EM Algorithm for 2-D Models

$$p(y, x) = p(y|x, \varphi) \frac{1}{Z(\theta)} e^{-u(x|\theta)}$$

- Need to estimate φ, θ from observations of y

$$Q(\theta', \varphi', \theta, \varphi) = E\left[\log p(y, X | \theta', \varphi') | Y=y, \theta, \varphi\right]$$

$$= E\left[\log p(y|X, \varphi') | Y=y, \theta, \varphi\right]$$

$$+ E\left[-u(X|\theta) | Y=y, \theta, \varphi\right] - \log Z(\theta)$$

$$\hat{\theta} = \arg\min_{\theta'} \left\{ E\left[u(X|\theta) | Y=y, \theta\right] + \log Z(\theta') \right\}$$

$$\hat{\varphi} = \arg\min_{\varphi} \left\{ E\left[-\log p(y|X, \varphi') | Y=y, \theta, \varphi\right] \right\}$$

Problem: E step is hard to compute

Note! M step is similar to case when X is known

Approach: Stochastic simmulation

E step: Use use simmulation method to generate samples from $p(x|Y,\theta)$

$$X^{(1)}, \dots, X^{(N)} \sim p(x|Y,\theta)$$

then

$$E[u(X)|Y=y,\theta] \cong \frac{1}{N}\sum_{i=1}^{N} u(X^{(i)}|\theta)$$

M step:

$$\hat{\theta}^{(k+1)} = \arg\min_{\theta} \left\{ E[u(X|\theta)|Y=y,\hat{\theta}^{(k)}] + \log Z(\theta) \right\}$$

$$\cong \arg\min_{\theta} \left\{ \frac{1}{N}\sum_{i=1}^{N} u(X^{(i)}|\theta) + \log Z(\theta) \right\}$$

If $N=1$ then

$$\hat{\theta}^{(k+1)} \cong \arg\min_{\theta} \left\{ u(X^{(1)}|\theta) + \log Z(\theta) \right\}$$

$$\hat{\varphi} = \underset{\varphi}{\operatorname{argmin}} \left\{ \frac{1}{N} \sum_{i=1}^{N} -\log p(y \mid x^{(i)}, \varphi) \right\}$$

For $N=1$

$$\hat{\varphi} = \underset{\varphi}{\operatorname{argmin}} \left\{ -\log p(y \mid x^{(1)}, \varphi) \right\}$$

Example 1)

$$X \sim \frac{1}{Z(1)\sigma^N} \exp\left\{ -\frac{1}{p\sigma^p} u(X) \right\}$$

where $u(\alpha X) = \alpha^p u(X) \quad \forall \alpha > 0$

Y_s are conditionally idependent given X with $\sim N(X_s, \gamma)$

$$p(y|x) = \left(\prod_{s \in S} \frac{1}{\sqrt{2\pi\gamma}} \right) \exp\left\{ -\frac{1}{2\gamma} \sum_{s \in S} (Y_s - X_s)^2 \right\}$$

$$= \frac{1}{Z'} \exp\left\{ -\frac{1}{2\gamma} \|Y - X\|^2 \right\}$$

$$p(y, x|\sigma) = \frac{1}{Z'} \exp\left\{ -\frac{1}{2\gamma} \|Y - X\|^2 \right\} \cdot \frac{1}{Z(1)\sigma^N} \exp\left\{ -\frac{1}{p\sigma^p} u(x) \right\}$$

$$p(x|y, \sigma) = \frac{p(y, x|\sigma)}{p(y|\sigma)}$$

$$= \underbrace{\frac{1}{Z''}}_{\text{function of } y \text{ and } \sigma} \exp \underbrace{\left\{ -\frac{1}{2\gamma} \|Y - X\|^2 - \frac{1}{p\sigma^p} u(x) \right\}}_{\text{Gibbs distribution with same neighborhood as } X \leftarrow MRF}$$

Comments

1) Can generate sample from $p(x|y,\sigma)$ using Metropolis algorithm

2) Warning! More than one Metropolis algorithm. Some are <u>much</u> faster than others!!

Remember:

$$\hat{\sigma}_{ML} = \arg\min_{\sigma} \left\{ \frac{1}{p\sigma^p} U(x) + \log z(1) + N \log \sigma \right\}$$

$$= \left(\frac{1}{N} U(x) \right)^{1/p}$$

For this example, and $X^{(1)}, \ldots, X^{(M)} \sim p(x|y,\sigma^{(k)})$

$$Q'(\sigma, \sigma^{(k)}) \cong -\frac{1}{p\sigma^p} \sum_{i=1}^{M} U(X^{(i)}) - N \log \sigma$$

$$\hat{\sigma}^{(k+1)} = \arg\min_{\sigma} \left\{ -\frac{1}{Mp\sigma^p} \sum_{i=1}^{M} U(X^{(i)}) - N \log \sigma \right\}$$

$$\hat{\sigma}^{(k+1)} = \left(\frac{1}{MN} \sum_{i=1}^{M} U(X^{(i)}) \right)^{1/p}$$

Comment

1. A "reasonable" approach to estimating σ is

$$\hat{X}^{(k)}_{MAP} = \arg\max_{X} p(X, Y \mid \hat{\sigma}^{(k)})$$

$$\hat{\sigma}^{(k+1)} = \arg\max_{\sigma} p(\hat{X}^{(k)}, Y \mid \sigma)$$

this is moves toward to solution

$$(\hat{X}, \hat{\sigma}) = \arg\max_{X, \sigma} p(Y, X \mid \sigma)$$

2. Warning! This solution is not consistent
   the solution may not converge!!!

too smooth σ $\Rightarrow$ too smooth X $\Rightarrow$ too smooth σ $\Rightarrow$ ...

Example 2)

$$p(X \mid T, p) = \frac{1}{Z(T,p)} \exp\left\{-\frac{1}{T} U(X,p)\right\}$$

$$U(\alpha X, p) = \alpha^p U(X,p)$$

$$Z(p\sigma^p, p) = \sigma^N p^{N/p} Z(1,p)$$

$$(\hat{\sigma}, \hat{p}) = \arg\min_{(\sigma,p)} \left\{ \frac{1}{p\sigma^p} U(X,p) + N \log\sigma + \frac{N}{p}\log p + \log Z(1,p) \right\}$$

$$= \arg\min_{(\sigma,p)} \left\{ \frac{1}{Np\sigma^p} U(X,p) + \log\sigma + \frac{\log p}{p} + \underbrace{\frac{\log Z(1,p)}{N}}_{f(p)} \right\}$$

$$f(p) \triangleq \frac{1}{N} \log Z(1,p)$$

$$\hat{\sigma}(x,p) = \arg\min_{\sigma} \left\{ \frac{1}{Np\sigma^p} U(X,p) + \log\sigma + \frac{\log p}{p} + f(p) \right\}$$

$$\hat{\sigma}(x,p) = \left(\frac{1}{N} U(x,p) \right)^{1/p}$$

M-step

$$\hat{p}(x) = \arg\min_{p} \left\{ \log\hat{\sigma}(x,p) + \frac{1}{p} + \frac{\log p}{p} + f(p) \right\}$$

$$\hat{\sigma}(x) = \hat{\sigma}(x, \hat{p}(x))$$

How do you compute $f(\rho)$?

$$f(\rho) = \frac{1}{N} \log z(1,\rho)$$

$$= \frac{1}{N} \log \int_{x \in \mathbb{R}^N} \exp\{-u(x,\rho)\} dx$$

$$\frac{d f(\rho)}{d\rho} = \frac{1}{N} \frac{1}{z(1,\rho)} \int_{x \in \mathbb{R}^N} -u(x,\rho) \exp\{-u(x,\rho)\} dx$$

$$= \underbrace{\frac{-1}{N} E\left[u(x,\rho) \,\middle|\, \sigma=1, \rho \right]}$$

evaluates this expectation using stochastic simmulation

$$x^{(1)}, \ldots, x^{(M)} \sim p(x \mid \sigma=1, \rho)$$

$$f'(\rho) = \frac{-1}{NM} \sum_{i=1}^{M} u(x^{(i)}, \rho)$$

Compute $f(\rho) = \int_{\rho_0}^{\rho} f'(t)\, dt + f(\rho_0)$

don't care about constant.

not explore these other alternatives to MAP estimation.

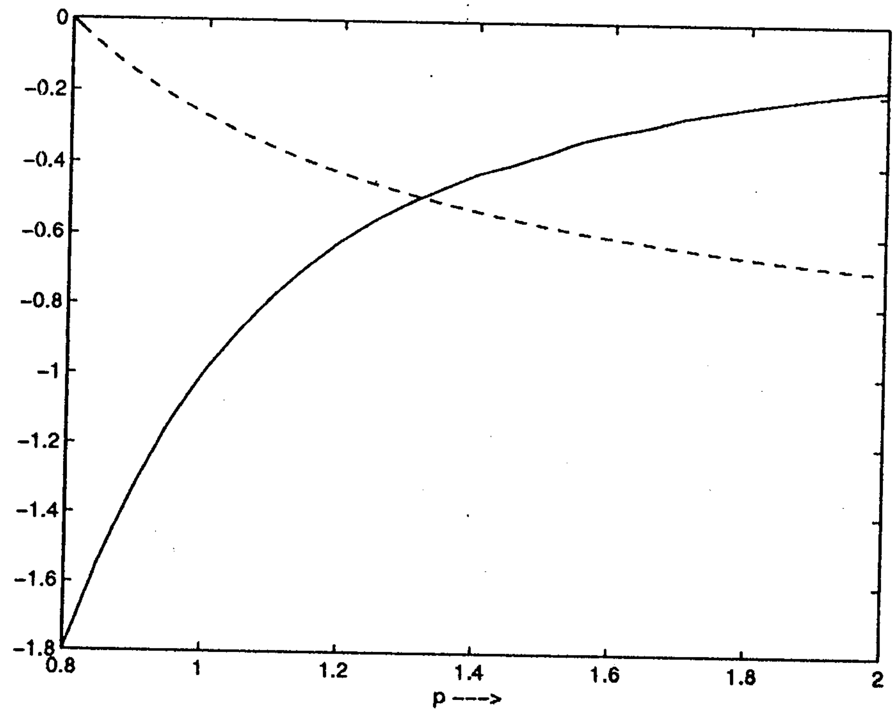


Figure 2: The solid line shows $f'(p)$ and the dashed line shows $f(p)$.

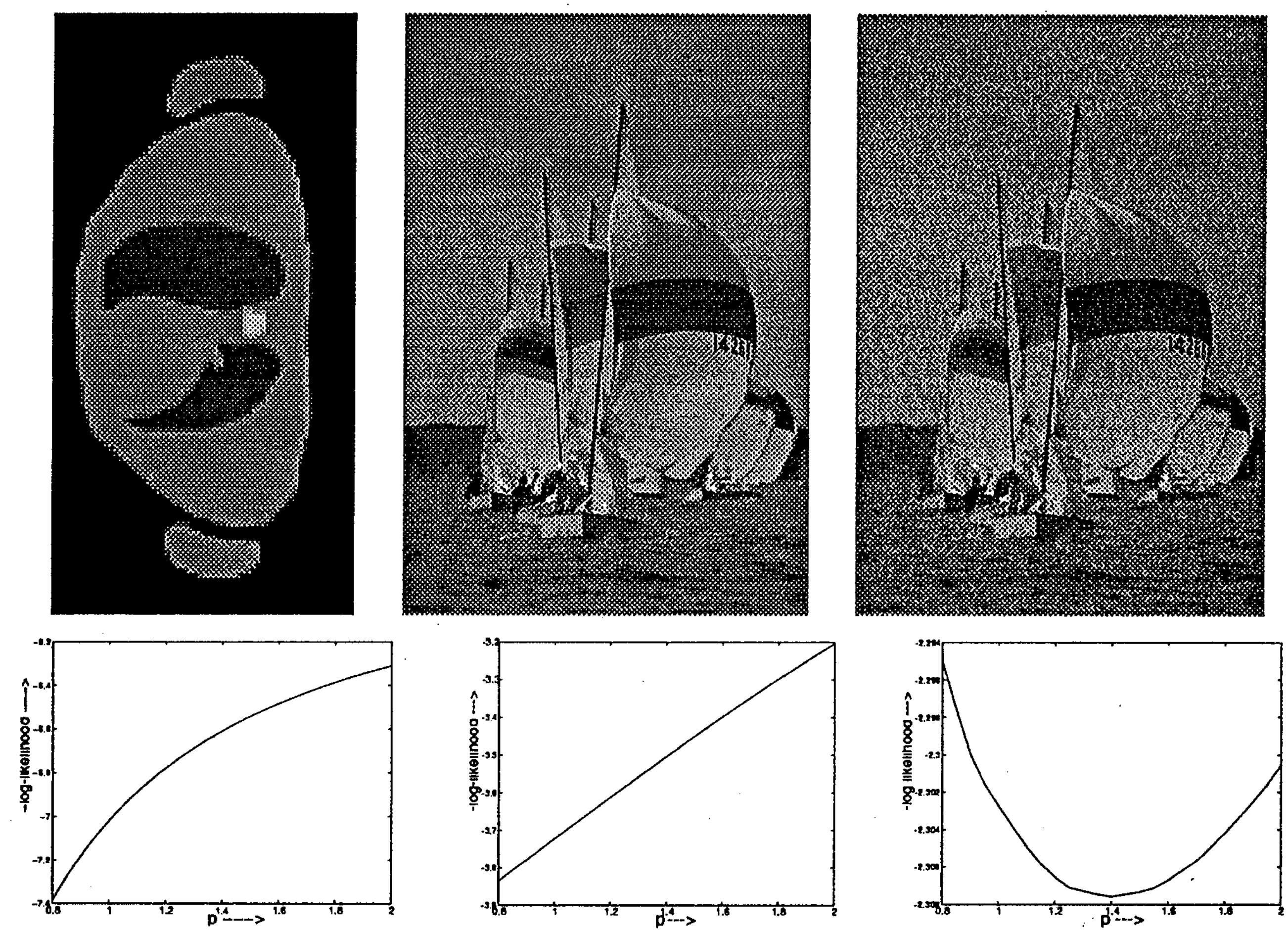
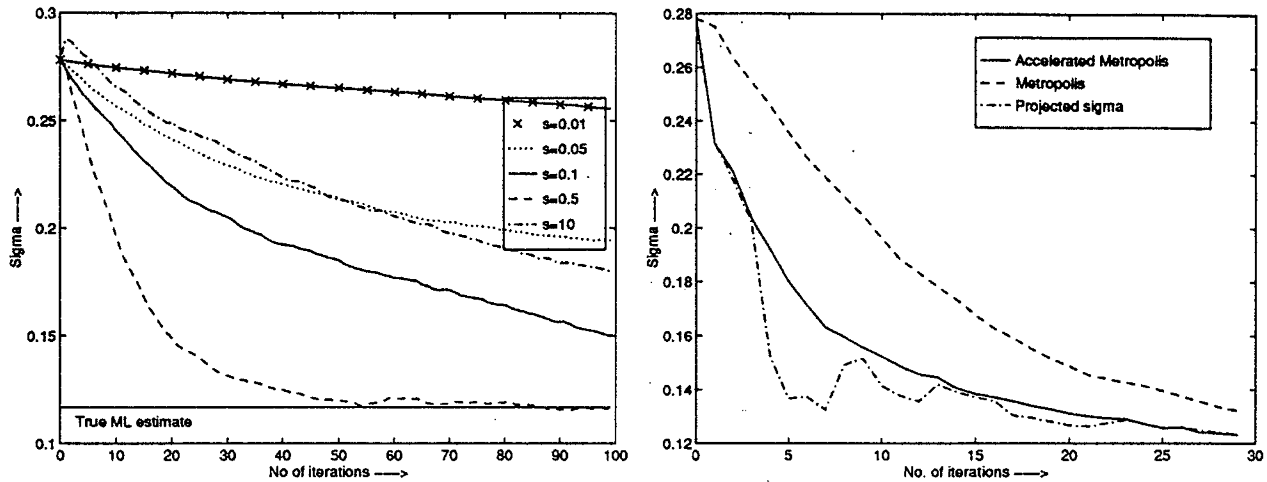

(a) (b) (c)

Figure 3: ML estimation of p for (a) transmission phantom (b) natural image (c) image (b) corrupted with Gaussian noise. The plot below each image shows the corresponding negative log-likelihood as a function of p. The ML estimate is the value of p that minimizes the plotted function.



(a) (b)

Figure 7: The above plots show the EM updates for σ for the emission phantom modeled by a GGMRF prior ($p = 1.1$) using (a) conventional Metropolis (CM) method, (b) accelerated Metropolis (AM) and the extrapolation method. The parameter s denotes the standard deviation of the symmetric transition distribution for the CM method. All the updates are done using a single sample of X to compute the expectation. The true ML estimate is the converged value of σ when 50 samples are used to compute the expectation.

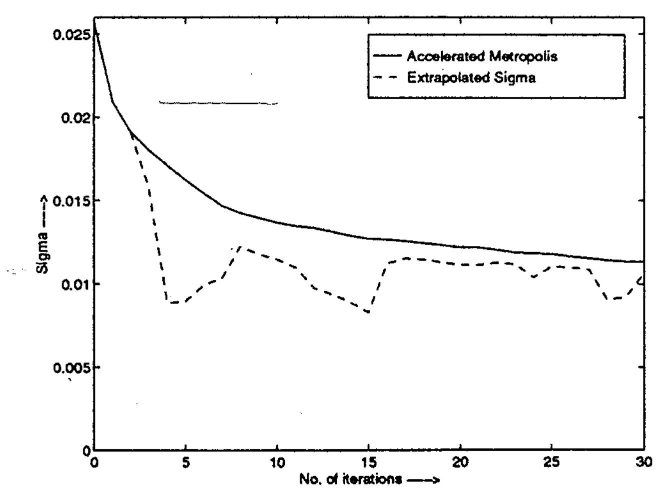


Figure 8: The above plots shows the EM updates for σ using the accelerated Metropolis method and the extrapolated value of σ for the emission phantom using the logcosh$(\cdot)$ prior with $T = 10$.

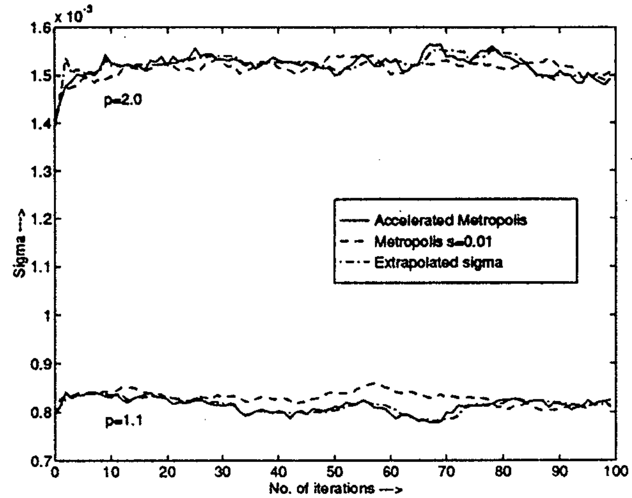
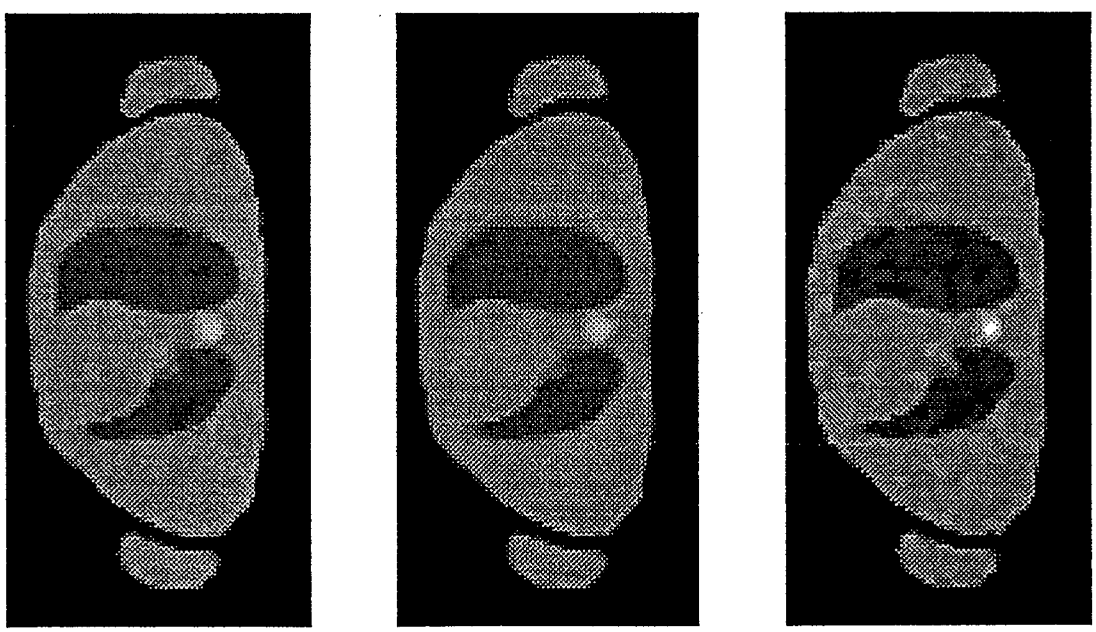


Figure 9: The above plots shows the EM updates for σ using the Metropolis method, accelerated Metropolis method, and the extrapolated value of σ for the transmission phantom using the GGMRF prior.



a	b	c

Figure 10: Reconstructed transmission phantom using GGMRF prior with $p = 1.1$ The scale parameter σ is (a) $\hat{\sigma}_{ML} \approx \hat{\sigma}_{CBP}$, (b) $\frac{1}{2}\hat{\sigma}_{ML}$, and (c) $2\hat{\sigma}_{ML}$

Example 3)

X is an Ising model with fixed β

$$p(x) = \frac{1}{z} e^{-\beta \sum_{\{r,s\}} \delta(x_r \neq x_s)}$$

$$p(y \mid x) = \prod_{s \in S} p(y_s \mid x_s)$$

$$p(y_s \mid x_s) \sim N(\mu_{x_s}, \gamma_{x_s})$$

$$\varphi = \{(\mu_i, \gamma_i)\}_{i=1}^{M}$$

$$\log p(y \mid x, \varphi) = \sum_{s \in S} \left\{ \frac{-1}{2\gamma_{x_s}} (y_s - \mu_{x_s})^2 - \frac{1}{2} \log(2\pi \gamma_{x_s}) \right\}$$

$$= \sum_{i=1}^{M} \sum_{s \in S_i} \left\{ -\frac{1}{2\gamma_i} (y_s - \mu_i)^2 - \frac{1}{2} \log(2\pi \gamma_i) \right\}$$

$$S_i = \{ s \in S : x_s = i \}$$

$$\arg\max_{\varphi} \log p(y \mid x, \varphi) =$$

$$N_i = |S_i|$$

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{s \in S_i} y_s$$

$$\hat{\gamma}_i = \frac{1}{N_i} \sum_{s \in S_i} (y_s - \mu_i)^2$$

also

(*) $$p(x|y) = \frac{1}{z} e^{-u(x)}$$

$$u(x) = -\beta \sum_{\{r,s\}} \delta(x_r \neq x_s) + \log p(y_s|x_s)$$

algorithm

1) Generate a sample x from (*)

2) Compute updates

$$N_i = |S_i|$$

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{s \in S_i} y_s$$

$$\hat{\gamma}_i = \frac{1}{N_i} \sum_{s \in S_i} (y_s - \mu_i)^2$$

3) Go to step 1)