Unsupervised Deconvolution-Segmentation of Textured Image

Bayesian approach: optimal strategy and stochastic sampling

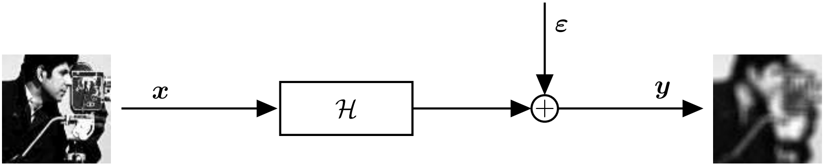
Cornelia Vacar and Jean-François Giovannelli

Audrey Giremus, Roxana Rosu, Andrei Barbos

Groupe Signal – Image
Laboratoire de l'Intégration du Matériau au Système
Université de Bordeaux – CNRS – BINP

New mathematical methods in computational imaging
Maxwell Insitute for Mathematical Sciences
School of Mathematical and Computer Sciences
Heriot-Watt University
2017, 29-th June

Inversion: standard question



Direct model / Inverse problem

- Direct model — Do degradations: noise, blur, mixing,...

$$\boldsymbol{y} = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{\varepsilon} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} = \boldsymbol{h} \star \boldsymbol{x} + \boldsymbol{\varepsilon}$$

- Inverse problem — Undo: denoising, deblurring, unmixing,...

$$\widehat{\boldsymbol{x}} = \mathcal{F}(\boldsymbol{y})$$

Various fields, modalities, problems,...

Fields

- Medical: diagnosis, prognosis, theranostics,...
- Astronomy, geology, hydrology,...
- Thermodynamics, fluid mechanics, transport phenomena,...
- Remote sensing, airborne imaging,...
- Surveillance, security,...
- Non destructive evaluation, control,...
- Computer vision, under bad conditions,...
- Photography, games, recreational activities, leisures,...
- ...

$\rightsquigarrow$ Health, knowledge, leisure,...
$\rightsquigarrow$ Aerospace, aeronautics, transport, energy, industry,...

Various fields, modalities, problems,. . .

Modalities

- Magnetic Resonance Imaging
- Tomography (X-ray, optical wavelength, tera-Hertz,. . .)
- Thermography,. . .
- Echography, Doppler echography,. . .
- Ultrasonic imaging, sound,. . .
- Microscopy, atomic force microscopy
- Interferometry (radio, optical, coherent,. . .)
- Multi-spectral and hyper-spectral,. . .
- Holography
- Polarimetry: optical and other
- Synthetic aperture radars
- . . .

$\rightsquigarrow$ Essentially “wave $\leftrightarrow$ matter” interaction

Various fields, modalities, problems,. . .

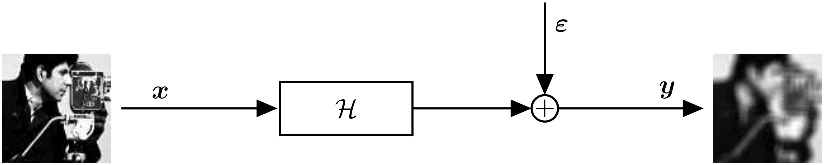
"Signal – Image" problems

- Denoising
- **Deconvolution**
- Inverse Radon
- Fourier synthesis
- Resolution enhancement, super-resolution
- Inter / extra-polation, **inpainting**
- Component unmixing / source separation
- . . .

- And also:
  - **Segmentation, labels and contours**
  - Detection of impulsions, salient points,. . .
  - Classification, clustering,. . .
  - . . .

Inversion: standard question

$$\boldsymbol{y} = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{\varepsilon} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} = \boldsymbol{h} \star \boldsymbol{x} + \boldsymbol{\varepsilon}$$



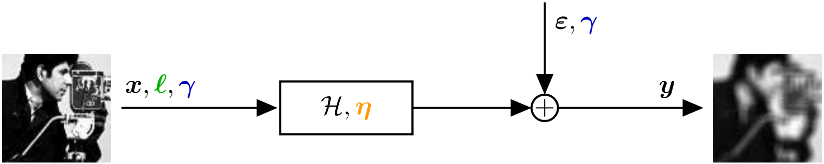
$$\widehat{\boldsymbol{x}} = \mathcal{F}(\boldsymbol{y})$$

Restoration, deconvolution, inter / extra-polation

- Issue: inverse problems
- Difficulty: ill-posed problems, *i.e., lack of information*
- Methodology: regularisation, *i.e., information compensation*
  - Specificity of the inversion methods
  - Compromise and tuning parameters

Inversion: advanced questions

$$\boldsymbol{y} = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{\varepsilon} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} = \boldsymbol{h} \star \boldsymbol{x} + \boldsymbol{\varepsilon}$$



$$\left[\, \widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\ell}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\eta}} \,\right] = \mathcal{F}(\boldsymbol{y})$$

Additional estimation problems

- Hyperparameters, tuning parameters: *self-tuned, unsupervised*
- Instrument parameters (resp. response): *myopic* (resp. *blind*)
- Hidden variables: edges, regions / labels, peaks,. . . : *augmented*

- Different models for image, noise, response,. . . : *model selection*

Some historical landmarks

- Quadratic approaches and linear filtering ~ 60
  - Phillips, Twomey, Tikhonov
  - Kalman
  - Hunt (and Wiener ~ 40)
- Extension: discrete hidden variables ~ 80
  - Kormylo & Mendel (impulsions, peaks,...)
  - Geman & Geman, Blake & Zisserman (lines, contours, edges,...)
  - Besag, Graffigne, Descombes (regions, labels,...)
- Convex penalties (also hidden variables,...) ~ 90
  - $\mathrm{L}_2 - \mathrm{L}_1$, Huber, hyperbolic: Sauer, Blanc-Fraud, Idier...
  - L_1: Alliney-Ruzinsky, Taylor ~ 79, Yarlagadda ~ 85 ...
  - And... L_1-boom ~ 2005
- Back to more complex approaches ~ 2000
  - Problems: unsupervised, semi-blind / blind, latent / hidden variables
  - Models: stochastic and hierarchical models
  - Methodology: Bayesian approaches and optimality
  - Algorithms: stochastic sampling (MCMC, Metropolis-Hastings,...)

$$\left[\,\widehat{\boldsymbol{\ell}}, \boldsymbol{x}_k, \widehat{\boldsymbol{\theta}}_k, \widehat{\gamma}_\varepsilon\,\right] = \mathcal{F}(\boldsymbol{y})$$

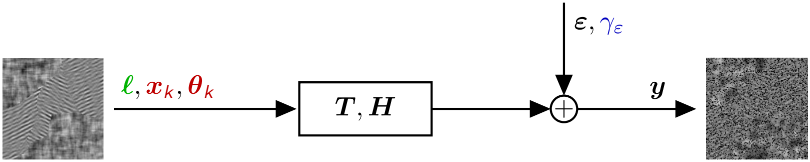


Image specificities

- Piecewise homogeneous
- Textured images, oriented textures
- Defined by: label $\boldsymbol{\ell}$ and texture parameters $\boldsymbol{\theta}_k$ for $k = 1, \ldots, K$

Observation: triple complication

1. Convolution
2. Missing data, truncation, mask
3. Noise

Outline

- Image model
  - Textured images, orientation
  - Piecewise homogeneous images, labels
- Observation system model
  - Convolution and missing data
  - Noise
- Hierarchical model
  - Conditional dependancies / independancies
  - Joint distribution
- Estimation / decision strategy and computations
  - Cost, risk and optimality $\rightsquigarrow$ Posterior distribution and estimation
  - Convergent computations: stochastic sampler $\oplus$ empirical estimates
    - Gibbs loop
    - Inverse cumulative density function
    - Metropolis-Hastings
- First numerical assessment
  - Behaviour, convergence,. . .
  - Labels, texture parameters and hyperparameters
  - Quantification of errors

Texture model: stationary Gauss Random Field

Original image $\boldsymbol{x} \sim \mathcal{N}(\,\mathbf{0}, \boldsymbol{R}_x\,)$, in $\mathbb{C}^P$

- Parametric covariance $\boldsymbol{R}_x = \boldsymbol{R}_x(\gamma_x, \boldsymbol{\theta})$
  - Natural parametrization: $\boldsymbol{R}_x(\gamma_x, \boldsymbol{\theta}) = \gamma_x^{-1} \boldsymbol{P}_x^{-1}(\boldsymbol{\theta})$
  - Parameters: scale γ_x and shape $\boldsymbol{\theta}$

$$f(\boldsymbol{x}|\boldsymbol{\theta}, \gamma_x) = (2\pi)^{-P} \gamma_x^P \det\left[\boldsymbol{P}_x(\boldsymbol{\theta})\right] \exp\left[-\gamma_x\, \boldsymbol{x}^\dagger \boldsymbol{P}_x(\boldsymbol{\theta}) \boldsymbol{x}\right]$$

- Whittle (circulant) approximation
  - Matrix $\boldsymbol{P}_x(\boldsymbol{\theta}) \longleftrightarrow$ eigenvalues $\lambda_p(\boldsymbol{\theta}) \longleftrightarrow$ field inverse PSD

$$f(\boldsymbol{x}|\boldsymbol{\theta}, \gamma_x) = \prod (2\pi)^{-1}\, \gamma_x\, \lambda_p(\boldsymbol{\theta})\; \exp\left[-\gamma_x\, \lambda_p(\boldsymbol{\theta})\, |\mathring{x}_p|^2\right]$$

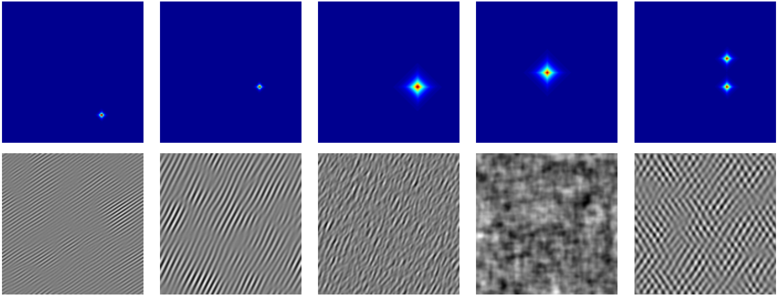
  - Separablilty w.r.t. the Fourier coefficients $\mathring{x}_p$
  - Precision parameter of the Fourier coefficients $\mathring{x}_p$: $\gamma_x\, \lambda_p(\boldsymbol{\theta})$
  - Any PSD, *e.g.*, Gaussian, Laplacian, Lorentzian,... more complex,...

- ...and K such models (PSD): $\boldsymbol{x}_k$ for $k = 1, \ldots, K$

Examples: Power Spectral Density and texture

- Laplacian PSD
- $\boldsymbol{\theta} = \left[(\nu_x^0, \nu_y^0)\,,\, (\omega_x, \omega_y) \right]$: central frequency and widths

$$\lambda^{-1}(\nu_x, \nu_y, \boldsymbol{\theta}) = \exp - \left[\frac{|\nu_x - \nu_x^0|}{\omega_x} + \frac{|\nu_y - \nu_y^0|}{\omega_y} \right]$$



Labels: a Markov field

Usual Potts model: favors large homogeneous regions

- Piecewise homogeneous image
  - P pixels in K classes (K is given)
  - Labels ℓ_p for $p = 1, \dots, P$ with discrete value in $\{1, \dots K\}$
- Count pairs of identical neighbour, "parcimony of a gradient"

$$\nu(\boldsymbol{\ell}) = \sum_{p \sim q} \delta(\ell_p; \ell_q) = \text{" } \| \, \mathit{Grad} \ \boldsymbol{\ell} \, \|_0 \text{ "}$$

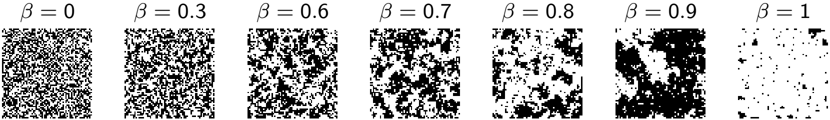
  - $\sim$: four nearest neighbours relation
  - δ: Kronecker function
- Probability law (exponential family)

$$\Pr[\boldsymbol{\ell} | \beta] = C(\beta)^{-1} \exp[\beta \nu(\boldsymbol{\ell})]$$

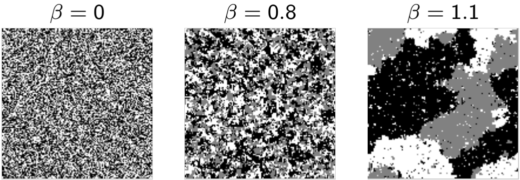
  - β: "correlation" parameter (mean number / size of the regions)
  - $C(\beta)$: normalization constant
- Various extensions: neighbour, interaction

Labels: a Potts field

- Example of realizations: Ising ($K = 2$)



- Example of realizations for $K = 3$



Labels: a Potts field

Partition function

$$\Pr[\boldsymbol{\ell}|\beta] = C(\beta)^{-1} \exp\left[\, \beta\nu(\boldsymbol{\ell}) \,\right]$$

Partition function (normalizing coefficient)

$$C(\beta) = \sum_{\boldsymbol{\ell} \in \{1,\dots K\}^P} \exp\left[\, \beta\nu(\boldsymbol{\ell}) \,\right]$$

$$\bar{C}(\beta) = \log\left[C(\beta)\right]$$

- Crucial in order to estimate β
- No closed-form expression (except for $K = 2$, $P = +\infty$)
- Enormous summation over the K^P configurations

Partition: an expectation computed as an empirical mean

- Distribution and partition

$$\Pr\left[\boldsymbol{\ell}|\beta\right] = C(\beta)^{-1} \exp\left[\beta\nu(\boldsymbol{\ell})\right] \quad \text{with} \quad C(\beta) = \sum \exp\left[\beta\nu(\boldsymbol{\ell})\right]$$

- A well-known result [Mac Kay] for exponential family:

$$C'(\beta) = \sum \nu(\boldsymbol{\ell}) \exp\left[\beta\nu(\boldsymbol{\ell})\right]$$

- Yields the log-partition derivative:

$$\bar{C}'(\beta) = \sum \nu(\boldsymbol{\ell}) \; C(\beta)^{-1} \exp\left[\beta\nu(\boldsymbol{\ell})\right] = \mathrm{E}\left[\, \nu(\boldsymbol{\ell}) \,\right]$$

- Approximated by an empirical mean

$$\bar{C}'(\beta) \simeq \frac{1}{Q} \sum \nu(\boldsymbol{\ell}^{(q)})$$

where the $\boldsymbol{\ell}^{(q)}$ are realizations of the field (given β)

Only few weeks of computations. . . but once for all !

Partition

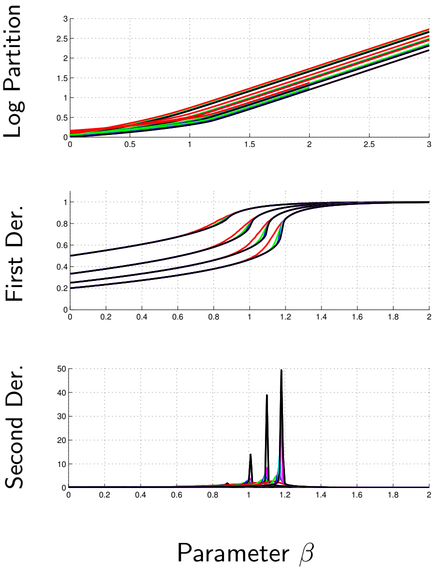


Image formation model

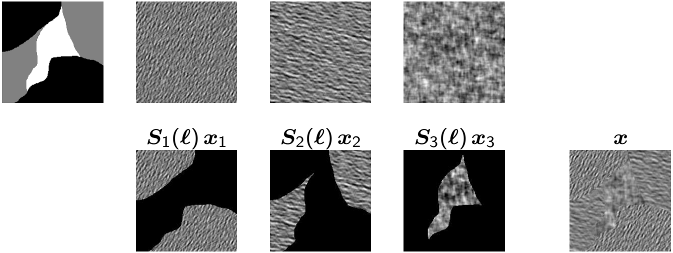
- Image $\boldsymbol{x}$ writes

$$\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{S}_k(\boldsymbol{\ell})\, \boldsymbol{x}_k$$

  - $\boldsymbol{x}_k$ for $k = 1, \ldots, K$: textured images (previous models)
  - $\boldsymbol{S}_k(\boldsymbol{\ell})$ for $k = 1, \ldots, K$: binary diagonal indicator of region k

$$\boldsymbol{S}_k(\boldsymbol{\ell}) = \mathrm{diag}\left[\, s_k(\ell_1), \ldots s_k(\ell_P) \,\right]$$

$$s_k(\ell_p) = \delta(\ell_p; k) = \begin{cases} 1 & \text{if the pixel } p \text{ is in the class } k \\ 0 & \text{if not} \end{cases}$$



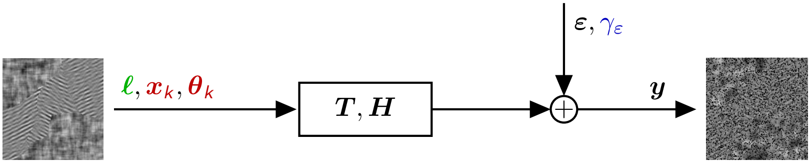
Outline

- Image model
  - Textured images, orientation
  - Piecewise homogeneous images
- Observation system model
  - Convolution and missing data
  - Noise
- Hierarchical model
  - Conditional dependancies / independancies
  - Joint distribution
- Estimation / decision strategy and computations
  - Cost, risk and optimality $\rightsquigarrow$ Posterior distribution and estimation
  - Convergent computations: stochastic sampler $\oplus$ empirical estimates
    - Gibbs loop
    - Inverse cumulative density function
    - Metropolis-Hastings
- First numerical assessment
  - Behaviour, convergence,. . .
  - Labels, texture parameters and hyperparameters
  - Quantification of errors

Observation model

Observation: triple complication

- Convolution: low-pass filter $\boldsymbol{H}$
- Missing data: truncation matrix $\boldsymbol{T}$, size $M \times P$
- Noise: $\boldsymbol{\varepsilon}$ accounts for measure and model errors



$$\boldsymbol{y} = \boldsymbol{T}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} = \begin{cases} y_m = [\boldsymbol{H}\boldsymbol{x}]_m + \varepsilon_m & \text{for observed pixels} \\ \text{Nothing} & \text{for missing pixels} \end{cases}$$

Noise

- Usual model
  - White and homogeneous
  - Zero-mean
  - Gaussian
  - Precision γ_ε

$$
\begin{aligned}
f(\boldsymbol{\varepsilon}|\gamma_\varepsilon) &= \mathcal{N}(\boldsymbol{\varepsilon}; \mathbf{0}, \gamma_\varepsilon^{-1} \boldsymbol{I}_M) \\
&= \pi^{-M}\, \gamma_\varepsilon^M\, \exp\left[\, -\gamma_\varepsilon \|\boldsymbol{\varepsilon}\|^2 \,\right]
\end{aligned}
$$

- Possible advanced models
  - Non gaussian (*e.g.*, Cauchy)
  - Poisson
  - Correlated, but...
  - ...

Hyperparameters

Precision parameter

- Model poorly informative
  - Conjugate prior: Gamma with parameter a_0, b_0
  - Nominal value (expected value) $\gamma = 1$
  - Very large variance

$$f(\gamma) = \mathcal{G}(\gamma; a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)} \gamma^{a_0 - 1} \exp\left[-b_0 \gamma\right] \mathbb{1}_+(\gamma)$$

Correlation parameter

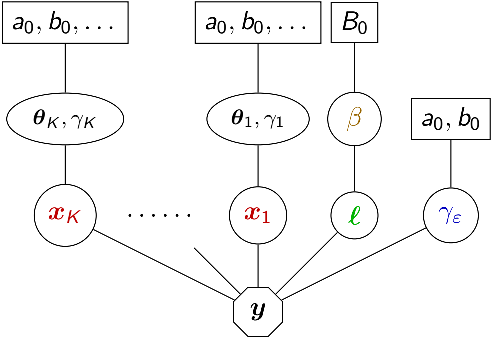
- Model poorly informative
  - No simple conjugate prior
  - Uniform prior on $[0, B_0]$
  - B_0 is the maximum authorised value, *e.g.*, $B_0 = 3$

$$f(\beta) = \mathcal{U}_{[0, B_0]}(\beta)$$

Outline

- Image model
  - Textured images, orientation
  - Piecewise homogeneous images
- Observation system model
  - Convolution and missing data
  - Noise
- Hierarchical model
  - Conditional dependancies / independancies
  - Joint distribution
- Estimation / decision strategy and computations
  - Cost, risk and optimality $\rightsquigarrow$ Posterior distribution and estimation
  - Convergent computations: stochastic sampler $\oplus$ empirical estimates
    - Gibbs loop
    - Inverse cumulative density function
    - Metropolis-Hastings
- First numerical assessment
  - Behaviour, convergence,...
  - Labels, texture parameters and hyperparameters
  - Quantification of errors

Hierarchy and distributions



Total joint distribution

$$f(\boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta) = f(\boldsymbol{y}|\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \gamma_\varepsilon)$$
$$\Pr[\boldsymbol{\ell}|\beta] \prod f(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \gamma_k)\; f(\gamma_\varepsilon)\; f(\beta) \prod f(\boldsymbol{\theta}_k) \prod f(\gamma_k)$$

- And then: “Total joint distribution”
  - Likelihood
  - Marginal distributions
  - Posterior and conditional posteriors

Optimal estimation / decision function

Usual Bayesian strategy: cost, risk, optimum

- Estimation / decision function

$$\begin{array}{rccl} \mathcal{F}: & \mathbb{R}^M & \longrightarrow & \mathbb{P} = \mathbb{R}, \mathbb{C}, \mathbb{K} \\ & \boldsymbol{y} & \longmapsto & \mathcal{F}(\boldsymbol{y}) = \widehat{\boldsymbol{p}} \end{array}$$

- Cost function

$$\begin{array}{rccl} \mathcal{C}: & \mathbb{P} \times \mathbb{P} & \longrightarrow & \mathbb{R} \\ & (\boldsymbol{p}, \boldsymbol{p}') & \longmapsto & \mathcal{C}[\boldsymbol{p}, \boldsymbol{p}'] \end{array}$$

- Risk as a mean cost under the joint law

$$\rho(\mathcal{F}) = \mathbb{E}_{\boldsymbol{Y},\boldsymbol{P}} \{ \mathcal{C}(\boldsymbol{P}, \mathcal{F}(\boldsymbol{Y})) \}$$

- Optimal estimation / decision function

$$\mathcal{F}_{\text{opt}} = \arg\min_{\mathcal{F}} \rho(\mathcal{F})$$

Optimal estimation / decision function

Continuous parameters: estimation

- Quadratic cost

$$\mathcal{C}\,[\,\boldsymbol{p}\,,\,\boldsymbol{p}'\,] = \|\,\boldsymbol{p} - \boldsymbol{p}'\,\|^2$$

- Optimal estimation function $\equiv$ Posterior Mean

$$\widehat{\boldsymbol{p}} = \mathbb{E}_{\boldsymbol{P}|\boldsymbol{Y}}\,\{\,\boldsymbol{P}\,\} = \int_{\boldsymbol{p}} \boldsymbol{p}\;\pi(\boldsymbol{p}|\boldsymbol{y})\;\mathrm{d}\boldsymbol{p}$$

Discrete parameters: decision

- Binary cost

$$\mathcal{C}\,[\,\boldsymbol{p}\,,\,\boldsymbol{p}'\,] = 1 - \delta\,(\boldsymbol{p}\,,\,\boldsymbol{p}') = \begin{cases} 0 & \text{for correct decision} \\ 1 & \text{for erroneous decision} \end{cases}$$

- Optimal decision function $\equiv$ Posterior Maximizer

$$\widehat{\boldsymbol{p}} = \arg\max_{\boldsymbol{p}} \pi(\boldsymbol{p}|\boldsymbol{y})$$

Posterior estimate / decision and computations

- Numerical computations (convergent)

  1 – For $n = 1, 2, \ldots, N$, sample

  $$[\,\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta\,]^{(n)} \quad \text{under} \quad \pi(\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta | \boldsymbol{y})$$

  2 – Compute. . .

  2-a . . . empirical mean

  $$[\widehat{\boldsymbol{x}}_{1..K}, \widehat{\boldsymbol{\theta}}_{1..K}, \widehat{\gamma}_{1..K}, \widehat{\gamma}_\varepsilon, \widehat{\beta}] \simeq \frac{1}{N} \sum_n [\,\boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta\,]^{(n)}$$

  2-b . . . empirical marginal maximiser

  $$\widehat{\ell}_p \simeq \arg\max_k \frac{1}{N} \sum_n \delta(\ell_p^{(n)}, k)$$

- As a bonus:
  - Exploration and knowledge of the posterior
  - Posterior variances / probabilities and uncertainties
  - Marginal distributions
  - . . . and model selection

Posterior sampling

- Sampling $\pi(\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta | \boldsymbol{y})$
  - Impossible directly
  - Gibbs algorithm: sub-problems
    - Standard
    - Inverse cumulative density function
    - Metropolis-Hastings

- Gibbs loop: Draw iteratively
  - γ_ε under $\pi(\gamma_\varepsilon | \boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \beta)$
  - γ_k under $\pi(\gamma_k | \boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_l, l \neq k, \gamma_\varepsilon, \beta)$ for $k = 1, \ldots K$
  - $\boldsymbol{\ell}$ under $\pi(\boldsymbol{\ell} | \boldsymbol{y}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta)$
  - $\boldsymbol{x}_k$ under $\pi(\boldsymbol{x}_k | \boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_l, l \neq k, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta)$ for $k = 1, \ldots K$
  - $\boldsymbol{\theta}_k$ under $\pi(\boldsymbol{\theta}_k | \boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_l, l \neq k, \gamma_{1..K}, \gamma_\varepsilon, \beta)$ for $k = 1, \ldots K$
  - β under $\pi(\beta | \boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon)$

Sampling the noise parameter γ_ε

$$f(\boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta) = f(\boldsymbol{y}|\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \gamma_\varepsilon)$$
$$\Pr[\boldsymbol{\ell}|\beta] \prod f(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \gamma_k)\; f(\gamma_\varepsilon)\; f(\beta) \prod f(\boldsymbol{\theta}_k) \prod f(\gamma_k)$$

- Conditional density for γ_ε

$$\begin{aligned}
\pi(\gamma_\varepsilon|\star) &\propto f(\boldsymbol{y}|\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \gamma_\varepsilon)\; f(\gamma_\varepsilon) \\
&= \gamma_\varepsilon^M \exp\left[-\gamma_\varepsilon \|\boldsymbol{y} - \boldsymbol{T}\boldsymbol{H}\boldsymbol{x}\|^2\right]\; \gamma_\varepsilon^{a_0 - 1} \exp\left[-b_0 \gamma_\varepsilon\right] \mathbb{1}_+(\gamma_\varepsilon) \\
&= \gamma_\varepsilon^{a_0 + M - 1}\; \exp\left[-\gamma_\varepsilon \left(b_0 + \|\boldsymbol{y} - \boldsymbol{T}\boldsymbol{H}\boldsymbol{x}\|^2\right)\right]\; \mathbb{1}_+(\gamma_\varepsilon)
\end{aligned}$$

- It is a Gamma distribution

$$\gamma_\varepsilon \sim \mathcal{G}(a, b)$$

$$\begin{cases} a &= a_0 + M \\ b &= b_0 + \|\boldsymbol{y} - \boldsymbol{T}\boldsymbol{H}\boldsymbol{x}\|^2 \end{cases}$$

Sampling the texture scale parameters γ_k

$$f(\boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta) = f(\boldsymbol{y}|\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \gamma_\varepsilon)$$
$$\Pr[\boldsymbol{\ell}|\beta] \prod f(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \gamma_k)\ f(\gamma_\varepsilon)\ f(\beta) \prod f(\boldsymbol{\theta}_k) \prod f(\gamma_k)$$

- Conditional density for γ_k

$$\begin{aligned}\pi(\gamma_k|\star) &\propto f(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \gamma_k)\ f(\gamma_k) \\ &= \gamma_k^P \exp\left[-\gamma_k\, \boldsymbol{x}_k^\dagger \boldsymbol{P}_x(\boldsymbol{\theta}_k)\boldsymbol{x}_k\right]\ \gamma_k^{a_0-1} \exp\left[-b_0\gamma_k\right] \mathbb{1}_+(\gamma_k) \\ &= \gamma_k^{a_0+P-1}\ \exp\left[-\gamma_k\left(b_0 + \boldsymbol{x}_k^\dagger \boldsymbol{P}_x(\boldsymbol{\theta}_k)\boldsymbol{x}_k\right)\right]\ \mathbb{1}_+(\gamma_k)\end{aligned}$$

- It is also a Gamma distribution

$$\gamma_k \sim \mathcal{G}(a, b)$$

$$\begin{cases} a = a_0 + P \\ b = b_0 + \boldsymbol{x}_k^\dagger \boldsymbol{P}_x(\boldsymbol{\theta}_k)\boldsymbol{x}_k = b_0 + \sum \lambda_p(\boldsymbol{\theta}_k)\, |\mathring{x}_p|^2 \end{cases}$$

Sampling the labels $\boldsymbol{\ell}$ (1)

$$f(\boldsymbol{y},\boldsymbol{\ell},\boldsymbol{x}_{1..K},\boldsymbol{\theta}_{1..K},\gamma_{1..K},\gamma_\varepsilon,\beta) = f(\boldsymbol{y}|\boldsymbol{\ell},\boldsymbol{x}_{1..K},\gamma_\varepsilon)$$

$$\Pr\left[\boldsymbol{\ell}|\beta\right] \prod f(\boldsymbol{x}_k|\boldsymbol{\theta}_k,\gamma_k)\; f(\gamma_\varepsilon)\; f(\beta) \prod f(\boldsymbol{\theta}_k) \prod f(\gamma_k)$$

- Conditional probability for the set of labels $\boldsymbol{\ell}$

$$\begin{aligned}\Pr\left[\boldsymbol{\ell}|\star\right] &\propto f(\boldsymbol{y}|\boldsymbol{\ell},\boldsymbol{x}_{1..K},\gamma_\varepsilon)\,\Pr\left[\boldsymbol{\ell}|\beta\right] \\ &\propto \exp\left[-\gamma_\varepsilon\|\boldsymbol{y}-\boldsymbol{T}\boldsymbol{H}\sum \boldsymbol{S}_k(\boldsymbol{\ell})\,\boldsymbol{x}_k\|^2\right]\;\exp\left[\beta\nu(\boldsymbol{\ell})\right]\end{aligned}$$

- Conditional categorical probability for one label ℓ_p

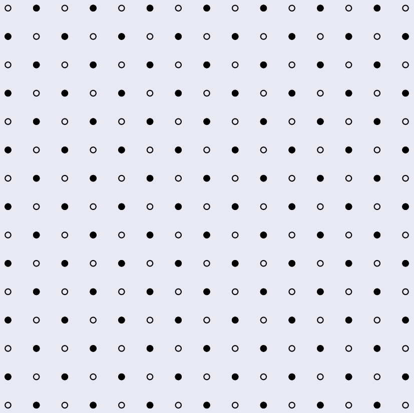
$$\pi\left[\ell_p = k|\star\right] \propto \begin{cases}\text{observed:} & \exp\left[-\gamma_\varepsilon|y_p - \;\cdots\; |^2 + \beta N_{p,k}\right] \\ \text{unobserved:} & \exp\left[\beta N_{p,k}\right]\end{cases}$$

- Joint structure: no convolution case
  - Conditional independance
  - Parallel sampling (two subsets: ebony and ivory)

Sampling the labels ℓ (2)

Markov field: conditional independance

- No convolution case



- Including convolution: More complex neighbour system...

Sampling the textured images $\boldsymbol{x}_k$ (1)

$$f(\boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta) = f(\boldsymbol{y} | \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \gamma_\varepsilon)$$
$$\Pr[\boldsymbol{\ell} | \beta] \prod f(\boldsymbol{x}_k | \boldsymbol{\theta}_k, \gamma_k)\ f(\gamma_\varepsilon)\ f(\beta) \prod f(\boldsymbol{\theta}_k) \prod f(\gamma_k)$$

- Conditional density for the textured image $\boldsymbol{x}_k$

$$\pi(\boldsymbol{x}_k | \star) \propto f(\boldsymbol{y} | \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \gamma_\varepsilon)\, f(\boldsymbol{x}_k | \boldsymbol{\theta}_k, \gamma_k)$$
$$\propto \exp\left[-\gamma_\varepsilon \|\boldsymbol{y} - \boldsymbol{T}\boldsymbol{H} \sum \boldsymbol{S}_k(\boldsymbol{\ell})\, \boldsymbol{x}_k\|^2\right] \exp\left[-\gamma_k\, \boldsymbol{x}_k^\dagger \boldsymbol{P}_x(\boldsymbol{\theta}_k) \boldsymbol{x}_k\right]$$

- Gaussian distribution

$$\begin{aligned} \boldsymbol{C}_k^{-1} &= \gamma_\varepsilon \boldsymbol{S}_k(\boldsymbol{\ell}) \boldsymbol{H}^\dagger \boldsymbol{T}^{\mathrm{t}} \boldsymbol{T} \boldsymbol{H} \boldsymbol{S}_k(\boldsymbol{\ell}) + \gamma_k \boldsymbol{P}_x(\boldsymbol{\theta}_k) \\ \boldsymbol{m}_k &= \gamma_\varepsilon \boldsymbol{C}_k \boldsymbol{S}_k(\boldsymbol{\ell}) \boldsymbol{H}^\dagger \boldsymbol{T}^{\mathrm{t}} \bar{\boldsymbol{y}}_k \\ \bar{\boldsymbol{y}}_k &= \boldsymbol{y} - \boldsymbol{T}\boldsymbol{H} \sum_{k' \neq k} \boldsymbol{S}_{k'}(\boldsymbol{\ell})\, \boldsymbol{x}_{k'} \end{aligned}$$

Sampling the textured images $\boldsymbol{x}_k$ (2)

- Gaussian distribution

$$\begin{aligned} \boldsymbol{C}_k^{-1} &= \gamma_\varepsilon \boldsymbol{S}_k(\boldsymbol{\ell}) \boldsymbol{H}^\dagger \boldsymbol{T}^{\mathrm{t}} \boldsymbol{T} \boldsymbol{H} \boldsymbol{S}_k(\boldsymbol{\ell}) + \gamma_k \boldsymbol{P}_x(\boldsymbol{\theta}_k) \\ \boldsymbol{m}_k &= \gamma_\varepsilon \boldsymbol{C}_k \boldsymbol{S}_k(\boldsymbol{\ell}) \boldsymbol{H}^\dagger \boldsymbol{T}^{\mathrm{t}} \bar{\boldsymbol{y}}_k \end{aligned}$$

- Standard approaches
  - Covariance factorization $\boldsymbol{C} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{t}}$
  - Precision factorisation $\boldsymbol{C}^{-1} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{t}}$
  - Diagonalization $\boldsymbol{C} = \boldsymbol{P}\boldsymbol{\Delta}\boldsymbol{P}^{\mathrm{t}}$ et $\boldsymbol{C}^{-1} = \boldsymbol{P}\boldsymbol{\Delta}^{-1}\boldsymbol{P}^{\mathrm{t}}$
  - Parallel Gibbs sampling
- Large dimension
  - Linear system solvers
  - Optimization: Quadratic criterion minimization
  - **Perturbation – Optimization**
    1. P: produce a adequately perturbed criterion
    2. O: minimize the perturbed criterion
  - . . .

Sampling texture parameters $\boldsymbol{\theta}_k$ (1)

$$f(\boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta) = f(\boldsymbol{y}|\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \gamma_\varepsilon)$$
$$\Pr[\boldsymbol{\ell}|\beta] \prod f(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \gamma_k)\; f(\gamma_\varepsilon)\; f(\beta) \prod f(\boldsymbol{\theta}_k) \prod f(\gamma_k)$$

- Conditional density for the texture parameters $\boldsymbol{\theta}_k$

$$\begin{aligned}\pi(\boldsymbol{\theta}_k|\star) &\propto f(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \gamma_k)\, f(\boldsymbol{\theta}_k) \\ &\propto \exp\left[-\gamma_k\, \boldsymbol{x}_k^\dagger \boldsymbol{P}_x(\boldsymbol{\theta}_k)\boldsymbol{x}_k\right] \;\; \mathcal{U}_{[\boldsymbol{\theta}_k^{\mathrm{m}}, \boldsymbol{\theta}_k^{\mathrm{M}}]}(\boldsymbol{\theta}_k) \\ &\propto \prod \lambda_p(\boldsymbol{\theta}_k)\;\; \exp\left[-\gamma_x\, \lambda_p(\boldsymbol{\theta}_k)\, |\mathring{x}_p|^2\right] \;\; \mathcal{U}_{[\boldsymbol{\theta}_k^{\mathrm{m}}, \boldsymbol{\theta}_k^{\mathrm{M}}]}(\boldsymbol{\theta}_k)\end{aligned}$$

- Metropolis-Hastings: Propose and accept or not
  - Independant or not, *e.g.*, random walk
  - Metropolis-adjusted Langevin algorithm
  - **Directional algorithms**
    - Gradient
    - Hessien, **Fisher matrix**
    - ...
  - ...

Sampling texture parameters $\boldsymbol{\theta}_k$ (2)

- Principe de Metropolis-Hastings indépendant
  - Simuler θ sous f ...
  - ... en simulant θ sous g

- Algorithme itératif produisant des $\theta^{(n)}$
  - Initialiser
  - Itérer, pour $n = 1, 2, \ldots,$
    - Simuler θ_{p} sous la loi $g(\theta)$
    - Calculer la probabilité

$$\alpha = \min\left(1\ ;\ \frac{f(\theta_{\mathrm{p}})}{f(\theta^{(n-1)})}\frac{g(\theta^{(n-1)})}{g(\theta_{\mathrm{p}})}\right)$$

    - Acceptation / conservation

$$\begin{cases} \theta^{(n)} = \theta_{\mathrm{p}} & \text{accepte avec la probabilité } \alpha \\ \theta^{(n)} = \theta^{(n-1)} & \text{conserve avec la probabilité } 1-\alpha \end{cases}$$

Sampling the correlation parameter β

$$f(\boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \boldsymbol{\theta}_{1..K}, \gamma_{1..K}, \gamma_\varepsilon, \beta) = f(\boldsymbol{y}|\boldsymbol{\ell}, \boldsymbol{x}_{1..K}, \gamma_\varepsilon) \\ \Pr[\boldsymbol{\ell}|\beta] \prod f(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \gamma_k)\ f(\gamma_\varepsilon)\ f(\beta) \prod f(\boldsymbol{\theta}_k) \prod f(\gamma_k)$$

- Conditional density for the correlation parameter β

$$\begin{aligned} \pi(\beta|\star) &\propto \Pr[\boldsymbol{\ell}|\beta]\ f(\beta) \\ &\propto C(\beta)^{-1}\ \exp[\beta\nu(\boldsymbol{\ell})]\ \mathcal{U}_{[0,B_0]}(\beta) \end{aligned}$$

- Sampling itself
  - Partition function $C(\beta)$ pre-computed (previous part)
  - Conditional cdf $F(\beta)$ through numerical integration / interpolation
  - Inverse the cdf to generate a sample

$$\begin{aligned} &\text{Sample } u \sim \mathcal{U}_{[0,1]}(u) \\ &\text{Compute } \beta = F^{-1}(u) \end{aligned}$$

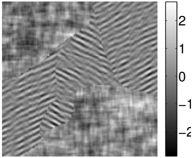
Outline

- Image model
  - Textured images, orientation
  - Piecewise homogeneous images
- Observation system model
  - Convolution and missing data
  - Noise
- Hierarchical model
  - Conditional dependancies / independancies
  - Joint distribution
- Estimation / decision strategy and computations
  - Cost, risk and optimality $\rightsquigarrow$ Posterior distribution and estimation
  - Convergent computations: stochastic sampler $\oplus$ empirical estimates
    - Gibbs loop
    - Inverse cumulative density function
    - Metropolis-Hastings
- First numerical assessment
  - Behaviour, convergence,...
  - Labels, texture parameters and hyperparameters
  - Quantification of errors

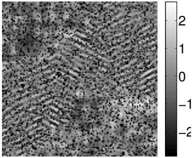
Numerical illustration: problem

- A first toy example

True label $\boldsymbol{\ell}^{\star}$



True image $\boldsymbol{x}^{\star}$

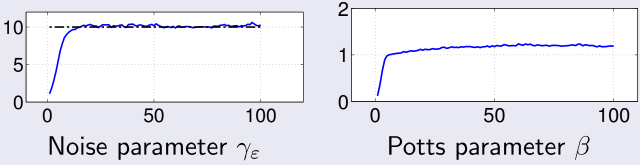


Observation $\boldsymbol{y}$

- Parameters
  - $P = 256 \times 256$, $K = 3$
  - No convolution here
  - Missing : 20 %
  - Noise level: $\gamma_{\varepsilon} = 10$ (standard deviation: 0.3, SNR: 10dB)

Numerical results: parameters

Simulated chains



Quantitative assessment

Parameter	γ_ε	β
True value	10.0	–
Estimate	10.2	1.19

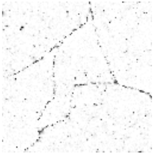
Computation time: one minute

Numerical results: classification

True label $\ell^\star$

Estimated $\widehat{\ell}$

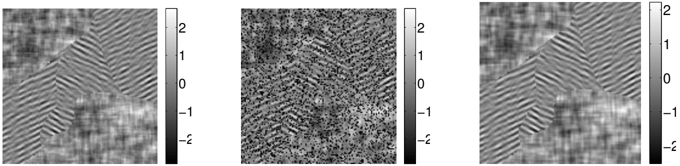
Misclassification



Probability

- Performances
    - Correct classification, including unoberved pixels
    - Only about 150 misclassifications, *i.e.*, less than 1%
    - Remark: maximizers of the marginal posteriors
- Quantification of errors
    - Probabilities (marginal)
    - Indication/warning of misclassification

Numerical results: restored image



True $\boldsymbol{x}^\star$ Observation $\boldsymbol{y}$ Estimated $\widehat{\boldsymbol{x}}$

- Performances
  - Correct restauration of textures
  - Correct restauration of edges (thanks to correct classification)
  - Including interpolation of missing pixels
- Quantification of errors
  - ...onging work...
  - Posterior standard deviation, credibility intervals

Conclusion

Synthesis

- Addressed problem: segmentation
  - Piecewise textured images
  - Triple difficulty: missing data + noise + convolution
  - Including all hyperparameter estimation
- Bayesian approach
  - Optimal estimation / decision
  - Convergent computation
- Numerical evaluation

Perspectives

- Ongoing: inversion-segmentation (*e.g.*, convolution, Radon,. . .)
- Non-Gaussian noise: Latent variables (*e.g.*, Cauchy), Poisson,. . .
- Correlated, structured, textured noise
- Myopic problem: estimation of instrument parameters
- Model selection, choice of K
- Application to real data