On the TAP equations, and a Morita type derivation of the RS-solution for the SK model

Erwin Bolthausen, University of Zürich

Eindhoven, August 26, 2009

Sherrington-Kirkpatrick model:

$$\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N) \in \Sigma_N \stackrel{\text{def}}{=} \{-1, 1\}^N .$$

Random interactions: Independent centered Gaussians $g_{ij},\ i < j$, with variance $1/N$, $g_{ji} = g_{ij},\ g_{ii} \stackrel{\text{def}}{=} 0$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$.
Inverse temperature $\beta > 0$, strength $h > 0$ of the external field.
Hamiltonian:

$$-H_{N,\omega}(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \frac{\beta}{2} \sum_{i,j=1}^{N} g_{ij}(\omega)\, \sigma_i \sigma_j + h \sum_{i=1}^{N} \sigma_i.$$

Partition function:

$$Z_{N,\omega} \stackrel{\text{def}}{=} \sum_{\boldsymbol{\sigma}} 2^{-N} \exp\left[-H_{N,\omega}(\boldsymbol{\sigma})\right] .$$

Gibbs measure:

$$\mathcal{G}_{N,\omega}(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \frac{2^{-N}\exp\left[-H_{N,\omega}(\boldsymbol{\sigma})\right]}{Z_{N,\omega}}.$$

Free energy:

$$f(\beta,h) \stackrel{\text{def}}{=} \lim_{N\to\infty}\frac{1}{N}\log Z_N = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_N.$$

The **replica symmetric "solution"**, given by SK:

$$\begin{aligned}\mathrm{RS}(\beta,h) &= \int \log\cosh\left(h+\beta\sqrt{q}z\right)\phi(dz)\\ &+\frac{\beta^2}{4}(1-q)^2,\end{aligned}$$

$\phi(dz)$: the standard normal distribution. $q = q(\beta,h)$ satisfies:

$$q = \int \tanh^2\left(h+\beta\sqrt{q}z\right)\phi(dz).$$

$h > 0$, $\forall\beta$, and $h = 0$, $\beta \leq 1$: unique q. $h = 0$, $\beta > 1$, there is positive solution.

Theorem 1 (Talagrand, Guerra-Toninelli) For small enough β :

$$f(\beta, h) = \mathrm{RS}(\beta, h)$$

*Remark 1 The equation is believed to be correct for β below the **de Almayda–Thouless-line***

$$\mathrm{AT}(h) \stackrel{\text{def}}{=} \sup \left\{ \beta : \beta^2 \int \frac{\phi(dz)}{\cosh^4\left(h + \beta\sqrt{q}z\right)} \leq 1 \right\}$$

$\underline{h=0}$: $\mathrm{AT}(0)=1$. $f(\beta,0)=\mathrm{RS}(\beta,0)=\beta^2/4$ for $\beta\leq 1$ by a 2nd moment computation:

$$\mathbb{E}Z_N=\exp\left[\frac{\beta^2(N-1)}{4}\right],$$

and one easily checks for $\beta<1$ that $\mathbb{E}Z_N^2\leq C(\beta)(\mathbb{E}Z_N)^2$

$$\begin{aligned}\mathbb{E}Z_N^2 &= \sum_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}2^{-2N}\mathbb{E}\exp\left[\beta\sum\nolimits_{i<j}g_{ij}\left(\sigma_i\sigma_j+\sigma_i'\sigma_j'\right)\right]\\ &= \sum_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}2^{-2N}\exp\left[\frac{\beta^2}{2N}\sum\nolimits_{i<j}\left(\sigma_i\sigma_j+\sigma_i'\sigma_j'\right)^2\right]\\ &= (\mathbb{E}Z_N)^2\sum_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}2^{-2N}\exp\left[\frac{\beta^2}{2N}\left(\sum\nolimits_i\sigma_i\sigma_i'\right)^2-\frac{\beta^2}{2}\right]\leq C(\beta)(\mathbb{E}Z_N)^2\end{aligned}$$

by Curie-Weiss. By Gaussian isoperimetry.

$$f(\beta,0)=f_{\mathrm{an}}(\beta,0)\overset{\mathrm{def}}{=}\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}Z_N=\frac{\beta^2}{4}$$

However: For $h > 0, \ \forall \beta > 0$

$$\mathrm{RS}\,(\beta, h) \neq f_{\mathrm{an}}\,(\beta, h)\,.$$

Talagrand's first proof: Based on

$$\mathbb{E}\left(\mathrm{cov}_{\mathcal{G}}(\sigma_i, \sigma_j)^2\right) \leq C/N, \ i \neq j$$

proved by induction on N. Then prove that the m_i become uncorrelated and satisfy

$$\lim_{N\to\infty} \mathbb{E}\left(m_1^2 m_2^2\right) = q^2.$$

Finally using this, derive an asymptotic differential equation

$$\lim_{N\to\infty} \frac{1}{N} \frac{\partial \log Z_N}{\partial \beta} = \frac{\beta}{2}\left(1 - q\,(\beta, h)^2\right),$$

and from that, the claim follows.
Guerra-Toninelli: Use of the Guerra interpolation method and a clever "replica coupling".
Simplest proof: Recent one by **Talagrand** (to appear in PTRF). Clever extension of the class of considered models + interpolation, and recursion.

Morita type correction: Random Hamiltonian $H_{N,\omega}(\boldsymbol{\sigma})$, where $f < f_{\mathrm{an}}$: Sometimes, it is possible to subtract a $\boldsymbol{\sigma}$-independent sequence

$$\hat{H}_{N,\omega}(\boldsymbol{\sigma}) = H_{N,\omega}(\boldsymbol{\sigma}) - \psi_{N,\omega},$$

for which:

$$\frac{1}{N}\psi_{N,\omega} \to \alpha \text{ a.s.},$$

then

$$f = \lim_{N\to\infty} \frac{1}{N} \log \sum_{\boldsymbol{\sigma}} 2^{-N} \mathrm{e}^{-H_{N,\omega}(\boldsymbol{\sigma})} = \hat{f} - \alpha.$$

Therefore, if we can prove $\hat{f} = \hat{f}_{\mathrm{an}}$, and we can evaluate the latter, we are done.

Thouless-Anderson-Palmer equations: Consider the Gibbs-expectations of the σ_i :

$$m_i(\omega) \stackrel{\text{def}}{=} \langle \sigma_i \rangle_{N,\omega} .$$

$$m_i = \tanh\left(h + \beta \sum\nolimits_j g_{ij} m_j - \beta^2 (1-q)\, m_i\right).$$

Can be true only in an asymptotic sense $N \to \infty$.
Physicists claim:

- True for all β.
- Solutions encode for the “pure states”.
- High temperature: One solution.

Mathematical proofs only for high temperature: Talagrand, Chatterjee.
$i = N$: Write $m_i^{(N-1)}$ for the mean of $\sigma_i,\ i \le N-1$ in the $N-1$-system with Hamiltonian $H^{(N-1)}$.Putting $\boldsymbol{\sigma}^{(N-1)} = (\sigma_1, \ldots, \sigma_{N-1})\,, y\left(\boldsymbol{\sigma}^{(N-1)}\right) \stackrel{\text{def}}{=} \sum_{i=1}^{N-1} g_{i,N}\sigma_i$:

$$-H(\boldsymbol{\sigma}) = -H^{(N-1)}\left(\boldsymbol{\sigma}^{(N-1)}\right) + h\sigma_N + \beta y\left(\boldsymbol{\sigma}^{(N-1)}\right)\sigma_N,$$

$$m_N = \frac{\left\langle \sinh\left(h + \beta y\left(\boldsymbol{\sigma}^{(N-1)}\right)\right)\right\rangle_{N-1}}{\left\langle \cosh\left(h + \beta y\left(\boldsymbol{\sigma}^{(N-1)}\right)\right)\right\rangle_{N-1}},$$

Using $\mathbb{E}\left(\mathrm{cov}_{\mathcal{G}}(\sigma_i, \sigma_j)^2\right) \leq C/N$, a not too difficult argument gives

$$\begin{aligned}\frac{\left\langle \sinh\left(h + \beta y\left(\boldsymbol{\sigma}^{(N-1)}\right)\right)\right\rangle_{N-1}}{\left\langle \cosh\left(h + \beta y\left(\boldsymbol{\sigma}^{(N-1)}\right)\right)\right\rangle_{N-1}} &= \frac{\sinh\left(h + \beta y\left(\left\langle \boldsymbol{\sigma}^{(N-1)}\right\rangle_{N-1}\right)\right)}{\cosh\left(h + \beta y\left(\left\langle \boldsymbol{\sigma}^{(N-1)}\right\rangle_{N-1}\right)\right)} + o(1) \\ &= \tanh\left(h + \beta \sum_{j=1}^{N-1} g_{N,j} m_j^{(N-1)}\right) + o(1),\end{aligned}$$

i.e.

$$m_N = \tanh\left(h + \beta \sum_{j \leq N-1} g_{N,j} m_j^{(N-1)}\right) + o(1).$$

Replacing $m_j^{(N-1)}$ by m_j leads to a correction as given in the TAP equations.

Direct construction of TAP without reference to the Gibbs measure: Define recursively $m_i^{(k)}$ by

$$m_i^{(0)} \stackrel{\mathrm{def}}{=} 0,\ m_i^{(1)} \stackrel{\mathrm{def}}{=} \sqrt{q},$$

$$h_i^{(k)} \stackrel{\mathrm{def}}{=} h + \beta \sum_{j=1}^{N} g_{ij} m_j^{(k-1)} - \beta^2 (1-q)\, m_i^{(k-2)},\ k \geq 2$$

$$m_i^{(k)} \stackrel{\mathrm{def}}{=} \tanh\left(h_i^{(k)}\right),\ k \geq 2.$$

Not too difficult to prove that the scheme converges for small enough β, maybe up to the AT-line.

Proposition 1 a)

$$\lim \frac{1}{N}\sum_{i=1}^{N}\psi\left(h_i^{(k)}\right) = \int \psi\left(h+\beta\sqrt{q}z\right)\phi\left(dz\right),\ \psi \text{ nice},\ k \geq 2$$

b)

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} m_i^{(k)2} = q,\ \text{a.s.},\ \forall k \geq 1.$$

c) β *small enough* $\Longrightarrow\ \exists\, C\left(\beta\right) > 0,\ 0 < \rho\left(\beta\right) < 1,$ *s.th.*

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(m_i^{(k)} - m_i^{(k-1)}\right)^2 \leq C\rho^k.$$

d)

$$\lim_{k\to\infty}\lim_{N\to\infty}\frac{1}{N}\sum_{i,j=1}^{N} g_{ij} m_i^{(k)} m_j^{(k)} = 2\beta q\left(1-q\right).$$

Proof by an alternative representation of $h_i^{(k)}$:

$$h_i^{(2)} = h + \beta\sqrt{q}\xi_i,\ \xi_i \overset{\text{def}}{=} \sum_j g_{ij},$$

$$h_i^{(3)} = h + \beta \sum_j g_{ij} \tanh\left(h + \beta\sqrt{q}\xi_j\right) - \beta^2\sqrt{q}\,(1-q)\,.$$

Idea: Replace g_{ij} by ones which are independent of the ξ_i :

$$g_{ij}^{(2)} \overset{\text{def}}{=} g_{ij} - \frac{\xi_i + \xi_j}{N}$$

$\approx$ independent of the ξ_k (up to corrections which are negligible for $N \to \infty$).
Then correct $g_{ij}^{(2)}$ s.th. $\mathcal{L} = \mathcal{L}\left(\{g_{ij}\}\right)$: Independent $\overline{\xi}_i$ with $\mathcal{L}\left(\{\overline{\xi}_i\}\right) = \mathcal{L}\left(\{\xi_i\}\right)$, and put

$$\overline{g}_{ij}^{(2)} \overset{\text{def}}{=} g_{ij}^{(2)} + \frac{\overline{\xi}_i + \overline{\xi}_j}{N}$$

$$h_i^{(3)} \approx h + \beta \sum_j \overline{g}_{ij}^{(2)} \left(m_j^{(2)} - \left\langle \mathbf{m}^{(2)}, \mathbf{1} \right\rangle \right) + \beta \left\langle \mathbf{m}^{(2)}, \mathbf{1} \right\rangle \xi_i,$$

where $\mathbf{m}^{(2)} \stackrel{\text{def}}{=} \left(m_1^{(2)}, \ldots, m_N^{(2)} \right)$, $\langle \mathbf{x}, \mathbf{y} \rangle \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N x_i y_i$.
$\approx$ means equality up to terms which are negligible in the $N \to \infty$ limit.
General scheme: $\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \ldots$. $\Pi^{(k)} : \mathbb{R}^N \to \operatorname{span}\left(\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(k)} \right)$ projections.

$$\hat{\mathbf{m}}^{(k)} \stackrel{\text{def}}{=} \frac{\mathbf{m}^{(k)} - \Pi^{(k-1)} \left(\mathbf{m}^{(k)} \right)}{\left\| \mathbf{m}^{(k)} - \Pi^{(k-1)} \left(\mathbf{m}^{(k)} \right) \right\|}, \quad \hat{\mathbf{m}}^{(1)} \stackrel{\text{def}}{=} \mathbf{1}.$$

Recursive construction of σ-fields $\mathcal{G}_1, \mathcal{G}_2, \ldots, g_{ij}^{(1)}, g_{ij}^{(2)}, \ldots$

$$\xi_i^{(k)} \stackrel{\text{def}}{=} \sum_j g_{ij}^{(k)} \hat{m}_j^{(k)}, \quad \mathcal{G}_k \stackrel{\text{def}}{=} \sigma \left(\boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(k)} \right),$$

$$g_{ij}^{(k+1)} \stackrel{\text{def}}{=} g_{ij}^{(k)} - \frac{\xi_i^{(k)} \hat{m}_j^{(k)} + \xi_j^{(k)} \hat{m}_i^{(k)}}{N}.$$

Conditionally on $\mathcal{G}_{k-1}$, $\left\{g_{ij}^{(k+1)}\right\}$ and $\left\{\xi_i^{(k)}\right\}$ are $\approx$ independent.
Recovering the "correct" distribution:

$$\overline{g}_{ij}^{(k+1)} \stackrel{\text{def}}{=} \overline{g}_{ij}^{(k)} - \frac{\overline{\xi}_i^{(k)} \hat{m}_j^{(k)} + \overline{\xi}_j^{(k)} \hat{m}_i^{(k)}}{N},$$

where *conditionally on* $\mathcal{G}_{k-1}$, $\left\{\overline{\xi}_i^{(k)}\right\}_i$ and $\left\{\xi_i^{(k)}\right\}_i$ are independent, and have the same conditional Gaussian law.
Then

$$h_i^{(k)} \approx h + \beta \sum_j \overline{g}_{i,j}^{(k-1)} \left(m_j^{(k-1)} - \Pi^{(k-2)} \left(\mathbf{m}^{(k-1)} \right)_j \right) + \beta \sum_{r=1}^{k-2} \left\langle \hat{\mathbf{m}}^{(r)}, \mathbf{m}^{(k-1)} \right\rangle \xi_i^{(r)}.$$

Using this representation, the proposition follows.

The transformation: New i-th spin distribution

$$p_i(\sigma_i) \stackrel{\text{def}}{=} \frac{1}{2} \frac{\mathrm{e}^{h_i \sigma_i}}{\cosh(h_i)},$$

$$p(\boldsymbol{\sigma}) \stackrel{\text{def}}{=} \prod_{i=1}^{N} p_i(\sigma_i).$$

h_i should be $h_i^{(k)}$ for large k such that that $h_i^{(k)}$ is close to $h_i^{(k-1)}$.

$$\begin{aligned} Z_N &= \sum_{\boldsymbol{\sigma}} 2^{-N} \exp[-H_N(\boldsymbol{\sigma})] \\ &= \exp\left[\sum\nolimits_i \log\cosh(h_i)\right] \sum_{\boldsymbol{\sigma}} p(\boldsymbol{\sigma}) \exp\left[-H(\boldsymbol{\sigma}) - \sum\nolimits_i h_i \sigma_i\right]. \end{aligned}$$

$$\sum\nolimits_i \log\cosh(h_i) \approx N \int \log\cosh(h + \beta\sqrt{q} z)\, \phi(dz).$$

$$
\begin{aligned}
-H(\boldsymbol{\sigma}) - \sum\nolimits_i h_i\sigma_i &= \frac{\beta}{2}\sum_{i,j} g_{ij}\sigma_i\sigma_j + h\sum_i \sigma_i - h\sum_i \sigma_i \\
&\quad - \sum_{i,j} g_{ij}\sigma_i m_j + \beta^2(1-q)\sum_i \sigma_i m_i \\
&= \frac{\beta}{2}\sum_{i,j} g_{ij}\hat{\sigma}_i\hat{\sigma}_j - \frac{\beta}{2}\sum_{i,j} g_{ij} m_i m_j + \beta^2(1-q)\sum_i \sigma_i m_i \\
&\approx \frac{\beta}{2}\sum_{i,j} g_{ij}\hat{\sigma}_i\hat{\sigma}_j - N\beta^2 q(1-q) + \beta^2(1-q)\sum_i \sigma_i m_i \\
&\approx \frac{\beta}{2}\sum_{i,j} g_{ij}\hat{\sigma}_i\hat{\sigma}_j + \beta^2(1-q)\sum_i \hat{\sigma}_i m_i,
\end{aligned}
$$

where

$$
\hat{\sigma}_i \stackrel{\text{def}}{=} \sigma_i - m_i.
$$

On the other hand $\hat{\sigma}_i^2 = (\sigma_i - m_i)^2 = 1 - m_i^2 - 2\hat{\sigma}_i m_i$, and therefore

$$\begin{aligned}\frac{\beta^2}{4N}\sum_{i,j}\hat{\sigma}_i^2\hat{\sigma}_j^2 &= \frac{\beta^2}{4N}\sum_{i,j}\left(1-m_i^2\right)\left(1-m_j^2\right) - \frac{\beta^2}{N}\sum_{i,j}\hat{\sigma}_i m_i\left(1-m_j^2\right) + \frac{\beta^2}{N}\left(\sum_i \hat{\sigma}_i m_i\right)^2 \\ &\approx \frac{\beta^2 N}{4}(1-q)^2 - \beta^2(1-q)\sum_i \hat{\sigma}_i m_i + \frac{\beta^2}{N}\left(\sum_i \hat{\sigma}_i m_i\right)^2.\end{aligned}$$

Combining:

$$-H(\boldsymbol{\sigma}) - \sum\nolimits_i h_i\sigma_i \approx \frac{\beta^2 N}{4}(1-q)^2 + \frac{\beta}{2}\sum_{i,j} g_{ij}\hat{\sigma}_i\hat{\sigma}_j - \frac{\beta^2}{4N}\sum_{i,j}\hat{\sigma}_i^2\hat{\sigma}_j^2 + \frac{\beta^2}{N}\left(\sum\nolimits_i m_i\hat{\sigma}_i\right)^2,$$

The last term is annoying, but harmless for small β :CW-type term. Leaving it out:

$$f(\beta,h) = \mathrm{RS}(\beta,h) + \lim_{N\to\infty}\frac{1}{N}\log\sum_{\boldsymbol{\sigma}} p(\boldsymbol{\sigma})\exp\left[\frac{\beta}{2}\sum_{i,j} g_{ij}\hat{\sigma}_i\hat{\sigma}_j - \frac{\beta^2}{4N}\sum_{i,j}\hat{\sigma}_i^2\hat{\sigma}_j^2\right].$$

Claim: The second summand vanishes by a 2nd moment computation, if β is small.

Problem: $p(\boldsymbol{\sigma})$ depends on $\{g_{ij}\}$.

Solution: Conditional second moment, given $\mathcal{G}_k$, by a replacement of $\{g_{ij}\}$ by $\left\{\overline{g}_{ij}^{(k)}\right\}$. The replacement introduces CW-type summands which are handled quenched. The CW-computation is quite messy here.

Other possible applications:

Generalized SK-model: σ_i with values in an arbitrary finite set S, with an arbitrary a priori measure π.

$$-H_N(\boldsymbol{\sigma}) \stackrel{\mathrm{def}}{=} \sum g_{ij}(\sigma_i, \sigma_j),$$

where $\{g_{ij}(s,t)\}_{i<j,\ s,t\in S}$ is a Gaussian field, independent for different (i,j). For that, we know what the correct TAP equations should be. The TAP equations are for the full (quenched) Gibbs distribution of σ_i on S. (joint work with Philipp Thomann).

Most general perceptron model: $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N) \in \{-1, 1\}^N$, g_{ij}, $1 \leq i, j \leq N$, i.i.d. standard Gaussians.

$$y_{\sigma,i} \stackrel{\text{def}}{=} \sum_{j=1}^{N} g_{ij}\sigma_j, \; L_{N,\sigma} \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{y_{\sigma,i}}.$$

We search for a rate function $J : \mathcal{M}_1^+(\mathbb{R}) \to [0, \infty]$, such that

$$\limsup_{N\to\infty} \frac{1}{N} \log \# \{\sigma : L_{N,\sigma} \in A\} - \log 2 \leq - \inf_{\mu\in A} J(\mu), \; A \text{ closed},$$
$$\liminf_{N\to\infty} \frac{1}{N} \log \# \{\sigma : L_{N,\sigma} \in A\} - \log 2 \geq - \inf_{\mu\in A} J(\mu), \; A \text{ open}.$$

The “annealed” rate function is the relative entropy $I(\mu|\phi)$ w.r.t. standard Gaussian ϕ, because by Sanov

$$\mathbb{E}\# \{\sigma : L_{N,\sigma} \in A\} = 2^N \mathbb{P}(L_{N,\sigma} \in A) \approx 2^N \exp[-NI(\mu|\phi)]$$

Up to now, we have not been able to find the correct TAP equation.
I expect that $J(\mu) = I(\mu|\phi)$ on a hypersurface in $\mathcal{M}_1^+(\mathbb{R})$ of codimension 1, and μ close to ϕ.