Lectures notes for STSCI6940: Topics in high-dimensional inference

Ahmed El Alaoui

February 18, 2021

1 Lecture 1: The additive Gaussian noise channel

Let P_0 be a (Borel) probability distribution on $\mathbb{R}^n$ with finite second moment. Consider a "signal" vector $x \sim P_0$. We observe a noisy version of x given by

$$y = \sqrt{\lambda}x + z,$$

where $z \sim N(0, I)$ independent of x and $\lambda \geq 0$ is a fixed scalar playing the role of a "signal-to-noise" ratio (snr). The task is to estimate x from y. We consider the setting where the prior P_0 and the snr λ are known. We measure the performance of an estimator $\widehat{x} : \mathbb{R}^n \to \mathbb{R}^n$ by its mean squared error

$$\mathrm{MSE}(\widehat{x}) := \mathbb{E}\left[\|\widehat{x}(y) - x\|_2^2\right].$$

We observe that the MSE is minimized by the *Bayes-optimal* estimator

$$\widehat{x}^{bayes} = \mathbb{E}[x|y],$$

and has a corresponding minimal mean squared error

$$\mathrm{MMSE}_{P_0}(\lambda) := \mathbb{E}\left[\|x - \mathbb{E}[x|y]\|_2^2\right].$$

The proof follows from the simple decomposition

$$\begin{aligned}
\mathrm{MSE}(\widehat{x}) &= \mathbb{E}\left[\|\widehat{x}(y) - \mathbb{E}[x|y] + \mathbb{E}[x|y] - x\|_2^2\right] \\
&= \mathbb{E}\left[\|\widehat{x}(y) - \mathbb{E}[x|y]\|_2^2\right] + \mathbb{E}\left[\|x - \mathbb{E}[x|y]\|_2^2\right] + 2\,\mathbb{E}\left[(\widehat{x}(y) - \mathbb{E}[x|y])^\top(\mathbb{E}[x|y] - x)\right],
\end{aligned}$$

and by noticing that the cross product is zero as a consequence of the tower property of expectations:

$$\begin{aligned}
\mathbb{E}\left[(\widehat{x}(y) - \mathbb{E}[x|y])^\top(\mathbb{E}[x|y] - x)\right] &= \mathbb{E}\left[\,\mathbb{E}\left[(\widehat{x}(y) - \mathbb{E}[x|y])^\top(\mathbb{E}[x|y] - x)|y\right]\right] \\
&= \mathbb{E}\left[(\widehat{x}(y) - \mathbb{E}[x|y])^\top(\mathbb{E}[x|y] - \mathbb{E}[x|y])\right] = 0.
\end{aligned}$$

It is clear now that in order to understand the MMSE, we need to study the posterior distribution of x given y. By Bayes' rule,

$$\mathrm{d}\,\mathbb{P}(x|y) = \frac{f(y|x)}{\tilde{Z}}\mathrm{d}P_0(x),$$

where $f(y|x)$ is the conditional density of y given x, which is the multivariate Gaussian $N(\sqrt{\lambda}x, I)$; i.e.,

$$f(y|x) = \frac{1}{\sqrt{2\pi}^n}e^{-\|y-\sqrt{\lambda}x\|_2^2/2}.$$

Therefore the posterior measure reads (after expanding the square and simplifying constant terms)

$$\mathrm{d}\,\mathbb{P}(x|y) = \frac{e^{\sqrt{\lambda} y^\top x - \lambda \|x\|_2^2/2}}{Z(y;\lambda)} \mathrm{d}P_0(x),$$

where

$$Z(y;\lambda) = \int e^{\sqrt{\lambda} y^\top x - \lambda \|x\|_2^2/2} \mathrm{d}P_0(x)$$

is the normalizing constant, so that $\mathbb{P}(\cdot|y)$ is a probability measure. Borrowing language from statistical physics, the posterior measure $\mathbb{P}(\cdot|y)$ is called a Boltzmann, or Gibbs distribution, Z is called the *partition function*, and $h(x) = h(x; y, \lambda) = \sqrt{\lambda} y^\top x - \lambda \|x\|_2^2/2$ is called the *Hamiltonian* function. The expected log-partition function

$$F(\lambda) = \mathbb{E} \log Z(y;\lambda)$$

is called the *free energy*, and will be an important object as it encapsulates precious information about the behavior of $\mathbb{P}(\cdot|y)$.

I-MMSE relation. The *mutual information* between two random variables (x, y) is defined as

$$I(x, y) = d_{\mathrm{KL}}(P_{x,y} \,||\, P_x \times P_y),$$

where $P_{x,y}$ is the joint distribution of x and y, and P_x and P_y are the respective marginals, d_{KL} is the Kullback-Liebler divergence between probability distributions. (Observe that the mutual information is symmetric in its arguments.)

When y is obtained from x according to the additive Gaussian noise channel as above, there is a simple relationship between the mutual information and the the free energy:

Lemma 1.

$$I(x, y) = \frac{\lambda}{2} \mathbb{E}\left[\|x\|_2^2\right] - F(\lambda).$$

Proof. By the Bayes rule,

$$\begin{aligned} I(x, y) &= \mathbb{E} \log \frac{\mathrm{d}P_{x,y}}{\mathrm{d}(P_x \times P_y)}(x, y) \\ &= \mathbb{E} \log \frac{\mathrm{d}\,\mathbb{P}}{\mathrm{d}P_0}(x|y) \\ &= \mathbb{E} \log \frac{\exp\left(\sqrt{\lambda} y^\top x - \lambda \|x\|_2^2/2\right)}{Z(y;\lambda)} \\ &= \sqrt{\lambda}\, \mathbb{E}[y^\top x] - \frac{\lambda}{2} \mathbb{E}\left[\|x\|_2^2\right] - F(\lambda) \\ &= \frac{\lambda}{2} \mathbb{E}\left[\|x\|_2^2\right] - F(\lambda). \end{aligned}$$

□

The I-MMSE relation is a statement about the derivative of $I(x, y)$ (or equivalently, F) with respect to λ, which as the name indicates, relates it to the MMSE:

Proposition 2. *For all $\lambda \geq 0$,*

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} I(x,y) = \frac{1}{2}\mathrm{MMSE}_{P_0}(\lambda).$$

Equivalently,

$$F'(\lambda) = \frac{1}{2}\,\mathbb{E}\left[\|\,\mathbb{E}[x|y]\|_2^2\right].$$

The proof relies on Gaussian integration by parts (a defining property of the Gaussian distribution) which we state below:

Lemma 3. *Let $z \sim N(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function such that $|f(x)|e^{-x^2/2} \to 0$ as $|x| \to \infty$. Then*

$$\mathbb{E}[zf(z)] = \mathbb{E}[f'(z)].$$

The proof of this lemma is an exercise in integration by parts.

We can now prove the I-MMSE relation:

Proof of Proposition 2. Let us write $y = \sqrt{\lambda}x_0 + z$ where $x_0 \sim P_0$ is the signal vector. (We are writing x_0 so we can distinguish it from the 'dummy' variable x which appears in the integral below.) We have

$$\begin{aligned} F'(\lambda) &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathbb{E}\log\int e^{\sqrt{\lambda}z^\top x + \lambda x_0^\top x - \lambda\|x\|_2^2/2}\mathrm{d}P_0(x) \\ &= \mathbb{E}\Big[\frac{1}{Z(y;\lambda)}\int\Big(\frac{1}{2\sqrt{\lambda}}z^\top x + x_0^\top x - \frac{1}{2}\|x\|_2^2\Big)e^{H(x)}\mathrm{d}P_0(x)\Big]. \end{aligned}$$

We pay special attention to the first term in the above expression. By Gaussian integration by parts, we have

$$\begin{aligned} \frac{1}{2\sqrt{\lambda}}\,\mathbb{E}\Big[z^\top\int xe^{H(x)}\mathrm{d}P_0(x)/Z(y;\lambda)\Big] &= \frac{1}{2\sqrt{\lambda}}\sum_{i=1}^n\mathbb{E}\Big[\frac{\partial}{\partial z_i}\Big(\int x_i e^{\sqrt{\lambda}z^\top x + \lambda x_0^\top x - \lambda\|x\|_2^2/2}\mathrm{d}P_0(x)/Z(y;\lambda)\Big)\Big] \\ &= \frac{1}{2}\,\mathbb{E}\left[\,\mathbb{E}\left[\|x\|_2^2\,|\,y\right]\right] - \frac{1}{2}\,\mathbb{E}\left[\left\|\,\mathbb{E}[x\,|\,y]\right\|_2^2\right] \\ &= \frac{1}{2}\,\mathbb{E}\left[\|x\|_2^2\right] - \frac{1}{2}\,\mathbb{E}\left[\left\|\,\mathbb{E}[x\,|\,y]\right\|_2^2\right]. \end{aligned}$$

Plugging this back in the expression of F', we obtain

$$\begin{aligned} F'(\lambda) &= \frac{1}{2}\,\mathbb{E}\left[\|x\|_2^2\right] - \frac{1}{2}\,\mathbb{E}\left[\left\|\,\mathbb{E}[x\,|\,y]\right\|_2^2\right] + \mathbb{E}\left[x_0^\top\,\mathbb{E}[x\,|\,y]\right] - \frac{1}{2}\,\mathbb{E}\left[\|x\|_2^2\right] \\ &= \frac{1}{2}\,\mathbb{E}\left[\left\|\,\mathbb{E}[x\,|\,y]\right\|_2^2\right], \end{aligned}$$

where the last equality follows again from the tower property of conditional expectations: $\mathbb{E}\left[x_0^\top\,\mathbb{E}[x|y]\right] = \mathbb{E}\left[\,\mathbb{E}[x_0|y]^\top\,\mathbb{E}[x|y]\right] = \mathbb{E}\left[\left\|\,\mathbb{E}[x|y]\right\|_2^2\right]$. The statement about the mutual information follows from the relation between I and F, Lemma 1. □

Using the same kind of computations, we can also compute the second derivative of F and find

$$F''(\lambda) = \frac{1}{2}\,\mathbb{E}\left[\|\,\mathrm{cov}(x|y)\|_F^2\right], \tag{1}$$

where $\mathrm{cov}(x|y)$ is the covariance matrix of x conditional on y. We leave this computation as an exercise!

A few key properties of F follow from the above computations:

Proposition 4.

1. *F is nondecreasing.*

2. *F is nonnegative.*

3. *F is convex.*

The first and last statements follow from the formulas for the derivatives. the second statement follows from the first one together with $F(0) = 0$.

2 Lecture 2:

2.1 The Gaussian prior is the hardest of all

We consider the case of a normal prior in one dimension: $P_0 = N(0,1)$. We compute the free energy, the mutual information and the MMSE. We have

$$\begin{aligned} F(\lambda) &= \mathbb{E}\log\Big(\int e^{\sqrt{\lambda}yx-\lambda x^2/2}e^{-x^2/2}\frac{\mathrm{d}x}{\sqrt{2\pi}}\Big) \\ &= \mathbb{E}\log\Big(\frac{1}{\sqrt{1+\lambda}}e^{\lambda y^2/(2(1+\lambda))}\Big) \\ &= \frac{\lambda}{2(1+\lambda)}\,\mathbb{E}[y^2] - \frac{1}{2}\log(1+\lambda) \\ &= \frac{\lambda}{2} - \frac{1}{2}\log(1+\lambda), \end{aligned}$$

and therefore, using Lemma 1,

$$I(x,y) = \frac{1}{2}\log(1+\lambda).$$

The posterior mean $\mathbb{E}[x|y]$ is the orthogonal projection of x onto y:

$$\mathbb{E}[x|y] = \frac{\mathbb{E}[xy]}{\mathbb{E}[y^2]}y = \frac{\sqrt{\lambda}}{1+\lambda}y,$$

and the MMSE is

$$\mathrm{MMSE}_{N(0,1)} = \mathbb{E}\left[\big(x - \frac{\sqrt{\lambda}}{1+\lambda}y\big)^2\right] = \frac{1}{1+\lambda}.$$

Now let's consider a general prior P_0 with unit second moment over $\mathbb{R}$. We can upper bound the MMSE by considering the estimator $\widehat{x}(y) = \frac{\sqrt{\lambda}}{1+\lambda}y$, which is not equal to $\mathbb{E}[x|y]$ is general. So via the same calculation as above, we have the upper bound

$$\mathrm{MMSE}_{P_0}(\lambda) \le \mathrm{MSE}(\widehat{x}) = \frac{1}{1+\lambda} = \mathrm{MMSE}_{N(0,1)}(\lambda).$$

We conclude that **among all distributions with unit second moment, the Gaussian $N(0,1)$ maximizes the minimal mean squared error**.

Remark 1. *Notice that by rescaling λ, the condition of unit second moment can be replaced by a fixed second moment σ_x^2, in which case the Gaussian $N(0,\sigma_x^2)$ is the one maximizing the MMSE.*

A similar extremality result can be proved for the mutual information: Using the I-MMSE relation, we have

$$I(x,y)_{|\lambda} - I(x,y)_{|\lambda=0} = \frac{1}{2}\int_0^\lambda \mathrm{MMSE}_{P_0}(\tau)\mathrm{d}\tau.$$

At $\lambda = 0$, x and y are independent, so $I(x,y)_{|\lambda=0} = 0$. Moreover, we obtained that $\mathrm{MMSE}_{P_0}(\lambda) \le \frac{1}{1+\lambda}$. Therefore,

$$I(x,y) \le \frac{1}{2}\int_0^\lambda \frac{1}{1+\tau}\mathrm{d}\tau = \frac{1}{2}\log(1+\lambda).$$

We have just computed the *channel capacity* of the the additive Gaussian channel, i.e., the maximal mutual information of this channel over priors of a given variance: **among priors of unit second moment, the mutual information of the additive Gaussian channel is maximized by the Gaussian distribution.** This is called the Shannon-Hartley theorem in information theory.

2.2 A simple Gaussian mean estimation problem

We consider another special case of the additive Gaussian channel, this time in high dimension. And this will be our first encounter with a *phase transition.* Consider a random integer σ_0 uniformly drawn from the set $\{1, \cdots, 2^n\}$, and the random vector y defined by

$$y = \sqrt{\lambda n}\, e_{\sigma_0} + z,$$

where $(e_1, \cdots, e_{2^n})$is the standard basis of $\mathbb{R}^{2^n}$ and z is a standard Gaussian vector in $\mathbb{R}^{2^n}$. In words, one observes a Gaussian vector in high dimension, where the mean of an unknown coordinate is elevated by an amount of $\sqrt{\lambda n}$. The task is to identify this special coordinate, i.e., to estimate σ_0.

First, a natural estimator to consider here is the *Maximum Likelihood Estimator* (MLE), which in this case is

$$\widehat{\sigma} = \arg\max_{1 \le i \le 2^n} y_i.$$

Since the maximum of 2^n independent standard normal variables concentrates very tightly around $\sqrt{2\log(2^n)} = \sqrt{(2\log 2)n}$, we see that the MLE estimator recovers σ_0 with high probability if $\lambda > 2\log 2$. Now the question is whether estimation is still possible below $2\log 2$. To answer this, we will compute the free energy, which reads

$$F_n(\lambda) = \mathbb{E}\log\Big(2^{-n}\sum_{\sigma=1}^{2^n} e^{\sqrt{\lambda n}y_\sigma - \lambda n/2}\Big).$$

Theorem 5. *We have for all $\lambda \ge 0$,*

$$\lim_{n\to\infty}\frac{F_n(\lambda)}{n} = \begin{cases} 0 & \text{if } \lambda \le 2\log 2, \\ \frac{\lambda}{2} - \log 2 & \text{otherwise.} \end{cases}$$

Proof. Let f_* be the function on the right-hand side of the above formula. We will prove matching upper and lower bounds on F_n. Let's start with the upper bound, which relies on Jensen's inequality conditioned on the random variables σ_0 and z_{σ_0}:

$$\begin{aligned} F_n(\lambda) &\le \mathbb{E}\log\Big(\mathbb{E}\Big[2^{-n}\sum_{\sigma=1}^{2^n} e^{\sqrt{\lambda n}y_\sigma - \lambda n/2}\Big|\sigma_0, z_{\sigma_0}\Big]\Big) \\ &= \mathbb{E}\log\big(2^{-n}(2^n-1) + 2^{-n}e^{\sqrt{\lambda n}z_{\sigma_0} + \lambda n/2}\big) \\ &\le \mathbb{E}\log\Big(1 + \exp\Big(\sqrt{\lambda n}z_{\sigma_0} + (\frac{\lambda}{2} - \log 2)n\Big)\Big). \end{aligned}$$

We see that the dominant term inside the exponential is $(\frac{\lambda}{2} - \log 2)n$. Indeed, we can further bound the above expression as

$$\begin{aligned} &\mathbb{E}\log\Big(1 + \exp\Big(\sqrt{\lambda n}z_{\sigma_0} + (\frac{\lambda}{2} - \log 2)n\Big)\Big)\mathbf{1}\{z_{\sigma_0} \le 0\} \\ &+ \mathbb{E}\log\Big(e^{-\sqrt{\lambda n}z_{\sigma_0}} + \exp\Big((\frac{\lambda}{2} - \log 2)n\Big)\Big)\mathbf{1}\{z_{\sigma_0} \ge 0\} + \sqrt{\lambda n}\,\mathbb{E}\left[z_{\sigma_0}\mathbf{1}\{z_{\sigma_0} \ge 0\}\right] \\ &\le \mathbb{E}\log\Big(1 + \exp\Big((\frac{\lambda}{2} - \log 2)n\Big)\Big) + \sqrt{\frac{\lambda n}{\pi}}. \end{aligned}$$

Therefore we obtain the upper bound

$$\limsup_{n\to\infty}\frac{F_n(\lambda)}{n} \le f_*(\lambda).$$

A matching lower bound can be obtained by only considering the term corresponding to $\sigma = \sigma_0$ in the expression of the free energy:

$$\begin{aligned} F_n(\lambda) &\geq \mathbb{E} \log \Big(2^{-n} e^{\sqrt{\lambda n} y_{\sigma_0} - \lambda n/2} \Big) \\ &= \Big(\frac{\lambda}{2} - \log 2 \Big) n. \end{aligned}$$

Since $F_n(\lambda)$ is also nonnegative, the matching lower bound follows. □

Now we deduce from the above theorem the impossibility of estimation for $\lambda \leq 2\log 2$. Let

$$Q_n(\lambda) = \mathbb{E}\left[\big\| \mathbb{E}[\sqrt{n} e_\sigma | y] \big\|_2^2 \right] = n \sum_{i=1}^{2^n} \mathbb{E}\left[\mathbb{P}(\sigma = i | y)^2 \right].$$

It follows from the I-MMSE relation that

$$\frac{1}{2} \int_0^\lambda \frac{Q_n(\tau)}{n} \mathrm{d}\tau = \frac{F_n(\lambda)}{n} \longrightarrow 0 \quad \text{if } \lambda \leq 2\log 2.$$

Since Q_n is nonnegative it follows that $\lim_{n\to\infty} Q_n(\lambda)/n = 0$ for almost all $\lambda \in [0, 2\log 2]$. Further since $\lambda \mapsto Q_n(\lambda)$ is non-decreasing, then $\lim_{n\to\infty} Q_n(\lambda)/n = 0$ for *all* $\lambda \in [0, 2\log 2]$. We deduce that the (rescaled) MMSE is maximal, i.e., it is asymptotically equal to the error of a "dummy" estimator which return the mean vector $\mathbb{E}[e_{\sigma_0}] = 2^{-n}\mathbf{1}$ without looking at the data y:

$$\lim_{n\to\infty} \frac{\mathrm{MMSE}_n(\lambda)}{n} = 1 \quad \text{for all } \lambda \leq 2\log 2.$$

The value of conditioning... What if we did not condition on σ_0 and z_{σ_0} when we used Jensen's inequality to upper bound $F_n(\lambda)$? We would have gotten

$$\frac{F_n(\lambda)}{n} \leq \frac{1}{n} \log \big(2^{-n}(2^n - 1) + 2^{-n} e^{\lambda n} \big),$$

and the upper bound would tend to zero only up to $\lambda = \log 2$ instead of $2\log 2$, and would lead to a loose upper bound above $\lambda = \log 2$. The explanation of this rather generic phenomenon is that the fluctuations of z_{σ_0} do not affect the typical value of the partition function $Z(y;\lambda)$ by more than a factor $e^{\Theta(\sqrt{n})}$, but they have an outsize influence on its expectation $\mathbb{E}[Z(y;\lambda)]$ with an effect of order $e^{\Theta(n)}$, as can be seen from the above computations. An other way of seeing this as follows: Consider the event

$$E_\tau = \{ z_{\sigma_0} \geq \sqrt{\tau n} \}, \quad \tau > 0.$$

This event is extremely unlikely: $\mathbb{P}(E_\tau) \sim e^{-\tau n/2}$. However, conditional on E_τ, the value of Z is atypically large:

$$Z(y;\lambda) \geq 2^{-n} e^{\sqrt{\lambda\tau} n + \lambda n/2} + 2^{-n} \sum_{\sigma \neq \sigma_0} e^{\sqrt{\lambda n} z_\sigma - \lambda n/2}.$$

Therefore,

$$\mathbb{E}[Z(y;\lambda)] \geq \mathbb{E}[Z | E_\tau]\, \mathbb{P}(E_\tau) \sim e^{n f_*(\lambda)} e^{n(\sqrt{\lambda\tau} - \tau/2)}.$$

So for $\tau < 4\lambda$, the event E_τ alone contributes an order of $e^{\Theta(n)}$ to the expectation of Z compared to its typical value which is $e^{n f_*(\lambda) + o(n)}$.

3 Lecture 3:

3.1 The rank-one spiked Wigner model

Let $x \in \mathbb{R}^n$ with coordinates x_i drawn i.i.d. from a prior P_0 with zero mean and unit variance. We consider the observation model

$$Y_{ij} = \sqrt{\frac{\lambda}{n}} x_i x_j + W_{ij}, \quad 1 \le i < j \le n.$$

where $W_{ij} \sim N(0,1)$ independently for $i < j$. The posterior measure is

$$\mathrm{d}\,\mathbb{P}(x|Y) = \frac{1}{Z_n(Y;\lambda)} \exp\Big(\sum_{i<j} \sqrt{\frac{\lambda}{n}} Y_{ij} x_i x_j - \frac{\lambda}{2n} x_i^2 x_j^2\Big) \mathrm{d}P_0^n(x),$$

and the associated free energy is

$$F_n(\lambda) = \mathbb{E} \log Z_n(Y;\lambda) = \mathbb{E} \log \int e^{H_n(x)} \mathrm{d}P_0^n(x),$$

where $H_n(x) = H_n(x; Y, \lambda) = \sqrt{\frac{\lambda}{n}} Y_{ij} x_i x_j - \frac{\lambda}{2n} x_i^2 x_j^2$ is the Hamiltonian.

We will be interested in computing the asymptotic behavior of the free energy. The limit $F_n(\lambda)/n$ has been computed, first heuristically in the statistical physics literature, and then it was rigorously proved by various authors. It takes the following "replica-symmetric" form:

Theorem 6. *Assume that P_0 has bounded support. Then for all $\lambda \ge 0$,*

$$\lim_{n\to\infty} \frac{F_n(\lambda)}{n} = \sup_{q \ge 0} \Big\{ \psi(\lambda q) - \frac{\lambda}{4} q^2 \Big\},$$

where $\psi(r)$ is the free energy associated to the scalar Gaussian channel $y = \sqrt{r} x_0 + z$, where $x_0 \sim P_0$ and $z \sim N(0,1)$ independently. In other words,

$$\psi(r) := \mathbb{E} \log \int e^{\sqrt{r} z x + r x x_0 - r x^2/2} \mathrm{d}P_0(x).$$

3.2 Heuristic derivation

It is instructive to first derive the formula on heuristic grounds, but with an argument which has a chance of being made rigorous. We choose to do this via a variant of the *cavity method*. The first idea is that is suffices to compute the limit of the difference $F_{n+1}(\lambda) - F_n(\lambda)$. (Indeed, if a sequence of real numbers (a_n) converges to a limit ℓ, then its running average $(\frac{1}{n}\sum_{i=1}^n a_i)$ will also converge to ℓ.)

Let's consider an experiment where independent random variables $x_{0i} \sim P_0$, $1 \le i \le n+1$ and $W_{ij} \sim N(0,1)$, $1 \le i < j \le n+1$ have been drawn in advance. The observation model in n dimensions is $Y_{ij}^{(n)} = \sqrt{\frac{\lambda}{n}} x_{0i} x_{0j} + W_{ij}$, for $1 \le i < j \le n$, and the model in $n+1$ dimensions is $Y_{ij}^{(n+1)} = \sqrt{\frac{\lambda}{n+1}} x_{0i} x_{0j} + W_{ij}$, for $1 \le i < j \le n+1$. Crucially, the two matrices share the same realizations of $(x_{0i})_{i=1}^n$ and $(W_{ij})_{1 \le < j \le n}$. We will write $x = (x_i)_{i=1}^n \in \mathbb{R}^n$ and $x_{n+1} = \varepsilon$,

$x_{0n+1} = \varepsilon_0$. The free energy difference can be written as

$$\begin{aligned}
F_{n+1}(\lambda) - F_n(\lambda) &= \mathbb{E}\log \frac{Z_{n+1}(\lambda)}{Z_n(\lambda)} \\
&= \mathbb{E}\log\Big(\frac{\int e^{H_{n+1}(x,\varepsilon)}\mathrm{d}P_0(\varepsilon)\mathrm{d}P_0^n(x)}{\int e^{H_n(x)}\mathrm{d}P_0^n(x)}\Big) \\
&= \mathbb{E}\log\Big(\underbrace{\frac{\int e^{H_{n+1}(x,\varepsilon)}\mathrm{d}P_0(\varepsilon)\mathrm{d}P_0^n(x)}{\int e^{\widetilde{H}_{n+1}(x,\varepsilon)}\mathrm{d}P_0(\varepsilon)\mathrm{d}P_0^n(x)}}_{A}\cdot\underbrace{\frac{\int e^{\widetilde{H}_{n+1}(x,\varepsilon)}\mathrm{d}P_0(\varepsilon)\mathrm{d}P_0^n(x)}{\int e^{H_n(x)}\mathrm{d}P_0^n(x)}}_{B}\Big) \\
&= \mathbb{E}\log A + \mathbb{E}\log B,
\end{aligned}$$

where

$$\begin{aligned}
H_n(x) &:= \sum_{1\le i<j\le n} \sqrt{\frac{\lambda}{n}}Y_{ij}^{(n)}x_ix_j - \frac{\lambda}{2n}x_i^2x_j^2 \\
&= \sum_{1\le i<j\le n} \sqrt{\frac{\lambda}{n}}W_{ij}x_ix_j + \frac{\lambda}{n}x_ix_jx_{0i}x_{0j} - \frac{\lambda}{2n}x_i^2x_j^2, \\
H_{n+1}(x,\varepsilon) &:= \sum_{1\le i<j\le n+1} \sqrt{\frac{\lambda}{n+1}}Y_{ij}^{(n+1)}x_ix_j - \frac{\lambda}{2(n+1)}x_i^2x_j^2 \\
&= \sum_{1\le i<j\le n} \sqrt{\frac{\lambda}{n+1}}W_{ij}x_ix_j + \frac{\lambda}{n+1}x_ix_jx_{0i}x_{0j} - \frac{\lambda}{2(n+1)}x_i^2x_j^2 \\
&\quad + \sum_{i=1}^{n} \sqrt{\frac{\lambda}{n+1}}W_{i,n+1}x_i\varepsilon + \frac{\lambda}{n+1}x_ix_{0i}\varepsilon\varepsilon_0 - \frac{\lambda}{2(n+1)}x_i^2\varepsilon^2,
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{H}_{n+1}(x,\varepsilon) &:= \sum_{1\le i<j\le n} \sqrt{\frac{\lambda}{n}}W_{ij}x_ix_j + \frac{\lambda}{n}x_ix_jx_{0i}x_{0j} - \frac{\lambda}{2n}x_i^2x_j^2 \\
&\quad + \sum_{i=1}^{n} \sqrt{\frac{\lambda}{n}}W_{i,n+1}x_i\varepsilon + \frac{\lambda}{n}x_ix_{0i}\varepsilon\varepsilon_0 - \frac{\lambda}{2n}x_i^2\varepsilon^2.
\end{aligned}$$

Notice that the only difference between H_{n+1} and $\widetilde{H}_{n+1}$ is the way λ is normalized. Let $\lambda_n := \lambda(1+\frac{1}{n})$. Then

$$\mathbb{E}\log A = F_{n+1}(\lambda) - F_{n+1}(\lambda_n).$$

We perform a Taylor expansion to first order in λ to obtain

$$\mathbb{E}\log A = F'_{n+1}(\lambda)\cdot(\lambda-\lambda_n) + o_n(1) = -\frac{\lambda}{n}F'_{n+1}(\lambda) + o_n(1).$$

We stopped at the first order because the second derivative of F_{n+1} is of order n as cam be seen from Eq. (1), and therefore the expansion error is actually $O(1/n)$. Now it remains to compute the derivative of F_{n+1}. Using the I-MMSE relation (Lemma 1), we have

$$\begin{aligned} F'_{n+1}(\lambda) &= \frac{1}{2(n+1)} \sum_{1\le i<j\le n+1} \mathbb{E}\left[\mathbb{E}[x_i x_j | Y^{(n+1)}]^2\right] \\ &= \frac{1}{2(n+1)} \sum_{1\le i<j\le n+1} \mathbb{E}\left[\langle x_i x_j\rangle^2\right] \\ &= \frac{n+1}{4}\mathbb{E}\left\langle\Big(\frac{1}{n+1}\sum_{i=1}^{n+1} x_i^1 x_i^2\Big)^2\right\rangle - \frac{1}{4(n+1)}\sum_{i=1}^{n+1}\mathbb{E}\left[\langle x_i^2\rangle^2\right], \end{aligned}$$

where x^1 and x^2 are two random vectors drawn independently from the posterior measure $\mathbb{P}(\cdot|Y^{n+1})$. The brackets in the first term in the above expression is the average with respect to the product measure $\mathbb{P}(\cdot|Y^{n+1})^{\otimes 2}$. Since we are considering a prior P_0 of bounded support, the second term in the above expression is bounded by a constant.

Now, the main unverified assumption we will make is that the normalized overlap $R_{1,2} = \frac{1}{n+1}\sum_{i=1}^{n+1} x_i^1 x_i^2$ between two replicas drawn from the posterior measure concentrates around a deterministic value $q \ge 0$. This is referred to as a *replica-symmetric* (RS) assumption. It follows from this assumption that

$$\mathbb{E}\log A \longrightarrow -\frac{\lambda}{4}q^2.$$

Note that we have not used the full strength of the RS assumption, as it would have sufficed to assume that $\mathbb{E}\langle R_{1,2}^2\rangle$ converges to a constant; a much weaker assumption. But the RS assumption will also be used is a more non trivial way in the computation of $\mathbb{E}\log B$, which we now undertake. We have

$$\begin{aligned} B &= \frac{\int e^{\widetilde{H}_{n+1}(x,\varepsilon)}\mathrm{d}P_0(\varepsilon)\mathrm{d}P_0^n(x)}{\int e^{H_n(x)}\mathrm{d}P_0^n(x)} \\ &= \left\langle \int e^{h_{n+1}(x,\varepsilon)}\mathrm{d}P_0(\varepsilon)\right\rangle_n, \end{aligned}$$

where

$$h_{n+1}(x,\varepsilon) := \sum_{i=1}^n \sqrt{\frac{\lambda}{n}}W_{i,n+1}x_i\varepsilon + \frac{\lambda}{n}x_i x_{0i}\varepsilon\varepsilon_0 - \frac{\lambda}{2n}x_i^2\varepsilon^2.$$

Now the pair (x^1, x^2) has the same law as (x_0, x), the RS assumption also implies that the overlap $R_{1,0} = \frac{1}{n}\sum_{i=1}^n x_i x_0$ concentrates around q. So the middle term in h_{n+1} can be replaced by $\lambda q\varepsilon\varepsilon_0$. Next, consider the Gaussian process

$$G_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n W_{i,n+1}x_i, \quad x\in\mathbb{R}^n.$$

It acts as a local magnetic field which "fills the cavity left by taking the variable $x_{n+1} = \varepsilon$ out of system", hence the name *cavity method*. This process is independent of $Y^{(n)}$, it is centered and it has covariance

$$\mathbb{E}\left[G_n(x^1)G_n(x^2)\right] = \frac{1}{n}\sum_{i=1}^n x_i^1 x_i^2 = R_{1,2}.$$

So under the RS assumption, we should expect that $G_n(x)$ can be replaced by a univariate Gaussian $\sqrt{q}z$, where $z \sim N(0,1)$, i.e., the field felt by x_{n+1} is Gaussian with a constant variance. Hence we can replace the term $\sum_{i=1}^n \sqrt{\frac{\lambda}{n}}W_{i,n+1}x_i\varepsilon - \frac{\lambda}{2n}x_i^2\varepsilon^2$ by $\sqrt{\lambda q}z\varepsilon - \lambda q\varepsilon^2/2$. (Note that the last term in both expressions acts as a normalization which ensure that the exponential has expectation one.) From this is deduce that under our RS assumption we have,

$$\mathbb{E} \log B \longrightarrow \mathbb{E} \log \int e^{\sqrt{\lambda q} z\varepsilon + \lambda q \varepsilon \varepsilon_0 - \lambda q \varepsilon^2/2} \mathrm{d}_0(\varepsilon) = \psi(\lambda q).$$

Putting things together we obtain the limit of the free energy:

$$\lim_{n\to\infty} \frac{F_n(\lambda)}{n} = \lim_{n\to\infty} F_{n+1}(\lambda) - F_n(\lambda) = \psi(\lambda q) - \frac{\lambda}{4} q^2,$$

where the value q is determined by the RS assumption: $x^\top x_0/n \to q$ (say, in probability) when $x \sim \mathbb{P}(\cdot|Y^{(n)})$.