$$\Sigma(x) = \begin{bmatrix} \sigma_x^2(x) & \rho\,\sigma_x(x)\,\sigma_y(x) \\ \rho\,\sigma_x(x)\,\sigma_y(x) & \sigma_y^2(x) \end{bmatrix},$$

where

$$\sigma_x(x) = L\Big(1 + \alpha\,\frac{x_1 - \bar{x}}{\Delta x}\Big), \quad \sigma_y(x) = L\Big(1 + \alpha\,\frac{x_2 - \bar{y}}{\Delta y}\Big).$$

Here:

- $L$ is the base length scale ("base_lengthscale_space"),
- $\alpha$ is a modulation factor,
- $\bar{x}, \bar{y}$ are the mean coordinates,
- $\Delta x, \Delta y$ are the ranges of the training data.

For two points $x$ and $x'$, let

$$\Sigma_i = \Sigma(x), \quad \Sigma_j = \Sigma(x').$$

Define the generalized Mahalanobis distance as

$$Q_{ij} = (x - x')^\top \Big(\frac{\Sigma_i + \Sigma_j}{2}\Big)^{-1} (x - x').$$

Also, define the prefactor

$$P(x, x') = \frac{\det\big(\Sigma(x)\big)^{\frac{1}{4}} \det\big(\Sigma(x')\big)^{\frac{1}{4}}}{\det\Big(\frac{\Sigma(x)+\Sigma(x')}{2}\Big)^{\frac{1}{2}}}.$$

Then the nonstationary Matérn kernel is given by

$$k(x, x') = \sigma_f^2\, P(x, x')\, \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\sqrt{2\nu\, Q_{ij}}\right)^{\nu} K_\nu\Big(\sqrt{2\nu\, Q_{ij}}\Big).$$

## (a) Exponential Kernel ($\nu = \frac{1}{2}$)

For $\nu = \frac{1}{2}$, we use the identity

$$K_{\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2z}}\, e^{-z}.$$

Also, note $\Gamma\Big(\frac{1}{2}\Big) = \sqrt{\pi}$ and $2^{-\frac{1}{2}} = \frac{1}{\sqrt{2}}$. Substituting $\nu = \frac{1}{2}$ into the kernel expression gives:

$$k_{0.5}(x, x') = \sigma_f^2\, P(x, x')\, \frac{1}{\Gamma\Big(\frac{1}{2}\Big) 2^{-\frac{1}{2}}} \left(\sqrt{2 \cdot \frac{1}{2}\, Q_{ij}}\right)^{\frac{1}{2}} K_{\frac{1}{2}}\Big(\sqrt{2 \cdot \frac{1}{2}\, Q_{ij}}\Big).$$

Since $\sqrt{2 \cdot \frac{1}{2}\, Q_{ij}} = \sqrt{Q_{ij}}$, we have

$$k_{0.5}(x, x') = \sigma_f^2\, P(x, x')\, \frac{1}{\sqrt{\pi}/\sqrt{2}} \left(\sqrt{Q_{ij}}\right)^{\frac{1}{2}} \sqrt{\frac{\pi}{2\sqrt{Q_{ij}}}}\, e^{-\sqrt{Q_{ij}}}.$$

In the stationary case (i.e. when $P(x, x') = 1$ and $Q_{ij} = \frac{r^2}{\ell^2}$, with $r = \|x - x'\|$), after simplification we obtain:

$$k_{0.5}(r) \propto \sigma_f^2\, e^{-r/\ell}.$$

This is the familiar Exponential kernel.

### (b) Squared Exponential Kernel ($\nu \to \infty$)

As $\nu \to \infty$, it can be shown that

$$\lim_{\nu\to\infty} \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu}\frac{r}{\ell}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{r}{\ell}\right) = \exp\left(-\frac{r^2}{2\ell^2}\right).$$

Thus, in the stationary case, the kernel becomes

$$k_{\mathrm{SE}}(r) = \sigma_f^2 \exp\left(-\frac{r^2}{2\ell^2}\right),$$

which is the Squared Exponential (RBF) kernel and is infinitely differentiable.

### (c) Intermediate Smoothness (e.g., $\nu = 2.5$)

For an intermediate value such as $\nu = 2.5$, the kernel expression involves $K_{2.5}$ and, while more complex algebraically, the essential property is that the GP sample paths will be twice mean-square differentiable.

### Direct Differentiation

**Exponential Kernel ($\nu = \frac{1}{2}$):** In the stationary case, write

$$k_{0.5}(r) \propto e^{-r/\ell}.$$

Differentiate with respect to $r$:

$$k'_{0.5}(r) = -\frac{1}{\ell}\, e^{-r/\ell}.$$

However, because $r = |x - x'|$, we must consider the directional derivatives:

$$\lim_{r\to 0^+} k'_{0.5}(r) = -\frac{1}{\ell}, \quad \lim_{r\to 0^-} k'_{0.5}(r) = +\frac{1}{\ell}.$$

These unequal limits show a discontinuity at $r = 0$, so the kernel is not differentiable there. Hence, its RKHS contains only rough functions and the corresponding GP sample paths are non-differentiable.

**Squared Exponential Kernel:**

$$k_{\mathrm{SE}}(r) = \sigma_f^2 \exp\left(-\frac{r^2}{2\ell^2}\right).$$

Differentiate with respect to $r$:

$$k'_{\mathrm{SE}}(r) = -\frac{\sigma_f^2\, r}{\ell^2} \exp\left(-\frac{r^2}{2\ell^2}\right).$$

At $r = 0$, clearly

$$k'_{\mathrm{SE}}(0) = 0.$$

Higher-order derivatives exist and are continuous, so the SE kernel is infinitely differentiable and its RKHS contains smooth functions; thus, GP sample paths are smooth.

### Mean-Square Differentiability and the RKHS Perspective

A Gaussian Process $f(x)$ is *mean-square differentiable* if

$$\lim_{h \to 0} \mathbb{E}\left[\left(\frac{f(x+h)-f(x)}{h} - f'(x)\right)^2\right] = 0.$$

This condition is satisfied if the kernel $k(x, x')$ has continuous mixed partial derivatives, e.g.,

$$\frac{\partial^2}{\partial x\, \partial x'} k(x, x').$$

The RKHS $\mathcal{H}_k$ associated with $k$ is the Hilbert space of functions for which the inner product satisfies the reproducing property:

$$f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}_k}.$$

If $k$ is $m$-times continuously differentiable, then every function in $\mathcal{H}_k$ is $m$-times differentiable in the mean-square sense. Therefore:

- For $\nu = \frac{1}{2}$, the lack of continuous derivative implies the RKHS contains only non-smooth functions, so GP sample paths are rough everywhere.
- For $\nu = 2.5$, the RKHS functions are twice differentiable, hence the GP sample paths are smoother.
- For $\nu \to \infty$, the RKHS comprises infinitely smooth functions, and so are the GP sample paths.

Given training data $\{x_i, y_i\}_{i=1}^N$, construct the kernel matrix

$$K_{ij} = k(x_i, x_j).$$

After adding noise variance, we set

$$K \leftarrow K + \sigma_n^2 I.$$

Compute the Cholesky decomposition:

$$K = L\, L^\top,$$

where $L$ is a lower triangular matrix. To solve

$$K\, \alpha = y,$$

we first solve:

$$L\, v = y,$$

and then:

$$L^\top\, \alpha = v.$$

This provides a numerically stable solution for

$$\alpha = K^{-1} y.$$

**Summary:**

(1) **Kernel Construction:** The nonstationary Matérn kernel is defined by

$$k(x,x') = \sigma_f^2\, P(x,x')\, \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu\, Q_{ij}}\right)^{\nu} K_\nu\!\left(\sqrt{2\nu\, Q_{ij}}\right),$$

with

$$Q_{ij} = (x-x')^\top \left(\frac{\Sigma(x)+\Sigma(x')}{2}\right)^{-1}(x-x')$$

and

$$P(x,x') = \frac{\det\!\big(\Sigma(x)\big)^{\frac{1}{4}} \det\!\big(\Sigma(x')\big)^{\frac{1}{4}}}{\det\!\left(\frac{\Sigma(x)+\Sigma(x')}{2}\right)^{\frac{1}{2}}}.$$

(2) **Substitution of $\nu$:**

- For $\nu = \frac{1}{2}$, the kernel reduces to

  $$k_{0.5}(r) \propto \sigma_f^2\, e^{-r/\ell},$$

  the Exponential kernel, which is not differentiable (its derivative has a discontinuity at $r = 0$).
- For $\nu \to \infty$, the kernel converges to

  $$k_{\mathrm{SE}}(r) = \sigma_f^2\, e^{-r^2/(2\ell^2)},$$

  the Squared Exponential kernel, which is infinitely differentiable.
- For intermediate values (e.g. $\nu = 2.5$), the kernel yields GP sample paths that are twice mean-square differentiable.

(3) **Differentiability Analysis:**

- *Exponential Kernel:* With

  $$k_{0.5}(r) \propto e^{-r/\ell}, \quad k'_{0.5}(r) = -\frac{1}{\ell}\, e^{-r/\ell},$$

  we have

  $$\lim_{r\to 0^+} k'_{0.5}(r) = -\frac{1}{\ell}, \quad \lim_{r\to 0^-} k'_{0.5}(r) = +\frac{1}{\ell}.$$

  Thus, the derivative is discontinuous at $r = 0$, and the kernel (and hence the GP) is non-differentiable.
- *Squared Exponential Kernel:* With

  $$k_{\mathrm{SE}}(r) = \sigma_f^2\, \exp\!\left(-\frac{r^2}{2\ell^2}\right),$$

  its derivative

  $$k'_{\mathrm{SE}}(r) = -\frac{\sigma_f^2\, r}{\ell^2}\, \exp\!\left(-\frac{r^2}{2\ell^2}\right)$$

  is continuous for all $r$, ensuring infinite differentiability.

(4) **RKHS and Mean-Square Differentiability:** The RKHS $\mathcal{H}_k$ associated with a kernel $k$ consists of functions $f$ such that

$$f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}_k}.$$

If $k(x, x')$ is $m$-times continuously differentiable, then every function in $\mathcal{H}_k$ is $m$-times mean-square differentiable. Therefore:

- For $\nu = \frac{1}{2}$, the discontinuity in the first derivative implies the RKHS contains only rough functions (GP sample paths are non-differentiable).
- For higher $\nu$, the RKHS consists of smoother functions, and the GP sample paths become smoother accordingly.

(5) **GP Inference via Cholesky Decomposition:** For training data $\{x_i, y_i\}_{i=1}^N$, construct the kernel matrix $K$ with $K_{ij} = k(x_i, x_j)$, then add noise:

$$K \leftarrow K + \sigma_n^2 I.$$

Decompose:

$$K = L\,L^\top,$$

and solve

$$L\,v = y, \quad L^\top\,\alpha = v,$$

yielding $\alpha = K^{-1}y$ in a numerically stable manner.