

スパースモデリングによる首都圏における地震動分布推定手法の開発

水迫 覚信¹・長尾 大道²・廣瀬 慧³・加納 将行⁴・堀 宗朗⁵

¹正会員 東京大学 学生 東京大学地震研究所 (〒113-0032 東京都文京区弥生一丁目1-1)

E-mail: mizusako@eri.u-tokyo.ac.jp

²非会員 東京大学准教授 博士 (理学) 東京大学地震研究所 (〒113-0032 東京都文京区弥生一丁目1-1)

³非会員 大阪大学助教 博士 (理学) 大阪大学大学院基礎工学研究科 (〒560-8531 大阪府豊中市待兼山町1-3)

⁴非会員 東京大学 博士 (理学) 東京大学地震研究所 (〒113-0032 東京都文京区弥生一丁目1-1)

⁵正会員 東京大学教授 Ph. D 東京大学地震研究所 (〒113-0032 東京都文京区弥生一丁目1-1)

構造物の即時被害予測は、大地震により損傷を受けた構造物を数値シミュレーションにより即座に同定し、復旧を効率化する仕組みである。シミュレーションの入力となる構造物における地震動は観測データに基づき推定され、高精度な被害予測には観測データよりも圧倒的に高空間分解能を持つ地震動分布を推定する必要がある。発表者らは首都圏地震観測網 (MeSO-net)¹⁾ で得られる観測データから、高空間分解能を持つ地震動分布を推定する手法を開発している。

水迫 [2013, 卒業論文]²⁾ は、テイラー展開に基づく手法を提案し、東北地方太平洋沖地震発生時に MeSO-net で観測されたデータに手法を適用した結果、0.15 Hz 以上の地震動分布を得ることが難しいことを示した。一般的な構造物の固有周波数は 1-10 Hz であるため、手法の実用化のためには 0.15 Hz 以上の周波数成分を含む地震動分布推定が求められる。テイラー展開に基づく手法を用いる際、テイラー展開の打ち切り次数、及びモデルを適用する観測点 (以降、クラスタと称する) をチューニングする必要があるが、水迫 [2013] は恣意的に一次でテイラー展開を打ち切り、ターゲットとする構造物の近傍 5 観測点をクラスタとした。

本研究では、lasso (least absolute shrinkage and selection operator)³⁾ を応用し、打ち切り次数とクラスタを客観的にチューニングする手法の開発に取り組んでいる。lasso は係数の L1 ノルムを正則化項としており、有効な係数を自動的に選択する。本研究で提案する手法では、所与のクラスタに対する打ち切り次数を lasso によって自動判定し、情報量規準に基づき適切なクラスタを選択する。初期的な結果として、0~1 Hz において、本手法によって既往研究よりも高い精度で地震動分布が推定されるという結果が得られている。

Key Words: Imaging of Ground Motion, Seismic Data, Sparse Modeling, Lasso

1. はじめに

構造物の即時被害推定は、大地震発生後の復旧を効率化するための仕組みであり、二次災害を防ぐうえで重要な役割を果たす。構造物の被害予測シミュレーションの入力値として、地震観測網で観測される地震動データからターゲットとする構造物における地震動が推定されるが、高精度な被害予測のためには、観測点よりも圧倒的に高空間分解能を持つ地震動分布を推定する必要がある。首都圏では、約 2~5km 間隔に約 300 観測点を有した首都圏地震観測網 (MeSO-net、図 1)¹⁾ が整備されており、東西・南北・上下方向の 200Hz サンプルングの加速度データをリアルタイムに得ることができる。発表者らは MeSO-net で得られる観測データから高空間分解能を持つ地震動分布を推定する手法を開発している。

水迫 [2013, 卒業論文]²⁾ はテイラー展開に基づく次の回帰式によって地震動分布を推定する手法を提案

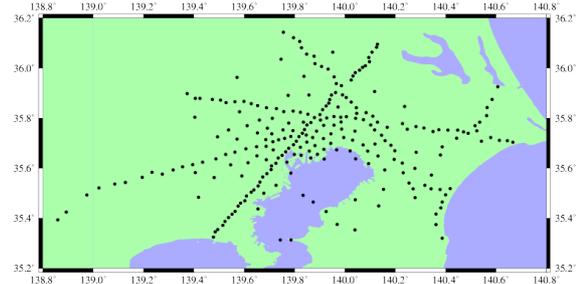


図-1 MeSO-net 観測点

した。

$$u_i(x_0 + \Delta x, \omega) = u_i(x_0, \omega) + \frac{\partial u_i(x_0, \omega)}{\partial x_j} \Delta x_j + \frac{1}{2!} \frac{\partial^2 u_i(x_0, \omega)}{\partial x_j \partial x_k} \Delta x_j \Delta x_k + \dots \quad (1)$$

ただし、 u は周波数領域に変換された地震動、 x_0 は地

震動を推定したい地点（ターゲット地点とする）の位置座標、 $x_0 + \Delta x$ は近傍観測点の座標、 ω は周波数、 i, j, k, \dots は x, y, z 方向に対応している。式 (1) において、左辺の $u_i(x_0 + \Delta x, \omega)$ は観測点における地震動であり既知、右辺の $\Delta x_i \Delta x_j \dots$ も観測点とターゲット地点との位置座標差で表されるので既知である。従って、式 (1) はターゲット地点における地震動 $u_i(x_0, \omega)$ と、同地点における偏微分係数を未知数とした線形回帰式 $y = x\beta$ を解く問題に帰着される。

回帰式 (1) を使用するあたり、テイラー展開の打ち切り次数と係数の決定に用いる観測点（クラスタと称する）を設定する必要がある。回帰式にテイラー展開近似を用いているため、クラスタの半径が小さく、打ち切り次数が高次であるほど推定精度が良いが、打ち切り次数が大きくなると係数を優決定するために必要な観測点数が増加し（打ち切り次数が Q のとき、未知数の個数 P は $P = \frac{Q(Q^2 + 6Q + 11)}{6}$ と表される）、クラスタの半径が大きくなるというトレードオフ関係が存在する。また、MeSO-net は首都圏に設置されており、観測ノイズを多く含むため、未知数の個数 P に対してクラスタサイズ N が小さいと、ノイズにより推定精度が低下する恐れがある。

水迫 [2013] は、 $P = N + 2$ の仮定のもとで水平成層構造の解析解を用いてチューニングした結果、打ち切り次数を 1 次、クラスタを近傍 5 観測点と定め、東北地方太平洋沖地震時の MeSO-net データに手法を適用したが、0.15Hz 以上周波数成分を含む地震動分布を推定することは難しいことが示された。なお、係数の推定には通常最小二乗法（OLS）が採用されている。

2. 手法

本研究では lasso (least absolute shrinkage and selection operator) ³⁾ に基づいた手法を提案する。 D 次元、データ数 N の観測データに対する線形回帰式 $y = x\beta$ (ただし、 $y \in \mathbb{R}^{D \times N}$ は観測データ、 $x \in \mathbb{R}^{N \times P}$ は説明変数、 $\beta \in \mathbb{R}^{D \times P}$ は係数) に対して、lasso は損失関数に係数の L1 ノルムを正則化項として加えた次の評価関数を最適化する。

$$J = N \log \sqrt{(2\pi)^P |R|} + \frac{1}{2} \text{tr}(y - x\beta)^T (y - x\beta) R^{-1} + N\lambda \|\beta\|_1 \quad (2)$$

ただし、 R は分散共分散行列、 $\lambda > 0$ は正則化パラメータであり、 λ が大きいほど選択される係数の個数は少なくなり、十分に大きい λ において選択される係数の個数は 0 個となる。lasso の評価関数を最適化すると、 $P > N$ の際にも分散の小さい推定を得ることが可能であることに加え、有効な係数以外を正確に 0 としたス

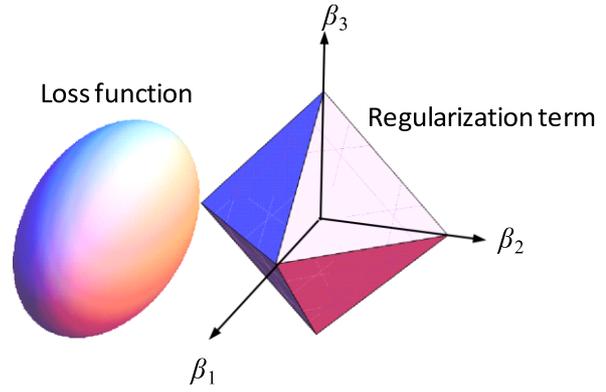


図-2 lasso の幾何学的解釈：3次元空間において、損失関数に対応する楕円球が正則化項に対応する正八面体の頂点あるいは辺上で接しやすいため、係数に 0 を含んだスパースな推定値が得られやすいことが分かる。

パースな解が得られるという特徴がある。従って、今回の問題設定において、lasso は所与のクラスタにおけるテイラー展開打ち切り次数を自動的に決定すると言い換えることができる。本研究では lasso による打ち切り次数選択と合わせ、情報量規準によるクラスタ選択によって、これらのチューニングを行う。また、式 (2) の最適化には加速付き近接勾配法⁴⁾を用いる。

複数の係数の組をグループとして扱うための lasso を拡張した手法として、共通のグループ g に属する係数の L2 ノルムを正則化項とする group lasso⁵⁾が提案されている。group lasso で用いる正則化項は次の式 (3) で表される。

$$\lambda \sum_{g \in G} \sqrt{|g|} \|\beta_g\|_2 \quad (3)$$

ただし、 $|g|$ はグループ g に属する係数の個数、 β_g は g に属する係数を並べたベクトルである。group lasso を用いることにより、共通のグループに属する係数が同時に 0 または非ゼロと推定されることに加え、座標系の選び方に依存しない推定値を得ることができる。本研究においては、group lasso を適用する座標変換に不変であるべきグループとして、フーリエ変換された地震動成分の実部・虚部、及び x, y, z 成分、同じ階数の偏微分係数群が挙げられる。推定の結果は時間領域における時刻 0 の定め方に対して不変であるべきであり、時間領域における $t \rightarrow t + \Delta t$ なる変換は、周波数領域における変換：

$$\begin{pmatrix} \text{Re} \\ \text{Im} \end{pmatrix} \rightarrow \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \text{Re} \\ \text{Im} \end{pmatrix} \quad (4)$$

に対応するため、フーリエ変換された地震動成分の実部・虚部を座標変換に不変であるべきグループとしている。

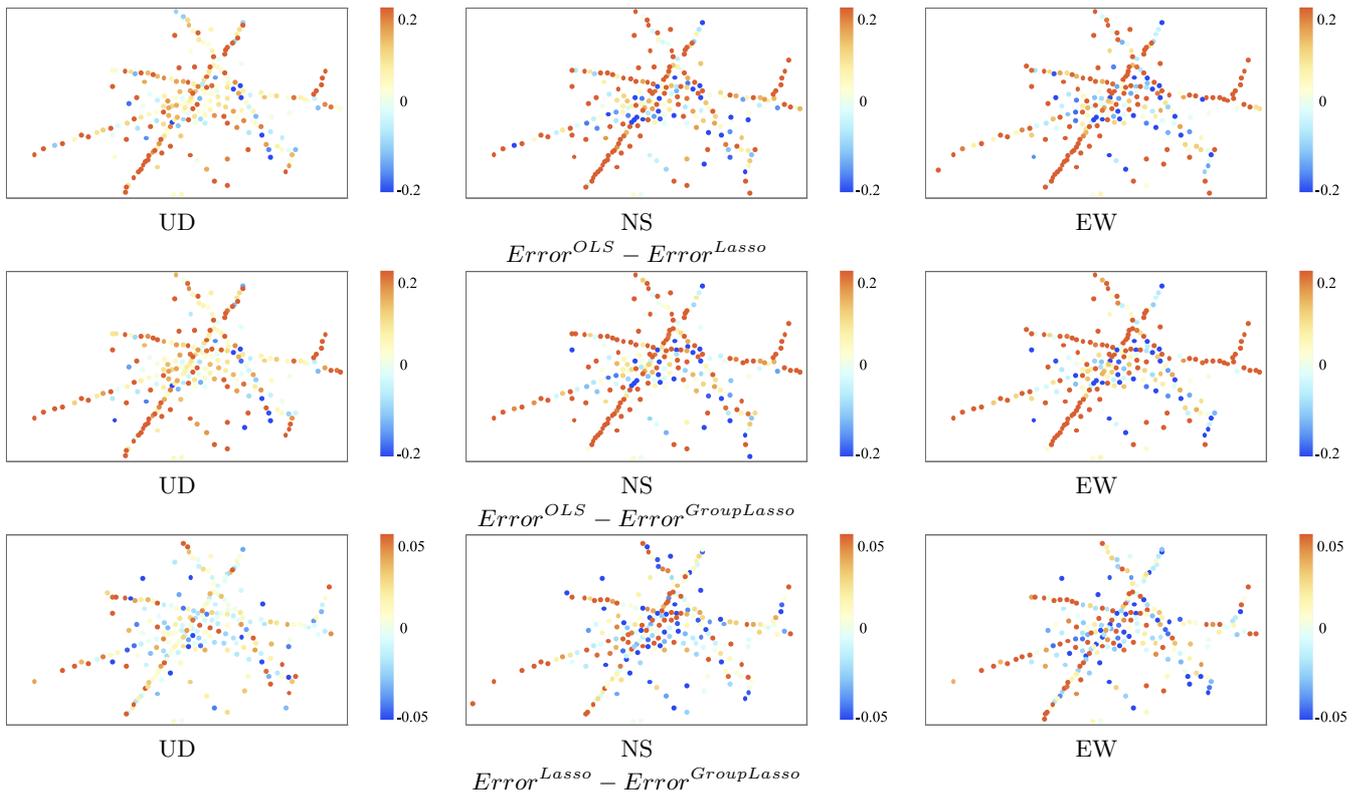


図-3 0~0.15 Hz における推定精度比較 (EBIC を最小化する λ を選択した場合)

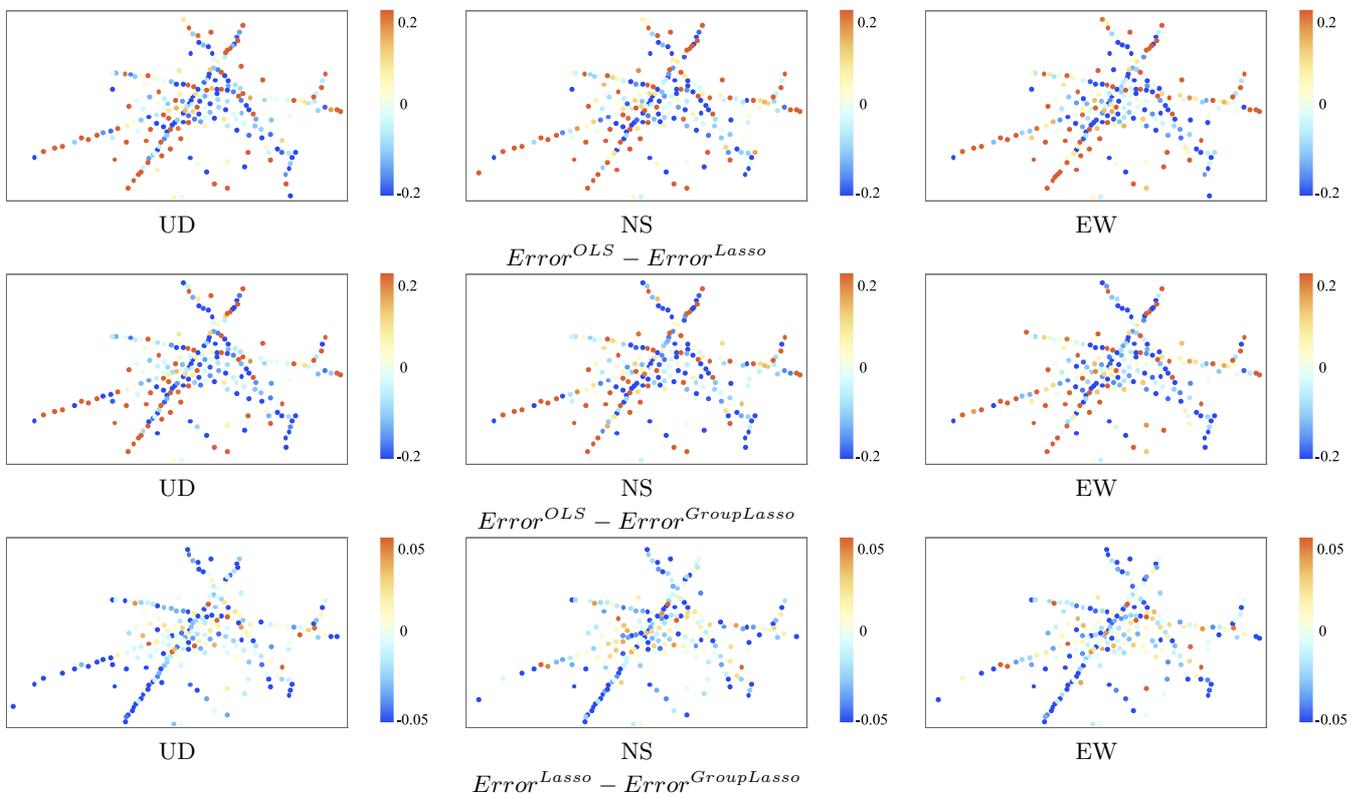


図-4 0~1 Hz における推定精度比較 (EBIC を最小化する λ を選択した場合)

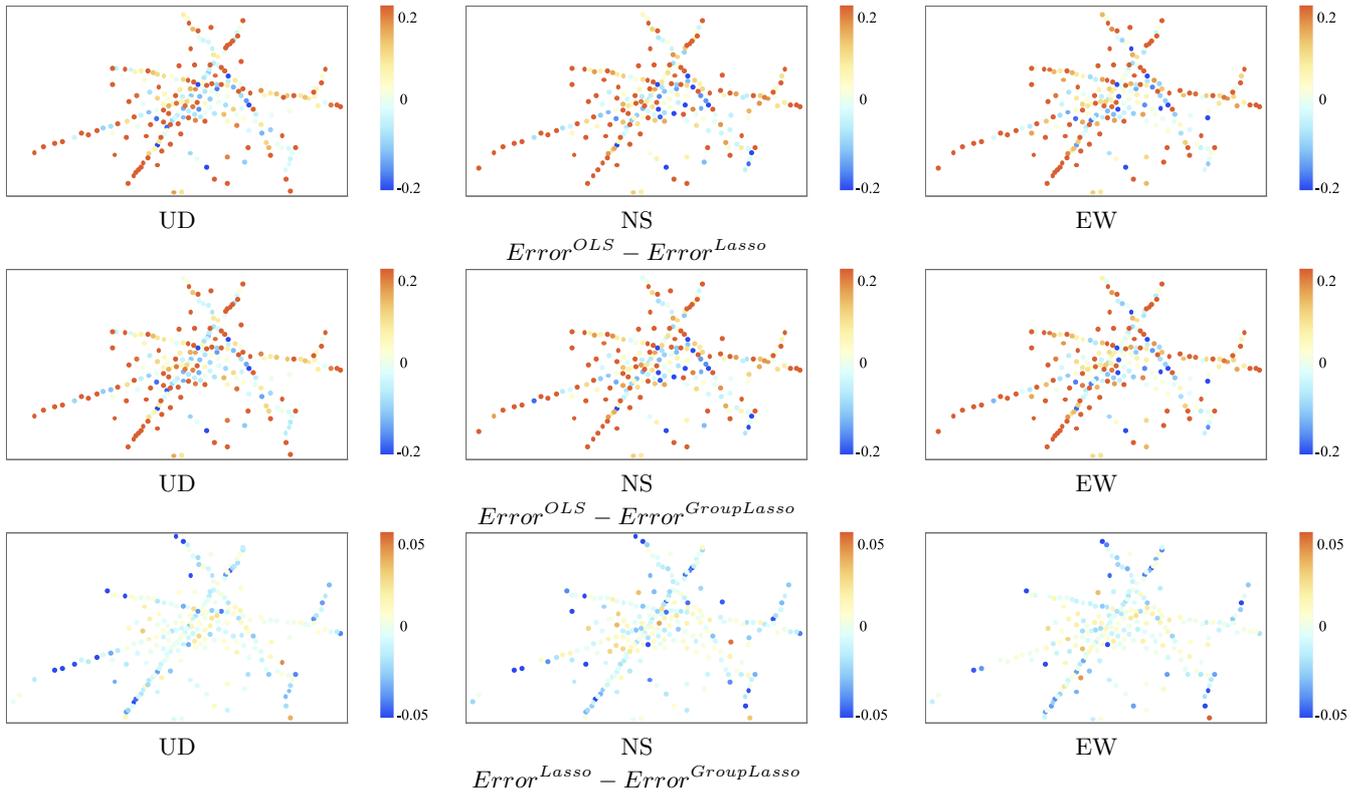


図-5 0~1 Hz における推定精度比較 ($\lambda = 0.178\lambda_0$ と設定した場合)

3. 数値実験

(1) 問題設定

初期結果として、水迫 [2013] と同様にクラスタサイズを $N = 5$ と固定して、本手法を東北地方太平洋沖地震時の MeSO-net データに適用した際の結果を報告する。水迫 [2013] は 1 次のテイラー展開モデルを採用したが、lasso を適用する際には、3 次のテイラー展開モデルを用いて有効な偏微分係数を推定する。ただし、クラスタサイズ N が未知パラメータ数 P に比べて小さい場合、分散共分散行列 R を変動させて評価関数 (2) を最適化すると解が不安定になるため、今回は R を単位行列として固定する。推定精度を議論するため、地震動を推定する地点として MeSO-net 観測点がある地点を選択し、推定された地震動と観測点で実際に観測される地震動を比較する。適用する手法は、水迫 [2013] (OLS) による手法、lasso による手法、地震動成分の実部・虚部をグループ化した group lasso による手法である。

(2) 初期結果

lasso や group lasso を適用する際、式 (2) や式 (3) における正則化パラメータ λ を設定する必要がある。例えば、未知パラメータ数 P が観測データの個数 N に比

べて大きい場合 ($N < P$) に有効な情報量規準として次式の EBIC^(6), 7) が提案されている。

$$EBIC = -2 \log L + df \log N + 2\gamma \log P C_{df} \quad (5)$$

ただし、 L はモデルの尤度、 df は自由度 (lasso によって推定される非ゼロのパラメータ数がモデルの自由度 df の不偏推定量である⁸⁾)、 N はクラスタサイズ、 P は未知パラメータ数、 γ は $0 \leq \gamma \leq 1$ で定義されたパラメータである。図-3、図-4 は、 $\gamma = 1 - \frac{1}{2 \log_N P}$ と設定し、EBIC を最小化する λ を選択した際に、水迫 [2013] で提案された手法 (OLS)、lasso、group lasso による推定値の誤差を比較した結果である。便宜上、推定値の誤差として式 (6) で定義された、誤差ノルムを観測値のノルムで除して無次元化したものを用いている。

$$Error \equiv \frac{\sqrt{\sum_i |u^{estimate}(\omega_i) - u^{observe}(\omega_i)|^2}}{\sqrt{\sum_i |u^{observe}(\omega_i)|^2}} \quad (6)$$

図-3 より、多くに観測点で 0~0.15 Hz においては lasso 及び group lasso が水迫 [2013] による誤差 $Error^{OLS}$ を改善していることが観察できるが、図-4 より、0~1 Hz において lasso および group lasso があまり有効ではないことが示唆される。EBIC は $N < P$ において有効な情報量規準であるとされているものの、本研究では EBIC の最小化によって、極端に大きいあるいは小さい λ が選択されることを確認しており、正則化パラメー

タを適切に設定できていないと考えられる。例えば図-5は、恣意的に $\lambda = 0.001^{50/200}\lambda_0 \approx 0.178\lambda_0$ と設定した場合の結果であるが、lasso 及び group lasso によって 0~1 Hz においても多くの観測点で $Error^{OLS}$ が改善されていることが分かる（ただし、 λ_0 は全ての係数を 0 と推定する最小の λ の値である）。高い推定精度を与える正則化パラメータの値は観測データに依存し、 $\lambda = 0.001^{50/200}\lambda_0$ のように一律に決定できるものではない。そのため、正則化パラメータのチューニングによって、図-5の結果をさらに改善することが可能であると考えられる。また、図-3、図-4、図-5より、lasso と、実部・虚部をグループ化した group lasso の優劣は判定し難く、特に図-5では $Error$ の大きさが同程度であることが分かる。

4. おわりに

本研究では、lasso による地震動分布推定手法を提案し、0~1 Hz において従来の手法よりも精度良く地震動を推定できることを示した。今後、以下に述べる数理的なアプローチと物理的なアプローチによって本手法を発展させたいと考えている。例えば数理的なアプローチとして、上述の正則化パラメータのチューニングに加え、観測値の x, y, z 成分、同階の偏微分係数をグループ化した group lasso の実装、説明変数の多重共線性が強い場合に安定した推定値を得るための elastic net⁹⁾ の実装が考えられる。また、地表面における境界条件や波動方程式、地盤の地下構造といった物理情報の組み込みによって、本手法をさらに発展させられる可能性がある。

参考文献

- 1) 笠原敬司、酒井慎一、森田裕一、平田直、鶴岡弘、中川茂樹、楠城一嘉、小原一成：首都圏地震観測網 (MeSONet) の展開、地震研究所彙報、Bull. Earthq. Inst. Vol.84 (2009) pp.71-88
- 2) 水迫覚信：高密度地震観測網を用いた時刻歴地震動分布推定、卒業論文 (2013)
- 3) Tibshirani: Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society. Series B (Methodological), Volume 58, Issue (1996), 267-288.
- 4) 富岡亮太：スパース正則化学習の理論とアルゴリズム、日本応用数理学会論文集、Vol. 23, No. 3, 2013, pp. 485-515
- 5) Yuan and Lin: Model selection and estimation in regression with grouped variables, J. R. Statist. Doc. B (2006) 68, Part1, pp. 49-67
- 6) Chen and Chen: Extended Bayesian Information Criteria for Model Selection with Large Model Spaces, Biometrika (2008) 95 (3), pp. 759-771
- 7) Chen and Chen: Extended BIC for small-n-large-P sparse GLM, Statistica Sinica 22 (2012), pp. 555-574
- 8) Zou, Hastie, Tibshirani: ON THE “DEGREES OF FREEDOM” OF THE LASSO, The Annals of Statistics 2007, Vol. 35, No. 5, pp. 2173-2192
- 9) Zou and Hastie: Regularization and variable selection via the elastic net, J. R. Statist. Soc. B (2005) 67, Part2, pp. 301-320
- 10) Friedman, Hastie, Tibshirani: Regularization Paths for Generalized Linear Models, Journal of Statistical Software, 2010, Volume 33, Issue, pp. 1-22
- 11) Chand: On Tuning Parameter Selection of Lasso-Type Method - A Monte Carlo Study, Applied Sciences and Technology (IBCAST), 2012 9th International Bhurban Conference on, pp. 120-129
- 12) Wang, Li, Leng: Shrinkage tuning parameter selection with a diverging number of parameters, J. R. Statist. Soc. B (2009), 71, Part 3, pp. 671-683
- 13) Tibshirani: The Lasso Problem and Uniqueness, Electronic Journal of Statistics (Impact Factor: 0.79). 06/2012; 7. DOI: 10.1214/13-EJS815
- 14) 川口健一著：一般逆行列と構造工学への応用、コロナ社、2003

IMAGING GROUND MOTIONS IN THE TOKYO METROPOLITAN AREA BASED ON MeSO-net USING LASSO

Sadanobu MIZUSAKO, Hiromichi NAGAO, Kei HIROSE, Masayuki KANO and
Muneo HORI

A rapid prediction of damage to structures due to a large earthquake is important to prevent secondary disasters. Ground motion at the base of each construction to be input to a numerical simulation is to be estimated from seismograms. An accurate damage prediction requires such ground motions with spatially-high resolution although the seismometers much sparsely distribute comparing with the required resolution. We have been developing a procedure based on sparse modeling to image the ground motions from seismic array data. Our target is the Tokyo metropolitan area of Japan, in which the seismic array “MeSO-net” (Metropolitan Seismic Observation network) is in operation.

Mizusako[2013, graduation thesis] applied an algorithm based on the Taylor expansion to MeSO-net data when the Earthquake off the Pacific coast of Tohoku occurred on March 11, 2011. This method was found to never account for ground motions in frequencies higher than 0.15 Hz, which was insufficient taking into consideration that the typical eigenfrequency of a construction is usually between 1-10 Hz. Moreover, this method requires a priori assumed truncation order in differential and groups of observatories called “cluster”, in order to determine the unknown partial differential coefficients. Mizusako[2013] suggested that the truncation order was one and each cluster included five nearest observatories.

We propose a new algorithm using lasso (least absolute shrinkage and selection operator, Tibshirani[1996]) in order to obtain an image of spatially-high-resolution ground motions objectively determining the truncation differential order and clusters. The truncation order for a given cluster is automatically determined by lasso owing to the L1-norm regularization, and an appropriate cluster is selected based on an information criterion. The initial result indicates that the proposed method is capable of estimating the ground motions in frequency band 0-1 Hz much better than the previous study.