ASPECTS OF STATISTICAL IMAGE MODELLING AND RESTORATION

D.M. Titterington
Department of Statistics
University of Glasgow
Glasgow, G12 QQ
Scotland, UK

mike@stats.gla.ac.uk

Notation (for pixellated images):

- x, an $N \times 1$ vector denoting the true scene
- y, an $N\times1$ vector denoting the noisy image
- N, the number of pixels

The Statistical Approach:

Introduce probability models for $y|x$ and possibly for x and use standard statistical paradigms to generate inference about x.

SUMMARY

A regularisation approach to image restoration

'Bayesian' image analysis

Deconvolution

(Links will Physics will be highlighted)

Notes

- It will turn out that these three parts overlap!

- There is some personal bias in the material in the talk (!), with some approaches to statistical image analysis not included, e.g. pattern theory (Grenander) and shape analysis (Dryden, Mardia and others)

A REGULARISATION APPROACH

A simple model: $y = Hx + \epsilon,$

where H is a blurring matrix and ϵ is additive noise; often $\epsilon \sim N(0, \sigma^2 I)$. Suitable for continuous intensity/grey-level images.

Least squares/maximum-likelihood estimator of x:

$$\widehat{x} := H^{-1}y = x + H^{-1}\epsilon,$$

the minimiser of

$$\Delta(x, y) := ||y - Hx||^2.$$

If the model is correct, $\widehat{x}$ is an *unbiased* estimator of x, but can be very unstable - high variance. Instead, minimise

$$\Delta(x, y) + \beta\Phi(x),$$

where Φ penalises 'roughness'.

A simple version has scalar $\beta > 0$ and $\Phi(x) = x^\top C x$, giving

$$\widehat{x}_\beta := (H^\top H + \beta C)^{-1} H^\top y.$$

(ridge regression, Tikhonov regularisation, spline smoothing,...)

How to choose β? Many approaches but we mention just a few with statistical connotations.

- **Minimum risk**:

  $$\widehat{\beta} = \text{arg min}_\beta E_{y|x} \delta(x, \widehat{x}_\beta),$$

  where δ is a measure of distance. Problem: $\widehat{\beta}$ is a function of the true x!

- **Generalised crossvalidation**:

$$\widehat{\beta} = \text{arg min}_{\beta} GCV(\beta),$$

where

$$GCV(\beta) := RSS(\beta)/[\text{tr}\{I - K(\beta)\}^2],$$

in which $K(\beta) = H(H^\top H + \beta C)^{-1}H^\top$, $K(\beta)y = E(y|\widehat{x}_\beta)$ denotes the set of 'fitted values', and $RSS(\beta) := ||\{I - K(\beta)\}y||^2$ is the residual sum of squares.

- **Empirical degrees of freedom choice**:

Solve $\quad RSS(\beta)/\text{tr}\{I - K(\beta)\} = \sigma^2.$

‘BAYESIAN’ IMAGE ANALYSIS

(Geman & Geman, 1984; Besag, 1986)

Propose models for $p(y|x, \theta)$ and $p(x|\beta)$, e.g. Ising/Potts models, with $\psi = (\theta, \beta)$ the total set of parameters. Typically, the models correspond to y coming from a **hidden Markov random field (MRF)** model.

‘Simplest’ example : for binary images, with each $x_i = -1$ or $+1$, take $p(x|\beta)$ to correspond to the Ising model:

$$p(x|\beta) = \{C(\beta)\}^{-1} \text{exp}(\beta \sum_{i \sim j} x_i x_j),$$

in which the sum is over neighbouring pairs of pixels.

For inference, key items of interest are

$$\begin{aligned} p(x|y,\psi) &\propto p(x,y|\psi) = p(y|x,\theta)p(x|\beta) \\ p(y|\psi) &= \sum_x p(y|x,\theta)p(x|\beta) \end{aligned}$$

Inference about x should be based on $p(x|y,\psi)$; $p(y|\psi)$ is the observed-data log-likelihood.

(How Bayesian is 'Bayesian' image analysis?)

INFERENCE ABOUT THE TRUE SCENE

If ψ is known, for restoration use some point estimate from $p(x|y,\psi)$, such as MAP (G&G), ICM (Besag).

More refined inferences, e.g. interval estimates, are hard to come by (MCMC - Physics links).

If ψ not known, we should estimate it from y, using Likelihood or Bayesian methods.

MAXIMUM LIKELIHOOD ESTIMATION OF ψ

Objective is to find ψ to max. $p(y|\psi)$.

General algorithm for incomplete data - **EM algorithm**

Define

$$L(x, y|\psi) = \log\{p(y|x, \theta)p(x|\beta)\},$$

and apply the following two-stage iterative step to the current iterate $\psi^{(m-1)}$.

- **E-step**: calculate $Q(\psi) = E_m L(x, y|\psi)$, with the averaging based on $p(x|y, \psi^{(m-1)})$.

- **M-step**: find $\psi = \psi^{(m)}$ to maximise $Q(\psi)$.

As a general rule, $p(y|\psi^{(m)}) \geq p(y|\psi^{(m-1)})$.

However, neither E-step nor M-step is easy for hidden MRF data. Various approximate methods exist, e.g. versions of mean-field approximations, Besag's pseudolikelihood, etc.

SOME ILLUSTRATIONS

- **Example 1**: four-band satellite image.

- **Example 2**: transition boundary on Co-Ni evaporated tape.

- **Example 3**: magnetc domain in TEM image.

DIRECT USE OF MEAN-FIELD-LIKE APPROXNS FOR COMPLICATED LIKELIHOODS

$$\begin{aligned}\text{log } p(y|\psi) &= \text{log}\{\sum_x p(x,y|\psi)\} \\ &\geq \sum_x q(x) \text{ log}\{p(x,y|\psi)/q(x)\},\end{aligned}$$

by Jensen's inequality, where $q(x)$ is any probability distribution for x. If $p(x|y,\psi)$ is too complicated, use q of a simple form, e.g. fully factorised, such that $q(x) = \prod_i q_i(x_i)$ (mean-field approximation), with certain 'hyper'parameters, chosen to maximise the lower bound. This q minimises

$$KL(q,p) := \sum_x q(x) \text{ log}\{q(x)/p(x|y,\psi)\}.$$

FULLY BAYESIAN ANALYSIS

Propose 'hyper'priors on $\psi = (\theta, \beta)$, and base inference about θ, β *and* x on $p(x, \theta, \beta|y)$:

$$p(x, \theta, \beta|y) \propto p(y|x, \theta)p(x|\beta)p(\theta)p(\beta).$$

Problems with image-analysis data:

- no practically useful closed form for $p(x, \theta, \beta|y)$
- MCMC simulation from $p(x, \theta, \beta|y)$ not easy

Solutions (?)

- fudges tried within the MCMC procedures
- attempts at variational approximations - find manageable approximation $q(x, \theta, \beta)$ to $p(x, \theta, \beta|y)$ to minimise $KL(q, p)$.

‘BAYESIAN’ INTERPRETATION OF REGULARISATION

If $y = Hx + \epsilon$ with $\epsilon \sim N(0, \sigma^2 I)$ and, marginally, $x \sim N(0, \sigma^2 \beta^{-1} C^{-1})$, then the negative of the logarithm of $p(x|y)$ is, ‘more or less’

$$||y - Hx||^2 + \beta x^\top C x,$$

so that the mode (or mean) of x is

$$\widehat{x}_\beta = (H^\top H + \beta C)^{-1} H^\top y.$$

This motivates choosing β to maximise the likelihood

$$p(y|\beta) = \int p(y|x) p(x|\beta) dx,$$

under the assumption that H and σ^2 are known.

DECONVOLUTION

Consider the inverse problem represented by

$$y = Hx,$$

interpretable as a 'discrete' deconvolution problem. Formal 'solution' for a square matrix H is $x = H^{-1}y$, but often ill-posed and also often x should be nonnegative.

Vardi and Lee (1993) note that, by scaling, can assume that elements of x, y and columns of H sum to 1. They propose the algorithm

$$x_i^{(m)} = x_i^{(m-1)} \sum_j (h_{ij} / \sum_k x_k^{(m-1)} h_{kj}) y_j,$$

for each i and $m = 1, \ldots$. This algorithm converges to the x^* that minimises

$$\sum_i y_i \log(y_i / z_i) = KL(y, z),$$

where $z_i = (\sum_j h_{ij} x_j)$.

The algorithm has an EM interpretation with

- **E-step**: for each i and j calculate

$$z_{ij}^{(m-1)} = \frac{x_i^{(m-1)} h_{ij}}{\sum_k x_k^{(m-1)} h_{kj}} y_j.$$

- **M-step**: for each i, calculate

$$x_i^{(m)} = \sum_j z_{ij}^{(m-1)}.$$

Applications include emission tomography image reconstruction and motion de-blurring.

Modifications include smoothed EM (Silverman et al., 1990) and modified EM, with roughness penalty (Green, 1990).

Another algorithm is the Iterative **Image Space Restoration Algorithm**, for which the iteration is

$$x_i^{(m)} = x_i^{(m-1)}(\sum_j h_{ij}y_j)/\{\sum_j h_{ij}(\sum_k x_k^{(m-1)}h_{kj})\},$$

for each i. References include Daube-Witherspoon and Muehllehner (1986), De Pierro (1987) and Titterington (1987) - interpretation as iterative way of calculating least squares estimates of x.

Example 4 (Vardi and Lee, 1993; Archer and Titterington, 1995). An example of motion de-blurring.

DECONVOLVING MULTIPERIODIC FUNCTIONS (Hall and Yin, 2003)

Observed signal: for $i = 1, \ldots, n$,

$$y_i = g(t_i) + \epsilon_i = \mu + \sum_{j=1}^{r} g_j(t_i) + \epsilon_i,$$

where the g_j are periodic components with minimal periods $0 < \theta_1 < \ldots < \theta_r$.

Objectives: estimate the $\theta = \{\theta_j\}$ and the $\{g_j\}$, nonparametrically, as follows:

- calculate $\theta = \widehat{\theta}$ to minimise

  $$S(\theta) = \sum_i \{y_i - \widehat{g}(t_i|\theta)\}^2,$$

  with

  $$\widehat{g}(t|\theta) = \{\sum_i y_i K(t, t_i)\} / \sum_i K(t, t_i),$$

  in which the kernel function $K(t, t')$ is a function of θ;

- given the $\{\widehat{\theta}_j\}$, write

$$g(t) = \mu + \sum_{j=1}^{r} \sum_{k=1}^{m} a_{jk} \psi_k(t/\widehat{\theta}_j),$$

  in which the $\{a_{jk}\}$ are generalised Fourier coefficients, μ is a constant and m is a truncation point;

- estimate μ and the $\{a_{jk}\}$ by least squares.

Example 5 (Hall and Yin, 2003)

Fit the model to radiation measurements from the slowly-pulsating B-star HD 123515; a multiperiodic function with $r = 4$ periods gives a good fit.