussian).
2. Generate u samples (**Metropolis Hastings** algorithm) according to:

$$p_{\boldsymbol{\lambda}}(u|w, g) \propto \exp\left(-\frac{1}{2\sigma^2\mu}\|Fu - Fw\|^2 + \frac{1}{\sigma^2}\langle Fu, FA^* g\rangle - \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \phi_{\mathbf{m}}(F_{\mathbf{m}} u)\right).$$

Accelerating Gradient Method

$$\text{Accelerate gradient methods} \Longrightarrow \begin{cases} \text{adaptive step-size rules} \\ \text{line-search strategies} \end{cases}$$

Main difficulty: objective function can't be evaluated

Idea: **two phases** (2Ph) gradient method.

1. Sequence of *simple gradient steps* (objective function not needed).
2. Using *line-search* to ensure *reduction of the gradient norm*.

Both phases exploit adaptive alternation of **Barzilai-Borwein** stepsizes:

$$\alpha_n^{BB1} = -({s_n}^T s_n)/({s_n}^T y_n), \qquad \alpha_n^{BB2} = -({s_n}^T y_n)/({y_n}^T y_n),$$

where $s_n = \boldsymbol{\lambda}^{(n)} - \boldsymbol{\lambda}^{(n-1)}$ and $y_n = \nabla p_{\boldsymbol{\lambda}^{(n)}}(g) - \nabla p_{\boldsymbol{\lambda}^{(n-1)}}(g)$.

Two Phases (2Ph) Gradient Method

Set $G(\boldsymbol{\lambda}) = \nabla p_{\boldsymbol{\lambda}}(g)$, $n = 1$, $\boldsymbol{\lambda}^{(1)} = \boldsymbol{\lambda}^{(0)} + \alpha_0 G(\boldsymbol{\lambda}^{(0)})$, $f(\boldsymbol{\lambda}) = \frac{1}{2}\|G(\boldsymbol{\lambda})\|^2$,
$gr_0 = f(\boldsymbol{\lambda}^{(0)})$, $gr = f(\boldsymbol{\lambda}^{(1)})$, $c\_check = 0$ (evidence of *local concavity*),
$\bar{q}_n(\alpha_n) = -(G(\boldsymbol{\lambda}^{(n)} + \alpha_n G(\boldsymbol{\lambda}^{(n)})) - G(\boldsymbol{\lambda}^{(n)}))$, $\theta, \gamma \in (0, 1)$ and integers $P, N \geq 1$.

Phase 1: (BB-like Gradient step)

WHILE $\left(\frac{gr}{gr_0} > \tau_g \text{ or } \frac{\|\boldsymbol{\lambda}^{(n)} - \boldsymbol{\lambda}^{(n-1)}\|}{\|\boldsymbol{\lambda}^{(n)}\|} > \tau_\lambda\right)$ and $n < N$ and $c\_check = 0$

1.1 Choose α_n and update c_check for *local concavity*;

1.2 Gradient Step: $\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n)} + \alpha_n G(\boldsymbol{\lambda}^{(n)})$, $n = n + 1$;

1.3 Set $gr = \max_{0 \leq j \leq \min(n, P-1)} f(\boldsymbol{\lambda}^{(n-j)})$;

ENDWHILE

Phase 2: (Stabilization with line-search)

WHILE $\left(\frac{gr}{gr_0} > \tau_g \text{ or } \frac{\|\boldsymbol{\lambda}^{(n)} - \boldsymbol{\lambda}^{(n-1)}\|}{\|\boldsymbol{\lambda}^{(n)}\|} > \tau_\lambda\right)$

2.1 Choose α_n;

2.2 Line-search:

IF $f\left(\boldsymbol{\lambda}^{(n)} + \alpha_n G(\boldsymbol{\lambda}^{(n)})\right) \leq gr - \gamma\|\alpha_n^2 G(\boldsymbol{\lambda}^{(n)})\|^2$ THEN

set $\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n)} + \alpha_n G(\boldsymbol{\lambda}^{(n)})$ and $n = n + 1$;

ELSE IF $f\left(\boldsymbol{\lambda}^{(n)} - \alpha_n \bar{q}_n(\alpha_n)\right) \leq gr - \gamma\|\alpha_n^2 G(\boldsymbol{\lambda}^{(n)})\|^2$ THEN

set $\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n)} - \alpha_n \bar{q}_n(\alpha_n)$ and $n = n + 1$;

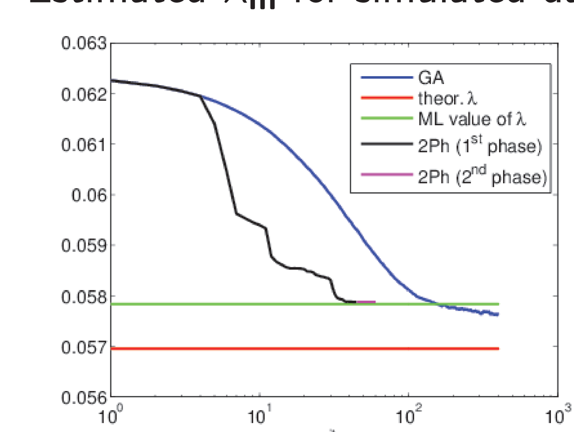
ELSE set $\alpha_n = \theta\alpha_n$ and go to Step 2.2; ENDIF

2.3 Set $gr = \max_{0 \leq j \leq \min(n, P-1)} f(\boldsymbol{\lambda}^{(n-j)})$;

ENDWHILE

Numerical Results

Estimated $\lambda_{\mathbf{m}}$ for simulated data



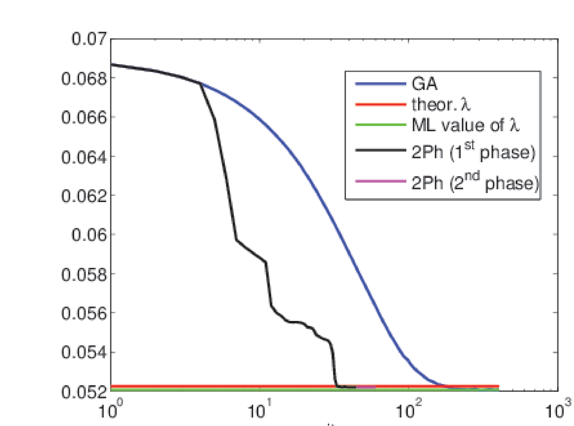
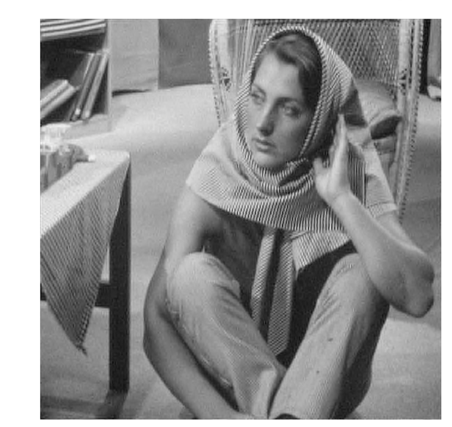
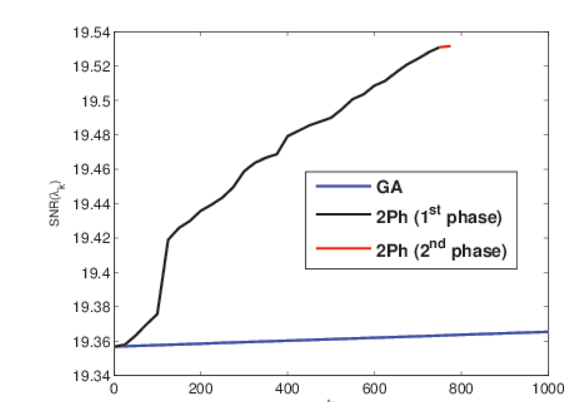


Table: Simulated Data

Alg.	It.	Grad.	Err_{ML}	$\mathrm{Err}_{th.}$	time
2Ph	61	90	0.007	0.021	431.8
GA	400	400	0.007	0.020	1920.5

Reconstructed Image





Problem	Alg.	It.	SNR_{init}	SNR_{fin}
Mandrill 256^2	2Ph	883	12.1616	14.2324
	GA	1000	12.1616	14.0051
Barbara 512^2	2Ph	785	18.5272	19.5316
	GA	1000	18.5272	19.3655