Sparse Modeling in Geotechnical Engineering

Takayuki Shuku[1]

[1]Graduate School of Environmental and Life Science, Okayama University, 3-1-1 Tsushima naka, Kita-ku, Okayama 700-8530, Japan. E-mail: shuku@cc.okayama-u.ac.jp

Abstract: Estimation and prediction problems in geotechnical engineering belong to the class of inverse problems. Many different approaches to analyze these problems have been reported in the literature. In recent years, the use of machine learning has become increasingly common in many research fields because of the rapid increase of computational capacity and advances in algorithms. In particular, the methodology for solving inverse problems known as "sparse modeling" has been receiving considerable attention. Sparse modeling is a statistical method which exploits specific features/structures in data based on solution sparsity, and has a great potential for application to geotechnical problems. This paper demonstrates the potential of sparse modeling for solving geotechnical engineering problems by means of two practical examples: a cross-borehole tomography, and a stratigraphic soil profiling based on cone penetration test.

Keywords: Sparse modeling; lasso; inverse analysis; tomography; stratigraphic soil profiling.

1 Introduction

Inverse analysis is a methodology which permits to infer unknown properties of a system using a model of the system and (usually noisy) observation data. The unknown system properties in inverse analysis often include unknown governing equations, boundary and initial conditions, and material properties. Some of the applications of inverse analysis include X-ray tomography in material/medical science, image analysis/processing, and geophysical methods such as electromagnetic monitoring, seismic tomography, and ground penetrating radar. In geotechnical engineering, inverse analysis has been applied to geotechnical problems for the past four decades. For instance, applications have been reported for tunnel excavation in rock, consolidation, pile settlement, and retaining walls (e.g., Sakurai and Takeuchi 1983; Asaoka 1978; Honjo et al. 1993; Ledesma et al. 1996; Finno and Carvello 2005).

Recently, inverse analysis has been receiving attention due to the rapid increase and availability of computational power on the one hand, and advances in machine learning algorithms on the other hand. Although machine learning tasks are classified into several categories, the main common goal is to infer an estimator using a finite set of data or samples. This goal is same as that of inverse analysis. Moreover, the methodologies utilized in both inverse analysis and machine learning rely on common theories (DeVito et al. 2005). The machine learning method known as "sparse modeling", in particular, has received much attention for its ability of managing several types of inverse problems. According to the general principle of sparsity, a phenomenon should be represented with as few variables as possible. This approach, which essentially favors simple theories over more complex ones, is central to many research fields. One of the most widely adopted methodologies for achieving sparse modeling is the least absolute shrinkage and selection operator (lasso) proposed by Tibshirani (1996). Some researchers have demonstrated the effectiveness of sparse modeling in solving inverse problems (Lustig et al. 2007; Honma et al. 2014).

Sparse modeling can be applied not only for solving ill-posed inverse problems, but also for exploiting internal structures in the data, and automatic selecting simpler but accurate statistical models. Clearly, it has a great potential for application to geotechnical problems, and existing technical issues in geotechnical engineering can be overcome through its use.

This paper demonstrates the potential of sparse modeling for solving geotechnical engineering problems by showing two practical examples: a cross-borehole tomography (CBT), and stratigraphic a soil profiling based on cone penetration test (CPT). The paper is structured as follows: in Section 2, the theoretical fundamentals of inverse analysis, classification and solution of inverse problems are presented; in Section 3, the concept of sparse modeling and its mathematical fundamentals are outlined; in Section 4, numerical algorithms to solve sparse modeling problems are summarized. Then, the two practical examples of sparse modeling and their results are shown in Section 5 and, finally a summary of the results of sparse modeling in geotechnical engineering are presented in Section 6.

2 Introduction to Inverse Analysis

Let us consider the following linear system model and the corresponding inverse problem:

Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR)
Editors: Jianye Ching, Dian-Qing Li and Jie Zhang

Published by Research Publishing, Singapore.
ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0_bs3-cd

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{1}$$

where, $\mathbf{y}$ is an m-dimensional observation vector, $\mathbf{x}$ is an n-dimensional (unknown) parameter vector, and $\mathbf{A}$ is an $m \times n$ matrix representing a linear operator. We want to estimate the unknown vector $\mathbf{x}$ using the observation vector $\mathbf{y}$. This is a typical example of inverse analysis. If the $\mathbf{A}$ matrix has full rank, the inverse problem can be classified into three categories depending on the values of m and n:

$m > n$: Over-determined problem
$m = n$: Even-determined problem
$m < n$: Under-determined problem

If solutions exist, even-determined problems have unique solutions, and the error vector $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{x}$ is a zero vector. With more observation data than unknown parameters, there is no solution that can fit exactly with the observation data. However, least square solutions can be defined by minimizing the quantity:

$$\min_{\mathbf{x}} \ \frac{1}{2}\left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2^2 \tag{2}$$

These two problems are hardly encountered in inverse problems in geotechnics, as observation data are usually much less than the unknown parameters. Most of the inverse problems in practice might be under-determined (ill-posed) problems, therefore the study on the methods for solving under-determined problems is the central topic of inverse analysis. One approach to solve under-determined problems is to use some kind of regularization. The most commonly used method consists of minimizing the quantity:

$$\min \left\|\mathbf{x}\right\|_2 \quad \text{s.t.} \quad \left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2 \le t \tag{3}$$

where $\|\cdot\|_2$ identifies the ℓ_2 norm, and t is a user-specified tuning parameter. The above optimization problem can also be written in the following unconstrained form:

$$\min_{\mathbf{x}} \ \frac{1}{2}\left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2 + \lambda\left\|\mathbf{x}\right\|_2 \tag{4}$$

where λ is the regularization parameter, which controls the intensity of the regularization term $\|\mathbf{x}\|_2$ and the least square term $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$. There are several advantages of using this objective function: the function is strictly convex, therefore it always has a unique solution; moreover, the solution to the problem is available in close form, which is defined by:

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y} \tag{5}$$

This method is termed "ridge regression" (Hoerl and Kennard, 1970), or "weight decay" in the context of neural network (e.g., Bishop 2006). In the context of Bayesian framework, the regularization term can be interpreted as "prior information" of the solution, which corresponds to saying that the "ℓ_2 norm of the solution vector should be small."

3 Sparse Modeling

3.1 Solution sparsity

Solution sparsity is a property in which the solution vector $\boldsymbol{x}$ has $x_j = 0$ in many components $j \in \{1, \ldots, n\}$. In other words, it is assumed that only a relatively small subset of $\boldsymbol{x}$ is truly important in a specific context: e.g., usually only a small number of simultaneous faults occurs in a system; a small number of nonzero Fourier coefficients is sufficient for an accurate representation of various signal types; and a small number of predictive variables is most relevant to the response variable, and is sufficient for learning an accurate predictive model. In all these examples, the solution we seek can be viewed as a sparse high-dimensional vector with only a few nonzero coordinates.

3.2 Best subset selection and its approximation

Solution sparsity is a useful prior information to solve under-determined problems. The role of this property in solving an underdetermined problem is shown in Figure 1. In the figure, colored cells indicate non-zero components (also called active-set), and white cells indicate zero components. When the solution $\mathbf{x}$ is sparse, and if we know how many zeros are and which components are nonzero in $\mathbf{x}$, the under-determined problem can be solved by minimizing the following

$$\min \left\|\mathbf{x}\right\|_0 \quad \text{s.t.} \quad \left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2 \le t \tag{6}$$

where $\|\cdot\|_0$ is an ℓ_0 norm (this is not a proper "norm", though), and indicates the number of non-zero components in **x**. This procedure is called the "best subset selection". The corresponding Lagrangian form of Eq. (6) is given by:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 + \lambda\|\mathbf{x}\|_0 \tag{7}$$

where λ is the regularization parameter, and its role is the same as that in Eq. (4). By minimizing Eq. (7), the sparse solution can be obtained. In general, however, finding a minimum-cardinality solution satisfying linear constraints is an NP-hard combinatorial problem (Natarajan 1995). Thus, an approximation is necessary to achieve computational efficiency, and it turns out that, under certain conditions, approximate approaches can recover the exact solution.

An alternative approach of best subset selection is provided by the following equation:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 + \lambda\|\mathbf{x}\|_1 \tag{8}$$

where $\|\cdot\|_1$ is an ℓ_1 norm, and stands for sum of the absolute values of **x**. A famous schematic of comparison between Eqs. (4) and (7) is shown in Figure 2. The constraint region for ridge regression is the disk $x_1^2 + x_2^2$, while that for lasso is the diamond $|x_1| + |x_2|$. Both methods find the first point where the red line hits the constraint region. Unlike the disk, the diamond has corners; if the solution occurs at a corner, then it has one parameter x_1 equal to zero. When $n > 2$, the diamond becomes a rhomboid, and has many corners, flat edges, and faces; there are many more opportunities for the estimated parameters to be zero. This idea can be applied in many different statistical models. In statistical literature, the problem of Eq. (8) is widely known as the least absolute shrinkage selection operator (lasso, Tibshirani 1996).

There have been many works on lasso since it was first proposed in 1996, and it has become clear that the ℓ_1 penalty has the following advantages.

1. The ℓ_1 penalty provides a natural way to encourage sparsity and simplicity in the solution. The lasso enables simultaneous model selection and parameter estimations and gives interpretable models.
2. The ℓ_1-based penalties are convex. This fact and the assumed sparsity can lead to significant computational advantages. For example, if we have to estimate one million non-zero parameters with 100 observation data, the computation is very challenging. However, if we apply the lasso, then at most 100 parameters can be nonzero in the solution, and this makes the computation much easier.

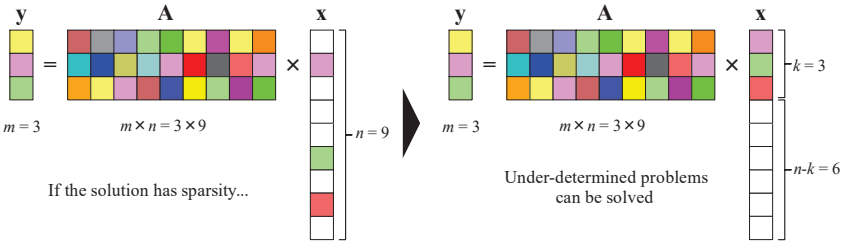


Figure 1. The role of solution sparsity in solving an under-determined problem.

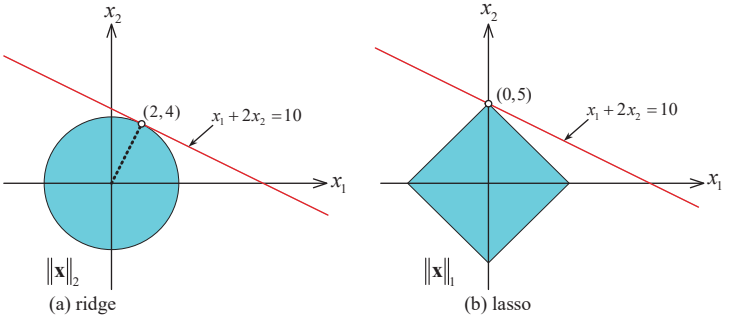


Figure 2. Schematic comparison between ridge and lasso

3.3 Structured sparsity

The basic lasso does not perform well when the solution is not sparse. In other words, the basic lasso has certain limitations in exploiting inherent structures that arise from underlying index sets, such as time and space, in the data. The unknown target parameters might each have an associated time stamp, and we might then ask for time-neighboring coefficients to be the same or similar. The sparse modeling performs well even in the problems by enforcing smoothness of neighboring unknown parameters. The approach is called "fused lasso" (Tibshirani et al. 2005), and can exploit such structure within a data. The fused lasso is the solution of the following optimization problem:

$$\min_{\mathbf{x}} \ \frac{1}{2}\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|_2+\lambda\left\|\mathbf{B}\mathbf{x}\right\|_1 \tag{9}$$

where $\boldsymbol{B}$ is a $(n-1)\times n$ matrix, and a commonly used form of $\mathbf{B}$ is:

$$\mathbf{B}=\mathbf{D}^1=\begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \tag{10}$$

This regularization term enforces the sparsity in the first-order differences of neighboring solutions, and is called "total variation". The total variation in Eq. (9) can be generalized to use a higher-order difference leading to the problem:

$$\min_{\mathbf{x}} \ \frac{1}{2}\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|_2+\lambda\left\|\mathbf{D}^{(k+1)}\mathbf{x}\right\|_1 \tag{11}$$

where $\mathbf{D}^{(k+1)}$ is a matrix of dimension $(n-k-1)\times n$ that computes discrete differences of order $k+1$. This method deals with different kinds of structures in the data in natural ways.

3.4 Bayesian view of Lasso estimates

In a Bayesian statistical framework, the lasso estimates can be derived as the Bayes posterior mode under Laplacian prior for the x_j, as:

$$p(x_j)=\frac{\lambda}{2\tau}\exp\left(-\frac{|x_j|}{\tau}\right) \tag{12}$$

where $\tau = 1/\lambda$. It is favorable to perform a Bayesian analysis for assessing the detailed uncertainty in the lasso solution. In this regard, Park and Casella (2008) proposed the "Bayesian lasso", which computes the posterior mean and median estimates from a Gaussian regression model with Laplacian prior, but the estimates are not sparse. If one wants to obtain sparse solutions from standard Bayesian analysis, prior has to be defined so that some mass is at zero, such as using the spike-and-slab model (George and McCulloch 1993). However, this method leads to non-convex problems that are computationally intensive, and does not have the advantages the basic lasso has.

4 Numerical Algorithm for Lasso Problems

Solving Eqs. (8) and (9) is a convex minimization problem. A standard approach to this minimization problem is to take the gradient with respect to $\mathbf{x}$ and set it to zero. However, one of the central difficulties in solving the problems is the presence of a non-smooth ℓ_1 norm. In other words, the absolute value function $|x_j|$ does not have a derivative at $x_j = 0$. Nevertheless, this problem can be solved by applying a soft-thresholding operator (Donoho 1995) to x_j, which is defined as:

$$S_\lambda=\begin{cases} x_j-\lambda & x_j>\lambda \\ 0 & -\lambda\le x_j\le\lambda \\ x_j+\lambda & x_j<\lambda \end{cases} \tag{13}$$

where S_λ is a soft-thresholding function. This operator translates x toward zero by an amount λ, and sets it to zero if $|x| < \lambda$. When $\lambda = 0$, the solution of Eq. (8) becomes the solution for the ordinary least squares problem.

The general approach for solving the lasso problem can be summarized as: 1) minimize the first term in the objective function; 2) apply the soft-thresholding operator to x; and 3) repeat steps 1 and 2. Of the many

reconstruction algorithms proposed for solving convex problems, the alternating direction method of multipliers (ADMM, Boyd et al. 2010), which blends the decomposability of the dual-ascent method with the superior convergence properties of the method of multipliers, is used in this paper to solve the lasso problem. This algorithm solves problems in the form:

$$\min_{\mathbf{x}} \ \frac{1}{2}\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|_2+\lambda\left\|\mathbf{z}\right\|_2 \quad \text{s.t.} \quad \mathbf{B}\mathbf{x}=\mathbf{z} \tag{14}$$

and the ADMM updates take the form of:

$$\mathbf{x}^{k+1}=\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}+\rho\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\left\{\frac{1}{\lambda}\mathbf{A}^{\mathrm{T}}\mathbf{z}+\rho\mathbf{B}^{\mathrm{T}}(\mathbf{z}-\mathbf{u})\right\} \tag{15}$$

$$\mathbf{z}^{k+1}=S_{1/\rho}(\mathbf{B}\mathbf{x}^{k+1}+\mathbf{u}) \tag{16}$$

$$\mathbf{u}^{k+1}=\mathbf{u}^{k}+(\mathbf{B}\mathbf{x}^{k+1}-\mathbf{z}^{k+1}) \tag{17}$$

where the ρ is the penalty parameter, the $\mathbf{u}$ vectors are Lagrange multipliers associated with the constraint, and $S_{1/\rho}$ is the soft-thresholding operator defined in Eq. (13). The algorithm involves a ridge regression update for $\mathbf{x}$, a soft-thresholding step for $\mathbf{z}$, and a simple linear update for $\mathbf{u}$.

5 Application Examples

5.1 CBT

The CBT is a method for inferring properties of the ground between two parallel boreholes (Figure 3). It has been used in geological and geotechnical practice since the early 1970s. A common problem in geophysical methods is that only a limited number of transmitters and receivers can be used. In addition, in contrast to X-ray computed tomography (CT) scanning, in which transmitters and receivers can be placed arbitrarily around the targets, the ray paths in CBT are restricted. These limitations affect the quality of image reconstruction, which tends to be an ill-posed problem from the perspective of inverse analysis. This subsection demonstrates the performance of the sparse modeling for CBT through the corresponding numerical tests.

5.1.1 Mathematical fundamentals of CBT

Image reconstruction techniques for CBT can be classified into two broad categories: transform methods and series expansion methods. We focused on the latter methods, as they are the most widely used in geotechnical applications. Figure 4 shows the schematic view of the image reconstruction procedure based on the series expansion method. In the figure, the dashed line indicates the i-th ray path (i = 1, 2,…, m), and the reconstruction area is divided into n cells. Each cell has a material property, i.e. the inverse of wave propagation velocity, defined as x_j. The a_{ij} indicates the length of the i-th ray path crossing the j-th cell. Assuming that the ray paths are straight, the wave observation vector can be discretely approximated by (Honjo and Kashiwagi 1991):

$$\mathbf{y}=\mathbf{A}\mathbf{x}+\boldsymbol{\varepsilon} \tag{18}$$

where $\mathbf{y}$ is an m-dimensional observation vector, $\mathbf{x}$ is an n-dimensional parameter vector, $\boldsymbol{\varepsilon}$ is an m-dimensional noise vector, and $\mathbf{A}$ is an m-by-n observation matrix defined as:

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{19}$$

In this problem, the error $\boldsymbol{\varepsilon}$ is assumed to be an independently and identically distributed random variable that follows a normal distribution with a mean of 0 and a standard deviation of σ_ε.

We applied the sparse modeling to two numerical examples of CBT to investigate the reconstruction accuracy. Figure 5 shows true images of the ground, in which (a) is the ground with some cavities (or potential piping holes), and (b) is a ground with layered structure. The inverse problem to solve here is to estimate the material properties of the cells ($\mathbf{x}$) using the observations ($\mathbf{y}$).

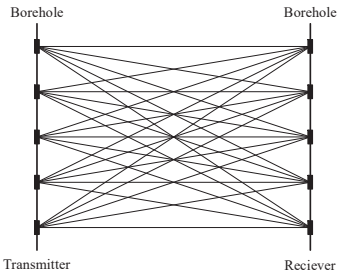


Figure 3. Cross-borehole tomography.

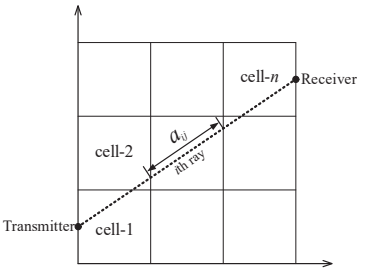
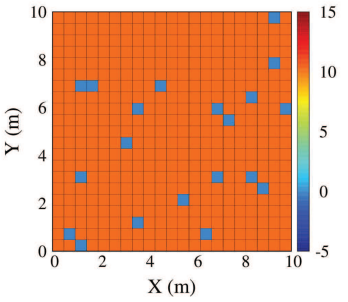
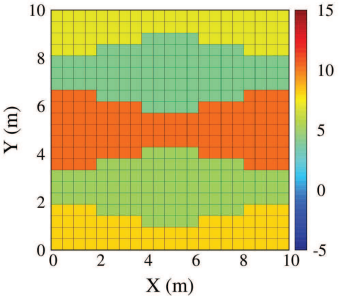


Figure 4. Image reconstruction procedure.



(a) Problem 1: Ground with cavities



(b) Problem 2: Layered ground

Figure 5. True images for the numerical tests.

5.1.2 Numerical setup

The total number of cells (n) is 441. Each cell has a value assigned, which can be considered a physical property of the ground, such as stiffness or the inverse of wave propagation velocity. Considering 21 receivers and transmitters, there are 21×21=441 ray paths in total. Although the dimension of the parameter vector n and that of observation vector m are identical in this problem, the observation matrix $\mathbf{A}$ is not a full-rank matrix because of rank deficiency. Hence, the reconstruction problems here are under-determined problems. The synthetic observation data ($\mathbf{y}$) used for the inverse analysis is generated as follows:

1. Assuming that all ray paths are straight, e the observation matrix $\mathbf{A}$ is computed (Eq. 1).
2. $\mathbf{Ax}$ is computed; the parameter vector $\mathbf{x}$ is the vector of true values given in Figure 5.
3. Gaussian observation noise with mean 0 and standard deviation σ_ε is added to the computed $\mathbf{Ax}$.

The basic sparse modeling, Eq. (8), cannot function well in the two problems because neither of them have sparse solutions. The reconstruction of the layered structure (Figure 5b) can be solved by structured lasso considering total variation (Eq. 9), and the image of the ground with cavities can be reconstructed well by minimizing the following objective function:

$$\min_{\mathbf{x}} \ \frac{1}{2}\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|_2+\lambda\left\|\mathbf{x}-\mathbf{x}_0\right\|_1 \tag{20}$$

where $\mathbf{x}_0$ represents the geotechnical parameters of the "healthy" areas. It is assumed here that vector $\mathbf{x}_0$ can be obtained through geotechnical boring when the transmitter and receiver are installed into the ground, so that vector $\mathbf{x}_0$ can be also considered as an additional prior information of the ground.

In this numerical test, we investigated two cases: without observation noise, or with observation noise (σ_ε = 2.0), and compared the results by sparse modeling with those of ridge regression. The choice of the regularization parameter λ is important in sparse modeling to obtain reasonable results. Although there have been several approaches, such as information criteria (e.g., Ninomiya and Kawano 2016), cross-validation (e.g., Bishop 2005), and stability selection (Meinshausen 2010), we used λ = 0.1 for noiseless problems, and λ = 1.0 for problems with noise. The accuracy of the reconstructed image is evaluated in terms of the root-mean-square error (RMSE):

$$\mathrm{RMSE}=\sqrt{1/n\sum_{i=1}^{n}(x_i-\hat{x}_i)^2} \tag{21}$$

where x_i is the true value of the geotechnical parameter for the i-th cell, shown in Figure 5, and $\hat{x}_i$ is the reconstructed (estimated) value for the i-th cell obtained through the tomography.

5.1.3 Results

Figure 6 shows the reconstructed images of the ground by ridge regression and sparse modeling. Figures 6 (a) and (b) are the results by the ridge regression, and (c) and (d) are those obtained by sparse modeling. In problem 1, the ridge regression can roughly identify the location of the cavities, but an image with some noise was obtained even in the noiseless case. On the other hand, the sparse modeling almost accurately identifies the cavities, and a considerably clear image was obtained. In problem 2, the ridge regression approximately detects the layers and also estimates the properties of the cells; on the other hand, the sparse modeling more accurately reconstructs the image and perfectly detects the soil layers. In terms of quantitative accuracy, the RMSE of sparse modeling (0.0164) is smaller than that of ridge regression (0.0537).

Figure 7 shows the results taking the observation noise into account. In problem 1 (Figures 7a and 7b), both of the two methods seem difficult to detect the cavities with high accuracy, but sparse modeling yields a clearer image than ridge regression. A similar trend can be seen in the layered-ground problem (Figures 7c and 7d), for which the sparse modeling reconstructs the image more accurately than ridge regression, and almost perfectly detects the layer boundaries.

In summary, sparse modeling can perform well in solving the cross-borehole tomography.

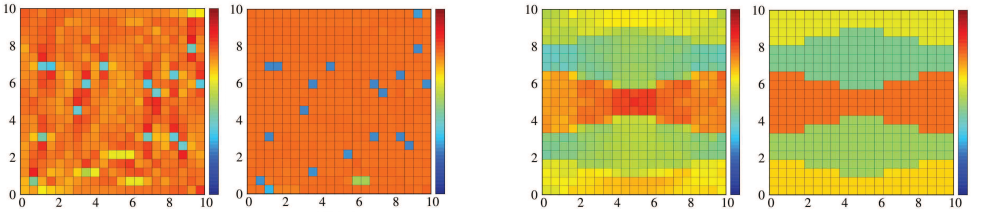


Figure 6. Reconstructed images of the ground (without observation noise).

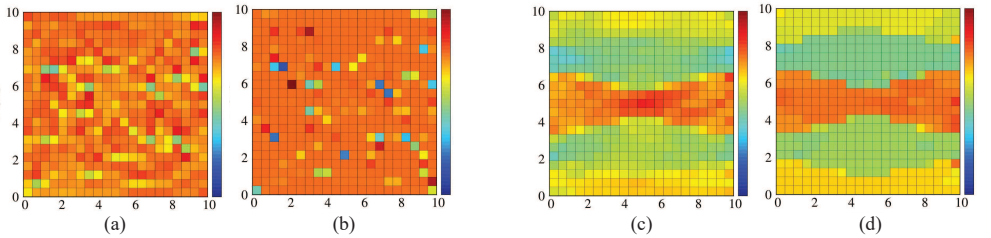


Figure 7. Reconstructed images of the ground (with observation noise).

5.2 Stratigraphic soil profiling

Stratigraphic profiling is a central task in geotechnical site investigation. Some studies have demonstrated the importance of soil stratification on the design of shallow foundations (e.g., Burd and Frydman 1997; Lee et al. 2013), deep foundations (e.g., Pardon et al. 2008), and structures on soft grounds (Huang and Griffiths 2010). In order to quantify the underground stratification, the number of layers (e.g., soil types) and their thickness (and hence their boundaries) should be identified.

In recent years, stratigraphic profiling based on the cone penetration test (CPT) has attracted attention because the CPT can provide data (tip resistance, sleeve friction, and pore pressure) with high spatial resolution (0.01 ~ 0.05 m) within a reasonable time. Because no soil samples are extracted, the CPT-based stratigraphic profiling is performed by means of a soil classification system. Among the available systems, the soil behavior type (SBT) and the I_c index are widely used. Nevertheless, due to the heterogeneity of soils, SBT and I_c vary spatially, and the interpretation of soil stratigraphy can be very difficult. An example of profile of I_c with depth is shown in Figure 8 (a) (Ching et al. 2015). Small- and large-scale fluctuations can be seen in the profiles. Figure 8 (b) shows the SBT profile based on the direct use of the I_c – SBT mapping. The SBT profile with depth also shows small-scale fluctuations in the data, suggesting the presence of many thing layers, less than 10 cm-thick, were identified. This results in a typical “unreasonable” soil stratification and cannot be used in geotechnical practice. To address the difficulty, several methods have been developed for CPT-based stratigraphic profiling using machine learning techniques (Jung et al. 2008; Wang et al. 2013).

As noted, the main advantage of sparse modeling is its ability to exploit some "structures" in the data, therefore it is expected that sparse modeling would be useful in stratigraphic soil stratification. The objective of this section is to conduct soil stratification using a sparse modeling-based method, and compare it with past studies in order to discuss the applicability of the proposed approach.

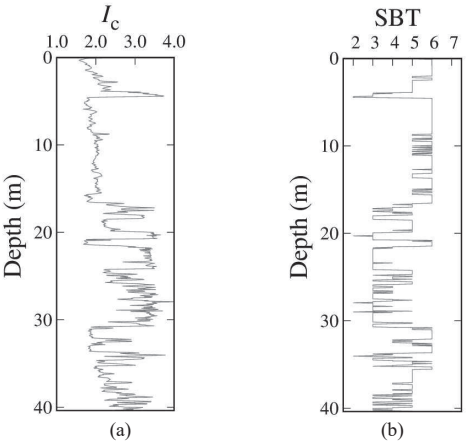


Figure 8. Profiles of I_c and SBT with depth.

5.2.1 Numerical setup

Stratigraphic soil profiling was conducted for the data shown in Figure 8. First, a hidden structure in the depth profile of I_c (Figure 8a) can be exploited using trend filtering with a TV regularization term, and the objective function Eq. (9) can be rewritten as:

$$\min_x \frac{1}{2}\sum_{i=1}^{m}(y_i - x_i)^2 + \lambda\sum_{i=1}^{m-1}|x_i - x_{i-1}| \tag{22}$$

where y_i is the value of I_c at depth i, x_i is the value of I_c at depth i (to be estimated), and m is the number of data values. Subsequently, the soil stratification is identified directly through the I_c - SBT mapping based on the filtered profile.

Although the lasso-based sparse modeling has many desirable properties, it is a biased estimator, for which the bias does necessarily disappear as $m \rightarrow \infty$. Moreover, this bias becomes evident in stratigraphic soil profiling problems. In order to deal with this bias, we used a two-stage algorithm called "relaxed lasso (Meinshausen 2008)". For more details on bias in lasso and the relaxed lasso, the reader is referred to Fan and Li (2001) and Meinshausen (2008).

5.2.2 Results

The estimated depth profiles of I_c and SBT are shown in Figure 9 and Figure 10, respectively. The red lines indicate filtered data, and the gray lines indicate original data. A 5-fold cross validation was used to determine the regularization parameter, and yielded a value of $\lambda = 0.2$. The filtered depth profile of I_s with $\lambda = 0.2$ (Figure 9a), however, still presents small and large fluctuations, and does not change significantly after trend filtering. The corresponding SBT profile (Figure 10a) seems unrealistic in soil stratification practice because many thin layers are identified. A possible technical reason for this failure is that the I_c profile has non-stationary noise. For example, although the upper part (depth from 0 to 20 m) of the I_c profile features a relatively small noise, the noise level becomes large in the lower part (depth > 20 m; Figure 8a). Analyzing data including non-stationary noise might be beyond the capabilities of the lasso, thus more flexible approaches, such as the Bayesian lasso mentioned in section **3.4**, should be employed to achieve a more reasonable stratigraphic soil profiling.

Nevertheless, sparse modeling has an attractive property, which is that of reasonably exploiting structures in the data without computational complexity. We analyzed the data using different regularization parameters, namely λ = 1.0, 2.0, 5.0, and 10.0 to investigate the performance of sparse modeling in stratigraphic soil profiling. Figures 8 and 9 show the filtered I_c and SBT profiles with different choices of λ. The larger the regularization parameter λ, the simpler trend is identified. We assumed the result with $\lambda = 5.0$ to represent a reasonable soil stratification, and compared it with results of past studies. The comparison is shown in Figure 11. No widely accepted or quantitative criteria exist to define "what is the reasonable soil stratification";

nevertheless, sparse modeling seems to yield a reasonable soil stratification which can be used in geotechnical practice.

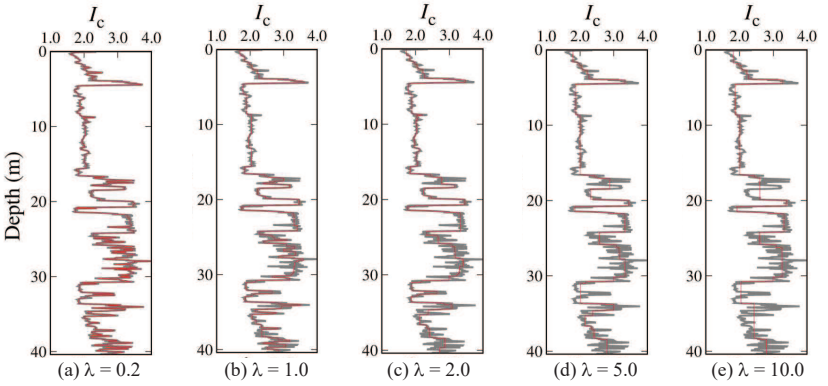


Figure 9. Filtered profile of I_c with depth, obtained using different regularization parameters.

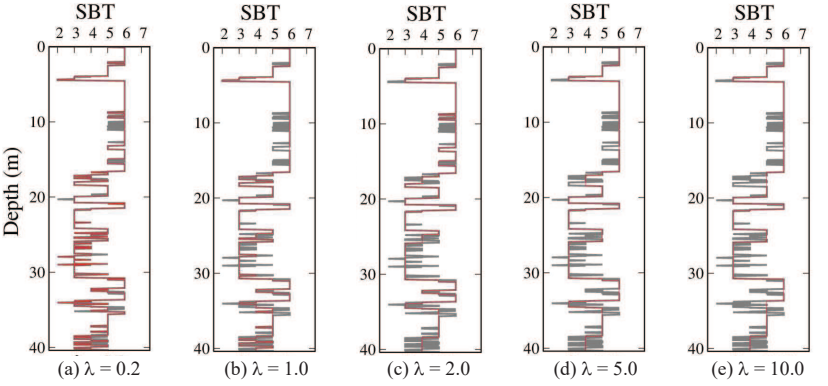


Figure 10. Profile of SBT with depth based on the filtered I_c.

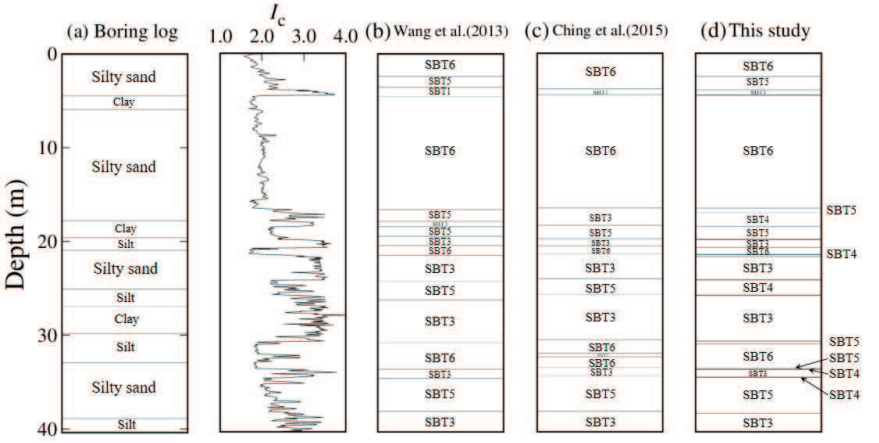


Figure 11. Comparison of stratification results (Modified from Ching et al. 2015).

6 Summary

This paper demonstrated the potential of sparse modeling in geotechnical engineering applications by means of two practical examples: cross-borehole tomography, and stratigraphic soil profiling based on cone penetration testing. If a "solution sparsity" can be properly induced, i.e., the regularization term is properly modeled depending on the problems, sparse modeling performs well in geotechnical engineering application. In particular, the reconstruction accuracy in cross-borehole tomography can be considerably improved using sparse modeling. The trend filtering via sparse modeling is also promising in stratigraphic soil profiling. However, sparse modeling is not exempt from technical limitations, such as in the case of non-stationary noise in the data, and further research is needed to investigate the applicability of sparse modeling in geotechnical engineering.

Acknowledgments

I would like to thank Prof. Yoshida (Tokyo City University), Prof. Ching (National Taiwan University) and Prof. Phoon (National University of Singapore) for valuable discussions and comments. This research was partly supported by JSPS KAKENHI Grant Numbers JP18K05880 and JP16H02577.

References

Asaoka, A. (1978). Observational procedure of settlement prediction. *Soils and Foundations*, 18(4), 87-101.

Bishop, C.M. (2005). *Pattern Recognition and Machine Learning*, Springer.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2010). Distributed optimization and statistical learning via alternating direction method of multipliers. *Found. Trends Machine Learning*, 3(1), 122.

Burd, H.J. and Frydman, S. (1997). Bearing capacity of plane-strain footings on layered soils. *Canadian Geotechnical Journal*, 34, 241-253.

Ching, J., Wang, J.S., Juang, C.H., and Ku, C.S. (2015). CPT-based stratigraphic profiling using the wavelet transform modulus maxima. *Canadian Geotechnical Journal*, 52(12), 1993-2007.

DeVito, E., Rosasco, L., Caponnetto, A., DeGiovannini, U., and Odone, F. (2005). Learning from examples as an inverse problem. *Journal of Machine Learning Research*, 6, 883-904.

Donoho, D.L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3), 613-627.

Fang, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association*, 96, 1348-1360.

Finno, R.J. and Calvello, M. (2005). Supported excavations: observational method and inverse modeling, *J. Geotech. Geoenviron. Eng.*, 131(7), 826-836.

George, E. and McCulloch, R.E. (1993). Variable selection via Gibbs sampling. *The Journal of the American Statistical Association*, 88, 881-889.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge Regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.

Honjo, Y. and Kashiwagi, N. (1991). On the optimum design of a smoothing filter for geophysical tomography. *Soils and Foundations*, 31(1), 131-144.

Honjo, Y., Benny, L., and Liu, W.T. (1993). Prediction of single pile settlement based on inverse analysis. *Soils and Foundations*, 33(2), 126-144.

Honma, M., Akiyama, K., Uemura, M., and Ikeda, S. (2014). Super-resolution imaging with radio interferometer using sparse modeling. *Publ. Astron. Soc. Jpn*, 66 (5), 95-1-14.

Huang, J. and Griffiths (2010). One-dimensional consolidation theories for layered soil and coupled and uncoupled solutions by the finite-element method. *Geotechnique*, 60(9), 709-713.

Jung, B.-C., Gardoni, P., and Biscontin, G. (2008). Probabilistic soil identification based on cone penetration tests. *Geotechnique*, 58(7), 591-603.

Lustig, M., Donoho, D., and Pauly, J.M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6), 1182-1195.

Meinshausen, N. (2008). Relaxed lasso. *Computational Statistics & Data Analysis*, 52, 374-393.

Meinshausen, N. (2010). Stability selection. *Journal of the Royal Statistical Society, Series B*, 72(4), 417-473.

Natarajan, B.K. (1998). Sparse approximate solutions to linear systems. *SIAM J, Comput.*, 24(2), 227-234.

Ninomiya, Y. and Kawano, S. (2016). AIC for the LASSO in generalized linear models. *Electronic Journal of Statistics*, 10, 2537-2560.

Padron, L.A., Aznarez, J.J., and Maeso, O. (2008). Dynamic analysis of piled foundations in stratified soils by a BEM-DEM model. *Soil Dynamics and Earthquake Engineering*, 28, 333-346.

Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association,* 103(482), 681-686.

Sakurai, S. and Takeuchi, K. (1983). Back analysis of measured displacement of tunnels. *Rock Mechanics and Rock Engineering*, 16(3), 173-180.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc. B.*, 58(1), 267-288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knit, K. (2005). Sparsity and smoothness via the fused lasso, *J. R. Statist. Soc. B*, 67(1), 91-108.

Wang, Y., Huang, K., and Cao, Z. (2013). Probabilistic identification of underground soil stratification using cone penetration tests. *Canadian Geotechnical Journal*, 50, 766-776.