XIX International Conference on Water Resources
CMWR 2012
University of Illinois at Urbana-Champaign
June 17-22,2012

DIMENSIONALITY REDUCTION IN THE GEOSTATISTICAL APPROACH FOR HYDRAULIC TOMOGRAPHY

Arvind K. Saibaba[*], **Peter K. Kitanidis**[*,†]

[*]Institute for Computational and Mathematical Engineering
[†]Department of Civil and Environmental Engineering
Stanford University, CA-94305

Key words: Inverse problems, Karhunen Loeve expansion, Gauss-Newton method.

Summary. The implementation of geostatistical approach to solve inverse problems [8], such as estimating hydraulic conductivity from measurements of head, is expensive for problems discretized on fine grids. Dimensionality reduction techniques such as representing the random field via Karhunen-Loève expansion, are frequently used in such a context. We show how to combine an efficient method to compute the expansion on an unstructured grid using Hierarchical matrices and a Gauss-Newton-Krylov approach for solving the inverse problem. We present preliminary results on a synthetic model problem arising from Hydraulic tomography.

1 Introduction

Hydraulic tomography is an aquifer characterization approach that uses a joint analysis of distributed aquifer pressure (head) response data collected during a series of pumping tests to produce estimates of aquifer variability of properties (for eg. conductivity, specific storage). A popular method to deal with such problems is the Geostatistical approach. Performing hydraulic tomography using the Geostatistical approach is a computationally challenging problem. In this approach, inference is made through the posterior distribution of the parameters, which is composed of the likelihood of the measurements, which controls the misfit of the data and the second term, the prior distribution, which usually controls the structure of the solution. The second term, is typically enforced via a dense covariance matrix, which can be computationally expensive to perform operations such as matrix-vector and matrix-matrix products.

The bottleneck for scalability of computational algorithms are two-fold: 1) Working with the covariance matrix, corresponding to a discretized spatial random variable, that is large and dense, and 2) computing the Jacobian, which is computationally infeasible, when the number of measurements are high. For structured grids, discrete spectral methods have had great success in reducing the computational and storage costs involving

matrix products of the covariance matrix. Many practical applications, need an approach that works on unstructured grids. We use the Karhunen-Loève expansion (KLE), in which the covariance function corresponding to the spatially distributed random variable is expanded in terms of basis functions. The inverse problem is then to compute the weights of these basis functions. For the second issue, we avoid explicit computation of the Jacobian. After appropriate discretization, the reduced Hessian system is solved using an appropriate Krylov subspace method .

The KLE is a way to represent the prior information of the Gaussian random field. The random spatial variable is parametrized by a series of eigenfunctions, weighted by the square root of the eigenvalues, which are derived from an integral eigenvalue problem involving the covariance function. The eigenvalues are known to decay depending on the smoothness of the kernel [13] and hence, the expansion can be truncated, resulting in an approximation to the random field. The number of eigenvalues retained are typically much smaller than the number of grid points after discretization, and this may result in significant reduction in computational cost. Previous attempts at using the KLE for dimensionality reduction of the random field in the context of inverse problems include [9, 10]. Our approach for solving the inverse problem, having computed the KLE closely follows [9] but the method of computing the KLE follows [7], employing Hierarchical matrix approach to accelerate matrix vector products involving the dense covariance matrix.

2 Problem Formulation

2.1 Karhunen-Loève representation

Consider the random field $s(\mathbf{x})$, with mean $\mu(\mathbf{x})$ and covariance $\kappa(\mathbf{x}, \mathbf{y})$, on the bounded domain $\mathbf{x} \in \mathcal{D}$. By the assumptions on s, the covariance kernel is symmetric and positive semi-definite. The KLE can now be written as

$$s(\mathbf{x}) = \mu(\mathbf{x}) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\phi_i(\mathbf{x})\xi_i \quad \text{with,} \tag{1}$$

$$\mu(\mathbf{x}) = E[s(\mathbf{x})], \qquad \xi_i = \frac{1}{\sqrt{\lambda_i}} \int_{\mathcal{D}} (s(\mathbf{x}) - \mu(\mathbf{x}))\phi_i(\mathbf{x})$$

Here, ξ_i are uncorrelated random variables, $(\lambda_i, \phi_i(\mathbf{x}))$ are the eigenpair obtained as the solution to the Fredholm integral equation of the second kind

$$\int_{\mathcal{D}} \kappa(\mathbf{x}, \mathbf{y})\phi(\mathbf{y})d\mathbf{y} = \lambda\phi(\mathbf{x}) \tag{2}$$

Since the covariance $\kappa(\cdot, \cdot)$ is symmetric and positive definite, the eigenfunctions $\phi_i(\cdot)$ are mutually orthogonal and form a basis for $L^2(\mathcal{D})$ and the eigenvalues λ_i are real, non-negative and can be arranged in decreasing order $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. If the random field

is Gaussian, then $\xi_i \sim \mathcal{N}(0,1)$. The KLE is typically truncated to a finite number terms K, typically far fewer than the number of grid points and independent of it. The number of terms retained in the series depends on the decay of the eigenvalues, which, in turn depends on the smoothness of the covariance kernel [13].

The eigenpair $\lambda_i, \phi_i(\mathbf{x})$ in the KLE, can be computed by first discretizing the weak form of system of equations 2 (i.e. performing a Galerkin projection) using piecewise linear basis functions and, subsequently solving the linear eigensystem using a generalized eigenvalue solver for symmetric matrices, that requires only matrix-vector products involving the discretized operator. The matrix-vector products involving the dense covariance matrix can be computed in $\mathcal{O}(N \log N)$, where N is the number of grid points after discretization, using the $\mathcal{H}$-matrix approach. For further details, the reader is referred to [7].

2.2 Geostatistical approach

In this section, we review the geostatistical approach for solving inverse problems. After discretization (e.g., using finite differences, or finite element models), the random field $s(\mathbf{x})$ is represented by the vector $\mathbf{s} \in \mathbb{R}^m$, from a set of noisy measurements $\mathbf{y} \in \mathbb{R}^n$

$$\mathbf{y} = \mathbf{h}(\mathbf{s}) + \mathbf{v} \qquad \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{3}$$

where, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ is the measurement operator that yields data as a function of parameters and $\mathbf{y} \in \mathbb{R}^m$ is a set of observable quantities. Thus, $p(\mathbf{y}|\mathbf{s}) \propto \exp\left(-\frac{1}{2}\|\mathbf{y} - \mathbf{h}(\mathbf{s})\|_{\mathbf{R}^{-1}}\right)$, where $\|\mathbf{x}\|_{\mathbf{M}} = \mathbf{x}^T\mathbf{M}\mathbf{x}$ is a vector norm, when $\mathbf{M}$ is symmetric positive definite. Using Bayes' rule, we can define the posterior probability density of the parameters $\mathbf{s}$, given the measurements $\mathbf{y}$

$$p(\mathbf{s}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{s})p(\mathbf{s})$$

We focus our attention on Gaussian prior for $s(\mathbf{x})$, with mean $\mu(\mathbf{x})$ and covariance $\kappa(\mathbf{x}, \mathbf{y})$. We use the K term KLE for $s(\mathbf{x})$, so that the truncated expansion, $s_K(\mathbf{x}) = \mu(\mathbf{x}) + \sum_{i=1}^{K} \sqrt{\lambda_i}\phi_i(\mathbf{x})\xi_i$, converges to $s(\mathbf{x})$ in both pointwise and in mean-squared sense as $K \to \infty$. After appropriate discretization, denoting by $\mathbf{s}_K$ and $\boldsymbol{\mu}$, the discrete representation of $s_K(\mathbf{x})$ and $\mu(\mathbf{x})$ respectively, we have $\mathbf{s}_K = \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{\xi}$, where the columns of $\boldsymbol{\Phi}$ correspond to the discrete representation of the weighted eigenfunctions $\sqrt{\lambda_i}\phi_i(\mathbf{x})$ for $i = 1, \ldots, K$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)^T$. We are now in a position to write the posterior probability density for $\boldsymbol{\xi}$,

$$\begin{aligned} p(\boldsymbol{\xi}|\mathbf{y}) &\propto p(\mathbf{s}_K(\boldsymbol{\xi})|\mathbf{y})p(\boldsymbol{\xi}) \\ &\propto p(\mathbf{s}_K(\boldsymbol{\xi})|\mathbf{y}) \exp\left(-\frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi}\right) \end{aligned} \tag{4}$$

The posterior mean $\hat{\boldsymbol{\xi}}$, obtained by maximizing the negative log-likelihood of the posterior

distributions, is given by the maximum a posteriori (MAP) estimate

$$\arg\min_{\boldsymbol{\xi}} \frac{1}{2}\|\mathbf{y} - \mathbf{h}(\mathbf{s}_K(\boldsymbol{\xi}))\|_{\mathbf{R}^{-1}} + \frac{\alpha}{2}\boldsymbol{\xi}^T\boldsymbol{\xi} \tag{5}$$

which can be calculated by solving the non-linear least squares problem using a Gauss-Newton method. Here α controls the relative importance between the data fitting and prior information. The posterior covariance matrix is bounded from below by the Fischer information matrix, as a result of the Cramer-Rao bound. Thus,

$$\mathbf{V} \geq \left(\mathbf{J}^T\mathbf{R}^{-1}\mathbf{J} + \alpha\mathbf{I}\right)^{-1} \tag{6}$$

where, $\mathbf{J} = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\xi}}$ which is evaluate at the MAP estimate.

3 A Gauss-Newton-Krylov method for computing MAP estimate

The optimization approach that we adopt is very similar to [4, 3, 5]. We adopt a slightly different notation, however to remain consistent with the rest of this paper. We assume that our forward problem, the partial differential equation, after discretization using a suitable method, such as finite differences or finite elements, can be written as follows

$$\mathbf{A}(\mathbf{s})\mathbf{u}_i = \mathbf{b}_i(\mathbf{s}) \qquad i = 1, \ldots, n_s \tag{7}$$

corresponding to n_s sources. For example, consider the system of equations arising from Hydraulic tomography [5]. The governing equations are

$$\begin{aligned} -\nabla(K(\mathbf{x})\nabla u_i) &= q_i \exp\left(-\|\mathbf{x} - \mathbf{x}_i^2\|/\gamma^2\right) \quad \mathbf{x} \in \mathcal{D} \\ \mathbf{n}.\nabla u_i &= 0 \qquad \mathbf{x} \in \partial\mathcal{D} \\ \int_{\mathcal{D}} u_i d\mathbf{x} &= 0 \end{aligned} \tag{8}$$

for $i = 1, \ldots, n_s$. Here u_i are the pressure heads, $K(\cdot)$ is the hydraulic conductivity that we want to estimate and q_i are the the flow rates corresponding to the recharge and $\mathbf{x}_i$ are the corresponding well locations. To ensure the positivity of $K(\cdot)$, for the well-posedness of the governing equations, it is more convenient to work with $s = \log K$. After appropriate discretization, this becomes the set of model parameters $\mathbf{s}$, that we want to estimate. The measurement operator that relates the model parameters to the observations, can then be written as

$$\mathbf{h}_i(\mathbf{s}) = \mathbf{H}\mathbf{u}_i = \mathbf{H}\mathbf{A}^{-1}(\mathbf{s})\mathbf{b}_i(\mathbf{s}) \qquad i = 1, \ldots, n_s \tag{9}$$

where, $\mathbf{H}$ is a sparse matrix that incorporates information regarding the receiver measurement locations. Following the approach in [4], we can write the system (5) as the

solution of a constrained optimization problem

$$\begin{aligned} \min_{\boldsymbol{\xi}} \quad & \tfrac{1}{2}\sum_{i=1}^{n_s}\|\mathbf{H}\mathbf{u}_i - \mathbf{y}_i\|^2_{\mathbf{R}^{-1}} + \tfrac{\alpha}{2}\boldsymbol{\xi}^T\boldsymbol{\xi} \\ s.t \quad & \mathbf{A}(\mathbf{s}_K(\boldsymbol{\xi}))\mathbf{u}_i = \mathbf{b}_i(\mathbf{s}_K(\boldsymbol{\xi})) \qquad i = 1,\ldots,n_s \end{aligned} \tag{10}$$

For simplicity of notation, we will derive the expressions assuming that $n_s = 1$. This can be easily extended to higher number of sources. Also, for convenience, we drop the dependence on $\mathbf{s}$ and $\boldsymbol{\xi}$. We introduce the Lagrangian $\mathcal{L}(\mathbf{u},\mathbf{s},\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a set of Lagrange multipliers,

$$\mathcal{L}(\mathbf{u},\mathbf{s},\boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{H}\mathbf{u}-\mathbf{y}\|^2_{\mathbf{R}^{-1}} + \frac{\alpha}{2}\boldsymbol{\xi}^T\boldsymbol{\xi} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{u}-\mathbf{b}) \tag{11}$$

We are in a position to write down the stationary points of the Lagrangian, which also form the first-order necessary conditions for minimizers to exist

$$\begin{aligned} \mathcal{L}_{\mathbf{u}} &= \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{u}-\mathbf{y}) + \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0} \\ \mathcal{L}_{\boldsymbol{\xi}} &= \alpha\mathbf{I}\boldsymbol{\xi} + \mathbf{G}^T\boldsymbol{\lambda} = \mathbf{0} \\ \mathcal{L}_{\boldsymbol{\lambda}} &= \mathbf{A}\mathbf{u}-\mathbf{b} = \mathbf{0} \end{aligned} \tag{12}$$

where,

$$\mathbf{G} = \frac{\partial(\mathbf{A}\mathbf{u})}{\partial\boldsymbol{\xi}} - \frac{\partial\mathbf{b}}{\partial\boldsymbol{\xi}} = \left(\frac{\partial(\mathbf{A}\mathbf{u})}{\partial\mathbf{s}_K} - \frac{\partial\mathbf{b}}{\partial\mathbf{s}_K}\right)\frac{\partial\mathbf{s}_K}{\partial\boldsymbol{\xi}} \tag{13}$$

The reduced Hessian is obtained by first linearizing the system of equations 12, setting the second order derivatives to zero and eliminating $\delta\mathbf{u}$ and $\delta\boldsymbol{\lambda}$, from the system of equations, to obtain the solution for $\delta\boldsymbol{\xi}$

$$\left(\mathbf{J}^T\mathbf{R}^{-1}\mathbf{J} + \alpha\mathbf{I}\right)\delta\boldsymbol{\xi} = -\alpha\boldsymbol{\xi} - \mathbf{J}^T(\mathbf{H}\mathbf{A}^{-1}\mathbf{b} - \mathbf{y}) \tag{14}$$

where, $\mathbf{J} = -\mathbf{H}\mathbf{A}^{-1}\mathbf{G}$ is the Jacobian matrix. We also define $\mathbf{H}_{red} \stackrel{\text{def}}{=} \mathbf{J}^T\mathbf{R}^{-1}\mathbf{J} + \alpha\mathbf{I}$ as the reduced Hessian, which is symmetric and positive definite. Constructing the Jacobian matrix could be very expensive in practice, because it requires repeated number of solutions of the forward problem. Instead, forming matrix vector products, $\mathbf{J}\mathbf{x}$ and $\mathbf{J}^T\mathbf{x}$ on appropriate sized vectors $\mathbf{x}$ can be performed in three steps - (1) Form $\mathbf{z} \leftarrow \mathbf{G}\mathbf{x}$, (2) solve the system $\mathbf{A}\mathbf{y} = \mathbf{z}$ and (3) form $\mathbf{q} \leftarrow -\mathbf{H}\mathbf{y}$. $\mathbf{J}^T\mathbf{x}$ can be computed in three steps, in a similar fashion. Thus, the computation of the entries of $\mathbf{J}$ is not necessary. Therefore, we solve the system (14) using a matrix-free Krylov subspace solver, such as Conjugate Gradient (CG) or MINRES. $\mathbf{J}^T\mathbf{R}^{-1}\mathbf{J}$ acts like a compact operator with eigenvalues clustering at zero but the spectrum of $\mathbf{H}_{red}$ is bounded from below by α. Therefore, we expect CG

to converge super-linearly for this problem, independent of the grid [3]. We emphasize here that the size of the system 14 is K, the number of terms retained in the KLE, which is much smaller than the number of grid points. Since, in exact arithmetic, the maximum number of iterations taken by is the size of the system, we expect the number of iterations to be far fewer than without the dimensionality reduction. This results in a huge savings of computational cost, because each iteration requires the solution of the forward problem $2 \times n_s$ times, which is very expensive.

In practice, a line search is necessary to ensure the global convergence of the Gauss-Newton approach. We perform a line search using a simple backtracking method, to determine a step length β that satisfies the strong Wolfe's condition.

$$\begin{aligned} f(\boldsymbol{\xi} + \beta\delta\boldsymbol{\xi}) &\leq f(\boldsymbol{\xi}) + \mu\beta\delta\boldsymbol{\xi}^T\nabla f(\boldsymbol{\xi}) \\ |\delta\boldsymbol{\xi}^T\nabla f(\boldsymbol{\xi} + \beta\delta\boldsymbol{\xi})| &\leq \eta|\delta\boldsymbol{\xi}^T\nabla f(\boldsymbol{\xi})| \end{aligned} \tag{15}$$

where, f is the objective function in equation (5), $\boldsymbol{\xi}_k$ is the current step, $\delta\boldsymbol{\xi}$ is the Gauss-Newton step that we obtain from the solution of (14) and μ, η are parameters which we choose to be $10^{-4}, 0.9$ respectively based on [11, 2].

4 Results and discussion

We first discuss some aspects regarding the implementation of the algorithm described in sections (2.1) and (3). We have implemented the computation of the KLE in C++ on top of PETSc[1]. The implementation of the Hierarchical matrices closely follows [7] and the integral equation (2) was discretized using linear Finite Elements after first performing a Galerkin projection. Subsequently, the generalized eigenvalue problem was solved using a Krylov-Schur algorithm accessed through SLEPc [6], a package also built on top of PETSc. In the case of symmetric problems, the Krylov-Schur method is equivalent to the thick-restart Lanczos algorithm. The matrix-vector products involving the covariance matrix accelerated using the Hierarchical matrix approach, as described in [7].

The covariance function is chosen to be $\kappa(\mathbf{x}, \mathbf{y}) = \exp(-r^2/\eta^2)$, where $r = \|\mathbf{x} - \mathbf{y}\|$ and $\eta = 0.5$. For the inverse problem, we focus our attention on a synthetic problem from hydraulic tomography. We are interested in imaging the hydraulic conductivity K in the domain $[-1, 1]^2$ from measurements of the change in head at various locations, due to pumping tests. The governing equations are described in (7). The locations of the pumping tests and the sensors are given in the figure 4. We pick $\gamma = 0.05$ and $q_i = 10^2$. The PDE's are discretized using a cell-centered Finite Volume method and are solved using a block-CG iterative method [12], which is sometimes advantageous over the repeated use of the Conjugate gradient method for solving a problem with multiple right hand sides. Theoretically, the number of iterations are $\lceil n/s \rceil$, where n is the system size and s is the

number of right hand side, which is smaller than the n iterations that conjugate gradient requires per right hand side, although the cost per iteration of the block-version is higher.

The true model that we assume is illustrated in figure 4. The synthetic measurements are generated by solving the forward problem to a much higher accuracy than what is used in the inversion, with the true model as the log-conductivity field and 0.1% Gaussian random noise was added to the measurements. For the inversion, we assume that $K = 100$ and $\alpha = 5 \times 10^{-3}$. We present very preliminary results. The results of the inversion are described in figure 4.

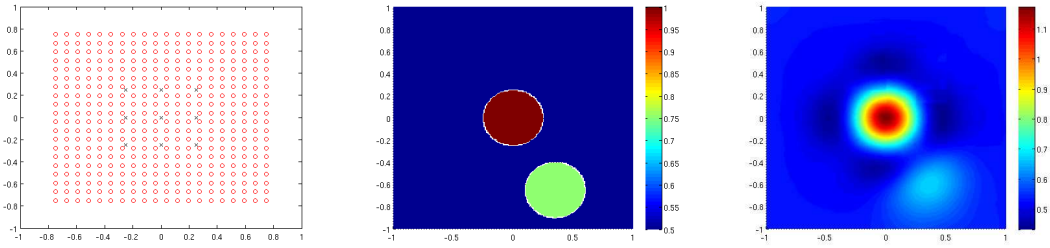

Figure 1: (left) Locations of 9 sources(crosses) and 400 measurement sensors(circles). (center) the true field. (right) reconstructed field on a 200×200 grid. 100 terms were used in the expansion. The reconstruction took 3 Gauss-Newton iterations. The relative L^2 error was about 10%.

Future work includes extension to irregular grids and 3D, uncertainty quantification via conditional realizations and identifying structural parameters. While the algorithm itself is not restricted to regular grids, our implementation is.

5 Acknowledgements

The authors were supported by NSF Award 0934596, Subsurface Imaging and Uncertainty Quantification. We would also like to thank E. Haber, for making his code [5] available on his website, which gave us a starting point for our Gauss-Newton implementation.

REFERENCES

[1] Satish Balay, Kris Buschelman, Victor Eijkhout, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Barry F. Smith, and Hong Zhang. PETSc users manual. Technical Report ANL-95/11 - Revision 3.0.0, Argonne National Laboratory, 2008.

[2] P.E. Gill, W. Murray, and M.H. Wright. Practical optimization. 1981.

[3] E. Haber and U.M. Ascher. Preconditioned all-at-once methods for large, sparse parameter estimation problems. *Inverse Problems*, 17:1847, 2001.

[4] E. Haber, U.M. Ascher, and D. Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse problems*, 16:1263, 2000.

[5] E. Haber and L. Hanson. Model problems in pde-constrained optimization. Technical report, Citeseer, 2007.

[6] V. Hernandez, J.E. Roman, and V. Vidal. Slepc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Transactions on Mathematical Software (TOMS)*, 31(3):351–362, 2005.

[7] BN Khoromskij, A. Litvinenko, and HG Matthies. Application of hierarchical matrices for computing the karhunen–loève expansion. *Computing*, 84(1):49–67, 2009.

[8] P.K. Kitanidis. Quasi-linear geostatistical theory for inversing. *Water Resources Research*, 31(10):2411–2419, 1995.

[9] W. Li and O.A. Cirpka. Efficient geostatistical inverse methods for structured and unstructured grids. *Water resources research*, 42(6):W06402, 2006.

[10] Y.M. Marzouk and H.N. Najm. Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902, 2009.

[11] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer verlag, 1999.

[12] D.P. O'Leary. The block conjugate gradient algorithm and related methods. *Linear Algebra and its Applications*, 29:293–322, 1980.

[13] C. Schwab and R.A. Todor. Karhunen-loeve approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, 2006.