Sparse Gaussian Graphical Models with Unknown Block Structure

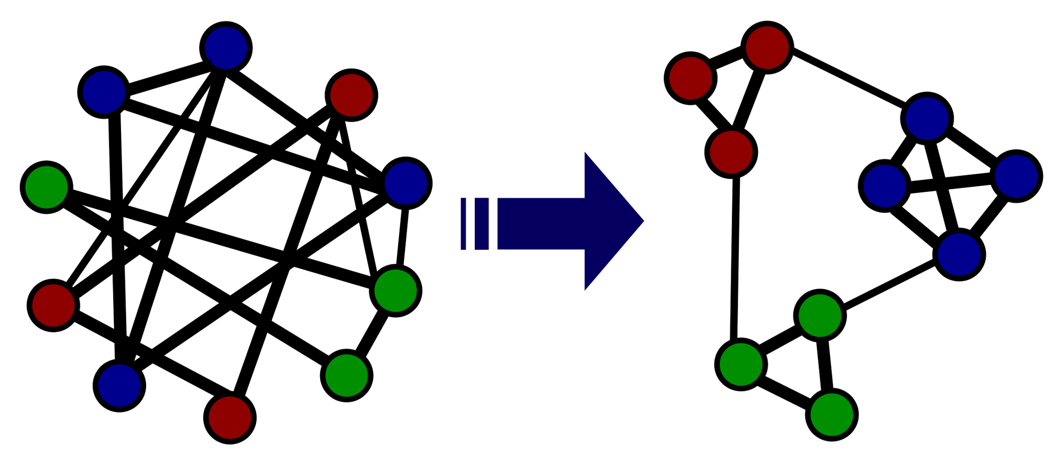
Benjamin M. Marlin and Kevin P. Murphy

Department of Computer Science, University of British Columbia, Vancouver, Canada

1.0 Introduction

Problem: In this work we consider the problem of sparse, block-structured Gaussian precision matrix (inverse covariance matrix) estimation when the blocks are not known *a priori*.

Motivation: Estimating a covariance matrix from high dimensional data using a small number of samples is known to be statistically challenging, and yet it is a problem that arises frequently in practice. For some kinds of data, it is reasonable to assume that the variables can be clustered or grouped into types that share similar connectivity or correlation patterns. For example, genes can be grouped into pathways, and connections within a pathway might be more likely than connections between pathways.



Our interest is in devising methods that simultaneously infer the block structure and a block-sparse precision matrix to provide improved regularization when there is no known block structure.

2.0 Related Work

Tikhonov Regularization: A very simple approach, which we shall call Tikhonov regularization, is to increase the diagonal of the empirical covariance matrix by adding a scalar multiple of the identity matrix.

$$\hat{\Sigma} = S + \nu I$$

L1 Regularized Precision Estimation: Sparse precision matrix estimation can be cast as a convex optimization problem in the penalized maximum likelihood framework. An L1 penalty is imposed on the elements of the precision matrix [Yuan07, Banerjee06].

$$\hat{\Omega} = \operatorname*{argmax}_{\Omega \in S^{++}} \log\det(\Omega) - \mathrm{tr}(S\Omega) - \lambda \sum_{i=1}^{D} \sum_{j \neq i}^{D} |\Omega_{ij}| - \nu \sum_{i=1}^{D} |\Omega_{ii}|$$

Group L1 Regularized Precision Estimation: If the group structure is known, one can extend the L1 penalized likelihood framework in a straightforward way, by penalizing the infinity norm [Duchi08] or the two-norm [Schmidt09] of each block separately. The resulting objective function is still convex, and encourages block-wise sparse graphs.

$$\hat{\Omega} = \arg\max_{\Omega \in S^{++}} \log\det(\Omega) - \mathrm{tr}(S\Omega) - \sum_{kl} \lambda_{kl} ||\{\Omega_{ij} : i \in G_k, j \in G_l\}||_{p_{kl}}$$

Sparse Dependency Networks: An alternative approach to sparse precision estimation is to learn the underlying graph by regressing each node on all the others using an L1 penalty [Meinshausen06].

$$\hat{\mathbf{w}}_j = \arg\max_{\mathbf{w}} \sum_{n=1}^{N} \log p(x_{nj} | x_{n,-j}, \mathbf{w}, \sigma_j^2) + \lambda \sum_{i \neq j} |w_i|$$

3.0 Model & Algorithm

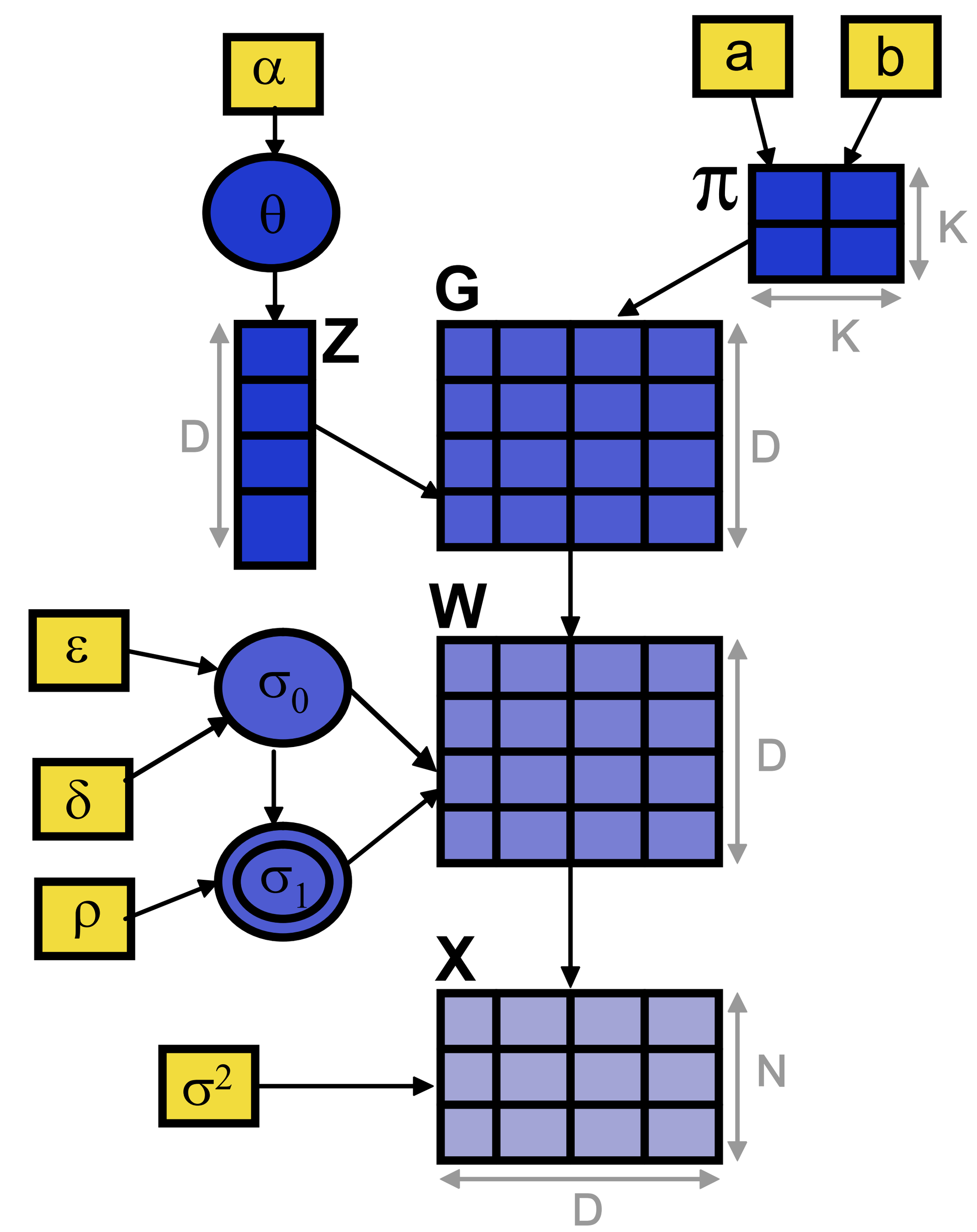
Overview: We propose a two-stage method for learning sparse precision matrices or GGMs with unknown block structure:

(1) Optimize a pseudolikelihood criterion combined with a sparsity promoting prior on the weights, similar to the approach of Meinshausen & Buhlmann. The sparsity level of each edge (i,j) is controlled by the clusters to which nodes i and j belong, as well as the probability of an edge between these cluster types.

(2) Having identified the clusters, we then estimate the precision matrix using the group L1/L2 method [Schmidt09].

The Model:

- $\theta \sim$ Dir$(\frac{\alpha}{K})$
- $z_d \sim$ Multi$(\theta, 1)$
- $\pi_{k,k'} \sim$ Beta$(a_{k,k'}, b_{k,k'})$
- $G_{d,d'} \sim$ Ber$(\pi_{z_d, z_{d'}})$
- $\sigma_0^2 \sim$ Ga(ϵ, δ) and $\sigma_1^2 = \rho\sigma_0^2$.
- $w_{d,d'} \sim \mathcal{N}(0, \sigma_0^2)^{G_{d,d'}} \mathcal{N}(0, \sigma_1^2)^{1-G_{d,d'}}$



Variational Approximation: $Q(Z, \theta, \pi, G, W, \sigma_0) = Q(Z)Q(\theta)Q(\pi)Q(G)Q(W)Q(\sigma_0)$

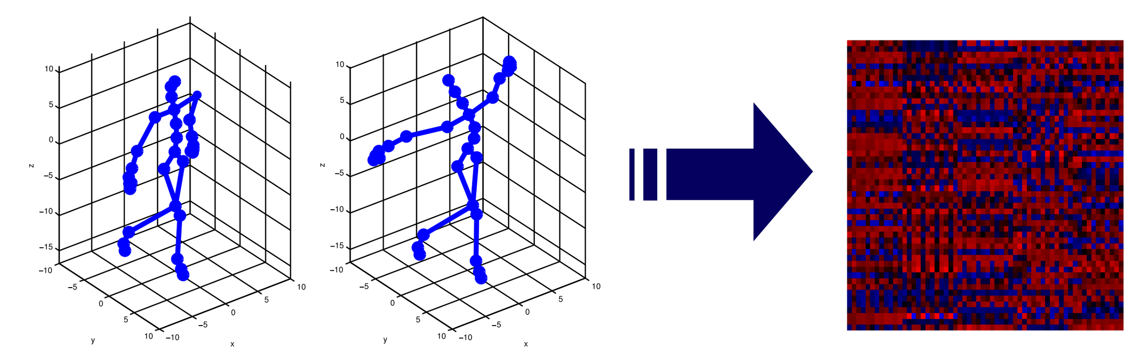
$$Q(Z_d) = \mathrm{Multi}(\phi_d, 1) \qquad Q(G_{d,d'}) = \mathrm{Ber}(\gamma_{d,d'})$$
$$Q(\theta) = \mathrm{Dir}(\alpha^*) \qquad Q(1/\sigma_0^2) = \mathrm{Ga}(\epsilon^*, \delta^*)$$
$$Q(\pi_{k,k'}) = \mathrm{Beta}(a^*_{k,k'}, b^*_{k,k'}) \qquad Q(W) = \delta(W - \hat{w})$$

Learning Algorithm:

[1] Initialize with all nodes in the same cluster
[2] While not converged
[3] Iterate variational updates to convergence
[4] Use graph cut on W to propose cluster splits
[5] Accept first split that increases objective function
[6] Extract MAP clustering. Apply group L1/L2 to estimate precision

4.0 Experiments & Results

CMU Motion Capture Data Set (D= 60):



Gasch et al. Gene Expression Data Set: (N=174, D=667)

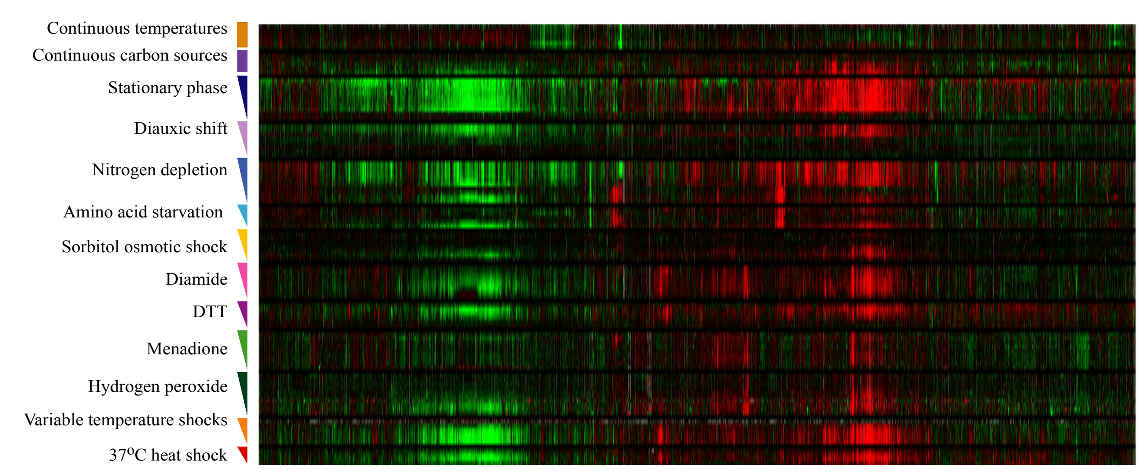
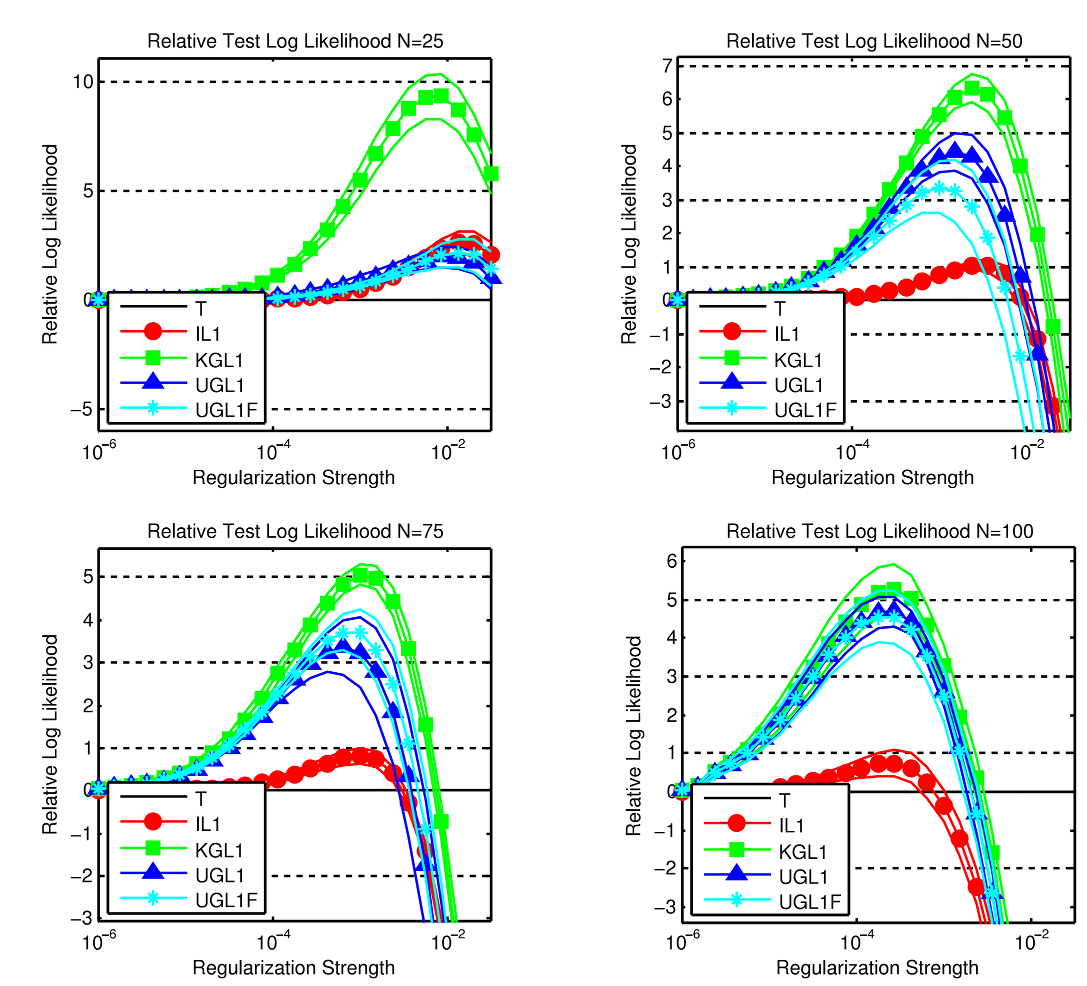


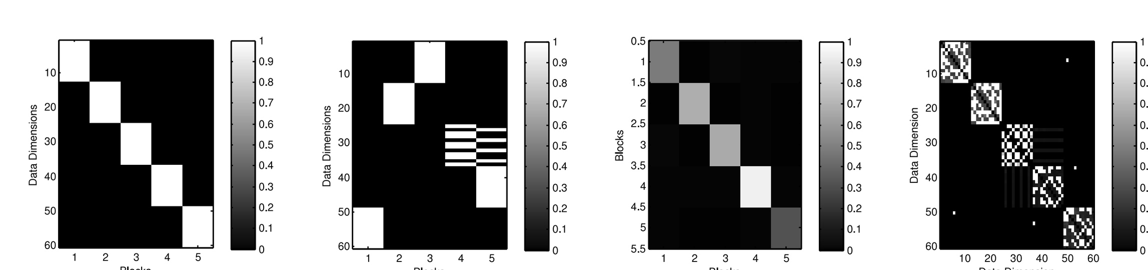
Image adapted from Gasch et al. (2000) supplemental materials.

4.1 CMU Results

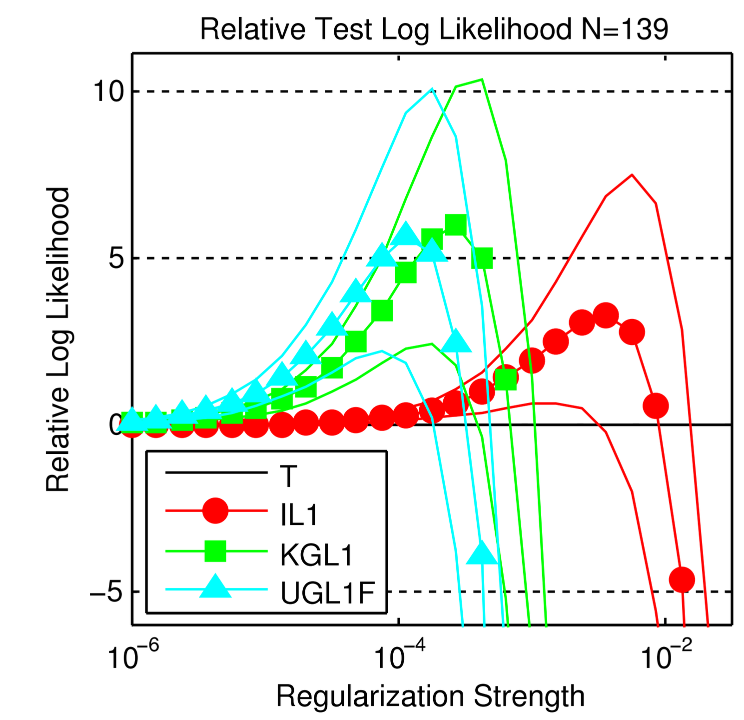
Test Log Likelihood vs Regularization Strength



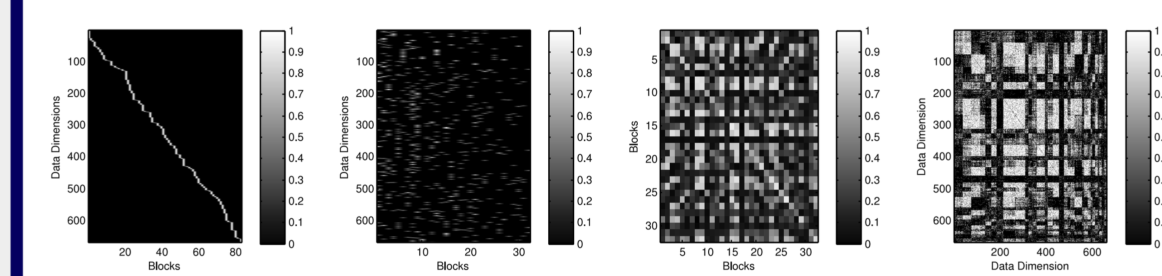
Inferred Structure for N=50



4.2 Genes Results



Inferred Structure



- **T:** Tikhonov regularization
- **IL1:** Independent L1 penalized likelihood
- **UGL1:** Our dependency network model (unknown groups)
- **UGL1F:** Our dependency network model (unknown groups, fast updates)
- **KGL1:** Group L1/L2 penalized likelihood (known groups)

5.0 Conclusions

Conclusions: We have shown how to estimate a block-sparse precision matrix while simultaneously estimating the blocks. Our two stage approach uses a hierarchical dependency network model to infer the blocks, and fast convex methods to estimate the precision matrix given the blocks.

Future Work: In work appearing at UAI 2009, we present an alternative approach that avoids the dependency network by converting the L1 and group L1/L2 regularization functions into priors on the space of positive definite matrices. We deal with the intractable normalization constants using novel lower bounds [Marlin09].

References

Banerjee, O., Ghaoui, L. E., & d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. J. of Machine Learning Research, 9, 485–516.

Dempster, A. (1972). Covariance selection. Biometrics, 28, 157-175.

Dobra, D., Hans, C., Jones, B., Nevins, J., Yao, G., & West, M. (2004). Sparse graphical models for exploring gene expression data. J. Multivariate analysis, 90, 196-212.

Duchi, J., Gould, S., & Koller, D. (2008). Projected subgradient methods for learning sparse gaussians. Proc. of the Conf. on Uncertainty in AI.

Friedman, J., Hastie, T., & Tibshirani, R. (2007). Sparse inverse covariance estimation the graphical lasso. Biostatistics, 432.441.

Ghahramani, Z., & Beal, M. (2000). Propagation algorithms for variational Bayesian learning. Advances in Neural Info. Proc. Systems (pp. 507.513).

Heckerman, D., Chickering, D., Meek, C., Rounthwaite, R., & Kadie, C. (2000). Dependency networks for density estimation, collaborative filtering, and data visualization. JMLR.

A. P. Gasch, P. T. Spellman, C. M. Kao, O. C. Harel, M. B. Eisen, G. Storz, D. Botstein, P. O Brown. (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Molecular Biology of the Cell, 11. 4241-4257.

Marlin, B., Schmidt, M., and Murphy, K. (2009) Group Sparse Priors for Covariance Estimation. Proc. of the Conf. on Uncertainty in AI.

Meinshausen, N., & Buhlmann, P. (2006). High dimensional graphs and variable selection with the lasso. The Annals of Statistics, 34, 1436-1462.

Schmidt, M., van den Berg, E., Friedlander, M., & Murphy, K. (2009). Optimizing Costly Functions with Simple Constraints: A Limited-Memory Projected Quasi-Newton Algorithm. AI & Statistics.

Yuan, M., & Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. Biometrika, 94, 19.35.