Particle Swarm Optimization for Spatial Design

Matthew Simpson[1]

Abstract

KEY WORDS:

[1](to whom correspondence should be addressed) Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100, themattsimpson@gmail.com

1 The Problem

Suppose we are interested in the latent spatial field of some response variable $Y(\boldsymbol{s})$, $\boldsymbol{s} \in \mathcal{D} \subseteq \Re^2$. Specifically, we are interested in predicting $Y(\boldsymbol{s})$ at a set of target locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M \in \mathcal{D}$. We have the ability to sample N locations anywhere in $\mathcal{D}$, and we wish to place them in order to optimize some design criterion. Let $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N \in \mathcal{D}$ denote the N sampled locations. Suppose that $Y(\boldsymbol{s})$ is a geostatistical process with mean function $\mu(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta}$ for some known covariate $\boldsymbol{x}(\boldsymbol{s})$ and covariance function $C(\boldsymbol{s}, \boldsymbol{t})$ for $\boldsymbol{s}, \boldsymbol{t} \in \mathcal{D}$. Once the design points are selected, we observe $Z(\boldsymbol{d}_i)$ for $i = 1, 2, \ldots, N$ where $Z(\boldsymbol{d}) = Y(\boldsymbol{d}) + \varepsilon(\boldsymbol{d})$ and $\varepsilon(\boldsymbol{d})$ is mean zero white noise with variance σ^2_ε, representing measurement error. An intuitive criterion to minimize is average mean square prediction error (MSPE) from kriging across each of the target locations.

1.1 Simple Kriging

In simple kriging, $C(.,.)$, $\boldsymbol{\beta}$, and σ^2_ε are all treated as known. Let $\boldsymbol{Z} = (Z(\boldsymbol{d}_1), Z(\boldsymbol{d}_2), \ldots, Z(\boldsymbol{d}_N))'$, $\boldsymbol{X} = (\boldsymbol{x}(\boldsymbol{d}_1), \boldsymbol{x}(\boldsymbol{d}_2), \ldots, \boldsymbol{x}(\boldsymbol{d}_N))'$, $\boldsymbol{C}_Z = \mathrm{cov}(\boldsymbol{Z})$ where $\mathrm{cov}(Z(\boldsymbol{d}_i), Z(\boldsymbol{d}_j)) = C(\boldsymbol{d}_i, \boldsymbol{d}_j) + \sigma^2_\varepsilon 1(\boldsymbol{d}_i = \boldsymbol{d}_j)$, and $\boldsymbol{c}_Y(\boldsymbol{s}_0) = \mathrm{cov}(Y(\boldsymbol{s}_0), \boldsymbol{Z})$ where $\mathrm{cov}(Y(\boldsymbol{s}_0), Z(\boldsymbol{d}_i)) = C(\boldsymbol{s}_0, \boldsymbol{d}_i)$. Then the simple kriging predictor of $Y(\boldsymbol{s}_0)$ is the linear predictor $\boldsymbol{Y}^*(\boldsymbol{s}_0) = \boldsymbol{\lambda}'\boldsymbol{Z} + k$ that MSPE conditional on sampled locations, $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N$. Specifically it minimizes $\mathrm{E}[Y(\boldsymbol{s}_0) - Y^*(\boldsymbol{s}_0)]^2$ over $\boldsymbol{\lambda}$ and k such that $\boldsymbol{\lambda}'\boldsymbol{1} = 1$ where $\boldsymbol{1}$ is a column vector of ones. The simple kriging predictor is easily derived as

$$Y^*(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta})$$

with MSPE

$$S^2_{sk}(\boldsymbol{s}_0) = C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0).$$

The simple kriging MSPE a function of the design points, $\boldsymbol{D} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N)$, through $\boldsymbol{c}_Y(\boldsymbol{s}_0)$ and $\boldsymbol{C}_Z^{-1}$.

From a Bayesian perspective, this predictor can be rationalized by a particular Bayesian hierachical model (HM) that has a bit more structure. Specifically, we now assume that $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ and $\{\varepsilon(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ are independent Gaussian processes. Then the posterior predictive distribution for $Y(\boldsymbol{s}_0)$ can be derived as

$$[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] = \int [Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z}][\boldsymbol{Y}|\boldsymbol{Z}]d\boldsymbol{Y}$$

where it is easy to derive that

$$\begin{aligned}
Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z} &\sim N\left[\boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right], \\
\boldsymbol{Y}|\boldsymbol{Z} &\sim N\left[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}), \boldsymbol{C}_Y - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}\boldsymbol{C}_Y\right]
\end{aligned}$$

so that $Y(\boldsymbol{s}_0)|\boldsymbol{Z}$ is Gaussian with

$$\begin{aligned}
\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= \mathrm{E}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} = \boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}), \\
\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= \mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} + \mathrm{var}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} \\
&= C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0).
\end{aligned}$$

Then the design criterion we minimize is average simple kriging variance, $\overline{S}^2_{sk} = \sum_{i=1}^N S^2_{sk}(\boldsymbol{s}_i)/N$. Note that in order to $\overline{S}^2_{sk}$, we need to know $\boldsymbol{x}(\boldsymbol{s})$ a priori for each location $\boldsymbol{s} \in \mathcal{D}$. Often for this to be feasible, $\boldsymbol{x}(\boldsymbol{s})$ must be some know function of $\boldsymbol{s}$ and not a covariate that must be measured at location $\boldsymbol{s}$. For example if $\boldsymbol{s} = (u, v)$, then $\boldsymbol{x}(\boldsymbol{s}) = (1, u, v)'$ requires no additional measurement.

Naively, it appears strange to minimize a *posterior* predictive variance prior to seeing the data. However, in the simple kriging setup, it turns out that $\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}]$ does not depend on $\boldsymbol{Z}$. More generally, a Bayesian would minimize the *expected* posterior predictive variance, $\mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}]\right\}$.

1.2 Universal Kriging

A major limitation of simple kriging is that typically $\boldsymbol{\beta}$, σ^2_ε, and $C(.,.)$ are unknown. Universal kriging attempts to remedy this by allowing $\boldsymbol{\beta}$ to be unknown. The universal kriging predictor is

$$\widehat{Y}(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)'\widehat{\boldsymbol{\beta}}_{gls} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{gls})$$

where $\widehat{\boldsymbol{\beta}}_{gls} = (\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{Z}$ is the generalized least squares estimate of $\boldsymbol{\beta}$, and the MSPE of $\widehat{Y}(\boldsymbol{s}_0)$ is

$$\begin{aligned}\overline{S}^2_{uk}(\boldsymbol{s}_0) &= C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\ &\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'[\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X}][\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)].\end{aligned}$$

This can also be justified via a Bayesian HM by using the same Gaussian assumptions as before, but further assuming that $\boldsymbol{\beta}$ has the improper uniform prior, i.e. $\boldsymbol{\beta} \sim U(-\infty, \infty)$. Let $\boldsymbol{C}_Y = \mathrm{cov}(\boldsymbol{Y})$ where $\mathrm{cov}(Y(\boldsymbol{d}_i), Y(\boldsymbol{d}_j)) = C(\boldsymbol{d}_i, \boldsymbol{d}_j)$. Now we have

$$[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] = \int\int [Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z}][\boldsymbol{Y}|\boldsymbol{Z}]d\boldsymbol{Y}d\boldsymbol{\beta}$$

where

$$\begin{aligned}Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z} &\sim N\left[\boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right], \\ \boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{Z} &\sim N\left[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}), \boldsymbol{C}_Y - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}\boldsymbol{C}_Y\right], \\ \boldsymbol{\beta}|\boldsymbol{Z} &\sim N\left[\widehat{\boldsymbol{\beta}}_{gls}, (\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\right],\end{aligned}$$

so that $Y(\boldsymbol{s}_0)|\boldsymbol{Z}$ is again Gaussian with

$$\begin{aligned}\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= \mathrm{E}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} = \boldsymbol{x}(\boldsymbol{s}_0)'\widehat{\boldsymbol{\beta}}_{gls} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{gls}), \\ \mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= \mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} + \mathrm{var}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} \\ &= C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) + \mathrm{var}\left\{[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{Y}|\boldsymbol{Z}\right\}.\end{aligned}$$

To compute the last variance term, note that

$$\begin{aligned}\operatorname{cov}(\boldsymbol{Y}|\boldsymbol{Z}) &= \mathrm{E}[\operatorname{cov}(\boldsymbol{Y}|\boldsymbol{\beta},\boldsymbol{Z})|\boldsymbol{Z}] + \operatorname{cov}[\mathrm{E}(\boldsymbol{Y}|\boldsymbol{\beta},\boldsymbol{Z})|\boldsymbol{Z}] \\ &= \boldsymbol{C}_Y - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}\boldsymbol{C}_Y + (\boldsymbol{I} - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{C}_Z^{-1}\boldsymbol{C}_Y)\end{aligned}$$

and

$$\begin{aligned}\operatorname{cov}(\boldsymbol{Y},\boldsymbol{\beta}|\boldsymbol{X}) &= \mathrm{E}[\operatorname{cov}(\boldsymbol{Y},\boldsymbol{\beta}|\boldsymbol{\beta},\boldsymbol{Z})|\boldsymbol{Z}] + \operatorname{cov}[\mathrm{E}(\boldsymbol{Y}|\boldsymbol{\beta},\boldsymbol{Z}),\boldsymbol{\beta}|\boldsymbol{Z}] \\ &= (\boldsymbol{I} - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}.\end{aligned}$$

Then we have

$$\begin{aligned}&\operatorname{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] = \\ &\quad C(\boldsymbol{s}_0,\boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\ &\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'[\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X}][\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)] \\ &\quad + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\left\{\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1} + (\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1})\right\}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\ &\quad + \boldsymbol{c}_Y(\boldsymbol{s}_0)'(\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)] \\ &\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}(\boldsymbol{s}_0)'(\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1})\boldsymbol{c}_Y(\boldsymbol{s}_0),\end{aligned}$$

which reduces to

$$\begin{aligned}\operatorname{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= C(\boldsymbol{s}_0,\boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\ &\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'[\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X}][\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)].\end{aligned}$$

Then the design criterion in this case is $\overline{S}_{uk}^2 = \sum_{i=1}^N S_{uk}^2(\boldsymbol{s}_i)/N$, and once again $\mathrm{E}\{\operatorname{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}]\} = \operatorname{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}]$, which is constant in $\boldsymbol{Z}$.

1.3 Full Uncertainty Kriging

In both simple and universal kriging the predictors and their MSPEs are the same as the posterior predictive means and posterior predictive variances for a suitably chosen

Bayesian HM. In both cases if we wish to choose the sampled locations, $\boldsymbol{d}_1, \boldsymbol{d}_2, \dots, \boldsymbol{d}_N$, in order to minimize the MSPE from the kriging estimate, we are effectively minimizing $\sum_{i=1}^{N} \mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_i)|\boldsymbol{Z}(\boldsymbol{D})]\right\}/N$ in $\boldsymbol{D}$. We can easily extend this to the case where both σ_ε^2 and $C(.,.)$ are unknown, though it may no longer correspond to the optimal linear predictor. Suppose the covariance function is parameterized by $\boldsymbol{\theta} \in \Theta$, denoted by $C_{\boldsymbol{\theta}}(.,.)$, and let $\boldsymbol{\phi} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$. Assume that a priori $\boldsymbol{\phi} \sim [\boldsymbol{\phi}]$ — we relax the restriction that $\boldsymbol{\beta} \sim U(-\infty, \infty)$. Then we can write the expected posterior predictive variance of $Y(\boldsymbol{s}_0)$ as

$$\mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}]\right\} = \mathrm{E}\Bigg(\mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}]|\boldsymbol{Z}\right\}\Bigg) + \mathrm{E}\Bigg(\mathrm{var}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}]|\boldsymbol{Z}\right\}\Bigg)$$

where from the universal kriging case we know

$$\begin{aligned}
\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}] &= \left\{\left[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right](\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\right\}\boldsymbol{C}_Z^{-1}\boldsymbol{Z}, \\
\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}] &= C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\
&\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'[\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X}][\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)].
\end{aligned}$$

[PROBLEM: $\mathrm{var}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}]|\boldsymbol{Z}\right\}$ IS NOT AVAILABLE IN CLOSED FORM]

2 other stuff

In practice both σ_ε^2 and $C(.,.)$ are unknown. Assume that the covariance function is parameterized, $C_{\boldsymbol{\theta}}(.,.)$, $\boldsymbol{\theta} \in \Theta$. Standard practice is to estimate $\boldsymbol{\theta}$ and σ_ε^2 by some means, then plug in those estimates to the formula for $\widehat{\boldsymbol{\beta}}_{gls}$, to obtain the feasible GLS estimate, and into the universal kriging formula. However in the Bayesian context, we can continue to minimize the posterior predictive variance. Let $\boldsymbol{\phi} = (\sigma_\varepsilon^2, \boldsymbol{\theta})$ and assume that $\boldsymbol{\phi}$ is independent of $\boldsymbol{\beta}$ in the prior with $\boldsymbol{\phi} \sim [\boldsymbol{\phi}]$. Then we can write

[NOTE: THIS IS NOW A FUNCTION OF Z! So do we minimize the expected prior variance...?]

Then the generalized least squares estimator of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}_{gls} = (\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{Z}$, the universal kriging predictor of $Y(\boldsymbol{s}_0)$ is $\widehat{Y}(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)'\widehat{\boldsymbol{\beta}}_{gls} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{gls})$, and its mean square prediction error is

$$\sigma^2_{\widehat{Y}}(\boldsymbol{d};\boldsymbol{s}_0) = \boldsymbol{C}_Y(\boldsymbol{s}_0,\boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z(\boldsymbol{d})^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) +$$
$$\left[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}(\boldsymbol{d})'\boldsymbol{C}_Z(\boldsymbol{d})^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right]' \left[\boldsymbol{X}(\boldsymbol{d})'\boldsymbol{C}_Z(\boldsymbol{d})^{-1}\boldsymbol{X}(\boldsymbol{d})\right]^{-1} \left[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}(\boldsymbol{d})'\boldsymbol{C}_Z(\boldsymbol{d})^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right]$$

where $\boldsymbol{d} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N)'$ and $\boldsymbol{d}_i = (u_i, v_i)'$ for $i = 1, 2, \ldots, N$. Then the design criterion is

$$U(\boldsymbol{d}) = \frac{1}{M}\sum_{j=1}^{M}\sigma^2_{\widehat{Y}}(\boldsymbol{d};\boldsymbol{s}_j)$$

where $\{\boldsymbol{s}_j\}$ are the M locations we wish to predict and our goal is to minimize U in $\boldsymbol{d}$. Alternatively if we wish to learn about the entire spatial domain, we can minimize

$$U_C(\boldsymbol{d}) = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}}\sigma^2_{\widehat{Y}}(\boldsymbol{d};\boldsymbol{s})d\boldsymbol{s},$$

though this integral is unlikely to be available in closed form and so in practice we would approximate with a criterion with the form of $U(\boldsymbol{d})$. We could also modify $U(\boldsymbol{d})$ by attaching weights to the spatial locations if some locations are more important than others.

We can consider two versions of this optimizatio problem: when $M > N$, i.e. we want to predict at more locations than we can observe, or the opposite case when $M < N$. When $M < N$, it is sensible to restrict ourself to designs where the first M observed locations are exactly the M locations at which we want to predict [CAN WE PROVE THIS?]. When $M > N$, it is no longer necessarily the case that putting a design location at a prediction location is a good idea.

[SHOULD WE CONSIDER OTHER OBJECTIVE FUNCTIONS? SOMETHING DEPENDING ON ENTROPY?]

Also note that $U(\boldsymbol{d})$ depends on the covariance function, $C(\boldsymbol{s},\boldsymbol{t})$, which may depend on unknown parameters. In that case, we can put a prior on those unknown parameters and instead minimize $\mathrm{E}_{\boldsymbol{\theta}}[U(\boldsymbol{d};\boldsymbol{\theta})] = \int_\Theta U(\boldsymbol{d};\boldsymbol{\theta})[\boldsymbol{\theta}]d\boldsymbol{\theta}$ where $[\boldsymbol{\theta}]$ is the prior on $\boldsymbol{\theta}$. [CONNECTION TO FACT THAT KRIGING CAN BE DERIVED FROM A BAYESIAN HIERARCHICAL LINEAR MODEL]