Inverse Problems: from Regularization to Bayesian inference solutions

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

Seminar given at
School of Mathematics and Computational Science
Sun Yat-sen University, Guangzhou, China, December 5, 2013

Content

1. Direct and indirect observation
2. Inverse Problems examples
3. Analytical methods
4. Discretization and algebraic deterministic methods
5. Regularization
6. Limitations of the deterministic methods
7. Probabilistic methods
8. Bayesian approach for inverse problems
9. Prior models
   - Gaussian, Generalized Gaussian (GG), Gamma, Beta,
   - Gauss-Markov, GG-Marvov
   - Sparsity enforcing priors (GG, Laplace, Student-t, Cauchy)
10. Full Bayesian approach (Estimation of hyperparameters)
11. Conclusions

1. 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 temeprature 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.

2. Inverse problems 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

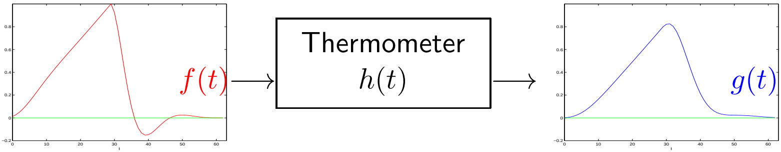
- Inverse problem: Deconvolution

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

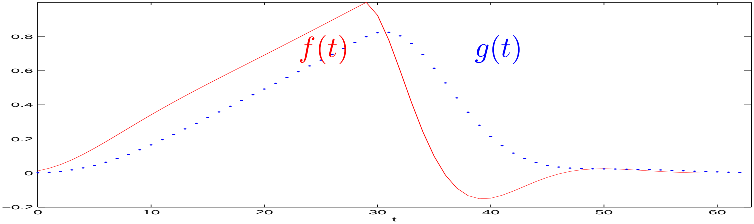
Measuring variation of temperature with a therometer

Forward model: Convolution

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

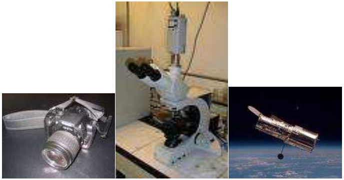


Inversion: Deconvolution



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

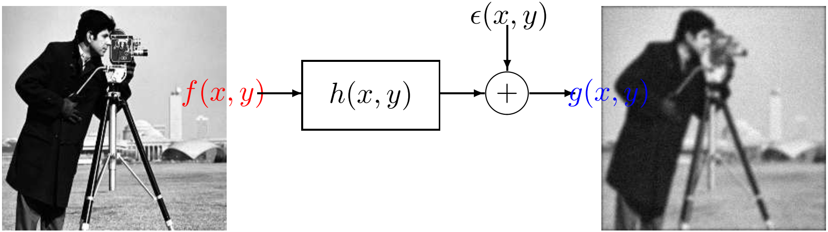
- Inverse problem: Image restoration

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 $f(x,y)$

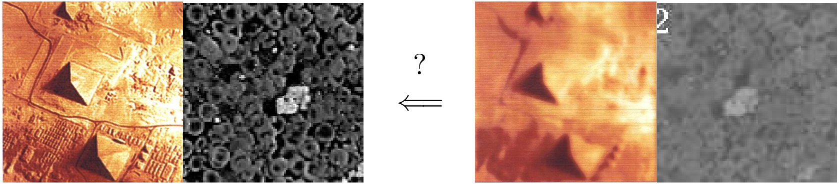
Making an image with an unfocused camera

Forward model: 2D Convolution

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

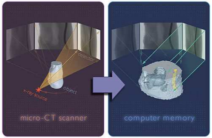


Inversion: Image Deconvolution or Restoration



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 $g_\phi(r,z)$



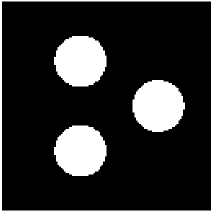
- Forward model:
  Line integrals or Radon Transform

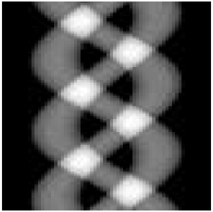
$$
\begin{aligned}
g_\phi(r) &= \int_{L_{r,\phi}} f(x,y)\, \mathrm{d}l + \epsilon_\phi(r) \\
&= \iint f(x,y)\, \delta(r - x\cos\phi - y\sin\phi)\, \mathrm{d}x\, \mathrm{d}y + \epsilon_\phi(r)
\end{aligned}
$$

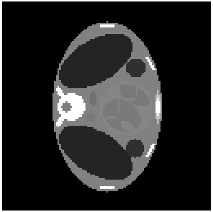
- Inverse problem: Image reconstruction

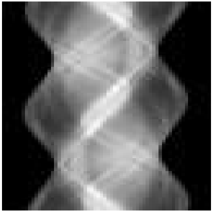
  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x,y)$

Computed Tomography: Radon Transform





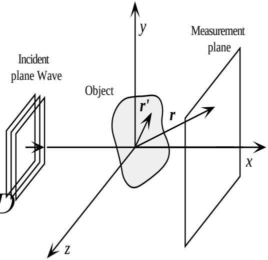




Forward:	$f(x,y)$	$\longrightarrow$	$g(r,\phi)$
Inverse:	$f(x,y)$	$\longleftarrow$	$g(r,\phi)$

Microwave or ultrasound imaging

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

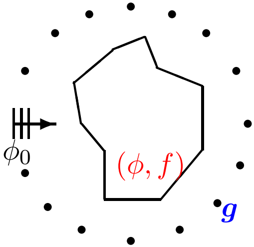


$$\phi_d(\boldsymbol{r}_i) = \iint_D G_m(\boldsymbol{r}_i, \boldsymbol{r}')\phi(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}', \ \boldsymbol{r}_i \in S$$

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

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

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



Discretization :

$$\begin{cases} \boldsymbol{\phi}_d = \boldsymbol{G}_m \boldsymbol{F} \boldsymbol{\phi} \\ \boldsymbol{\phi} = \phi_0 + \boldsymbol{G}_o \boldsymbol{F} \boldsymbol{\phi} \end{cases} \longrightarrow \begin{cases} \boldsymbol{\phi}_d = \boldsymbol{H}(\boldsymbol{f}) \\ \text{with} \quad \boldsymbol{F} = \mathrm{diag}(\boldsymbol{f}) \\ \boldsymbol{H}(\boldsymbol{f}) = \boldsymbol{G}_m \boldsymbol{F}(\boldsymbol{I} - \boldsymbol{G}_o \boldsymbol{F})^{-1} \phi_0 \end{cases}$$

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

3. Inverse problems: Analytical methods

Two communities working on Inverse problems:

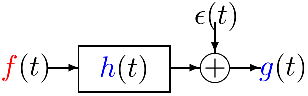
- Mathematical departments:
  Analytical methods: Existance and Uniqueness
  Differential equations, PDE
- Engineering and Computer sciences:
  Algebraic methods: Discretization, Uniqueness and Stability
  Integral equations, Discretization using Moments method, Galerkin, ...

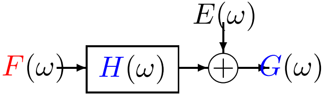
Two examples:

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

Deconvolution: Analytical methods

Time domain	Fourier domain
Forward model:	
$g(t) = h(t) * f(t) + \epsilon(t)$	$G(\omega) = H(\omega)\, F(\omega) + E(\omega)$





Deconvolution:

$$g(t) \rightarrow \boxed{w(t) = IFT\{\frac{1}{H(\omega)}\}} \rightarrow \widehat{f}(t)$$

Inverse filtering

$$G(\omega) \rightarrow \boxed{\frac{1}{H(\omega)}} \rightarrow \widehat{F}(\omega)$$

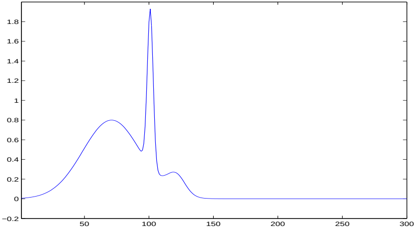
Deconvolution:

$$g(t) \rightarrow \boxed{W(\omega)} \rightarrow \widehat{f}(t)$$

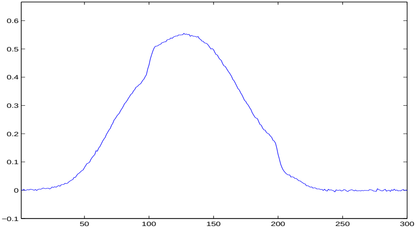
Wiener filtering

$$G(\omega) \rightarrow \boxed{\frac{H^*(\omega)}{|H(\omega)|^2 + \frac{S_\epsilon(\omega)}{S_f(\omega)}}} \rightarrow \widehat{F}(\omega)$$

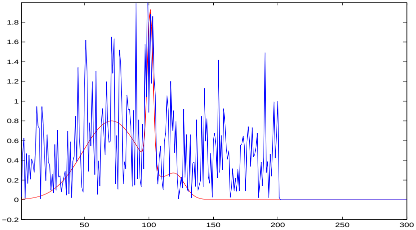
Deconvolution example



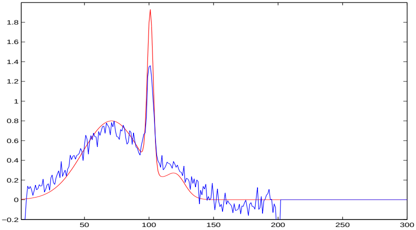
$f(t)$



$g(t)$

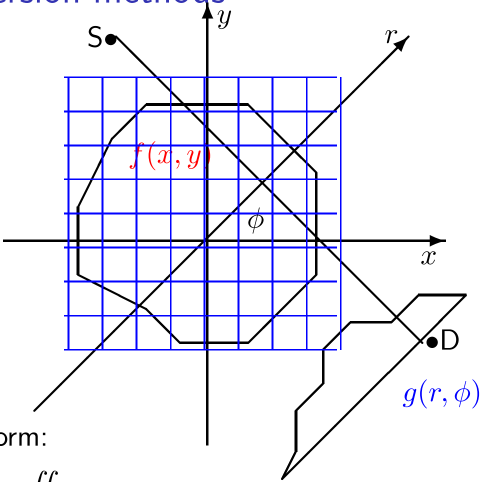


Inverse Filtering



Wiener Filtering

Analytical Inversion methods



Radon Transform:

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

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

Filtered Backprojection method

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

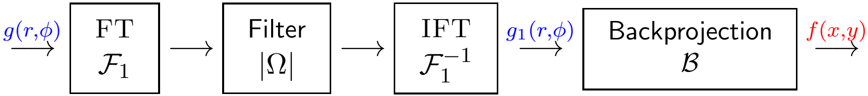
Derivation $\mathcal{D}$: $\quad \overline{g}(r,\phi) = \frac{\partial g(r,\phi)}{\partial r}$

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

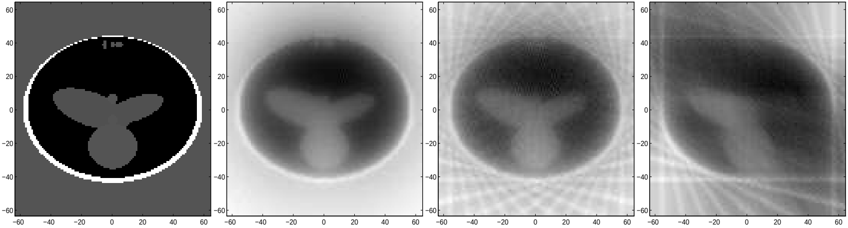
Backprojection $\mathcal{B}$: $\quad f(x,y) = \frac{1}{2\pi} \int_0^{\pi} g_1(r' = x\cos\phi + y\sin\phi, \phi) \, \mathrm{d}\phi$

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

- Backprojection of filtered projections:



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, ...

4. Inverse problems: Discretization and Algebraic 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

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

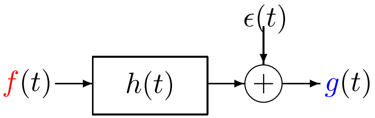
with $\{b_j(\boldsymbol{r})\}$ a basis or any other set of known functions

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

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

- $\boldsymbol{H}$ 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.

Convolution: Discretization



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

- The signals $f(t)$, $g(t)$, $h(t)$ are discretized with the same sampling period $\Delta T = 1$,
- The impulse response is finite (FIR) : $h(t) = 0$, for t such that $t < -q\Delta T$ or $\forall t > p\Delta T$.

$$g(m) = \sum_{k=-q}^{p} h(k)\, f(m-k) + \epsilon(m), \quad m = 0, \cdots, M$$

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

Convolution: Discretized matrix vector forms

$$
\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{bmatrix}
=
\begin{bmatrix}
h(p) & \cdots & h(0) & \cdots & h(-q) & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \ddots & & \ddots & & \ddots & & & & \vdots \\
\vdots & & & & & & & & & \vdots \\
\vdots & & h(p) & \cdots & h(0) & \cdots & h(-q) & & & \vdots \\
\vdots & & & & & & & & & \vdots \\
\vdots & & & \ddots & & & \ddots & & & \vdots \\
\vdots & & & & \ddots & & & & & 0 \\
0 & \cdots & \cdots & & 0 & h(p) & \cdots & h(0) & \cdots & h(-q)
\end{bmatrix}
\begin{bmatrix} f(-p) \\ \vdots \\ f(0) \\ f(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ f(M) \\ f(M+1) \\ \vdots \\ f(M+q) \end{bmatrix}
$$

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

- $\boldsymbol{g}$ is a $(M+1)$-dimensional vector,
- $\boldsymbol{f}$ has dimension $M+p+q+1$,
- $\boldsymbol{h} = [h(p), \cdots, h(0), \cdots, h(-q)]$ has dimension $(p+q+1)$
- $\boldsymbol{H}$ has dimensions $(M+1) \times (M+p+q+1)$.

Convolution: Discretized matrix vector form

- If system is causal ($q=0$) we obtain

$$\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{bmatrix} = \begin{bmatrix} h(p) & \cdots & h(0) & 0 & \cdots & \cdots & 0 \\ 0 & & & & & & \vdots \\ \vdots & & & & & & \vdots \\ \vdots & & h(p) & \cdots & h(0) & & \vdots \\ \vdots & & & & & & \vdots \\ \vdots & & & & & & 0 \\ 0 & \cdots & \cdots & 0 & h(p) & \cdots & h(0) \end{bmatrix} \begin{bmatrix} f(-p) \\ \vdots \\ f(0) \\ f(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ f(M) \end{bmatrix}$$

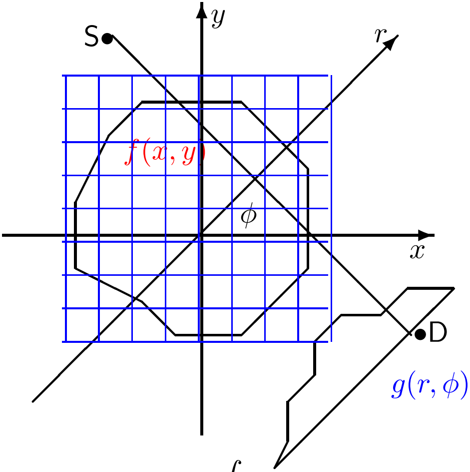
- $\boldsymbol{g}$ is a $(M+1)$-dimensional vector,
- $\boldsymbol{f}$ has dimension $M+p+1$,
- $\boldsymbol{h}=[h(p),\cdots,h(0)]$ has dimension $(p+1)$
- $\boldsymbol{H}$ has dimensions $(M+1)\times(M+p+1)$.

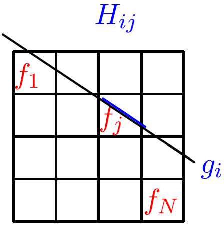
Convolution: Causal system and causal input

$$\begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{bmatrix} = \begin{bmatrix} h(0) & & & & & \\ h(1) & \ddots & & & & \\ \vdots & & & & & \\ h(p) & \cdots & & h(0) & & \\ 0 & \ddots & & & \ddots & \\ \vdots & & & & & \\ 0 & \cdots & 0 & h(p) & \cdots & h(0) \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ f(M) \end{bmatrix}$$

- $\boldsymbol{g}$ is a $(M+1)$-dimensional vector,
- $\boldsymbol{f}$ has dimension $M+1$,
- $\boldsymbol{h} = [h(p), \cdots, h(0)]$ has dimension $(p+1)$
- $\boldsymbol{H}$ has dimensions $(M+1) \times (M+1)$.

Discretization of Radon Transfrom in CT





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

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

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

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

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

Deterministic methods

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

- Backprojection: $\widehat{\boldsymbol{f}} = \boldsymbol{H}'\boldsymbol{g}$, but $\boldsymbol{H}'\boldsymbol{H} \neq \boldsymbol{I}$
- Minimum norm solution of $\boldsymbol{H}\boldsymbol{f} = \boldsymbol{g}$:

  $$\text{minimize } \|\boldsymbol{f}\|^2 \text{ s.t. } \boldsymbol{H}\boldsymbol{f} = \boldsymbol{g} \longrightarrow \widehat{\boldsymbol{f}} = \boldsymbol{H}'(\boldsymbol{H}\boldsymbol{H}')^{-1}\boldsymbol{g}$$

  (Filtered Backprojection)
- Least Squares:

  $$\text{minimize } \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 \longrightarrow \widehat{\boldsymbol{f}} = (\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{g}$$

  (Backprojection followed by image deconvolution)
- Constraint optimization:

  $$\text{minimize } \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2 \text{ s.t. } \|\boldsymbol{f}\|^2 < c \longrightarrow \widehat{\boldsymbol{f}} = (\boldsymbol{H}'\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}'\boldsymbol{g}$$

  $$\text{minimize } \|\boldsymbol{f}\|^2 \text{ s.t. } \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2 < c \longrightarrow \widehat{\boldsymbol{f}} = (\mu\boldsymbol{H}'\boldsymbol{H}+\boldsymbol{I})^{-1}\boldsymbol{H}'\boldsymbol{g}$$
- Minimum norme LS (MNLS):

  $$J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda||\boldsymbol{f}||^2$$

Deterministic methods: General Data matching methods

- Observation model: $\boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{\epsilon}$
  $g_i = h_i(\boldsymbol{f}) + \epsilon_i, \quad i = 1, \ldots, M$
- 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.

5. Inverse problems: Regularization theory

Inverse problems = Ill posed problems

$\longrightarrow$ Need for prior information

- Functional space (Tikhonov):

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

$$\widehat{f}(t) = (\mathcal{H}^*\mathcal{H} + \lambda\mathcal{D}^*\mathcal{D})^{-1}\mathcal{H}g(t)$$

- Finite dimensional space (Philips & Towmey):

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

$$\widehat{\boldsymbol{f}} = (\boldsymbol{H}^*\boldsymbol{H} + \lambda\boldsymbol{D}^*\boldsymbol{D})^{-1}\boldsymbol{H}\boldsymbol{g}$$

- Minimum norme LS (MNLS):

$$J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda||\boldsymbol{f}||^2 \longrightarrow \widehat{\boldsymbol{f}} = (\boldsymbol{H}^*\boldsymbol{H} + \lambda\boldsymbol{I})^{-1}\boldsymbol{H}\boldsymbol{g}$$

- More general regularization:

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

or

$$J(\boldsymbol{f}) = \Delta_1(\boldsymbol{g}, \boldsymbol{H}(\boldsymbol{f})) + \lambda\Delta_2(\boldsymbol{f}, \boldsymbol{f}_0)$$

Algebraic Regularization methods

$$J(\boldsymbol{f}) = ||\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})||^2 + \lambda\Omega(\boldsymbol{f})$$

- Classical quadratic regularization:

$$\Omega(\boldsymbol{f}) = ||\boldsymbol{D}\boldsymbol{f}||^2 = \sum_j |f_j - f_{j-1}|^2$$

- Sparsity enforcing regularization:

$$\Omega(\boldsymbol{f}) = ||\boldsymbol{f}||_1 = \sum_j |f_j|$$

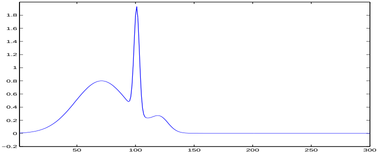
- Total Variation regularization:

$$\Omega(\boldsymbol{f}) = ||\boldsymbol{D}\boldsymbol{f}||_1 = \sum_j |f_j - f_{j-1}|$$

- Combined regularization:

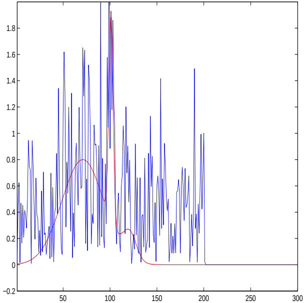
$$\Omega(\boldsymbol{f}) = \sum_j \lambda_1 |f_j|^2 + \lambda_1 |f_j - f_{j-1}|$$

Deconvolution example

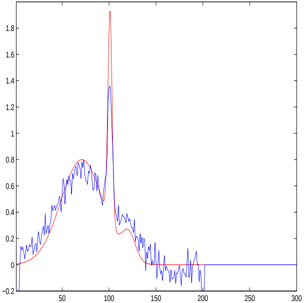


$f(t)$

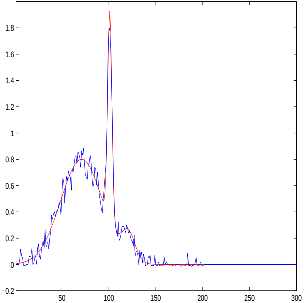
$g(t) = h(t) * f(t) + \epsilon(t)$



Inverse Filtering



Wiener



Regularization

6. Limitations of deterministic regularization methods

General regularization methods:

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

or

$$J(\boldsymbol{f}) = \Delta_1(\boldsymbol{g}, \boldsymbol{H}\boldsymbol{f}) + \lambda\Delta_2(\boldsymbol{D}\boldsymbol{f}, \boldsymbol{D}\boldsymbol{f}_0)$$

Limitations:

- Errors are considered implicitly white and Gaussian
- Limited prior information on the solution
- Lack of tools for the determination of the regularization parameters, even if there are a few methods: L curve, Cross Validation
- Lack of tools for handeling uncertainties

$\Downarrow$

Probabilistic Methods and Bayesian inference

7. Probabilistic methods for inversion

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

8. 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 : $$p(\boldsymbol{f}|\boldsymbol{g};\mathcal{M}) = \frac{p(\boldsymbol{g}|\boldsymbol{f};\mathcal{M})\, p(\boldsymbol{f}|\mathcal{M})}{p(\boldsymbol{g}|\mathcal{M})}$$

Link with regularization :

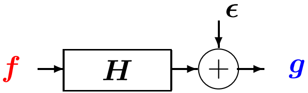
Maximum A Posteriori (MAP) :

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

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

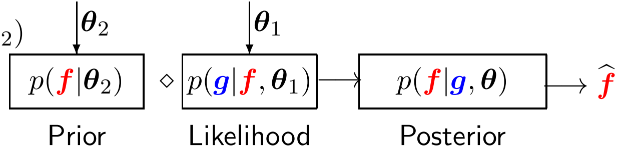
- Linear Inverse problems: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$



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



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

Case of linear models and Gaussian priors

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

- Hypothesis on the noise: $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{I})$

$$p(\boldsymbol{g}|\boldsymbol{f}) \propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right]$$

- Hypothesis on $\boldsymbol{f}$: $\boldsymbol{f} \sim \mathcal{N}(0, \sigma_f^2 \boldsymbol{I})$

$$p(\boldsymbol{f}) \propto \exp\left[-\frac{1}{2\sigma_f^2}\|\boldsymbol{f}\|^2\right]$$

- A posteriori:

$$p(\boldsymbol{f}|\boldsymbol{g}) \propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}\left[\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \frac{\sigma_\epsilon^2}{\sigma_f^2}\|\boldsymbol{f}\|^2\right]\right]$$

- MAP : $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$

  with $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2, \qquad \lambda = \frac{\sigma_\epsilon^2}{\sigma_f^2}$

- Advantage : characterization of the solution

$\boldsymbol{f}|\boldsymbol{g} \sim \mathcal{N}(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{P}})$ with $\widehat{\boldsymbol{f}} = \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g}$, $\widehat{\boldsymbol{P}} = \sigma_\epsilon^2\left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{I}\right)^{-1}$

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

Markovian models:

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

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
  - ...

MAP estimation and Compressed Sensing

$$\begin{cases} \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \\ \boldsymbol{f} = \boldsymbol{W}\boldsymbol{z} \end{cases}$$

- $\boldsymbol{W}$ a code book matrix, $\boldsymbol{z}$ coefficients
- Gaussian:

$$p(\boldsymbol{z}) = \mathcal{N}(0, \sigma_z^2 \boldsymbol{I}) \propto \exp\left[-\frac{1}{2\sigma_z^2}\sum_j |z_j|^2\right]$$
$$J(\boldsymbol{z}) = -\ln p(\boldsymbol{z}|\boldsymbol{g}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{W}\boldsymbol{z}\|^2 + \lambda\sum_j |z_j|^2$$

- Generalized Gaussian (sparsity, $\beta = 1$):

$$p(\boldsymbol{z}) \propto \exp\left[-\lambda\sum_j |z_j|^\beta\right]$$
$$J(\boldsymbol{z}) = -\ln p(\boldsymbol{z}|\boldsymbol{g}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{W}\boldsymbol{z}\|^2 + \lambda\sum_j |z_j|^\beta$$

- $\boldsymbol{z} = \arg\min_{\boldsymbol{z}} \{J(\boldsymbol{z})\} \longrightarrow \widehat{\boldsymbol{f}} = \boldsymbol{W}\widehat{\boldsymbol{z}}$

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

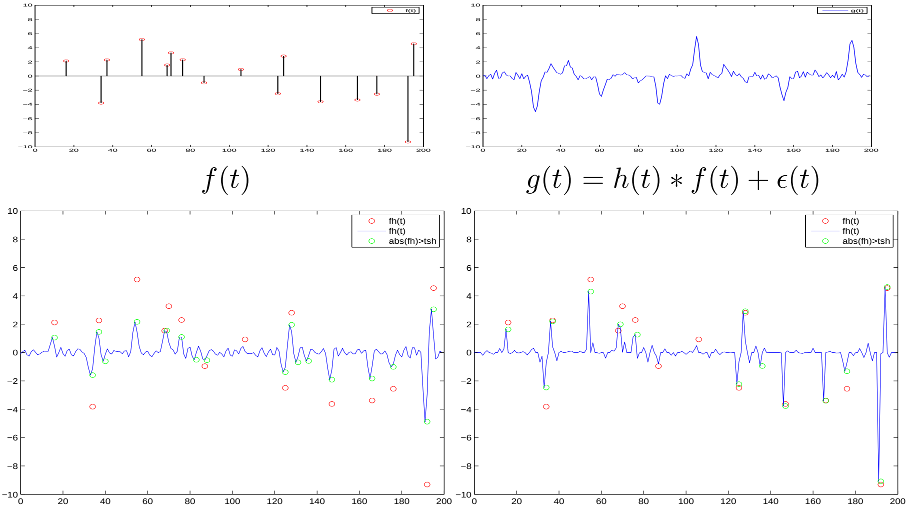
$$\left\{\begin{array}{l} \widehat{\boldsymbol{f}}=(\boldsymbol{H}'\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{H}'\boldsymbol{g} \\ \widehat{\boldsymbol{P}}=\theta_1(\boldsymbol{H}'\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}, \quad \lambda=\frac{\theta_1}{\theta_2} \end{array}\right.$$

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

- Example 2: Double Exponential prior & MAP:

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

Deconvolution example



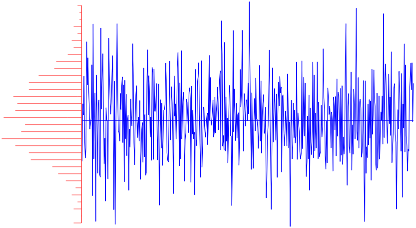
$f(t)$ $\quad$ $g(t) = h(t) * f(t) + \epsilon(t)$

Quadratic Reg. (Gaussian) $\quad$ L_1 Reg. (Laplace)

Two main steps in the Bayesian approach

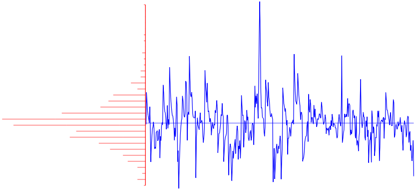
- Prior modeling
  - Separable:
    Gaussian, Gamma,
    Sparsity enforcing: Generalized Gaussian, mixture of Gaussians, mixture of Gammas, ...
  - Markovian:
    Gauss-Markov, GGM, ...
  - Markovian with hidden variables
    (contours, region labels)
- Choice of the estimator and computational aspects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

9. Prior modeling of signals



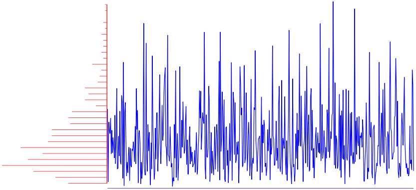
Gaussian

$$p(f_j) \propto \exp\left[-\alpha |f_j|^2\right]$$



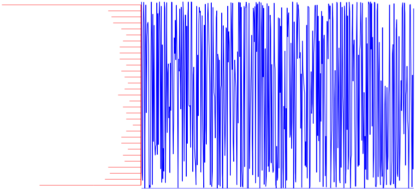
Generalized Gaussian

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



Gamma

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

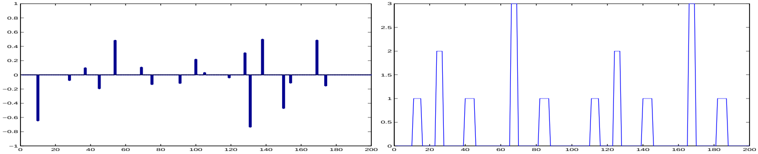


Beta

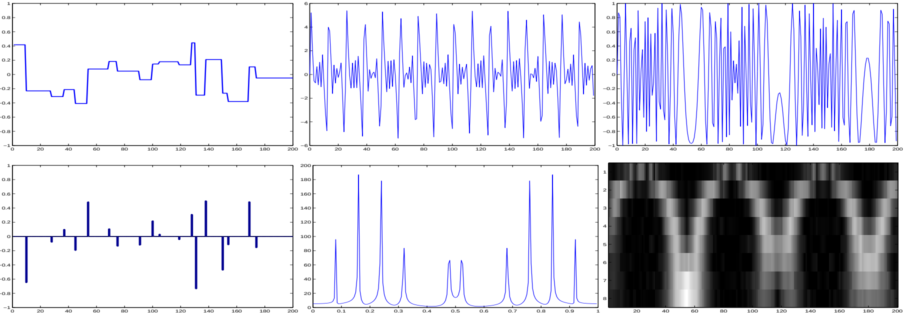
$$p(f_j) \propto f_j^{\alpha} (1 - f_j)^{\beta}$$

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
  - Student-t, Cauchy

  - Elastic net
  - Symmetric Weibull, Symmetric Rayleigh
  - Generalized hyperbolic

- 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.

- 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 $\nu = 1$.

Mixture models

- Mixture of two Gaussians (MoG2) model

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

- Bernoulli-Gaussian (BG) model

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

- 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

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

10. Full Bayesian: Estimation of Hyper Parameters

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

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

- Examples:
  Gaussian noise, Gaussian prior and MAP:

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

  Gaussian noise, Double Exponential prior and 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: Joint Posterior:

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

- Joint MAP:

$$(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f}, \boldsymbol{\theta})} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\}$$

Main conclusions

- Determinsitic Regularization methods have been very successful.
- Bayesian approach is an appropriate tool with much greater possibilities
- More details tomorrow with advanced Bayesian approach