

Quadratic Minimisation Problems in Statistics

C.J. Albers^{1,2}, F. Critchley¹, J.C. Gower¹

¹⁾ Department of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, U.K.

²⁾ Corresponding author. c.j.albers@open.ac.uk, tel. (+44)1908652680, fax (+44)1908655515

Abstract. We consider the problem $\min_{\mathbf{x}} (\mathbf{x} - \mathbf{t})' \mathbf{A} (\mathbf{x} - \mathbf{t})$ subject to $\mathbf{x}' \mathbf{B} \mathbf{x} + 2\mathbf{b}' \mathbf{x} = k$ where \mathbf{A} is positive definite or positive semi-definite. Commonly occurring statistical variants of this problem are discussed within the framework of a general unifying methodology. These include non-trivial considerations that arise when (i) \mathbf{A} and/or \mathbf{B} are not of full rank and (ii) \mathbf{t} takes special forms (especially $\mathbf{t} = 0$ which, under further conditions, reduces to the well-known two-sided eigenvalue solution). Special emphasis is placed on insights provided by geometrical interpretations. Algorithmic considerations are discussed and examples given.

Keywords. canonical analysis, constraints, geometry, Hardy-Weinberg, minimisation, optimal scaling, Procrustes analysis, quadratic forms, ratios, reduced rank, splines.

1 Introduction

Our main objectives in writing this paper are: (1) to bring to a statistical readership what may be some unfamiliar results to be found in a dispersed literature, (2) to unify this area, so leading to a general purpose algorithm of wide applicability, and (3) to provide some new results of our own.

The problem

$$\left. \begin{aligned} & \min_{\mathbf{x}} (\mathbf{x} - \mathbf{t})' \mathbf{A} (\mathbf{x} - \mathbf{t}) \\ & \text{subject to } \mathbf{x}' \mathbf{B} \mathbf{x} + 2\mathbf{b}' \mathbf{x} = k \end{aligned} \right\} \quad (1)$$

is common in statistics. Examples covered by this formulation – some of which we develop in Section 3 – arise, for example, in (i) canonical variate analysis with fewer samples than variables, (ii) normal linear models with quadratic and/or linear constraints, (iii) Hardy-Weinberg estimation, (iv) spline fitting, (v) various forms of oblique Procrustes analysis, (vi) Fisher/Guttmann estimation of optimal scores, (vii) handling size and location constraints in iterative missing value procedures in Procrustes analysis, (viii) Bayesian decision theory and (ix) minimum distance estimation.

We term the first part of (1) the *objective function* and the second part *the constraint*. It might arise as a problem in itself or as part of a bigger problem as, for example, when iteratively minimising a general convex objective function via a series of local quadratic approximations (see Sections 3.2 and 3.4.2). Again, (1) may arise either as a direct optimisation problem with strong constraints, or as a Lagrangian form with weak constraints derived from optimising a ratio (see Gower, 1998 and Section 3.7 for a discussion). In the constraint, equality can be relaxed to inequality (see *Exact solutions*, Section 2), while additional linear equality or inequality constraints are easily subsumed. To provide ad hoc solutions

to each of a large number of special cases is inefficient. Thus, a solution to (1) backed up by a reliable accessible algorithm would provide a useful supplement to the statistician's toolkit of existing linear algebra algorithms.

Without loss of generality, \mathbf{A} and \mathbf{B} of order p , may be assumed to be symmetric. The linearly constrained case being well-known, there is no real loss in also assuming that \mathbf{B} is non-zero. The only restriction on the forms we allow is that \mathbf{A} must be positive definite (p.d.) or positive semi-definite (p.s.d.). This condition, which renders the objective function convex, is shown to have good statistical justification. \mathbf{B} is unrestricted while, often, \mathbf{b} is zero, but without loss we may assume that \mathbf{B} is not negative definite or negative semi-definite, if necessary by changing the signs of \mathbf{B} , \mathbf{b} , and k . The methodology developed below is for general k but when $k = 0$ special considerations may apply, as discussed in the algorithm section of Appendix B.

The minimisation problem (1) may be reparameterised in several ways. For example, it is trivial that (1) includes

$$\min_{\mathbf{x}} (\mathbf{x} - \mathbf{t})' \mathbf{A} (\mathbf{x} - \mathbf{t}) \text{ subject to } (\mathbf{x} - \mathbf{s})' \mathbf{B} (\mathbf{x} - \mathbf{s}) + 2\mathbf{b}' (\mathbf{x} - \mathbf{s}) = k$$

merely by setting $\mathbf{x}^* = \mathbf{x} - \mathbf{s}$ and $\mathbf{t}^* = \mathbf{t} - \mathbf{s}$. Less obvious is that the quadratically constrained full-rank regression problem below is also included:

$$\min_{\mathbf{x}} \|\mathbf{X}\mathbf{x} - \mathbf{y}\|^2 \text{ subject to } \mathbf{x}' \mathbf{B} \mathbf{x} + 2\mathbf{b}' \mathbf{x} = k$$

but on expanding $\|\mathbf{X}\mathbf{x} - \mathbf{y}\|^2 = \mathbf{x}' \mathbf{X}' \mathbf{X} \mathbf{x} - 2\mathbf{y}' \mathbf{X} \mathbf{x} + \mathbf{y}' \mathbf{y}$ and then replacing $\mathbf{X}' \mathbf{X}$ by any decomposition $\mathbf{L}' \mathbf{L}$ where \mathbf{L} is non-singular (e.g. use the Choleski decomposition), we may write:

$$\|\mathbf{X}\mathbf{x} - \mathbf{y}\|^2 = \left(\mathbf{L}\mathbf{x} - (\mathbf{L}^{-1})' \mathbf{X}' \mathbf{y} \right)' \left(\mathbf{L}\mathbf{x} - (\mathbf{L}^{-1})' \mathbf{X}' \mathbf{y} \right) + \text{constant}$$

which, on defining $\mathbf{x}^* = \mathbf{L}\mathbf{x}$, $\mathbf{t} = (\mathbf{L}^{-1})' \mathbf{X}' \mathbf{y}$ and $\mathbf{A} = \mathbf{I}$, gives the basic form of the objective function of (1), parallel changes being made to the constraint.

We may also consider a multidimensional form of (1):

$$\left. \begin{array}{l} \min_{\mathbf{X}} \text{trace } (\mathbf{X} - \mathbf{T})' \mathbf{A} (\mathbf{X} - \mathbf{T}) \\ \text{subject to } \text{trace } (\mathbf{X}' \mathbf{B} \mathbf{X} + 2\mathbf{G}' \mathbf{X}) = k \end{array} \right\}$$

where \mathbf{X} , \mathbf{T} and \mathbf{G} have dimensions $p \times n$. By writing $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ and $\mathbf{x}' = (\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_n)$ and similarly for \mathbf{T} and \mathbf{G} , and defining a block-diagonal matrix $\mathbf{A}^* = \text{diag}(\mathbf{A}, \mathbf{A}, \dots, \mathbf{A})$ with \mathbf{A} repeated n times, and similarly for \mathbf{B}^* , the problem becomes:

$$\left. \begin{array}{l} \min_{\mathbf{x}} (\mathbf{x} - \mathbf{t})' \mathbf{A}^* (\mathbf{x} - \mathbf{t}) \\ \text{subject to } \mathbf{x}' \mathbf{B}^* \mathbf{x} + 2\mathbf{g}' \mathbf{x} = k \end{array} \right\}$$

and so is subsumed in (1).

Several special cases of (1) merit consideration: (i) $\mathbf{t} = 0$, (ii) $\mathbf{b} = 0$ and, more especially, (iii) the intersection of (i), (ii) and $k = 1$. Case (iii) is related to optimising a ratio of quadratic forms:

$$\frac{\mathbf{x}' \mathbf{A} \mathbf{x}}{\mathbf{x}' \mathbf{B} \mathbf{x}}$$

as in canonical analysis, leading to the well-known two-sided eigenvalue solution, at least when one of the matrices is of full rank, but also subsuming the case where both matrices are indefinite: see Sections 3.3 and 3.7.

Thus, not only does (1) include a very wide class of statistical problems, but they may manifest themselves in various equivalent forms. Some classical problems involve the simultaneous diagonalisation of \mathbf{A} and \mathbf{B} , at least when \mathbf{A} is of full rank. Perhaps for this reason, much of the mathematical literature is concerned with conditions under which simultaneous diagonalisation is possible, but this is not always possible (see Sections 3.3 and 3.7). It is useful to distinguish between purely algebraic problems, such as simultaneously diagonalising two matrices, and their subsequent use in optimisation problems. In the latter case, we have to further distinguish between \mathbf{t} being zero or non-zero in (1).

In this paper, Appendix A develops a potentially non-diagonal canonical form for \mathbf{A} and \mathbf{B} , also possibly including linear terms, which we term the *General Canonical Form* (GCF) and examines some important special cases of the GCF. It turns out that minimisation of all these special cases requires the solution to an analogue of a characteristic equation, the *Fundamental Canonical Equation* (FCE), discussed in Appendix B.

The solution to (1) is greatly simplified by recasting the problem in terms of the GCF. This depends on a simple affine transformation:

$$\mathbf{z} = \mathbf{T}^{-1}\mathbf{x} + \mathbf{m} \quad (2)$$

where the explicit forms of the non-singular matrix \mathbf{T} or (A.4), and the translation \mathbf{m} are given in Appendix A. We use the inverse \mathbf{T}^{-1} here to simplify the appearance of \mathbf{T} in Appendix A. Underlying the GCF is the interplay of the range and null spaces of \mathbf{A} and \mathbf{B} . The vector \mathbf{z} is decomposed into components $\mathbf{z} = (\mathbf{z}'_{11}, \mathbf{z}'_{10}, \mathbf{z}'_{01}, \mathbf{z}'_{00})$ where the first suffix position refers to \mathbf{A} and the second to \mathbf{B} ; a suffix 1 denotes the range space and 0 the null space (but see the comments after (A.1) of Appendix A). We shall term the variables \mathbf{z}_{01} and \mathbf{z}_{00} that occur only in the constraint, *extraneous variables*. With this notation, (1) simplifies to its GCF form:

$$\left. \begin{aligned} & \min_{\mathbf{z}} \left(\|\mathbf{z}_{11} - \mathbf{s}_{11}\|^2 + \|\mathbf{z}_{10} - \mathbf{s}_{10}\|^2 \right) \\ & \text{subject to:} \\ & \left(\begin{array}{c} \mathbf{z}_{11} \\ \mathbf{z}_{10} \\ \mathbf{z}_{01} \\ \mathbf{z}_{00} \end{array} \right)' \left(\begin{array}{cc} \boldsymbol{\Gamma}_1 & \mathbf{D}_{10} \\ \mathbf{0} & \mathbf{D}_{00} \\ & \boldsymbol{\Gamma}_0 \\ \mathbf{D}'_{10} & \mathbf{D}'_{00} \end{array} \right) \left(\begin{array}{c} \mathbf{z}_{11} \\ \mathbf{z}_{10} \\ \mathbf{z}_{01} \\ \mathbf{z}_{00} \end{array} \right) + 2 \left(\begin{array}{c} \mathbf{g}_{10} \\ \mathbf{z}_{10} \\ \mathbf{g}_{00} \\ \mathbf{z}_{00} \end{array} \right)' \left(\begin{array}{c} \mathbf{z}_{10} \\ \mathbf{z}_{00} \end{array} \right) = k \end{aligned} \right\} \quad (3)$$

where $\mathbf{s} = \mathbf{T}^{-1}\mathbf{t} + \mathbf{m}$. $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_0$ are diagonal matrices in the range space of \mathbf{B} associated with the range and null spaces, respectively, of \mathbf{A} . All terms in (3) are linear functions of the parameters of (1)

and are defined in Appendix A. Thus to minimise (1) it is sufficient to minimise (3) and substitute into (2) to derive the optimal setting of \mathbf{x} . For most problems encountered in practice, the full form of (3) simplifies, as shown under ‘the main special cases’ in Appendix A. Exact solutions, $\mathbf{z}_{11} = \mathbf{s}_{11}$, $\mathbf{z}_{10} = \mathbf{s}_{10}$ may exist, and when they do, this suggests that too loose constraints are being applied. Appendix A addresses these issues in the context of the main special cases that occur in practice.

The FCE has a unique solution lying in a simply defined feasible region or, in some special cases, at one of the extremes of this feasible region. Appendix B outlines how these results may be used to develop a general algorithm. Geometrical insights discussed in Section 2 illuminate and motivate the detailed algebraic discussion of Appendix A. Section 3 discusses specific statistical examples using the preceding methodology. Section 4 gives a discussion of related problems and extensions, some of which require further work. Proofs of some of the more technical details are given in a companion paper (Albers *et al.*, 2007).

2 Geometry

Geometrical considerations give insight into difficulties that may be encountered when minimising (3). The generality of the criterion makes it quite difficult to illustrate the full geometry in the two dimensions of a sheet of paper. We have to cope with the possibilities that \mathbf{A} and \mathbf{B} may be of different ranks and have different null spaces, requiring a minimum of four dimensions, as implied by the parameters \mathbf{z}_{11} , \mathbf{z}_{10} , \mathbf{z}_{01} , \mathbf{z}_{00} of the GCF(s). Further we need to consider the effects of including the linear terms in the constraint. Thus, we can never visually represent the full generality. Nevertheless, much can be done with two dimensions and attempts at representing three dimensions in two.

In this section we refer to the geometrical object representing the

constraint by B , reserving \mathbf{B} for its matrix form. If \mathbf{B} is p.s.d., the quadratic surface is ellipsoidal but in general B will have elliptic, hyperbolic or, when there are linear terms in the constraint, parabolic cross-sections. Because the left-hand-side of (3) is a simple squared Euclidean distance, the basic criterion is equivalent to finding the foot, S , of the shortest normal from \mathbf{s} to the surface of B , thus putting (1) firmly into the class of constrained least-squares problems; any linear terms make no material difference to this interpretation. Without the condition that \mathbf{A} has no negative eigenvalues, $(\mathbf{z}_1 - \mathbf{s}_1)'(\mathbf{z}_1 - \mathbf{s}_1)$ would have to be replaced by a hyperbolic distance and the least-squares rationale sacrificed.

The simplest case

The simplest case is when \mathbf{A} and \mathbf{B} share the same dimensions, in which case \mathbf{B} has no extraneous variables and only $\boldsymbol{\Gamma}_1$ is non-null in the GCF. Note that this diagonality implies that B is referred to its principal axes. This is illustrated in Figure C.1 where B is an ellipse and \mathbf{s} may be any point in the same plane. The shortest normals for two settings of \mathbf{s} are shown in Figure C.1a. The hyperbolic case is shown in Figure C.1b and introduces no additional problems.

There may be more than one normal, but we require only the shortest and this is usually unique. The other normals are of interest in understanding the FCE and its solution as discussed in Appendix B. Generally, there is a unique minimum but when \mathbf{s}_k lies on a principal axis of B there are two solutions.

The star shape (the evolute of the ellipse) shown in Figure C.1 separates the regions inside B for which there are two and four real normals. For any point on the minor axis, the foot of the shortest normal is always at one end of this axis. On the major axis there are two regions: (i) between the origin and the cusp there are varying normals to B and (ii) at the cusp there are three coincident normals at one end of the major axis and another

normal at the other end; beyond the cusp there are only two normals (either end of the axis). These geometrical properties are associated with the “phantom asymptote” effect appearing in the solution to the FCE, discussed in Appendix B. In the special case where \mathbf{s} is at the origin, we have conventional eigenvalue problems, and when B is circular, any of the infinitely many points on B is a solution.

Exact solutions and inequality constraints

To understand the geometry of exact solutions it is important to distinguish points that are inside from those that are outside B . We focus on elliptical B .

One way exact solutions can occur is when the equality constraint in (1) is replaced by inequality, such as $\mathbf{x}'\mathbf{B}\mathbf{x} + 2\mathbf{b}'\mathbf{x} \leq k$. Then, if \mathbf{s} is inside B the constraint is satisfied by the exact solution $\mathbf{x} = \mathbf{s}$ (see E in Figure C.2). If \mathbf{s} is outside B , we require the nearest normal, as before, in which case the solution is inexact but the constraint is satisfied with equality. Similar remarks apply when $\mathbf{x}'\mathbf{B}\mathbf{x} + 2\mathbf{b}'\mathbf{x} \geq k$. It follows that inequality constraints can be handled within the same framework as equality constraints.

In the second part of Appendix A we discuss several other ways in which (1) may have exact solutions. In particular, we shall be interested in when \mathbf{s} is constrained to lie in a subspace of B and is inside B . We call this subspace the *target space*. In Figure C.2, B is represented as an ellipse, with its major axis representing a target space, to which \mathbf{s} is confined; that is, \mathbf{A} has rank one. The apparently small change to the geometry has a fundamental affect on the minimisation problem, which becomes:

$$\min(z_1 - s_1)^2 \text{ subject to } \gamma_1 z_1^2 + \gamma_2 z_2^2 = c^2.$$

Thus, the constraint now includes an extraneous variable (z_2) that is not part of the objective function. Indeed, $\gamma_1 \in \Gamma_1$ and $\gamma_2 \in \Gamma_2$ so now the GCF includes both diagonal matrices. Clearly, we may set $z_1 = s_1$ provided $\gamma_2 z_2^2 = c^2 - \gamma_2 s_1^2$ has a real solution for z_2 ; this is the condition that $\mathbf{s} = s_1$ lies inside B .

With \mathbf{s}_1 as shown inside B , S_1 is a point on B that projects to a point that fits \mathbf{s} exactly in one dimension. In this way, $\mathbf{z} = \mathbf{s}_1$. There is a second equally valid choice S_1^* at the other side of the major axis but it corresponds to the same $\mathbf{z}_1 = \mathbf{s}_1$ in the target space; z_2 is relevant only in ensuring that the constraint is satisfied. If \mathbf{s} were not confined to the subspace, the shortest normal would be as indicated by the dotted line. When $\mathbf{z} = \mathbf{s}_1^*$ is outside B but in the target space, there is a unique solution at S_2 at an end of the major axis.

When B is three-dimensional, but \mathbf{A} remains one-dimensional, then the set of points S that project into a given interior point \mathbf{s} form an elliptical cross-section of the ellipsoid B , providing an infinity of exact solutions. If \mathbf{s} were outside B , but remaining in the subspace, exact solutions would not exist and the approximate solution S would continue to be at one end or the other of the major axis as shown by \mathbf{s}_1 and S_1^* in Figure C.2. When B is three-dimensional and \mathbf{s}_2 is confined to a target space of two of these dimensions, exact solutions remain available when \mathbf{s} is inside B . When \mathbf{s}_2 is outside B we require the shortest normal S_2 from \mathbf{s}_2 , to the projection of B onto the target subspace as shown in Figure C.3. This solution is not exact and, indeed, coincides with that given by the FCE, replacing B by its projection onto the target space (Appendix C).

Essentially exact solutions

Consider now $\min((z_1 - s_1)^2 + (z_2 - s_2)^2)$ subject to $z_1 z_2 = c^2$, a rectangular hyperbola. This is shown in Figure C.4 but not referred to principal axes, so is not in its GCF. The GCF may be obtained by setting $\eta_1 = z_1 + z_2$ and $\eta_2 = z_1 - z_2$, so transforming the problem into $\min((\eta_1 - \frac{1}{2}(s_1 + s_2))^2 + (\eta_2 - \frac{1}{2}(s_1 - s_2))^2)$ subject to $\eta_1^2 - \eta_2^2 = 4c^2$. This is now in the GCF with the same parameters in the constraint as in the objective function so that only Γ_1 is non-null. It raises no problems and has well-defined shortest normals as indicated in Figure C.4 for various settings s_1 . Note that the two-dimensional solution for the origin is well-defined, up to a reflection. However, when the term $(z_2 - s_2)^2$ is excluded from

the objective function, so that we seek a one-dimensional target solution, $\min((z_1 - s_1)^2)$ subject to $z_1 z_2 = c^2$, the specification is already in GCF but the diagonal matrices Γ_1 and Γ_0 both vanish and are replaced by a cross-product term corresponding to \mathbf{D}_{11} . This is a simple instance of where \mathbf{A} and \mathbf{B} are not simultaneously diagonable. The parameter z_2 of the constraint is not in the target space, so the usual exact solutions are available (e.g. s_2 in Figure C.4) with the exception that for $\mathbf{s} = 0$, the origin, S is only defined asymptotically (an essentially exact solution; see Albers *et al.*, 2007). Thus, what may seem a trivial difference between two simple minimisation problems, can have a profound effect on the geometry, and hence the algebraic structure, of its solution.

Albers *et al.* (2007) give necessary and sufficient conditions for essentially exact solutions to occur. In particular, they show that they can occur only when \mathbf{B} is indefinite and \mathbf{D}_{11} and \mathbf{D}_{10} occur. In practice, such pathological solutions are probably mainly of interest in indicating that one may be attempting to fit an inappropriate model. Algorithms should trap and report on these situations.

Including linear terms

Next we consider a simple example that includes a linear term: $\min((z_1 - s_1)^2 + (z_2 - s_2)^2)$ with the constraint $z_2^2 = 4cz_1$. B is a parabola as shown in Figure C.5. We have not included a constant term but if we had this would merely shift the vertex of the parabola away from the origin O. We see that without extraneous variables the shortest normal is well defined (\mathbf{s}_1). If we require $\min(z_1 - s_1)^2$ subject to $z_2^2 = 4cz_1$, so we have a one-dimensional target space and z_2 is extraneous, we have exact solutions when s_1 is non-negative, as shown for S_2 , while if s_1 is negative we have the shortest normal at O ($z_1 = z_2 = 0$). The linear term has not raised any new problems and, as shown in Appendix B, has little effect on algorithms.

General projection of a quadratic onto a subspace

The matrix $\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$ occurs in equation (A.2) of Ap-

pendix A as a step in developing the GCF, and is familiar in various forms of canonical analysis. Its geometrical interpretation is that the quadratic form

$$P : \mathbf{x}'_1 (\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}) \mathbf{x}_1 = 1$$

is the projection of the quadratic form

$$Q : \begin{pmatrix} \mathbf{x}'_1 & \mathbf{x}'_2 \end{pmatrix} \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = 1$$

onto the subspace spanned by \mathbf{x}_1 ; a proof is given in appendix C. Figure C.6a shows this for Q in three dimensions and P in two dimensions. Note that P contains, and indeed is tangential to, the intersection $\mathbf{x}'_1 \mathbf{C}_{11} \mathbf{x}_1 = 1$ of Q with the subspace. This implies that when \mathbf{s} lies in the same subspace as P and \mathbf{x}_2 corresponds to extraneous variables, then exact solutions to (1) arise not only from those \mathbf{s} that are inside the intersection of Q with the subspace but also from those that lie between P and the projection of Q , as is shown in Figure C.6. When Q is p.s.d. we show in Appendix A that only diagonal values appear in the GCF and then \mathbf{C}_{12} vanishes. Then, Q is normal to the subspace and the projection and intersection coincide as in Figure C.6b. Thus, the full generality occurs only when there are extraneous variables and \mathbf{D}_{10} and \mathbf{D}_{00} do not vanish, in which case B is not p.s.d. and must have hyperbolic cross-sections. We cannot show this in three-dimensional form but Figure C.7 gives two-dimensional contour plots. Figure C.7a is the contour version of Figure C.6a, while Figure C.7b shows contours when Q is hyperbolic. As with the rectangular hyperbola of Figure C.4, exact solutions will exist everywhere in the two dimensional subspace, as accords with Case 3 in Appendix A.

Summary

If we denote the space occupied by \mathbf{s} as A and the constraint space by B , now using B_Q for the quadratic surface itself, the above geometry may be summarised as follows. Recalling that a

proper subspace is one of strictly lower dimension, we distinguish the following four mutually exclusive and exhaustive cases:

- (1) When A and B are the same space, then the solution \mathbf{z} is given by foot of the shortest normal from \mathbf{s} to B_Q .
- (2) When B is a proper subspace of A , then project \mathbf{s} onto B to obtain $\mathbf{s} = \mathbf{s}_B + \mathbf{s}_{B^\perp}$, where \mathbf{s}_{B^\perp} is the part of \mathbf{s} that is orthogonal to B . Then \mathbf{s}_B is in B and we may proceed as in 1 to obtain \mathbf{z}_B and, finally, $\mathbf{z} = \mathbf{z}_B + \mathbf{s}_{B^\perp}$.
- (3) When A is a proper subspace of B , then (a) if \mathbf{s} is outside B_Q then \mathbf{z} is the foot of the shortest normal from \mathbf{s} onto B_Q (on a principal axis or in a principal subspace), (b) if \mathbf{s} is inside B_Q then the solution is exact and \mathbf{z} is the set of all points on B_Q that project into \mathbf{s} (*not* the shortest normal from \mathbf{s} to B_Q).
- (4) When A and B intersect in a proper subspace of each of them, then project \mathbf{s} onto $A \cap B$ to obtain $\mathbf{s} = \mathbf{s}_{A \cap B} + \mathbf{s}_{B^\perp}$ where \mathbf{s}_{B^\perp} is again orthogonal to B . Proceed as in 3 for $\mathbf{s}_{A \cap B}$ to give $\mathbf{z} = \mathbf{z}_{A \cap B} + \mathbf{s}_{B^\perp}$.

3 Applications

In this section we give some examples of where (1) is to be minimised. We have tried to cover the major varieties of the problem, both from the statistical and algebraic points of view. Thus, Section 3.1 discusses Hardy-Weinberg estimation partly because it includes a linear term but also because it has constraints additional to the quadratic in \mathbf{B} . Section 3.2 gives an example of where \mathbf{B} is indefinite and $k = 0$. In Section 3.3, we turn to reduced rank canonical variate analysis, in which both matrices are p.s.d. but without the vector \mathbf{t} or the linear terms of (1). We shall see that there are subtle issues concerning the interplay between minimising a ratio of quadratic forms and viewing its Lagrangian form as a problem in constrained estimation. This is such an important problem in statistics that we return to it in Section 3.7,

investigating what happens when the usual conditions pertaining to canonical analysis are relaxed and discussing the status of multiple solutions. Section 3.4 reinstates \mathbf{t} in (1), discussing two problems from Procrustes Analysis, in the first of which (1) arises in a substantive way and in the second in an algorithmic context. Sections 3.5 and 3.6 are concerned with the more mainstream statistical issues of quadratically constrained regression, possibly including linear constraints, and with spline fitting.

3.1 Hardy–Weinberg equilibrium

In genetics, we may have three genotypes denoted by AA, BB, AB occurring in proportions $\mathbf{p} = (p_1, p_2, p_3)$. Under random mating, these proportions remain unchanged when the Hardy–Weinberg condition for equilibrium are satisfied

$$p_3^2 = 4p_1p_2. \quad (4)$$

When proportions $\mathbf{q} = (q_1, q_2, q_3)$ are observed, we may wish to estimate \mathbf{p} . Maximum likelihood estimation based on a multinomial distribution might be used (see below) but first we examine least-squares estimation:

$$\min_{\mathbf{p}} \|\mathbf{p} - \mathbf{q}\|^2 \text{ subject to } p_3^2 = 4p_1p_2, \quad (5)$$

which is in the form of equation 1, so the results of Appendices A and B should be immediately available. However, we have the additional constraints $\mathbf{1}'\mathbf{p} = 1$ and $\mathbf{p} \geq 0$ that need to be accommodated. Before showing how to do this, we note that for fixed p_3 , (4) represent a rectangular hyperbola in the plane of p_1, p_2 , so the surface described by the constraint is a series of increasingly large rectangular hyperbolae as one moves away from the origin in the direction of p_3 .

The constraint $\mathbf{1}'\mathbf{p} = 1$ can be handled by transforming into coordinates \mathbf{z} in the plane $\mathbf{1}'\mathbf{p} = 1$. Thus $\mathbf{z} = \mathbf{H}\mathbf{p}$ where \mathbf{H} is the

orthogonal matrix:

$$\mathbf{H} = \begin{pmatrix} \frac{1}{\sqrt{2}} & & \\ & \frac{1}{\sqrt{6}} & \\ & & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix}$$

the diagonal matrix giving the normalisers. With this change of axes we find that the constraint becomes:

$$z_2^2 = \frac{\sqrt{6}}{3} z_1 + \frac{1}{6} \quad (6)$$

which represents a parabola in the plane $\mathbf{1}'\mathbf{p} = 1$, termed the equilibrium parabola. We find it interesting that this slice through the rectangular hyperbolic structure described above is parabolic. Next, consider the effect of the transformation on the objective function. We have:

$$\|\mathbf{p} - \mathbf{q}\|^2 = (\mathbf{p} - \mathbf{q})'(\mathbf{p} - \mathbf{q}) = (\mathbf{p} - \mathbf{q})'\mathbf{H}'\mathbf{H}(\mathbf{p} - \mathbf{q}) = (\mathbf{z} - \mathbf{r})'(\mathbf{z} - \mathbf{r})$$

where $\mathbf{r} = \mathbf{H}\mathbf{q}$. Now, because \mathbf{p} and \mathbf{q} are proportions then $z_3 = r_3 = 1/\sqrt{3}$ and one of the terms vanishes giving the transformed version of the objective function:

$$(z_1 - r_1)^2 + (z_2 - r_2)^2. \quad (7)$$

Thus, the problem is transformed into minimising (7) subject to (6). This is in the form of (1) while ensuring that the constraint $\mathbf{1}'\mathbf{p} = 1$ is satisfied. Furthermore, (6) is in the GCF (with z_1 as linear variable), which was not so for (5). Thus, provided \mathbf{q} satisfies these constraints (see below), it follows that the minimisation of (7) automatically delivers a result that also satisfies the constraints including a non-extraneous linear term. Thus, the problem is now in the GCF form described at the end of Section 1.

Contours of equal least-squares estimates are normal to the equilibrium parabola, as shown on the right-hand-side of Figure C.8.

This diagram confirms that so long as \mathbf{q} is within the triangle, then so is its estimate \mathbf{p} , so guaranteeing that $\mathbf{p} \geq 0$ as well as $\mathbf{1}'\mathbf{p} = 1$.

The multinomial maximum likelihood estimates are:

$$p_1 = q_1 + a, \quad p_2 = q_2 + a, \quad p_3 = q_3 - 2a,$$

where $a = \frac{1}{4}(q_3^2 - 4q_1q_2)$. Thus, $p_1 - p_2 = q_1 - q_2 = \rho$, say. These results, together with (6), give:

$$p_1 = \frac{1}{4}(1 + \rho)^2, \quad p_2 = \frac{1}{4}(1 - \rho)^2, \quad p_3 = \frac{1}{2}(1 - \rho^2)$$

showing that ρ uniquely determines a point on the equilibrium parabola. Contours of constant ρ are linear, being given by the intersection of the planes $\mathbf{1}'\mathbf{q} = 1$ and $q_1 - q_2 = \rho$. These contours are shown on the left hand side of Figure C.8 where they may be compared with the contours normal to the equilibrium parabola, given by least squares.

3.2 Regression with an indefinite constraint

Cases where \mathbf{B} is indefinite are uncommon but do exist in the literature. Thus, Gower and Dijksterhuis (2004) require $\mathbf{x}'\mathbf{B}\mathbf{x} = -1$ where \mathbf{B} is the Householder transformation $\mathbf{I} - 2\mathbf{e}\mathbf{e}'$ where \mathbf{e} is a zero vector apart from a single unit value. In this section, we outline a problem with its origins in the ALSCAL algorithm. ALSCAL (Takane *et al.*, 1977) is a well-known alternating least-squares multidimensional scaling algorithm for minimising the metric SSTRESS criterion $\sum_{i < j} (d_{ij}^2 - \delta_{ij}^2)^2$ where d_{ij} are observed distance-like quantities and δ_{ij} are Euclidean distances generated by points in some small number of dimensions, whose coordinates are required. Ten Berge (1983) has noted that the algorithm requires the solution to a constrained regression problem and gave a bespoke solution. In our notation, Ten Berge's specification re-

quires

$$\left. \begin{aligned} & \min_{\mathbf{x}} \|\mathbf{d} - \mathbf{Kx}\|^2 \\ & \text{subject to } x_2^2 = 4x_1x_3 \end{aligned} \right\}$$

where

$$\mathbf{d} = \begin{pmatrix} .6533 \\ .2706 \\ .2706 \\ .6533 \end{pmatrix} \quad \text{and} \quad \mathbf{K} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}.$$

The constraint looks like (4) but, not being subject to the further conditions of Hardy–Weinberg, cannot be reduced to the two-dimensional parabolic form discussed in Section 3.1. We may write the objective function in the form (1), as follows:

$$\min_{\mathbf{x}} \|\mathbf{d} - \mathbf{Kx}\|^2 = \min_{\mathbf{x}} (\mathbf{x} - \mathbf{t})' \mathbf{A} (\mathbf{x} - \mathbf{t}) + \text{constant},$$

where $\mathbf{A} = \mathbf{K}'\mathbf{K}$ and $\mathbf{t} = (\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{d}$. The matrix form of the constraint is

$$\mathbf{x}' \mathbf{B} \mathbf{x} = \mathbf{x}' \begin{pmatrix} -2 \\ 1 \\ -2 \end{pmatrix} \mathbf{x} = 0,$$

where \mathbf{B} has eigenvalues $(-2, 1, 2)$, showing that we have an indefinite constraint. Both \mathbf{A} and \mathbf{B} are of full rank so the transformation \mathbf{T} (A.4) to the GCF simplifies to $\mathbf{T}'\mathbf{B}\mathbf{T} = \text{diag}(\boldsymbol{\Gamma}_1) = (1-\sqrt{2}, 1/2, 1+\sqrt{2})$ and $\mathbf{s} = \mathbf{T}\mathbf{t} = (0, 0, -1)$ on a principal axis of B . Asymptotes of the GCE (Appendix B) are at $\lambda = 1/\gamma$ with values, in ascending order, of $\text{diag}(-(\sqrt{2}+1), \sqrt{2}-1, 2)$. The first two elements contain the origin and hence define the feasible region. Then the FCE $f(\lambda) = 0$ has two phantom asymptotes, one at the upper boundary of the feasible region, shown shaded in Figure C.9. The optimal λ is given by the boundary of the feasible region at $\lambda = 1/\gamma_3 = \sqrt{2} - 1$. Then, \mathbf{z} is computed according to (B.5), yielding $\mathbf{z} = (-\frac{1}{4}(2 + \sqrt{2}), 0, \pm 8^{-1/2})' = (-.8536, 0, \pm .3528)',$

thus having two solutions. The original parameters are given by $\mathbf{x} = \mathbf{T}^{-1}\mathbf{z} = (2.6132, -2.6132, .6533)'$ and $\mathbf{x} = (.467, 0, 0)'$, both giving the same minimum, in agreement with Ten Berge (1983). Although the two solutions for \mathbf{z} differ only in sign, after transformation to \mathbf{x} they have quite different appearances.

3.3 Canonical analysis

In canonical analysis we require to minimise the ratio:

$$\frac{\mathbf{x}'\mathbf{W}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} \quad (8)$$

where \mathbf{W} and \mathbf{B} are within and between-group sums-of-squares-and-products matrices. Unless both are p.s.d. there is a well known solution in terms of the two-sided eigenvalue problem

$$\mathbf{B}\mathbf{x} = \lambda\mathbf{W}\mathbf{x}.$$

When both matrices are p.s.d. this solution is not immediately available and there may be trivial minima of (8) where $\mathbf{x}'\mathbf{W}\mathbf{x} = 0$ and $\mathbf{x}'\mathbf{B}\mathbf{x} \neq 0$. Suppose \mathbf{X} is the centred data matrix where $\mathbf{X}'\mathbf{X} = \mathbf{T}$ is the total sums-of-squares-and-products matrix. Suppose also that \mathbf{G} is an indicator matrix whose k th column gives membership of the k th group ($k = 1, 2, \dots, K$). Then the group-means are given by $(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{X}$ and the deviations from these means by $(\mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}')\mathbf{X}$. These contribute to the usual orthogonal between/within analysis of variance $\mathbf{T} = \mathbf{B} + \mathbf{W}$ as follows:

$$\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{X} + \mathbf{X}'(\mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}')\mathbf{X}. \quad (9)$$

Now, when \mathbf{v} is a null vector of \mathbf{X} , then $\mathbf{X}\mathbf{v} = 0$. It follows from (9) that any null-vector of \mathbf{X} is also a null vector of \mathbf{T} , \mathbf{B} and \mathbf{W} . Of course, \mathbf{B} and \mathbf{W} may have additional null vectors.

Krzanowski *et al.* (1995) consider the analysis of spectroscopic data where \mathbf{X} has n (30—200) rows representing samples and

p (200—4000) columns representing frequencies, treated as variables. With data like these the ranks of \mathbf{T} , \mathbf{B} , and \mathbf{W} will be much less than the order of the matrices p .

Because all matrices are p.s.d. we may simultaneously diagonalise any two of them (Case 2 of Appendix A). Also, because all matrices share the same null space as \mathbf{T} , then if we choose \mathbf{T} to play the role of \mathbf{A} in the GCF then, whichever plays the role of \mathbf{B} , the matrix $\boldsymbol{\Gamma}_0$ will vanish. Furthermore, because $\mathbf{T} = \mathbf{B} + \mathbf{W}$, having diagonalised \mathbf{T} and \mathbf{B} say, then the same transformation must also diagonalise the third matrix \mathbf{W} . Taking all these things into account we have the following simultaneous diagonalisations:

$$\begin{aligned} \mathbf{T} &= \mathbf{B} + \mathbf{W} \\ \begin{pmatrix} \mathbf{I} \\ \mathbf{I} \\ \mathbf{I} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\Gamma}_r \\ \mathbf{I} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{I} - \boldsymbol{\Gamma}_r \\ \mathbf{0} \\ \mathbf{I} \end{pmatrix}. \end{aligned} \tag{10}$$

In (10) only the rows and columns associated with \mathbf{z}_{11} and \mathbf{z}_{10} of the GCF are represented, the remaining terms are zero, all being in the null space of \mathbf{T} . $\boldsymbol{\Gamma}_1$ is partitioned into two parts, $\boldsymbol{\Gamma}_r$ and \mathbf{I} , the latter being necessary to pair with the possible, but unlikely, occurrences of zero values in \mathbf{W} other than those associated with the unshown null space shared with \mathbf{T} and \mathbf{B} . Normally, \mathbf{B} will have rank K , though it may be less, and normally $r = K$, but (10) allows for all possibilities. The matrices in (10) have been labelled \mathbf{T} , \mathbf{B} and \mathbf{W} though they are now in canonical form. In this section, until further notice, we use these labels to refer to the canonical forms.

Associate the vector $\mathbf{v}' = (\mathbf{v}'_1, \mathbf{v}'_2, \mathbf{v}'_3)$ with the columns of the matrices in (10) then we have the quadratic forms:

$$(\mathbf{v}'_1 \mathbf{v}_1 + \mathbf{v}'_2 \mathbf{v}_2 + \mathbf{v}'_3 \mathbf{v}_3) = (\mathbf{v}'_1 \boldsymbol{\Gamma}_r \mathbf{v}_1 + \mathbf{v}'_2 \mathbf{v}_2) + (\mathbf{v}_1 (\mathbf{I} - \boldsymbol{\Gamma}_r) \mathbf{v}_1 + \mathbf{v}'_3 \mathbf{v}_3). \tag{11}$$

To minimise $\mathbf{v}' \mathbf{W} \mathbf{v}$, we merely have to choose $\mathbf{v} = (\mathbf{0}, \mathbf{v}_2, \mathbf{0})$

giving, $\mathbf{v}'\mathbf{T}\mathbf{v} = \mathbf{v}'\mathbf{B}\mathbf{v} = \mathbf{v}'_2\mathbf{v}_2$, $\mathbf{v}'\mathbf{W}\mathbf{v} = 0$. Thus we have an absolute minimum whichever of $\mathbf{v}'\mathbf{T}\mathbf{v}$ and $\mathbf{v}'\mathbf{B}\mathbf{v}$ may be used as a denominator and whatever normalisation may be used. Thus, if \mathbf{W} has a zero term in (10), no better can be done. Usually, \mathbf{W} will not have this zero term and then (11) simplifies to:

$$(\mathbf{v}'_1\mathbf{v}_1 + \mathbf{v}'_3\mathbf{v}_3) = (\mathbf{v}'_1\boldsymbol{\Gamma}_r\mathbf{v}_1) + (\mathbf{v}_1(\mathbf{I} - \boldsymbol{\Gamma}_r)\mathbf{v}_1 + \mathbf{v}'_3\mathbf{v}_3).$$

Now, minimising $\mathbf{v}'\mathbf{W}\mathbf{v}$ relative to either $\mathbf{v}'\mathbf{T}\mathbf{v}$ or $\mathbf{v}'\mathbf{B}\mathbf{v}$ is equivalent to requiring the maximisation of:

$$\frac{\mathbf{v}'_1\boldsymbol{\Gamma}_r\mathbf{v}_1}{\mathbf{v}'_1\mathbf{v}_1 + \mathbf{v}'_3\mathbf{v}_3}. \quad (12)$$

Whatever the value of \mathbf{v}_1 in (12), the ratio is always greatest when $\mathbf{v}_3 = 0$, the maximum γ_{\max} occurring when \mathbf{v}_1 is zero except for the position corresponding to γ_{\max} . The scaling of \mathbf{v}_1 is immaterial except when multiple solutions are sought (see Section 3.7).

The canonical variables \mathbf{v} are related to the original variables \mathbf{x} by the transformation (A.4) of Appendix A, which in the current context greatly simplifies to:

$$\mathbf{T}_{\text{trans}} = \left(\mathbf{U}_1 \boldsymbol{\Delta}^{-1} \mathbf{V} \mathbf{U}_0 \right),$$

where here we have used the suffixed form $\mathbf{T}_{\text{trans}}$ to distinguish it from the total sum-of-squares-and-products matrix \mathbf{T} , used in this section.

The above justifies what Krzanowski *et al.* (1995) term the PPC method and other methods they describe where \mathbf{W} is modified to reduce the effects of dimensions in which variation, if not actually zero, is deemed to be sufficiently small to be ignored. Equation (10) is particularly simple. This is a consequence of the between and within groups formulation and it should not be thought that these results extend to general ratios of p.s.d. quadratic forms. Newcomb (1961) is usually regarded as the first to discuss the

simultaneous diagonalisation of two symmetric semi-definite matrices. See Uhlig (1979) for a survey. Conditions under which (8) has acceptable optima for general p.s.d. matrices \mathbf{B} and \mathbf{W} are discussed by Rao and Mitra (1971), McDonald *et al.* (1979) and De Leeuw (1982) where the effects of the matrix Γ_0 must also be considered. We discuss this further in section 3.7.

3.4 Problems in Procrustes analysis

3.4.1 Oblique axes

The origins of Procrustes analysis in psychometrics are concerned with transformations to oblique axes. The simplest problem of this kind is to minimise:

$$||\mathbf{X}\mathbf{C} - \mathbf{Y}||^2$$

where \mathbf{X} gives one set of coordinates and \mathbf{Y} another set (termed the target) and direction cosines \mathbf{C} of oblique axes are sought to give the best match. Thus \mathbf{C} is a direction cosine matrix and therefore has columns with unit sum-of-squares. It is clear that the optimisation can be done column by column, so that it is only necessary to solve the problem (13):

$$\min_{\mathbf{c}} ||\mathbf{X}\mathbf{c} - \mathbf{y}||^2 \text{ subject to } \mathbf{c}'\mathbf{c} = 1. \quad (13)$$

Reference to Section 1 confirms that (13) is in the form of one of the simple reparameterisations of (1) so our methods are immediately applicable. A full solution to minimising (13) was given by Browne (1967) using methods that are direct precursors to those described in Appendix B. It was in this context that Cramer (1974) first recognised the potential problem of what we term phantom asymptotes (Appendix B) and proposed methods subsequently improved by Ten Berge and Nevels (1977).

Gower and Dijksterhuis (2004, Chapter 6) discuss oblique axes variants in which \mathbf{C} is replaced by \mathbf{C}' , \mathbf{C}^{-1} or $(\mathbf{C}')^{-1}$ and in which

the constraint need not be positive definite but the GCF remains diagonal. In these variants the columns of \mathbf{C} cannot be estimated independently so our methods have to be used iteratively.

3.4.2 Missing values

The generalised Procrustes problem is:

$$\min_{\mathbf{T}_k} \sum_{k=1}^K \|\mathbf{X}_k \mathbf{T}_k - \mathbf{G}\|^2 \text{ where } \mathbf{G} = \frac{1}{K} \sum_{k=1}^K \mathbf{X}_k \mathbf{T}_k \quad (14)$$

where \mathbf{T}_k is constrained to belong to some matrix-class, typically orthogonal. It may be assumed that each \mathbf{X}_k is column centred, $\mathbf{1}' \mathbf{X}_k = 0$. We shall need an alternative, but equivalent, form of (14):

$$\min_{\mathbf{T}_k} \left(\frac{K-1}{K} \right)^2 \sum_{k=1}^K \|\mathbf{X}_k \mathbf{T}_k - \mathbf{G}_k\|^2 \text{ where } \mathbf{G}_k = \frac{1}{K-1} \sum_{i \neq k}^K \mathbf{X}_i \mathbf{T}_i. \quad (15)$$

\mathbf{G} is termed the group-average and \mathbf{G}_k the k -excluded group-average. We assume that a method is available for minimising (14) and (15) but suppose \mathbf{X}_k ($k = 1, 2, \dots, K$) has missing values. Because the k -excluded group-average does not depend on \mathbf{X}_k , the missing values for each \mathbf{X}_k may be estimated independently. We assume an iterative process with current estimates of the missing values and seek to find an updating matrix \mathbf{X} that is zero except for values, to be calculated, in positions corresponding to the missing values in the k th set. The update should preserve the centring property so we wish to find \mathbf{X} satisfying:

$$\min_{\mathbf{X}} \|(\mathbf{I} - \mathbf{N})(\mathbf{X}_k - \mathbf{X}) \mathbf{T}_k - \mathbf{G}_k\|^2 \quad (16)$$

where \mathbf{N} is the matrix $\mathbf{1}\mathbf{1}'/n$. An additional requirement is that the size of \mathbf{X}_k should be the same before and after updating. Thus:

$$\text{trace} (\mathbf{X}_k - \mathbf{X})' (\mathbf{I} - \mathbf{N}) (\mathbf{X}_k - \mathbf{X}) = \text{trace} (\mathbf{X}_k' \mathbf{X}). \quad (17)$$

Clearly, both the objective function (16) and the constraint (17) are quadratic in the elements of \mathbf{X} , so we have a problem of our basic type (1). We do not give the detailed manipulations to get these forms to coincide with (1) but a first step is to define $\mathbf{x} = \text{vec}(\mathbf{X})$ and $\mathbf{x}_k = \text{vec}(\mathbf{X}_k)$ and then express the quadratic functions as functions of \mathbf{x} . Gower and Dijksterhuis (2004) give further details that fully cover the centring requirement but handle the size constraint by ad hoc methods; the rather heavy, but basically straightforward, algebra required for the full solution will be reported elsewhere.

We note that if this method were to be used when $\mathbf{T}_k = \mathbf{C}_k$, an oblique axis direction cosine matrix, then (1) arises both in the estimation of \mathbf{C}_k and in the estimation of the missing values.

3.5 Normal linear models with quadratic constraints

Consider the following normal linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2)$ is a vector comprising the intercept and two non-zero slopes, $\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2)$ consists of a column of ones and two columns of exogenous variables, and $\boldsymbol{\varepsilon}$ is a vector of i.i.d. $\mathcal{N}(0, \sigma^2)$ residuals. By definition, the OLS estimator $\hat{\boldsymbol{\beta}}$ minimises $\boldsymbol{\beta}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, which has the form of the objective function (1).

Gregory and Veall (1985) were interested in whether $\beta_1\beta_2 = 1$, so considered testing (a) $H_0^a : g^a(\boldsymbol{\beta}) = \beta_1 - 1/\beta_2 = 0$ or, algebraically equivalent, (b) $H_0^b : g^b(\boldsymbol{\beta}) = \beta_1\beta_2 - 1 = 0$. Again, it is immediately clear that H_0^b can be written in the quadratic form $\boldsymbol{\beta}'\mathbf{B}\boldsymbol{\beta}$ with \mathbf{B} a zero matrix, except for $(\mathbf{B})_{23} = 1/2$; hence this example is of the form of (1).

Both hypotheses were studied because each yields a different Wald statistic

$$W = g(\hat{\boldsymbol{\beta}})' \left(\left(\frac{\partial g(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right) I(\hat{\boldsymbol{\beta}})^{-1} \left(\frac{\partial g(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right)' \right)^{-1} g(\hat{\boldsymbol{\beta}}),$$

where I represents the Fisher information matrix, due to the nonlinear relationship under investigation (cf. Lafontainte and White, 1986). Critchley *et al.* (1996) explain the difference in behaviour via differential geometry. They outline a general approach to minimum distance estimation based on the Fisher geodesic statistic, obtained by computing the squared distance between $\hat{\beta}$ and the constraint using the Fisher information matrix as the metric tensor. In the special case of the normal linear model, the resulting methodology takes the form of (1) above.

3.6 Smoothing by splines with bounded roughness

Splines can be used to construct an estimate of a smooth (i.e. twice differentiable) curve for which a number of values x_1, \dots, x_n ($n > 3$) have been observed at locations $t_1 < t_2 < \dots < t_n$ in (a, b) . Splines need to fulfill two contradictory tasks: (i) to fit the data as well as possible; and (ii) to be as smooth as possible. Reinsch (1967) introduced the following approach, which fits into our methodology. His splines $g(\cdot)$ are the solution to the following optimisation problem

$$\min_g \sum_{i=1}^n (x_i - g(t_i))^2 \text{ subject to } \int_a^b (g''(x))^2 dx \leq k, \quad (18)$$

i.e. the goodness-of-fit is measured by the residual sum of squares, and the smoothness by the integral of the squared second derivative. Changing the inequality sign of (18) to an equality sign does not affect the solution(s).

It can be shown that the optimal \hat{g} has the form of a natural cubic spline (NCS) (see, e.g., Green and Silverman, 1994). Without going into too much detail, a function $g(\cdot)$ is called a NCS if (i) on each interval (t_i, t_{i+1}) , g is a cubic polynomial, (ii) $\lim_{t \uparrow t_i} g(t) = \lim_{t \downarrow t_i} g(t) \forall i$, similar properties holding for g' and g'' , and (iii) some regularity conditions hold for $g(t_1)$ and $g(t_n)$. It is possible to completely specify a NCS by two vectors, \mathbf{g} with $g_i = g(t_i)$ and $\boldsymbol{\gamma}$ with $\gamma_i = g''(t_i)$. After introducing matrices \mathbf{Q} ($n \times (n - 2)$)

and \mathbf{R} ($(n - 2) \times (n - 2)$), both defined purely by the positions t_i , we obtain \mathbf{K} via $\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}'$. Now (18) can be rewritten in the form (1) as

$$\left. \begin{array}{l} \min_{\mathbf{g}} (\mathbf{x} - \mathbf{g})'(\mathbf{x} - \mathbf{g}) \\ \text{subject to } \mathbf{g}'\mathbf{K}\mathbf{g} \leq k \end{array} \right\}. \quad (19)$$

Nowadays a slightly different optimisation problem,

$$\min_{\mathbf{g}} ((\mathbf{x} - \mathbf{g})'(\mathbf{x} - \mathbf{g}) + \alpha \mathbf{g}'\mathbf{K}\mathbf{g}) \quad (20)$$

for some $\alpha \geq 0$, is often used (cf. Green and Silverman, 1994). This problem has an explicit solution for given α , but in general the choice of α will be data-dependent and decided via e.g. cross validation, resulting in a methodology outside the scope of (1).

3.7 Ratios, scaling, constraints and multiple solutions

These seemingly miscellaneous topics are all aspects of problems expressed as ratios of quadratic forms

$$\rho = \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}}. \quad (21)$$

We examined this ratio in Section 3.3 in the context of canonical analysis when the matrices \mathbf{A} and \mathbf{B} are both p.s.d. and we shall return to this and related themes later in this section. We distinguish between *multiple solutions* and *multidimensional solutions* to (1). We have seen that (1) generally has a unique minimum, yet in the special case of canonical analysis, it is commonplace to accept solutions based on two, or more, eigenvalues which, with suitable scaling of the associated eigenvectors, generate useful multidimensional visualisations. We regard these as true multidimensional solutions and examine their status in the following, distinguishing them from other multiple solutions derived from different scaling of canonical vectors and, more generally, from multiple solutions of the GCE. First we rehearse the full-rank

case which, although ‘well known’, raises some interesting issues and is needed for understanding similar issues in the non-full rank case.

Full rank

Direct differentiation of (21) yields the two-sided eigenvalue problem:

$$\mathbf{Ax} = \rho \mathbf{Bx} \quad (22)$$

where ρ is minimised (maximised) by choosing the eigenvector associated with the smallest (biggest) eigenvalue. The scaling of \mathbf{x} is arbitrary, though for convenience it is usual to choose $\mathbf{x}'\mathbf{Bx} = 1$. This is a weak constraint that has no effect on the ratio and we could equally well have chosen, for example, $\mathbf{x}'\mathbf{x} = 1$ or $\mathbf{x}'\mathbf{1} = 1$ (provided $\mathbf{x}'\mathbf{1}$ is not zero, as it might be in Optimal Scaling). A seemingly simple alternative Lagrangian form is:

$$\min_{\mathbf{x}} ((\mathbf{x}'\mathbf{Ax}) - \lambda(\mathbf{x}'\mathbf{Bx} - 1)) \quad (23)$$

which again yields (22), so all seems well. However, if a different constraint were used we would now get a different minimum. Indeed, if we imposed a non-quadratic constraint, we would not even have an eigenvalue problem. Thus, in the Lagrangian form, the weak constraint becomes a strong constraint. By regarding (23) as the fundamental statement of the problem of minimising (21), it might seem natural to consider constraints other than $\mathbf{x}'\mathbf{Bx} = 1$. This seems to be the root of the problem reported by Healy and Goldstein (1984) and discussed in detail by Gower (1998). Another difference is that the distinction between objective function and constraint in the Lagrangian form suggests a degree of asymmetry absent from (21).

Throughout we have been concerned with a single multidimensional minimiser \mathbf{x} of (1). In many statistical applications including Principal Component Analysis, Canonical Variate Analysis, Correspondence Analysis, Multiple Correspondence Analysis, Canonical Correlation Analysis, Optimal Scores, we are interested in higher dimensional solutions. In eigenvalue problems

these are given by choosing eigenvectors corresponding to the successive largest (or smallest) eigenvalues, sometimes justified by saying that the k th dimension is chosen optimally, conditional on being orthogonal to the solutions for the first $k - 1$ dimensions. This is unsatisfactory on three grounds: (i) we would prefer a guaranteed global optimum, (ii) the different vectors could each be scaled arbitrarily and (iii) the Hessian \mathbf{H} discussed in Appendix B remains valid, so secondary eigenvalues correspond to saddle points and not maxima or minima. More satisfactory justifications appeal to the best k dimensional least-squares approximation to a matrix (Eckart and Young, 1936) or to its generalisations in terms of weighted least-squares, employing metrics other than the identity metric. One generalisation (see e.g. Gower and Hand, 1995 for details) is based on a generalised singular value decomposition: given a rectangular matrix \mathbf{X} and a symmetric p.d. matrix \mathbf{B} then we may write \mathbf{X} in the form:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{W}^{-1}$$

where \mathbf{U} is orthogonal and $\mathbf{W}'\mathbf{B}\mathbf{W} = \mathbf{I}$ is orthogonal in the metric \mathbf{B} . \mathbf{W} (here not a within-group matrix) is given as the full set of eigenvector solutions to (22) with $\mathbf{A} = \mathbf{XX}'$ and the singular values matrix Σ has the same shape as \mathbf{X} . The normalisation $\mathbf{W}'\mathbf{B}\mathbf{W} = \mathbf{I}$ implies that $\mathbf{B} = (\mathbf{WW}')^{-1}$ which in turn, implies that the outer product of \mathbf{XW} is $\mathbf{XW}(\mathbf{XW})' = \mathbf{XB}^{-1}\mathbf{X}'$, showing that the distance between any two rows of ‘canonical variables’ \mathbf{XW} is a Mahalanobis distance in the metric \mathbf{B} . To preserve the metric approximately in r dimensions requires a generalised Eckart-Young theorem that \mathbf{X}_r , the best rank r approximation to \mathbf{X} in the metric \mathbf{B} , that minimises $\text{trace}(\mathbf{X} - \mathbf{X}_r)\mathbf{B}^{-1}(\mathbf{X} - \mathbf{X}_r)'$ is given by $\mathbf{X}_r = \mathbf{XW}_r\mathbf{W}^r$ where \mathbf{W}_r denotes the first r columns of \mathbf{W} and \mathbf{W}^r the first r rows of \mathbf{W}^{-1} . Geometrically, this approximation is a projection in the metric \mathbf{B} and, because $\mathbf{W}'\mathbf{X}'\mathbf{XW}$ is diagonal, amounts to taking the first r principal components of the canonical variables \mathbf{XW} . Because we are dealing with orthogonal projections onto an r -dimensional subspace, \mathbf{X}_r gives a genuine minimum and not a saddle point solution. The point

of all this is that the scaling $\mathbf{W}'\mathbf{B}\mathbf{W} = \mathbf{I}$ associated with (22) is fundamental to giving a global minimum to approximating Mahalanobis distance in r dimensions, thus giving a basis for multidimensional interpretation that may be used to justify multiple solutions. We do not know of other interpretable scalings but they could exist.

In summary, we have seen that, in the full rank case, the distinction between weak and strong constraints is important and if we are interested in multiple solutions, then this has implications on acceptable relative scaling of the eigenvectors of (22). Next we shall examine how these considerations manifest themselves when ranks are deficient.

Deficient rank

We saw in Section 3.3 that canonical analysis expressed as a problem of minimising ratios of quadratic forms, even when both matrices are p.s.d., has a fairly straightforward solution in terms of the vector \mathbf{x}_1 . Because the criterion was expressed as a ratio, the scaling of \mathbf{x}_1 is arbitrary, although for convenience we may scale so that $\|\mathbf{x}\|^2 = 1$ or $\mathbf{x}'\mathbf{B}\mathbf{x} = 1$. As in the full rank case, these are weak constraints, that have no effect on the value of the ratio.

When \mathbf{A} is of full rank, the GCF gives $\mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{I}$ and $\mathbf{T}'\mathbf{B}\mathbf{T} = \boldsymbol{\Gamma}$ where $\boldsymbol{\Gamma}$ is diagonal but not necessarily of full rank. It follows that $(\mathbf{T}'\mathbf{B}\mathbf{T}) = (\mathbf{T}'\mathbf{A}\mathbf{T})\boldsymbol{\Gamma}$ and because \mathbf{T} is non-singular we have the two-sided eigenvalue expression $(\mathbf{B}\mathbf{T}) = (\mathbf{A}\mathbf{T})\boldsymbol{\Gamma}$, as in the higher dimensional form of (22), allowing \mathbf{T} and $\boldsymbol{\Gamma}$ to be calculated by standard algorithms. When \mathbf{A} is not of full rank and \mathbf{B} is p.s.d. then $\mathbf{T}'\mathbf{B}\mathbf{T} = \boldsymbol{\Gamma}$ remains diagonal but now includes the diagonal matrices $\boldsymbol{\Gamma}_1$ in the range space of \mathbf{A} corresponding to a unit matrix in part of the diagonal of $\mathbf{T}'\mathbf{A}\mathbf{T}$ and $\boldsymbol{\Gamma}_0$ in the null space of \mathbf{A} , corresponding to a null matrix in the remaining part of the diagonal of $\mathbf{T}'\mathbf{A}\mathbf{T}$. Now, the operation $(\mathbf{T}'\mathbf{A}\mathbf{T})\boldsymbol{\Gamma}$ annihilates $\boldsymbol{\Gamma}_0$ so differs from $\mathbf{T}'\mathbf{B}\mathbf{T}$, showing that a necessary and sufficient condition that the two-sided eigenvalue problem formu-

lation remains available is $\boldsymbol{\Gamma}_0 = 0$. We saw that this condition is satisfied in canonical analysis.

When $\boldsymbol{\Gamma}_0$ is not null, the ratio of the quadratic forms becomes:

$$\frac{\mathbf{z}'_1 \mathbf{z}_1 + \mathbf{z}'_2 \mathbf{z}_2}{\mathbf{z}'_1 \boldsymbol{\Gamma}_1 \mathbf{z}_1 + \mathbf{z}'_3 \boldsymbol{\Gamma}_0 \mathbf{z}_3}.$$

Now, by setting \mathbf{z}_1 zero, this ratio may be made indefinitely large or indefinitely small by choosing \mathbf{z}_2 and \mathbf{z}_3 appropriately. However, by applying a strong constraint to the numerator or denominator, (21) may have a nontrivial solution. For example the constraint $\mathbf{z}'_1 \mathbf{z}_1 + \mathbf{z}'_2 \mathbf{z}_2 = 1$ ensures that the ratio has a maximum of $1/\gamma_1$. The position is now reversed from the full rank case where the fundamental ratio form (21) may, with care, be handled by the Lagrangian (22). Now the ratio form is ill-defined and only the Lagrangian form with its strong constraint has any validity. One may query whether such a strong constraint has any justification, other than enabling a non-trivial solution to be found to an artificial problem.

When \mathbf{B} is indefinite we may still have that $\mathbf{T}'\mathbf{B}\mathbf{T}$ is diagonal with either or both of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_0$ indefinite. Further, we may consider the off-diagonal matrices \mathbf{D}_1 and \mathbf{D}_0 of the GCF. We echo the comment of De Leeuw (1982): “If the pair \mathbf{A} and \mathbf{B} is not simultaneously diagonalisable, then the situation becomes considerably more complicated [and] the relevance of this case for practical data analysis is limited”; see also Case 3 of Appendix A.

4 Discussion

At the outset we stated that our objectives were to review, extend and unify a rich class of optimisation problems subsumed in (1). Our examples and citations establish that there are many applications in statistics as there probably are in other fields of science.

Previous work may be dichotomised into (i) the algebraic problems concerned with the simultaneous diagonalisation (or not) of \mathbf{A} and \mathbf{B} and (ii) numerical optimisation problems. The former group can be couched in quite formal mathematical language which we have tried to minimise in this article. Rao and Mitra (1971) place the origins of the algebraic approach with Weierstrass and Kronecker, stating that their own approach is simpler. Yet it is formulated in terms of Hermitian matrices and generalised inverses and is concerned with both cogredient and contragredient transformations. This is admirably complete but is more than we have found necessary for ordinary statistical work. Thus we are concerned only with real symmetric matrices, consider only cogredient transformations $\mathbf{T}'\mathbf{A}\mathbf{T}$ paired with $\mathbf{T}'\mathbf{B}\mathbf{T}$, and subsume generalised inverse considerations by appealing to the partition of orthogonal matrices $\mathbf{V} = (\mathbf{V}_1 \ \mathbf{V}_0)$ where \mathbf{V}_0 represents an arbitrary set of orthogonal column-vectors spanning a null space. With these simplifications we have been able to give an explicit representation (A.1) for \mathbf{T} and its inverse (A.5).

These results have been mainly used in the literature to optimise ratio criteria such as (21) arising from a variety of statistical canonical variable problems. De Leeuw (1982), extending work by McDonald *et al.* (1979), discusses the optimisation of $\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{B}\mathbf{x}$ where \mathbf{A} and \mathbf{B} are p.s.d. Critchley (1990) performed preliminary studies for our problem in the case where \mathbf{A} is positive definite. Gower and Dijksterhuis (2004) address the problem in the context of Procrustes analysis and give a preliminary algorithm.

Gander (1981) studied exact solutions and their properties of the optimisation problem $\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|^2$ subject to $\|\mathbf{Cx} - \mathbf{d}\|^2 = 0$, that is strongly related to ours (see Section 1). Moré (1993) studied exact solutions of minimising quadratic functions subject to ellipsoidal constraints. Minimising quadratic functions under quadratic constraints is an active topic in optimisation theory (see, for example, Tuy and Hoai-Phuong, 2007).

We have seen that when some elements of \mathbf{t} are zero (1) has

eigenvalue-like characteristics, reducing to this type of problem when $\mathbf{t} = 0$. The case $\mathbf{t} = 0$ is discussed in sections 3.3 and 3.7 where our approach may be compared with previous work cited there. These problems may usually be expressed as two-sided eigenvalue problems, though we have seen that this is not always possible. We have taken the opportunity to relate this field of study to the problems arising from expressing criteria in ratio or Lagrangian form, with associated, and often disregarded, considerations of weak and strong constraints, and to the status of multiple solutions. These issues are more subtle when \mathbf{A} and \mathbf{B} are not of full rank.

Notwithstanding the importance of the case $\mathbf{t} = 0$, our main concern has been with the more general problem where \mathbf{t} is a given non-zero vector. Special cases have arisen in the literature but we have given a unified approach that leads to a general purpose algorithm. We believe that not only does this contribute to a better understanding of this class of problems, but also greatly helps in the formulation and solution to special cases of this class that may arise in the future. There is generally a unique minimum to (1) but, like the algebraic eigenvalue problem, pathological solutions occur under unlikely practical circumstances. It is important that they be taken into account to ensure robust algorithms.

We have dealt with the general case where \mathbf{A} is p.s.d. and \mathbf{B} possibly indefinite. When \mathbf{B} is indefinite, there are genuine applications when the GCF is diagonal but rarely in the more general case that includes the theoretically possible off-diagonal matrices \mathbf{D} . We have also included the linear term $2\mathbf{b}'\mathbf{x}$ in (1) but have seen that in many cases reparameterisation eliminates this term. The main exception is when there is a linear term in the constraint with no quadratic counterpart, in which case the geometric form of the constraint is parabolic, which we have seen gives rise to no special problem. The other exception is when there is a linear term in the extraneous variable \mathbf{z}_{00} , also associated with \mathbf{D} . Recall that \mathbf{z}_{00} represents a variable in the null spaces of \mathbf{A} and \mathbf{B} , so would normally be immaterial. However it could occur as an

extraneous linear variable in the constraint in which case it may always be chosen to give an exact fit. This draws attention to a limitation of (1) where, if the full generality is used, the resulting overparameterisation leads to exact solutions.

Problem (1) is a vector minimisation problem. In Section 1 we showed that even an apparent multidimensional matrix generalisation involving terms such as $\min \text{trace}(\mathbf{X} - \mathbf{T})'\mathbf{A}(\mathbf{X} - \mathbf{T})$ subject to $\text{trace}(\mathbf{X}'\mathbf{B}\mathbf{X}) = k$ is subsumed in (1). This type of multidimensional generalisation should be distinguished from higher dimensional solutions associated with many well-known multivariate methods, some of which have been discussed above, that are based on combining solutions given by several eigenvalues. Minimising (1), as with eigenvalue problems, also has several extrema but we do not know whether these multiple solutions may be of practical interest; we suspect not because even when \mathbf{B} is p.s.d. some roots of the FCE (B.3) may not be real. Finally, we should mention a totally different kind of generalisation of (21), involving functions of quadratic forms, with or without constraints. Thus, Kiers (1995) discusses minimising/maximising sums of ratios of quadratic forms, offering algorithmic solutions to several variants of the problem.

Acknowledgements. We thank Chris Jones for his helpful comments.

Appendices

A The general canonical form

We are concerned with the simultaneous simplification of \mathbf{A} and \mathbf{B} . In some cases this becomes simultaneous diagonalisation but in general, there exists a non-singular transformation \mathbf{T} that gives

the General Canonical Form:

$$\mathbf{T}'\mathbf{A}\mathbf{T} = \begin{pmatrix} \mathbf{I} & & \\ & \mathbf{I} & : \\ & : & \\ \dots & \dots & : & \dots & \dots \\ & & : & & \\ & & : & & \end{pmatrix}; \mathbf{T}'\mathbf{B}\mathbf{T} = \begin{pmatrix} \mathbf{\Gamma}_1 & & & & \mathbf{D}_{10} \\ & \vdots & & & \mathbf{D}_{00} \\ & & \ddots & & \\ \dots & \dots & : & \dots & \dots \\ & & & & \mathbf{\Gamma}_0 \\ \mathbf{D}'_{10} & \mathbf{D}'_{00} & : & & \end{pmatrix} \quad (\text{A.1})$$

In (A.1) we have indicated a partition into the range and null spaces of \mathbf{A} . The implicit partitions corresponding to the diagonal matrices $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_0$ are in the range space of $\mathbf{T}'\mathbf{B}\mathbf{T}$. For p.s.d. matrices, $\mathbf{z}'\mathbf{B}\mathbf{z} = 0$ implies that $\mathbf{B}\mathbf{z} = 0$, so then the zero diagonals correspond to the null space of $\mathbf{T}'\mathbf{B}\mathbf{T}$. However, for indefinite matrices this is not so and the partitions corresponding to the zero diagonals may include parts of the range space of \mathbf{B} . To be more explicit and modifying the notation of Section 1, so that now \mathbf{z}_{ij} denotes the vector $\mathbf{z}' = (\mathbf{z}'_{11}, \mathbf{z}'_{10}, \mathbf{z}'_{01}, \mathbf{z}'_{00})$ with all elements zero except \mathbf{z}_{ij} ($i = 0, 1; j = 0, 1$), we note that $(\mathbf{T}'\mathbf{B}\mathbf{T})\mathbf{z}_{10}$ and $(\mathbf{T}'\mathbf{B}\mathbf{T})\mathbf{z}_{00}$ need not be null when \mathbf{D}_{00} and/or \mathbf{D}_{10} do not vanish, even though $\mathbf{z}'_{10}(\mathbf{T}'\mathbf{B}\mathbf{T})\mathbf{z}_{10}$ and $\mathbf{z}'_{00}(\mathbf{T}'\mathbf{B}\mathbf{T})\mathbf{z}_{00}$ both vanish.

The transformation \mathbf{T} may be written explicitly in terms of three spectral decompositions, the decomposition of \mathbf{A} being used to define \mathbf{C} in (A.3) and, thereby, the other two decompositions. Denoting non-null eigenvectors by a unit suffix (e.g. \mathbf{U}_1) and null eigenvectors by a zero suffix (e.g. \mathbf{U}_0) with $\mathbf{U} = (\mathbf{U}_1 \ \mathbf{U}_0)$ giving a complete orthogonal matrix, the required spectral decomposi-

tions are as follows:

$$\left. \begin{aligned} \mathbf{A} &= \mathbf{U}_1 \Delta^2 \mathbf{U}'_1 \\ \mathbf{C}_{22} &= \mathbf{W}_1 \Gamma_0 \mathbf{W}'_1 \\ \mathbf{C}_{11} - \mathbf{C}_{12} (\mathbf{W}_1 \Gamma_0^{-1} \mathbf{W}'_1) \mathbf{C}_{21} &= \mathbf{V}_1 \Gamma_1 \mathbf{V}'_1 \end{aligned} \right\} \quad (\text{A.2})$$

where

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} = \begin{pmatrix} \Delta^{-1} \mathbf{U}'_1 \mathbf{B} \mathbf{U}_1 \Delta^{-1} & \Delta^{-1} \mathbf{U}'_1 \mathbf{B} \mathbf{U}_0 \\ \mathbf{U}'_0 \mathbf{B} \mathbf{U}_1 \Delta^{-1} & \mathbf{U}'_0 \mathbf{B} \mathbf{U}_0 \end{pmatrix}. \quad (\text{A.3})$$

Δ^2 , Γ_1 , Γ_0 are all non-singular diagonal matrices of eigenvalues; we have used Δ^2 to emphasise the positivity of the non-zero eigenvalues of \mathbf{A} . Also $\mathbf{D}_{10} = \mathbf{V}'_1 \mathbf{C}_{12} \mathbf{W}_0$ and $\mathbf{D}_{00} = \mathbf{V}'_0 \mathbf{C}_{12} \mathbf{W}_0$.

With the notation given in (A.2) and (A.3), the transformation may be written explicitly as:

$$\mathbf{T} = \left(\mathbf{U}_1 \Delta^{-1} \mathbf{V} - \mathbf{U}_0 \mathbf{W}_1 \Gamma_0^{-1} \mathbf{W}_1 \mathbf{C}_{21} \mathbf{V} \mathbf{U}_0 \mathbf{W} \right) \quad (\text{A.4})$$

with inverse

$$\mathbf{T}^{-1} = \begin{pmatrix} \mathbf{V}' \Delta \mathbf{U}'_1 \\ \mathbf{W}' \mathbf{W}_1 \Gamma_0^{-1} \mathbf{W}'_1 \mathbf{C}_{21} \Delta \mathbf{U}'_1 + \mathbf{W}' \mathbf{U}'_0 \end{pmatrix}. \quad (\text{A.5})$$

We do not give a detailed derivation of (A.1) here, but on substituting for \mathbf{T} from (A.4) it is easily, if somewhat tediously, verified to give the result claimed. A constructive proof may be found in Albers *et al.* (2007). Note that because the null eigenvectors (those with zero suffices) are determined only up to spanning the null space, there is a degree of non-uniqueness in (A.4) and (A.5). This has no substantive effect in applications.

In the above, we have assumed that Γ_0 is present. However, when \mathbf{C}_{22} is absent, then so is Γ_0 and the above formulae have to be modified by replacing all Γ_0^{-1} by zero and setting $\mathbf{W}_0 = 0$, $\mathbf{W}_1 =$

$\mathbf{W} = \mathbf{I}$. Indeed, \mathbf{W} may be any arbitrary orthogonal matrix, a fact that allows $(\mathbf{D}'_{10} \mathbf{D}'_{00})'$ to be represented only by as many non-zero columns as its rank. We do not pursue this simplification further here.

The transformation \mathbf{T} also operates on \mathbf{x} to give $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x} = (\mathbf{z}'_{11}, \mathbf{z}'_{10}, \mathbf{z}'_{01}, \mathbf{z}'_{00})'$. We may also write $\mathbf{b}'\mathbf{x} = \mathbf{g}'\mathbf{z}$ where $\mathbf{g}' = \mathbf{b}'\mathbf{T}$. The effect of the transformation \mathbf{T} is that (1) has been simplified into the following form:

$$\left. \begin{aligned} & \min_{\mathbf{z}} \|\mathbf{z}_{11} - \mathbf{s}_{11}\|^2 + \|\mathbf{z}_{10} - \mathbf{s}_{10}\|^2 \\ & \text{subject to } \mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11} + \mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01} + 2\mathbf{z}'_{11}\mathbf{D}_{10}\mathbf{z}_{00} + 2\mathbf{z}'_{10}\mathbf{D}_{00}\mathbf{z}_{00} + 2\mathbf{g}'\mathbf{z} = k \end{aligned} \right\} \quad (\text{A.6})$$

The linear term in (A.6) may be simplified further. Examining the individual contributions gives $\mathbf{g}'\mathbf{z} = \mathbf{g}'_{11}\mathbf{z}_{11} + \mathbf{g}'_{10}\mathbf{z}_{10} + \mathbf{g}'_{01}\mathbf{z}_{01} + \mathbf{g}'_{00}\mathbf{z}_{00}$ and focussing on the term $\mathbf{g}'_{11}\mathbf{z}_{11}$, we observe that:

$$\mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11} + 2\mathbf{g}'_{11}\mathbf{z}_{11} - k = (\mathbf{z}_{11} + \Gamma_1^{-1}\mathbf{g}_{11})'\Gamma_1(\mathbf{z}_{11} + \Gamma_1^{-1}\mathbf{g}_{11}) - (k + \mathbf{g}'_{11}\Gamma_1^{-1}\mathbf{g}_{11}).$$

A similar observation applies to $\mathbf{g}'_{01}\mathbf{z}_{01}$, allowing the linear terms in \mathbf{g}_{11} and \mathbf{g}_{01} to be absorbed into the quadratic terms, by the translation

$$\mathbf{z} \rightarrow \mathbf{z} + \mathbf{m} \text{ where } \mathbf{m}' = \left(\mathbf{g}'_{11}\Gamma_1^{-1}, \mathbf{0}, \mathbf{g}'_{01}\Gamma_0^{-1}, \mathbf{0} \right).$$

The same translation must be made to \mathbf{s}_{11} but not \mathbf{s}_{01} (as it is not defined). Combining all these refinements, (A.6) becomes:

$$\left. \begin{aligned} & \min_{\mathbf{z}} \|\mathbf{z}_{11} - \mathbf{s}_{11}\|^2 + \|\mathbf{z}_{10} - \mathbf{s}_{10}\|^2 \\ & \text{subject to } \\ & \mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11} + \mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01} + 2\mathbf{z}'_{11}\mathbf{D}_{10}\mathbf{z}_{00} + 2\mathbf{z}'_{10}\mathbf{D}_{00}\mathbf{z}_{00} + 2\mathbf{g}'_{10}\mathbf{z}_{10} + 2\mathbf{g}'_{00}\mathbf{z}_{00} = k \end{aligned} \right\} \quad (\text{A.7})$$

where it is assumed that, when appropriate, the translational adjustments are included in \mathbf{z}_{11} , \mathbf{z}_{01} , \mathbf{s}_{11} and k .

The main special cases

Equation (A.7) seems to be the most simple general form of (1) though further simplifications occur in important special cases, considered in the following. Nothing is lost by considering the matrix in (3) the transformed version of \mathbf{B} , so in the remainder of this Appendix references to \mathbf{B} are to its transformed, or canonical, form.

Case 1. \mathbf{A} of full rank

When \mathbf{A} is of full rank, it has no null space so the extraneous variables \mathbf{z}_{01} and \mathbf{z}_{00} do not occur. This is the most important practical case because it implies that, as is usually the case, the constraint contains no variables that are not in the objective function. Thus $\mathbf{z}' = (\mathbf{z}'_{11}, \mathbf{z}'_{10})$, and the minimisation problem (3) reduces to:

$$\left. \begin{array}{l} \min_{\mathbf{z}} (||\mathbf{z}_{11} - \mathbf{s}_{11}||^2 + ||\mathbf{z}_{10} - \mathbf{s}_{10}||^2) \\ \text{subject to } \mathbf{z}'_{11} \boldsymbol{\Gamma}_1 \mathbf{z}_{11} + 2\mathbf{g}'_{10} \mathbf{z}_{10} = k \end{array} \right\}.$$

The normal equations derived from the Lagrangian are:

$$\left. \begin{array}{l} \mathbf{z}_{11} - \mathbf{s}_{11} = \lambda \boldsymbol{\Gamma}_1 \mathbf{z}_{11} \\ \mathbf{z}_{10} - \mathbf{s}_{10} = \lambda \mathbf{g}_{10} \end{array} \right\}. \quad (\text{A.8})$$

When there is no linear term, the second equation of (A.8) gives $\mathbf{z}_{10} = \mathbf{s}_{10}$ and the problem is in what we term the fundamental canonical form (FCF) discussed in Appendix B. Appendix B gives the solution to the minimisation of the FCF when exact solutions are unavailable. Usually, but not always, the associated fundamental canonical equation (FCE) has a unique easily computed minimum (see Appendix B). When the linear term \mathbf{g}_{10} is

included, the constraint becomes:

$$\mathbf{z}'_{11}\boldsymbol{\Gamma}_1\mathbf{z}_{11} + 2\lambda\mathbf{g}'_{10}\mathbf{g}_{10} = k - 2\mathbf{g}'_{10}\mathbf{s}_{10}. \quad (\text{A.9})$$

Note the positive coefficient of λ .

Case 2. \mathbf{A} not of full rank, \mathbf{B} is diagonal (including \mathbf{B} positive semi-definite)

So far we have assumed that there are no extraneous variables, \mathbf{z}_{01} and \mathbf{z}_{00} . In Case 2 we begin to relax this condition. Suppose that \mathbf{B} is p.s.d. The GCF simplifies because any zero diagonal values of a p.s.d. matrix induce corresponding entire zero rows and columns. Thus, in (3) the matrices \mathbf{D}_{10} and \mathbf{D}_{00} are zero. Also, both $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_0$ must be positive definite. Thus, in this case (3) represents a simultaneous diagonalisation of \mathbf{A} and \mathbf{B} (Newcomb, 1961). Diagonal forms of (3) may also occur when \mathbf{B} is not p.s.d. in which case $\boldsymbol{\Gamma}_1$ and/or $\boldsymbol{\Gamma}_0$ must be indefinite. We now consider all these situations.

We first consider the case where \mathbf{B} is p.s.d. There is an exact solution with $\mathbf{z}_{11} = \mathbf{s}_{11}$, $\mathbf{z}_{10} = \mathbf{s}_{10}$ when the constraint

$$\mathbf{s}'_{11}\boldsymbol{\Gamma}_1\mathbf{s}_{11} + 2\mathbf{g}'_{10}\mathbf{s}_{10} + \mathbf{z}'_{01}\boldsymbol{\Gamma}_0\mathbf{z}_{01} + 2\mathbf{g}'_{00}\mathbf{z}_{00} = k \quad (\text{A.10})$$

has solutions for \mathbf{z}_{01} and \mathbf{z}_{00} . For any given setting of \mathbf{z}_{01} , and when $\mathbf{g}_{00} \neq 0$, (A.10) always has a solution for \mathbf{z}_{00} . Therefore, we regard the problem as over-parameterised if the linear term in \mathbf{g}_{00} is admitted. Even when $\mathbf{g}_{00} = 0$ (A.10) may still have exact solutions but in general cannot be satisfied and we must seek non-exact solutions. Then, the normal equations of the Lagrangian simplify to:

$$\left. \begin{array}{l} \mathbf{z}_{11} - \mathbf{s}_{11} = \lambda\boldsymbol{\Gamma}_1\mathbf{z}_{11} \\ \mathbf{z}_{10} - \mathbf{s}_{10} = \lambda\mathbf{g}_{10} \\ 0 = \lambda\boldsymbol{\Gamma}_0\mathbf{z}_{01} \\ 0 = \lambda\mathbf{g}_{00} \end{array} \right\}. \quad (\text{A.11})$$

When there is no exact solution, the first two equations of (A.11) show that we must have $\lambda \neq 0$. Then, the last equation of (A.11) shows that $\mathbf{g}_{00} = 0$, confirming what we've already seen above that $\mathbf{g}_{00} = 0$ is a necessary condition for a non-exact solution. When present, the third equation of (A.11) gives $\mathbf{z}_{01} = 0$. Thus, the extraneous variables are either absent, zero (\mathbf{z}_{01}) or irrelevant (\mathbf{z}_{00}). Overall, we arrive back at the equations (A.8) of the FCF.

When \mathbf{B} is diagonal but not p.s.d. the only difference is that (A.10) has exact solutions whenever $\boldsymbol{\Gamma}_0$ is indefinite as well as when $\mathbf{g}_{00} \neq 0$. We are therefore led to regard this situation to be over-parameterised and admit only $\boldsymbol{\Gamma}_1$ to be indefinite and $\boldsymbol{\Gamma}_0$ to be either positive or negative definite, depending on the sign of $k - \mathbf{s}'_{11}\boldsymbol{\Gamma}_1\mathbf{s}_{11} - 2\mathbf{g}'_{10}\mathbf{s}_{10}$.

Case 3. \mathbf{A} not of full rank, \mathbf{B} not diagonal

In this case we must treat the full GCF, first examining the possibility of exact solutions. We have seen that when \mathbf{B} is p.s.d., it must be diagonal. Hence, when \mathbf{B} includes non-zero \mathbf{D}_{10} and/or \mathbf{D}_{00} it must be indefinite. The constraint may be written:

$$\mathbf{z}'_{11}\boldsymbol{\Gamma}_1\mathbf{z}_{11} + \mathbf{z}'_{01}\boldsymbol{\Gamma}_0\mathbf{z}_{01} + 2\mathbf{g}'_{10}\mathbf{z}_{10} + 2\mathbf{p}'(\mathbf{z})\mathbf{z}_{00} = k, \quad (\text{A.12})$$

where $\mathbf{p}(\mathbf{z}) = \mathbf{D}'_{10}\mathbf{z}_{11} + \mathbf{D}'_{00}\mathbf{z}_{00} + \mathbf{g}_{00}$. For an exact solution, $\mathbf{z}_{11} = \mathbf{s}_{11}$ and $\mathbf{z}_{10} = \mathbf{s}_{10}$, and (A.12) becomes:

$$\mathbf{s}'_{11}\boldsymbol{\Gamma}_1\mathbf{s}_{11} + \mathbf{z}'_{01}\boldsymbol{\Gamma}_0\mathbf{z}_{01} + 2\mathbf{g}'_{10}\mathbf{s}_{10} + 2\mathbf{p}'(\mathbf{s})\mathbf{z}_{00} = k, \quad (\text{A.13})$$

which must have real solutions for \mathbf{z}_{00} and \mathbf{z}_{01} . For any setting of \mathbf{z}_{01} , (A.13) is linear in \mathbf{z}_{00} and, unless $\mathbf{p}(\mathbf{s}) = 0$, will always have a solution. When $\mathbf{p}(\mathbf{s}) = 0$, \mathbf{z}_{00} becomes arbitrary but there remains the possibility that $\mathbf{z}'_{01}\boldsymbol{\Gamma}_0\mathbf{z}_{01} = k - \mathbf{s}'_{11}\boldsymbol{\Gamma}_1\mathbf{s}_{11} - 2\mathbf{g}'_{10}\mathbf{s}_{10}$ has solutions for \mathbf{z}_{01} ; this is trivially true when $\boldsymbol{\Gamma}_0$ is itself indefinite and remains a possibility when $\boldsymbol{\Gamma}_0$ is definite. It follows that for exact solutions not to exist it is necessary, though not sufficient, for $\boldsymbol{\Gamma}_0$ to be definite or absent, i.e. \mathbf{z}_{01} is excluded from the con-

straint. Furthermore, we require $\mathbf{p}(\mathbf{s}) = 0$. With these settings, \mathbf{z}_{00} enters neither into the objective function nor the constraint. Nevertheless, as we see below in (A.14), \mathbf{z}_{00} still enters the normal equations. For a full treatment of essentially exact solutions, see Albers *et al.* (2007).

When there is a non-exact solution, the normal equations will take the following form for non-zero λ :

$$\left. \begin{array}{l} \mathbf{z}_{11} - \mathbf{s}_{11} = \lambda(\mathbf{\Gamma}_1 \mathbf{z}_{11} + \mathbf{D}_{10} \mathbf{z}_{00}) \\ \mathbf{z}_{10} - \mathbf{s}_{10} = \lambda(\mathbf{D}_{00} \mathbf{z}_{00} + \mathbf{g}_{10}) \\ 0 = \lambda \mathbf{\Gamma}_0 \mathbf{z}_{01} \\ 0 = \lambda(\mathbf{D}'_{10} \mathbf{z}_{11} + \mathbf{D}'_{00} \mathbf{z}_{10} + \mathbf{g}_{00}) \end{array} \right\}. \quad (\text{A.14})$$

The third of the equations (A.14) shows that, when present, $\mathbf{z}_{01} = 0$ is necessary for a non-exact solution. The fourth equation gives $\mathbf{p}(\mathbf{z}) = 0$, to which must be added the necessary condition $\mathbf{p}(\mathbf{s}) = 0$ for a non-exact solution. The constraint simplifies to:

$$\mathbf{z}'_{11} \mathbf{\Gamma}_1 \mathbf{z}_{11} + 2\mathbf{z}'_{10} \mathbf{g}_{10} = k. \quad (\text{A.15})$$

From $\mathbf{p}(\mathbf{z}) - \mathbf{p}(\mathbf{s}) = 0$ we have that $\mathbf{D}'_{10}(\mathbf{z}_{11} - \mathbf{s}_{11}) + \mathbf{D}'_{00}(\mathbf{z}_{10} - \mathbf{s}_{10}) = 0$ which shows that $\mathbf{z}_{11} - \mathbf{s}_{11}$ and $\mathbf{z}_{10} - \mathbf{s}_{10}$ must lie in the null row space of \mathbf{D}_{10} and \mathbf{D}_{00} . That is

$$\begin{pmatrix} \mathbf{z}_{11} - \mathbf{s}_{11} \\ \mathbf{z}_{10} - \mathbf{s}_{10} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1 \mathbf{h} \\ \mathbf{H}_0 \mathbf{h} \end{pmatrix} \text{ where } \mathbf{H}' \mathbf{D} = \begin{pmatrix} \mathbf{H}'_1 & \mathbf{H}'_0 \end{pmatrix} \begin{pmatrix} \mathbf{D}_{10} \\ \mathbf{D}_{00} \end{pmatrix} = 0 \quad (\text{A.16})$$

for some \mathbf{h} . The matrices \mathbf{H}_1 and \mathbf{H}_0 may be found from the SVD of \mathbf{D} and without loss of generality \mathbf{H} may be assumed to be column orthonormal. The first two equations of (A.14) then become:

$$\left. \begin{array}{l} \mathbf{H}_1 \mathbf{h} = \lambda(\mathbf{\Gamma}_1(\mathbf{H}_1 \mathbf{h} + \mathbf{s}_{11}) + \mathbf{D}_{10} \mathbf{z}_{00}) \\ \mathbf{H}_0 \mathbf{h} = \lambda(\mathbf{D}_{00} \mathbf{z}_{00} + \mathbf{g}_{10}) \end{array} \right\}.$$

Multiplying the first equation by \mathbf{H}'_1 the second by \mathbf{H}'_0 and adding, using (A.16) and orthonormality of \mathbf{H} , gives:

$$\mathbf{h} = \lambda ((\mathbf{H}'_1 \Gamma_1 \mathbf{H}_1) \mathbf{h} + \mathbf{g}) \quad (\text{A.17})$$

where $\mathbf{g} = \mathbf{H}'_1 \Gamma_1 \mathbf{s}_{11} + \mathbf{H}'_0 \mathbf{g}_{10}$.

The constraint (A.15) becomes:

$$(\mathbf{H}_1 \mathbf{h} + \mathbf{s}_{11})' \Gamma_1 (\mathbf{H}_1 \mathbf{h} + \mathbf{s}_{11}) + 2\mathbf{g}'_{10} (\mathbf{H}_0 \mathbf{h} + \mathbf{s}_{10}) = k,$$

i.e.

$$\mathbf{h}' (\mathbf{H}'_1 \Gamma_1 \mathbf{H}_1) \mathbf{h} + 2\mathbf{h}' \mathbf{g} = k - \mathbf{s}'_{11} \Gamma_1 \mathbf{s}_{11} - 2\mathbf{g}'_{10} \mathbf{s}_{10} = k^*, \quad (\text{A.18})$$

say.

With the spectral decomposition $\mathbf{H}'_1 \Gamma_1 \mathbf{H}_1 = \mathbf{K}' \Delta \mathbf{K}$, (A.17) and (A.18) give:

$$\left. \begin{aligned} \mathbf{K}\mathbf{h} &= \lambda(\Delta \mathbf{K}\mathbf{h} + \mathbf{K}\mathbf{g}) \\ (\mathbf{K}\mathbf{h})' \Delta (\mathbf{K}\mathbf{h}) + 2(\mathbf{K}\mathbf{h})' (\mathbf{K}\mathbf{g}) &= k^*. \end{aligned} \right\} \quad (\text{A.19})$$

With $\mathbf{z} = \mathbf{K}\mathbf{h}$ and $\mathbf{b} = \mathbf{K}\mathbf{g}$, these are the normal equations arising from minimisation of $||\mathbf{z}||^2$ subject to $\mathbf{z}' \Delta \mathbf{z} + 2\mathbf{b}' \mathbf{z} = k^*$. Since Δ is diagonal, this is an instance of Case 2 above. Its solution $\hat{\mathbf{z}}$ gives $\hat{\mathbf{h}} = \mathbf{K}' \hat{\mathbf{z}}$ which may be substituted into (A.16) to obtain $\hat{\mathbf{z}}_{11}$ and $\hat{\mathbf{z}}_{10}$.

In general, the constraint $\mathbf{p}(\mathbf{s}) = 0$ for non-exact solution seems unrealistic. An exception is when \mathbf{D}_{10} and \mathbf{D}_{00} are both zero, in which case we return to Case 2.

B Solving the fundamental canonical equation

We have seen in the previous appendix that the basic minimisation problem required by all forms of (1) may be expressed in the

Fundamental Canonical Form (FCF)

$$\left. \begin{aligned} & \min_{\mathbf{z}} \|\mathbf{z} - \mathbf{s}\|^2 \\ & \text{subject to } \mathbf{z}'\boldsymbol{\Gamma}\mathbf{z} = k \end{aligned} \right\}, \quad (\text{B.1})$$

with $\boldsymbol{\Gamma}$ non-singular (noting that k here differs from k in (1) unless $\mathbf{b} = 0$).

The Lagrangian form is to minimise:

$$\|\mathbf{z} - \mathbf{s}\|^2 - \lambda(\mathbf{z}'\boldsymbol{\Gamma}\mathbf{z} - k)$$

which on differentiation, gives:

$$(\mathbf{z} - \mathbf{s}) - \lambda\boldsymbol{\Gamma}\mathbf{z} = 0.$$

When $\mathbf{s}'\boldsymbol{\Gamma}\mathbf{s} = k$ the constraint is satisfied for the exact solution $\mathbf{z} = \mathbf{s}$ and $\lambda = 0$. For approximate solutions,

$$\mathbf{z} = (\mathbf{I} - \lambda\boldsymbol{\Gamma})^{-1}\mathbf{s},$$

which on substitution into the constraint gives:

$$\mathbf{s}'(\mathbf{I} - \lambda\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\mathbf{I} - \lambda\boldsymbol{\Gamma})^{-1}\mathbf{s} = k. \quad (\text{B.2})$$

Equation (B.2) gives the basic Fundamental Canonical Equation (FCE) that has to be solved for λ . It is convenient to write (B.2) in the non-matrix form:

$$f(\lambda) = \sum_{i=1}^p \frac{\gamma_i s_i^2}{(1 - \lambda\gamma_i)^2} = k. \quad (\text{B.3})$$

The FCE could be expanded as a polynomial of degree $2p$, but it is more convenient to retain its original form. Provided $s_i \neq 0$, there are vertical asymptotes at $\lambda = 1/\gamma_i$ ($i = 1, 2, \dots, p$); the case where some, or all s_i are zero, is discussed below. Assuming that the eigenvalues γ_i are given in increasing order, for indefinite \mathbf{B} γ_1 will be the smallest negative eigenvalue and γ_p the largest positive eigenvalue. However, for p.d matrices \mathbf{B} , all eigenvalues will be positive.

Figure C.10a indicates the general shape of $f(\lambda)$ when \mathbf{B} is indefinite. We see that the origin is contained in the shaded interval:

$$F : \lambda \in \left(\frac{1}{\gamma_1} \leq 0 \leq \frac{1}{\gamma_p} \right),$$

termed the *feasible region* for reasons about to be explained. Furthermore the Figure shows that $f(\lambda)$ is monotonically increasing in this interval and not in any other interval determined by adjacent asymptotes. This is an easily proved general result. Because of the monotonicity, there is at most one root in F . There may be real pairs of roots in other intervals but from the Hessian $\mathbf{H} = \mathbf{I} - \lambda \boldsymbol{\Gamma}$, it follows that it is only when $\lambda \in F$ are all the diagonal values of \mathbf{H} positive, indicating the existence of a minimum; other roots all refer to saddle-point solutions or, possibly, a root indicating a maximum in one of the end branches of $f(\lambda)$. Thus, it suffices to focus on the feasible region F , shaded in Figure C.10a. When \mathbf{B} is p.s.d. no γ_i is negative and then Figure C.10b illustrates the behaviour of $f(\lambda)$. The only change is that F now extends to $-\infty$. The Algorithm Section in Appendix B discusses algorithms for computing the unique root, when one exists. The pathological situation when no root exists in F is described next. First, we note that equation (A.9) showed that the effect of including a linear term $\mathbf{g}'_{10}\mathbf{z}_{10}$ in the constraint was to add to $f(\lambda)$ the term $2\lambda\mathbf{g}'_{10}\mathbf{g}_{10}$ which, being linear in λ with a positive slope, has no essential effect on the geometry of Figures C.10a and C.10b or on solutions to $f(\lambda) = k$.

Zero values of s_i

In the above, repeated values of any γ_i cause no problem and neither do zero values. There is, however, one pathological situation that deserves consideration. This is when γ_1 and/or γ_p are associated with zero values of s_1 and/or s_p . Then, the first asymptotes occur at $\lambda = 1/\gamma_a$ and $1/\gamma_b$, say, bounding an enlarged region G : $\lambda = 1/\gamma_a \leq 0 \leq 1/\gamma_b$ containing F . We refer to those asymptotes

that disappear as *phantom asymptotes*. The region of admissible solutions remains unchanged and $f(\lambda)$ remains monotone increasing but now in the region G. There is, at most, one root λ_0 in G and if this root is also in F, it generates the desired minimum. If λ_0 is outside F, this root is a saddle point. When λ_0 does not exist, or is outside F, it may be shown that the unique minimum is given by setting/replacing λ_0 by $\lambda = 1/\gamma_1$ or $1/\gamma_p$ (see Albers *et al.*, 2007 for details). This entails setting $z_i = 0$ for all $s_i = 0$, other than z_1 or z_p , whichever is selected and whose value is to be determined. For the selected z the constraint becomes

$$f(1/\gamma) + \gamma z^2 = k, \quad (\text{B.4})$$

so determining z . If λ_0 is not in F then both $f(1/\gamma_1) - k$ and $f(1/\gamma_p) - k$ must be of the same sign. When this sign is positive, γz^2 in (B.4) must be negative, so we choose $\gamma = \gamma_1$, so estimating z_1 , else $\gamma = \gamma_p$, so estimating z_p . Thus (B.4) has a solution for either z_1 or z_p but not both.

An important special case is when $\mathbf{s} = 0$, so that all s_i vanish, giving $\gamma z^2 = k$. Thus, for the constraint to be satisfied, γ must have the same sign as k . We have arranged that k be positive so $\lambda = 1/\gamma_p$ with $z_p^2 = k/\gamma_p$ else $z_i = 0$ is the only possible solution. This case corresponds to the classical two-sided algebraic eigenvalue problem with minimum at the extremity of the shortest real principal axis.

These results derive from the properties of the normal equations $\mathbf{z} - \mathbf{s} = \lambda \boldsymbol{\Gamma} \mathbf{z}$ which now divide into two forms:

- (i) $z_i - s_i = \lambda \gamma_i z_i, s_i \neq 0$
- (ii) $z_j = \lambda \gamma_j z_j, s_j = 0$.

The first is of the kind discussed above, requiring solutions to $f(\lambda) = k$ (see below for $k = 0$). The second occurs only when $s_j = 0$ and is a simple eigenvalue expression, giving $\lambda = 1/\gamma_j$ with z_j undetermined, or $z_j = 0$ with λ undetermined. When λ

is determined from (i), it follows that $z_j = 0$ is the only possible solution to (ii), but when λ is determined from (ii), solutions for z_i may be derived from (i), choosing, if possible, z_j to satisfy the constraint (B.4). Where no admissible z_j exists, the value of λ is not in the feasible region and must be rejected. A value of λ in the feasible region always exists that either satisfies (i) or (ii) but not both. Thus, as well as the previously discussed generalisations, (i) includes elements of conventional eigenvalue problems, with which it coincides when $\mathbf{s} = 0$.

In the above, should γ_1 (or γ_p) be a repeated root, not all of whose manifestations correspond to phantom asymptotes, then the real asymptote characteristics dominate.

The case $k = 0$

Essentially, this case is already covered by our previous development. However, there are two particular instances of $k = 0$ in (1) where it is possible, and computationally more convenient, to avoid constructing the GCF.

If \mathbf{B} is p.d. and \mathbf{b} is zero, the constraint $\mathbf{x}'\mathbf{B}\mathbf{x} = 0$ is only satisfied by $\mathbf{x} = 0$ which is therefore uniquely optimal. Again, if \mathbf{B} is p.s.d. and \mathbf{b} is zero, the constraint $\mathbf{x}'\mathbf{B}\mathbf{x} = 0$ is equivalent to the linear constraint $\mathbf{B}\mathbf{x} = 0$. That is, to $\mathbf{x} = \mathbf{V}\mathbf{v}$ for some \mathbf{v} , where the columns of \mathbf{V} form a basis for the null space of \mathbf{B} . Thus, we have only to solve the generalised least squares problem of minimising $(\mathbf{V}\mathbf{v} - \mathbf{t})'\mathbf{A}(\mathbf{V}\mathbf{v} - \mathbf{t})$, giving $\hat{\mathbf{v}} = (\mathbf{V}'\mathbf{A}\mathbf{V})^{-1}\mathbf{V}'\mathbf{A}'\mathbf{t}$ and, hence, $\hat{\mathbf{x}} = \mathbf{V}\hat{\mathbf{v}}$, when \mathbf{A} is p.d. When \mathbf{B} is indefinite our solution of Section 3.2 remains valid since, although $\mathbf{B}\mathbf{x} = 0$ implies $\mathbf{x}'\mathbf{B}\mathbf{x} = 0$, there are other vectors, not in the null space of \mathbf{B} , satisfying the constraint.

Figure C.1a shows a star shape separating the region of an ellipse where there are four normals from the region where there are only two real normals. In general, the normals occur in pairs accounting for the parabolic-like branches contained in the different regions bounded by successive asymptotes of $f(\lambda)$ shown in Figure C.10. A parabolic-like branch that crosses or touches the axis in Figure C.10 gives a pair of real normals, else the paired normals are not real. Any point on the boundary of the star shaped region generates a pair of equal length normals and gives equal roots manifested by the curve $f(\lambda) = k$ touching the axis in Figure C.10. Non-parabolic branches occur at either extreme region and, crucially, in the feasible region containing the origin. This normal behaviour is shown in Figure C.11, where a transect $s_1 = s_2$ is taken through the ellipse and studied. As $s_1 = s_2$ increases, the middle “parabola” initially crosses the λ -axis, then touches it at the point corresponding to B on the star, then moves away from the axis leaving only two real roots of $f(\lambda) = 1$. On the surface of the ellipse, point D, there is an exact fit at $\lambda = 0$, the root λ being positive in the interior and negative outside the ellipse. All roots are in the feasible region. If we plotted λ against $s_1 = s_2$ the curve would be smooth. The function is well-behaved but we have seen in the algebraic treatment that when $s_k = 0$ except for s_1 , the behaviour of the optimal value of z_1 is more complicated. Figure C.12 shows what happens as one traverses the major axis from the origin outwards. Figure C.12a shows three versions of $f(\lambda)$, one giving a root in the feasible region, one not, and one with the root at the boundary of the feasible region. In fact for the constraint $\gamma_1 s_1^2 + \gamma_2 s_2^2 = 1$ we have that:

$$f(\lambda) = \frac{\gamma_1 s_1^2}{(1 - \lambda \gamma_1)^2} = 1$$

so that $f(\lambda)$ increases linearly with s_1^2 . When $\lambda \leq 1/\gamma_2$, $f(\lambda) = 1$ has a root in the feasible region, i.e. when $s_1 \geq \delta = \sqrt{\gamma_1}(\frac{1}{\gamma_1} - \frac{1}{\gamma_2})$, the cusp of the star shaped region on the major axis. Figure C.12b

shows how the root λ changes with s_1 , being constant at $1/\gamma_2$ until the cusp δ is reached, after which it decreases linearly. However, the primary interest is in the fitted values of z_1 and z_2 . These are shown in Figure C.12c where z_2 falls elliptically until the cusp, after which it becomes zero (with $z_1 = \sqrt{1/\gamma_1}$) which is at the end of the major axis. Rather counter-intuitively, the initial region for constant λ corresponds to the region of the ellipse where there are equal pairs of shortest normals from s_1 to the perimeter of the ellipse, while with variable λ after the cusp, including exterior points on the major axis, the shortest normals all end at the end of the major axis. This, and similar pathological cases are all included in the analytical treatment and the algorithm derived from it. Animations of the behaviour of the FCE are available at <http://mcs.open.ac.uk/cja235/quadratic>.

An example of the geometry of $k = 0$ and \mathbf{B} of full rank with phantom asymptotes, is given in Figure C.9 (see Section 3.2).

Algorithmic issues

Algorithms have been published for special cases of (1). We have already discussed the method by Ten Berge (1983, see Section 3.2) and Browne's (1967) solution to the oblique Procrustes problem, see also Eldén (2002) .

Our objective is to provide a general algorithm that subsumes all the variants of (1) discussed above, those already published, and others that may arise in the future. The basic approach is to derive the GCF and solve the FCE to give the root λ of $f(\lambda) = k$, thus obtaining a minimum of (1). The algebraic basis for writing a general algorithm is provided in the above and in Appendix A.

Computational difficulties.

Various numerical difficulties can occur, due to e.g. very large/small numbers and near-singular matrices. For instance, Eldén (2002) considered the case where \mathbf{A} has a huge condition number (the

condition number of a matrix is the ratio of largest and the lowest eigenvalue, in absolute value). When this condition number is larger than, say, 10^{16} , computational difficulties can arise. In statistical practice this, however, is uncommon, but in our algorithm a warning will be given when the condition number lies above a certain threshold. Also when similar numerical problems (nearly) occur, warnings will be provided.

Finding the root of the Lagrangian.

The root λ in (B.3) has to be found via numerical methods (at least when $p > 3$). A simple bisection method is preferred above the Newton–Raphson method because the latter has the potential to converge to the wrong branch of $f(\lambda)$ due to the flatness of $f(\lambda)$ for small values of λ (cf. Cramer, 1974). Experience after using the bisection algorithm for various examples indicates that usually a root is found within a second.

For the p.s.d. case (see Figure C.10), a lower bound is not directly available. Gower and Dijksterhuis (2004) provide a lower bound to the feasible region, allowing the bisection method to be initiated.

The algorithm, designed for *R* (<http://www.r-project.org>), is available from <http://mcs.open.ac.uk/cja235/quadratic>. Full details of the algorithm are given in Albers (2007).

C Projection of a quadratic onto a subspace

Theorem. The projection of:

$$Q : \begin{pmatrix} \mathbf{x}'_1 & \mathbf{x}'_2 \end{pmatrix} \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = 1 \quad (\text{C.1})$$

onto the space spanned by \mathbf{x}_1 is given by

$$P : \mathbf{x}'_1 (\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}) \mathbf{x}_1 = 1,$$

whenever \mathbf{C}_{22} is non-singular.

Proof. The projection is defined by the points of Q that have normals in the \mathbf{x}_1 -space. The projection of $(\mathbf{x}'_1, \mathbf{x}'_2)'$ at this point of Q is \mathbf{x}_1 . The normal at $(\mathbf{x}'_1, \mathbf{x}'_2)'$ to Q is proportional to the vector:

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}.$$

The component of this vector that is orthogonal to the the \mathbf{x}_1 -space must vanish, i.e.

$$\mathbf{C}_{21}\mathbf{x}_1 + \mathbf{C}_{22}\mathbf{x}_2 = 0. \quad (\text{C.2})$$

Thus, from (C.1) and (C.2):

$$\mathbf{x}'_1 \mathbf{C}_{11} \mathbf{x}_1 - 2\mathbf{x}'_1 \mathbf{C}_{12} (\mathbf{C}_{22}^{-1} \mathbf{C}_{21} \mathbf{x}_1) + (\mathbf{x}'_1 \mathbf{C}_{12} \mathbf{C}_{22}^{-1}) \mathbf{C}_{22} (\mathbf{C}_{22}^{-1} \mathbf{C}_{21} \mathbf{x}_1) = 1$$

i.e.

$$\mathbf{x}'_1 (\mathbf{C}_{11} - \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \mathbf{C}_{21}) \mathbf{x}_1 = 1$$

which is P , as was to be shown. \square

Remark. In the above, we have arranged that the subspace concerned is given by the first set of variables defining Q . Any other subspace could by accommodated by first orthogonally transforming it to the leading position and using the inverse transformation to return to the original parameterisation.

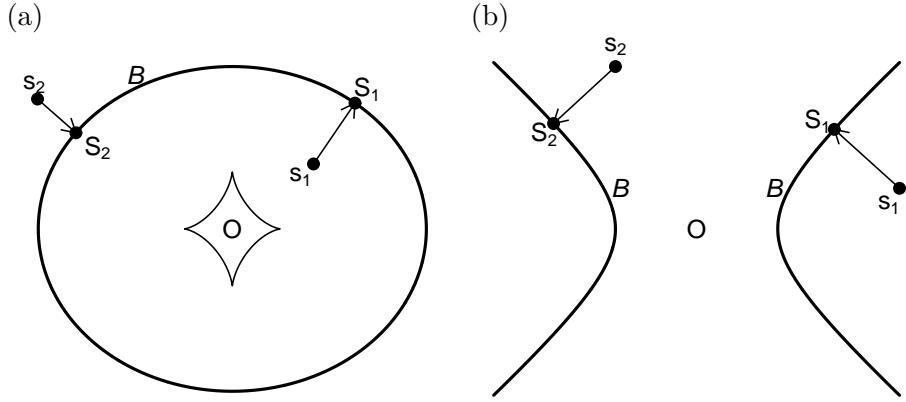


Fig. C.1. In C.1a, B is elliptical and \mathbf{B} is positive definite. In C.1b, B is hyperbolic and \mathbf{B} is indefinite. Shortest normals are indicated for internal (s_1) and external (s_2) settings of \mathbf{s} . The arrows at the end of the normal indicate the constrained solutions \mathbf{z} . Points inside the star-shaped region have four real normals to B , those outside have two.

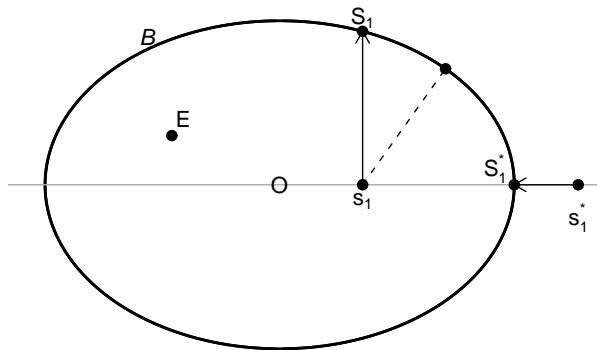


Fig. C.2. The target space is one-dimensional (the major axis of the ellipse). When \mathbf{s} is inside there is an exact solution S that projects onto \mathbf{s} . Although S itself is two-dimensional, only its first coordinate is relevant. E is inside B so it gives an exact solution when the constraint is $\mathbf{x}'\mathbf{B}\mathbf{x} \leq k$.

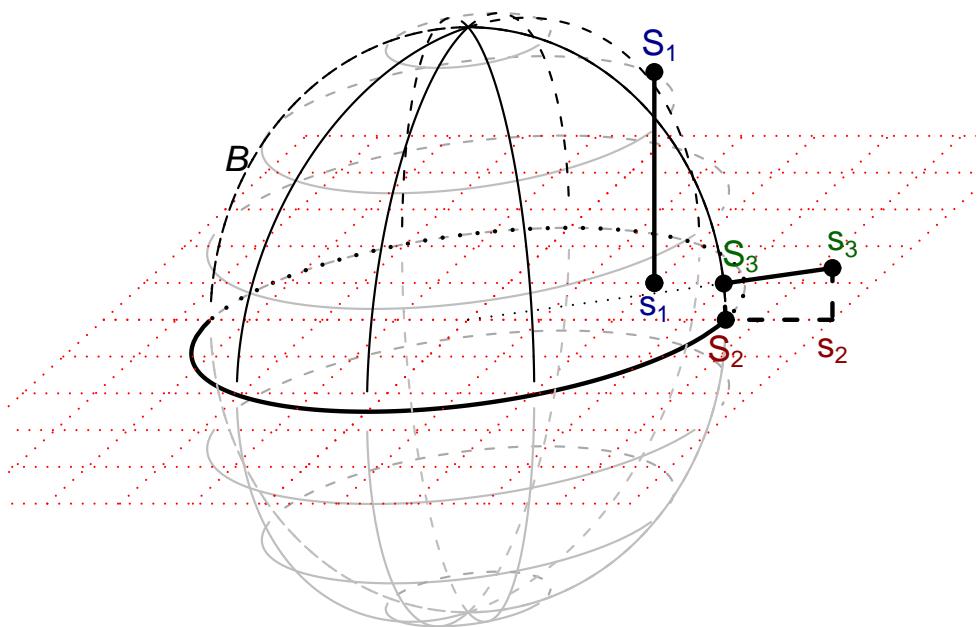


Fig. C.3. B is three-dimensional while the target space containing \mathbf{s} is two-dimensional (shown as the grid-plane). When \mathbf{s} is inside B there is an exact solution. When \mathbf{s} is outside we require the shortest normal to the projection of B onto the target space. S_2 and s_2 are in a one-dimensional target space, S_3 and s_3 are in two dimensions.

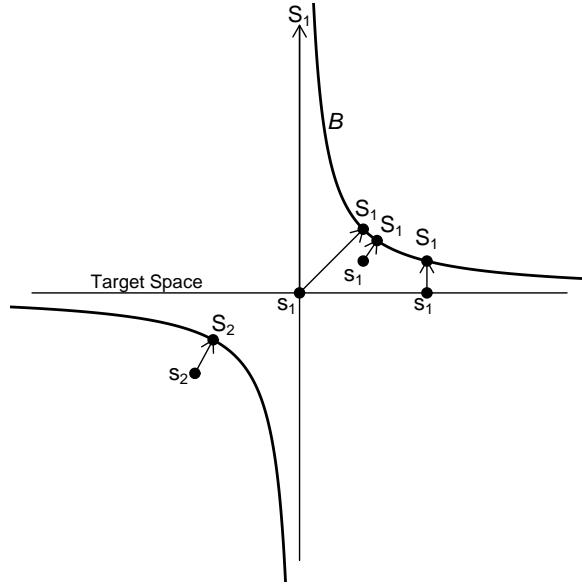


Fig. C.4. A rectangular hyperbola of the form $z_1 z_2 = c^2$ not referred to principal axes. Assorted two-dimensional solutions are shown together with exact solutions when the target space is restricted to the horizontal axis. The particular one-dimensional solution for $s = 0$ only gives an asymptotic “exact” solution. The diagram does not correspond to the GCF for two-dimensional solutions but does for the one-dimensional solution.

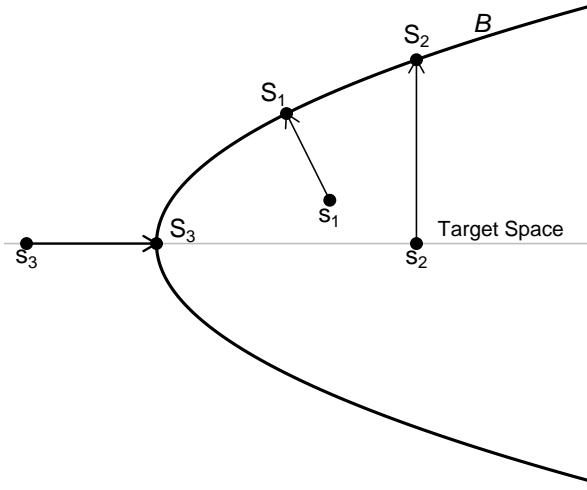


Fig. C.5. B is the parabola $z_2^2 = 4cz_1$ containing a linear term. The shortest normal remains well-defined (as for s_1) and when there are extraneous variables with z confined to the horizontal axis (as for s_2), exact solutions exist. When s is outside the parabola shortest normals remain available but when extraneous and negative only the non-exact solution at the origin O (as for s_3) is available. The linear term has had no substantive effect.

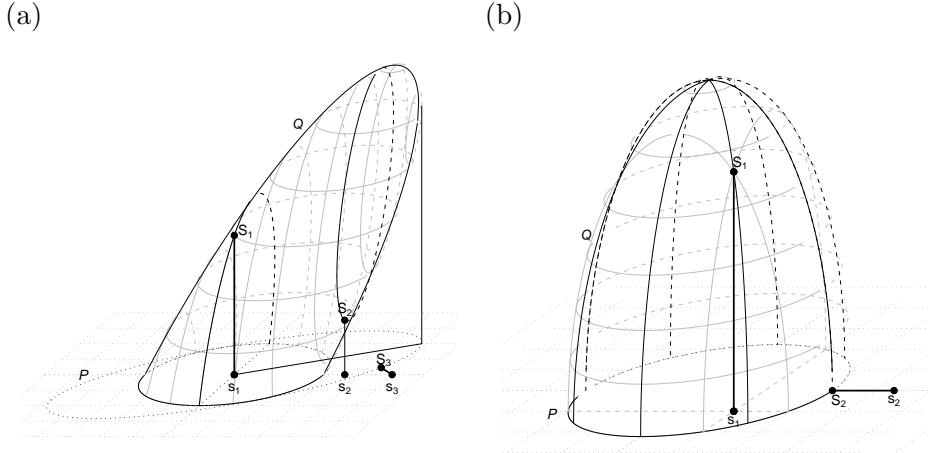


Fig. C.6. The projection P of Q onto a subspace. The large dotted ellipse indicates the projection P and the smaller one the intersection of Q with a two-dimensional subspace. The algebraic form of P is given by (A.2) of Appendix A. In C.6a, if the vertical axis represents an extraneous variable, points s_1 and s_2 indicate exact solutions when \mathbf{s} lies within the intersection, though possibly outside Q itself (s_2). The point s_3 in the solution subspace is outside the intersection and its solution is given by the shortest normal onto P . In C.6b is the case when \mathbf{C}_{12} vanishes, as when \mathbf{B} is p.s.d. Now Q is normal to the subspace P .

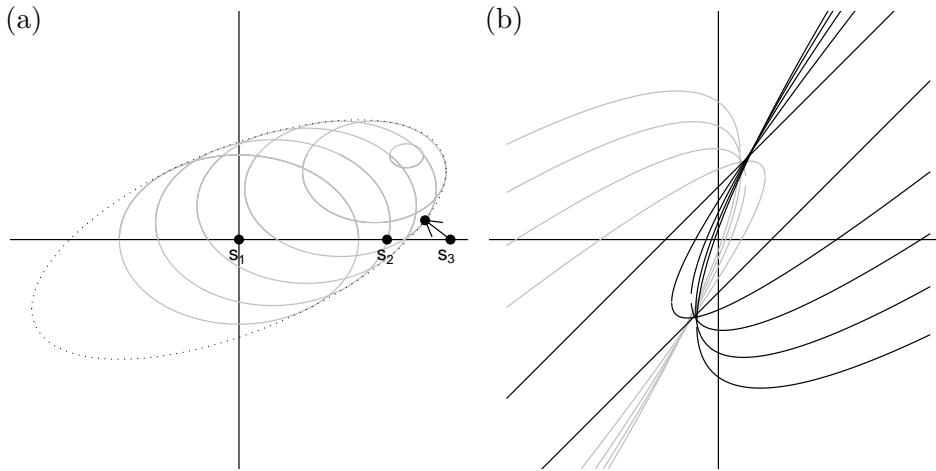


Fig. C.7. Figure C.7a is a contour version of C.6a for z_3 positive (to include z_3 negative would entail confusing overlap). Figure C.7b is a contour version when Q is hyperbolic. The darker contours are for z_3 positive and the lighter contours for z_3 negative.

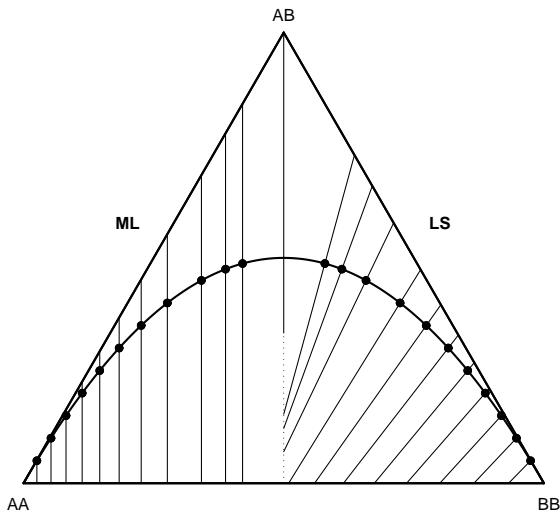


Fig. C.8. Contours of same equilibrium prediction. Left-hand-side: Maximum Likelihood contours, perpendicular to base-line. Right-hand-side: Least squares contours, normal to the equilibrium parabola.

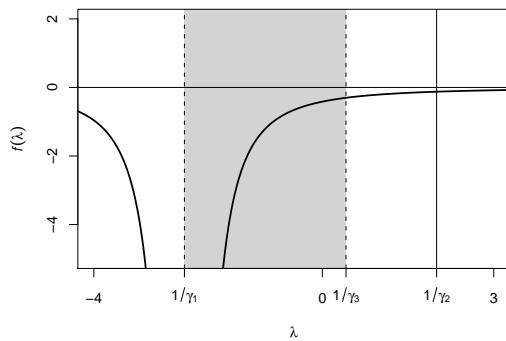


Fig. C.9. Lagrangian $f(\lambda)$ for the example in Section 3.2 with zero values of s_i . There are phantom asymptotes at $1/\gamma_2$ and at $1/\gamma_3$, the upper bound of the feasible region, so the minimum occurs at $\lambda = 1/\gamma_3$.

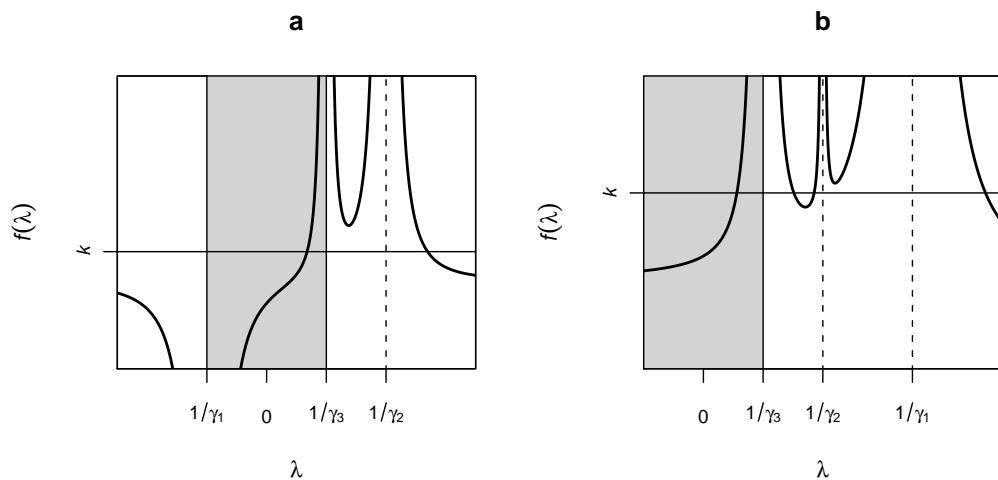


Fig. C.10. Figure C.10a: The form of $f(\lambda)$ when \mathbf{B} is indefinite. Asymptotes occur for positive and negative values of λ . In the shaded region F , $f(\lambda)$ is monotone increasing so contains a unique root. Figure C.10b: The form of $f(\lambda)$ when \mathbf{B} is p.s.d. All asymptotes are for positive λ and F stretches to $-\infty$.

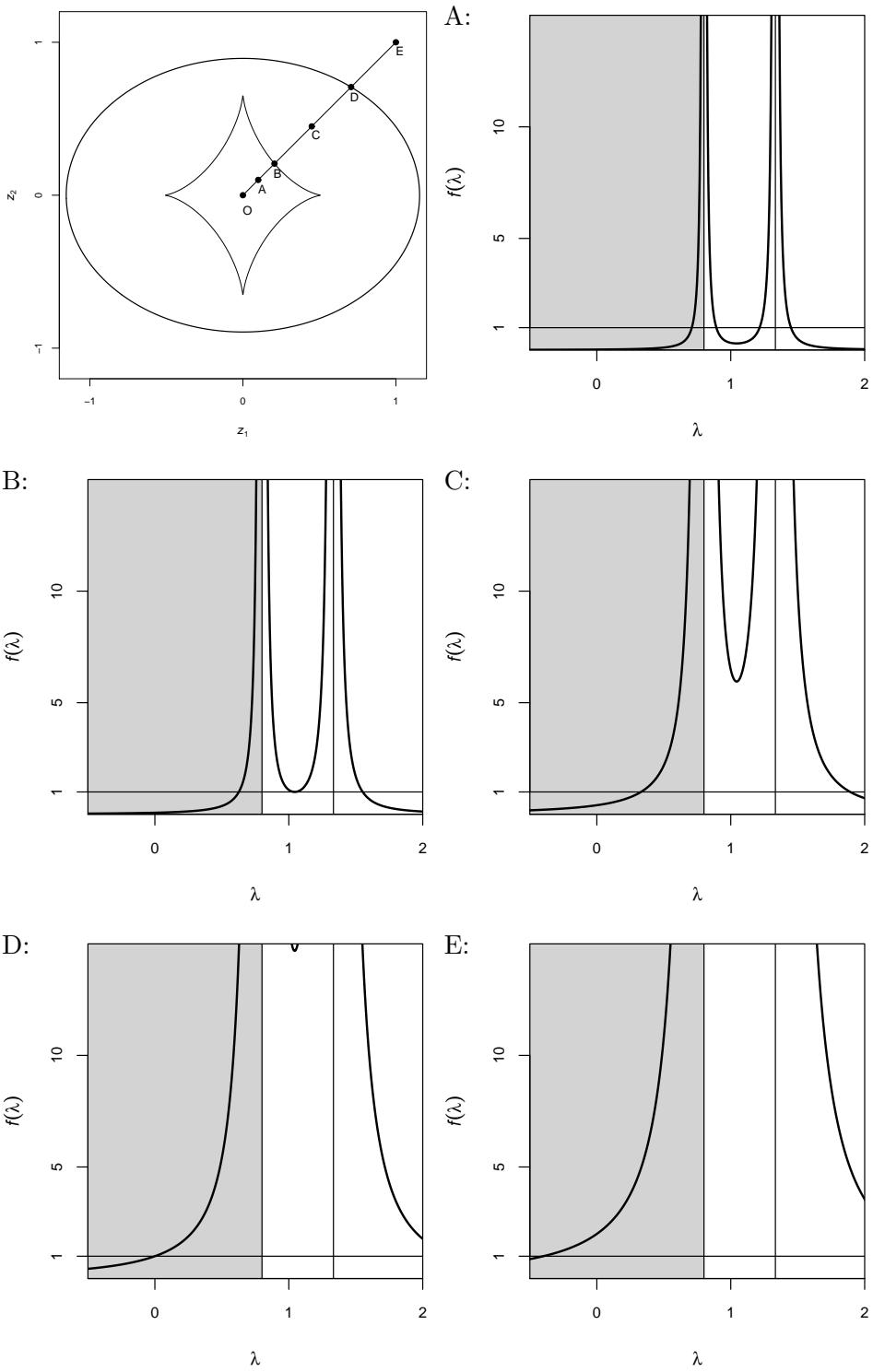


Fig. C.11. Geometrical (top-left) and Lagrangian (other plots) view of the two-dimensional elliptic case, exemplified via the ellipse $\frac{3}{4}z_1^2 + \frac{5}{4}z_2^2 = 1$. The star shape in the top-left graph depicts the boundary between two and four normals.

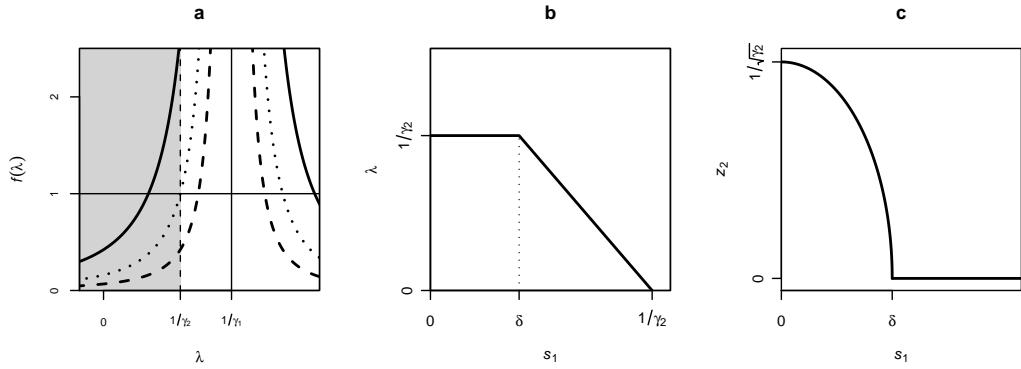


Fig. C.12. Figure C.12a shows $f(\lambda)$ for different choices of $\mathbf{s} = (s_1, 0)$. The solid, dotted and dashed line correspond to $|s_1|$ greater than, equal to, or smaller than $\delta = \sqrt{\gamma_1}(1/\gamma_1 - 1/\gamma_2)$, respectively. There is a phantom asymptote at $\lambda = 1/\gamma_2$. Figure C.12b displays the relation between λ and s_1 : the root of (1) is obtained at $\lambda = 1/\gamma_2$ for $s_1 < \delta$, after which is linearly decreases. Figure C.12c displays the relation between z_2 and s_1 .

References

- [1] C.J. Albers, *An algorithm for quadratic minimisation problems in statistics (working title)*, In preparation, 2007.
- [2] C.J. Albers, F. Critchley, and J.C. Gower, *Explicit minimisation of a convex quadratic under a general quadratic constraint*, Submitted for publication, 2007.
- [3] J.M.F. ten Berge, *A generalization of Verhelst's solution for a constrained regression problem in ALSCAL and related MDS-algorithms*, Psychometrika **48** (1983), no. 4, 631–638.
- [4] J.M.F. ten Berge and K. Nevels, *A general solution for Mosier's oblique Procrustes problem*, Psychometrika **42** (1977), 593–600.
- [5] M.W. Browne, *On oblique Procrustes rotation*, Psychometrika **32** (1967), 125–132.
- [6] E.M. Cramer, *On Browne's solution for oblique Procrustes rotation*, Psychometrika **39** (1974), 139–163.
- [7] F. Critchley, *On the minimisation of a positive definite quadratic form under a quadratic constraint: analytical solution and statistical applications*, Technical Report, Department of Statistics, University of Warwick, 1990.
- [8] F. Critchley, P. Marriott, and M.H. Salmon, *On the differential geometry of the Wald test with nonlinear restrictions*, Econometrica **64** (1996), 1213–1222.
- [9] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika **1** (1936), 211–218.
- [10] L. Eldén, *Solving quadratically constrained least squares problems using a differential-geometric approach*, BIT **42** (2002), 323–335.
- [11] W. Gander, *Least squares with a quadratic constraint*, Numerische Mathematik **36** (1981), 291–307.
- [12] J.C. Gower, *The role of constraints in determining optimal scores*, Statistics in Medicine **17** (1998), no. 23, 2709–2721.
- [13] J.C. Gower and G.B. Dijksterhuis, *Procrustes problems*, Oxford Statistical Science Series, vol. 30, Oxford University Press, 2004.
- [14] J.C. Gower and D.J. Hand, *Biplots*, Monographs on Statistics and Applied Probability, vol. 54, London: Chapman & Hall, 1995.
- [15] P.J. Green and B.W. Silverman, *Nonparametric regression and generalized linear models - a roughness penalty approach*, Monographs on Statistics and Applied Probability, vol. 58, London: Chapman & Hall, 1994.
- [16] A.W. Gregory and M.R. Veall, *Formulating Wald test of nonlinear restrictions*, Econometrica **53** (1985), no. 6, 1465–1468.

- [17] M.J.R. Healy and H. Goldstein, *An approach to the scaling of categorical attributes*, Biometrika **63** (1976), 219–229.
- [18] H.A.L. Kiers, *Maximization of sums of quotients of quadratic forms and some generalizations*, Psychometrika **60** (1995), 221–245.
- [19] W.J. Krzanowski, P. Jonathan, W.V. McCarthy, and M.R. Thomas, *Discriminant analysis with singular covariance matrices: methods and applications to spectroscopic data*, Journal of the Royal Statistical Society, Series C **44** (1995), 101–115.
- [20] F. Lafontaine and K.J. White, *Obtaining any Wald statistic you want*, Economics Letters **21** (1986), no. 1, 35–40.
- [21] J. de Leeuw, *Generalized eigenvalue problems with positive semidefinite matrices*, Psychometrika **47** (1982), 87–93.
- [22] R.P. McDonald, Y. Torii, and S. Nishisato, *Some results on proper eigenvalues and eigenvectors with applications to scaling*, Psychometrika **44** (1979), 211–227.
- [23] J.J. Moré, *Generalizations of the trust region problem*, Optimization methods and Software, Volume II (1993), no. 189–209.
- [24] R.W. Newcomb, *On the simultaneous diagonalization of two semi-definite matrices*, Quarterly Journal of Applied Mathematics **19** (1961), 144–146.
- [25] C.R. Rao and S.K. Mitra, *Generalized inverse of matrices and its applications*, New York: John Wiley & Sons, 1971.
- [26] C. Reinsch, *Smoothing by spline functions*, Numerische Mathematik **10** (1967), 177–183.
- [27] Y. Takane, F.W. Young, and J. de Leeuw, *Nonmetric individual differences multidimensional scaling: An alternating least-squares method with optimal scaling features*, Psychometrika **42** (1977), 7–67.
- [28] H. Tuy and N.T. Hoai-Phuong, *A robust algorithm for quadratic optimization under quadratic constraints*, Journal of global optimization **37** (2007), 557–569.
- [29] F. Uhlig, *A recurring theorem about pairs of quadratic forms and extensions: a survey*, Linear Algebra and Applications **25** (1979), 219–237.