

Preamble 1: Gaussian Processes for Vector Valued Functions

In this section, we fix some notation for the rest of the document, by describing Gaussian Processes (GPs) for vector valued functions as in Alvarez et al. (*Kernels for vector valued functions: a Review*, 2011).

Recall that a GP can be defined as a stochastic process such that any finite number of random variables taken from a realization of the GP follows a joint Gaussian distribution. In the 1D output case, this characterization implies that for any finite set $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N$, if we let $f(\mathbf{X}) = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)]^T$, then

$$f(\mathbf{X}) \sim \mathcal{N}(m(\mathbf{X}), k(\mathbf{X}, \mathbf{X}))$$

where $m(\mathbf{X}) = [m(\mathbf{x}_1), \dots, m(\mathbf{x}_N)]^T$ and $k(\mathbf{X}, \mathbf{X})$ is a $N \times N$ kernel matrix.

In the multiple output case, the random variables are associated to different processes $\{f_d(\mathbf{X})\}_{d=1}^D$, evaluated at input values \mathbf{X} . Note that here we're making the simplifying assumption that each component has the same cardinality, and the input to all the components are the same, as in our problem of interest. Then, the GP characterization generalizes to

$$f(\mathbf{X}) \sim \mathcal{N}(\mathbf{m}(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X}))$$

where (i) $\mathbf{m}(\mathbf{X})$ is a vector of dimension ND that concatenates the mean vectors associated to the outputs $\{m_d(\mathbf{X})\}_{d=1}^D$, i.e. $\mathbf{m}(\mathbf{X}) = [m_1(\mathbf{x}_1), \dots, m_D(\mathbf{x}_1), \dots, m_1(\mathbf{x}_N), \dots, m_D(\mathbf{x}_N)]^T$, and (ii) $\mathbf{K}(\mathbf{X}, \mathbf{X})$ is a block-partitioned $ND \times ND$ matrix,

$$\mathbf{K}(\mathbf{X}, \mathbf{X}) = \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X})_{1,1} & \dots & \mathbf{K}(\mathbf{X}, \mathbf{X})_{1,D} \\ \mathbf{K}(\mathbf{X}, \mathbf{X})_{2,1} & \dots & \mathbf{K}(\mathbf{X}, \mathbf{X})_{2,D} \\ \vdots & \dots & \vdots \\ \mathbf{K}(\mathbf{X}, \mathbf{X})_{D,1} & \dots & \mathbf{K}(\mathbf{X}, \mathbf{X})_{D,D} \end{bmatrix}$$

where each block $\mathbf{K}(\mathbf{X}, \mathbf{X})_{d,d'}$ is an $N \times N$ kernel matrix such that the entries $\mathbf{K}(\mathbf{x}, \mathbf{x}')_{d,d'}$ correspond to the covariances between the outputs $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$, and express the degree of correlation or similarity between them.

Question: what is the correct order for expressing both $f(\mathbf{X})$ and $m(\mathbf{x})$ as 1D column vectors?

Preamble 2: Helmholtz decomposition

Here, review main concepts about Helmholtz decomposition in a somewhat concise way, following the review paper "<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6365629>".

Framework introduction

In our problem of interest, we are given a 2D vector field $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, mapping spatial locations, (X, Y) , into horizontal and vertical velocities, (U, V) , representing buoys floating across the oceans. In terms of data, we have N observations, at locations $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N = \{(x_1, y_1), \dots, (x_N, y_N)\}$, and the corresponding vector field velocity observations, $F(\mathbf{X}) = \{F(\mathbf{x}_n)\}_{n=1}^N = \{(u_1, v_1), \dots, (u_N, v_N)\}$. Referring to the above notation, we have $D = 2$, and $f_1(\mathbf{x}) := u(\mathbf{x})$, $f_2(\mathbf{x}) := v(\mathbf{x})$. Our ultimate goal is to reconstruct ocean currents given these data, that is, predict the velocity field at spatial locations where we don't observe any buoy, with some measure of uncertainty about the quality of our prediction.

Independently for each n , we model the likelihood for our quantities of interest as

$$(u_n, v_n) \mid (x_n, y_n) \stackrel{\text{ind}}{\sim} \mathcal{N} \left(F(x_n, y_n), \begin{bmatrix} \sigma_{\text{obs}}^2 & 0 \\ 0 & \sigma_{\text{obs}}^2 \end{bmatrix} \right)$$

where σ_{obs}^2 represent the variance of our observations.

In order to characterize the posterior distribution, we need to specify a prior for (u_n, v_n) . Here is where the Helmholtz decomposition comes into play. According to this result, any sufficiently smooth, rapidly decaying vector field in two dimensions can be resolved into the sum of an irrotational (curl-free) scalar vector field and a solenoidal (divergence-free) vector field.

As an irrotational vector field has a scalar potential and a solenoidal vector field has a vector potential, we can formally write down the decomposition as

$$\mathbf{F} = -\nabla\phi + \nabla \times \boldsymbol{\psi}$$

where ϕ is the scalar potential, and $\boldsymbol{\psi}$ the vector potential.

In our work, we model these two quantities **independently**, so that we can naturally specify two distinct GP priors on them, that is

$$\phi \sim \text{GP}(0, K_\phi)$$

$$\boldsymbol{\psi} \sim \text{GP}(0, K_\psi)$$

so that

$$\phi(\mathbf{X}) \sim \mathcal{N}(0, K_\phi(\mathbf{X}, \mathbf{X}))$$

$$\boldsymbol{\psi}(\mathbf{X}) \sim \mathcal{N}(0, K_\psi(\mathbf{X}, \mathbf{X}))$$

We can show (see Theorem below - *when it'll be ready...*) that since (i) the Helmholtz decomposition is based on partial (mixed) derivatives, and (ii) differentiation is a linear operator, this kernel choice implies that the prior on the original vector field \mathbf{F} is a GP as well.

Consequently, if we consider two observations' indices n and n' , and we denote with Cov the *cross-covariance*, we have that

$$\begin{aligned}\text{Cov}((u_n, v_n), (u_{n'}, v_{n'})) &= \text{Cov}(F(x_n, y_n), F(x_{n'}, y_{n'})) = \\ &= \text{Cov}(-\nabla\phi(x_n, y_n) + \nabla \times \psi(x_n, y_n), -\nabla\phi(x_{n'}, y_{n'}) + \nabla \times \psi(x_{n'}, y_{n'})) = \\ &= \text{Cov}(\nabla\phi(x_n, y_n), \nabla\phi(x_{n'}, y_{n'})) + \text{Cov}(\nabla \times \psi(x_n, y_n), \nabla \times \psi(x_{n'}, y_{n'}))\end{aligned}$$

where the last equality follows from the independence of ψ and ϕ . Therefore, we have that our prior on the original vector field F is a GP with the following structure

$$F(\mathbf{X}) \sim \text{GP}(0, \nabla K_\phi(\mathbf{X}) + \nabla \times K_\psi(\mathbf{X}))$$

which is easy to deal with, once we specify the prior kernels on ϕ and ψ .

Inferring ϕ and ψ components of the posterior

Let F be our vector field of interest. Through the Helmholtz decomposition, we write it as

$$F = \nabla\phi + \nabla \times \psi$$

where ϕ is the irrotational, curl-free, component, and ψ a divergence-free vector potential. In our analysis, we evaluate our predictions on a grid of size $N \times N$, and we want to know ϕ and ψ at each point of the grid. That is, for each $n, n' \in \{1, 2, \dots, N\}$, we are interested in $\phi_{n,n'}, \psi_{n,n'} \in \mathbb{R}$, where $\phi_{n,n'} := \phi((x_n, y_n), (x_{n'}, y_{n'}))$, and similarly for $\psi_{n,n'}$.

Our goal is to plot $\nabla\phi \in \mathbb{R}^2$ and $\nabla \times \psi \in \mathbb{R}^2$, for each pair of interest. To achieve this, we simply need to derive the posterior distribution of $\phi \mid u, v$. In other words, we are interested in retrieving $\mathbb{E}[\phi \mid u, v]$ and $\text{Cov}[\phi \mid u, v]$. For simplicity, in this exposition we focus only on the ϕ component, but the following holds in a very similar fashion for ψ as well.

Recall the following basic probability facts.

Fact 1. *If*

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

then $X \mid Y = y \sim \mathcal{N}(\bar{\mu}, \bar{\Sigma})$, with $\bar{\mu} = \Sigma_{12}\Sigma_{22}^{-1}y$ and $\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Fact 2. *If $a \sim \mathcal{N}(\mu_a, \Sigma_a)$ and $a' \sim \mathcal{N}(\mu_{a'}, \Sigma_{a'})$ are two independent random variables, and $b \mid a, a' \sim \mathcal{N}(\mu_a + \mu_{a'}, \Sigma_b)$, then $b \sim \mathcal{N}(\mu_a + \mu_{a'}, \Sigma_a + \Sigma_{a'} + \Sigma_b)$.*

In order to use these in our problem of interest, we need to express the distribution of $\begin{pmatrix} \phi \\ u \\ v \end{pmatrix}$.

We begin considering the random vector ϕ . By construction, ϕ is a vector of size N^2 , with mean $\mathbf{0} \in \mathbb{R}^{N^2}$ and covariance matrix of size $N^2 \times N^2$ determined by K_ϕ .

Next, we consider u, v . Given

$$\begin{bmatrix} u_m \\ v_m \end{bmatrix} \mid \phi, \psi \sim \mathcal{N}(\nabla \phi(x_m, y_m) + \nabla \times \psi(x_m, y_m), \sigma_{\text{obs}}^2 \mathbb{I}_2)$$

we can use fact 2 above with

$$\begin{aligned} \nabla \phi(\mathbf{X}) &\sim \mathcal{N}(0, \nabla K_\phi(\mathbf{X}, \mathbf{X})) \\ \nabla \times \psi(\mathbf{X}) &\sim \mathcal{N}(0, \nabla \times K_\psi(\mathbf{X}, \mathbf{X})) \end{aligned}$$

to get

$$\begin{bmatrix} u(\mathbf{X}) \\ v(\mathbf{X}) \end{bmatrix} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_{2M} + \nabla K_\phi(\mathbf{X}, \mathbf{X}) + \nabla \times K_\psi(\mathbf{X}, \mathbf{X}))$$

with $\begin{pmatrix} u \\ v \end{pmatrix}$ of dimension $2M$, and each of the matrices of dimension $2M \times 2M$, where each entry is obtained by evaluating, e.g., $\nabla K_\phi((x_m, y_m), (x_{m'}, y_{m'}))$.

ISSUE 1.1: I am abusing notation for \mathbf{X} . Here it's of dimension M , data observed, whereas before it was intended of dimension N^2 , predicted grid. Fix this.

ISSUE 1.2: By using the "theorem" below, if we unpack the actual $\nabla K_\phi((x_m, y_m), (x_{m'}, y_{m'}))$ we see that it looks more like a second derivative rather than a first derivative (each entry is a second derivative w.r.t. the corresponding components). Wouldn't it be better to call it H , for Hessian? Rather than using the ∇ symbol for the gradient.

ISSUE 1.3: I only have m points/observations. The covariance matrices have dimension $2M \times 2M$ because the gradient "doubles" the dimension, since it is partial derivative w.r.t. both entries. Question: if I would like to be super explicit in writing down these matrices, how would they be? Alternatively, considering a more computational approach, how are they stored in a computer memory?

This concludes two blocks of the covariance matrix of $\begin{pmatrix} \phi \\ u \\ v \end{pmatrix}$. Now we need to take into account the cross terms, i.e. the cross covariance matrix $\text{Cov}(\phi, \begin{pmatrix} u \\ v \end{pmatrix})$. Since we have N^2 points on the grid, and M observations, this matrix will have size $N^2 \times 2M$, properly fitting in the "hole" in the general covariance matrix of interest.

Overall, we have

$$\begin{aligned} \text{Cov} \left(\phi, \begin{bmatrix} u \\ v \end{bmatrix} \right) &= \text{Cov} \left(\begin{bmatrix} \phi(x_1, y_1) \\ \phi(x_1, y_2) \\ \vdots \\ \phi(x_N, y_{N-1}) \\ \phi(x_N, y_N) \end{bmatrix}, \begin{bmatrix} u(x'_1, y'_1) \\ v(x'_1, y'_1) \\ \vdots \\ u(x'_M, y'_M) \\ v(x'_M, y'_M) \end{bmatrix} \right) = \\ &= \begin{bmatrix} \text{Cov}(\phi(x_1, y_1), u(x'_1, y'_1)) & \dots & \text{Cov}(\phi(x_1, y_1), v(x'_M, y'_M)) \\ \vdots & & \vdots \\ \text{Cov}(\phi(x_N, y_N), u(x'_1, y'_1)) & \dots & \text{Cov}(\phi(x_N, y_N), v(x'_M, y'_M)) \end{bmatrix} \end{aligned}$$

ISSUE 2.1: again, abuse of notation here. We need to be more explicit about the two different X's.

ISSUE 2.2: is it correct the way in which we represent the vector $\begin{pmatrix} u \\ v \end{pmatrix}$? Or should it be first all the u's and then all the v's?

Now we just need to characterize each of these entries, as follows:

$$\begin{aligned} \text{Cov} \left(\phi(x_i, y_i), \begin{bmatrix} u(x'_j, y'_j) \\ v(x'_j, y'_j) \end{bmatrix} \right) &= \text{Cov} \left(\phi(x_i, y_i), F(x'_j, y'_j) \right) = \\ &= \text{Cov} \left(\phi(x_i, y_i), -\nabla \phi(x'_j, y'_j) + \nabla \times \psi(x'_j, y'_j) \right) = \\ &= \text{Cov} \left(\phi(x_i, y_i), -\nabla \phi(x'_j, y'_j) \right) = \\ &= \text{Cov} \left(\phi(x_i, y_i), \begin{bmatrix} \frac{\partial}{\partial x'_j} \phi(x'_j, y'_j) \\ \frac{\partial}{\partial y'_j} \phi(x'_j, y'_j) \end{bmatrix} \right) = \\ &= \begin{bmatrix} \frac{\partial}{\partial x'_j} \text{Cov}(\phi(x_i, y_i), \phi(x'_j, y'_j)) \\ \frac{\partial}{\partial y'_j} \text{Cov}(\phi(x_i, y_i), \phi(x'_j, y'_j)) \end{bmatrix} = \\ &= \begin{bmatrix} \frac{\partial}{\partial x'_j} K_\phi((x_i, y_i), (x'_j, y'_j)) \\ \frac{\partial}{\partial y'_j} K_\phi((x_i, y_i), (x'_j, y'_j)) \end{bmatrix} := \nabla_{D_M} K_\phi((x_i, y_i), (x'_j, y'_j)) \end{aligned}$$

where we used this notation because we refer to the test points (grid) with $D_N := (x_n, y_n)_{n,n'}^N$, and to the observation with $D_M := (x_m, y_m)_{m,m'}^M$.

So overall we have

$$\begin{bmatrix} \phi \\ u \\ v \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_\phi & \nabla_{D_M} K_\phi \\ \nabla_{D_M} K_\phi^T & \sigma^2 \mathbb{I}_{2m} + \nabla K_\phi + \nabla \times K_\psi \end{bmatrix} \right)$$

This vector has size $N^2 + 2M$, with the caveat that ϕ is evaluated at test points D_N whereas $\begin{pmatrix} u \\ v \end{pmatrix}$ at the observations D_M .

We can then, finally, get the quantity of interest, $\phi | u, v$, applying fact 1 above. We obtain

$$\mathbb{E}(\phi | u, v) = \nabla_{D_M} K_\phi (\sigma^2 \mathbb{I}_{2m} + \nabla K_\phi + \nabla \times K_\psi)^{-1} \begin{bmatrix} u \\ v \end{bmatrix}$$

which has dimensionality $N^2 \times 1$, exactly as desired, and

$$\text{Var}(\phi | u, v) = K_\phi - \nabla_{D_M} K_\phi (\sigma^2 \mathbb{I}_{2m} + \nabla K_\phi + \nabla \times K_\psi)^{-1} \nabla_{D_M} K_\phi^T$$

with dimensionality $N^2 \times N^2$, again as desired.

The same exact analysis can then be generalized to ψ , leading to

$$\phi | u, v \sim \mathcal{N}(\bar{\mu}_\psi, \bar{\Sigma}_\psi)$$

with

$$\begin{aligned} \bar{\mu}_\psi &= \nabla_{D_M} K_\psi (\sigma^2 \mathbb{I}_{2m} + \nabla K_\phi + \nabla \times K_\psi)^{-1} \begin{bmatrix} u \\ v \end{bmatrix} \\ \bar{\Sigma}_\psi &= K_\psi - \nabla_{D_M} K_\psi (\sigma^2 \mathbb{I}_{2m} + \nabla K_\phi + \nabla \times K_\psi)^{-1} \nabla_{D_M} K_\psi^T \end{aligned}$$

for $\nabla_{D_M} K_\psi$ defined similarly to what we did for ϕ above. This concludes our analysis for this section.

Final issue: for the ψ , we have the curl, but we want that to be a gradient as well. How can we formalize that? Still not clear to me.

Derivative observations on Gaussian Processes - notation to be fixed, proof to be fixed...

Theorem 1. Consider a Gaussian process with mean μ and covariance kernel K and let ϕ be a differentiable function, mapping \mathbb{R}^2 into \mathbb{R} , drawn from this GP, i.e. $\phi \sim GP(\mu, K)$. Given two pairs of inputs, (x_i, y_i) for $i = 1, 2$, such that $\phi(x_1, y_1) \sim GP(\mu(x_1, y_1), K((x_1, y_1), (x_2, y_2)))$, then $\nabla\phi$ is itself a Gaussian process with mean $\nabla\mu$ and covariance

$$\begin{bmatrix} \frac{\partial^2 K((x_1, y_1), (x_2, y_2))}{\partial x_1 \partial x_2} & \frac{\partial^2 K((x_1, y_1), (x_2, y_2))}{\partial x_1 \partial y_2} \\ \frac{\partial^2 K((x_1, y_1), (x_2, y_2))}{\partial y_1 \partial x_2} & \frac{\partial^2 K((x_1, y_1), (x_2, y_2))}{\partial y_1 \partial y_2} \end{bmatrix}$$

Proof. First of all, observe that since differentiation is a linear operator, the derivative of a Gaussian process is another Gaussian process. In particular, according to Rasmussen and Williams, a covariance function $K(\cdot, \cdot)$ on function values implies the following covariance between function values and partial derivatives, and between partial derivatives:

$$\begin{aligned} \text{Cov}\left(f_i, \frac{\partial f_j}{\partial x_{dj}}\right) &= \frac{\partial K(x_i, x_j)}{\partial x_{dj}} \\ \text{Cov}\left(\frac{\partial f_i}{\partial x_{di}}, \frac{\partial f_j}{\partial x_{dj}}\right) &= \frac{\partial^2 K(x_i, x_j)}{\partial x_{di} \partial x_{dj}} \end{aligned}$$

If we apply this result in our setting, where $\nabla\phi = \begin{bmatrix} \frac{\partial\phi}{\partial x} & \frac{\partial\phi}{\partial y} \end{bmatrix}^T$, we obtain:

$$\begin{aligned} \text{Cov}(\nabla\phi(x_1, y_1), \nabla\phi(x_2, y_2)) &= \text{Cov}\left(\begin{bmatrix} \frac{\partial\phi(x_1, y_1)}{\partial x_1} \\ \frac{\partial\phi(x_1, y_1)}{\partial y_1} \end{bmatrix}, \begin{bmatrix} \frac{\partial\phi(x_2, y_2)}{\partial x_2} \\ \frac{\partial\phi(x_2, y_2)}{\partial y_2} \end{bmatrix}\right) = \\ &= \begin{bmatrix} \text{Cov}\left(\frac{\partial\phi(x_1, y_1)}{\partial x_1}, \frac{\partial\phi(x_2, y_2)}{\partial x_2}\right) & \text{Cov}\left(\frac{\partial\phi(x_1, y_1)}{\partial x_1}, \frac{\partial\phi(x_2, y_2)}{\partial y_2}\right) \\ \text{Cov}\left(\frac{\partial\phi(x_1, y_1)}{\partial y_1}, \frac{\partial\phi(x_2, y_2)}{\partial x_2}\right) & \text{Cov}\left(\frac{\partial\phi(x_1, y_1)}{\partial y_1}, \frac{\partial\phi(x_2, y_2)}{\partial y_2}\right) \end{bmatrix} = \\ &= \begin{bmatrix} \frac{\partial^2 K((x_1, y_1), (x_2, y_2))}{\partial x_1 \partial x_2} & \frac{\partial^2 K((x_1, y_1), (x_2, y_2))}{\partial x_1 \partial y_2} \\ \frac{\partial^2 K((x_1, y_1), (x_2, y_2))}{\partial y_1 \partial x_2} & \frac{\partial^2 K((x_1, y_1), (x_2, y_2))}{\partial y_1 \partial y_2} \end{bmatrix} \end{aligned}$$

as desired. □