

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} . Define $f(\mathbf{X}) = [f_1(\mathbf{X}), \dots, f_D(\mathbf{X})]^T$, a $N \times D$ dimensional vector, where each entry $f_d(\mathbf{X})$ is N -dimensional as in the 1D framework above. 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. $\mathbf{m}(\mathbf{X}) = [m_1(\mathbf{x}_1), \dots, m_1(\mathbf{x}_N), m_2(\mathbf{x}_1), \dots, m_D(\mathbf{x}_1), \dots, m_D(\mathbf{x}_N)]^T$, and (ii) $\mathbf{K}(\mathbf{X}, \mathbf{X})$ is a block-partitioned $ND \times ND$ matrix,

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

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.

Preamble 2: Helmholtz decomposition

Here, review main concepts about Helmholtz decomposition in a somewhat concise way, following [this](#) review paper.

Gaussian Processes for 2D oceans vector fields

In our problem of interest, we are given a 2D vector field $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, mapping spatial locations, (x^1, x^2) , into horizontal and vertical velocities, (F^1, F^2) , representing buoys floating across the oceans. In terms of data, we have M observations, at locations $\mathbf{X} = \{\mathbf{x}_m\}_{m=1}^M = \{(x_1^1, x_1^2), \dots, (x_M^1, x_M^2)\}$, and the corresponding vector field velocity observations, $\mathbf{F}(\mathbf{X}) = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_M] := [(F^1(\mathbf{x}_1), F^2(\mathbf{x}_1)), \dots, (F^1(\mathbf{x}_M), F^2(\mathbf{x}_M))]$, where $F^1(\mathbf{x}_m)$ represents the horizontal velocity at location \mathbf{x}_m , and $F^2(\mathbf{x}_m)$ the vertical one.

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 m , we model the likelihood for our quantities of interest as

$$(F^1(\mathbf{x}_m), F^2(\mathbf{x}_m)) \stackrel{\text{ind}}{\sim} \mathcal{N}\left(\mathbf{F}_m, \begin{bmatrix} \sigma_{\text{obs}}^2 & 0 \\ 0 & \sigma_{\text{obs}}^2 \end{bmatrix}\right)$$

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

In order to characterize the posterior distribution, we need to specify a prior for \mathbf{F}_m . 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 $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ is the scalar potential, and $\boldsymbol{\psi}$ the vector potential. In abuse of notation, we consider $\boldsymbol{\psi} : \mathbb{R}^2 \rightarrow \mathbb{R}$, and define $(\nabla \times \boldsymbol{\psi})(\mathbf{x}) := [\partial\boldsymbol{\psi}(\mathbf{x})/\partial x^2, -\partial\boldsymbol{\psi}(\mathbf{x})/\partial x^1]^T$.

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

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

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

Note that - to ease notation - from now on we will denote this kernel matrices with $K_\phi(\mathbf{X}) := K_\phi(\mathbf{X}, \mathbf{X})$.

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 F is a GP as well. To formalize this statement, we need the following definition and lemma.

Definition 1. L is a linear operator if $\forall c \in \mathbb{R}$, and $F, F' \in L^2$, $L(cF + F') = cL(F) + L(F')$.

Lemma 1. Let L be a linear operator. If $F \sim GP(\mu, K)$, then $L(F) \sim GP(\mu', K')$ for some (μ', K') .

Proof. Proof (in L^2). □

Corollary 1. Let $F \sim GP(\mu, K)$. Then $\nabla F \sim GP(\nabla\mu, \nabla K)$, where $\nabla\mu = \dots$, $\nabla K = \dots$.

Proof. Use above lemma plus Rasmussen plus theorem at the end of this document (that will then be useless?). □

As a consequence of this, if we consider two observations' indices m and m' , and we denote with Cov the *cross-covariance*, we have that

$$\begin{aligned} \text{Cov}(\mathbf{F}_m, \mathbf{F}_{m'}) &= \text{Cov}(-\nabla\phi(\mathbf{x}_m) + \nabla \times \psi(\mathbf{x}_m), -\nabla\phi(\mathbf{x}_{m'}) + \nabla \times \psi(\mathbf{x}_{m'})) \\ &= \text{Cov}(\nabla\phi(\mathbf{x}_m), \nabla\phi(\mathbf{x}_{m'})) + \text{Cov}(\nabla \times \psi(\mathbf{x}_m), \nabla \times \psi(\mathbf{x}_{m'})) \end{aligned}$$

where the last equality follows from the independence of ψ and ϕ .

Furthermore, the following result on the mean of a Gaussian process can be shown.

Lemma 2. Let $c \in \mathbb{R}$, $F \sim GP(\mu, K)$, and $F' \sim GP(\mu', K')$. Under "some" (which?) assumptions, $cF + F'$ is a GP itself, with mean $c\mu + \mu'$.

Proof. even though trivial, for this we might need some assumptions on when to switch expectation and derivative? maybe this is an issue of the lemma above actually... The condition should be something like Lipschitz continuity with high probability. Can't we just use the above lemma on linear operators? □

Putting everything together, our prior on the original vector field F is a GP with the following structure

$$F(\mathbf{X}) \sim GP(0, \nabla K_\phi(\mathbf{X}) + \nabla \times K_\psi(\mathbf{X})) \quad (1)$$

where we use $\nabla K_\phi(\mathbf{X})$ as a shorthand for the $2N \times 2N$ matrix with entries $\nabla K_\phi(\mathbf{X})_{n,n'} := \text{Cov}\left(\frac{\partial}{\partial x_n^1}\phi(\mathbf{x}_n), \frac{\partial}{\partial x_{n'}^1}\phi(\mathbf{x}_{n'})\right)$, $\nabla K_\phi(\mathbf{X})_{n,N+n'} := \text{Cov}\left(\frac{\partial}{\partial x_n^1}\phi(\mathbf{x}_n), \frac{\partial}{\partial x_{n'}^2}\phi(\mathbf{x}_{n'})\right)$, $\nabla K_\phi(\mathbf{X})_{N+n,n'} := \text{Cov}\left(\frac{\partial}{\partial x_n^2}\phi(\mathbf{x}_n), \frac{\partial}{\partial x_{n'}^1}\phi(\mathbf{x}_{n'})\right)$, and $\nabla K_\phi(\mathbf{X})_{N+n,N+n'} := \text{Cov}\left(\frac{\partial}{\partial x_n^2}\phi(\mathbf{x}_n), \frac{\partial}{\partial x_{n'}^2}\phi(\mathbf{x}_{n'})\right)$,

for $n, n' \in [N]$.

Note: little abuse of notation in the derivative of second component in the covariance terms above. It should be more something like $\frac{\partial}{\partial x_n'^2} \phi(\mathbf{x}_n')$.

Inferring ϕ and ψ components of the posterior

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

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

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 $G \times G := N$, and we want to know ϕ and ψ at each point of the grid, e.g. for visualization purposes. That is, we are interested in $\phi \in \mathbb{R}^N$, where $\phi = [\phi_1, \dots, \phi_N]$, $\phi_n := \phi(\mathbf{x}_n)$ for each $n \in [N]$, and similarly for $\psi \in \mathbb{R}^n$.

Our goal is to plot $\nabla\phi \in \mathbb{R}^2$ and $\nabla \times \psi \in \mathbb{R}^2$, for each grid point of interest. To achieve this, we need to derive the posterior distribution of $\phi \mid \mathbf{F}'$, where $\mathbf{F}' := [F^1(\mathbf{x}'_1), \dots, F^1(\mathbf{x}'_M), F^2(\mathbf{x}'_1), \dots, F^2(\mathbf{x}'_M)] \in \mathbb{R}^{2M}$ is a vector containing the observations of our vector field F at M given spatial points. In other words, we are simply interested in retrieving $\mathbb{E}[\phi \mid \mathbf{F}']$ and $\text{Cov}[\phi \mid \mathbf{F}']$ from $\begin{pmatrix} \phi \\ \mathbf{F}' \end{pmatrix}$. For simplicity, in this exposition we focus only on the ϕ component, but at the end of the section we will show how these results can be adapted to the ψ component.

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 $\begin{pmatrix} \phi \\ \mathbf{F} \end{pmatrix}$.

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

Next, we consider \mathbf{F}' . We have

$$\mathbf{F}'_m \mid \phi, \psi \sim \mathcal{N}(\nabla\phi(\mathbf{x}_m) + \nabla \times \psi(\mathbf{x}_m), \sigma_{\text{obs}}^2 \mathbb{I}_2)$$

where \mathbb{I}_2 represents the 2×2 identity matrix. Then, we can use fact 2 above with

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

to get

$$\mathbf{F}' \sim \mathcal{N}\left(0, \sigma_{\text{obs}}^2 \mathbb{I}_{2M} + \nabla K_\phi(\mathbf{X}) + \nabla \times K_\psi(\mathbf{X})\right)$$

where each of the matrices has dimension $2M \times 2M$, and the entries are obtained as in Equation (1) from previous section.

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

Overall, we have

$$\begin{aligned}\text{Cov}(\phi, \mathbf{F}') &= \text{Cov}\left(\begin{bmatrix} \phi(\mathbf{x}_1) \\ \phi(\mathbf{x}_2) \\ \vdots \\ \phi(\mathbf{x}_N) \end{bmatrix}, \begin{bmatrix} [F^1(\mathbf{x}'_1)] \\ \vdots \\ F^1(\mathbf{x}'_M) \\ F^2(\mathbf{x}'_1) \\ \vdots \\ F^2(\mathbf{x}'_M) \end{bmatrix}\right) = \\ &\begin{bmatrix} \text{Cov}(\phi(\mathbf{x}_1), F^1(\mathbf{x}'_1)) & \dots & \text{Cov}(\phi(\mathbf{x}_1), F^2(\mathbf{x}'_M)) \\ \vdots & \vdots & \vdots \\ \text{Cov}(\phi(\mathbf{x}_n), F^1(\mathbf{x}'_1)) & \dots & \text{Cov}(\phi(\mathbf{x}_n), F^2(\mathbf{x}'_M)) \end{bmatrix}\end{aligned}$$

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

$$\begin{aligned}\text{Cov}(\phi(\mathbf{x}_n), F^1(\mathbf{x}'_m)) &= \text{Cov}\left(\phi(\mathbf{x}_n), -\frac{\partial}{\partial x_m^1} \phi(\mathbf{x}'_m) + \frac{\partial}{\partial x_m^2} \psi(\mathbf{x}'_m)\right) \\ &\# \text{ by independence of } \phi \text{ and } \psi \\ &= \text{Cov}\left(\phi(\mathbf{x}_n), -\frac{\partial}{\partial x_m^1} \phi(\mathbf{x}'_m)\right) \\ &= -\frac{\partial}{\partial x_m^1} \text{Cov}(\phi(\mathbf{x}_n), \phi(\mathbf{x}'_m)) \\ &\# \text{ by lemma above (cite number)} \\ &= -\frac{\partial}{\partial x_m^1} K_\phi(\mathbf{x}_n, \mathbf{x}'_m)\end{aligned}$$

and similarly

$$\begin{aligned}\text{Cov}(\phi(\mathbf{x}_n), F^2(\mathbf{x}'_m)) &= \text{Cov}\left(\phi(\mathbf{x}_n), -\frac{\partial}{\partial x_m^2}\phi(\mathbf{x}'_m) - \frac{\partial}{\partial x_m^1}\psi(\mathbf{x}'_m)\right) \\ &= -\frac{\partial}{\partial x_m^2}K_\phi(\mathbf{x}_n, \mathbf{x}'_m)\end{aligned}$$

To ease notation, we define

$$\nabla_{D_M}K_\phi(\mathbf{x}_n, \mathbf{x}'_m) := \text{Cov}(\phi(\mathbf{x}_n), F(\mathbf{x}'_m)) = \begin{bmatrix} -\frac{\partial}{\partial x_m^1}K_\phi(\mathbf{x}_n, \mathbf{x}'_m) \\ -\frac{\partial}{\partial x_m^2}K_\phi(\mathbf{x}_n, \mathbf{x}'_m) \end{bmatrix}$$

where we used this notation because we refer to the test points (grid) with $D_N := (\mathbf{x}_n)_{n=1}^N$, and to the observation with $D_M := (\mathbf{x}_m)_{m=1}^M$.

So overall we have

$$\begin{bmatrix} \phi \\ \mathbf{F}' \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_\phi & \nabla_{D_M}K_\phi \\ \nabla_{D_M}K_\phi^T & \sigma^2\mathbb{I}_{2M} + \nabla K_\phi + \nabla \times K_\psi \end{bmatrix}\right)$$

This vector has size $N + 2M$, with the caveat that ϕ is evaluated at test points D_N whereas \mathbf{F}' at the observations D_M .

We can then, finally, get the quantity of interest, $\phi \mid \mathbf{F}'$, applying fact 1 above. We obtain

$$\mathbb{E}(\phi \mid \mathbf{F}') = \nabla_{D_M}K_\phi(\sigma^2\mathbb{I}_{2M} + \nabla K_\phi + \nabla \times K_\psi)^{-1}\mathbf{F}'$$

which is of dimension N , exactly as desired, and

$$\text{Var}(\phi \mid \mathbf{F}') = K_\phi - \nabla_{D_M}K_\phi(\sigma^2\mathbb{I}_{2M} + \nabla K_\phi + \nabla \times K_\psi)^{-1}\nabla_{D_M}K_\phi^T$$

with dimension $N \times N$, again as desired.

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

$$\psi \mid \mathbf{F}' \sim \mathcal{N}(\bar{\mu}_\psi, \bar{\Sigma}_\psi)$$

with

$$\begin{aligned}\bar{\mu}_\psi &= \nabla_{D_M}K_\psi(\sigma^2\mathbb{I}_{2M} + \nabla K_\phi + \nabla \times K_\psi)^{-1}\mathbf{F}' \\ \bar{\Sigma}_\psi &= K_\psi - \nabla_{D_M}K_\psi(\sigma^2\mathbb{I}_{2M} + \nabla K_\phi + \nabla \times K_\psi)^{-1}\nabla_{D_M}K_\psi^T\end{aligned}$$

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

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. \square