The Journal of Open Source Software

cxreg: An R package for complex-valued lasso and graphical lasso

Younghoon Kim[1*], **Navonil Deb**[1*], **and Sumanta Basu**[1¶]

1 Department of Statistics and Data Science, Cornell University, United States ¶ Corresponding author
* These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: George K. Thiruvathukal

Reviewers:

- @adallak
- @tambstat

Submitted: 25 July 2025
Published: unpublished

Summary

Pathwise coordinate descent (Friedman et al. 2007) is a widely used iterative optimization method for solving penalized linear regression [*lasso*; Tibshirani (1996)] and penalized Gaussian log-likelihood [*graphical lasso*; e.g., Banerjee, El Ghaoui, and d'Aspremont (2008)] problems. For a given sequence of tuning parameters λ in the penalty term, the algorithm updates one coefficient at a time by computing partial residuals.

The coordinate descent algorithm for complex-valued lasso and Gaussian graphical lasso, implemented in the cxreg R package, extends the standard pathwise coordinate descent method to complex-valued data (Deb, Kuceyeski, and Basu 2024). The algorithm operates by *realifying* complex numbers through a ring isomorphism (Herstein 1991). Specifically, for a complex number $z \in \mathbb{C}$, we define the map:

$$\varphi(z) = \begin{pmatrix} \mathrm{Re}(z) & -\mathrm{Im}(z) \\ \mathrm{Im}(z) & \mathrm{Re}(z) \end{pmatrix}.$$

We show that the φ is a field isomorphism between $\mathbb{C}$ and the set $\mathcal{M}^{2\times 2}$ whose entries consists of $a, b \in \mathbb{R}$ to form 2 by 2 matrices,

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Under this mapping, algebraic operations in the complex domain correspond directly to those in the real domain. We extend these operations to the p-dimensional case $(p > 1)$; see Section 3.1 of Deb, Kuceyeski, and Basu (2024) for details.

Using this correspondence, we transform complex-valued linear regression problems into equivalent real-valued formulations. The same idea applies to the Gaussian log-likelihood for complex-valued variables. For the ℓ_1 penalty used in lasso and graphical lasso, we note that the modulus of a complex coefficient equals the entrywise ℓ_2 norm of its real and imaginary parts. Consequently, the lasso for p-dimensional complex variables is equivalent to a $2p$-dimensional group lasso (Yuan and Lin, 2006), where each group contains the real and imaginary components of one complex variable.

We also incorporate standard computational enhancements, including warm starts, using solutions from previous λ values, and active set selection to limit updates to relevant predictors (see Hastie, Tibshirani, and Wainwright, 2015, Chapter 5.4). The R package 'glmnet' adopts a familiar interface similar to glmnet (Friedman, Hastie, and Tibshirani, 2010), and a detailed vignette is available Here.

Statement of need

Solving complex-valued penalized Gaussian likelihood problems (and, as a component, penalized linear regression) is central to high-dimensional time series analysis. In the frequency domain, examining partial spectral coherence is analogous to studying partial correlations in Gaussian graphical models (e.g., Priestley, 1988). With the development of the local Whittle likelihood approximation (Whittle, 1951), the inverse spectral density matrix, known as the spectral precision matrix, provides a way to characterize associations among variables in high-dimensional time series data.

These associations, represented by the spectral precision matrix, are evaluated at fixed frequencies. However, because the spectral density matrix estimator is inconsistent at a single frequency (see Brockwell and Davis, 1991, Chapter 10), it is common to use an averaged, smoothed periodogram, obtained by aggregating periodogram matrices across neighboring frequencies, as a consistent sample analog. Consequently, spectral precision matrices must be computed across multiple neighboring frequencies around each frequency of interest. Such analyses involve repeated high-dimensional optimization, as required for statistical inference on spectral precision matrices (Krampe and Paparoditis, 2025). Therefore, developing efficient algorithms for complex-valued penalized Gaussian likelihood estimation is crucial for scalable and accurate frequency-domain analysis.

State of the field

Although coordinate descent is computationally efficient and widely used to solve the lasso (Friedman et al., 2007) and graphical lasso (Friedman, Hastie, and Tibshirani, 2008), particularly under sparsity assumptions, it has not been extended to complex-valued settings. This gap stems mainly from the lack of methods for updating complex-valued partial residuals within the coordinate descent framework.

The most notable approach to complex-valued graphical modeling is that of Fiecas et al. (2019), which employs the CLIME algorithm (Cai, Liu, and Luo, 2011) based on linear programming. Another alternative is the alternating direction method of multipliers (ADMM), as described by Baek, Düker, and Pipiras (2023). However, neither method explicitly leverages sparsity, despite being applied in high-dimensional settings. Our proposed algorithm, in contrast, is specifically designed to exploit sparsity, achieving substantial gains in both estimation accuracy and computational efficiency. Detailed comparisons with benchmark methods are provided in Deb, Kuceyeski, and Basu (2024).

Acknowledgements

ND and SB acknowledge partial support from NIH award R21NS120227. In addition, SB acknowledges partial support from NSF awards DMS-1812128, DMS-2210675, DMS-2239102 and NIH award R01GM135926.

References

Baek, Changryong, Marie-Christine Düker, and Vladas Pipiras. 2023. "Local Whittle Estimation of High-Dimensional Long-Run Variance and Precision Matrices." *The Annals of Statistics* 51 (6): 2386–2414.

Banerjee, Onureena, Laurent El Ghaoui, and Alexandre d'Aspremont. 2008. "Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data." *The Journal of Machine Learning Research* 9: 485–516.

Brockwell, Peter J, and Richard A Davis. 1991. *Time Series: Theory and Methods*. Springer science & business media.

Cai, Tony, Weidong Liu, and Xi Luo. 2011. "A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation." *Journal of the American Statistical Association* 106 (494): 594–607.

Deb, Navonil, Amy Kuceyeski, and Sumanta Basu. 2024. "Regularized Estimation of Sparse Spectral Precision Matrices." *arXiv Preprint arXiv:2401.11128*.

Fiecas, Mark, Chenlei Leng, Weidong Liu, and Yi Yu. 2019. "Spectral Analysis of High-Dimensional Time Series." *Electronic Journal of Statistics* 13 (2): 4079–4101.

Friedman, Jerome, Trevor Hastie, Holger Höfling, and Robert Tibshirani. 2007. "Pathwise Coordinate Optimization." *The Annals of Applied Statistics*, 302–32.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33: 1–22.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2008. "Sparse Inverse Covariance Estimation with the Graphical Lasso." *Biostatistics* 9 (3): 432–41.

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. "Statistical Learning with Sparsity." *Monographs on Statistics and Applied Probability* 143 (143): 8.

Herstein, Israel N. 1991. *Topics in Agebra*. John Wiley & Sons.

Krampe, Jonas, and Efstathios Paparoditis. 2025. "Frequency Domain Statistical Inference for High-Dimensional Time Series." *Journal of the American Statistical Association*, 1–13.

Priestley, Maurice Bertram. 1988. *The Spectral Analysis of Time Series*. Oxford University Press.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58 (1): 267–88.

Whittle, Peter. 1951. "Hypothesis Testing in Time Series Analysis." *(No Title)*.

Yuan, Ming, and Yi Lin. 2006. "Model Selection and Estimation in Regression with Grouped Variables." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68 (1): 49–67.