Some quadratic optimisation problems in psychometrics

Casper J. Albers
Department of Statistics, The Open University, Walton Hall, Milton Keynes MK6 7AA, United Kingdom, `c.j.albers@open.ac.uk`

Abstract

In this paper, I discuss various examples arising from different areas in psychometrics. I will show that they have a common background and can be solved using a constrained quadratic optimisation algorithm developed by Albers *et al.* (2008a,b).

1. Introduction

In this paper, I discuss some examples arising from different areas in psychometrics. Although the three problems are from different backgrounds and are quite different in their form, nevertheless they all require the solution to similar constrained optimisation problems. For all problems, both objective function and constraint are quadratic functions, and the matrix in the objective function is non-negative definite. Consider, for instance, Oblique Procrustes analysis (Browne , 1967; Cramer , 1974), where one aims to minimise

$$\min_{\mathbf{t}} ||\mathbf{F}\mathbf{t} - \boldsymbol{\varphi}||^2 \quad \text{subject to } \mathbf{t}'\mathbf{t} = 1.$$

Here $\mathbf{F}$ is a factor matrix (not necessary of full rank), $\boldsymbol{\varphi}$ a vector from a specified factor structure, and $\mathbf{t}$ a vector from the transformation matrix $\mathbf{T}$ required to minimise the least squares distance between $\mathbf{FT}$ and the factor structure. The decomposition $\mathbf{F}'\mathbf{F} = \mathbf{UCU}'$ (where $\mathbf{C}$ is diagonal with non-negative elements) can be used to restate the oblique Procrustes problem as follows (cf. Ten Berge and Nevels , 1977)

$$\left.\begin{array}{c} \min_{\mathbf{w}} \left(\mathbf{w} - \mathbf{C}^{-1}\mathbf{x}\right)' \mathbf{C} \left(\mathbf{w} - \mathbf{C}^{-1}\mathbf{x}\right) + \text{constant} \\ \text{subject to } \mathbf{w}'\mathbf{w} = 1, \end{array}\right\} \qquad (1)$$

where $\mathbf{w} = \mathbf{U}'\mathbf{t}$ and $\mathbf{x} = \mathbf{U}'\mathbf{F}'\boldsymbol{\varphi}$. In its rewritten state, the problem is a optimisation problem where both objective function and constraint are in quadratic form. Furthermore, the objective function is convex, since $\mathbf{C}$ is a positive (semi-)definite (p(s)d) matrix.

This type of problems is quite common, as I shall show with three examples. Examples 1 and 2 are concerned with Procrustes analysis: the first dealing with missing value estimation and the second with a Procrustes problem associated with the Parafac-model. Example 3, having its roots in multidimensional scaling, discusses a constrained regression problem in the ALSCAL-algorithm. After recognising the common structure, I will apply the methodology of Albers *et al.* (2008a,b) to solve Example 3, and to outline how to solve such problems in general. The examples will be discussed in order of increasing complexity.

2. Example 1: Missing values in Generalised Procrustes Analysis

In Generalised Procrustes Analysis (GPA; Gower , 1975; Gower and Dijksterhuis , 2004) one is, given matrices $\mathbf{X}_k, k = 1, \ldots, K$, and one seeks matrices $\mathbf{T}_k$ such that the K transformations $\mathbf{X}_k\mathbf{T}_k$ lie as close to each other as possible, in a least squares sense. Thus, we require $\mathbf{T}_k$ that minimise

$$\sum_{i<j}^{K} ||\mathbf{X}_i\mathbf{T}_i - \mathbf{X}_j\mathbf{T}_j||^2 .$$

Equivalently, one can minimise

$$\sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{T}_k - \mathbf{G}||^2 \quad \text{or} \quad \sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{T}_k - \mathbf{G}_k||^2$$

where $\mathbf{G} = K^{-1}\sum_{k=1}^{K}\mathbf{X}_k\mathbf{T}_k$ is the group average matrix and $\mathbf{G}_k = (K-1)^{-1}((\sum_{j=1}^{K}\mathbf{X}_j\mathbf{T}_j) - \mathbf{X}_k\mathbf{T}_k)$ is the k-excluded group average matrix. The matrices $\mathbf{T}_k$ are required to be members of some well-defined type, classically orthogonal matrices ($\mathbf{T}_k'\mathbf{T}_k = \mathbf{I}$), but here we proceed generally. Using an iterative algorithm, one can find the optimal $\mathbf{T}_k$.

If data are missing, which is not uncommon in psychometric analyses, straightforward utilisation of the algorithm is not possible, and adaptations are necessary. For certain special cases of GPAs the solutions are available in the literature. E.g. Ten Berge *et al.* (1993) present full solutions when the $\mathbf{T}_k$ are required to be orthogonal.

Albers and Gower (in preparation) present a general method for missing value estimation that allows the $\mathbf{T}_k$ to have any structure. After each iterative step in the generalised Procrustes problem algorithm, a step to estimate missing values is introduced. One creates a matrix $\mathbf{X}$ whose elements are zero, except for the elements corresponding to the missing values in $\mathbf{X}_k$. Then, the aim is to estimate the missing values by optimising

$$\min_{\mathbf{X}} ||(\mathbf{I} - \mathbf{1}'\mathbf{1}/n)(\mathbf{X}_k - \mathbf{X})\mathbf{T}_k - \mathbf{G}_k||^2$$

(the $(\mathbf{I}-\mathbf{1}'\mathbf{1}/n)$ part is added to preserve centring). With some classes of $\mathbf{T}_k$ there is a trivial solution $\mathbf{T}_k = 0$ unless some additional constraint in size is imposed. Fore example, we may require that the updated matrices $\mathbf{X}_k$ remain of constant size,

$$\text{tr } (\mathbf{X}_k - \mathbf{X})'(\mathbf{I} - \mathbf{1}'\mathbf{1}/n)(\mathbf{X}_k - \mathbf{X}) = \text{tr } \mathbf{X}_k'\mathbf{X}_k.$$

The resulting problem, is of the form discussed in this paper. The quadratic form is rather easily obtained after applying the vec-operation. The constraint, which is in trace-form, can be rewritten as a quadratic vector-constraint similarly to what I will do in Example 2. Albers and Gower (in preparation) prove that the quadratic objective function thus obtained is indeed convex.

3. Example 2: a Procrustes approach to Parafac

In his discussion of asymmetry models, (Gower , 2008, this volume) discusses a 2×2 Parafac-model (cf. Harshman and Lundy , 1984). Gower requires the solution to the following Procrustes problem (with adapted notation):

$$\min_{\mathbf{T}} ||\mathbf{X}_1\mathbf{T} - \mathbf{X}_2||^2$$
$$\text{subject to det } \mathbf{T} = 1$$

where $\mathbf{T} = \begin{pmatrix} t_1 & t_2 \\ t_3 & t_4 \end{pmatrix}$ and, hence, det $\mathbf{T} = t_1 t_4 - t_2 t_3$.

As in the previous example, it is possible to recognise a quadratic form in the objective function and constraint. For the constraint det $\mathbf{T} = 1$ this is directly clear after writing det $\mathbf{T} = \mathbf{t}'\mathbf{B}\mathbf{t} = 1$ with $\mathbf{t} = \text{vec}\mathbf{T}$ and

$$\mathbf{B} = \frac{1}{2}\begin{pmatrix} & & & 1 \\ & & -1 & \\ & -1 & & \\ 1 & & & \end{pmatrix}.$$

The objective function $||\mathbf{X}_1\mathbf{T} - \mathbf{X}_2||^2$ is quadratic in $\mathbf{T}$ and thus one can easily determine some $\mathbf{A}$, $\mathbf{b}$, and c such that it can be written as

$$\text{tr } ||\mathbf{X}_1\mathbf{T} - \mathbf{X}_2||^2 = \mathbf{t}'\mathbf{A}\mathbf{t} - 2\mathbf{b}'\mathbf{t} + c.$$

Hence,

$$\text{tr } ||\mathbf{X}_1\mathbf{T} - \mathbf{X}_2||^2 = \left(\mathbf{t} - \mathbf{A}^{-1}\mathbf{b}\right)' \mathbf{A} \left(\mathbf{t} - \mathbf{A}^{-1}\mathbf{b}\right) + \textit{const.}$$

This problem is exactly similar to that of Example 1, provided the objective function is convex. $\mathbf{A}$ is obtained from $\mathbf{t}'\text{tr}(\mathbf{T}'\mathbf{X}_1'\mathbf{X}_1\mathbf{T})\ \mathbf{t} = \mathbf{t}'\mathbf{A}\mathbf{t}$ with

$$\mathbf{A} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_1'\mathbf{X}_1 \end{pmatrix}$$

which is p(s)d, since $\mathbf{X}_1'\mathbf{X}_1$ is p(s)d because of symmetry.

Thus, again, this problem can be rewritten in quadratic form with convex objective function, this time with an indefinite constraint (i.e. for general $\mathbf{x}$, $-\infty < \mathbf{x}'\mathbf{B}\mathbf{x} < \infty$).

4. Example 3: Indefinitely constrained multidimensional scaling

The example in this section has its origins in the ALSCAL (Takane *et al.* , 1977) algorithm. ALSCAL 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.

Generalising Verhelst (1981), who was studying a small part of this algorithm in detail, Ten Berge (1983) noted that the ALSCAL algorithm requires the solution to a constrained regression problem and gave an explicit solution.

In the notation of this paper, Ten Berge's specification requires

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

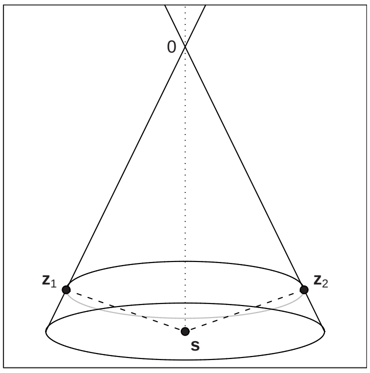

Fig. 1 Geometry of the Fundamental Canonical Form of Example 3.

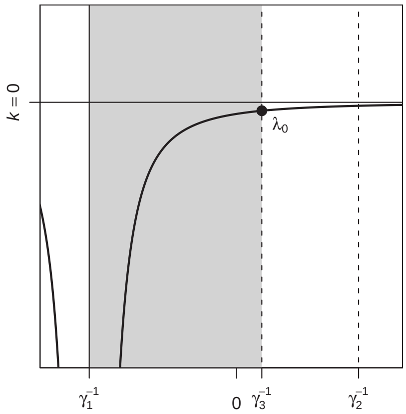

Fig. 2 Fundamental Canonical Equation of Example 3.

where

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

At first sight, this problem doesn't seem very remarkable. However, on rewriting it, it turns out to be more interesting.

We can rewrite the objective function in a form, similar to that of Examples 1 and 2, as follows:

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

where $\mathbf{A} = \mathbf{K}'\mathbf{K}$, which is pd, and $\mathbf{t} = (\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{d} = (1.8014, -1.5308, .3827)'$.

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 $(-2, 1, 2)$. As in Example 2, there is an indefinite constraint.

5. Details of solution to Example 3

Essentially, it was possible to reparameterise all three examples in the following form:

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

with $\mathbf{A}$ either a positive definite or positive semi-definite matrix and $\mathbf{B}$ without any restrictions. Accidentally, in all three examples, $\mathbf{b} = 0$ and, except for Example 2, $k = 0$.

Albers *et al.* (2008a,b) provide a general unifying framework to solve all problems of type (2). I shall analyse Example 3 using this framework and, in so doing, explain the methodology.

In this example $\mathbf{A}$ is pd and $\mathbf{B}$ is of full rank. Thus, there exists a matrix $\mathbf{T}$ that simultaneously diagonalises the matrices, in this case

$$\mathbf{T} = \begin{pmatrix} 3.04 & 1.41 & -1.80 \\ -3.70 & -.71 & 1.53 \\ .92 & 0 & -.38 \end{pmatrix},$$

giving $\mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{I}$ and $\mathbf{T}'\mathbf{B}\mathbf{T} = \mathbf{\Gamma} = \text{diag}(1+\sqrt{2}, \frac{1}{2}, 1-\sqrt{2})$.

After computing $\mathbf{s} = \mathbf{T}^{-1}\mathbf{t} = (0, 0, -1)$, we can study the transformation $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x} + \mathbf{m}$ to the problem, instead of the original problem (2). This gives, for Example 3, what Albers *et al.* (2008a) call the General Canonical Form

$$\left.\begin{array}{r} \min_{\mathbf{z}} ||\mathbf{z} - \mathbf{s}||^2 \\ \text{subject to } \mathbf{z}'\mathbf{\Gamma}\mathbf{z} = k = 0. \end{array}\right\} \tag{3}$$

(When $\mathbf{A}$ or $\mathbf{B}$ is not of full rank, or when $\mathbf{b} \neq 0$, computation is somewhat more elaborate.)

In this canonical form, a special structure of $\mathbf{B}$ and $\mathbf{s}$ suddenly becomes clear, see also Figure 1: B, the elliptical double-cone corresponding to the constraint, takes the shape of an 'hour glass'. Note that Verhelst (1981) worked with a general vector $\mathbf{s}$ not on a principal axis of B. Because here $\mathbf{s}$ is on a principal axis, the problem has two solutions: the symmetry of B ensures that, if (z_1, z_2, z_3) is a solution minimising $||\mathbf{z} - \mathbf{s}||^2$, then so is $(-z_1, z_2, z_3)$. Note: Ten Berge reported his $\mathbf{d}$ up to four digits; the exact values in terms of $\sqrt{2}$ in $\mathbf{\Gamma}$ and $\mathbf{s}$ above are after 'expert intervention': although the values found for $\mathbf{\Gamma}$ by the algorithm differ by less than 10^{-13} from the 'algebraic interpretation', and the values in $\mathbf{s}$ differ by less than 10^{-5} from $(0, 0, -1)$, it does affect the analysis considerably: $\mathbf{s}$ is no longer on a principal axis, so symmetry is lost and only one solution, instead of two, will be found.

A solution to (3) is found via the Lagrangian $||\mathbf{z} - \mathbf{s}||^2 - \lambda(\mathbf{z}'\mathbf{\Gamma}\mathbf{z})$, which yields the 'Fundamental Canonical Equation'

$$f(\lambda) = \sum_{i=1}^{3} \frac{\gamma_i s_i^2}{(1 - \lambda\gamma_i)^2} = 0,$$

where $1/\gamma_i$ are the diagonal eigenvalues of $\mathbf{\Gamma}$ in ascending order, namely $-(\sqrt{2}+1), \sqrt{2}-1$, and 2. Usually, $f(\lambda)$ has asymptotes at $\lambda = 1/\gamma$, however, since the first two elements of $\mathbf{s}$ are zero, the asymptotes, corresponding to γ_2 and γ_3, vanish. The origin is contained in the first two elements, hence a solution to the Fundamental Canonical Equation is found in $[-(\sqrt{2}+1), \sqrt{2}-1]$, the Feasible Region (Albers *et al.* , 2008a). Because of the zero values in $\mathbf{s}$, $f(\lambda) = 0$ reduces to $f(\lambda) = (-\sqrt{2}-1)(\sqrt{2}+1+\lambda)^{-2}$ and has no solution within the boundary of the Feasible Region, see Figure 2. Then the solution lies on the boundary, giving $\lambda_0 = \sqrt{2} - 1$.

From $\lambda_0 = \sqrt{2} - 1$, we can compute our solutions. First, $\mathbf{z}$ is computed according to $\mathbf{z} = (\mathbf{I} - \lambda_0 \mathbf{\Gamma})^{-1}\mathbf{s}$, yielding $\mathbf{z} = (\pm 8^{-1/2}, 0, -\frac{1}{4}(2 + \sqrt{2}))' = (\pm.3528, 0, -.8536)'$. Indeed, there are two solutions. Secondly, the original parameters are given by $\mathbf{x} = \mathbf{T}^{-1}\mathbf{z} = (2.6132, -2.6132, .6533)'$ and $\mathbf{x} = (.467, 0, 0)'$, in agreement with Ten Berge (1983). It is not evident from solutions $\mathbf{x}$ that they are the same, up to sign, in the Fundamental Canonical Form.

Acknowledgements.
John Gower suggested Example 2. Both John Gower and Frank Critchley critically read versions of this paper.

References

Albers, C. J. & Gower, J. C. (2008). Missing value estimation in generalised Procrustes problems. *(In preparation).*

Albers, C. J., Critchley, F. & Gower, J. C. (2008a). Quadratic minimisation problems in statistics. *Submitted for publication, preprint available from http://mcs.open.ac.uk/cja235.*

Albers, C. J., Critchley, F. & Gower, J. C. (2008b). Explicit minimisation of a convex quadratic under a general quadratic constraint. *Submitted for publication, preprint available from http://mcs.open.ac.uk/cja235.*

Ten Berge, J. M. F. (1977). Orthogonal Procrustes rotation for two or more matrices. *Psychometrika, 42:2*, 267–276.

Ten Berge, J. M. F. (1983). A generalization of Verhelst's solution for a constrained regression problem in ALSCAL and related MDS–algorithms. *Psychometrika, 48:4*, 631–638.

Ten Berge, J. M. F., Kiers, H. A. L. & Commandeur, J. J. F. (1993). Orthogonal Procrustes rotation for matrices with missing values. *British Journal of Mathematical and Statistical Psychology, 46*, 119–134.

Ten Berge, J. M. F. & Nevels, K. (1977) A general solution to Mosier's oblique Procrustes problem. *Psychometrika, 42:4*, 593–600.

Browne, M. W. (1967). On oblique Procrustes rotation. *Psychometrika, 32*, 125–132.

Cramer, E. M. (1974). On Browne's solution for oblique Procrustes rotation. *Psychometrika, 39*, 159-163.

Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes Analysis*, Oxford: Oxford University Press.

Gower, J. C. (1975). Generalized Procrustes problems. *Psychometrika, 40:1*, 33–51

Gower, J. C. (2008). Asymmetry Analysis: The Place of Models. In: *New Trends In Psychometrics, this volume.* Tokyo: Universal Academy Press

Harshman, R. A. & Lundy, M. A. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In: Law, H. G. Snyder, C. W. , Hattie, J. A. & McDonald, R. P. *Research methods for multimode data analysis*, New York: Praeger, pp. 122–215.

Takane, Y., Young, F. W. & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least-squares method with optimal scaling features *Psychometrika, 42*, 7–67.

Verhelst, N. D. (1981). A note on ALSCAL: the estimation of the additive constant. *Pshychometrika, 46:4*, 465–468.