

# 1<sup>ST</sup> CHAPTER ISING MODEL

N BINARY SPIN VARIABLES  $(s_i)_{i=1,\dots,n}$

MAGNETIC FIELDS  $(h_i)$

COUPLINGS MATRIX  $(J_{ij})_{ij}$ ,  $J_{ij} = J_{ji}$

HAMILTONIAN  $H(s) = -\sum_i h_i s_i - \sum_{i < j} J_{ij} s_i s_j$

BOLTZMANN MEASURE  $P(s) = \frac{1}{Z} \exp[-H(s)]$

PARTITION FUNCTION  $Z(J, h) = \sum_s \exp[-H(s)]$

SPIN VARIABLE:  $\sigma_i$ . REALIZATION:  $\alpha_i$ .

THERMAL AVERAGE  $\langle Q(\sigma) \rangle = \sum_{\alpha} P_{\alpha} Q(\alpha)$ .

"FIRST MOMENTS"  $m_i = \langle \sigma_i \rangle$  [MAGNETISATIONS]

"SECOND MOMENTS"  $X_{ij} = \langle \sigma_i \sigma_j \rangle$  [CORRELATIONS]

FACT:  $\frac{\partial \log Z}{\partial h_i} = \langle \sigma_i \rangle$ .       $\frac{\partial \log Z}{\partial J_{ij}} = \langle \sigma_i \sigma_j \rangle$ .

"HELMOLTZ FREE ENERGY"  $F(S, h) = -\log Z(S, h)$

FACT:  $F(S, h)$  IS CONCAVE AND  $\frac{-\partial F}{\partial h_i}(S, h) = m_i$ ,  $-\frac{\partial F}{\partial J_{ij}}(S, h) = X_{ij}$

FACT: "FUNCTIONAL MAXIMUM ENTROPY PRINCIPLE" I

$$\begin{aligned} F(S, h) &= \min_{q \in P_B} \left\{ \langle H \rangle_q(S, h) - \langle \log q \rangle_q \right\} \\ &= \min_{q \in P_B} \left\{ \mathcal{U}[q] - \mathcal{S}[q] \right\}. \\ &= \min_{q \in P_B} \left\{ \mathcal{F}[q] \right\} \end{aligned}$$

$P_B$ : IS THE SET OF BINARY DISTRIBUTIONS ON SPINS.

WE CALL  $\mathcal{F}[q]$  THE "FUNCTIONAL HELMHOLTZ FREE ENERGY"

# 2<sup>ND</sup> CHAPTER INVERSE PROBLEM AT EQUILIBRIUM?

WE SEE  $N$  iid SAMPLES  $D = \{\delta^n\}, n=1, \dots, N$

I MODEL THE REAL DISTRIBUTION WITH A BOLTZMANN,  
I WANT TO INFERENCE  $\beta, h$ .

[I WANT TO INFERENCE A SPARSE  $\beta$ ]

$$\langle Q \rangle^D = \frac{1}{N} \sum_n Q(\delta^n)$$

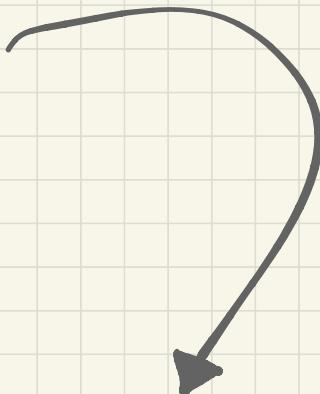
# WHY DO WE ASSUME A BOLTZMANN? MAXIMUM ENTROPY DISTRIBUTION

$$\underset{P \in P_{bi}}{\text{MAX}} - \sum_i P(\mathbf{x}_i) \log P(\mathbf{x}_i)$$

SUCH THAT  $\sum_i P(\mathbf{x}_i) = 1$

$$\sum_i P(\mathbf{x}_i) x_i = m_i^*$$

$$\sum_i P(\mathbf{x}_i) x_i x_j = X_{ij}^*$$



IMPLIES THAT  $P(\mathbf{x}) = \frac{1}{Z} \exp \left[ - \sum_i h_i x_i - \sum_{i,j} J_{ij} x_i x_j \right]$

WITH  $J, h$  THAT MAKE  $m_i^* = \langle \sigma_i \rangle$   
 $X_{ij}^* = \langle \sigma_i \sigma_j \rangle$

$$\underset{P \in P_{S_i}}{\text{MAX}} - \sum_s P(s_i) \log P(s_i)$$

$$= \underset{S, m}{\text{MAX}} - \sum_s P(s_i) \left[ + \sum_i h_i s_i + \sum_{i < j} S_{ij} s_i s_j - \log Z(S, m) \right]$$

$$= \underset{S, m}{\text{MAX}} - \sum_i h_i m_i^* - \sum_{i < j} S_{ij} X_{ij}^* + \log Z(S, m)$$

MAX IS GIVEN WITH  $S, m$  SUCH THAT  $m_i^* = \langle \sigma_i \rangle$   
 $X_{ij}^* = \langle \sigma_i, \sigma_j \rangle$

FACT: THIS IS EQUIVALENT TO SEARCHING FOR THE MAX.  
LIKELIHOOD DISTRIBUTION AMONG DOLTMANN FAMILR  
GIVEN SAMPLES

$$L_D(S, \theta) = \frac{1}{M} \log P(D | S, \theta)$$

$$= \sum_{i \in S} S_{ij} \frac{1}{M} \sum_{n=1}^N \gamma_i^n \gamma_j^n + \sum_i m_i \frac{1}{M} \sum_n \gamma_i^n - \log Z(S, \theta)$$

$$= \sum_{i \in S} S_{ij} \langle \sigma_i \sigma_j \rangle^D + \sum_i m_i \langle \sigma_i \rangle^D - \log Z(S, \theta).$$

FACT: THIS IS EQUIVALENT TO SEARCHING FOR THE DISTRIB.  
THAT MINIMIZES  $D_{KL}(\cdot || P_D)$  AMONG DOLTMANN FAMILR

# CRUDE MAXIMUM LIKELIHOOD

$$\{\hat{s}^m, \hat{h}^m\} = \text{ARGMAX } \langle D(s, h) \rangle$$

FACT:  $\frac{\partial \langle D(s, h) \rangle}{\partial h_i} = \langle \sigma_i \rangle^D - \langle \sigma_i \rangle^{J, h}$

$$\frac{\partial \langle D(s, h) \rangle}{\partial s_{ij}} = \langle \sigma_i \sigma_j \rangle^D - \langle \sigma_i \sigma_j \rangle^{J, h}$$

SINCE  $\langle D(s, h), s \rangle$  IS CONCAVE  $\rightarrow$  GRADIENT ASCENT [BML].

$$h_i^{n+1} = h_i^n + \eta \frac{\partial \langle D(s, h) \rangle}{\partial h_i} (\hat{s}^n, \hat{h}^n)$$

$$s_{ij}^{n+1} = s_{ij}^n + \eta \frac{\partial \langle D(s, h) \rangle}{\partial s_{ij}} (\hat{s}^n, \hat{h}^n)$$

BUT COMPUTING  $\langle \sigma_i \rangle^D, \langle \sigma_i \sigma_j \rangle^D$  IS INFEASIBLE

## MAXIMUM LIKELIHOOD + MCMC

$$\text{APPROXIMATE } \langle \sigma_i \rangle \approx \frac{1}{3} \sum_{\delta=1}^3 \sigma_i^\delta, \quad \langle \sigma_i \sigma_j \rangle \approx \frac{1}{3} \sum_{\delta=1}^3 \sigma_i^\delta \sigma_j^\delta$$

GENERATED BY MCMC.

GLAUBER

METROPOLIS HASTINGS

WOLFF

SIMULATED ANNEALING

TEMPERING

LANGEVIA...

## MAXIMUM LIKELIHOOD + MCMC + BARES

SAME AS BEFORE, BUT ADDING A PRIOR ON  $\beta, \alpha$ .

$$\pi(\beta, \alpha | D) \propto \pi(\beta, \alpha) p(D | \beta, \alpha).$$

INSTEAD OF MAXIMIZING  $L_D(\beta, \alpha)$ , I MAXIMIZE  
 $L_D(\beta, \alpha) + \log \pi(\beta, \alpha)$

[YOU CAN GET SPARSE REGULARIZATION:

$$L_D^{l_1} = L_D - \gamma \sum_{i < j} |\beta_{i,j}|$$

$$L_D^{l_2} = L_D - \gamma \sum_{i < j} \beta_{i,j}^2$$

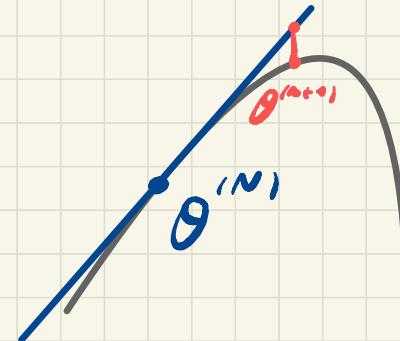
## MAXIMUM LIKELIHOOD + NEWTON

IN CLASSIC GRADIENT ASCENT,

$$f(\theta^{(n)} + \delta\theta^{(n)}) \approx f(\theta^{(n)}) + \nabla f(\theta^{(n)}) / \delta\theta^{(n)}$$

$$\delta\theta^{(n)} = \eta \nabla f(\theta^{(n)})$$

$$\theta^{(n+1)} = \theta^{(n)} + \delta\theta^{(n)}$$



IN NEWTON,

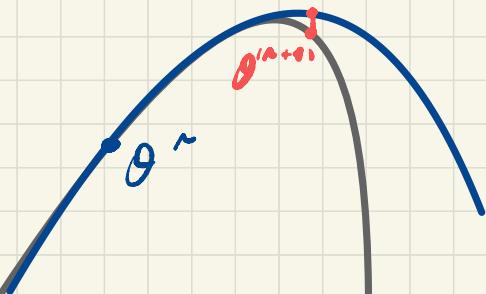
$$\nabla f(\theta^{(n)} + \delta\theta^{(n)}) \approx \nabla f(\theta^{(n)}) + \nabla^2 f(\theta^{(n)}) / \delta\theta^{(n)}$$

$$\nabla f(\theta^{(n)} + \delta\theta^{(n)}) = 0,$$

$$\delta\theta^{(n)} = -[\nabla^2 f(\theta^{(n)})]^{-1} \nabla f(\theta^{(n)})$$

$$\theta^{(n+1)} = \theta^{(n)} + \delta\theta^{(n)}$$

↗ JACOBIAN.



EFFICIENT ALGORITHM TO EVALUATE  $[\nabla^2 f(\theta^{(n)})]^{-1}$ : BFGS

## MAXIMUM LIKELIHOOD + MOMENTUM

IT HELPS GAINING VELOCITY IN THE FIRST STEPS OF BML.

CLASSIC MOMENTUM:  $\delta\theta^{(n)} = \eta \nabla f(\theta^{(n)}) + \beta \delta\theta^{(n-1)} + \gamma \delta\theta^{(n-2)}$

NESTEROV MOMENTUM:  $\delta\theta^{(n)} = \eta \nabla f(\theta^{(n)}) + \beta \delta\theta^{(n-1)} + \gamma \delta\theta^{(n-2)}$ .

## SCORE MATCHING

IN ABSTRACT: DATA COMES FROM  $g(x) d\pi(x)$  [UNKNOWN]

WE MODEL IT WITH  $f(x, \theta) d\pi(x)$  [FOR US, BOLTZMANN].

HEURISTICS:  $\theta^{\text{sm}} = \underset{\theta}{\text{ARGMIN}} \mathbb{E}_g [\|\nabla_x \log g(x) - \nabla_x \log f(x, \theta)\|^2]$

AFTER INTEGRATING BY PARTS:

$$\theta^{\text{sm}} = \underset{\theta}{\text{ARGMIN}} \mathbb{E}_g [(\nabla_x \log f(x, \theta))^2 + 2 \Delta_x \log f(x, \theta)]$$

$$\approx \underset{\theta}{\text{ARGMIN}} \mathbb{E}_D [(\nabla_x \log f(x, \theta))^2 + 2 \Delta_x \log f(x, \theta)]$$

$$\approx \underset{\theta}{\text{ARGMIN}} \mathbb{E}_D [(\nabla_x H)^2 - 2 \Delta_x H] \text{ DOES NOT REQUIRE } Z.$$

PROBLEM [IN MR OPINION]:  $\nabla_x$  DOES NOT MAKE SENSE FOR BOLTZMANN.

## — WASSERSTEIN VIEW —

FACT: IF  $X_0 \sim \nu_0$  AND  $\dot{X}_t = \nabla_{\nu} F(\nu_t)$ , THEN  $\nu_t = \text{Law}(X_t)$   
SATISFIES  $\Delta_t \nu_t + \nabla \cdot (\nabla_{\nu} F(\nu_t)) = 0$  [IN WEAK SENSE]

GIVEN  $\mathcal{F}: P_2 \rightarrow \mathbb{R}$  I CAN DEFINE  $\nabla_{W_2} \mathcal{F}[\nu](x): \mathbb{R}^d \rightarrow \mathbb{R}$   
 $\nu \mapsto \nabla \mathcal{F}[\nu]$   $x \mapsto \nabla_x \mathcal{F}[\nu](x)$

IDEA:  $\dot{\nu}_t = -\nabla_{W_2} \mathcal{F}[\nu_t]$ .

$\Delta_t \nu_t + \nabla \cdot (\nabla_{W_2} \mathcal{F}[\nu_t] \nu_t) = 0$  IS THE "WGF" OF  $\mathcal{F}$ .

FACT: IF  $\mathcal{F}$  IS (STRICTLY) CONVEX ALONG (WASSERSTEIN) GEODESICS,  
THEN  $\nu_t$  GOES EXPONENTIALLY FAST AND MONOTONICALLY  
TOWARDS  $\nu^* = \underset{\nu}{\operatorname{argmin}} \mathcal{F}[\nu]$  POLYAK THEOREM.

FACT: IF  $d\tilde{\pi}_0 = f_\theta ds$ ,  $f_\theta = \frac{1}{Z} e^{-H_\theta}$ ,  $d\mu_g = g ds$

THEN.  $\nabla_{W_2} D_{KL}(\mu_g \parallel \tilde{\pi}_0)(x) = \nabla_x H_\theta(x) + \nabla_x \log g(x)$

FACT:  $D_{KL}(\cdot \parallel \tilde{\pi}_0)$  IS STRICTLY CONVEX ALONG WASSERSTEIN GEODESICS, PROVIDED  $H_\theta$  IS STRICTLY CONVEX.

IDEA LOOK AT WGF OF  $D_{KL}(\cdot \parallel \tilde{\pi}_0)$  WHERE  $\tilde{\pi}_0$  IS BOLTZMANN, STARTING FROM THE REAL DISTRIBUTION  $\mu_g$

IF  $\tilde{\pi}_0 \approx \mu_g$ , THEN THE FLOW IS ALMOST STATIONARY SO THE (WASSERSTEIN) GRADIENT IS SMALL.



I WANT THE PARTICLES TO NOT MOVE MUCH

I WANT  $E_D [(-\nabla_x H_\theta - \nabla_x \log g)^2]$  TO BE SMALL

$$= E_D [\|\nabla_x \log f(x, \theta) - \nabla_x \log g(x)\|^2].$$

## MINIMUM PROBABILITY FLOW

SIMILAR HEURISTIC, FOR DISCRETE DOMAINS (LIKE ISING).

DATA COMES FROM  $P^{(0)}$ .  $P_i^{(0)} = P(\sigma_i)$

I WANT TO APPROXIMATE IT WITH BOLTZMANN  $P_\theta^{(\infty)}$ .  $(P_\theta^{(\infty)})_i = P_\theta(\sigma_i)$

I CREATE A DYNAMICS THAT CONVERGES TO  $P_\theta^{(\infty)}$ .

[I HAVE TO IMPOSE DETAILED BALANCE:  $(\Gamma_\theta)_{ij}(P_\theta^{(\infty)})_j = (\Gamma_\theta)_{ji}(P_\theta^{(\infty)})_i$   
WITH  $(\Gamma_\theta)_{ij} = \#\{j \rightarrow i\}$ ]

I CHOOSE  $(\Gamma_\theta)_{ij} = g_{ij} \exp[\frac{1}{2}(H_\theta(\sigma_j) - H_\theta(\sigma_i))]$

$g_{ij} = 1$  IF AND ONLY IF  $\sigma_i, \sigma_j$  DIFFERS BY ONE FLIP.

I FOLLOW THE DYNAMICS STARTING FROM  $\hat{P}^{(0)}$ .

$$\dot{\hat{P}}_i^{(\infty)} = \sum_{j \neq i} (\Gamma_0)_{ij} \hat{P}_j^{(\infty)} - \sum_{j \neq i} (\Gamma_0)_{ji} \hat{P}_i^{(\infty)}$$

IDEA:  $\hat{P}^{(0)} \approx \hat{P}_0^{(\infty)} \Rightarrow D_{KL}(\hat{P}^{(0)} \parallel \hat{P}_0^{(\infty)}) \approx 0$ .

$$D_{KL}(\hat{P}^{(0)} \parallel \hat{P}^{(\epsilon)}) \approx D_{KL}(\hat{P}^{(0)} \parallel \hat{P}^{(\epsilon)}) \Big|_{\epsilon=0} + \epsilon \frac{\partial D_{KL}(\hat{P}^{(0)} \parallel \hat{P}_0^{(\epsilon)})}{\partial \epsilon} \Big|_{\epsilon=0}$$

$$= \sum_{i \in D} \dot{\hat{P}}_i^{(0)}$$

$$= \frac{\epsilon}{M} \sum_{i \in D} \sum_{j \in D} (\Gamma_0)_{ij}$$

$$= \frac{\epsilon}{M} \sum_{j \in D} \sum_{i \in D} g_{ij} \exp \left[ \frac{1}{2} [H_0(g_i) - H_0(g_j)] \right]$$

MINIMIZE THIS AND FIND  $\theta^{\text{MPF}}$ .

# VARIATIONAL PRINCIPLES PRELUDE

THE LEGENDRE TRANSFORM OF  $F(S, M)$  IS

$$S(X, m) = \min_{S, M} \left[ -\sum_i m_i m_i - \sum_{i,j} S_{ij} X_{ij} - F(S, M) \right] \quad \text{CONCAVE}$$

FACT:  $S(X, m)$  IS THE ENTROPY OF A BOLTZMANN WITH  
FIXED  $\langle \sigma_i \rangle = m_i$ ,  $\langle \sigma_i \sigma_j \rangle = X_{ij}$ . [UP TO THE SIGNS]

$$\text{FACT: } \frac{-\Delta S}{\Delta m_i} (X, m) = m_i, \quad \frac{\Delta^2 S}{\Delta m_i \Delta m_j} (X, m) = S_{ij} \quad [S, M \text{ FROM CORRESP. BOLT.}]$$

THE Gibbs FREE ENERGY IS

$$G(S, M) = \max_h \left[ \sum_i m_i m_i + F(S, h) \right] \quad \text{CONCAVE}$$

$$\frac{\Delta G}{\Delta m_i} (S, M) = m_i$$

$$\begin{aligned} \frac{\Delta^2 G}{\Delta m_i \Delta m_j} (S, M) &= (C^{-1})_{ij} \\ &= [(X_{ij} - m_i m_j)^{-1}]_{ij} \end{aligned}$$

**FACT:** "FUNCTIONAL MAXIMUM ENTROPY PRINCIPLE" II.

$$S(X, m) = \max_{q \in S} \{ f(q) \}$$

WITH  $S = P_0 \cap \{ q : \langle \sigma_i \rangle_q = m_i, \langle \sigma_i \sigma_j \rangle_q = X_{ij} \}$

$$G(S, m) = \min_{q \in G} \left\{ - \sum_{i < j} S_{ij} \langle \sigma_i \sigma_j \rangle_q - f(q) \right\}$$

$$= \min_{q \in G} \{ g(q) \}$$

WITH  $G = P_0 \cap \{ q : \langle \sigma_i \rangle_q = m_i \}$ .

AND  $g(q)$  IS SAID AS THE FUNCTIONAL GIBBS FREE ENERGY

# VARIATIONAL PRINCIPLES + MEAN FIELD

IDEA [VERY GENERAL]

$$G(J, m) = \min_{q \in \underline{Q}} \left\{ - \sum_{i < j} J_{ij} \langle \sigma_i \sigma_j \rangle_q - f(q) \right\}$$

$$\approx \min_{q \in \underline{Q}^*} \left\{ - \sum_{i < j} J_{ij} \langle \sigma_i \sigma_j \rangle_q - f(q) \right\}$$

$$\underline{G}^* = \underline{G} \cap \dots$$

FIND A CLOSED FORM FOR  $G(J, m)$  UNDER THIS RESTRICTION.

$$\text{USE } \frac{\delta G}{\delta m_i}(J, m^D) = h_i, \quad \frac{\delta^2 G}{\delta m_i \delta m_j}(J, m^D) = (C^{-1})_{ij}^D$$

AND FIND  $J^{MF}, h^{MF}$ . (UNDERLYING PRINCIPLE: MAX. ENTROPY  $\rightarrow$  PMF)

IN THIS CASE  $\{ \dots \} = \left\{ P^{MF}; P_{(i,j)}^{MF} = \prod_i \frac{1+m_i}{2} \right\}$

THIS FORCES  $m_i = \tilde{m}_i$  AND SO  $|\hat{G}^{MF}| = |\underline{G} \cap \{\dots\}| = 1$ .

$$\hat{G}(m, \mathcal{I}) = \hat{G}^{MF}(m, \mathcal{I})$$

$$= - \sum_{i \in \mathcal{I}} \mathcal{I}_{i,i} m_i m_j + \sum_i \left[ \frac{1+m_i}{2} \log \frac{1+m_i}{2} + \frac{1-m_i}{2} \log \frac{1-m_i}{2} \right]$$

WHICH ENDS UP IN:

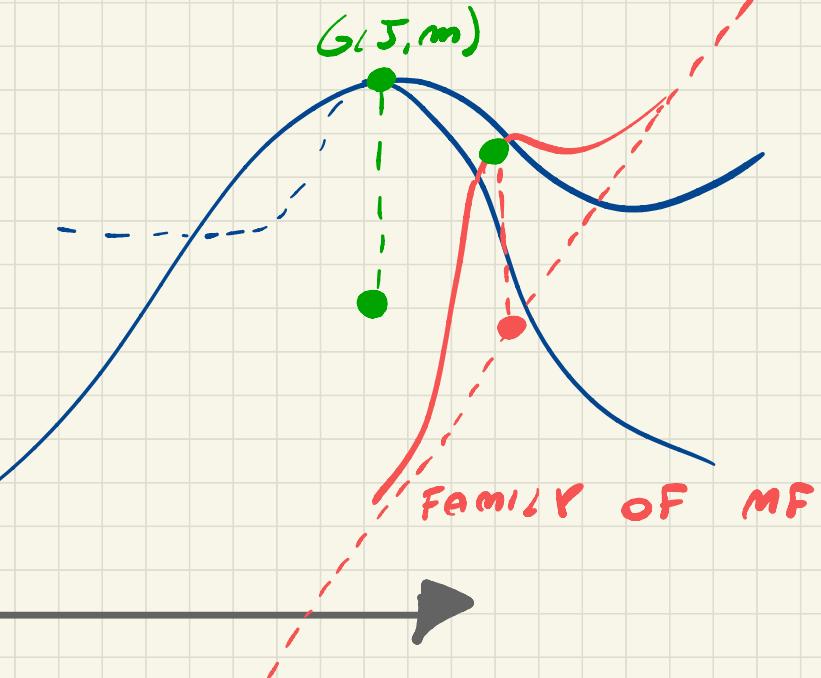
$$(C^{-1})_{ij}^{MF} = -\mathcal{I}_{ij}^{MF} \quad (i \neq j)$$

$$h_i^{MF} = -\sum_{j \neq i} \mathcal{I}_{ij}^{MF} + \operatorname{ARCTANH} m_i^D$$

A BIT FORCED  
[IN MY OPINION]

[YOU STILL NEED TO DO MATRIX INVERSION, POLYTIME].

$\mathcal{S}[q]$



# VARIATIONAL PRINCIPLES + ONSAGER

SAME AS BEFORE BUT YOU ADD

$$G^{TAD}(J, m) = G^M(J, m) - \frac{1}{2} \sum_{i>j} J_{ij}^2 (1-m_i^2)(1-m_j^2)$$

GIVING NOW

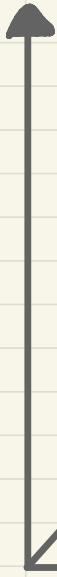
$$(C^{-1})_{ij}^D = -J_{ij}^{TAD} - 2 J_{ij}^{TAD} m_i^D m_j^D .$$

$$m_i^{TAD} = \arctan m_i^D - \sum_{j \neq i} J_{ij}^{TAD} m_j^D + m_i^D \sum_{j \neq i} (J_{ij}^{TAD})^2 / (1-(m_j^D)^2).$$

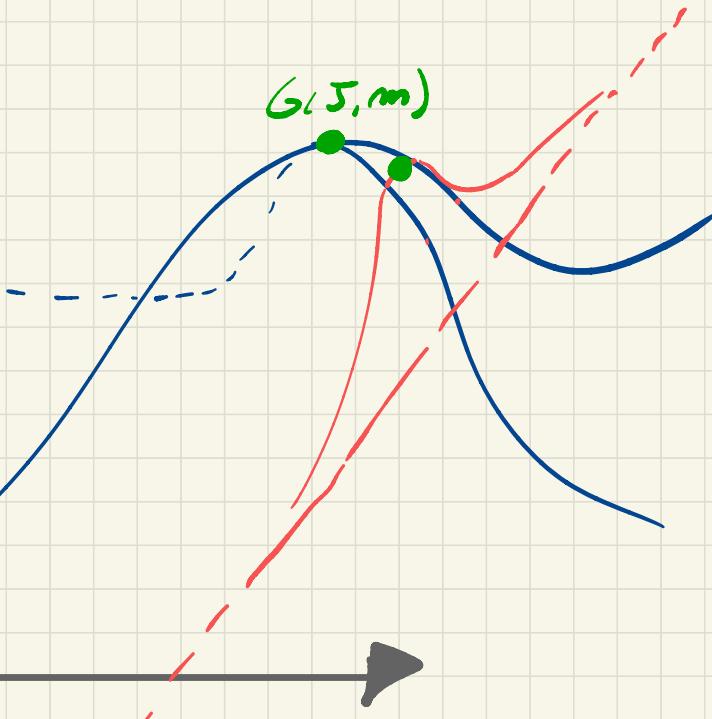
WHICH ALSO DESCRIBES "FLUCTUATIONS"

[ IT'S A BETTER APPROX. WHY? LATER ].

$\mathcal{S}[9]$



$G(J, m)$



## TREE ISING

IDEA: IF  $\Sigma$  DESCRIBES A TREE (OR FOREST), i.e. IF THERE ARE NO LOOPS, THEN:

$$P_T = \overline{\prod_{i \in V_T} P_i(\gamma_i)} \overline{\prod_{(i,j) \in E_T} \frac{P_{ij}(\gamma_i, \gamma_j)}{P_i(\gamma_i) P_j(\gamma_j)}}$$

$$= \overline{\prod_{(i,j) \in E_T} P_{ij}(\gamma_i, \gamma_j)} \overline{\prod_{i \in V_T} (P_i(\gamma_i))^{1 - |\Sigma_i|}}$$

$$|\Sigma_i| = \#\{j \mid \Sigma_{ij} \neq 0\}$$

FACT  $D_{KL}(P^D || P)$  IS MINIMIZED WHEN  $P_i = P_i^D$ ,  $P_{ij} = P_{ij}^D$

$$\text{SO: } \min_{P_i, P_{ij}} D_{KL}(P^D || P) = -H^D + \sum_{i \in V_T} H_i - \sum_{(i,j) \in E_T} I_{ij}$$

$$I_{ij} = \sum_{\gamma_i, \gamma_j} P_{ij}^D(\gamma_i, \gamma_j) \log \frac{P_{ij}^D(\gamma_i, \gamma_j)}{P_{ij}^D(\gamma_i) P_{ij}^D(\gamma_j)}$$

IN  $D_{KL}$ , WE MINIMIZE OVER THE LAST TERM AMONG  
TREE TOPOLOGIES.

$$E_{T_{\text{opt}}} = \underset{E_T}{\operatorname{argmin}} - \sum_{(i,j) \in E_T} I_T.$$

SINCE  $I_{i,j} \geq 0$ , I WILL GET A CONNECTED GRAPH

I USE KRUSKAL\PRIM FOR MST POLYTIME.

ONCE I HAVE  $T_{\text{opt}}$  IT IS EASY TO GET  $P_i, P_{i,j}$ .

WILSON...

# VARIATIONAL METHODS + BP

SAME STEPS OF MF, BUT EVEN LESS FORMAL.

DEFINE  $\tilde{P}_i(\theta_i) = \frac{1 + \tilde{m}_i(\theta_i)}{2}$

$$\tilde{P}_{ij}(\theta_i, \theta_j) = \frac{(1 + \tilde{m}_i(\theta_i))(1 + \tilde{m}_j(\theta_j)) + \tilde{C}_{ij}(\theta_i, \theta_j)}{4}$$

WHERE:  $-1 \leq \tilde{m}_i \leq 1$

$$-1 + |\tilde{m}_i + \tilde{m}_j| \leq \tilde{C}_{ij} + \tilde{m}_i \tilde{m}_j \leq 1 - |\tilde{m}_i - \tilde{m}_j|$$

$$\left\{ \begin{array}{l} P_{\theta_i}^{BP} = \prod_{j \in \text{children}(i)} \frac{\tilde{P}_{ij}(\theta_i, \theta_j)}{\tilde{P}_i(\theta_i) \tilde{P}_j(\theta_j)} \end{array} \right\} = BP$$

$$\underline{G}^* = BP \cap \left\{ q : \langle \sigma_i \rangle = m_i \right\} \rightarrow \text{GET CLOSED FORM, TAKE DERIVATIVE}$$

$P^{BP}$  IS NORMALIZED ONLY ON TREES.

## CAVITY METHODS

IT IS A TRICK TO CALCULATE  $Z_{(j,k)}^n$ ,  $\langle \sigma_j \rangle^n$ ,  $\langle \sigma_i \sigma_j \rangle^n$  EFFICIENTLY ON TREES

CAN BE EXTENDED TO GRAPHS. [SUBSCEPTIBILITY]

$Z_{i-\rightarrow j}(\alpha_i)$  IS THE PARTITION FUNCTION OF THE PART OF THE SYSTEM CONTAINING  $i$  WHERE  $(i,j)$  IS BROKEN. AND  $i$  IS CONSTRAINED TO BE  $\alpha_i$ .

$$Z_{i-\rightarrow j}(\alpha_i) = e^{h_i \alpha_i} \prod_{k \in \delta(i) \setminus j} \left[ \sum_{\alpha_k} Z_{k-\rightarrow i}(\alpha_k) e^{J_{ik} \alpha_i \alpha_k} \right]$$

FAST IF YOU START FROM LEAVES.

$$Z_i(\alpha_i) = e^{h_i \alpha_i} \prod_{k \in \delta(i)} \left[ \sum_{\alpha_k} Z_{k-\rightarrow i}(\alpha_k) e^{J_{ik} \alpha_i \alpha_k} \right] \text{PARTITION FUNCTION WITH } i \text{ CONSTRAINED ON } \alpha_i.$$

YOU CAN COMPUTE  $P_i(\alpha_i)$  BY NORMALIZING  $P_i(\alpha_i) \propto Z_i(\alpha_i)$

## ADAPTIVE CLUSTER EXPANSION

APPROXIMATE THE ENTROPY  $S(X, m)$  AND TAKE DERIV.

IF WE HAD A SINGLE SPIN,  $S_{\text{one}}(m_i) = \sum_{\sigma_i} \frac{1+m_i \sigma_i}{2} \log \frac{1+m_i \sigma_i}{2}$

IF WE HAD TWO SPINS,

$$S_{\text{two}}(m_i, m_j, \chi_{ij}) = \sum_{\sigma_i, \sigma_j} \frac{1+m_i \sigma_i + m_j \sigma_j + \chi_{ij} \sigma_i \sigma_j}{4} + \log \frac{1+m_i \sigma_i + m_j \sigma_j + \chi_{ij} \sigma_i \sigma_j}{4}$$

IF  $i \perp j$ ,  $S_{\text{two}} = S_{\text{one}}(m_i) + S_{\text{one}}(m_j)$

$$\Delta S_{\text{two}}(m_i, m_j, \chi_{ij}) := S_{\text{two}} - S_{\text{one}}(m_i) - S_{\text{one}}(m_j)$$

DEFINE ACCORDINGLY  $\Delta S_m = S_m$

$$\Delta S_{\text{two}} = [\text{TRIPLET-COUPLES} + \text{SINGLETORS}]$$

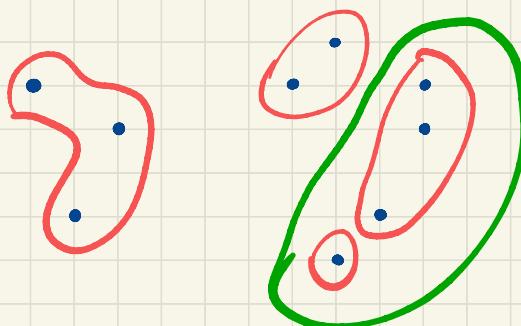
$$\text{IDEA } S(X, m) \approx \sum_i \Delta S_{m,i}(m_i) + \sum_{(i,j)} \Delta S_{m,i}(m_i, m_j, X_{i,j}) + \dots$$

FIX A THRESHOLD  $\delta$ .

ADD A CLUSTER IF ITS  $\Delta S > \delta$ .

STOP.

YOU GET  $\tilde{S}$ , TAKE DERIVATIVES, FIND  $\tilde{J}$ ,  $\tilde{h}$ .



## PLEFKA'S EXPANSION

SIMILAR TO PERTURBATION TRICKS

LET  $H(\mathbf{s}) = H_0(\mathbf{s}) + \lambda V(\mathbf{s})$

$$H_0(\mathbf{s}) = - \sum_i h_i s_i$$

$$V(\mathbf{s}) = - \sum_{i,j} J_{ij} s_i s_j$$

$$F(\lambda) = F^{(0)} - \lambda \langle V \rangle_0 + \frac{\lambda^2}{2} [\langle V^2 \rangle_0 - \langle V \rangle_0^2] + \dots$$

→ TAYLOR EXPANSION OF  $F(\lambda)$  AROUND ZERO.

$$F^{(0)} = \sum_i 2 \cosh h_i$$

LEGENDRE TRANSFORM TO GET

$$G^* = G^{(0)} + \lambda G^{(1)} + \dots + \lambda^2 G^{(2)} \quad [ G^* = \sum_i h_i m_i + F(\lambda), \quad m_i = -\frac{\partial F}{\partial h_i}(\lambda) ]$$

USE  $h^{\infty} = h^{(0)} + \lambda h^{(1)} + \dots + \lambda^2 h^{(2)} + \dots$

INTO  $m_i = -\frac{\partial F}{\partial h_i}(S, h)$  TO FIND  $h^{(1)}$  AS A FUNCTION OF  $m_i, S$ .

$$\rightsquigarrow G^{(0)}(S, m) = \sum_i \frac{1+m_i}{2} \log \frac{1+m_i}{2} + \frac{1-m_i}{2} \log \frac{1-m_i}{2}$$

$$G^{(1)}(S, m) = -\sum_{i \neq j} S_{ij} m_i m_j;$$

$$G^{(2)}(S, m) = \sum_{i \neq j} S_{ij}^2 (1-m_i^2)(1-m_j^2)$$

IMPOSE OBSERVATIONS AND TAKE DERIVATIVES.

# PSEUDO LIKELIHOOD I

$$H(\theta) = H_i(\theta_i) + H_{\setminus i}(\theta \setminus \theta_i)$$

$$= -h_i(\theta_i) - \theta_i \sum_{j \neq i} S_{ij} \theta_j + H_{\setminus i}(\theta \setminus \theta_i)$$

$$Z(S, \theta) = \prod_{i \in S} 2 \cos \left[ \theta_i + \sum_{j \neq i} S_{ij} \theta_j \right] e^{-H_i(\theta \setminus \theta_i)}$$

→  $\langle \sigma_i \rangle = \langle \tanh \left[ \theta_i + \sum_{k \neq i} S_{ik} \sigma_k \right] \rangle$

$$\langle \sigma_i \sigma_j \rangle = \langle \sigma_i \tanh \left[ \theta_i + \sum_{k \neq i} S_{ik} \sigma_k \right] \rangle$$

IDEA: IMPOSE  $\langle \sigma_i \rangle^D = \langle \tanh \left[ \theta_i^D + \sum_{k \neq i} S_{ik}^D \sigma_k \right] \rangle^D$

$$\langle \sigma_i \sigma_j \rangle^D = \langle \sigma_j \tanh \left[ \theta_i^D + \sum_{k \neq i} S_{ik}^D \sigma_k \right] \rangle^D$$

GRADIENT DESCENT

SOLVE FOR  $m_i^{PL}$ ,  $\sigma_{ij}^{PL}$ ,  $\sigma_{ji}^{PL}$

$$\sigma_{ij}^{PL} \leftarrow \frac{1}{2} (\sigma_{ij}^{PL} + \sigma_{ji}^{PL}).$$

## PSEUDOLOGLIKELIHOOD II

NOTICE THAT  $P(\gamma_i | (\gamma_j)_{j \neq i}) = \frac{1}{2} [1 + \text{TH}[m_i + \sum_{j \neq i} \sigma_{ij} \gamma_j]]$

$$= \frac{1}{1 + \exp[m_i + \sum_{j \neq i} \sigma_{ij} \gamma_j]}$$

$$\mathcal{L}_D^i(\sigma_{ij}, m_i) = \frac{1}{M} \sum_{\mu} \log P(\gamma_i^{\mu} | (\gamma_j^{\mu})_{j \neq i})$$

$$= \frac{1}{M} \sum_{\mu} \log \frac{1}{2} [1 + \text{TH}[m_i + \sum_{j \neq i} \sigma_{ij} \gamma_j^{\mu}]]$$

SETTING DERIVATIVES TO ZERO RECOVERS \*

BUT WE CAN WRITE  $L^P(L, S, M) = \sum_i L_i^P(S_{i,i}, M_i)$   
AND WE CAN MAXIMIZE OVER  $S, M$ .

IF WE FORCE  $S$  SYMMETRIC, HARD.

## RESULTS

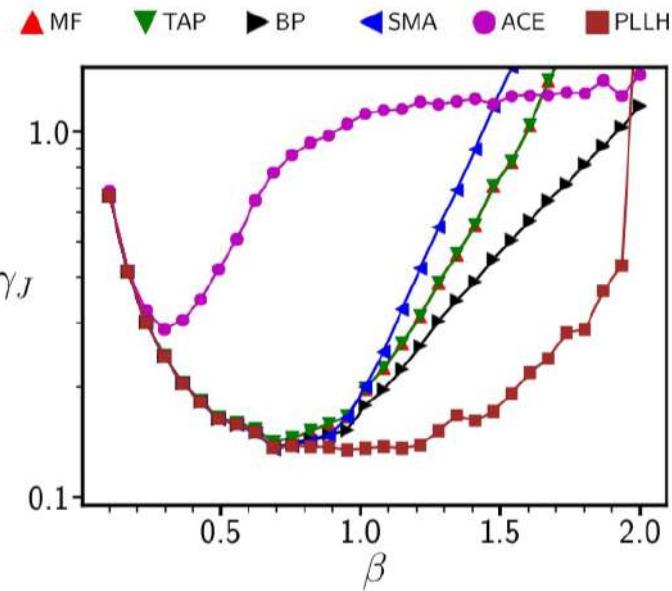
GROUND

TRUTH

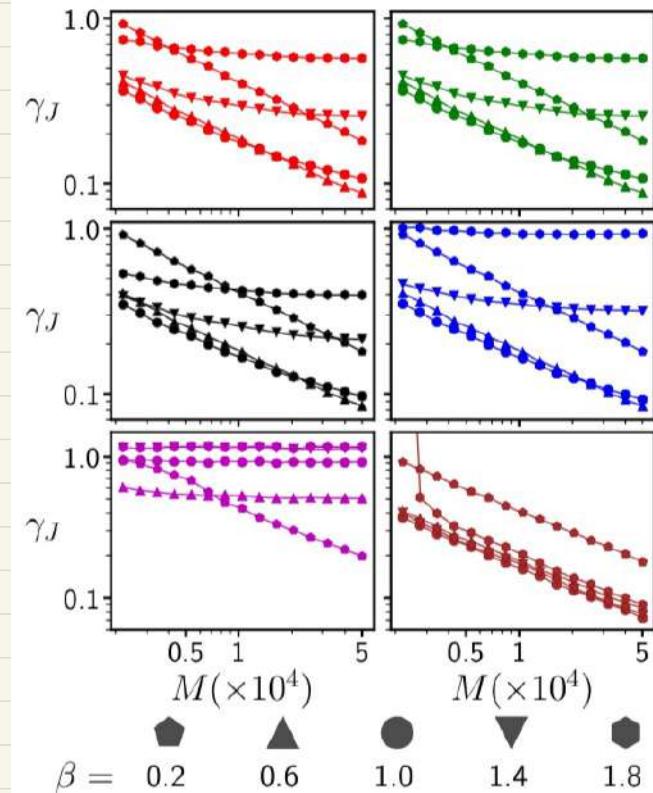
- FULLY CONNECTED GRAPH  $J_{ij}^o \sim N(0, \beta/\sqrt{n})$  I
- LONG LOOPS [RANDOM GRAPH,  
FIXED DEGREE]  $J_{ij}^o \sim \text{UNIF}[-\beta, \beta]$  II
- SHORT LOOPS [2D SQUARE  
LATTICE]  $J_{ij}^o \sim \text{UNIF}[-\beta, \beta]$ . III

$$h_i^o \sim \text{UNIF}[-0.3\beta, 0.3\beta]$$

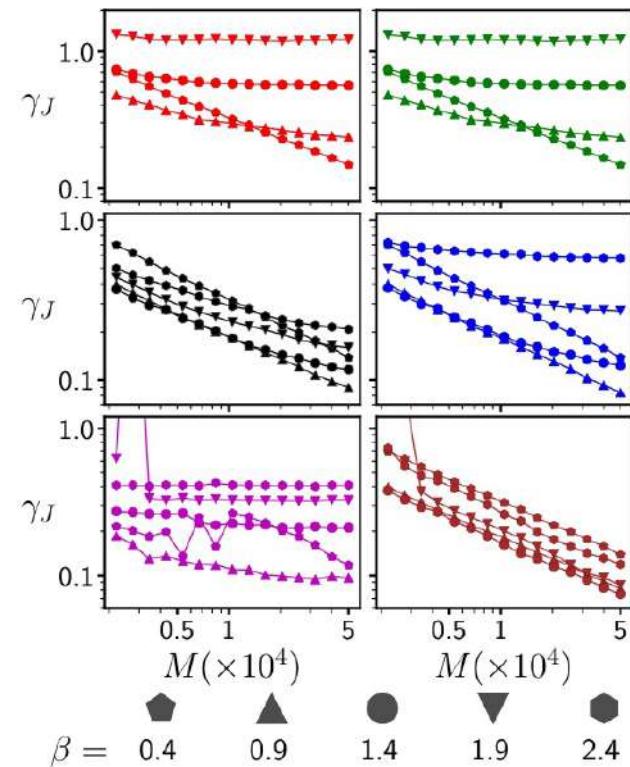
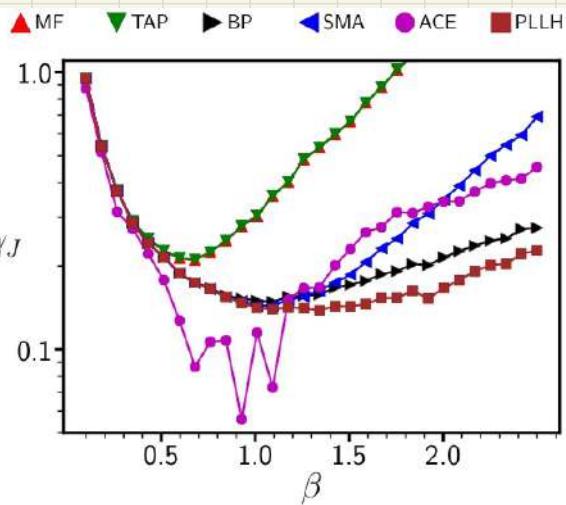
工



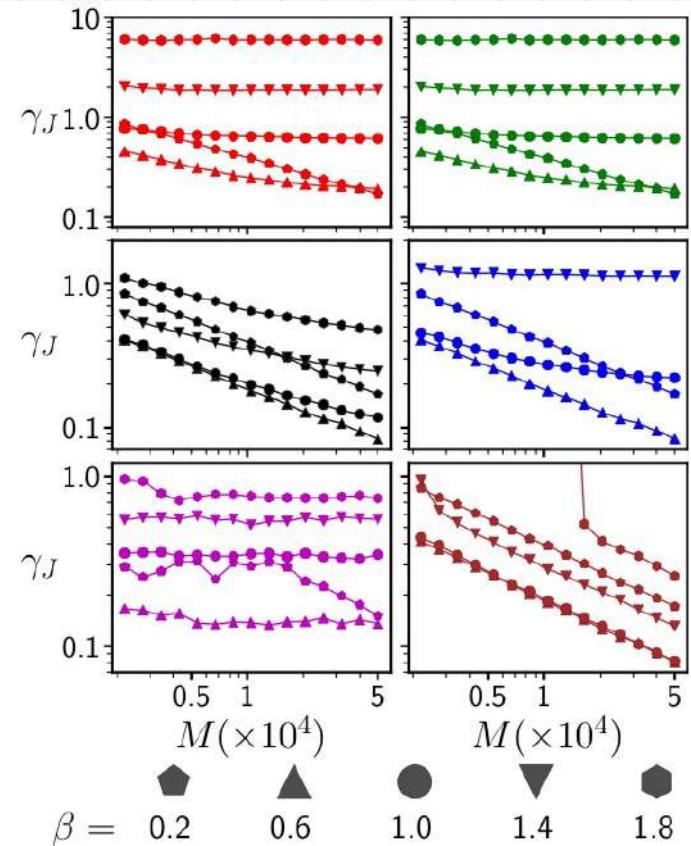
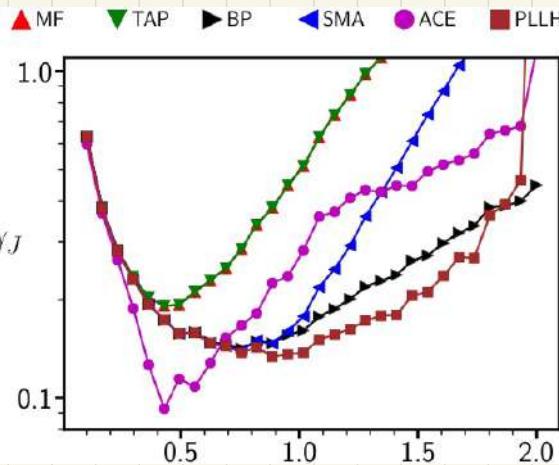
$$\gamma_j = \sqrt{\frac{\sum_{i \neq j} (\beta_{ij} - \beta_{ij}^*)^2}{\sum_{i \neq j} (\beta_{ij}^*)^2}}$$



II



III



# 3<sup>RD</sup> CHAPTER INVERSE PROBLEM [NOT EQUILIBRIUM]

IN A BOLTZMANN,  $\sigma_{ij} = \sigma_{ji}$ .

I CAN START A GLAUBER THAT SATISFIES

DETAILED BALANCE  $\Gamma_{ji} P_i = \Gamma_{ij} P_j$

CONVERGES TO EQUILIBRIUM  $(\dot{P}_i)^{(e)} = 0$ .

IF  $\sigma_{ij} \neq \sigma_{ji}$ , GLAUBER WILL LEAD TO A NESS:

$$(\dot{P}_i)^{(e)} = 0, \quad \Gamma_{ji} P_i \neq \Gamma_{ij} P_j$$

## GLAUBER I

- PICK RANDOM i.

- $P(\delta_i(t+1) | \delta_i(t)) = \frac{\exp[\beta_i(t+1)\theta_i(t)]}{2\cosh\theta_i(t)}, \quad \theta_i(t) = \sum_j J_{ij} \delta_j(t) + m_i$ .

IF  $J_{ij} = J_{ji}$  WE CONVERGE TO BOLTZMANN [THERE VELOCITY RESULTS].

IF  $J_{ij} \neq J_{ji}$  WE CONVERGE TO A NESS. DIFFICULT TO CHARACTERIZE (NO CLOSED FORM).

IF I PICK i,  $m_i(t+1)\delta_i(t) = \langle \delta_i(t+1) \rangle \delta_i(t) = T H(\theta_i(t))$ .

IF I DO NOT PICK i,  $m_i(t+1)\delta_i(t) = \delta_i(t)$ .

IN NESS

$$m_i = \langle \sigma_i \rangle = \langle T_H \theta_i \rangle = \langle T_H(\sum_j S_{ij} \sigma_j + n_i) \rangle$$

$$\chi_{ij} = \langle \sigma_i \sigma_j \rangle = \frac{1}{2} \langle \sigma_i T_H(\theta_j) \rangle + \frac{1}{2} \langle \sigma_j T_H(\theta_i) \rangle.$$

$$\phi_{ij} = \langle \sigma_i (\epsilon + 1) \sigma_j (\epsilon) \rangle = \frac{1}{N} \langle T_H(\theta_i) \sigma_j \rangle + \frac{N-1}{N} \chi_{ij}$$

## GLAUBER II

- ALL SPIN VARIABLES CAN CHANGE AT  $t+1$ :

$$\bullet P(\theta_t | \theta_{t+1}) / P(\theta_t) = \frac{\exp \left[ \sum_i \delta_i (\theta_{t+1}) \theta_i (\epsilon) \right]}{\prod_i 2 \cosh \theta_i (\epsilon)} .$$

NESS

$$m_i = \langle \tanh \theta_i \rangle$$

$$\chi_{ij} = \langle \tanh \theta_i \rangle$$

$$\phi_{ij} = \langle \tanh(\theta_i) \theta_j \rangle .$$

[ NESS II  $\neq$  NESS I. ]

## TIMES SERIES LIKELIHOOD

FOR GLAUBER II,

$$L_D(S, \theta) = \frac{1}{M} \sum_{i=1}^{M-1} \sum_j \left\{ g_i(\epsilon + \gamma) \theta_j(\epsilon) - \log 2 \cosh \theta_j(\epsilon) \right\}$$

↳  $\frac{\partial L_D}{\partial \theta_i}(S, \theta) = \frac{1}{M} \sum_{i=1}^{M-1} [g_i(\epsilon + \gamma) - \text{TH} \theta_i(\epsilon)]$

$$\frac{\partial L_D}{\partial \theta_j}(S, \theta) = \frac{1}{M} \sum_{i=1}^{M-1} [g_i(\epsilon + \gamma) / g_j(\epsilon) - \text{TH} \theta_i(\epsilon) g_j(\epsilon)]$$

CAN BE EVALUATE IN POLYNOMIAL TIME [GRADIENT ASCENT SAME]

## MEAN FIELD IN NESS

P IS A NESS OF A GLAUBER II.

LET SUPPOSE WE ARE ALREADY IN NESS P.

LET  $q_{(18)} = \overline{\prod_i \frac{1+m_i \beta_i}{2}}$  BE CLOSE TO P..

q IS THE NESS OF A DIFFERENT GLAUBER II:

$$m_i^q = \operatorname{arctanh} m_i;$$

$$\beta_{ij}^q = 0.$$

$$m_i^q = \langle \tanh \left( \sum_j \beta_{ij} \sigma_j + h_i \right) \rangle$$

IDEA: PERTURB THESE AND SEE HOW GLAUBER CHANGES  
 [BY TAKING DERIVATIVES OF  $m_i = \langle \text{TH}(\sum_j J_{ij} \sigma_j + h_i) \rangle$ ]

$$\begin{aligned}\Delta m_i &= \sum_j \frac{\partial \langle \sigma_j \rangle}{\partial h_i} \Big|_0 \Delta h_j + \sum_{k,j} \frac{\partial \langle \sigma_j \rangle}{\partial J_{kj}} \Big|_0 \Delta J_{kj} + \dots \\ &= (1 - m_i^2) \Delta h_i + (1 - m_i^2) \sum_j \Delta J_{ij} m_j + \dots\end{aligned}$$

AND IMPOSE  $\Delta h_i = h_i - h_i^q$

$$\Delta J_{ij} = J_{ij} - J_{ij}^q = J_{ij}.$$

IMPOSE  $\Delta m_i = 0$  [WE WERE ALREADY CLOSE]

$$h_i^q = h_i + \sum_j J_{ij} m_j \longrightarrow m_i = \text{TH}(h_i + \sum_j J_{ij} m_j). \quad \clubsuit$$

THEN, DO THE SAME PROCEDURE FOR  $\phi_{ij}$

$$\rightarrow \phi_{ij} = m_i m_j + (1 - m_i^2) \sum_k J_{ik} / (\chi_{di} - m_j m_i)$$

IN 8, 20 IMPOSE  $m$  AND  $\chi$  ARE THE SAMPLE AVERAGES OR NESS...

SOLVE:  $m^P, J^M$ .

## RESULTS

GROUND

TRUTH

$$J_{ij}^o \sim N(0, \beta/\sqrt{n})$$

$$J_{ji}^o \sim N(0, \beta/\sqrt{n})$$

$$J_{ij}^o \perp J_{ji}^o$$

$$h_i^o \sim \text{unif}[-0.3\beta, 0.3\beta].$$

