MINIMAX PREDICTION DESIGNS, ROBUST AGAINST MISSPECIFIED RESPONSE AND ERROR STRUCTURES

Douglas P. Wiens

April 27, 2017

Abstract blah.

Key words and phrases Replication;

1 Introduction

Some motivation (in these examples from some of my old notes, $h$ is the correlation function and $g$ the variance function):

Example 1: Spatial design. Here, the covariates $x$ will contain information about the physical locations, and observations at different locations are expected to be correlated. The predictions $\hat{Y}(x)$ are obtained by (universal) kriging; this presupposes prior knowledge of, or some model for the covariance function $h(x, x')$ and for $g(x)$. The “classical” spatial design problem, in which the choices of $h$ and $g$, with $\psi \equiv 0$, are assumed to be exactly correct, is well studied – ... . An experimenter wishing robustness against misspecifications in these choices might seek predictions that minimize $IMSE$ in neighbourhoods of the assumed values; see Wiens (2005).

Example 2: Computer experimentation. In this context $x$ is often viewed as a vector of control variables. Models commonly used are as for those in spatial design, but with $\sigma^2 = 0$, reflecting the deterministic nature of the response. See Santner, Williams & Notz (2003), ... ..

Example 3: Model robust design of experiments. Here, $\psi(x)$ is adopted as a convenient method of characterizing the model misspecification. There is a considerable literature on this problem under independence, (Wiens, ...) have recently studied the problem with a variety of choices of correlation functions, but with $\psi \equiv 0$. In any case, the interpretation is again that the estimation is carried out assuming that $E[Y(x)] = f'(x)\theta$ exactly, and that any variation is due solely to the additive errors. The estimate $\hat{\theta}$ is then computed by least squares, and $\hat{Y}(x) = f'(x)\hat{\theta}$. Again however, the experimenter wishes some protection against errors in these assumptions.

---

1Department of Mathematical and Statistical Sciences; University of Alberta, Edmonton, Alberta, Canada T6G 2G1. e-mail: doug.wiens@ualberta.ca
2 Problem formulation

We consider design problems in 'approximately linear' models, for which the response variable \( Y \) is observed subject to random error at covariates \( x \in \chi \subset \mathbb{R}^q \), for a finite 'design space' \( \chi = \{ x_1, \ldots, x_N \} \). The response has 'approximate' mean \( f'(x) \theta \) for \( p \)-dimensional regressors \( f(x) \) and some unknown parameter \( \theta_{p \times 1} \).

More precisely, we consider a collection \( \{ Y_i = Y(x_i) \}_{i=1}^N \) of r.v.s, satisfying

\[
E[Y(x_i)] = f'(x_i) \theta_{p \times 1} + \psi(x_i), \quad i = 1, \ldots, N, \quad (1)
\]

for a function \( \psi(\cdot) \) quantifying the approximate nature of the experimenter’s, possibly incorrect, model assumption that \( E[Y(x)] = f'(x) \theta \). For identifiability we define

\[
\theta = \arg \min_{\eta} \sum_{x \in \chi} (E[Y(x)] - f'(x) \eta)^2.
\]

Defining \( \psi(x) = E[Y(x)] - f'(x) \theta \) then leads to (1) and to the orthogonality requirement

\[
F' \psi_N = 0_{p \times 1}; \quad (2)
\]

here \( \psi_N \) is the \( N \times 1 \) vector with elements \( \{ \psi(x_i) \}_{i=1}^N \) and \( F \) is the \( N \times p \) matrix with rows \( \{ f'(x_i) \}_{i=1}^N \). We assume that \( F \) has full column rank.

The \( Y_i \) are possibly dependent; with \( Y = (Y_1, \ldots, Y_N)' \) we represent the covariance matrix as

\[
C_N = \text{COV}[Y] \overset{df}{=} (\sigma_{ij})_{i,j=1,\ldots,N}.
\]

The intention is to predict 'future' values of \( \{ Y_i \} \). To this end, for each \( i \) we may observe \( n_i \geq 0 \) copies \( \{ Y_{ik}(x_i) \}_{ik=1}^{n_i} \) distributed in the same manner as \( Y_i \), but observed with i.i.d. measurement error with common variance \( \sigma_\varepsilon^2 \).

We suppose that \( \text{COV}[Y_{ik}, Y_{jl}] = \sigma_{ij} \) and that

\[
\text{COV}[Y_{ik}, Y_{jl}] = \sigma_{ij} + \begin{cases} \sigma_\varepsilon^2, & (i,k) = (j,l), \\ 0, & \text{otherwise}. \end{cases}
\]

It is convenient to express the various parameters in terms of \( \sigma_\varepsilon \). Thus, we impose the bound

\[
\| \psi_N \|^2 \leq \sigma_\varepsilon^2 \alpha_n^2, \quad (3)
\]

where \( \alpha_n^2 \) is possibly dependent on the study size \( n \) and \( \| \cdot \| \) is the Euclidean norm. For an induced matrix norm \( \| \cdot \|_M \) we impose a bound

\[
\| C_N \|_M \leq \sigma_\varepsilon^2 \beta_n^2, \quad (4)
\]
with again a possible dependence of $\beta_n$ on $n$, discussed below (see the remark following Theorem 2). A common choice is the spectral radius, which for symmetric matrices is the maximum eigenvalue $\|C_N\|_M = \text{ch}_{\text{max}}(C_N)$. In this case the interpretation is that the variance of linear combinations $\sum_{i=1}^N t_i Y_i$, with $\sum_{i=1}^N t_i^2 = 1$, is bounded by $\sigma^2_{\beta_n}$.

We note that the possible dependence of $\alpha_n$ and $\beta_n$ on $n$ is only for the asymptotics; for finite $n$ it can be absorbed into $\psi_N$ and $C_N$.

We define $\Psi$ to be the class of functions $\psi(\cdot)$ satisfying (2) and (3), and $C$ to be the class of positive semi-definite matrices satisfying (4).

Let $y$ be the $n \times 1$ vector of observations with subvectors $\{y_i\}_{n_i>0}$, where, when $n_i > 0$, $y_i = (Y_{i1}, ..., Y_{in_i})'$. Then $n = \sum n_i$ and $E[y] = X_{n \times p} \theta$ for

$$X_{n \times p} = \begin{pmatrix} 1_{n_1} f'(x_1) \\ \vdots \\ 1_{n_N} f'(x_N) \end{pmatrix}_{n_i>0}.$$

The covariances between $Y$ and the data form the matrix

$$C_{N,n} = \text{cov} [Y, y] : N \times n;$$

this has $i^{th}$ row

$$\text{cov} [Y_i, y'] = (\sigma_{i1} 1'_{n_1}, ..., \sigma_{ii} 1'_{n_i}, ..., \sigma_{iN} 1'_{n_N}),$$

with the understanding that if $n_j = 0$ then the $j^{th}$ block is absent.

It is useful to introduce the incidence matrix $E_{N \times n} = (e_{ij})$, with

$$e_{ij} = I \left( \text{the } j^{th} \text{ element of } y \text{ is observed at } x_i \right).$$

An alternate expression is

$$E = \begin{pmatrix} e'_1 \\ \vdots \\ e'_N \end{pmatrix}, \text{ where } e'_i = \begin{cases} 0'_{1 \times n}, & n_i = 0, \\ 0'_{\sum_{j<i} n_j} : 1'_{n_i} : 0'_{\sum_{j>i} n_j}, & n_i > 0. \end{cases}$$

In this notation

$$X = E'F \text{ and } C_{N,n} = C_N E.$$  \hspace{1cm} (5)

The matrix $E$ determines the design: if $\{\xi_i = n_i/n\}$ are the design weights – the proportion of observations made at $x_i$ – then

$$D_\xi \overset{\text{def}}{=} \bigoplus_{i=1}^N \xi_i = n^{-1} EE'.$$

We define the covariance matrix of the data by

$$C_n = \text{cov} [y] : n \times n.$$
This is a block matrix in which most blocks – those corresponding to \( n_i = 0 \) or \( n_j = 0 \) – are typically absent; if \( n_i, n_j > 0 \) the \((i, j)\)th block is

\[
C_{n,ij} = \text{COV}[y_i, y'_j] = \sigma_{ij}^2 1_{n_i} 1'_{n_j} + \begin{cases} 
\sigma^2 \delta_{ii}, & i = j, \\
0, & i \neq j.
\end{cases}
\]  

Thus

\[
C_n = \sigma^2 \left( E' C_N E + I_n \right) .
\]  

(7)

We suppose that the investigator computes estimates and inferences under the assumption that \( \psi \equiv 0 \), so that \( E[Y(x)] = f'(x) \theta \), and the assumption that the true covariance matrix is a particular \( C_{0, N} \) of \( C \).

The investigator seeks a set of linear predictors \( \hat{Y} = L_0 y \) of \( Y = (Y_1, \ldots, Y_N)' \) that are unbiased – \( E[\hat{Y}] = E[Y] \) – and that minimize the prediction mean squared error (pmse), defined as

\[
\text{PMSE} = \sum E \left[ (Y_i - \hat{Y}_i)^2 \right] = E \left[ \| Y - \hat{Y} \|^2 \right].
\]

The solution to this problem is the universal kriging estimate – see, e.g. Cressie (1993) – given in part (i) in the following theorem. Part (ii) gives the pmse under the general mean/covariance structures discussed above; in part (iii) this pmse is expressed explicitly in terms of the design. In Theorem 2 we give the maximum of this pmse, over \( \Psi \) and \( C \).

**Theorem 1** (i) The linear predictors \( \hat{Y} = L_0 y \) minimizing the pmse (under the experimenter’s model assumptions) are given by \( L_0 = L_1 + L_2 \), with

\[
L_1 = F \left( X' C_{0,n}^{-1} X \right)^{-1} X' C_{0,n}^{-1},
\]

\[
L_2 = C_{0,N,n} C_{0,n}^{-1} \left[ I_n - X (X' C_{0,n}^{-1} X)^{-1} X' C_{0,n}^{-1} \right];
\]

thus

\[
\hat{Y} = F \hat{\theta}_0 + C_{0,N,n} C_{0,n}^{-1} \left( y - X \hat{\theta}_0 \right),
\]

where \( \hat{\theta}_0 = (X' C_{0,n}^{-1} X)^{-1} X' C_{0,n}^{-1} y \) is the generalized least squares estimator.

(ii) If \( \psi \) and \( \text{COV}[Y] = C_N \) are arbitrary members of \( \Psi \) and \( C \) respectively, then

\[
\text{PMSE} = \| (I_N - L_0 E') \psi_N \|^2 + tr \left\{ (I_N - L_0 E') C_N (I_N - L_0 E')' \right\} + \sigma^2 \| E L_0 \|_2.
\]  

(8)

Thus, under the model assumptions, the minimized pmse is

\[
\text{PMSE}_0 = tr \left\{ (I_N - L_0 E') C_{0,N} (I_N - E L_0)' \right\} + \sigma^2 \| L_0 L_0' \|.
\]  

\[
\text{PMSE}_0 = tr \left\{ (I_N - L_0 E') C_{0,N} (I_N - E L_0)' \right\} + \sigma^2 \| L_0 L_0' \|.
\]
(iii) Define $V_0 = E \left( C_{0,n}/\sigma_\varepsilon^2 \right)^{-1} E'$; recall (6). Then:

$$V_0 = \sqrt{n} D_{\xi}^{1/2} \left( I_N + n D_{\xi}^{1/2} \frac{C_{0,N}}{\sigma_\varepsilon^2} D_{\xi}^{1/2} \right)^{-1} \sqrt{n} D_{\xi}^{1/2},$$

$$I_N - L_0 E' = \left( I_N + n \frac{C_{0,N}}{\sigma_\varepsilon^2} D_{\xi} \right)^{-1} \left( I_N - F (F' V_0 F)^{-1} F' V_0 \right),$$

$$tr L_0 L_0' = tr \left\{ \left[ \frac{C_{0,N}}{\sigma_\varepsilon^2} + \left( I_N + n \frac{C_{0,N}}{\sigma_\varepsilon^2} D_{\xi} \right)^{-1} F (F' V_0 F)^{-1} F' \right] \cdot \sqrt{n} D_{\xi}^{1/2} \left( I_N + n D_{\xi}^{1/2} \frac{C_{0,N}}{\sigma_\varepsilon^2} D_{\xi}^{1/2} \right)^{-2} \sqrt{n} D_{\xi}^{1/2}, \right\}$$

$$\left[ \frac{C_{0,N}}{\sigma_\varepsilon^2} + \left( I_N + n \frac{C_{0,N}}{\sigma_\varepsilon^2} D_{\xi} \right)^{-1} F (F' V_0 F)^{-1} F' \right]' \right\}.$$ 

Note that

$$(I_N - L_0 E') F = 0_{N \times p}.$$ 

The maximum value of $\text{PMSE}$ at (8), over $\Psi$ and $C$, is developed in the following two lemmas and summarized in Theorem 2. The proof of Lemma 1 is given in the Appendix. Lemma 2 is proved in Welsh and Wiens (2013), but was previously noted in Wiens and Zhou (2008).

**Lemma 1** For an $N \times N$ matrix $A$ satisfying $A F = 0_{N \times p}$, the maximum of $\psi_N A' A \psi_N$ over $\psi_N$, subject to (2) and (3), is

$$\sigma_\varepsilon^2 \alpha_0^2 c_{\text{max}} A A'.$$

**Lemma 2** Suppose that $\mathcal{L}(C)$ is a function of positive semi-definite matrices $C_{N \times N}$ which is monotonic with respect to the ordering by positive semi-definiteness, in that $C_1 \geq C_2 \Rightarrow \mathcal{L}(C_1) \geq \mathcal{L}(C_2)$. For any induced matrix norm $\| \cdot \|$, define a class of matrices

$$C = \left\{ C \mid C \text{ positive semi-definite and } \| C \| \leq \tau^2 \right\},$$

and define the class

$$C' = \left\{ C \mid C \text{ positive semi-definite and } 0 \leq C \leq \tau^2 I_N \right\}.$$ 

Then in all such classes

$$\max_C \mathcal{L}(C) = \max_{C'} \mathcal{L}(C) = \mathcal{L} \left( \tau^2 I_N \right).$$

An implication of Lemma 2 is that the least favourable model of dependence in this problem is in fact independence.

The following is now immediate.
Theorem 2 Recall (10) and (12) and put
\[ A_0 = I_N - L_0 E'. \]

Then the maximum, over \( \Psi \) and \( C \), of the PMSE at (8) is \( \sigma^2 (\alpha_n^2 + \beta_n^2) \) times
\[ L(\xi) = (1 - \nu) ch_{\max} A_0 A_0' + \nu tr A_0 A_0' + \omega_n tr L_0 L_0', \tag{14} \]
where \( \nu = \beta_n^2 / (\alpha_n^2 + \beta_n^2) \) and \( \omega_n = 1 / (\alpha_n^2 + \beta_n^2) \).

Remark 1: We can now specify the dependence of \( \alpha_n, \beta_n \) on \( n \). These are determined by the requirement that all three components of \( L(\xi) \) be of the same asymptotic order. The designer chooses \( \nu \) and \( \omega_n \) according to the emphasis that he/she wishes to place on the various contributors to PMSE. We consider two cases:

Case 1: no replication In this case, as typically occurs in spatial studies, all \( n_i \in \{0, 1\} \) and so \( nD_\xi = O(1) \). We take (\( \alpha_n, \beta_n \)) = (\( \alpha, \beta \), hence \( nD_\xi, V_0, A_0 = I_N - L_0 E' \) and \( tr L_0 L_0 \) are \( O(1) \), as are \( \nu = \beta^2 / (\alpha^2 + \beta^2) \) and \( \omega_n \). Hence all three components of \( L(\xi) \) are \( O(1) \).

Case 2: replication If the setting allows for unrestricted replications then \( n_i = O(n) \), so that \( D_\xi = O(1) \). We take (\( \alpha_n, \beta_n \)) = (\( \alpha/\sqrt{n}, \beta/\sqrt{n} \)), so that again \( nC_{0,N}/\sigma^2 \) is \( O(1) \). Then \( V_0 = O(n) \), \( A_0 = I_N - L_0 E' \) is \( O(1) \) and \( tr L_0 L_0 \) is \( O(1/n) \). Since \( \nu = \beta^2 / (\alpha^2 + \beta^2) = O(1) \) and \( \omega_n = O(n) \), it is again the case that all three components of \( L(\xi) \) are \( O(1) \).

Remark 2: An alternate class of covariance structures is
\[ C_\gamma = \{(1 - \gamma) C_{N,0} + \gamma C_N \mid \| C_N \|_M \leq \sigma^2 (\alpha_n^2 + \beta_n^2) \}, \]
for \( \gamma \in [0, 1] \). For this class the theory above continues to hold, and yields that
\[ \max_{\Psi, C_\gamma} \text{PMSE} = \text{PMSE}_0 + \sigma^2 \left[ \alpha_n^2 ch_{\max} A_0 A_0' + \gamma tr A_0 \left( \beta_n^2 I_N - \frac{C_{0,N}}{\sigma^2} \right) A_0' \right]. \]

Remark 3: We might restrict to correlation structures which are isotropic, i.e. structures for which \( \text{CORR}[Y(x_1), Y(x_2)] \) depends only on the distance \( \| x_1 - x_2 \| \). In this case the theory above continues to hold, since the maximizing structure \( \tau^2 I_N \) in Lemma 2 is a scale multiple of an isotropic correlation matrix.

In the next section we discuss designs minimizing \( L(\xi) \) given by (14).
3 Computations

Note: The computations involve

\[ G_1 = \left( I_N + nD^{1/2}_\xi \frac{C_{0,N}}{\sigma^2_\xi} D^{1/2}_\xi \right)^{-1} \overset{\text{def}}{=} (I_N + S_1)^{-1}, \]
\[ G_2 = \left( I_N + n \frac{C_{0,N}}{\sigma^2_\xi} D_\xi \right)^{-1} \overset{\text{def}}{=} (I_N + S_2)^{-1}. \]

Typically \( D_\xi \) is quite sparse; denote by ‘pos’ the locations of the \( N_+ \) non-zero elements (\( N_+ \leq \min(N, n) \)). Let \( J : N \times N_+ \) consist of those columns of \( I_N \) in ‘pos’. Let

\[ S_{1+} = S_1 [\text{pos}, \text{pos}] : N_+ \times N_+ \]

consist only of these rows and columns, and let

\[ S_{2+} = S_2 [:, \text{pos}] : N \times N_+ \]

consist only of these columns. Then \( J'J = I_{N+} \) and

\[ S_1 = JS_{1+}J', \quad S_2 = S_{2+}J'; \]

thus

\[ G_1 = (I_N + JS_{1+}J')^{-1} = I_N - J (I_{N_+} + S_{1+})^{-1} S_{1+}J', \]
\[ G_2 = (I_N + S_{2+}J')^{-1} = I_N - S_{2+} (I_{N_+} + J'S_{2+})^{-1} J'. \]

These involve only the inversion of matrices of order \( N_+ \); this reduces the computing time substantially.

Appendix: Derivations

Proof of Theorem 1. (i) For notational convenience we temporarily drop the subscript ‘0’. The requirement of unbiasedness is \( E \left[ \hat{Y} \right] = L \theta = F \theta = E \left[ Y \right] \) for all \( \theta \); this entails \( LX = F \) and so \( L = L_1 + L_2 \), where \( L_1 = F \left( X'C_n^{-1}X \right)^{-1} X'C_n^{-1} \) and

\[ L_2X = 0. \]  \hspace{1cm} (A.1)

Thus

\[ \hat{Y} = F\hat{\theta}_{GLS} + L_2y, \]

where \( \hat{\theta}_{GLS} = \left( X'C_n^{-1}X \right)^{-1} X'C_n^{-1}y \) is unbiased for \( \theta \) and \( E[L_2y] = 0. \)
The PMSE to be minimized is

\[
\text{PMSE} = tr \left\{ \text{COV} \left[ Y - \hat{Y} \right] \right\} \\
= tr \text{COV} \left[ Y \right] + tr \text{COV} \left[ \hat{Y} \right] - 2tr \text{COV} \left[ Y, \hat{Y}' \right] \\
= tr C_N + tr L C_n L' - 2tr \text{COV} \left[ Y, y' L' \right] \\
= tr C_N + tr F \left( X' C_n^{-1} X \right)^{-1} F' - 2tr C_{N,n} L_1 + \left[ tr L_2 C_n L_2' - 2tr C_{N,n} L_2' \right];
\]

here we use that \( L_1 C_n L_1' = L_2 C_n L_1' = 0_{N \times N} \) and that \( L_1 C_n L_1' = F \left( X' C_n^{-1} X \right)^{-1} F' \).

We are now to minimize \( tr L_2 C_n L_2' - 2tr C_{N,n} L_2' \) subject to (A.1). This orthogonality condition, which we now write as \( L_2 C_n^{1/2} C_n^{-1/2} X = 0 \), states that the rows of \( L_2 C_n^{1/2} \) lie in the row space of the orthogonal projector \( I_n - H \), where \( H = C_n^{-1/2} X \left( X' C_n^{-1} X \right)^{-1} X' C_n^{-1/2} \), so that if the rows of \( \Pi : (n - p) \times n \) form an orthogonal basis for this space (so that \( \Pi \Pi' = I_{n - p} \) and \( \Pi' \Pi = I_n - H \)), we have that, for some \( M : N \times (n - p) \), \( L_2 C_n^{1/2} = M \Pi \).

Thus we minimize

\[
tr L_2 C_n L_2' - 2tr C_{N,n} L_2' \\
= tr \left[ MM' - C_{N,n} C_n^{-1/2} \Pi' M' - M \Pi C_n^{-1/2} C_n', n \right] \\
= tr \left[ \left( M - C_{N,n} C_n^{-1/2} \Pi \right) \left( M' - \Pi' C_n^{-1/2} C_n' \right) - C_{N,n} C_n^{-1/2} \Pi' \Pi C_n^{-1/2} C_n' \right]
\]

over \( A \), unconditionally. The solution is clearly \( M = C_{N,n} C_n^{-1/2} \Pi' \), whence

\[
L_2 = C_{N,n} C_n^{-1} \left[ I_n - X \left( X' C_n^{-1} X \right)^{-1} X' C_n^{-1} \right].
\]

(ii) Put \( \psi_n = (1_n', \psi(x_1), \ldots, 1_n' \psi(x_N))' \), so that \( E [Y] = F \theta + \psi_n \),
\( E [y] = X \theta + \psi_n \). Define \( m = E \left[ Y - \hat{Y} \right] \), then PMSE = \( \| m \|^2 + tr \text{COV} \left[ Y - \hat{Y} \right] \).

Since \( \psi_n = E' \psi_N \) and \( L_0 X = F \), we have that

\[
m = E \left[ Y - L_0 y \right] = (I_N - L_0 E') \psi_N.
\]

Furthermore, and using (5) and (7),

\[
tr \text{COV} \left[ Y - \hat{Y} \right] = tr \text{COV} \left[ Y - L_0 y \right] \\
= tr \left\{ C_N - C_{N,n} L_0 - L_0 C_{N,n} + L_0 C_n L_0' \right\} \\
= tr \left\{ \left( I_N : - L_0 \right) \left[ \left( \begin{array}{c} I_N \\ E' \end{array} \right) C_N \left( \begin{array}{c} I_N : E \end{array} \right) \right] \left( \begin{array}{c} I_N \\ - L_0 \end{array} \right) \right\} \\
= tr \left\{ (I_N - L_0 E') C_N (I_N - E L_0') \right\} + \sigma^2 tr L_0 L_0'.
\]
(iii) We repeatedly use standard matrix identities for the inversion of matrices of the form $I - AB$. These identities together with

$$V_0 = E \left( E' \frac{C_{0;N}}{\sigma^2_\xi} E + I_N \right)^{-1} E'$$

and (6) yield (9).

To verify (10) first recall (5), which implies that $X'C_{0;n}^{-1}X = F'V_0F/\sigma^2_\xi$; apply (7) as well to get

$$L_1E' = F (F'V_0F)^{-1} F'V_0,$$
$$L_2E' = \frac{C_{0;N}}{\sigma^2_\xi} V_0 - \frac{C_{0;N}}{\sigma^2_\xi} V_0F (F'V_0F)^{-1} F'V_0,$$

so that, with $L_0 = L_1 + L_2$,

$$I_N - L_0E' = \left[ I_N - \frac{C_{0;N}}{\sigma^2_\xi} V_0 \right] \left[ I_N - F (F'V_0F)^{-1} F'V_0 \right].$$

Now (10) follows from

$$I_N - \frac{C_{0;N}}{\sigma^2_\xi} V_0 = \left( I_N + n \frac{C_{0;N}}{\sigma^2_\xi} D_\xi \right)^{-1}. \quad (A.2)$$

To obtain (11), first use (7) to verify that

$$E \left( \frac{C_{0;n}}{\sigma^2_\xi} \right)^{-1} = \left( I_N + nD_\xi \frac{C_{0;N}}{\sigma^2_\xi} \right)^{-1} E;$$

from this it follows that

$$L_1 = \left[ F (F'V_0F)^{-1} F' \right] \left( I_N + nD_\xi \frac{C_{0;N}}{\sigma^2_\xi} \right)^{-1} E,$$
$$L_2 = \frac{C_{0;N}}{\sigma^2_\xi} \left[ I_N - V_0F (F'V_0F)^{-1} F' \right] \left( I_N + nD_\xi \frac{C_{0;N}}{\sigma^2_\xi} \right)^{-1} E,$$

whence

$$L_0 = \left[ \frac{C_{0;N}}{\sigma^2_\xi} + \left( I_N + n \frac{C_{0;N}}{\sigma^2_\xi} D_\xi \right)^{-1} F (F'V_0F)^{-1} F \right] \left( I_N + nD_\xi \frac{C_{0;N}}{\sigma^2_\xi} \right)^{-1} E;$$

here we have used (A.2). Then

$$trL_0L_0' = tr \left\{ \left[ \frac{C_{0;N}}{\sigma^2_\xi} + \left( I_N + n \frac{C_{0;N}}{\sigma^2_\xi} D_\xi \right)^{-1} F (F'V_0F)^{-1} F \right] \left( I_N + nD_\xi \frac{C_{0;N}}{\sigma^2_\xi} \right)^{-1} \right\},$$

$$\cdot \left[ \frac{C_{0;N}}{\sigma^2_\xi} + \left( I_N + n \frac{C_{0;N}}{\sigma^2_\xi} D_\xi \right)^{-1} F (F'V_0F)^{-1} F \right]'.$$
and (11) follows from

\[
\left( I_N + n D_\xi \frac{C_{0:N}}{\sigma_\xi^2} \right)^{-1} n D_\xi \left( I_N + n \frac{C_{0:N}}{\sigma_\xi^2} D_\xi \right)^{-1} = \sqrt{n} D_\xi^{1/2} \left( I_N + n D_\xi^{1/2} \frac{C_{0:N}}{\sigma_\xi^2} D_\xi^{1/2} \right)^{-2} \sqrt{n} D_\xi^{1/2}.
\]

\[\square\]

**Proof of Lemma 1:** We first classify the solutions to (2). Let \( F = Q_1 R \) be the qr-decomposition of \( F \), so that \( Q_1 : N \times p \) satisfies \( Q_1^t Q_1 = I_p \) and \( R : p \times p \) is upper triangular and non-singular. Augment \( Q_1 \) by \( Q_2 : N \times (N - p) \) in such a way that \( Q = (Q_1; Q_2) \) is an orthogonal matrix. Then the columns of \( Q_2 \) form an orthogonal basis for the orthogonal complement of the column space of \( F \), to which \( N \) belongs by virtue of (2). Thus \( \psi_N = Q_2 c \) for some \( c \in \mathbb{R}^{N-p} \), and we maximize \( \psi_N^t A^t A \psi_N = c^t Q_2^t A^t A Q_2 c \), subject to \( \|\psi_N\| = \|c\| \leq \sigma_\xi \alpha_n \). The maximizing \( c \) is \( \sigma_\xi \alpha_n \) times the unit eigenvector of \( Q_2^t A^t A Q_2 \) corresponding to the maximum eigenvalue \( \chi_{\text{max}} Q_2^t A^t A Q_2 \), and then

\[
\max \psi_N^t A^t A \psi_N = \sigma_\xi^2 \alpha_n^2 \chi_{\text{max}} Q_2^t A^t A Q_2.
\]

Now (13) follows from

\[
\chi_{\text{max}} Q_2^t A^t A Q_2 = \chi_{\text{max}} A Q_2^t A' = \chi_{\text{max}} A (I_N - Q_1 Q_1^t) A' = \chi_{\text{max}} A \left( I_N - F (F' F)^{-1} F' \right) A' = \chi_{\text{max}} A A',
\]

\[\square\]

**Notes to myself:**

1. Put \( A_0 = C_{0:N,n} C_{0:n}^{-1} \) and \( B_0 = (F - C_{0:N,n} C_{0:n}^{-1} X) (X' C_{0:n}^{-1} X)^{-1} X' C_{0:n}^{-1} \), note that \( B_0 X (X' C_{0:n}^{-1} X)^{-1} X' C_{0:n}^{-1} = B_0 \). In these terms we have that \( L_0 = A_0 + B_0 \) and the error vector decomposes into uncorrelated components

\[
Y - \hat{Y} = (Y - A_0 y) - B_0 y.
\]

**Acknowledgements**

This work was carried out with the support of the Natural Sciences and Engineering Research Council of Canada.
References


