Variational Bayes and Mean Field Approximations for Markov Field Unsupervised Estimation

Ali Mohammad-Djafari and Hacheme Ayasso

Groupe Problèmes Inverses
Laboratoire des Signaux et Systèmes
UMR 8506 CNRS - SUPELEC - Univ Paris Sud 11
Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette, FRANCE.

djafari@lss.supelec.fr
http://djafari.free.fr
http://www.lss.supelec.fr

MLSP2009, 2-4 September 2009, Grenoble, France

Content

- Gibbs-Markov fields and ML parameter estimation
- Variational energies and entropy
- Markov fields in imaging systems
- ML and MAP parameter estimation
- Variational Bayesian Approximation
- Mean Field Approximation
- Three classes of Markov fields in imaging systems
- Comparison of performances of VBA and MFA and MCMC
- Conclusions and Work in Progress
- Questions and Discussion

Gibbs-Markov fields and ML parameter estimation

- Gibbs-Markov fields :

$$p(\boldsymbol{x}|\lambda) = \frac{1}{Z_p(\lambda)} \exp\left(-\lambda \mathcal{E}(\boldsymbol{x})\right),$$

- Energy :

$$\mathcal{E}(\boldsymbol{x}) = \sum_{c \in \mathcal{C}} \Phi_c(\boldsymbol{x}_c)$$

$\mathcal{C}$ set of cliques, $\Phi_c(.)$ potential

- Partition function :

$$Z_p(\lambda) = \int_{\mathcal{X}} \exp\left(-\lambda \mathcal{E}(\boldsymbol{x})\right) \, \mathrm{d}\boldsymbol{x}$$

- Maximum likelihood (ML) estimation of λ :

$$\widehat{\lambda}_{\mathrm{MV}} = \arg\max_{\lambda} \{\ln p(\boldsymbol{x}|\lambda)\} = \arg\max_{\lambda} \{-\ln Z(\lambda) - \lambda \mathcal{E}(\boldsymbol{x})\}$$

$$\longrightarrow \qquad \frac{-\partial \ln Z(\lambda)}{\partial \lambda} = \mathcal{E}(\boldsymbol{x})$$

Variational energies and entropy

Main objective : Approximate p by q minimizing

- Kullback-Leibler divergence

$$\mathrm{KL}(q|p) = < -\ln \frac{q}{p} >_q = -\int_{\mathcal{X}} q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x},$$

- Variational free energy :

$$F(q) = U(q) - H(q)$$

- Variational average energy :

$$U(q) = < \mathcal{E}(\boldsymbol{x}) >_q = \int_{\mathcal{X}} q(\boldsymbol{x}) \, \mathcal{E}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

- Variational entropy :

$$H(q) = < -\ln q >_q = -\int_{\mathcal{X}} q(\boldsymbol{x}) \ln q(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

Variational energies and entropy

- Two main relations :
  - Variational average energy :

$$U(q) = -\ln Z(\lambda) + < \ln p >_q$$

  - Variational free energy :

$$F(q) = -\ln Z(\lambda) + \mathrm{KL}(q|p) = F_{\text{Helmoltz}} + \mathrm{KL}(q|p).$$

- The main inequality :
  - $F(q) \geq F_{\text{Helmoltz}}$, with equality when $q = p$.
- Main conclusion :
  - Minimizing $F(q)$ is a good way to compute $F_{\text{Helmoltz}} = -\ln Z$ and use it where necessary.

Markov fields in imaging systems

- $\boldsymbol{x} = \{x(\boldsymbol{r}_i), \boldsymbol{r}_i \in \mathcal{R}\}$ represent the pixels of an image $x(\boldsymbol{r})$
- $\boldsymbol{r}_i$ spatial position of the pixel or voxel number i.
- Markov fields considered :

$$\mathcal{E}(\boldsymbol{x}) = \sum_i \sum_{j \in \mathcal{V}i} \Phi_i(x_i, x_j),$$

where

$$\Phi_i(x_i, x_j) = \Phi(x_i - x_j) \; \forall i.$$

- Expression :

$$p(\boldsymbol{x}|\lambda) = \prod_i p\left(x_i | x_j, j \in \mathcal{V}(i)\right)$$

$$p(x_i | x_j, \in \mathcal{V}(i)) \propto \exp\left(-\lambda \sum_{j \in \mathcal{V}(i)} \Phi(x_i - x_j)\right)$$

ML and MAP parameter estimation

- Maximum likelihood (ML) :

$$\begin{aligned}\widehat{\lambda}_{\text{MV}} &= \arg\max_{\lambda} \{\ln p(\boldsymbol{x}|\lambda)\} \\ &= \arg\max_{\lambda} \{-\ln Z(\lambda) - \lambda\mathcal{E}(\boldsymbol{x})\}\end{aligned}$$

$$\longrightarrow \qquad \frac{-\partial \ln Z(\lambda)}{\partial\lambda} = \mathcal{E}(\boldsymbol{x})$$

- Maximum A Posteriori (MAP) :

$$\begin{aligned}\widehat{\lambda} &= \arg\max_{\lambda} \{\ln p(\lambda|\boldsymbol{x})\} \\ &= \arg\max_{\lambda} \{-\ln Z(\lambda) - \lambda\mathcal{E}(\boldsymbol{x}) + \ln \pi(\lambda)\}\end{aligned}$$

$$\longrightarrow \qquad \frac{-\partial \ln \left(Z(\lambda)/\pi(\lambda)\right)}{\partial\lambda} = \mathcal{E}(\boldsymbol{x})$$

Three classes of Markov fields in imaging systems

- Generalized Gaussian (GG) :

$$\mathcal{E}(\boldsymbol{x}) = \sum_i \sum_{j \in \mathcal{V}(i)} \left| x_i - x_j \right|^\beta$$

- Entropic (I-Distribution family) :
  - First kind :

$$\mathcal{E}(\boldsymbol{x}) = \sum_i \sum_{j \in \mathcal{V}(i)} x_j \ln \frac{x_j}{x_i} - (x_j - x_i).$$

  - Second kind :

$$\mathcal{E}(\boldsymbol{x}) = \sum_i \sum_{j \in \mathcal{V}(i)} x_i \ln \frac{x_i}{x_j} - (x_i - x_j).$$

- Potts and Ising models :

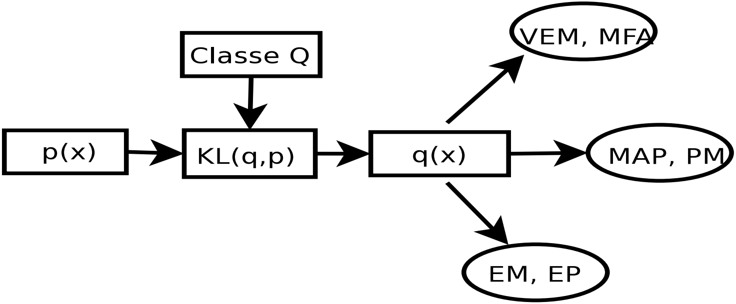
$$\mathcal{E}(\boldsymbol{x}) = -\sum_i \sum_{j \in \mathcal{V}(i)} \delta(x_i - x_j),$$

Variational Approximations Basics

Main objective : Approximate p by simpler $q \in \mathcal{Q}$ minimizing

$$q(\boldsymbol{x}) = \arg\min_{q \in \mathcal{Q}} KL(q : p).$$

and use it for all posterior computations.



- No constraint : $\mathcal{Q} = \{q : \int q = 1\} \longrightarrow q = p$
- Separable : $\mathcal{Q} = \{q : q(\boldsymbol{x}) = \sum_j q_j(x_j)\} \longrightarrow$ VBA
- Parametric : $\mathcal{Q} = \{q : q(\boldsymbol{x}) = q_{\boldsymbol{\theta}}(\boldsymbol{x})\} \longrightarrow$ Parametric VBA
- Separable and parametric $\longrightarrow$ MFA

Variational Bayesian Approximation

- Main objective : Approximate p by a separable

$$q \in \mathcal{Q} = \{q : q(\boldsymbol{x}) = \sum_j q_j(x_j)\}$$

minimizing $\mathrm{KL}(q : p)$.

- The solution has the form :

$$q_i(x_i) = \frac{1}{Z_i(\lambda)} \exp\left(-\lambda \left\langle \mathcal{E}(\boldsymbol{x}) \right\rangle_{\prod_{j \neq i} q_j}\right)$$

- Iterative
- Needs the computation of $\left\langle \mathcal{E}(\boldsymbol{x}) \right\rangle_{\prod_{j \neq i} q_j}$
- Analytic expression if $p(\boldsymbol{x})$ is in Exponential family
- In our application only parametric form is usable.

Mean Field Approximation

Example :
$p(\boldsymbol{x}|\lambda) = \prod_i p\left(x_i | x_j, j \in \mathcal{V}(i)\right)$ with

$$p(x_i | x_j, j \in \mathcal{V}(i)) = \frac{1}{Z_i(\lambda)} \exp\left(-\lambda \sum_{j \in \mathcal{V}(i)} \Phi(x_i - x_j)\right),$$

is approximated by
$q(\boldsymbol{x}|\lambda) = \prod_i q\left(x_i | \bar{x}_j, j \in \mathcal{V}(i)\right)$ with

$$q(x_i | \bar{x}_j, j \in \mathcal{V}(i)) = \frac{1}{Z_i(\lambda)} \exp\left(-\lambda \sum_{j \in \mathcal{V}(i)} \Phi(x_i - \bar{x}_j)\right),$$

$\bar{x}_j$ becomes a parameter to be estimated such that $\mathrm{KL}(q : p)$ be minimized.

VBA = MFA for Gaussian, Entropic, Potts and Ising

- Gaussian Case

$$\ln(q_i(x_i)) \propto -\frac{\lambda}{2} \sum_{j \in \mathcal{V}(i)} \left[x_i^2 - 2x_i\tilde{\mu}_j + \tilde{\mu}_j^2 + \tilde{v}_j \right] \Rightarrow q_i(x_i) = \mathcal{N}(\tilde{\mu}_i, \tilde{v}_i)$$

with

$$\tilde{\mu}_i = \frac{1}{|\mathcal{V}|} \sum_{j \in \mathcal{V}(i)} \tilde{\mu}_j \quad \text{and} \quad \tilde{v}_i = \frac{1}{|\mathcal{V}|\lambda}$$

Bayesian estimation of λ : $\pi(\lambda) = \Gamma(a, b) \longrightarrow p(\lambda|\boldsymbol{x}) = \Gamma(\hat{a}, \hat{b})$
where $\hat{a} = \left[\frac{1}{a} + \sum_{i \in \mathcal{R}} (x_i - \tilde{\mu}_i)^2\right]^{-1}$ and $\hat{b} = \frac{|\mathcal{R}|}{2} + b$.

VBA = MFA for Gaussian, Entropic, Potts and Ising

- Entropic Case 1

$$\ln(q(x_i)) \propto -\lambda \sum_{j\in\mathcal{V}(i)} \exp[x_j]_{q_j} \ln \frac{\exp[x_j]_{q_j}}{x_i} - x_i + \exp[x_j]_{q_j}$$

$$Z_i(\lambda) = e^{-\lambda\tilde{\mu}_i \ln \tilde{\mu}_i + \lambda\tilde{\mu}_i} \lambda^{-\lambda\tilde{\mu}_i - 1} \Gamma(\lambda\tilde{\mu}_i + 1)$$

with $\tilde{\mu}_i = \sum_{j\in\mathcal{V}(i)} \exp[x_j]_{q_j}$.

- Entropic case 2

$$\ln(q(x_i)) \propto -\lambda \sum_{j\in\mathcal{V}(i)} x_i \ln \frac{x_i}{e^{-\exp[\ln x_j]_{q_j}}} + x_i - \exp[x_j]_{q_j}$$

$$Z_i(\lambda) = \frac{e^{-\lambda\tilde{\mu}_i}}{\lambda} \alpha(\lambda\tilde{\mu}_i)$$

with :
$\tilde{\mu}_i = e^{-\sum_{j\in\mathcal{V}(i)} \exp[\ln x_j]_{q_j}}$ and $\alpha(a) = \int_0^\infty a^y e^{-y \ln y + y}\, dy$.

VBA = MFA for Gaussian, Entropic, Potts and Ising

- Ising and Potts

$$q_i(x_i = k) = \frac{\exp\left(-\lambda \sum_{j\in\mathcal{V}(i)} q_j(x_j = k)\right)}{\sum_{k=1}^{K} \mathrm{e}^{-\lambda \sum_{j\in\mathcal{V}(i)} q_j(x_j=k)}}$$

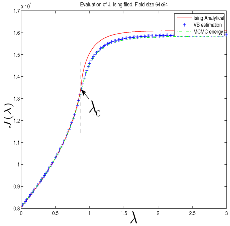
$x_j \in \{1, ..., K\}$ is the general case and $x_j \in \{0, 1\}$ is the Ising model case.

Simulation results : Comparison between VBA, MFA and MCMC

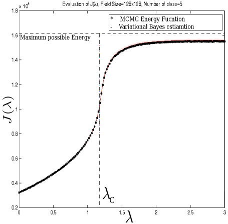
-

Simulations

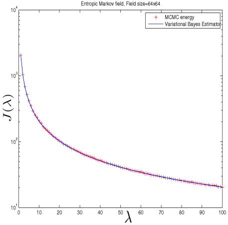
Ising	Pottz	Entropic



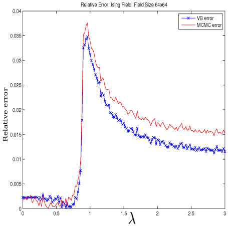
a



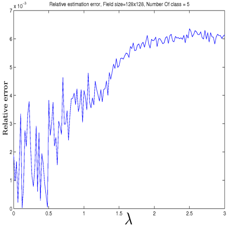
c



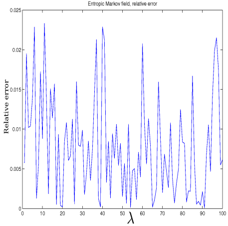
e



b

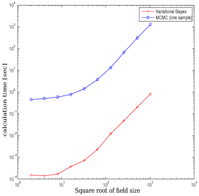


d

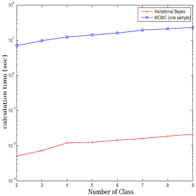


f

Simulations



a



b

Conclusions

- Gauss-Markov-Potts are useful prior models for images incorporating regions and contours
- Bayesian computation needs often pproximations (Laplace, MCMC, Variational Bayes)
- Application in different CT systems (X ray, Ultrasound, Microwave, PET, SPECT) as well as other inverse problems

Work in Progress and Perspectives :

- Efficient implementation in 2D and 3D cases using GPU
- Evaluation of performances and comparison with MCMC methods
- Application to other linear and non linear inverse problems :
  (PET, SPECT or ultrasound and microwave imaging)