ISING

$$\max S = -\sum_{k=1}^{2^m} P(\underset{\sim}{z}_k)\log P(\underset{\sim}{z}_k)$$

s.t.
$$\sum_{k=1}^{2^m} P(\underset{\sim}{z}_k) = 1$$

$$\langle z_i \rangle = \sum_{k=1}^{2^m} z_i^k P(\underset{\sim}{z}_k) = P_i^{obs}$$

$$\langle z_i z_j \rangle = \sum_{k=1}^{2^m} z_i^k z_j^k P(\underset{\sim}{z}_k) = P_{ij}^{obs}$$

$$F = -\sum_{k=1}^{2^m} P(\underset{\sim}{z}_k)\log P(\underset{\sim}{z}_k) + \alpha\left(1 - \sum_{k=1}^{2^m} P(\underset{\sim}{z}_k)\right)$$
$$+ \sum_{i=1}^{m}\left[h_i \left(P_i^{obs} - \sum_{k=1}^{2^m} z_i^k P(\underset{\sim}{z}_k)\right)\right]$$
$$+ \sum_{i=1}^{m}\sum_{j=i+1}^{m}\left[J_{ij}\left(P_{ij}^{obs} - \sum_{k=1}^{2^m} z_i^k z_j^k P(\underset{\sim}{z}_k)\right)\right]$$

$$\frac{\partial F}{\partial P(\underset{\sim}{z}_k)} = 0 = -\log P(\underset{\sim}{z}_k) - 1 + \alpha - \sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k$$

$$\Rightarrow P(\underset{\sim}{z}_k) = \frac{\exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right)}{\exp(1-\alpha)}$$

1) $$\sum_{k=1}^{2^m} P(\underset{\sim}{z}_k) = \frac{\sum_{k=1}^{2^m}\exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right)}{\exp(1-\alpha)} = 1$$

$$\Rightarrow \exp(1-\alpha) = \sum_{k=1}^{2^m}\exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right) = Z$$

$$\therefore P(\underset{\sim}{z}_k) = \frac{\exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right)}{Z}$$

2) $$\langle z_i \rangle = \frac{\sum_{k=1}^{2^m} z_i^k \exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right)}{Z} = \sum_{k=1}^{2^m} z_i^k P(\underset{\sim}{z}_k) = P_i^{obs}$$

3) $$\langle z_i z_j \rangle = \frac{\sum_{k=1}^{2^m} z_i^k z_j^k \exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right)}{Z} = \sum_{k=1}^{2^m} z_i^k z_j^k P(\underset{\sim}{z}_k) = P_{ij}^{obs}$$

NEWTON

①

$$\log L(\{h_i^*, J_{ij}^*\}|MSA) = \log L(\{h_i, J_{ij}\}|MSA) + \frac{\partial \log L(\{h_i, J_{ij}\}|MSA)}{\partial\{h_i, J_{ij}\}}$$
$$\times\left(\{h_i^*, J_{ij}^*\} - \{h_i, J_{ij}\}\right)$$
$$+ \left(\{h_i^*, J_{ij}^*\} - \{h_i, J_{ij}\}\right)^T \frac{\partial^2 \log L(\{h_i, J_{ij}\}|MSA)}{\partial\{h_i, J_{ij}\}^2}\left(\{h_i^*, J_{ij}^*\} - \{h_i, J_{ij}\}\right)$$
$$+ h.o.t.$$

$$\Rightarrow \left.\frac{\partial \log L(\{h_i^*, J_{ij}^*\}|MSA)}{\partial\{h_i, J_{ij}\}}\right|_{\{h_i^*, J_{ij}^*\}} = 0$$

$$= \frac{\partial \log L(\{h_i, J_{ij}\}|MSA)}{\partial\{h_i, J_{ij}\}}$$
$$+ \frac{\partial^2 \log L(\{h_i, J_{ij}\}|MSA)}{\partial\{h_i, J_{ij}\}^2}\left(\{h_i^*, J_{ij}^*\} - \{h_i, J_{ij}\}\right)$$
$$+ h.o.t.$$

$$\Rightarrow \{h_i^*, J_{ij}^*\} = \{h_i, J_{ij}\} - \frac{\dfrac{\partial \log L(\{h_i, J_{ij}\}|MSA)}{\partial\{h_i, J_{ij}\}}}{\dfrac{\partial^2 \log L(\{h_i, J_{ij}\}|MSA)}{\partial\{h_i, J_{ij}\}^2}}$$
$$+ h.o.t.$$

$$\Rightarrow \{h_i, J_{ij}\}^{n+1} = \{h_i, J_{ij}\}^{n} - \gamma\,\underset{\sim}{H}^{-1}\underset{\sim}{J}$$

iteration

softening parameter

ISING

observations in MSA, $p^{obs}(z_k) K = n^{obs}(z_k)$

$$L\left(\{h_i, J_{ij}\} \mid \overset{MSA}{\cancel{\{p_i^{obs}, p_{ij}^{obs}\}}}\right) = \prod_{k=1}^{2^m} P(z_k)^{p^{obs}(z_k) K}$$

$$\parallel$$

$$P(MSA \mid \{h_i, J_{ij}\})$$

$$\frac{1}{K} \log L\left(\{h_i, J_{ij}\} \mid \overset{MSA}{\cancel{\{p_i^{obs}, p_{ij}^{obs}\}}}\right) = \sum_{k=1}^{2^m} p^{obs}(z_k) \log P(z_k) = \sum_{k=1}^{2^m} p^{obs}(z_k)\left[-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k - \log Z\right]$$

$$= \sum_{k=1}^{2^m} p^{obs}(z_k)\left[-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right] - \log Z$$

$$\parallel$$

$$\mathcal{L}$$

$$\frac{\partial \mathcal{L}}{\partial h_r} = -\frac{\partial \log Z}{\partial h_i} - \sum_{k=1}^{2^m} p^{obs}(z_k) z_r^k$$

$$= -p_r^{obs} - \frac{1}{Z}\frac{\partial}{\partial h_r}\sum_{k=1}^{2^m} \exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right)$$

$$= -p_r^{obs} + \frac{1}{Z}\sum_{k=1}^{2^m} z_r^k \exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right)$$

$$= \langle z_r \rangle - p_r^{obs}$$

$$\frac{\partial \mathcal{L}}{\partial J_{rs}} = -\frac{\partial \log Z}{\partial J_{rs}} - \sum_{k=1}^{2^m} p^{obs}(z_k) z_r^k z_s^k$$

$$= -p_{rs}^{obs} + \frac{1}{Z}\sum_{k=1}^{2^m} z_r^k z_s^k \exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right)$$

$$= \langle z_r z_s \rangle - p_{rs}^{obs}$$

$$\frac{\partial^2 \mathcal{L}}{\partial h_r \partial h_s} = \frac{-Z\sum_{k=1}^{2^m} z_r^k z_s^k \exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right) + \sum_{k=1}^{2^m} z_s^k \exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij} z_i^k z_j^k\right) \cdot \sum_{k=1}^{2^m} z_r^k \exp\left(-\sum_{i=1}^{m} h_i z_i^k - \sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij}\right.}{Z^2}$$

$$= \langle z_r \rangle \langle z_s \rangle - \langle z_r z_s \rangle = \frac{\partial^2 \mathcal{L}}{\partial h_s \partial h_r}$$

$$\frac{\partial^2 \mathcal{L}}{\partial J_{rs} \partial h_t} = \langle z_r z_s \rangle \langle z_t \rangle - \langle z_r z_s z_t \rangle = \frac{\partial^2 \mathcal{L}}{\partial h_t \partial J_{rs}}$$

$$\frac{\partial^2 \mathcal{L}}{\partial J_{rs} \partial J_{uv}} = \langle z_r z_s \rangle \langle z_u z_v \rangle - \langle z_r z_s z_u z_v \rangle = \frac{\partial^2 \mathcal{L}}{\partial J_{uv} \partial J_{rs}}$$

$$P(\{\mathbf{h}_i, \mathbf{J}_{ij}\} \mid MSA) = \frac{P(MSA \mid \{\mathbf{h}_i, \mathbf{J}_{ij}\})\, P(\{\mathbf{h}_i, \mathbf{J}_{ij}\})}{P(MSA)} \qquad \text{(Bayes)}$$

$$P(\mathbf{s}_k \mid \{\mathbf{h}_i, \mathbf{J}_{ij}\}) = \frac{1}{Z} \exp\left(-\sum_{i=1}^{m} \mathbf{h}_i^T \mathbf{z}_i^{(k)} - \sum_{i=1}^{m}\sum_{j=i+1}^{m} \mathbf{z}_i^{(k)T} \mathbf{J}_{ij} \mathbf{z}_j^{(k)}\right)$$

sequence, elements $s_k^i \in [0, Q]$, where Q is alphabet size

Indicator vector of length Q (where Q is size of alphabet) containing one 1 and $(Q-1)$ zeros indicating identity of aa residue in position i
$\mathbf{z}_i(\alpha) = 1$ if $s_i = \alpha$, $= 0$ else

$$Z = \sum_{k=1}^{Q^m} \exp\left(-\sum_{i=1}^{m} \mathbf{h}_i^T \mathbf{z}_i^{(k)} - \sum_{i=1}^{m}\sum_{j=i+1}^{m} \mathbf{z}_i^{(k)T} \mathbf{J}_{ij} \mathbf{z}_j^{(k)}\right)$$

$$P(\{\mathbf{h}_i, \mathbf{J}_{ij}\}) = \prod_{i=1}^{m}\prod_{q=1}^{Q} \exp(-\lambda_h \|h_i^q\|) \prod_{i=1}^{m}\prod_{j=i+1}^{m}\prod_{q=1}^{Q}\prod_{r=1}^{Q} \exp(-\lambda_J \|J_{ij}^{qr}\|)$$

$$P(MSA \mid \{\mathbf{h}_i, \mathbf{J}_{ij}\}) = \prod_{k=1}^{Q^m} P(\mathbf{s}_k)^{p^{obs}(\mathbf{s}_k) K}$$

observations in MSA, $p^{obs}(\mathbf{z}_k) K = n^{obs}(\mathbf{s}_k)$

$$\therefore\ P(\{\mathbf{h}_i, \mathbf{J}_{ij}\} \mid MSA)\, P(MSA) = \prod_{k=1}^{Q^m} P(\mathbf{s}_k)^{p^{obs}(\mathbf{s}_k) K} \cdot \prod_{i=1}^{m}\prod_{q=1}^{Q} e^{-\lambda_h \|h_i^q\|} \cdot \prod_{i=1}^{m}\prod_{j=i+1}^{m}\prod_{q=1}^{Q}\prod_{r=1}^{Q} e^{-\lambda_J \|J_{ij}^{qr}\|}$$

$$\Rightarrow \log\left(P(\{\mathbf{h}_i, \mathbf{J}_{ij}\} \mid MSA)\, P(MSA)\right) = K \sum_{k=1}^{Q^m} p^{obs}(\mathbf{s}_k) \log P(\mathbf{s}_k) - \lambda_h \sum_{i=1}^{m}\sum_{q=1}^{Q} \|h_i^q\| - \lambda_J \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{q=1}^{Q}\sum_{r=1}^{Q} \|J_{ij}^{qr}\|$$

$$\Rightarrow \mathcal{L}(\{\mathbf{h}_i, \mathbf{J}_{ij}\}) = \frac{1}{K} \log\left(P(\{\mathbf{h}_i, \mathbf{J}_{ij}\} \mid MSA)\, P(MSA)\right) = \sum_{k=1}^{Q^m} P^{obs}(\mathbf{s}_k) \log P(\mathbf{s}_k) - \frac{\lambda_h}{K} \sum_{i=1}^{m}\sum_{q=1}^{Q} \|h_i^q\| - \frac{\lambda_J}{K} \sum_{i=1}^{m}\sum_{j=i+1}^{m}\sum_{q=1}^{Q}\sum_{r=1}^{Q} \|J_{ij}^{qr}\|$$

$$\Rightarrow \mathcal{L}(\{\underset{\sim}{h}_i, \underset{\approx}{J}_{ij}\}) = \sum_{k=1}^{Q^m} P^{obs}(\underset{\sim}{s}_k)\left[-\sum_{i=1}^{m} \underset{\sim}{h}_i^T \underset{\sim}{z}_i^{(k)} - \sum_{i=1}^{m}\sum_{j=i+1}^{m} \underset{\sim}{z}_i^T \underset{\approx}{J}_{ij} \underset{\sim}{z}_j^{(k)}\right] - \log Z$$

$$- \frac{\lambda_h}{K}\sum_{i=1}^{m}\sum_{q=1}^{Q} \| h_i^q \| - \frac{\lambda_J}{K}\sum_{i=1}^{m}\sum_{j=i+1}^{m}\sum_{q=1}^{Q}\sum_{r=1}^{Q} \| J_{ij}^{qr} \| \qquad \sum_{k=1}^{Q^m} P^{obs}(\underset{\sim}{s}_k) = 1$$

$$\frac{\partial \mathcal{L}}{\partial h_l^\alpha} = -\frac{1}{Z}\frac{\partial Z}{\partial h_l^\alpha} - \sum_{k=1}^{Q^m} P^{obs}(\underset{\sim}{s}_k)\, z_l^{\alpha(k)} - \frac{\lambda_h}{K}\left[\operatorname{sgn}(h_l^\alpha)\right] \quad \text{for } \|h_i^q\| = |h_i^q|$$

$$- \frac{\lambda_h}{K}\left[2h_l^\alpha\right] \quad \text{for } \|h_i^q\| = (h_i^q)^2$$

$$= \frac{\sum_{k=1}^{Q^m} z_l^{\alpha(k)} \exp\left(-\sum_{i=1}^{m} h_i^T z_i^{(k)} - \sum_{i=1}^{m}\sum_{j=i+1}^{m} \underset{\sim}{z}_i^{(k)T} \underset{\approx}{J}_{ij} \underset{\sim}{z}_j^{(k)}\right)}{Z} - \sum_{k=1}^{Q^m} P^{obs}(\underset{\sim}{s}_k)\, z_l^{\alpha(k)} - \frac{\lambda_h}{K} g(h_l^\alpha)$$

$$g(h_l^\alpha) = \operatorname{sgn}(h_l^\alpha) \text{ for } L1 \text{ reg}$$
$$= 2h_l^\alpha \text{ for } L2 \text{ reg}$$

$$= \langle z_l^\alpha \rangle - P^{obs}(z_l^\alpha) - \frac{\lambda_h}{K} g(h_l^\alpha)$$

model probability of residue α in position l

observed —— " ——

$$\frac{\partial \mathcal{L}}{\partial J_{ls}^{\alpha\delta}} = -\frac{1}{Z}\frac{\partial Z}{\partial J_{ls}^{\alpha\delta}} - \sum_{k=1}^{Q^m} P^{obs}(\underset{\sim}{s}_k)\, z_l^{\alpha(k)} z_s^{\delta(k)} - \frac{\lambda_J}{K} g(J_{ls}^{\alpha\delta})$$

$$= \langle z_l^\alpha z_s^\delta \rangle - P^{obs}(z_l^\alpha z_s^\delta) - \frac{\lambda_J}{K} g(J_{ls}^{\alpha\delta})$$

$$\frac{\partial^2 \mathcal{L}}{\partial h_l^\alpha \partial h_s^\delta} = \frac{-\sum_{k=1}^{Q^m} z_l^{\alpha(k)} z_s^{\delta(k)} \exp(\sim) . Z + \sum_{k=1}^{Q^m} z_s^{\delta(k)} \exp(\sim) . \sum_{k=1}^{Q^m} z_l^{\alpha(k)} \exp(\sim)}{Z^2}$$

$$- \frac{\lambda_h}{K} \frac{\partial g(h_l^\alpha)}{\partial h_s^\delta}$$

$$= \langle z_l^\alpha \rangle \langle z_s^\delta \rangle - \langle z_l^\alpha z_s^\delta \rangle - \frac{\lambda_h}{K} \frac{\partial g(h_l^\alpha)}{\partial h_s^\delta}$$

$\frac{\partial g(h_l^\alpha)}{\partial h_s^\delta} = 0$ for $\alpha \neq \delta$ & $s \neq l$

$= 0$ for $\alpha = \delta$ & $s = l$ & L1 reg
$\boxed{\text{BUT}}$ undefined at $h_l^\alpha = 0$

$= 2$ for $\alpha = \delta$ & $s = l$ & L2 reg

$$\frac{\partial^2 \mathcal{L}}{\partial h_l^\alpha \partial J_{st}^{\delta\sigma}} = \langle z_l^\alpha \rangle \langle z_s^\delta z_t^\sigma \rangle - \langle z_l^\alpha z_s^\delta z_t^\sigma \rangle$$

$$\frac{\partial^2 \mathcal{L}}{\partial J_{ls}^{\alpha\delta} \partial J_{tu}^{\sigma\nu}} = \langle z_l^\alpha z_s^\delta \rangle \langle z_t^\sigma z_u^\nu \rangle - \langle z_l^\alpha z_s^\delta z_t^\sigma z_u^\nu \rangle - \frac{\lambda_J}{K} \frac{\partial g(J_{ls}^{\alpha\delta})}{\partial J_{tu}^{\sigma\nu}}$$

$\frac{\partial g(J_{ls}^{\alpha\delta})}{\partial J_{tu}^{\sigma\nu}} = 0$ for $\alpha \neq \sigma$ & $\delta \neq \nu$
& $l = t$ & $s = u$

$= 0$ for $\alpha = \sigma$ & $\delta = \nu$
& $l = t$ & $s = u$
& L1 reg
$\boxed{\text{BUT}}$ undefined at $J_{ls}^{\alpha\delta} = 0$

$= 2$ for $\alpha = \sigma$ & $\delta = \nu$
& $l = t$ & $s = u$
& L2 reg