Inferring the Underlying Structure of Spin Glass Model

Jialei Wu, Stephen Zhang, Yong See Foo, Michael Stumpf

The University of Melbourne, School of Mathematics and Statistics

Introduction

The Ising Model arises from the field of Statistical Mechanics. It has been applied to study gene regulatory networks by representing genes as binary variables (on/off states) and modeling the interactions between them. A spin glass model consists of N binary spins s_i, $i = 1, 2, ..., N$ with $s_i = \pm 1$, These spins are coupled by a coupling matrix J_{ij} and are subject to an external magnetic field H_i, which is the bias term in our application. The energy of such configuration is given by the Hamiltonian:

$$H(\vec{s}) = -\sum_{i,j} s_i J_{ij} s_j - \sum_i H_i s_i$$

The configuration probability is given by $\frac{e^{-\beta H}}{Z}$ where $\beta > 0$ is the inverse temperature and Z is the normalising constant.

Spin Glass Model Inference

Given sample of size N, $\vec{s}^{\mu}$, $\mu = 1, 2, ..., N$ from a spin glass model with unknown parameters, to compute the coupling matrix and the bias term, we use the maximum likelihood approach:

$$\begin{aligned} L(\vec{J}, \vec{H}) &= \sum_{i,j} J_{ij} \frac{1}{N} \sum_{\mu} s_i^{\mu} s_j^{\mu} + \sum_i H_i \frac{1}{N} \sum_{\mu} s_i^{\mu} - \log Z(\vec{J}, \vec{H}) \\ &= \sum_{i,j} J_{ij} \langle \sigma_i \sigma_j \rangle^D + \sum_i H_i \langle \sigma_i \rangle^D - \log Z(\vec{J}, \vec{H}) \end{aligned}$$

Differentiated with respect to J and H we have:

$$\frac{\partial L}{\partial J_{ij}} = \langle \sigma_i \sigma_j \rangle^D - \langle \sigma_i \sigma_j \rangle \qquad \frac{\partial L}{\partial H_i} = \langle \sigma_i \rangle^D - \langle \sigma_i \rangle$$

The $\langle \sigma \rangle^D$ is the sample average and the $\langle \sigma \rangle$ is the expectation of the model fitted. At the maximum of the log-likelihood these derivatives are zeros. Two algorithms **Gradient Descent** and **Proximal Gradient Descent** are then applied to search for the optimal parameters

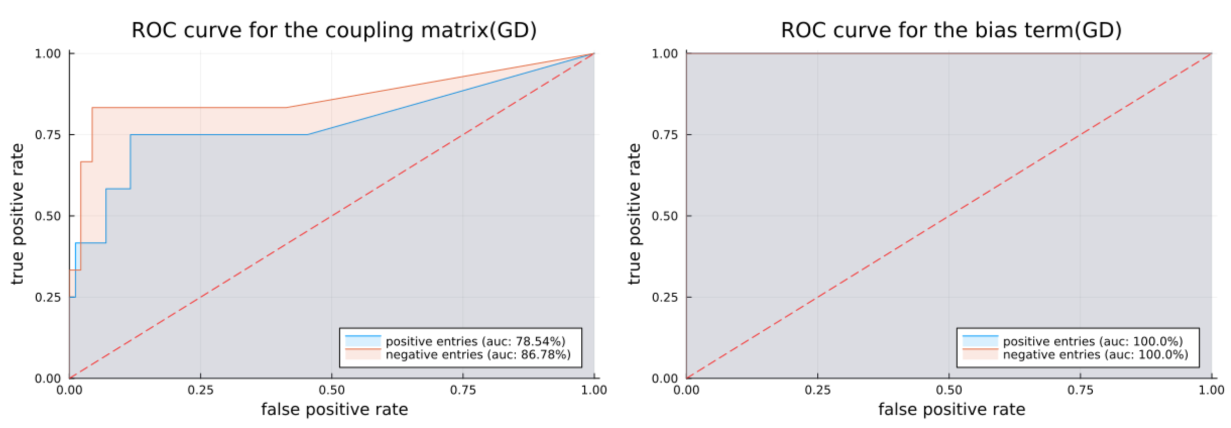
Gradient Descent(GD)

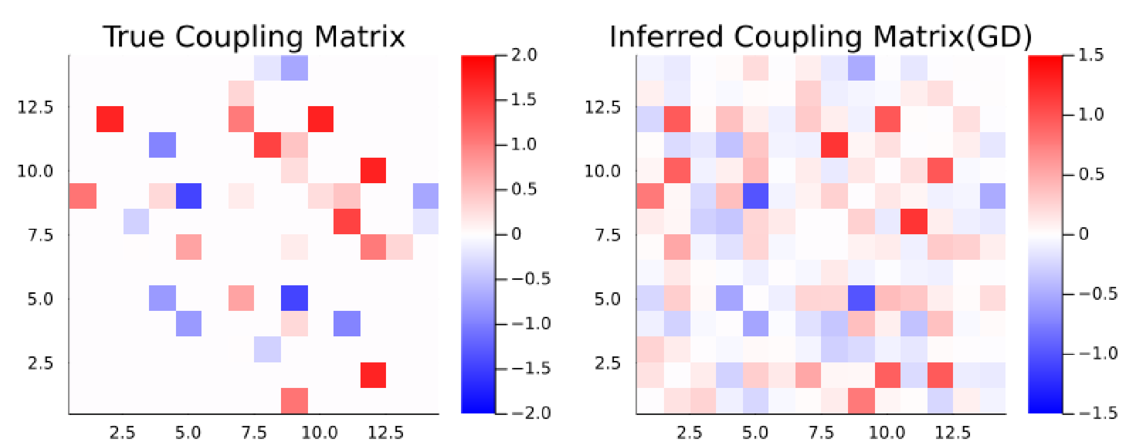
The log-likelihood is a concave function, therefore, we use a gradient descent approach. With learning rate λ, at each step, the coupling matrix and the bias are updated according to:

$$J_{ij}^{t+1} = J_{ij}^t + \lambda \frac{\partial L}{\partial J_{ij}}(\vec{J}^n, \vec{H}^n) \qquad H_i^{t+1} = H_i^t + \lambda \frac{\partial L}{\partial H_i}(\vec{J}^n, \vec{H}^n)$$

Given our focus on discerning positive and negative interactions among genes, our model's efficacy is assessed within a categorical spectrum.

The Area Under the Curve(AUC) of the Receiver Operating Characteristic(ROC) is used to evaluate model performance, with values between 1 and 0.5, where a larger AUC indicates better performance.





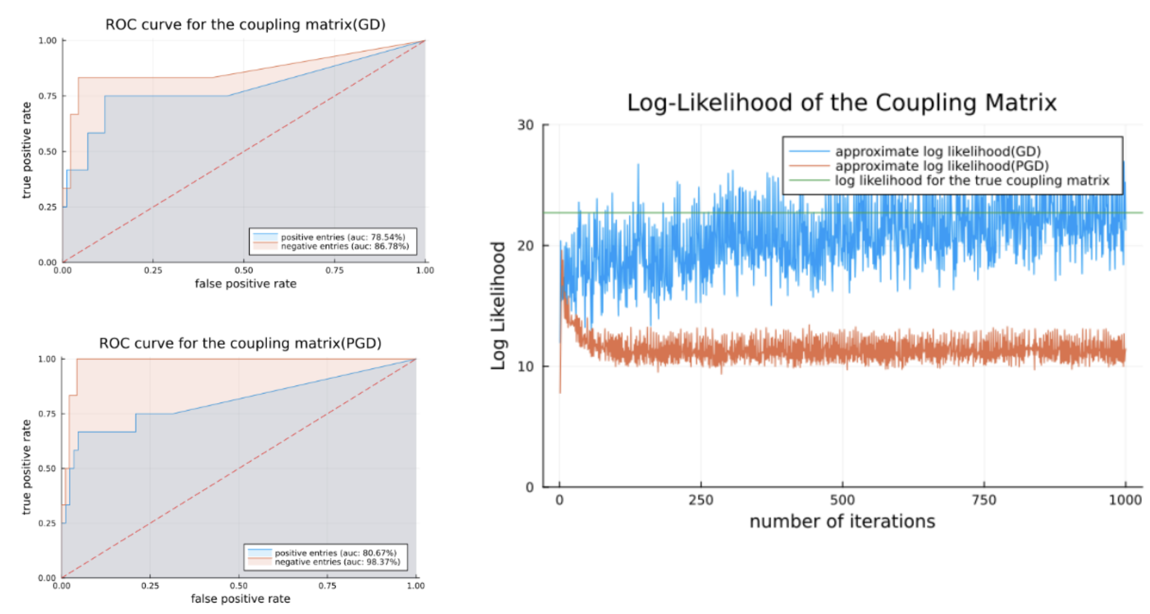
Proximal Gradient Descent(PGD)

Proximal Gradient Descent with Lasso Regularization promote sparsity in the inferred coupling matrix and improve model accuracy under the sparsity assumption.

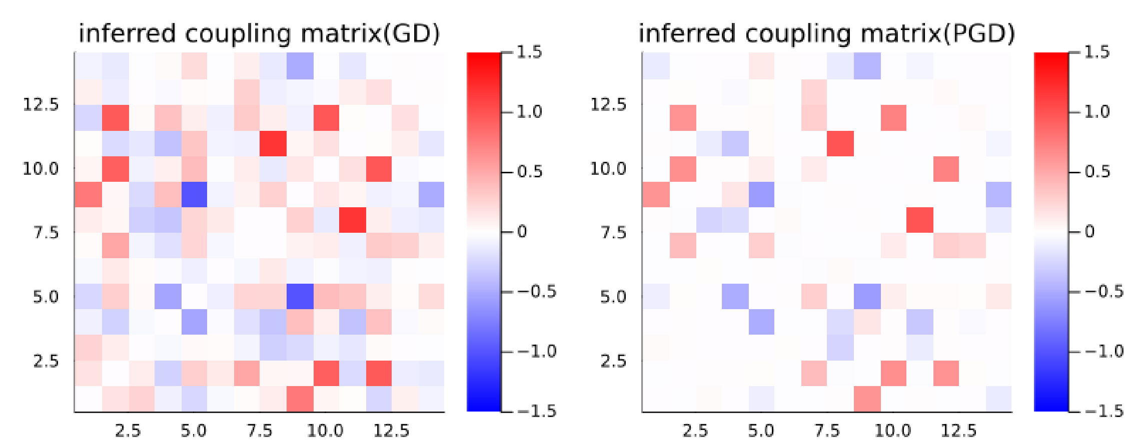
$$J_{ij}^{t+1} = \mathbf{prox}_{\beta_1 h}\left(J_{ij}^t + \lambda \frac{\partial L}{\partial J_{ij}}(\vec{J}^n, \vec{H}^n)\right) \qquad H_i^{t+1} = \mathbf{prox}_{\beta_2 h}\left(H_i^t + \lambda \frac{\partial L}{\partial H_i}(\vec{J}^n, \vec{H}^n)\right)$$

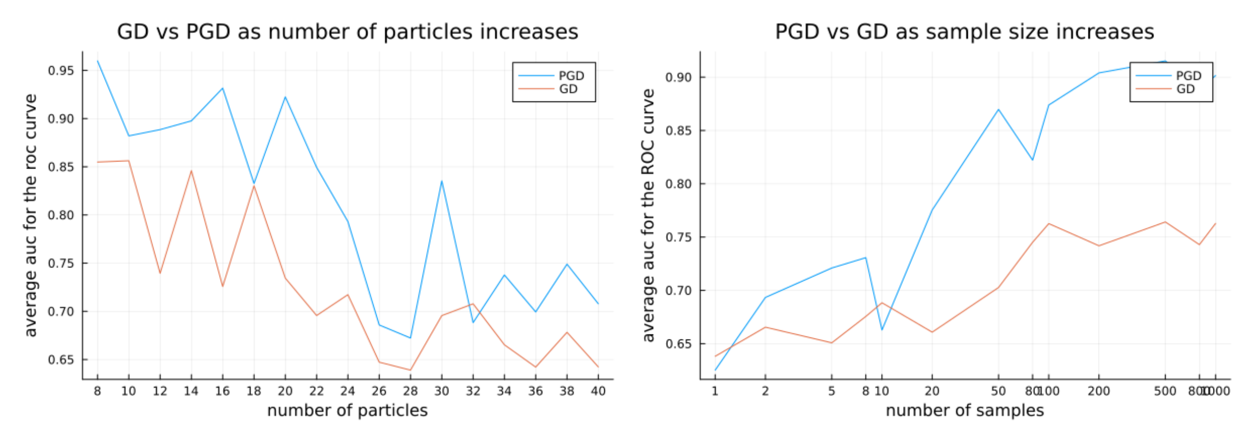
The proximal function is defined as

$$\mathbf{prox}_{\beta h}(u) = \begin{cases} u - \beta & u > \beta \\ 0 & -\beta < u \leq \beta \\ u + \beta & u < -\beta \end{cases}$$



PGD prevents the model from overfitting to the noise of the data and help recover the true underlying coupling between spins.





References

[1] H.C.Nguyen, R.Zecchina, J.Berg. Inverse Statistical Problems: From the Inverse Ising problem to Data Science(2017)

[2] N.Parikh, S.Boyd. **148** Proximal Algorithms(2013)