LA-UR- 09-003208

Approved for public release; distribution is unlimited.

Title: A REGULARIZATION APPROACH TO HYDROFACIES DELINEATION

Author(s): DANIEL TARTAKOVSKY
BRENDT WOHLBERG

Intended for: PROCEEDINGS:
SECOND INTERNATIONAL SYMPOSIUM ON COMPUTATIONAL MECHANICS

Los Alamos
NATIONAL LABORATORY
EST. 1943

A Regularization Approach to Hydrofacies Delineation

Daniel M. Tartakovsky[a] and Brendt E. Wohlberg[b, *]

[a]*Department of Mechanical and Aerospace Engineering, University of California, San Diego, USA*
[b]*Theoretical Division, Los Alamos National Laboratory, USA (*brendt@tmail.lanl.gov)*

Abstract. We consider an inverse problem of identifying complex internal structures of composite (geological) materials from sparse measurements of system parameters and system states. Two conceptual frameworks for identifying internal boundaries between constitutive materials in a composite are considered. A sequential approach relies on support vector machines, nearest neighbor classifiers, or geostatistics to reconstruct boundaries from measurements of system parameters, and then uses system states data to refine the reconstruction. A joint approach inverts the two data sets simultaneously by employing a regularization approach.

Keywords: Subsurface, Inverse modeling, Statistical learning theory, Total Variation regularization.
PACS: 02.30.Zz, 92.40.Kf, 92.40.-t

INTRODUCTION

Many complex phenomena in science and engineering take place in composite media, whose internal structures are uncertain. For example, heterogeneous groundwater aquifers typically consist of multiple lithofacies, the spatial arrangement of which significantly affects flow and transport in the subsurface. The estimation of these lithofacies is complicated by the sparsity of data and by the lack of a clear correlation between identifiable geologic indicators and attributes (e.g. hydraulic conductivity and porosity). This so-called zonation problem has been studied in [1-4], among others.

Data which are used in geomaterials classification procedures are typically obtained from core samples that often disturb soils and are by necessity sparse, thus contributing to predictive uncertainty associated with the location of different geomaterials. Within a stochastic framework, this uncertainty is quantified by treating a formation's properties as random fields that are characterized by multivariate probability density functions or, equivalently, by their joint ensemble moments. Geostatistics has become an invaluable tool for estimating facies distributions at points in a computational domain where data are not available, as well as for quantifying the corresponding uncertainty [5].

Recently we demonstrated that both Support Vector Machine (SVM) [6,7] and nearest-neighbor classification [8] techniques provide a viable alternative to geostatistical frameworks by allowing one to delineate lithofacies in the absence of sufficient data parameterization, without treating geologic parameters as random and, hence, without the need for the ergodicity assumption. These approaches have been used to reconstruct interfaces from both well [7,8] and poorly [9] differentiated parameter data.

In this study, we present two alternative approaches to hydrofacies delineation from both system parameter (e.g., porosity, hydraulic or electrical conductivity, etc.) and system state (e.g., hydraulic head, solute concentration, temperature, etc.) data. The first approach relies on a sequential use of data, wherein system state data are used to refine an SVM-based reconstruction [6,7] of interfaces from parameter data. The second approach relies on a functional minimization approach to reconstruct interfaces from both system state and system parameter data *simultaneously*. We demonstrate these ideas on a synthetic case of steady-state flow through a composite domain Ω consisting of two materials separated by highly irregular boundaries (see Fig. 1). For simplicity, the hydraulic conductivity of each material is assumed to be constant.

The facies delineation problem can be formulated as follows. Given system-parameter data $K_i \equiv K(\mathbf{x}_i)$, $i = 1,\dots,N_k$, system-state data $h_j \equiv h(\mathbf{x}_j)$, $j = 1,\dots,N_h$, and a model connecting the two—e.g., a steady-state flow equation $\nabla\cdot(K\nabla h)=0$ for $\forall \mathbf{x}\in\Omega$ with appropriate boundary conditions—find the boundaries between subdomains Ω_1 and Ω_2 ($\Omega=\Omega_1\cup\Omega_2$) consisting of materials M_1 and M_2, respectively. To simplify the presentation, we assume that both data sets are collected at the same N locations $\mathbf{x}_i$, $i=1,\ldots,N$ where $N=N_k=N_h$.

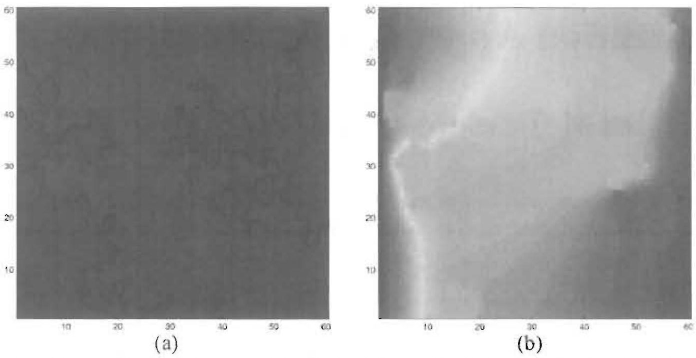

FIGURE 1. Flow domain consisting of two contrasting lithofacies (a). A highly conducting material is shown in red and a low conducting material in blue. The corresponding hydraulic head distribution (b).

SEQUENTIAL INVERSION WITH SVM

Inversion of System Parameter Data

Since parameter data $K_i\equiv K(\mathbf{x}_i)$, $i=1,\ldots,N$ uniquely characterize constitutive materials M_1 and M_2, the points where they are taken can be labeled by means of the indicator function

$$I_i\equiv I(\mathbf{x}_i)=\begin{cases}+1 & \mathbf{x}_i\in\Omega_1\\ -1 & \mathbf{x}_i\in\Omega_2\end{cases}. \tag{1}$$

This step typically involves an analysis of a data histogram, which is often nontrivial when constitutive materials (e.g., geological lithofacies) are heterogeneous. Here we assume that the available parameter data $K_i\equiv K(\mathbf{x}_i)$, $i=1,\ldots,N$ are well differentiated, so that the process of assigning the values of the indicator functions to points $\mathbf{x}_i$, $i=1,\ldots,N$ does not introduce interpretive errors. This assumption can be relaxed to account for poor differentiation of data [9].

The SVM approach to delineation of boundaries between Ω_1 and Ω_2 is to minimize the quadratic functional

$$\max_{\gamma}\left[\sum_{i=1}^{N}\gamma_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_i\gamma_j I_i I_j R(\mathbf{x}_i,\mathbf{x}_j)\right], \tag{2}$$

where $R(\mathbf{x}_i,\mathbf{x}_j)$ is a given Mercer kernel, subject to the constraints

$$0\le\gamma_i\le C,\quad \sum_{i=1}^{N}\gamma_i I_i=0. \tag{3}$$

This optimization problem has a well-defined global minimum that is influenced by the choice of the fitting parameter C. Let γ_i^* ($i=1,\ldots,N$) denote a solution of the optimization problem (2)-(3). Then the indicator function $I(\mathbf{x})$ at any point $\mathbf{x}$, and hence the boundary separating the two materials, is given by [6,7]

$$I(\mathbf{x})=\text{sign}\left[\sum_{i=1}^{N}\gamma_i^* I_i R(\mathbf{x}_i,\mathbf{x})+b^*\right],\qquad b^*=I_r-\sum_{i=1}^{N}\gamma_i^* I_i R(\mathbf{x}_i,\mathbf{x}_r). \tag{4}$$

for some j such that $\gamma_j > 0$.

Inversion of System State Data

Incorporation of system state data $h_i \equiv h(\mathbf{x}_i)$, $i = 1,...,N$ into the SVM framework is challenging, since one cannot assign the indicator function to such data, and the relationship between the two data types is nonlinear. Following [10], we use the parameter field reconstructed with the SVM procedure (2)-(4) as an initial guess for the optimization problem

$$\min_{\gamma}\left[-\sum_{i=1}^{N}\gamma_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_i\gamma_j I_i I_j R(\mathbf{x}_i,\mathbf{x}_j) + \lambda\sqrt{\frac{1}{N}\sum_{i=1}^{N}[h_i - h^*(\mathbf{x}_i)]}\right] \quad (5)$$

subject to the constraints (3) and fixed $\lambda > 0$. The system state $h^*(\mathbf{x})$ is a solution of $\nabla\cdot(K\nabla h) = 0$ in which the hydraulic conductivity $K(\mathbf{x})$ is determined by the current state of γ_i, $i = 1,...,N$. The SVM inversion aims to retain the maximization of the SVM margin based on conductivity data (Step 1), while minimizing the difference between the measured and computed heads (Step 2). This balance is controlled by the choice of the parameter λ in (5). The higher its value, the more weight one assigns to the head measurements relative to the conductivity measurements, and vice versa.

JOINT INVERSION VIA FUNCTIONAL MINIMIZATION

In the SVM approach, hydraulic head data affect only the radial functions weights, which, in principle, might provide too few degrees of freedom. Indeed, one can expect the estimated subdomain boundary to be overly smooth, when a conductivity data set is small. As an alternative, we introduce a hydrofacies delineation approach that is based on minimization of the following functional

$$T(\mathbf{k},\mathbf{h}) = \frac{1}{2}\|F(\mathbf{k}) - \mathbf{h}\|_2^2 + \frac{\lambda_1}{2}\|M_k\mathbf{k} - \hat{\mathbf{k}}\|_2^2 + \frac{\lambda_2}{2}\|M_h\mathbf{h} - \hat{\mathbf{h}}\|_2^2 + \lambda_3 R_k(\mathbf{k}) + \lambda_4 R_h(\mathbf{h}). \quad (6)$$

Here $\mathbf{k}$ and $\mathbf{h}$ are the M-dimensional arrays representing $K(\mathbf{x})$ and $h(\mathbf{x})$ on the flow domain Ω discretized into M elements; $\hat{\mathbf{k}}$ and $\hat{\mathbf{h}}$ are the N-dimensional arrays containing the measurements $K_i \equiv K(\mathbf{x}_i)$ and $h_i \equiv h(\mathbf{x}_i)$ ($i = 1,...,N$), respectively; operators M_k and M_h extract the solution values at measurement points $\mathbf{x}_i$ ($i = 1,...,N$); $F(\mathbf{k})$ is a (finite-elements) representation of the flow equation $\nabla\cdot(K\nabla h) = 0$ on the flow domain Ω discretized into M elements; and regularization functions R_k and R_h represent prior knowledge of the expected form of $\mathbf{k}$ and $\mathbf{h}$ solutions, respectively. Since the head field $h(\mathbf{x})$ is smooth, and the conductivity field $K(\mathbf{x})$ may have discontinuities, we choose

$$R_h(\mathbf{h}) = \frac{1}{2}\|\nabla\mathbf{h}\|_2^2, \qquad R_k(\mathbf{k}) = \|\nabla\mathbf{k}\|_1, \quad (7)$$

the latter being the Total Variation norm [11].

Minimization of the nonlinear functional (6)-(7) is clearly more computationally expensive than solving the optimization problem (5). Yet its flexibility is assigning relative degrees of importance to the two data types (system parameter data and system state data) contains the promise of a better performance. Below we conduct numerical experiments on the fields shown in Fig. 1 to demonstrate the performance of the SVM-based approach. The performance of our joint inversion algorithm will be reported at the time of the ISCM- EPMESC meeting.

COMPUTATIONAL EXAMPLE

We employ the SVM-based algorithm to reconstruct boundaries between the two materials shown in Fig. 1(a) from N randomly selected data points $\mathbf{x}_i$ ($i = 1,...,N$). At these data points, both the system parameter K and the system state h are sampled. The values of the system state in Fig. 1(b) are obtained by solving the flow equation

$\nabla \cdot (K\nabla h) = 0$. This equation is subject to the Dirichlet boundary conditions $h = H_1$ and $h = H_2$ prescribed along the left and right vertical boundaries, respectively. The lower and upper horizontal boundaries are impermeable.

The first step in the proposed algorithm consists of the use of an SVM to reconstruct the boundaries from $N = 100$ conductivity measurements. The location of these measurements, the reconstructed boundaries, and the corresponding hydraulic head distribution are shown in Fig. 2(a). We used the Gaussian kernel

$$R(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}\right), \tag{8}$$

with $\sigma = 5$, and set the SVM parameter $C = 1000$. These fixed values were chosen for good hydrofacies delineation performance based on our previous experience [7], but in actual applications these would be chosen via a cross-validation method.

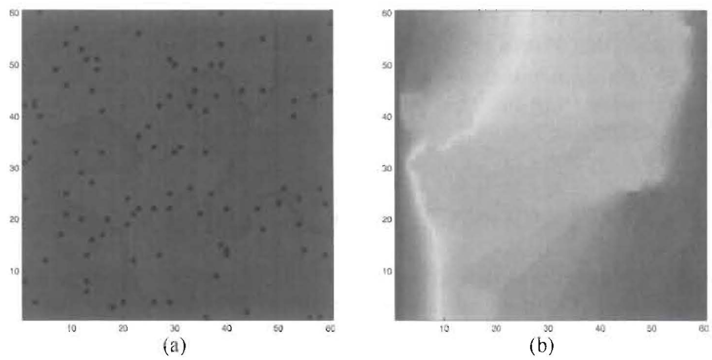

FIGURE 2. The system parameter $K(\mathbf{x})$ reconstructed from $N = 100$ measurements, whose locations are shown by the black dots (a) and the corresponding system state $h(\mathbf{x})$ (b).

Using the parameter field $K(\mathbf{x})$ in Fig. 2(a) as an initial guess in the second step, and choosing $\lambda = 1000$ to provide appropriate weighting to each term, we obtain the reconstructed boundaries and the corresponding hydraulic head distribution in Fig. 3. A comparison of the reconstructed parameter fields $K(\mathbf{x})$ in Fig. 2(a) and Fig. 3(a) with the reference parameter field in Fig. 1(a) reveals that the use of both system-parameter and system-state data noticeably improves the boundary reconstruction.

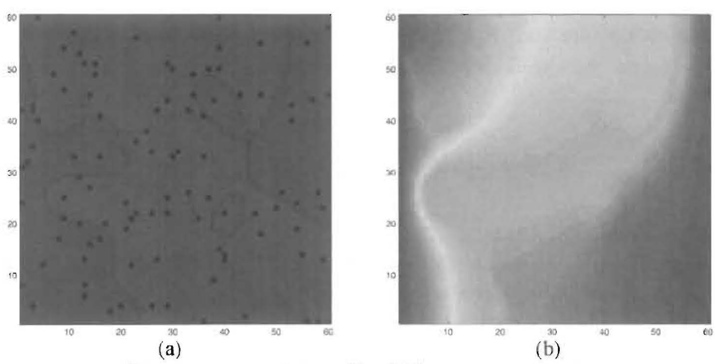

FIGURE 3. The system parameter $K(\mathbf{x})$ reconstructed from $N = 100$ measurements of both the system parameter and the system state, whose locations are shown by the black dots (a) and the corresponding system state $h(\mathbf{x})$ (b).

CONCLUSIONS

We presented two approaches for identification of internal composition of composite materials from sparse measurements of system parameters (material properties) and system states (physical quantities). The first, the

Support Vector Machines (SVM), assimilates data sequentially, starting with parameter data. The second, total variation inversion, assimilates both data types simultaneously. The SVM-based approach has been used to identify the internal structure of a synthetic porous medium; the second approach is currently under development. Our preliminary results, including those presented here, demonstrate the potential of the proposed approaches.

ACKNOWLEDGMENTS

This study was supported by the Office of Science of the U.S. Department of Energy, Advanced Scientific Computing Research (ASCR) program in Applied Mathematical Sciences.

REFERENCES

1. N.-Z. Sun and W. W.-G. Yeh, *Water Resour. Res.* **21**, 869-883 (1985).
2. J. Carrera and S. P. Neuman, *Water Resour. Res.* **22**, 228-242 (1986).
3. M. J. Eppstein and D. E. Dougherty, *Water Resour. Res.* **32**, 3321-3336 (1996).
4. F. T.-C. Tsai and W. W.-G. Yeh, *Water Resour. Res.* **40**, doi:10.1029/2003WR002893 (2004).
5. L. Guadagnini, A. Guadagnini and D. M. Tartakovsky, *J. Hydrol.* **294**, 57-67 (2004).
6. D. M. Tartakovsky and B. E. Wohlberg, *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020864 (2004).
7. B. E. Wohlberg, D. M. Tartakovsky and A. Guadagnini, *IEEE Trans. Geosci. Remote Sens.* **44**, 47-57 (2006).
8. D. M. Tartakovsky, B. E. Wohlberg and A. Guadagnini, *Water Resour. Res.* **34**, doi:10.1029/2007GL029245 (2007).
9. B. E. Wohlberg and D. M. Tartakovsky, *Adv. Water Resour.* **32**, 225-230 (2009).
10. D. M. Tartakovsky, A. Guadagnini and B. E. Wohlberg, "Machine learning methods for inverse modeling" in *GeoENV VI – Geostatistics for Environmental Applications*, edited by A. Soares et al., Springer Science+Business Media B.V., 2008, pp. 117-125.
11. L. Rudin, S. J. Osher and E. Fatemi, *Physica D*, **60**, 259-268 (1992).