Improving Inference via Perturbations

by

Ying-Ying (Brittany) Lu

Undergraduate Thesis Supervised by
Dr. David Sivak

Department of Physics
Faculty of Science

SIMON FRASER UNIVERSITY Spring 2019

Template modified from the sfuthesis.cls class.

Abstract

Cellular networks in biological systems are complex and as such, identifying the molecular interactions that give rise to the complex behavior observed can require an immense amount of data. Often, statistical and machine learning techniques are used to analyze this data and extract a global picture of network dynamics. One of the challenges of network analysis in systems biology is finding the connections between genes, proteins, or both, and predicting additional ones that have not yet been detected experimentally.

This problem is easily mappable to the inverse problem of statistical physics: inferring the microscopic particle-particle interactions given macroscopic observations of a system. In particular, the focus of this work is to investigate whether perturbations can be introduced into the system so as to improve the output data quality. Specifically, we explore how perturbations in the form of magnetic field can be used to improve the inference of interactions for a three-spin Ising system. Utilizing a maximum likelihood approach, we empirically show that there exists an optimal field where learning is most efficient. Such a field seems to counteract the individual interactions between spins, allowing for optimal inference.

Keywords: Ising Model, Statistical Inference, Fisher Information

Acknowledgements

First and most importantly, I would like to thank Dr. Sivak for giving me the opportunity to work under his supervision. I am extremely grateful for his help and insight. I would also like to thank Joseph Lucero and Suemin Lee for helping me get started on Compute Canada. Thanks again to Joseph for the suggestions he has given me throughout the past year.

Table of Contents

List of Figures

Chapter 1

Introduction

Stem cells have the potential to develop into different types of cells. Under certain conditions, they can be induced to become tissue or organ-specific cells with specialized functions [1]. In 2006, researchers at Kyoto University demonstrated that mature, specialized cells could be reprogrammed into stem cell-like states, by introducing genetic perturbations to the cell cultures [2]. The ability to reprogram cells gives potential to treat illnesses such as diabetes, cancer and heart diseases. However, to successfully design a stem cell requires an understanding of the regulatory networks that interact to control the level of gene expression. Today, there exists high-throughput techniques such as Drop-seq to simultaneously analyze mRNA expression in thousands of cells [3]. Nevertheless, cellular networks are complex, and identifying molecular interactions that causes this complex behavior, either through experimental or computational methods, may require extensive quantitative information, which is not always available [4].

Statistical physics seeks to study the behavior and interactions of microscopic particles in a macroscopic system. In the inverse problem, given a set of observations of a system whose microscopic parameters are unknown to us, we want to infer the parameters describing the system. In this work, we aim to understand how a biological network can be perturbed so as to better infer the interactions within the network. We use a modeling framework from statistical physics, and represent a biological network as a Ising spin system, which will be discussed next. The ultimate goal is to efficiently reconstruct the network of interactions within biological systems by improving data quality.

1.1 Ising model

The Ising model was originally introduced to study magnetic materials, but its application was quickly extended as a model for studying interacting, many-body systems [5]. In the Ising model, each microscopic degree of freedom in a system is represented by a spin σ_i. In our model, spins assume discrete binary values. Each spin may point either "up" ($\sigma_i = 1$) or "down" ($\sigma_i = -1$) relative to a local magnetic field h_i that pushes a spin σ_i to either

point up or down. The interaction strength between spins σ_i and σ_j is represented by the pairwise coupling J_{ij}. The interactions within a network can be represented by a matrix $\mathbf{J}$. When $J_{ij} < 0$, the spins prefer to be anti-aligned, and when $J_{ij} > 0$, the spins prefer to be aligned. The case where $J_{ij} = 0$ means that the spins do not interact. We impose the condition that there are no self interactions ($J_{ii} = 0$).

Since each spin can be in one of two states, a spin network with m spins can be in one of 2^m possible configurations. The Hamiltonian that specifies the energy of the spin system having particular configurations $\{s_k\},\ k\ \epsilon\ 1,\ ..,\ 2^m$, and coupling parameters $\mathbf{J}$ is

$$\mathcal{H}(s_k|\mathbf{J}) = -\sum_i h_i\sigma_i - \sum_{i,j} J_{i,j}\sigma_i\sigma_j. \tag{1.1}$$

For an Ising system in equilibrium, probability distribution is

$$p(s_k|\mathbf{J}) = \frac{1}{\mathcal{Z}} \exp\left[\frac{1}{T}\left(\sum_{i,j} J_{i,j}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)\right]. \tag{1.2}$$

where

$$\mathcal{Z} = \sum_i e^{-\beta\mathcal{H}(s_k|\mathbf{J})}, \tag{1.3}$$

is the partition function (sums over all configurations), and $\beta = (k_{\mathrm{B}}T)^{-1}$. Equation 1.2 gives the probability of observing state s_k in a spin system with interactions described by the matrix $\mathbf{J}$. We may view the temperature T as a temperature that controls the thermal fluctuations [6]. Throughout, we set T and Boltzmann's constant to unity, $k_{\mathrm{B}}T = 1$.

For a non-equilibrium system, there does not exist a Boltzmann distribution with a known Hamiltonian to describe the steady state [5].

In the context of reconstructing genetic networks, each spin can be thought of as the expression of a particular gene. Gene expression can be on or off, similar to how spins can point up or down. Expression is controlled by transcription factors, proteins that bind near the regulatory region of a gene. A gene targeted by a transcription factor may encode yet another transcription factor, leading to a chain of regulatory interactions [5].

The inverse Ising problem is the determination of interactions within the spin network. We allow the network to fluctuate between configurations, and from these observations we estimate the coupling parameters $\mathbf{J}$. However, we would like to understand how inference can be improved. Perturbing a biological system, such as by knocking out, silencing, or over-expressing a gene, offers the possibility of improving data quality. Here, our perturbation is represented by the field h_i. Fields that align or anti-align two spins reinforce either the positive or negative interactions that exist between the spins, analogous to silencing and over-expressing a gene.

Chapter 2

Information and motivation

2.1 Maximum likelihood

The Boltzmann distribution for our Ising spin system given by Eq. 1.2 describes the probability distribution of the states s_k for a given coupling matrix $\mathbf{J}$. When the coupling matrix $\mathbf{J'}$ is inferred from the observation data $\{x_{2^m}\} = x_{1,2}, \ldots, x_n$ (where m is the number of spins in the system), the likelihood function $\mathcal{L}(\mathbf{J'}|\{s_k\})$ describes the probability of observing the given data $\{s_k\}$ as a function of the unknown parameters $\mathbf{J'}$ [7]:

$$\mathcal{L}(\mathbf{J'}|\{s_k\}) = p(s_1|\mathbf{J'})\ p(s_2|\mathbf{J'}) \ldots p(s_{2^m}|\mathbf{J'}) \tag{2.1a}$$

$$= \prod_{k=1}^{2^m} p(s_k|\mathbf{J'}). \tag{2.1b}$$

Suppose we make N observations such that each of the states $\{s_k\} = s_1, s_2, \ldots, s_n$ occurs $\{x_k\}$ times, where $\{x_k\} = x_1, x_2, \ldots, x_n$, the x_k's are *not* independent. The multinomial likelihood function is then,

$$\mathcal{L}(\mathbf{J'}|s_1, \ldots, s_{2^m}) = \frac{N!}{x_1!\ x_2! \ldots x_{2^m}!}\ p(s_1|\mathbf{J'})^{x_1}\ p(s_2|\mathbf{J'})^{x_2} \ldots p(s_n|\mathbf{J'})^{x_{2^m}}, \tag{2.2}$$

Values of $\mathbf{J'}$ that are plausible should have a relatively high likelihood. The maximum likelihood estimate $\hat{\mathbf{J}}$ of a parameter $\mathbf{J}$ is the value that maximizes the likelihood.

Maximizing the likelihood function can be quite tedious. Since the logarithm is an increasing function, maximizing the likelihood is equivalent to maximizing the log-likelihood

function. Taking the logarithm on both sides of Eq. 2.2, we get,

$$\log \mathcal{L} = \log N! - \log x_1! - \ldots - \log x_{2^m}! + x_1 \log p(s_1|\mathbf{J'}) + \ldots + x_{2^m} \log p(s_{2^m}|\mathbf{J'}) \tag{2.3a}$$

$$= C + \sum_k^{2^m} x_k \log p(s_k|\mathbf{J'}) \tag{2.3b}$$

$$= C - \beta \sum_k^{2^m} x_k \mathcal{H}(s_k|\mathbf{J'}) + \sum_k^{2^m} x_k \log Z \ , \tag{2.3c}$$

where $C = \log N! - \log x_1! - \ldots - \log x_{2^m}!$, and we have used the definition of $p(s_k|\mathbf{J})$ in Eq. 1.2.

To illustrate the idea of maximizing likelihood, shown in Fig. 2.1a is a spin system with the following coupling matrix,

$$\mathbf{J} = \begin{bmatrix} J_{11} & J_{12} & J_{13} & J_{14} & J_{15} & J_{16} \\ J_{21} & J_{22} & J_{23} & J_{24} & J_{25} & J_{26} \\ J_{31} & J_{32} & J_{33} & J_{34} & J_{35} & J_{36} \\ J_{41} & J_{42} & J_{43} & J_{44} & J_{45} & J_{46} \\ J_{51} & J_{52} & J_{53} & J_{54} & J_{55} & J_{56} \\ J_{61} & J_{62} & J_{63} & J_{64} & J_{65} & J_{66} \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & -1 & 1 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} . \tag{2.4}$$

We impose the conditions that the matrix is symmetric, $J_{ij} = J_{ji}$, and that there is no self interaction, $J_{ii} = 0$. The coupling matrix was specifically chosen to show two examples, one where J_{mn} is easy to learn, and the other where J_{mn} is hard to learn.

Suppose we do not know the value of J_{mn}, the interaction between spins σ_m and σ_n, but we know all other coupling parameters $J_{ij},\ i, j \neq m, n$. Ignoring the term $C = \frac{N!}{x_1! x_2! \ldots x_n!}$ in Eq. 2.3a that is independent of $\mathbf{J}$, we get

$$\log \mathcal{L}(J'_{mn}|\{s_k\}) = \sum_{k=1}^{n} x_k \cdot \log P(s_k|J'_{mn}) \ . \tag{2.5}$$

Here, $P(s_k|J'_{mn})$ is the probability of observing the system in state s_k as a function of J'_{mn}.

Rather than observing the system N times and then tallying the number of occurrences for each of the states in $\{s_k\}$, we will assume that x_k in Eq. 2.2 is proportional to the true probability distribution of the system with coupling matrix given by Eq. 2.4,

$$x_k \propto P(s_k|\mathbf{J}), \tag{2.6}$$

where $P(s_k|\mathbf{J})$ is the actual probability of the system found in state s_k.

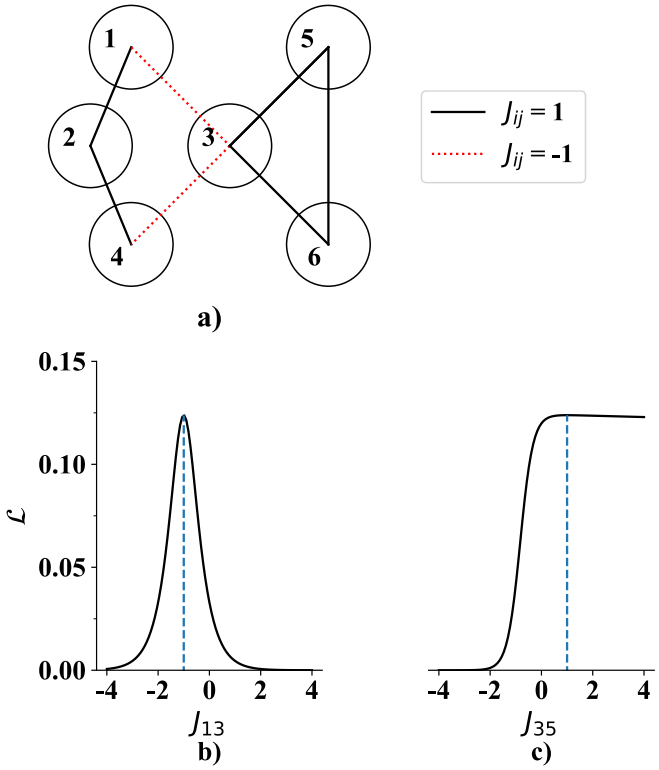


Figure 2.1: Example likelihood function of an arbitrary six-spin network. a) A six-spin network with positive (solid black line) and negative (dotted red line) unit interactions. b) Likelihood as a function of coupling coefficient J_{13}. The function peaks at $J_{13} = -1$. c) Likelihood as a function of coupling coefficient J_{35}. The function levels off at $J_{35} \approx 0$. In (b) and (c), the dotted blue line indicates the value of the true coupling coefficient, corresponding to the diagram above.

Doing so, we normalize the likelihood function by fixing $\sum_k x_k = 1$. It follows that

$$\log \mathcal{L}(J'_{mn}|\{s_k\}) \propto \sum_{k=1}^{n} P(s_k|\mathbf{J}) \cdot \log P(s_k|J'_{mn}). \tag{2.7}$$

The overall shape of the actual value of the function will be the same as what is plotted. We choose not to sample because each time we do so, the distribution for the count changes, which affects our estimate. In Fig. 2.1b, all of the values of J_{ij} are fixed, except for J_{13}. The function peaks at $J_{13} = -1$, meaning that this is the value that best fits with the observations $\{x_k\}$. Indeed, the actual value is $J_{13} = -1$. In contrast, when all values but J_{35} are fixed, the function in Fig. 2.1c does not peak at the actual value of $J_{35} = 1$; instead, the function begins to level off at around $J_{35} = 0$. Given this data, it is not possible to say which of the values where $J_{35} > 0$ is the most plausible one, as they are all equally likely to describe that parameter of the system. Moreover, in real systems, we likely do not have

prior knowledge of any of the parameters. Hence, the likelihood function depends on more than one parameter, and may peak at different combinations of $\{J_{ij}\}$, leading to a difficult inference problem.

2.2 Kullback-Leibler divergence

In the inverse problem, we aim to infer a coupling matrix $\hat{\mathbf{J}}$ giving a distribution $p(s_k|\hat{\mathbf{J}})$ that is as close as possible to the true distribution $p(s_k|\mathbf{J})$ of our system. To assess the quality of our inference, we require a way of measuring the dissimilarity between two different distributions. The "distance" between the two distributions is given by the Kullback-Leibler divergence (relative entropy) D_{KL}, [8],

$$D_{\mathrm{KL}}\left(p(\{s_k\}|\mathbf{J}) \;||\; p(\{s_k\}|\hat{\mathbf{J}})\right) = \sum_k p(\{s_k\}|\mathbf{J}) \;\log\frac{p(\{s_k\}|\mathbf{J})}{p(\{s_k\}|\hat{\mathbf{J}})}. \tag{2.8}$$

Note that the Kullback-Leibler divergence is not a distance measure in the usual sense as is it not generally symmetric, i.e. $D_{\mathrm{KL}}(p||q) \neq D_{\mathrm{KL}}(q||p)$.

Expanding Eq. 2.8 gives

$$\sum_k^n p(s_k|\mathbf{J}) \;\log\frac{p(s_k|\mathbf{J})}{p(s_k|\hat{\mathbf{J}})} = \sum_k^n p(s_k|\mathbf{J})\cdot\log p(s_k|\mathbf{J}) - \sum_k^n p(s_k|\mathbf{J})\cdot\log p(s_k|\hat{\mathbf{J}}). \tag{2.9}$$

When two distributions are identical, $\hat{\mathbf{J}} = \mathbf{J}$, and it follows that $D_{\mathrm{KL}} = 0$. Thus, we look for a matrix $\hat{\mathbf{J}}$ such that the Kullback-Leibler divergence is minimized.

The two summation terms appear to be log-likelihood functions. In fact, we can show that maximizing the likelihood function is equivalent to minimizing the Kullback-Leibler divergence. Since we want D_{KL} to be as small as possible, this amounts to maximizing the second term (the first term is independent of $\hat{\mathbf{J}}$), which is also the expectation value of $\log p(\{s_k\}|\hat{\mathbf{J}})$,

$$\mathrm{argmax}\sum_k^n p(s_k|\mathbf{J})\cdot\log p(s_k|\hat{\mathbf{J}}) = \mathrm{argmax}\;\mathbb{E}\left[\log p(\{s_k\}|\hat{\mathbf{J}})\right]. \tag{2.10}$$

If n is very large, then $\mathbb{E}\left[\log p(\{s_k\}|\hat{\mathbf{J}})\right] \approx \frac{1}{n}\sum_k^n \log p(s_k|\hat{\mathbf{J}})$. Therefore,

$$\begin{aligned}\mathrm{argmax}\;\mathbb{E}\left[\log p(\{s_k\}|\hat{\mathbf{J}})\right] &= \mathrm{argmax}\;\frac{1}{n}\sum_k^n \log p(\{s_k\}|\hat{\mathbf{J}}) \\ &= \mathrm{argmax}\;\sum_k^n \log p(\{s_k\}|\hat{\mathbf{J}}) \\ &= \mathrm{argmax}\;\log\mathcal{L}(\mathbf{J}|\{s_k\})\;.\end{aligned} \tag{2.11}$$

Thus minimizing the Kullback-Leibler divergence is equivalent to maximizing the likelihood function as a function of the parameter $\hat{\mathbf{J}}$.

2.3 Fisher information

The sharpness of the peak in the likelihood function provides information on the goodness of the estimate. We want to quantify the amount of information that an observed variable s_k carries about an unknown parameter $\mathbf{J}$ with the Fisher information $\mathcal{I}$ [8]:

$$\mathcal{I}(\mathbf{J}) = \mathbb{E}\left[\frac{\partial}{\partial J_{ij}} \log \mathcal{L}(\mathbf{J}|s_k)\right]^2. \tag{2.12}$$

The derivative $\frac{\partial}{\partial J_{ij}} \log \mathcal{L}(\mathbf{J}|\{s_k\})$ is known as the score function, describing how sensitive the likelihood is to changes in $\mathbf{J}$ at a particular value of $\mathbf{J}$. The Fisher information is also the variance of $\frac{\partial \log \mathcal{L}}{\partial J_{ij}}$,

$$\mathcal{I}(\mathbf{J}) = \mathrm{Var}\left[\frac{\partial}{\partial J_{ij}} \log \mathcal{L}(\mathbf{J}|s_k)\right]. \tag{2.13}$$

If the likelihood function is twice differentiable as a function of J_{ij}, and the integral $\int \mathcal{L}\, dx$ can be differentiated twice under the integral sign as a function of J_{ij}, then it can be shown that the (ij, mn) element of the Fisher information for a system in a particular state s_k is [9]

$$\mathcal{I}_{ij,mn}(\mathbf{J}) = -\mathbb{E}\left[\frac{\partial^2 \log \mathcal{L}(\mathbf{J}|s_k)}{\partial J_{ij} \partial J_{mn}}\right]. \tag{2.14}$$

Viewed this way, the Fisher information can be interpreted as the curvature of the likelihood function at a particular value of $\mathbf{J}$. The more negative the Fisher information, the sharper the curve, and the more information s_k contains about the estimate of $\mathbf{J}$. The less negative the Fisher information is, the shallower the curve, and the less information s_k contains about the estimate of $\mathbf{J}$. The (ij, mn) entry gives the mixed partial derivative of the log-likelihood with respect to the coupling coefficient J_{ij} between spins i and j and the coupling coefficient J_{mn} between spins m and n [10].

To calculate an analytical form of the (ij, mn) element of the Fisher information at the true parameter value $\mathbf{J}$, $\mathcal{I}_{ij,mn}$, recall Eq. 2.3a:

$$\frac{\partial}{\partial J_{ij}} \log \mathcal{L}(\mathbf{J}|s_k) \propto -\frac{\partial}{\partial J_{ij}} \left(\mathcal{H}(s_k|\mathbf{J}) - \log Z\right) \tag{2.15}$$

where the term C has been omitted, since it is independent of J_{ij}. With $\mathcal{H}$ given by Eq. 1.1,

$$\frac{\partial}{\partial J_{ij}}\mathcal{H}(s_k|\mathbf{J}) = \frac{\partial}{\partial J_{ij}}\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right) \tag{2.16}$$

$$= \sigma_i\sigma_j. \tag{2.17}$$

And since

$$\log \mathcal{Z} = \log\left[\sum_{\sigma_1=\pm 1}\cdots\sum_{\sigma_m=\pm 1}\exp\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)\right], \tag{2.18}$$

the derivative of $\log \mathcal{Z}$ with respect to J_{ij} is

$$-\frac{\partial}{\partial J_{ij}}\log\mathcal{Z} = -\frac{\partial}{\partial J_{ij}}\log\left[\sum_{k=1}^{n}\exp\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)\right] \tag{2.19a}$$

$$= \frac{-\sum_{k=1}^{n}\frac{\partial}{\partial J_{ij}}\exp\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)}{\sum_{k=1}^{n}\exp\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)} \tag{2.19b}$$

$$= \frac{-\sum_{k=1}^{n}\exp\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)}{\sum_{k=1}^{n}\exp\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)} \times (\sigma_i\sigma_j) \tag{2.19c}$$

$$= -\sum_{k=1}^{n}\frac{\exp\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)}{\sum_{k=1}^{n}\exp\left(\sum_{i,j} J_{ij}\sigma_i\sigma_j + \sum_i h_i\sigma_i\right)} \times \sigma_i\sigma_j \tag{2.19d}$$

$$= -\sum_{k=1}^{n} P(s_k|\mathbf{J}) \times \sigma_i\sigma_j \tag{2.19e}$$

$$= -\langle\sigma_i\sigma_j\rangle\ , \tag{2.19f}$$

where $\langle\cdot\rangle$ denotes an average over the number of samples, and $\sum_{\sigma_1=\pm1}\cdots\sum_{\sigma_m=\pm1}$ as $\sum_{k=1}^{n}$ denotes summation over all possible spin configurations. Note that the exponential terms in the third line (Eq. 2.22c) do not cancel out because the second term is multiplied by $\sigma_i\sigma_j$ then summed over all possible states (the values of σ_i and σ_j depend on the subscript k).

With these results, we now return to the definition of the Fisher information, Eq. 2.12:

$$\mathbb{E}\left[\frac{\partial}{\partial J_{ij}}\log\mathcal{L}(\mathbf{J}|s_k)\right]^2 \propto \mathbb{E}\left[\left(\sigma_i\sigma_j - \langle\sigma_i\sigma_j\rangle\right)\left(\sigma_m\sigma_n - \langle\sigma_m\sigma_n\rangle\right)\right]. \tag{2.20}$$

Therefore, at the true parameters, the (ij, mn) element of the Fisher information is

$$\mathcal{I}_{ij,mn}(\mathbf{J}) \propto \langle\sigma_i\sigma_j\sigma_m\sigma_n\rangle - \langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle. \tag{2.21}$$

Because the spins hold binary values of ± 1, along the diagonal $\sigma_i \sigma_i$ is always equal to 1, and it follows that $\langle \sigma_i \sigma_j \rangle = 1$ and $\langle \sigma_i \sigma_j \sigma_i \sigma_j \rangle = 1$. Thus,

$$\mathcal{I}_{ij,ij}(\mathbf{J}) \propto \mathrm{Var}\left[\sigma_i \sigma_j\right] = 1 - \langle \sigma_i \sigma_j \rangle^2. \tag{2.22}$$

If the interactions in a spin network are symmetric and there are no self-coupling terms, then the Fisher information matrix has dimensions $\frac{m(m-1)}{2}$ by $\frac{m(m-1)}{2}$, where m is the number of spins in the system.

2.3.1 Eigenvectors

For a spin network comprised of more than two spins, the likelihood function depends on more than one parameter. For example, in a three-spin network, the likelihood function depends on $\mathbf{J} = (J_{12}, J_{13}, J_{23})$. As such, there exist many combinations of J_{12}, J_{13} and J_{23} that give rise to different likelihood "surfaces". We want to estimate the best set of combinations to explain our observations. For a three-spin case, we look for a peak in four-dimensional space. However, in general, for any given point (J_{12}, J_{13}, J_{23}), there will be directions that correspond to either "stiff" modes (sharp peaks) or "sloppy" modes (broad peaks).

The normal curvatures of a surface help characterize the behavior of the likelihood function as a function of the parameters of interest. These directions of the normal curvatures are given by the eigenvectors of the Fisher information matrix. The eigenvector $\mathbf{v}^*$ corresponding to the largest eigenvalue λ^* gives the direction that is most sensitive to changes in $\mathbf{J}$. In other words, $\mathbf{v}^*$ can be used to study the the likelihood surface in the direction corresponding to the most extreme curvature (sharpest peak).

As an example, we examine three different three-spin networks shown in Fig. 2.2. The magnitude of the coupling strengths are all set to unity. In one of the networks, the strength coupling coefficients are uniform. In the second network, two of the coupling coefficients are positive while the other is negative, and in the third network, two of the coupling coefficients are negative while the other is positive. Given that the true network coupling parameters are uniform in the first network (Fig. 2.2a), we expect that the eigenvector for the first network with the largest eigenvalue has all of its elements equal: $\mathbf{v}^* = (0.57, 0.57, 0.57)$, meaning that the direction where J_{12}, J_{13}, J_{23} are changed by the same amount has highest curvature. Similarly, in Figs. 2.2b and c, the directions of highest curvature occur where $J_{13} = J_{23}$.

The higher the eigenvalues are, the easier it is to learn the parameters of the network using the maximum-likelihood approach, whereas the lower the eigenvalues are, the harder it is to learn. Higher eigenvalues correspond to lower inference error, as we will see in the next section. When the eigenvalue is zero, the observations carry no information about the parameter.

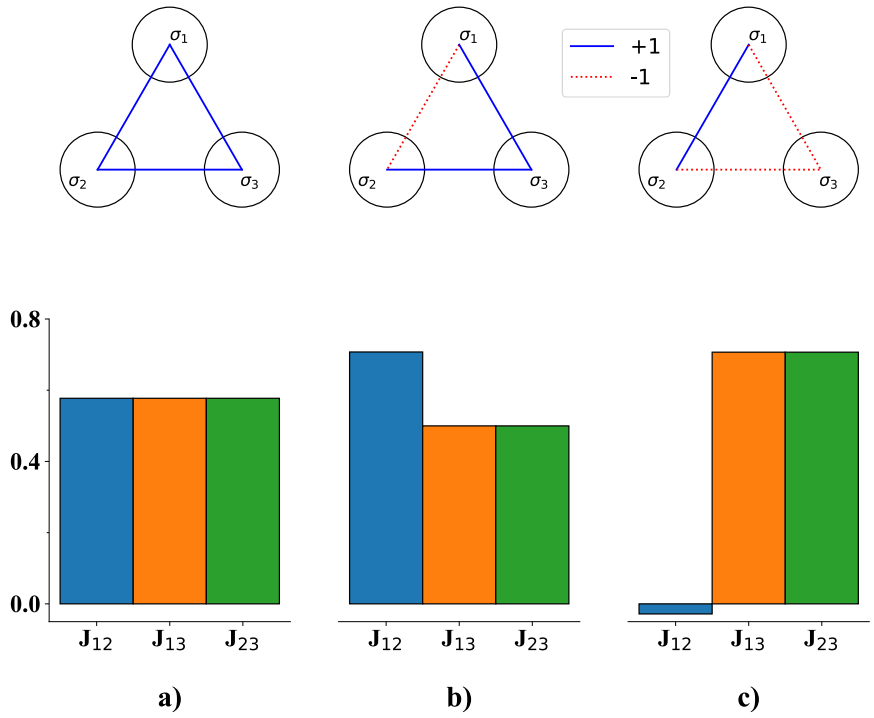


Figure 2.2: Eigenvectors (bottom row) corresponding to the largest eigenvalue of the Fisher information matrix of three different spin networks (top row) with a) uniform, positive couplings; b) one negative ($J_{12} = -1$) and two positive couplings ($J_{13} = J_{23} = 1$); and c) one positive ($J_{12} = 1$) and two negative couplings ($J_{13} = J_{23} = -1$).

2.3.2 Cramér-Rao Bound

Increasing the Fisher information of a system tends to increase the maximum precision (the amount of relevant information) that can be obtained from an estimation algorithm. According to the Cramér-Rao inequality [8], the mean-squared error of any unbiased estimator is lower bounded by the reciprocal of the Fisher information:

$$\mathrm{Var}(\hat{\mathbf{J}}) \geq \mathcal{I}(\hat{\mathbf{J}})^{-1} = \left(-\mathbb{E}\left[\frac{\partial^2 \log \mathcal{L}(\mathbf{J}|s_k)}{\partial J_{ij} \partial J_{mn}} \right] \right)^{-1} . \tag{2.23}$$

The bias of an estimator $\hat{J}$ for the parameter J is the expected value of the error of the estimator (the error of the estimator is the difference between the true parameter and the estimator), taken with respect to the probability distribution of the true parameter:

$$\mathbb{E}\left[\hat{J} - J \right] . \tag{2.24}$$

The estimator is unbiased if the expected value of the estimator is equal to the parameter. When the bound is achieved, the estimator is said to be efficient.

From Eq. 2.23, the lower bound for the variance of the maximum-likelihood estimator is inversely proportional to the curvature of the log-likelihood function. If the curvature is low, then there will be a wider spread of values of J' that are closer to the maximum likelihood estimator $\hat{J}$, whereas if the curvature is high, there will be a smaller number of values that are closer to $\hat{J}$. For this reason, in Ch. 3 we will assess the learning efficiency by the harmonic mean of the eigenvalues of the Fisher information matrix, which is the inverse of the sum of the the inverse of the eigenvalues.

Chapter 3

Three-Spin Network

We now return to the three-spin network introduced in Sec. 2.3.1. Each of the spins can point either up or down, giving a total of $2^3 = 8$ possible network configurations. The three-spin network is interesting in that the number of spin states is small enough for calculations and analysis, yet it allows us to study frustration in a network, which occurs when a spin cannot satisfy all interactions, regardless of the direction it is pointing. Figure 3.1 shows all of the possible states that a three-spin system can assume, with each state having a different probability of occurring, depending on the coupling strengths of the network.

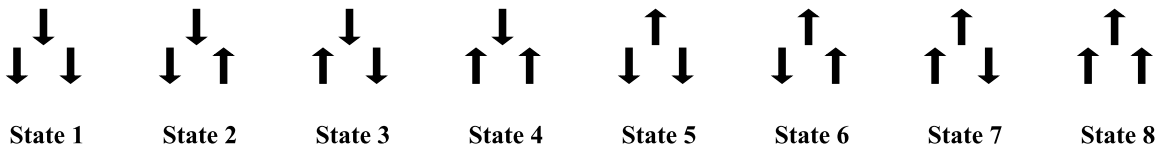


Figure 3.1: The eight different states that a three-spin network can assume, when each spin is restricted to point either up ($\sigma_i = 1$) or down ($\sigma_j = -1$).

3.1 Unit, Uniform Coupling Strengths

If the coupling parameters $J_{ij},\ J_{mn}$ between spins σ_i and σ_j, σ_m and σ_n, respectively, are restricted to be uniform in magnitude, i.e. $J_{ij} = \pm J_{mn}$, then there are ten possible spin networks each with different coupling matrices, as shown in Fig. 3.2. Note that this number accounts for permutations: for example, we would get the same set of networks if the labels of σ_1 and σ_2 were flipped.

Each of these ten networks have different coupling matrices. As before, we will restrict ourselves to symmetric interactions with no self-interaction. For simplicity, we will refer to each network by its coupling matrix, denoted as J^l. So J^0 refers to the network with

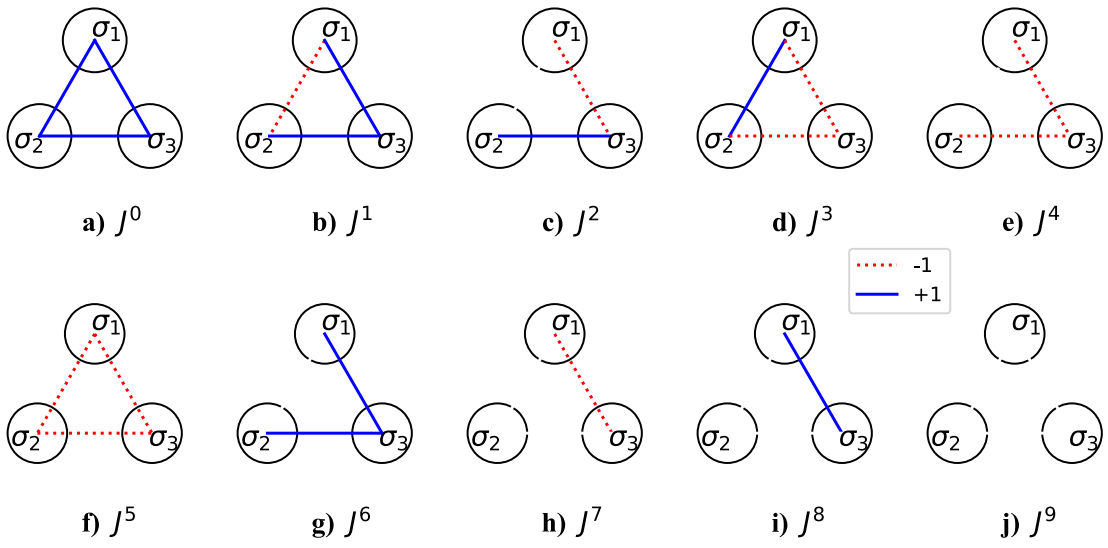


Figure 3.2: The ten different networks a spin network can assume if the interaction terms are restricted to be uniform in magnitude, $J_{ij} = \pm J_{mn}$. Each of these networks has a different coupling matrix, labeled as J^l. A dashed red line indicates negative coupling, and a blue solid line indicates positive coupling.

$J_{12} = J_{13} = J_{23} = 1$ (Fig. 3.2a), J^1 refers to the network with $J_{12} = -1, J_{12} = J_{13} = 1$ (Fig. 3.2b), etc.

3.1.1 Probability distribution

As a check, in Fig. 3.3, the probability distributions of each of the ten networks in Fig. 3.2 are plotted as a function of the eight states shown in Fig. 3.1. Figure 3.3a shows that two states, where all spins are aligned, exist with equal probability when the interactions between the spins are positive: the positive interactions induce the spins to align themselves. For spin networks J^2, J^3, J^4 and J^6, there are also two states that occur with equal probabilities, close to 1/2.

From these diagrams, we can see that J^1 corresponds to a frustrated system. The positive couplings between σ_1 and σ_3, as well as σ_2 and σ_3, tend to align both σ_1 and σ_2 with σ_3. However, the negative interaction between σ_1 and σ_2 causes these two spins to prefer to be anti-aligned. For that reason, there is no single state that the system prefers to assume, as shown in Fig. 3.3b. States 1, 3–6, 8 occur with the highest probability; these are the ones in which spin σ_3 is aligned with spins σ_1 or σ_2. Conversely, states 2 and 7 occur with the lowest probability, as these states align spins σ_1 and σ_2.

In Fig. 3.3, spin network J^5 has all negative coupling coefficients, hence the probabilities of states 1 and 8, where the spins are all aligned, are near zero. Since spin networks J^7 and J^8 are opposite images of each other, their probability distributions complement, with the

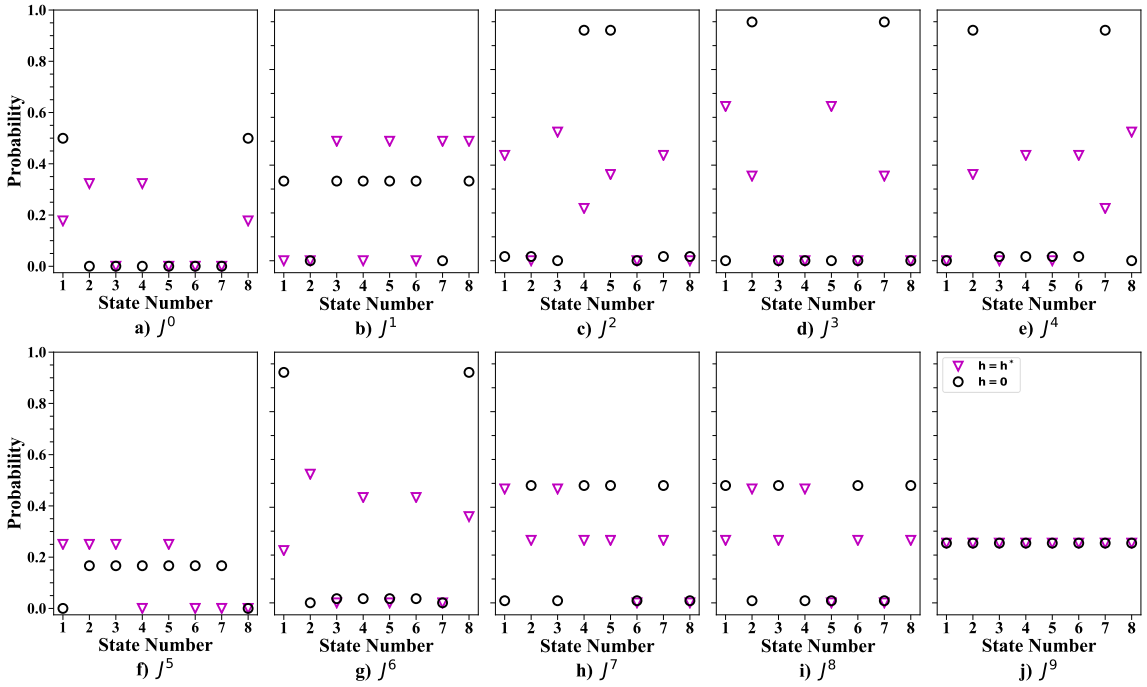


Figure 3.3: Probability distribution of the ten networks over the eight states shown in Fig. 3.2 with no applied field $\mathbf{h} = \mathbf{0}$ (black circles), or with the optimal field $\mathbf{h} = (\mathbf{h_1^*}, \mathbf{h_2^*}, \mathbf{h_3^*})$ (magenta triangles).

highest probabilities occurring in the states where the spins σ_1 and σ_3 are anti-aligned and aligned, respectively.

The last spin network, J^9 has no coupling between spins; thus all states occur with equal probability (Fig. 3.3j) as the exponential term in Eq. 1.2 is zero.

3.1.2 Harmonic mean

Figure 3.4 plots the inverse of the trace of the inverse of the Fisher information for each of the ten networks. The trace of the Fisher information matrix is the sum of its eigenvalues. It follows that the inverse of the trace of the inverse of the Fisher information can be written in terms of the sum of the eigenvalues λ_i of the Fisher information matrix,

$$\mathrm{Tr}\left[\mathcal{I}^{-1}\right]^{-1} = \frac{1}{\sum_{i=1}^{n} \frac{1}{\lambda_i}}, \tag{3.1}$$

where the (ij, lm)-element of the Fisher information is given by Eq. 2.21.

Equation 3.1 is the harmonic mean of the eigenvalues of the Fisher information. The lower the value of the harmonic mean, the harder it is to estimate the true coupling parameters, while the higher the value, the easier it is, since the harmonic mean is the weighted sum of the variances along the eigen-directions. We will refer to the harmonic mean as

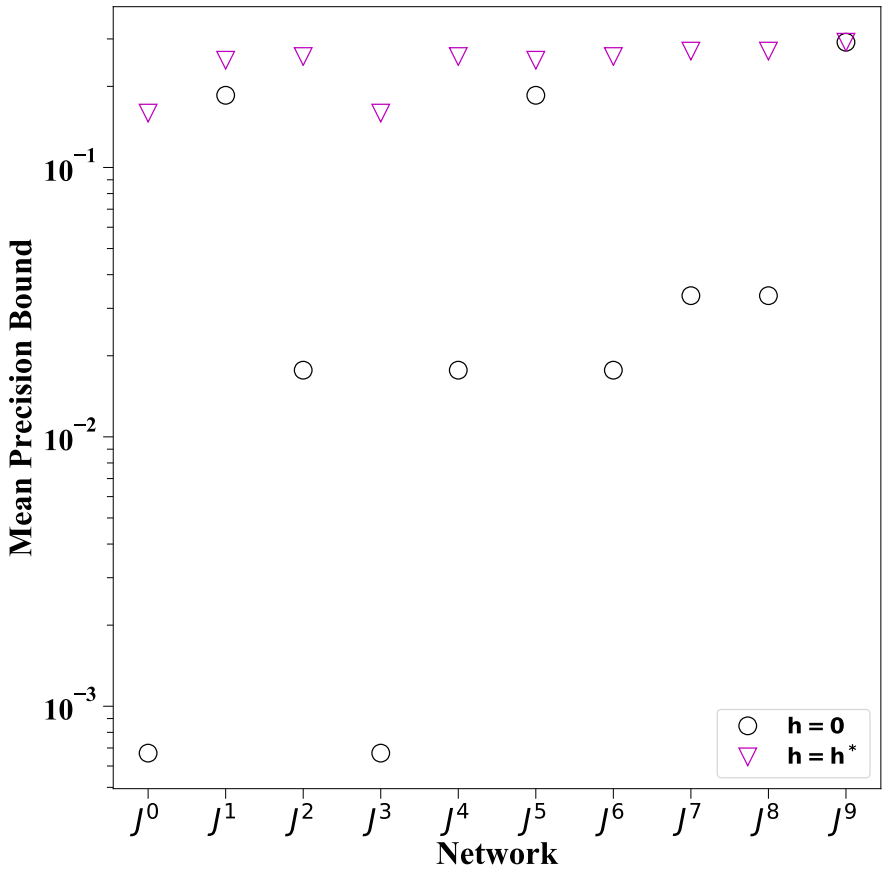


Figure 3.4: Inverse precision bound of the ten three-spin networks with zero field, $\mathbf{h} = \mathbf{0}$ (black circles) and with optimal field, $\mathbf{h} = \mathbf{h}^*$ (magenta triangles).

the "inverse precision bound", as it is related to how precisely a model parameter can be measured (refer to Sec. 2.3.2).

Figure 3.4 shows that the easiest spin network to learn is spin network J^9, where there are no interactions between the spins, and where the probabilities are equal across all states. By contrast, the hardest networks to learn are J^0 and J^3. These two networks have two states with high probabilities while the remaining six states have equally low probabilities. It appears that networks that are easier to learn have a more uniform probability distribution, with fewer excited states and less energy in the ground states, as in networks J^1, J^5 and J^9 (see Fig. 3.5 for plots of the energy distributions.).

Networks J^0 and J^3 may be harder to learn because two out of the eight states occur with probabilities close to 50%, while the other states occur with essentially zero probability. While networks J^2, J^4 and J^6 also have two states with high probabilities, some of the other states have significantly non-zero probabilities.

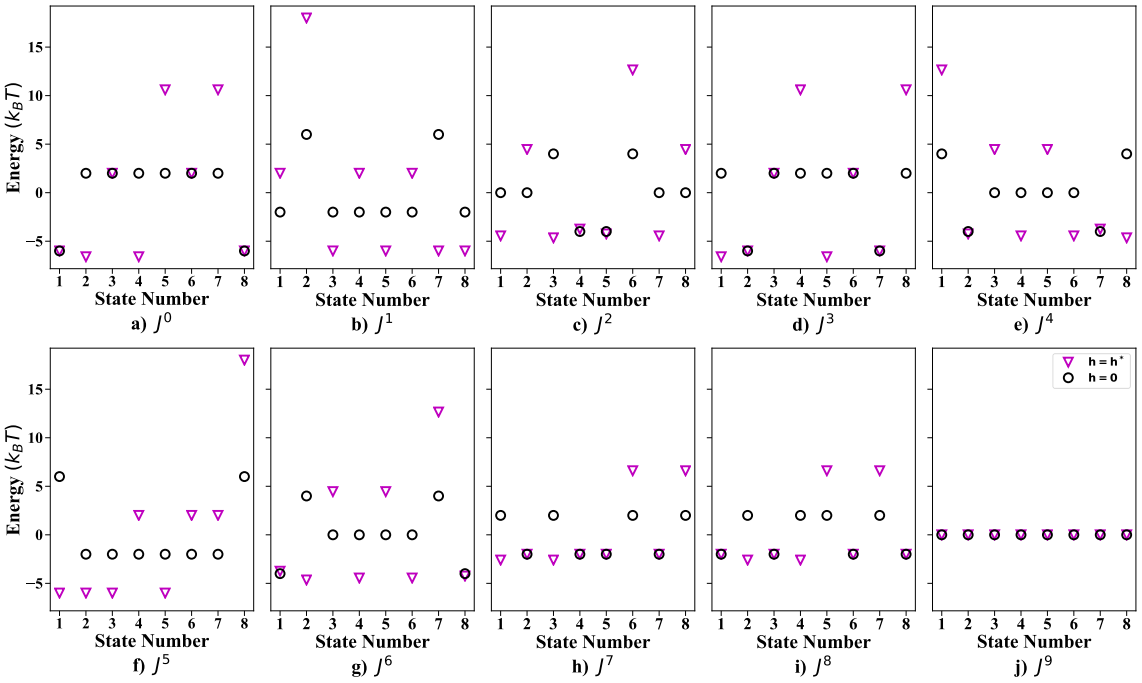


Figure 3.5: Energy distribution of the ten three-spin networks across the eight states shown in Fig. 3.2, with no field $\mathbf{h} = \mathbf{0}$ (black circles) and the optimal fields $\mathbf{h}^* = (h_1^*, h_2^*, h_3^*)$ (magenta triangles).

3.2 Optimal perturbation

We now return to our original question: does there exist a perturbation field in which learning efficiency is improved? To answer this, for each of the ten networks a range of fields h_i in Eq. 1.1 are individually applied to each of the spins σ_1, σ_2, and σ_3. We then search for a combination of fields $\mathbf{h}^* = (h_1^*, h_2^*, h_3^*)$ that maximizes the harmonic mean of the Fisher information eigenvalues (Eq. 3.1).

Figure 3.6 shows contour plots of the inverse precision bound as a function of the fields h_1 and h_2, with h_3 set to the optimal value. Note that only the relative signs between the fields matter: the inverse precision bound is identical for applied fields $(h_1, -h_2, h_3)$ or for applied fields $-(h_1, -h_2, h_3) = (-h_1, +h_2, -h_3)$.

Fig. 3.4 shows the maximal values of inverse precision bound (triangles) attained for the optimal fields. J^9 is still the easiest network to learn.

Figure 3.7 shows the directions of the optimal fields for each of the ten networks, for the couplings given in Fig. 3.6. In all of these cases, the fields seem to counteract the interactions between spins. Where there is positive coupling between spins σ_i and σ_j, the optimal fields h_i and h_j are opposite in direction, and where there is negative coupling between spins, the optimal fields h_i and h_j are in the same direction. For example, in the frustrated network J^1 (Fig. 3.7b), the optimal fields push σ_1 and σ_2 to align (against the negative coupling

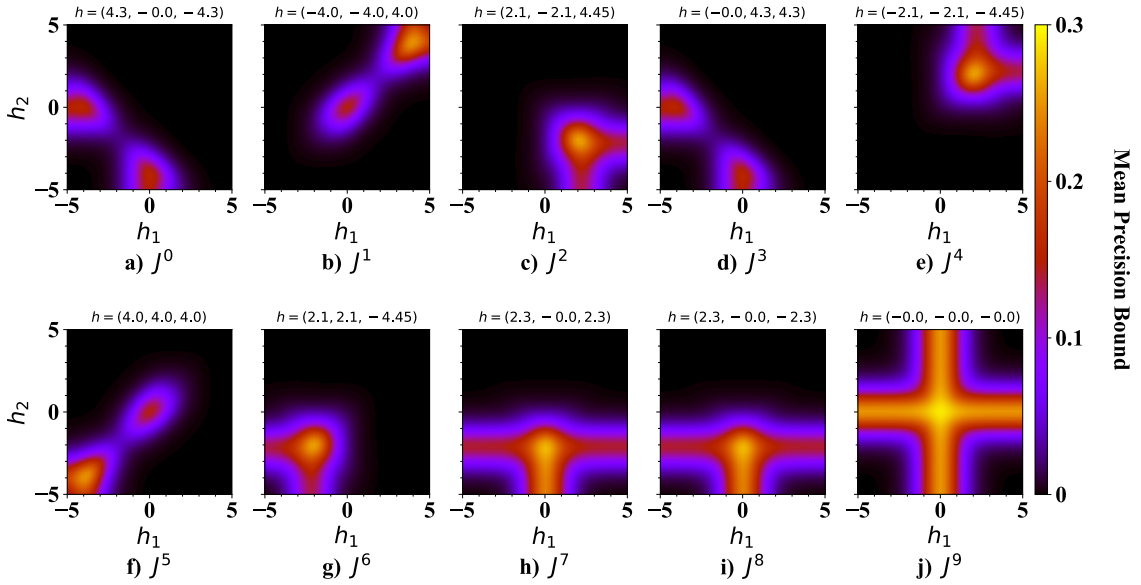


Figure 3.6: Contour plot showing the inverse precision bound (Eq. 3.1) as a function of fields h_1 and h_2, at the optimal field h_3^*. The optimal fields $\mathbf{h}^* = (h_1^*, h_2^*, h_3^*)$ that maximize the inverse precision bound are given on the top of each subplot.

between them), σ_1 and σ_3 to anti-align (against the positive coupling), and σ_2 and σ_3 to anti-align (again, against the positive coupling).

Although we have shown that the optimal fields $\mathbf{h}^* = (0, -4.3, -4.3)$ counteract the negative coupling between spins σ_2 and σ_3 in Fig. 3.7d, the Fisher information is equally maximized when fields of the same magnitude are applied to spins σ_1 and σ_2, but in opposite directions: $\mathbf{h}^* = \pm(4.3, -4.3, 0)$. In this case, the field counteracts the positive coupling between σ_1 and σ_2.

In Fig. 3.8 the variance terms $\langle\sigma_i\sigma_j\sigma_m\sigma_n\rangle$ and $\langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle$ that make up the analytical form of the Fisher information (Eqs. 2.21, 2.22) are plotted. Since the Fisher information matrix is symmetric, only four terms are shown: $\mathcal{I}_{ij,ij}$, $\mathcal{I}_{12,13}$, $\mathcal{I}_{12,23}$, and $\mathcal{I}_{13,23}$.

Focusing on the black and magenta circles that represent the value of the first term in Eq. 2.21, $\langle\sigma_i\sigma_j\sigma_m\sigma_n\rangle$, with and without the field, respectively, we see that the field brings this value closer to zero. Similarly, application of the optimal field reduces the magnitude of the second term, $\langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle$ (black triangles for $\mathbf{h} = 0$ and magenta triangles for $\mathbf{h} = \mathbf{h}^*$, respectively).

After the application of the field, the inverse precision bound for all the networks have approximately the same order of magnitude, because the elements of the Fisher information have all been reduced. Not only is the inverse precision bound larger when the trace is maximized, but also when the off-diagonal entries approach zero, i.e. when $\langle\sigma_i\sigma_j\sigma_m\sigma_n\rangle$

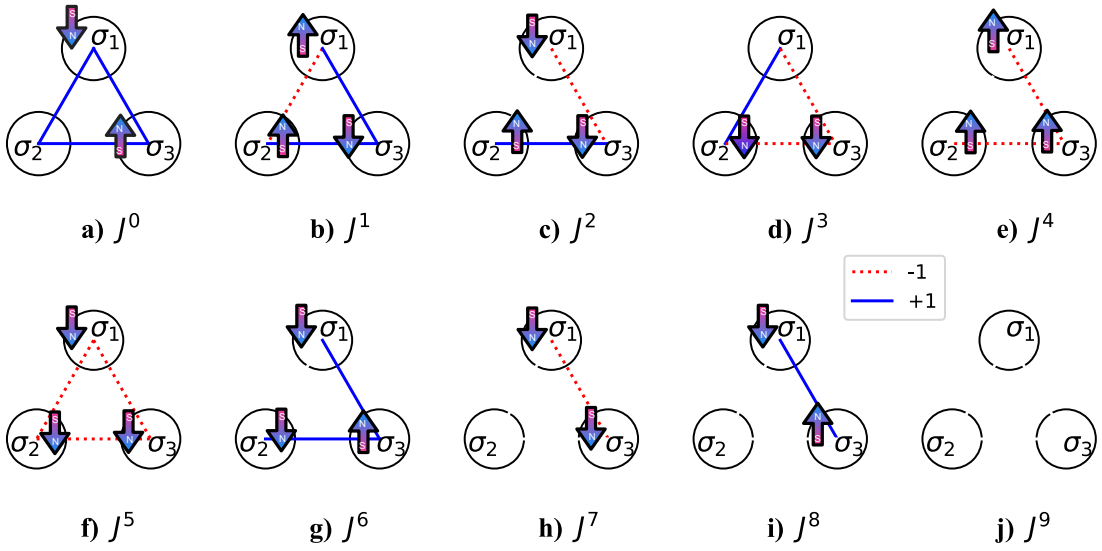


Figure 3.7: The relative directions of the optimal fields on the ten networks of a three-spin network with unit couplings. The arrows are not drawn to scale. Note that the directions of the fields can be simultaneously reversed to give the same maximal value of Tr $\left[\mathcal{I}^{-1}\right]^{-1}$.

is close to $\langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle$ (recall Eq. 2.21), as in network $J9$ (Fig. 3.8j). The optimal field seems to reduce correlations between the spins in the system, thereby increasing $\mathcal{I}_{ij,ij}$ (since $\langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle$ becomes closer to zero) and decreasing $\mathcal{I}_{ij,mn}$ (since both the magnitudes of $\langle\sigma_i\sigma_j\sigma_m\sigma_n\rangle$ and $\langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle$ are smaller). In network J^9, where the inverse precision bound has the highest value, the off-diagonal terms are equal to zero, while the diagonal element $\langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle$ is small ($\langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle = 0.125$). For this network, spins fluctuate depending on their thermal environment and do not interact.

3.2.1 Probability distribution

Referring back to the triangles in Fig. 3.3, we can see that when the optimal field is applied to any given network, certain states increase in probability, while others decrease. In general, application of the field seems to reduce the probability difference between states for all ten networks, and increases the probability of otherwise unlikely states. For example, states 2 and 4 in spin network J^0 increase in probability while states 1 and 8, which initially have probabilities of $\sim$50%, decrease in probability. States 2 and 4 correspond to states with spins σ_1 and σ_3 aligned with the direction of the optimal field shown in Fig. 3.7a, while states 3, 5, 6, and 7 have spins σ_1 and σ_3 anti-aligned with the direction of the optimal field or both anti-aligned with spin σ_2 (σ_2 still experiences positive interactions with σ_1 and σ_3).

In the frustrated network in Fig. 3.7b, the optimal field shown is $\mathbf{h}^* = (4.0, 4.0, -4.0)$. States 1, 2, 4, and 6, which correspond to the network having two or more spins anti-aligned with the applied field, have the lowest probabilities. Conversely, states 3, 5, 7 and 8 have

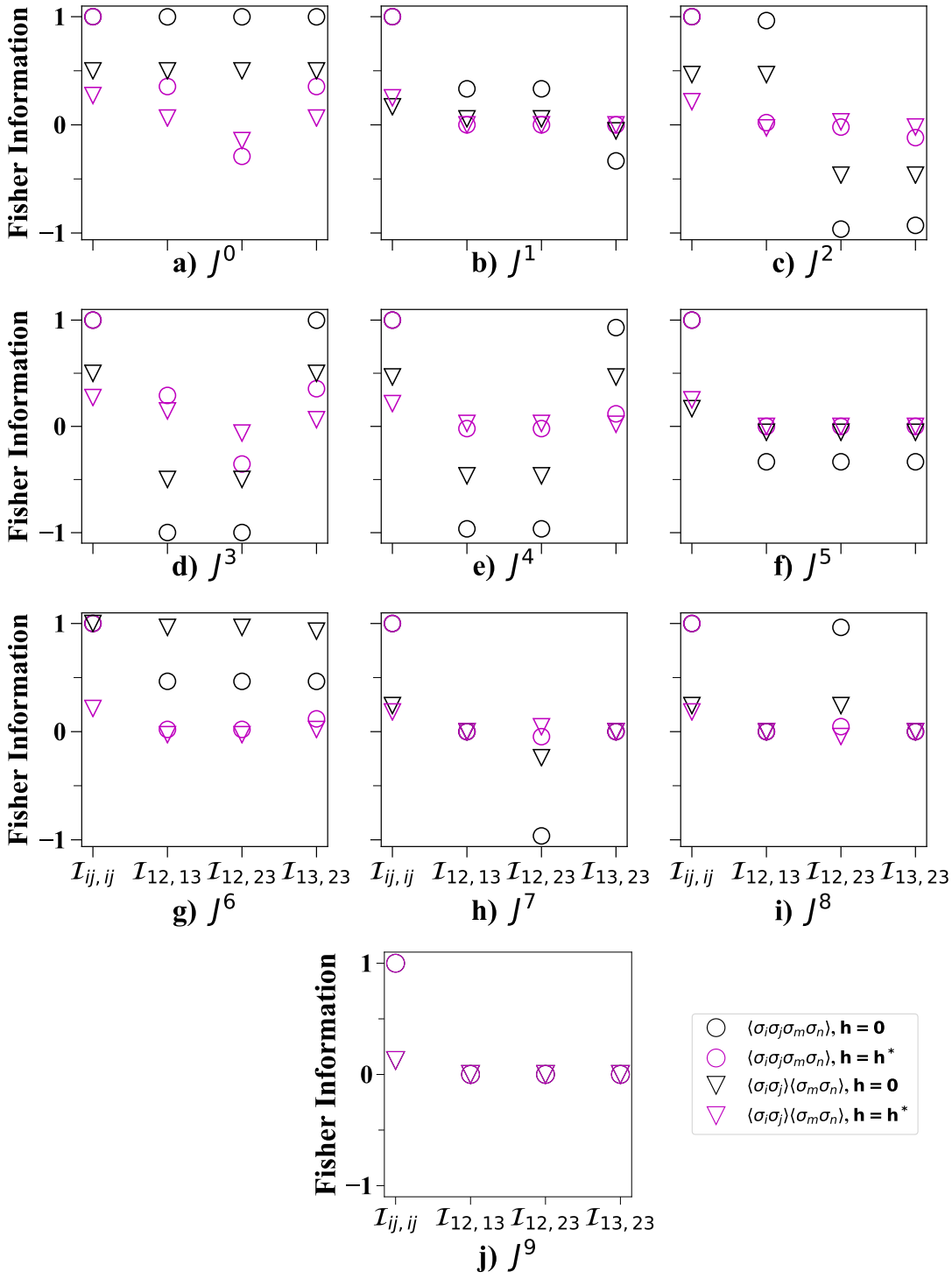


Figure 3.8: Components of the Fisher information for each of the ten networks: $\langle\sigma_i\sigma_j\sigma_m\sigma_n\rangle$ (circles) and $\langle\sigma_i\sigma_j\rangle\langle\sigma_m\sigma_n\rangle$ (triangles), with $\mathbf{h} = 0$ (black) and $\mathbf{h} = h^*$ (magenta).

the highest probabilities – these correspond to states that have two or more spins aligned with the applied field.

Learning occurs best for network J^9 when there is no field applied. This is reasonable because the lack of interactions between the spins allows the spins to fluctuate independently. Applying a field does not provide information on how tightly coupled the spins are and destroys the uniformity of the probability distribution. Moreover, application of an incorrect field can result in erroneous predictions, because the probability distributions of systems (and hence the likelihood function) at incorrect fields may resemble each other. For example, in Fig. 3.9, the likelihood functions and the probability distributions for spin networks J^8 and J^9, which have an improperly applied field of $\mathbf{h} = (-5, 5, -5)$, appear identical. Given this data and assuming that the range of parameter values were unknown, the observer would be unable to infer the couplings, despite knowing the strength of the applied field.

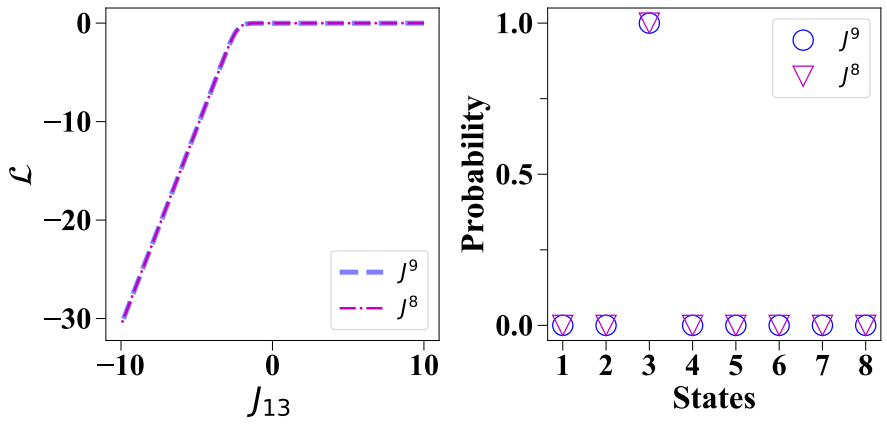


Figure 3.9: Likelihood function (a) and probability distribution (b) of networks J^8 (magenta triangles) and J^9 (blue circles) with $\mathbf{h} = (-5, 5, -5)$.

3.2.2 Energy distribution

We can also look at the energy distribution of the networks before and after the application of the optimal field (Fig. 3.5). At equilibrium, a system prefers to be in the lowest energy state(s). For an Ising spin network, the optimal field seems to generally reduce the energy of excited states. With the field applied, certain states that are unfavorable due to the positive/negative interactions between the spins as well as states that are aligned with the field have a reduction in energy, while states that have two or more spins anti-aligned with the direction the field are excited.

But how much does the Fisher information increase after the application of the field? The purple triangles in Fig. 3.4 show the values of the harmonic mean after the optimal field is applied. Networks (such as J^0 and J^3) that have greater increases in the harmonic

mean have fewer states that increase in energy. Excluding network J^9, networks J^1 and J^5 show the smallest increase in the harmonic mean. The two networks each have 4 states that increase in energy, and 4 states that are reduced in energy. The difference in energy between the most excited states and the ground states is largest in these two networks.

Even with the optimal field, as in the absence of any applied field, networks J^0 and J^3 are still the hardest networks to learn. The next two easiest networks to learn after network J^9 are networks J^7 and J^8 (which only have spins σ_1 and σ_3 coupled). Excluding network J^9, these two networks have the smallest energy difference between the ground states and the highest excited states, and as mentioned previously in Sec. 3.1.2, have higher energy ground states. Networks J^0 and J^3 have the lowest energy ground states. While network J^5, which is easier to learn than network J^0 and J^3 (but slightly harder to learn that J^7 and J^8) also has the same low energy ground states, the energy difference between most of its states is smaller, with one high energy state that occurs with essentially zero probability.

The energy difference between the original states with no applied field and the states with the applied field provides a way to measure how much energy is required to decouple the interactions between spins, to better learn the strength of the couplings within the networks.

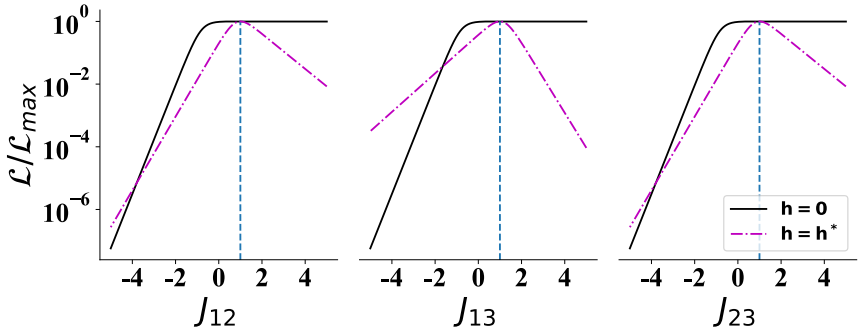
3.2.3 Likelihood function

Figure 3.10 shows the likelihood functions (on a log scale) with and without the application of the optimal field for the two networks J^0 and J^3 that are hardest to learn.

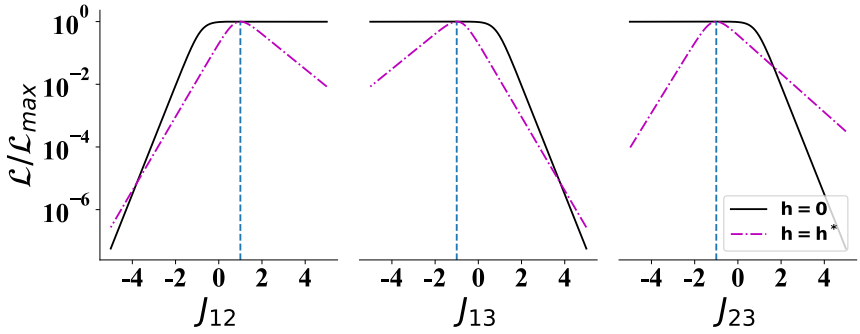
The likelihoods are marginal likelihoods as a function of unknown parameter J_{ij}, with all the other parameters set to their true values.

For both networks, when the field is zero, the likelihood function is flat around the true parameters. This shows why spin network J^0 is difficult to learn: although spins prefer to be aligned most of the time, it is difficult to estimate how strong the couplings are. Values of the coupling constants that are positive are equally likely to explain why we observe states 1 and 8 the majority of the time. Similarly, for network J^3, there are two states that occur with probabilities close to 50%. Since the network is overwhelmingly likely to occupy one of those two states, it is also difficult to estimate the strength of the interaction. However, applying $\mathbf{h} = \mathbf{h}^*$ reveals how tightly bound the spins originally were, and the curvature of the likelihood functions increase near the true parameters. Clearly, inference has been improved, as indicated by the thousand-fold increase in harmonic mean (refer back to Fig. 3.4). We have $\mathrm{Tr}\left[\mathcal{I}^{-1}\right]^{-1} = 6.7 \times 10^{-3}$ when $\mathbf{h} = 0$ and $\mathrm{Tr}\left[\mathcal{I}^{-1}\right]^{-1} = 1.6 \times 10^{-1}$ when $\mathbf{h} = \mathbf{h}^*$, for both networks.

Without the field, the likelihood function for network J^0 (Fig. 3.10a) is the same for J_{12}, J_{13} and J_{23} because of the symmetry in the network: all coupling parameters are equal to zero. With the field on, the likelihood function for J_{13} differs from the likelihood functions for J_{12} and J_{23}, because the field is only applied to spins σ_1 and σ_3.



a)



b)

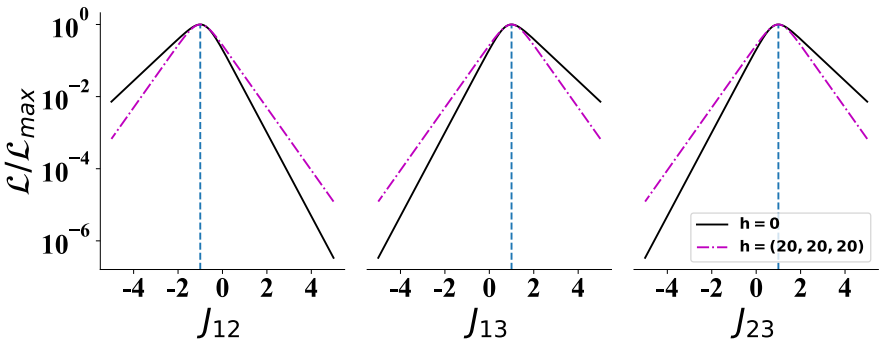
Figure 3.10: Likelihood function for networks a) J^0 and b) J^3, with $\mathbf{h} = 0$ (solid black line) and $\mathbf{h} = \mathbf{h}^*$ (dashed-dot magenta line). The true coupling coefficient is given by the dashed blue line.

Figure 3.11 shows the likelihood functions of J^1 and J^5, the networks that, upon application of their optimal field, show the smallest increases in the harmonic mean of the Fisher information eigenvalues (from 0.19 when $\mathbf{h} = 0$ to 0.25 when $\mathbf{h} = \mathbf{h}^*$). The increase is much smaller compared to networks J^0 and J^3.

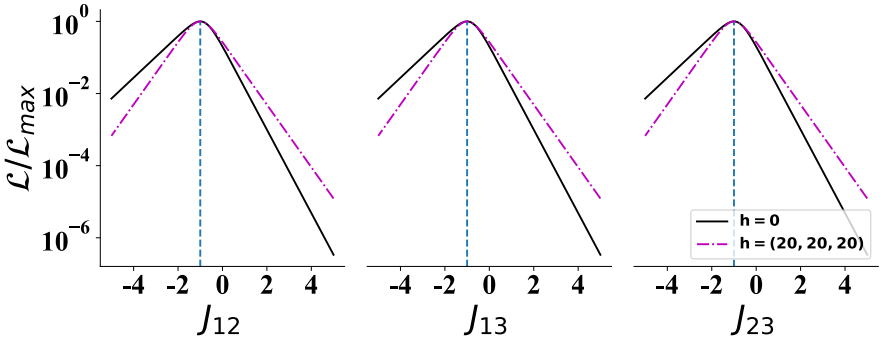
Upon application of the optimal field, the likelihood functions appears to rotate slightly, with not much difference between the curvatures of the likelihood function of the networks with and without the optimal fields.

As a comparison, Fig. 3.12 shows the likelihood function for spin network J^9, which has the best learning efficiency out of all networks, with the optimal field and at zero field. The likelihood functions are symmetric about zero, and have the sharpest curvature out of all ten networks.

Also shown in Fig. 3.12 is the likelihood function at very large fields; the likelihood is flat everywhere (dashed magenta line), indicating that inference is poor. In general, for all spin networks, the learning efficiency is worst at when $\sum h_i\sigma_i \gg \sum J_{ij}\sigma_i\sigma_j$, i.e. at high field strengths.



a)



b)

Figure 3.11: Likelihood function for networks a) J^1 and b) J^5, with $\mathbf{h} = 0$ (solid black line) and $\mathbf{h} = \mathbf{h}^*$ (dashed-dot magenta line). The true coupling coefficient is given by the dashed blue line. The likelihoods have been scaled so that the largest value is 1.

3.3 Non-unit, uniform coupling strengths

Section 3.1 presented results for unit coupling strengths, $J_{ij} = \pm 1$. In this section, we examine networks that have non-unit, but still uniform, coupling strengths of magnitude $c = 0.1$ to $c = 2.0$. For each of these values, we find the fields that maximize the inverse precision bound.

Fig. 3.13 shows the field $\mathbf{h}^* = (h_1^*, h_2^*, h_3^*)$ as a function of the coupling strength. To save computational time, the values of (h_1^*, h_2^*, h_3^*) were incremented in steps of 0.05 for coupling strengths greater than 0.5, or in steps of 0.01 for coupling strengths less than 0.5. For network J^5 however, the fields were searched in steps of 0.01 at all coupling strengths, because this network was analyzed separately.

Since there is more than one possible combination of (h_1^*, h_2^*, h_3^*) for certain networks (because only the relative directions are important), the directions of the fields were chosen to match the directions shown in Fig. 3.7. For all networks the field increases linearly with the coupling strengths.

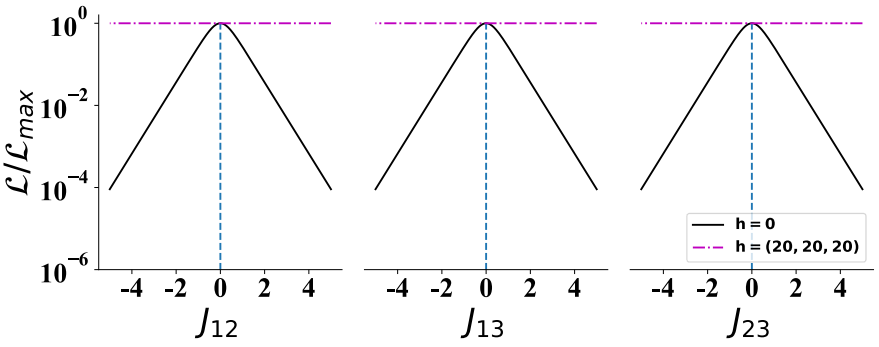


Figure 3.12: Likelihood function for network J^9, where the interactions between spins are all zero, $J_{ij} = 0$ without the field $\mathbf{h} = 0$ (black solid line) and with a very large field $\mathbf{h} = (20, 20, 20)$ (magenta dashed line). The true coupling parameters are given by the dashed blue lines. The likelihood has been scaled so that the largest value is 1.

Figure 3.14 shows the inverse precision bound for both the perturbed (magenta triangles) and the unperturbed (black circles) cases at different couplings strengths, for each of the ten networks shown in Fig. 3.2. Without the field, the Fisher information decreases dramatically for most of the networks. In particular, for networks J^0, J^2, J^3, J^4, J^6, J^7 and J^8. the Fisher information tends to zero when $J_{ij} > 1.5$. The Fisher information approaches zero for these networks because there are fewer low energy states. As the coupling strength increases, so do the energies of the excited states. From the analysis in the previous sections, it seems that systems that are easier to learn have a more uniform probability distribution; thus when coupling strengths increase, the systems are more likely to stay in their ground states, increasing the differences in energies between the ground and the excited states.

With the optimal field applied, the Fisher information decreases much less as the coupling strength is increased. At around $J_{ij} = 0.75$, the Fisher information begins to level off. Networks J^0 and J^3 have the largest decrease – these are the networks that are hardest to learn, as noted in the previous section (Fig. 3.4 shows the unit coupling case). The harmonic mean for network J^9 (the network that is easiest to learn) is 0.292, whereas the smallest value of the Fisher information is 0.159 for network J^0 with $c > 0.75$.

When the coupling strengths increase, inference becomes more difficult, since the curvature of the log-likelihood decreases. Past a certain value, inference is almost impossible since the harmonic mean of the Fisher information eigenvalues is near zero. The effect of the optimal field seems to minimize the change in the curvature of the log-likelihood function near the true coupling parameter, i.e. minimize the overall change in the Fisher information. Hence, as the coupling strength increases, inference is possible with the application of the optimal field.

In Fig. 3.15, the energy and the probability distributions for networks J^5, J^1 and J^0 are shown for four different coupling strengths, $c = 0.5, 1.0, 1.5, 2.0$. To compare the difference

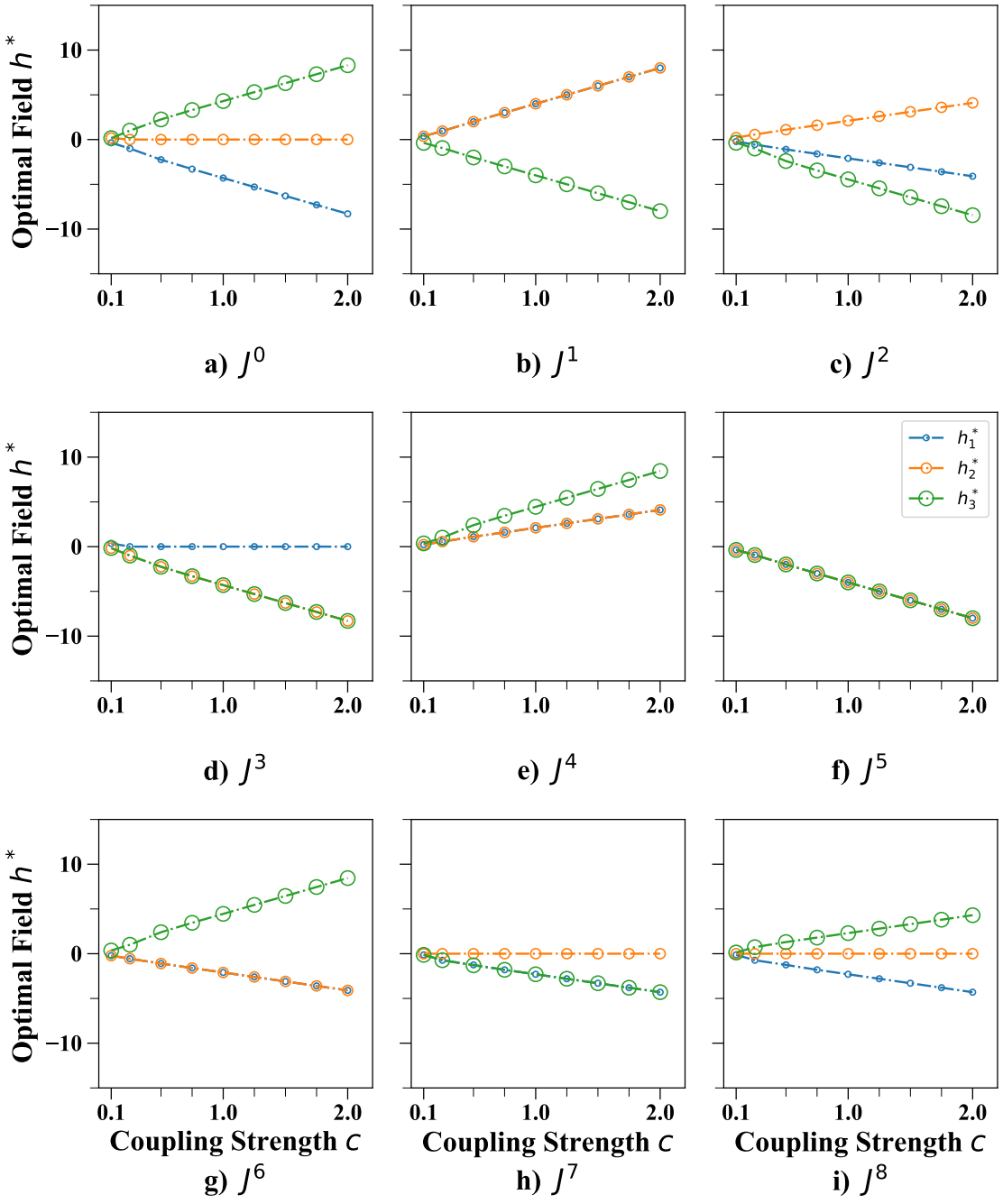


Figure 3.13: Values of the optimal field $\mathbf{h}^*$ for networks J^0 through J^8, (network J^9 is omitted, because the optimal fields are zero for all values of c shown on this figure), as a function of coupling strength c. The directions of the fields are chosen to match the directions shown in Fig. 3.7. The field applied onto spin h_1^* is shown in blue; on h_2^* in orange; and on h_3^* in green.

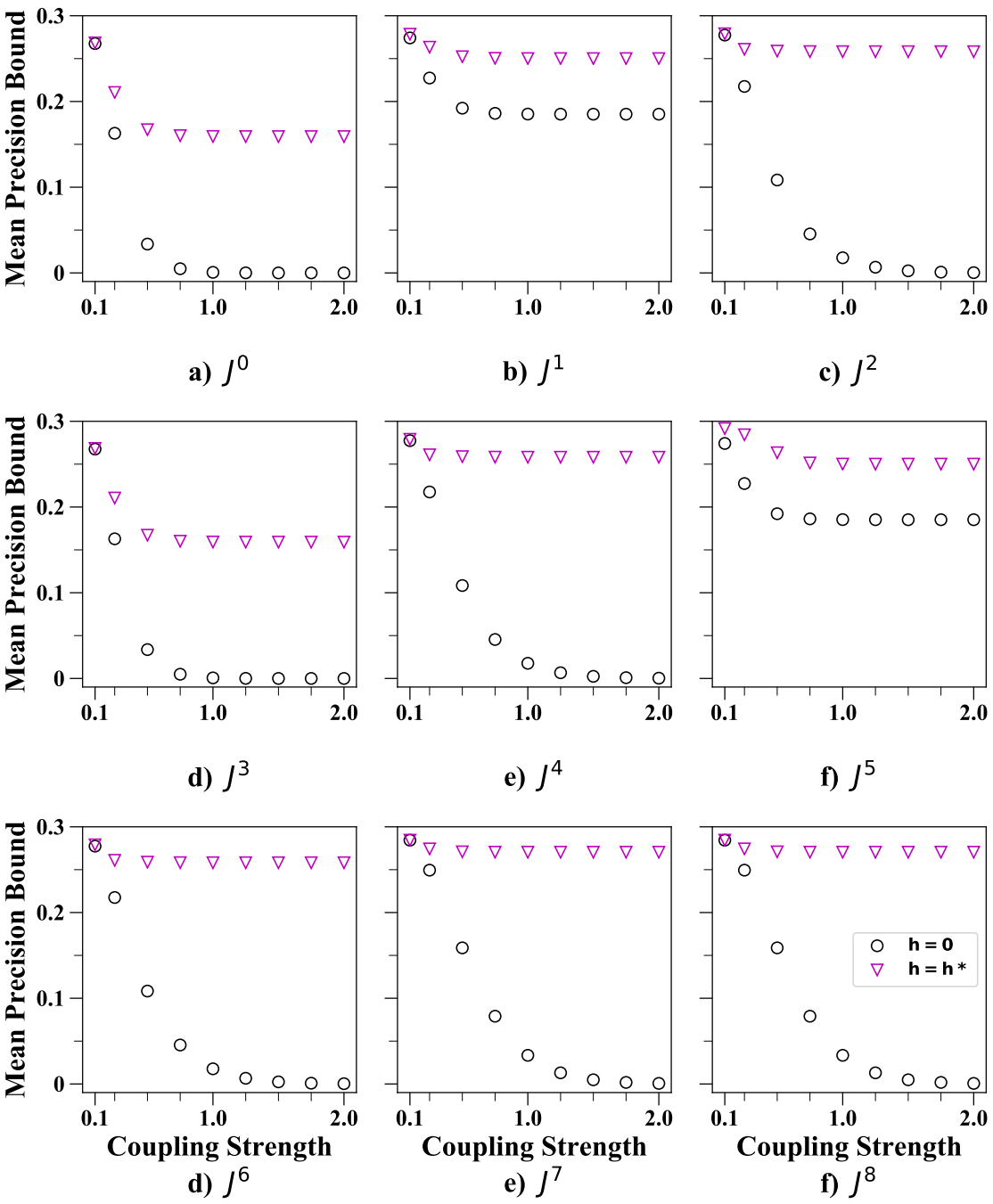


Figure 3.14: Inverse precision bound as a function of coupling strength for network J^0 though J^8 in Fig. 3.2. Black circles show the inverse precision bound for the unperturbed case, while magenta triangles show the maximized Fisher information for the optimal field.

in energies before and after the application of the optimal field, we also plot the distributions for coupling strength $c = 2.0$, but with zero field. Changes in the coupling strength does

not appear to change the probability distribution of the networks that are perturbed with the optimal field.

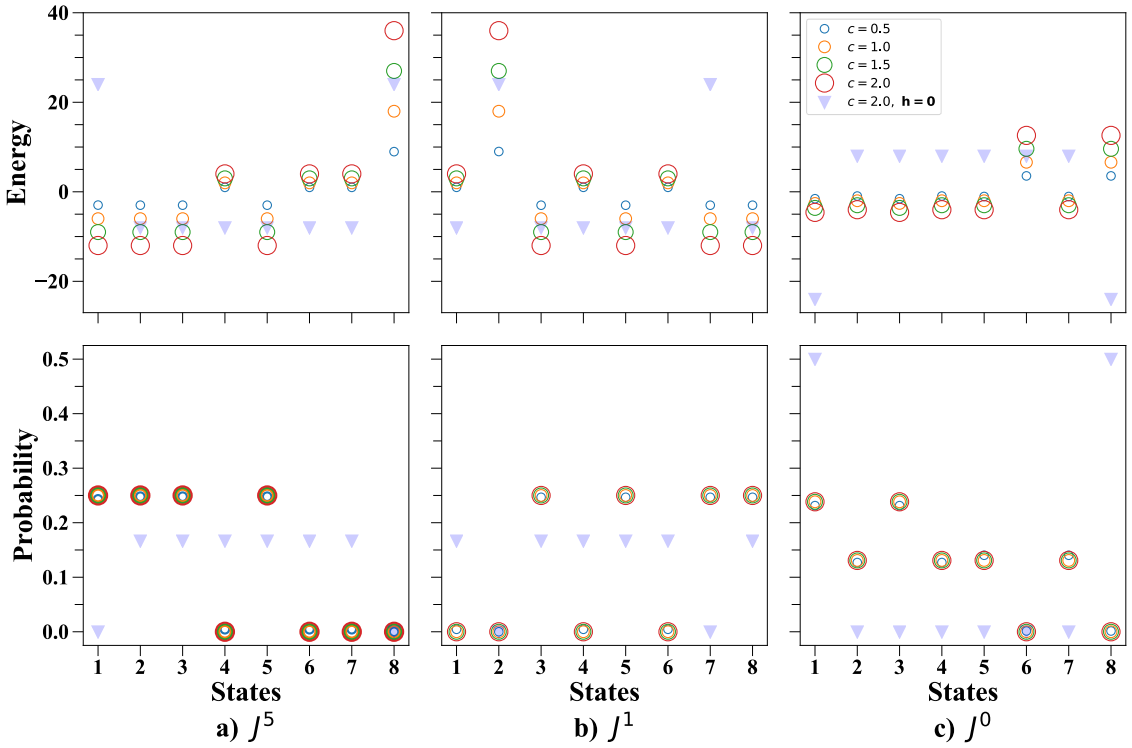


Figure 3.15: Energy (top row) and probability distribution (bottom row) for various coupling strengths c (circles) for a) network J^5, b) network J^1 and c) network J^0. For reference, analogous results for $c = 2.0$ and no field, $\mathbf{h} = \mathbf{0}$, are also plotted (light blue triangles).

Chapter 4

Conclusion and Future Work

One of the main challenges in computational biology is the reconstruction of biological networks from high-throughput data. Understanding the interconnections within biological networks, such as genetic or signaling pathways, allows the potential for scientists to reprogram or design cells that can proliferate into specialized cells. In this work, we sought to understand how to better predict the interactions within biological networks by modeling them with a binary Ising spin system. Using the maximum-likelihood approach, we introduce perturbations in the form of fields to various three-spin networks so as to optimize inference. The optimal perturbations maximize the Fisher information (the curvature of the likelihood function) near the true parameters. We show that inference is improved with the application of particular fields, as judged by the increase in curvature of the likelihood function. By reducing their energies, such fields increase the probability of certain states that would otherwise not be observed. This optimal field seems to decouple and de-correlate the interactions between the spins, allowing us to see how tightly bound the spins are.

For networks with uniform couplings, we show that optimal field strength depends on the coupling strength between spins, increasing linearly with coupling strengths. We find that at higher coupling strengths, with the optimal field, the quality of inference is not reduced significantly, and perhaps stabilizes beyond a given coupling strength. At the optimal fields, the probability distribution does not change much when the coupling strength is varied.

Questions arise, however, when finding the optimal fields for large systems, as well as for large coupling strengths. As the number of spins in the network increases, so do the number of fields that need to be optimized. With the addition of each field comes an extra dimension when calculating the Fisher information for a particular value of the field. It would be time consuming to precisely search through all combinations of the fields for the best Fisher information in a very large system. In real biological systems, the number of “spins" is very large, as there are many compounds, some undiscovered, that affect DNA transcription and protein production. Enhancement or suppression of gene expressions, represented by the magnetic field, depend on many factors. Moreover, the magnitude of interactions (spin couplings) will vary for different compounds. In a cellular setting, we can imagine finding the

optimal field by introducing or eliminating specialized proteins or other genes that changes how the expression of (or lack of) one gene affects the expression another.

To tackle this issue, we require methods to predict optimal fields, which begins with understanding how field affects simple networks such as the ones studied in this work. Future work could include the study of non-uniform systems that have various coupling strengths between different spins, as well as system that have non-symmetric interactions. It would also be interesting to see how the optimal field scales with coupling strength for systems that exhibit self-coupling.

Bibliography

[1] National Institutes of Health, U.S. Department of Health and Human Services, NIH Stem Cell Information Home Page in *Stem Cell Information* [World Wide Web site], Bethesda, MD, 2016.

[2] K.Takahashi and S. Yamanaka,"Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors", Cell **126** (2006).

[3] E. Macosko *et al.*"Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets" Cell **161** (2015).

[4] T. Gardner, D. di Bernardo, D. Lorenz, and J. Collins, "Inferring Genetic Networks and Identifying Compound Mode of Action via Expression Profiling", Science **301** (2003).

[5] H. Nguyen, R. Zecchina, and J. Berg,"Inverse Statistical Problems: from the Inverse Ising problem to Data Science", Adv. in Phys. **66** 3 (2017).

[6] S. Haykin, *Neural Networks and Learning Machines* (Pearson Education Inc., New Jersey, 2009).

[7] L. Held and D. Bové, *Applied Statistical Inference Likelihood and Bayes* (Springer-Verlag, Berlin Heidelberg, 2014).

[8] T. Cover and J. Thomas, *Elements of Information Theory* (John Wiley & Sons, New Jersey, 2006).

[9] R. Hogg, J. McKean, and A. Craig *Introduction to Mathematical Statistics* (Pearson, Boston, 2019) p. 363.

[10] A. Huang, B. Sheldan, D. Sivak and M. Thomson, "Physically Optimizing Inference", arXiv:1805.07512 (2018).

[11] E. Jaynes,"Information Theory and Statistical Mechanics", Phys. Rev. **106** 4 (1957).