Bayesian computations and efficient algorithms for computing functions of large, sparse matrices

M. Ilić   I. W. Turner   A. N. Pettitt[*]

(received 8 August 2003, revised 30 January 2004)

Abstract

The need for computing functions of large, sparse matrices arises in Bayesian spatial models where the computations using Gaussian Markov random fields require the evaluation of $\mathbf{G}^{-1}$ and $\mathbf{G}^{-1/2}$ for the precision matrix $\mathbf{G}$ and in the geostatistical approach where approximations of $\mathbf{R}^{-1}$ and $\mathbf{R}^{1/2}$ are needed for the covariance matrix $\mathbf{R}$. In both cases, good approximations to the desired matrix functions are required over a range of probable values of a vector $\mathbf{v}$ drawn randomly from a given population, as occurs in simulation techniques for finding posterior distributions such as Markov chain Monte Carlo. Consequently, it is preferable that the complete matrix function approximation be determined rather than for its action on a given $\mathbf{v}$.

---

[*]School of Mathematical Sciences, Queensland University of Technology, Australia
mailto:i.turner@fsc.qut.edu.au

The aim of this work is to find low degree polynomial approximations $p(\mathbf{A})$ such that $e = \|f(\mathbf{A}) - p(\mathbf{A})\|_2$ is small in some sense on the spectral interval $[a, b]$, where the extreme eigenvalues a and b are calculated using Krylov subspace approximation. Algorithms based on low order near-minimax polynomial approximations are proposed for the required matrix functions for a typical case study in computational Bayesian statistics, where a good balance between accuracy and computationally efficiency is achieved.

Contents

1 Introduction

In Bayesian spatial models using a generalised linear mixed model hierarchical structure [4], Gaussian Markov random field GMRF models are often

utilised. The distribution can be defined explicitly so that the density function for the Gaussian random n-variable $\mathbf{y}$ is given by $p\left(\mathbf{y}\right) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T \mathbf{G}\mathbf{y}\right)$ for the symmetric $n \times n$ matrix $\mathbf{G}$, which is known as the precision matrix. For the GMRF, $\mathbf{G}$ is defined to be sparse with off-diagonal elements being non-zero for so-called spatial neighbours. An example of such a structure arises in species abundance mapping when data are observed for regions of a country and neighbours are defined in terms of geographical adjacencies [16]. Let $\mathcal{N}_i$ denote the set of neighbours of the ith region or, more generally, node or vertex i (since the definition can be extended to a undirectional graph), then $G_{ij} = 0$ *iff* $j \notin \mathcal{N}_i$. The sparseness, or Markov property, is demonstrated by considering the full conditional distributions $p\left(\mathbf{y}_i \mid \mathbf{y}_{-i}\right) \propto p\left(\mathbf{y}\right)$ where $\mathbf{y}_{-i}$ denotes the vector $\mathbf{y}$ without y_i. In many applications [4, e.g.] the pair-wise difference distribution

$$p\left(\mathbf{y}\right) \propto \exp\left\{-\frac{1}{2}\gamma \sum_{i,j \in \mathcal{N}_i} (y_i - y_j)^2\right\}$$

has been used which defines $\mathbf{G}$ to be singular, whereas Pettitt et al. [10] introduce a class of distributions where $\mathbf{G}$ is positive definite. The approach utilising a GMRF appears reasonable when a spatial distribution is being defined for a fixed discrete structure such as a map of regions and centroids are considered as sites. However, when the distribution is defined on a continuous set, then generally a Gaussian random field is defined in terms of moments, means and covariances. The latter are defined in terms of the distance d between two locations. Often it is sufficient to take the covariances to be stationary and isotropic. For example, Diggle et al. [5] use the covariance function $\sigma^2 \exp\left(-d^\delta/r\right)$, while the spherical function

$$\sigma^2 \begin{cases} 1 - \frac{3d}{2r} + \frac{d^3}{2r^3}, & d \leq r, \\ 0, & d > r, \end{cases}$$

gives zero covariance for locations at greater distance than r apart. This approach is sometimes called the geostatistical approach where the $n \times n$ covariance matrix $\mathbf{R}$ is defined explicitly. If $\mathbf{R}$ is positive definite then $\mathbf{G} =$

$\mathbf{R}^{-1}$. Rue [11] gives details of some of the computations required when using a GMRF. These include sampling from a Gaussian density $p(\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{G}\mathbf{y} + \mathbf{b}^T\mathbf{y}\right)$ with $\mathbf{G}$ and $\mathbf{b}$ explicitly defined. A general approach is to generate a vector $\mathbf{z}$ of independently distributed normal, zero mean, unit variance random variables and construct $\mathbf{G}^{-1}\mathbf{b} + \mathbf{G}^{-1/2}\mathbf{z}$, noting $\mathbf{y}$ has mean $\mathbf{G}^{-1}\mathbf{b}$. The approach in [11] is to consider the Cholesky decomposition $\mathbf{G} = \mathbf{L}\mathbf{L}^T$ and solve the systems $\mathbf{L}\mathbf{v} = \mathbf{b}$, $\mathbf{L}^T\mu = \mathbf{v}$ and $\mathbf{L}^T\mathbf{e} = \mathbf{z}$ and sample $\mathbf{y} = \mu + \mathbf{e}$. Thus in matrix terms, one can see that these matrix computations involve the operators $\mathbf{G}^{-1}$ and $\mathbf{G}^{-1/2}$.

When the geostatistical approach is adopted and $\mathbf{R}$ is defined explicitly, then Bayesian computations require explicit evaluation of the likelihood and full-conditionals and simulation from the distributions. For these we require the quadratic form $\mathbf{y}^T\mathbf{R}^{-1}\mathbf{y}$ and $\mathbf{R}^{1/2}\mathbf{z}$, and again we see the need to consider the matrix function $\mathbf{R}^{-1}$ and, in this case, $\mathbf{R}^{1/2}$.

Note that in both approaches discussed above, good approximations to the desired matrix functions are required over a range of probable values of a vector, $\mathbf{z}$, drawn randomly from a given population. A popular method for approximating $f(\mathbf{A})\mathbf{v}$ is to use Arnoldi (Lanczos for symmetric matrices) decomposition [6, 13]:

$$\mathbf{A}\mathbf{Q}_m = \mathbf{Q}_m\mathbf{H}_m + \beta_m\hat{\mathbf{q}}_{m+1}\mathbf{e}_m^T,$$

where $\mathbf{Q}_m$ is an ON basis for the m-dimensional Krylov subspace $\mathcal{K}_m(\mathbf{A}, \mathbf{v})$ generated by $\mathbf{A}$ and $\mathbf{v}$. Thus, $f(\mathbf{A})\mathbf{v} \approx \mathbf{Q}_m f(\mathbf{H}_m)\mathbf{e}_1$ where $\mathbf{v} = \mathbf{Q}_m\mathbf{e}_1$. However, such strategies do not provide the matrix function $f(\mathbf{A})$, only the result of its action on vector $\mathbf{v}$. In this work preference is given to determining low order polynomial approximations $p(\mathbf{A})$ to the complete matrix function, which can be used thereafter to compute $p(\mathbf{A})\mathbf{v}$ for a number of vectors $\mathbf{v}$ drawn randomly from some population.

For the class of statistical matrices introduced above it can be shown that the spectrum $\sigma(\mathbf{A}) \subset [1, 2]$, which for the functions $f(t) = t^{-1}$, $t^{-1/2}$, $t^{1/2}$ enables low order *near-minimax* polynomials to be exploited for the matrix

function approximation. Hence, all that is required is to obtain the extreme eigenvalues and to fit a polynomial approximation $p(\mathbf{A})$ to the function. The former is discussed in Section 2 by using Krylov subspace methods and thick restart for refinement of the approximate eigenpairs. The latter is investigated in Section 3 via two key strategies — the first uses least squares approximations based on Chebyshev polynomials and the second uses interpolation at the Chebyshev nodes. The performance of the proposed algorithms for typical large sparse precision matrices is given in Section 4 and finally, the main conclusions of the work are presented in Section 5.

2   Krylov subspace approximation

As mentioned in the introduction, one needs to find the smallest and largest (extreme) eigenvalues of a large, sparse symmetric positive definite matrix. The best known method to do this is to use a Krylov subspace and although ARPACK [8] could be used to determine the desired eigenvalues, we have opted here to use a customised algorithm based on Thick restart that rapidly converges to the extreme eigenvalues of the matrix in question. Let $\mathcal{K}_\ell(\mathbf{A}, \mathbf{v}) = \mathrm{span}\left\{\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{\ell-1}\mathbf{v}\right\}$ be the Krylov subspace generated by $\mathbf{v}$, where ℓ is the chosen analytic grade. The analytic grade was introduced in [7] as a useful indicator of the effectiveness of the iterative process. The Arnoldi process produces the decomposition

$$\mathbf{A}\mathbf{Q}_\ell = \mathbf{Q}_\ell\mathbf{H}_\ell + \beta_\ell\hat{\mathbf{q}}_{\ell+1}\mathbf{e}_\ell^T \,, \tag{1}$$

where for symmetric, positive definite $\mathbf{A}$ produces a Hessenberg matrix $\mathbf{H}_\ell$ that is symmetric tridiagonal and positive definite. The Lanczos algorithm that is usually used to accomplish this decomposition invariably produces vectors that quickly lose their orthogonality and strategies like reorthogonalization have to be implemented [12]. To avoid this problem, the Householder algorithm as explained in [17] was used. Extreme eigenvalues of $\mathbf{A}$ are approximated by the extreme eigenvalues of $\mathbf{H}_\ell$.

2.1 Refinement using Thick restart

To obtain the extreme eigenvalues more accurately, filter polynomials as discussed in [2] can be used. We have found that the following thick restart procedure works well (see [3, 15] for more details on restarting). Let

$$\mathbf{H}_\ell \mathbf{Y}_\ell = \mathbf{Y}_\ell \mathbf{\Lambda}_\ell \,,$$

be the eigenvalue decomposition of $\mathbf{H}_\ell$, where $\mathbf{\Lambda}_\ell$ is real and diagonal and $\mathbf{Y}_\ell$ is ON. Since $\mathbf{H}_\ell$ is of small dimension, one can obtain this eigenvalue decomposition using standard methods. Next, choose k extreme eigenvalues (for example, a cluster of the smallest and largest eigenvalues at the ends of the spectrum). Set $\mathbf{P}_k = \mathbf{Q}_\ell \mathbf{Y}_k$. It is straightforward to show that

$$\begin{aligned}
\mathbf{A}\mathbf{P}_k &= \mathbf{P}_k \mathbf{\Lambda}_k + \beta_\ell \hat{\mathbf{q}}_{\ell+1} \mathbf{s}_k^T \qquad \mathbf{s}_k = \mathbf{Y}_k^T \mathbf{e}_\ell \\
&= [\mathbf{P}_k, \hat{\mathbf{q}}_{\ell+1}] \begin{bmatrix} \mathbf{\Lambda}_k \\ \beta_\ell \mathbf{s}_k^T \end{bmatrix} \,.
\end{aligned}$$

Obtain the Householder transformations $\mathcal{H}_i$ so that

$$\mathcal{H}_k \cdots \mathcal{H}_1 \mathbf{P}_k = \mathbf{E}_k \quad \text{or} \quad \mathbf{P}_k = \mathcal{H}_1 \cdots \mathcal{H}_k \mathbf{E}_k \,,$$

where $\mathbf{E}_k$ denotes the first k columns of $\mathbf{I}_n$. To restart the process, calculate $\mathcal{H}_{k+1}$ for $\mathcal{H}_k \cdots \mathcal{H}_1 \hat{\mathbf{q}}_{\ell+1}$ and set

$$\hat{\mathbf{p}}_{k+1} = \hat{\mathbf{q}}_{\ell+1} = \mathcal{H}_1 \cdots \mathcal{H}_{k+1} \mathbf{e}_{k+1} \,.$$

Then as usual: compute $\mathcal{H}_{k+2}$ for $\mathcal{H}_{k+1} \cdots \mathcal{H}_1 \mathbf{A} \hat{\mathbf{p}}_{k+1}$ and set

$$\mathbf{h}_{k+2} = \mathcal{H}_{k+2} \cdots \mathcal{H}_1 \mathbf{A} \hat{\mathbf{p}}_{k+1} = \beta_\ell \mathbf{s}_k \,.$$

Define $\hat{\mathbf{p}}_{k+2} = \mathcal{H}_1 \cdots \mathcal{H}_{k+2} \mathbf{e}_{k+2}$. Expand up to ℓ to reproduce equation (1) with $\mathbf{H}_\ell$ replaced by a matrix that is symmetric and tridiagonal, except for the principal $(k+1) \times (k+1)$ leading submatrix. This submatrix has the current approximations of the extreme eigenvalues on its diagonal and

$\mathbf{h}_{k+2}^T$ in the $(k+1)$st row and $\mathbf{h}_{k+2}$ in the $(k+1)$st column. Note that care must be taken with the construction of $\mathcal{H}_1, \ldots, \mathcal{H}_k$ to ensure that the signs of the eigenvalues remain positive. Thereafter, the extreme eigenpairs of this matrix are computed and the entire restart process is repeated until $\|\mathbf{A}\mathbf{p}_i - \lambda\mathbf{p}_i\| = |\beta_\ell| \left|\mathbf{e}_\ell^T\mathbf{y}_i\right| \leq \varepsilon$, $i = 1, n$ is satisfied for the smallest $(\lambda_1, \mathbf{p}_1)$ and largest $(\lambda_n, \mathbf{p}_n)$ eigenpairs.

3   Polynomial approximation

In this section numerical strategies for approximating $f(\mathbf{A})$ using Chebyshev polynomials which provide near-minimax approximations are analysed. We begin with the following proposition.

Proposition 1 *For positive definite matrix $\boldsymbol{A}$*

$$\|f(\mathbf{A}) - p(\mathbf{A})\|_2 = \max_{\lambda_i}|f(\lambda_i) - p(\lambda_i)| \leq \max_{a \leq t \leq b}|f(t) - p(t)| = \|f - p\|_\infty.$$

This proposition shows that the polynomial p that interpolates f at all eigenvalues λ_i would be exact. However, this would be impractical for matrices that may have thousands of eigenvalues. One option is to interpolate on a handful of eigenvalues chosen at the ends of the spectrum. According to the usual wisdom the interpolation error at the omitted eigenvalues in the middle of the spectrum will be minimised. An alternative is to construct polynomial approximations that may not give an exact match at any point on the interval, but do provide a good approximation at every point in the interval. Ideally for a given degree n, the polynomial approximation q_n should have the best possible accuracy, that is, $\|f - q_n\|_\infty = \inf_{\deg(p) \leq n} \|f - p\|_\infty$. These so called minimax polynomial approximations (also called best uniform approximations or Chebyshev approximations) are quite difficult to produce [1]. An intermediate approximation is the least squares approximation, which is known to be a fairly good uniform approximation [1].

3.1   Near-minimax approximations

Although there is a method called the *Remes algorithm* [1] that will compute a best approximation to high accuracy, the following two *near-minimax* approximations give a good enough estimate of the best polynomial approximation.

3.1.1   Interpolation at the Chebyshev Nodes

If $p_n(t)$ is the polynomial of degree $\leq n$ that interpolates $f(t)$ at the zeros of the Chebyshev polynomial $T_{n+1}(t)$ on $[-1, 1]$, then [1] gives

$$\|f - p_n(t)\|_\infty \leq \frac{1}{(n+1)!\, 2^n} \left\| f^{(n+1)} \right\|_\infty .$$

This approximation can be constructed by either linearly mapping the interval $[a, b]$ onto $[-1, 1]$ and using standard results, or by mapping the zeros of the Chebyshev polynomial on $[-1, 1]$ to the interval $[a, b]$. We choose the latter strategy and map the Chebyshev nodes $\bar{t}_j = \cos(2j + 1\pi/k)$, $j = 0, 1, \ldots, k$ to the interval $[a, b]$ using the linear transformation $t_j = \frac{1}{2}[(b-a)\bar{t}_j + (b+a)]$. The above error bound on $[a, b]$ becomes

$$E_n^{(1)} = \frac{1}{(n+1)!\, 2^n} \left(\frac{b-a}{2} \right)^{n+1} \left\| f^{(n+1)} \right\|_\infty .$$

3.1.2   Chebyshev Least Squares Approximation

If $C_n(t)$ is the least squares approximation, that is, $C_n(t) = \sum_{i=0}^{n}{}' c_i T_i(t)$, with the coefficients

$$c_i = \frac{2}{\pi} \int_{-1}^{1} \frac{f(t)\, T_i(t)}{\sqrt{1 - t^2}}\, dt ,$$

and the prime on the summation indicating that the first term ($i = 0$) should be halved, then [1] gives

$$\|f - C_n(t)\|_\infty \leq \left(4 + \frac{4}{\pi^2}\ln n\right) E_n^{(1)} = E_n^{(2)}\,.$$

In this work the Maple software[1] has been used to evaluate the coefficients c_i. However, note that these coefficients could be approximated using trapezoidal integration (see for example Snyder [14]).

4   Results and discussion

In this section the accuracy of the matrix function approximations outlined in Section 3 for computing the function of a large sparse matrix with compact spectrum is studied for a typical case study in computational Bayesian statistics. An application to tree biodiversity is presented, where the covariance matrix and its square root require approximation. The chosen case study is motivated by a data set that arose in collaborative work with the Queensland Department of Natural Resources [9]. The data involved the presence or absence of a number of native tree species at about 24,000 sites in south-east Queensland. A number of explanatory variables, such as soil type, rainfall and terrain, were available for each site. A computational approach to make a Bayesian solution feasible on a subset of the data of 474 sites uses the precision matrix **G** generated as follows (see for example [10]):

Define $\mathcal{N}_i = \{j \neq i \mid j \text{ is in a neighbourhood of } i\}$ and n_i to be the cardinality of $\mathcal{N}_i$, then

---

[1]©Maplesoft, Waterloo Maple Inc.

- Form the sparse matrix $\mathbf{C}$ such that

$$C_{ij} = \begin{cases} n_i, & i = j\,, \\ -1, & j \in N_i\,, \\ 0, & j \notin N_i\,. \end{cases}$$

  Note that the matrix $\mathbf{C}$ is rank 1 deficient, and thus symmetric positive semidefinite.

- Form the precision matrix $\mathbf{G} = \mathbf{I} + \gamma\mathbf{C}$, with γ chosen to ensure that $\mathbf{G}$ has the spectrum $\sigma(\mathbf{G}) \subset [1, \alpha]$, $\alpha > 1$.

The neighbourhood consisted of all sites within a $1\,\mathrm{km}$ radius of site i. The precision matrix $\mathbf{G}$ gives independence when $\gamma = 0$ and a degree of spatial dependence for positive γ. Alternative definitions of the precision matrix might take the distance between sites explicitly into account. For a Bayesian analysis, the computational requirement is to approximate $\mathbf{G}^{-1}$ and $\mathbf{G}^{-1/2}$. It can be shown that the best choice of the parameter γ to ensure positive definiteness and to guarantee that the spectrum $\sigma(\mathbf{G}) \subseteq [1, \alpha]$ is

$$\gamma \leq \frac{\alpha - 1}{\|\mathbf{C}\|_\infty} \leq \frac{\alpha - 1}{\lambda_{max}}, \quad \alpha > 1\,.$$

The choice of $\gamma = 1/\|\mathbf{C}\|_\infty$ led to a precision matrix of dimension 474×474 having its spectrum $\sigma(\mathbf{G}) \subseteq [1, 1.543]$ and $\|\mathbf{G}\|_\infty = 2$. However, although this particular restriction on γ provided good matrix function approximations, see Tables 1–2, there is some scope in the choice of γ to enable good approximations for a larger parameter space. However, for larger values of γ the precision matrix $\mathbf{G}$ tends to the singular pair-wise difference distribution described in the Introduction.

Tables 1–2 present the main findings of the study. The tables list the method used for the approximation, the measures of accuracy $\|f(\mathbf{G}) - p(\mathbf{G})\|_2$, $\|f - p\|_\infty$ and either $E_n^{(1)}$ or $E_n^{(2)}$. Note that $f(\mathbf{G}) = \mathbf{S}f(\mathbf{D})\mathbf{S}^{-1}$ uses the complete diagonalisation of the original matrix $\mathbf{G}$ available from MATLAB.

TABLE 1: Summary of errors for precision matrix function $f(\mathbf{G}) = \mathbf{G}^{-1}$.

Approximation method	$\|f(\mathbf{G}) - p(\mathbf{G})\|_2$	$\|f - p\|_\infty$	$E_n^{(1)}$ or $E_n^{(2)}$
ChebyshevLS-2	2.343420e-3	2.343420e-3	2.141839002e-2
ChebyshevN-2	2.599147e-3	2.599147e-3	5.003218970e-3
ChebyshevLS-3	2.557314e-4	2.557314e-4	3.019156437e-3
ChebyshevN-3	2.836387e-4	2.836387e-4	6.791869750e-4
ChebyshevLS-4	2.790731e-5	2.790731e-5	4.206003321e-4
ChebyshevN-4	3.095276e-5	3.095276e-5	9.219963188e-5

TABLE 2: Summary of errors for precision matrix function $f(\mathbf{G}) = \mathbf{G}^{-1/2}$.

Approximation method	$\|f(\mathbf{G}) - p(\mathbf{G})\|_2$	$\|f - p\|_\infty$	$E_n^{(1)}$ or $E_n^{(2)}$
ChebyshevLS-2	1.950813e-4	1.950813e-4	6.693246879e-3
ChebyshevN-2	2.084965e-4	2.084965e-4	1.563505928e-3
ChebyshevLS-3	1.341534e-5	1.341534e-5	8.255505884e-4
ChebyshevN-3	1.444581e-5	1.444581e-5	1.857151885e-4
ChebyshevLS-4	1.030461e-6	1.030461e-6	1.035071130e-4
ChebyshevN-4	1.115135e-6	1.115135e-6	2.268975316e-5

The Householder based Lanczos scheme outlined in Section 2 was continued until an $\varepsilon_{\mathrm{grade}} = 10^{-25}$ was reached, which is the value given in [7] for determining the analytic grade of the Krylov subspace $\mathcal{K}_\ell(\mathbf{G}, \mathbf{v})$. For this precision matrix $\mathbf{G}$ the analytic grade was estimated as $\ell = 26$ and it took three further restarts to converge the desired extreme eigenpairs to an accuracy of less that 10^{-10}. A randomly chosen vector $\mathbf{v}$ was used to start the subspace generation.

The results exhibited in the tables highlight that the best approximation for both matrix functions $f(\mathbf{G})$ is offered by the Chebyshev least squares method, followed by interpolation using the Chebyshev nodes, which contradicts the predicted theoretical bounds given by $E_n^{(1)}$ and $E_n^{(2)}$. Undoubtedly the good agreement between the polynomial approximations and the exact result can be attributed to the rather short interval (approximately $[1, 1.543]$) over which the approximate polynomial $p(\mathbf{G})$ was constructed. The least squares approximations are particular suitable in this case when one notes that there are 474 eigenvalues confined to this interval. Another conclusion drawn from the tables is that the 4th degree polynomials provide quite accurate approximations, however one could easily justify the use of the quadratic and cubic polynomial approximations because it is felt that the errors offered by these low degree matrix polynomial approximations are acceptable for most applications in computational Bayesian statistics. Most importantly however, the computational effort required to evaluate the quadratic or cubic $p(\mathbf{G})$ would be minimal. Clearly, one must balance the accuracy and computational efficiency of these approximations when addressing the underlying statistical problem. For the precision matrix under study in this section the Chebyshev cubic least squares approximation is highly recommended.

5 Conclusions

We proposed methods for computing functions of large sparse matrices that arise in Bayesian statistical modelling. The theory and algorithms presented throughout the text are well suited to such matrices, since they are frequently symmetric positive definite with fairly compact spectrum. For the class of sparse matrices examined here, the thick restarted householder based Lanczos scheme enabled rapid convergence to the extreme eigenvalues. These Ritz values are used to construct low order polynomial approximations. The theoretical bounds $E_n^{(1)}$ and/or $E_n^{(2)}$ provided the perfect mechanism by which to gauge the expected accuracy of the required matrix functions approximations by analysing norms involving only scalar functions.

The accuracy of the computed polynomial approximations was measured against the complete diagonalisation method. It was found that the near-minimax Chebyshev least squares approximation of cubic order offers a good balance between accuracy and computational effort. We suggest that any of the algorithms given in this paper can be used with confidence in Bayesian statistical modelling for the purposes of matrix function approximation. In terms of simplicity and ease of implementation, interpolation at the Chebyshev nodes is highly recommended.

References

[1] K. E. Aitkinson. *An introduction to Numerical Analysis second edition*, Wiley, 1989. C510, C511, C512

[2] J. Baglama, D. Calvetti, G. H. Golub and L. Reichel. Adaptively Preconditioned GMRES Algorithms, *Siam J. Sci. Comput.*, vol 20, 1, pp.243–269, 1998. C509

[3] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe and H. van der Vorst, editors. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide.* SIAM, Philadelphia, 2000. C509

[4] J. E. Besag and D. Higdon. Bayesian analysis of agricultural field experiments (with discussion), *J. Roy. Statist. Soc. B*, 61, 691–746, 1999. C505, C506

[5] P. J. Diggle, J. A. Tawn and R. A. Moyeed. Model-based geostatistics (with discussion). *Appl. Statist.*, 47, 299–350, 1998. C506

[6] V. L. Druskin and L. A. Knizhnerman. Krylov subspace approximations of eigenpairs and matrix functions in exact and computer arithmetic. *Numerical Linear Algebra with Applications*, 2:205–217, 1995. C507

[7] M. Ilic and I. W. Turner, Krylov Subspaces and the Analytic Grade, submitted to the *Journal Numerical Linear Algebra and its Applications*, 2002. C508, C515

[8] R. B. Lehoucq, D. C. Sorensen and C. Yang, ARPACK Users' Guide Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods SIAM Publications, Philadelphia, USA, 1998. C508

[9] B. McCormack. Timber inventory manual for the native forests of Queensland. Technical report, Queensland Department of Primary Industries - Forest Services, Brisbane, Queensland, 1995. C512

[10] A. N. Pettitt, I. S. Weir and A. Hart. A conditional autoregressive Gaussian process for irregularly spaced multivariate data, *Statistics & Computing* 12, 353–367, 2002. C506, C512

[11] H. Rue. Fast sampling of Gaussian Markov random fields. *J. R. Statist. Soc. B* 63, 325–338, 2001. C507

[12] A. Ruhe. Lanczos Method (Section 4.4). In Z. Bai, J. Demmel, J. Dongarra, A. Ruhe and H. van der Vorst, editors, *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide.* SIAM, Philadelphia, 2000. C508

[13] Y. Saad. *Iterative Methods for Sparse Linear Systems*, PWS publishing Company, ITS, 1996. C507

[14] M. A. Snyder, Chebyshev Methods in Numerical Approximation, Prentice Hall, Englewood Cliffs, New Jersey USA, 1996. C512

[15] A. Stathopoulos, Y. Saad, and K. Wu. Dynamic Thick Restarting of the Davidson, and the Implicitly Restarted Arnoldi Methods, SIAM J. Scientific Computing. (19:1) pp.229–245, 1998. C509

[16] I. S. Weir and A. N. Pettitt. Binary probability maps using a hidden conditional autoregressive Gaussian process with an application to Finnish common toad data. *Appl. Statist.*, 49, 473–484, 2000. C506

[17] H. F. Walker and L. Zhou. A Simpler GMRES, Numer. *Linear Algebra Appl.*, 1, pp.571–581, 1994. C508