

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

- ▶ Inverse 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

- ▶ Inverse problems examples
- ▶ Classification of Inversion methods: Analytical, Parametric and Non Parametric algebraic methods
- ▶ Regularization theory
- ▶ Bayesian inference for inverse 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 temperature 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.
 - ▶ Inferring 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 thermometer
 - ▶ $f(t)$ 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 using 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 radiograph $g_\phi(r, z)$
- ▶ Example 1: **Deconvolution**
- ▶ Example 2: **Image restoration**
- ▶ Example 3: **Image reconstruction**

Measuring variation of temperature with a thermometer

- ▶ $f(t)$ variation of temperature over time
- ▶ $g(t)$ variation of length of the liquid in thermometer
- ▶ Forward model: Convolution

$$g(t) = \int f(t') h(t - t') dt' + \epsilon(t)$$

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

- ▶ Inverse problem: Deconvolution

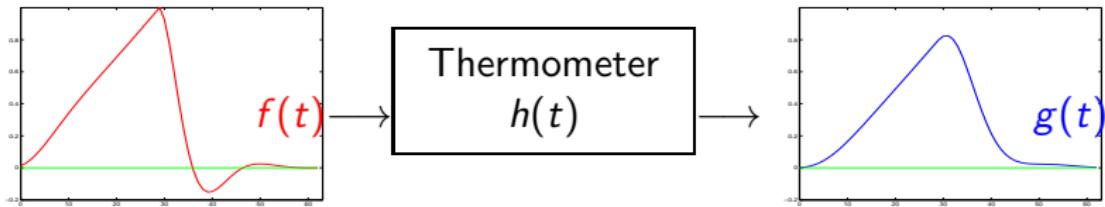
Given the forward model \mathcal{H} (impulse response $h(t)$))
and a set of data $g(t_i), i = 1, \dots, M$
find $f(t)$



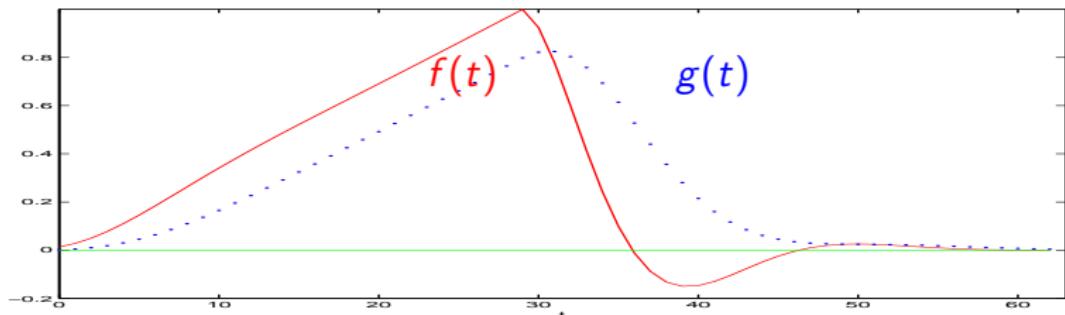
Measuring variation of temperature with a thermometer

Forward model: Convolution

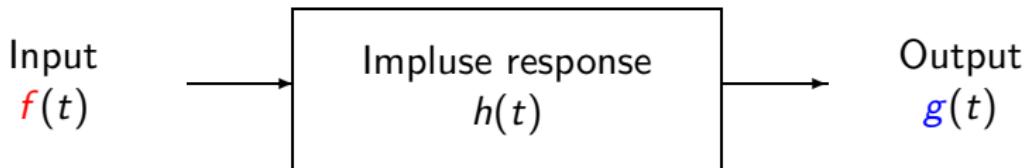
$$g(t) = \int f(t') h(t - t') dt' + \epsilon(t)$$



Inversion: Deconvolution



Instrumentation



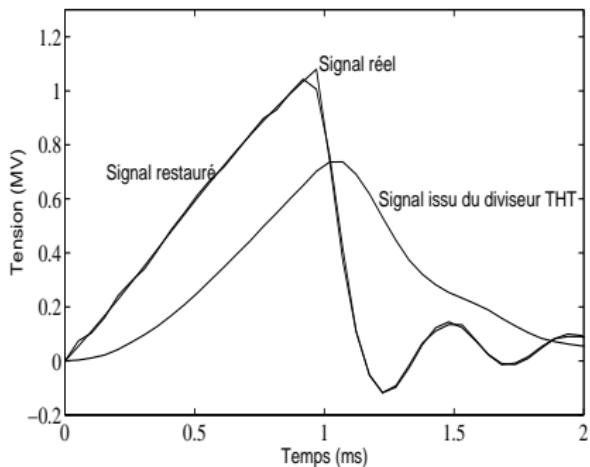
- ▶ Ideal Instrument $g(t) = f(t)$ does not exist.
- ▶ A linear and time invariant instrument is characterized by its impulse response $h(t)$.
- ▶ Ideal Instrument $h(t) = \delta(t)$ does not exist.
- ▶ **Forward problem:** $f(t), h(t) \rightarrow g(t) = h(t) * f(t)$
- ▶ Two linked problems in instrumentation: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

Ex1: Isolators resistivity against lightning strike

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

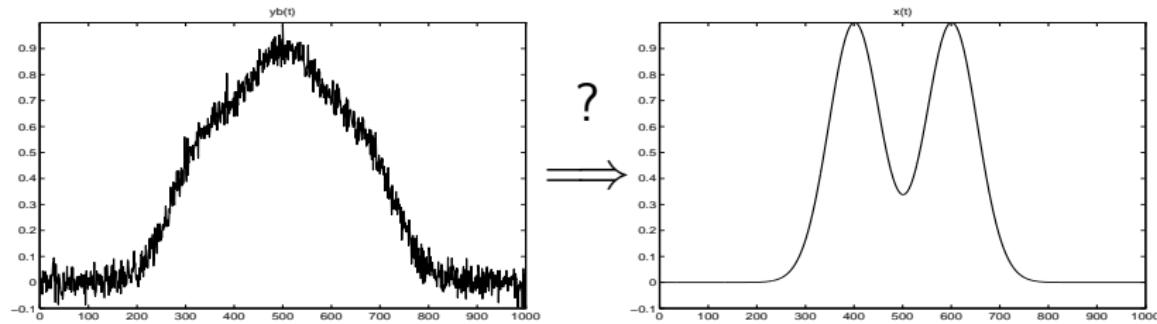


EDF – Les Renardières

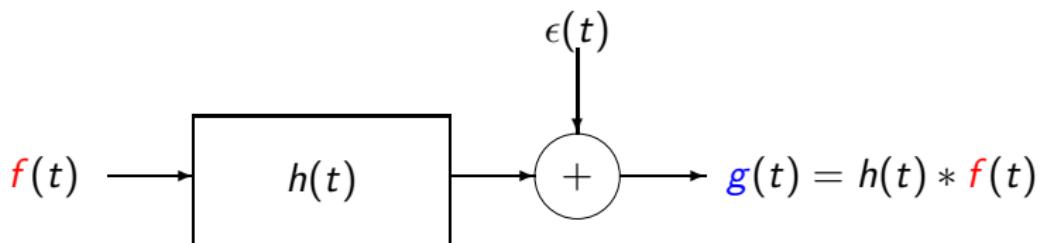


Real and Estimated

Ex2: Radio-astronomy

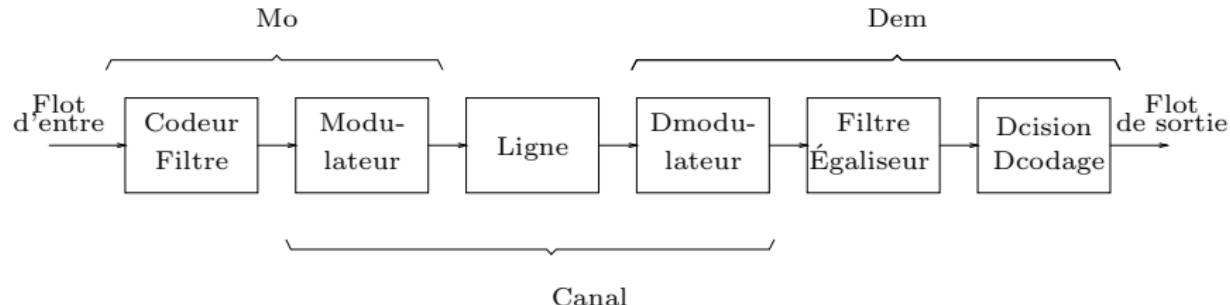


Forward model:

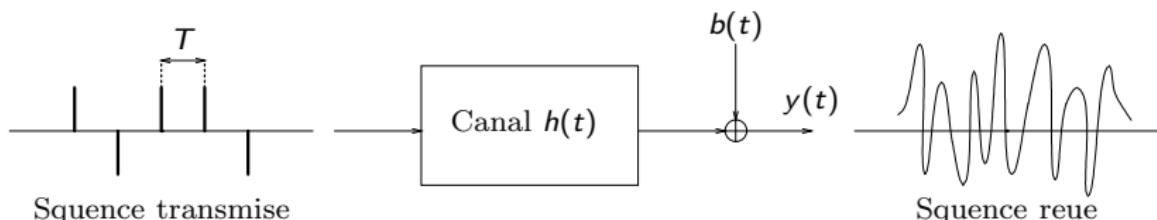


Telecommunication: transmission channel compensation

► 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



$$g(x, y) = \iint f(x', y') h(x - x', y - y') dx' dy' + \epsilon(x, y)$$

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

- ▶ Inverse problem: Image restoration

Given the forward model \mathcal{H} (PSF $h(x, y)$))

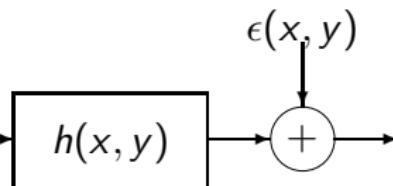
and a set of data $g(x_i, y_i), i = 1, \dots, M$

find $f(x, y)$

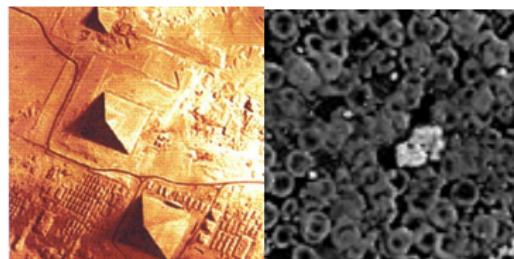
Making an image with an unfocused camera

Forward model: 2D Convolution

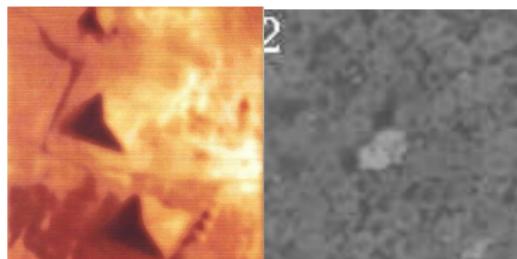
$$g(x, y) = \iint f(x', y') h(x - x', y - y') dx' dy' + \epsilon(x, y)$$

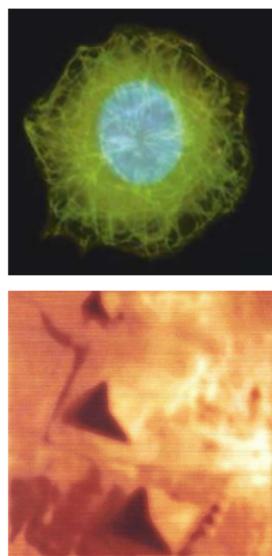
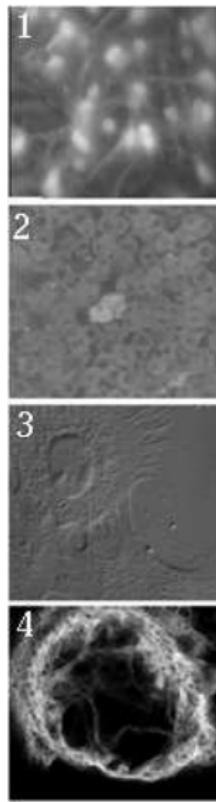


Inversion: Image Deconvolution or Restoration



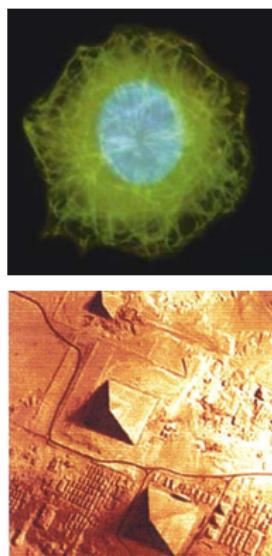
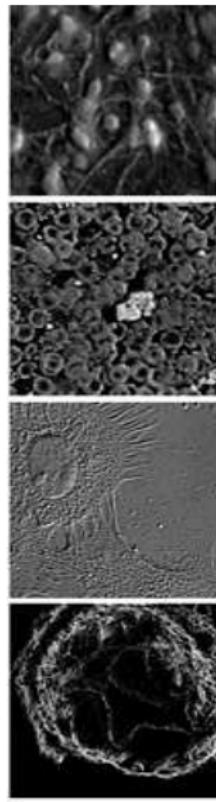
?
↔





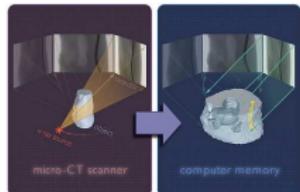
?

⇒



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 radiograph $g_\phi(r, z)$
- ▶ Forward model:
Line integrals or Radon Transform



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

- ▶ Inverse problem: Image reconstruction

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

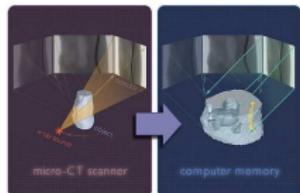
Making an image of the interior of a body

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



- ▶ Forward model:
Line integrals or Radon Transform

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

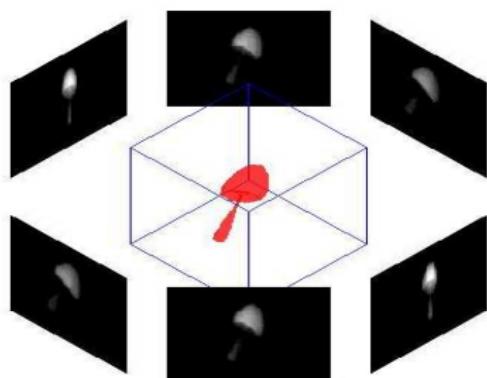


- ▶ Inverse problem: Image reconstruction

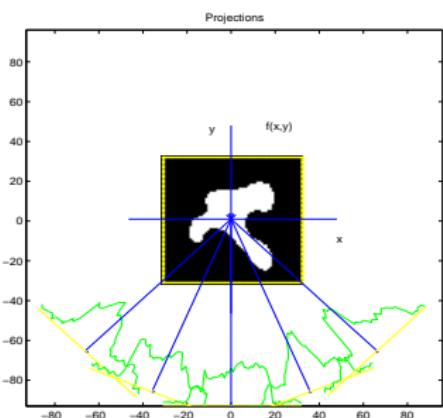
Given the forward model \mathcal{H} (Radon Transform) and
a set of data $g_{\phi_i}(r), i = 1, \dots, M$
find $f(x, y)$

2D and 3D Computed Tomography

3D



2D

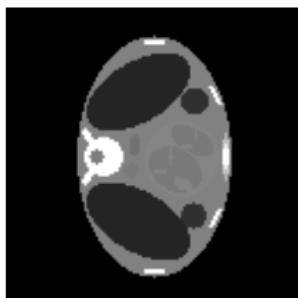
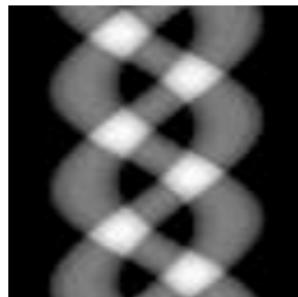


$$g_\phi(r_1, r_2) = \int_{\mathcal{L}_{r_1, r_2, \phi}} f(x, y, z) \, dl \quad g_\phi(r) = \int_{\mathcal{L}_{r, \phi}} f(x, y) \, dl$$

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

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

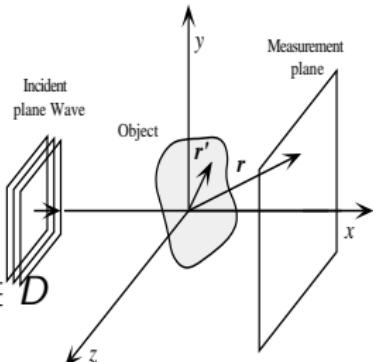
Measrs: diffracted wave by the object $g(\mathbf{r}_i)$

Unknown quantity: $f(\mathbf{r}) = k_0^2(n^2(\mathbf{r}) - 1)$

Intermediate quantity : $\phi(\mathbf{r})$

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

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

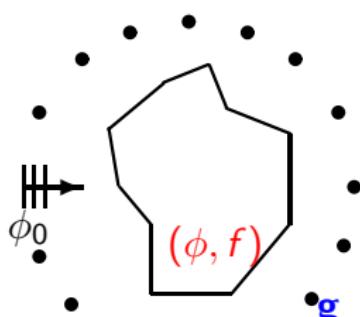


Born approximation ($\phi(\mathbf{r}') \simeq \phi_0(\mathbf{r}')$):

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

Discretization :

$$\begin{cases} \mathbf{g} = \mathbf{G}_m \mathbf{F} \phi \\ \phi = \phi_0 + \mathbf{G}_o \bar{\mathbf{F}} \phi \end{cases} \xrightarrow{\text{with } \mathbf{F} = \text{diag}(\mathbf{f})} \begin{cases} \mathbf{g} = \mathbf{H}(\mathbf{f}) \\ \mathbf{H}(\mathbf{f}) = \mathbf{G}_m \mathbf{F} (\mathbf{I} - \mathbf{G}_o \mathbf{F})^{-1} \phi_0 \end{cases}$$



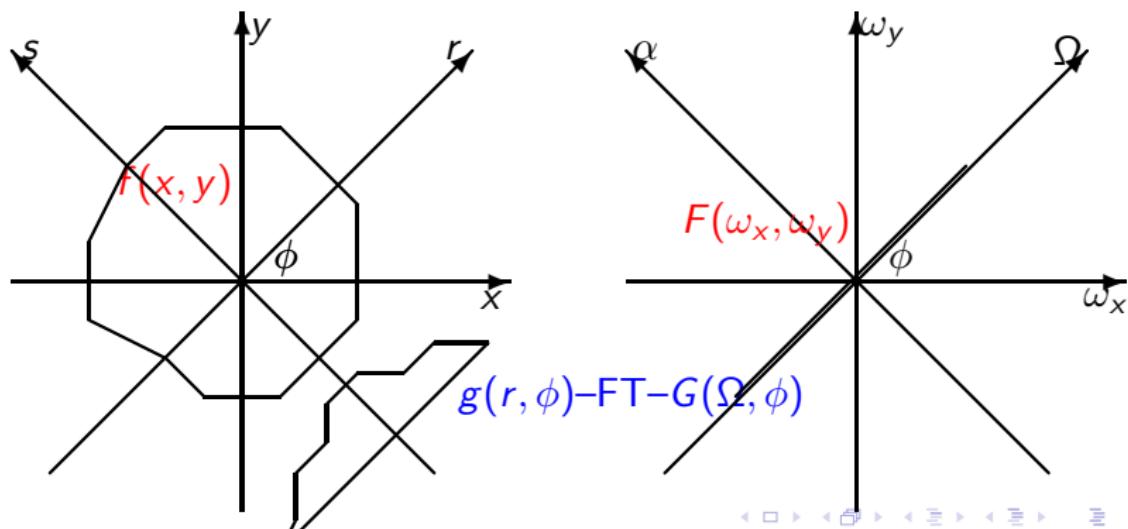
Fourier Synthesis in X ray Tomography

$$g(r, \phi) = \iint f(x, y) \delta(r - x \cos \phi - y \sin \phi) dx dy$$

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

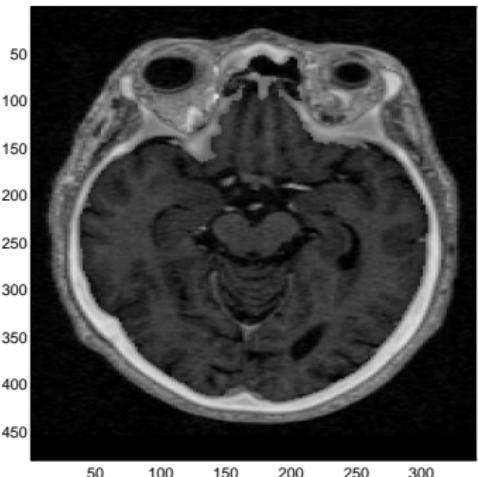
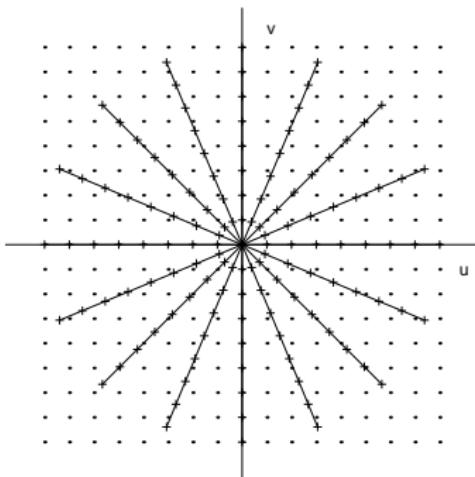
$$F(\omega_x, \omega_y) = \iint f(x, y) \exp \{-j\omega_x x, \omega_y y\} dx dy$$

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



Fourier Synthesis in X ray tomography

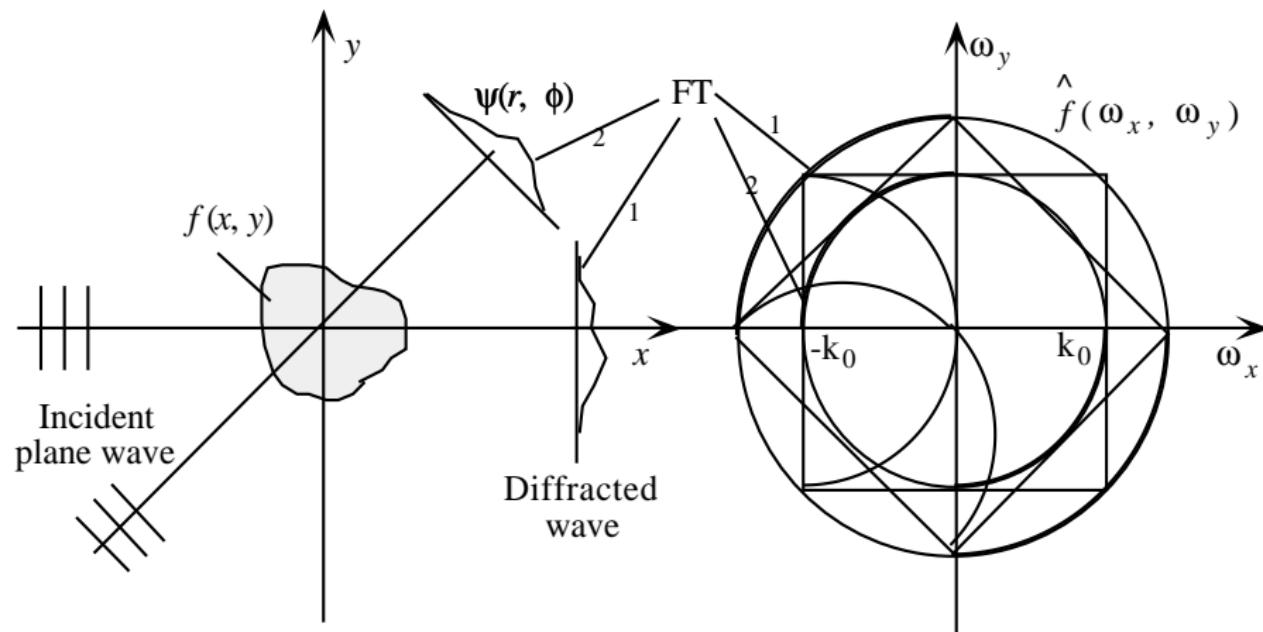
$$G(\omega_x, \omega_y) = \iint f(x, y) \exp \{-j(\omega_x x + \omega_y y)\} dx dy$$



Forward problem: Given $f(x, y)$ compute $G(\omega_x, \omega_y)$

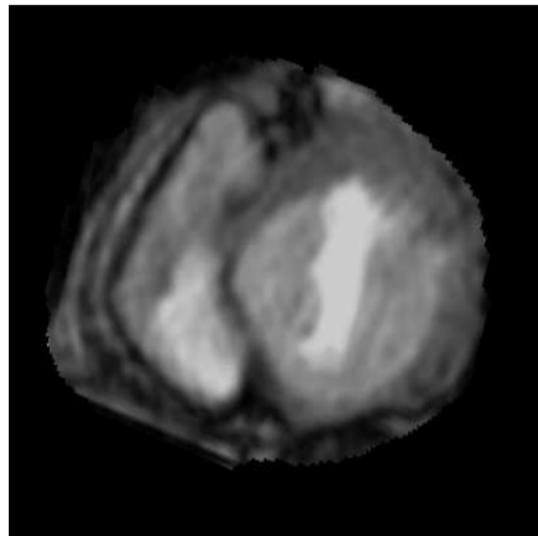
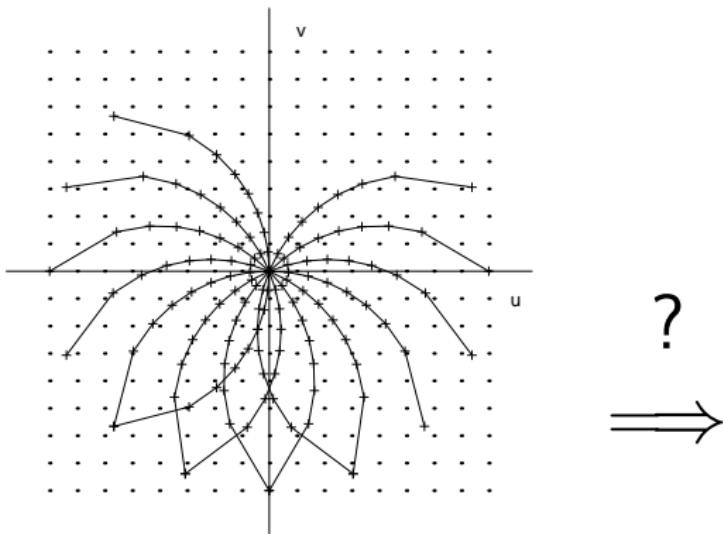
Inverse problem: Given $G(\omega_x, \omega_y)$ on those lines
estimate $f(x, y)$

Fourier Synthesis in Diffraction tomography



Fourier Synthesis in Diffraction tomography

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

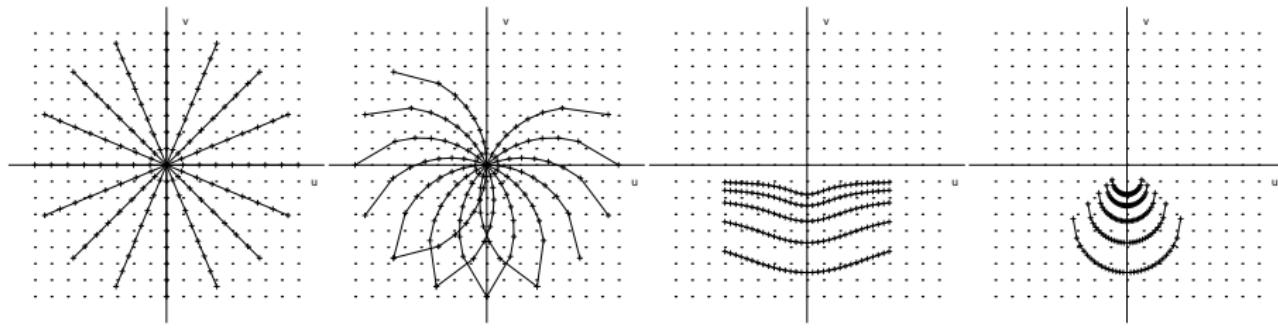


Forward problem: Given $f(x, y)$ compute $G(\omega_x, \omega_y)$

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

Fourier Synthesis in different imaging systems

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



X ray Tomography

Diffraction

Eddy current

SAR & Radar

Forward problem: Given $f(x, y)$ compute $G(\omega_x, \omega_y)$

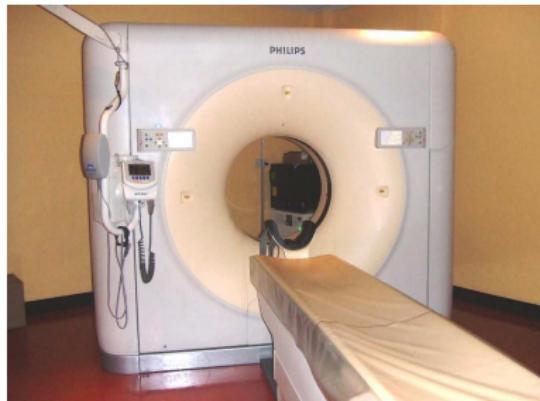
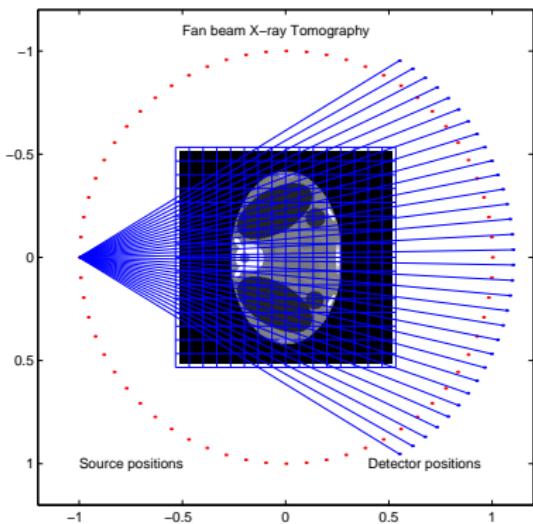
Inverse problem : Given $G(\omega_x, \omega_y)$ on those algebraic lines, circles or curves, estimate $f(x, y)$

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



$$g(s_i) = \int_{L_i} f(\mathbf{r}) \, d\mathbf{l}_i + \epsilon(s_i)$$

Discretization

$$\mathbf{g} = \mathbf{Hf} + \boldsymbol{\epsilon}$$

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

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

model measured data unknown quantity intermediate quantity errors and noise

Particular cases:

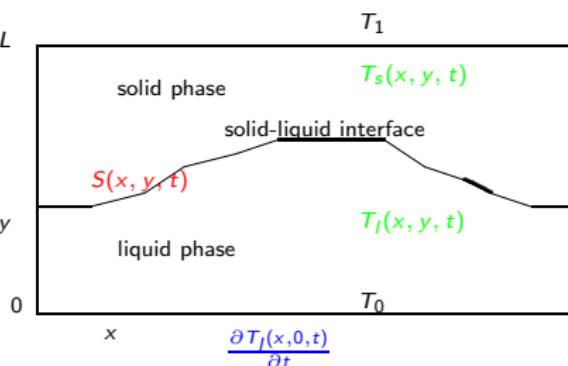
- Implicit model linking \mathbf{f} and \mathbf{z} :
$$\begin{cases} \mathbf{g} = \mathcal{H}_1(\mathbf{f}, \mathbf{z}) + \boldsymbol{\epsilon} \\ \mathcal{H}_2(\mathbf{f}, \mathbf{z}) = 0 \end{cases}$$
- Simple non linear model:
$$\mathbf{g} = \mathcal{H}(\mathbf{f}) + \boldsymbol{\epsilon}$$
- Linear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

Time evolution of liquid-solid fusion interface

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

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

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



\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

unknown quantity:

temperature field $T_s(x, y, t)$ et $T_l(x, y, t)$

General formulation of inverse problems

- ▶ General non linear inverse problems:

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

- ▶ Linear models:

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

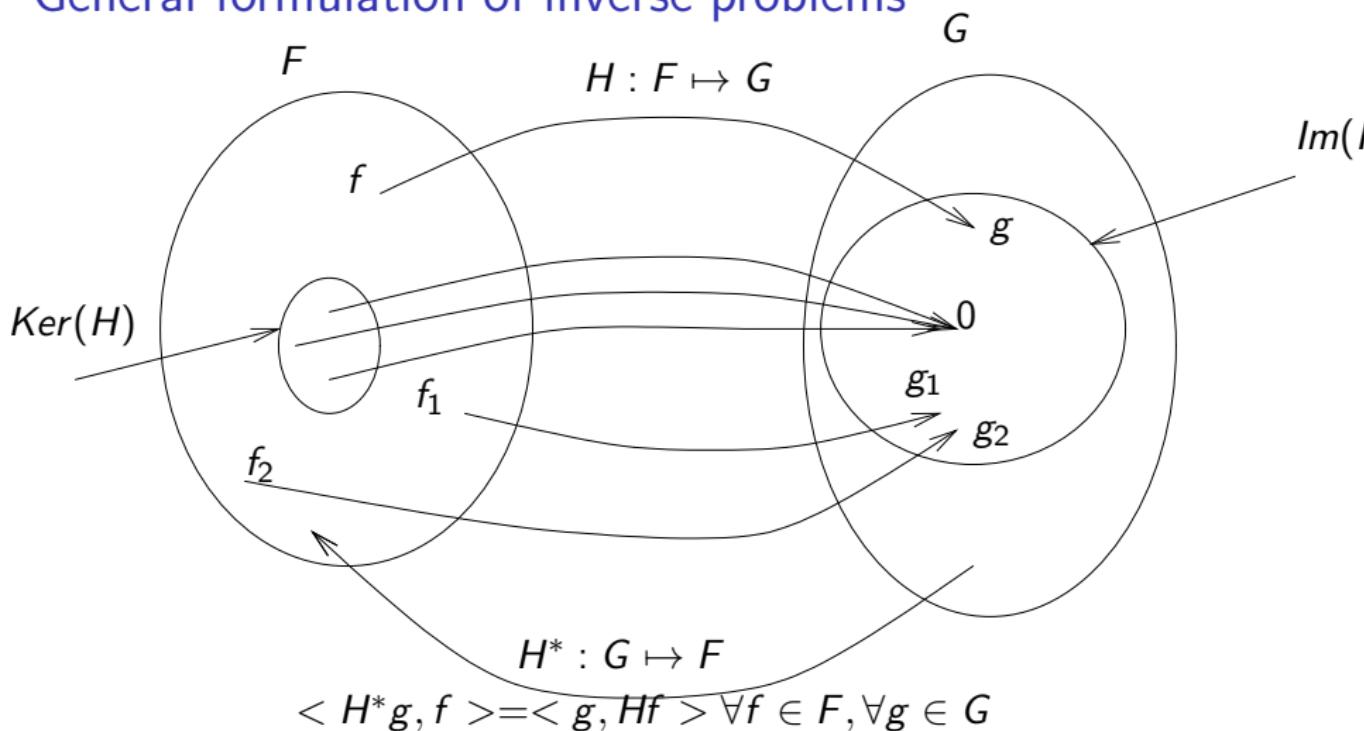
If $h(\mathbf{r}, \mathbf{s}) = h(\mathbf{r} - \mathbf{s}) \rightarrow$ Convolution.

- ▶ Discrete data:

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

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

General formulation of inverse problems



Inverse problems scientific communities

Two communities working on Inverse problems:

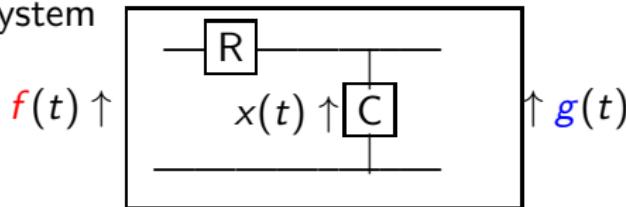
- ▶ Mathematical departments:
Analytical methods: Existence and Uniqueness
Differential equations, PDE
- ▶ Engineering and Computer sciences:
Algebraic methods: Discretization, Uniqueness and Stability
Integral equations, Discretization using Moments method,
Galerkin, ...

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

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

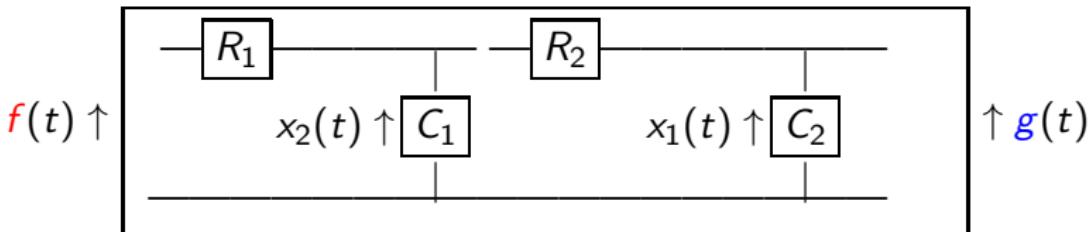
- ▶ State Space Modelling

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

- ▶ 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

$$\begin{cases} \begin{bmatrix} \frac{\partial x_1(t)}{\partial t} \\ \frac{\partial x_2(t)}{\partial t} \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} f(t) \\ g(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} x_1(t) \end{cases}$$

- ▶ 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 $\theta = a$ 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 $\theta = a$ and $f(t), t > 0$, find $x(t), t > 0$
 - ▶ Inverse: Given $\theta = a$ 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 $\theta = (m, c, k)$, 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 $\theta = (m, c, k)$, x_0, v_0 and $F(t)$, $t > 0$, find $x(t)$, $t > 0$
 - ▶ Inverse: Given $\theta = (m, c, k)$, v_0 and $x(t)$, $t > 0$, find $F(t)$

Input-Output model

- ▶ Linear Systems

- ▶ Single Input Single Output (SISO) systems

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

- ▶ Multi Input Multi Output (MIMO) systems

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

- ▶ Linear Time Invariant System

- ▶ SISO Convolution

$$y(t) = h(t) * u(t) = \int h(t - \tau) u(\tau) d\tau$$

- ▶ MIMO Convolution

$$\mathbf{y}(t) = \int \mathbf{H}(t - \tau) \mathbf{u}(\tau) d\tau$$

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

State space model: Continuous case

Dynamic systems:

- ▶ Single Input Single Output (SISO) system:

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

- ▶ Multiple Input Multiple Output (MIMO) system:

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

H, **B**, **C** and **D** are the matrices of the system.

Analytical methods (mathematical physics)

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

$$g(\mathbf{s}) = \int h(\mathbf{s}, \mathbf{r}) \mathbf{f}(\mathbf{r}) \, d\mathbf{r}$$

$$\widehat{\mathbf{f}}(\mathbf{r}) = \int w(\mathbf{s}, \mathbf{r}) g(\mathbf{s}) \, d\mathbf{s}$$

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

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

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

Analytical methods

- Trivial solution:

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

Example: Fourier Transform:

$$g(\mathbf{s}) = \int f(\mathbf{r}) \exp\{-j\mathbf{s} \cdot \mathbf{r}\} d\mathbf{r}$$

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

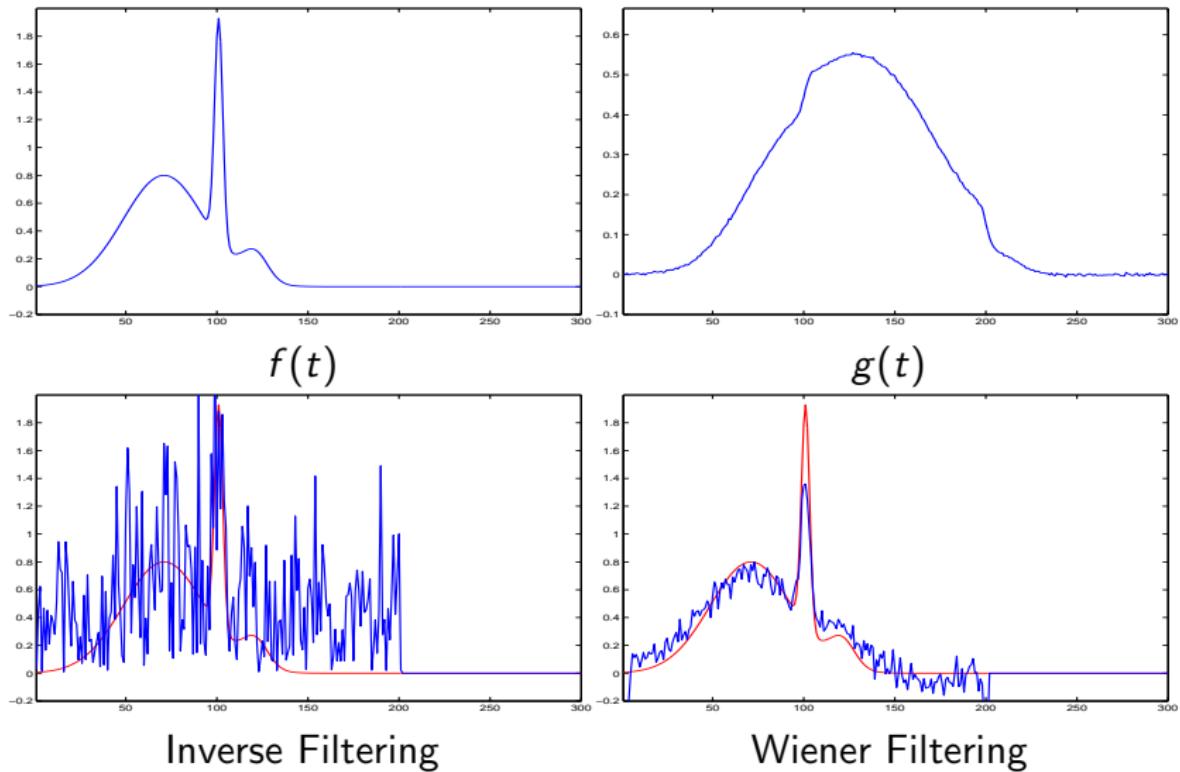
$$\hat{f}(\mathbf{r}) = \int g(\mathbf{s}) \exp\{+j\mathbf{s} \cdot \mathbf{r}\} d\mathbf{s}$$

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

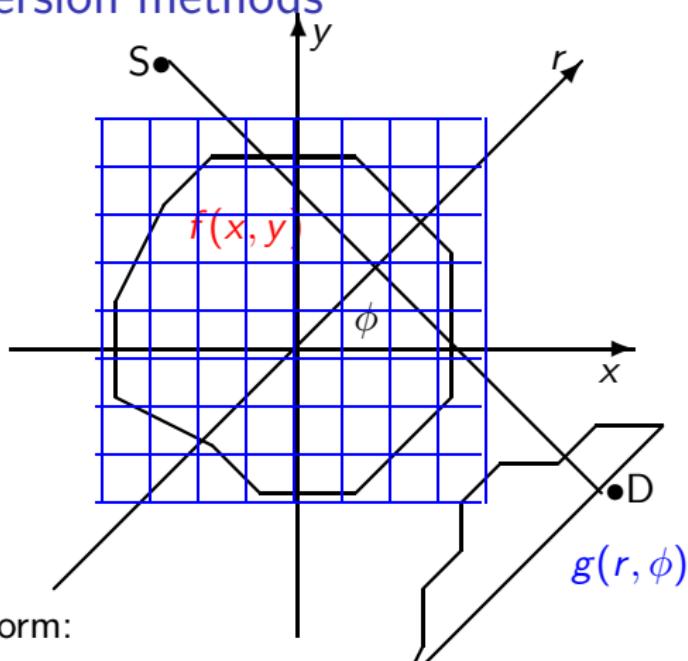
Deconvolution: Analytical methods

Time domain	Fourier domain
<p>Forward model:</p> $g(t) = h(t) * f(t) + \epsilon(t)$	$G(\omega) = H(\omega) F(\omega) + E(\omega)$
<p>Deconvolution:</p> $g(t) \rightarrow \boxed{w(t) = IFT\left\{\frac{1}{H(\omega)}\right\}} \rightarrow \hat{f}(t)$	<p>Inverse filtering</p> $G(\omega) \rightarrow \boxed{\frac{1}{H(\omega)}} \rightarrow \hat{F}(\omega)$
<p>Deconvolution:</p> $g(t) \rightarrow \boxed{W(\omega)} \rightarrow \hat{f}(t)$	<p>Wiener filtering</p> $G(\omega) \rightarrow \boxed{\frac{H^*(\omega)}{ H(\omega) ^2 + \frac{S_\epsilon(\omega)}{S_f(\omega)}}} \rightarrow \hat{F}(\omega)$

Deconvolution example



Analytical Inversion methods

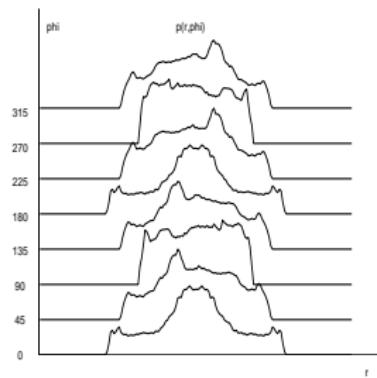
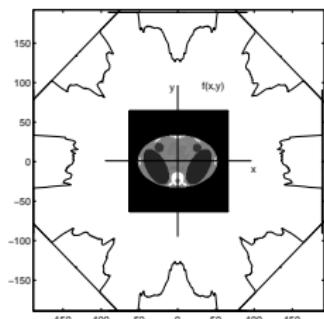


Radon Transform:

$$g(r, \phi) = \iint_D f(x, y) \delta(r - x \cos \phi - y \sin \phi) dx dy$$

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

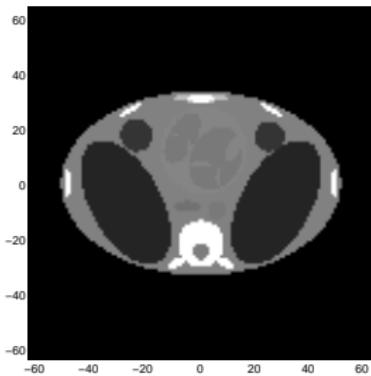
X ray Tomography



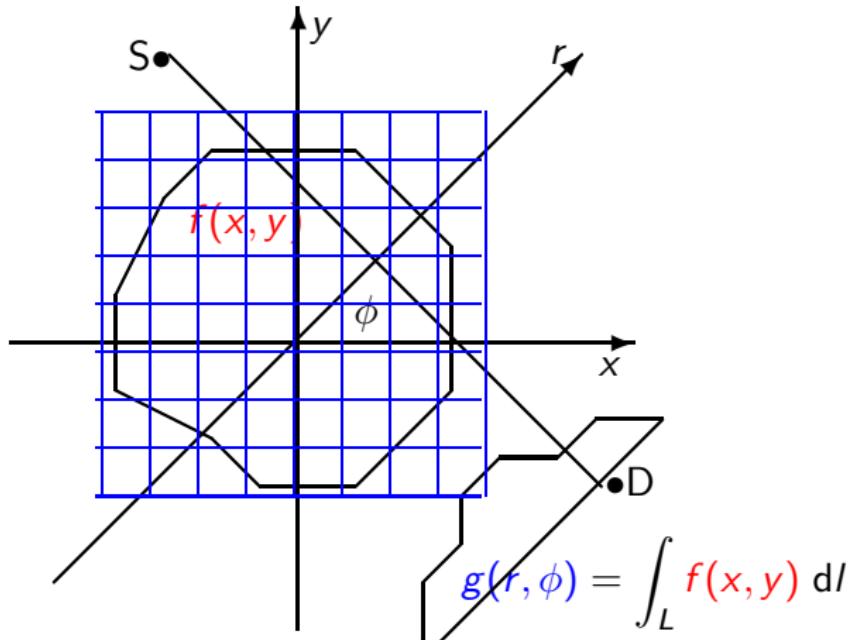
$$g(r, \phi) = -\ln \left(\frac{I}{I_0} \right) = \int_{L_{r,\phi}} f(x, y) \, dx$$
$$g(r, \phi) = \iint_D f(x, y) \delta(r - x \cos \phi - y \sin \phi) \, dx \, dy$$



IRT ?



Analytical Inversion methods



$$g(r, \phi) = \iint_D f(x, y) \delta(r - x \cos \phi - y \sin \phi) \, dx \, dy$$

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

Filtered Backprojection method

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

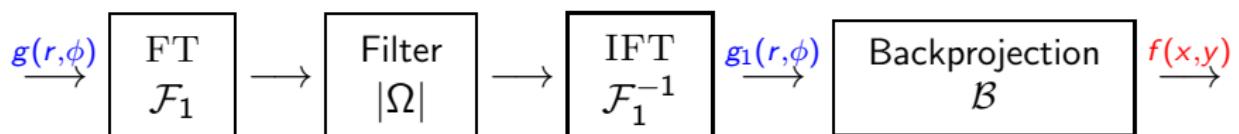
Derivation \mathcal{D} : $\bar{g}(r, \phi) = \frac{\partial g(r, \phi)}{\partial r}$

Hilbert Transform \mathcal{H} : $g_1(r', \phi) = \frac{1}{\pi} \int_0^\infty \frac{\bar{g}(r, \phi)}{(r - r')} dr$

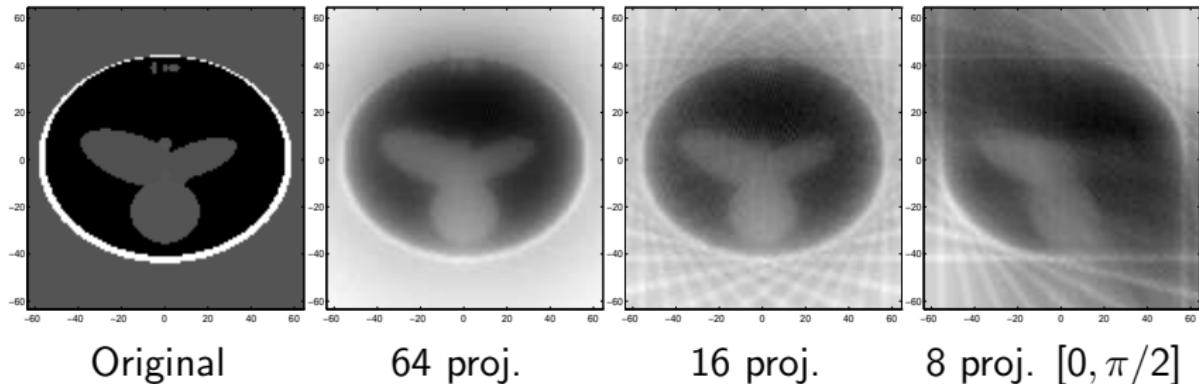
Backprojection \mathcal{B} : $f(x, y) = \frac{1}{2\pi} \int_0^\pi g_1(r' = x \cos \phi + y \sin \phi, \phi) d\phi$

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

- Backprojection of filtered projections:

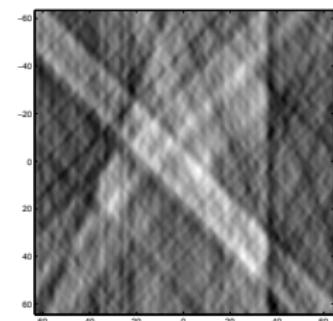
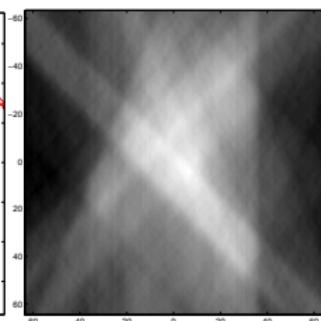
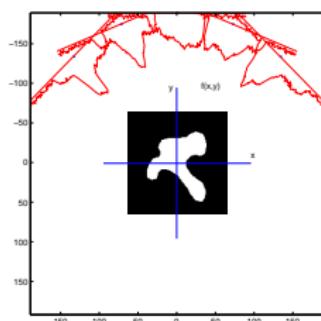
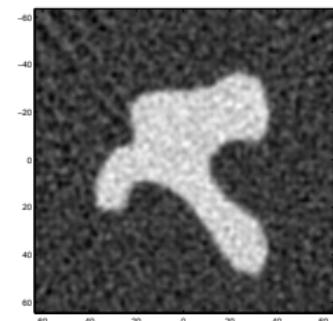
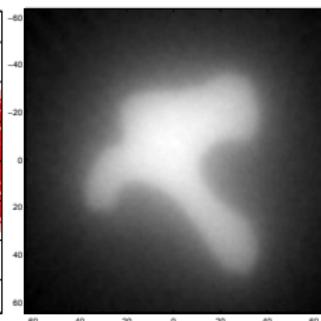
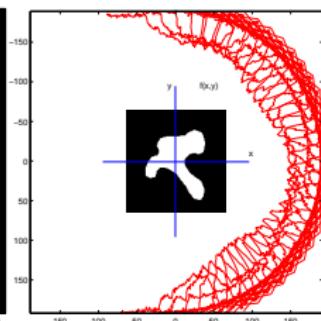
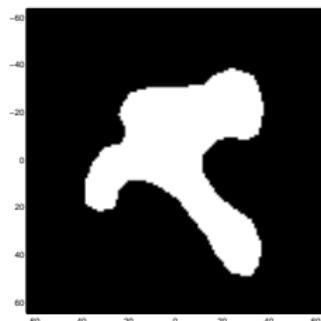


Limitations : Limited angle or noisy data



- ▶ 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

- ▶ $f(\mathbf{r})$ is described in a parametric form with a very few number of parameters $\boldsymbol{\theta}$ and one searches $\hat{\boldsymbol{\theta}}$ 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 :
with different functions ϕ (L_1 , Hubert, ...).
$$Q(\boldsymbol{\theta}) = \sum_i \phi(|g_i - [\mathcal{H} f(\boldsymbol{\theta})]_i|)$$
- ▶ Likelihood :
$$\mathcal{L}(\boldsymbol{\theta}) = -\ln p(\mathbf{g}|\boldsymbol{\theta})$$
- ▶ Penalized likelihood :
$$\mathcal{L}(\boldsymbol{\theta}) = -\ln p(\mathbf{g}|\boldsymbol{\theta}) + \lambda \Omega(\boldsymbol{\theta})$$

Examples:

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

Non parametric methods

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

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

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

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

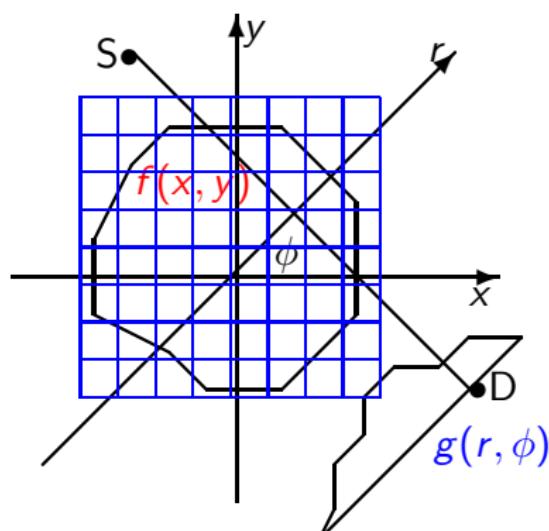
$$g(\mathbf{s}_i) = g_i \simeq \sum_{j=1}^N f_j \int h(\mathbf{s}_i, \mathbf{r}) b_j(\mathbf{r}) d\mathbf{r}, \quad i = 1, \dots, M$$

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

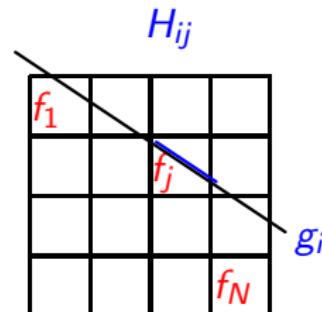
- \mathbf{H} is huge dimensional
- LS solution : $\widehat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{Q(\mathbf{f})\}$ with
$$Q(\mathbf{f}) = \sum_i |g_i - [\mathbf{H}\mathbf{f}]_i|^2 = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2$$

does not give satisfactory result.

Algebraic methods: Discretization



$$g(r, \phi) = \int_L f(x, y) \, dl$$



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

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

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

Inversion: Deterministic methods

Data matching

- ▶ Observation model

$$g_i = h_i(\mathbf{f}) + \epsilon_i, \quad i = 1, \dots, M \longrightarrow \mathbf{g} = \mathbf{H}(\mathbf{f}) + \boldsymbol{\epsilon}$$

- ▶ Misatch between data and output of the model $\Delta(\mathbf{g}, \mathbf{H}(\mathbf{f}))$

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{\Delta(\mathbf{g}, \mathbf{H}(\mathbf{f}))\}$$

- ▶ Examples:

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

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

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

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

Regularization theory

Inverse problems = III posed problems
→ Need for prior information

Functional space (Tikhonov):

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

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

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

$$J(\mathbf{f}) = Q(\mathbf{g} - \mathbf{H}(\mathbf{f})) + \lambda \Omega(\mathbf{Df})$$

or

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

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 → 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

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

- ▶ Observation model \mathcal{M} + Hypothesis on the noise $\boldsymbol{\epsilon}$ $\rightarrow p(\mathbf{g}|\mathbf{f}; \mathcal{M}) = p_{\boldsymbol{\epsilon}}(\mathbf{g} - \mathbf{H}\mathbf{f})$
- ▶ A priori information $p(\mathbf{f}|\mathcal{M})$
- ▶ Bayes :
$$p(\mathbf{f}|\mathbf{g}; \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathcal{M}) p(\mathbf{f}|\mathcal{M})}{p(\mathbf{g}|\mathcal{M})}$$

Link with regularization :

Maximum A Posteriori (MAP) :

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

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

Case of linear models and Gaussian priors

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

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

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

MAP estimation with other priors:

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

Separable priors:

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

Markovian models:

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

MAP estimation with markovien priors:

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

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

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:

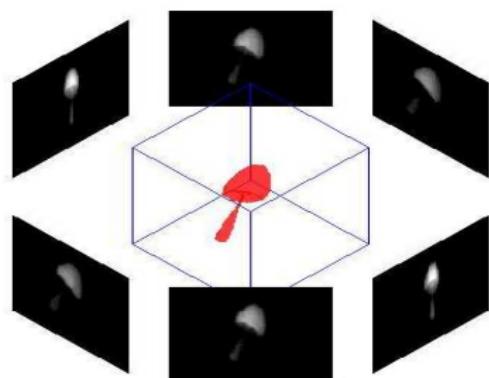
$$\log(1+t^2), \frac{t^2}{1+t^2}, \arctan(t^2), \begin{cases} t^2 & |t| \leq T \\ T^2 & |t| > T \end{cases}$$

Main advantages of the Bayesian approach

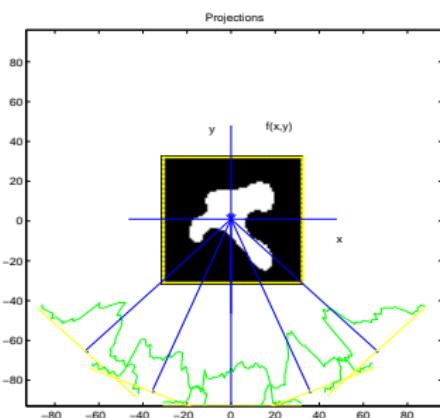
- ▶ MAP = Regularization
- ▶ Posterior mean ? Marginal MAP ?
- ▶ More information in the posterior law than only its mode or its mean
- ▶ Meaning and tools for estimating hyper parameters
- ▶ Meaning and tools for model selection
- ▶ More specific and specialized priors, particularly through the hidden 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



$$g_\phi(r_1, r_2) = \int_{\mathcal{L}_{r_1, r_2, \phi}} f(x, y, z) \, dl \quad g_\phi(r) = \int_{\mathcal{L}_{r, \phi}} f(x, y) \, dl$$

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

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

Inverse problems: Discretization

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

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

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

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

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

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

- \mathbf{H} is huge dimensional
- LS solution : $\widehat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{Q(\mathbf{f})\}$ with
$$Q(\mathbf{f}) = \sum_i |g_i - [\mathbf{H}\mathbf{f}]_i|^2 = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2$$

does not give satisfactory result.

Inverse problems: Deterministic methods

Data matching

- ▶ Observation model

$$g_i = h_i(\mathbf{f}) + \epsilon_i, \quad i = 1, \dots, M \longrightarrow \mathbf{g} = \mathbf{H}(\mathbf{f}) + \boldsymbol{\epsilon}$$

- ▶ Misatch between data and output of the model $\Delta(\mathbf{g}, \mathbf{H}(\mathbf{f}))$

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{\Delta(\mathbf{g}, \mathbf{H}(\mathbf{f}))\}$$

- ▶ Examples:

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

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

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

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

Inverse problems: Regularization theory

Inverse problems = III posed problems

→ Need for prior information

Functional space (Tikhonov):

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

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

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

$$J(\mathbf{f}) = Q(\mathbf{g} - \mathbf{H}(\mathbf{f})) + \lambda \Omega(\mathbf{Df})$$

or

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

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

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

- ▶ Observation model \mathcal{M} + Hypothesis on the noise $\boldsymbol{\epsilon}$ $\rightarrow p(\mathbf{g}|\mathbf{f}; \mathcal{M}) = p_{\boldsymbol{\epsilon}}(\mathbf{g} - \mathbf{H}\mathbf{f})$
- ▶ A priori information $p(\mathbf{f}|\mathcal{M})$
- ▶ Bayes :
$$p(\mathbf{f}|\mathbf{g}; \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}; \mathcal{M}) p(\mathbf{f}|\mathcal{M})}{p(\mathbf{g}|\mathcal{M})}$$

Link with regularization :

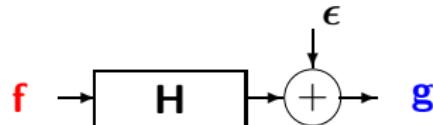
Maximum A Posteriori (MAP) :

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

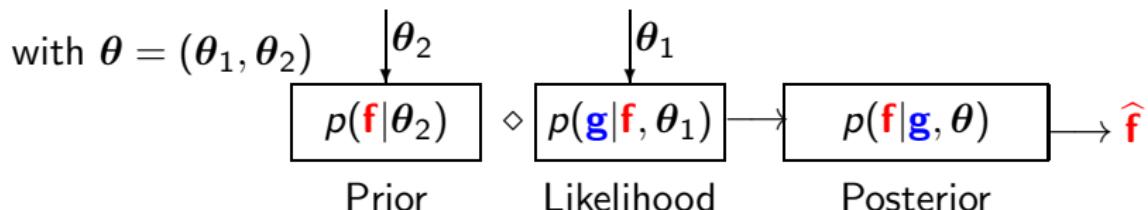
with $Q(\mathbf{g}, \mathbf{H}\mathbf{f}) = -\ln p(\mathbf{g}|\mathbf{f})$ and $\lambda\Omega(\mathbf{f}) = -\ln p(\mathbf{f})$

Bayesian inference for inverse problems

- ▶ Linear Inverse problems: $\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$
- ▶ Bayesian inference:



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



- ▶ Point estimators:
 - ▶ Maximum A Posteriori (MAP): $\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \{p(\mathbf{f}|\mathbf{g}, \boldsymbol{\theta})\}$
 - ▶ Posterior Mean (PM): $\hat{\mathbf{f}} = E_{p(\mathbf{f}|\mathbf{g}, \boldsymbol{\theta})} \{\mathbf{f}\} = \int \mathbf{f} p(\mathbf{f}|\mathbf{g}, \boldsymbol{\theta}) d\mathbf{f}$

Bayesian Estimation: Two simple priors

- ▶ Example 1: Linear Gaussian case:

$$\begin{cases} p(\mathbf{g}|\mathbf{f}, \theta_1) = \mathcal{N}(\mathbf{H}\mathbf{f}, \theta_1\mathbf{I}) \\ p(\mathbf{f}|\theta_2) = \mathcal{N}(0, \theta_2\mathbf{I}) \end{cases} \longrightarrow p(\mathbf{f}|\mathbf{g}, \boldsymbol{\theta}) = \mathcal{N}(\hat{\mathbf{f}}, \hat{\mathbf{P}})$$

with

$$\begin{cases} \hat{\mathbf{P}} = (\mathbf{H}'\mathbf{H} + \lambda\mathbf{I})^{-1}, & \lambda = \frac{\theta_1}{\theta_2} \\ \hat{\mathbf{f}} = \hat{\mathbf{P}}\mathbf{H}'\mathbf{g} \end{cases}$$

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

- ▶ Example 2: Double Exponential prior & MAP:

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

Full Bayesian approach

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

- ▶ Forward & errors model: $\rightarrow p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}_1; \mathcal{M})$
- ▶ Prior models $\rightarrow p(\mathbf{f}|\boldsymbol{\theta}_2; \mathcal{M})$
- ▶ Hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \rightarrow p(\boldsymbol{\theta}|\mathcal{M})$
- ▶ Bayes: $\rightarrow p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}; \mathcal{M}) p(\mathbf{f}|\boldsymbol{\theta}; \mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathbf{g}|\mathcal{M})}$
- ▶ Joint MAP: $(\hat{\mathbf{f}}, \hat{\boldsymbol{\theta}}) = \arg \max_{(\mathbf{f}, \boldsymbol{\theta})} \{p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M})\}$
- ▶ Marginalization:
$$\begin{cases} p(\mathbf{f}|\mathbf{g}; \mathcal{M}) &= \int p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M}) d\boldsymbol{\theta} \\ p(\boldsymbol{\theta}|\mathbf{g}; \mathcal{M}) &= \int p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M}) d\mathbf{f} \end{cases}$$
- ▶ Posterior means:
$$\begin{cases} \hat{\mathbf{f}} &= \int \int \mathbf{f} p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M}) d\boldsymbol{\theta} d\mathbf{f} \\ \hat{\boldsymbol{\theta}} &= \int \int \boldsymbol{\theta} p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M}) d\mathbf{f} d\boldsymbol{\theta} \end{cases}$$
- ▶ Evidence of the model:

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

Full Bayesian: Marginal MAP and PM estimates

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

$$p(\theta|g) = \int p(f, \theta|g) df \propto p(g|\theta) p(\theta)$$

and then $\hat{f} = \arg \max_f \left\{ p(f|\hat{\theta}, g) \right\}$ or

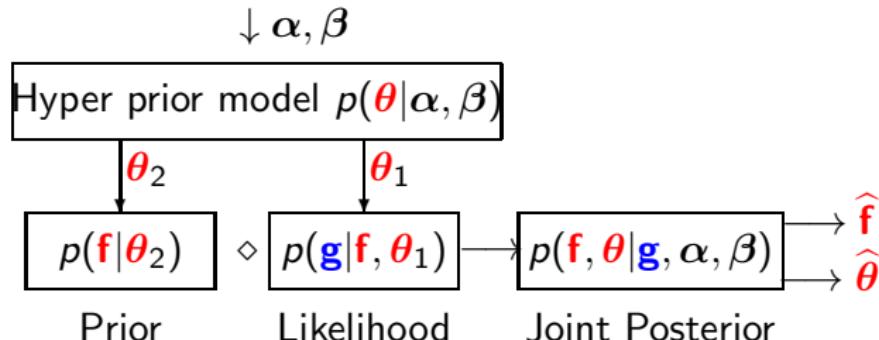
Posterior Mean: $\hat{f} = \int f p(f|\hat{\theta}, g) df$

- Needs the expression of the Likelihood:

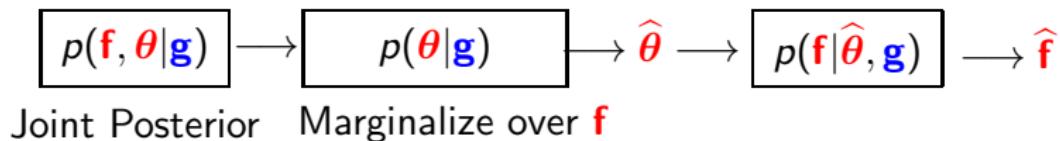
$$p(g|\theta) = \int p(g|f, \theta_1) p(f|\theta_2) df$$

Not always analytically available \rightarrow EM, SEM and GEM algorithms

Full Bayesian Model and Hyperparameter Estimation



Full Bayesian Model and Hyperparameter Estimation scheme



Marginalization for Hyperparameter Estimation

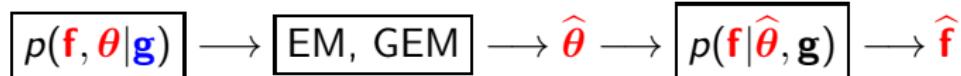
Full Bayesian: EM and GEM algorithms

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

$$\begin{cases} \text{E-step: } Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) = E_{p(\mathbf{f}|\mathbf{g}, \hat{\boldsymbol{\theta}}^{(k)})} \{ \ln p(\mathbf{g}, \mathbf{f}|\boldsymbol{\theta}) \} \\ \text{M-step: } \hat{\boldsymbol{\theta}}^{(k)} = \arg \max_{\boldsymbol{\theta}} \left\{ Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) \right\} \end{cases}$$

- ▶ GEM (Bayesian) algorithm:

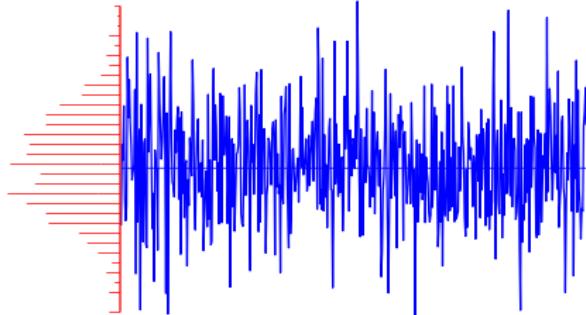
$$\begin{cases} \text{E-step: } Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) = E_{p(\mathbf{f}|\mathbf{g}, \hat{\boldsymbol{\theta}}^{(k)})} \{ \ln p(\mathbf{g}, \mathbf{f}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta}) \} \\ \text{M-step: } \hat{\boldsymbol{\theta}}^{(k)} = \arg \max_{\boldsymbol{\theta}} \left\{ Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k-1)}) \right\} \end{cases}$$



Two main steps in the Bayesian approach

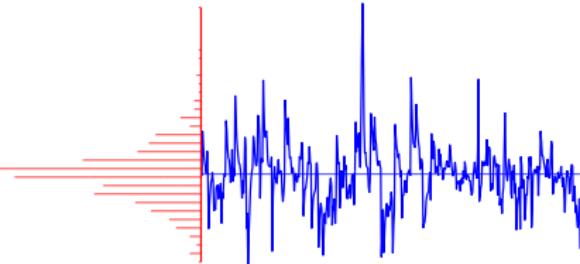
- ▶ Prior modeling
 - ▶ Separable:
Gaussian, Gamma,
Sparsity enforcing: Generalized Gaussian, mixture of Gaussians, mixture of Gammas, ...
 - ▶ Markovian:
Gauss-Markov, GGM, ...
 - ▶ Markovian with **hidden variables**
(contours, region labels)
- ▶ Choice of the estimator and computational aspects
 - ▶ 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



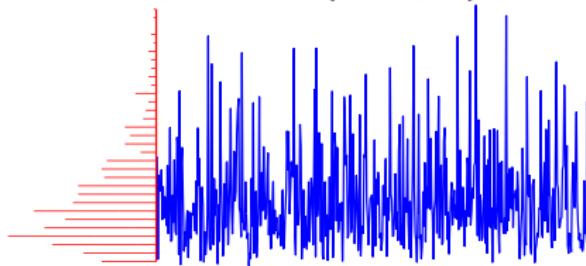
Gaussian

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



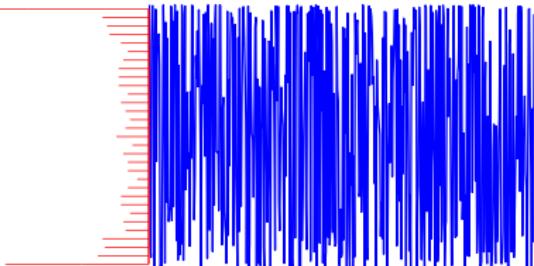
Generalized Gaussian

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



Gamma

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

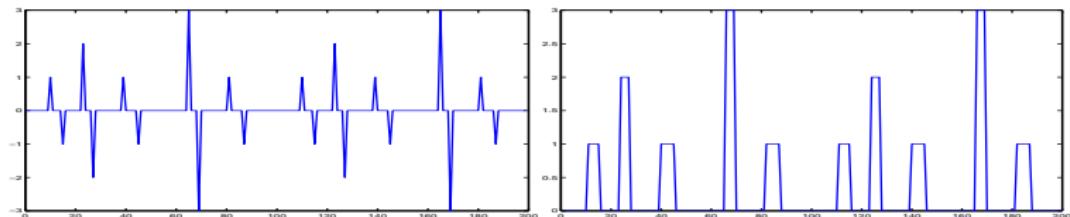


Beta

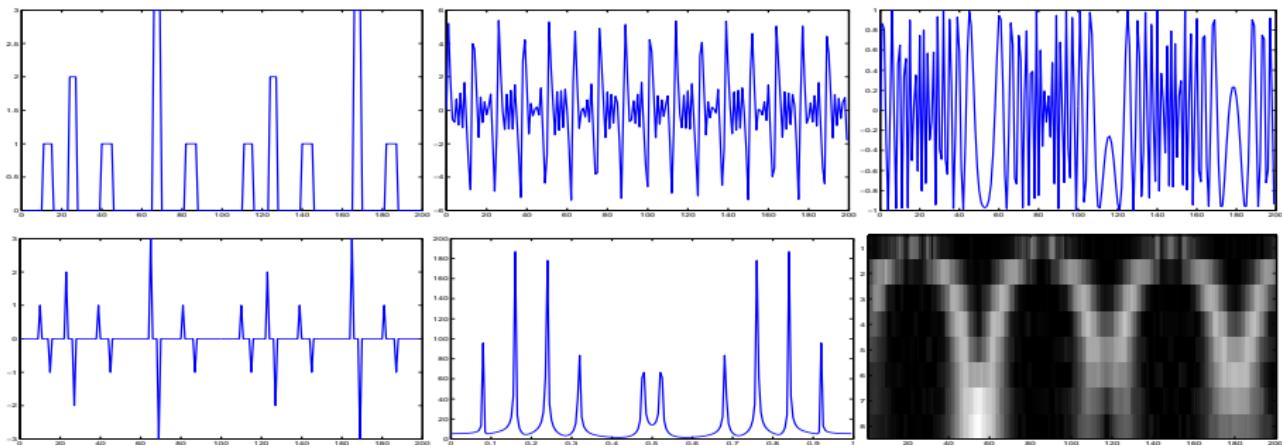
$$p(f_j) \propto f_j^\alpha (1 - f_j)^\beta$$

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

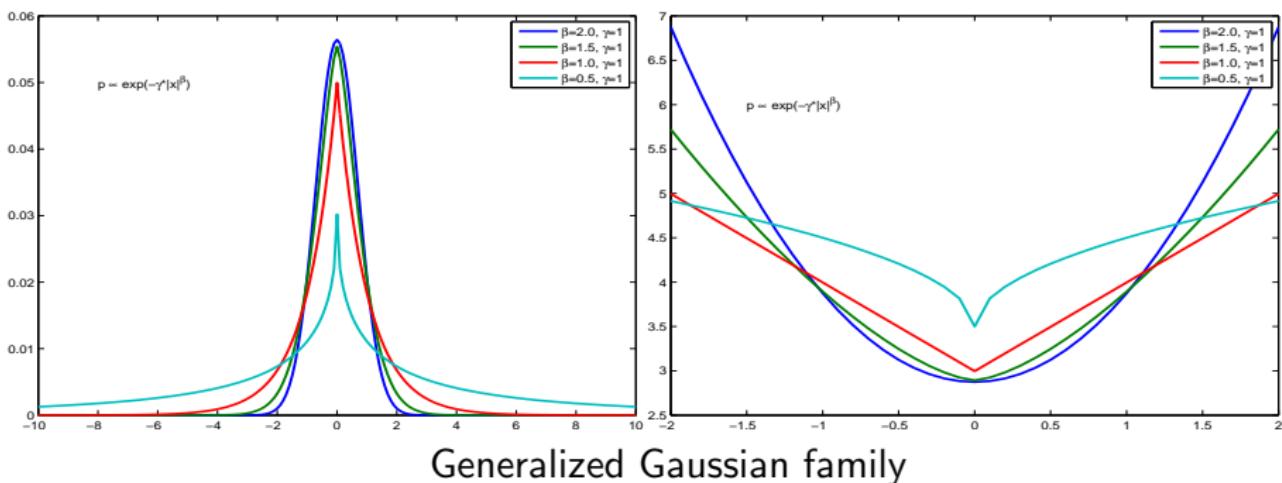
Simple heavy tailed models

- Generalized Gaussian, Double Exponential

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

$\beta = 1$ Double exponential or Laplace.

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



Generalized Gaussian family

Simple heavy tailed models

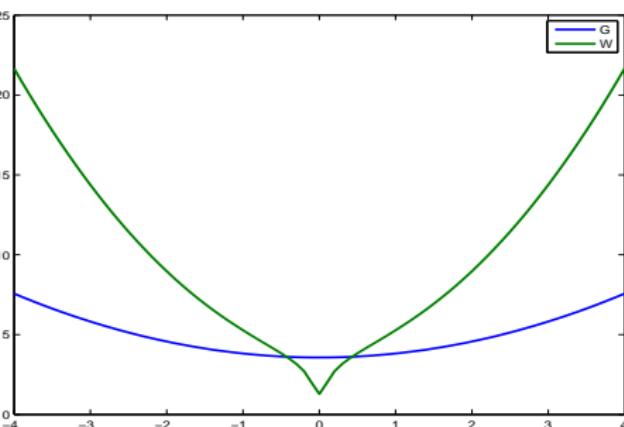
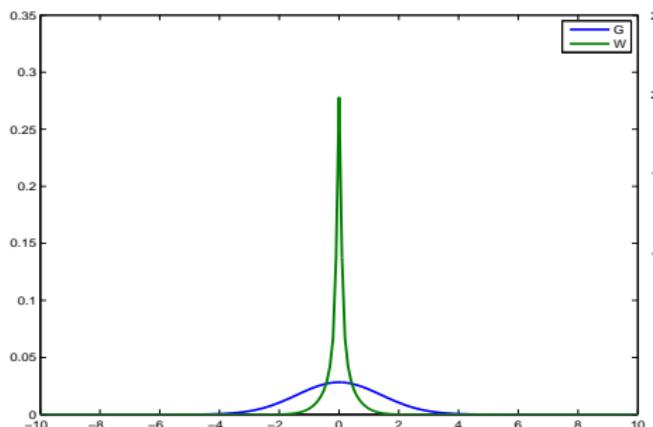
- Symmetric Weibull

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

$\beta = 2$ is the Symmetric Rayleigh distribution.

$\beta = 1$ is the Double exponential and

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

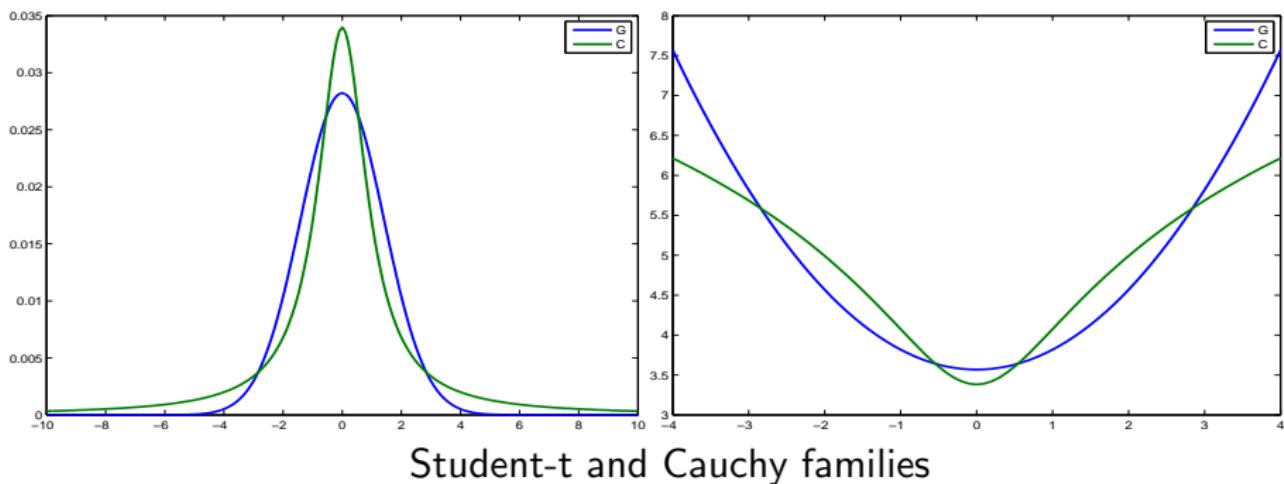


Simple heavy tailed models

- Student-t and Cauchy models

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

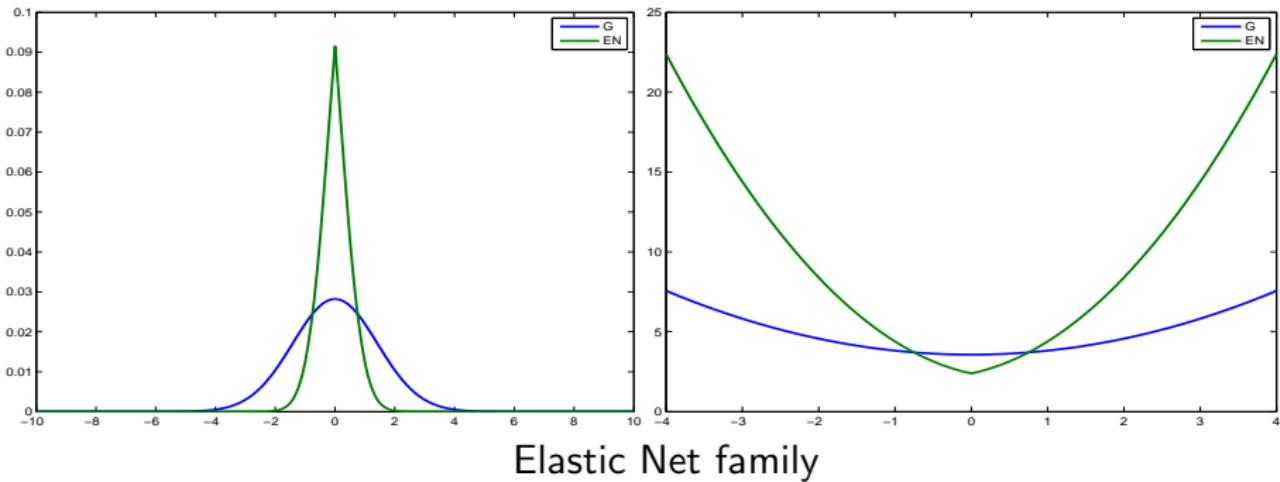
Cauchy model is obtained when $\nu = 1$.



Simple heavy tailed models

- Elastic net prior model

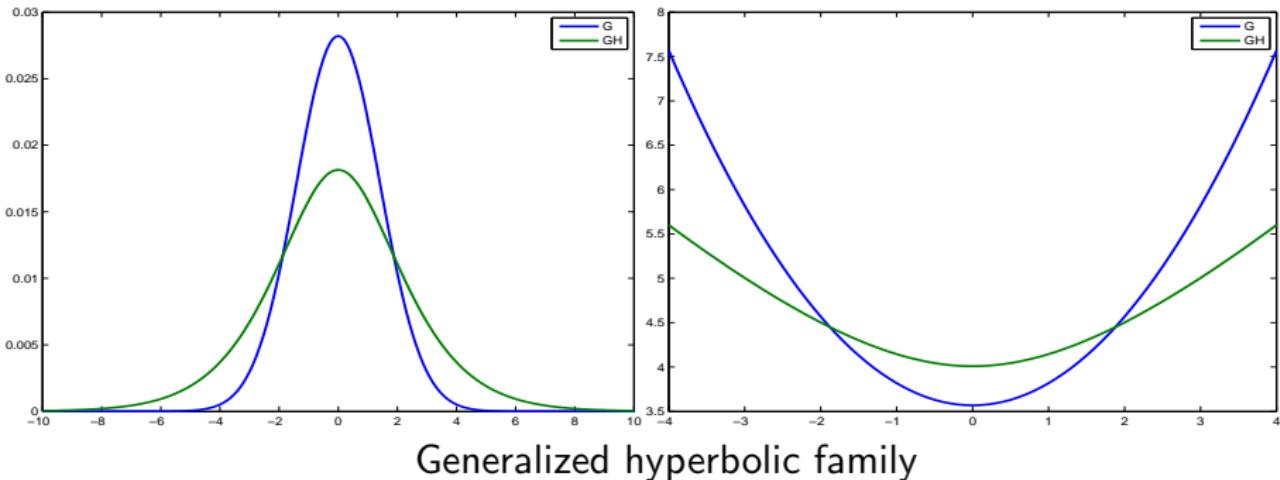
$$p(\mathbf{f}|\nu) = \prod_j \mathcal{EN}(f_j|\nu) \propto \exp \left\{ - \sum_j (\gamma_1 |f_j| + \gamma_2 f_j^2) \right\}$$



Simple heavy tailed models

- Generalized hyperbolic (GH) models

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



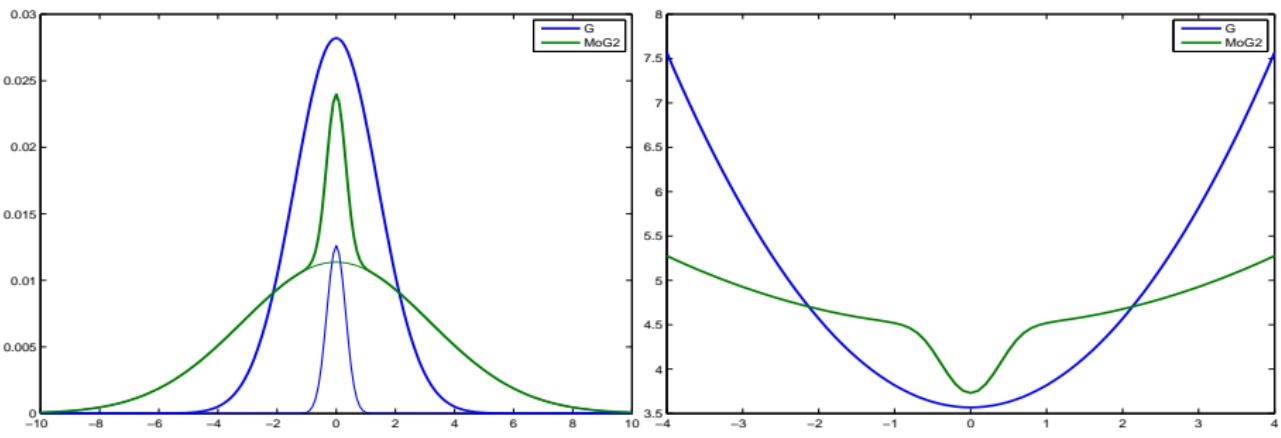
Mixture models

- Mixture of two Gaussians (MoG2) model

$$p(\mathbf{f}|\alpha, \nu_1, \nu_0) = \prod_j [\alpha \mathcal{N}(f_j|0, \nu_1) + (1 - \alpha) \mathcal{N}(f_j|0, \nu_0)]$$

- Bernoulli-Gaussian (BG) model

$$p(\mathbf{f}|\alpha, \nu) = \prod_j p(f_j) = \prod_j [\alpha \mathcal{N}(f_j|0, \nu) + (1 - \alpha) \delta(f_j)]$$



Mixture of 2 Gaussians families

- Mixture of Gammas

$$p(\mathbf{f}|\lambda, \nu_1, \nu_0) = \prod_j [\lambda \mathcal{G}(f_j|\alpha_1, \beta_1) + (1 - \lambda) \mathcal{G}(f_j|\alpha_2, \beta_2)]$$

- Bernoulli-Gamma model

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

- Mixture of Dirichlets model

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

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

where $\mathbf{H} = \{a_1, \dots, a_K\}$ and $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_K\}$

with $\sum_k \alpha_k = \alpha$ and $\sum_k a_k = 1$.

- Bernoulli-Multinomial (BMultinomial) model

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

Hierarchical models and hidden variables

- All the mixture models and some of simple models can be modeled via **hidden variables \mathbf{z}** .

$$p(f) = \sum_{k=1}^K \alpha_k p_k(f) \longrightarrow \begin{cases} p(f|\mathbf{z} = k) = p_k(f), \\ P(\mathbf{z} = k) = \alpha_k, \quad \sum_k \alpha_k = 1 \end{cases}$$

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

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

$$\begin{cases} p(f_j|\mathbf{z}_j = 0, v_0) = \mathcal{N}(f_j|0, v_0), \\ p(f_j|\mathbf{z}_j = 1, v_1) = \mathcal{N}(f_j|0, v_1), \end{cases} \text{ and } \begin{cases} P(\mathbf{z}_j = 0) = \lambda, \\ P(\mathbf{z}_j = 1) = 1 - \lambda \end{cases}$$

$$\begin{cases} p(\mathbf{f}|\mathbf{z}) = \prod_j p(f_j|\mathbf{z}_j) = \prod_j \mathcal{N}(f_j|0, v_{\mathbf{z}_j}) \propto \exp \left\{ -\frac{1}{2} \sum_j \frac{f_j^2}{v_{\mathbf{z}_j}} \right\} \\ P(\mathbf{z}_j = 1) = \lambda, \quad P(\mathbf{z}_j = 0) = 1 - \lambda \end{cases}$$

Hierarchical models and hidden variables

- ▶ Example 2: Student-t model

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

- ▶ Infinite mixture

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

$$\begin{cases} p(\mathbf{f}|\mathbf{z}) &= \prod_j p(f_j|\textcolor{red}{z}_j) = \prod_j \mathcal{N}(f_j|0, 1/\textcolor{red}{z}_j) \propto \exp \left\{ -\frac{1}{2} \sum_j \textcolor{red}{z}_j f_j^2 \right\} \\ p(\mathbf{z}|\alpha, \beta) &= \prod_j \mathcal{G}(\textcolor{red}{z}_j|\alpha, \beta) \propto \prod_j \textcolor{red}{z}_j^{(\alpha-1)} \exp \{-\beta \textcolor{red}{z}_j\} \\ &\propto \exp \left\{ \sum_j (\alpha-1) \ln \textcolor{red}{z}_j - \beta \textcolor{red}{z}_j \right\} \\ p(\mathbf{f}, \mathbf{z}|\alpha, \beta) &\propto \exp \left\{ -\frac{1}{2} \sum_j \textcolor{red}{z}_j f_j^2 + (\alpha-1) \ln \textcolor{red}{z}_j - \beta \textcolor{red}{z}_j \right\} \end{cases}$$

Hierarchical models and hidden variables

- ▶ Example 3: Laplace (Double Exponential) model

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

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

- ▶ With these models we have:

$$p(\mathbf{f}, z, \theta | \mathbf{g}) \propto p(\mathbf{g}|\mathbf{f}, \theta_1) p(\mathbf{f}|z, \theta_2) p(z|\theta_3) p(\theta)$$

Bayesian Computation and Algorithms

- ▶ Often, the expression of $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g})$ 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
 $p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}, \mathbf{g})$, $p(\mathbf{z} | \mathbf{f}, \boldsymbol{\theta}, \mathbf{g})$, and $p(\boldsymbol{\theta} | \mathbf{f}, \mathbf{z}, \mathbf{g})$
- ▶ VBA: Approximate $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g})$ by a separable one

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

and do any computations with these separable ones.

MCMC based algorithm

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

General scheme:

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

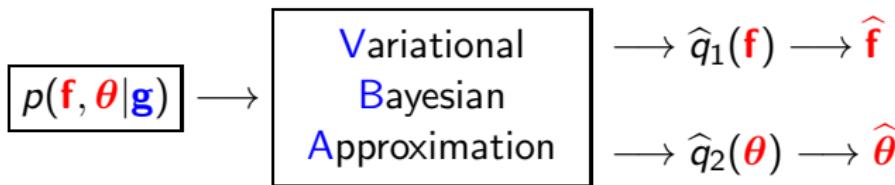
- ▶ Estimate \mathbf{f} using $p(\mathbf{f} | \hat{\mathbf{z}}, \hat{\boldsymbol{\theta}}, \mathbf{g}) \propto p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta}) p(\mathbf{f} | \hat{\mathbf{z}}, \hat{\boldsymbol{\theta}})$
When Gaussian, can be done via optimisation of a quadratic criterion.
- ▶ Estimate \mathbf{z} using $p(\mathbf{z} | \hat{\mathbf{f}}, \hat{\boldsymbol{\theta}}, \mathbf{g}) \propto p(\mathbf{g} | \hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\boldsymbol{\theta}}) p(\mathbf{z})$
Often needs sampling (hidden discrete variable)
- ▶ Estimate $\boldsymbol{\theta}$ using
$$p(\boldsymbol{\theta} | \hat{\mathbf{f}}, \hat{\mathbf{z}}, \mathbf{g}) \propto p(\mathbf{g} | \hat{\mathbf{f}}, \sigma_\epsilon^2 \mathbf{I}) p(\hat{\mathbf{f}} | \hat{\mathbf{z}}, (m_k, v_k)) p(\boldsymbol{\theta})$$

Use of Conjugate priors \longrightarrow analytical expressions.

Variational Bayesian Approximation

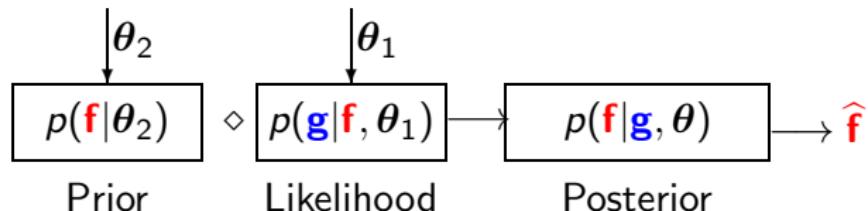
- ▶ Approximate $p(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g})$ by $q(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g}) = q_1(\mathbf{f} | \mathbf{g}) q_2(\boldsymbol{\theta} | \mathbf{g})$ and then continue computations.
- ▶ Criterion $\text{KL}(q : p) : p(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g})$
- ▶ $\text{KL}(q : p) = \int \int q \ln q / p = \int \int q_1 q_2 \ln \frac{q_1 q_2}{p} = \int q_1 \ln q_1 + \int q_2 \ln q_2 - \int \int q \ln p = -H(q_1) - H(q_2) - \langle \ln p \rangle_q$
- ▶ Iterative algorithm $q_1 \rightarrow q_2 \rightarrow q_1 \rightarrow q_2, \dots$

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

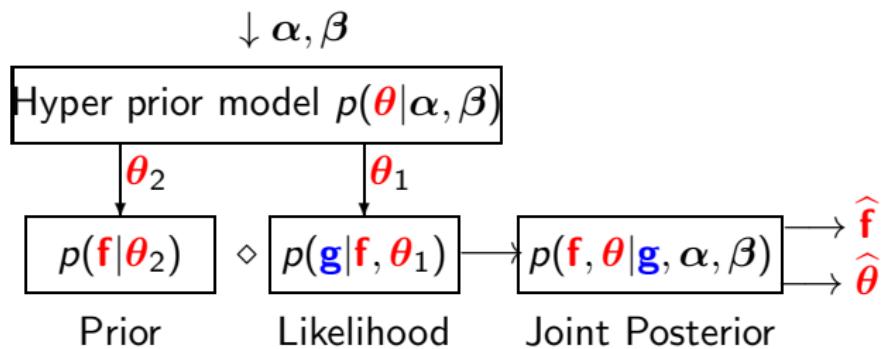


Summary of Bayesian estimation 1

- ▶ Simple Bayesian Model and Estimation

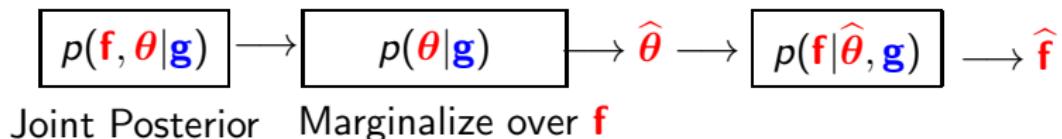


- ▶ Full Bayesian Model and Hyperparameter Estimation

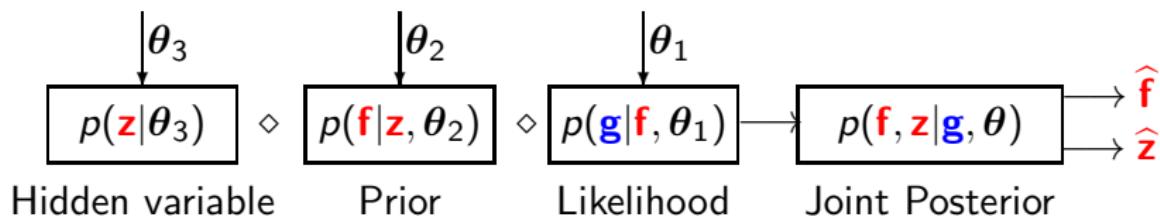


Summary of Bayesian estimation 2

- ▶ Marginalization for Hyperparameter Estimation



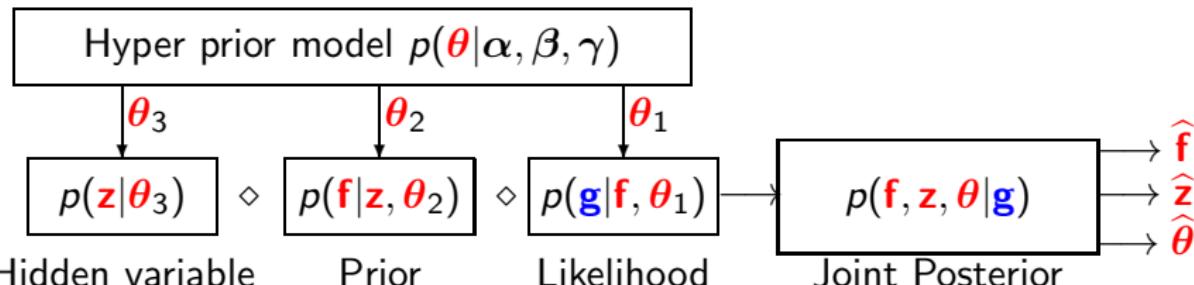
- ▶ Full Bayesian Model with a Hierarchical Prior Model



Summary of Bayesian estimation 3

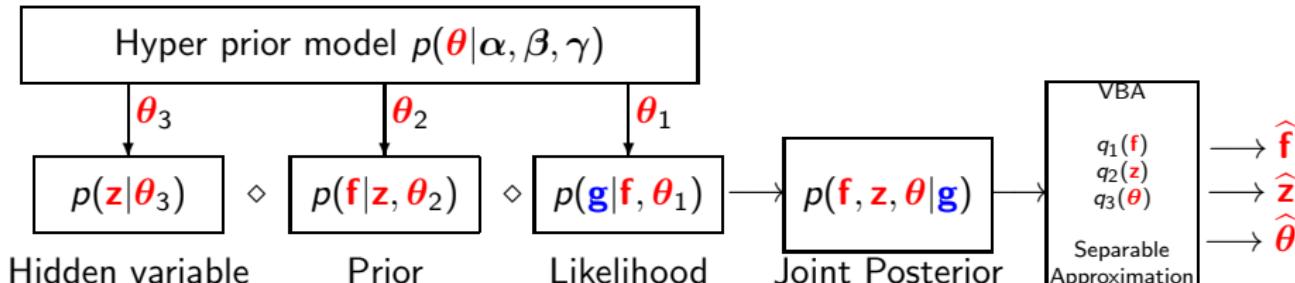
- Full Bayesian Hierarchical Model with Hyperparameter Estimation

$$\downarrow \alpha, \beta, \gamma$$

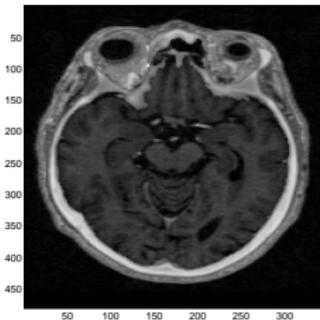


- Full Bayesian Hierarchical Model and Variational Approximation

$$\downarrow \alpha, \beta, \gamma$$



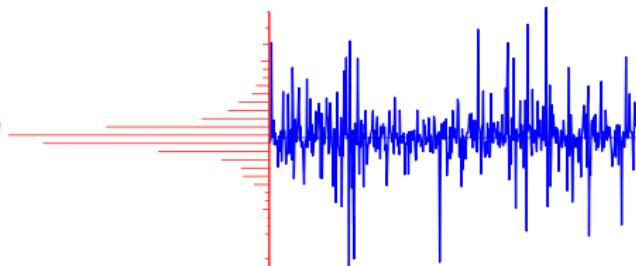
Which images I am looking for?



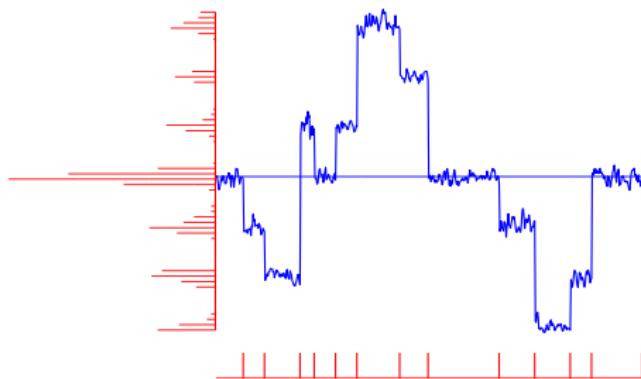
Which image I am looking for?



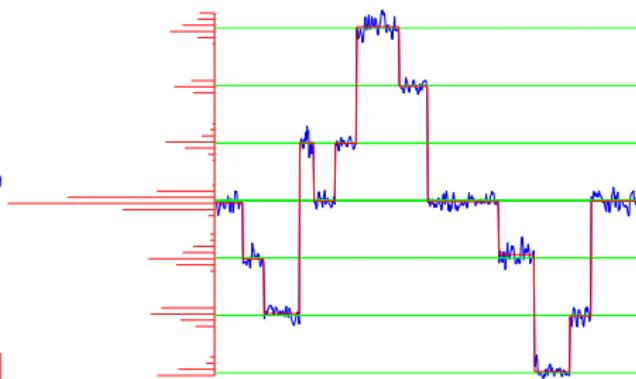
Gauss-Markov



Generalized GM

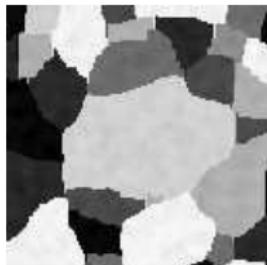
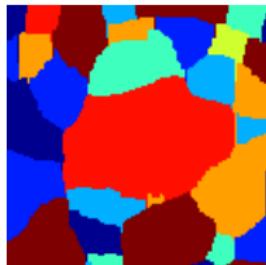


Piecewise Gaussian



Mixture of GM

Gauss-Markov-Potts prior models for images

 $f(\mathbf{r})$  $z(\mathbf{r})$ 

$$c(\mathbf{r}) = 1 - \delta(z(\mathbf{r}) - z(\mathbf{r}'))$$

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

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

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

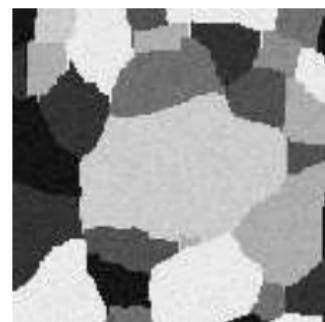
$$p(z(\mathbf{r})|z(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \propto \exp \left\{ \gamma \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right\}$$
$$p(\mathbf{z}) \propto \exp \left\{ \gamma \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})}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

Four different cases

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

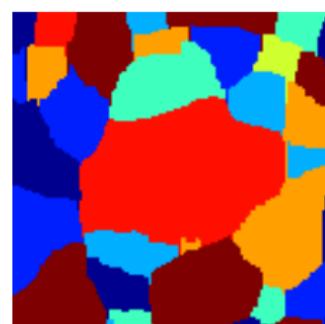
- ▶ $f|z$ Gaussian iid, z iid :

Mixture of Gaussians



- ▶ $f|z$ Gauss-Markov, z iid :

Mixture of Gauss-Markov

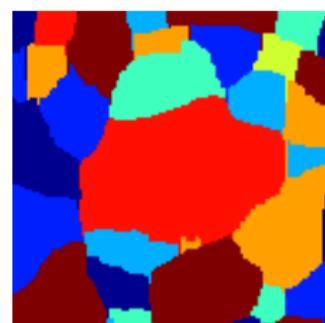


- ▶ $f|z$ Gaussian iid, z Potts-Markov :

Mixture of Independent Gaussians
(MIG with Hidden Potts)

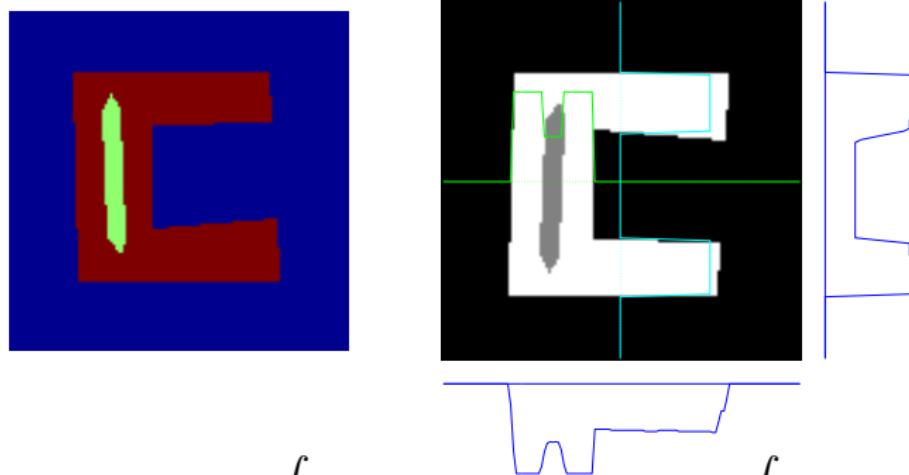
- ▶ $f|z$ Markov, z Potts-Markov :

Mixture of Gauss-Markov
(MGM with hidden Potts)



Application of CT in NDT

Reconstruction from only 2 projections

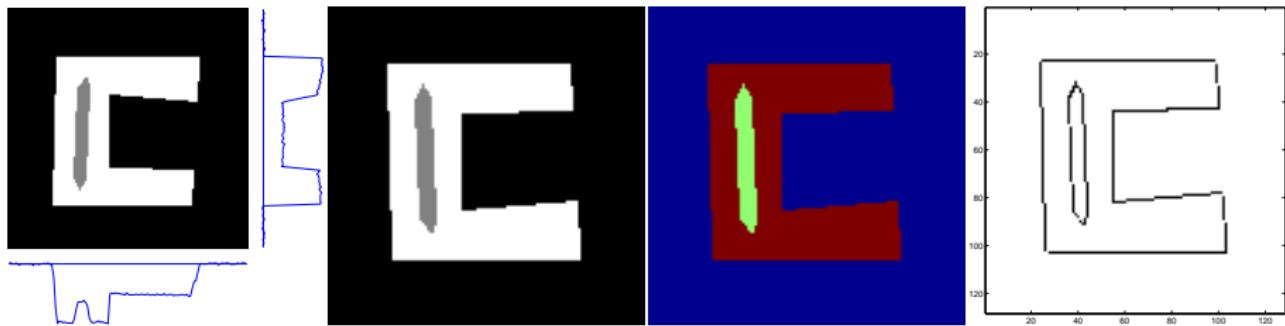


$$g_1(x) = \int f(x, y) dy, \quad g_2(y) = \int f(x, y) dx$$

- 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:

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

Application in CT



$$\begin{aligned} \mathbf{g} | \mathbf{f} & \\ \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon} & \\ \mathbf{g} | \mathbf{f} \sim \mathcal{N}(\mathbf{H}\mathbf{f}, \sigma_\epsilon^2 \mathbf{I}) & \\ \text{Gaussian} & \end{aligned}$$

$\mathbf{f} | \mathbf{z}$
iid Gaussian
or
Gauss-Markov

\mathbf{z}
iid
or
Potts

\mathbf{c}
 $c(\mathbf{r}) \in \{0, 1\}$
 $1 - \delta(z(\mathbf{r}) - z(\mathbf{r}'))$
binary

Proposed algorithm

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

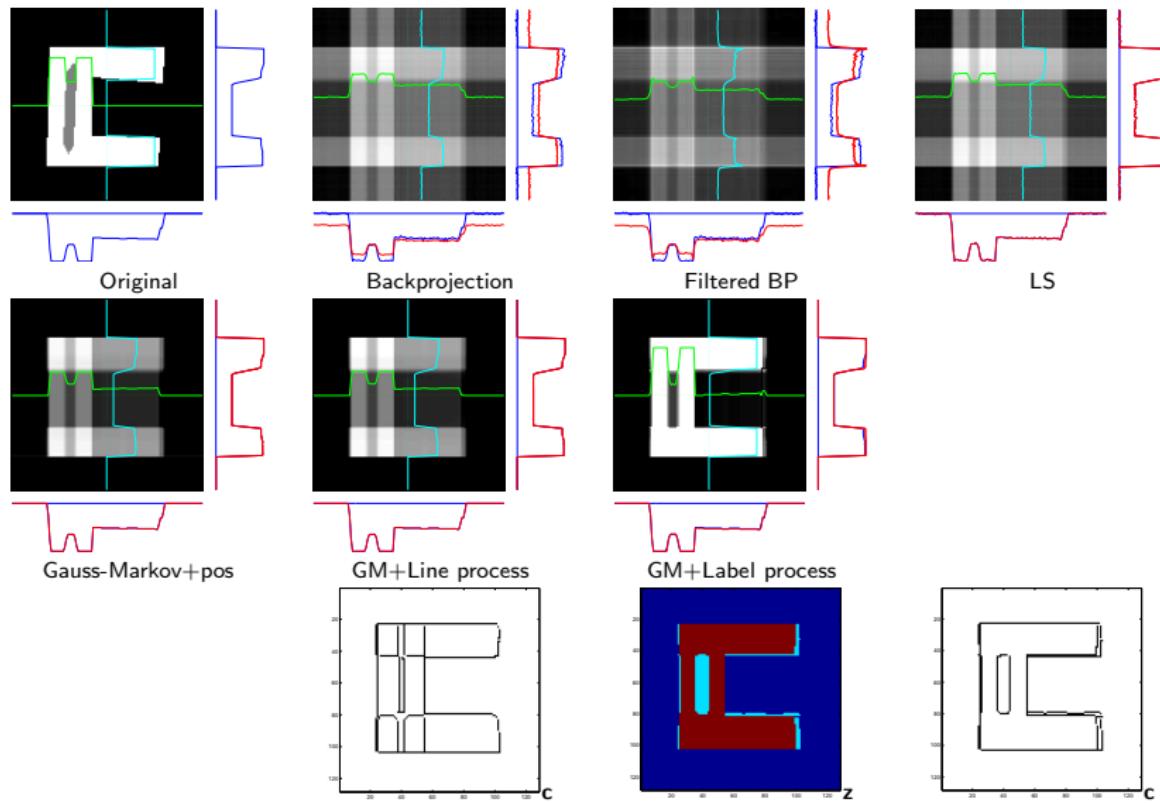
General scheme:

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

Iterative algorithme:

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

Results

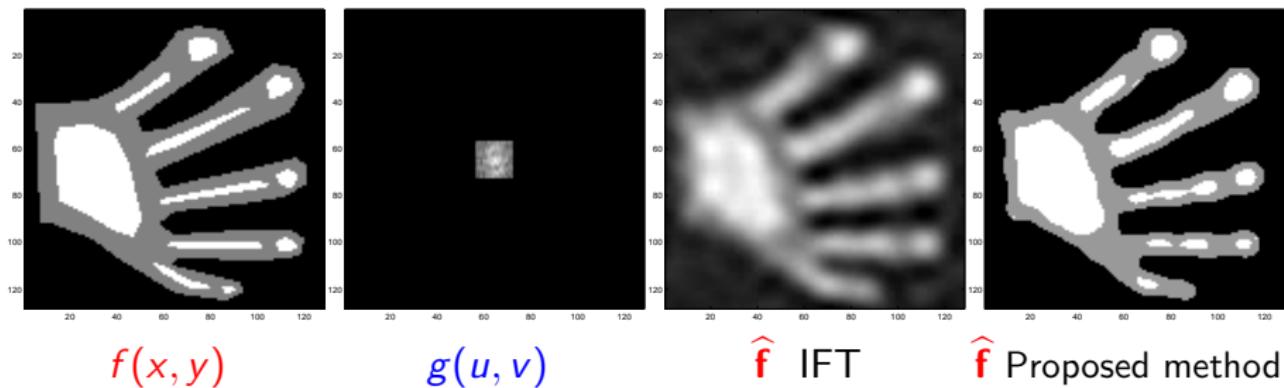


Application in Microwave imaging

$$g(\omega) = \int f(\mathbf{r}) \exp \{-j(\omega \cdot \mathbf{r})\} \, d\mathbf{r} + \epsilon(\omega)$$

$$g(u, v) = \iint f(x, y) \exp \{-j(ux + vy)\} \, dx \, dy + \epsilon(u, v)$$

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



Conclusions

- ▶ Bayesian Inference for inverse problems
- ▶ Different prior modeling for signals and images:
Separable, Markovian, without and with hidden variables
- ▶ Sparsity 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)