Contents

Chapter 1

Analysis Techniques

1.1 The Linear Gaussian Inversion [DONE1]

The *linear Gaussian inversion* (LGI)[Jakob: do you have you an original ref?] is an extremely useful numerical tool used in a specific situation to calculate the posterior distribution in a single relatively simple operation. Although occurances of the exact conditions are rare, many problems can be well approximated in a compatible way.

The method is applicable when the likelihood of the data vector $\underline{\mathrm{D}}$ can be expressed as a multivariate Gaussian with some known covariance $\underline{\underline{\sigma_{\mathrm{D}}}}$, and that the mean $\underline{\mathrm{D}_0}$ is a known linear combination of the unknown parameters $\underline{\mathrm{D}_0} = \underline{\underline{\mathrm{M}}}\,\underline{\mu}$. The *response matrix* $\underline{\underline{\mathrm{M}}}$ is the linear forward function in this case. The final requirement is that the prior must also be expressed as multivariate Gaussian with mean $\underline{\mu_{\mathrm{p}}}$ and covariance $\underline{\underline{\sigma_{\mathrm{p}}}}$:

$$
\begin{aligned}
P\left(\underline{\mathrm{D}} \,|\, \underline{\mu}\right) &= \mathcal{G}\left(\underline{\mathrm{D}};\ \underline{\underline{\mathrm{M}}}\,\underline{\mu},\ \underline{\underline{\sigma_{\mathrm{D}}}}\right) && (1.1)\\
P\left(\underline{\mu}\right) &= \mathcal{G}\left(\underline{\mu};\ \underline{\mu_{\mathrm{p}}},\ \underline{\underline{\sigma_{\mathrm{p}}}}\right) && (1.2)
\end{aligned}
$$

Through Bayes theorem, the posterior, with an arbitary normalisation constant C_1 is:

$$
\begin{aligned}
P\left(\underline{\mu} \,|\, \underline{\mathrm{D}}\right) &\propto P\left(\underline{\mathrm{D}} \,|\, \underline{\mu}\right) P\left(\underline{\mu}\right) \\
&\propto \mathcal{G}\left(\underline{\mathrm{D}};\ \underline{\underline{\mathrm{M}}}\,\underline{\mu},\ \underline{\underline{\sigma_{\mathrm{D}}}}\right) \mathcal{G}\left(\underline{\mu};\ \underline{\mu_{\mathrm{p}}},\ \underline{\underline{\sigma_{\mathrm{p}}}}\right) \\
\log P\left(\underline{\mu} \,|\, \underline{\mathrm{D}}\right) &= -\tfrac{1}{2}\left(\underline{\mathrm{D}} - \underline{\underline{\mathrm{M}}}\,\underline{\mu}\right)^T \underline{\underline{\sigma_{\mathrm{D}}}}^{-1} \left(\underline{\mathrm{D}} - \underline{\underline{\mathrm{M}}}\,\underline{\mu}\right) \qquad (1.3)\\
&\quad -\tfrac{1}{2}\left(\underline{\mu} - \underline{\mu_{\mathrm{p}}}\right)^T \underline{\underline{\sigma_{\mathrm{p}}}}^{-1} \left(\underline{\mu} - \underline{\mu_{\mathrm{p}}}\right) + C_1
\end{aligned}
$$

Since the posterior is a multiplication of the two Gaussians, it can be written as a single Gaussian with mean $\underline{\mu_0}$ and covariance $\underline{\underline{\sigma}}$:

$$
\begin{aligned}
P\left(\underline{\mu} \,|\, \underline{\mathrm{D}}\right) &\propto \mathcal{G}\left(\underline{\mu};\ \underline{\mu_0},\ \underline{\underline{\sigma}}\right) \\
\log P\left(\underline{\mu} \,|\, \underline{\mathrm{D}}\right) &= -\tfrac{1}{2}\left(\underline{\mu} - \underline{\mu_0}\right)^T \underline{\underline{\sigma}}^{-1} \left(\underline{\mu} - \underline{\mu_0}\right) + C_2 \qquad (1.4)
\end{aligned}
$$

Expanding equations 1.3 and 1.4 and equating like terms in $\underline{\mu}$:

$$\underline{\mu}^T \underline{\underline{\sigma}}^{-1} \underline{\mu} = \underline{\mu}^T \left[\underline{\underline{\mathrm{M}}}^T \underline{\underline{\sigma_{\mathrm{D}}}}^{-1} \underline{\underline{\mathrm{M}}} + \underline{\underline{\sigma_{\mathrm{p}}}}^{-1} \right] \underline{\mu} \quad (1.5)$$

$$\underline{\mu_0}^T \underline{\underline{\sigma}}^{-1} \underline{\mu} = \left[\underline{\mathrm{D}}^T \underline{\underline{\sigma_{\mathrm{D}}}}^{-1} \underline{\underline{\mathrm{M}}} + \underline{\mu_{\mathrm{p}}}^T \underline{\underline{\sigma_{\mathrm{p}}}}^{-1} \right] \underline{\mu} \quad (1.6)$$

$$\underline{\mu}^T \underline{\underline{\sigma}}^{-1} \underline{\mu_0} = \underline{\mu}^T \left[\underline{\underline{\mathrm{M}}}^T \underline{\underline{\sigma_{\mathrm{D}}}}^{-1} \underline{\mathrm{D}} + \underline{\underline{\sigma_{\mathrm{p}}}}^{-1} \underline{\mu_{\mathrm{p}}} \right] \quad (1.7)$$

$$\underline{\mu_0}^T \underline{\underline{\sigma}}^{-1} \underline{\mu_0} + C_2 = \underline{\mathrm{D}}^T \underline{\underline{\sigma_{\mathrm{D}}}}^{-1} \underline{\mathrm{D}} - \tfrac{1}{2} \underline{\mu_{\mathrm{p}}}^T \underline{\underline{\sigma_{\mathrm{p}}}}^{-1} \underline{\mu_{\mathrm{p}}} + C_1 \quad (1.8)$$

From equations 1.6 and 1.7, the posterior mean and covariance are:

$$\underline{\underline{\sigma}} = \left[\underline{\underline{\mathrm{M}}}^T \underline{\underline{\sigma_{\mathrm{D}}}}^{-1} \underline{\underline{\mathrm{M}}} + \underline{\underline{\sigma_{\mathrm{p}}}}^{-1} \right]^{-1} \quad (1.9)$$

$$\underline{\mu_0} = \underline{\underline{\sigma}} \left[\underline{\underline{\mathrm{M}}}^T \underline{\underline{\sigma_{\mathrm{D}}}}^{-1} \underline{\mathrm{D}} + \underline{\underline{\sigma_{\mathrm{p}}}}^{-1} \underline{\mu_{\mathrm{p}}} \right] \quad (1.10)$$

The power of the method can be seen in these last two equations - that the full mean and covariance of the posterior distribution can be calculated in a single matrix inversion given the inverse covariances of the prior and likelihood distributions and the response matrix.

In general (where the LGI conditions are not met), the posterior maximum is found by iterative numerical algorithms and the shape and extent investigated by the drawing random samples, a procedure which usually involves a gradual random walk. Both of these processes can take many thousands of times the parameter dimensionality of forward function evaluations to complete and for very high dimensionality ($N(\underline{\mu}) > \sim 200$) becomes prohibitively expensive.

For the LGI, the maximum posterior $\underline{\mu_0}$ is immediately avaiable and the drawing of random samples from a multivariate Gaussian is a well known, relatively trivial proceedure. With the general methods, the marginal distribution for a given subset of the parameters is calculated by the random walk algorithms. Here, the marginal distributions for each parameter are 1D Gaussians with variances given by the reciprocal of the diagonal elements of $\underline{\underline{\sigma}}$.

1.1.1 Practical Application and Parallel Implementation

Often, the forward function is easily written as a linear combination of the parameters and the coefficients are used to directly construct $\underline{\underline{\mathrm{M}}}$. For more complex cases, $\underline{\underline{\mathrm{M}}}$ is found by first

calculating the data $\underline{\mathrm{D_i}}$ for some initial set of parameters $\underline{\mu_\mathrm{i}}$ about which the forward function is assumed linear. Each parameter in turn is then modified and the relevant row of $\underline{\underline{\mathrm{M}}}$ filled with the difference between the new predicted data and $\underline{\mathrm{D_i}}$. The calculation requires only one evaluation of the forward function for each parameter so is relativly low cost.

However, in some cases this work requires repeated application of the LGI to very large problems ($N(\underline{\mu}) \sim N(\underline{\mathrm{D}}) > 4000$) with relatively slow forward functions ($t \sim 100ms$). To reduce the inversion time, a general parallel implementation of the LGI was developed. The determination of $\underline{\underline{\mathrm{M}}}$ is ideally suited to this since each machine can independantly calculate a selection of the rows. Once complete, $\underline{\mathrm{M}}$ is distributed over the involved machines so after distribution of $\underline{\underline{\sigma_\mathrm{D}}}^{-1}$, $\underline{\underline{\sigma_\mathrm{p}}}^{-1}$ and $\underline{\mu_\mathrm{p}}$, equations 1.9 and 1.10 are performed using freely avalaible parallel matrix libraries (PBLAS, BLACS, scaLAPACK [1], [2]). The developed software allows the calculation of such large problems in a few minutes, where the serial implemention can take hours.

1.1.2 Truncated Gaussians

In many cases where the priors are not Gaussian, they can be easily represented as truncated Gaussians. Typical examples are densities and temperatures where parameters are restricted to being positive $P(n_e) = 0$ for $n_e < 0$. In such cases, the LGI proceedure is applied as if the truncations were not present and the truncation simply applied directly to the posterior (There should also be a modification to the normalisation, but the correct normalisation of the posterior is rarely of any practical use).

If the Gaussian centre lies inside the truncation limits so $P(\underline{\mu_0}) \neq 0$ then it is the posterior maximum, otherwise the maximum will lie somewhere on one of the truncation hyper-planes and must be found by one of the general iterative algorithms.

The best known way of drawing random samples from the truncated posterior is a Monte-Carlo process, so is much slower than in the standard LGI, but it is much less costly than the general Monte-Carlo methods. The proceedure is based on the Gibbs-sampler [Jakob: do you have a ref for Gibbs in general??] where a random sample is drawn from the conditional distribution over one parameter given the current position in all others $P(\mu_i \,|\, \mu_{j\neq i})$. The sampler

moves to the position in μ_i of the sample and the procedure is repeated for each parameter in turn. The whole process repeated many times so that position, moving in steps along each axis, gradually explores the entire joint distribution.

The proceedure is favourable for the truncated multivariate Gaussian because the conditional distributions are always truncated univariate Guassians whose mean and variance are easily calcuated from $\underline{\mu_0}$ and $\underline{\underline{\sigma}}$ and methods exist to efficiently draw samples from these[3], so the overall proceedure is fairly efficient. While this has been done previously[4], the posteriors in this work are often very highly correlated making the parameter-by-parameter Gibbs sampling relativly slow to move along the correlations. To mitigate this, a new Gibbs-sampler was developed where each step is taken along the next eigenvector of $\underline{\underline{\sigma}}$. The variance of the conditional along this line (still univariate Gaussian) is given by the associated eigenvalue. The distance along the line where it intersects the truncation hyper-plane on each real parameter is found and the most constrictive of these give the limits on that conditional. For the situations where it is used in this work, the proceedure is significantly more efficient than that in [4].

The principal behind both the existing (parameter-space aligned) and the eigenvector aligned methods are shown in figure 1.1.

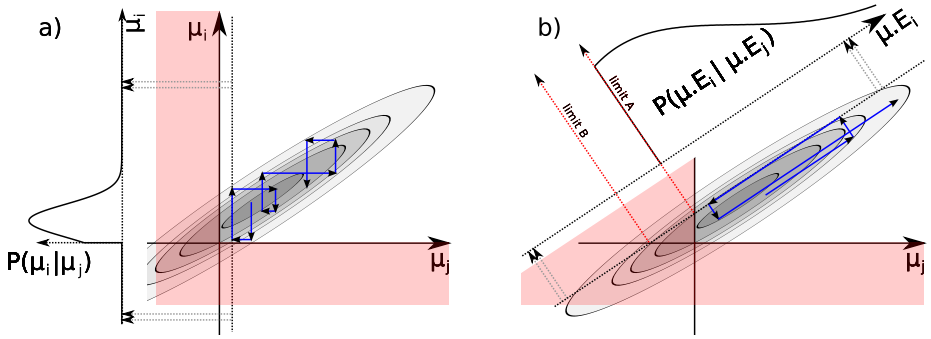

Figure 1.1: The principal of the Gibbs-sampler applied to a 2D truncated Gaussian using a) parameter-space aligned conditionals/steps as in [4] and b) using eigenvector aligned conditionals as in this work. The green arrowed lines show some arbitary example jumps and the 1D graphs show the conditional PDFs for one them in each case. $\underline{\mu}$ are the paremeters and $\underline{\mathrm{E}}$ the eigenvectors of the Gaussian covariance.

Bibliography

[1] Jack J. Dongarra and David W. Walker. Software libraries for linear algebra computations on high performance computers. *SIAM Rev.*, 37(2):151–180, 1995.

[2] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley. *ScaLAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

[3] Pseudo random variable generation. Technical report, University of Chicago, Econ 319, 2002.

[4] J. Geweke. Efficient simulation from the multivariate normal and student-t dsitributions subject to linear constraints and the evaluation of constraint probabilities. *?*, ?:?, 1991.