CPSC 664, Spring 2023 — Scribe: Felix Zhou
February 14, 2023 — Lecturer: Tim Kunisky

LECTURE #9: FIXME

1 A Model

Recall the Kac-Rice formula for counting critical points of "nice" functions (e.g. Gaussian processes).

$$\mathbb{E}\,\mathsf{Crit}(f, A) = \int_A \mathbb{E}\left[\mathsf{det}\, \boldsymbol{\nabla}^2 f(x) \mid \boldsymbol{\nabla} f(x) = 0\right] p_{\boldsymbol{\nabla} f(x)}(0) dx.$$

We now introduce the model which we analyze (c.f. "Elastic Manifold" in physics).

$$f(x) = \frac{\alpha}{2}\|x\|^2 + g(x)$$

where g is the some centered, stationary Gaussian process, which we explain below.

Recall from probability theory that a stochastic process $G_x = g(x)$ is a *Gaussian process (GP)* if every finite-dimensional distribution of G_x satisfies

$$(G_{x_1}, \dots, G_{x_m}) \sim \mathcal{N}(\mu_{x_1,\dots,x_m}, \Sigma_{x_1,\dots,x_m}).$$

For instance, Brownian motion is a Gaussian process.

We will specifically concern ourselves with centered Gaussian processes with a very specific covariance function.

$$\begin{aligned}(G_{x_1}, \dots, G_{x_m}) &\sim \mathcal{N}(0, \Sigma) \\ \Sigma_{ij} &= K(x_i, x_j) \\ &= K(x_i - x_j).\end{aligned}$$

Here $K(\cdot)$ is a *kernel* function (e.g. rbf kernel). This choice of covariance ensures that our Gaussian process is *stationary*, i.e.

$$(g(x))_{x \in \mathbb{R}^N} \overset{d}{=} (g(x - x_0))_{x \in \mathbb{R}^N}.$$

The particular Gaussian process we examine is given by

$$g(x) = \sum_{i_1,\dots,i_d \in [N], s_1,\dots,s_d \in \{0,1\}} \prod_{a=1}^{d} \left[\mathsf{cos}(x_{i_a}) \mathbb{1}\{s_i = 0\} + \mathsf{sin}(x_{i_a}) \mathbb{1}\{s_i = 1\}\right] W_{i_1,\dots,i_d,s_1,\dots,s_d}.$$

Here $W_{i_1,\dots,i_d,s_1,\dots,s_d} \sim \mathcal{N}(0,1)$ i.i.d.

2 Applying the Kac-Rice Formula

In order to apply the Kac-Rice formula, we will need up to compute second-order information about $f(x)$. As a preliminary, let us first consider $g(x)$. We have

$$\begin{aligned}
K(x,y) &= \mathbb{E} g(x) g(y) \\
&= \sum_{i_1,\ldots,i_d,s_1,\ldots,s_d} \prod_{a=1}^{d} \left[\cos(x_{i_a})\cos(y_{i_a})\mathbb{1}\{s_a = 0\} + \sin(x_{i_a})\sin(y_{i_a})\mathbb{1}\{s_a = 1\}\right] \\
&= \left[\sum_{i=1}^{N} \cos(x_i)\cos(y_i) + \sin(x_i)\sin(y_i)\right]^d \\
&= \left[\sum_{i=1}^{N} \cos(x_i - y_i)\right]^d \\
&=: S(x,y)^d
\end{aligned}$$

The second equality comes from staring at the terms in the expanded summation and realizing that if any indices are not precisely the same, then the expectation of the term is 0.

Note that this computation shows that $g(x)$ is indeed stationary. In particular, for $x = y$, we have $K(x,x) = N^d$,

In order to apply the Kac-Rice formula, we need to understand the joint distribution of $(f(x), \nabla f(x), \nabla^2 f(x)) \in \mathbb{R} \times \mathbb{R}^N \times \mathbb{S}^N$. It is known that the derivative of a centered Gaussian process with differentiable kernel is another Gaussian process whose kernel is just the derivative of original kernel. We have

$$\begin{aligned}
f(x) &= \frac{\alpha}{2}\|x\|^2 + g(x) \\
\nabla f(x) &= \alpha x + \nabla g(x) \\
\nabla^2 f(x) &= \alpha I_N + \nabla^2 g(x).
\end{aligned}$$

This is a "massive Gaussian vector" say with parameters (μ_x, Σ_x). By inspection,

$$\begin{aligned}
\mu_x &= \left(\frac{\alpha}{2}\|x\|^2, \alpha x, \alpha I_N\right) \\
\Sigma_x &= \mathsf{Cov}\left(g(x), \nabla g(x), \nabla^2 g(x)\right).
\end{aligned}$$

It remains to compute these covariances depicted in Figure 1.

2.1 Covariance Computations

Now, taking an expectation is simply an integral. In the case of "nice" functions, we know from elementary calculus that we are able to exchange the order of the integral and differential. Staring long enough yields the following identity.

$$\mathbb{E}\left[\frac{\partial^a g}{\partial x_{i_1} \ldots \partial x_{i_a}}(x) \frac{\partial^b g}{\partial y_{j_1} \ldots \partial y_{j_b}}(y)\right] = \frac{\partial^{a+b} K}{\partial x_{i_1} \ldots \partial x_{i_a} \partial y_{j_1} \ldots \partial y_{j_b}}(x,y).$$

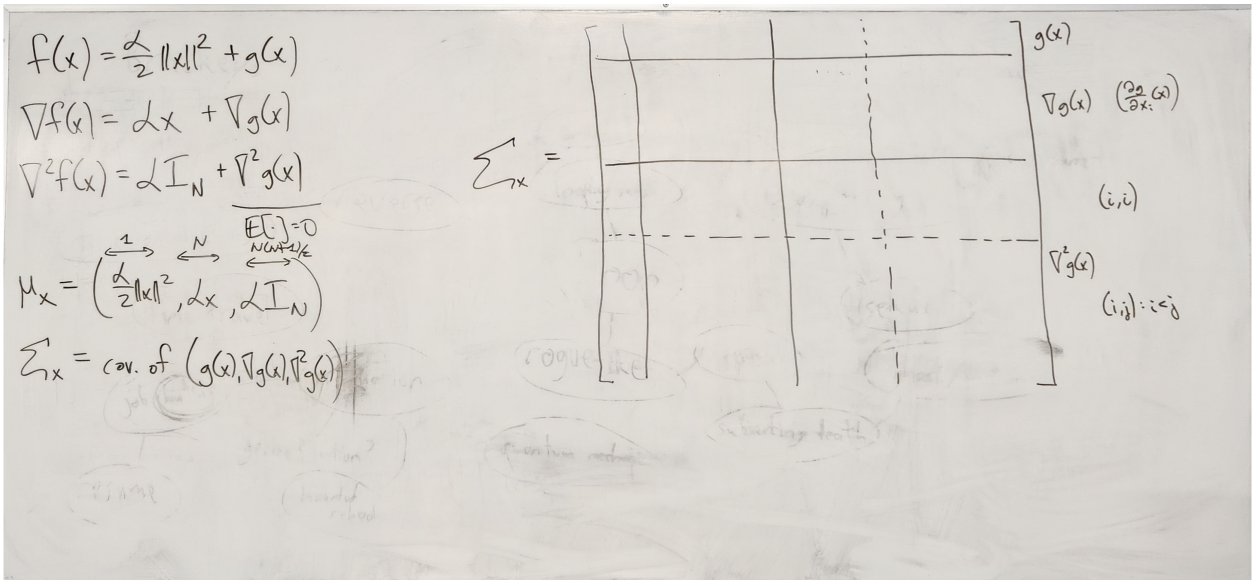

Figure 1: A big covariance matrix.

Using the identity above, let us compute a few blocks of the big covariance matrix.

We begin with $\mathsf{Cov}[g(x), g(x)]$. This corresponds to applying the identity with $a = b = 0$. we already know that

$$\mathbb{E}g(x)^2 = K(x, x) = N^d.$$

Now, let us consider $\mathsf{Cov}[g(x), \boldsymbol{\nabla} g(x)]$. This corresponds to applying the identity with $a = 0, b = 1$. We have

$$\begin{aligned} \mathbb{E}g(x)\frac{\partial g}{\partial x_j} &= \frac{\partial K}{\partial y_j}(x, y)\Big|_{x=y} \\ &= dS(x, y)^{d-1}\left[-\mathsf{sin}(x_j - y_j)\right](-1)\Big|_{x=y} \\ &= 0. \end{aligned}$$

Another example is the $\mathsf{Cov}[\boldsymbol{\nabla} g(x), \boldsymbol{\nabla} g(x)]$. We apply the identity with $a = 1, b = 1$ and deduce that

$$\begin{aligned} &\mathbb{E}\frac{\partial g}{\partial x_i}(x)\frac{\partial g}{\partial x_j}(x) \\ &= \frac{\partial^2 K}{\partial x_i \partial y_j}(x, y)\Big|_{x=y} \\ &= d(d-1)S(x, y)^{d-2}\,\mathsf{sin}(x_j - y_j)\left[-\mathsf{sin}(x_i - y_i)\right] \cdot \mathbb{1}\{i = j\} dS(x, y)^{d-1}\,\mathsf{cos}(x_i - y_i)\Big|_{x=y}. \end{aligned}$$

The extra term occurring when $i = j$ corresponds to $\frac{\partial}{\partial x_i}\mathsf{sin}(x_j - y_j)$.

Continuing with these computations eventually fills the blocks of our big covariance matrix as in Figure 2.

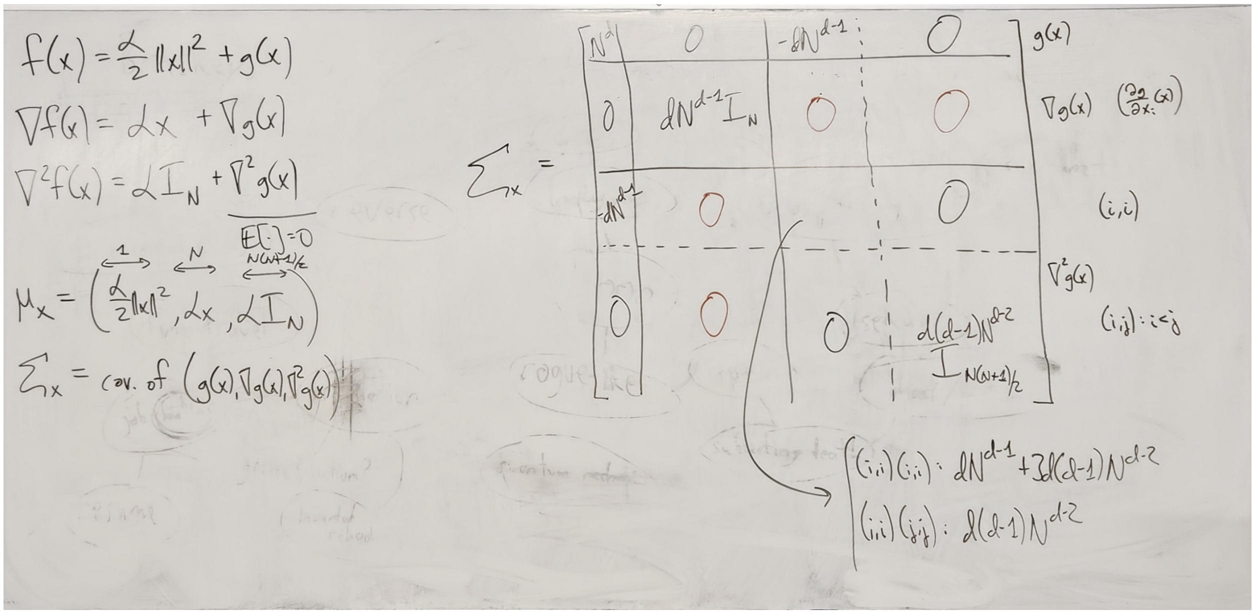

Figure 2: A filled out covariance matrix.

Remark 2.1. *The following hold.*

1. $\boldsymbol{\nabla} g(x), \boldsymbol{\nabla}^2 g(x)$ *are independent.*
2. *The joint distribution of* $(g(x), \boldsymbol{\nabla} g(x), \boldsymbol{\nabla}^2 g(x))$ *is identical for all* x*.*
3. *The joint distribution of* $\boldsymbol{\nabla}^2 f(x)$ *is identical for all* x*.*

Returning to the Kac-Rice formula, we have

$$\mathbb{E}\left[\boldsymbol{\nabla}^2 f(x) \mid \boldsymbol{\nabla} f(x) = 0\right] = \mathbb{E}\left[\boldsymbol{\nabla}^2 f(x)\right]$$

by independence,. Moreover, the Hessian does not depend on the value of x and we can pull this entire term out of the integral. It follows that there is some random matrix H for which

$$\begin{aligned} \mathbb{E}\,\mathsf{Crit}(f, A) &= \mathbb{E}_H\left[|\mathsf{det}(H)|\right] \int_A p_{\mathcal{N}(\alpha x, dN^{d-1} I_N)}(0, \ldots, 0)dx \\ &= \mathbb{E}_H\left[|\mathsf{det}(H)|\right] \int_A p_{\mathcal{N}(\alpha x, \sigma^2 I_N)}(0, \ldots, 0)dx. \qquad \sigma^2 = dN^{d-1} \end{aligned}$$

Let us ignore the constant term and focus on the integral.

$$\begin{aligned}
&\int_A \left[\mathsf{det}\left(2\pi\sigma^2 I\right)\right]^{-\frac{1}{2}} \mathsf{exp}\left[(\alpha x)^T(\sigma^2 I)^{-1}(\alpha x)/2\right] dx \\
&= \int_A \frac{1}{\alpha^N}\left[\mathsf{det}\left(2\pi\sigma^2/\alpha^2 I\right)\right]^{-\frac{1}{2}} \mathsf{exp}\left[(\alpha x)^T(\sigma^2/\alpha^2 I)^{-1}(\alpha x)/2\right] dx \\
&= \frac{1}{\alpha^N}\mathbb{P}\{g \in A\} && g \sim \mathcal{N}(0, \sigma^2/\alpha^2 I) \\
&= \frac{1}{\alpha^N}\mathbb{P}\left\{\frac{1}{\alpha} g \in A\right\} && g \sim \mathcal{N}(0, \sigma^2 I) \\
&= \frac{1}{\alpha^N}\mathbb{P}\{g \in \alpha A\}. && g \sim \mathcal{N}(0, \sigma^2 I).
\end{aligned}$$

As $\alpha \to 0$, we have

$$\frac{1}{\alpha^N}\,\mathsf{vol}(\alpha A) = \frac{1}{\alpha^N}\alpha^N\,\mathsf{vol}(A) = \mathsf{vol}(A).$$

As $\alpha \to \infty$, since probability measures are normalized,

$$\frac{1}{\alpha^N}\,\mathsf{vol}(\alpha A) \to 0.$$

This agrees with out intuition that the number of critical points should tend to 0 as we increase the parameter α. Furthermore, at $\alpha = 0$, the number of critical points is completely determined by the Gaussian volume of A, which again agrees with out intuition.

3 An Explicit Random Matrix Model

It remains to explicitly construct a random matrix H which has the desired covariance structure. We claim that it suffices to take

$$H = \alpha I_n + \sqrt{d(d-1)N^{d-2}}hI_N + \sqrt{d(d-1)N^{d-1}}W + \sqrt{dN^{d-1}}D.$$

Here $h \sim \mathcal{N}(0,1)$ independently of W, D and D, W are diagonal, symmetric matrices, respectively, satisfying

$$\begin{aligned}
D_{ii} &\overset{iid}{\sim} \mathcal{N}(0,1) \\
W_{ji} = W_{ij} &\overset{iid}{\sim} \mathcal{N}(0,1/N) && i < j \\
W_{ii} &\overset{iid}{\sim} \mathcal{N}(0,2/N).
\end{aligned}$$

Thus W is drawn from the so called *Gaussian Orthogonal Ensemble.*

As a sanity check, we have

$$\begin{aligned}
\mathsf{Var}[H_{ii}] &= d(d-1)N^{d-2} + d(d-1)N^{d-1}\frac{2}{N} \\
&= dN^{d-1} + 3d(d-1)N^{d-2} \\
\mathsf{Cov}[H_{ii}, H_{jj}] &= d(d-1)N^{d-2}
\end{aligned}$$

as desired.

In order to determine the value of the determinant term we wrote above, we need to determine the product of the spectrum of H. In particular, we would like to understand the eigenvalues of $W + D$.