



Sensors, Measurement systems Signal processing and Inverse problems

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>

Files: http://djafari.free.fr/Cours/Master_MNE/Cours/Cours_MNE_2014_01.pdf

Contents

1. Sensors and Measurement systems
 - ▶ Direct and indirect measurement sensors
 - ▶ Primary sensor characteristics
2. Basic sensors design and their mathematical models
 - ▶ R, L, C models and equations
 - ▶ Forward model and simulation
3. Basic signal and image processing of the sensors output
 - ▶ Averaging, Convolution
 - ▶ Fourier Transform, Filtering
4. Indirect measurement and inverse problems
 - ▶ Deconvolution
 - ▶ X ray Computed Tomography
 - ▶ Ultrasound, Microwave and Eddy current NDT

Contents

- 5. Inversion methods
 - ▶ Analytical methods
 - ▶ Algebraic methods
 - ▶ Regularization
- 6. Bayesian estimation approach
 - ▶ Basics of Bayesian estimation
 - ▶ Bayesian inversion
- . Multivariate data analysis
 - ▶ Principal Component Analysis (PCA)
 - ▶ Independent Component Analysis (ICA)
- 8. Blind sources separation
 - ▶ Classical methods
 - ▶ Bayesian approach

1. Sensors and measurement systems

Direct and indirect measurement

- ▶ Direct measurement: Length, Time, Frequency
- ▶ Indirect measurement: All the other quantities
 - ▶ Temperature
 - ▶ Sound
 - ▶ Vibration
 - ▶ Position and Displacement
 - ▶ Pressure
 - ▶ Force
 - ▶ ...
 - ▶ Resistivity, Permeability, Permittivity,
 - ▶ Magnetic inductance
 - ▶ Surface, Volume, Speed, Acceleration
 - ▶ ...

Sensors and measurement systems

Different sensors:

- ▶ Fluid Property Sensors
- ▶ Force Sensors
- ▶ Humidity Sensors
- ▶ Mass Air Flow Sensors
- ▶ Photo Optic Sensors
- ▶ Piezo Film Sensors
- ▶ Position Sensors
- ▶ Pressure Sensors
- ▶ Scanners and Systems
- ▶ Temperature Sensors
- ▶ Torque Sensors
- ▶ Traffic Sensors
- ▶ Vibration Sensors
- ▶ Water Resources Monitoring

Sensors and measurement systems

Main Glossary and Nomenclature:

- ▶ Sensor:

Primary sensing element

(example: thermistor which translates changes in temperature to changes to resistance)

- ▶ Transducer:

Changes one instrument signal value to another instrument signal value

(example: resistance to volts through an electrical circuit)

- ▶ Transmitter:

Contains the transducer and produces an amplified, standardized instrument signal

(example: A/D conversion and transmission)

Primary sensor characteristics

- ▶ Range:

The extreme (min and max) values over which the sensors can make correct measurement over controlled variable.

- ▶ Response time:

The amount of time required for a sensor to completely respond to a change in its input.

- ▶ Accuracy:

Closeness of the sensor output to indicating the actual value of the measured variable.

- ▶ Precision:

The consistency of the sensor output in measuring the same value under the same operating conditions over a period of time.

Primary sensor characteristics

- ▶ Sensitivity:
The minimum small change in the controlled variable that the sensor can measure.
- ▶ Dead band:
The minimum amount of a change to the process which is required before the sensor responds to the change.
- ▶ Costs:
Not simply the purchase cost, but also the installed/operating costs?
- ▶ Installation problems:
Special installation problems, e.g., corrosive fluids, explosive mixtures, size and shape constraints, remote transmission questions, etc.

Signal transmission

- ▶ Pneumatic:
Pneumatic signals are normally 3-15 pounds per square inch (psi).
- ▶ Electronic:
Electronic signals are normally 4-20 milliamp (mA).
- ▶ Optic:
Optical signals are also used with fiber optic systems or when a direct line of sight exists.
- ▶ Hydraulic
- ▶ Radio
- ▶
- ▶ Glossary:
<http://lorien.ncl.ac.uk/ming/procmeas/glossary.htm>
<http://www.sensorland.com/GlossaryPage001.html>
<http://www.sensorland.com/>

2. Basic sensors designs and their mathematical models

- ▶ We can easily measure electrical quantities:
 - ▶ Resistance: $U = RI$ or $u(t) = Ri(t)$
 - ▶ Capacitance: $\frac{\partial u(t)}{\partial t} = \frac{1}{C}i(t)$ or $i(t) = C\frac{\partial u(t)}{\partial t}$
 - ▶ Inductance: $u(t) = L\frac{\partial i(t)}{\partial t}$
- ▶ Sensors and transducers are used to convert many physical quantities to changes in R , C or L .
- ▶ Resistance:
 - ▶ Resistive Temperature Detectors (Thermistors)
 - ▶ Strain Gauges (Pressure to resistance)
- ▶ Capacitance: Capacitive Pressure Sensor
- ▶ Inductance: Inductive Displacement Sensor
- ▶ Thermoelectric Effects: Temperature Measurement
- ▶ Hall Effect: Electric Power Meter
- ▶ Photoelectric Effect: Optical Flux-meter

Resistivity/Conductivity

- ▶ Resistance R : $R = \rho l/s$ (Ohm)
 - ▶ ρ : Resistivity *ohm/meter*
 - ▶ $1/\rho$: conductivity *Siemens/meter*
 - ▶ l : length *meter*
 - ▶ s : section surface *meter*²

- ▶ Dipole model:

$$u(t) = R i(t)$$

- ▶ Impedance

$$U(\omega) = R I(\omega) \longrightarrow Z(\omega) = \frac{U(\omega)}{I(\omega)} = R$$

- ▶ Power dissipation

$$P(t) = R i^2(t) = u^2(t)/R$$

Capacity C

- ▶ Capacitance: $C = \frac{Q}{U} = \varepsilon_0 \frac{\Phi}{U}$ (Farads)
 - ▶ Q Electric charge (coulombs)
 - ▶ U Potential (volts)
 - ▶ ε_0 Electrical permittivity
 - ▶ Φ Electric charge flux (weber)
- ▶ Dipole model:

$$u(t) = \frac{1}{C} \int_0^t i(t') dt'$$

$$\frac{\partial u(t)}{\partial t} = \frac{1}{C} i(t) \quad \text{or} \quad i(t) = C \frac{\partial u(t)}{\partial t}$$

$$I(\omega) = j\omega C U(\omega)$$

- ▶ Impedance

$$Z(\omega) = \frac{1}{j\omega C}$$

Inductance L

- ▶ Inductance: $L = \frac{\Phi}{I}$ (Henri)
 - ▶ Φ Magnetic flux (Weber)
 - ▶ I Current (Amp)
- ▶ Dipole model (Faraday) :

$$u(t) = L \frac{\partial i(t)}{\partial t}$$

$$U(\omega) = j\omega L I(\omega)$$

- ▶ Impedance

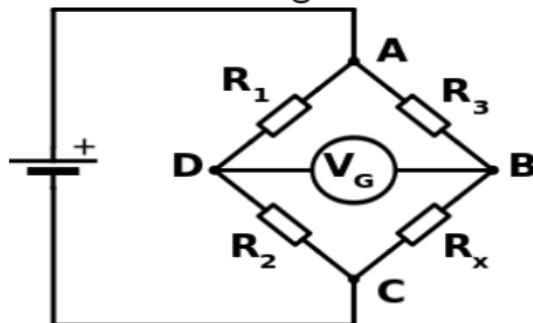
$$U(\omega) = j\omega L I(\omega) \longrightarrow Z(\omega) = j\omega L$$

Measuring R, C and L

- ▶ Measuring R:
 - ▶ Simple voltage divider
- ▶ Bridge measurement systems
 - ▶ Single-Point Bridge
 - ▶ Two-Point Bridge (Wheatstone Bridge)
 - ▶ Four-Point Bridge
- ▶ Measuring C and L
 - ▶ AC voltage dividers and Bridges
(Maxwell Bridge)
 - ▶ Resonant circuits
(R L C circuits)

Measuring R

- ▶ Wheatstone bridge:



At the point of balance:

$$\frac{R_2}{R_1} = \frac{R_x}{R_3} \Rightarrow R_x = \frac{R_2}{R_1} \cdot R_3$$

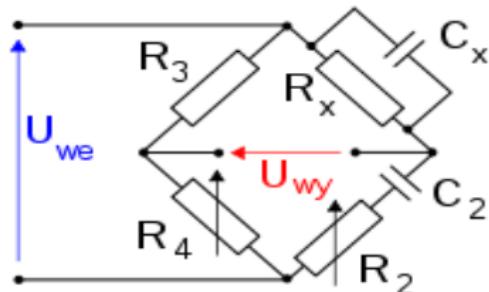
$$V_G = \left(\frac{R_x}{R_3 + R_x} - \frac{R_2}{R_1 + R_2} \right) V_s$$

- ▶ See Demo here:

<http://www.magnet.fsu.edu/education/tutorials/java/wheatstonebridge/index.html>

Measuring R

- ▶ The Wien bridge:



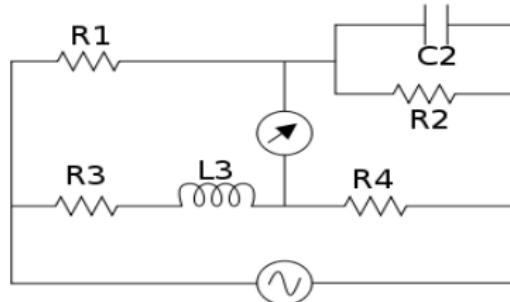
At some frequency, the reactance of the series R_2C_2 arm will be an exact multiple of the shunt R_xC_x arm. If the two R_3 and R_4 arms are adjusted to the same ratio, then the bridge is balanced.

$$\omega^2 = \frac{1}{R_x R_2 C_x C_2} \quad \text{and} \quad \frac{C_x}{C_2} = \frac{R_4}{R_3} - \frac{R_2}{R_x}.$$

The equations simplify if one chooses $R_2 = R_x$ and $C_2 = C_x$; the result is $R_4 = 2R_3$.

Measuring C

- ▶ Maxwell Bridge:



- ▶ R_1 and R_4 are known fixed entities. R_2 and C_2 are adjusted until the bridge is balanced.

$$R_3 = \frac{R_1 \cdot R_4}{R_2} \rightarrow L_3 = R_1 \cdot R_4 \cdot C_2$$

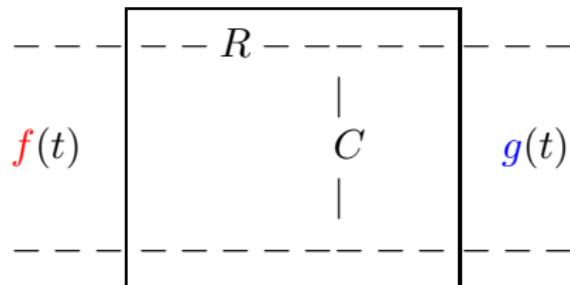
To avoid the difficulties associated with determining the precise value of a variable capacitance, sometimes a fixed-value capacitor will be installed and more than one resistor will be made variable.

Forward modeling and simulation of circuits

- ▶ Input-Output model
- ▶ R-C and R-L circuits
- ▶ L-C and R-L-C circuits
- ▶ Transfert function and impulse response
- ▶ General linear systems

Forward modeling and simulation of circuits

Input-Output model: Examples of $R - C$ circuits:



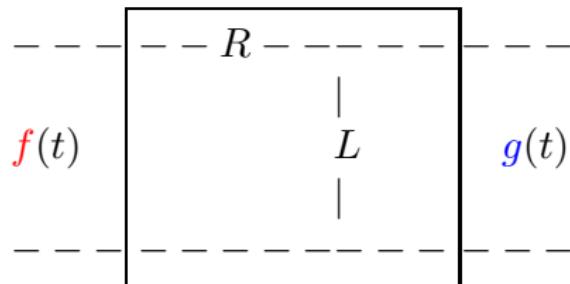
$$\frac{\partial \mathbf{g}(t)}{\partial t} = \frac{1}{C} i(t), \quad i(t) = \frac{(\mathbf{f}(t) - \mathbf{g}(t))}{R}$$

$$\frac{\partial \mathbf{g}(t)}{\partial t} = \frac{1}{RC} (\mathbf{f}(t) - \mathbf{g}(t)) \longrightarrow \mathbf{g}(t) + RC \frac{\partial \mathbf{g}(t)}{\partial t} = \mathbf{f}(t)$$

$$\mathbf{G}(\omega) + RCj\omega \mathbf{G}(\omega) = \mathbf{F}(\omega) \longrightarrow H(\omega) = \frac{\mathbf{G}(\omega)}{\mathbf{F}(\omega)} = \frac{1}{1 + jRC\omega}$$

Forward modeling and simulation of circuits

Input-Output model: Examples of $R - L$ circuits



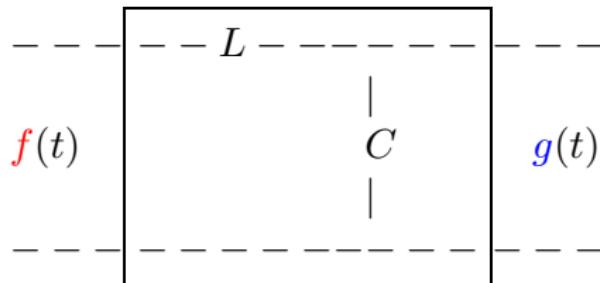
$$L \frac{\partial i(t)}{\partial t} = g(t), \quad i(t) = \frac{(f(t) - g(t))}{R}$$

$$\frac{L}{R} \frac{\partial(f(t) - g(t))}{\partial t} = g(t) \rightarrow g(t) + \frac{L}{R} \frac{\partial g(t)}{\partial t} = \frac{L}{R} j\omega f(t)$$

$$G(\omega) + \frac{L}{R} j\omega G(\omega) = \frac{L}{R} F(\omega) \rightarrow H(\omega) = \frac{G(\omega)}{F(\omega)} = \frac{1 + j\omega L/R}{L/R}$$

Forward modeling and simulation of circuits

$L - C$ circuits



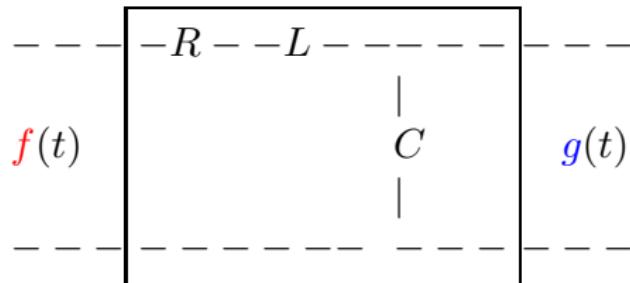
$$\frac{\partial \mathbf{g}(t)}{\partial t} = \frac{1}{C} \mathbf{i}(t), \quad L \frac{\partial \mathbf{i}(t)}{\partial t} = (\mathbf{f}(t) - \mathbf{g}(t))$$

$$LC \frac{\partial^2 \mathbf{g}(t)}{\partial t^2} = (\mathbf{f}(t) - \mathbf{g}(t)) \longrightarrow \mathbf{g}(t) + LC \frac{\partial^2 \mathbf{g}(t)}{\partial t^2} = \mathbf{f}(t)$$

$$\mathbf{G}(\omega) - LC\omega^2 \mathbf{G}(\omega) = \mathbf{F}(\omega) \longrightarrow H(\omega) = \frac{\mathbf{G}(\omega)}{\mathbf{F}(\omega)} = \frac{1}{1 - LC\omega^2}$$

Forward modeling and simulation of circuits

$R - L - C$ circuits



$$\frac{\partial \mathbf{g}(t)}{\partial t} = \frac{1}{C} i(t), \quad R i(t) + L \frac{\partial i(t)}{\partial t} = (\mathbf{f}(t) - \mathbf{g}(t))$$

$$RC \frac{\partial \mathbf{g}(t)}{\partial t} + LC \frac{\partial^2 \mathbf{g}(t)}{\partial t^2} = (\mathbf{f}(t) - \mathbf{g}(t))$$

$$\mathbf{g}(t) + RC \frac{\partial \mathbf{g}(t)}{\partial t} + LC \frac{\partial^2 \mathbf{g}(t)}{\partial t^2} = \mathbf{f}(t)$$

$$\mathbf{G}(\omega) + RCj\omega \mathbf{G}(\omega) - LC\omega^2 \mathbf{G}(\omega) = \mathbf{F}(\omega) \longrightarrow H(\omega) = \frac{\mathbf{G}(\omega)}{\mathbf{F}(\omega)} = \frac{1}{1 + jRC\omega - LC\omega^2}$$

Resonant circuits

- ▶ The resonant pulsation is:

$$\omega_0 = \sqrt{\frac{1}{LC}}$$

which gives:

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}}$$

Forward modeling and simulation of circuits

General linear systems:

$$\textcolor{red}{f}(t) \longrightarrow \boxed{H(\omega)} \longrightarrow \textcolor{blue}{g}(t)$$

- ▶ $H(\omega)$ is called Transfert function of the system.
- ▶ $h(t) = \text{IFT}\{H(\omega)\}$ is the impulse response of the system.
- ▶ Given $\textcolor{red}{f}(t)$ and $h(t)$ or $H(\omega)$ we can compute $\textcolor{blue}{g}(t)$.

$$\textcolor{blue}{G}(\omega) = H(\omega) \textcolor{red}{F}(\omega) \longrightarrow \textcolor{blue}{g}(t) = h(t) * \textcolor{red}{f}(t)$$

- ▶ $\textcolor{red}{f}(t) = \delta(t) \longrightarrow \textcolor{blue}{g}(t) = h(t)$
- ▶ $\textcolor{red}{f}(t) = u(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases} \longrightarrow \textcolor{blue}{g}(t) = \int_0^t h(t) \, dt$
- ▶ $\textcolor{red}{f}(t) = \sin(\omega t) \longrightarrow \textcolor{blue}{g}(t) = |H(\omega)| \sin(\omega t + \angle H(\omega))$

3- Data and signal processing of sensors output

- ▶ Some measurement systems defaults:
 - ▶ Full-scale error: Calibration
 - ▶ Offset error: Offset elimination
 - ▶ Drift: changes with temperature
 - ▶ Non-linearity
- ▶ In the following, we do not deal with these points. But, we are going to do:
 - ▶ Dealing with noise: Averaging, Filtering
 - ▶ Considering linear measurement systems, to model their input-output and then do inversion.
- ▶ Fixed averaging
- ▶ Moving Average (MA) filtering
- ▶ Autoregressive (AR) filtering
- ▶ Moving Average Autoregressive (ARMA) filtering
- ▶ Errors, noise and uncertainties → Background on Probability theory

Background on Probability theory

- ▶ Why we need Probability theory ?
- ▶ What is the significance of probability ?
- ▶ What means a random variable ?

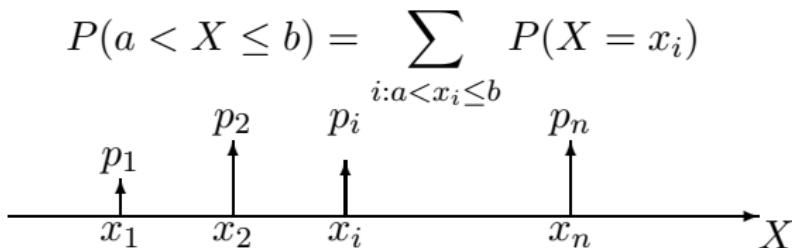
- ▶ Discrete variables $\{x_1, \dots, x_n\}$
Probability distribution: $\{p_1, \dots, p_n\}$ with $\sum p_n = 1$
- ▶ Continuous variables $x \in \mathcal{R}$ or $x \in \mathcal{R}_+$ or $x \in [a, b]$
Probability density function $p(x)$ with $\int_{-\infty}^{+\infty} p(x) dx = 1$,
Partition function: $F(x) = P(X \leq x) = \int_{-\infty}^x p(x) dx$
- ▶ Expected value: $E\{X\} = \int x p(x) dx$
- ▶ Variance value: $\text{Var}\{X\} = \int (x - E\{X\})^2 p(x) dx$
- ▶ Mode value $\text{Mode} = \arg \max_x \{p(x)\}$

- ▶ Normal distribution $\mathcal{N}(x|m, v)$
- ▶ Gamma distribution $\mathcal{G}(x|\alpha, \beta)$

Discrete random variables

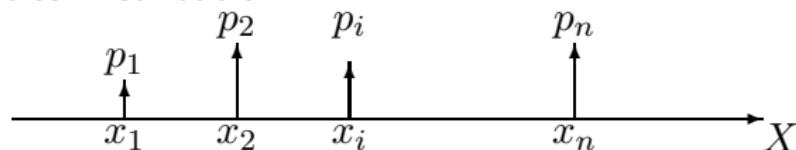
- ▶ X takes values x_i with probabilities p_i , $i = 1, \dots, n$.
- ▶ $P(X = x_i) = p_i$, $i = 1, \dots, n$ is probability distribution (pd).
- ▶ If we sort x_i in such a way that $x_1 \leq x_2 \leq \dots \leq x_n$, then we can define the "probability cumulative distribution (pcd)":

$$F(x) = P(X \leq x) = \sum_{i:x_i \leq x} P(X = x_i)$$

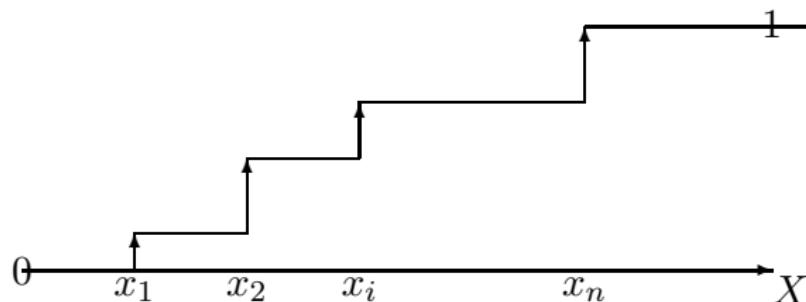


Discrete events

- ▶ Probabilities Distribution



- ▶ Cumulative Probability Distribution



Discrete events

- ▶ Expected value

$$\mathbb{E}\{X\} = \langle X \rangle = \sum_i p_i x_i$$

- ▶ Variance

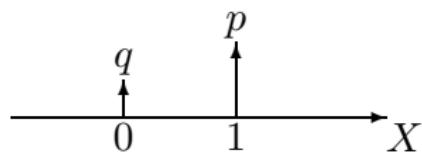
$$\text{Var}\{X\} = \sum_i p_i (x_i - \mathbb{E}\{X\})^2 = \sum_i p_i (x_i - \langle X \rangle)^2$$

- ▶ Entropy

$$H(X) = - \sum_i p_i \ln p_i$$

Discrete variables probability distributions

- ▶ Bernoulli distribution: A variable with two outcomes only
 $X = \{0, 1\}$, $P(X = 1) = p$, $P(X = 0) = q = 1 - p$



- ▶ Bernoulli trial $B(n, p)$: n independent trials of an experiment with two outcomes only 0010001100000010
 - ▶ p probability of success
 - ▶ $q = 1 - p$ probability of failure
- ▶ Binomial distribution $\text{Bin}(\cdot | n, p)$:
The probability of k successes in n trials:

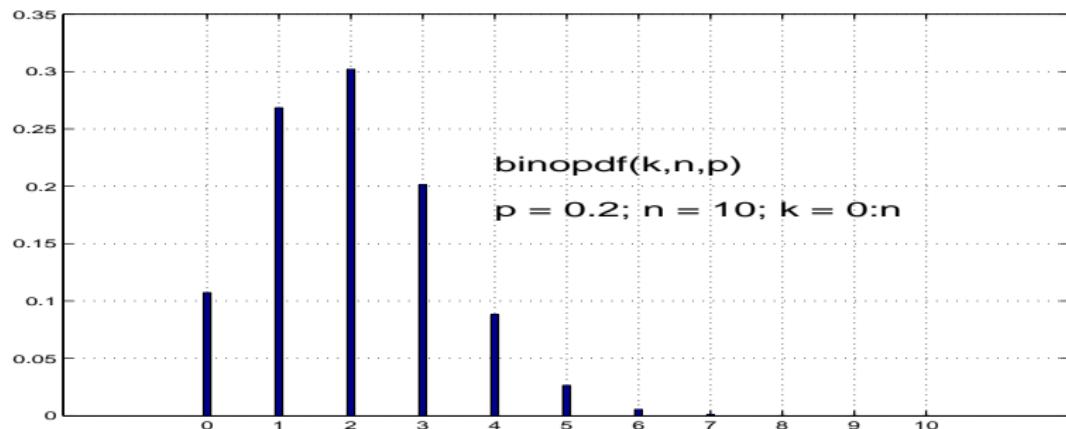
$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Binomial distribution $\text{Bin}(\cdot|n, p)$

The probability of k successes in n trials:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, k = 0, 1, \dots, n$$

$$\mathbb{E}\{X\} = np, \quad \text{Var}\{X\} = npq = np(1 - p)$$



Poisson distribution

- ▶ The Poisson distribution can be derived as a limiting case to the binomial distribution as the number of trials goes to infinity and the expected number of successes remains fixed

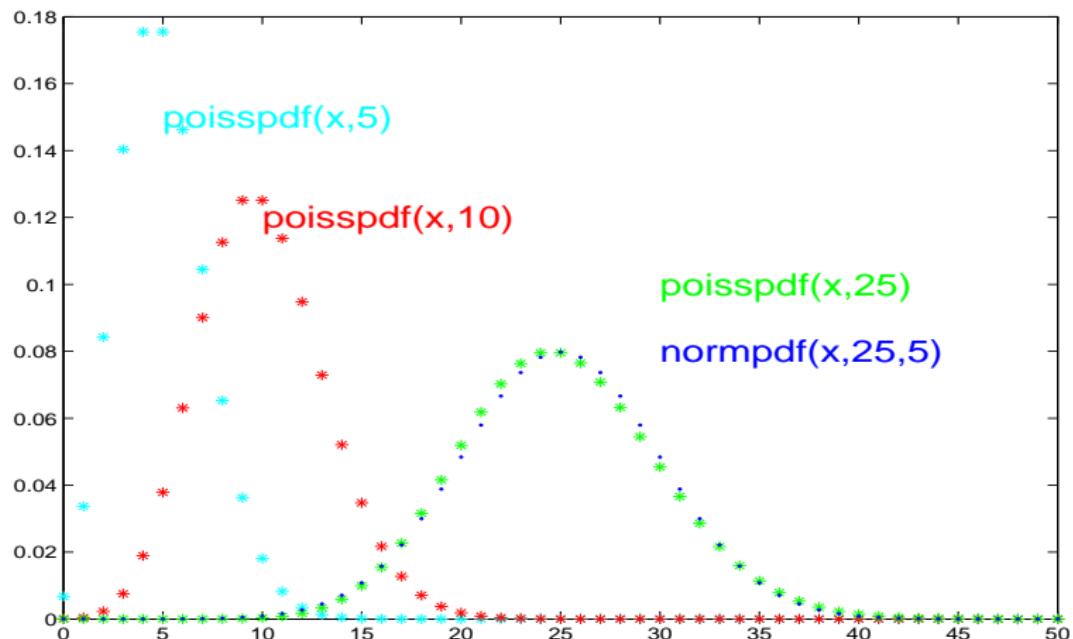
$$X \sim \text{Bin}(n, p) \underset{n \rightarrow \infty, np \rightarrow \lambda}{\lim} X \sim \mathcal{P}(\lambda)$$

$$P(X = k | \lambda) = \frac{\lambda^k \exp[-\lambda]}{k!}$$

$$\mathbb{E}\{X\} = \lambda, \quad \text{Var}\{X\} = \lambda$$

- ▶ If $X_n \sim \text{Bin}(n, \lambda/n)$ and $Y \sim \mathcal{P}(\lambda)$ then for each fixed k , $\lim_{n \rightarrow \infty} P(X_n = k) = P(Y = k)$.

Poisson distribution



Continuous case

- ▶ Cumulative Distribution Function (cdf): $F(x) = P(X < x)$
- ▶ Measure theory

$$P(a \leq X < b) = F(b) - F(a)$$

$$P(x \leq X < x + dx) = F(x + dx) - F(x) = dF(x)$$

- ▶ If $F(x)$ is a continuous function

$$p(x) = \frac{\partial F(x)}{\partial x}$$

- ▶ $p(x)$ probability density function (pdf)

$$P(a < X \leq b) = \int_a^b p(x) \, dx$$

- ▶ Cumulative distribution function (cdf)

$$F(x) = \int_{-\infty}^x p(x) \, dx$$

Continuous case

- ▶ Expected value

$$\mathsf{E}\{X\} = \int x p(x) \, dx = \langle X \rangle$$

- ▶ Variance

$$\mathsf{Var}\{X\} = \int (x - \mathsf{E}\{X\})^2 p(x) \, dx = \langle (x - \mathsf{E}\{X\})^2 \rangle$$

- ▶ Entropy

$$\mathsf{H}(X) = \int -\ln p(x) p(x) \, dx = \langle -\ln p(X) \rangle$$

- ▶ Mode: $\text{Mode}(X) = \arg \max_x \{p(x)\}$
- ▶ Median $\text{Med}(X)$:

$$\int_{-\infty}^{\text{Med}(X)} p(x) \, dx = \int_{\text{Med}(X)}^{+\infty} p(x) \, dx$$

Uniform and Beta distributions

- ▶ Uniform:

$$X \sim U(.|a, b) \longrightarrow p(x) = \frac{1}{b-a}, \quad x \in [a, b]$$

$$\mathbb{E}\{X\} = \frac{a+b}{2}, \quad \text{Var}\{X\} = \frac{(b-a)^2}{12}$$

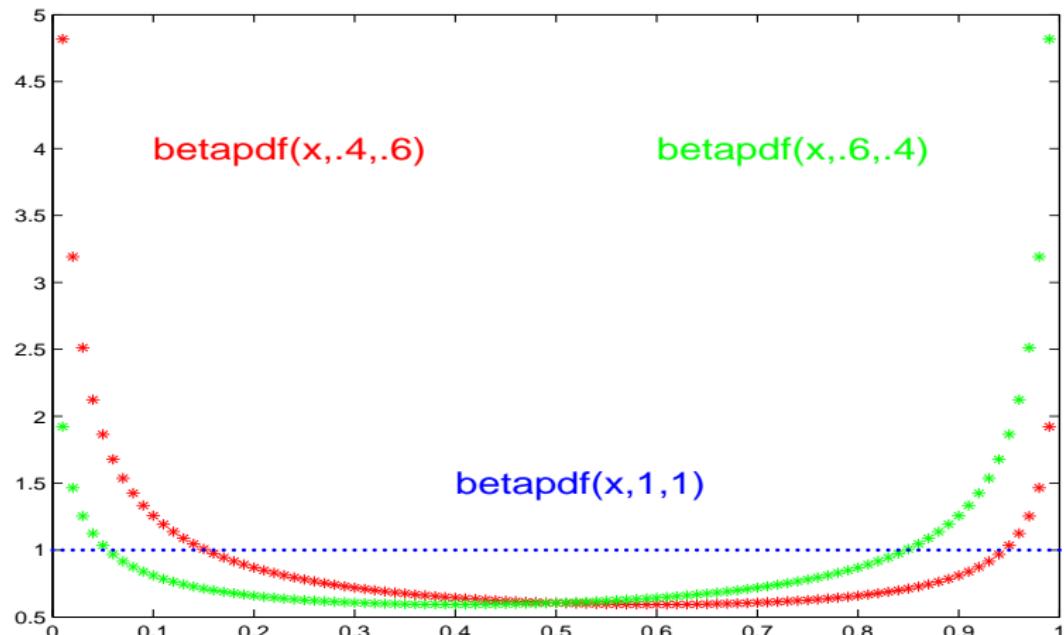
- ▶ Beta:

$$X \sim \text{Beta}(.|\alpha, \beta) \longrightarrow p(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, x \in [0, 1]$$

$$\mathbb{E}\{X\} = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}\{X\} = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

- ▶ $\text{Beta}(.|1, 1) = U(.|0, 1)$

Uniform and Beta distributions



Gaussian distributions

Different notations:

- ▶ classical one with mean and variance:

$$X \sim \mathcal{N}(\cdot | \mu, \sigma^2) \longrightarrow p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2}(x - \mu)^2 \right]$$

$$\mathbb{E}\{X\} = \mu, \quad \text{Var}\{X\} = \sigma^2$$

- ▶ mean and precision parameters:

$$X \sim \mathcal{N}(\cdot | \mu, \lambda) \longrightarrow p(x) = \frac{\lambda}{\sqrt{2\pi}} \exp \left[-\frac{\lambda}{2}(x - \mu)^2 \right]$$

$$\mathbb{E}\{X\} = \mu, \quad \text{Var}\{X\} = \sigma^2 = \frac{1}{\lambda}$$

Generalized Gaussian distributions

- ▶ Gaussian:

$$X \sim \mathcal{N}(\cdot | \mu, \sigma^2) \longrightarrow p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2\right]$$

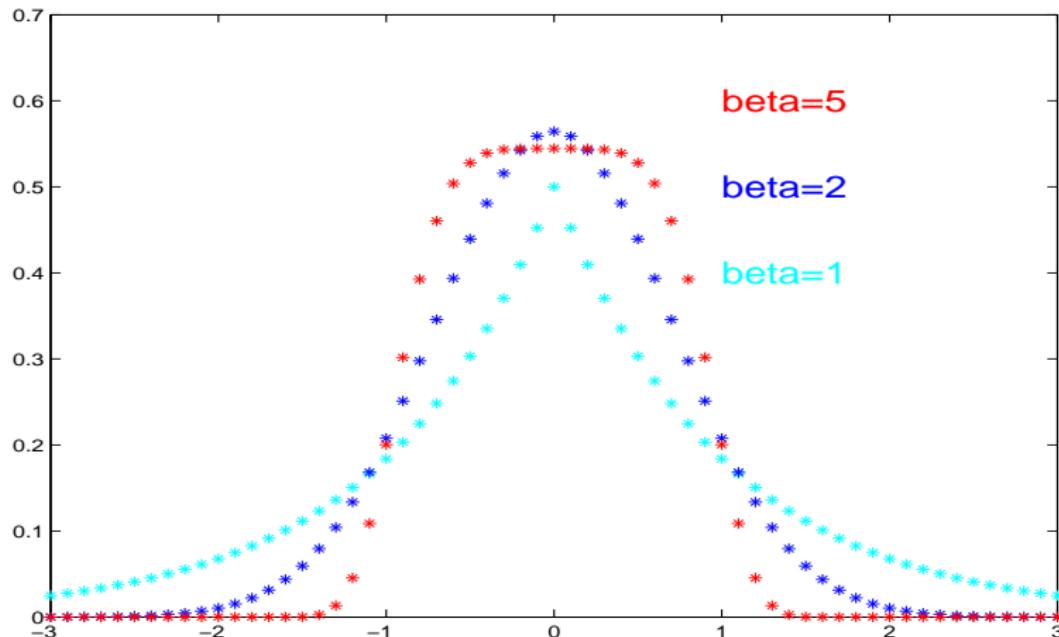
- ▶ Generalized Gaussian:

$$X \sim \mathcal{GG}(\cdot | \alpha, \beta) \longrightarrow p(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left[-\left(\frac{|x-\mu|}{\alpha}\right)^\beta\right]$$

$$\mathbb{E}\{X\} = \mu, \quad \text{Var}\{X\} = \frac{\alpha^2\Gamma(3/\beta)}{\gamma(1/\beta)}$$

- ▶ $\beta > 0$, $\beta = 1$ Laplace, $\beta = 2$: Gaussian, $\beta \mapsto \infty$: Uniform

Gaussian and Generalized Gaussian distributions



Gamma distributions

- ▶ Forme 1:

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \text{ for } x \geq 0$$

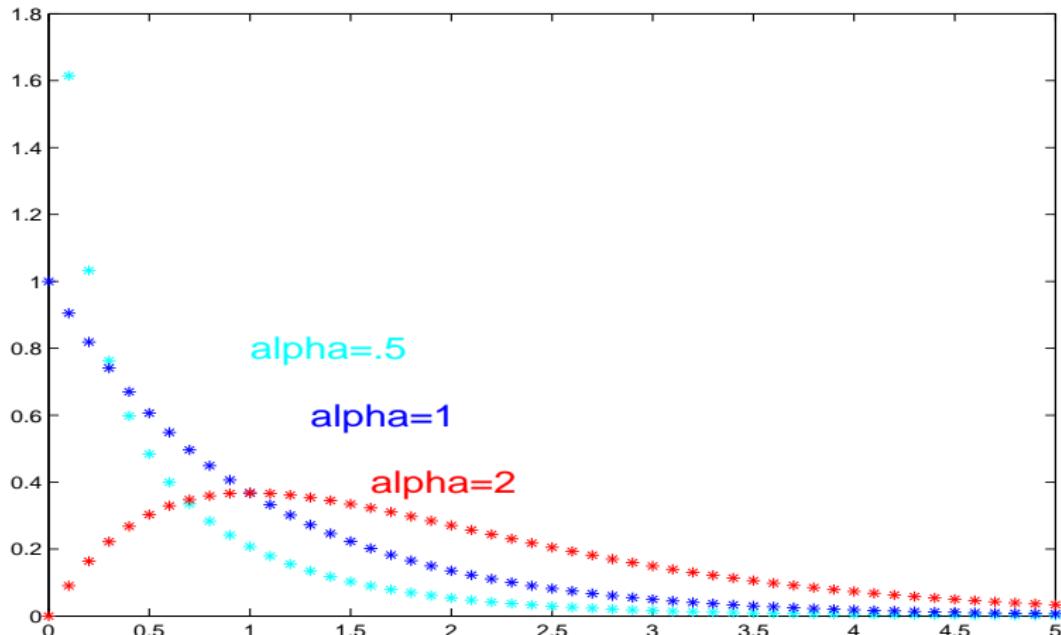
$$\mathsf{E}\{X\} = \frac{\alpha}{\beta}, \quad \mathsf{Var}\{X\} = \frac{\alpha}{\beta^2}, \quad \mathsf{Mod}(X) = \frac{\alpha - 1}{\alpha + \beta - 2}$$

- ▶ Forme 2: $\theta = 1/\beta$

$$p(x|\alpha, \theta) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta} \text{ for } x \geq 0$$

- ▶ $\alpha = 1$: Exponential,
- ▶ $0 < \alpha < 1$: decreasing,
- ▶ $\alpha > 1$: Mode = $\frac{\alpha-1}{\beta}$

Gamma distributions



Student-t and Cauchy distributions

- ▶ Student's t-distribution has the probability density function:

$$p(x|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\sqrt{\nu} B(\frac{1}{2}, \frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where

- ▶ ν is the number of degrees of freedom,
- ▶ Γ is the Gamma function and
- ▶ B is the Beta function.
- ▶ $\nu = 1$ gives Cauchy distribution.

$$p(x) = \frac{\pi}{1+x^2}$$

- ▶ Cauchy distribution:

$$p(t|\mu) = \frac{\pi}{1+(x-\mu)^2}$$

Student-t and Cauchy distributions

- ▶ Three parameters location (μ) / scale (λ) / degree of freedom (ν) version

$$p(x|\mu, \lambda, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{\lambda}{\pi\nu} \right)^{\frac{1}{2}} \left[1 + \frac{\lambda(x-\mu)^2}{\nu} \right]^{-\frac{\nu+1}{2}}$$

$$\begin{aligned}\mathbb{E}\{X\} &= \mu && \text{for } \nu > 1, \\ \text{Var}\{X\} &= \frac{1}{\lambda} \frac{\nu}{\nu-2} && \text{for } \nu > 2, \\ \text{mode}(X) &= \mu.\end{aligned}$$

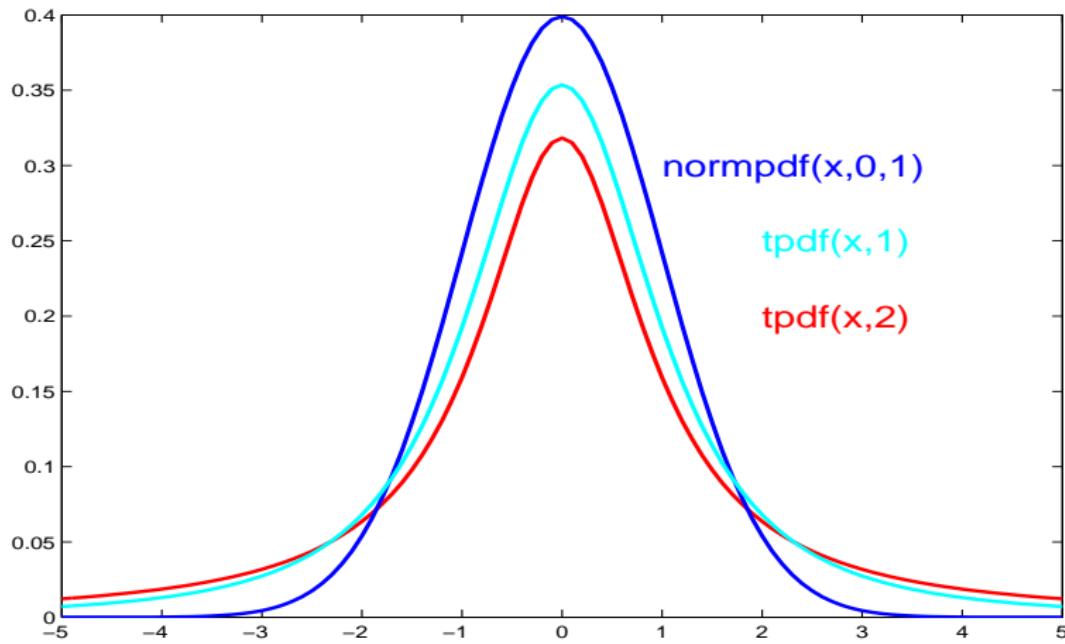
- ▶ Interesting relation between Student-t, Normal and Gamma distributions:

$$\mathcal{S}(x|\mu, 1, \nu) = \int \mathcal{N}(x|\mu, 1/\lambda) \mathcal{G}(\lambda|\nu/2, \nu/2) d\lambda$$

$$\mathcal{S}(x|0, 1, \nu) = \int \mathcal{N}(x|0, 1/\lambda) \mathcal{G}(\lambda|\nu/2, \nu/2) d\lambda$$

Student and Cauchy

$$p(x|\nu) \propto \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$



Vector variables

- ▶ Vector variables: $\mathbf{X} = [X_1, X_2, \dots, X_n]'$
- ▶ $p(\mathbf{x})$ probability density function (pdf)
- ▶ Expected value

$$\mathbb{E}\{\mathbf{X}\} = \int \mathbf{x} p(\mathbf{x}) \, d\mathbf{x} = \langle \mathbf{X} \rangle$$

- ▶ Covariance

$$\begin{aligned}\text{cov}[\mathbf{X}] &= \int (\mathbf{X} - \mathbb{E}\{\mathbf{X}\})(\mathbf{X} - \mathbb{E}\{\mathbf{X}\})' p(\mathbf{x}) \, d\mathbf{x} \\ &= \langle (\mathbf{X} - \mathbb{E}\{\mathbf{X}\})(\mathbf{X} - \mathbb{E}\{\mathbf{X}\})' \rangle\end{aligned}$$

- ▶ Entropy

$$\mathbb{E}(\mathbf{X}) = \int -\ln p(\mathbf{x}) p(\mathbf{x}) \, d\mathbf{x} = \langle \ln p(\mathbf{X}) \rangle$$

- ▶ Mode: $\text{Mode}(p(\mathbf{x})) = \arg \max_{\mathbf{x}} \{p(\mathbf{x})\}$

Vector variables

- ▶ Case of a vector with 2 variables: $\mathbf{X} = [X_1, X_2]'$
- ▶ $p(\mathbf{x}) = p(x_1, x_2)$ joint probability density function (pdf)
- ▶ Marginals

$$\begin{aligned} p(x_1) &= \int p(x_1, x_2) \, dx_2 \\ p(x_2) &= \int p(x_1, x_2) \, dx_1 \end{aligned}$$

- ▶ Conditionals

$$\begin{aligned} p(x_1|x_2) &= \frac{p(x_1, x_2)}{p(x_2)} \\ p(x_2|x_1) &= \frac{p(x_1, x_2)}{p(x_1)} \end{aligned}$$

Multivariate Gaussian

Different notations:

- ▶ mean and covariance matrix (classical): $\mathbf{X} \sim \mathcal{N}(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$p(\mathbf{x}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

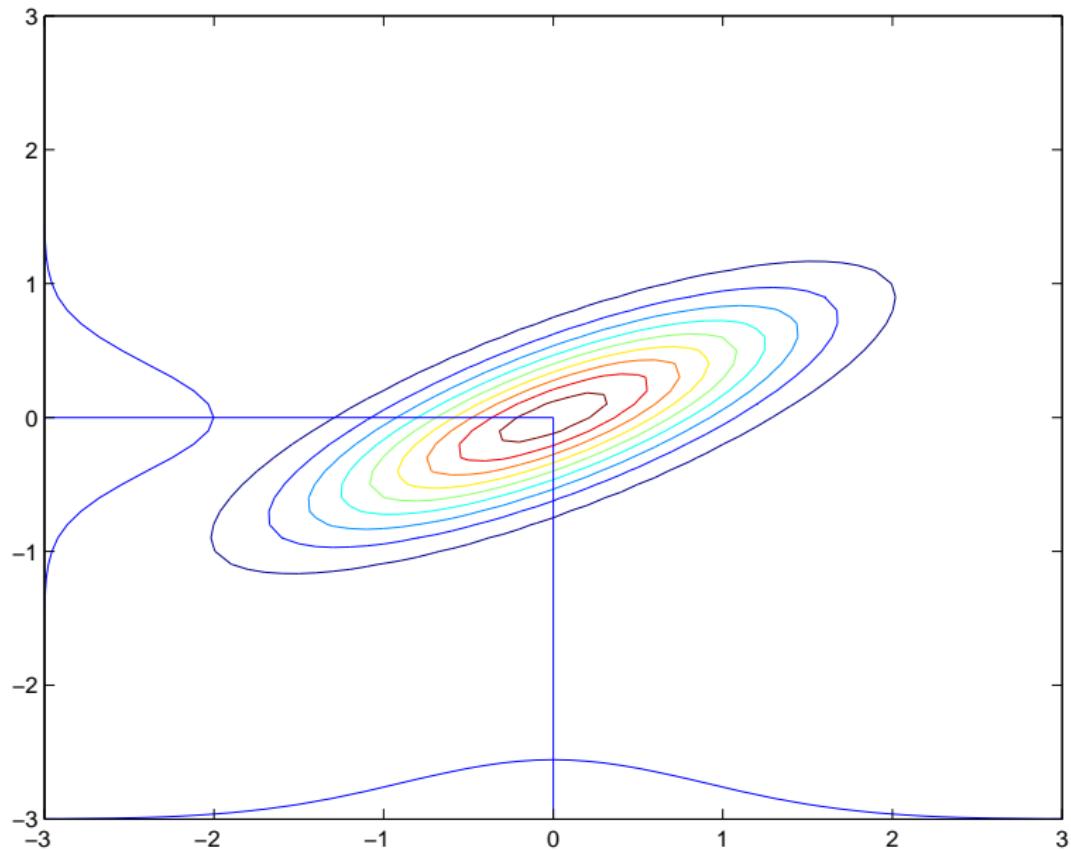
$$\mathbb{E}\{\mathbf{X}\} = \boldsymbol{\mu}, \quad \text{cov}[\mathbf{X}] = \boldsymbol{\Sigma}$$

- ▶ mean and precision matrix: $\mathbf{X} \sim \mathcal{N}(\cdot | \boldsymbol{\mu}, \boldsymbol{\Lambda})$

$$p(\mathbf{x}) = (2\pi)^{-n/2} |\boldsymbol{\Lambda}|^{1/2} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Lambda} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

$$\mathbb{E}\{\mathbf{X}\} = \boldsymbol{\mu}, \quad \text{cov}[\mathbf{X}] = \boldsymbol{\Lambda}^{-1}$$

Multivariate normal distributions



Multivariate Student-t

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \propto |\boldsymbol{\Sigma}|^{-1/2} \left[1 + \frac{1}{\nu} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{-(\nu+p)/2}$$

- $p = 1$

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}} (1+t^2/\nu)^{\frac{-(\nu+1)}{2}}$$

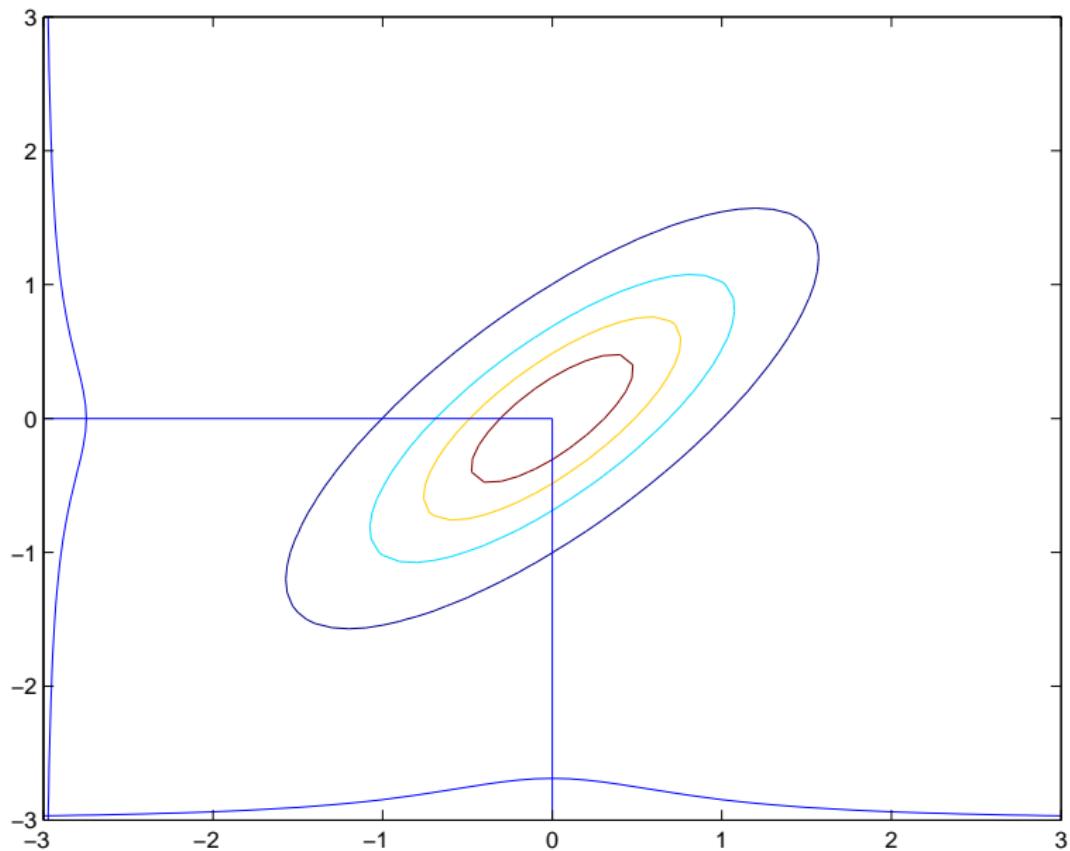
- $p = 2, \boldsymbol{\Sigma}^{-1} = \mathbf{A}$

$$f(t_1, t_2) = \frac{\Gamma((\nu+p)/2)}{\Gamma(\nu/2)\sqrt{\nu^p\pi^p}} \frac{|\mathbf{A}|^{1/2}}{2\pi} \left(1 + \sum_{i=1}^p \sum_{j=1}^p A_{ij} t_i t_j / \nu \right)^{\frac{-(\nu+2)}{2}}$$

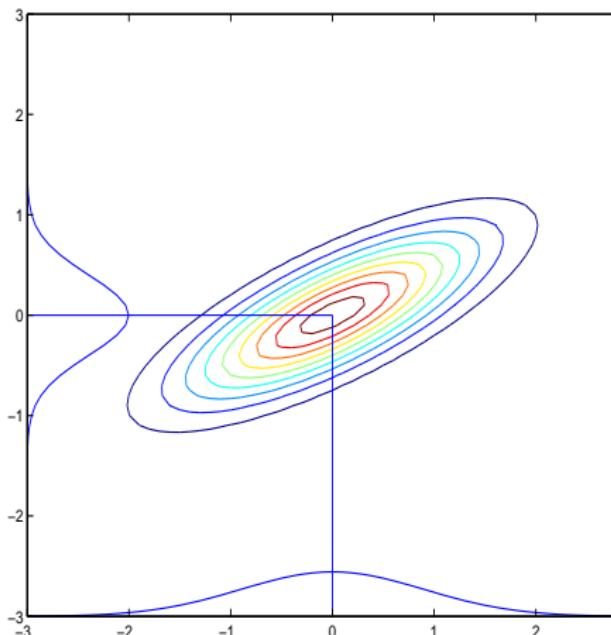
- $p = 2, \boldsymbol{\Sigma} = \mathbf{A} = \mathbf{I}$

$$f(t_1, t_2) = \frac{1}{2\pi} (1 + (t_1^2 + t_2^2)/\nu)^{\frac{-(\nu+2)}{2}}$$

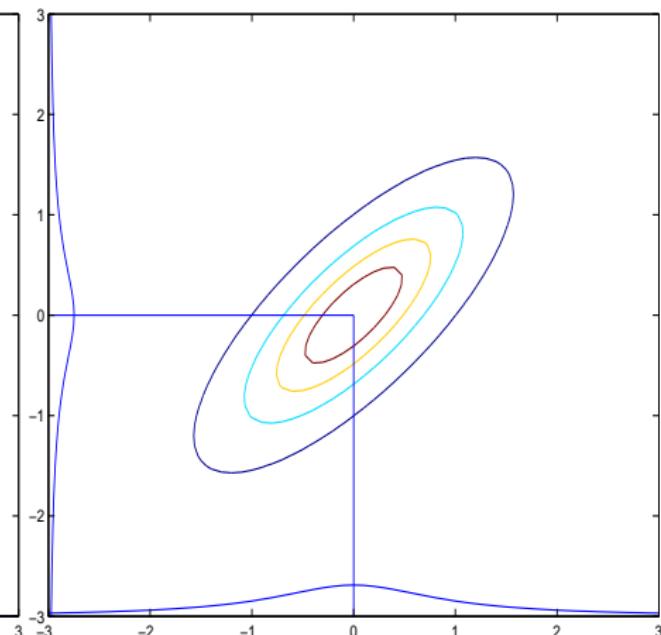
Multivariate Student-t distributions



Multivariate normal distributions



Normal



Student-t

Dealing with noise, errors and uncertainties

- ▶ Sample averaging: mean and standard deviation

$$\bar{x} = \frac{1}{n} \sum_{n=1}^N x_n$$

$$S = \sqrt{\frac{1}{n-1} \sum_{n=1}^N (x_n - \bar{x})^2}$$

- ▶ Recursive computation: moving average

$$\bar{x}_k = \frac{1}{n} \sum_{i=k-n+1}^k x_i, \quad \bar{x}_{k-1} = \frac{1}{n} \sum_{i=k-n}^{k-1} x_i$$

$$\bar{x}_k = \bar{x}_{k-1} + \frac{1}{n}(x_k - x_{k-n})$$

Dealing with noise

- ▶ Exponential moving average

$$\bar{x}_k = \frac{1}{n} \sum_{i=k-n+1}^k x_i, \quad \bar{x}_{k+1} = \frac{1}{n+1} \sum_{i=k-n+1}^{k+1} x_i$$

$$\bar{x}_{k+1} = \frac{n}{n+1} \bar{x}_k + \frac{1}{n+1} x_{k+1}$$

$$\bar{x}_k = \frac{n}{n+1} \bar{x}_{k-1} + \frac{1}{n+1} x_k = \alpha \bar{x}_{k-1} + (1 - \alpha) x_k$$

- ▶ The Exponentially Weighted Moving Average filter places more importance to more recent data by discounting older data in an exponential manner

$$\bar{x}_k = \alpha \bar{x}_{k-1} + (1 - \alpha) x_k = \alpha [\alpha \bar{x}_{k-2} + (1 - \alpha) x_{k-1}] (1 - \alpha) x_k$$

$$\bar{x}_k = \alpha \bar{x}_{k-1} + (1 - \alpha) x_k = \alpha^2 \bar{x}_{k-2} + \alpha (1 - \alpha) x_{k-1} (1 - \alpha) x_k$$

$$\bar{x}_k = \alpha^3 \bar{x}_{k-3} + \alpha^2 (1 - \alpha) x_{k-2} + \alpha (1 - \alpha) x_{k-1} + (1 - \alpha) x_k$$

Dealing with noise

- ▶ Exponential moving average

$$\bar{x}_k = \alpha \bar{x}_{k-1} + (1 - \alpha) x_k$$

$$\bar{x}_k = \alpha^2 \bar{x}_{k-2} + \alpha(1 - \alpha)x_{k-1}(1 - \alpha)x_k$$

$$\bar{x}_k = \alpha^3 \bar{x}_{k-3} + \alpha^2(1 - \alpha)x_{k-2} + \alpha(1 - \alpha)x_{k-1} + (1 - \alpha)x_k$$

- ▶ The Exponentially Weighted Moving Average filter is identical to the discrete first-order low-pass filter:
- ▶ Consider the Laplace transform function of a first-order low-pass filter, with time constant τ :

$$\frac{\bar{x}(s)}{x(s)} = \frac{1}{1 + \tau s} \rightarrow \tau \frac{\partial \bar{x}(t)}{\partial t} + \bar{x}(t) = x(t)$$

$$\frac{\partial \bar{x}(t)}{\partial t} = \frac{\bar{x}_k - \bar{x}_{k-1}}{T_s} \rightarrow \bar{x}_k = \left(\frac{\tau}{\tau + T_s} \right) \bar{x}_{k-1} + \left(\frac{T_s}{\tau + T_s} \right) x_k$$

Other Filters

- ▶ First order filter:

$$\frac{\bar{x}(s)}{x(s)} = H(s) = \frac{1}{(1 + \tau s)}$$

- ▶ Second order filter:

$$H(s) = \frac{1}{(1 + \tau s)^2}$$

- ▶ Third order filter:

$$H(s) = \frac{1}{(1 + \tau s)^3}$$

- ▶ Bode diagram of the filter transfer function as a function of τ and as a function of the order of the filter.

Background on linear invariant systems

- ▶ A linear and invariant system: Time representation

$$f(t) \longrightarrow \boxed{h(t)} \longrightarrow g(t)$$

- ▶ A linear and invariant system: Fourier Transform representation

$$F(\omega) \longrightarrow \boxed{H(\omega)} \longrightarrow G(\omega)$$

- ▶ A linear and invariant system: Laplace Transform representation

$$F(s) \longrightarrow \boxed{H(s)} \longrightarrow G(s)$$

Sampling theorem and digital linear invariant systems

- ▶ Link between the FTs of a continuous signal and its sampled version
- ▶ Sampling theorem: If a Band limited signal ($|F(\omega)| = 0, \forall \omega > \Omega_0$) is sampled with a sampling frequency $f_s = \frac{1}{T_s}$ two times greater than its maximum frequency ($2\pi f_s \geq 2\Omega_0$), its can be reconstructed without error from its samples by an ideal low pass filtering.
- ▶ Z-Transform is used in place of Laplace Transform to handle with digital signals
- ▶ A numerical or digital linear and invariant system:

$$f(n) \longrightarrow \boxed{h(n)} \longrightarrow g(n)$$

$$F(z) \longrightarrow \boxed{H(z)} \longrightarrow G(z)$$

Moving Average (MA)



- ▶ Convolution
 - ▶ Continuous

$$g(t) = h(t) * f(t) = \int h(\tau)f(t - \tau) d\tau$$

- ▶ Discrete

$$g(n) = \sum_{k=0}^q h(k)f(n - k), \quad \forall n$$

- ▶ Filter transfer function

$$f(n) \longrightarrow \boxed{H(z) = \sum_{k=0}^q h(k)z^{-k}} \longrightarrow g(n)$$

Autoregressive (AR)

- Continuous

$$g(t) = \sum_{k=1}^p a(k) g(t - k\Delta t) + f(t)$$

- Discrete

$$g(n) = \sum_{k=1}^p a(k) g(n - k) + f(n), \quad \forall n$$

- Filter transfer function

$$f(n) \longrightarrow \boxed{H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{k=1}^p a(k) z^{-k}}} \longrightarrow g(n)$$

Autoregressive Moving Average (ARMA)

- Continuous

$$g(t) = \sum_{k=1}^p a(k) g(t - k\Delta t) + \sum_{l=0}^q b(l) f(t - l\Delta t) dt$$

- Discrete

$$g(n) = \sum_{k=1}^p a(k) g(n - k) + \sum_{l=0}^q b(l) f(n - l)$$

$$\epsilon(n) \rightarrow \boxed{H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^q b(k)z^{-k}}{1 + \sum_{k=1}^p a(k) z^{-k}}} \rightarrow f(n)$$

- Filter transfer function

$$\epsilon(n) \rightarrow \boxed{B_q(z)} \rightarrow \boxed{\frac{1}{A_p(z)}} \rightarrow f(n)$$

Parameter estimation

We observe n samples $\boldsymbol{x} = \{x_1, \dots, x_n\}$ of a quantity X whose pdf depends on certain parameters $\boldsymbol{\theta}$: $p(\boldsymbol{x}|\boldsymbol{\theta})$.

The question is to determine $\boldsymbol{\theta}$.

- Moments method:

$$\mathbb{E}\left\{x^k\right\} = \int x^k p(x|\boldsymbol{\theta}) \, dx \approx \frac{1}{n} \sum_{i=1}^n x_i^k, \quad k = 1, \dots, K$$

- Maximum Likelihood

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^n p(x_i|\boldsymbol{\theta}) \text{ or } \ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \ln p(x_i|\boldsymbol{\theta})$$

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{\mathcal{L}(\boldsymbol{\theta})\} = \arg \min_{\boldsymbol{\theta}} \{-\ln \mathcal{L}(\boldsymbol{\theta})\}$$

- Bayesian approach

Bayesian Parameter estimation

- ▶ Likelihood

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n p(x_i|\boldsymbol{\theta})$$

- ▶ A priori

$$p(\boldsymbol{\theta})$$

- ▶ A posteriori

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

- ▶ Infer on $\boldsymbol{\theta}$ using $p(\boldsymbol{\theta}|\mathbf{x})$.

For example:

- ▶ Maximum A Posteriori (MAP)

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\mathbf{x})\}$$

- ▶ Posterior Mean

$$\hat{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{x}) \, d\boldsymbol{\theta}$$

Parameter estimation: Normal distribution

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$p(\mu, \sigma^2 | \mathbf{x}) = \frac{p(\mu, \sigma^2)}{p(\mathbf{x})} \prod_{i=1}^N p(x_i | \mu, \sigma)$$

$$p(\mu, \sigma^2 | \mathbf{x}) = \frac{p(\mu, \sigma^2)}{p(\mathbf{x})} \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{and} \quad s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$p(\mu, \sigma^2 | \mathbf{x}) = \frac{p(\mu, \sigma^2)}{p(\mathbf{x})} \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{(\mu - \bar{x})^2 + s^2}{2\sigma^2/N}\right]$$

Parameter estimation: Normal distribution: σ known

- σ^2 known: $p(\mu, \sigma^2) = p(\mu) \delta(\sigma - \sigma_0)$

$$\begin{aligned} p(\mu|\mathbf{x}) &= \frac{p(\mu)}{p(\mathbf{x})} \frac{1}{(2\pi\sigma_0^2)^{N/2}} \exp \left[-\sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma_0^2} \right] \\ &= \frac{p(\mu)}{p(\mathbf{x})} \frac{1}{(2\pi\sigma_0^2)^{N/2}} \exp \left[-\frac{(\mu - \bar{x})^2 + s^2}{2\sigma_0^2/N} \right] \\ &\propto p(\mu) \exp \left[-\frac{(\mu - \bar{x})^2}{2\sigma_0^2/N} \right] \end{aligned}$$

- $p(\mu) = c \longrightarrow p(\mu|\mathbf{x}) = \mathcal{N}(\mu|\bar{x}, \sigma_0^2/N)$

$$\mu = \bar{x} \pm \frac{\sigma_0}{\sqrt{N}}$$

- $p(\mu) = \mathcal{N}(\mu_0, v_0) \longrightarrow p(\mu|\mathbf{x}) = \mathcal{N}(\mu|\hat{\mu}, \hat{v})$

$$\hat{\mu} = \frac{v_0}{v_0 + \sigma_0^2} \bar{x} + \frac{\sigma_0^2}{v_0 + \sigma_0^2} \mu_0, \quad \hat{v} = \frac{v_0 + \sigma_0^2}{v_0 \sigma_0^2}$$

Parameter estimation

We observe n samples $\boldsymbol{x} = \{x_1, \dots, x_n\}$ of a quantity X whose pdf depends on certain parameters $\boldsymbol{\theta}$: $p(\boldsymbol{x}|\boldsymbol{\theta})$.

The question is to determine $\boldsymbol{\theta}$.

- Moments method:

$$\mathbb{E}\left\{x^k\right\} = \int x^k p(x|\boldsymbol{\theta}) dx \approx \frac{1}{n} \sum_{i=1}^n x_i^k, \quad k = 1, \dots, K$$

- Maximum Likelihood

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^n p(x_i|\boldsymbol{\theta}) \text{ or } \ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \ln p(x_i|\boldsymbol{\theta})$$

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{\mathcal{L}(\boldsymbol{\theta})\} = \arg \min_{\boldsymbol{\theta}} \{-\ln \mathcal{L}(\boldsymbol{\theta})\}$$

- Bayesian approach

Bayesian Parameter estimation

- ▶ Likelihood

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n p(x_i|\boldsymbol{\theta})$$

- ▶ A priori

$$p(\boldsymbol{\theta})$$

- ▶ A posteriori

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

- ▶ Infer on $\boldsymbol{\theta}$ using $p(\boldsymbol{\theta}|\mathbf{x})$.

For example:

- ▶ Maximum A Posteriori (MAP)

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\mathbf{x})\}$$

- ▶ Posterior Mean

$$\hat{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{x}) \, d\boldsymbol{\theta}$$

Parameter estimation: Normal distribution

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$p(\mu, \sigma^2 | \mathbf{x}) = \frac{p(\mu, \sigma^2)}{p(\mathbf{x})} \prod_{i=1}^N p(x_i | \mu, \sigma)$$

$$p(\mu, \sigma^2 | \mathbf{x}) = \frac{p(\mu, \sigma^2)}{p(\mathbf{x})} \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{and} \quad s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$p(\mu, \sigma^2 | \mathbf{x}) = \frac{p(\mu, \sigma^2)}{p(\mathbf{x})} \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{(\mu - \bar{x})^2 + s^2}{2\sigma^2/N}\right]$$

Parameter estimation: Normal distribution: σ known

- σ^2 known: $p(\mu, \sigma^2) = p(\mu) \delta(\sigma - \sigma_0)$

$$\begin{aligned} p(\mu | \mathbf{x}) &= \frac{p(\mu)}{p(\mathbf{x})} \frac{1}{(2\pi\sigma_0^2)^{N/2}} \exp \left[-\sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma_0^2} \right] \\ &= \frac{p(\mu)}{p(\mathbf{x})} \frac{1}{(2\pi\sigma_0^2)^{N/2}} \exp \left[-\frac{(\mu - \bar{x})^2 + s^2}{2\sigma_0^2/N} \right] \\ &\propto p(\mu) \exp \left[-\frac{(\mu - \bar{x})^2}{2\sigma_0^2/N} \right] \end{aligned}$$

- $p(\mu) = \mathcal{N}(\mu_0, v_0) \longrightarrow p(\mu | \mathbf{x}) = \mathcal{N}(\mu | \hat{\mu}, \hat{v})$

$$\hat{\mu} = \frac{v_0}{v_0 + \sigma_0^2} \bar{x} + \frac{\sigma_0^2}{v_0 + \sigma_0^2} \mu_0, \quad \hat{v} = \frac{v_0 + \sigma_0^2}{v_0 \sigma_0^2}$$

Conjugate priors

Observation law $p(x \theta)$	Prior law $p(\theta \tau)$	Posterior law $p(\theta x, \tau) \propto p(\theta \tau)p(x \theta)$
Binomial $\text{Bin}(x n, \theta)$	Beta $\text{Bet}(\theta \alpha, \beta)$	Beta $\text{Bet}(\theta \alpha + x, \beta + n - x)$
Negative Binomial $\text{NegBin}(x n, \theta)$	Beta $\text{Bet}(\theta \alpha, \beta)$	Beta $\text{Bet}(\theta \alpha + n, \beta + x)$
Multinomial $\mathbf{M}_k(x \theta_1, \dots, \theta_k)$	Dirichlet $\text{Di}_k(\theta \alpha_1, \dots, \alpha_k)$	Dirichlet $\text{Di}_k(\theta \alpha_1 + x_1, \dots, \alpha_k + x_k)$
Poisson $\text{Pn}(x \theta)$	Gamma $\text{Gam}(\theta \alpha, \beta)$	Gamma $\text{Gam}(\theta \alpha + x, \beta + 1)$

Conjugate priors

Observation law $p(x \theta)$	Prior law $p(\theta \tau)$	Posterior law $p(\theta x,\tau) \propto p(\theta \tau)p(x \theta)$
Gamma Gam ($x \nu, \theta$)	Gamma Gam ($\theta \alpha, \beta$)	Gamma Gam ($\theta \alpha + \nu, \beta + x$)
Beta Bet ($x \alpha, \theta$)	Exponential Ex ($\theta \lambda$)	Exponential Ex ($\theta \lambda - \log(1-x)$)
Normal N ($x \theta, \sigma^2$)	Normal N ($\theta \mu, \tau^2$)	Normal N $\left(\mu \mid \frac{\mu\sigma^2 + \tau^2 x}{\sigma^2 + \tau^2}, \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2} \right)$
Normal N ($x \mu, 1/\theta$)	Gamma Gam ($\theta \alpha, \beta$)	Gamma Gam $(\theta \alpha + \frac{1}{2}, \beta + \frac{1}{2}(\mu - 1))$
Normal N ($x \theta, \theta^2$)	Generalized inverse Normal INg ($\theta \alpha, \mu, \sigma$) \propto $ \theta ^{-\alpha} \exp \left[-\frac{1}{2\sigma^2} \left(\frac{1}{\theta} - \mu \right)^2 \right]$	Generalized inverse Normal INg ($\theta \alpha_n, \mu_n, \sigma_n$)

3- 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 radiography $g_\phi(r, z)$

- ▶ 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

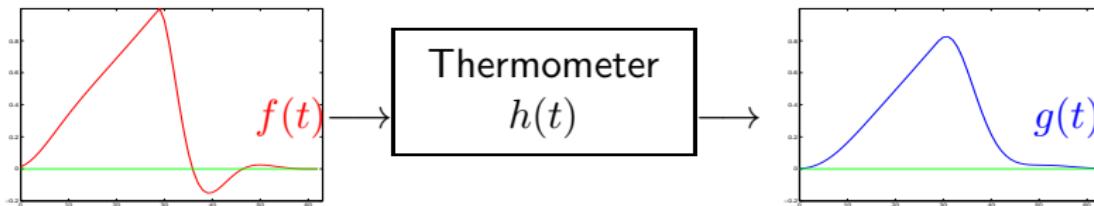
Given the forward model \mathcal{H} (impulse response $h(t)$)
and a set of data $g(t_i), i = 1, \dots, M$
find $f(t)$



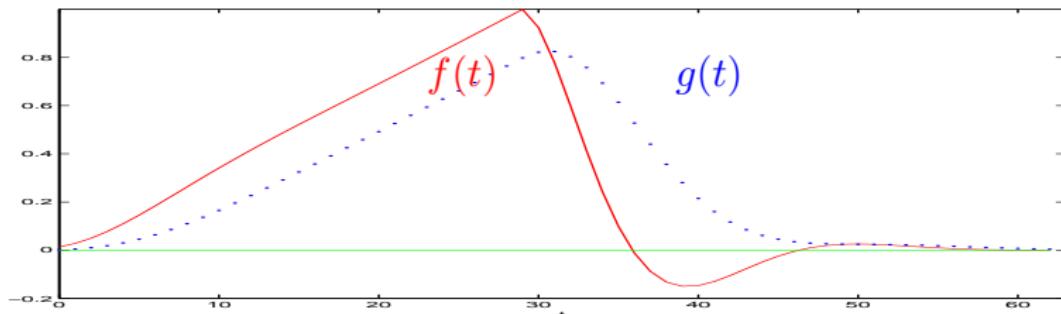
Measuring variation of temperature with a thermometer

Forward model: Convolution

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



Inversion: Deconvolution



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 $f(x, y)$

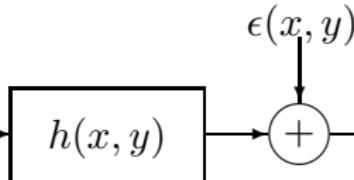
Making an image with an unfocused camera

Forward model: 2D Convolution

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



$$f(x, y)$$

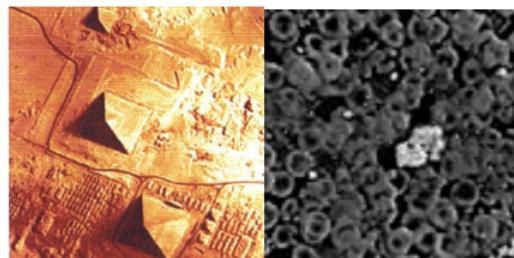


$$\epsilon(x, y)$$

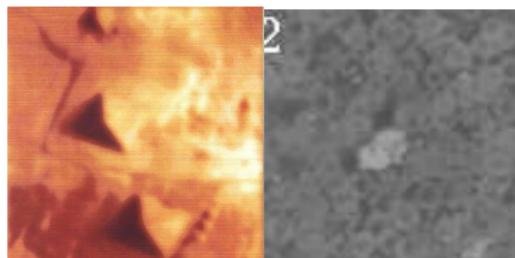
$$+$$

$$g(x, y)$$

Inversion: Image Deconvolution or Restoration



?
↔



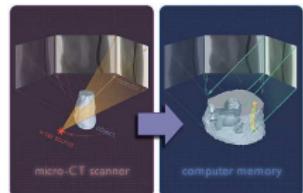
Seeing inside of a body: Computed Tomography

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



- ▶ Forward model:
Line integrals or Radon Transform

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

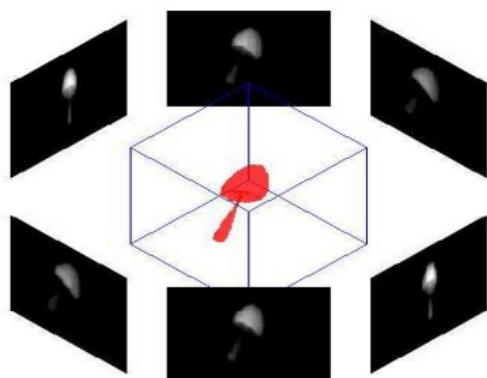


- ▶ Inverse problem: Image reconstruction

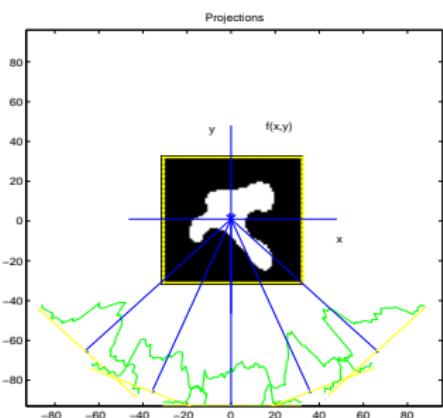
Given the forward model \mathcal{H} (Radon Transform) and
a set of data $g_{\phi_i}(r), i = 1, \dots, M$
find $f(x, y)$

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 $f(x, y, z)$

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

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

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

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

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

- \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

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

- ▶ Mismatch 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 = Ill posed problems

→ Need for prior information

Functional space (Tikhonov):

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

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

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

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

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 hyper-parameters

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 :

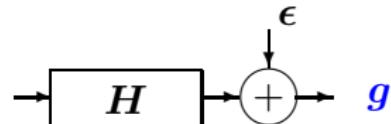
Maximum A Posteriori (MAP) :

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

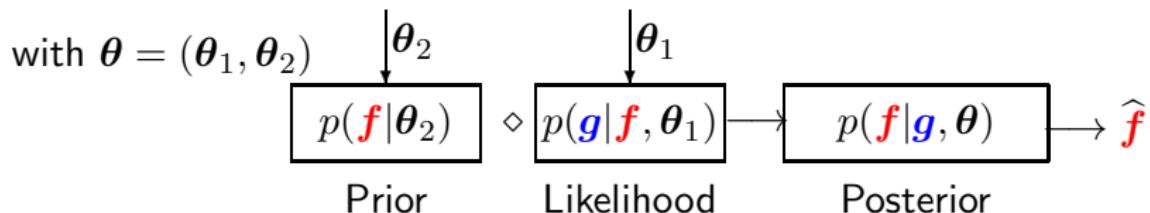
with $Q(\mathbf{g}, \mathbf{H}\mathbf{f}) = -\ln p(\mathbf{g}|\mathbf{f})$ and $\lambda\Omega(\mathbf{f}) = -\ln p(\mathbf{f})$

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 $\rightarrow p(\mathbf{f}|\boldsymbol{\theta}_2; \mathcal{M})$
- ▶ Hyper-parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \rightarrow p(\boldsymbol{\theta}|\mathcal{M})$
- ▶ 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: $(\widehat{\mathbf{f}}, \widehat{\boldsymbol{\theta}}) = \arg \max_{(\mathbf{f}, \boldsymbol{\theta})} \{p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M})\}$
- ▶ 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:
$$\begin{cases} \widehat{\mathbf{f}} &= \int \int \mathbf{f} p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M}) d\boldsymbol{\theta} d\mathbf{f} \\ \widehat{\boldsymbol{\theta}} &= \int \int \boldsymbol{\theta} p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M}) d\mathbf{f} d\boldsymbol{\theta} \end{cases}$$
- ▶ 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: $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\mathbf{g})\}$ where

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

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

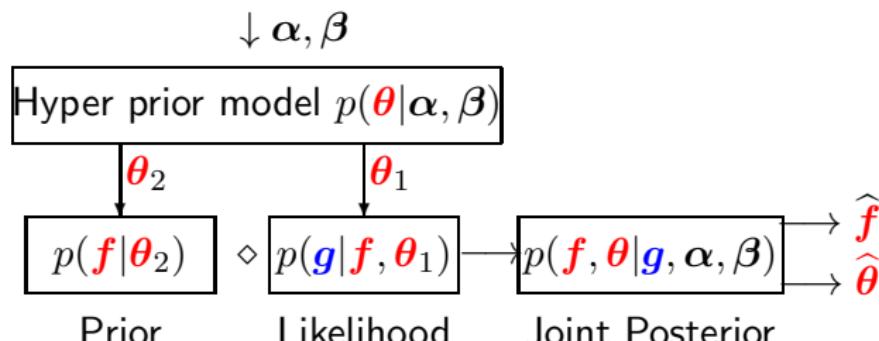
Posterior Mean: $\hat{\mathbf{f}} = \int \mathbf{f} p(\mathbf{f}|\hat{\boldsymbol{\theta}}, \mathbf{g}) \, d\mathbf{f}$

- ▶ Needs the expression of the Likelihood:

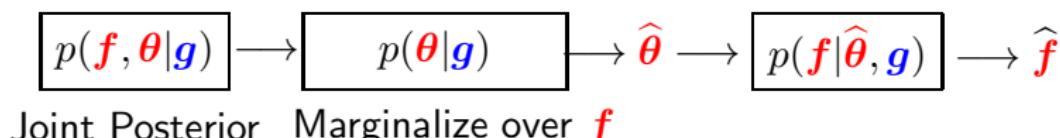
$$p(\mathbf{g}|\boldsymbol{\theta}) = \int p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}_1) p(\mathbf{f}|\boldsymbol{\theta}_2) \, d\mathbf{f}$$

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

Full Bayesian Model and Hyper-parameter Estimation



Full Bayesian Model and Hyper-parameter Estimation scheme



Marginalization for Hyper-parameter Estimation

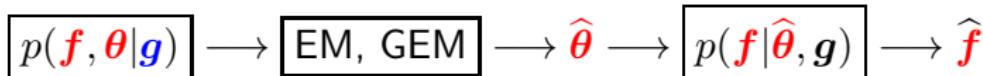
Full Bayesian: EM and GEM algorithms

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

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

- ▶ GEM (Bayesian) algorithm:

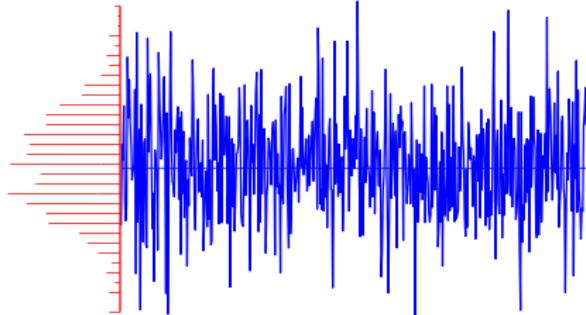
$$\begin{cases} \text{E-step: } Q(\theta, \hat{\theta}^{(k)}) = E_{p(f|g, \hat{\theta}^{(k)})} \{ \ln p(g, f|\theta) + \ln p(\theta) \} \\ \text{M-step: } \hat{\theta}^{(k)} = \arg \max_{\theta} \left\{ Q(\theta, \hat{\theta}^{(k-1)}) \right\} \end{cases}$$



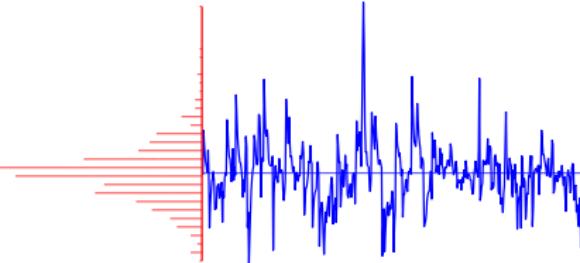
Two main steps in the Bayesian approach

- ▶ 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 Hyper-parameter estimation need **integration and optimization**
 - ▶ Approximations:
 - ▶ Gaussian approximation (Laplace)
 - ▶ Numerical exploration MCMC
 - ▶ Variational Bayes (**Separable approximation**)

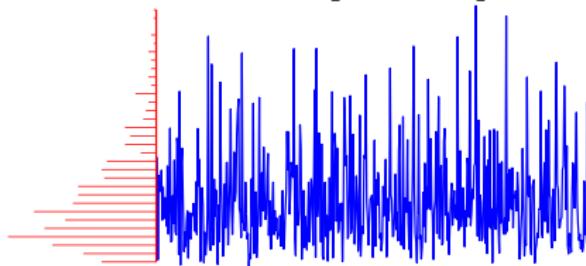
Different prior models for signals and images: Separable



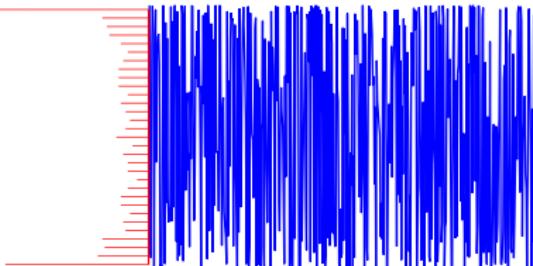
Gaussian
 $p(f_j) \propto \exp [-\alpha |f_j|^2]$



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



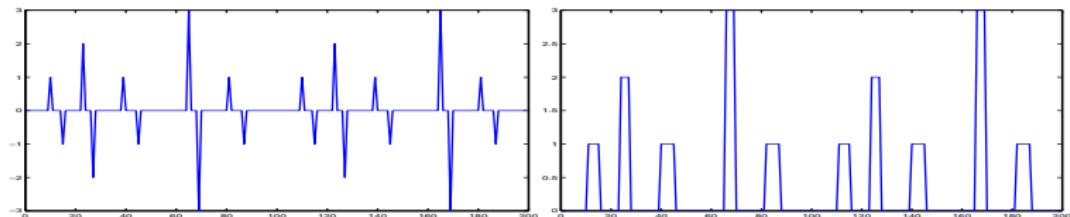
Gamma
 $p(f_j) \propto f_j^\alpha \exp [-\beta f_j]$



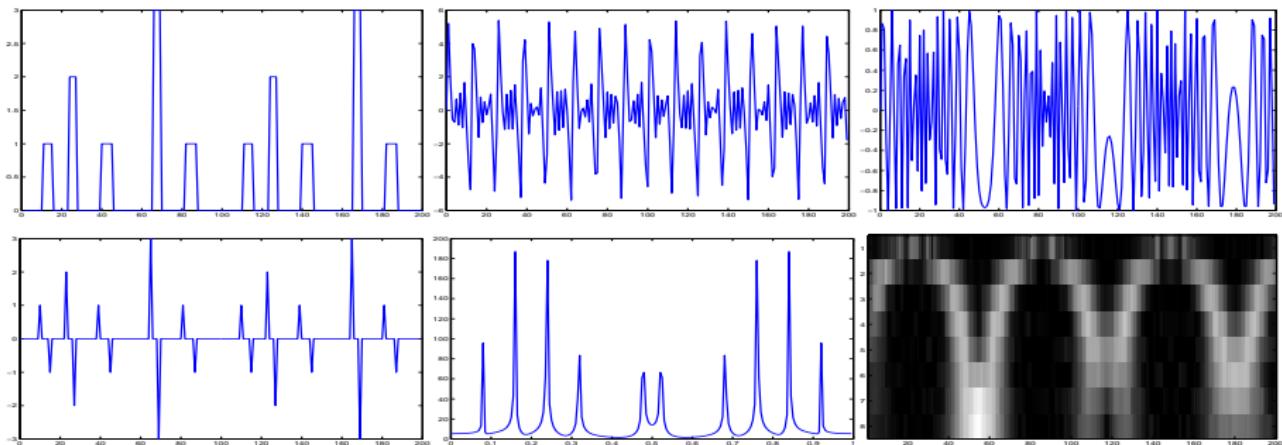
Beta
 $p(f_j) \propto f_j^\alpha (1 - f_j)^\beta$

Sparsity enforcing prior models

- Sparse signals: Direct sparsity



- Sparse signals: Sparsity in a Transform domain



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

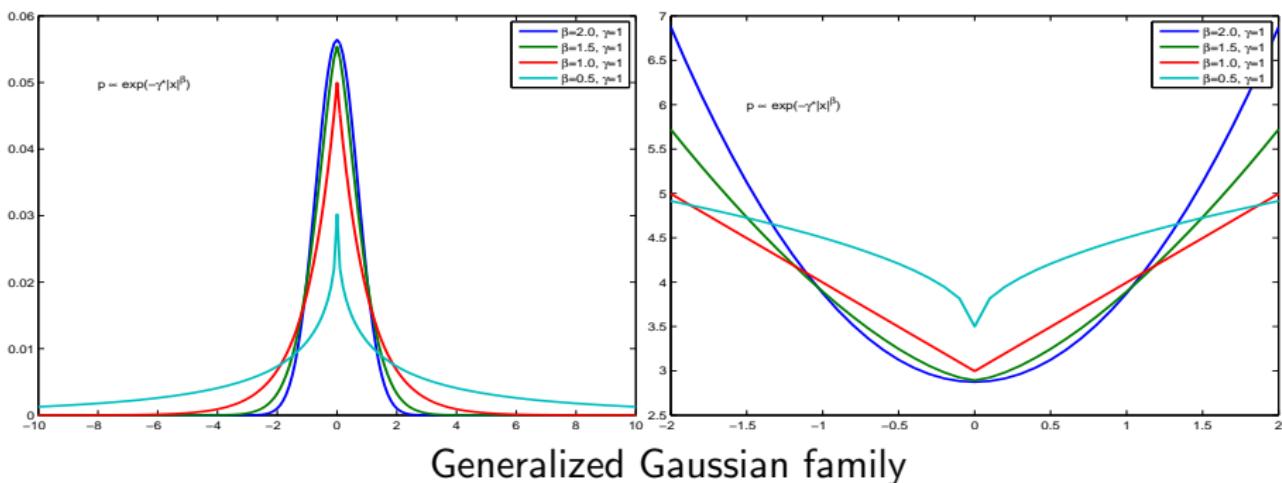
Simple heavy tailed models

- 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.



Simple heavy tailed models

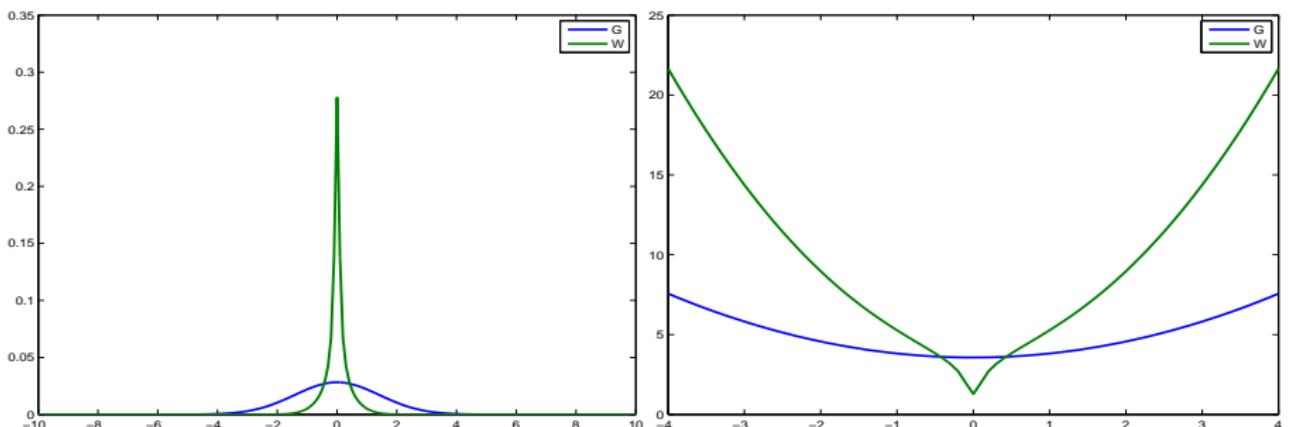
- Symmetric Weibull

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

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

$\beta = 1$ is the Double exponential and

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



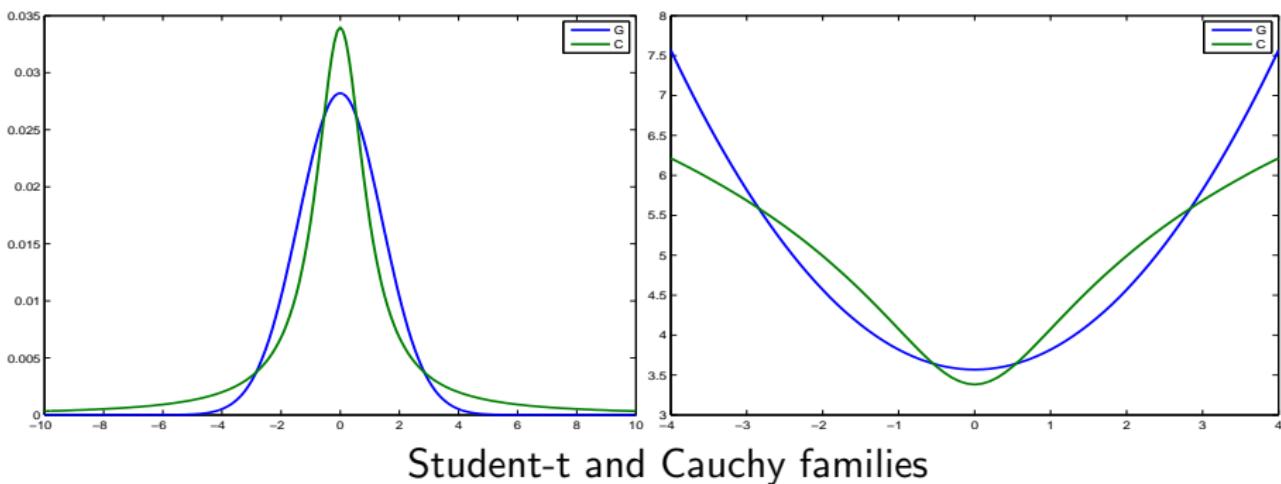
Symmetric Weibull family

Simple heavy tailed models

- Student-t and Cauchy models

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

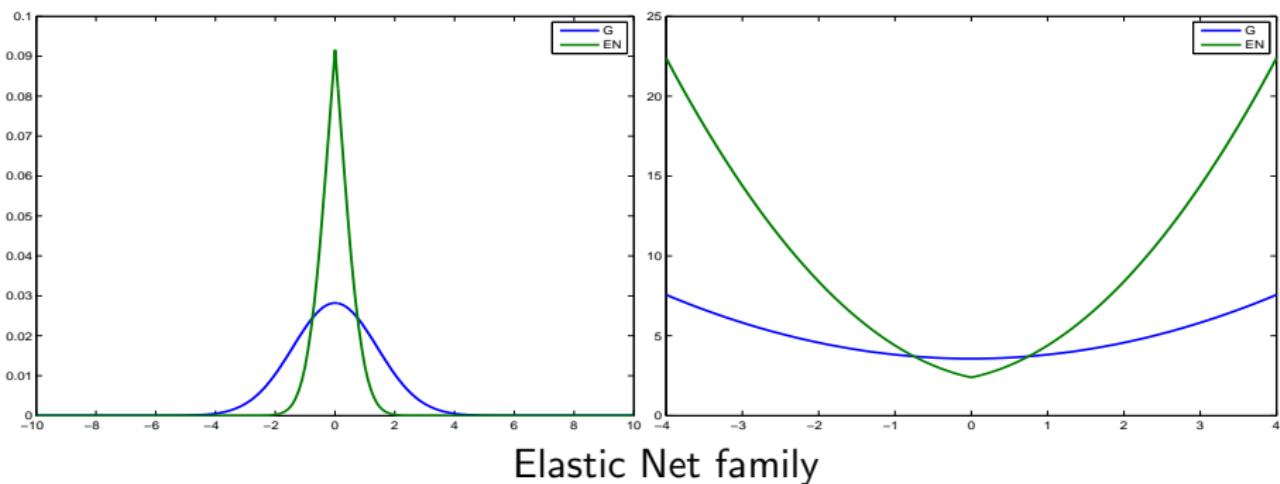
Cauchy model is obtained when $\nu = 1$.



Simple heavy tailed models

- Elastic net prior model

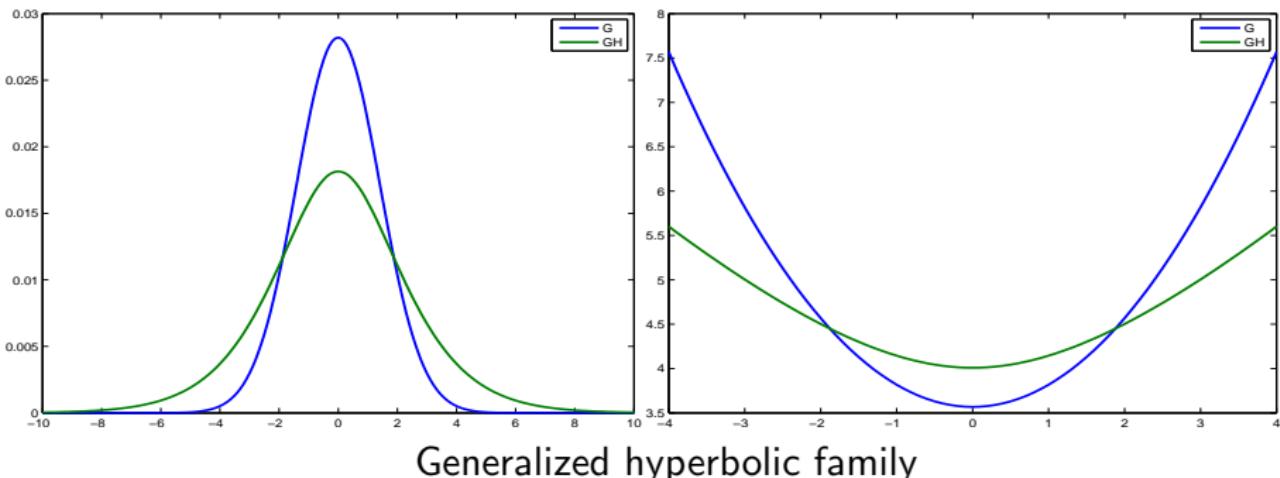
$$p(\mathbf{f}|\nu) = \prod_j \mathcal{EN}(f_j|\nu) \propto \exp \left[- \sum_j (\gamma_1 |f_j| + \gamma_2 f_j^2) \right]$$



Simple heavy tailed models

- Generalized hyperbolic (GH) models

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



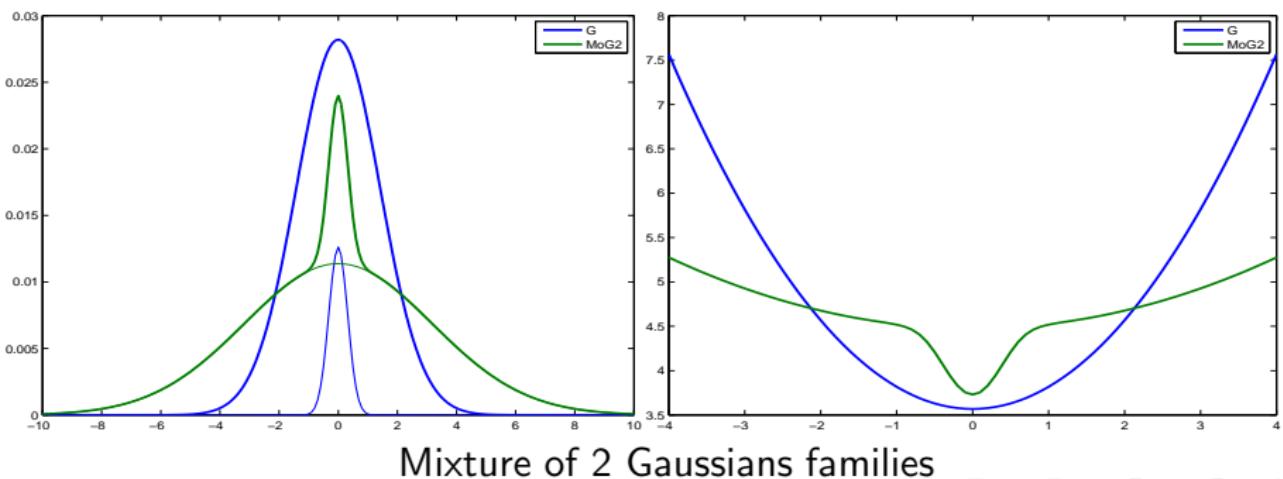
Mixture models

- Mixture of two Gaussians (MoG2) model

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

- Bernoulli-Gaussian (BG) model

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



- Mixture of Gammas

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

- 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 z** .

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

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

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

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

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

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

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

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

Hierarchical models and hidden variables

- ▶ Example 3: Laplace (Double Exponential) model

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

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

- ▶ With these models we have:

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

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:

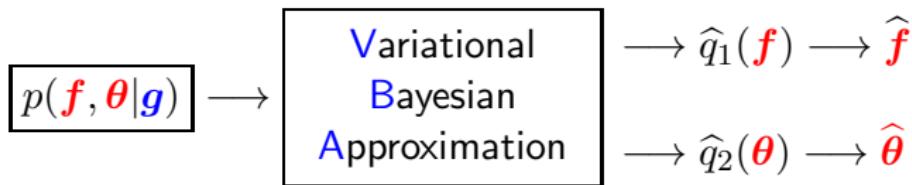
$$\hat{\mathbf{f}} \sim p(\mathbf{f} | \hat{\mathbf{z}}, \hat{\boldsymbol{\theta}}, \mathbf{g}) \longrightarrow \hat{\mathbf{z}} \sim p(\mathbf{z} | \hat{\mathbf{f}}, \hat{\boldsymbol{\theta}}, \mathbf{g}) \longrightarrow \hat{\boldsymbol{\theta}} \sim (\boldsymbol{\theta} | \hat{\mathbf{f}}, \hat{\mathbf{z}},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

- ▶ 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 optimization 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

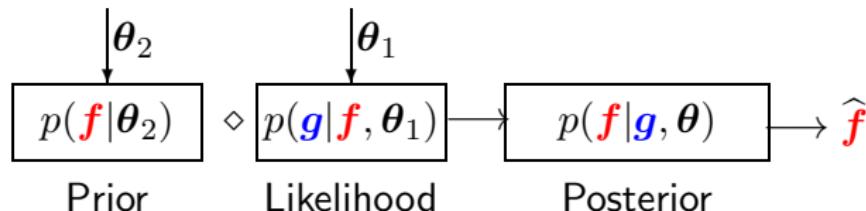
- ▶ Approximate $p(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g})$ by $q(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g}) = q_1(\mathbf{f} | \mathbf{g}) q_2(\boldsymbol{\theta} | \mathbf{g})$ and then continue computations.
- ▶ Criterion $\text{KL}(q(\mathbf{f}, \boldsymbol{\theta} | \mathbf{g}) : p(\mathbf{f},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

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

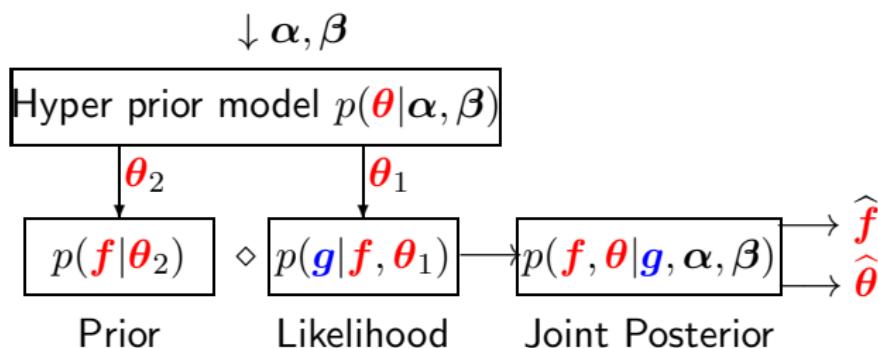


Summary of Bayesian estimation 1

- ▶ Simple Bayesian Model and Estimation

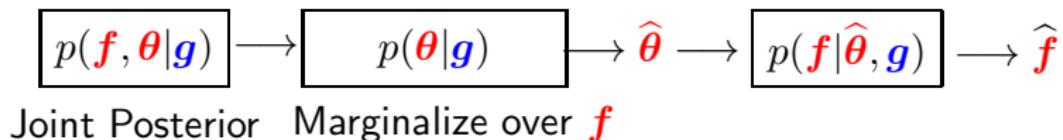


- ▶ Full Bayesian Model and Hyper-parameter Estimation

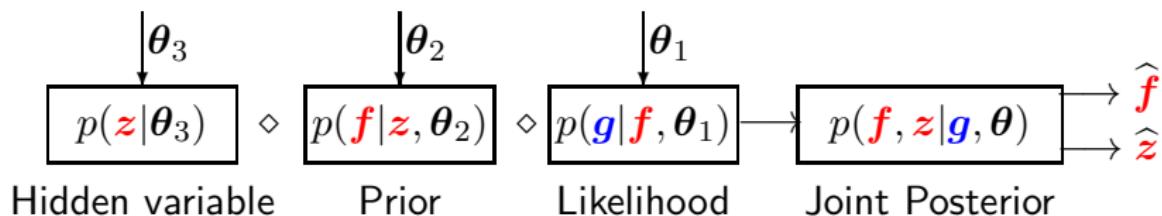


Summary of Bayesian estimation 2

- ▶ Marginalization for Hyper-parameter Estimation



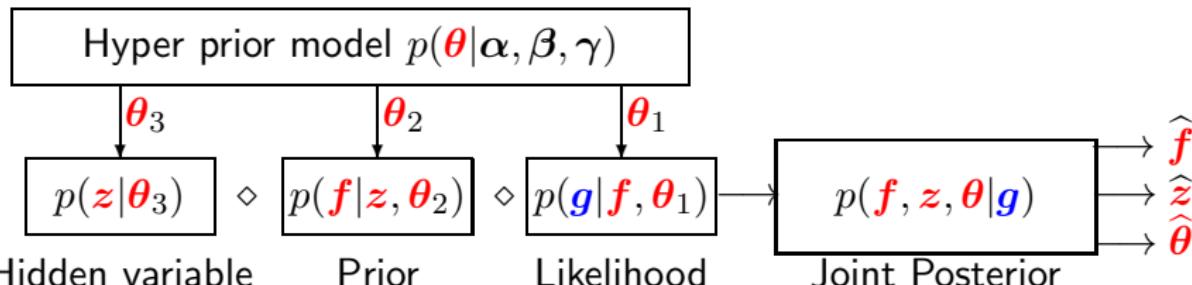
- ▶ Full Bayesian Model with a Hierarchical Prior Model



Summary of Bayesian estimation 3

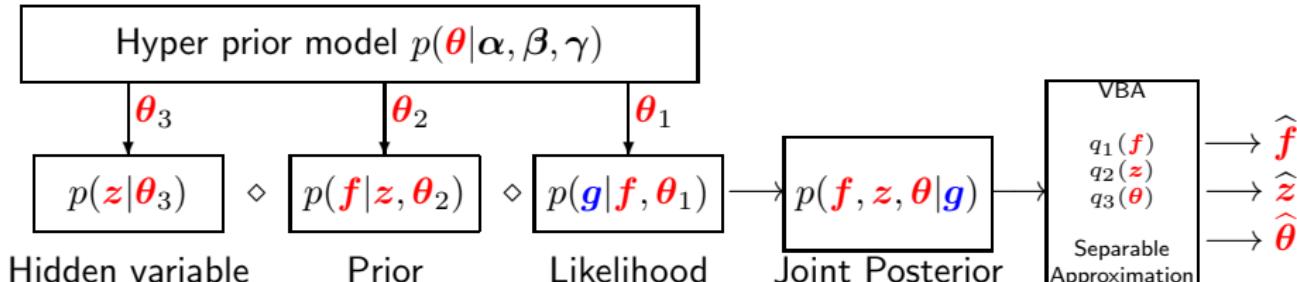
- Full Bayesian Hierarchical Model with Hyper-parameter Estimation

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

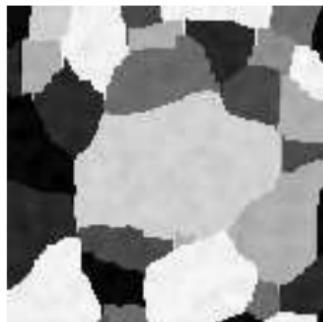
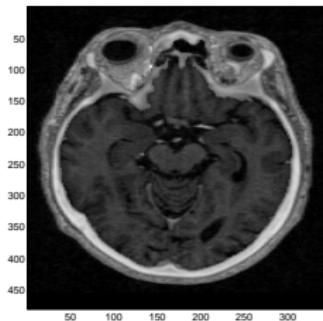


- Full Bayesian Hierarchical Model and Variational Approximation

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



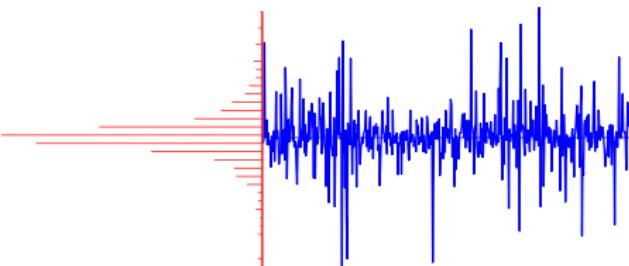
Which images I am looking for?



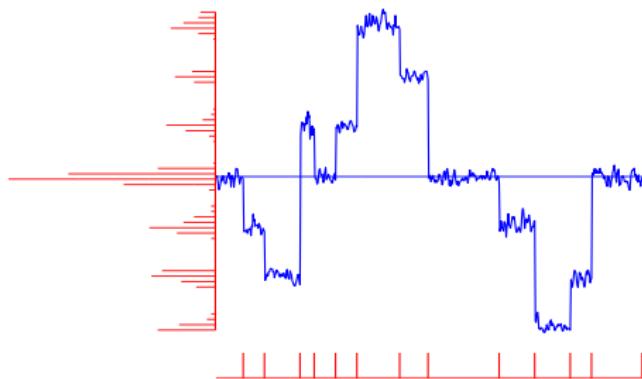
Which image I am looking for?



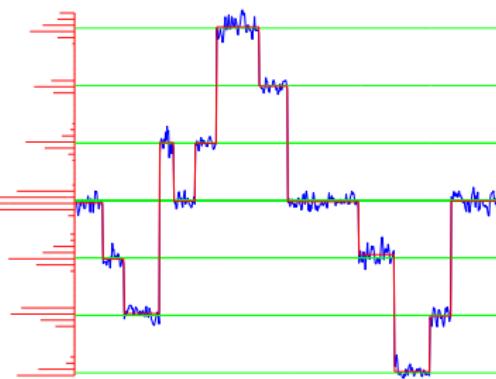
Gauss-Markov



Generalized GM

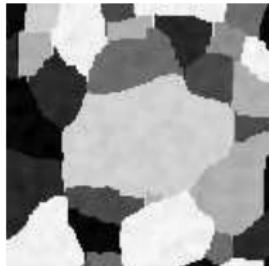
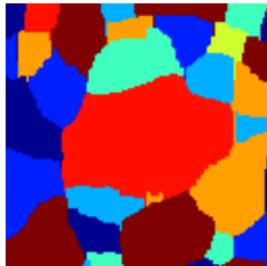


Piecewise Gaussian



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 :

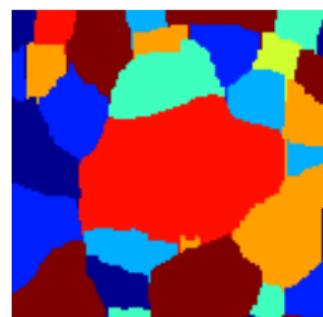
Mixture of Gaussians



$f(\mathbf{r})$

- ▶ $f|z$ Gauss-Markov, z iid :

Mixture of Gauss-Markov

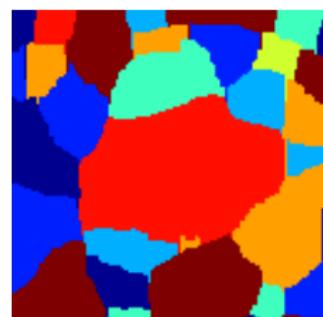


$z(\mathbf{r})$

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

Mixture of Independent Gaussians

(MIG with Hidden Potts)

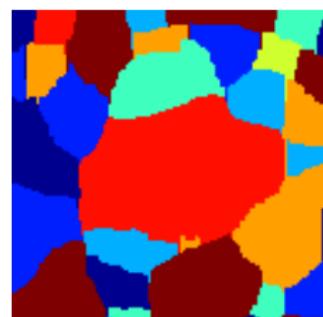


$z(\mathbf{r})$

- ▶ $f|z$ Markov, z Potts-Markov :

Mixture of Gauss-Markov

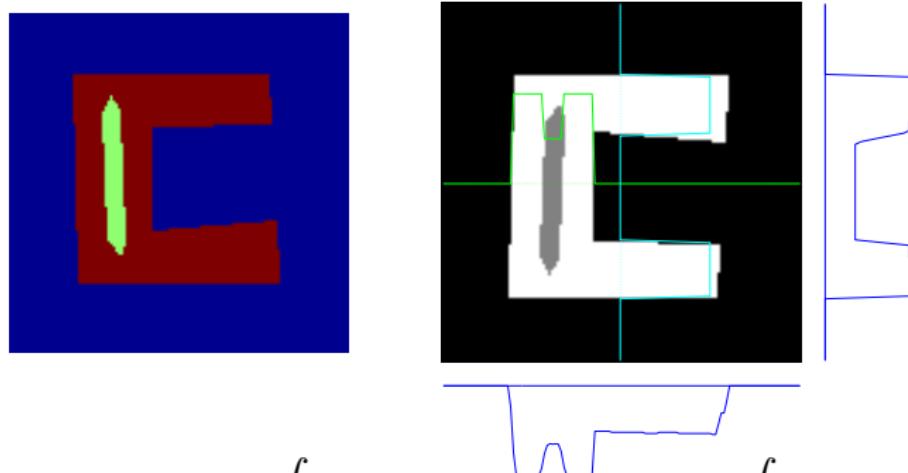
(MGM with hidden Potts)



$z(\mathbf{r})$

Application of CT in NDT

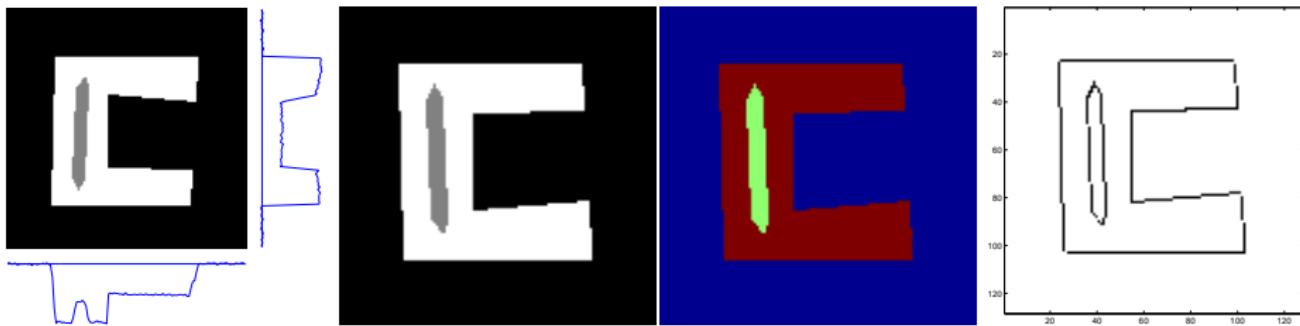
Reconstruction from only 2 projections



- 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



$\mathbf{g} \mathbf{f}$ $\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$ $\mathbf{g} \mathbf{f} \sim \mathcal{N}(\mathbf{H}\mathbf{f}, \sigma_\epsilon^2 \mathbf{I})$ Gaussian	$\mathbf{f} \mathbf{z}$ 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

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

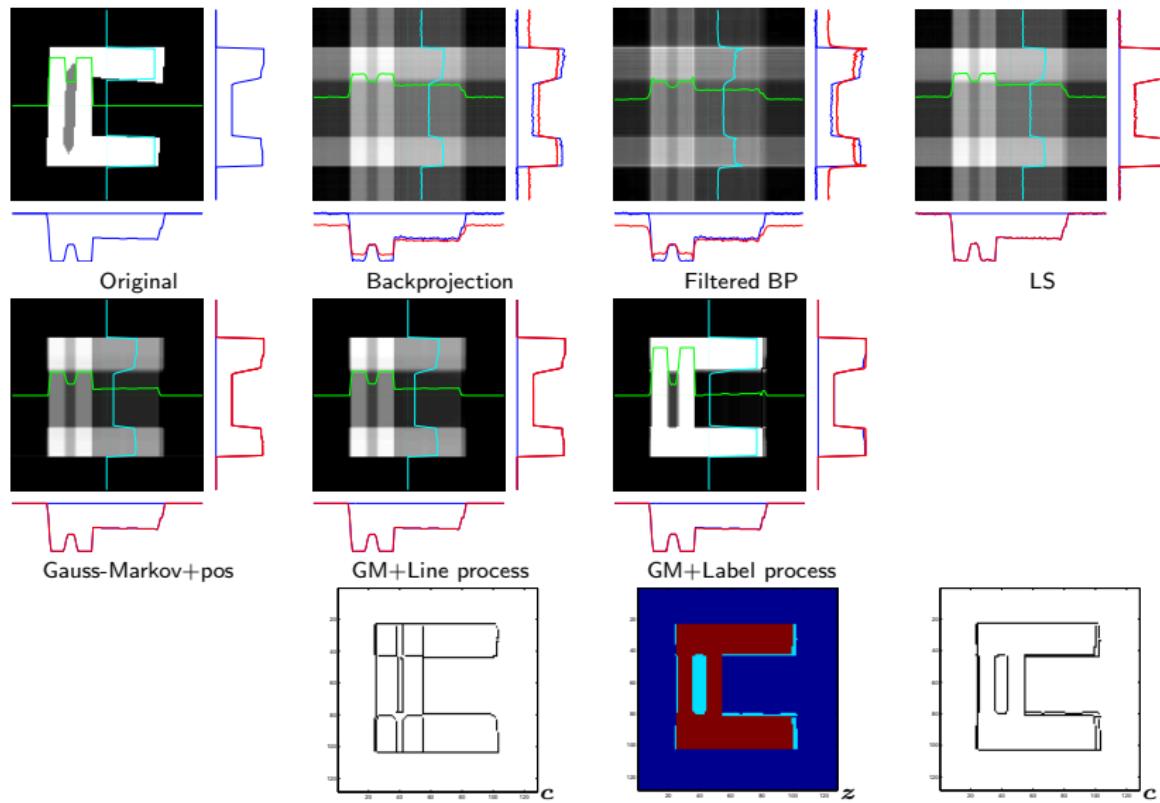
General scheme:

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

Iterative algorithm:

- ▶ 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 optimization 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

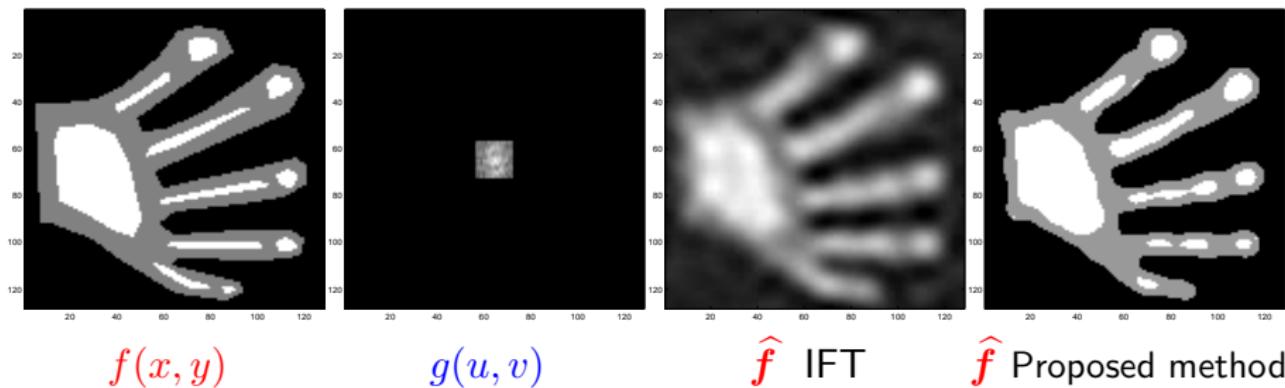


Application in Microwave imaging

$$g(\omega) = \int f(r) \exp[-j(\omega \cdot r)] dr + \epsilon(\omega)$$

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

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



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)