Auckland University – Department of Mathematics

http://www.math.auckland.ac.nz/~fox

A comparison of mode and mean in binary image reconstruction

Colin Fox (and thanks to Geoff Nicholls, Andrés Christen, CIMAT)

fox@math.auckland.ac.nz

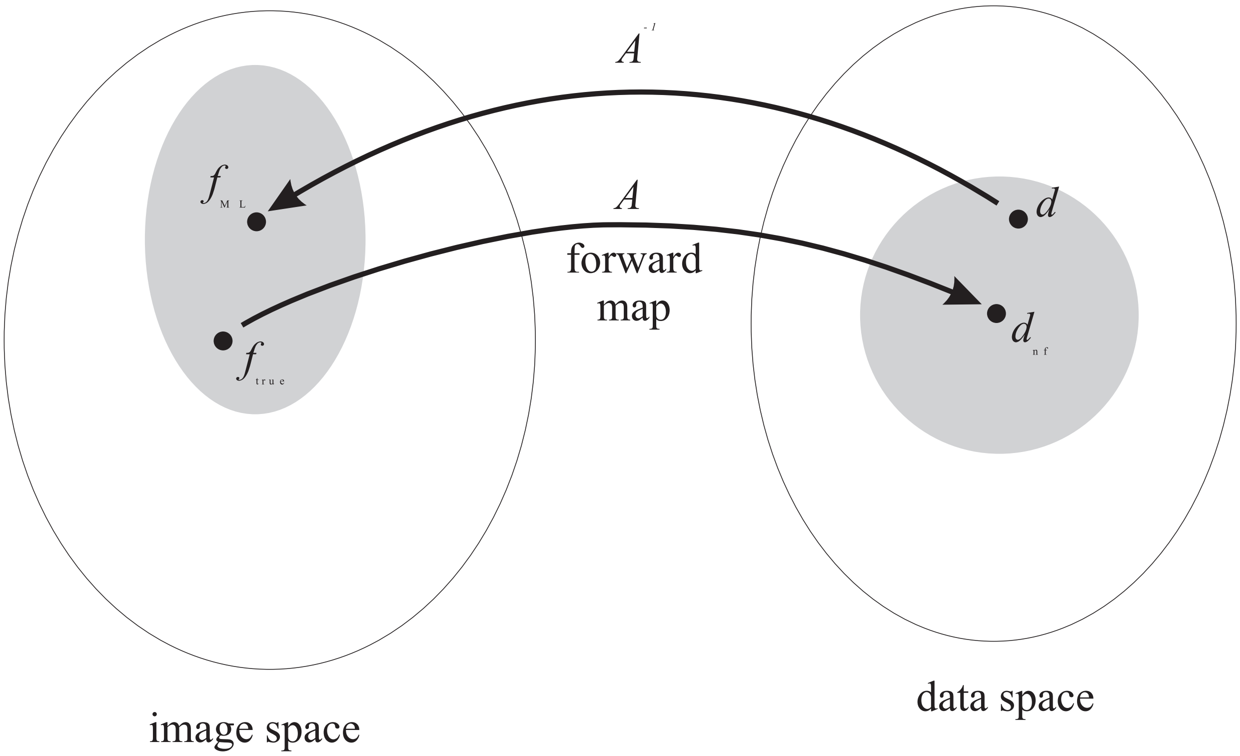
Overview

- Bayesian Image Recovery
- Binary images
- Exact MAP state (GPS)
- Perfect Sampling
- Numerical Example (Mode, Mean, Sample, MPM)

Bayesian Model for Image Recovery

data d, image f, measurement noise n, forward map$A$

$$d = Af + n$$



If $n \sim N(0, s^2)$ then $d \sim N(Af, s^2)$

Given measurements d, the likelihood for f is

$$L_d(f) \equiv \Pr(d|f) \propto \exp(|d - Af|^2/2s^2)$$

Posterior distribution for f conditional on d

$$\Pr(f|d) = \frac{\Pr(d|f)\Pr(f)}{\Pr(d)} \qquad \text{(Bayes rule)}$$

$\Pr(\sigma)$is the prior distribution

Traditional and Statistical Solutions

All information contained in posterior distribution $\Pr(f|d)$

Traditional Solutions - modes

$$\begin{aligned} \hat{f}_{\mathsf{MLE}} &= \arg\max L_d(f) \equiv \arg\max \Pr(d|f) \\ \hat{f}_{\mathsf{MAP}} &= \arg\max \Pr(f|d) \equiv \arg\max \Pr(d|f)\Pr(f) \end{aligned}$$

e.g. simple Gaussian prior: $\Pr(f) \propto \exp\left(-|f|^2/2\lambda^2\right)$

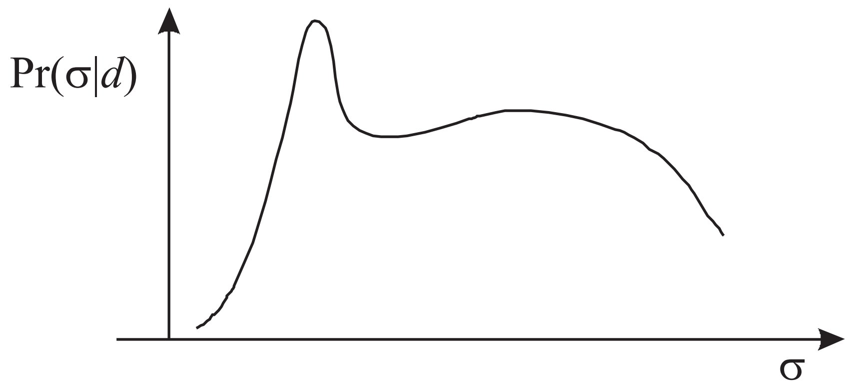
$$\hat{f}_{\mathsf{MAP}} = \arg\min |d - Af|^2 + \alpha |f|^2 \qquad \alpha = s^2/\lambda^2$$

- Tikhonov regularization, Kalman filtering, Backus-Gilbert
- $\alpha \to 0$ Moore-Penrose inverse, $\alpha = 0$ least-squares

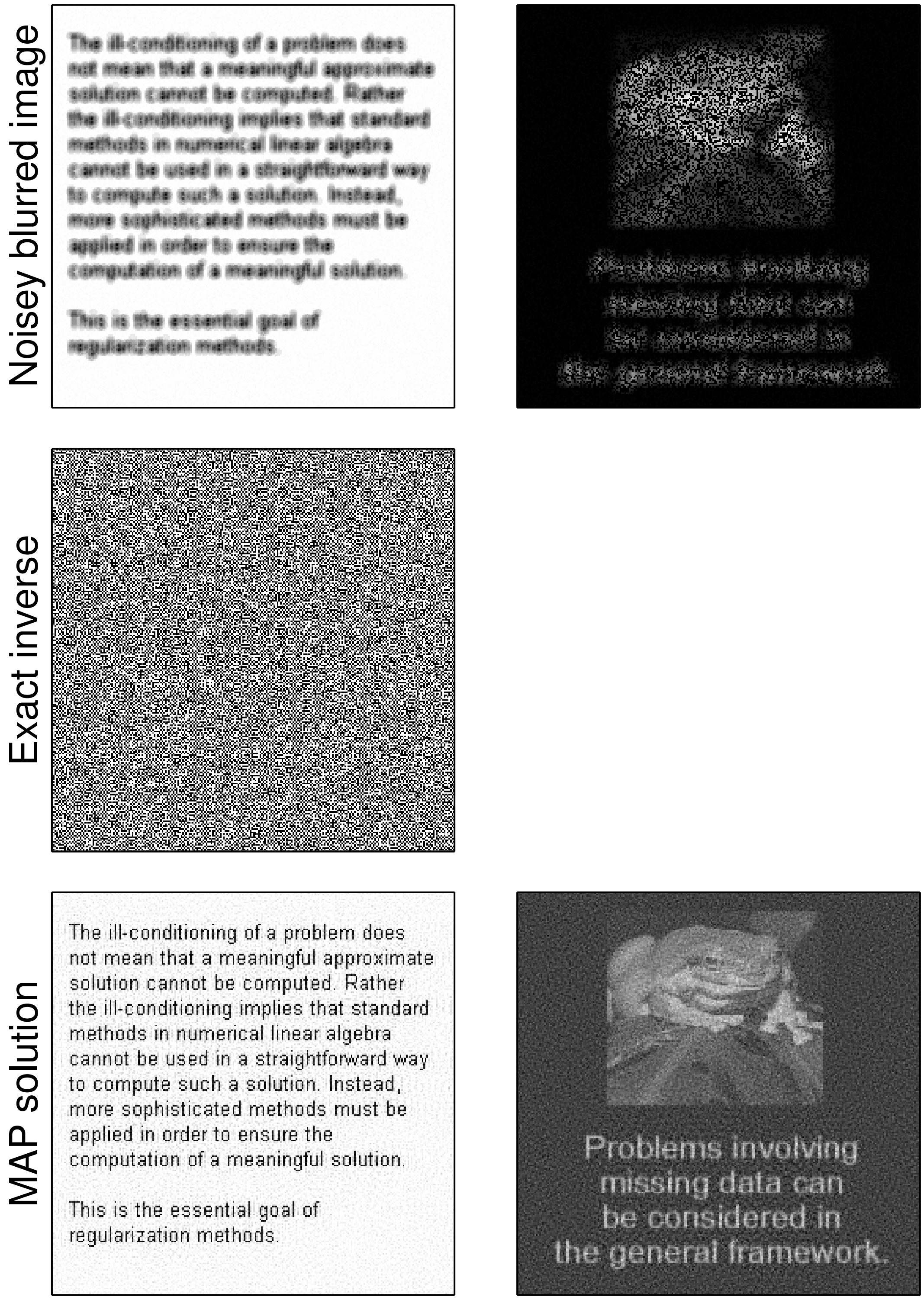
Inferential Solutions

"Answers" are expectations over the posterior

$$\mathsf{E}[g(f)] = \int \Pr(f|d)\, g(f)\, df$$



Examples of Modes = Regularization



α chosen using "L-curve"

Bayesian recovery of binary images

$N \times N$ binary image $f = (f_1, f_2 \dots f_{N^2})$ where each $f_i \in \{0, 1\}$.
We measure $g = (g_1, g_2 \dots g_{N^2})$ where $g_i = f_i + \epsilon_i$ with
$\epsilon_i \sim \mathrm{N}(0, \sigma^2)$

Likelihood

$$\begin{aligned}
\mathcal{L}(x) &= \prod_{i=1}^{N^2} \Pr\{g_i | x_i\} = \prod_{i=1}^{N^2} (\Pr\{g_i|1\})^{x_i} (\Pr\{g_i|0\})^{1-x_i} \\
&= \prod_{i=1}^{N^2} \left(\frac{\Pr\{g_i|1\}}{\Pr\{g_i|0\}}\right)^{x_i} \prod_{i=1}^{N^2} \Pr\{g_i|0\}
\end{aligned}$$

$$\lambda_i = \log\left(\frac{\Pr\{g_i|1\}}{\Pr\{g_i|0\}}\right) = \frac{2g_i - 1}{2\sigma^2}$$

Prior

$$\Pr(x) \propto \exp\left(-\theta \sum_{i=1}^{N^2} \sum_{j \sim i} (2x_i - 1)(2x_j - 1)\right)$$

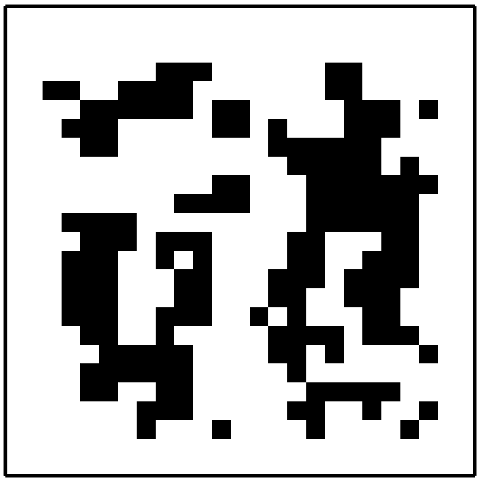
Joint posterior

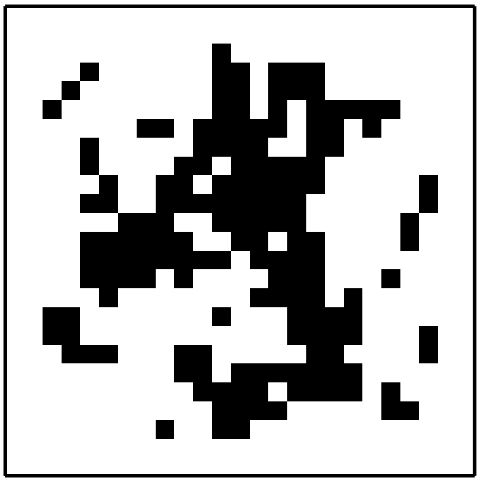
$$\begin{aligned}
\Pr\{x|g\} &\propto \mathcal{L}(x) \Pr(x) \\
&\propto \exp\left(\sum_{i=1}^{N^2} \lambda_i x_i - \theta \sum_{i=1}^{N^2} \sum_{j \sim i} (2x_i - 1)(2x_j - 1)\right)
\end{aligned}$$

Some samples from the prior

The prior is an Ising MRF

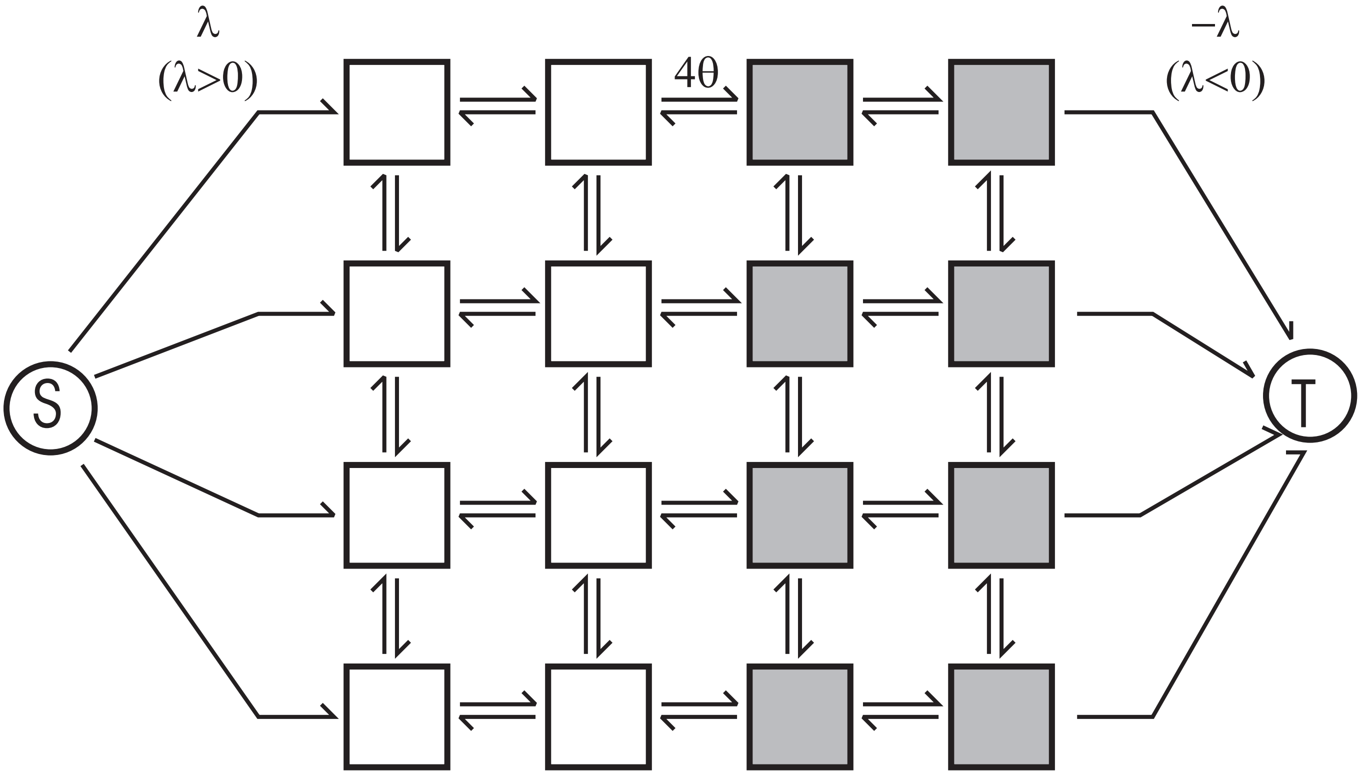
$$\Pr(x) \propto \exp\left(-\theta \sum_{i=1}^{N^2} \sum_{j \sim i} (2x_i - 1)(2x_j - 1) \right)$$





Network formulation for MAP

$$x_{\mathsf{MAP}} = \arg\max \Pr\{x|g\}$$



Cut capacity

$$c(X) = -\sum_{i=1}^{N^2} x_j \lambda_j + \theta \sum_{i=1}^{N^2} \sum_{j \sim i} 2(x_i + x_j) - 4x_i x_j$$

Label-setting

$\underline{X, p = \mathrm{label}(c, w, s)}$

$T \leftarrow \{s\}$

$X \leftarrow \{\}$

while $T \neq \{\}$

  $U \leftarrow \{\}$

  for $i \in T$

    for $k \in \{k = (i, j)$ is an arc : $j \notin X$ and $w_k < c_k\}$

      $U \leftarrow U \cup \{j\}$, $p_j = k$

    for $k \in \{k = (j, i)$ is an arc : $j \notin X$ and $w_k > 0\}$

      $U \leftarrow U \cup \{j\}$, $p_j = k$

  $X \leftarrow X \cup T$

  $T \leftarrow U$

Ford-Fulkerson successive-labelling

$\underline{X, w = \mathrm{FordFulk}(c, s, t)}$

$w \leftarrow 0$

$X, p \leftarrow \mathrm{label}(c, w, s)$

while $t \in X$

  Construct a flow-augmenting path P using p

  $w \leftarrow w+$ maximum increase of flow along P

  $X, p \leftarrow \mathrm{label}(c, w, s)$

Sampling a binary MRF

$$\ln(\pi_x) = -\ln(Z) - \theta \sum_{i=1}^{N^2} \sum_{j \sim i} (2x_i - 1)(2x_j - 1) - \sum_{i=1}^{N^2} (2x_i - 1) g_i / \sigma^2$$

Stochastic update (MH with dumb proposal)

$\underline{y = \phi(x, v)}$

$i \leftarrow \lceil v_1 N^2 \rceil$

$b \leftarrow \lceil v_2 - \frac{1}{2} \rceil$

$x' \leftarrow (x_1, x_2 \ldots x_{i-1}, b, x_{i+1} \ldots x_{N^2})$

if $v_3 \leq \min(1, \pi_{x'}/\pi_x)$

$\quad y = x'$

else

$\quad y = x$

$$x \leq y \Rightarrow \phi(x, v) \leq \phi(y, v) \quad \text{whenever} \quad \theta > 0$$

The stochastic update preserves partial ordering

Perfectly sampling a binary MRF

Fix $T > 0$, initialize two realizations with $\check{x}_{-T}(-T) = \mathbf{0}$ and $\hat{x}_{-T}(-T) = \mathbf{1}$. Simulate for $-T \leq t \leq 0$. If the upper and lower chains have coalesced, the sample is perfect.

Perfect Sampling

$\underline{x(0) = \psi(u, T)}$:

$t \leftarrow -T$

$\check{x}_{-T}(-T) \leftarrow \mathbf{0}$

$\hat{x}_{-T}(-T) \leftarrow \mathbf{1}$

do

  $\check{x}_{-T}(t+1) = \phi(\check{x}_{-T}(t), u^{(t)})$

  $\hat{x}_{-T}(t+1) = \phi(\hat{x}_{-T}(t), u^{(t)})$

  $t \leftarrow t + 1$

while $t < 0$

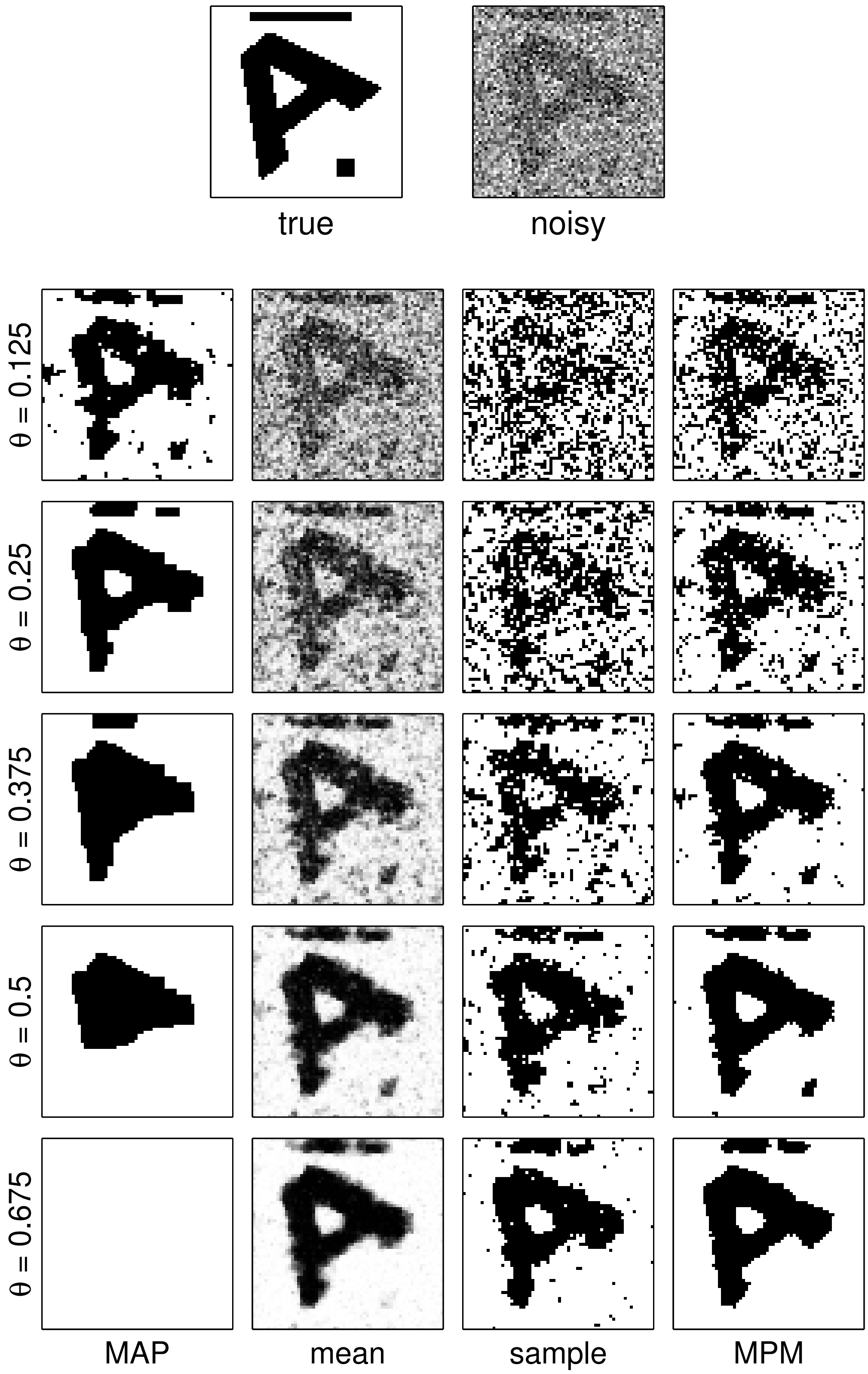
if $\hat{x}_{-T}(0) = \check{x}_{-T}(0)$

  $x(0) = \hat{x}_{-T}(0)$

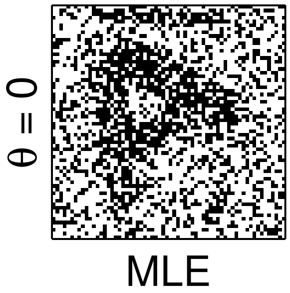
else

  $x(0) = \psi(u, 2T)$

Numerical experiments



Maximum likelihood estimate



Summary

- Ising MRF causes bulk of posterior probability to be centered on good reconstructions
- Good estimators should be available (e.g. Baddeley's delta loss)
- MAP estimate is sensitive to θ
- and is outlier for large θ
- hence not a good estimator
- Conjecture that this is true in most image recovery

References

[1] Fox, C., and Nicholls, G.K. *Exact MAP states and expectations from perfect sampling: Greig, Porteous and Seheult revisited.* Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 20th International Workshop (Gif-sur-Yvette), 8-13 July 2000, edited by Ali Mohammad-Djafari, AIP Conference Proceedings volume 568, 252-263, American Institute of Physics, New York, 2001.