Stratigraphic Profiling using Sparse Modeling

Y. Tsuda[1] and M. Kagehira[2]

[1]*Tokyo City University. Email: y.tsuda28@gmail.com*
[2]*Okayama University. Email: pnul4v67@s.okayama-u.ac.jp*

Abstract: In recent years, sparse modeling has attracted attention as an effective analysis method for ill-posed inverse problems, and has achieved successful results in various fields. In general, the cone penetration test (CPT) is conducted when geological survey is conducted. Soil behavior test (SBT) and I_c index are widely used for stratigraphic profiling based on CPT. However, since SBT and I_c contains lots of noise and it is very difficult to interpret the soil stratigraphy, there is a need to use an advanced inverse analysis method such as sparse modelling. After introducing basic concept and formulation of sparse modelling, we show practical examples of soil stratification using sparse modeling with actual ground data. Fused lasso, which is one of popular formulation in sparse modelling, assumes sparsity in first order difference space. We apply this fused lasso to stratification problem and have good results. The regularization parameter, which is a key parameter in lasso, is determined empirically. It is pointed out that cross-validation which is the most popular way to determine the regularization parameter has a problem when correlated noise is contained in the data.

Keywords: sparse modelling, inverse problems, fused lasso, sparsity, cross validation .

1. Introduction

Stratigraphic profiling is a central task in geotechnical site investigation. Some studies have demonstrated the importance of soil stratification on the design of geotechnical structures. 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 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 1 (a). Small- and large-scale fluctuations can be seen in the profiles. Figure 1 (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 thin 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).

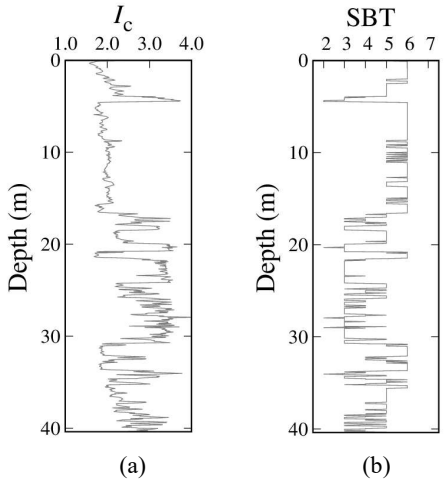


Figure 1. Profiles of I_c and SBT with depth

The machine learning method known as "sparse modeling" 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). In this paper, lasso is referred to as sparse modeling. 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. Therefore, it is expected that sparse modeling would be useful in stratigraphic soil stratification.

This paper demonstrates the potential of sparse modeling for solving geotechnical engineering problems by showing practical examples. 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, its mathematical fundamentals and numerical algorithms to solve sparse modeling problems are summarized are outlined; in Section 4, Then, describes future topics in this study are shown in Section 5 and, finally a summary of the results of sparse modeling in geotechnical engineering are presented in Section 6.

2. Stratigraphic Profiling as Inverse Analysis

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

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

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. In stratigraphic profiling, $\mathbf{y}$ is the measured depth profile of Ic, and $\mathbf{x}$ is the exploited structure (a trend function) from $\mathbf{y}$. 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

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 2. In the figure, colored cells indicate non-zero components (also called active-set), and white cells indicate zero components. When the solution **x** is sparse, and if we know how many zeros are and which components are nonzero in **x**, the under-determined problem can be solved by minimizing the following

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

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:

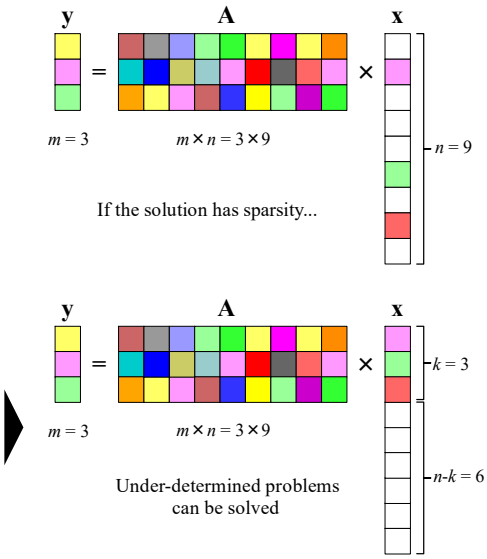


Figure 2. The role of solution sparsity in solving an under-determined problem.

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

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 (8) is shown in Figure 3. 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).

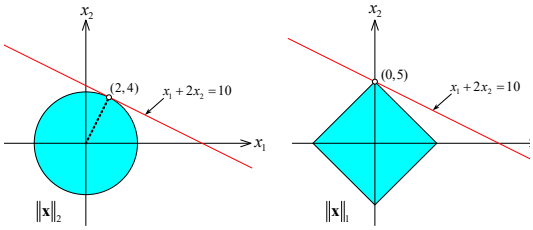


Figure 3. Schematic comparison between ridge and lasso.

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.

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:

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

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:

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

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.

This study applies fused lasso to exploit a hidden structure in the depth profile of I_c (Figure 1a), and it can be done using the fused lasso with a TV regularization term. The target problem is one-dimensional soil profile and the objective function Eq. (9) can be rewritten as the following equation for simplicity:

$$\min_{x} \frac{1}{2}\sum_{i=1}^{m}(y_i - x_i)^2 + \lambda\sum_{i=1}^{m-1}|x_i - x_{i-1}| \tag{12}$$

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. Eq. (11) is particularly called trend filtering (Kim et al. 2009).

3.4 Relaxed lasso

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).

3.5 Bayesian view of lasso

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

$$p(x_j) = \frac{\lambda}{2\tau} \exp\left(-\frac{|x_j|}{\tau}\right) \tag{13}$$

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. Application Examples

4.1 Training data

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 4 show the filtered I_c profiles with different choices of λ. The larger the regularization parameter λ, the simpler trend is identified. We empirically determined the reasonable regularization parameters in lasso for the test data, which is λ =7, and compared it

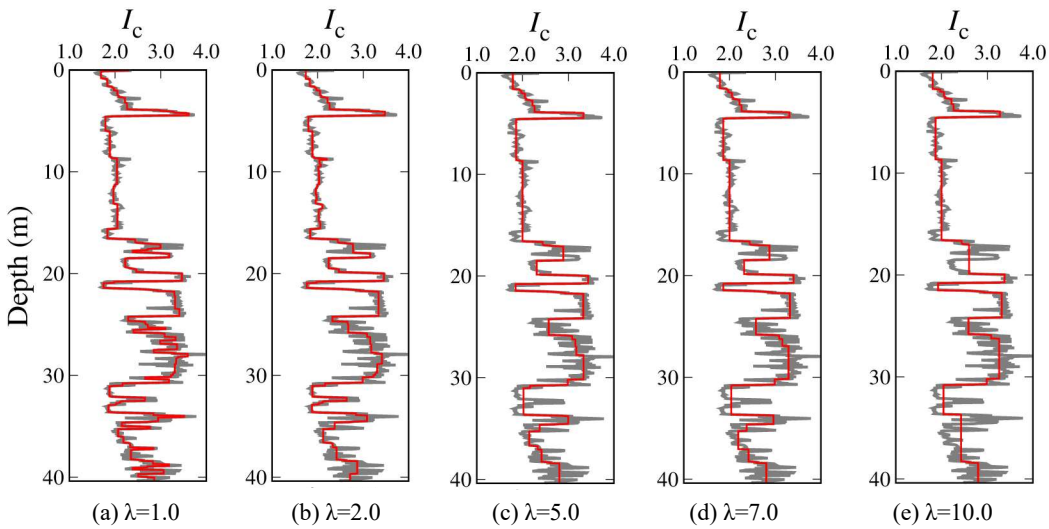


Figure 4. Filtered profile of I_c with depth, obtained using different regularization parameters.

with results of past studies. The comparison is shown in Figure 5. It matches very well with past studies shown in Figure 6(b) and (c). sparse modeling seems to yield a reasonable soil stratification which can be used in geotechnical practice.

4.2 Test data

We empirically determined the reasonable regularization parameters in lasso for the test data, which are λ=7 for (a), 5 for (b), and 7 for (c) respectively. Figure 6 shows the estimated trend of I_c. In the figure, red line indicates estimated trend with the selected λ, and the gray line indicates the original Ic profile. Since the noise in test data (a) is very small, there is no big difference between estimated and original I_c profiles. In test data (b), the noise seems large compared with data (a). Sparse modeling approach works well and gives reasonable trend. The proposed method works well for data (c) also and capture the trend of distribution generally. However, the proposed method could not capture the "spike" around z = 20m. It might

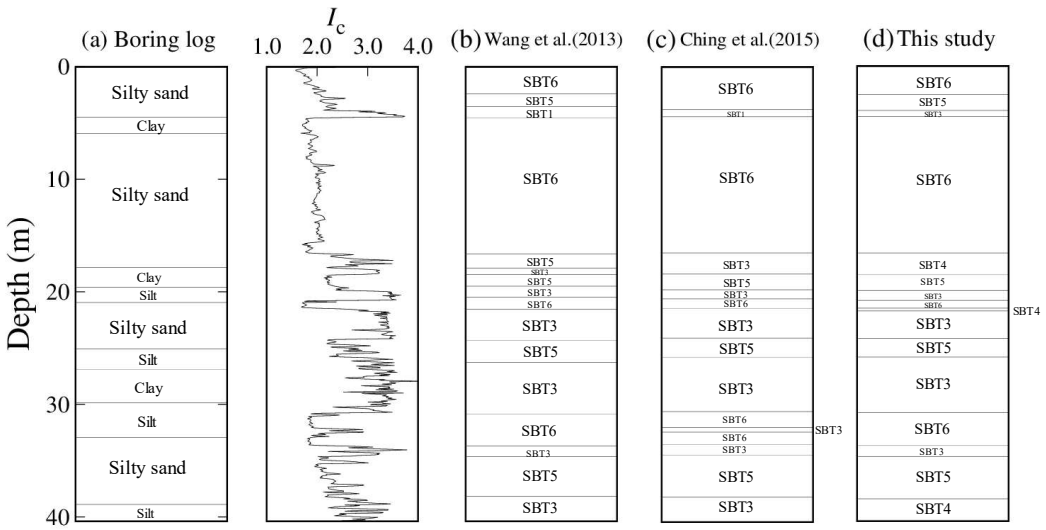
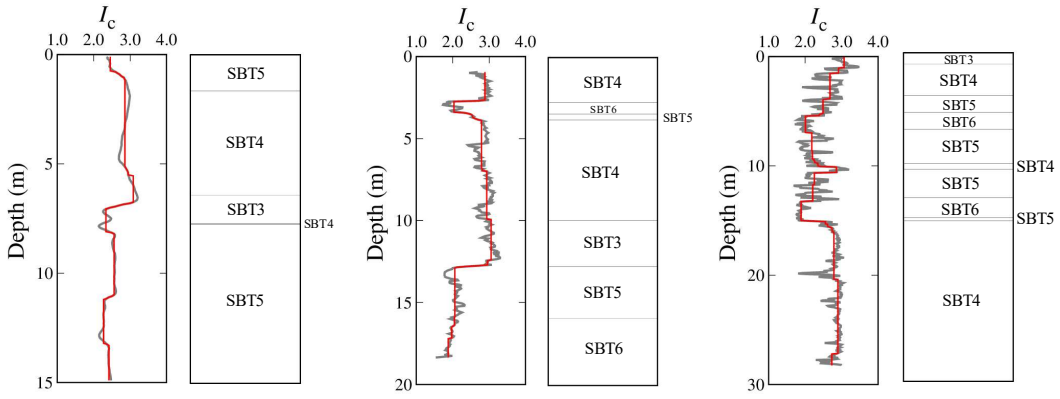


Figure 5. Comparison of stratification results (Modified from Ching et al. 2015).



(a) Test data (a), λ=7.0 (b) Test data (b), λ =5.0 (c) Test data (c), λ =7.0

Figure 6. Stratified I_c with λ determined empirically for the test data a, b, c.

be due to non-stationarity in the noise. It seems the noise in upper part is larger than that of lower part.

5. Future topics

In Chap.4, the regularization parameter λ is determined empirically for stratification of test data. The most popular method to determine the regularization parameter λ is cross-validation. We also tried *K*-fold cross validation (CV) with one-standard error rule to select the best λ, which 0.1 for (a) and 0.2 for (b) respectivery. Figures 7 shows the stratified Ic with λ selected by CV for test data (a) and (b).

The stratified result does not change significantly after trend filtering. Figure 7(a) seems reasonable because there is very small noise in the data. However, Figure 7(b) is against our intuition. A possible reason is that the I_c profile contains not white noise but correlated noise. We tried formulation considering correlated noise (Yoshida et al. 2017) and have stratification results shown in Figure 8. Stratification for test data (a) is almost same because the noise level is very small irrespective of correlated or uncorrelated. Stratification for test data (b) is improved but shows only two layers. It suggests the importance of consideration of correlation in the noise for stratification problems though we need more improvement of the proposed method.

Non-stationarity is another challenging research topic. The characteristics of noise (level and correlation) in test data (c) and training data change depending on depth. We need special procedure to perform stratification of data with such non-stationarity.

6. Conclusions

This study shows a practical example of soil stratification using sparse modeling, fused lasso for actual ground data. It is confirmed that the stratification can be performed flexibly according to the characteristics of the obtained soil data by a simple procedure. However, determination of regularization parameter λ is remaining research topics as stated in Chap.5.

References

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.

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

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.

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.

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

Kim, S.J., Koh, K., Boyd, S. and Gorinevsky, D. (2009). ℓ_1 trend filtering, *SIAM REVIEW*, 51(2), 339-360.

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.

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

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

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.

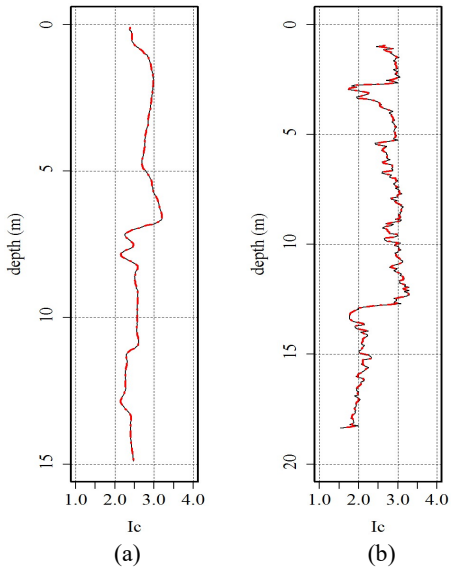


Figure 7. Stratified Ic with λ selected by CV for test data (a) and (b).

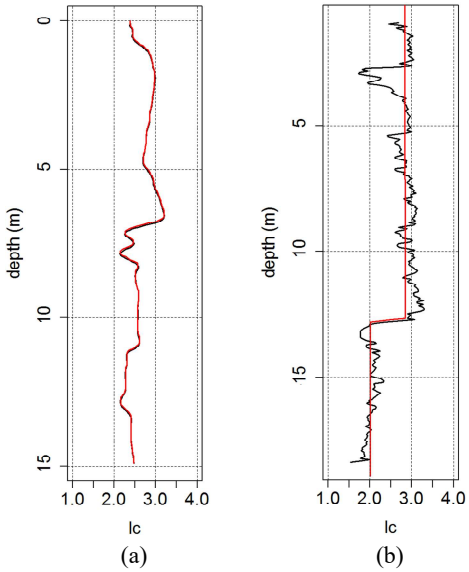


Figure 8. Stratified Ic considering correlated noise for test data (a) and (b).

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

Yoshida, I., Tasaki, Y., Nishimura, S. and Imaide, K. 2019. Basic Study on Conditional Random Field with Sparse Modelling, Proceedings of 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019), 2019.12