Sensors, Measurement systems 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

Contents

- Sensors
- Measurement systems
- Basic sensors designs and their mathematical models
- Signal and image processing of the sensors output
- Indirect measurement and inverse problems
- Regularization and Bayesian inversion
- Case studies:
  - Deconvolution
  - X ray Computed Tomography
  - Eddy current NDT

Basic sensors designs and their mathematical models

- 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
  - ...

Basic sensors designs and their mathematical models

- 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

Basic sensors designs and their mathematical models

- 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 (variance):
  Closeness of the sensor output to indicating the actual value of the measured variable.
- Precision (bias):
  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/`

Physical principles of sensors

- We can easily measures 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
  - U Electric charge flux (weber)
- Dipole model:

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

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

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

- 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

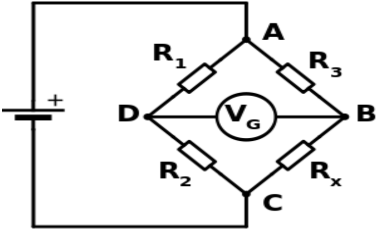
$$U(\omega) = j\omega L \, I(\omega) \longrightarrow Z(\omega) = j\omega \, L$$

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:

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

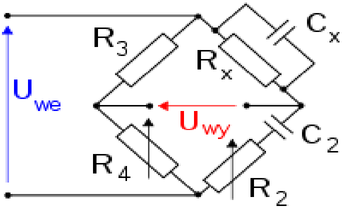
$$V_G = \left(\frac{R_x}{R_3 + R_x} - \frac{R_2}{R_1 + R_2} \right) V_s$$

- See Demo here:

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

Measuring R

- The Wien bridge:



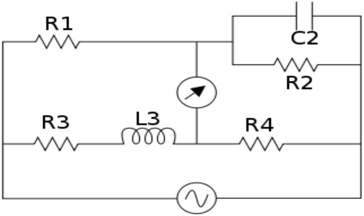
At some frequency, the reactance of the series R2C2 arm will be an exact multiple of the shunt RxCx arm. If the two R3 and R4 arms are adjusted to the same ratio, then the bridge is balanced.

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

The equations simplify if one chooses R2 = Rx and C2 = Cx; the result is R4 = 2 R3.

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.

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

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.

Resonant circuits

- The resonant pulsation is:

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

which gives:

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

2- Signal processing of sensors output

- Full-scale error: Calibration
- Offset error: Offset elimination
- Drift: changes with temperature
- Non-linearity

- Dealing with noise $\longrightarrow$ Filtering
  - Analog filtering
  - Digital filtering
    - Fixed averaging
    - Moving Average (MA) filtering
    - Autoregressive (AR) filtering
    - Moving Average Autoregressive (ARMA) filtering

Dealing with noise, errors and uncertainties

- Errors, noise and uncertainties $\longrightarrow$ Probability theory
- Background on Probability theory:
  - Discrete variables $\{x_1, \cdots, x_n\}$
    Probability distribution: $\{p_1, \cdots, p_n\}$ 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 $\int_{-\infty}^{+\infty} p(x) \, \mathrm{d}x = 1$,
    Partition function: $F(x) = P(X \leq x) = \int_{\infty}^{x} p(x) \, \mathrm{d}x$
  - Expected value: $\mathsf{E}\{X\} = \int x \, p(x) \, \mathrm{d}x$
  - Variance value: $\mathsf{Var}\{X\} = \int (x - \mathsf{E}\{X\})^2 \, p(x) \, \mathrm{d}x$
  - Mode value Mode $= \arg\max_x \{p(x)\}$

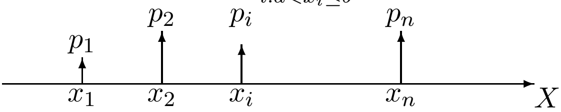
  - Normal distribution $\mathcal{N}(x|m, v)$
  - Gamma distribution $\mathcal{G}(x|\alpha, \beta)$

Discrete events

- X takes values x_i with probabilities $p_i,\ i = 1, \cdots, n$.
- $P(X = x_i) = p_i,\ i = 1, \cdots, n$ is probability distribution (pd).
- If we sort x_i in such a way that $x1 \leq x_2 \leq \cdots \leq x_n$, then we can define the "probability cumulative distribution (pcd)":

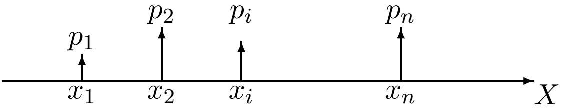
$$F(x) = P(X \leq x) = \sum_{i:x_i \leq x} P(X = x_i)$$

$$P(a < X \leq b) = \sum_{i:a < x_i \leq b} P(X = x_i)$$

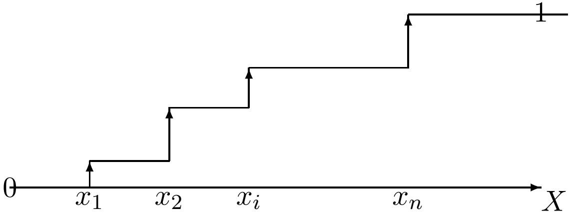


Discrete events

- Probabilities Distribution



- Cumulative Probablity Distribution



Discrete events

- Expected value

$$\mathsf{E}\{X\} = < X > = \sum_i p_i \, x_i$$

- Variance

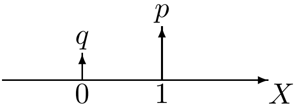
$$\mathsf{Var}\{X\} = \sum_i p_i \, (x_i - \mathsf{E}\{X\})^2 = \sum_i p_i \, (x_i - < X >)^2$$

- Entropy

$$\mathsf{H}(X) = -\sum_i p_i \, \ln p_i$$

Discrete variables probability distributions

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



- 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 $\mathcal{B}in(.|n, p)$:
  The probability of k successes in n trials:

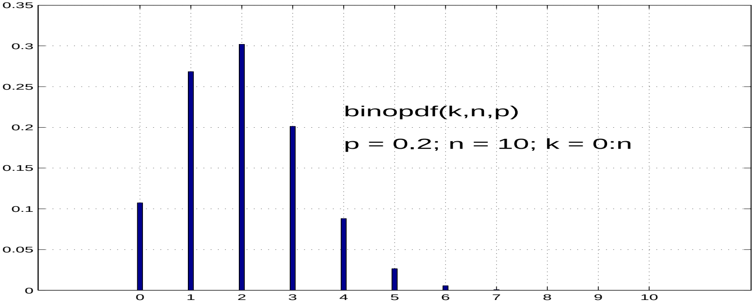
$$P(X = k) = \binom{n}{k} p^k \, (1 - p)^{n-k}$$

Binomial distribution $\mathcal{B}in(.|n,p)$

The probability of k successes in n trials:

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

$$\mathsf{E}\{X\} = n\,p, \quad \mathsf{Var}\{X\} = n\,p\,q = n\,p\,(1-p)$$



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

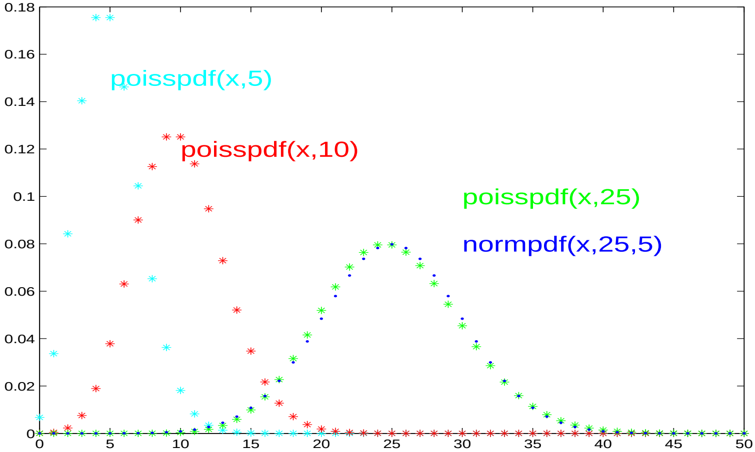
$$X \sim \mathcal{B}in(n,p) \lim_{n \mapsto \infty, np \mapsto \lambda} X \sim \mathcal{P}(\lambda)$$

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

$$\mathsf{E}\{X\} = \lambda, \quad \mathsf{Var}\{X\} = \lambda$$

- If $X_n \sim \mathcal{B}in(n, \lambda/n)$ and $Y \sim \mathcal{P}(\lambda)$ then for each fixed k, $\lim_{n\to\infty} P(X_n = k) = P(Y = k)$.

Poisson distribution



Continuous case

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

$$\begin{aligned} P(a \leq X < b) &= F(b) - F(a) \\ P(x \leq X < x + dx) &= F(x + dx) - F(x) = \mathrm{d}F(x) \end{aligned}$$

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

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

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

$$P(a < X \leq b) = \int_a^b p(x) \, \mathrm{d}x$$

- Cumulative distribution function (cdf)

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

Continuous case

- Expected value

$$\mathsf{E}\{X\} = \int x\, p(x)\, \mathrm{d}x =< X >$$

- Variance

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

- Entropy

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

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

$$\int_{-\infty}^{\mathsf{Med}(X)} p(x)\, \mathrm{d}x = \int_{\mathsf{Med}(X)}^{+\infty} p(x)\, \mathrm{d}x$$

Uniform and Beta distributions

- Uniform:

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

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

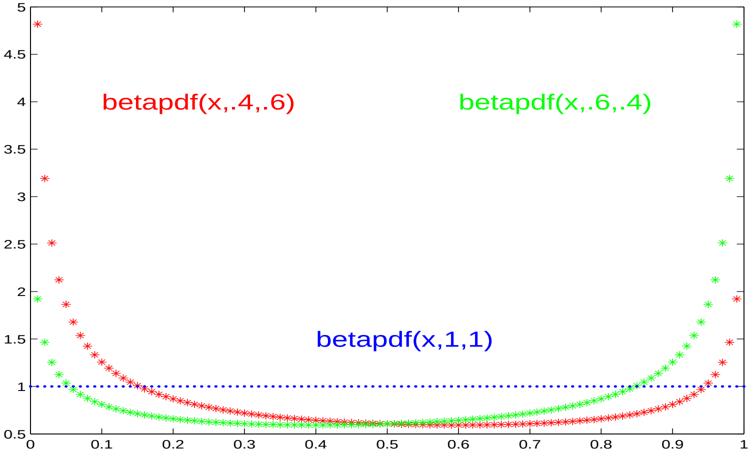
- Beta:

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

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

- $\mathsf{Beta}(.|1,1) = \mathsf{U}(.|0,1)$

Uniform and Beta distributions



Gaussian distributions

Different notations:

- classical one with mean and variance:

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

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

- mean and precision parameters:

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

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

Generalized Gaussian distributions

- Gaussian:

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

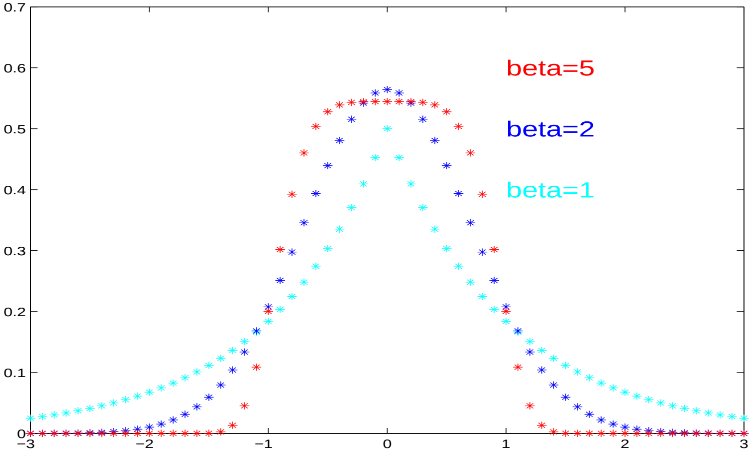
- Generalized Gaussian:

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

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

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

Gaussian and Generalized Gaussian distributions



Gamma distributions

- Forme 1:

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

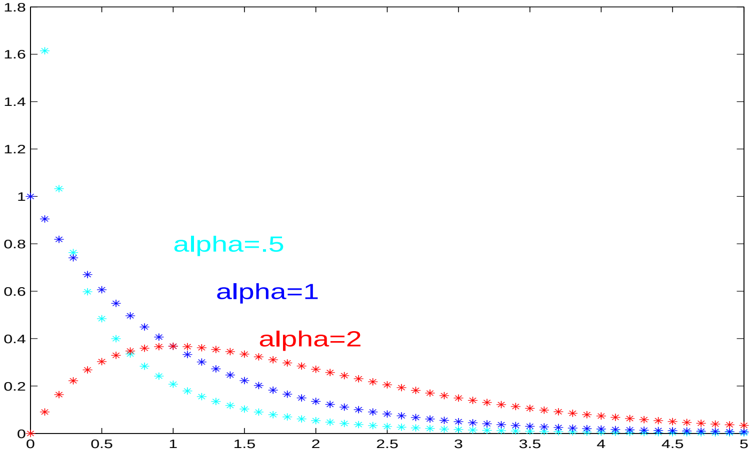
$$\mathsf{E}\{X\} = \frac{\alpha}{\beta}, \quad \mathsf{Var}\{X\} = \frac{\alpha}{\beta^2}, \quad \mathsf{Mod}(X) = \frac{\alpha-1}{\alpha+\beta-2}$$

- Forme 2: $\theta = 1/\beta$

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

- $\alpha = 1$: 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:

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

  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

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

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

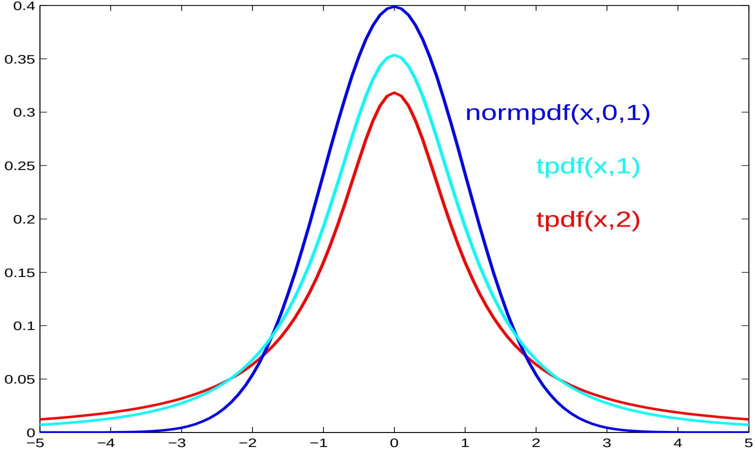
- Interesting relation between Student-t, Normal and Gamma distributions:

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

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

Student and Cauchy

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



Vector variables

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

$$\mathsf{E}\{\boldsymbol{X}\} = \int \boldsymbol{x}\, p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} =< \boldsymbol{X} >$$

- Covariance

$$\begin{aligned} \mathrm{cov}[\boldsymbol{X}] &= \int (\boldsymbol{X} - \mathsf{E}\{\boldsymbol{X}\})(\boldsymbol{X} - \mathsf{E}\{\boldsymbol{X}\})'\, p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\ &= \left\langle (\boldsymbol{X} - \mathsf{E}\{\boldsymbol{X}\})(\boldsymbol{X} - \mathsf{E}\{\boldsymbol{X}\})' \right\rangle \end{aligned}$$

- Entropy

$$\mathsf{E}(\boldsymbol{X}) = \int -\ln p(\boldsymbol{x})\, p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \langle \ln p(\boldsymbol{X}) \rangle$$

- Mode: $\mathrm{Mode}(p(\boldsymbol{x})) = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x})\}$

Vector variables

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

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

- 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): $\boldsymbol{X} \sim \mathcal{N}(.|\boldsymbol{\mu}, \boldsymbol{\sigma})$

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

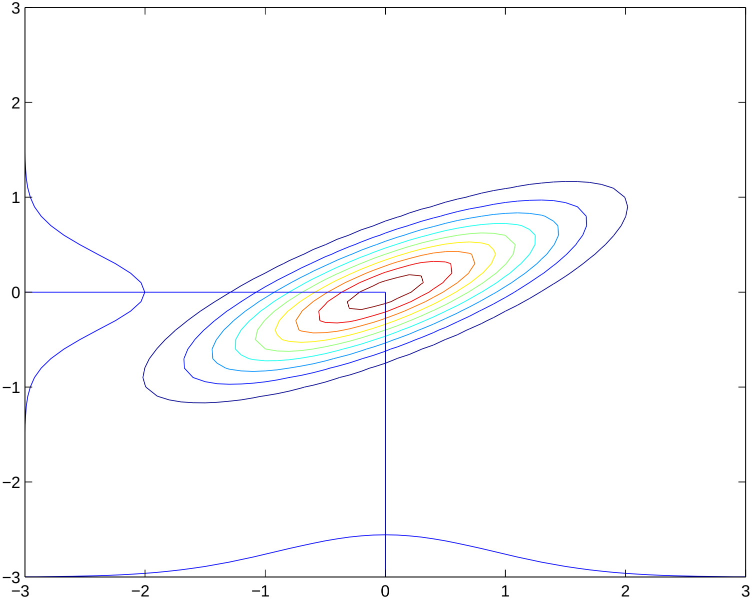
$$\mathsf{E}\left\{\boldsymbol{X}\right\} = \boldsymbol{\mu}, \quad \mathsf{cov}[\boldsymbol{X}] = \boldsymbol{\Sigma}$$

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

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

$$\mathsf{E}\left\{\boldsymbol{X}\right\} = \boldsymbol{\mu}, \quad \mathsf{cov}[\boldsymbol{X}] = \boldsymbol{\Lambda}^{-1}$$

Multivariate normal distributions



Multivariate Student-t

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

- $p = 1$

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

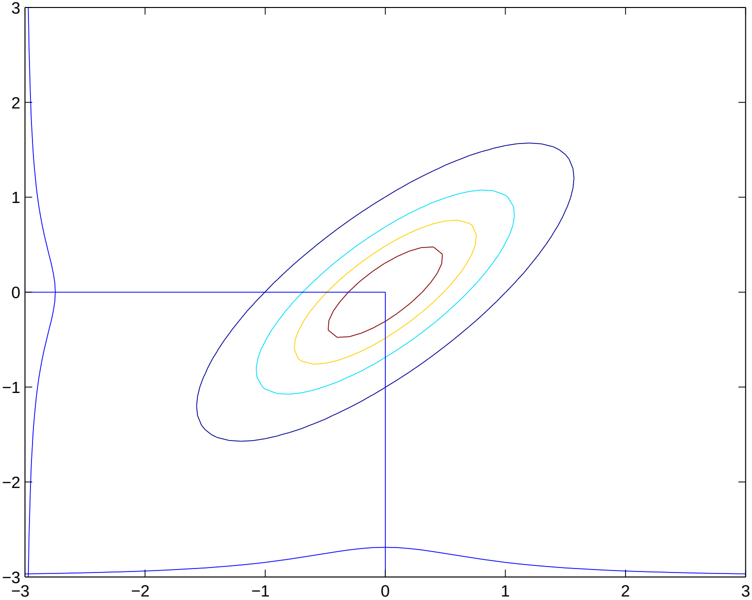
- $p = 2,\ \boldsymbol{\Sigma}^{-1} = \boldsymbol{A}$

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

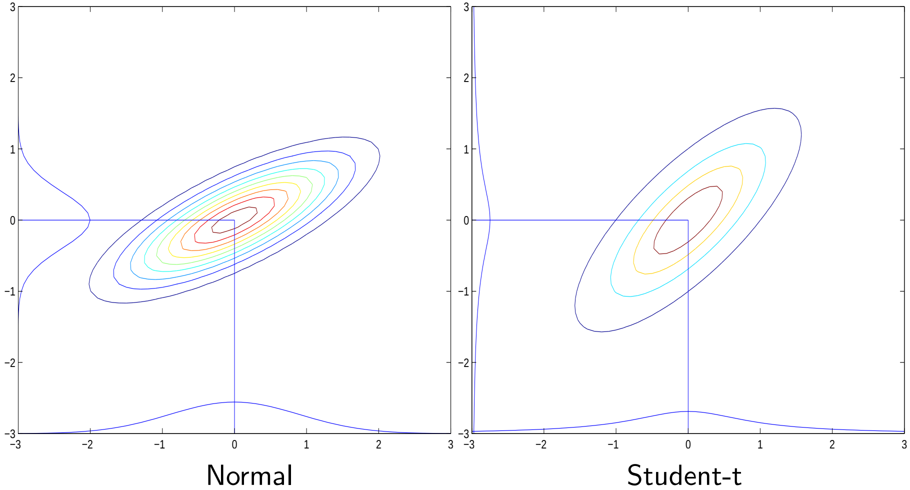
- $p = 2,\ \boldsymbol{\Sigma} = \boldsymbol{A} = \boldsymbol{I}$

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

Multivariate Student-t distributions



Multivariate normal distributions



Parameter estimation

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

- Moments method:

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

- Maximum Likelihood

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

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

- Bayesian approach

Bayesian Parameter estimation

- Likelihood

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

- A priori

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

- A posteriori

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

- Infer on $\boldsymbol{\theta}$ using $p(\boldsymbol{\theta}|\boldsymbol{x})$.
  For example:
  - Maximum A Posteriori (MAP)

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

  - Posterior Mean

$$\widehat{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta}$$

Parameter estimation: Normal distribution

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

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

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

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

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

Parameter estimation: Normal distribution: σ known

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

$$p(\mu|\boldsymbol{x}) = \frac{p(\mu)}{p(\boldsymbol{x})} \frac{1}{\left(2\pi\sigma_0^2\right)^{N/2}} \exp\left[-\sum_{i=1}^{N} \frac{(x_i-\mu)^2}{2\sigma_0^2}\right]$$

$$= \frac{p(\mu)}{p(\boldsymbol{x})} \frac{1}{\left(2\pi\sigma_0^2\right)^{N/2}} \exp\left[-\frac{(\mu-\bar{x})^2 + s^2}{2\sigma_0^2/N}\right]$$

$$\propto p(\mu) \exp\left[-\frac{(\mu-\bar{x})^2}{2\sigma_0^2/N}\right]$$

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

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

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

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

Parameter estimation

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

- Moments method:

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

- Maximum Likelihood

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

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

- Bayesian approach

Bayesian Parameter estimation

- Likelihood

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

- A priori

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

- A posteriori

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

- Infer on $\boldsymbol{\theta}$ using $p(\boldsymbol{\theta}|\boldsymbol{x})$.
  For example:
  - Maximum A Posteriori (MAP)

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

  - Posterior Mean

$$\widehat{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta}$$

Parameter estimation: Normal distribution

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

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

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

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

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

Parameter estimation: Normal distribution: σ known

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

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

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

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

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

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

Conjugate priors

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

Conjugate priors

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

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

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

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

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

Exercise 1

Let note
$$\bar{x}_N = \frac{1}{N}\sum_{n=1}^{N} x(n), \qquad v_N = \frac{1}{N}\sum_{n=1}^{N}(x(n)-\bar{x}_N)^2$$
$$\bar{x}_{N-1} = \frac{1}{N-1}\sum_{n=1}^{N-1} x(n), \qquad v_{N-1} = \frac{1}{N-1}\sum_{n=1}^{N-1}(x(n)-\bar{x}_N)^2$$

Show that

- Updating mean and variance:

$$\bar{x}_N = \frac{N-1}{N}\bar{x}_{N-1} + \frac{1}{N}x(n) = \bar{x}_{N-1} + \frac{1}{N}(x(n)-\bar{x}_{N-1})$$
$$v_N = \frac{N-1}{N}v_{N-1} + \frac{N-1}{N^2}(x(n)-\bar{x}_N)^2$$

- Updating inverse of the variance:

$$v_N^{-1} = \frac{N}{N-1}v_{N-1}^{-1} + \frac{N}{(N-1)(N+\rho_N)}(x(n)-\bar{x}_N)^2 v_{N-1}^{-2}$$
with $\rho_N = (x(n)-\bar{x}_N)^2 v_{N-1}^{-1}$

- Vectorial data $\boldsymbol{x}_n$

$$\bar{\boldsymbol{x}}_N = \frac{N-1}{N}\bar{\boldsymbol{x}}_{N-1} + \frac{1}{N}\boldsymbol{x}(n) = \bar{\boldsymbol{x}}_{N-1} + \frac{1}{N}(\boldsymbol{x}(n)-\bar{\boldsymbol{x}}_{N-1})$$
$$\boldsymbol{V}_N = \frac{N-1}{N}\boldsymbol{V}_{N-1} + \frac{N-1}{N^2}(\boldsymbol{x}(n)-\bar{\boldsymbol{x}}_N)(\boldsymbol{x}(n)-\bar{\boldsymbol{x}}_N)'$$
$$\boldsymbol{V}_N^{-1} = \frac{N}{N-1}\boldsymbol{V}_{N-1}^{-1} + \frac{N}{(N-1)(N+\rho_N)}\boldsymbol{V}_{N-1}^{-1}(x(n)-\bar{x}_N)(x(n)-\bar{x}_N)'\boldsymbol{V}_{N-1}^{-1}$$
with $\rho_N = (\boldsymbol{x}(n)-\bar{\boldsymbol{x}}_N)'\boldsymbol{V}_{N-1}^{-1}(\boldsymbol{x}(n)-\bar{\boldsymbol{x}}_N)'$

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:

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

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

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:

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

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

Moving Average (MA)

$$f(t) \longrightarrow \boxed{\text{Filter}} \longrightarrow g(t)$$

- Convolution
  - Continuous

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

  - 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

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

- Discrete

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

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

- Filter transfer function

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

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

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

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

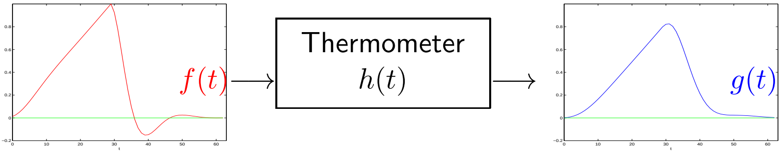
- Inverse problem: Deconvolution

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

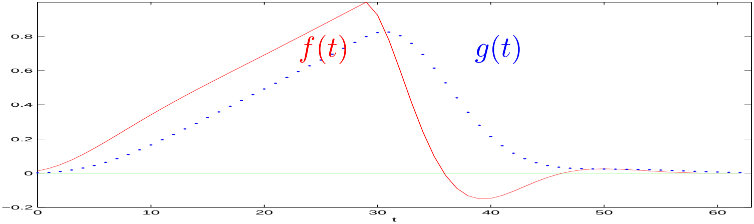
Measuring variation of temperature with a thermometer

Forward model: Convolution

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

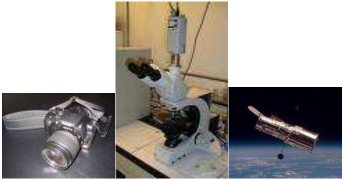


Inversion: Deconvolution



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

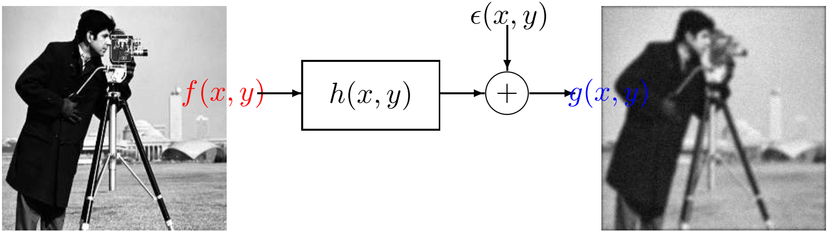
- Inverse problem: Image restoration

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 $f(x, y)$

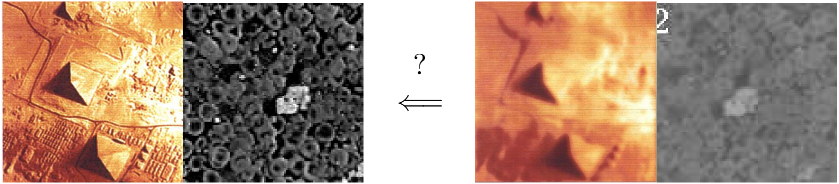
Making an image with an unfocused camera

Forward model: 2D Convolution

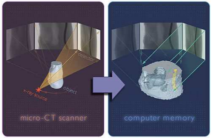
$$g(x,y) = \iint f(x',y')\, h(x-x',y-y')\,\mathrm{d}x'\,\mathrm{d}y' + \epsilon(x,y)$$



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

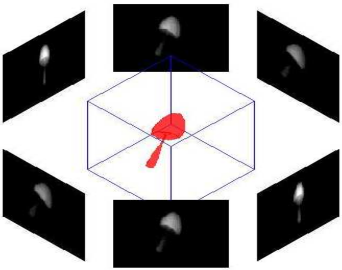
$$
\begin{aligned}
g_\phi(r) &= \int_{L_{r,\phi}} f(x, y) \,\mathrm{d}l + \epsilon_\phi(r) \\
&= \iint f(x, y)\, \delta(r - x\cos\phi - y\sin\phi) \,\mathrm{d}x\, \mathrm{d}y + \epsilon_\phi(r)
\end{aligned}
$$

- Inverse problem: Image reconstruction

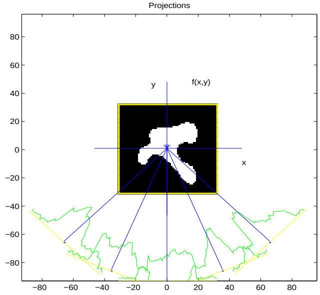
  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x, y)$

2D and 3D Computed Tomography

3D



2D



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

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

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

Inverse problems: Discretization

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

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

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

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

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

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

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

Inverse problems: Deterministic methods

Data matching

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

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

- Examples:

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

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

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

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

Inverse problems: Regularization theory

Inverse problems = Ill posed problems

$\longrightarrow$ Need for prior information

Functional space (Tikhonov):

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

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

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

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

or

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

Limitations:

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

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 :

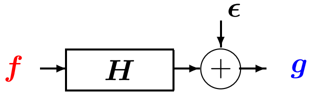
Maximum A Posteriori (MAP) :

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

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

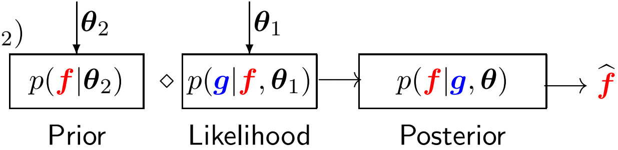
- Linear Inverse problems: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$



- 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): $\widehat{\boldsymbol{f}} = \mathsf{E}_{p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})}\{\boldsymbol{f}\} = \int \boldsymbol{f}\; p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{f}$

Bayesian Estimation: Two simple priors

- Example 1: Linear Gaussian case:

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

with

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

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

- Example 2: Double Exponential prior & MAP:

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

Full Bayesian approach

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

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

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

Full Bayesian: Marginal MAP and PM estimates

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

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

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

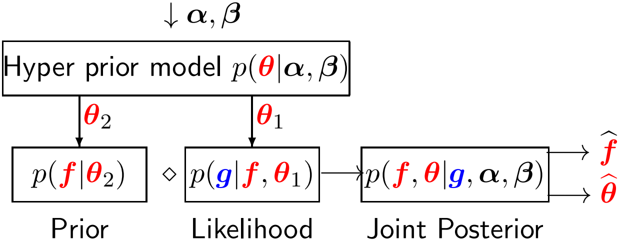
  Posterior Mean: $\widehat{\boldsymbol{f}} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \,\mathrm{d}\boldsymbol{f}$

- Needs the expression of the Likelihood:

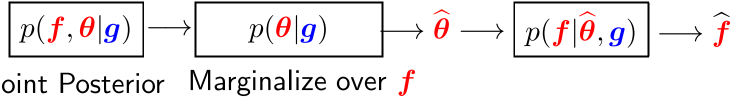
$$p(\boldsymbol{g}|\boldsymbol{\theta}) = \int p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{\theta}_2) \,\mathrm{d}\boldsymbol{f}$$

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

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

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

- GEM (Bayesian) algorithm:

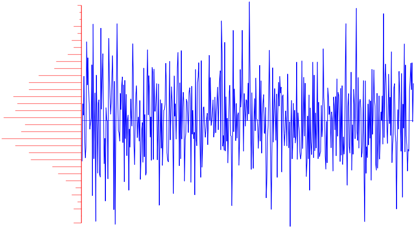
$$\left\{\begin{array}{ll} \text{E-step:} & Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(k)}) = \mathsf{E}_{p(\boldsymbol{f}|\boldsymbol{g},\widehat{\boldsymbol{\theta}}^{(k)})}\{\ln p(\boldsymbol{g}, \boldsymbol{f}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta})\} \\ \text{M-step:} & \widehat{\boldsymbol{\theta}}^{(k)}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

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

Two main steps in the Bayesian approach

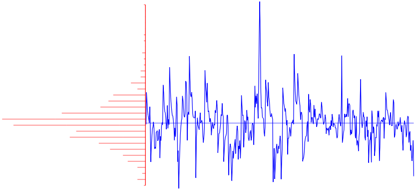
- Prior modeling
  - Separable:
    Gaussian, Gamma,
    Sparsity enforcing: Generalized Gaussian, mixture of Gaussians, mixture of Gammas, ...
  - Markovian:
    Gauss-Markov, GGM, ...
  - Markovian with hidden variables
    (contours, region labels)
- Choice of the estimator and computational aspects
  - MAP, Posterior mean, Marginal MAP
  - MAP needs optimization algorithms
  - Posterior mean needs integration methods
  - Marginal MAP and 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



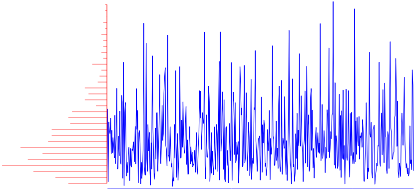
Gaussian

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



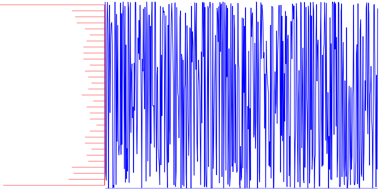
Generalized Gaussian

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



Gamma

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

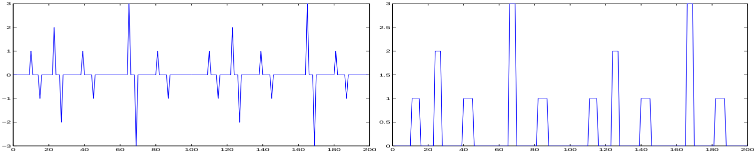


Beta

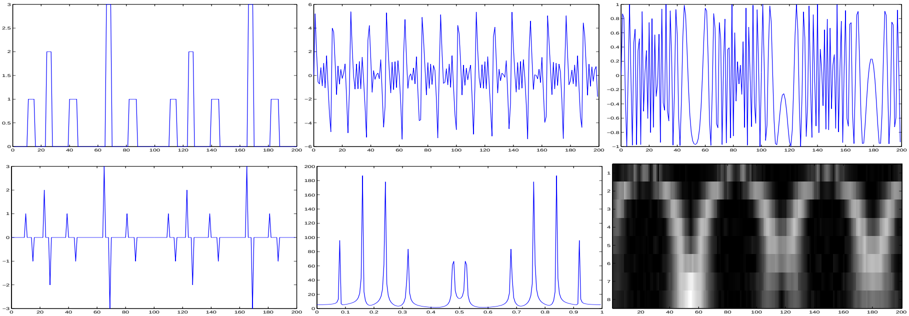
$$p(f_j) \propto f_j^\alpha (1 - f_j)^\beta$$

Sparsity enforcing prior models

- Sparse signals: Direct sparsity



- Sparse signals: Sparsity in a Transform domain



Sparsity enforcing prior models

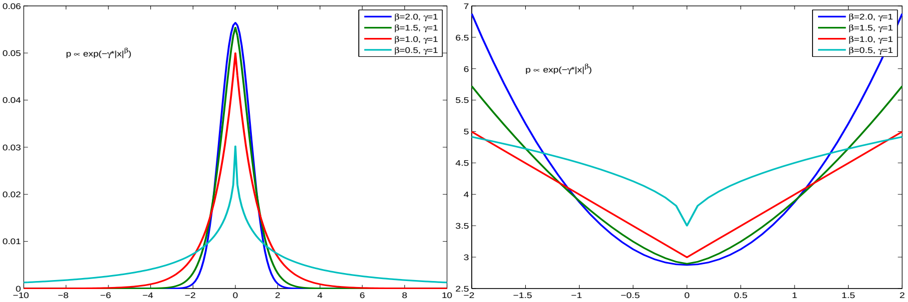
- Simple heavy tailed models:
  - Generalized Gaussian, Double Exponential
  - Symmetric Weibull, Symmetric Rayleigh
  - Student-t, Cauchy
  - Generalized hyperbolic
  - Elastic net
- Hierarchical mixture models:
  - Mixture of Gaussians
  - Bernoulli-Gaussian
  - Mixture of Gammas
  - Bernoulli-Gamma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

Simple heavy tailed models

- Generalized Gaussian, Double Exponential

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

$\beta = 1$ Double exponential or Laplace.
$0 < \beta \leq 1$ are of great interest for sparsity enforcing.



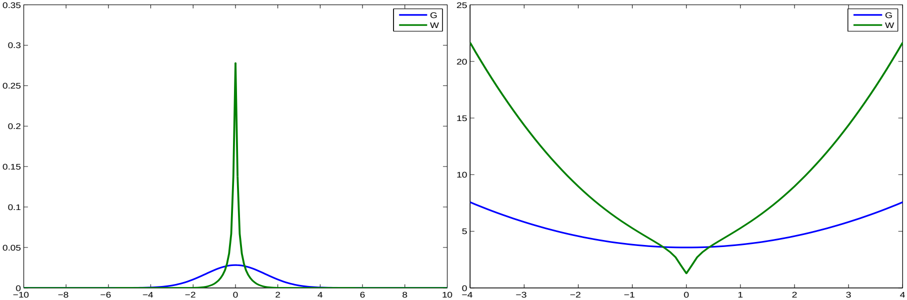
Generalized Gaussian family

Simple heavy tailed models

- Symmetric Weibull

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

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



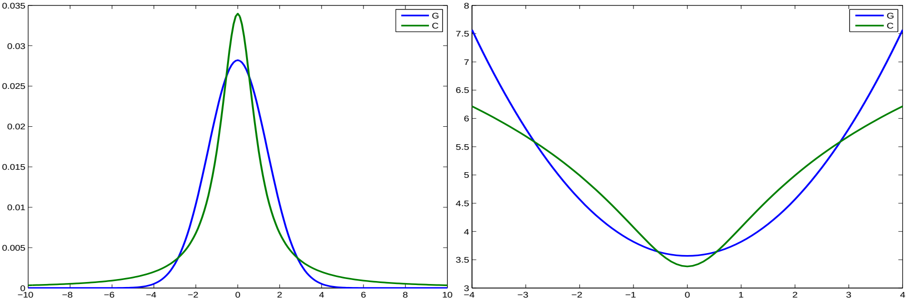
Symmetric Weibull family

Simple heavy tailed models

- Student-t and Cauchy models

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

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

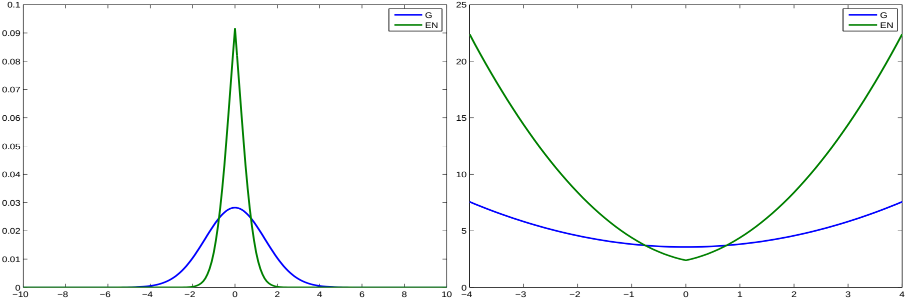


Student-t and Cauchy families

Simple heavy tailed models

- Elastic net prior model

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

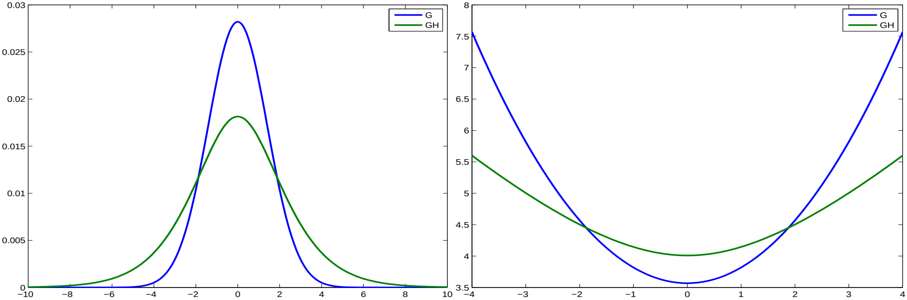


Elastic Net family

Simple heavy tailed models

- Generalized hyperbolic (GH) models

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



Generalized hyperbolic family

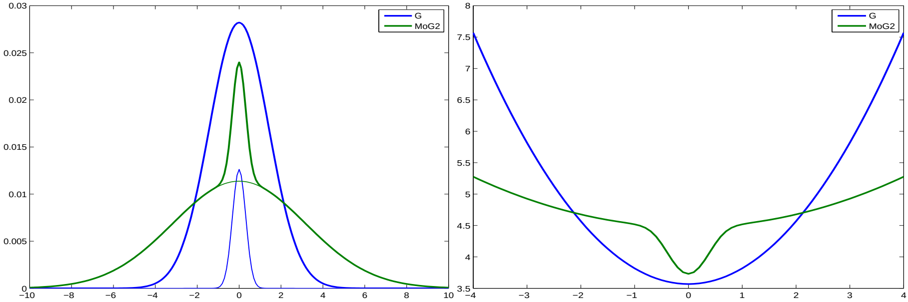
Mixture models

- Mixture of two Gaussians (MoG2) model

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

- Bernoulli-Gaussian (BG) model

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



Mixture of 2 Gaussians families

- Mixture of Gammas

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

- Bernoulli-Gamma model

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

- Mixture of Dirichlets model

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

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

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

- Bernoulli-Multinomial (BMultinomial) model

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

Hierarchical models and hidden variables

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

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

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

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

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

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

Hierarchical models and hidden variables

- Example 2: Student-t model

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

- Infinite mixture

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

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

Hierarchical models and hidden variables

- Example 3: Laplace (Double Exponential) model

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

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

- With these models we have:

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

Bayesian Computation and Algorithms

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

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

  and do any computations with these separable ones.

MCMC based algorithm

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

General scheme:

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

- Estimate $\boldsymbol{f}$ using $p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta})\, p(\boldsymbol{f}|\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}})$
  When Gaussian, can be done via optimization of a quadratic criterion.
- Estimate $\boldsymbol{z}$ using $p(\boldsymbol{z}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \widehat{\boldsymbol{\theta}})\, p(\boldsymbol{z})$
  Often needs sampling (hidden discrete variable)
- Estimate $\boldsymbol{\theta}$ using
  $p(\boldsymbol{\theta}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\widehat{\boldsymbol{f}}, \sigma_\epsilon^2 \boldsymbol{I})\, p(\widehat{\boldsymbol{f}}|\widehat{\boldsymbol{z}}, (m_k, v_k))\, p(\boldsymbol{\theta})$
  Use of Conjugate priors $\longrightarrow$ analytical expressions.

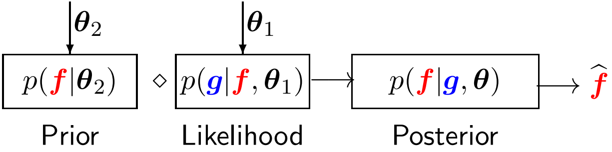
Variational Bayesian Approximation

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

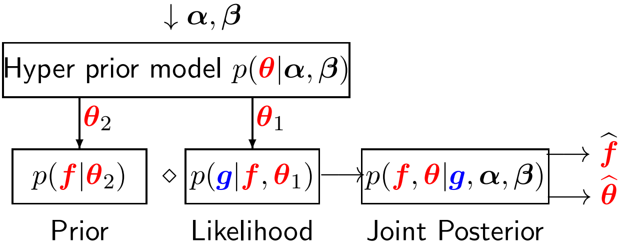
$$\left\{ \begin{array}{ll} q_1(\boldsymbol{f}) & \propto \exp\left[\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_2(\boldsymbol{\theta})}\right] \\ q_2(\boldsymbol{\theta}) & \propto \exp\left[\langle \ln p(\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_1(\boldsymbol{f})}\right]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

Summary of Bayesian estimation 1

- Simple Bayesian Model and Estimation

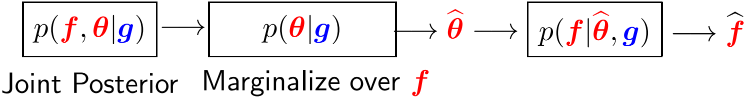


- Full Bayesian Model and Hyper-parameter Estimation

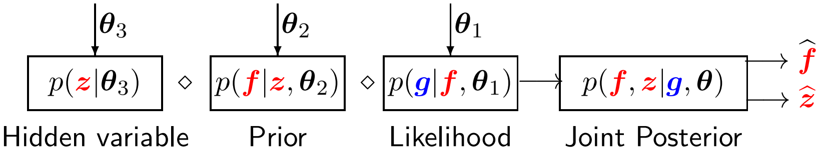


Summary of Bayesian estimation 2

- Marginalization for Hyper-parameter Estimation

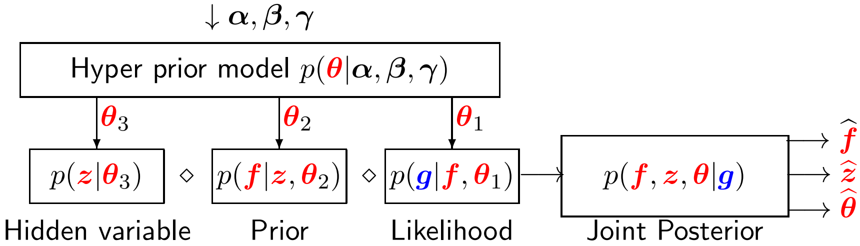


- Full Bayesian Model with a Hierarchical Prior Model

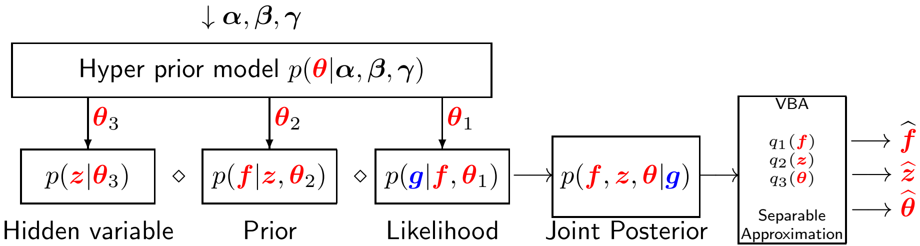


Summary of Bayesian estimation 3

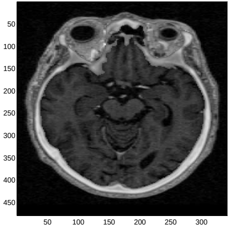
- Full Bayesian Hierarchical Model with Hyper-parameter Estimation

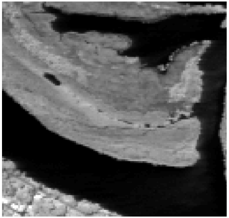


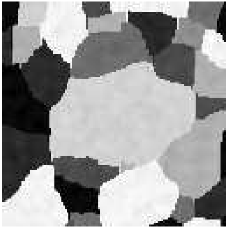
- Full Bayesian Hierarchical Model and Variational Approximation

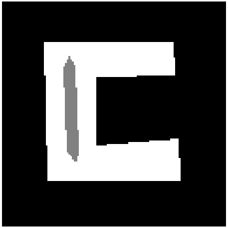


Which images I am looking for?

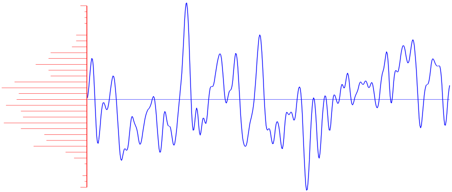




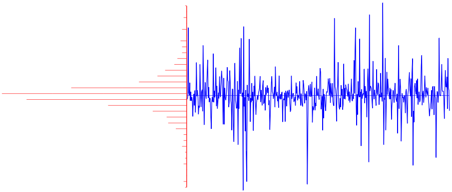




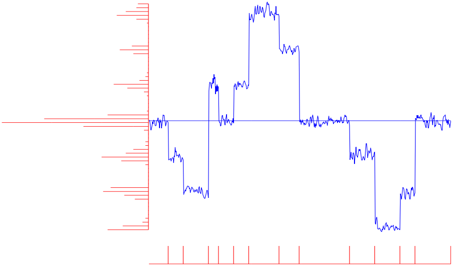
Which image I am looking for?



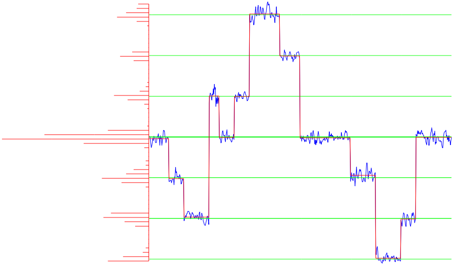
Gauss-Markov



Generalized GM

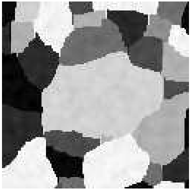


Piecewize Gaussian

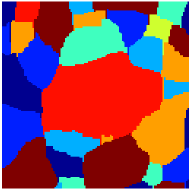


Mixture of GM

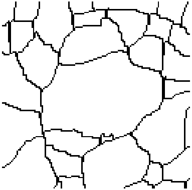
Gauss-Markov-Potts prior models for images



$f(\boldsymbol{r})$



$z(\boldsymbol{r})$



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

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

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

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

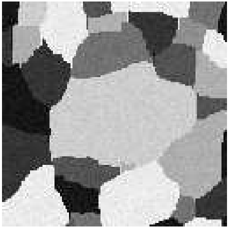
$$p(z(\boldsymbol{r})|z(\boldsymbol{r}'), \boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})) \propto \exp\left[\gamma \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right]$$

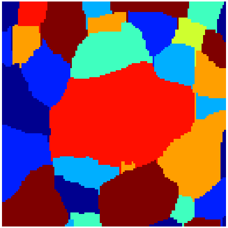
$$p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r} \in \mathcal{R}} \sum_{\boldsymbol{r}' \in \mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right]$$

Four different cases

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

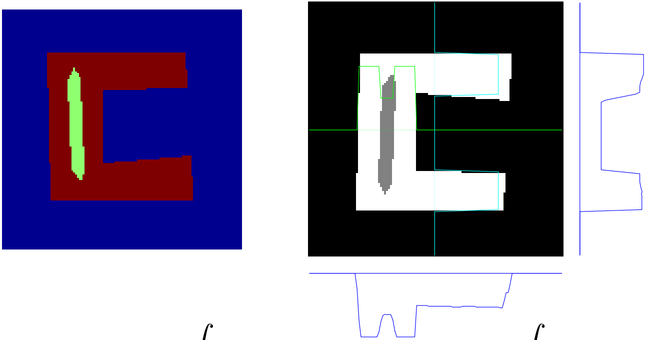
- $\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid, $\boldsymbol{z}$ iid :
  Mixture of Gaussians
- $\boldsymbol{f}|\boldsymbol{z}$ Gauss-Markov, $\boldsymbol{z}$ iid :
  Mixture of Gauss-Markov
- $\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid, $\boldsymbol{z}$ Potts-Markov :
  Mixture of Independent Gaussians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


$f(\boldsymbol{r})$


$z(\boldsymbol{r})$

Application of CT in NDT

Reconstruction from only 2 projections

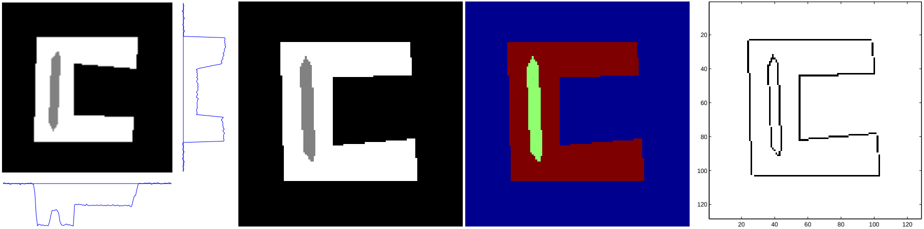


$$g_1(x) = \int f(x,y)\, dy, \qquad g_2(y) = \int f(x,y)\, dx$$

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

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

Application in CT



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

Proposed algorithm

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

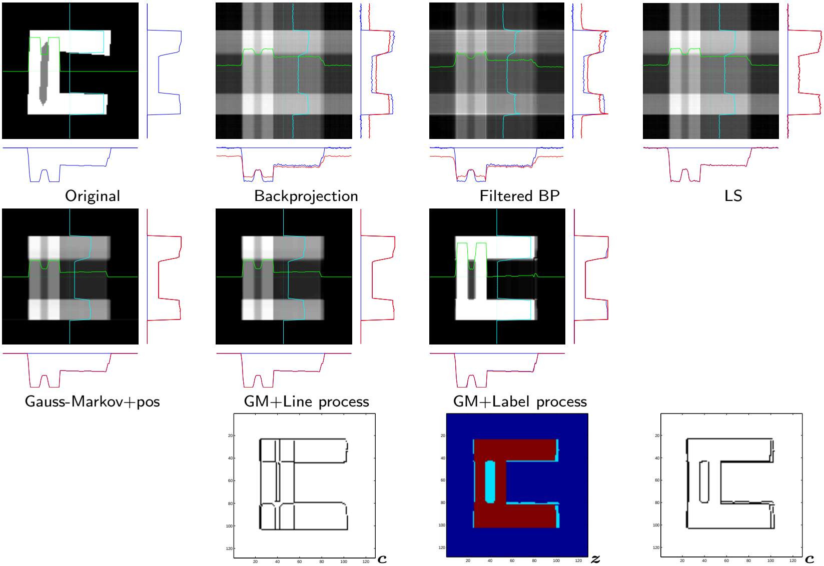
General scheme:

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

Iterative algorithm:

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

Results

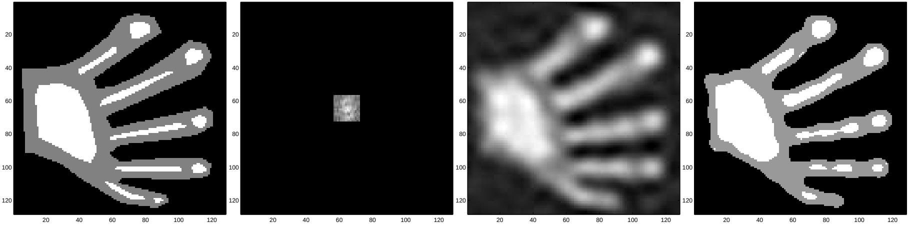

Original
Backprojection
Filtered BP
LS
Gauss-Markov+pos
GM+Line process
GM+Label process
c
z
c

Application in Microwave imaging

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

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

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



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

Conclusions

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