Inverse problems, Deconvolution and Parametric Estimation

Ali Mohammad-Djafari
Laboratoire des Signaux et Systèmes,
UMR8506 CNRS-SUPELEC-UNIV PARIS SUD 11
SUPELEC, 91192 Gif-sur-Yvette, France
http://lss.supelec.free.fr

Email: djafari@lss.supelec.fr
http://djafari.free.fr

Contents

- Invese problems examples: Deconvolution, Image restoration, Image reconstruction, Fourier synthesis, ...
- Classification of Invesion methods: Analytical, Parametric and Non Parametric algebraic methods
- Regularization theory
- Bayesian inference for invese problems
- Full Bayesian with hyperparameter estimation
- Two main steps in Bayesian approach:
  Prior modeling and Bayesian computation
- Priors which enforce sparsity
  - Heavy tailed: Double Exponential, Generalized Gaussian, ...
  - Mixture models: Mixture of Gaussians, Student-t, ...
  - Gauss-Markov-Potts
- Computational tools:
  MCMC and Variational Bayesian Approximation
- Some results and applications
  - X ray Computed Tomography, Microwave and Ultrasound imaging, Sattelite Image separation, Hyperspectral image processing, Spectrometry, CMB, ...

Contents

- Invese problems examples
- Classification of Invesion methods: Analytical, Parametric and Non Parametric algebraic methods
- Regularization theory
- Bayesian inference for invese problems
- Full Bayesian with hyperparameter estimation
- Two main steps in Bayesian approach:
  Prior modeling and Bayesian computation
- Priors which enforce sparsity
  - Heavy tailed: Double Exponential, Generalized Gaussian, ...
  - Mixture models: Mixture of Gaussians, Student-t, ...
  - Gauss-Markov-Potts
- Computational tools:
  MCMC and Variational Bayesian Approximation
- Applications: X ray Computed Tomography, Microwave and Ultrasound imaging, ...

Direct and indirect observation

- Direct observation of a few quantities are possible: length, time, electrical charge, number of particles
- For many others, we only can measure them by transforming them.
  Example: Thermometer transforms variation of temeperature f to variation of length g.

- Relating measurable quantity g to the desired quantity f is called Forward modeling: $g = \mathcal{H}(f)$.
- Predicting the measurements g if we knew the desired quantity f and the measurement system is called Forward problem.
- Infering on the desired quantity f from the measurement g is called Inverse problem.

Inverse problems : 3 main examples

- Example 1:
  Measuring variation of temperature with a therometer
  - $f(t)$ variation of temperature over time
  - $g(t)$ variation of length of the liquid in thermometer
- Example 2: Seeing outside of a body: Making an image using a camera, a microscope or a telescope
  - $f(x, y)$ real scene
  - $g(x, y)$ observed image
- Example 3: Seeing inside of a body: Computed Tomography usng X rays, US, Microwave, etc.
  - $f(x, y)$ a section of a real 3D body $f(x, y, z)$
  - $g_\phi(r)$ a line of observed radiographe $g_\phi(r, z)$

- Example 1: Deconvolution
- Example 2: Image restoration
- Example 3: Image reconstruction

Measuring variation of temperature with a therometer

- $f(t)$ variation of temperature over time
- $g(t)$ variation of length of the liquid in thermometer

- Forward model: Convolution

$$g(t) = \int f(t')\, h(t-t')\, \mathrm{d}t' + \epsilon(t)$$

$h(t)$: impulse response of the measurement system

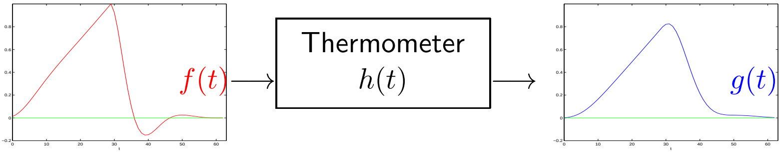
- Inverse problem: Deconvolution

Given the forward model $\mathcal{H}$ (impulse response $h(t)$))
and a set of data $g(t_i), i = 1, \cdots, M$
find $f(t)$

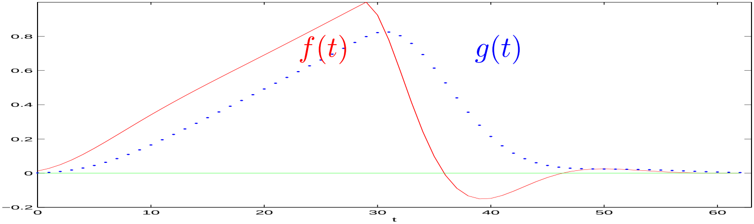
Measuring variation of temperature with a therometer

Forward model: Convolution

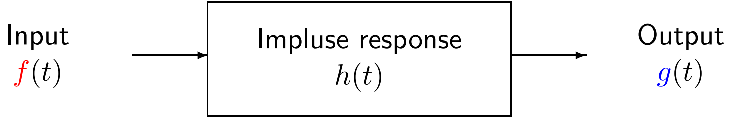
$$g(t) = \int f(t')\, h(t-t')\, \mathrm{d}t' + \epsilon(t)$$



Inversion: Deconvolution



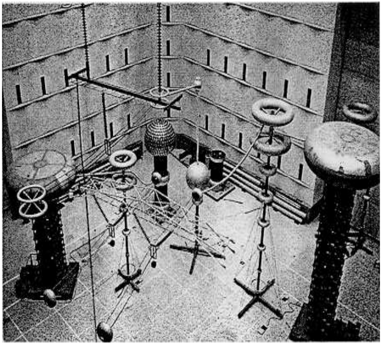
Instrumentation



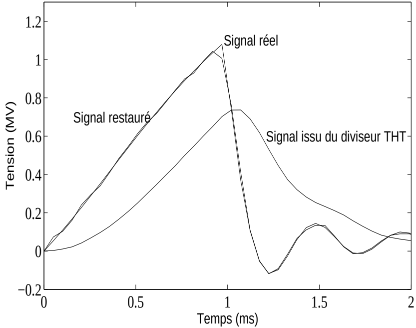
- Ideal Instrument $\quad g(t) = f(t) \quad$ does not exist.
- A linear and time invariant instrument is characterized by its impulse response $\quad h(t)$.
- Ideal Instrument $\quad h(t) = \delta(t) \quad$ does not exist.
- Forward problem: $\quad f(t),\, h(t) \longrightarrow g(t) = h(t) * f(t)$
- Two linked problems in instrumentation:
  - Inversion: $\quad g(t),\, h(t) \longrightarrow f(t)$
  - Identification: $\quad g(t),\, f(t) \longrightarrow h(t)$

Ex1: Isolators resistivity against lightning strike

An instrument giving the possibility to apply very high voltage to simulate lightning strike

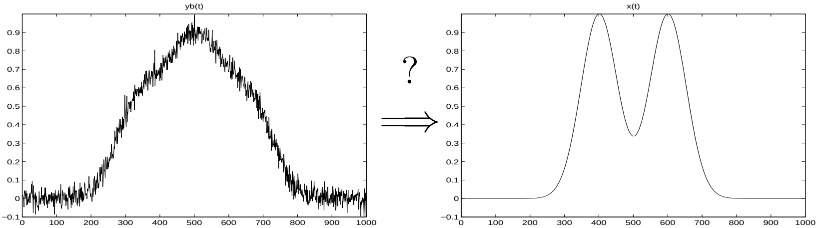


EDF– Les Renardières

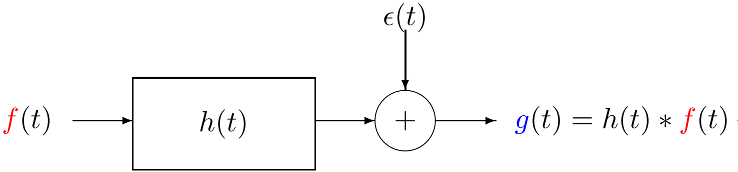


Real and Estimated

Ex2: Radio-astronomy

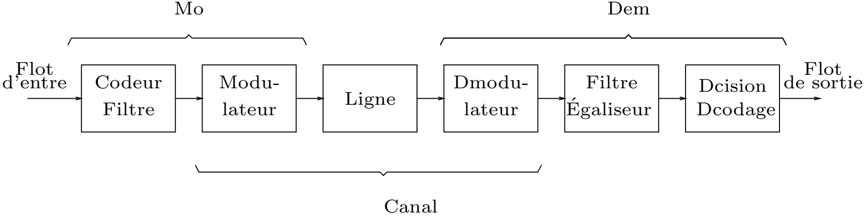


Forward model:

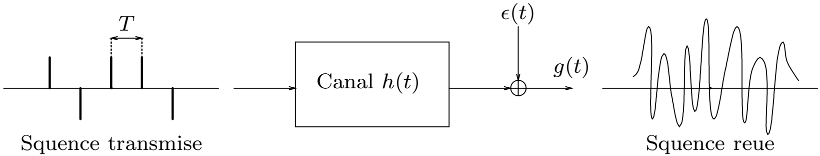


Telecommunication: transmission channel compensation

- Data transmission System

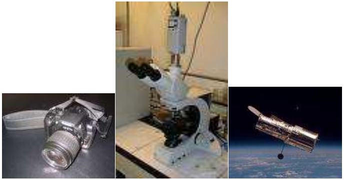


- Channel Model: convolution + noise



Seeing outside of a body: Making an image with a camera, a microscope or a telescope

- $f(x,y)$ real scene
- $g(x,y)$ observed image



- Forward model: Convolution

$$g(x,y) = \iint f(x',y')\, h(x-x', y-y')\, \mathrm{d}x'\, \mathrm{d}y' + \epsilon(x,y)$$

  $h(x,y)$: Point Spread Function (PSF) of the imaging system

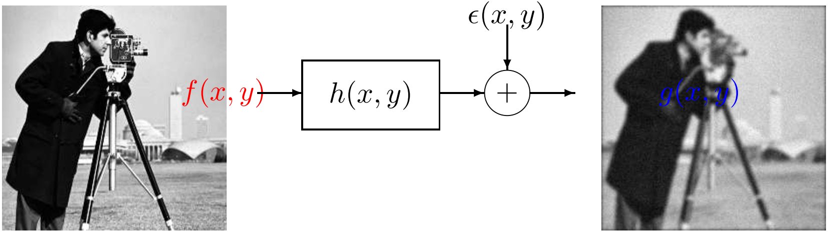
- Inverse problem: Image restoration

  Given the forward model $\mathcal{H}$ (PSF $h(x,y)$)
  and a set of data $g(x_i, y_i), i = 1, \cdots, M$
  find $f(x,y)$

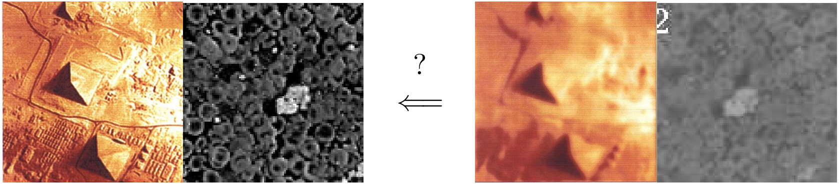
Making an image with an unfocused camera

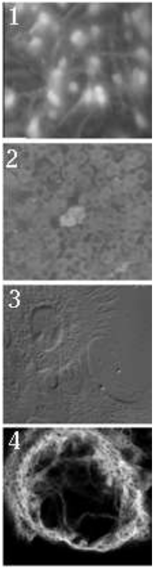
Forward model: 2D Convolution

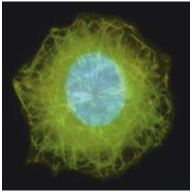
$$g(x,y) = \iint f(x',y')\, h(x-x',y-y')\, \mathrm{d}x'\, \mathrm{d}y' + \epsilon(x,y)$$

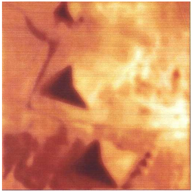


Inversion: Image Deconvolution or Restoration



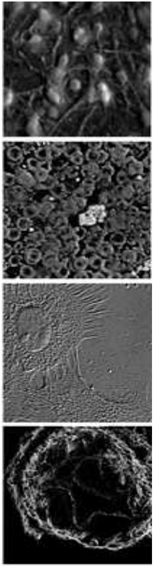


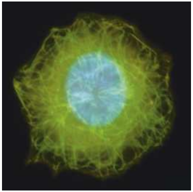


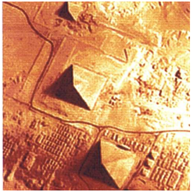


?

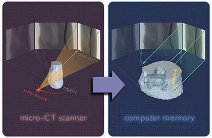
$\Longrightarrow$







Seeing inside of a body: Computed Tomography



- $f(x, y)$ a section of a real 3D body $f(x, y, z)$
- $g_\phi(r)$ a line of observed radiographe $g_\phi(r, z)$
- Forward model:
  Line integrals or Radon Transform

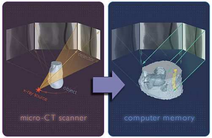
$$\begin{aligned} g_\phi(r) &= \int_{L_{r,\phi}} f(x, y) \, \mathrm{d}l + \epsilon_\phi(r) \\ &= \iint f(x, y) \, \delta(r - x\cos\phi - y\sin\phi) \, \mathrm{d}x \, \mathrm{d}y + \epsilon_\phi(r) \end{aligned}$$

- Inverse problem: Image reconstruction

  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x, y)$

Making an image of the interior of a body

- $f(x,y)$ a section of a real 3D body $f(x,y,z)$
- $g_\phi(r)$ a line of observed radiographe $g_\phi(r,z)$



- Forward model:
  Line integrals or Radon Transform

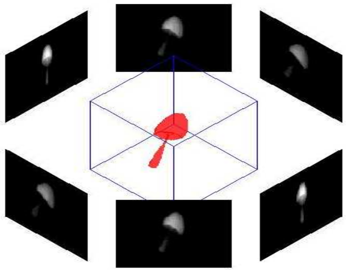
$$
\begin{aligned}
g_\phi(r) &= \int_{L_{r,\phi}} f(x,y)\, \mathrm{d}l + \epsilon_\phi(r) \\
&= \iint f(x,y)\, \delta(r - x\cos\phi - y\sin\phi)\, \mathrm{d}x\, \mathrm{d}y + \epsilon_\phi(r)
\end{aligned}
$$

- Inverse problem: Image reconstruction

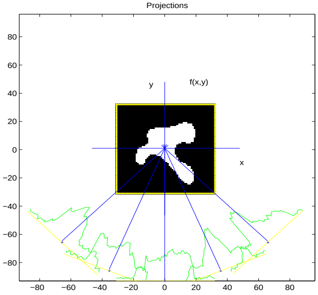
  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x,y)$

2D and 3D Computed Tomography

3D



2D

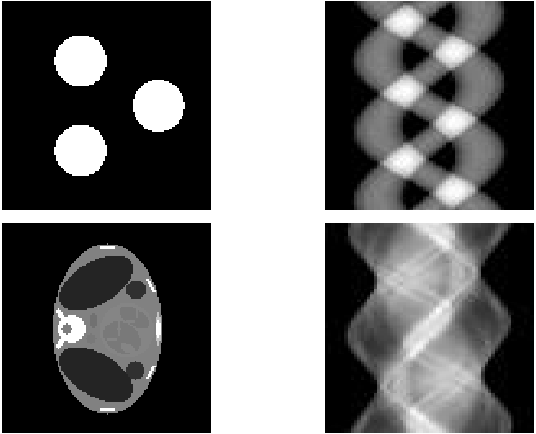


$$g_\phi(r_1,r_2)=\int_{\mathcal{L}_{r_1,r_2,\phi}} f(x,y,z)\,\mathrm{d}l \qquad g_\phi(r)=\int_{\mathcal{L}_{r,\phi}} f(x,y)\,\mathrm{d}l$$

Forward probelm: $f(x,y)$ or $f(x,y,z) \longrightarrow g_\phi(r)$ or $g_\phi(r_1,r_2)$

Inverse problem: $g_\phi(r)$ or $g_\phi(r_1,r_2) \longrightarrow f(x,y)$ or $f(x,y,z)$

Computed Tomography: Radon Transform



Forward: $f(x,y) \longrightarrow g(r,\phi)$

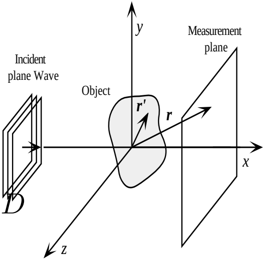
Inverse: $f(x,y) \longleftarrow g(r,\phi)$

Microwave or ultrasound imaging

Measurs: diffracted wave by the object $g(\boldsymbol{r}_i)$
Unknown quantity: $f(\boldsymbol{r}) = k_0^2(n^2(\boldsymbol{r}) - 1)$
Intermediate quantity : $\phi(\boldsymbol{r})$

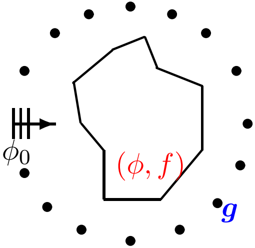
$$g(\boldsymbol{r}_i) = \iint_D G_m(\boldsymbol{r}_i, \boldsymbol{r}')\phi(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \;\; \boldsymbol{r}_i \in S$$

$$\phi(\boldsymbol{r}) = \phi_0(\boldsymbol{r}) + \iint_D G_o(\boldsymbol{r}, \boldsymbol{r}')\phi(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \;\; \boldsymbol{r} \in D$$



Born approximation $(\phi(\boldsymbol{r}') \simeq \phi_0(\boldsymbol{r}'))$ **):**

$$g(\boldsymbol{r}_i) = \iint_D G_m(\boldsymbol{r}_i, \boldsymbol{r}')\phi_0(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \;\; \boldsymbol{r}_i \in S$$



Discretization :

$$\begin{cases} \boldsymbol{g} = \boldsymbol{G}_m \boldsymbol{F} \boldsymbol{\phi} \\ \boldsymbol{\phi} = \boldsymbol{\phi}_0 + \boldsymbol{G}_o \boldsymbol{F} \boldsymbol{\phi} \end{cases} \longrightarrow \begin{cases} \boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) \\ \text{with} \quad \boldsymbol{F} = \mathrm{diag}(\boldsymbol{f}) \\ \boldsymbol{H}(\boldsymbol{f}) = \boldsymbol{G}_m \boldsymbol{F} (\boldsymbol{I} - \boldsymbol{G}_o \boldsymbol{F})^{-1} \boldsymbol{\phi}_0 \end{cases}$$

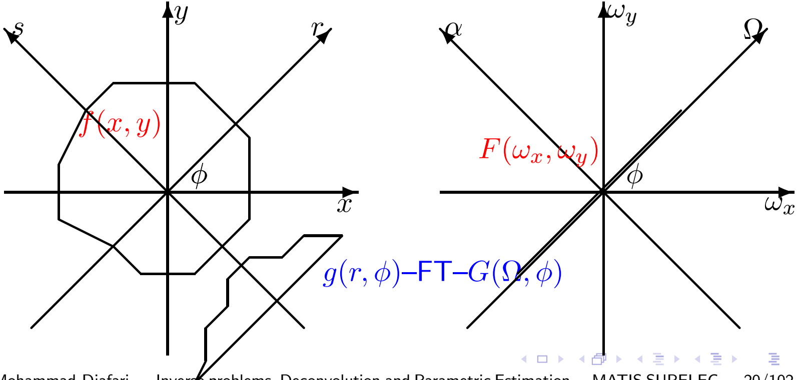
Fourier Synthesis in X ray Tomography

$$g(r,\phi) = \iint f(x,y)\, \delta(r - x\cos\phi - y\sin\phi)\, \mathrm{d}x\, \mathrm{d}y$$

$$G(\Omega,\phi) = \int g(r,\phi) \exp\{-j\Omega r\}\, \mathrm{d}r$$

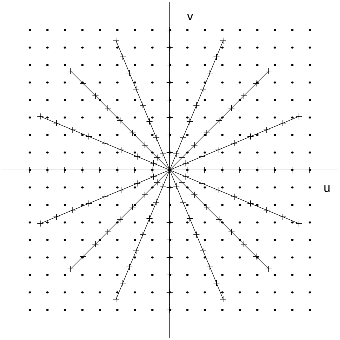
$$F(\omega_x,\omega_y) = \iint f(x,y) \exp\{-j\omega_x x, \omega_y y\}\, \mathrm{d}x\, \mathrm{d}y$$

$$F(\omega_x,\omega_y) = G(\Omega,\phi) \quad \text{for} \quad \omega_x = \Omega\cos\phi \quad \text{and} \quad \omega_y = \Omega\sin\phi$$

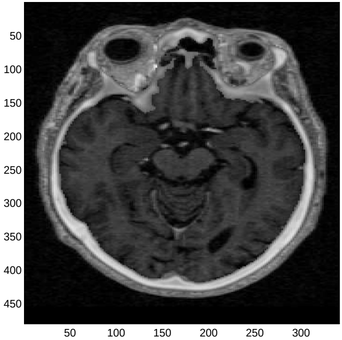


Fourier Synthesis in X ray tomography

$$G(\omega_x, \omega_y) = \iint f(x, y) \exp\{-j(\omega_x x + \omega_y y)\} \, \mathrm{d}x \, \mathrm{d}y$$

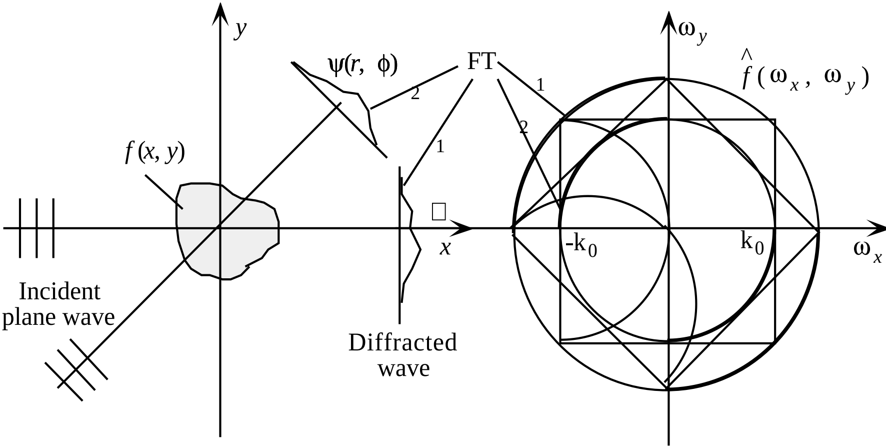


$$? \Longrightarrow$$



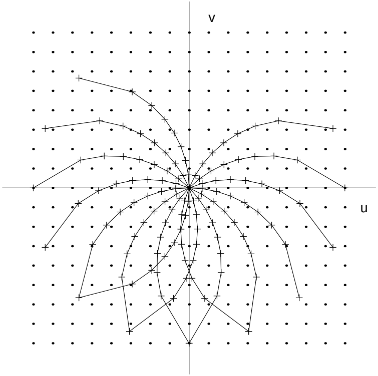
Forward problem: Given $f(x, y)$ compute $G(\omega_x, \omega_y)$
Inverse problem: Given $G(\omega_x, \omega_y)$ on those lines estimate $f(x, y)$

Fourier Synthesis in Diffraction tomography



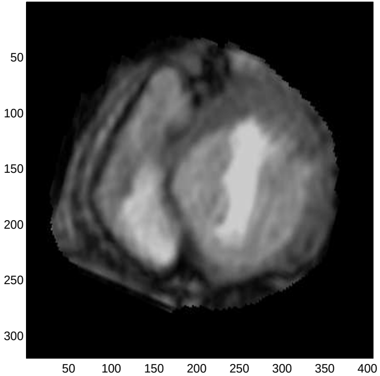
Fourier Synthesis in Diffraction tomography

$$G(\omega_x, \omega_y) = \iint f(x, y) \exp\{-j(\omega_x x + \omega_y y)\} \, \mathrm{d}x \, \mathrm{d}y$$



$?$

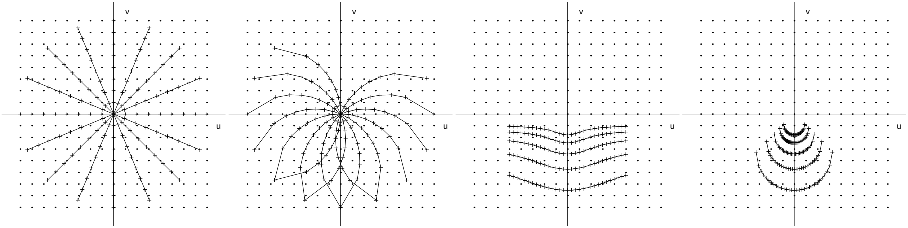
$\Longrightarrow$



Forward problem: Given $f(x, y)$ compute $G(\omega_x, \omega_y)$
Inverse problem : Given $G(\omega_x, \omega_y)$ on those semi cercles estimate $f(x, y)$

Fourier Synthesis in different imaging systems

$$G(\omega_x, \omega_y) = \iint f(x, y) \exp\{-j(\omega_x x + \omega_y y)\} \, \mathrm{d}x \, \mathrm{d}y$$



X ray Tomography | Diffraction | Eddy current | SAR & Radar

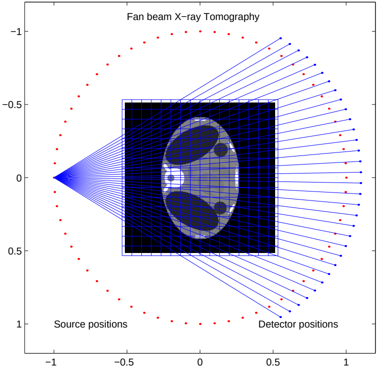
Forward problem: Given $f(x, y)$ compute $G(\omega_x, \omega_y)$
Inverse problem : Given $G(\omega_x, \omega_y)$ on those algebraic lines, cercles or curves, estimate $f(x, y)$

Invers Problems: other examples and applications

- X ray, Gamma ray Computed Tomography (CT)
- Microwave and ultrasound tomography
- Positron emission tomography (PET)
- Magnetic resonance imaging (MRI)
- Photoacoustic imaging
- Radio astronomy
- Geophysical imaging
- Non Destructive Evaluation (NDE) and Testing (NDT) techniques in industry
- Hyperspectral imaging
- Earth observation methods (Radar, SAR, IR, ...)
- Survey and tracking in security systems

Computed tomography (CT)

A Multislice CT Scanner



$$g(s_i) = \int_{L_i} f(\boldsymbol{r}) \; \mathrm{d}l_i + \epsilon(s_i)$$

Discretization

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \epsilon$$

Positron emission tomography (PET)

Magnetic resonance imaging (MRI)

Nuclear magnetic resonance imaging (NMRI), Para-sagittal MRI of the head

Radio astronomy (interferometry imaging systems)

The Very Large Array in New Mexico, an example of a radio telescope.

General inverse problems

$$\mathcal{H} \quad (\quad \boldsymbol{g} \quad , \quad \boldsymbol{f} \quad , \quad \boldsymbol{z} \quad , \quad \boldsymbol{\epsilon} \quad) \quad = \quad 0$$

model	measured data	unknown quantity	intermediate quantity	errors and noise

Particular cases:

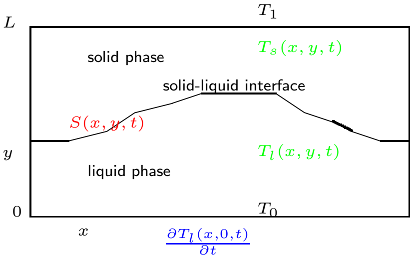
- Implicite model linking $\boldsymbol{f}$ and $\boldsymbol{z}$: $\begin{cases} \boldsymbol{g} = \mathcal{H}_1(\boldsymbol{f}, \boldsymbol{z}) + \boldsymbol{\epsilon} \\ \mathcal{H}_2(\boldsymbol{f}, \boldsymbol{z}) = 0 \end{cases}$
- Simple non linear model: $\boldsymbol{g} = \mathcal{H}(\boldsymbol{f}) + \boldsymbol{\epsilon}$
- Linear model with additive noise: $\boldsymbol{g} = \mathcal{H}\boldsymbol{f} + \boldsymbol{\epsilon}$

Time evolution of liquid-solid fusion interface

Solid : $\frac{\partial T_s}{\partial t} = \alpha_s \left(\frac{\partial^2 T_s}{\partial x^2} + \frac{\partial^2 T_s}{\partial x^2} \right)$

Liquid : $\frac{\partial T_l}{\partial t} = \alpha_l \left(\frac{\partial^2 T_l}{\partial x^2} + \frac{\partial^2 T_l}{\partial x^2} \right)$

Energy conservation $k_s \frac{\partial T_s}{\partial n} - k_l \frac{\partial T_l}{\partial n} = \rho L_f \vec{v}.\vec{n}$



$\vec{v}$: speed of solid-liquid interface

$\vec{n}$: normal vector on the interface

Observed quantity : heat flux on the heating surface $\frac{\partial T_s(x,0,t)}{\partial t}$

Unknown quantity : solid-liquid surface evolution $S(x,y,t)$

Intermediate uknown quantity: temperature field $T_s(x,y,t)$ et $T_l(x,y,t)$

General formulation of inverse problems

- General non linear inverse problems:

$$g(\boldsymbol{s}) = [\mathcal{H} f(\boldsymbol{r})](\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \quad \boldsymbol{r} \in \mathcal{R}, \quad \boldsymbol{s} \in \mathcal{S}$$

- Linear models:

$$g(\boldsymbol{s}) = \int f(\boldsymbol{r})\, h(\boldsymbol{r}, \boldsymbol{s})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s})$$

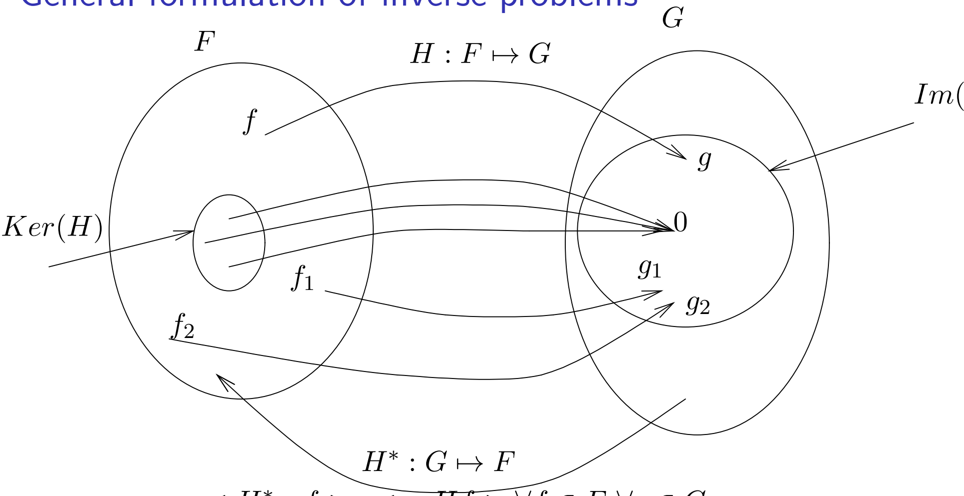
If $h(\boldsymbol{r}, \boldsymbol{s}) = h(\boldsymbol{r} - \boldsymbol{s}) \longrightarrow$ Convolution.

- Discrete data:

$$g(\boldsymbol{s}_i) = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s}_i), \quad i = 1, \cdots, m$$

- Inversion: Given the forward model $\mathcal{H}$ and the data $\boldsymbol{g} = \{g(\boldsymbol{s}_i), i = 1, \cdots, m)\}$ estimate $f(\boldsymbol{r})$
- Well-posed and Ill-posed problems (Hadamard): existance, uniqueness and stability
- Need for prior information

General formulation of inverse problems



$$< H^* g, f >=< g, Hf > \forall f \in F, \forall g \in G$$

Inverse problems scientific communities

Two communities working on Inverse problems:

- Mathematical departments:
  Analytical methods: Existance and Uniqueness
  Differential equations, PDE

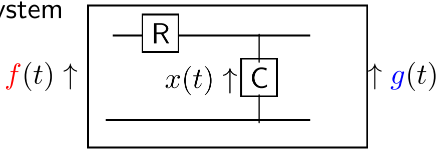
- Engineering and Computer sciences:
  Algebraic methods: Discretization, Uniqueness and Stability
  Integral equations, Discretization using Moments method, Galerkin, ...

Two examples:

- Deconvolution: Inverse filtering and Wiener filtering
- X ray Computed Tomography: Radon transform:
  Direct Inversion or Filtered Backprojection methods

Differential Equation, State Space and Input-Output

A simple electric system



$$f(t) = R\,i(t) + v_c(t) = RC\,\frac{\partial x(t)}{\partial t} + x(t), \quad RC = 1$$

- Differential Equation Modelling

$$\frac{\partial x(t)}{\partial t} + x(t) = f(t), \qquad x(t) = g(t)$$

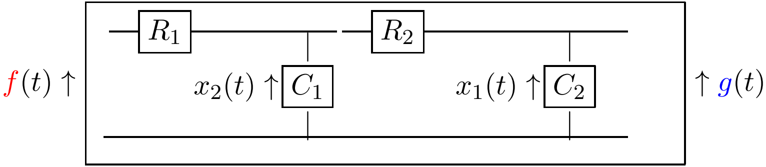
- State Space Modelling

$$\begin{cases} \frac{\partial x(t)}{\partial t} & = -x(t) + f(t) \\ g(t) & = x(t) \end{cases}$$

- Input-Output Modelling

$$\begin{cases} \frac{\partial x(t)}{\partial t} = -x(t) + f(t) \\ g(t) \;= x(t) \end{cases} \rightarrow \begin{cases} pX(p) = -X(p) + F(p) \rightarrow X(p) = \frac{1}{p+1} F(p) \\ g(t) = x(t) = h(t) * f(t), \quad h(t) = \exp\{-t\} \end{cases}$$

A more complex electric system example



$$f(t) = \frac{\partial x_2(t)}{\partial t} + x_2(t), \quad x_2(t) = \frac{\partial x_1(t)}{\partial t} + x_1(t)$$

- Differential Equation model: $\frac{\partial^2 x_1(t)}{\partial t^2} + 2\frac{\partial x_1(t)}{\partial t} + x_1(t) = f(t)$
- State space model

$$\left\{ \begin{array}{ll} \left[\begin{array}{c} \frac{\partial x_1(t)}{\partial t} \\ \frac{\partial x_2(t)}{\partial t} \end{array} \right] & = \left[\begin{array}{rr} -1 & 1 \\ 0 & -1 \end{array} \right] \left[\begin{array}{c} x_1(t) \\ x_2(t) \end{array} \right] + \left[\begin{array}{c} 0 \\ 1 \end{array} \right] f(t) \\ g(t) & = \left[\begin{array}{c} 1 \\ 0 \end{array} \right] x_1(t) \end{array} \right.$$

- Input-Output Model: $g(t) = h(t) * f(t)$

Design/Control Inverse problems examples

Simple Electrical system:

$$a\frac{\partial x(t)}{\partial t} + x(t) = f(t), \quad x(0) = x_0, \quad g(t) = x(t)$$

- Design: $\theta = a = RC$
  - Forward: Given $\boldsymbol{\theta} = a$ and $f(t), t > 0$, find $x(t), t > 0$
  - Inverse: Given $x(t)$ and $f(t)$ find $\theta = a$
- Control: $f(t)$
  - Forward: Given $\boldsymbol{\theta} = a$ and $f(t), t > 0$, find $x(t), t > 0$
  - Inverse: Given $\boldsymbol{\theta} = a$ and $x(t), t > 0$, find $f(t)$

More complex Electrical system:

$$f(t) = b\frac{\partial x_2(t)}{\partial t} + x_2(t), \quad x_2(t) = a\frac{\partial x_1(t)}{\partial t} + x_1(t), \quad g(t) = x_1(t)$$

$\theta = (a = R_1 C_1, b = R_2 C_2)$

Design/Control Inverse problems examples

Mass-spring-dashpot system

$$m\frac{\partial^2 x(t)}{\partial t^2} + c\frac{\partial x(t)}{\partial t} + k = F(t), \quad x(0) = x_0, \quad \frac{\partial x}{\partial t}(0) = v_0$$

- Design: $\theta = (m, c, k)$
  - Forward: Given $\boldsymbol{\theta} = (m, c, k)$, x_0, v_0 and $F(t), t > 0$, find $x(t), t > 0$
  - Inverse: Given $x(t)$ for $t > 0$, v_0, $F(t)$ find $\theta = (m, c, k)$
- Control: $F(t)$
  - Forward: Given $\boldsymbol{\theta} = (m, c, k)$, x_0, v_0 and $F(t), t > 0$, find $x(t), t > 0$
  - Inverse: Given $\boldsymbol{\theta} = (m, c, k)$, v_0 and $x(t), t > 0$, find $F(t)$

Input-Output model

- Linear Systems
  - Single Input Single Output (SISO) systems

$$y(t) = \int h(t,\tau)\, u(\tau) \, \mathrm{d}\tau$$

  - Multi Input Multi Output (MIMO) systems

$$\boldsymbol{y}(t) = \int \boldsymbol{H}(t,\tau)\, \boldsymbol{u}(\tau) \, \mathrm{d}\tau$$

- Linear Time Invariant System
  - SISO Convolution

$$y(t) = h(t) * u(t) = \int h(t-\tau)\, u(\tau) \, \mathrm{d}\tau$$

  - MIMO Convolution

$$\boldsymbol{y}(t) = \int \boldsymbol{H}(t-\tau)\, \boldsymbol{u}(\tau) \, \mathrm{d}\tau$$

  - Impulse response $h(t)$ or $\boldsymbol{H}(t) = \begin{bmatrix} . & . & . \\ . & h_{ij}(t) & . \\ . & . & . \end{bmatrix}$

State space model: Continuous case

Dynamic systems:

- Single Input Single Output (SISO) system:

$$\begin{cases} \dot{x}(t) & = A\,x(t) + B\,u(t) \quad \text{State equation} \\ y(t) & = C\,x(t) + D\,v(t) \quad \text{Observation equation} \end{cases}$$

- Multiple Input Multiple Output (MIMO) system:

$$\begin{cases} \dot{\boldsymbol{x}}(t) & = \boldsymbol{H}\,\boldsymbol{x}(t) + \boldsymbol{B}\,\boldsymbol{u}(t) \quad \text{State equation} \\ \boldsymbol{y}(t) & = \boldsymbol{C}\,\boldsymbol{x}(t) + \boldsymbol{D}\,\boldsymbol{v}(t) \quad \text{Observation equation} \end{cases}$$

$\boldsymbol{H}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ are the matrices of the system.

Analytical methods (mathematical physics)

$$g(\boldsymbol{s}_i) = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s}_i), \quad i = 1, \cdots, m$$
$$g(\boldsymbol{s}) = \int h(\boldsymbol{s}, \boldsymbol{r})\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}$$
$$\widehat{f}(\boldsymbol{r}) = \int w(\boldsymbol{s}, \boldsymbol{r})\, g(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}$$

$w(\boldsymbol{s}, \boldsymbol{r})$ minimizing a criterion:

$$\begin{aligned} Q(w(\boldsymbol{s}, \boldsymbol{r})) &= \left\| g(\boldsymbol{s}) - [\mathcal{H}\, \widehat{f}(\boldsymbol{r})](\boldsymbol{s}) \right\|_2^2 = \int \left| g(\boldsymbol{s}) - [\mathcal{H}\, \widehat{f}(\boldsymbol{r})](\boldsymbol{s}) \right|^2 \mathrm{d}\boldsymbol{s} \\ &= \int \left| g(\boldsymbol{s}) - \int h(\boldsymbol{s}, \boldsymbol{r})\, \widehat{f}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \right|^2 \mathrm{d}\boldsymbol{s} \\ &= \int \left| g(\boldsymbol{s}) - \int\!\!\int h(\boldsymbol{s}, \boldsymbol{r}) w(\boldsymbol{s}, \boldsymbol{r})\, g(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{r} \right|^2 \mathrm{d}\boldsymbol{s} \end{aligned}$$

Trivial solution: $\quad h(\boldsymbol{s}, \boldsymbol{r}) w(\boldsymbol{s}, \boldsymbol{r}) = \delta(\boldsymbol{r})\delta(\boldsymbol{s})$

Analytical methods

- Trivial solution:

$$w(\boldsymbol{s},\boldsymbol{r}) = h^{-1}(\boldsymbol{s},\boldsymbol{r})$$

Example: Fourier Transform:

$$g(\boldsymbol{s}) = \int f(\boldsymbol{r}) \exp\{-j\boldsymbol{s}.\boldsymbol{r}\} \, \mathrm{d}\boldsymbol{r}$$

$$h(\boldsymbol{s},\boldsymbol{r}) = \exp\{-j\boldsymbol{s}.\boldsymbol{r}\} \longrightarrow w(\boldsymbol{s},\boldsymbol{r}) = \exp\{+j\boldsymbol{s}.\boldsymbol{r}\}$$

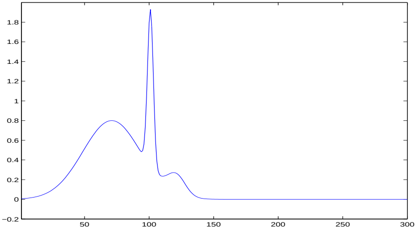
$$\hat{f}(\boldsymbol{r}) = \int g(\boldsymbol{s}) \exp\{+j\boldsymbol{s}.\boldsymbol{r}\} \, \mathrm{d}\boldsymbol{s}$$

- Known classical solutions for specific expressions of $h(\boldsymbol{s},\boldsymbol{r})$:
  - 1D cases: 1D Fourier, Hilbert, Weil, Melin, ...
  - 2D cases: 2D Fourier, Radon, ...

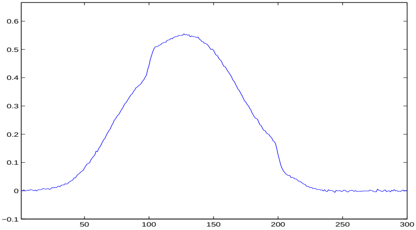
Deconvolution: Analytical methods

Time domain	Fourier domain
Forward model:	
$g(t) = h(t) * f(t) + \epsilon(t)$	$G(\omega) = H(\omega)\, F(\omega) + E(\omega)$
$f(t) \to \boxed{h(t)} \to \overset{\epsilon(t)}{\oplus} \to \quad g(t)$	$F(\omega) \to \boxed{H(\omega)} \to \overset{E(\omega)}{\oplus} \to \quad G(\omega)$
Deconvolution:	Inverse filtering
$g(t) \to \boxed{w(t) = IFT\{\frac{1}{H(\omega)}\}} \to \widehat{f}(t)$	$G(\omega) \to \boxed{\frac{1}{H(\omega)}} \to \widehat{F}(\omega)$
Deconvolution:	Wiener filtering
$g(t) \to \boxed{W(\omega)} \to \widehat{f}(t)$	$G(\omega) \to \boxed{\frac{H^*(\omega)}{\lvert H(\omega)\rvert^2 + \frac{S_\epsilon(\omega)}{S_f(\omega)}}} \to \widehat{F}(\omega)$

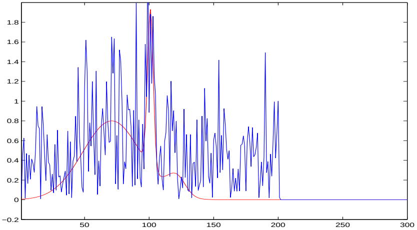
Deconvolution example



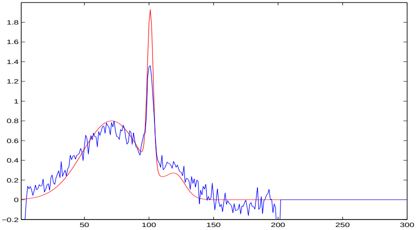
$f(t)$



$g(t)$

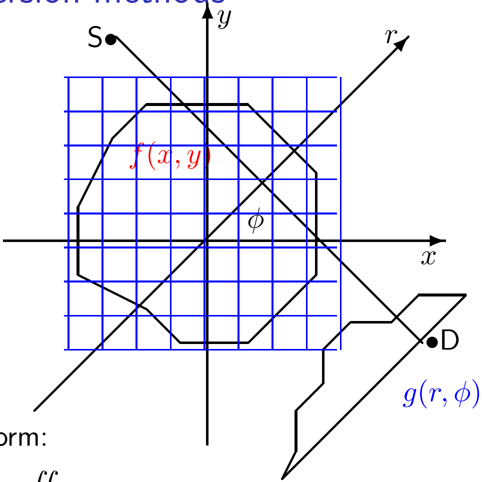


Inverse Filtering



Wiener Filtering

Analytical Inversion methods



Radon Transform:

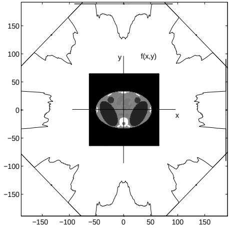
$$g(r,\phi) = \iint_D f(x,y)\, \delta(r - x\cos\phi - y\sin\phi)\, \mathrm{d}x\ \mathrm{d}y$$

$$f(x,y) = \left(-\frac{1}{2\pi^2}\right) \int_0^{\pi} \int_{-\infty}^{+\infty} \frac{\frac{\partial}{\partial r} g(r,\phi)}{(r - x\cos\phi - y\sin\phi)}\, \mathrm{d}r\ \mathrm{d}\phi$$

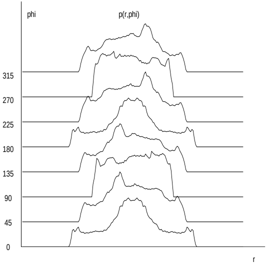
X ray Tomography

$$g(r,\phi) = -\ln\left(\frac{I}{I_0}\right) = \int_{L_{r,\phi}} f(x,y)\;\mathrm{d}l$$

$$g(r,\phi) = \iint_D f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\;\mathrm{d}x\;\mathrm{d}y$$

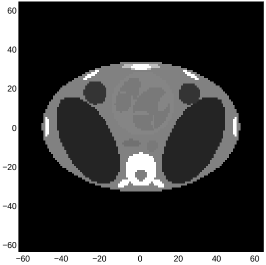


$$f(x,y) \longrightarrow \boxed{\text{RT}} \longrightarrow g(r,\phi)$$

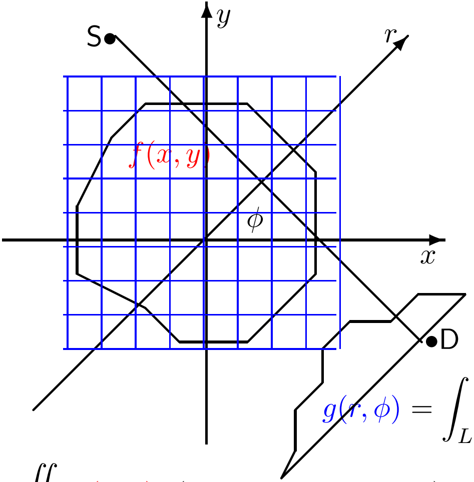


IRT ?

$$\Longrightarrow$$



Analytical Inversion methods



$$g(r,\phi) = \int_L f(x,y) \, \mathrm{d}l$$

Radon:

$$g(r,\phi) = \iint_D f(x,y)\, \delta(r - x\cos\phi - y\sin\phi) \,\mathrm{d}x \,\mathrm{d}y$$

$$f(x,y) = \left(-\frac{1}{2\pi^2}\right) \int_0^{\pi} \int_{-\infty}^{+\infty} \frac{\frac{\partial}{\partial r} g(r,\phi)}{(r - x\cos\phi - y\sin\phi)} \,\mathrm{d}r \,\mathrm{d}\phi$$

Filtered Backprojection method

$$f(x,y) = \left(-\frac{1}{2\pi^2}\right) \int_0^{\pi} \int_{-\infty}^{+\infty} \frac{\frac{\partial}{\partial r} g(r,\phi)}{(r - x\cos\phi - y\sin\phi)} \, \mathrm{d}r \, \mathrm{d}\phi$$

$$\text{Derivation } \mathcal{D}: \quad \overline{g}(r,\phi) = \frac{\partial g(r,\phi)}{\partial r}$$

$$\text{Hilbert Transform} \mathcal{H}: \quad g_1(r',\phi) = \frac{1}{\pi} \int_0^{\infty} \frac{\overline{g}(r,\phi)}{(r - r')} \, \mathrm{d}r$$

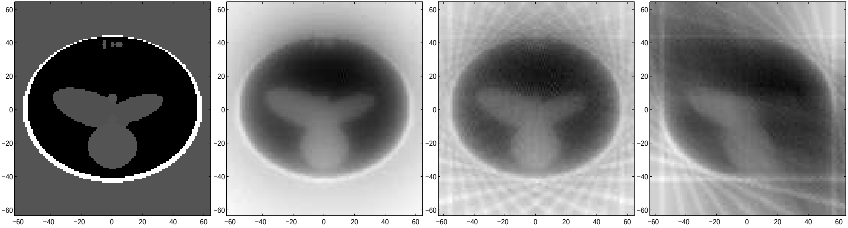
$$\text{Backprojection } \mathcal{B}: \quad f(x,y) = \frac{1}{2\pi} \int_0^{\pi} g_1(r' = x\cos\phi + y\sin\phi, \phi) \, \mathrm{d}\phi$$

$$f(x,y) = \mathcal{B}\,\mathcal{H}\,\mathcal{D}\, g(r,\phi) = \mathcal{B}\,\mathcal{F}_1^{-1}\,|\Omega|\,\mathcal{F}_1\, g(r,\phi)$$

- Backprojection of filtered projections:

$$\xrightarrow{g(r,\phi)} \boxed{\begin{matrix}\mathrm{FT} \\ \mathcal{F}_1\end{matrix}} \longrightarrow \boxed{\begin{matrix}\text{Filter} \\ |\Omega|\end{matrix}} \longrightarrow \boxed{\begin{matrix}\mathrm{IFT} \\ \mathcal{F}_1^{-1}\end{matrix}} \xrightarrow{g_1(r,\phi)} \boxed{\begin{matrix}\text{Backprojection} \\ \mathcal{B}\end{matrix}} \xrightarrow{f(x,y)}$$

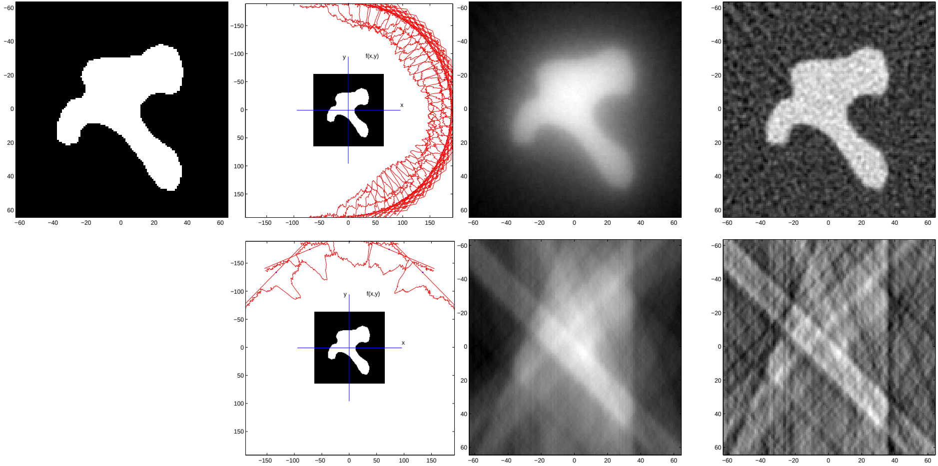
Limitations : Limited angle or noisy data



Original | 64 proj. | 16 proj. | 8 proj. $[0, \pi/2]$

- Limited angle or noisy data
- Accounting for detector size
- Other measurement geometries: fan beam, ...

Limitations : Limited angle or noisy data



Original   Data   Backprojection   Filtered Backprojection

Parametric methods

- $f(\boldsymbol{r})$ is described in a parametric form with a very few number of parameters $\boldsymbol{\theta}$ and one searches $\widehat{\boldsymbol{\theta}}$ which minimizes a criterion such as:
- Least Squares (LS): $Q(\boldsymbol{\theta}) = \sum_i |g_i - [\mathcal{H}\, f(\boldsymbol{\theta})]_i|^2$
- Robust criteria : $Q(\boldsymbol{\theta}) = \sum_i \phi\left(|g_i - [\mathcal{H}\, f(\boldsymbol{\theta})]_i|\right)$
  with different functions ϕ (L_1, Hubert, ...).
- Likelihood : $\mathcal{L}(\boldsymbol{\theta}) = -\ln p(\boldsymbol{g}|\boldsymbol{\theta})$
- Penalized likelihood : $\mathcal{L}(\boldsymbol{\theta}) = -\ln p(\boldsymbol{g}|\boldsymbol{\theta}) + \lambda\Omega(\boldsymbol{\theta})$

Examples:

- Spectrometry: $f(t)$ modelled as a sum og gaussians
  $f(t) = \sum_{k=1}^{K} a_k \mathcal{N}(t|\mu_k, v_k)$ $\quad \boldsymbol{\theta} = \{a_k, \mu_k, v_k\}$
- Tomography in CND: $f(x,y)$ is modelled as a superposition of circular or elleiptical discs $\quad \boldsymbol{\theta} = \{a_k, \mu_k, r_k\}$

Non parametric methods

$$g(\boldsymbol{s}_i) = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s}_i), \quad i = 1, \cdots, M$$

- $f(\boldsymbol{r})$ is assumed to be well approximated by

$$f(\boldsymbol{r}) \simeq \sum_{j=1}^{N} f_j\, b_j(\boldsymbol{r})$$

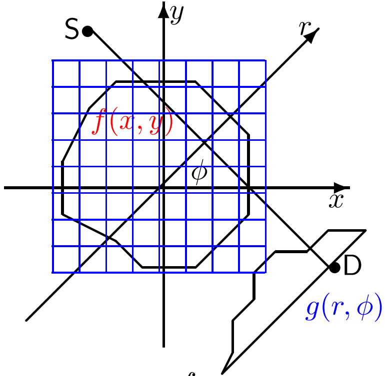
with $\{b_j(\boldsymbol{r})\}$ a basis or any other set of known functions

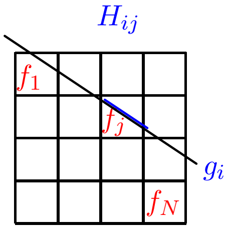
$$g(\boldsymbol{s}_i) = g_i \simeq \sum_{j=1}^{N} f_j \int h(\boldsymbol{s}_i, \boldsymbol{r})\, b_j(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}, \quad i = 1, \cdots, M$$

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \;\text{ with }\; H_{ij} = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, b_j(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}$$

- $\boldsymbol{H}$ is huge dimensional
- LS solution : $\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{Q(\boldsymbol{f})\}$ with
  $Q(\boldsymbol{f}) = \sum_i |g_i - [\boldsymbol{H}\boldsymbol{f}]_i|^2 = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$
  does not give satisfactory result.

Algebraic methods: Discretization





$$f(x,y) = \sum_j f_j \, b_j(x,y)$$

$$b_j(x,y) = \begin{cases} 1 & \text{if } (x,y) \in \text{ pixel } j \\ 0 & \text{else} \end{cases}$$

$$g(r,\phi) = \int_L f(x,y) \, \mathrm{d}l$$

$$g_i = \sum_{j=1}^{N} H_{ij} \, f_j + \epsilon_i$$

$$\boldsymbol{g} = \boldsymbol{H} \boldsymbol{f} + \boldsymbol{\epsilon}$$

Inversion: Deterministic methods

Data matching

- Observation model
  $g_i = h_i(\boldsymbol{f}) + \epsilon_i, \quad i = 1, \ldots, M \longrightarrow \boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{\epsilon}$
- Misatch between data and output of the model $\Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f}))$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{\Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f}))\}$$

- Examples:

– LS $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \|\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})\|^2 = \sum_i |g_i - h_i(\boldsymbol{f})|^2$

– L_p $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \|\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})\|^p = \sum_i |g_i - h_i(\boldsymbol{f})|^p, \quad 1 < p < 2$

– KL $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \sum_i g_i \ln \frac{g_i}{h_i(\boldsymbol{f})}$

- In general, does not give satisfactory results for inverse problems.

Regularization theory

Inverse problems = Ill posed problems

$\longrightarrow$ Need for prior information

Functional space (Tikhonov):

$$g = \mathcal{H}(f) + \epsilon \longrightarrow J(f) = ||g - \mathcal{H}(f)||_2^2 + \lambda||\mathcal{D}f||_2^2$$

Finite dimensional space (Philips & Towmey): $\boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{\epsilon}$

- Minimum norme LS (MNLS): $J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda||\boldsymbol{f}||^2$
- Classical regularization: $J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda||\boldsymbol{D}\boldsymbol{f}||^2$
- More general regularization:

$$J(\boldsymbol{f}) = \mathcal{Q}(\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})) + \lambda\Omega(\boldsymbol{D}\boldsymbol{f})$$

or

$$J(\boldsymbol{f}) = \Delta_1(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) + \lambda\Delta_2(\boldsymbol{f}, \boldsymbol{f}_\infty)$$

Limitations:

- Errors are considered implicitly white and Gaussian
- Limited prior information on the solution
- Lack of tools for the determination of the hyperparameters

Inversion: Probabilistic methods

Taking account of errors and uncertainties $\longrightarrow$ Probability theory

- Maximum Likelihood (ML)
- Minimum Inaccuracy (MI)
- Probability Distribution Matching (PDM)
- Maximum Entropy (ME) and Information Theory (IT)
- Bayesian Inference (Bayes)

Advantages:

- Explicit account of the errors and noise
- A large class of priors via explicit or implicit modeling
- A coherent approach to combine information content of the data and priors

Limitations:

- Practical implementation and cost of calculation

Bayesian estimation approach

$$\mathcal{M}: \qquad \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$

- Observation model $\mathcal{M}$ + Hypothesis on the noise $\boldsymbol{\epsilon} \longrightarrow$
  $p(\boldsymbol{g}|\boldsymbol{f};\mathcal{M}) = p_{\boldsymbol{\epsilon}}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})$
- A priori information $\quad p(\boldsymbol{f}|\mathcal{M})$
- Bayes : $\quad p(\boldsymbol{f}|\boldsymbol{g};\mathcal{M}) = \dfrac{p(\boldsymbol{g}|\boldsymbol{f};\mathcal{M})\, p(\boldsymbol{f}|\mathcal{M})}{p(\boldsymbol{g}|\mathcal{M})}$

Link with regularization :

Maximum A Posteriori (MAP) :

$$\begin{aligned}\widehat{\boldsymbol{f}} &= \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f})\} \\ &= \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f}) - \ln p(\boldsymbol{f})\}\end{aligned}$$

with $\quad Q(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) = -\ln p(\boldsymbol{g}|\boldsymbol{f}) \quad$ and $\quad \lambda\Omega(\boldsymbol{f}) = -\ln p(\boldsymbol{f})$

Case of linear models and Gaussian priors

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$

- Hypothesis on the noise: $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{I}) \longrightarrow$
  $p(\boldsymbol{g}|\boldsymbol{f}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right\}$
- Hypothesis on $\boldsymbol{f}$: $\boldsymbol{f} \sim \mathcal{N}(0, \sigma_f^2 (\boldsymbol{D}'\boldsymbol{D})^{-1}) \longrightarrow$
  $p(\boldsymbol{f}) \propto \exp\left\{-\frac{1}{2\sigma_f^2}\|\boldsymbol{D}\boldsymbol{f}\|^2\right\}$
- A posteriori:
  $$p(\boldsymbol{f}|\boldsymbol{g}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 - \frac{1}{2\sigma_f^2}\|\boldsymbol{D}\boldsymbol{f}\|^2\right\}$$
- MAP : $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$
  with $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|^2, \qquad \lambda = \frac{\sigma_\epsilon^2}{\sigma_f^2}$
- Advantage : characterization of the solution

$$\boldsymbol{f}|\boldsymbol{g} \sim \mathcal{N}(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{P}}) \quad \text{with} \quad \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{P}}\boldsymbol{H}'\boldsymbol{g}, \quad \widehat{\boldsymbol{P}} = \left(\boldsymbol{H}'\boldsymbol{H} + \lambda\boldsymbol{D}'\boldsymbol{D}\right)^{-1}$$

MAP estimation with other priors:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \quad \text{with} \quad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\Omega(\boldsymbol{f})$$

Separable priors:

- Gaussian: $p(f_j) \propto \exp\{-\alpha|f_j|^2\} \longrightarrow \Omega(\boldsymbol{f}) = \alpha\sum_j |f_j|^2$
- Gamma:
  $p(f_j) \propto f_j^\alpha \exp\{-\beta f_j\} \longrightarrow \Omega(\boldsymbol{f}) = \alpha\sum_j \ln f_j + \beta f_j$
- Beta:
  $p(f_j) \propto f_j^\alpha (1-f_j)^\beta \longrightarrow \Omega(\boldsymbol{f}) = \alpha\sum_j \ln f_j + \beta\sum_j \ln(1-f_j)$
- Generalized Gaussian: $p(f_j) \propto \exp\{-\alpha|f_j|^p\}, \quad 1 < p < 2 \longrightarrow \quad \Omega(\boldsymbol{f}) = \alpha\sum_j |f_j|^p,$

Markovian models:

$$p(f_j|\boldsymbol{f}) \propto \exp\left\{-\alpha \sum_{i\in N_j} \phi(f_j, f_i)\right\} \longrightarrow \quad \Omega(\boldsymbol{f}) = \alpha \sum_j \sum_{i\in N_j} \phi(f_j, f_i),$$

MAP estimation with markovien priors:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \quad \text{with} \quad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\Omega(\boldsymbol{f})$$

$$\Omega(\boldsymbol{f}) = \sum_j \phi(f_j - f_{j-1})$$

with $\phi(t)$:

Convex functions:

$$|t|^\alpha, \; \sqrt{1+t^2} - 1, \; \log(\cosh(t)), \; \begin{cases} t^2 & |t| \leq T \\ 2T|t| - T^2 & |t| > T \end{cases}$$

or Non convex functions:

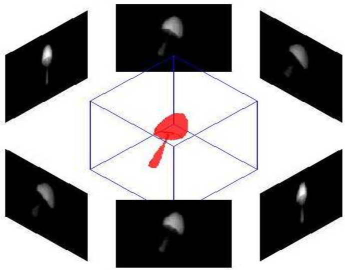
$$\log(1+t^2), \quad \frac{t^2}{1+t^2}, \quad \arctan(t^2), \quad \begin{cases} t^2 & |t| \leq T \\ T^2 & |t| > T \end{cases}$$

Main advantages of the Bayesian approach

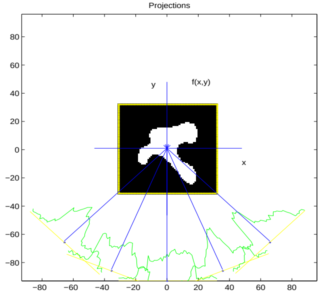
- MAP = Regularization
- Posterior mean ? Marginal MAP ?
- More information in the posterior law than only its mode or its mean
- Meaning and tools for estimating hyper parameters
- Meaning and tools for model selection
- More specific and specialized priors, particularly through the hidden variables
- More computational tools:
  - Expectation-Maximization for computing the maximum likelihood parameters
  - MCMC for posterior exploration
  - Variational Bayes for analytical computation of the posterior marginals
  - ...

2D and 3D Computed Tomography

3D



2D



$$g_\phi(r_1,r_2) = \int_{\mathcal{L}_{r_1,r_2,\phi}} f(x,y,z)\,\mathrm{d}l \qquad g_\phi(r) = \int_{\mathcal{L}_{r,\phi}} f(x,y)\,\mathrm{d}l$$

Forward probelm: $f(x,y)$ or $f(x,y,z) \longrightarrow g_\phi(r)$ or $g_\phi(r_1,r_2)$

Inverse problem: $g_\phi(r)$ or $g_\phi(r_1,r_2) \longrightarrow f(x,y)$ or $f(x,y,z)$

Inverse problems: Discretization

$$g(\boldsymbol{s}_i) = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{s}_i), \quad i = 1, \cdots, M$$

- $f(\boldsymbol{r})$ is assumed to be well approximated by

$$f(\boldsymbol{r}) \simeq \sum_{j=1}^{N} f_j\, b_j(\boldsymbol{r})$$

with $\{b_j(\boldsymbol{r})\}$ a basis or any other set of known functions

$$g(\boldsymbol{s}_i) = g_i \simeq \sum_{j=1}^{N} f_j \int h(\boldsymbol{s}_i, \boldsymbol{r})\, b_j(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}, \quad i = 1, \cdots, M$$

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \;\text{ with }\; H_{ij} = \int h(\boldsymbol{s}_i, \boldsymbol{r})\, b_j(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}$$

- $\boldsymbol{H}$ is huge dimensional
- LS solution : $\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{Q(\boldsymbol{f})\}$ with
  $Q(\boldsymbol{f}) = \sum_i |g_i - [\boldsymbol{H}\boldsymbol{f}]_i|^2 = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$
  does not give satisfactory result.

Inverse problems: Deterministic methods

Data matching

- Observation model
  $g_i = h_i(\boldsymbol{f}) + \epsilon_i, \quad i = 1, \ldots, M \longrightarrow \boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{\epsilon}$
- Misatch between data and output of the model $\Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f}))$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{\Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f}))\}$$

- Examples:

– LS $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \|\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})\|^2 = \sum_i |g_i - h_i(\boldsymbol{f})|^2$

– L_p $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \|\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})\|^p = \sum_i |g_i - h_i(\boldsymbol{f})|^p, \quad 1 < p < 2$

– KL $\quad \Delta(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) = \sum_i g_i \ln \frac{g_i}{h_i(\boldsymbol{f})}$

- In general, does not give satisfactory results for inverse problems.

Inverse problems: Regularization theory

Inverse problems = Ill posed problems

$\longrightarrow$ Need for prior information

Functional space (Tikhonov):

$$g = \mathcal{H}(f) + \epsilon \longrightarrow J(f) = ||g - \mathcal{H}(f)||_2^2 + \lambda||\mathcal{D}f||_2^2$$

Finite dimensional space (Philips & Towmey): $\boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{\epsilon}$

- Minimum norme LS (MNLS): $J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda||\boldsymbol{f}||^2$
- Classical regularization: $J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda||\boldsymbol{D}\boldsymbol{f}||^2$
- More general regularization:

$$J(\boldsymbol{f}) = \mathcal{Q}(\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})) + \lambda\Omega(\boldsymbol{D}\boldsymbol{f})$$

or

$$J(\boldsymbol{f}) = \Delta_1(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) + \lambda\Delta_2(\boldsymbol{f}, \boldsymbol{f}_\infty)$$

Limitations:

- Errors are considered implicitly white and Gaussian
- Limited prior information on the solution
- Lack of tools for the determination of the hyperparameters

Bayesian inference for inverse problems

$$\mathcal{M}: \qquad \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$

- Observation model $\mathcal{M}$ + Hypothesis on the noise $\boldsymbol{\epsilon}$ $\longrightarrow$ $p(\boldsymbol{g}|\boldsymbol{f};\mathcal{M}) = p_{\boldsymbol{\epsilon}}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})$
- A priori information $\quad p(\boldsymbol{f}|\mathcal{M})$
- Bayes : $\quad p(\boldsymbol{f}|\boldsymbol{g};\mathcal{M}) = \dfrac{p(\boldsymbol{g}|\boldsymbol{f};\mathcal{M})\, p(\boldsymbol{f}|\mathcal{M})}{p(\boldsymbol{g}|\mathcal{M})}$

Link with regularization :

Maximum A Posteriori (MAP) :

$$\begin{aligned}\widehat{\boldsymbol{f}} &= \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{g}|\boldsymbol{f})\ p(\boldsymbol{f})\} \\ &= \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f}) - \ln p(\boldsymbol{f})\}\end{aligned}$$

with $\quad Q(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) = -\ln p(\boldsymbol{g}|\boldsymbol{f}) \quad$ and $\quad \lambda\Omega(\boldsymbol{f}) = -\ln p(\boldsymbol{f})$

Bayesian inference for inverse problems

- Linear Inverse problems: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$

$$\boldsymbol{f} \longrightarrow \boxed{\boldsymbol{H}} \longrightarrow \bigoplus^{\downarrow \boldsymbol{\epsilon}} \longrightarrow \boldsymbol{g}$$

- Bayesian inference:

$$p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta}) = \frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2)}{p(\boldsymbol{g}|\boldsymbol{\theta})}$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$

$$\underset{\text{Prior}}{\boxed{p(\boldsymbol{f}|\boldsymbol{\theta}_2)}}^{\downarrow \boldsymbol{\theta}_2} \diamond \underset{\text{Likelihood}}{\boxed{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)}}^{\downarrow \boldsymbol{\theta}_1} \longrightarrow \underset{\text{Posterior}}{\boxed{p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})}} \longrightarrow \widehat{\boldsymbol{f}}$$

- Point estimators:
  - Maximum A Posteriori (MAP): $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})\}$
  - Posterior Mean (PM): $\widehat{\boldsymbol{f}} = \mathrm{E}_{p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})}\{\boldsymbol{f}\} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{f}$

Bayesian Estimation: Two simple priors

- Example 1: Linear Gaussian case:

$$\left\{\begin{array}{l} p(\boldsymbol{g}|\boldsymbol{f},\theta_1) = \mathcal{N}(\boldsymbol{H}\boldsymbol{f},\theta_1\boldsymbol{I}) \\ p(\boldsymbol{f}|\theta_2) = \mathcal{N}(0,\theta_2\boldsymbol{I}) \end{array}\right. \longrightarrow p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta}) = \mathcal{N}(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{P}})$$

with

$$\left\{\begin{array}{l} \widehat{\boldsymbol{P}} = (\boldsymbol{H}'\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}, \quad \lambda = \frac{\theta_1}{\theta_2} \\ \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{P}}\boldsymbol{H}'\boldsymbol{g} \end{array}\right.$$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}}\{J(\boldsymbol{f})\} \text{ with } J(\boldsymbol{f}) = \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2 + \lambda\|\boldsymbol{f}\|_2^2$$

- Example 2: Double Exponential prior & MAP:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}}\{J(\boldsymbol{f})\} \text{ with } J(\boldsymbol{f}) = \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2 + \lambda\|\boldsymbol{f}\|_1$$

Full Bayesian approach

$$\mathcal{M}: \qquad \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$

- Forward & errors model: $\longrightarrow p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1; \mathcal{M})$
- Prior models $\longrightarrow p(\boldsymbol{f}|\boldsymbol{\theta}_2; \mathcal{M})$
- Hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \longrightarrow p(\boldsymbol{\theta}|\mathcal{M})$
- Bayes: $\longrightarrow p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M}) = \frac{p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}; \mathcal{M})\, p(\boldsymbol{f}|\boldsymbol{\theta}; \mathcal{M})\, p(\boldsymbol{\theta}|\mathcal{M})}{p(\boldsymbol{g}|\mathcal{M})}$
- Joint MAP: $\qquad (\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f}, \boldsymbol{\theta})} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\}$
- Marginalization: $\left\{\begin{array}{ll} p(\boldsymbol{f}|\boldsymbol{g}; \mathcal{M}) & = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{\theta} \\ p(\boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M}) & = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{f} \end{array}\right.$
- Posterior means: $\left\{\begin{array}{ll} \widehat{\boldsymbol{f}} & = \int\int \boldsymbol{f}\, p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{f} \\ \widehat{\boldsymbol{\theta}} & = \int\int \boldsymbol{\theta}\, p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}; \mathcal{M})\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta} \end{array}\right.$
- Evidence of the model:

$$p(\boldsymbol{g}|\mathcal{M}) = \iint p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}; \mathcal{M}) p(\boldsymbol{f}|\boldsymbol{\theta}; \mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M})\, \mathrm{d}\boldsymbol{f}\, \mathrm{d}\boldsymbol{\theta}$$

Full Bayesian: Marginal MAP and PM estimates

- Marginal MAP: $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\boldsymbol{g})\}$ where

$$p(\boldsymbol{\theta}|\boldsymbol{g}) = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) \,\mathrm{d}\boldsymbol{f} \propto p(\boldsymbol{g}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

  and then $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g})\right\}$ or

  Posterior Mean: $\widehat{\boldsymbol{f}} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \,\mathrm{d}\boldsymbol{f}$

- Needs the expression of the Likelihood:

$$p(\boldsymbol{g}|\boldsymbol{\theta}) = \int p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2) \,\mathrm{d}\boldsymbol{f}$$

  Not always analytically available $\longrightarrow$ EM, SEM and GEM algorithms

Full Bayesian Model and Hyperparameter Estimation

$$\downarrow \boldsymbol{\alpha}, \boldsymbol{\beta}$$

$$\boxed{\text{Hyper prior model } p(\boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\beta})}$$

$$\downarrow \boldsymbol{\theta}_2 \qquad\qquad \downarrow \boldsymbol{\theta}_1$$

$$\boxed{p(\boldsymbol{f}|\boldsymbol{\theta}_2)} \diamond \boxed{p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)} \longrightarrow \boxed{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \begin{array}{l} \longrightarrow \widehat{\boldsymbol{f}} \\ \longrightarrow \widehat{\boldsymbol{\theta}} \end{array}$$

Prior | Likelihood | Joint Posterior

Full Bayesian Model and Hyperparameter Estimation scheme

$$\boxed{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow \boxed{p(\boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow \widehat{\boldsymbol{\theta}} \longrightarrow \boxed{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g})} \longrightarrow \widehat{\boldsymbol{f}}$$

Joint Posterior | Marginalize over $\boldsymbol{f}$

Marginalization for Hyperparameter Estimation

Full Bayesian: EM and GEM algorithms

- EM and GEM Algorithms: $\boldsymbol{f}$ as hidden variable, $\boldsymbol{g}$ as incomplete data, $(\boldsymbol{g}, \boldsymbol{f})$ as complete data
  $\ln p(\boldsymbol{g}|\boldsymbol{\theta})$ incomplete data log-likelihood
  $\ln p(\boldsymbol{g}, \boldsymbol{f}|\boldsymbol{\theta})$ complete data log-likelihood
- Iterative algorithm:

$$\left\{\begin{array}{ll} \text{E-step:} & Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(k)}) = \mathsf{E}_{p(\boldsymbol{f}|\boldsymbol{g}, \widehat{\boldsymbol{\theta}}^{(k)})} \{\ln p(\boldsymbol{g}, \boldsymbol{f}|\boldsymbol{\theta})\} \\ \text{M-step:} & \widehat{\boldsymbol{\theta}}^{(k)} = \arg\max_{\boldsymbol{\theta}} \left\{Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(k-1)})\right\} \end{array}\right.$$

- GEM (Bayesian) algorithm:

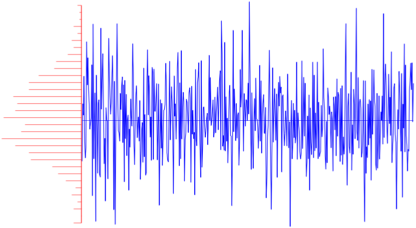
$$\left\{\begin{array}{ll} \text{E-step:} & Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(k)}) = \mathsf{E}_{p(\boldsymbol{f}|\boldsymbol{g}, \widehat{\boldsymbol{\theta}}^{(k)})} \{\ln p(\boldsymbol{g}, \boldsymbol{f}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta})\} \\ \text{M-step:} & \widehat{\boldsymbol{\theta}}^{(k)} = \arg\max_{\boldsymbol{\theta}} \left\{Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(k-1)})\right\} \end{array}\right.$$

$$\boxed{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow \boxed{\text{EM, GEM}} \longrightarrow \widehat{\boldsymbol{\theta}} \longrightarrow \boxed{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g})} \longrightarrow \widehat{\boldsymbol{f}}$$

Two main steps in the Bayesian approach

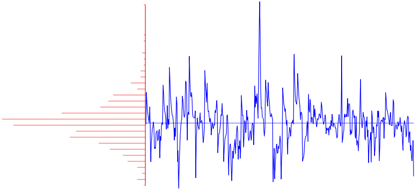
- Prior modeling
  - Separable:
    Gaussian, Gamma,
    Sparsity enforcing: Generalized Gaussian, mixture of Gaussians, mixture of Gammas, ...
  - Markovian:
    Gauss-Markov, GGM, ...
  - Markovian with hidden variables
    (contours, region labels)
- Choice of the estimator and computational aspects
  - MAP, Posterior mean, Marginal MAP
  - MAP needs optimization algorithms
  - Posterior mean needs integration methods
  - Marginal MAP and Hyperparameter estimation need integration and optimization
  - Approximations:
    - Gaussian approximation (Laplace)
    - Numerical exploration MCMC
    - Variational Bayes (Separable approximation)

Different prior models for signals and images: Separable



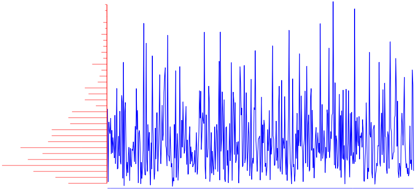
Gaussian

$$p(f_j) \propto \exp\left\{-\alpha|f_j|^2\right\}$$



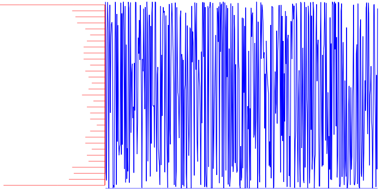
Generalized Gaussian

$$p(f_j) \propto \exp\left\{-\alpha|f_j|^p\right\}, \quad 1 \leq p \leq 2$$



Gamma

$$p(f_j) \propto f_j^\alpha \exp\left\{-\beta f_j\right\}$$

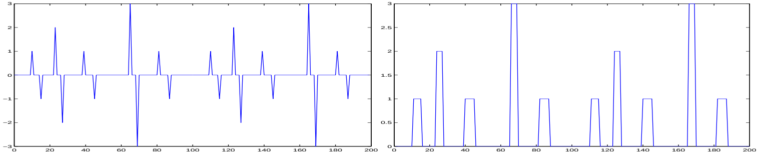


Beta

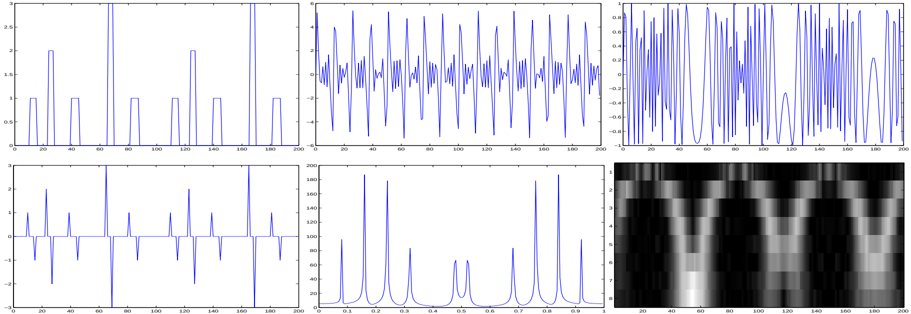
$$p(f_j) \propto f_j^\alpha (1-f_j)^\beta$$

Sparsity enforcing prior models

- Sparse signals: Direct sparsity



- Sparse signals: Sparsity in a Transform domaine



Sparsity enforcing prior models

- Simple heavy tailed models:
  - Generalized Gaussian, Double Exponential
  - Symmetric Weibull, Symmetric Rayleigh
  - Student-t, Cauchy
  - Generalized hyperbolic
  - Elastic net
- Hierarchical mixture models:
  - Mixture of Gaussians
  - Bernoulli-Gaussian
  - Mixture of Gammas
  - Bernoulli-Gamma
  - Mixture of Dirichlet
  - Bernoulli-Multinomial

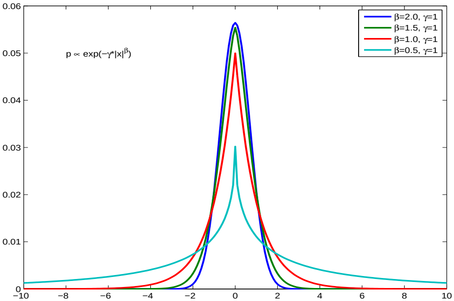
Simple heavy tailed models

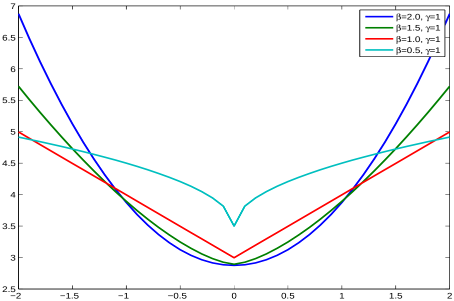
- Generalized Gaussian, Double Exponential

$$p(\boldsymbol{f}|\gamma,\beta) = \prod_j \mathcal{GG}(f_j|\gamma,\beta) \propto \exp\left\{-\gamma \sum_j |f_j|^\beta\right\}$$

$\beta = 1$ Double exponential or Laplace.

$0 < \beta \leq 1$ are of great interest for sparsity enforcing.





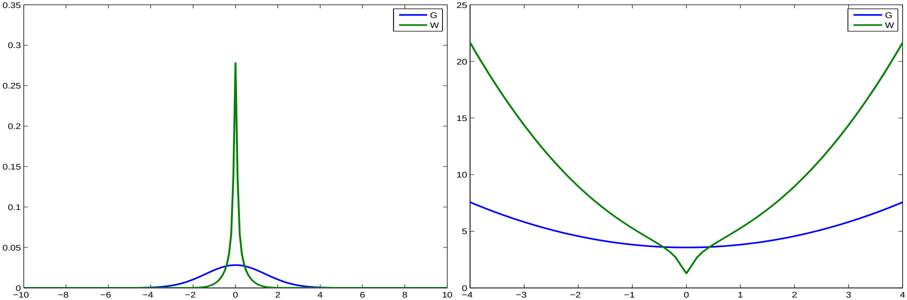
Generalized Gaussian family

Simple heavy tailed models

- Symmetric Weibull

$$p(\boldsymbol{f}|\gamma,\beta) = \prod_j \mathcal{W}(f_j|\gamma,\beta) \propto \exp\left\{-\gamma \sum_j |f_j|^\beta + (\beta-1)\log|f_j|\right\}$$

$\beta = 2$ is the Symmetric Rayleigh distribution.
$\beta = 1$ is the Double exponential and
$0 < \beta \leq 1$ are of great interest for sparsity enforcing.



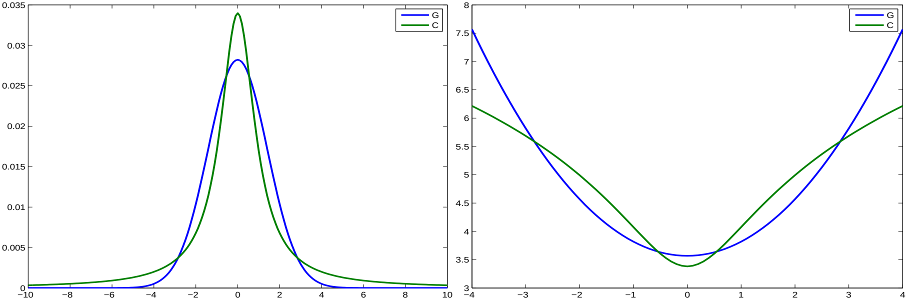
Symmetric Weibull family

Simple heavy tailed models

- Student-t and Cauchy models

$$p(\boldsymbol{f}|\nu) = \prod_j \mathcal{S}t(f_j|\nu) \propto \exp\left\{ -\frac{\nu+1}{2} \sum_j \log\left(1 + f_j^2/\nu\right) \right\}$$

Cauchy model is obtained when $\nu = 1$.

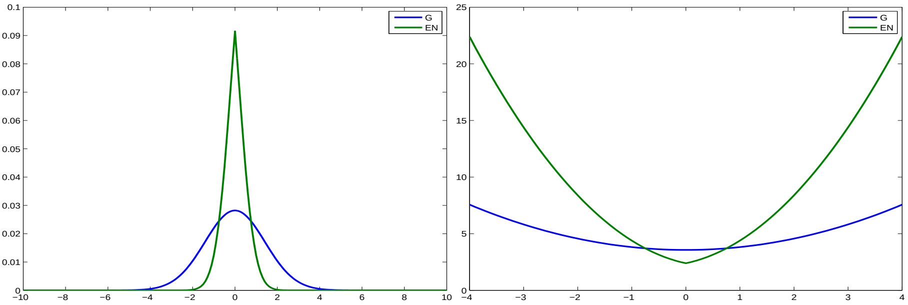


Student-t and Cauchy families

Simple heavy tailed models

- Elastic net prior model

$$p(\boldsymbol{f}|\nu) = \prod_j \mathcal{EN}(f_j|\nu) \propto \exp\left\{-\sum_j (\gamma_1|f_j| + \gamma_2 f_j^2)\right\}$$

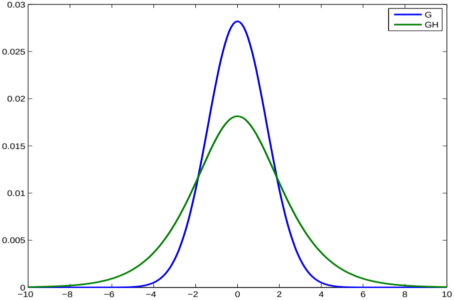


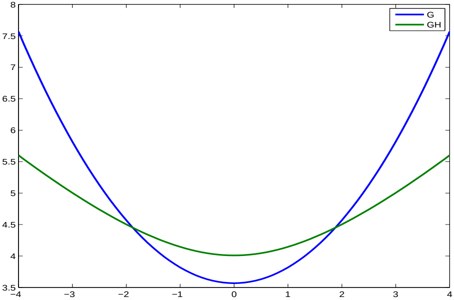
Elastic Net family

Simple heavy tailed models

- Generalized hyperbolic (GH) models

$$p(\boldsymbol{f}|\delta,\nu,\beta) = \prod_j (\delta^2 + f_j^2)^{(\nu-1/2)/2} \exp\{\beta x)\} K_{\nu-1/2}(\alpha\sqrt{\delta^2+f_j^2})$$





Generalized hyperbolic family

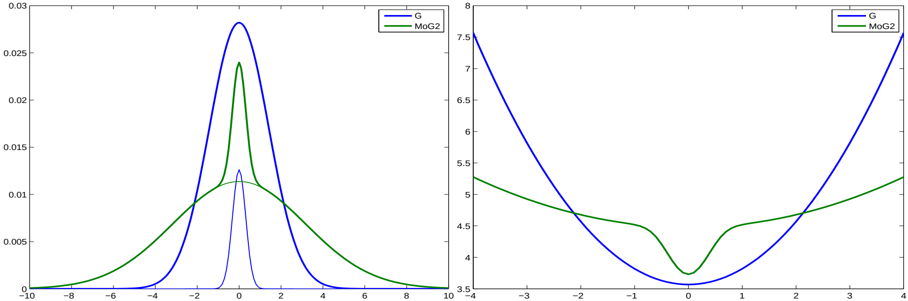
Mixture models

- Mixture of two Gaussians (MoG2) model

$$p(\boldsymbol{f}|\alpha, v_1, v_0) = \prod_j \left[\alpha \mathcal{N}(f_j|0, v_1) + (1-\alpha)\mathcal{N}(f_j|0, v_0)\right]$$

- Bernoulli-Gaussian (BG) model

$$p(\boldsymbol{f}|\alpha, v) = \prod_j p(f_j) = \prod_j \left[\alpha \mathcal{N}(f_j|0, v) + (1-\alpha)\delta(f_j)\right]$$



Mixture of 2 Gaussians families

- Mixture of Gammas

$$p(\boldsymbol{f}|\lambda, v_1, v_0) = \prod_j \left[\lambda \mathcal{G}(f_j|\alpha_1, \beta_1) + (1-\lambda)\mathcal{G}(f_j|\alpha_2, \beta_2)\right]$$

- Bernoulli-Gamma model

$$p(\boldsymbol{f}|\lambda, \alpha, \beta) = \prod_j \left[\lambda \mathcal{G}(f_j|\alpha, \beta) + (1-\lambda)\delta(f_j)\right]$$

- Mixture of Dirichlets model

$$p(\boldsymbol{f}|\lambda, \boldsymbol{H}_1, \boldsymbol{\alpha}_1, \boldsymbol{H}_2, \boldsymbol{\alpha}_2) = \prod_j \left[\lambda \mathcal{D}(f_j|\boldsymbol{H}_1, \boldsymbol{\alpha}_1) + (1-\lambda)\mathcal{D}(f_j|\boldsymbol{H}_2, \boldsymbol{\alpha}_2)\right]$$

$$\mathcal{D}(f_j|\boldsymbol{H}, \boldsymbol{\alpha}) = \prod_{k=1}^{K} \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)\Gamma(\alpha_K)} a_k^{\alpha_k - 1}, \quad \alpha_k \geq 0, \quad a_k \geq 0$$

where $\boldsymbol{H} = \{a_1, \cdots, a_K\}$ and $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_K\}$
with $\sum_k \alpha_k = \alpha$ and $\sum_k a_k = 1$.

- Bernoulli-Multinomial (BMultinomial) model

$$p(\boldsymbol{f}|\lambda, \boldsymbol{H}, \boldsymbol{\alpha}) = \prod_j \left[\lambda \delta(f_j) + (1-\lambda)\mathcal{M}ult(f_j|\boldsymbol{H}, \boldsymbol{\alpha})\right]$$

Hierarchical models and hidden variables

- All the mixture models and some of simple models can be modeled via hidden variables z.

$$p(f)=\sum_{k=1}^{K}\alpha_k p_k(f) \longrightarrow \left\{\begin{array}{l} p(f|z=k)=p_k(f), \\ P(z=k)=\alpha_k, \quad \sum_k \alpha_k=1 \end{array}\right.$$

- Example 1: MoG model: $p_k(f)=\mathcal{N}(f|m_k,v_k)$
  2 Gaussians: $p_0=\mathcal{N}(0,v_0), p_1=\mathcal{N}(0,v_1), \alpha_0=\lambda, \alpha_1=1-\lambda$

$$p(f_j|\lambda,v_1,v_0)=\lambda\mathcal{N}(f_j|0,v_1)+(1-\lambda)\mathcal{N}(f_j|0,v_0)$$

$$\left\{\begin{array}{l} p(f_j|z_j=0,v_0)=\mathcal{N}(f_j|0,v_0), \\ p(f_j|z_j=1,v_1)=\mathcal{N}(f_j|0,v_1), \end{array}\right. \text{ and } \left\{\begin{array}{l} P(z_j=0)=\lambda, \\ P(z_j=1)=1-\lambda \end{array}\right.$$

$$\left\{\begin{array}{l} p(\boldsymbol{f}|\boldsymbol{z})=\prod_j p(f_j|z_j)=\prod_j \mathcal{N}\left(f_j|0,v_{z_j}\right) \propto \exp\left\{-\frac{1}{2}\sum_j \frac{f_j^2}{v_{z_j}}\right\} \\ P(z_j=1)=\lambda, \qquad P(z_j=0)=1-\lambda \end{array}\right.$$

Hierarchical models and hidden variables

- Example 2: Student-t model

$$\mathcal{S}t(f|\nu) \propto \exp\left\{-\frac{\nu+1}{2}\log\left(1+f^2/\nu\right)\right\}$$

- Infinite mixture

$$\mathcal{S}t(f|\nu) \propto= \int_0^\infty \mathcal{N}(f|,0,1/z)\,\mathcal{G}(z|\alpha,\beta)\,\mathrm{d}z, \quad \text{with } \alpha=\beta=\nu/2$$

$$\left\{\begin{array}{ll} p(\boldsymbol{f}|\boldsymbol{z}) & = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0,1/z_j) \propto \exp\left\{-\frac{1}{2}\sum_j z_j f_j^2\right\} \\ p(\boldsymbol{z}|\alpha,\beta) & = \prod_j \mathcal{G}(z_j|\alpha,\beta) \propto \prod_j {z_j}^{(\alpha-1)} \exp\left\{-\beta z_j\right\} \\ & \propto \exp\left\{\sum_j (\alpha-1)\ln z_j - \beta z_j\right\} \\ p(\boldsymbol{f},\boldsymbol{z}|\alpha,\beta) & \propto \exp\left\{-\frac{1}{2}\sum_j z_j f_j^2 + (\alpha-1)\ln z_j - \beta z_j\right\} \end{array}\right.$$

Hierarchical models and hidden variables

- Example 3: Laplace (Double Exponential) model

$$\mathcal{DE}(f|a) = \frac{a}{2} \exp\left\{-a|f|\right\} = \int_0^\infty \mathcal{N}(f|,0,z)\,\mathcal{E}(z|a^2/2)\;\mathrm{d}z, \quad a>0$$

$$\left\{\begin{array}{ll} p(\boldsymbol{f}|\boldsymbol{z}) & = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0,z_j) \propto \exp\left\{-\frac{1}{2}\sum_j f_j^2/z_j\right\} \\ p(\boldsymbol{z}|\frac{a^2}{2}) & = \prod_j \mathcal{E}(z_j|\frac{a^2}{2}) \propto \exp\left\{\sum_j \frac{a^2}{2} z_j\right\} \\ p(\boldsymbol{f},\boldsymbol{z}|\frac{a^2}{2}) & \propto \exp\left\{-\frac{1}{2}\sum_j f_j^2/z_j + \frac{a^2}{2} z_j\right\} \end{array}\right.$$

- With these models we have:

$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\; p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_2)\; p(\boldsymbol{z}|\boldsymbol{\theta}_3)\; p(\boldsymbol{\theta})$$

Bayesian Computation and Algorithms

- Often, the expression of $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ is complex.
- Its optimization (for Joint MAP) or
  its marginalization or integration (for Marginal MAP or PM)
  is not easy
- Two main techniques:
  MCMC and Variational Bayesian Approximation (VBA)
- MCMC:
  Needs the expressions of the conditionals
  $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})$, $p(\boldsymbol{z}|\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{g})$, and $p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{g})$
- VBA: Approximate $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ by a separable one

  $$q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{\theta})$$

  and do any computations with these separable ones.

MCMC based algorithm

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta})\, p(\boldsymbol{z})\, p(\boldsymbol{\theta})$$

General scheme:

$$\widehat{\boldsymbol{f}} \sim p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{z}} \sim p(\boldsymbol{z}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{\theta}} \sim (\boldsymbol{\theta}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \boldsymbol{g})$$

- Estimate $\boldsymbol{f}$ using $p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta})\, p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}})$
  When Gaussian, can be done via optimisation of a quadratic criterion.

- Estimate $\boldsymbol{z}$ using $p(\boldsymbol{z}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}})\, p(\boldsymbol{z})$
  Often needs sampling (hidden discrete variable)

- Estimate $\boldsymbol{\theta}$ using
  $p(\boldsymbol{\theta}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}}, \sigma_\epsilon^2 \boldsymbol{I})\, p(\widehat{\boldsymbol{f}}|\widehat{\boldsymbol{z}}, (m_k, v_k))\, p(\boldsymbol{\theta})$
  Use of Conjugate priors $\longrightarrow$ analytical expressions.

Variational Bayesian Approximation

- Approximate $p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})$ by $q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\boldsymbol{g})\, q_2(\boldsymbol{\theta}|\boldsymbol{g})$ and then continue computations.
- Criterion $\mathrm{KL}(q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) : p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}))$
- $\mathrm{KL}(q:p) = \int\int q \ln q/p = \int\int q_1 q_2 \ln \frac{q_1 q_2}{p} =$
  $\int q_1 \ln q_1 + \int q_2 \ln q_2 - \int\int q \ln p = -H(q_1) - H(q_2) - < \ln p >_q$
- Iterative algorithm $q_1 \longrightarrow q_2 \longrightarrow q_1 \longrightarrow q_2, \cdots$

$$\left\{\begin{array}{ll} q_1(\boldsymbol{f}) & \propto \exp\left\{\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_2(\boldsymbol{\theta})}\right\} \\ q_2(\boldsymbol{\theta}) & \propto \exp\left\{\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_1(\boldsymbol{f})}\right\} \end{array}\right.$$

$$\boxed{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})} \longrightarrow \boxed{\begin{array}{c}\text{Variational}\\ \text{Bayesian}\\ \text{Approximation}\end{array}} \begin{array}{l} \longrightarrow \widehat{q}_1(\boldsymbol{f}) \longrightarrow \widehat{\boldsymbol{f}} \\ \\ \longrightarrow \widehat{q}_2(\boldsymbol{\theta}) \longrightarrow \widehat{\boldsymbol{\theta}} \end{array}$$

Summary of Bayesian estimation 1

- Simple Bayesian Model and Estimation

$$\begin{array}{ccccccc} \downarrow \boldsymbol{\theta}_2 & & \downarrow \boldsymbol{\theta}_1 & & & & \\ \boxed{p(\boldsymbol{f}|\boldsymbol{\theta}_2)} & \diamond & \boxed{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)} & \longrightarrow & \boxed{p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})} & \longrightarrow & \widehat{\boldsymbol{f}} \\ \text{Prior} & & \text{Likelihood} & & \text{Posterior} & & \end{array}$$

- Full Bayesian Model and Hyperparameter Estimation

$$\downarrow \boldsymbol{\alpha},\boldsymbol{\beta}$$

$$\boxed{\text{Hyper prior model } p(\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta})}$$

$$\begin{array}{ccccccc} \downarrow \boldsymbol{\theta}_2 & & \downarrow \boldsymbol{\theta}_1 & & & & \\ \boxed{p(\boldsymbol{f}|\boldsymbol{\theta}_2)} & \diamond & \boxed{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)} & \longrightarrow & \boxed{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{\alpha},\boldsymbol{\beta})} & \begin{array}{c}\longrightarrow \\ \longrightarrow\end{array} & \begin{array}{c}\widehat{\boldsymbol{f}} \\ \widehat{\boldsymbol{\theta}}\end{array} \\ \text{Prior} & & \text{Likelihood} & & \text{Joint Posterior} & & \end{array}$$

Summary of Bayesian estimation 2

- Marginalization for Hyperparameter Estimation

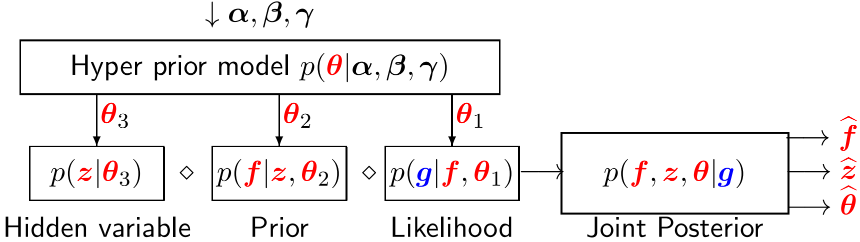
$$\begin{array}{ccccccccc} \boxed{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})} & \longrightarrow & \boxed{p(\boldsymbol{\theta}|\boldsymbol{g})} & \longrightarrow & \widehat{\boldsymbol{\theta}} & \longrightarrow & \boxed{p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}},\boldsymbol{g})} & \longrightarrow & \widehat{\boldsymbol{f}} \\ \text{Joint Posterior} & & \text{Marginalize over } \boldsymbol{f} & & & & & & \end{array}$$

- Full Bayesian Model with a Hierarchical Prior Model

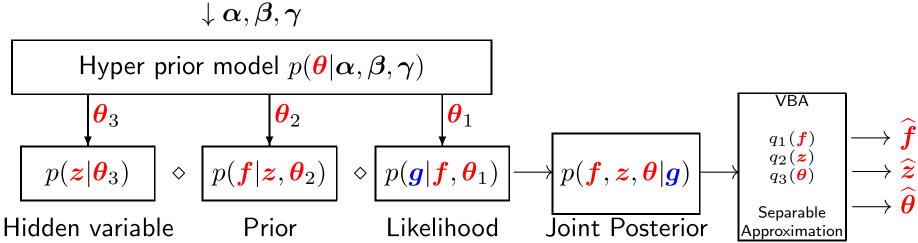
$$\begin{array}{ccccccc} \downarrow \boldsymbol{\theta}_3 & & \downarrow \boldsymbol{\theta}_2 & & \downarrow \boldsymbol{\theta}_1 & & \\ \boxed{p(\boldsymbol{z}|\boldsymbol{\theta}_3)} & \diamond & \boxed{p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_2)} & \diamond & \boxed{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)} & \longrightarrow & \boxed{p(\boldsymbol{f},\boldsymbol{z}|\boldsymbol{g},\boldsymbol{\theta})} \begin{array}{l} \longrightarrow \widehat{\boldsymbol{f}} \\ \longrightarrow \widehat{\boldsymbol{z}} \end{array} \\ \text{Hidden variable} & & \text{Prior} & & \text{Likelihood} & & \text{Joint Posterior} \end{array}$$

Summary of Bayesian estimation 3

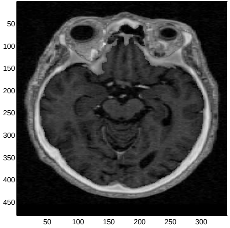
- Full Bayesian Hierarchical Model with Hyperparameter Estimation

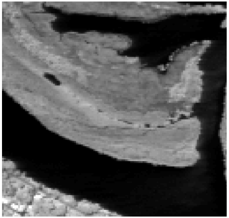


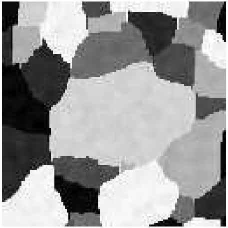
- Full Bayesian Hierarchical Model and Variational Approximation

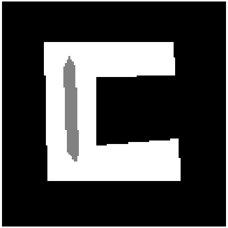


Which images I am looking for?

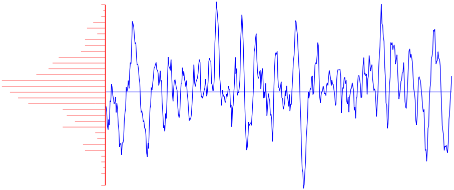




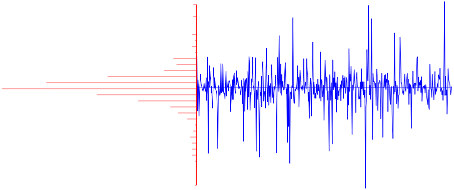




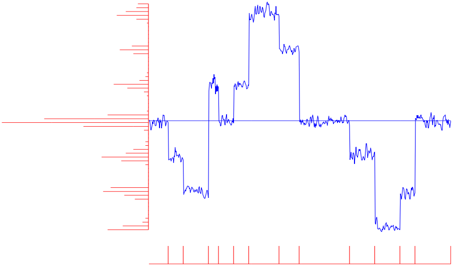
Which image I am looking for?



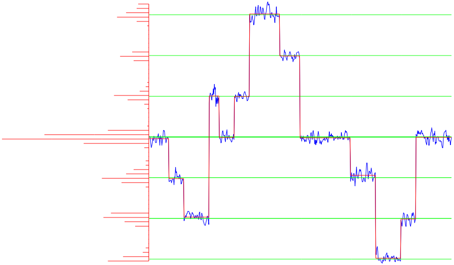
Gauss-Markov



Generalized GM

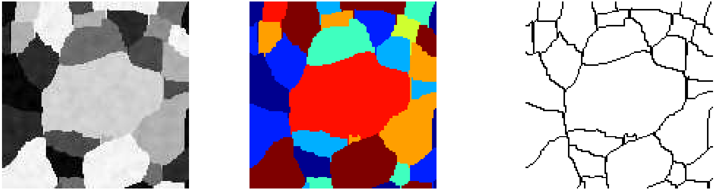


Piecewize Gaussian



Mixture of GM

Gauss-Markov-Potts prior models for images



$f(\boldsymbol{r})$ $\qquad$ $z(\boldsymbol{r})$ $\qquad$ $c(\boldsymbol{r}) = 1 - \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))$

$$p(f(\boldsymbol{r})|z(\boldsymbol{r}) = k, m_k, v_k) = \mathcal{N}(m_k, v_k)$$

$$p(f(\boldsymbol{r})) = \sum_k P(z(\boldsymbol{r}) = k)\, \mathcal{N}(m_k, v_k) \quad \text{Mixture of Gaussians}$$

- Separable iid hidden variables: $\quad p(\boldsymbol{z}) = \prod_{\boldsymbol{r}} p(z(\boldsymbol{r}))$
- Markovian hidden variables: $\quad p(\boldsymbol{z})$ Potts-Markov:

$$p(z(\boldsymbol{r})|z(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})) \propto \exp\left\{\gamma \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}$$

$$p(\boldsymbol{z}) \propto \exp\left\{\gamma \sum_{\boldsymbol{r} \in \mathcal{R}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right\}$$

Four different cases

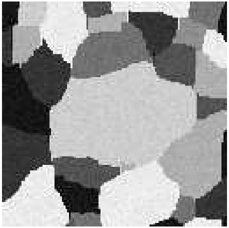
To each pixel of the image is associated 2 variables $f(\boldsymbol{r})$ and $z(\boldsymbol{r})$

- $\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid, $\boldsymbol{z}$ iid :
  Mixture of Gaussians

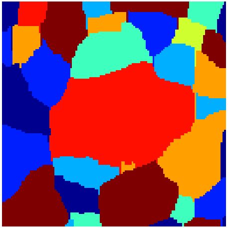
- $\boldsymbol{f}|\boldsymbol{z}$ Gauss-Markov, $\boldsymbol{z}$ iid :
  Mixture of Gauss-Markov

- $\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid, $\boldsymbol{z}$ Potts-Markov :
  Mixture of Independent Gaussians
  (MIG with Hidden Potts)

- $\boldsymbol{f}|\boldsymbol{z}$ Markov, $\boldsymbol{z}$ Potts-Markov :
  Mixture of Gauss-Markov
  (MGM with hidden Potts)



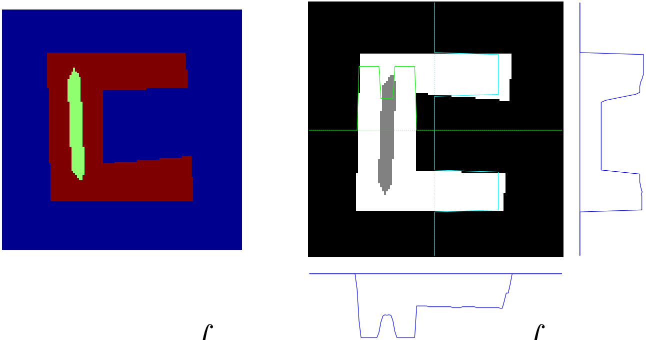
$f(\boldsymbol{r})$



$z(\boldsymbol{r})$

Application of CT in NDT

Reconstruction from only 2 projections

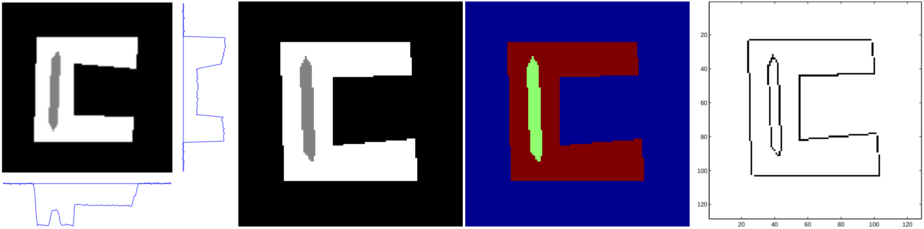


$$g_1(x) = \int f(x,y)\,\mathrm{d}y, \qquad g_2(y) = \int f(x,y)\,\mathrm{d}x$$

- Given the marginals $g_1(x)$ and $g_2(y)$ find the joint distribution $f(x,y)$.
- Infinite number of solutions : $f(x,y) = g_1(x)\,g_2(y)\,\Omega(x,y)$
  $\Omega(x,y)$ is a Copula:

$$\int \Omega(x,y)\,\mathrm{d}x = 1 \quad \text{and} \quad \int \Omega(x,y)\,\mathrm{d}y = 1$$

Application in CT



$g \mid f$	$f \mid z$	z	c
$g = Hf + \epsilon$	iid Gaussian	iid	$c(r) \in \{0, 1\}$
$g \mid f \sim \mathcal{N}(Hf, \sigma_\epsilon^2 I)$	or	or	$1 - \delta(z(r) - z(r'))$
Gaussian	Gauss-Markov	Potts	binary

Proposed algorithm

$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\, p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

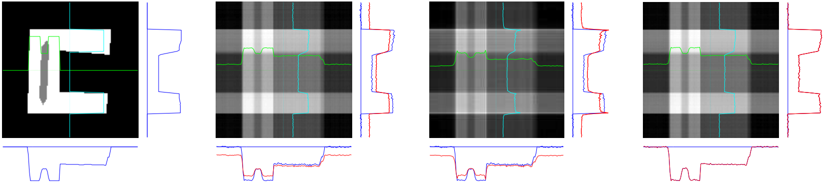
General scheme:

$$\widehat{\boldsymbol{f}} \sim p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{z}} \sim p(\boldsymbol{z}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \longrightarrow \widehat{\boldsymbol{\theta}} \sim (\boldsymbol{\theta}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\boldsymbol{g})$$

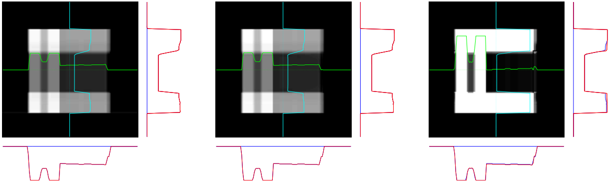
Iterative algorithme:

- Estimate $\boldsymbol{f}$ using $p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta})\, p(\boldsymbol{f}|\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}})$
  Needs optimisation of a quadratic criterion.
- Estimate $\boldsymbol{z}$ using $p(\boldsymbol{z}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\widehat{\boldsymbol{\theta}})\, p(\boldsymbol{z})$
  Needs sampling of a Potts Markov field.
- Estimate $\boldsymbol{\theta}$ using
  $p(\boldsymbol{\theta}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}},\sigma_\epsilon^2\boldsymbol{I})\, p(\widehat{\boldsymbol{f}}|\widehat{\boldsymbol{z}},(m_k,v_k))\, p(\boldsymbol{\theta})$
  Conjugate priors $\longrightarrow$ analytical expressions.

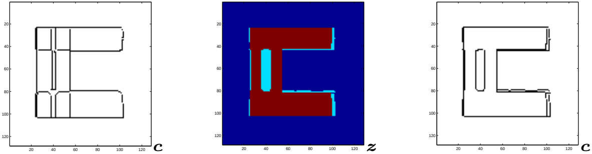
Results



Original Backprojection Filtered BP LS



Gauss-Markov+pos GM+Line process GM+Label process

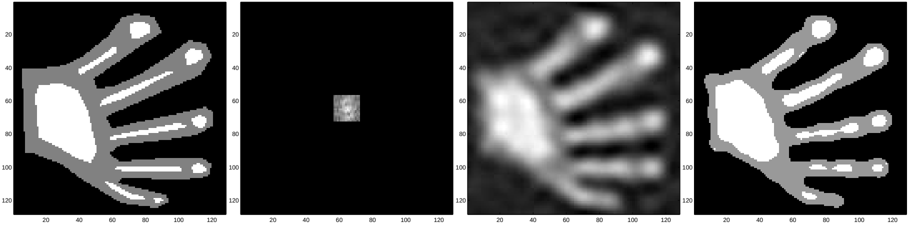


Application in Microwave imaging

$$g(\boldsymbol{\omega}) = \int f(\boldsymbol{r}) \exp\left\{-j(\boldsymbol{\omega}.\boldsymbol{r})\right\} \; \mathrm{d}\boldsymbol{r} + \epsilon(\boldsymbol{\omega})$$

$$g(u,v) = \iint f(x,y) \exp\left\{-j(ux+vy)\right\} \; \mathrm{d}x \; \mathrm{d}y + \epsilon(u,v)$$

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$



$f(x,y)$ $\quad$ $g(u,v)$ $\quad$ $\widehat{\boldsymbol{f}}$ IFT $\quad$ $\widehat{\boldsymbol{f}}$ Proposed method

Conclusions

- Bayesian Inference for inverse problems
- Different prior modeling for signals and images: Separable, Markovian, without and with hidden variables
- Sprasity enforcing priors
- Gauss-Markov-Potts models for images incorporating hidden regions and contours
- Two main Bayesian computation tools: MCMC and VBA
- Application in different CT (X ray, Microwaves, PET, SPECT)

Current Projects and Perspectives :

- Efficient implementation in 2D and 3D cases
- Evaluation of performances and comparison between MCMC and VBA methods
- Application to other linear and non linear inverse problems: (PET, SPECT or ultrasound and microwave imaging)