OPERATIONS RESEARCH
DOI 10.1287/opre.1090.0754ec

e-companion

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"Stochastic Kriging for Simulation Metamodeling"
by Bruce Ankenman, Barry L. Nelson, and Jeremy Staum,
Operations Research, DOI 10.1287/opre.1090.0754.

Proofs

EC.1. MSE-optimal Stochastic Kriging Predictor

Here we prove the results in Section 2 for stochastic kriging when $\mathbf{\Sigma}_{\mathsf{M}}, \mathbf{\Sigma}_{\varepsilon}$ and β_0 are known. Most of these results parallel similar derivations in Stein (1999) for standard kriging, to which the reader can refer for missing details.

We consider linear predictors of the form $\lambda_0(\mathbf{x}_0) + \boldsymbol{\lambda}(\mathbf{x}_0)^\top \bar{\mathcal{Y}}$ where $\lambda_0(\mathbf{x}_0)$ and $\boldsymbol{\lambda}(\mathbf{x}_0)$ are weights that depend on $\mathbf{x}_0$ and will be chosen to give minimum MSE for predicting $\mathsf{Y}(\mathbf{x}_0) = \beta_0 + \mathsf{M}(\mathbf{x}_0)$. To simplify the notation we drop the dependence of λ_0 and $\boldsymbol{\lambda}$ on $\mathbf{x}_0$.

The MSE of the linear predictor is

$$
\begin{aligned}
\mathrm{MSE} &= \mathrm{E}\left[(\mathsf{Y}(\mathbf{x}_0) - \lambda_0 - \boldsymbol{\lambda}^\top \bar{\mathcal{Y}})^2\right] \\
&= \mathrm{E}\left[(\beta_0 - \lambda_0 - \boldsymbol{\lambda}^\top \mathbf{1}_k \beta_0)^2\right] + \mathrm{E}\left[(\mathsf{M}(\mathbf{x}_0) - \boldsymbol{\lambda}^\top \bar{\mathcal{Y}} + \boldsymbol{\lambda}^\top \mathbf{1}_k \beta_0)^2\right] \\
&\quad + 2\mathrm{E}\left[(\beta_0 - \lambda_0 - \boldsymbol{\lambda}^\top \mathbf{1}_k \beta_0)(\mathsf{M}(\mathbf{x}_0) - \boldsymbol{\lambda}^\top \bar{\mathcal{Y}} + \boldsymbol{\lambda}^\top \mathbf{1}_k \beta_0)\right] \\
&= (\beta_0 - \lambda_0 - \boldsymbol{\lambda}^\top \mathbf{1}_k \beta_0)^2 + \mathrm{E}\left[(\mathsf{M}(\mathbf{x}_0) - \boldsymbol{\lambda}^\top \bar{\mathcal{Y}} + \boldsymbol{\lambda}^\top \mathbf{1}_k \beta_0)^2\right] \\
&= (\beta_0 - \lambda_0 - \boldsymbol{\lambda}^\top \mathbf{1}_k \beta_0)^2 + \Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{x}_0) + \boldsymbol{\lambda}^\top \left[\mathbf{\Sigma}_{\mathsf{M}} + \mathbf{\Sigma}_{\varepsilon}\right] \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^T \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot).
\end{aligned}
$$

Now following the same steps as Stein (1999, Chapter 1.2), we can show that the MSE-optimal weights are $\lambda_0^\star = \beta_0 - \boldsymbol{\lambda}^\top \mathbf{1}_k \beta_0$ and $\boldsymbol{\lambda}^\star = \left[\mathbf{\Sigma}_{\mathsf{M}} + \mathbf{\Sigma}_{\varepsilon}\right]^{-1} \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot)$. With these weights, the estimator is

$$
\widehat{\mathsf{Y}}(\mathbf{x}_0) = \beta_0 + \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot)^\top \left[\mathbf{\Sigma}_{\mathsf{M}} + \mathbf{\Sigma}_{\varepsilon}\right]^{-1} \left(\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_k\right).
$$

By direct substitution the MSE is

$$
\begin{aligned}
\mathrm{MSE}^\star &= \Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{x}_0) - \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot)^\top \left[\mathbf{\Sigma}_{\mathsf{M}} + \mathbf{\Sigma}_{\varepsilon}\right]^{-1} \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot) \\
&= \Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{x}_0) - \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot)^\top \mathbf{\Sigma}_{\mathsf{M}}^{-1} \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot) + \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot)^\top \mathbf{\Xi}\, \mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot). \qquad \text{(EC.1)}
\end{aligned}
$$

The final step follows from Henderson and Searle (1981), which shows that $\mathbf{\Xi}$ will be a positive definite matrix that depends on $\mathbf{\Sigma}_{\varepsilon}$ and $\mathbf{\Sigma}_{\mathsf{M}}$ and will disappear if $\mathbf{\Sigma}_{\varepsilon} = \mathbf{0}$.

EC.2. Impact of CRN for $k > 2$ Design Points

As a second illustration of the impact of CRN on stochastic kriging, suppose that

$$\mathbf{\Sigma}_{\mathsf{M}} + \mathbf{\Sigma}_{\varepsilon} = \tau^2\ \mathbf{I} + \mathsf{V}\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

This matrix is invertible in closed form using Theorem 8.3.3 in Graybill (1969) (see also Appendix EC.5). Therefore, the MSE of the stochastic kriging estimator (everything known) can be obtained by plugging $\mathbf{\Sigma}_{\mathsf{M}}, \mathbf{\Sigma}_{\varepsilon}$ and $\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot) = \tau^2(r_0, r_0, \ldots, r_0)^\top$ into (7) to obtain

$$\mathrm{MSE}^\star = \tau^2\left(1 - \frac{kr_0^2\tau^2}{(1+(k-1)\rho)\mathsf{V} + \tau^2}\right)$$

which is increasing in ρ.

EC.3. Joint Distribution

The multivariate normality of $(\mathsf{Y}(\mathbf{x}_0), \bar{\mathcal{Y}}(\mathbf{x}_1), \ldots, \bar{\mathcal{Y}}(\mathbf{x}_k))^\top$ follows directly from the response surface model (3) and Assumption 1. They imply that

$$\begin{aligned}&\mathrm{Cov}[\mathcal{Y}_j(\mathbf{x}_i), \mathcal{Y}_\ell(\mathbf{x}_h)] = \\ &\quad \mathrm{Cov}[\mathsf{M}(\mathbf{x}_i) + \varepsilon_j(\mathbf{x}_i), \mathsf{M}(\mathbf{x}_h) + \varepsilon_\ell(\mathbf{x}_h)] = \begin{cases} \tau^2 + \mathsf{V}(\mathbf{x}_i), & i = h, j = \ell \\ \tau^2, & i = h, j \neq \ell \\ \tau^2 R_{\mathsf{M}}(\mathbf{x}_i - \mathbf{x}_j; \boldsymbol{\theta}) & i \neq h \end{cases}\end{aligned}$$

and similarly $\mathrm{Cov}[\mathsf{Y}(\mathbf{x}_0), \mathcal{Y}_j(\mathbf{x}_i)] = \mathrm{Cov}[\mathsf{M}(\mathbf{x}_0), \mathsf{M}(\mathbf{x}_i) + \varepsilon_j(\mathbf{x}_i)] = \tau^2 R_{\mathsf{M}}(\mathbf{x}_0 - \mathbf{x}_i; \boldsymbol{\theta})$. Since the $\varepsilon_j(\mathbf{x}_i)$ are independent across replications and design points, averaging the n_i replications at design point $\mathbf{x}_i$ only affects

$$\mathrm{Cov}[\bar{\mathcal{Y}}(\mathbf{x}_i), \bar{\mathcal{Y}}(\mathbf{x}_i)] = \mathrm{Var}[\bar{\mathcal{Y}}(\mathbf{x}_i)] = \tau^2 + \mathsf{V}(\mathbf{x}_i)/n_i.$$

Therefore,

$$\begin{pmatrix} \mathsf{Y}(\mathbf{x}_0) \\ \bar{\mathcal{Y}}(\mathbf{x}_1) \\ \vdots \\ \bar{\mathcal{Y}}(\mathbf{x}_k) \end{pmatrix} \sim \mathrm{MVN}\left[\beta_0 \mathbf{1}_{k+1}, \begin{pmatrix} \tau^2 & \mathbf{R}_{\mathsf{M}}(\mathbf{x}_0, \cdot; \boldsymbol{\theta})^\top \\ \mathbf{R}_{\mathsf{M}}(\mathbf{x}_0, \cdot; \boldsymbol{\theta}) & \tau^2 \mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta}) + \mathrm{Diag}\{\mathsf{V}(\mathbf{x}_i)/n_i\} \end{pmatrix}\right]. \tag{EC.2}$$

The conditional expectation of $\mathsf{Y}(\mathbf{x}_0)$ given $(\bar{\mathcal{Y}}(\mathbf{x}_1), \ldots, \bar{\mathcal{Y}}(\mathbf{x}_k))^\top$ then follows immediately from standard results for the multivariate normal distribution (e.g., Stein 1999, Appendix A).

EC.4. Proof of Theorem 1

We first show that for any fixed positive definite covariance matrix $\boldsymbol{\Sigma}'_{\varepsilon}$, the predictor

$$\widehat{\mathsf{Y}}'(\mathbf{x}_0) = \beta_0 + \boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^\top \left[\boldsymbol{\Sigma}_{\mathsf{M}} + \boldsymbol{\Sigma}'_{\varepsilon}\right]^{-1} \left(\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_k\right)$$

is an unbiased predictor. This follows immediately from $\mathrm{E}[\widehat{\mathsf{Y}}'(\mathbf{x}_0) - \mathsf{Y}(\mathbf{x}_0)] = \beta_0 + 0 - \beta_0 = 0$.

Next notice that under Model (3) we have

$$\mathcal{S}^2(\mathbf{x}_i) = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(\mathcal{Y}_j(\mathbf{x}_i) - \bar{\mathcal{Y}}(\mathbf{x}_i)\right)^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(\varepsilon_j(\mathbf{x}_i) - \bar{\varepsilon}_j(\mathbf{x}_i)\right)^2.$$

Therefore, under Assumption 1, $\widehat{\mathsf{V}}(\mathbf{x}_i) = \mathcal{S}^2(\mathbf{x}_i)$ is independent of $\bar{\mathcal{Y}}(\mathbf{x}_i) = \beta_0 + \mathsf{M}(\mathbf{x}_i) + \bar{\varepsilon}_j(\mathbf{x}_i)$ by well-known properties of the normal distribution (recall that M is also independent of $\varepsilon_j(\mathbf{x}_i)$). Then since $\widehat{\boldsymbol{\Sigma}}_{\varepsilon} = \mathrm{Diag}\left\{\widehat{\mathsf{V}}(\mathbf{x}_1)/n_1, \widehat{\mathsf{V}}(\mathbf{x}_2)/n_2, \ldots, \widehat{\mathsf{V}}(\mathbf{x}_k)/n_k\right\}$ we have

$$\mathrm{E}\left[\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0) - \mathsf{Y}(\mathbf{x}_0)\right] = \mathrm{E}\left[\mathrm{E}\left(\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0) - \mathsf{Y}(\mathbf{x}_0)\,\middle|\,\widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right)\right] = \mathrm{E}[\beta_0 - \beta_0] = 0.$$

EC.5. Variance Inflation Example of Section 3.1

As in §3.1, suppose that

$$\boldsymbol{\Sigma}_{\mathsf{M}} = \tau^2 \begin{pmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \vdots & \vdots & \ddots & \vdots \\ r & r & \cdots & 1 \end{pmatrix},$$

$\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot) = \tau^2(r_0, r_0, \ldots, r_0)^\top$ with $r_0, r \geq 0$, $\boldsymbol{\Sigma}_{\varepsilon} = (\mathsf{V}/n)\mathbf{I}$, and we have an independent estimator $\widehat{\mathsf{V}} \sim \mathsf{V}\chi^2_{n-1}/(n-1)$. Let $\gamma = \mathsf{V}/\tau^2$ be the ratio of the intrinsic variance to the extrinsic variance. A key to the result is noting that we can write

$$\boldsymbol{\Sigma}_{\mathsf{M}} + \boldsymbol{\Sigma}_{\varepsilon} = \mathbf{q}\mathbf{q}^T + \tau^2 \mathrm{Diag}\{1-r, 1-r, \ldots, 1-r\} + \frac{\mathsf{V}}{n}\,\mathbf{I}$$

where $\mathbf{q}^T = (\sqrt{r}, \sqrt{r}, \ldots, \sqrt{r})$. This matrix is invertible in closed form using Theorem 8.3.3 of Graybill (1969). Specialized to this case

$$\begin{aligned} [\boldsymbol{\Sigma}_{\mathsf{M}} + \boldsymbol{\Sigma}_{\varepsilon}]^{-1} &= \mathrm{Diag}\left\{\frac{1}{\tau^2(1-r) + \mathsf{V}/n}, \ldots, \frac{1}{\tau^2(1-r) + \mathsf{V}/n}\right\} \\ &\quad - \frac{\tau^2}{\tau^2(1-r) + k\tau^2 r + \mathsf{V}/n}\left\|\frac{r}{\tau^2(1-r) + \mathsf{V}/n}\right\| \end{aligned} \tag{EC.3}$$

where $||\cdot||$ indicates a matrix of appropriate dimension with all elements the same.

We obtain the MSE of $\widehat{\mathsf{Y}}(\mathbf{x}_0)$, the stochastic kriging estimator with $\boldsymbol{\Sigma}_\varepsilon$ (i.e., V) known,

$$\mathrm{MSE}^\star = \tau^2\left(1-\frac{kr_0^2}{1+(k-1)r+\frac{\gamma}{n}}\right)$$

by plugging $\boldsymbol{\Sigma}_{\mathsf{M}}, \boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)$ into (7) and using the known inverse (EC.3).

The MSE of $\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0)$ is derived by substituting $\widehat{\mathsf{V}}$ for V and noting that since $\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0)$ is unbiased for any $\widehat{\mathsf{V}}$ independent of $\bar{\mathcal{Y}}$ (see Theorem 1 and the proof in Section EC.4)

$$\begin{aligned}
\mathrm{MSE}\left[\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0)\right] &= \mathrm{Var}\left[\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0)\right] \\
&= \mathrm{E}\left[\mathrm{Var}\left(\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0)\,\middle|\widehat{\mathsf{V}}\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0)\,\middle|\widehat{\mathsf{V}}\right)\right] \\
&= \mathrm{E}\left[\mathrm{Var}\left(\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0)\,\middle|\widehat{\mathsf{V}}\right)\right] \\
&= \mathrm{E}\left[\mathrm{Var}\left(\beta_0+\tau^2(r_0,r_0,\ldots,r_0)^\top\left[\boldsymbol{\Sigma}_{\mathsf{M}}+\frac{\widehat{\mathsf{V}}}{n}\mathbf{I}\right]^{-1}\left(\bar{\mathcal{Y}}-\beta_0\mathbf{1}_k\right)\middle|\widehat{\mathsf{V}}\right)\right] \\
&= \tau^2\mathrm{E}\left[\left(1+\frac{\left(1+(k-1)r+\frac{\gamma}{n}\right)kr_0^2}{\left(1+(k-1)r+\frac{\gamma}{n}\frac{\widehat{\mathsf{V}}}{\mathsf{V}}\right)^2}-\frac{2kr_0^2}{\left(1+(k-1)r+\frac{\gamma}{n}\frac{\widehat{\mathsf{V}}}{\mathsf{V}}\right)}\right)\right].
\end{aligned}$$

The last step requires writing out the conditional variance, recalling that $\bar{\mathcal{Y}}$ and $\widehat{\mathsf{V}}$ are independent, and employing several tedious applications of Theorem 8.3.3 of Graybill (1969).

EC.6. Maximum Likelihood Estimators

The log likelihood function for $(\beta_0,\tau^2,\boldsymbol{\theta})$ given $(\bar{\mathcal{Y}}(\mathbf{x}_1),\ldots,\bar{\mathcal{Y}}(\mathbf{x}_k))^\top$ follows directly from the joint normal distribution (EC.2):

$$\begin{aligned}
&\ell(\beta_0,\tau^2,\boldsymbol{\theta}) = \\
&\quad -\ln\left[(2\pi)^{k/2}\right]-\frac{1}{2}\ln\left[\left|\tau^2\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta})+\boldsymbol{\Sigma}_\varepsilon\right|\right]-\frac{1}{2}\left(\bar{\mathcal{Y}}-\beta_0\mathbf{1}_k\right)^\top\left[\tau^2\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta})+\boldsymbol{\Sigma}_\varepsilon\right]^{-1}\left(\bar{\mathcal{Y}}-\beta_0\mathbf{1}_k\right).
\end{aligned}$$

For brevity, let $\boldsymbol{\Sigma}=\tau^2\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta})+\boldsymbol{\Sigma}_\varepsilon$. Taking derivatives, and applying the matrix calculus results (22) and (23) gives

$$\frac{\partial\ell(\beta_0,\tau^2,\boldsymbol{\theta})}{\partial\beta_0} = -\beta_0\mathbf{1}_k^T\boldsymbol{\Sigma}^{-1}\mathbf{1}_k+\mathbf{1}_k^T\boldsymbol{\Sigma}^{-1}\bar{\mathcal{Y}}=\mathbf{1}_k^T\boldsymbol{\Sigma}^{-1}(\bar{\mathcal{Y}}-\beta_0\mathbf{1}_k)=0 \qquad \text{(EC.4)}$$

$$\frac{\partial \ell(\beta_0,\tau^2,\boldsymbol{\theta})}{\partial \tau^2} = -\frac{1}{2}\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta})\right] - \frac{1}{2}\left(\bar{\mathcal{Y}} - \beta_0\mathbf{1}_k\right)^\top \left\{-\boldsymbol{\Sigma}^{-1}\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}\right\}\left(\bar{\mathcal{Y}} - \beta_0\mathbf{1}_k\right) = 0 \tag{EC.5}$$

$$\frac{\partial \ell(\beta_0,\tau^2,\boldsymbol{\theta})}{\partial \theta_p} = -\frac{1}{2}\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial \tau^2\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta})}{\partial \theta_p}\right] - \frac{1}{2}\left(\bar{\mathcal{Y}} - \beta_0\mathbf{1}_k\right)^\top \left\{-\boldsymbol{\Sigma}^{-1}\frac{\partial \tau^2\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta})}{\partial \theta_p}\boldsymbol{\Sigma}^{-1}\right\}\left(\bar{\mathcal{Y}} - \beta_0\mathbf{1}_k\right) = 0 \tag{EC.6}$$

where θ_p is the pth element of $\boldsymbol{\theta}$. Observe that $\widehat{\beta}_0$ that solves (EC.4) is explicit given $\boldsymbol{\Sigma}$.

To obtain starting solutions to search for the MLEs we set $\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta}) = \mathbf{I}$ and simplify (EC.4) and (EC.5).

EC.7. Estimated MSE

To obtain the MSE estimator (25), we first assume $\boldsymbol{\Sigma}_{\mathsf{M}} = \tau^2\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_\varepsilon$ are known, so that only β_0 need be estimated. An expression for $\mathrm{MSE}(\mathbf{x}_0)$ then follows precisely the development in Stein (1999, §1.5) with $\tau^2\mathbf{R}_{\mathsf{M}}(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_\varepsilon$ playing the role of $\mathbf{K}$, the covariance matrix of the random field. Our estimator $\widehat{\mathrm{MSE}}(\mathbf{x}_0)$ is then obtained by plugging in $\widehat{\tau}^2\mathbf{R}_{\mathsf{M}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ obtained via maximum likelihood estimation.

EC.8. Derivation of IMSE-Optimal Design

The first step in deriving the IMSE-optimal allocation of replications is to note that

$$\begin{aligned}\mathrm{MSE}(\mathbf{x}_0;\mathbf{n}) &= \Sigma_{\mathsf{M}}(\mathbf{x}_0,\mathbf{x}_0) - \boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^\top\boldsymbol{\Sigma}(\mathbf{n})^{-1}\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)\\ &= \tau^2 - \tau^4\mathbf{r}(\mathbf{x}_0)^\top\boldsymbol{\Sigma}(\mathbf{n})^{-1}\mathbf{r}(\mathbf{x}_0)\\ &= \tau^2 - \tau^4\sum_{i,j=1}^{k}\left[\boldsymbol{\Sigma}(\mathbf{n})^{-1}\right]_{i,j} r_i(\mathbf{x}_0)r_j(\mathbf{x}_0)\end{aligned}$$

where $\tau^2\mathbf{r}(\mathbf{x}_0)^\top = (\Sigma_{\mathsf{M}}(\mathbf{x}_0,\mathbf{x}_1),\Sigma_{\mathsf{M}}(\mathbf{x}_0,\mathbf{x}_2),\ldots,\Sigma_{\mathsf{M}}(\mathbf{x}_0,\mathbf{x}_k))$. Therefore,

$$\begin{aligned}\mathrm{IMSE}(\mathbf{n}) &= \tau^2 - \tau^4\sum_{i,j=1}^{k}\left[\boldsymbol{\Sigma}(\mathbf{n})^{-1}\right]_{i,j}\int_{\mathbf{x}_0\in\mathcal{X}} r_i(\mathbf{x}_0)r_j(\mathbf{x}_0)\,d\mathbf{x}_0\\ &= \tau^2 - \tau^4\mathbf{1}^\top\left[\mathbf{W}\circ\boldsymbol{\Sigma}(\mathbf{n})^{-1}\right]\mathbf{1}\end{aligned}$$

where $\circ$ is the Hadamard product of matrices. From this we can form the Lagrangian

$$L(\mathbf{n}) = \tau^2 - \tau^4\mathbf{1}^\top\left[\mathbf{W}\circ\boldsymbol{\Sigma}(\mathbf{n})^{-1}\right] + \lambda(N - \mathbf{1}^\top\mathbf{n}).$$

The first-order optimality conditions are

$$\frac{\partial L(\mathbf{n})}{\partial n_i} = -\tau^4 \mathbf{1}^\top \left[\mathbf{W} \circ \frac{\partial}{\partial n_i} \boldsymbol{\Sigma}(\mathbf{n})^{-1}\right] \mathbf{1} - \lambda = 0$$

for $i = 1, 2, \ldots, k$. Careful application of matrix calculus gives

$$\begin{aligned} \frac{\partial}{\partial n_i} \boldsymbol{\Sigma}(\mathbf{n})^{-1} &= -\boldsymbol{\Sigma}(\mathbf{n})^{-1} \left[\frac{\partial}{\partial n_i} \boldsymbol{\Sigma}(\mathbf{n})\right] \boldsymbol{\Sigma}(\mathbf{n})^{-1} \\ \frac{\partial}{\partial n_i} \boldsymbol{\Sigma}(\mathbf{n}) &= -\frac{\mathsf{V}(\mathbf{x}_i)}{n_i^2} \mathbf{J}^{(ii)} \end{aligned}$$

where $\mathbf{J}^{(ii)}$ is a $k \times k$ matrix with 1 in position (i, i) and zeroes elsewhere. The result then follows by solving the first-order conditions.

References

Graybill, F. A. 1969. *Matrices with Applications in Statistics*, 2nd edition, Wadsworth, Belmont, CA.

Henderson, H. V. and S. R. Searle. 1981. On deriving the inverse of a sum of matrices. *SIAM Review* **23**, 53–60.

Stein, M. L. 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, NY.