Constrained linear prediction

Jack Fraser-Govil and Amery Gration

2023

So. I've changed the notation. Please don't be too cross—I've not done this gratuitously. It's mainly to make this paper consistent with the one I wrote with Sylvy. But it's probably best to explain my motivation face-to-face.

An introduction explaining our motivation for constructing the BSCLP and BSCLUP. To be written after we have found a tame statistician.

THE BEST LINEAR UNBIASED PREDICTOR

Consider the random processes $Z := \{Z_t\}_{t\in T}$, where T is an arbitrary index set, and $\mathbf{X} := \{Z_{t_i} + E_{t_i}\}_{i=1}^{n}$, where $t_1, \ldots, t_n \in T$ and E_{t_i} is a centred random variable. We think of $\mathbf{X}$ as a noisy sample of Z. A linear predictor, $\hat{Z}_t$, for $Z_t \in Z$ is a linear combination of the elements of $\mathbf{X}$. This predictor may be written as

$$\hat{Z}_t = \sum_{i=1}^{n} a_i X_{t_i} \tag{1}$$

for real numbers $a_1, \ldots, a_n$. The *best linear unbiased predictor* (BLUP) for Z_t is the linear predictor, $\hat{Z}_t^{\mathrm{BLUP}}$, that has least mean-square error among all linear predictors that are unbiased. The BLUP for Z_t is the linear predictor for which the mean-square error (MSE),

$$\mathrm{MSE}(\hat{Z}_t) = \mathrm{E}\big((Z_t - \hat{Z}_t)^2\big), \tag{2}$$

is a minimum among all linear predictors for which it is the case that

$$\mathrm{E}(Z_t - \hat{Z}_t) = 0. \tag{3}$$

We may write Z_t as the decomposition

$$Z_t = m(t) + Y_t \tag{4}$$

where $m : T \longrightarrow \mathbb{R}$ is the *mean-value function*, given by $m(t) = \mathrm{E}(Z_t)$ and Y_t is a centred random variable. Suppose that the mean-value function is an element of the linear vector space spanned by the basis functions $\varphi_1, \ldots, \varphi_p$. Then the mean-value function may be written as

$$m(t) = \sum_{i=1}^{p} \beta_i \varphi_i(t) \tag{5}$$

for some coefficients $\beta_1, \ldots, \beta_p$.

In matrix notation a linear predictor for Z_t may be written as

$$\hat{Z}_t = \mathbf{a}^{\mathrm{T}}\mathbf{X} \tag{6}$$

where $\mathbf{a} := (a_i)_{i=1}^n$ is a column vector of length n. Similarly, we may write

$$Z_t = \boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\beta} + Y_t \tag{7}$$

and

$$\mathbf{X} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{Y} + \mathbf{E} \tag{8}$$

where $\boldsymbol{\varphi} := (\varphi_i(t))_{i=1}^p$ is a column vector of length p, $\boldsymbol{\Phi} := (\varphi_i(t_j))_{i,j}$ is a matrix of size $m \times n$, $\boldsymbol{\beta} := (\beta_i)_{i=1}^{ps}$ is a column vector of length m, and both $\mathbf{Y} := (Y_{t_i})_{i=1}^n$ and $\mathbf{E} := (E_{t_i})_{i=1}^n$ are column vectors of length n. The MSE is then

$$\mathrm{MSE}(\hat{Z}_t) = \mathrm{var}(Z_t) + \mathbf{a}^{\mathrm{T}}\mathbf{K}\mathbf{a} - 2\mathbf{a}^{\mathrm{T}}\mathbf{k} \tag{9}$$

where $\mathbf{K} := (\mathrm{cov}(X_{t_i}, X_{t_j}))_{i,j}$ is a matrix of size $n \times n$ and $\mathbf{k} := (\mathrm{cov}(Z_t, X_{t_i}))_{i=1}^n$ is a column vector of length n.[1] The unbiasedness constraint is

$$\boldsymbol{\Phi}\mathbf{a} - \boldsymbol{\varphi} = \mathbf{0}. \tag{10}$$

Following Goldberger (1962) we may find the BLUP for Z_t using the method of Lagrangian multipliers. To find the BLUP for Z_t we minimize the MSE subject to the unbiasedness constraint by forming the Lagrangian,

$$L = \mathbf{a}^{\mathrm{T}}\mathbf{K}\mathbf{a} - 2\mathbf{a}^{\mathrm{T}}\mathbf{k} - 2\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{\Phi}\mathbf{a} - \boldsymbol{\varphi}), \tag{11}$$

where $\boldsymbol{\lambda} := (\lambda_i)_{i=1}^p$ is a column vector of Lagrangian multipliers of length m. By taking the appropriate partial derivatives we find that

$$\frac{\partial L}{\partial \mathbf{a}} = 2\mathbf{K}\mathbf{a} - 2\mathbf{k} - 2\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\lambda} \text{ and} \tag{12}$$

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = 2\boldsymbol{\Phi}\mathbf{a} - 2\boldsymbol{\varphi}. \tag{13}$$

In the case of the BLUP these partial derivatives vanish giving a pair of equations with a solution $(\mathbf{a}^{\mathrm{BLUP}}, \boldsymbol{\lambda}^{\mathrm{BLUP}})$. The first element of this pair is

$$\mathbf{a}^{\mathrm{BLUP}} = \mathbf{K}^{-1}(\mathbf{I} - \boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\mathbf{K}^{-1}\boldsymbol{\Phi}^{\mathrm{T}})^{-1}\boldsymbol{\Phi}\mathbf{K}^{-1})\mathbf{k} + \mathbf{K}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\mathbf{K}^{-1}\boldsymbol{\Phi}^{\mathrm{T}})^{-1}\boldsymbol{\varphi}. \tag{14}$$

Hence the BLUP for Z_t is

$$\hat{Z}_t^{\mathrm{BLUP}} = \boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{K}^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}\mathbf{K}^{-1}\mathbf{X} + \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{X} - \boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{K}^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}\mathbf{K}^{-1}\mathbf{X}). \tag{15}$$

with MSE

$$\mathrm{MSE}(\hat{Z}_t^{\mathrm{BLUP}}) = \mathrm{var}(Z_t) + (\mathbf{a}^{\mathrm{BLUP}})^{\mathrm{T}}\mathbf{K}\mathbf{a}^{\mathrm{BLUP}} - 2(\mathbf{a}^{\mathrm{BLUP}})^{\mathrm{T}}\mathbf{k} \tag{16}$$

$$= \mathrm{var}(Z_t) - \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{k} + (\boldsymbol{\varphi} - \boldsymbol{\Phi}\mathbf{K}^{-1}\mathbf{k})^{\mathrm{T}}(\boldsymbol{\Phi}\mathbf{K}^{-1}\boldsymbol{\Phi})^{-1}(\boldsymbol{\varphi} - \boldsymbol{\Phi}\mathbf{K}^{-1}\mathbf{k}). \tag{17}$$

The BLUP for Z_t exists if and only if m is contained by the span of the basis functions $\varphi_1, \ldots, \varphi_p$. Moreover, if the BLUP for Z_t exists then it is unique.

[1] We use to term 'covariance' to mean *proper covariance*, i.e. the covariance of the random variables X and Y is $\mathrm{cov}(X, Y) := \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y)))$. Accordingly, the variance of a random variable X is $\mathrm{var}(X) := \mathrm{E}((X - \mathrm{E}(X))^2)$.

Remark 1 (interpretation of the BLUP). As Goldberger pointed out, the BLUP for Z_t is a sum of two terms: the generalized least-squares estimator for the mean of Z_t and a weighted sum of the residuals of the generalized least-squares estimators of the elements of $\mathbf{X}$. Let us write the GLS estimator for the mean of Z_t as $\hat{\Theta}_t := \boldsymbol{\varphi}^{\mathrm{T}}(\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{K}^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}\mathbf{K}^{-1}\mathbf{X}$, the tuple of GLS estimators for the means of the elements of $\mathbf{X}$ as $\hat{\boldsymbol{\Theta}} := \boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{K}^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}\mathbf{K}^{-1}\mathbf{X}$ (a random vector of length n), and the tuple of the residuals of $\hat{\boldsymbol{\Theta}}$ as $\mathbf{D} := \mathbf{X} - \hat{\boldsymbol{\Theta}}$ (a random vector of length n). With these definitions made we may now write the BLUP for Z_t as

$$\hat{Z}_t^{\mathrm{BLUP}} = \hat{\Theta}_t + \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{D}. \tag{18}$$

Remark 2 (prediction intervals for the BSCLUP). Since $\hat{Z}_t^{\mathrm{BLUP}}$ is an unbiased predictor for Z_t it is the case that $\mathrm{var}(Z_t - \hat{Z}_t^{\mathrm{BLUP}}) = \mathrm{MSE}(\hat{Z}_t^{\mathrm{BLUP}})$. Therefore we may construct either (i) bounds for a prediction interval using Chebyshev's inequality or (ii) an equivalent Gaussian prediction interval using Gaussian critical values.

THE BEST LINEAR PREDICTOR

The *best linear predictor* (BLP) for Z_t is the linear predictor, $\hat{Z}_t^{\mathrm{BLP}}$, that has least mean-square error among all linear predictors. In this case, we make the definitions $\mathbf{R} := (\mathrm{E}(X_{t_i}X_{t_j}))_{i,j}$, which is a matrix of size $n \times n$, and $\mathbf{r} := (\mathrm{E}(Z_t X_{t_i}))_{i=1}^n$, which is a column vector of length n. Then the MSE is

$$\mathrm{MSE}(\hat{Z}_t) = \mathbf{a}^{\mathrm{T}}\mathbf{R}\mathbf{a} - 2\mathbf{a}^{\mathrm{T}}\mathbf{r}. \tag{19}$$

To find the BLP for Z_t we minimize the MSE directly. This has the derivative

$$\frac{\mathrm{d}\,\mathrm{MSE}(\hat{Z}_t)}{\mathrm{d}\mathbf{a}} = 2\mathbf{R}\mathbf{a} - 2\mathbf{r} \tag{20}$$

which vanishes giving an equation with the solution

$$\mathbf{a}^{\mathrm{BLP}} = \mathbf{R}^{-1}\mathbf{r}. \tag{21}$$

Hence the BLP for Z_t is

$$\hat{Z}_t^{\mathrm{BLP}} = \mathbf{r}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{X} \tag{22}$$

with MSE

$$\mathrm{MSE}(\hat{Z}_t^{\mathrm{BLP}}) = \mathrm{E}(Z_t^2) - \mathbf{r}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}. \tag{23}$$

The BLP for Z_t always exists and is unique.

Remark 3 (bias of the BLP). The BLP for Z_t is in general a biased predictor for Z_t. It is unbiased if and only if the mean of Z_t vanishes. In this case the BLP for Z_t is coincident with the BLUP for Z_t since $\boldsymbol{\Phi} = \mathbf{0}$ and $\mathrm{cov}(X_{t_i}, X_{t_j}) = \mathrm{E}(X_{t_i}X_{t_j})$.

THE BEST SEQUENCE OF CONSTRAINED LINEAR UNBIASED PREDICTORS

We are often interested in constructing sets or sequences of linear predictors. Alongside unbiasedness we might wish to impose additional constraints on the elements of these sets or sequences. For example, in the case of a set of linear predictors, we might wish to impose nonnegativity. Similarly, in the case of a sequence of linear predictors, we might wish impose monotonicity. This moves us to consider sets and sequences of predictors in their own right.

Consider again, then, the random processes $Z := \{Z_t\}_{t\in T}$ and $\mathbf{X} := \{Z_{t_i} + E_{t_i}\}_{i=1}^{n}$. A set of linear predictors, $\hat{\mathbf{Z}} := \{\hat{Z}_{t_j}\}_{j=1}^{m}$, for a set of random variables, $\mathbf{Z} := \{Z_{t_j}\}_{j=1}^{m}$, is a set of linear combinations of the elements of $\mathbf{X}$. The j -th element of this set may be written

$$\hat{Z}_{t_j} = \sum_{i=1}^{n} a_{ji} X_{t_i} \tag{24}$$

for real numbers $a_{j1}, \ldots, a_{jn}$. As before, we may write any element of Z as the sum of its mean and a centred random variable (Eq. 4) and then write the mean as a linear combination of basis functions (Eq. 5). We have called a predictor 'best' among some class of predictors if it has the least MSE of those predictors. Similarly, we will call a sequence of predictors 'best' among some class of predictors if it has least total MSE,

$$\sum_{j=1}^{m} \mathrm{MSE}(\hat{Z}_{t_j}) = \sum_{j=1}^{m} \mathrm{E}((Z_{t_j} - \hat{Z}_{t_j})^2), \tag{25}$$

of all sequences of those predictors. We may now define the *best sequence of constrained linear unbiased predictors* (BSCLUP) for $\mathbf{Z}$ to be the sequence of linear predictors, $\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} := (\hat{Z}_{t_j}^{\mathrm{BSCLUP}})_{j=1}^{m}$, that has least total mean-square error, among all sequences of linear predictors that are unbiased, such that

$$\mathrm{E}(Z_{t_j} - \hat{Z}_{t_j}) = 0 \text{ for all } j = 1, \ldots, m, \tag{26}$$

and satisfy the constraints

$$f_i(\hat{\mathbf{Z}}) = 0 \text{ and} \tag{27}$$

$$g_j(\hat{\mathbf{Z}}) \geq 0 \tag{28}$$

for functions f_i and g_j where $i = 1, \ldots, u$ and $j = 1, \ldots, v$. (Note that these are constraints on the realized values of $\hat{\mathbf{Z}}$.) In the spirit of Goldberger (1962) we may find the j -th element of the BSCLUP for $\mathbf{Z}$ using the method of Lagrangian multipliers. However, we have imposed constraints on $\hat{\mathbf{Z}}$ despite seeking an optimal value for the tuple $(a_{ji})_{i=1}^{n}$. We therefore limit ourselves to the case of linear constraints, where

$$f_i(\hat{\mathbf{Z}}) = \sum_{k=1}^{m} (b_{ik} \hat{\mathbf{Z}}_k) - c_i, \text{ and} \tag{29}$$

$$g_j(\hat{\mathbf{Z}}) = \sum_{k=1}^{m} (b_{jk} \hat{\mathbf{Z}}_k) - c_j. \tag{30}$$

When we come to form the Lagrangian we will be then be able to take the required derivatives. We could enforce the inequality constraints through slack variables and use the Karush–Kuhn–Tucker theorem. However, for our purposes it is better to parameterize the inequality constraints such that they become equality constraints. Specifically, we parameterize the constraint vector such that

$$c_j = \psi_j(\mathbf{w}) \tag{31}$$

for vector $\mathbf{w} = (w_i)_{i=1}^{\nu}$ and some function $\psi_k : \mathbb{R}^{\nu} \longrightarrow \mathbb{R}$. For example, if we wished to enforce positivity, we might choose $a_{jk} = \delta_{jk}$ (the Kronecker delta) and $c_j = \exp(w_j)$, whereupon the constraint would become $\sum_{k=1}^{m}(\delta_{jk}\hat{\mathbf{Z}}_k - \exp(w_j)) = 0$. Such a parameterization allows us to treat both inequality and equality constraints using the same formalism by rewriting the inequality constraints as

$$\sum_{k=1}^{m}(b_{jk}\hat{\mathbf{Z}}_k) - \psi_j(\mathbf{w}) = 0. \tag{32}$$

We may now restate the problem in matrix notation and proceed to perform the optimization, just as we did in the case of the BLUP. A linear predictor for Z_{t_j} (the j -th element of the sequence $\mathbf{Z}$) may be written as

$$\hat{Z}_{t_j} = \mathbf{a}_j^{\mathrm{T}}\mathbf{X} \tag{33}$$

where $\mathbf{a}_j := (a_{ji})_{i=1}^{n}$ is a column vector of length n. Similarly, we may write

$$Z_{t_j} = \boldsymbol{\varphi}_j^{\mathrm{T}}\boldsymbol{\beta} + Y_{t_j} \tag{34}$$

where $\boldsymbol{\varphi}_j := (\varphi_i(t_j))_{i=1}^{m}$ is a column vector of length m. The random vector $\mathbf{X}$ may be decomposed in the way already done (Equation 8). In matrix form the MSE is

$$\sum_{j=1}^{m}\mathrm{MSE}(\hat{Z}_{t_j}) = \sum_{j=1}^{m}(\mathrm{var}(Z_{t_j}) + \mathbf{a}_j^{\mathrm{T}}\mathbf{K}\mathbf{a}_j - 2\mathbf{a}_j^{\mathrm{T}}\mathbf{k}_j) \tag{35}$$

where $\mathbf{K} := (\mathrm{cov}(X_{t_i}, X_{t_j}))_{i,j}$ is a matrix of size $n \times n$, as before, and $\mathbf{k}_j := (\mathrm{cov}(Z_{t_j}, X_{t_i}))_{i=1}^{n}$ is a column vector of length n.[1] The unbiasedness constraints may be written as

$$\boldsymbol{\Phi}\mathbf{a}_j - \boldsymbol{\varphi}_j = \mathbf{0} \text{ for all } j = 1\ldots, m. \tag{36}$$

and the additional constraints may be written as

$$\mathbf{B}\hat{\mathbf{Z}} - \mathbf{c} = \mathbf{0} \tag{37}$$

where $\mathbf{B} = (b_{ij})_{i,j}$ is a matrix of size $q \times m$ and $\mathbf{c} = (c_i)_{i=1}^{m}$ is a column vector of length m. In matrix form the Lagrangian is

$$L = \sum_{j=1}^{m}(\mathrm{var}(Z_{t_j}) + \mathbf{a}_j^{\mathrm{T}}\mathbf{K}_j\mathbf{a}_j - 2\mathbf{a}_j^{\mathrm{T}}\mathbf{k}_j - 2\boldsymbol{\lambda}_j^{\mathrm{T}}(\boldsymbol{\Phi}\mathbf{a}_j - \boldsymbol{\varphi}_j)) - \boldsymbol{\mu}^{\mathrm{T}}(\mathbf{B}\hat{\mathbf{Z}} - \mathbf{c}) \tag{38}$$

where $\boldsymbol{\lambda}_j := (\lambda_{ji})_{i=1}^{p}$ and $\boldsymbol{\mu} := (\mu_i)_{i=1}^{q}$ are column vectors of Lagrangian multipliers of lengths p and q. The derivatives of the Lagrangian are

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{a}_j} &= 2\mathbf{K}\mathbf{a}_{t_i} - 2\mathbf{k}_i - \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\lambda} - \frac{\partial \hat{\mathbf{Z}}}{\partial \mathbf{a}_j}(\mathbf{B}^{\mathrm{T}}\boldsymbol{\mu}) \\ &= 2\mathbf{K}\mathbf{a}_{t_i} - 2\mathbf{k}_i - \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\lambda} - [\mathbf{B}^{\mathrm{T}}\boldsymbol{\mu}]_j,\end{aligned} \tag{39}$$

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = \boldsymbol{\Phi}\mathbf{a}_{t_i} - \boldsymbol{\varphi}_{t_i}, \text{ and} \tag{40}$$

$$\frac{\partial L}{\partial \boldsymbol{\mu}} = \mathbf{B}\hat{\mathbf{Z}} - \mathbf{c}. \tag{41}$$

In the case of the BSCLUP these vanish giving us a system of three equations with a solution $(\mathbf{a}_j^{\mathrm{BSCLUP}}, \boldsymbol{\lambda}_j^{\mathrm{BSCLUP}}, \boldsymbol{\mu}_j^{\mathrm{BSCLUP}})$. From the first equation we find that

$$\mathbf{a}_j^{\mathrm{BSCLUP}} = \mathbf{K}^{-1}\mathbf{k}_j - 2\mathbf{K}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\lambda} - \mathbf{K}^{-1}[\mathbf{B}^{\mathrm{T}}\boldsymbol{\mu}]_j\hat{\mathbf{Z}}. \tag{42}$$

The second and third equations are restatements of our constraints, and may be substituted into the first giving

$$\boldsymbol{\lambda}_j^{\mathrm{BSCLUP}} = \boldsymbol{\Phi}\mathbf{K}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{K}\mathbf{k}_{t_j} + [\mathbf{B}^{\mathrm{T}}\boldsymbol{\mu}]_j\boldsymbol{\Phi}\mathbf{K}^{-1}\mathbf{X} - \boldsymbol{\varphi}_{t_j}) \tag{43}$$

and

$$\boldsymbol{\mu}^{\mathrm{BSCLUP}} = \frac{1}{\mathbf{D}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{X}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BLUP}} - \mathbf{c}) \tag{44}$$

where

$$\hat{\mathbf{Z}}^{\mathrm{BLUP}} = (\hat{Z}_j^{\mathrm{BLUP}})_{j=1}^{m} \tag{45}$$

is the vector of BLUPs for the elements of $\mathbf{Z}$. We find that

$$\mathbf{a}_j^{\mathrm{BSCLUP}} = \mathbf{a}_j^{\mathrm{BLUP}} - \frac{1}{\mathbf{D}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{X}}[\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}(\mathbf{A}\mathbf{X}^{\mathrm{BLUP}} - \mathbf{c})]_j\mathbf{K}^{-1}\mathbf{D} \tag{46}$$

and hence that the BSCLUP is

$$\hat{Z}_j^{\mathrm{BSCLUP}} = (\mathbf{a}_j^{\mathrm{BSCLUP}})^{\mathrm{T}}\mathbf{X} \tag{47}$$

$$= \hat{Z}_j^{\mathrm{BLUP}} - [\mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} - \mathbf{c})]_j \tag{48}$$

with MSE

$$\mathrm{MSE}(\hat{Z}_j) = \mathrm{var}(Z_{t_j}) + (\mathbf{a}_j^{\mathrm{BSCLUP}})^{\mathrm{T}}\mathbf{K}\mathbf{a}_j^{\mathrm{BSCLUP}} - 2(\mathbf{a}_j^{\mathrm{BSCLUP}})^{\mathrm{T}}\mathbf{k}_j \tag{49}$$

$$= \ldots. \tag{50}$$

If the constraint is exact then $\mathbf{c}$ is given and we may use these formulae to compute $\hat{Z}_j^{\mathrm{BSCLUP}}$ and $\mathrm{MSE}(\hat{Z}_j^{\mathrm{BSCLUP}})$ directly. But if the constraint is inexact then $\mathbf{c}$, and hence $\mathbf{a}_j^{\mathrm{BSCLUP}}$, is a function of some unknown variable, $\mathbf{w}$ (Equation 31), and we must optimize the Lagrangian with respect to this variable also. After some manipulation (the details of which are included in the Appendix), we find that

$$\frac{\partial L}{\partial w_i} = -(2(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} - \mathbf{c}))^{\mathrm{T}}\frac{\partial \mathbf{c}}{\partial w_i}. \tag{51}$$

This vanishes, telling us that either

(1) $(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} - \mathbf{c})$ vanishes,

(2) $\partial\mathbf{c}/\partial w_i$ vanishes,

(3) both $(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} - \mathbf{c})$ and $\partial\mathbf{c}/\partial w_i$ vanish, or

(4) $(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}(\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} - \mathbf{c})$ and $\partial\mathbf{c}/\partial w_i$ are orthogonal.

The first case holds if and only if $\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} = \hat{\mathbf{Z}}^{\mathrm{BLUP}}$ since if $\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} = \hat{\mathbf{Z}}^{\mathrm{BSCLUP}}$ then $\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BLUP}} - \mathbf{c} = \mathbf{0}$ and, conversely, if $\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BLUP}} - \mathbf{c} = \mathbf{0}$ then $\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} = \hat{\mathbf{Z}}^{\mathrm{BLUP}}$ (see equation 37). To interpret the remaining three cases we should note that our constraints restrict the parameter $\mathbf{w}$ to taking values within some bounded region of parameter space. The second case holds when a constraint is satisfied by a parameter lying on the boundary of that region. The third case holds when $\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} = \hat{\mathbf{Z}}^{\mathrm{BLUP}}$ and the constraint is again satisified on a boundary. The fourth case holds when there are multiple constraints that cannot be satisfied at the boundary of the region but are instead satisfied by some point within it (i.e. by some critical point). In this fourth case we have a first-order nonlinear differential equation, which we must, in general, solve numerically, for example using gradient descent.[2]

The BSCLUP for $\mathbf{Z}$ exists if and only if the mean-value function, m, is contained by the span of the basis functions $\varphi_1, \ldots, \varphi_p$ and the constraints are consistent and expressed in minimal form. Moreover, if the BSCLUP for $\mathbf{Z}$ exists then it is unique.

Remark 4 (interpretation of the BSCLUP).

Remark 5 (reduction to the BLUP). By definition the BSCLUP is a sequence of constrained BLUPs. If a sequence of BLUPs obeys the specified constraints then that sequence is necessarily the BSCLUP. We can see this explicitly by noting that that a sequence of BLUPs obeys the constraints when $\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BLUP}} = \mathbf{0}$ (Equation 37). In this case the second term in Equation 47 vanishes to give $\hat{\mathbf{Z}}^{\mathrm{BSCLUP}} = \hat{\mathbf{Z}}^{\mathrm{BLUP}}$. Equivalently (as we have already observed) if $\mathbf{B}\hat{\mathbf{Z}}^{\mathrm{BLUP}} - \mathbf{c} = \mathbf{0}$ then the derivative $\partial L / \partial w_i$ (Equation 51) vanishes and the sequence of BLUPs is an extremum of the Lagrangian.

Remark 6 (prediction intervals for the BSCLUP).

Initialization

Validation

Optimization

THE BEST SEQUENCE OF CONSTRAINED LINEAR PREDICTORS

CASE STUDIES: CONSTRAINED CURVE FITTING

Positive constraints

Monotonic constraints

Riemann-sum constraint

Joint positivity and Riemann-sum constraint

REFERENCES

Goldberger, Arthur S. 1962. "Best Linear Unbiased Prediction in the Generalized Linear Regression Model." *Journal of the American Statistical Association* 57 (298): 369–75. `https://doi.org/10.2307/2281645`.

Kingma, Diederik P., and Jimmy Ba. 2017. "Adam: A Method for Stochastic Optimization." January 29, 2017. `https://doi.org/10.48550/arXiv.1412.6980`.

[2]In our implementation (see Case studies: constrained curve fitting) we use the method of adaptive moment estimation (ADAM) proposed by Kingma and Ba (2017).

APPENDIX