Evolution strategies for robust optimization

Kruisselbrink, J.W.

Appendix A

Test Problems for Noisy Optimization

A.1 Noisy Sphere Problem

The noisy sphere is a simple scalable test function for studying optimization of noisy real-valued objective functions using Evolution Strategies. It reads

$$f(\mathbf{x}) = \sum_{i=1}^{n} z_i^2, \tag{A.1}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*, \tag{A.2}$$

with $\mathbf{x}^* \in \mathbb{R}^n$ being the location of the optimum. The noisy sphere function, reads:

$$f(\mathbf{x}) = \sum_{i=1}^{n} z_i^2 + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.3}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.4}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 1$. Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-5, \ldots, -5]$ and $\mathbf{x}_\mathrm{u} = [5, \ldots, 5]$.

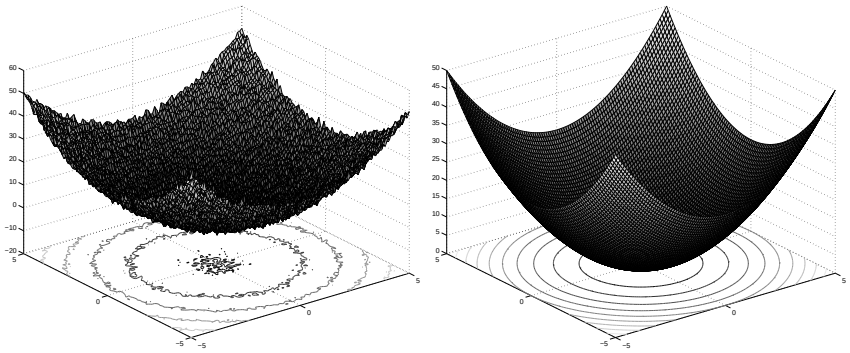

Figure A.1: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.2 Noisy Ellipsoid Problem

The ellipsoid function is a transformed (stretched and rotated) version of the sphere function.

$$f(\mathbf{x}) = \sum_{i=1}^{n} z_i^2, \tag{A.5}$$

$$\mathbf{z} = \mathbf{RD}\left(\mathbf{x} - \mathbf{x}^*\right). \tag{A.6}$$

with $\mathbf{x}^* \in \mathbb{R}^n$ being the location of the optimum, $\mathbf{R}$ being a rotation matrix, and $\mathbf{D}$ being a diagonal (scaling) matrix. The noisy ellipsoid function reads:

$$f(\mathbf{x}) = \sum_{i=1}^{n} z_i^2 + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.7}$$

$$\mathbf{z} = \mathbf{RD}\left(\mathbf{x} - \mathbf{x}^*\right). \tag{A.8}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 2$. The rotation matrix $\mathbf{R}$ is generated from normally distributed entries and scaling matrix $\mathbf{D}$ has a condition number of 10 with equally spaced eigenvalues. Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-1, \ldots, -1]$ and $\mathbf{x}_\mathrm{u} = [1, \ldots, 1]$.

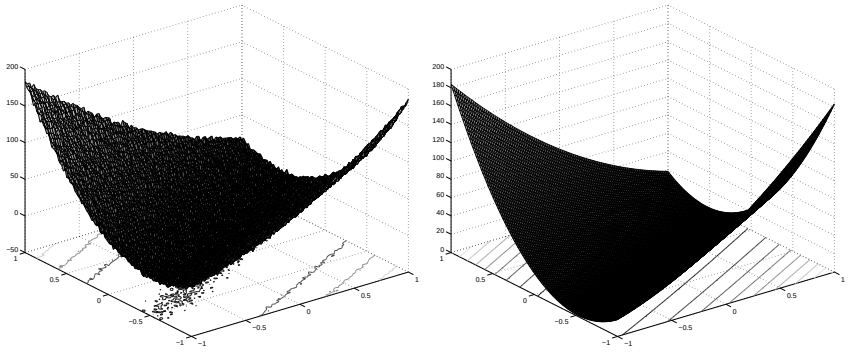

Figure A.2: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.3 Noisy Step Ellipsoid Problem

The step ellipsoid function is an ellipsoid function transformed such that it has a non-steady surface consisting of plateaus. It reads

$$f(\mathbf{x}) = \sum_{i=1}^{n} \lfloor z_i \rfloor^2, \tag{A.9}$$

$$\mathbf{z} = \mathbf{RD}\left(\mathbf{x} - \mathbf{x}^*\right). \tag{A.10}$$

Here $\mathbf{x}^* \in \mathbb{R}^n$ is the location of the optimum, $\mathbf{R}$ is a rotation matrix, and $\mathbf{D}$ is a diagonal (scaling) matrix. The noisy step ellipsoid function reads

$$f(\mathbf{x}) = \sum_{i=1}^{n} \lfloor z_i \rfloor^2 + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.11}$$

$$\mathbf{z} = \mathbf{RD}\left(\mathbf{x} - \mathbf{x}^*\right). \tag{A.12}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 2$. The rotation matrix $\mathbf{R}$ is generated from normally distributed entries and scaling matrix $\mathbf{D}$ has a condition number of 10 with equally spaced eigenvalues (these matrices are the same as for the noisy ellipsoid function). Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-1, \ldots, -1]$ and $\mathbf{x}_\mathrm{u} = [1, \ldots, 1]$.

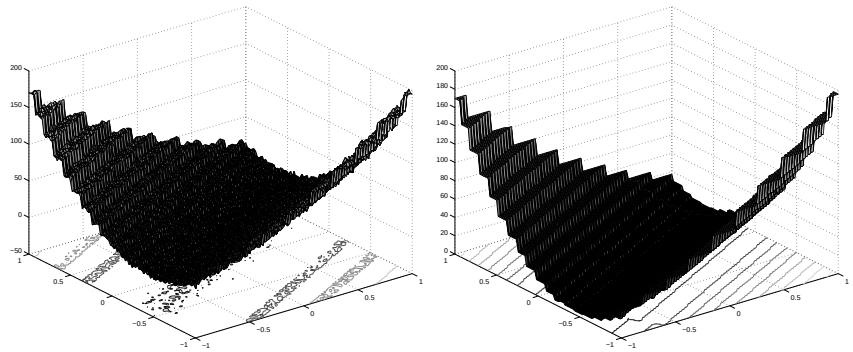

Figure A.3: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.4 Noisy Rosenbrock Problem

The Rosenbrock function is a unimodal test function for real-valued optimization. The particular characteristic of this function is that its fitness landscape shows a bent valley that leads towards the optimum and the direction of the steepest descent changes continuously when nearing the optimum. It is defined as

$$f(\mathbf{x}) = \sum_{i=1}^{n-1} \left(100\left(z_i^2 - z_{i+1}\right)^2 + \left(z_i - 1\right)^2\right), \tag{A.13}$$

$$\mathbf{z} = (\mathbf{x} - \mathbf{x}^*) + 1. \tag{A.14}$$

Here $\mathbf{x}^* \in \mathbb{R}^n$ is the location of the optimum. The noisy Rosenbrock function reads

$$f(\mathbf{x}) = \sum_{i=1}^{n-1} \left(100\left(z_i^2 - z_{i+1}\right)^2 + \left(z_i - 1\right)^2\right) + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.15}$$

$$\mathbf{z} = (\mathbf{x} - \mathbf{x}^*) + 1. \tag{A.16}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 2$. Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-2, \ldots, -2]$ and $\mathbf{x}_\mathrm{u} = [2, \ldots, 2]$.

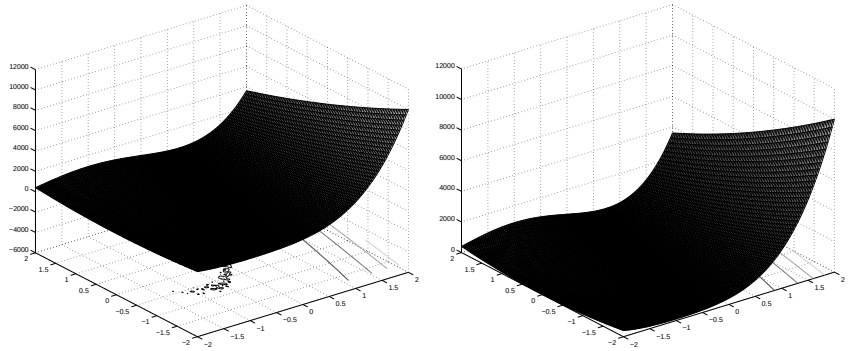

Figure A.4: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.5 Noisy Ackley Problem

The Ackley function is a multi-modal test function, defined as

$$\begin{aligned} f(\mathbf{x}) &= -c_1 \cdot \exp\left(-c_2\sqrt{\frac{1}{n}\sum_{i=1}^{n} z_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(c_3 \cdot z_i)\right) \\ &\quad + c_1 + \exp(1), \end{aligned} \tag{A.17}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.18}$$

Here, $c_1 = 20$, $c_2 = 0.2$, $c_3 = 2\pi$, and $\mathbf{x}^* \in \mathbb{R}^n$ is the location of the optimum. The noisy Ackley function reads

$$\begin{aligned} f(\mathbf{x}) &= -c_1 \cdot \exp\left(-c_2\sqrt{\frac{1}{n}\sum_{i=1}^{n} z_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(c_3 \cdot z_i)\right) \\ &\quad + c_1 + \exp(1) + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \end{aligned} \tag{A.19}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.20}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 1$. Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-5, \ldots, -5]$ and $\mathbf{x}_\mathrm{u} = [5, \ldots, 5]$.

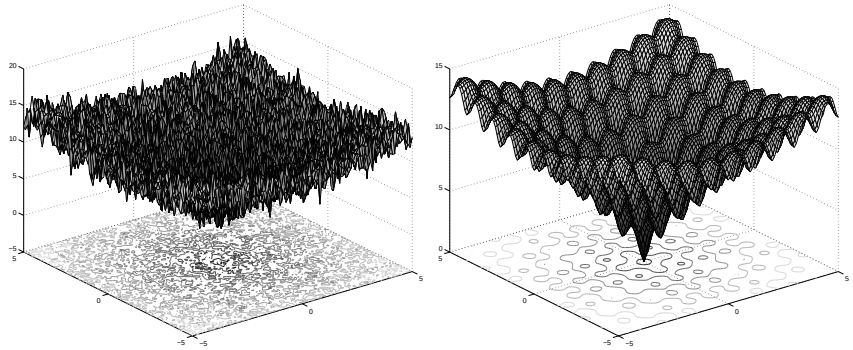

Figure A.5: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.6 Noisy Griewank Problem

The Griewank function is a multi-modal test function, defined as

$$f(\mathbf{x}) = 1 + \frac{1}{4000}\sum_{i=1}^{n} z_i^2 - \prod_{i=1}^{n} \cos\left(\frac{z_i}{\sqrt{i}}\right), \tag{A.21}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.22}$$

Here $\mathbf{x}^* \in \mathbb{R}^n$ is the location of the optimum. The noisy Griewank function reads

$$f(\mathbf{x}) = 1 + \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.23}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.24}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 0.5$. Unless stated otherwise, we will assume stationary noise. Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-60, \ldots, -60]$ and $\mathbf{x}_\mathrm{u} = [60, \ldots, 60]$.

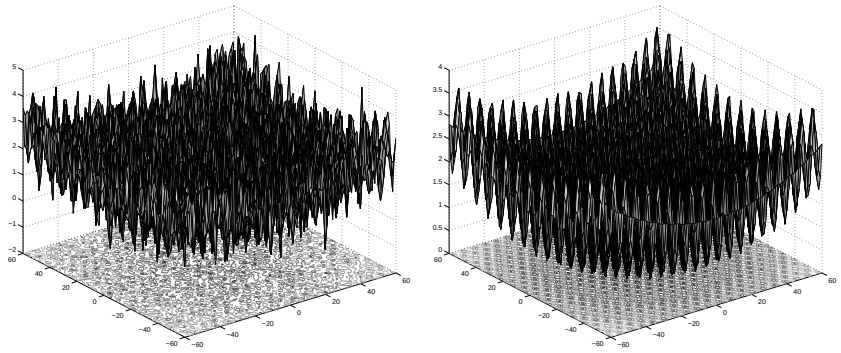

Figure A.6: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.7 Noisy Rastrigin Problem

The Rastrigin function is a multi-modal test function, defined as

$$f(\mathbf{x}) = 10n + \sum_{i=1}^{n}(z_i^2 - 10\cos(2\pi \cdot z_i)), \tag{A.25}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.26}$$

Here $\mathbf{x}^* \in \mathbb{R}^n$ is the location of the optimum. The noisy Griewank function reads

$$f(\mathbf{x}) = 10n + \sum_{i=1}^{n}(z_i^2 - 10\cos(2\pi \cdot z_i)) + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.27}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.28}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 2$. Unless stated otherwise, we will assume stationary noise. Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-5, \ldots, -5]$ and $\mathbf{x}_\mathrm{u} = [5, \ldots, 5]$.

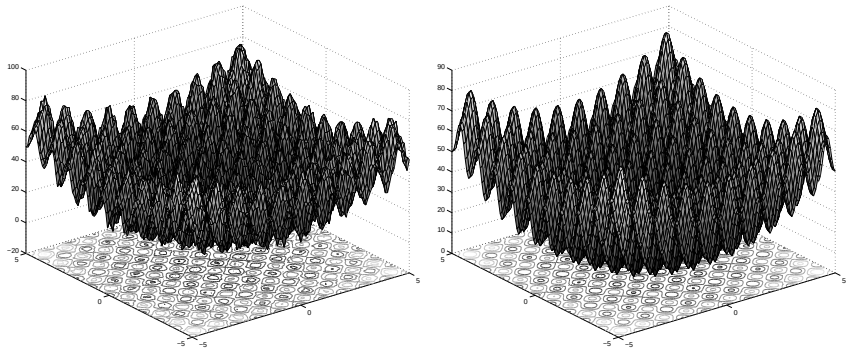

Figure A.7: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.8 Noisy Schaffer's F7 Problem

Schaffer's F7 function is a multi-modal test function, defined as

$$f(\mathbf{x}) = \sum_{i=1}^{n-1} (z_i^2 + z_{i+1}^2)^{0.25} (\sin^2(50(z_i^2 + z_{i+1}^2)^{0.1}) + 1), \tag{A.29}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.30}$$

Here $\mathbf{x}^* \in \mathbb{R}^n$ is the location of the optimum. The noisy Griewank function reads

$$f(\mathbf{x}) = \sum_{i=1}^{n-1} (z_i^2 + z_{i+1}^2)^{0.25} (\sin^2(50(z_i^2 + z_{i+1}^2)^{0.1}) + 1) + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.31}$$

$$\mathbf{z} = \mathbf{x} - \mathbf{x}^*. \tag{A.32}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 1$. Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-5, \ldots, -5]$ and $\mathbf{x}_\mathrm{u} = [5, \ldots, 5]$.

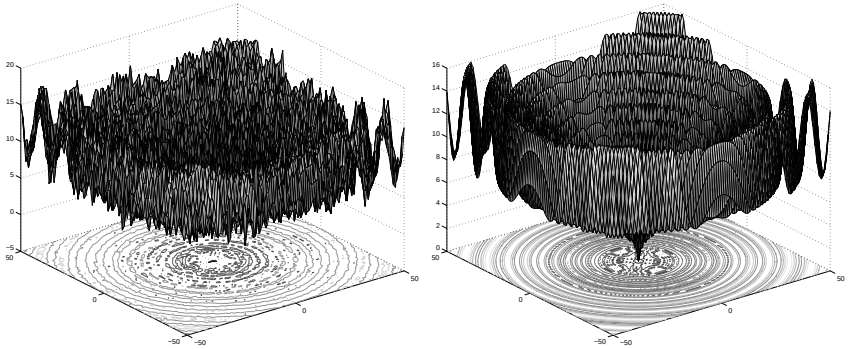

Figure A.8: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.9 Noisy Branke's Multipeak Problem

Branke's Multipeak function is a multi-modal test function with 2^n peaks, defined as

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i-1}^{n}\left(c - g(x_i)\right), \tag{A.33}$$

$$g(x_i) = \begin{cases} -\left(x_i+1\right)^2 + 1 & \text{if } -2 \leq x_i < 0 \\ c \cdot 2^{-8|x_i-1|} & \text{if } 0 \leq x_i \leq 2 \\ 0 & \text{otherwise} \end{cases}. \tag{A.34}$$

Here $c = 1.3$ and the global optimum is located at $\mathbf{x} = [1, \ldots, 1]$. The noisy version of Branke's multipeak function reads

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i-1}^{n}\left(c - g(x_i)\right) + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.35}$$

$$g(x) = \begin{cases} -\left(x+1\right)^2 + 1 & \text{if } -2 \leq x < 0 \\ c \cdot 2^{-8|x-1|} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}. \tag{A.36}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 0.1$. Unless stated otherwise, we will assume stationary noise. Furthermore, in experimental settings, the search interval is set to $\mathbf{x}_\mathrm{l} = [-2, \ldots, -2]$ and $\mathbf{x}_\mathrm{u} = [2, \ldots, 2]$.

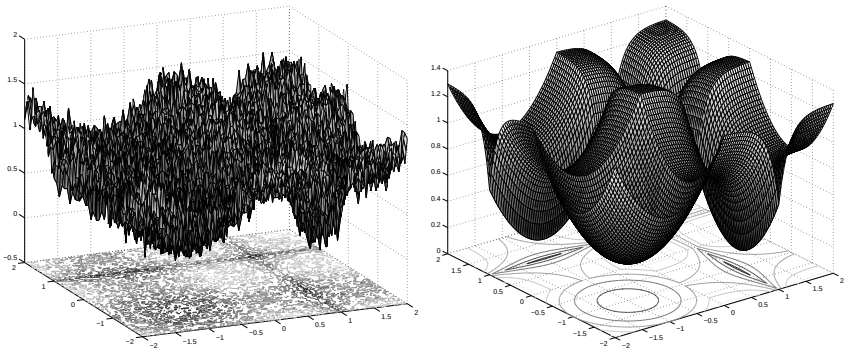

Figure A.9: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

A.10 Noisy Keane's Bump Problem

Keane's Bump function is a highly multi-modal test function, defined as

$$\min \qquad f(\mathbf{x}) = -\frac{|\sum_{i=1}^{n}\cos^4(x_i) - 2\prod_{i=1}^{n}(\cos^2(x_i))|}{\sqrt{\sum_{i=1}^{n} i \cdot x_i^2}}, \tag{A.37}$$

$$\text{s.t.} \qquad \prod_{i=1}^{n} x_i > 0.75 \ , \ \sum_{i=1}^{n} x_i < \frac{15n}{2} \ , \ x_i \in]0, 10[. \tag{A.38}$$

The global minimizer for this function is unknown. The noisy version adopted in this work uses a penalty mechanism to aggregate the constraints in one objective function. It reads

$$f(\mathbf{x}) = g(\mathbf{x}) + \mathcal{N}(0, \sigma_\epsilon^2(\mathbf{x})), \tag{A.39}$$

$$g(\mathbf{x}) = \begin{cases} -\frac{|\sum_{i=1}^{n}\cos^4(x_i) - 2\prod_{i=1}^{n}(\cos^2(x_i))|}{\sqrt{\sum_{i=1}^{n} i \cdot x_i^2}} & \text{if } \prod_{i=1}^{n} x_i > 0.75 \text{ and } \sum_{i=1}^{n} x_i < \frac{15n}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{A.40}$$

In this work, stationary noise is assumed with $\sigma_\epsilon^2(\mathbf{x}) = 0.05$. Unless stated otherwise, we will assume stationary noise. Furthermore, in experimental settings, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\text{l} = [0, \ldots, 0]$ and $\mathbf{x}_\text{u} = [10, \ldots, 10]$.

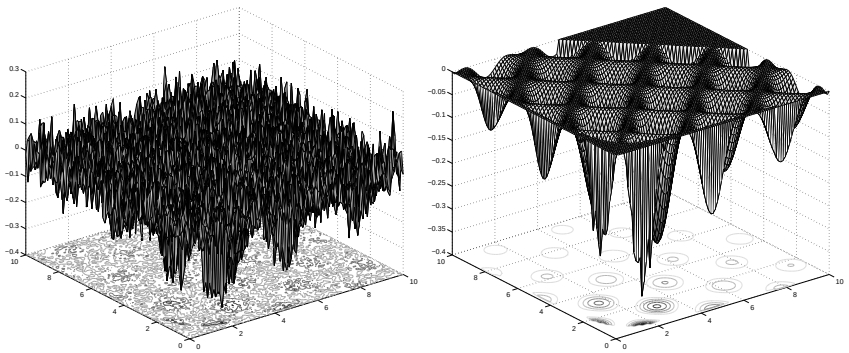

Figure A.10: Left: 2D visualization of the noisy objective function landscape. Right: 2D visualization of the noise-free underlying objective function landscape.

Appendix B

Test Problems for Finding Robust Optima

B.1 RO Sphere Problem

A simple unimodal test function for finding robust optima that can be used to assess the quality of a robust optimization algorithm w.r.t. zooming in on the robust optimum. For this function the robust optimizer and the optimizer of the original function are the same. The sphere function reads

$$f(\mathbf{x}) = \sum_{i=1}^{n} z_i^2 \quad , \quad \mathbf{z} = \mathbf{x} - \mathbf{x}^*, \tag{B.1}$$

with $\mathbf{x}^* \in \mathbb{R}^n$ being the location of the optimum. In this work, the optimum location is set to $\mathbf{x}^* = [0, \ldots, 0]$ and the search interval is set to $\mathbf{x}_\mathrm{l} = [-5, \ldots, -5]$ and $\mathbf{x}_\mathrm{u} = [5, \ldots, 5]$. The uncertainty in the design variables is of the form

$$\mathbf{x} = \mathbf{x} + \boldsymbol{\delta} \quad , \quad \boldsymbol{\delta} \sim \mathcal{U}(-\mathbf{1}, \mathbf{1}). \tag{B.2}$$

This function has as robust optimizer for the expected objective function

$$\mathbf{x}^*_\mathrm{exp} = \mathbf{x}^*. \tag{B.3}$$

Figure B.1 shows 2D visualizations of the original objective function landscape and the effective objective function landscape.

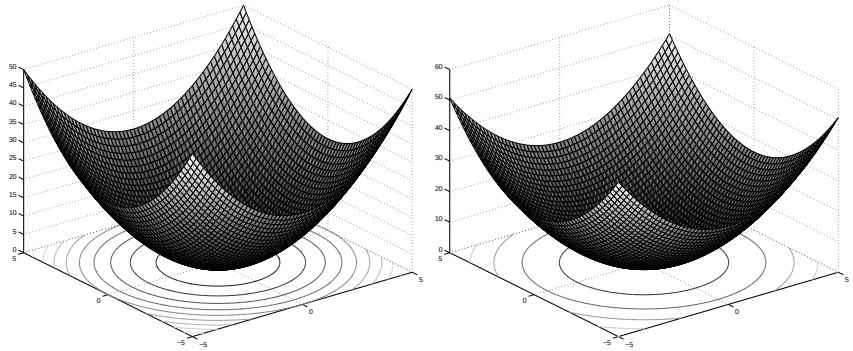

Figure B.1: Left: 2D visualization of the original objective function landscape. Right: 2D visualization of the effective objective function landscape.

B.2 RO Heaviside Sphere Problem

The Heaviside sphere function reads

$$f(\mathbf{x}) = \left(1 - \prod_{i=1}^{2} g(x_i)\right) + \sum_{i=1}^{n} \left(\frac{x_i}{10}\right)^2 \quad , \quad g(x_i) = \begin{cases} 0 & \text{if } x_i < 0 \\ 1 & \text{otherwise} \end{cases}, \tag{B.4}$$

with search interval $\mathbf{x}_\mathrm{l} = [-10, \ldots, -10]$ and $\mathbf{x}_\mathrm{u} = [10, \ldots, 10]$. The uncertainty in the design variables is of the form

$$\mathbf{x} = \mathbf{x} + \boldsymbol{\delta} \quad , \quad \boldsymbol{\delta} \sim U(-\mathbf{1}, \mathbf{1}). \tag{B.5}$$

This function has as robust optimizer for the expected objective function

$$\mathbf{x}^*_{\mathrm{exp}} = [1, \ldots, 1]^n. \tag{B.6}$$

Figure B.2 shows 2D visualizations of the original objective function landscape and the effective objective function landscape.

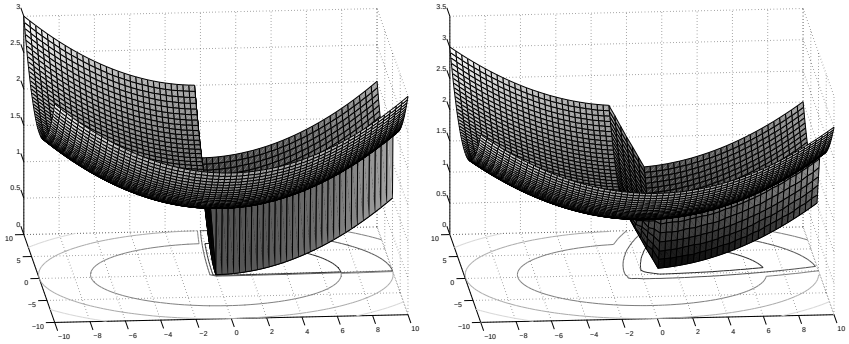

Figure B.2: Left: 2D visualization of the original objective function landscape. Right: 2D visualization of the effective objective function landscape.

B.3 RO Sawtooth Problem

The sawtooth function, originally proposed in [Bra01], reads

$$f(\mathbf{x}) = 1 - \frac{1}{n}\sum_{i=1}^{n} g(x_i) \quad , \quad g(x_i) = \begin{cases} x_i + 0.8 & \text{if } -0.8 \leq x_i < 0.2 \\ 0 & \text{otherwise} \end{cases}, \tag{B.7}$$

$$\tag{B.8}$$

with search interval $\mathbf{x}_\mathrm{l} = [-1, \ldots, -1]$ and $\mathbf{x}_\mathrm{u} = [1, \ldots, 1]$. The uncertainty in the design variables is of the form

$$\mathbf{x} = \mathbf{x} + \boldsymbol{\delta} \quad , \quad \boldsymbol{\delta} \sim U(-\mathbf{0.2}, \mathbf{0.2}). \tag{B.9}$$

This function has as robust optimizer for the expected objective function

$$\mathbf{x}^*_\mathrm{exp} = [0, \ldots, 0]^n. \tag{B.10}$$

Figure B.3 shows 2D visualizations of the original objective function landscape and the effective objective function landscape.

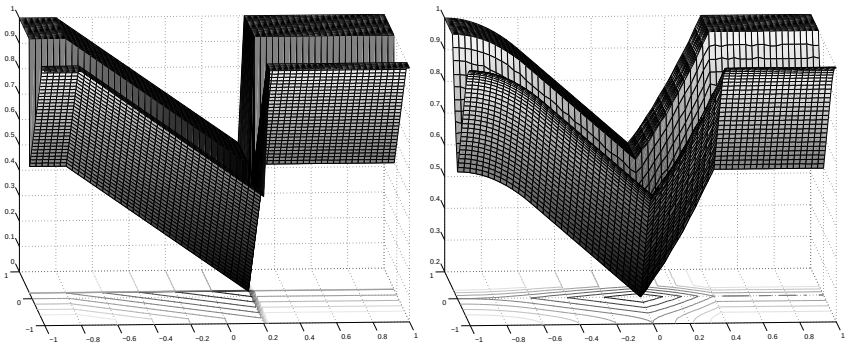

Figure B.3: Left: 2D visualization of the original objective function landscape. Right: 2D visualization of the effective objective function landscape.

B.4 RO Volcano Problem

The volcano function reads

$$f(\mathbf{x}) = \begin{cases} \sqrt{||\mathbf{x}||} - 1 & \text{if } ||\mathbf{x}|| > 1 \\ 0 & \text{otherwise} \end{cases}, \tag{B.11}$$

with search interval $\mathbf{x}_{\mathrm{l}} = [-10, \ldots, -10]$ and $\mathbf{x}_{\mathrm{u}} = [10, \ldots, 10]$. The uncertainty in the design variables is of the form

$$\mathbf{x} = \mathbf{x} + \boldsymbol{\delta} \quad , \quad \boldsymbol{\delta} \sim U(-\mathbf{1.5}, \mathbf{1.5}). \tag{B.12}$$

This function has as robust optimizer for the expected objective function

$$\mathbf{x}^*_{\mathrm{exp}} = [0, \ldots, 0]^n. \tag{B.13}$$

Figure B.4 shows 2D visualizations of the original objective function landscape and the effective objective function landscape.

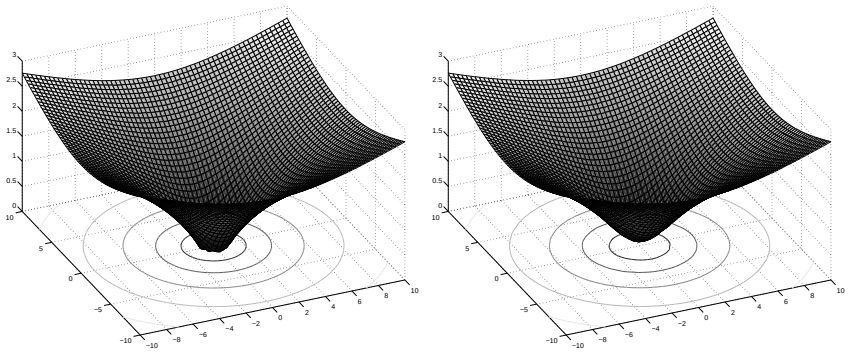

Figure B.4: Left: 2D visualization of the original objective function landscape. Right: 2D visualization of the effective objective function landscape.

B.5 RO Pickelhaube Problem

The objective function reads

$$f(\mathbf{x}) = \frac{5}{5-\sqrt{5}} - \max\left\{g_0(\mathbf{x}), g_{1a}(\mathbf{x}), g_{1b}(\mathbf{x}), g_2(\mathbf{x})\right\}, \tag{B.14}$$

$$g_0(\mathbf{x}) = \frac{1}{10} \cdot e^{-\frac{1}{2}\cdot||\mathbf{x}||}, \tag{B.15}$$

$$g_{1a}(\mathbf{x}) = \frac{5}{5-\sqrt{5}} \cdot \left(1 - \sqrt{\frac{||\mathbf{x}+\mathbf{5}||}{5\cdot\sqrt{n}}}\right), \tag{B.16}$$

$$g_{1b}(\mathbf{x}) = c_1 \cdot \left(1 - \left(\frac{||\mathbf{x}+\mathbf{5}||}{5\cdot\sqrt{n}}\right)^4\right), \tag{B.17}$$

$$g_2(\mathbf{x}) = c_2 \cdot \left(1 - \left(\frac{||\mathbf{x}+\mathbf{5}||}{5\cdot\sqrt{n}}\right)^{d_2}\right), \tag{B.18}$$

with $c_1 = 625/624$, $c_2 = 1.5975$, $d_2 = 1.1513$, and search interval $\mathbf{x}_\mathrm{l} = [-10, \ldots, -10]$ and $\mathbf{x}_\mathrm{u} = [10, \ldots, 10]$. The uncertainty in the design variables is of the form

$$\mathbf{x} = \mathbf{x} + \boldsymbol{\delta} \quad , \quad \boldsymbol{\delta} \sim \mathcal{U}(-\mathbf{1}, \mathbf{1}). \tag{B.19}$$

This function has as robust optimizer for the expected objective function

$$\mathbf{x}^*_{\mathrm{exp}} = [5, \ldots, 5]^n. \tag{B.20}$$

Figure B.5 shows 2D visualizations of the original objective function landscape and the effective objective function landscape.

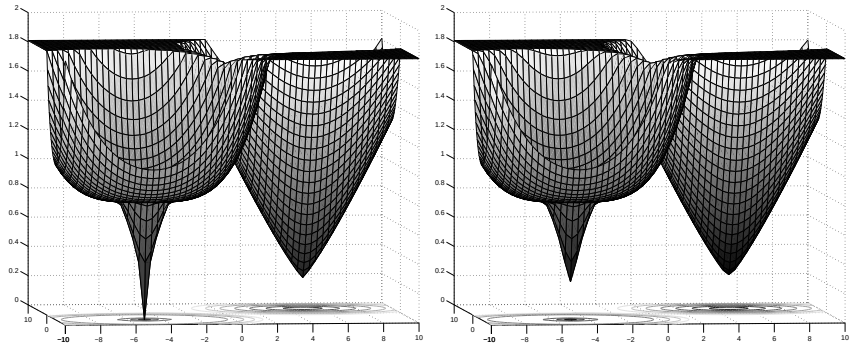

Figure B.5: Left: 2D visualization of the original objective function landscape. Right: 2D visualization of the effective objective function landscape.

B.6 RO Branke's Multipeak Problem

Objective function from [Bra98]. Originally posed as a maximization problem, converted to a minimization problem. The objective function reads

$$f(\mathbf{x}) = \max\{c_1, c_2\} - \frac{1}{n}\sum_{i-1}^{n} g(x_i), \tag{B.21}$$

$$g(x_i) = \begin{cases} c_1 \cdot \left(1 - \frac{4\left(x_i + \frac{b_1}{2}\right)^2}{(b_1)^2}\right) & \text{if } -b_1 \leq x_i < 0 \\ c_2 \cdot 16^{\frac{-2|b_2 - 2x_i|}{b_2}} & \text{if } 0 \leq x_i \leq b_2 \\ 0 & \text{otherwise} \end{cases}, \tag{B.22}$$

with $b_1 > 0$, $b_2 > 0$, $c_1 > 0$, $c_2 > 0$ and search interval $\mathbf{x}_\text{l} = [-b_1, \ldots, -b_1]$ and $\mathbf{x}_\text{u} = [b_2, \ldots, b_2]$. In this work, we use the settings $b_1 = 2$, $b_2 = 2$, $c_1 = 1$, $c_2 = 1.3$. The uncertainty in the design variables is of the form

$$\mathbf{x} = \mathbf{x} + \boldsymbol{\delta} \quad , \quad \boldsymbol{\delta} \sim \mathcal{U}(-\mathbf{0.5}, \mathbf{0.5}). \tag{B.23}$$

This function has as robust optimizer for the expected objective function

$$\mathbf{x}^*_\text{exp} = [-b_1/2, \ldots, -b_1/2]^n. \tag{B.24}$$

Figure B.6 shows 2D visualizations of the original objective function landscape and the effective objective function landscape.

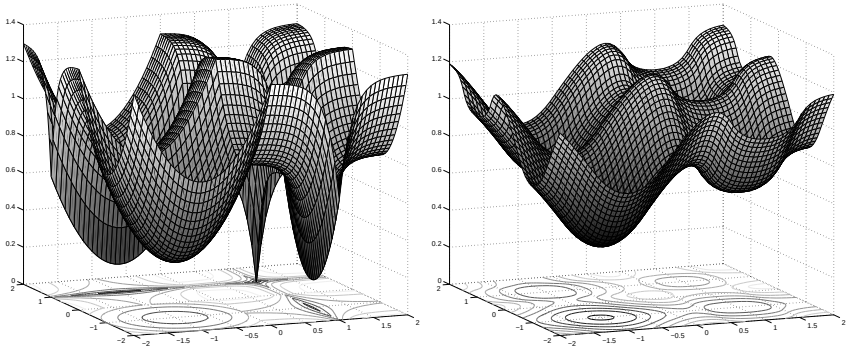

Figure B.6: Left: 2D visualization of the original objective function landscape. Right: 2D visualization of the effective objective function landscape.

B.7 RO Multipeak F1 Problem

A multimodal test problem used in, e.g., [TGF96, TG97]. Originally posed as maximization problem, converted to minimization. The objective function reads

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{x}), \tag{B.25}$$

$$g(x_i) = \begin{cases} e^{-2\ln 2\left(\frac{x-0.1}{0.8}\right)^2}\sqrt{|\sin(5\pi x_i)|}, & \text{if } 0.4 < x_i \leq 0.6 \\ e^{-2\ln 2\left(\frac{x-0.1}{0.8}\right)^2}\sin^6(5\pi x_i), & \text{otherwise} \end{cases}, \tag{B.26}$$

with search interval $\mathbf{x}_\mathrm{l} = [0,\ldots,0]$ and $\mathbf{x}_\mathrm{u} = [1,\ldots,1]$. The uncertainty in the design variables is of the form

$$\mathbf{x} = \mathbf{x} + \boldsymbol{\delta} \quad , \quad \boldsymbol{\delta} \sim \mathcal{U}(-\mathbf{0.0625}, \mathbf{0.0625}). \tag{B.27}$$

This function has as robust optimizer for the expected objective function

$$\mathbf{x}^*_{\mathrm{exp}} \approx [0.4911,\ldots,0.4911]^n. \tag{B.28}$$

Figure B.7 shows 2D visualizations of the original objective function landscape and the effective objective function landscape.

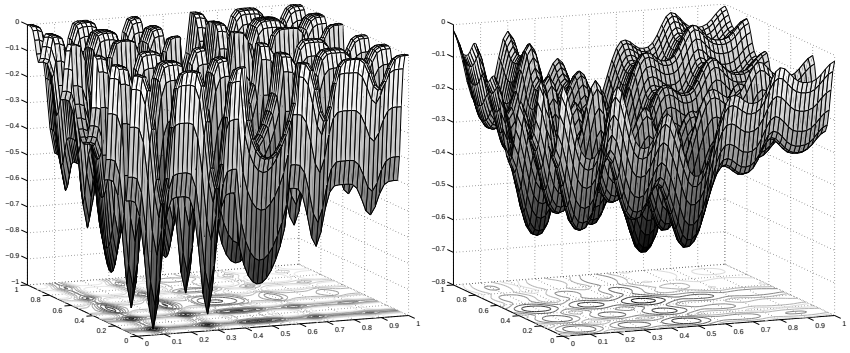

Figure B.7: Left: 2D visualization of the original objective function landscape. Right: 2D visualization of the effective objective function landscape.

B.8 RO Multipeak F2 Problem

A multimodal test problem used in, e.g., [PBJ06]. Originally posed as maximization problem, converted to minimization. The objective function reads

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{x}), \tag{B.29}$$

$$g(x_i) = 2\sin\left(10\exp(-0.2x_i)x_i\right)\exp\left(-0.25x_i\right), \tag{B.30}$$

with search interval $\mathbf{x}_\mathrm{l} = [0,\ldots,0]$ and $\mathbf{x}_\mathrm{u} = [10,\ldots,10]$. The uncertainty in the design variables is of the form

$$\mathbf{x} = \mathbf{x} + \boldsymbol{\delta} \quad , \quad \boldsymbol{\delta} \sim \mathcal{U}(-\mathbf{0.5}, \mathbf{0.5}). \tag{B.31}$$

This function has as robust optimizer for the expected objective function

$$\mathbf{x}^*_\mathrm{exp} \approx [3.5,\ldots,3.5]^n. \tag{B.32}$$

Figure B.8 shows 2D visualizations of the original objective function landscape and the effective objective function landscape.

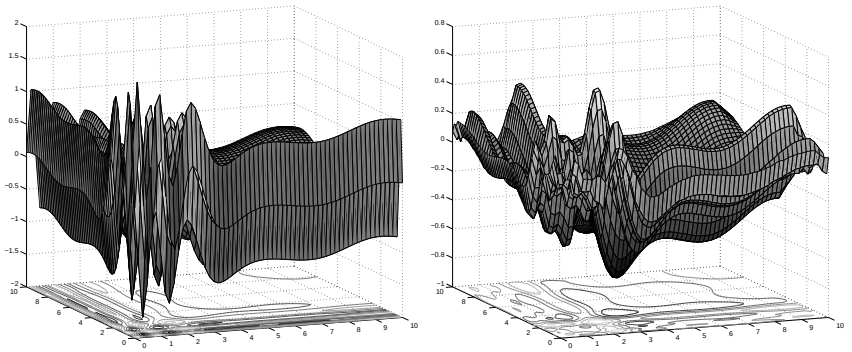

Figure B.8: Left: 2D visualization of the original objective function landscape. Right: 2D visualization of the effective objective function landscape.

Appendix C

Kriging Metamodeling

Kriging Metamodeling

A basic assumption of function modeling through Kriging is that the deviation of a function from a general trend or mean value can be modeled as a realization of a Gaussian random field $\mathcal{F}_{\mathbf{x}}$, $\mathbf{x} \in \mathbb{R}^n$, where $\mathcal{F}_{\mathbf{x}}$ is a Gaussian random variable indexed by space. Two random variables $\mathcal{F}_{\mathbf{x}}$ and $\mathcal{F}_{\mathbf{x}'}$ are correlated via a spatial correlation function $c(\mathbf{x}, \mathbf{x}')$. Often $c(\mathbf{x}, \mathbf{x}')$ is based on the difference $\mathbf{x} - \mathbf{x}'$ (stationary correlation) or on the distance $d(\mathbf{x}, \mathbf{x}')$ (isotropic correlation). When c is fully specified, then it is possible to compute the conditional distribution of $\mathcal{F}_{\mathbf{x}}$ for a new point $\mathbf{x}$ given a number of measured realizations $y_1 = \mathcal{F}_{\mathbf{x}_1}, \ldots, y_m = \mathcal{F}_{\mathbf{x}_m}$. The mean value of the conditional distribution can be interpreted as a predictor for the function value $f(\mathbf{x})$ and the standard deviation of $\mathcal{F}_{\mathbf{x}}$ can be interpreted as a measure of prediction uncertainty. *Ordinary Kriging*, used in this work, assumes a constant trend β. It thus estimates $f(\mathbf{x})$ as $\mathcal{F}_{\mathbf{x}} = \beta + \mathcal{R}_{\mathbf{x}}$, where $\mathcal{R}_{\mathbf{x}}$, $\mathbf{x} \in \mathbb{R}^n$ is a Gaussian random field model with mean zero and global variance s^2. The Kriging predictor $\hat{f}(\mathbf{x})$ for an unknown point $\mathbf{x}$ is

$$\hat{f}(\mathbf{x}) = \beta + (\mathbf{y} - \mathbf{1} \cdot \beta)^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(\mathbf{x}), \tag{C.1}$$

with $\mathbf{y} = [y_1, \ldots, y_m]^T$, $\mathbf{C} = [c(\mathbf{x}_i, \mathbf{x}_j)]_{i=1,\ldots,m, j=1,\ldots m}$, and $\mathbf{c}(\mathbf{x}) = [c(\mathbf{x}, \mathbf{x}_1), \ldots, c(\mathbf{x}, \mathbf{x}_n)]^T$. Here, β and s^2 are estimated using generalized least square estimates (see, [JSW98]),

$$\hat{\beta} = \frac{\mathbf{1}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{y}}{\mathbf{1}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{1}}, \tag{C.2}$$

$$\hat{s} = \frac{(\mathbf{y} - \mathbf{1} \cdot \hat{\beta})^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{y} - \mathbf{1} \cdot \hat{\beta})}{m}. \tag{C.3}$$

The correlation function $c(\mathbf{x}, \mathbf{x}')$ can have different shapes. The choice of the correlation function can be based on a-priori knowledge or on the correlation structure, e.g., maximum likelihood estimation. Typically, isotropic kernel functions are of the form

$$c_\theta(\mathbf{x}, \mathbf{x}') = \exp\left(-\theta \cdot |\mathbf{x} - \mathbf{x}'|^q\right) \ , q > 0. \tag{C.4}$$

Setting θ is done by maximizing the likelihood of $\mathcal{F}_{\mathbf{x}_1} = \mathbf{y}_1 \wedge \ldots \wedge \mathcal{F}_{\mathbf{x}_m} = \mathbf{y}_m$ by minimizing

$$m \log \hat{s}(\theta) + \log \det \mathbf{C}(\theta). \tag{C.5}$$

The same principle can be applied for multiparametric kernels, though this requires multi-dimensional optimization demanding considerably more evaluations of the log-likelihood expression.

The calibration procedure is algorithmically described in Algorithm C.1. In this description, the minimization of θ is omitted. In this work, a grid search method is adopted using a logarithmically scaled grid on the interval $[10^{-50}, 10^5]$.

Advantages of Kriging are that it is an exact interpolator (i.e., it returns the sample value as the estimate at the sample points) and that the prediction confidence range can be locally assessed.

Algorithm C.1: Ordinary Kriging Calibration

Input: Archive $A = \{(\mathbf{x}_1, f_1), \ldots, (\mathbf{x}_m, f_m)\}$

Output: Kriging model $f_{\text{krig}}(\mathbf{x})$

1: **Set parameters:** $q \leftarrow 2$, minimization method for θ

2: $\theta \leftarrow \min_\theta \{m \log \hat{s}(\theta) + \log\det \mathbf{C}(\theta)\}$, with

$$\mathbf{C} \leftarrow \begin{bmatrix} c_\theta(\mathbf{x}_1, \mathbf{x}_1) & \cdots & c_\theta(\mathbf{x}_1, \mathbf{x}_m) \\ \vdots & \ddots & \vdots \\ c_\theta(\mathbf{x}_m, \mathbf{x}_1) & \cdots & c_\theta(\mathbf{x}_m, \mathbf{x}_m) \end{bmatrix}, \; c_\theta(\mathbf{x}, \mathbf{x}') = \exp\left(-\theta \cdot |\mathbf{x} - \mathbf{x}'|^q\right),$$

$$\hat{s} = \frac{(\mathbf{y} - \mathbf{1} \cdot \hat{\beta})^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{y} - \mathbf{1} \cdot \hat{\beta})}{m}.$$

3: $f_{\text{krig}}(\mathbf{x}) \leftarrow \hat{\beta} + (\mathbf{y} - \mathbf{1} \cdot \hat{\beta})^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(\mathbf{x})$, with

$$\hat{\beta} = \frac{\mathbf{1}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{y}}{\mathbf{1}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{1}}, \; \mathbf{c}(\mathbf{x}) = [c_\theta(\mathbf{x}, \mathbf{x}_1), \ldots, c_\theta(\mathbf{x}, \mathbf{x}_n)]^T.$$

4: **return** $f_{\text{krig}}(\mathbf{x})$

A disadvantage is the computational effort (repeated inversion and determinant computation for $\mathbf{C}(\theta)$ within the likelihood minimization over θ). Besides, the points $\mathbf{x}_1, \ldots, \mathbf{x}_m$ do not only have to be unique so that the $\mathbf{C}$ is positive definite, but one should also ensure that these points are well distributed to prevent that $\mathbf{C}$ gets ill-conditioned, which could cause failure due to numerical errors.