Thresholding-based iterative selection procedures for model selection and shrinkage

Yiyuan She, *Florida State University*

Abstract

This paper discusses a class of thresholding-based iterative selection procedures (TISP) for model selection and shrinkage. People have long before noticed the weakness of the convex l_1-constraint (or the soft-thresholding) in wavelets and have designed many different forms of nonconvex penalties to increase model sparsity and accuracy. But for a nonorthogonal regression matrix, there is great difficulty in both investigating the performance in theory and solving the problem in computation. TISP provides a simple and efficient way to tackle this so that we successfully borrow the rich results in the orthogonal design to solve the nonconvex penalized regression for a general design matrix. Our starting point is, however, thresholding rules rather than penalty functions. Indeed, there is a universal connection between them. But a drawback of the latter is its non-unique form, and our approach greatly facilitates the computation and the analysis. In fact, we are able to build the convergence theorem and explore theoretical properties of the selection and estimation via TISP nonasymptotically. More importantly, a novel Hybrid-TISP is proposed based on hard-thresholding and ridge-thresholding. It provides a fusion between the l_0-penalty and the l_2-penalty, and adaptively achieves the right balance between shrinkage and selection in statistical modeling. In practice, Hybrid-TISP shows superior performance in test-error and is parsimonious.

AMS 2000 subject classifications: Primary 62J07, 62J05.

Keywords: Sparsity, nonconvex penalties, thresholding, model selection & shrinkage, lasso, ridge, SCAD.

Full Text: PDF

References

[1]   Antoniadis, A. Wavelets in statistics: a review (with discussion). Italian Journal of Statistics 6 (1997), 97–144.

[2]   Antoniadis, A. Wavelet methods in statistics: Some recent developments and their applications. Statistics Surveys 1 (2007), 16–55.

[3]   Antoniadis, A., and Fan, J. Regularization of wavelets approximations. JASA 96 (2001), 939–967. MR1946364

[4]   Browder, F.E., and Petryshyn, W.V. Construction of fixed points of nonlinear mappings in Hilbert space. Journal of Mathematical Analysis and Applications 20, 2 (1967), 197–228. MR0217658

[5]   Bunea, F., Tsybakov, A.B., and Wegkamp, M. Sparsity oracle inequalities for the lasso. Electronic Journal of Statistics 1 (2007), 169–194. MR2312149

[6]   Cai, J., Fan, J., Zhou, H., and Zhou, Y. Hazard models with varying coefficients for multivariate failure time data. Annals of Statistics 35 (2007), 324. MR2332278

[7]   Candès, E. Modern statistical estimation via oracle inequalities. Acta Numerica 15 (2006), 257–325. MR2269743

[8]   Candès, E., Romberg, J., and Tao, T. Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. 59 (2006), 1207–1223. MR2230846

[9]   Candès, E., and Tao, T. The Dantzig selector: statistical estimation when p is much smaller than n. Annals of Statistics 35 (2005), 2392–2404. MR2382651

[10]   Daubechies, I., Defrise, M., and De Mol, C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics 57 (2004), 1413–1457. MR2077704

[11]   Donoho, D., Elad, M., and Temlyakov, V. Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Transactions on Information Theory 52 (2006), 6–18. MR2237332

[12]   Donoho, D., and Johnstone, I. Ideal spatial adaptation via wavelet shrinkages. Biometrika 81 (1994), 425–455. MR1311089

[13]   Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. Least angle regression. Annals of Statistics 32 (2004), 407–499. MR2060166

[14]   Fan, J. Comment on 'Wavelets in Statistics: A Review' by A. Antoniadis. Italian Journal of Statistics 6 (1997), 97–144.

[15]   Fan, J., and Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96 (2001), 1348–1360. MR1946581

[16]   Friedman, J., Hastie, T., Hofling, H., and Tibshirani, R. Pathwise coordinate optimization. Annals of Applied Statistics 1 (2007), 302. MR2415737

[17]   Fu, W. Penalized regressions: the bridge vs the lasso. JCGS 7, 3 (1998), 397–416. MR1646710

[18]   Gannaz, I. Robust estimation and wavelet thresholding in partial linear models. Tech. rep., University Joseph Fourier, Grenoble, France, 2006. MR2409795

[19]   Gao, H.-Y. Wavelet shrinkage denoising using the non-negative garrote. J. Comput. Graph. Statist. 7 (1998), 469–488. MR1665666

[20]   Geman, D., and Reynolds, G. Constrained restoration and the recovery of discontinuities. IEEE PAMI 14, 3 (1992), 367–383.

[21]   Hunter, D.R., and Lange, K. Rejoinder to discussion of 'Optimization transfer using surrogate objective functions'. J. Comput. Graphical Stat 9 (2000), 52–59. MR1819866

[22]   Hunter, D.R., and Li, R. Variable selection using mm algorithms. Annals of Statistics 33 (2005), 1617–1642. MR2166557

[23]   Knight, K., and Fu, W. Asymptotics for lasso-type estimators. Annals of Statistics 28 (2000), 1356–1378. MR1805787

[24]   Meinshausen, N. Relaxed lasso. Computational Statistics and Data Analysis 52, 1 (2007), 374–393. MR2409990

[25]   Meinshausen, N., and Yu, B. Lasso-type recovery of sparse representations for high-dimensional data. Annals of Statistics, 720 (2009), 246–270.

[26]   Osborne, M., Presnell, B., and Turlach, B. On the LASSO and its dual. J. Comput. Graph. Statist. 9, 2 (2000), 319–337. MR1822089

[27]   She, Y. Sparse Regression with Exact Clustering. PhD thesis, Stanford University, 2008.

[28]   She, Y. Thresholding-based iterative selection procedures for model selection and shrinkage. Tech. rep., Statistics Department, Stanford University, June 2008.

[29]   Shimizu, K., Ishizuka, Y., and Bard, J. Nondifferentiable and Two-Level Mathematical Programming. Kluwer Academic Publishers, 1997. MR1429865

[30]   Tibshirani, R. Regression shrinkage and selection via the lasso. JRSSB 58 (1996), 267–288. MR1379242

[31]   Šidák, Z. Rectangular confidence regions for the means of multivariate normal distribution. JASA 62 (1967), 626–633. MR0216666

[32]   Wang, L., Chen, G., and Li, H. Group scad regression analysis for microarray time course gene expression data. Bioinformatics 23, 12 (2007), 1486–1494.

[33]   Wu, T., and Lange, K. Coordinate descent algorithm for lasso penalized regression. Ann. Appl. Stat. 2, 1 (2008), 224–244.

[34]   Yuan, M., and Lin, Y. Model selection and estimation in regression with grouped variables. JRSSB 68 (2006), 49–67. MR2212574

[35]   Zhang, C.-H., and Huang, J. The sparsity and bias of the Lasso selection in high-dimensional linear regression. Ann. Statist 36 (2008), 1567–1594. MR2435448

[36]   Zhang, H.H., Ahn, J., Lin, X., and Park, C. Gene selection using support vector machines with non-convex penalty. Bioinformatics 22, 1 (2006), 88–95.

[37]   Zhao, P., and Yu, B. On model selection consistency of lasso. Journal of Machine Learning Research 7 (2006), 2541–2563. MR2274449

[38]   Zou, H. The adaptive lasso and its oracle properties. JASA 101, 476 (2006), 1418–1429. MR2279469

[39]   Zou, H., and Hastie, T. Regularization and variable selection via the elastic net. JRSSB 67, 2 (2005), 301–320. MR2137327

[40]   Zou, H., and Li, R. One-step sparse estimates in nonconcave penalized likelihood models. Annals of Statistics (2008), 1509–1533. MR2435443