Deterministic and Bayesian Sparsity enforcing models in signal and image processing

Ali Mohammad-Djafari
Laboratoire des Signaux et Systèmes (L2S)
UMR8506 CNRS-CentraleSupélec-UNIV PARIS SUD
SUPELEC, 91192 Gif-sur-Yvette, France
http://lss.centralesupelec.fr
Email: djafari@lss.supelec.fr
http://djafari.free.fr
http://publicationslist.org/djafari

Keynote talk at SITIS 2016, Napoli, Italy

Contents

1. Sparsity ?
2. Sparsity in data, signal and image processing
3. Modelling for sparse representation
4. Sparsity as a deterministic Regularizer
5. Bayesian approach:
   Maximum A Posteriori (MAP) and link with Regularization
6. Prior models for enforcing (promoting) sparsity
   - Heavy tailed: Double Exponential, Generalized Gaussian, ...
   - Mixture models: Mixture of Gaussians, Student-t, ...
   - Hierarchical models with hidden variables
   - General Gauss-Markov-Potts models
7. Bayesian Computational tools:
   Joint Maximum A Posteriori (JMAP), MCMC and Variational Bayesian Approximation (VBA)
8. Applications in Inverse Problems:
   X ray Computed Tomography, Microwave and Ultrasound imaging, Sattelite and Hyperspectral image processing, ...

Sparsity?

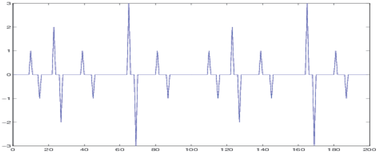
- Signal and image representation and modelling:
  Many real world signals, sounds, images can be represented by a sparse model.

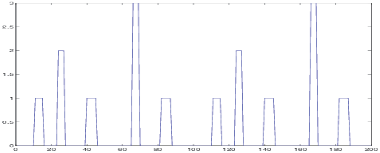
- Sparse modelling in Inverse problems and Machine learning:
  Sparsity can be used as regularizer to avoid over fitting in many machine learning problems:
  Feature selection, SVMs, ...

- Sparsity as a tool for fast algorithms:
  Sparsity can be exploited for fast computations
  Matrix factorisation for recommender systems
  Sparse solutions in kernel machines

Sparse signals and images

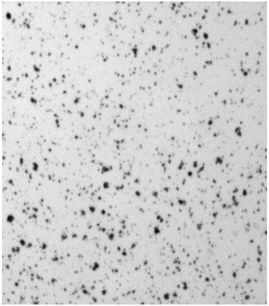
- Sparse signals: Direct sparsity



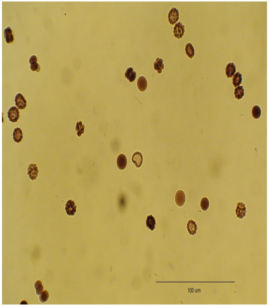


- Sparse images: Direct sparsity

B& W text



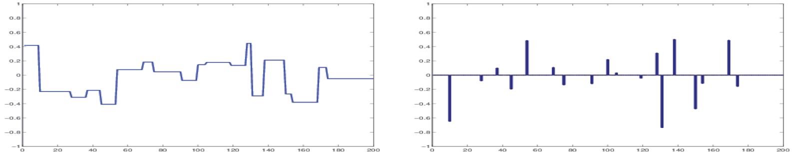
Cell Colony



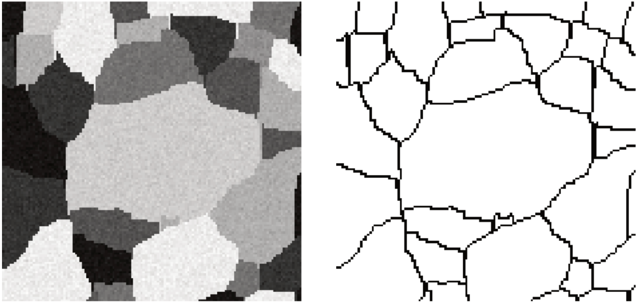
embryos

Sparse signals and images

- Sparse signals in a Transform domain



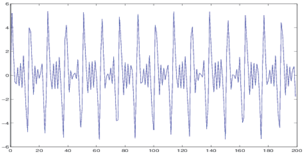
- Sparse images in a Transform domain



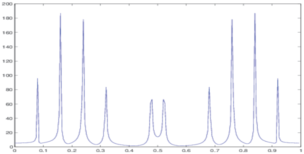
Sparse signals and images

- Sparse signals in Time and Fourier domain

Time domain

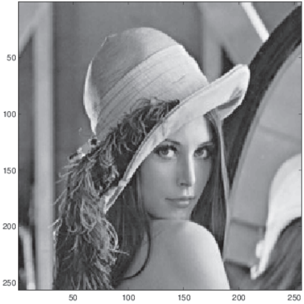


Fourier domain

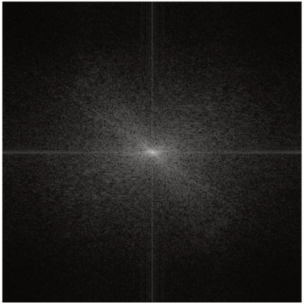


- Sparse images in Space and Fourier domain

Space domain

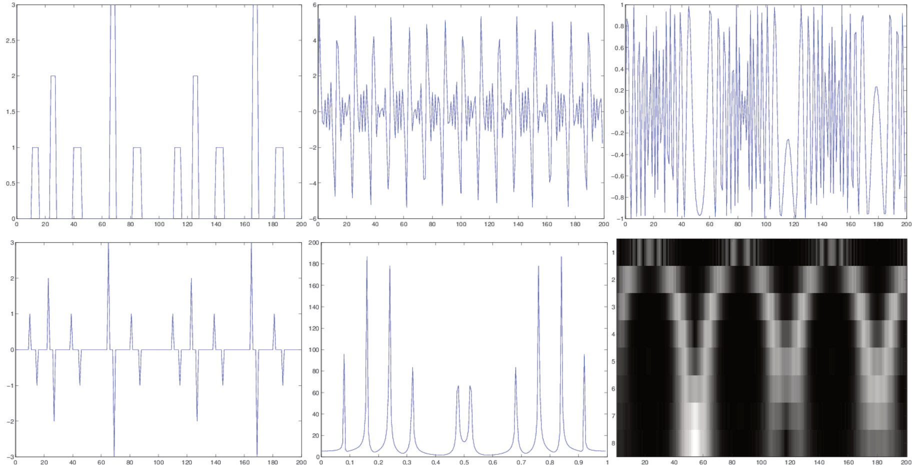


Fourier domain

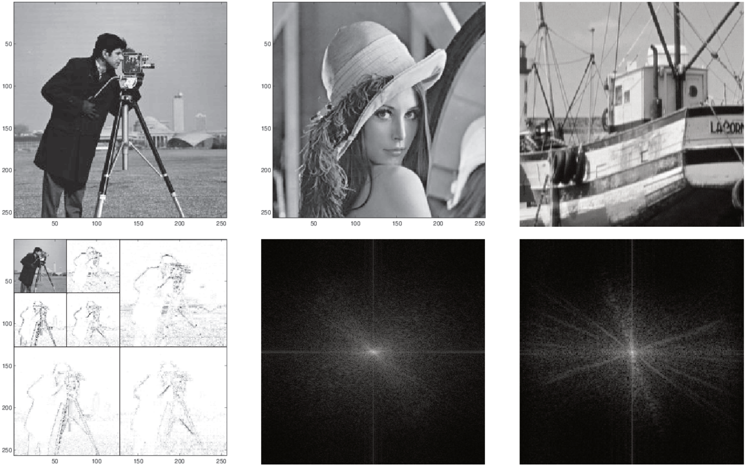


Sparse signals and images

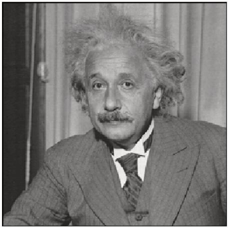
- Sparse signals: Sparsity in a Transform domaine



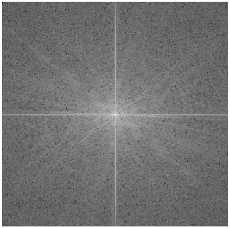
Sparse signals and images



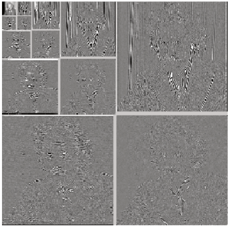
Sparse signals and images (Fourier and Wavelets domain)



Image



Fourier



Wavelets

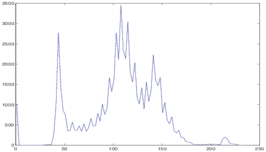
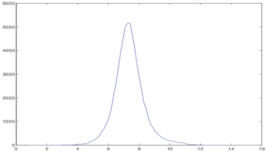
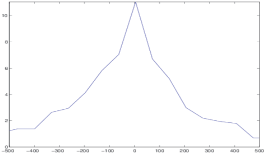


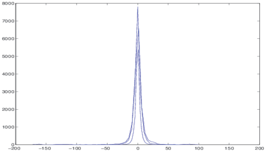
Image hist.



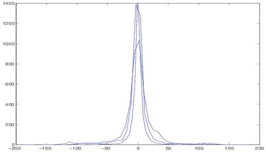
Fourier coeff. hist.



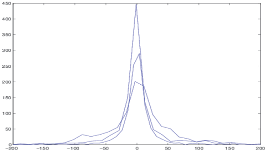
Wavelet coeff. hist.



bands 1-3



bands 4-6



bands 7-9

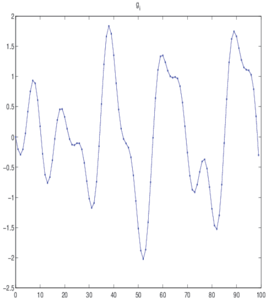
Finite and Sparse representation: some references

- 1948: Shannon:
  Sampling theorem and reconstruction of a band limited signal
- 1993-2007:
  - Mallat, Zhang, Candès, Romberg, Tao and Baraniuk:
    Non linear sampling, Compression and reconstruction,
  - Fuch: Sparse representation
  - Donoho, Elad, Tibshirani, Tropp, Duarte, Laska:
    Compressive Sampling, Compressive Sensing
- 2007-2016: Deterministic
  Algorithms for sparse representation and Compressive Sampling: Matching Pursuit (MP), Projection Pursuit Regression, Pure Greedy Algorithm, OMP, Basis Poursuit (BP), Dantzig Selector (DS), Least Absolute Shrinkage and Selection Operator (LASSO), Iterative Hard Thresholding...
- 2003-2016: Bayesian
  Bayesian approach to sparse modeling
  Tipping, Bishop: Sparse Bayesian Learning,
  Relevance Vector Machine (RVM), Sparsity enforcing priors,...

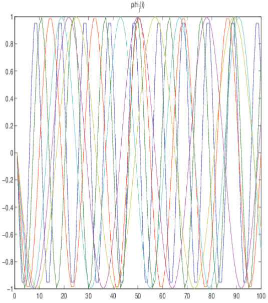
Modelling and representation

- Modelling via decomposition
  (basis, codebook, dictionary, Design Matrix)

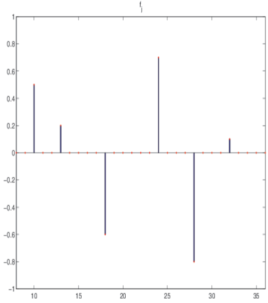
$$g(t) = \sum_{j=1}^{N} f_j \, \phi_j(t), \;\; t = 1, \cdots, T \longrightarrow \mathbf{g} = \mathbf{\Phi} \, \mathbf{f}$$



$T = 100$



$[100 \times 35]$

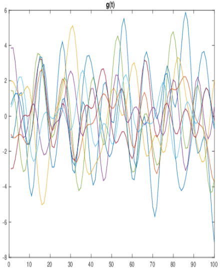


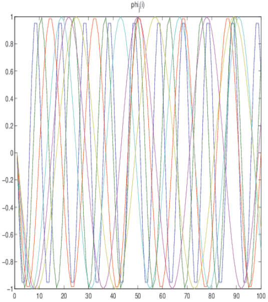
$N = 35 (7 \, nonzero)$

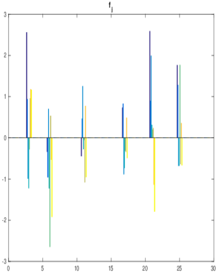
Modelling and representation

- Modelling via decomposition
  (basis, codebook, dictionary, Design Matrix,...)

$$g(t) = \sum_{j=1}^{N} f_j \, \phi_j(t), \; t = 1, \cdots, T \longrightarrow \mathbf{g} = \mathbf{\Phi} \, \mathbf{f}$$



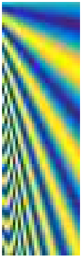




$T = 100$ $\qquad [100 \times 35]$ $\qquad N = 35(7 \, nonzero)$

Modelling and representation

$$g(t) \quad = \quad \sum_j \phi_j(t) \quad f_j$$

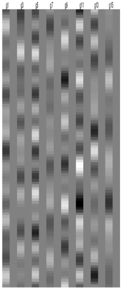


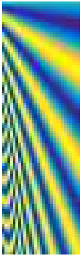
$$\mathbf{g} \quad = \quad \boldsymbol{\Phi} \quad \mathbf{f}$$

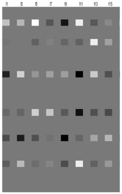
$T = 100$ $[100 \times 35]$ $N = 35$

Modelling and representation

$$g(t) \quad = \quad \sum_j \phi_j(t) \quad f_j$$







$$\mathbf{g} \quad = \quad \boldsymbol{\Phi} \quad \mathbf{f}$$

$T = 100$ $[100 \times 35]$ $N = 35$

Modelling and representation

- Modelling via a basis (codebook, dictionary, Design Matrix)

$$g(t)=\sum_{j=1}^{N} f_j\,\phi_j(t),\;\; t=1,\cdots,T \longrightarrow \mathbf{g}=\boldsymbol{\Phi}\,\mathbf{f}$$

- When $T \geq N$

$$\widehat{f}_j=\arg\min_{f_j}\left\{\sum_{t=1}^{T}\left|g(t)-\sum_{j=1}^{N} f_j\,\phi_j(t)\right|^2\right\} \longrightarrow$$

$$\widehat{\mathbf{f}}=\arg\min_{\mathbf{f}}\left\{\|\mathbf{g}-\boldsymbol{\Phi}\mathbf{f}\|_2^2\right\}=[\boldsymbol{\Phi}'\boldsymbol{\Phi}]^{-1}\boldsymbol{\Phi}'\mathbf{g}$$

- When orthogonal basis: $\boldsymbol{\Phi}'\boldsymbol{\Phi}=\mathbf{I} \longrightarrow \widehat{\mathbf{f}}=\boldsymbol{\Phi}'\mathbf{g}$

$$\widehat{f}_j=\sum_{t=1}^{N} g(t)\,\phi_j(t)=<g(t),\phi_j(t)>$$

- Application in Compression, Transmission and Decompression

Modelling and representation

- When over complete basis $N > T$: Infinite number of solutions for $\mathbf{\Phi f} = \mathbf{g}$. We have to select one.
  Minimum norm solution:
$$\widehat{\mathbf{f}} = \arg\min_{\mathbf{f}:\ \mathbf{\Phi f}=\mathbf{g}} \left\{ \|\mathbf{f}\|_2^2 \right\}$$
  or writing differently:
$$\text{minimize} \quad \|\mathbf{f}\|_2^2 \quad \text{subject to} \quad \mathbf{\Phi f} = \mathbf{g}$$
  resulting to:
$$\widehat{\mathbf{f}} = \mathbf{\Phi}'[\mathbf{\Phi\Phi}']^{-1}\mathbf{g}$$
- Again if $\mathbf{\Phi\Phi}' = \mathbf{I} \longrightarrow \widehat{\mathbf{f}} = \mathbf{\Phi}'\mathbf{g}$.
- No real interest if we have to keep all the N coefficients:
- Sparsity:
$$\text{minimize} \quad \|\mathbf{f}\|_0 \quad \text{subject to} \quad \mathbf{\Phi f} = \mathbf{g}$$
  or
$$\text{minimize} \quad \|\mathbf{f}\|_1 \quad \text{subject to} \quad \mathbf{\Phi f} = \mathbf{g}$$

Sparse decomposition (MP and OMP)

- Strict sparsity and exact reconstruction

$$\text{minimize} \quad \|\mathbf{f}\|_0 \quad \text{subject to} \quad \boldsymbol{\Phi}\mathbf{f} = \mathbf{g}$$

$\|\mathbf{f}\|_0$ is the number of non-zero elements of $\mathbf{f}$

- Matching Pursuit (MP) [Mallat & Zhang, 1993]
  - MP is a greedy algorithm that finds one atom at a time.
  - Find the one atom that best matches the signal;
    Given the previously found atoms, find the next one to best fit, Continue to the end.
- Orthogonal Matching Pursuit (OMP) [Lin, Huang et al., 1993]
  The Orthogonal MP (OMP) is an improved version of MP that re-evaluates the coefficients after each round.

Sparse decomposition (BP,PPR,BCR,IHT,...)

- Sparsity enforcing and exact reconstruction

$$\text{minimize} \quad \|\mathbf{f}\|_1 \quad \text{subject to} \quad \boldsymbol{\Phi}\mathbf{f} = \mathbf{g}$$

  - This problem is convex (linear programming).
  - Very efficient solvers has been deployed:
    - Interior point methods [Chen, Donoho & Saunders (95)],
    - Iterated shrinkage [Figuerido & Nowak (03), Daubechies, Defrise, & Demole (04), Elad (05), Elad, Matalon, & Zibulevsky (06), Marvasti et al].
  - Basis Pursuit (BP)
  - Projection Pursuit Regression
  - Block Coordinate Relaxation (BCR)
  - Greedy Algorithms
  - Iterative Hard Thresholding (IHT)

Sparse decomposition algorithms

- Strict sparsity and exact reconstruction

$$\text{minimize} \quad \|\mathbf{f}\|_0 \quad \text{subject to} \quad \mathbf{g} = \boldsymbol{\Phi}\mathbf{f}$$

- Strict sparsity and approximate reconstruction

$$\text{minimize} \quad \|\mathbf{f}\|_0 \quad \text{subject to} \quad \|\mathbf{g} - \boldsymbol{\Phi}\mathbf{f}\|_2^2 < c$$

- NP-hard. Looking for other solutions:
- Sparsity promoting and exact reconstruction:
  Basis Pursuit (BP)

$$\text{minimize} \quad \|\mathbf{f}\|_1 \quad \text{subject to} \quad \boldsymbol{\Phi}\mathbf{f} = \mathbf{g}$$

- Sparsity promoting and approximate reconstruction:

$$\text{minimize} \quad \|\mathbf{f}\|_1 \quad \text{subject to} \quad \|\mathbf{g} - \boldsymbol{\Phi}\mathbf{f}\|_2^2 < c$$

or equivalently (LASSO):

$$\widehat{\mathbf{f}} = \arg\min_{\mathbf{f}} \{J(\mathbf{f})\} \quad \text{with} \quad J(\mathbf{f}) = \frac{1}{2}\|\mathbf{g} - \boldsymbol{\Phi}\mathbf{f}\|_2^2 + \lambda\|\mathbf{f}\|_1$$

Sparse Decomposition Applications

- Denoising: $\mathbf{g} = \mathbf{f} + \epsilon$ with $\mathbf{f} = \boldsymbol{\Phi}\mathbf{z}$

$$J(\mathbf{z}) = \frac{1}{2}\|\mathbf{g} - \boldsymbol{\Phi}\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_1$$

  When $\widehat{\mathbf{z}}$ computed, we can compute $\widehat{\mathbf{f}} = \boldsymbol{\Phi}\widehat{\mathbf{z}}$.

- Compressed Sensing and Linear Inverse problems:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon \quad \text{with} \quad \mathbf{f} = \boldsymbol{\Phi}\mathbf{z}$$

$$J(\mathbf{z}) = \frac{1}{2}\|\mathbf{g} - \mathbf{H}\boldsymbol{\Phi}\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_1$$

  When $\widehat{\mathbf{z}}$ computed, we can compute $\widehat{\mathbf{f}} = \boldsymbol{\Phi}\widehat{\mathbf{z}}$.

- Linear Inverse problems with piecewise constant prior:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon \quad \text{with} \quad \mathbf{D}\mathbf{f} = \mathbf{z} \text{ and } z_j \sim \mathcal{DE}(\lambda) \text{ Sparse}$$

$$J(\mathbf{f}) = \frac{1}{2}\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda\|\mathbf{D}\mathbf{f}\|_1$$

Sparse Decomposition algorithms (Unitary decomposition)

$$J(\mathbf{z}) = \frac{1}{2}\|\mathbf{g} - \boldsymbol{\Phi}\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_1$$

- When $\boldsymbol{\Phi}\boldsymbol{\Phi}' = \boldsymbol{\Phi}'\boldsymbol{\Phi} = \mathbf{I}$

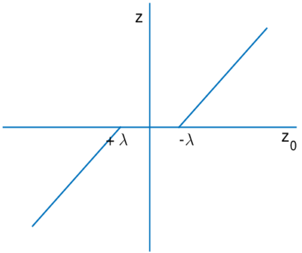
$$J(\mathbf{z}) = \frac{1}{2}\|\boldsymbol{\Phi}'\mathbf{g} - \boldsymbol{\Phi}'\boldsymbol{\Phi}\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_1 = \frac{1}{2}\|\mathbf{z}_0 - \mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_1 \text{ with } \mathbf{z}_0 = \boldsymbol{\Phi}'\mathbf{g}$$

which is a separable criterion:

$$J(\mathbf{f}) = \frac{1}{2}\|\mathbf{z} - \mathbf{z}_0\|_2^2 + \lambda\|\mathbf{z}\|_1 = \sum_j \frac{1}{2}|z_j - \mathbf{z}_{0j}|^2 + \lambda|z_j|_1$$

- Closed form solution: Shrinkage

$$z_j = \begin{cases} 0 & |z_{0j}| < \lambda \\ z_{0j} - \mathrm{sign}(z_{0j})\lambda & \text{otherwise} \end{cases}$$



Sparse Decomposition Algorithms (Lasso and extensions)

- LASSO:

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{\Phi f}\|_2^2 + \lambda \sum_j |f_j|$$

- Other Criteria
  - L_p

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{\Phi f}\|_2^2 + \lambda_1 \sum_j |f_j|^p, \quad 1 < p \leq 2$$

  - Elastic net

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{\Phi f}\|_2^2 + \sum_j \left(\lambda_1 |f_j| + \lambda_2 |f_j|^2\right)$$

  - Group LASSO

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{\Phi f}\|_2^2 + \lambda_1 \sum_j |f_j| + \lambda_2 \sum_j |f_j - f_{j-1}|^2$$

  - Weighted L1:

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{\Phi f}\|_2^2 + \lambda \sum_j |w_j f_j|$$

$w_j = \lambda$

Advanced Algorithms

Many optimisation algorithms have been developed based on the facts that: $J(\mathbf{f})$ is composed of a first term which is sum of strictly convex terms and the second term is convex but not derivable at the origin.

- Alternating Direction Method of Multipliers (ADMM) is based on the splitting variables and augmented Lagrangian [Boyd et al. 2011,...]
- Iterative Shrinkage Thresholding Algorithm (ISTA) and its accelerated version Fast ISTA (FISTA) are based on the duality, convex conjugate property, level set and minmax theorem [Beck and Teboulle 2009, Parikh and Boyd, 2014,...]
- Cyclic or Bloc Coordinate Descent method (CCD or BCD) are based on bloc coordinate optimization of the criterion [Saha and Tewari, 2010] .

Algorithms:
Alternating Direction Method of Multipliers (ADMM)

- ADMM is based on the splitting variables and augmented Lagrangian [Boyd et al., 2011]

$$J(\mathbf{f})=\frac{1}{2}\|\mathbf{g}-\boldsymbol{\Phi}\mathbf{f}\|_2^2+\lambda\|\mathbf{f}\|_1 \text{ s.t. } \mathbf{f}-\mathbf{z}=0$$

$$\mathcal{L}(\mathbf{f},\mathbf{z},\mu)=\frac{1}{2}\|\mathbf{g}-\boldsymbol{\Phi}\mathbf{f}\|_2^2+\lambda\|\mathbf{z}\|_1+\mu'(\mathbf{f}-\mathbf{z})+\frac{\rho}{2})\|\mathbf{f}-\mathbf{z}\|_2^2$$

Stationary point of $\mathcal{L}$ can be reached by an alternate optimization with respect to $\mathbf{f}$, $\mathbf{z}$ and μ gives the algorithm.

$$\left\{\begin{array}{l}
\mathbf{f}^{(k+1)}=\arg\min_{\mathbf{f}}\left\{Lc(\mathbf{f},\mathbf{z}^{(k)},\mu^{(k)})\right\}\\
\mathbf{z}^{(k+1)}=\arg\min_{\mathbf{z}}\left\{Lc(\mathbf{f}^{(k)},\mathbf{z},\mu^{(k)})\right\}\\
\mu^{(k+1)}=\mu^{(k)}+\tau(\mathbf{f}^{(k)}-\mathbf{z}^{(k)})
\end{array}\right.$$

Algorithms: Iterative Shrinkage Thresholding Algorithm (ISTA)

ISTA and its accelerated version Fast ISTA (FISTA) are based on the duality, convex conjugate property, level set and MinMax theorems and can be summarized in two steps:

$$J(\mathbf{f}) = J_0(\mathbf{f}) + \lambda\|\mathbf{f}\|_1 \text{ with } J_0(\mathbf{f}) = \frac{1}{2}\|\mathbf{g} - \boldsymbol{\Phi}\mathbf{f}\|_2^2$$

- Prox Linear approximation:

$$J(\mathbf{f}) = \lambda\|\mathbf{f}\|_1 + \left\langle \nabla J_0(\mathbf{f}^{(k)}), (\mathbf{f} - \mathbf{f}^{(k)}) \right\rangle + \frac{1}{2\delta^{(k)}}\|\mathbf{f} - \mathbf{f}^{(k)}\|_2^2$$

- Shrinkage:

$$\mathbf{f}^{(k+1)} = \mathcal{S}\left(\mathbf{f}^{(k)} - \delta^{(k)}\nabla J_0(\mathbf{f}), \mathbf{f}^{(k)}, \lambda\delta^{(k)}\right)$$

whre $\mathcal{S}(z_j, t) = max(|z_j| - t,\ 0)\text{sign}(z_j)$

Other criteria

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{\Phi f}\|_2^2 + \lambda \sum_j \phi(f_j)$$

Convex criteria:

- $L_p: \quad \phi(f_j) = |f_j|^p, \quad 1 \le p \le 2$
- Hubber: $\quad \phi(f_j) = \begin{cases} f_j^2 & |f_j| < s \\ s^2 + |f_j - s| & \text{otherwise} \end{cases}$

Non Convex criteria:

- fractional power: $\quad \phi(f_j) = |f_j|^p, \quad 0 < p < 1$
- Truncated quadratic: $\quad \phi(f_j) = \begin{cases} f_j^2 & |f_j| < s \\ s^2 & \text{otherwise} \end{cases}$
- Cauchy: $\quad \phi(f_j) = s \ln(1 + |f_j|/s)$
- $\phi(f_j) = |f_j| + s \ln(1 + |f_j|/s)$

Dictionary learning

- Given a set of training data $(g_k(t), f_{jk})$ related by:

$$g_k(t) = \sum_{j=1}^{N} \phi_j(t)\, f_{jk}, \; t = 1, \cdots, T, \; k = 1, \cdots, K$$

  or equivalently, given

$$\mathbf{g}_k = \boldsymbol{\Phi}\, \mathbf{f}_k, \; k = 1, \cdots, K$$

  determine $\boldsymbol{\Phi}$.

- Objective criterion:

$$J(\boldsymbol{\Phi}) = \sum_k \sum_t \left| g_k(t) - \sum_j \phi_j(t) f_{jk} \right|^2 + \lambda \sum_t \sum_j |\phi_j(t)|^2$$

$$J(\boldsymbol{\Phi}) = \sum_k \|\mathbf{g}_k - \boldsymbol{\Phi}\mathbf{f}_k\|_2^2 + \lambda \|\boldsymbol{\Phi}\|_2^2$$

Dictionary learning

- Optimizing

$$J(\boldsymbol{\Phi}) = \sum_k \|\mathbf{g}_k - \boldsymbol{\Phi}\mathbf{f}_k\|_2^2 + \lambda\|\boldsymbol{\Phi}\|_2^2$$

gives

$$\widehat{\boldsymbol{\Phi}} = \left[\sum_k \mathbf{g}_k\mathbf{g}_k' + \lambda\mathbf{I}\right]^{-1} \mathbf{g}_k'\mathbf{f}_k$$

- Looking for sparse dictionary, we can use

$$J(\boldsymbol{\Phi}) = \sum_k \|\mathbf{g}_k - \boldsymbol{\Phi}\mathbf{f}_k\|_2^2 + \lambda\|\boldsymbol{\Phi}\|_1$$

and we can again use an iterative algorithm to find the solution.

Joint Dictionary learning and sparse reconstruction

- Given a set of data $\mathbf{g}_k$ modelled as

$$\mathbf{g}_k = \boldsymbol{\Phi}\, \mathbf{f}_k, \;\; k = 1, \cdots, K$$

  determine both the dictionary $\boldsymbol{\Phi}$ and the compositions $\mathbf{f}_k$.
- Joint criterion

$$J(\boldsymbol{\Phi}, \mathbf{f}_k) = \sum_k \|\mathbf{g}_k - \boldsymbol{\Phi}\mathbf{f}_k\|_2^2 + \lambda_0 |\boldsymbol{\Phi}\|_2^2 + \lambda_1 \sum_k \|\mathbf{f}_k\|_2^2$$

- Alternate optimization:

$$\left\{ \begin{array}{l} \widehat{\boldsymbol{\Phi}} = [\sum_k \mathbf{g}_k \mathbf{g}_k' + \lambda_0 \mathbf{I}]^{-1} \; \mathbf{g}_k' \widehat{\mathbf{f}}_k \\ \widehat{\mathbf{f}}_k = \left[\widehat{\boldsymbol{\Phi}}' \widehat{\boldsymbol{\Phi}} + \lambda_1 \mathbf{I} \right]^{-1} \; \widehat{\boldsymbol{\Phi}}_k' \mathbf{g}_k \end{array} \right.$$

- Looking for sparse dictionary and sparse coefficients we can use

$$J(\boldsymbol{\Phi}) = \sum_k \|\mathbf{g}_k - \boldsymbol{\Phi}\mathbf{f}_k\|_2^2 + \lambda_0 \|\boldsymbol{\Phi}\|_1 + \lambda_1 \sum_k \|\mathbf{f}_k\|_1$$

  and we can again use an iterative algorithm to find the solution.

Multi Dimensional signals: PCA, SPCA, BSS, ...

$$g_i(t) = \sum_{j=1}^{N} \boldsymbol{\Phi}_{ij} f_j(t), \; i = 1, \cdots, M, \; t = 1, \cdots, T$$

$$\mathbf{g}(t) = \boldsymbol{\Phi}\, \mathbf{f}(t), \; t = 1, \cdots, T$$

$$\mathbf{G} = \boldsymbol{\Phi}\, \mathbf{F}, \quad \text{with } \mathbf{G}\ [M \times T], \quad \boldsymbol{\Phi}\ [M \times N], \quad \mathbf{F}\ [N \times T]$$

- $f_j(t)$ factors, sources, codes
- $\boldsymbol{\Phi}$ Loading matrix (Factor Analysis),
  Mixing matrix (Blind Sources Separation),
  Design matrix (Sparse coding, Compressed Sensing)
- Objective: Find $\boldsymbol{\Phi}$ and $f_j(t)$

$$J(\mathbf{f}(t), \boldsymbol{\Phi}) = \sum_t \|\mathbf{g}(t) - \boldsymbol{\Phi}\mathbf{f}(t)\|_2^2 + \lambda_1 \sum_i \sum_j |\boldsymbol{\Phi}_{ij}| + \lambda_2 \sum_t \sum_j |f_j(t)|$$

$$J(\mathbf{F}, \boldsymbol{\Phi}) = \|\mathbf{G} - \boldsymbol{\Phi}\mathbf{F}\|_2^2 + \lambda_1 \|\boldsymbol{\Phi}\|_1 + \lambda_2 \|\mathbf{F}\|_1$$

Matrix Decomposition or Approximation

- Matrix approximation:
  Find an approximate matrix $\widehat{\mathbf{G}} = \mathbf{\Phi F}$ for $\mathbf{G}$ with some degrees of sparsity in the elements of $\mathbf{\Phi}$ and $\mathbf{F}$.

$$J(\mathbf{F}, \mathbf{\Phi}) = \|\mathbf{G} - \mathbf{\Phi F}\|_2^2 + \lambda_1 \|\mathbf{\Phi}\|_1 + \lambda_2 \|\mathbf{F}\|_1$$

- Low rank Matrix decomposition:

$$\widehat{\mathbf{G}} = \sum_{k=1}^{K} d_k \mathbf{u}_k \mathbf{v}_k' = \mathbf{UDV}$$

  with some degrees of sparsity in the elements of $\mathbf{u}$ and $\mathbf{v}$.

$$J(\mathbf{U}, \mathbf{V}) = \|\mathbf{G} - \mathbf{UDV}\|_2^2 + \lambda_1 \|\mathbf{U}\|_1 + \lambda_2 \|\mathbf{V}\|_1$$

Bayesian approach

- Bayesian approach: $\mathbf{g} = \mathbf{\Phi}\mathbf{f} + \epsilon$

$$p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f})\, p(\mathbf{f})}{p(\mathbf{g})} \propto p(\mathbf{g}|\mathbf{f})\, p(\mathbf{f})$$

- Priors: Gaussian noise and Double Exp ($\mathcal{DE}$) for $\mathbf{f}$

$$\begin{cases} p(\mathbf{g}|\mathbf{f}) = \mathcal{N}(\mathbf{g}|\mathbf{\Phi}\mathbf{f}, \sigma_\epsilon^2) \propto \exp\left[\frac{-1}{2\sigma_\epsilon^2}\|\mathbf{g} - \mathbf{\Phi}\mathbf{f}\|_2^2\right] \\ p(\mathbf{f}) = \mathcal{DE}(\mathbf{f}|\gamma) \propto \exp\left[-\gamma\|\mathbf{f}\|_1\right] \end{cases}$$

- Maximum A Posteriori (MAP):

$$\widehat{\mathbf{f}} = \arg\max_{\mathbf{f}} \{p(\mathbf{f}|\mathbf{g})\} = \arg\min_{\mathbf{f}} \{J(\mathbf{f})\}$$

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{\Phi}\mathbf{f}\|_2^2 + \lambda\|\mathbf{f}\|_1 \quad \text{with} \quad \lambda = 2\gamma\sigma_\epsilon^2$$

- MAP = LASSO

Sparse decomposition: Regularization or MAP

- Regularization: $J(\mathbf{f}) = \|\mathbf{g} - \boldsymbol{\Phi}\mathbf{f}\|_2^2 + \lambda\|\mathbf{f}\|_1$
- With fixed λ: Find a good optimization algorithm
- How to choose λ ?
- L-Curve, Cross Validation, adhoc $\lambda = 1$, ...
- ---
- MAP: $p(\mathbf{f}|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f})\, p(\mathbf{f})$
  $\widehat{\mathbf{f}} = \arg\max_{\mathbf{f}} \{p(\mathbf{f}|\mathbf{g})\} = \arg\min_{\mathbf{f}} \{J(\mathbf{f})\}$ with
  $$J(\mathbf{f}) = \|\mathbf{g} - \boldsymbol{\Phi}\mathbf{f}\|_2^2 + \lambda\|\mathbf{f}\|_1 \quad \text{with} \quad \lambda = 2\gamma\sigma_\epsilon^2$$
- How to estimate γ and σ_ϵ^2? $\boldsymbol{\theta} = (\gamma, \sigma_\epsilon^2)$
- Bayesian: $p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f}, \sigma_\epsilon^2)\, p(\mathbf{f}|\gamma) p(\gamma)) p(\sigma_\epsilon^2)$
  Joint MAP, Expectation-Maximization, MCMC, Variational Bayesian Approximation,...
- Advantages of the Bayesian approach:
  - More probabilistic modelling for sparsity enforcing
  - Hyperparameter estimation
  - Uncertainty handling and quantification

Sparsity enforcing prior models

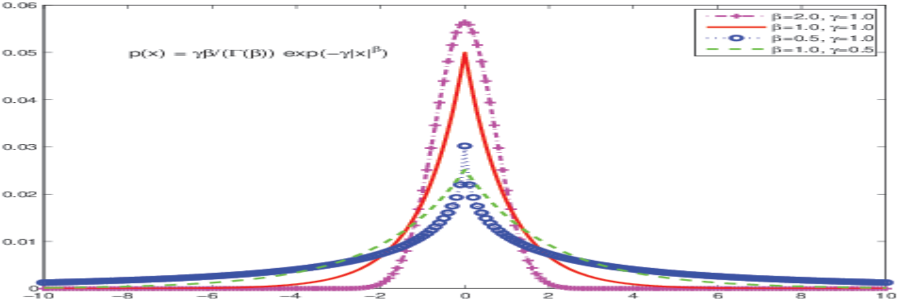
- Simple heavy tailed models:
  - Generalized Gaussian, Double Exponential
  - Student-t, Cauchy
  - Generalized hyperbolic
  - Symmetric Weibull, Symmetric Rayleigh
  - Elastic net
- Hierarchical mixture models:
  - Mixture of Gaussians
  - Bernoulli-Gaussian
  - Mixture of Gammas
  - Bernoulli-Gamma
  - 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.
$0 < \beta < 2$ are of great interest for sparsity enforcing.



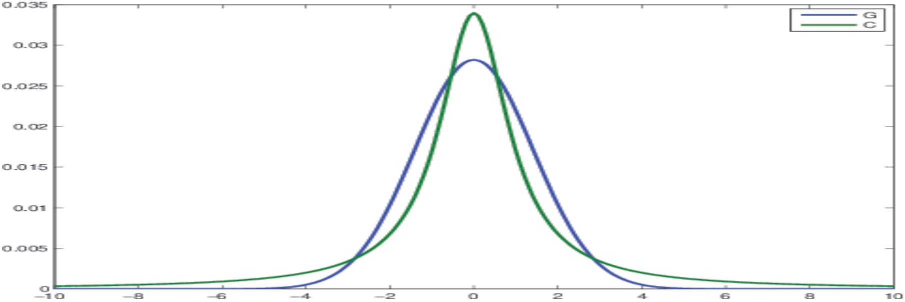
Generalized Gaussian family

Simple heavy tailed models

- Student-t and Cauchy models

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

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

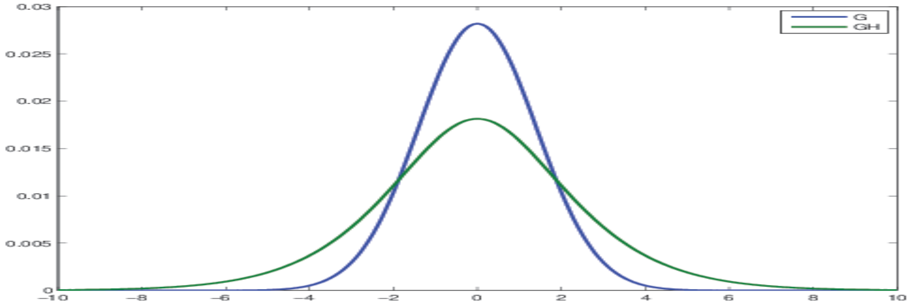


Student-t and Cauchy families

Simple heavy tailed models

- Generalized hyperbolic (GH) models

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



Generalized hyperbolic family

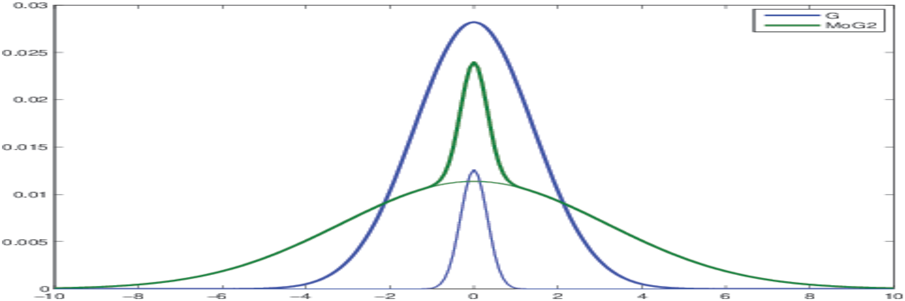
Mixture models

- Mixture of two Gaussians (MoG2) model

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

- Bernoulli-Gaussian (BG) model

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



Mixture of 2 Gaussians families

- Mixture of Gammas

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

- Bernoulli-Gamma model

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

- Mixture of Dirichlets model

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

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

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

- Bernoulli-Multinomial (BMultinomial) model

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

Hierarchical models and hidden variables

- All the mixture models can be modelled via hidden variables $\mathbf{z}$.

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

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

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

  Bernouilli-Gaussian model:

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

$$\begin{cases} p(\mathbf{f}|\mathbf{z}) = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}\left(f_j|0, v_{z_j}\right) \propto \exp\left[-\frac{1}{2}\sum_j \frac{f_j^2}{v_{z_j}}\right] \\ p(\mathbf{z}) = \lambda^{n_1}(1-\lambda)^{n_0}, \quad n_1 = \sum_j \delta(z_j - 1), \quad n_0 = \sum_j \delta(z_j) \end{cases}$$

Hierarchical models and hidden variables

- Example 2: Student-t model

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

- and its Infinite Gaussian Scaled Mixture IGSM model:

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

$$p(f|z)=\mathcal{N}(f|0,1/z), \quad p(z)=\mathcal{G}(z|\alpha,\beta)$$

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

Hierarchical models and hidden variables

$$p(f_j|z_j) = \mathcal{N}(f|0, 1/z_j), \quad p(z_j) = \mathcal{G}(z_j|\alpha, \beta)$$

More general hierarchical models:
(BG, IGSM, Gauss-Markov-Potts)

$$\begin{cases} p(\mathbf{g}|\mathbf{f}) & = \mathcal{N}(\mathbf{g}|\Phi\mathbf{f}, \sigma_\epsilon^2\mathbf{I}) \\ p(\mathbf{f}|\mathbf{z}) & = \prod_j p(f_j|z_j) \text{ or Markovian } \\ p(\mathbf{z}) & = \prod_j p(z_j) \text{ or Markovian (Potts)} \end{cases} \rightarrow p(\mathbf{f},\mathbf{z}|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f})p(\mathbf{f}|\mathbf{z})p(\mathbf{z})$$

- With Hyperparameters $\boldsymbol{\theta}$ we have:
  - Simple priors

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

  - Hierarchical priors

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

Bayesian Computation and Algorithms

- When the expression of $p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g})$ or of $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\mathbf{g})$ is obtained, we have following options:
- Joint MAP: (needs optimization algorithms)

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

- 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 $p(\boldsymbol{\theta}|\mathbf{f}, \mathbf{z}, \mathbf{g})$
- Variational Bayesian Approximation (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.

Joint MAP

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

- Objective:

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

- Alternate optimization:

$$\begin{cases} \widehat{\mathbf{f}} = \arg\max_{\mathbf{f}} \left\{p(\mathbf{f},\widehat{\theta}|\mathbf{g})\right\} \\ \widehat{\theta} = \arg\max_{\theta} \left\{p(\widehat{\mathbf{f}},\theta|\mathbf{g})\right\} \end{cases}$$

$$\begin{array}{ccccc} \theta^{(0)} \longrightarrow \widehat{\theta} \longrightarrow & \boxed{\widehat{\mathbf{f}} = \arg\max_{\mathbf{f}} \left\{p(\mathbf{f}|\widehat{\theta},\mathbf{g})\right\}} & \longrightarrow \widehat{\mathbf{f}} \\ \uparrow & & \downarrow \\ \widehat{\theta} \longleftarrow & \boxed{\widehat{\theta} = \arg\max_{\theta} \left\{p(\theta|\widehat{\mathbf{f}},\mathbf{g})\right\}} & \longleftarrow \widehat{\mathbf{f}} \end{array}$$

- Uncertainties are not propagated.

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 (Gibbs Sampling):

- Generate samples from the conditionals:

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

- Waite for convergency
- Compute empirical statistics (means, variances, modes, medians)
  from the samples $\{\mathbf{f}^{m+1}, \cdots, \mathbf{f}^{m+N}\}$

$$\widehat{\mathbf{f}} = \mathrm{E}\{\mathbf{f}\} \approx \frac{1}{N} \sum_{n=m+1}^{m+N} \mathbf{f}^{(n)}$$

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 $\mathrm{KL}(q(\mathbf{f},\boldsymbol{\theta}|\mathbf{g}) : p(\mathbf{f},\boldsymbol{\theta}|\mathbf{g}))$

$$\mathrm{KL}(q:p) = \iint q \ln \frac{q}{p} = \iint q_1 q_2 \ln \frac{q_1 q_2}{p}$$

- Iterative algorithm $q_1 \longrightarrow q_2 \longrightarrow q_1 \longrightarrow q_2, \cdots$

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

$$\begin{array}{ccccc} \widehat{q}_2^{(0)} \longrightarrow \widehat{q}_2 \longrightarrow & \boxed{q_1(\mathbf{f}) \propto \exp\left[\langle \ln p(\mathbf{g},\mathbf{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_2}\right]} & \longrightarrow \widehat{q}_1(\mathbf{f}) \longrightarrow \widehat{\mathbf{f}} \\ \uparrow & & \downarrow \\ \widehat{\boldsymbol{\theta}} \leftarrow \widehat{q}_2(\boldsymbol{\theta}) \longleftarrow & \boxed{q_2(\boldsymbol{\theta}) \propto \exp\left[\langle \ln p(\mathbf{g},\mathbf{f},\boldsymbol{\theta};\mathcal{M})\rangle_{q_1}\right]} & \longleftarrow \widehat{q}_1 \end{array}$$

- Uncertainties are propagated (Message Passing methods)

Summary of Bayesian approach

- Simple priors

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

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

$$\begin{array}{ccccccc} p(\boldsymbol{\theta}_2|\alpha_2,\beta_2) & & p(\boldsymbol{\theta}_1|\alpha_1,\beta_1) & & & & \\ \downarrow & & \downarrow & & & & \\ \boxed{p(\mathbf{f}|\boldsymbol{\theta}_2)} & \diamond & \boxed{p(\mathbf{g}|\mathbf{f},\boldsymbol{\theta}_1)} & \longrightarrow & \boxed{p(\mathbf{f},\boldsymbol{\theta}|\mathbf{g},\boldsymbol{\alpha},\boldsymbol{\beta})} & \longrightarrow & \boxed{\begin{array}{c}\text{JMAP}\\ \text{MCMC}\\ \text{VBA}\end{array}} \begin{array}{l}\longrightarrow \widehat{\mathbf{f}}\\ \longrightarrow \widehat{\boldsymbol{\theta}}\end{array} \\ \text{Prior} & & \text{Likelihood} & & \text{Joint Posterior} & & \end{array}$$

- Hierarchical priors

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

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

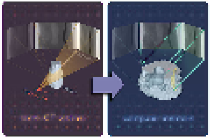
$$\begin{array}{ccccccccc} p(\boldsymbol{\theta}_3|\alpha_3,\beta_3) & & p(\boldsymbol{\theta}_2|\alpha_2,\beta_2) & & p(\boldsymbol{\theta}_1|\alpha_1,\beta_1) & & & & \\ \downarrow & & \downarrow & & \downarrow & & & & \\ \boxed{p(\mathbf{z}|\boldsymbol{\theta}_3)} & \diamond & \boxed{p(\mathbf{f}|\mathbf{z},\boldsymbol{\theta}_2)} & \diamond & \boxed{p(\mathbf{g}|\mathbf{f},\boldsymbol{\theta}_1)} & \longrightarrow & \boxed{p(\mathbf{f},\mathbf{z},\boldsymbol{\theta}|\mathbf{g})} & \longrightarrow & \boxed{\begin{array}{c}\text{JMAP}\\ \text{MCMC}\\ \text{VBA}\end{array}} \begin{array}{l}\longrightarrow \widehat{\mathbf{f}}\\ \longrightarrow \widehat{\mathbf{z}}\\ \longrightarrow \widehat{\boldsymbol{\theta}}\end{array} \\ \text{Hidden variable} & & \text{Prior} & & \text{Likelihood} & & \text{Joint Posterior} & & \end{array}$$

Advantages of the Bayesian Approach

- More possibilities to model sparsity
- More tools to handle hyperparameters
- More tools to account for uncertainties
- More possibilities to understand and to control many ad hoc deterministic algorithms
- Hierarchical models give still more modelling possibilities
  - Bernouilli-Gaussian: strict sparsity
  - Bernouilli-Gamma: strict sparsity + positivity
  - Bernouilli-Multinomial: strict sparsity + discrete values (finite states)
  - Independent Mixture models: sparsity enforcing
  - Mixture of multivariate models: group sparsity enforcing
  - Gauss-Markov-Potts models: indirect sparsity enforcing

Seeing inside of a body: Computed Tomography

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



- Forward model:
  Line integrals or Radon Transform

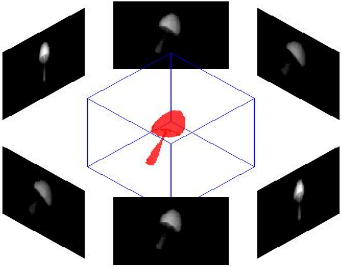
$$
\begin{aligned}
g_\phi(r) &= \int_{L_{r,\phi}} f(x, y)\,\mathrm{d}l + \epsilon_\phi(r) \\
&= \iint f(x, y)\,\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y + \epsilon_\phi(r)
\end{aligned}
$$

- Inverse problem: Image reconstruction

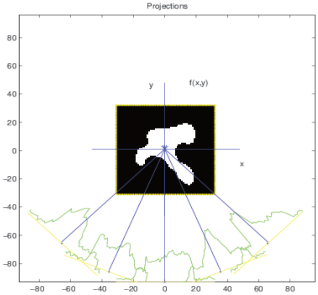
  Given the forward model $\mathcal{H}$ (Radon Transform) and
  a set of data $g_{\phi_i}(r), i = 1, \cdots, M$
  find $f(x, y)$

2D and 3D Computed Tomography

3D



2D



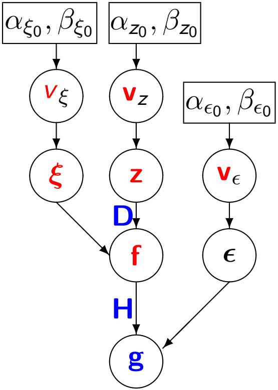
$$g_\phi(r_1, r_2) = \int_{\mathcal{L}_{r_1, r_2, \phi}} f(x, y, z) \,\mathrm{d}l \qquad g_\phi(r) = \int_{\mathcal{L}_{r, \phi}} f(x, y) \,\mathrm{d}l$$

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

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

Inverse Problems with:
non stationary noise and sparse dictionary prior

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}, \quad \mathbf{f} = \mathbf{D}\mathbf{z} + \boldsymbol{\xi}, \quad \mathbf{z} \text{ sparse}$$



$$\begin{cases} p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, \mathbf{V}_\epsilon), & \mathbf{V}_\epsilon = \text{diag}[\mathbf{v}_\epsilon] \\ p(\mathbf{f}|\mathbf{z}) = \mathcal{N}(\mathbf{f}|\mathbf{D}\mathbf{z}, v_\xi \mathbf{I}), & \\ p(\mathbf{z}|\mathbf{v}_z) = \mathcal{N}(\mathbf{z}|0, \mathbf{V}_z), & \mathbf{V}_z = \text{diag}[\mathbf{v}_z] \end{cases}$$

$$\begin{cases} p(\mathbf{v}_\epsilon) = \prod_i \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \\ p(\mathbf{v}_z) = \prod_i \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0}) \\ p(v_\xi) = \mathcal{IG}(v_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) \end{cases}$$

$$p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_z|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon)\, p(\mathbf{f}|\mathbf{z}_f)\, p(\mathbf{z}|\mathbf{v}_z)$$
$$p(\mathbf{v}_\epsilon)\, p(\mathbf{v}_z)\, p(v_\xi)$$

JMAP:

$$(\widehat{\mathbf{f}}, \widehat{\mathbf{z}}, \widehat{\mathbf{v}}_\epsilon, \widehat{\mathbf{v}}_z, \widehat{v}_\xi) = \arg\max_{(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_z, v_\xi)} \{p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_z, v_\xi|\mathbf{g})\}$$

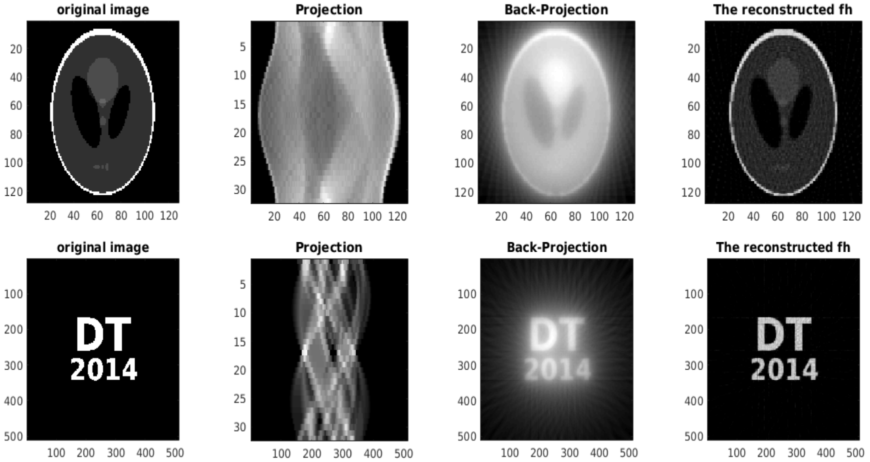
Alternate optimization.

VBA: Approximate

$p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_z, v_\xi|\mathbf{g})$ by $q_1(\mathbf{f})\, q_2(\mathbf{z})\, q_3(\mathbf{v}_\epsilon)\, q_4(\mathbf{v}_z)\, q_5(v_\xi)$

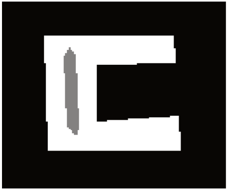
Alternate optimization.

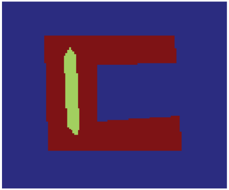
Results

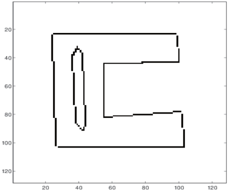


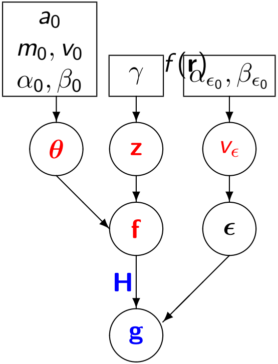
Original $\mathbf{f}$ Data $\mathbf{g}$ Initial. BP $\widehat{\mathbf{f}}$

Gauss-Markov-Potts prior models for images









$$\begin{cases} \mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon \\ p(\mathbf{g}|\mathbf{f}, v_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, v_\epsilon \mathbf{I}) \quad z(\mathbf{r}) \quad c(\mathbf{r}) = 1 - \delta(z(\mathbf{r}) - z(\mathbf{r}')) \\ p(v_\epsilon) = \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \end{cases}$$

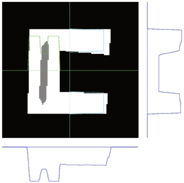
$$\begin{cases} p(f(\mathbf{r})|z(\mathbf{r}) = k, m_k, v_k) = \mathcal{N}(f(\mathbf{r})|m_k, v_k) \\ p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}) = \sum_k \prod_{\mathbf{r}\in\mathcal{R}_k} a_k \mathcal{N}(f(\mathbf{r})|m_k, v_k), \\ \qquad \boldsymbol{\theta} = \{(a_k, m_k, v_k), k = 1, \cdots, K\} \\ p(\boldsymbol{\theta}) = \mathcal{D}(\mathbf{a}|a_0)\mathcal{N}(\mathbf{a}|m_0, v0)\mathcal{IG}(\mathbf{v}|\alpha_0, \beta_0) \\ p(\mathbf{z}|\gamma) \propto \exp\left[\gamma \sum_{\mathbf{r}} \sum_{\mathbf{r}'\in\mathcal{N}(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{r}'))\right] \quad \text{Potts} \end{cases}$$

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

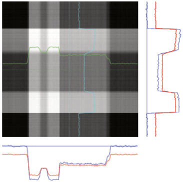
MCMC: Gibbs Sampling

VBA: Alternate optimization.

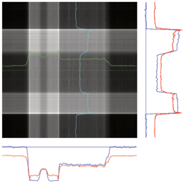
Results



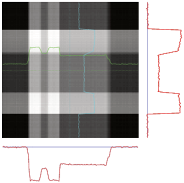
Original



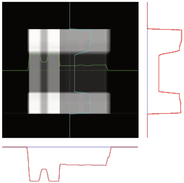
Backprojection



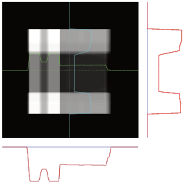
Filtered BP



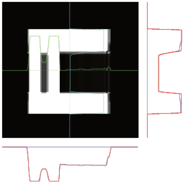
LS



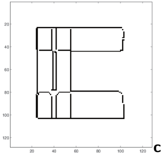
Gauss-Markov+pos

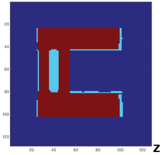


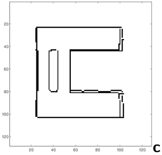
GM+Line process



GM+Label process





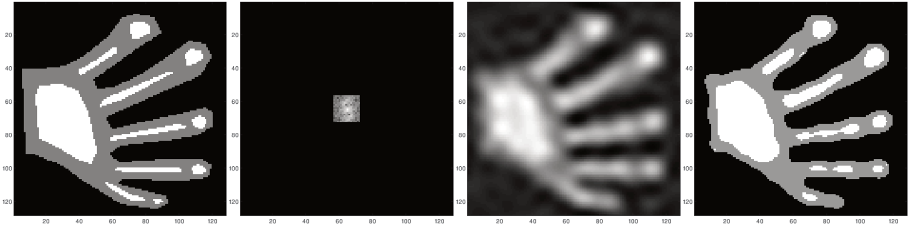


Application in Microwave imaging

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

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

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



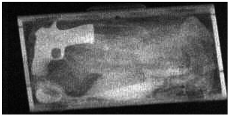
$f(x,y)$ $g(u,v)$ $\widehat{\mathbf{f}}$ IFT $\widehat{\mathbf{f}}$ Proposed method

Images fusion and joint segmentation

(with O. Féron)

$$\begin{cases} g_i(\mathbf{r}) = f_i(\mathbf{r}) + \epsilon_i(\mathbf{r}) \\ p(f_i(\mathbf{r})|z(\mathbf{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2) \\ p(\underline{\mathbf{f}}|\mathbf{z}) = \prod_i p(\mathbf{f}_i|\mathbf{z}) \end{cases}$$

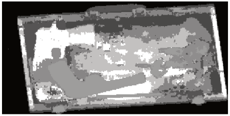
$\mathbf{g}_1$



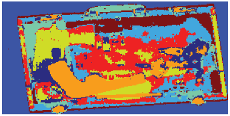
$\mathbf{g}_2$

$\longrightarrow$

$\widehat{\mathbf{f}}_1$



$\widehat{\mathbf{f}}_2$

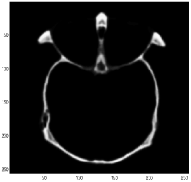


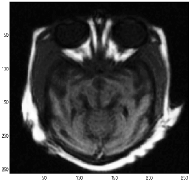
$\widehat{\mathbf{z}}$

Data fusion in medical imaging

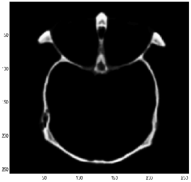
(with O. Féron)

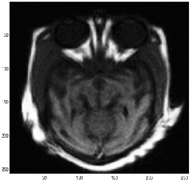
$$\begin{cases} g_i(\mathbf{r}) = f_i(\mathbf{r}) + \epsilon_i(\mathbf{r}) \\ p(f_i(\mathbf{r})|z(\mathbf{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma^2_{i\,k}) \\ p(\underline{\mathbf{f}}|\mathbf{z}) = \prod_i p(\mathbf{f}_i|\mathbf{z}) \end{cases}$$

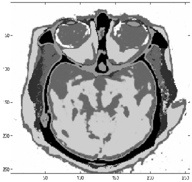
$\mathbf{g}_1$ 

$\mathbf{g}_2$ 

$\longrightarrow$

$\widehat{\mathbf{f}}_1$ 

$\widehat{\mathbf{f}}_2$ 

$\widehat{\mathbf{z}}$ 

Conclusions

- Sparsity: a great property to use in signal and image processing
- Origine: Sampling theory and reconstruction, modeling and representation Compressed Sensing, Approximation theory
- Deterministic Algorithms: Optimization of a two termes criterion, penalty term, regularization term
- Probabilistic: Bayesian approach
- Sprasity enforcing priors: Simple heavy tailed and Hierarchical with hidden variables.
- Gauss-Markov-Potts models for images incorporating hidden regions and contours
- Main Bayesian computation tools: JMAP, MCMC and VBA
- Application in different imaging system (X ray CT, Microwaves, PET, ultrasound and microwave imaging)

Current Projects:

- Efficient implementation in 2D and 3D cases
- Comparison between MCMC and VBA methods

Thanks to:

Present PhD students:

- Camille Chapdelaine (3D Tomography for industrial NDT)
- Guillaume Rvillon (Uncertainty in Radar signal processing and classification)
- Li Wang (Computed Tomography: 2D and 3D)

Freshly Graduated PhD students:

- M. Dumitru (Multivariate time series analysis for biological signals)
- L. Gharsali (Microwave imaging for Cancer detection)
- C. Cai (2013: Multispectral X ray Tomography)
- N. Chu (2013: Acoustic sources localization)
- Th. Boulay (2013: Non Cooperative Radar Target Recognition)
- R. Prenon (2013: Proteomic and Masse Spectrometry)
- Sh. Zhu (2012: SAR Imaging)
- D. Fall (2012: Emission Positon Tomography, Non Parametric Bayesian)
- D. Pougaza (2011: Copula and Tomography)
- H. Ayasso (2010: Optical Tomography, Variational Bayes)

Older Graduated PhD students:

- S. Fékih-Salem (2009: 3D X ray Tomography)
- N. Bali (2007: Hyperspectral imaging)
- O. Féron (2006: Microwave imaging)
- F. Humblot (2005: Super-resolution)
- M. Tahir (2005: Image separation in Wavelet domain)

Thanks to:

Older Graduated PhD students:

- H. Snoussi (2003: Sources separation)
- Ch. Soussen (2000: Geometrical Tomography)
- G. Montémont (2000: Detectors, Filtering)
- H. Carfantan (1998: Microwave imaging)
- S. Gautier (1996: Gamma ray imaging for NDT)
- M. Nikolova (1994: Piecewise Gaussian models and GNC)
- D. Prémel (1992: Eddy current imaging)

Post-Docs:

- J. Lapuyade (2011: Dimentionality Reduction and multivariate analysis)
- S. Su (2006: Color image separation)
- A. Mohammadpour (2004-2005: HyperSpectral image segmentation)

Colleagues:

- B. Duchêne & A. Joisel (L2S)& G. Perruson (Inverse scattering and Microwave Imaging)
- N. Gac (L2S) (GPU Implementation)
- Th. Rodet (L2S) (Computed Tomography)

Thanks to:

National Collaborators

- A. Vabre & S. Legoupil (CEA-LIST), (3D X ray Tomography)
- E. Barat (CEA-LIST) (Positon Emission Tomography, Non Parametric Bayesian)
- C. Comtat (SHFJ, CEA) (PET, Spatio-Temporal Brain activity)
- J. Picheral (SSE, Supélec) (Acoustic sources localization)
- D. Blacodon (ONERA) (Acoustic sources separation)
- J. Lagoutte (Thales Air Systems) (Non Cooperative Radar Target Recognition)
- P. Grangeat (LETI, CEA, Grenoble) (Proteomic and Masse Spectrometry)
- F. Lévi (CNRS-INSERM, Hopital Paul Brousse) (Biological rythms and Chronotherapy of Cancer)

International Collaborators

- K. Sauer (Notre Dame University, IN, USA) (Computed Tomography, Inverse problems)
- F. Marvasti (Sharif University), (Sparse signal processing)
- M. Aminghafari (Amir Kabir University) (Independent Components Analysis)
- A. Mohammadpour (AKU) (Statistical inference)
- Gh. Yari (Tehran Technological University) (Probability and Analysis)

References 1

- A. Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," EURASIP Journal on Advances in Signal Processing, vol. Special issue on Sparse Signal Processing, (2012).
- A. Mohammad-Djafari (Ed.) Problèmes inverses en imagerie et en vision (Vol. 1 et 2), Hermes-Lavoisier, Traité Signal et Image, IC2, 2009,
- A. Mohammad-Djafari (Ed.) Inverse Problems in Vision and 3D Tomography, ISTE, Wiley and sons, ISBN: 9781848211728, December 2009, Hardback, 480 pp.
- A. Mohammad-Djafari, Gauss-Markov-Potts Priors for Images in Computer Tomography Resulting to Joint Optimal Reconstruction and segmentation, International Journal of Tomography & Statistics 11: W09. 76-92, 2008.
- A Mohammad-Djafari, Super-Resolution : A short review, a new method based on hidden Markov modeling of HR image and future challenges, The Computer Journal doi:10,1093/comjnl/bxn005:, 2008.
- H. Ayasso and Ali Mohammad-Djafari Joint NDT Image Restoration and Segmentation using Gauss-Markov-Potts Prior Models and Variational Bayesian Computation, IEEE Trans. on Image Processing, TIP-04815-2009.R2, 2010.
- H. Ayasso, B. Duchene and A. Mohammad-Djafari, Bayesian Inversion for Optical Diffraction Tomography Journal of Modern Optics, 2008.
- N. Bali and A. Mohammad-Djafari, "Bayesian Approach With Hidden Markov Modelling and Mean Field Approximation for Hyperspectral Data Analysis," IEEE Trans. on Image Processing 17: 2. 217-225 Feb. (2008).
- H. Snoussi and J. Idier., "Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures," IEEE Trans. on Signal Processing, 2006.

References 2

- O. Féron, B. Duchène and A. Mohammad-Djafari, Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data, Inverse Problems, 21(6):95-115, Dec 2005.
- M. Ichir and A. Mohammad-Djafari,
  Hidden markov models for blind source separation, IEEE Trans. on Signal Processing, 15(7):1887-1899, Jul 2006.
- F. Humblot and A. Mohammad-Djafari,
  Super-Resolution using Hidden Markov Model and Bayesian Detection Estimation Framework, EURASIP Journal on Applied Signal Processing, Special number on Super-Resolution Imaging: Analysis, Algorithms, and Applications:ID 36971, 16 pages, 2006.
- O. Féron and A. Mohammad-Djafari,
  Image fusion and joint segmentation using an MCMC algorithm, Journal of Electronic Imaging, 14(2):paper no. 023014, Apr 2005.
- H. Snoussi and A. Mohammad-Djafari,
  Fast joint separation and segmentation of mixed images, Journal of Electronic Imaging, 13(2):349-361, April 2004.
- A. Mohammad-Djafari, J.F. Giovannelli, G. Demoment and J. Idier,
  Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems, Int. Journal of Mass Spectrometry, 215(1-3):175-193, April 2002.
- H. Snoussi and A. Mohammad-Djafari, “Estimation of Structured Gaussian Mixtures: The Inverse EM Algorithm,” IEEE Trans. on Signal Processing 55: 7. 3185-3191 July (2007).
- N. Bali and A. Mohammad-Djafari, “A variational Bayesian Algorithm for BSS Problem with Hidden Gauss-Markov Models for the Sources,” in: Independent Component Analysis and Signal Separation (ICA 2007) Edited by:M.E. Davies, Ch.J. James, S.A. Abdallah, M.D. Plumbley. 137-144 Springer (LNCS 4666) (2007).

References 3

- N. Bali and A. Mohammad-Djafari, "Hierarchical Markovian Models for Joint Classification, Segmentation and Data Reduction of Hyperspectral Images" ESANN 2006, September 4-8, Belgium. (2006)
- M. Ichir and A. Mohammad-Djafari, "Hidden Markov models for wavelet-based blind source separation," IEEE Trans. on Image Processing 15: 7. 1887-1899 July (2005)
- S. Moussaoui, C. Carteret, D. Brie and A Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov Chain Monte Carlo methods sampling," Chemometrics and Intelligent Laboratory Systems 81: 2. 137-148 (2005).
- H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images" Journal of Electronic Imaging 13: 2. 349-361 April (2004)
- H. Snoussi and A. Mohammad-Djafari, "Bayesian unsupervised learning for source separation with mixture of Gaussians prior," Journal of VLSI Signal Processing Systems 37: 2/3. 263-279 June/July (2004)
- F. Su and A. Mohammad-Djafari, "An Hierarchical Markov Random Field Model for Bayesian Blind Image Separation," 27-30 May 2008, Sanya, Hainan, China: International Congress on Image and Signal Processing (CISP 2008).
- N. Chu, J. Picheral and A. Mohammad-Djafari, "A robust super-resolution approach with sparsity constraint for near-field wideband acoustic imaging," IEEE International Symposium on Signal Processing and Information Technology pp 286–289, Bilbao, Spain, Dec14-17,2011

Questions, Discussions, Open mathematical problems

- Sparsity representation, low rank matrix decomposition
  - Sparsity and positivity or other constraints
  - Group sparsity
  - Algorithmic and implementation issues for great dimensional applications (Big Data)
  - Joint estimation of Dictionary and coefficients
- Optimization of the KL divergence for Variational Bayesian Approximation
  - Convergency of alternate optimization
  - Other possible algorithms
- Properties of the obtained approximation
  - Does the moments of q's corresponds to the moments of p?
  - How about any other statistics: entropy, ...
- Other divergency or Distance measures?
- Using Sparsity as a prior in Inverse Problems
- Applications in Medical imaging, Non Destructive Testing (NDT) Industrial Imaging, Communication, Geophysical imaging, Radio Astronomy, ...