Statistica Sinica: Supplement

SEQUENTIAL PARETO MINIMIZATION OF PHYSICAL SYSTEMS USING CALIBRATED COMPUTER SIMULATORS

The Ohio State University

Supplementary Material

S1 The Bivariate Distribution of the Calibrated Predictor

The bivariate distribution of the calibrated predictor which is used to calculate the proposed minimax fitness function in Section 4.3 can be derived using the result of the conditional distribution of the multivariate normal. First, the statement of the conditional distribution of the multivariate normal is given in Result 1, and the proof is given, for example, in Result 5.2.10 of Ravishanker and Dey (2001).

Result 1: Let $\boldsymbol{Y}_1$ be an $n_1 \times 1$ random vector and $\boldsymbol{Y}_2$ be an $n_2 \times 1$ random vector with

$$\left.\begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix}\right| \boldsymbol{\beta} \sim N\left(\begin{bmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \end{bmatrix}\boldsymbol{\beta}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right), \tag{S1.1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector, $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ are $n_1 \times p$ and $n_2 \times p$ matrices respectively with full column rank, and $\boldsymbol{\Sigma}_{ij} = Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ for $i, j = 1, 2$. Then,

$$[\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2] \sim N\left(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}\right), \tag{S1.2}$$

where

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{F}_1\boldsymbol{\beta} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{F}_2\boldsymbol{\beta}), \quad \text{and} \tag{S1.3}$$

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^{\top}. \tag{S1.4}$$

To derive the bivariate distribution of calibrated predictor, first note that the joint distribution of $[U_\ell(\boldsymbol{x}_i), U_\ell(\boldsymbol{x}), \mathcal{Y}_\ell]^\top$ is a jointly multivariate normal distribution as following:

$$\left.\begin{bmatrix} U_\ell(\boldsymbol{x}_i) \\ U_\ell(\boldsymbol{x}) \\ \mathcal{Y}_\ell \end{bmatrix}\right| \Omega \sim N\left(\begin{bmatrix} \eta_\ell \\ \eta_\ell \\ \eta_\ell \mathbf{1}_{n^p+n^s} \end{bmatrix}, \begin{bmatrix} \lambda_{\ell,Z}^{-1} + \lambda_{\ell,\delta}^{-1} & \Sigma_{\ell,\boldsymbol{x}_i,\boldsymbol{x}}^\top & \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}}^\top \\ \Sigma_{\ell,\boldsymbol{x}_i,\boldsymbol{x}} & \lambda_{\ell,Z}^{-1} + \lambda_{\ell,\delta}^{-1} & \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}}^\top \\ \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}} & \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}} & \Sigma_{\ell,\mathcal{Y}} \end{bmatrix}\right), \tag{S1.5}$$

where the definition of all notation in (S1.5) can be found in (21) and (22) of the paper.

By a straightforward application of Result 1, we can get that

$[(U_\ell(\boldsymbol{x}_i), U_\ell(\boldsymbol{x}))^\top \mid \mathcal{Y}^c, \boldsymbol{\Omega}] \sim N(\boldsymbol{\mu}_{\ell,\boldsymbol{x}_i,\boldsymbol{x},\mathcal{U}}, \Sigma_{\ell,\boldsymbol{x}_i,\boldsymbol{x},\mathcal{U}})$, where

$$\boldsymbol{\mu}_{\ell,\boldsymbol{x}_i,\boldsymbol{x},\mathcal{U}} = \begin{bmatrix} \eta_\ell \\ \eta_\ell \end{bmatrix} + \begin{bmatrix} \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}}^\top \\ \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}}^\top \end{bmatrix} \Sigma_{\ell,\mathcal{Y}}^{-1}(\mathcal{Y}_\ell - \eta_\ell \mathbf{1}_{n^p+n^s}), \quad \text{and} \tag{S1.6}$$

$$\begin{aligned} \Sigma_{\ell,\boldsymbol{x}_i,\boldsymbol{x},\mathcal{U}} &= \begin{bmatrix} \lambda_{\ell,Z}^{-1} + \lambda_{\ell,\delta}^{-1} & \Sigma_{\ell,\boldsymbol{x}_i,\boldsymbol{x}}^\top \\ \Sigma_{\ell,\boldsymbol{x}_i,\boldsymbol{x}} & \lambda_{\ell,Z}^{-1} + \lambda_{\ell,\delta}^{-1} \end{bmatrix} - \begin{bmatrix} \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}}^\top \\ \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}}^\top \end{bmatrix} \Sigma_{\ell,\mathcal{Y}}^{-1} \begin{bmatrix} \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}} & \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_{\ell,Z}^{-1} + \lambda_{\ell,\delta}^{-1} - \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}}^\top \Sigma_{\ell,\mathcal{Y}}^{-1} \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}} & \Sigma_{\ell,\boldsymbol{x}_i,\boldsymbol{x}}^\top - \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}}^\top \Sigma_{\ell,\mathcal{Y}}^{-1} \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}} \\ \Sigma_{\ell,\boldsymbol{x}_i,\boldsymbol{x}} - \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}}^\top \Sigma_{\ell,\mathcal{Y}}^{-1} \Sigma_{\ell,\boldsymbol{x}_i,\mathcal{Y}} & \lambda_{\ell,Z}^{-1} + \lambda_{\ell,\delta}^{-1} - \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}}^\top \Sigma_{\ell,\mathcal{Y}}^{-1} \Sigma_{\ell,\boldsymbol{x},\mathcal{Y}} \end{bmatrix}. \end{aligned} \tag{S1.7}$$

Thus, the mean (21) and the variance (22) of the distribution of $U_\ell^* = U_\ell^*(\boldsymbol{x}_i) - U_\ell^*(\boldsymbol{x})$ in Section 4.3 can be derived using (S1.6) and (S1.7) directly.

S2 A Property of the Minimax Fitness Funtion

The following theorem shows that when the mean function $\boldsymbol{\mu}(\boldsymbol{x})$ is known, the minimax fitness function (14) is zero at any $\boldsymbol{x}$ that has been previously observed.

Theorem 1. *Suppose the mean function $\boldsymbol{\mu}(\boldsymbol{x}) = (\mu_1(\boldsymbol{x}), \ldots, \mu_m(\boldsymbol{x}))$ of an m-output physical system is known. Let $\mathcal{X}^p = \{\boldsymbol{x}_1^p, \ldots, \boldsymbol{x}_{n^p}^p\}$ be the set of n^p control inputs for initial physical observations and $\mathcal{X}^s = \{\boldsymbol{x}_1^s, \ldots, \boldsymbol{x}_{n^s}^s\}$ be the set of n^s control inputs for initial simulator runs. Define $\mathcal{P}_X^{n^p+n^s}$ to be the current Pareto Set based on the known mean function $\boldsymbol{\mu}(\boldsymbol{x})$. If a new point $\boldsymbol{x}$ is in the set of previously explored points, i.e., $\boldsymbol{x} \in \mathcal{X}^p \bigcup \mathcal{X}^s$, then $I_{\mathcal{F}}(\boldsymbol{\mu}(\boldsymbol{x})) = 0$.*

Proof. Let $\boldsymbol{x} \in \mathcal{X}^p \bigcup \mathcal{X}^s$. First, if $\boldsymbol{x} \notin \mathcal{P}_X^{n^p+n^s}$, then based on the definition of the Pareto Front and Set, there exists at least one $\boldsymbol{x}_i \in \mathcal{P}_X^{n^p+n^s}$ satisfying $\mu_\ell(\boldsymbol{x}_i) - \mu_\ell(\boldsymbol{x}) \leq 0$ for every $\ell = \{1, \ldots, m\}$. Then, we have

$$\max_{\ell=1,\ldots,m} (\mu_\ell(\boldsymbol{x}_i) - \mu_\ell(\boldsymbol{x})) \leq 0.$$

Thus,

$$\min_{\boldsymbol{x}_i \in \mathcal{P}_X^{n^p+n^s}} \max_{\ell=1,\ldots,m} (\mu_\ell(\boldsymbol{x}_i) - \mu_\ell(\boldsymbol{x})) \leq 0 \quad \text{and} \quad I_{\mathcal{F}}(\boldsymbol{\mu}(\boldsymbol{x})) = 0.$$

Second, if $\boldsymbol{x} \in \mathcal{P}_X^{n^p+n^s}$, then taking $\boldsymbol{x}_i = \boldsymbol{x}$ gives $\mu_\ell(\boldsymbol{x}_i) - \mu_\ell(\boldsymbol{x}) = 0$ for every $\ell = \{1, \ldots, m\}$ and hence

$$\max_{\ell=1,\ldots,m} (\mu_\ell(\boldsymbol{x}_i) - \mu_\ell(\boldsymbol{x})) = 0.$$

For every other $\boldsymbol{x}_j \in \mathcal{P}_X^{n^p+n^s}$ with $j \neq i$, $\mu_\ell(\boldsymbol{x}_j) - \mu_\ell(\boldsymbol{x}) \geq 0$ for at least one $\ell = \{1, \ldots, m\}$. Thus

$$\max_{\ell=1,\ldots,m} (\mu_\ell(\boldsymbol{x}_j) - \mu_\ell(\boldsymbol{x})) \geq 0.$$

We conclude

$$\min_{\boldsymbol{x}_i \in \mathcal{P}_X^{n^p+n^s}} \max_{\ell=1,\ldots,m} (\mu_\ell(\boldsymbol{x}_i) - \mu_\ell(\boldsymbol{x})) = 0 \quad \text{and} \quad I_{\mathcal{F}}(\boldsymbol{\mu}(\boldsymbol{x})) = 0.$$

□

S3 Expected Improvement Function of the *MOP2* Function Example

As stated in Section 4.1, the expected minimax function (15) is not necessarily zero at inputs previously observed due to the presence of measurement error or bias in the output, depending on the experimental platform used. However, while this may lead to an additional $\boldsymbol{x}$ being selected close to a previously used input, our examples have not revealed exact duplication. For example, Figure 1 shows that, for the *MOP2* function of Section 5.1, the global maximum of the expected improvement function $E[I_F]$ is considerably larger than the values of the $E[I_F]$ of the 25 points in the initial design.

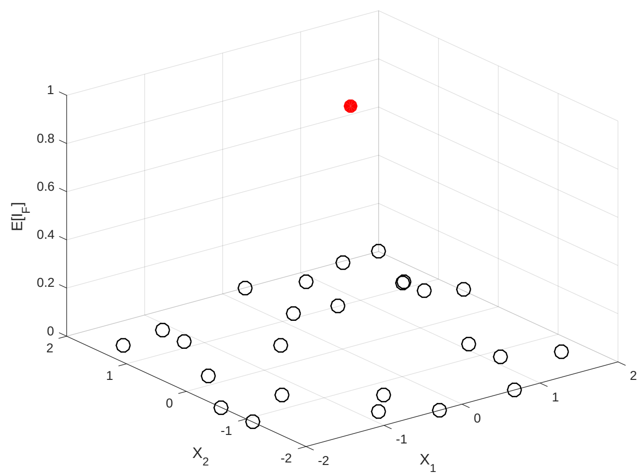

Figure 1: Values of the $E[I_F]$ for the 25 points in the initial design for the *MOP2* example of Section 5.1 together with the global maximum (red) of the $E[I_F]$.