Two Applications of Bayesian Statistics and Decision Theory

Simon Wilson

School of Computer Science and Statistics
Trinity College Dublin
simon.wilson@tcd.ie
Overview of this short course

• I will look at two applications of Bayesian statistics and decision theory:
  1. The content-based image retrieval problem;
  2. Source separation for multi-spectral images;

• Both applications were very interesting to me;
• The first is much simpler and easy to implement than the second;
• The second problem needed a lot of complicated computation that I will try to describe as simply as possible!
APPLICATION 1
CONTENT-BASED IMAGE RETRIEVAL

Joint work with Georgios Stefanou
Outline of first application

• What is content-based image retrieval?
• Some image processing basics;
• A Bayesian solution to CBIR (with a demo);
• Solving the “display” problem;
• (If time permits) Improving subsequent searches by modelling the search process.
Content-based image retrieval (CBIR)

- Problem: searching for images with certain content in a large digital database;
- Canon (2002): $\approx 10^{11}$ digital images taken per year, and total multi-media data generation $\approx 10^{18}$ bytes/year;
- These days search for text is very good (Google);
- What about image search in Google — it’s sometimes good but often bad!
Applications of CBIR

If you had a really good image searching system, the applications are many and lucrative:

- Entertainment (images to create an emotion etc.);
- Document preparation (incl. newspapers);
- Teaching;
- Graphical design and advertising;
- Medicine (looking for similar NMR images to a patient);
- Security (face recognition);
- Organising and searching your personal photograph collection;
- More generally, multi-media data retrieval (video, audio).
Types of image searching

Search may be:

1. *Specific*: “I want that photograph of a sailor and a woman kissing in Times Square, New York at the end of World War II”;

2. *Categorical*: “I want an image of a beach at sunset with palm trees on the left hand side”, “an image of the Taj Mahal”, etc.;

3. *Browsing*: “I’m need a painting to hang in my living room and I’m looking for suggestions”. 
Mental image search

- One common CBIR scenario is “query-based” search;
- The user provides an example of the sort of image required;
- Having such an image makes the search process easier;
- More difficult is “mental image” search;
- There is no query image — user has an image of the target image in his/her mind;
- An extension is mental image category search, where the user is interested in retrieving images in a certain category.
Text annotation

- A simple way to search a database is if each image is “annotated” by text that describes it;
- This is common in many commercial databases (media organisations, art libraries, etc.);
- However this is far from a perfect solution:
  - It’s expensive (humans must annotate);
  - It’s subjective;
  - It can’t possibly describe all of an image’s interpretations.
- ⇒ even very good annotations will not be able to match many queries.
Content-based image retrieval

A better way to search is *Content-based Image Retrieval.*

- CBIR consists of two elements:
  1. A *feature extraction* algorithm that describes the content of each image;
  2. A *retrieval algorithm* that uses the features to retrieve images according to a query.

- Successful retrieval algorithms always work interactively with the user by a process called *relevance feedback*;
- Many use statistical techniques to try to "learn" about what the user is really looking for;
A computer extracts features of an image, to do with colour, texture, location and shape of objects;

These features (hopefully) describe well the content (or semantics) of the image;

This can be done off-line and needs to be done only once;

Searching the database is based on these features and a “similarity measure” between them;

This is a decreasing function of a distance between their features.
Feature extraction 2

• An image $X$ is a matrix $\{X_{ij} | i = 1, \ldots, n_1; j = 1, \ldots, n_2\}$;
• $X_{ij}$ is colour of pixel $(i, j)$; colour is a 3-vector, for example in red/green/blue (RGB) -space $X_{ij} = (R_{ij}, G_{ij}, B_{ij}) \in \{0, \ldots, 255\}^3$;
• Feature vector of length $d$ is $f(X) \in \mathbb{R}^d$;
• Distance between images $X_1$ and $X_2$ is $d(X_1, X_2) = \| f(X_1) - f(X_2) \|$;
• Similarity measure $s(X_1, X_2) = \exp(-d(X_1, X_2))$ or $d(X_1, X_2)^{-1}$, etc.
Alternative colour representations: hue, saturation, value
HSV colour space

It is defined in terms of RGB as:

\[ V = \frac{R + G + B}{3}; \]

\[ S = 100 \left[ 1 - \frac{3 \min(R, G, B)}{R + G + B} \right]; \]

\[
H = \begin{cases} 
\text{undefined,} & \text{if } S = 0, \\
180 \frac{0.5(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)}}, & \text{if } BV \leq GV, \\
360 - 180 \frac{0.5(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)}}, & \text{if } BV > GV.
\end{cases}
\]
Feature extraction 3

• Typical features: histograms of colours (in HSV space), autocorrelograms at different pixel distances, colour coherence vectors, locations–directions–lengths of edges, location–shape–colour of objects;

• However, good automatic object detection and image segmentation is difficult to achieve.
The semantic gap: closest images in Feature Space
Relevance feedback

- An interactive process between user and system;
- Aids the retrieval process and attempts to bridge the semantic gap;
- CBIR system displays images and the user then judges them by marking one or more that he/she perceives to be really relevant;
- The system then updates its retrieval in light of this information;
- More sophisticated feedback involves: marking objects in the image, marking images as not relevant, ordering images by relevance, etc.
A Bayesian Solution to CBIR

1 Initial set of $N_D$ images from the database is displayed:
   - These may be randomly chosen;
   - More sophisticated systems allow the user to draw a simple representation of what is wanted, then $N_D$ most similar images in database are displayed.

2 Repeat until image found:
   - User chooses one or more image from display set that are most relevant to query;
   - System updates and displays a new set of $N_D$ images in light of this information.
Bayesian CBIR — PicHunter

• Developed by Cox et. al. (2000);
• Uses the simple case of target search and one image per display chosen;
• Assume database $\mathcal{X} = \{X_1, \ldots, X_N\}$ of images, one of which is target $T$;
• At $t$-th iteration, system displays a set $D_t \subseteq \mathcal{X}$ and user selects $A_t \in D_t$ as most relevant;
• Let $H_t = \{D_1, A_1, \ldots, D_t, A_t\}$;
• Objective: compute $\mathbb{P}(T \mid H_t), \ T \in \mathcal{X}$;
• We have extended this basic system to include extra unknowns.
We have

\[ P(T | H_t) \propto P(H_t | T) P(T) \]

\[ = \left\{ \prod_{i=1}^{t} P(A_i, D_i | T, H_{i-1}) \right\} P(T) \]

\[ = \left\{ \prod_{i=1}^{t} P(A_i | D_i, T) \right\} P(T); \]

\[ \mathbb{P}(A_i | D_i, T) \text{ is the likelihood of the user picking } A_i \text{ from } D_i, \text{ were the target } T \text{ to be known.} \]
• $P(A_i \mid D_i, T)$ is a function of the relative similarity between $A_i$ and $T$ in $D_i$, e.g.

$$P(A_i \mid D_i, T, \sigma) = \frac{\exp\left(-d(A_i, T)/\sigma\right)}{\sum_{X \in D_i} \exp\left(-d(X, T)/\sigma\right)}, \quad A_i \in D_i,$$

for a precision parameter $\sigma$;

• $\sigma$ measures how well the distance measure describes the user’s choices as most relevant image.
A Better distance measure

- We divide our features into 3 classes:
  - colour (CL);
  - texture (TX);
  - segmentation/objects (SG).

- Let $F$ be the class used, then:

\[
\mathbb{P}(A_i \mid D_i, T, \sigma, F) = \frac{\exp\left(-d_F(A_i, T)/\sigma\right)}{\sum_{X \in D_i} \exp\left(-d_F(X, T)/\sigma\right)}, \quad A_i \in D_i,
\]

where $d_F$ is distance measure on features in class $F$. 
Posterior computation 1

• Assume uniform priors:

\[ P(T = X_i) = N^{-1}, \ X_i \in \mathcal{X}; \]
\[ P(\sigma) = 1/\sigma_{\text{max}}, \ 0 \leq \sigma \leq \sigma_{\text{max}}; \]
\[ P(F) = 1/3, \ F \in \{\text{CL, TX, SG}\}. \]

• After the \( t \)-th iteration compute:

\[ P(T, \sigma, F | H_t) \propto \left\{ \prod_{i=1}^{t} \frac{\exp \left( -d_F(A_i, T)/\sigma \right)}{\sum_{X \in D_i} \exp \left( -d_F(X, T)/\sigma \right)} \right\} \times P(T)P(\sigma)P(F). \]
Posterior computation 2

• This computation is done on-line so must be quick (absolutely no MCMC!);
• We discretise $\sigma$ and compute over all $3NN_\sigma$ combinations;
• In MATLAB code, this allows $N_\sigma = 20$ and $N = 5000$ to be computed in 5 seconds;
• Computation is ideally suited to parallelisation.
Demo

- Once we have computed \( P(T, \sigma, F \mid H_t) \), we must display \( D_{t+1} \);
- Usual choice is \( N_D \) most probable images from

\[
P(T \mid H_t) = \sum_{F=GC, TX, SG} \int P(T, \sigma, F \mid H_t) \, d\sigma;
\]

- The system is run on a database of 1066 images of paintings from the Bridgeman Art Library, London;
- This is a challenging database for CBIR because it contains images with many different subjects and styles.
Solving the “display” problem

Choosing $D_{t+1}$

- Once we have computed $\mathbb{P}(T, \sigma, F | H_t)$ we must display $D_{t+1}$;
- Usual choice is $N_D$ most probable images from $\mathbb{P}(T | H_t)$;
- But this might not be sensible;
- Problem can be framed as a decision problem;
- Let $\mathcal{U}(D, T)$ be the utility of picking a set $D$ when $T$ is the target, then

$$D_{t+1} = \arg \max_{D \subseteq \mathcal{X}} \mathbb{E}(\mathcal{U}(D, T)) = \sum_{T \in \mathcal{X}} \mathcal{U}(D, T) \mathbb{P}(T | H_t).$$
Indicator utility

• Let

\[ U_I(D, T) = \begin{cases} 
1, & \text{if } T \in D, \\
0, & \text{otherwise.} 
\end{cases} \]

• Then \( E(U(D, T)) = \sum_{T \in D} P(T \mid H_t); \)

• So \( D_{t+1} \) is indeed the set of \( N_D \) most probable images.
We might want to choose $D$ to maximise information content; Try a utility based on negative expected entropy from choosing an image from $D$:

$$U_E(D, T) = - \sum_{A \in D} E(A, D) \mathbb{P}(A \mid D, T),$$

where

$$E(A, D) = - \sum_{T \in \mathcal{X}} \mathbb{P}(T \mid A, D) \log(\mathbb{P}(T \mid A, D)).$$

To speed up computation we say

$$\mathbb{P}(A \mid D, T) = \frac{\exp \left(-d(A, T)/\hat{\sigma} \right)}{\sum_{X \in D} \exp \left(-d(X, T)/\hat{\sigma} \right)},$$

where $\hat{\sigma} = \mathbb{E}(\sigma \mid H_t)$. 

**Entropy utility**
Evaluating $D_{t+1}$ with $U_E(D, T)$

- $E(U_E(D, T))$ not separable in $T$;
- Cannot evaluate all $\binom{\|x\|}{N_D}$ combinations;
- It’s an online computation so must be fast;
- Tried two approaches:
  1. Simulated Annealing;
Example: 15 image database with 2 features
$D_2$ under indicator utility
$D_2$ under entropy utility
Expected utilities for $D_2$ — comparison of computational methods (average over 100 trials)

<table>
<thead>
<tr>
<th>Utility</th>
<th>Exact</th>
<th>Random Generation of 100 subsets</th>
<th>Sim. Annealing 100 iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator</td>
<td>0.2643</td>
<td>0.2624</td>
<td>0.2591</td>
</tr>
<tr>
<td>Entropy</td>
<td>-2.6869</td>
<td>-2.6879</td>
<td>-2.6883</td>
</tr>
</tbody>
</table>
Example: Bridgeman art library

- A database of 1066 paintings;
- Some 500 features computed on each image;
- PCA done on each feature class - reduces to 140 principal components.
Initial display set (centre top image selected)
Next display set $D_2$ under indicator itility
Next display set $D_2$ under entropy utility
Expected utilities for $D_2$ in paintings example (average over 100 trials)

<table>
<thead>
<tr>
<th>Utility</th>
<th>Exact</th>
<th>Random Generation of 100 subsets</th>
<th>Simulated Annealing 100 iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator</td>
<td>0.0157</td>
<td>0.0106</td>
<td>0.0100</td>
</tr>
<tr>
<td>Entropy</td>
<td>—</td>
<td>-6.9642</td>
<td>-6.9647</td>
</tr>
</tbody>
</table>
Hybrid display strategy

- A good strategy might be to:
  - In early iterations we want to maximise information content ($U_E$);
  - Later we want to display images that we think are close to what the user wants ($U_I$);
- So might at iteration $t$ use

$$U(D, T) = \alpha_t U_E(D, T) + (1 - \alpha_t) U_I(D, T),$$

where $\alpha_1 = 1$ and $\alpha_t \to 0$. 
Improving subsequent searches by modelling the search process mental image category search

- This is applicable in mental image category search;
- Recall that the user is interested in retrieving images in a certain category, but has no query image;
- One solution to building up a category: Repeatedly run the system from different (random) $D_1$;
Mental image category search

• This builds up a category of images;
• Can we use information from previous runs to improve search?
• We could use posterior $\mathbb{P}(T, \sigma, F \mid H_t)$ from previous run as prior for new run
  • The retrieved image from each run: highly dependent on $D_1$, using it will eliminate other clusters in category
  • More robust to use $\mathbb{P}(\sigma, F \mid H_t)$. 
Mental image category search

- Assume that $M$ previous runs of a query have been done.
- Data from run $m$ is $H_t^{(m)} = (D_1^{(m)}, A_1^{(m)}, \ldots, D_t^{(m)})$.
- The posterior of $(T, \sigma, F)$:

$$
P(T, \sigma, F \mid H_t^{(1)}, \ldots, H_t^{(M)}) 
\propto \left( \prod_{m=1}^{M} \mathbb{P}(H_t^{(m)} \mid T, \sigma, F) \right) \mathbb{P}(T, \sigma, F),$$

$$
= \left( \prod_{m=1}^{M} \prod_{k=1}^{t^{(m)}} \mathbb{P}(A_k^{(m)} \mid D_k^{(m)}, T, \sigma, F) \right) \mathbb{P}(T, \sigma, F),$$

- The posterior distribution of $\sigma$ and $F$ is

$$
P(\sigma, F \mid H_t^{(1)}, \ldots, H_t^{(M)}) = \sum_{T \in \mathcal{X}} \mathbb{P}(T, \sigma, F \mid H_t^{(1)}, \ldots, H_t^{(M)}).$$
Mental image category search

- For the \((M+1)\)th search, we use prior:

\[
P(T) \, P(\sigma, F \mid H_t^{(1)}, \ldots, H_t^{(M)})
\]

- Posterior distribution after iteration \(t\) of \((M+1)\)th search is:

\[
P(T, \sigma, F \mid H_t^{(1)}, \ldots, H_t^{(M)}, H_t)
\propto \left\{ \prod_{i=1}^{t} \frac{\exp\left(-d_F(A_i, T)/\sigma\right)}{\sum_{X \in D_i} \exp\left(-d_F(X, T)/\sigma\right)} \right\}
\times \, P(T)P(\sigma, F \mid H_t^{(1)}, \ldots, H_t^{(M)}).\]
Evaluation

- We use the Bridgeman Art Library database;
- We look for images with different emotional content (romantic, sad, violent);
- Ten searches are done with the uniform prior;
- Then ten more searches are done with the prior for $(\sigma, F)$ as the posterior from the first 10 searches;
- We look at the no. of iterations to find a relevant image between the two sets of searches.
Romantic — first 10 iterations: $\bar{x} = 9.4$, $s = 6.2$
Romantic — second 10 iterations: $\bar{x} = 3.7, s = 2.1$
Evaluation

- In this case mean no. of iterations is significantly smaller ($t$-test);
- Test repeated with searches for “violent” and “sad” images;
- In these cases, mean no. for second stage is smaller (but not statistically significant);
- Results repeated for 2 different users;
- Clearly more testing is needed.
Posterior of $\sigma$ and $F$ after 1st and 2nd stages
Posterior of $\sigma$ and $F$ after 1st and 2nd Stages — Violent
Evaluation and conclusion

- Posterior is consistent between Stages 1 and 2;
- We find that we get different posteriors for:
  - Different queries with the same user;
  - Different users with the same query;
- Also tested where a user is given a posterior from another user;
  - No significant difference between mean no. of iterations with uniform prior and other-user posterior.
Conclusion to CBIR application

• Presented a Bayesian approach to CBIR;
• Described some of the issues that need to be solved;
• The principal impediment is still feature description and the semantic gap;
• Novel features about this approach:
  • Modelling framework permits easy extension to other forms of relevance feedback;
  • Natural measure of relevance through posterior probability;
  • Decision theory allows nice abstraction of display problem.
APPLICATION 2
SOURCE SEPARATION FOR THE COSMIC BACKGROUND RADIATION

Joint work with Jiwon Yoon (Trinity College Dublin), Ercan Kuruoğlu (CNR Pisa) and Alicia Quirós Carretero (URJC)
Cosmic microwave background (CMB)

- Discovered by accident in 1964 by researchers at Bell Laboratories, New Jersey;
- It is microwave radiation that:
  - exists at all points in the sky;
  - is very uniform;
- By 1970’s agreed to be an image of the first scattering of EM radiation at recombination \( \approx 300,000 \) years after Big Bang;
- Of great interest as an observation of the state of the early universe:
  - Accurate measurement of the small anisotropies place strong restrictions on theories of big bang, galaxy formation etc.;
- Cosmic expansion \( \Rightarrow \) radiation has cooled to 2.7K (microwave);
Source separation problem for the CMB

Unfortunately there are many other sources of microwaves in the universe:
Source separation

- 3 satellites have taken "all-sky" maps of microwaves (COBE, WMAP and Planck);
- From these images, want to "separate" CMB from the other sources;
- If we have one image of the microwaves then this is impossible uniquely;
- It is possible if we take many images of the microwaves at different frequencies;
- Different sources have different importance at different frequencies:
  - Therefore it should be possible to separate the sources.
Images of data at 4 different frequencies
Images of 3 sources (CMB, synchrotron, galactic dust)
Model for source separation

- $J$ pixels per image ($J = 10^6 - 10^7$), $n_f$ observed images at frequencies $\nu_1, \ldots, \nu_{n_f}$ ($n_f = 5$ for WMAP, 9 for Planck), $n_s$ sources (one is CMB);
- At pixel $j$, observe $y_j$, related to sources $s_j$:
  \[ y_j = As_j + \epsilon_j, \quad s_j \in \mathbb{R}^{n_s}, \]
  where $\epsilon_j \sim N(0, \text{diag}(\tau_1^{-1}, \ldots, \tau_{n_f}^{-1}))$, $A$ is a $n_f \times n_s$ matrix;
- This is factor analysis in statistics;
- More notation:
  - $S_i = (s_{1i}, \ldots, s_{ji})$, $i = 1, \ldots, n_s$ is the image of the $i$th source;
  - $Y_k = (y_{1k}, \ldots, y_{Jk})$, $k = 1, \ldots, n_f$ is $k$th observed image;
  - $Y = (Y_1, Y_2, \ldots, Y_{n_f})$, $S = (S_1, S_2, \ldots, S_{n_s})$;
Source separation for multi-channel images

Example with $n_s = 3$, $n_f = 2$ and $A = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \end{pmatrix}$.
Source separation for multi-channel images

Example with $n_s = 3$, $n_f = 2$ and $A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$.
Source separation problem for the CMB

\[ y_j = A s_j + \epsilon_j \]

- Goal: observe the \( y_j \) and infer the \( s_j \) and \( A \);
- \( A \) is called the mixing matrix;
- \( A_{ik} \) is the contribution of the \( k \)th source at the \( i \)th frequency;
- Different number \( n_j \) of observations at pixel \( j \);
- Model can be written:

\[ Y = BS + E, \]  (1)

where:
- \( B = A \otimes I_{J \times J} \);
- \( E \sim N(0, C^{-1}) \), \( C = \text{diag}(n_1 \tau_1, \ldots, n_J \tau_1, \ldots, n_1 \tau_n, \ldots, n_J \tau_n) \);
- \( y_{jk} \) is the average of the \( n_j \) observations of that pixel;
- \( n_j \) are assumed known and \( \tau_k \) are pretty much known.
Likelihood

Likelihood is:

\[ p(Y | S, A) \propto |C|^{0.5} \exp(-0.5(Y - BS)^T C (Y - BS)), \]

where

\[ B = A \otimes I_{J \times J}. \]
Prior information

- $S_i$, $i = 1, \ldots, n_s$ are the sources that make up the microwave signal received by the satellite:
  - One of these sources is the CMB (source 1);
  - Other important ones are synchrotron radiation and galactic dust;
  - There are many others.... is $n_s$ known?
  - A lot is known from physics about the properties of these sources e.g. their spectrum, mean, variance etc;

- $A$ is not known but the physics tell us a lot about it;
  - The physics of each source will tell us a lot about the $A_{ik}$;
  - For example, CMB is believed to be black body radiation at 2.725K;

- A lot of “prior” information $\Rightarrow$ a Bayesian approach looks promising.
Prior information about A: CMB is black body
Prior for $A$

- The CMB is black body radiation at $T_0 = 2.725$K, so response at $\nu_k$ is known:

$$A_{k_1} \propto \left( \frac{h \nu_k}{k_B T_0} \right)^2 \frac{e^{h \nu_k / k_B T_0}}{(e^{h \nu_k / k_B T_0} - 1)^2},$$

$h$ is Planck constant, $k_B$ is Boltzmann’s constant.
Prior for $A$

- For other sources, physical argument to say that approximately:

$$A_{ki} = \left(\frac{\nu_k}{\nu_4}\right)^{\theta_i}$$

for a parameter $\theta_i$;

- So one free parameter $\theta_i$ per column of $A$;

- So $A$ parameterised by $(n_s - 1)$ dimensional $\theta$;

- Physical theories give tight bounds on these $\theta$:\n  - $(-3.5, -2.0)$ for synchotron, $(0.5, 1.5)$ for galactic dust, etc.;
    - These lead to Gaussian priors e.g. $\theta_{\text{dust}} \sim N(1.0, 0.25^2)$. 

Prior model for sources

- Marginally, source distributions show skewness, multimodality;
- Sources exhibit within and across (spatial) pixel correlations:
  - Galactic sources, extra-galactic sources and CMB should be independent;
  - Sources within galaxy show dependencies, similarly extra-galactic;
  - Most sources show spatial smoothness.
Intrinsic GMRF

• These model the spatial smoothness of sources;
• We’ll ignore multimodality, between-source correlation for now!
• IGMRF: the differences

$$\sum_{j' \in c(j)} (S_{ij} - S_{ij'})$$

are $N(0, \phi_i^{-1})$, where $c(j)$ is four neighbours of pixel $j$. 
Intrinsic GMRF

- Distribution of $S_i$ is:

$$p(S_i | \phi_i) \propto |Q(\phi_i)|^{0.5} \exp \left( -0.5 S_i^T Q(\phi_i) S_i \right), \quad (2)$$

where $Q(\phi_i) = \phi_i D^T D$ with

$$D_{j_1,j_2} = \begin{cases} 1, & \text{if } j_2 \in c(j_1) \\ 0, & \text{otherwise,} \end{cases}$$

for $j_1 \neq j_2$ and main diagonal elements are

$$D_{jj} = - \sum_{l=1}^{J} D_{j,l}.$$

- $\Psi = (\theta_2, \ldots, \theta_{n_s}, \phi_1, \ldots, \phi_{n_s})$ is all hyperparameters
Finally we have

\[ p(S | \Psi) \propto |Q(\Psi)|^{0.5} \exp \left( -0.5 S^T Q(\Psi) S \right), \quad (3) \]

where

\[
Q(\Psi) = \begin{pmatrix}
Q(\phi_1) & 0 & \cdots & 0 \\
0 & Q(\phi_2) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & Q(\phi_{ns})
\end{pmatrix}.
\]

- \(Q(\Psi)\) is about a \(10^8 \times 10^8\) sparse band matrix.
Now try to compute:

\[ p(S, \Psi \mid Y) \propto p(Y \mid S, \Psi) p(S \mid \Psi) p(\Psi); \]

Could use MCMC but here we’ll try a Laplace approximation;

Observe that

\[ p(\Psi \mid Y) \propto \frac{p(Y \mid S, \Psi) p(S \mid \Psi) p(\Psi)}{p(S \mid Y, \Psi)}. \]

for any \( S \) such that \( p(S \mid Y, \Psi) > 0 \).
Denominator term is

\[ p(S \mid Y, \Psi) = \frac{p(Y, S, \Psi)}{p(Y, \Psi)} \propto p(Y, S, \Psi) \]

\[ = p(Y \mid S, \Psi) \, p(S \mid \Psi) \, p(\Psi) \propto p(Y \mid S, \Psi) \, p(S \mid \Psi) \]

\[ \propto |C|^{0/5} \exp(-0.5(Y - BS)^T C (Y - BS)) \]

\[ \times |Q(\Psi)|^{0.5} \exp(-0.5S^T Q(\Psi)S) \]

\[ = \cdots \text{some algebra (complete the square)} \]

\[ \propto |Q^*(\Psi)|^{0.5} \exp(-0.5(S - \mu^*(\Psi))^T Q^*(\Psi)(S - \mu^*(\Psi))) \]

so \( S \mid Y, \Psi \) is Gaussian with precision and mean:

\[ Q^*(\Psi) = Q(\Psi) + B^T C B \]

\[ \mu^*(\Psi) = Q^*(\Psi)^{-1} B^T C Y \]
Marginal Posterior of $\Psi$

- Evaluate at $S = \mu^*(\Psi)$ to get:

\[
p(\Psi | Y) \propto |Q^*(\Psi)|^{-0.5} \, p(Y | S = \mu^*(\Psi), \Psi) \times p(S = \mu^*(\Psi) | \Psi) \, p(\Psi) = q(\Psi | Y).
\]

- Approximation to $p(\Psi | Y)$ on discrete grid $\Psi \in Q$:

\[
p(\Psi | Y) \approx \tilde{p}(\Psi | Y) = \frac{q(\Psi | Y)}{\sum_{\Psi \in Q} q(\Psi | Y) \, \Delta \Psi}, \quad \Psi \in Q.
\]

- Possible because $\text{dim}(\Psi)$ small.
Posterior mean of $S$

By conditional expectation formula:

$$
E(S \mid Y) = E_{\Psi \mid Y}(E(S \mid Y, \Psi))
$$

$$
= \int_{\Psi} \mu^*(\Psi) \rho(\Psi \mid Y) \, d\Psi
$$

$$
\approx \sum_{\Psi \in Q} \mu^*(\Psi) \bar{\rho}(\Psi \mid Y).
$$
Posterior covariance of $S$

- $\Sigma^*(\Psi) = Q^*(\Psi)^{-1} = (Q(\Psi) + B^T C B)^{-1}$ is covariance of $S$ given $Y$ and $\Psi$.
- By conditional covariance formula:

$$\text{Cov}(S \mid Y) = \text{Cov}_{\psi \mid Y}(E(S \mid Y, \Psi)) + E_{\psi \mid Y}(\Sigma^*(\Psi))$$

$$= \int_\Psi \left[ (\mu^*(\Psi) - E(S \mid Y))(\mu^*(\Psi) - E(S \mid Y))^T \right] p(\Psi \mid Y) \, d\Psi$$

$$\approx \sum_{\Psi \in Q} \left[ (\mu^*(\Psi) - E(S \mid Y))(\mu^*(\Psi) - E(S \mid Y))^T \right] \tilde{p}(\Psi \mid Y) + \Sigma^*(\Psi)$$
Computational problems!

- To compute $q$ requires determinant and inverse of $Q^*(\Psi)$ — impossible! Too big! A 12.5 million $\times$ 12.5 million matrix!

- So far we’ve been able to implement on blocks of 512 pixels with 4 sources (so $Q^*(\Psi)$ is 2048 $\times$ 2048), and compute posterior expectations;

- WMAP data has $n_f = 5$ images of $J = 3 \times 2^{20} = 3.145.728$ pixels, divided into 6144 blocks of 512 pixels;

- Each block takes 40 seconds to compute $p(\Psi | Y)$ and $E(Y | S)$;
Computational problems!

- Total computation time is 3 days in MATLAB on an Intel Core2 Duo 3.16GHz, 3.25GB of RAM (but very parallelisable!);
- In the middle of implementing the posterior covariance formula (shouldn’t be a problem) for the same blocks;
- Currently we seem to need very strong priors on the $\phi_i$ for the calculation to be stable.
WMAP data again
$E(S|Y)$ with $E(\phi_i) = 10$

Clockwise from top left: CMB, synchotron, free-free emission, galactic dust.
$E(S | Y)$ with $E(\phi_i) = 5$

Clockwise from top left: CMB, synchotron, free-free emission, galactic dust.
$\mathbb{E}(S \mid Y)$ with $\mathbb{E}(\phi_i) = 1$

Clockwise from top left: CMB, synchotron, free-free emission, galactic dust.
Discussion

- Why do we need such a strong prior on the $\phi_i$?
- Model assessment e.g. residual analysis where we plot $Y$ against $E(\mathbf{B S} | Y)$;
- Alternative priors (GMM, IGMRFMM);
- Alternative ways to compute the answer — GPUs.
Reference