Hybrid Methods for Solving the Educational Testing Problem

Suliman Al-Homidan
Department of Mathematics, King Fahd University of Petroleum and Minerals, Dhahran 31261, PO Box 119, Saudi Arabia

Abstract

Methods for solving the educational testing problem are considered. One approach (Glunt [7]) is to formulate the problem as a linear convex programming problem in which the constraint is the intersection of three convex sets. This method is globally convergent but the rate of convergence is slow. However, the method does have the capability of determining the correct rank of the solution matrix, and this can be done in relatively few iterations. If the correct rank of the solution matrix is known, it is shown how to formulate the problem as a smooth nonlinear minimization problem, for which a rapid convergence can be obtained by l_1SQP method [6]. This paper studies hybrid methods that attempt to combine the best features of both types of method. An important feature concerns the interfacing of the component methods. Thus, it has to be decided which method to use first, and when to switch between methods. Difficulties such as these are addressed in the paper. Comparative numerical results are also reported.

Key words : Alternating projections, educational testing, non-smooth optimization, positive semi-definite matrix.
AMS (MOS) subject classifications 65F99, 99C25, 65F30

1 Introduction

The problem to be considered in this paper is the educational testing problem. Such optimization problems arise in many practical situations, particularly in statistics where we are given a matrix F which is usually a covariance matrix with varying elements. The educational testing problem is this: given a real symmetric positive definite $n \times n$ matrix F, how much can be subtracted from the diagonal of F and still retain a positive semi–definite matrix. This can be expressed as

$$\begin{aligned} &\text{maximize} \quad \mathbf{e}^T\boldsymbol{\theta} \qquad \boldsymbol{\theta} \in \mathbb{R}^n \\ &\text{subject to} \quad F - \operatorname{diag}\boldsymbol{\theta} \geq 0 \\ &\qquad\qquad \theta_i \geq 0 \quad i = 1, ..., n \end{aligned} \tag{1.1}$$

where $\mathbf{e} = (1, 1, ..., 1)^T$. An equivalent form of (1.1) is

$$\begin{aligned} &\text{minimize} \quad \mathbf{e}^T\mathbf{x} \qquad \mathbf{x} \in \mathbb{R}^n \\ &\text{subject to} \quad \bar{F} + \operatorname{diag}\mathbf{x} \geq 0 \\ &\qquad\qquad x_i \leq v_i \quad i = 1, ..., n \end{aligned} \tag{1.2}$$

where $\bar{F} = F - \operatorname{Diag} F$, and $\operatorname{diag}\mathbf{v} = \operatorname{Diag} F$.

An early approach in solving the educational testing problem is due to Bentler [2]. He writes $F - \operatorname{diag}\boldsymbol{\theta} = CC^T$, where C is unknown, and minimizes the trace of (CC^T), subject to certain conditions. He found that there are a large number of variables, and also it does not account for the bounds $\theta_i \geq 0 \;\; \forall\, i$. Furthermore, some difficulties in convergence to the optimum solution arise.

Woodhouse and Jackson [14] have given a method for solving the problem by searching in the space of $\boldsymbol{\theta}$. However, their method does not work efficiently and fails for particular examples.

Fletcher [5] has solved the problem by reducing the semi-definite constraint to an eigenvalue constraint, using standard nonlinear programming techniques. However, some difficulties still arise with the associated rates of convergence. Also, the presumption that the eigenvalue constraint would be smooth at the solution, except in rare cases, is not correct; in fact, a majority of such problems are nonsmooth at the solution.

In [6], Fletcher has developed a different algorithm for solving the educational testing problem. He gives various iterative methods for solving the nonlinear programming problem derived from the educational testing problem (1.2), using the Sequential Quadratic Programming (SQP) techniques. One of these algorithms is the use of the l_1-exact penalty function. This algorithm works well with second order convergence and the function converging to the optimal solution. The only problem in these algorithms is how to know the exact rank for the matrix $A^* = \bar{F} + \mathrm{diag}\, \mathbf{x}^*$ where $\mathbf{x}^*$ solves (1.2) .

Glunt [7] describes a projection method for solving the educational testing problem. His idea is to construct a hyperplane and then carry out the method of alternating projections ([12]) between the convex set K and the hyperplane. His method converges globally but the order of convergence is very slow.

New methods for solving the educational testing problem are introduced. The methods described here depend upon both the projection and the l_1SQP methods using a hybrid method. The hybrid method works in two stages. During the first stage, the projection method converges globally and, hence, is potentially reliable but often converges slowly. During the second stage, the l_1SQP method, has a second order convergence rate if the correct rank r^* is given. The main disadvantage of the l_1SQP method is that it requires the correct r^*. A hybrid method is one which switches between these methods and aims to combine their best features. To apply the l_1SQP method requires a knowledge of the rank r^* which can be gained from the progress of the projection method. Hybrid methods have often been used successfully in optimization, (e.g. Hald and Madsen [10] and Al-Homidan and Fletcher [1]).

The statistical background involved in the educational testing problem is described in Section 2. In Section 3 the educational testing problem is solved using the von Neumann algorithm. Section 4 contains a brief description of the l_1SQP method for solving (1.2). In Section 5 two new hybrid methods are described. Firstly, there is the projection–l_1SQP method, which starts with the projection method to determine the rank $r^{(k)}$ and continues with the l_1SQP method. Secondly, the l_1SQP–projection method is described which solves the problem by the l_1SQP method and uses the projection method to update the rank. Finally in Section 6, numerical comparisons of these methods are carried out.

2 The Educational Testing Problem

This section explains the educational testing problem which arises from statistics and gives rise to the nonlinear programming problem (1.1). The problem is to find lower bounds for the reliability of the total score on a test (or subtests) whose items are not parallel using data from a single test administration. The educational testing problem consists of a number of student (N) taking a test or examination consisting of (n) subtests. The problem is to find how reliable is the students's total score in the sense of being able to reproduce the same total on two independent occasions. Specifically, it is required to know what evidence about reliability can be obtained by carrying out a test on one occasion only. The discussion will closely follow that of Fletcher [5].

The given data for the problem is an $N \times n$ table of scores $[X_{ij}]$ (e.g. [5]) such that X_{ij} gives the observed score of student i on subject j. The student's total score is $X_i = \sum_j X_{ij}$, and $\underline{X}$ is the vector of total scores. The mean score for subject j

is $\overline{X}_j = 1/N \sum_i X_{ij}$, and the mean total score is $\overline{X} = \sum_j \overline{X}_j$. These observed scores are regarded as having been sampled from a universe of test, and $\mathcal{E}[\cdot]$ denotes the expected value on this universe. Then it is assumed that

$$X_{ij} = T_{ij} + E_{ij} \ \ \forall\, i, j \tag{2.1}$$

where

$$\mathcal{E}[E_{ij}] = 0 \ \ \forall\, i, j. \tag{2.2}$$

The quantities T_{ij} represent the hypothetical true scores where $\underline{T}, \overline{T}_j$, and $\overline{T}$ are defined as for X, and are the expected values of the corresponding quantities for the true scores.

The variance of the total scores from the expected mean scores is

$$\sigma_X^2 = \frac{1}{N-1} \sum_i \mathcal{E}[(X_i - \mathcal{E}[\overline{X}])^2] = \frac{1}{N-1} \sum_i \mathcal{E}[(X_i - \overline{T})^2]. \tag{2.3}$$

Reliability of the test may be regarded as the correlation in the student's total scores from two indepenent tests. Let $X^{(1)}$ and $X^{(2)}$ represent two such tests and $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ be the corresponding total scores. [9] defins the reliability coefficient ρ by

$$\rho^2 = \frac{1}{N-1} \frac{\sum_i \mathcal{E}[(X_i^{(1)} - \overline{T})(X_i^{(2)} - \overline{T})]}{\sigma_X^2}. \tag{2.4}$$

This is a correlation in the observed scores. In a completely reliable test, $X^{(1)} = X^{(2)}$ and it follows from (2.1) that $\rho = 1$. Assuming that the errors are uncorrelated such that $\mathcal{E}[E_i^{(1)} \cdot E_i^{(2)}] = \mathcal{E}[E_i^{(1)}]\mathcal{E}[E_i^{(2)}]$ it follows that

$$\rho^2 = \frac{\sigma_T^2}{\sigma_X^2} \tag{2.5}$$

where $\sigma_T^2 = \frac{1}{N-1} \sum_i (T_i - \overline{T})^2$. To determine how much information about ρ can be deduced from a single set of test scores, one can relate σ_X^2 and σ_T^2 to certain variance-covariance matrices. The variance-covariance matrix Σ_X of observed scores from the expected mean observed scores is defined as

$$\begin{aligned}[\Sigma_X]_{jk} &= \frac{1}{N-1}\sum_i \mathcal{E}[(X_{ij}-\mathcal{E}[\overline{X}_j])(X_{ik}-\mathcal{E}[\overline{X}_k]) \\ &= \frac{1}{N-1}\sum_i \mathcal{E}[(X_{ij}-\overline{T}_j)(X_{ik}-\overline{T}_k]). \qquad (2.6)\end{aligned}$$

Similarly for Σ_T and Σ_E

$$[\Sigma_T]_{jk} = \frac{1}{N-1}\sum_i (T_{ij}-\overline{T}_j)(T_{ik}-\overline{T_k}), \qquad (2.7)$$

and $[\Sigma_E]_{jk} = \frac{1}{N-1}\sum_i \mathcal{E}[E_{ij}E_{ik}]$. If we assume uncorrelated errors in the sense that

$$\mathcal{E}[E_{ij}T_{ik}] = \mathcal{E}[E_{ij}]T_{ik} \quad \forall\, i,j,k, \qquad (2.8)$$

and

$$\mathcal{E}[E_{ij}E_{ik}] = \mathcal{E}[E_{ij}]\mathcal{E}[E_{ik}] \quad \forall\, i,j,k, \quad j \neq k, \qquad (2.9)$$

then it follows from (2.2) that Σ_E is diagonal and from (2.7) and (2.1) that

$$\Sigma_X = \Sigma_T + \Sigma_E. \qquad (2.10)$$

It also follows from (2.3) and (2.7) that $\sigma_X^2 = \sum_{j,k}[\Sigma_E]_{jk} = \mathbf{e}^T\Sigma_T\mathbf{e}$ where $\mathbf{e} = (1,1,\dots 1)^T$, and that $\sigma_T^2 = \mathbf{e}^T\Sigma_T\mathbf{e}$. So writing $\theta_i = (\Sigma_E)_{ii}$, (2.5) becomes

$$\rho^2 = 1 - \frac{\sum_i \theta_i}{\sigma_X^2}. \qquad (2.11)$$

[9] shows that for large values of N, Σ_X may be estimated by

$$f_{jk} = [\Sigma_X]_{jk} \approx \frac{1}{N-1}\sum_i (X_{ij}-\overline{X}_j)(X_{ik}-\overline{X}_k). \qquad (2.12)$$

Σ_T and Σ_E are unknown, but being variance-covariance matrices, they are positive semi-definite. Using (2.10), these conditions may be written as

$$\Sigma_X - \Sigma_E \geq 0, \Sigma_E \geq 0, \qquad (2.13)$$

and may be regarded as constraints on the θ_i.

Obviously the θ_i satisfy

$$\sum_i \theta_i \leq \max_{\theta_i} \sum_i \theta_i \tag{2.14}$$

where the max is taken over all θ_i satifying (2.13). By (2.11),

$$\rho^2 \geq 1 - \frac{\phi}{\sigma_X^2}. \tag{2.15}$$

So by solving the optimization problem in (2.14) or equivalently (1.1), one obtains a lower bound on the value of ρ. This is the best that can be done on the basis of a single test.

3 A Projection Method

In this section a projection algorithm due to [7] for solving the educational testing problem is described. The method described here depends on the basic iterated projection algorithm by [12].

It is convenient to define three convex sets for the purpose of constructing the probem. The set of all $n \times n$ symmetric positive semi–definite matrices

$$K_{\mathbb{R}} = \{A : A \in \mathbb{R}^{n\times n}, \quad A^T = A \text{ and } \mathbf{z}^T A\mathbf{z} \geq 0 \quad \forall\, \mathbf{z} \in \mathbb{R}^n\} \tag{3.1}$$

is a convex cone of dimension $n(n+1)/2$. If $F \in \mathbb{R}^{n\times n}$ is any given symmetric positive definite matrix, then define

$$K_{\text{off}} = \{A : A \in \mathbb{R}^{n\times n},\ A - \text{Diag}\, A = \bar{F}\}. \tag{3.2}$$

where $\bar{F} = F - \text{Diag}\, F$. This is the set of matrices whose off–diagonal elements are equal to those of F. Also, let $\text{diag}\,\mathbf{v} = \text{Diag}\, F$, then define

$$K_{\text{b}} = \{A : A \in \mathbb{R}^{n\times n}, A = \bar{A} + \text{diag}\,\mathbf{x},\ x_i \leq v_i \ \ i = 1, 2, ...n\} \tag{3.3}$$

where $\bar{A} = A - \text{Diag } A$. This is the set of matrices that is obtained by reducing the diagonal of A. K_{off} and K_{b} are subspaces. Then (1.2) can be expressed as

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{e}^T\mathbf{x} \qquad \mathbf{x} \in \mathbb{R}^n \\ &\text{subject to} \quad \bar{F} + \text{diag } \mathbf{x} \in K_{\mathbb{R}} \cap K_{\text{off}} \cap K_b \end{aligned} \tag{3.4}$$

Let K_1 and K_2 be supspaces of Hilbert space and P_1 and P_2 be, respectively, the orthogonal projections onto K_1 and K_2. Then, the von Neumann method is given by

Algorithm 3.1 Given a point $\mathbf{f}$,

$$\begin{aligned} &\text{Set} \quad \mathbf{x}^{(0)} = \mathbf{f} \\ &\text{For} \quad k = 0, 1, 2, ... \\ &\quad \mathbf{x}^{(k+1)} = P_2 P_1(\mathbf{x}^{(k)}) \end{aligned}$$

The sequence in Algorithm 3.1 converges to $P_{K_1 \cap K_2}(\mathbf{f})$, which is the orthogonal projection onto the intersection of K_1 and K_2.

Glunt's idea is to take account of the function $\mathbf{e}^T\mathbf{x}$ by defining the hyperplane

$$\begin{aligned} L_\tau &= \{Y = \bar{Y} + \text{diag } \mathbf{y} \in \mathbb{R}^{n \times n} | \ \mathbf{e}^T\mathbf{y} = \tau\} \\ &= \{Y \in \mathbb{R}^{n \times n} | \ tr(Y) = \tau\} \end{aligned} \tag{3.5}$$

where $\text{Diag } Y = \text{diag } \mathbf{y}$ and τ is chosen such that

$$\tau < \min_{\mathbf{x} \in K} \mathbf{e}^T\mathbf{x}. \tag{3.6}$$

Then the sets $K = K_{\mathbb{R}} \cap K_{\text{off}} \cap K_{\text{b}}$ and L_τ are disjoint. Given a matrix $F \in \mathbb{R}^{n \times n}$, with $F = \bar{F} + \text{diag } \mathbf{f}$ and $A = \bar{A} + \text{diag } \mathbf{x}$, Glunt applies Algorithm 3.1 to the problem

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{f} - \mathbf{x}\|_2 \\ &\text{subject to} \quad A \in K \cap L_\tau \end{aligned} \tag{3.7}$$

which has no feasible solution. Now (3.7) generates the sequences $\{Y^{(k)}\} \in L_\tau$ and $\{A^{(k)}\} \in K$ converges to the points $Y^* \in L_\tau$, and $A^* \in K$ such that $\| Y - A\|_2$ attains the minimum distance between K and L_τ [3]. It can then be deduced from the relationship of L_τ and $\mathbf{e}^T\mathbf{x}$ that A^* solves (3.4).

The von Neumann algorithm involves computing alternately the projections onto L_τ and K. The projection onto L_τ is straightforward and is given by

$$P_{L_\tau}(Y) = Y + \frac{\tau - tr(Y)}{n} I, \tag{3.8}$$

see [7]. For (1.2), we need the projection $P_K(A)$ where $K = K_{\mathbb{R}} \cap K_{\text{off}} \cap K_{\text{b}}$ for any matrix A. The projection on $K = \bigcap_{i=1}^3 K_i$ is computed using an inner iteration based on the Dykstra algorithm [4] and is included as an inner iteration inside Algorithm 3.2, equations (3.9) and (3.10). It follows from [4] that the resulting method is globally convergent.

Algorithm 3.2 Given any positive definite matrix F, let $F^{(0)} = F$

For $k = 1, 2, \ldots$

$$B^{(k+1)} = P_{L_\tau}(F^{(k)})$$

$$\textit{For} \quad l = 1, 2, \ldots \tag{3.9}$$

$$A^{(0)} = B^{(k+1)}$$

$$A^{(l+1)} = A^{(l)} + P_{\text{b}} P_{\text{off}} P_{\mathbb{R}}(A^{(l)}) - P_{\mathbb{R}}(A^{(l)}) \tag{3.10}$$

$$F^{(k+1)} = P_{\text{b}} P_{\text{off}} P_{\mathbb{R}}(A^{(*)})$$

where A^* is the solution for the inner iteration.

The projection map $P_{\mathbb{R}}(A)$ formula on to $K_{\mathbb{R}}$ is given by [11]

$$P_{\mathbb{R}}(F) = U\Lambda^+ U^T. \tag{3.11}$$

where

$$\Lambda^{+} \;=\; \begin{bmatrix} \Lambda_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad (3.12)$$

and $\Lambda_r = \text{diag}\,[\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_r]$ is the diagonal matrix formed from the positive eigenvalues of F. Since K_{off} consists of all real symmetric $n \times n$ matrices, in which the off–diagonal elements are fixed to F (the given matrix), therefore

$$P_{\text{off}}\;(A) \;=\; \bar{F} \;+\; \text{Diag}\; A. \qquad (3.13)$$

Also, since K_{b} consists of all real symmetric $n \times n$ matrices, in which the diagonal elements are not greater than $\text{diag}\; \mathbf{v} =$ Diag F , we have

$$P_{\text{b}}\;(A) \;=\; \bar{A} \;+\; \text{diag}\,[h_1,\ h_2, ...,\ h_n], \qquad (3.14)$$

where

$$\mathbf{h} \;=\; \begin{Bmatrix} h_i \;=\; a_{ii} & if & a_{ii} \;\leq\; v_i \\ h_i \;=\; v_i & if & a_{ii} \;>\; v_i \end{Bmatrix}.$$

4 The l_1SQP Method

This section contains a brief description of the l_1SQP method for solving the educational testing problem [6].

Problem (1.2) can be expressed as

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{e}^T\mathbf{x} \qquad \mathbf{x} \;\in\; \mathbb{R}^n \\ &\text{subject to} \;\; \bar{A} \;+\; \text{diag}\,\mathbf{x} \;\in\; K_{\mathbb{R}} \,\cap K_{off}, \;\;\; \mathbf{x} \;\leq\; \mathbf{v} \end{aligned} \qquad (4.1)$$

where $\text{diag}\,\mathbf{v} \;=\; \text{Diag}\; A^{(0)}$. We can follow [6] for full details in solving (4.1). However in this section we give a summary of what has been given.

The first order necessary conditions can be stated as follows: If $\mathbf{x}^*$ solves (4.1), then $\mathbf{x}^*$ is feasible and there exists a matrix $\acute{B}^* \in \partial(K_{\mathbb{R}} \cap K_{off})(A^*)$ where $\partial(K_{\mathbb{R}} \cap$

$K_{off})(A^*)$ is the normal cone to $K_{\mathrm{IR}} \cap K_{\mathrm{off}}$ at A^* and a vector $\boldsymbol{\pi}^* \geq 0(\boldsymbol{\pi}^* \in \mathrm{IR}^n)$ such that

$$\mathbf{e} \; + \; \mathbf{b}^* \; + \; \boldsymbol{\pi}^* \; = \; 0. \tag{4.2a}$$

$$\boldsymbol{\pi}^{*T} \; (\mathbf{v} - \mathbf{x}^*) \; = \; 0 \tag{4.2b}$$

where diag $\mathbf{b}^* \; = \;$ Diag $\acute{B}^*$. This gives a characterization of the first order conditions for (4.1). However, it does not take into account second order effects, although it may be important to do this in order to obtain a second order rate of convergence in an algorithm. It is difficult to deal with the matrix cone constraints in (4.1) since it is not easy to specify if the elements are feasible or not. Using partial LDL^T factorization of A, this difficulty is rectified. Assume that r, the rank of A^*, is known, then for A sufficiently close to A^*, the partial factors $A = LDL^T$ can be calculated where

$$L \; = \; \begin{bmatrix} L_{11} & \\ L_{21} & I \end{bmatrix}, D \; = \; \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix}, A \; = \; \begin{bmatrix} A_{11} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix}.$$

Then

$$D_2(A) \; = \; A_{22} \; - \; A_{21} A_{11}^{-1} A_{21}^T, \tag{4.3}$$

and $D_2(\mathbf{x}) \; = \; D_2(\bar{A} \; + \; \mathrm{diag}\; \mathbf{x}) \; = \; D_2(A)$. Therefore an equivalent problem to (4.1) with the constraint $D_2 \; = \; \mathbf{0}$ is considered and expressed as

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{e}^T\mathbf{x} \qquad \mathbf{x} \; \in \; \mathrm{IR}^n \\ &\text{subject to} \;\; D_2(\mathbf{x}) \; = \; 0, \qquad \mathbf{x} \; \leq \; \mathbf{v} \end{aligned} \tag{4.4}$$

The Lagrangian for this problem is $\mathcal{L}(\mathbf{x}, \Lambda, \pi) = e^T\mathbf{x} - \Lambda : D_2(\mathbf{x}) + \pi^T(\mathbf{x} - \mathbf{v})$. To eliminate the variables $x_i, \;\; i \; = \; r+1, \ldots, n$, (4.3) is exploited by using the diagonal elements of $D_2(\mathbf{x})$

$$d_{ii}(\mathbf{x}) \; = \; x_i \; - \; \sum_{k,l=1}^{r} a_{ik} \; [A_{11}^{-1}]_{kl} \;\; a_{il} \; = \; 0 \qquad\qquad i \; = \; r+1, \ldots, n \tag{4.5}$$

where a_{ik} and a_{il} are elements in A_{21}. Therefore the unknown variables are reduced to $\mathbf{x} = [x_1, \ x_2, \ \ldots, \ x_r]^T \in \mathbb{R}^r$. This formulation will enable us to derive algorithms with a second order rate of convergence. Now, using the constraint $D_2 = \mathbf{0}$, will produce an equivalent problem to (4.4). The number of variables in this new problem can be reduced to r variables which gives the new reduced problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) = \sum_{k=1}^{r} x_k + \sum_{i=r+1}^{n} x_i(\mathbf{x}) \\ \text{subject to} & d_{ij}(\mathbf{x}) = 0, \quad i \neq j, \quad \mathbf{x} \leq \mathbf{v}. \quad i,j = r+1,\ldots,n \end{array} \tag{4.6}$$

where $x_i(\mathbf{x})$ indicates that x_i is the function of $\mathbf{x}$ determined by

$$x_i(\mathbf{x}) = \sum_{k,l=1}^{r} a_{ik}[A_{11}^{-1}]_{kl} \, a_{il} \qquad i = r+1,\ldots,n.$$

The expressions for the derivatives $\frac{\partial d_{ij}}{\partial x_s}$ and $\frac{\partial^2 d_{ij}}{\partial x_s \partial x_t}$ are given in [6] which enable us to find expressions for ∇f, $\nabla^2 f$ and $W = \nabla^2 \mathcal{L}(\mathbf{x}, \Lambda, \pi)$. Then using these expressions the QP subproblem

$$\begin{array}{lll} \underset{\boldsymbol{\delta}}{\text{minimize}} & f^{(k)} + \nabla f^{(k)} \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}^T W^{(k)} \boldsymbol{\delta} & \boldsymbol{\delta} \in \mathbb{R}^r \\ \text{subject to} & d_{ij}^{(k)} + \nabla d_{ij}^{(k)T} \boldsymbol{\delta} = 0 \qquad i \neq j \quad i,j = r+1,\ldots,n & \\ & \mathbf{x}^{(k)} + \boldsymbol{\delta} \leq \mathbf{v} & \end{array} \tag{4.7}$$

is defined. Thus the SQP method applied to (4.6) requires the solution of the QP subproblem (4.7). The matrix $W^{(k)}$ is positive semi–definite see [6].

5 Hybrid Methods

In this section, new methods for solving the educational testing problem are introduced. The methods described here depend upon both the projection and l_1SQP methods using a hybrid method. The hybrid method works in two stages. During the first stage, the projection method converges globally and, hence, is potentially reliable but often converges slowly. During the second stage, the l_1SQP method and the method, described in Section 4, has a second order convergence rate if the correct rank r^* is given. The main disadvantage of the l_1SQP method is that it requires the correct r^*. A hybrid method is one which switches between these methods and aims to combine their best features. To apply the l_1SQP method requires a knowledge of the rank r^* which can be gained from the progress of the projection method. This hybrid method can work well but there is one disadvantage: if the positive definite matrix has the same rank as the optimal positive semi–definite matrix in which the l_1SQP method works well, then most of the time will be taken up in the first stage, using the projection method. If this converges slowly, then the hybrid method will not solve the problem effectively. Thus, it is important to ensure that the second stage method is used to maximum effect. Hence, in the algorithm of Section 5.2, the l_1SQP method is applied first.

5.1 Projection–l_1SQP method

The main disadvantage of the l_1SQP method is finding the exact rank r^*. Since it is not known in advance, it is necessary to estimate it by an integer $r^{(k)}$. It is suggested that the best estimate of the matrix rank $r^{(k)}$ is obtained by carrying out some iterations of the projection method given in Section 3. This is because the projection method is a globally convergent method.

Considering Λ_r in (3.12), then at the solution, the number of eigenvalues in Λ_r is equal to the rank r^*. Thus

$$No. \quad \Lambda_r^* = r^*, \tag{5.1}$$

where $No.\ \Lambda$ is the number of positive eigenvalues in Λ. An equation similar to (5.1) is used to calculate an estimated rank $r^{(k)}$, given by

$$No. \quad \Lambda_r^{(k)} \;=\; r^{(k)},$$

where Λ_r is given by (3.12). Then, the l_1SQP method will be applied to solve the problem as described in Section 4.

Another consideration is how to choose τ. If τ is close to the boundary of (3.6), then the equation $No.\ \Lambda_r^{(k)} = r^*$ may be satisfied during the first few iterations. Experiments have proved this fact as shown in Table 6.1

The projection–l_1SQP algorithm can now be described as follows.

Algorithm 5.1 Given any positive definite matrix $F = F^T \in \mathbb{R}^{n\times n}$, let s be a positive integer. Then the following algorithm solves (1.2)

i. Let $F^{(0)} := F$.

ii. Choose τ to be close to the boundary of the condition (3.6).

iii. Apply Algorithm 3.2 until

$$No. \quad \Lambda_r^{(k)} = No.\ \Lambda_r^{(k+j)} \quad j \;=\; 1, 2,\ \ldots, s. \tag{5.2}$$

iv. $r^{(k)} \;=\; No.\ \Lambda_r^{(k)}$.

v. Use the result vector $\mathbf{x}$ from Algorithm 3.2 as an initial vector for the l_1SQP method.

vi. Apply the l_1SQP method to solve the problem with $r = r^{(k)}$.

If $\|D_2(\mathbf{x})\| \leq \epsilon$ for some small ϵ, then

$F^* := F^{(k)}$, $r^* := r^{(k)}$ and the algorithm is terminated

vii. Apply one inner iteration of Algorithm 3.2

viii. Go to (iv).

The integer s in Algorithm 5.1 can be any positive number. If s is small, then the rank $r^{(k)}$ may not be accurately estimated, but the number of iterations taken by projection method is small. On the other hand, if s is large, then a more accurate rank is obtained but the projection method needs more iterations.

The advantage of using the projection method as the first stage of the projection–l_1SQP method is that if $F^{(0)}$ is positive semi–definite and singular of rank r^*, then the projection method terminates at the first iteration. Moreover, it gives the best estimate for $r^{(k)}$. The singularity plays an important rule here, for if the matrix F has rank r, this means there is $n-r$ zeros eigenvalues. So, subtracting a small value from the diagonal leads to a matrix F which is indefinite. This implies that F is the optimal.

5.2 l_1SQP–Projection method

Starting with the projection method has the advantage that, if the given matrix is positive semi–definite and singular, the projection method converges in one step. However, sometimes it takes many iterations before equation (5.2) is satisfied, especially if τ is chosen to be small. This means slow convergence since the projection method is a slowly convergent method. In this method, an algorithm starting with the l_1SQP method with an estimated rank $r^{(k)}$ is considered. Then, one iteration of the projection method will be calculated after every stage of the l_1SQP–projection algorithm. The resulting vector $\mathbf{x}^{(k)}$ will be used as an initial vector for the next stage; thus the vector $\mathbf{x}^{(k)}$ is updated at every stage from the previous one.

Now the l_1SQP–projection algorithm can be described as follows:

Algorithm 5.2 Given any positive definite matrix $F = F^T \in \mathbb{R}^{n\times n}$, the following algorithm solves the educational testing problem:

i. Let $F^{(0)} = F$.

ii. Choose $r^{(k)}$ (as small as possible based on one of Section 5.1 strategies).

iii. Apply the l_1SQP method. When $\|D_2(\mathbf{x})\| \leq \epsilon$, for some small ϵ, then the algorithm is terminated.

iv. Use the result $\mathbf{x}^{(k)}$ as an initial vector for the projection method (Algorithm 3.2).

v. Choose τ to be close to the boundary of the condition (3.6), $(\tau = \sum x_i^{(k)})$.

vi. Apply one iteration of the projection method.

vii. $r^{(k)} = No.\ \Lambda_r^{(k)}$.

viii. Use the result $\mathbf{x}^{(k)}$ as an initial vector for the l_1SQP method.

ix. Go to (iii).

Another advantage of this algorithm is that if the rank is not correct, then instead of adding one to $r^{(k)}$, it goes back to the projection method to provide a better estimate to $r^{(k)}$. This will increase or decrease $r^{(k)}$, gives with the resulting value being nearer to r^*; therefore variables will be added to or subtracted from the problem. The new variables are estimated using the projection method. Another advantage is that at every stage only one iteration of projection method is used, giving a faster converging algorithm.

6 Numerical Results and Comparisons

In this section, numerical problems are obtained from the data given by [13]. The Woodhouse data set is a 64×20 data which corresponds to 64 students and 20 subtests. Various selections from the set of subsets of columns are used to give various test problems to form the matrix A. These subsets are given in the first columns of Tables 6.2–4, the value of n being the number of elements in each subset. Equation (2.12) gives the formula for calculating the educational testing problem.

In Algorithm 3.2, τ must satisfy condition (3.6). Since $\mathbf{x}^*$ is not known in advance and with elements $f_{ij} \overset{\simeq}{>} 100$, therefore it is clear that the diagonal elements $\bar{F} +$ diag $\mathbf{x}^{(k)}$ are greater than about 100 so $\mathbf{e}^T\mathbf{x} \overset{\simeq}{>} 100n$ as F is positive definite. Therefore from (3.6), the choice $\tau = 100$ is recommended. In fact, we recommend this choice since the elements f_{ij} are close to each either in magnitude. However, in general. the off-diagonal elements can play a role in making a better estimate for τ. If τ is chosen randomly and does not satisfy the condition (3.6), then the matrix $F -$ diag $\mathbf{x}^{(k)}$ is indefinite and the method is rerun with a different τ.

[7] and [6] tested their methods on the twelve test problems originally due to [13]. The same test problems are applied for the methods in this paper. In all the tables of this section, NOI gives the number of outer iteration. When solved by

the von Neumann Algorithm, TNII gives the total number of inner iteration used by von Neumann algorithm in Algorithm 3.2, and $r^{(0)}$ gives the number of positive eigenvalues in the first iteration of Algorithm 3.2.

The projection method is very expensive in the sense that it consumes a large number of iterations, while the l_1SQP method takes a very small number of iterations.

The NAG routine is used to find the eigenvalues and eigenvectors for the matrix $\bar{F} + \text{diag } \mathbf{x}^{(k)}$. This matrix is reduced to a real symmetric tridiagonal matrix by Householder's method. Then the eigenvalues and eigenvectors are calculated using the QL algorithm. The amount of work required by these algorithms is approximately $\frac{2}{3}n^3$ multiplications per one inner iteration ([8]). Again, the NAG routine is used for solving the QP subproblem (4.7) as one iteration of the SQP method. The NAG routine is used in our method to solve the QP subproblem which requires the solution for the system

$$Z^{(k)} \, W \, Z^{(k)T}\mathbf{p}^{(k)} \; = \; -\, Z^{(k)T}(\mathbf{c} \, + \, W \, \mathbf{x}^{(k)}) \tag{6.1}$$

where $\mathbf{c} = \nabla f$ and $Z^{(k)}$ is a matrix whose columns form a basis for the null space of $A^{(k)}$ (the matrix of coefficients of the bounds and active constraints). $\mathbf{p}^{(k)}$ is a search direction. The matrix $Z^{(k)}$ is obtained from the TQ factorization of $A^{(k)}$, in which $A^{(k)}$ is represented as

$$A^{(k)} \begin{bmatrix} Z^{(k)} \\ Q \end{bmatrix} \; = \; [\, \mathbf{0} \quad T^{(k)} \,]. \tag{6.2}$$

The Lagrange multipliers $\boldsymbol{\lambda}^{(k)}$ are defined as the solution of the system

$$A^{(k)} \, \boldsymbol{\lambda}^{(k)} \; = \; \mathbf{c} \, + \, W \, \mathbf{x}^{(k)}. \tag{6.3}$$

Equations (6.1) and (6.2) cost approximately $\frac{7}{3}n^3$ multiplications to be solved while (6.3) costs approximately $\frac{8}{3}n^3$ multiplications to be solved, see [8]. Thus one iteration of the SQP method costs approximately $\frac{15}{3}n^3$ multiplications. Hence, one iteration of the SQP method costs about 7 times as much as one iteration of the projection method. Nonetheless, the SQP method is much better than the projection method since the number of iterations taken by the projection method is about 60 times greater than that taken by the SQP method. However, the Hybrid methods, as shown in Table 6.4, use even fewer iterations.

Table 6.1 investigates the effect of varying τ. It shows the outcome from Algorithm 3.2 for the following example

$$\bar{F} = \begin{bmatrix} 0 & 1 & 2 & -2 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ -2 & 2 & 1 & 0 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} 2 \\ 4 \\ 8 \\ 10 \end{bmatrix}$$

with a different τ. From Table 6.1, it is clear that small τ increases the total number of iterations performed by the von Neumann algorithm, while a bigger τ decreases the total number of inner iterations and increases the number of outer iterations which are very cheap to calculate using the projection (3.8) which costs approximately n multiplications while one inner iteration costs approximately $\frac{2}{3}n^3$ multiplications. Hence, it is recommended to increase τ to be close to the boundary of condition (3.6) which is compatible with the choice in Table 6.1. The results obtained by the new method of Section 5.1 are tabulated in Table 6.2. In Table 6.2, the columns headed by NQP give the number of times the l_1SQP is solved.

In the projection–l_1SQP method, τ needs to be estimated very close to $\sum x_i^*$. This will give us a very good estimate of the rank. Since the average size of the educational testing problem elements is more than 100, $\tau = n \times 100$ is chosen as an initial value. In Table 6.2, it is clear that when $n > 10$, then τ becomes very small compared with $\sum x_i^*$, which makes the projection method estimate $r^{(k)}$ very small compared with the

τ	NOI	TNII	$\sum x_i^*$	$r^{(0)}$	r^*
-30.0	2	2679	15	0	2
-20.0	2	2215	15	1	2
-10.0	2	1734	15	2	2
-5.0	2	1571	15	2	2
0.0	2	1291	15	2	2
5.0	3	1308	15	2	2
10.0	3	960	15	2	2
14.0	6	787	15	2	2
14.9	15	891	15	2	2
15.0	30	792	15.0051	2	2

Table 6.1: Numerical comparisons for same example with different τ.

correct r^*. The results obtained by the new method of Section 5.2 are tabulated in Table 6.3. In the l_1SQP–projection method, $r^{(k)}$ is updated using one iteration of the projection method. In the projection method, τ is estimated using the result from the l_1SQP method. In the 1–10 case, the projection method estimates $r^{(k)} = 10$ instead of $r^{(k)} = 9$. In Tables 6.2 and 6.3, it can be seen that the our results are exactly the same as those of [6]. Also, one or two of the variables are adjusted so that the matrix $F - \text{diag}\,\boldsymbol{\theta}$ is exactly singular.

Finally, in Table 6.4, the four methods are compared. It clear from the data in Table 6.4 that the l_1SQP–projection method is the best for the problems considered since it requires fewer iterations in each problem for solving the QP subproblem.

7 Conclusions

In this paper we have studied certain problems involving the positive semi–definite matrix constraint. Two methods are used for solving the educational testing problem. One is the l_1SQP method [6], and the other is the projection method [7]. The

Columns which determine F	τ	TNII	$r^{(0)}$	r^*	NQP	$\sum \theta_i^*$
1,2,5,6	400	4	3	3	11	542.77356
1,3,4,5	400	2	2	2	12	633.15784
1,2,3,6,8,10	600	11	4	5	8	305.48170
1,2,4,5,6,8	600	4	4	4	13	564.46331
1–6	600	6	4	4	10	535.36227
1–8	800	13	5	6	14	641.83848
1–10	1000	15	7	8	21	690.78040
1–12	1200	23	9	9	9	747.48921
1–14	1400	25	10	12	34	671.27506
1–16	1600	22	11	14	44	663.46204
1–18	1800	20	12	15	27	747.50574
1–20	2000	29	14	18	39	820.34265

Table 6.2: Results for the educational testing problem from the projection–l_1SQP method of Section 5.1.

hybrid methods developed in Section 5 give a good rate of convergence, especially the l_1SQP–projection method, as compared with the methods of Section 4. The projection method is not very effective in determining the rank when $n \geq 12$. This is because a small value of s is chosen in Algorithms 5.1 and 5.2. On the other hand, if s is increased then a large number of iterations are consumed by the projection method. Hence, a suitable way of chosing the integer s needs further investigation.

Acknowledgements

This research paper is supported by King Fahd University of Petroleum and Minerals under Project FT/2005-07.

References

[1] Al-Homidan, S. and Fletcher, R. [1995]. Hybrid methods for finding the nearest Euclidean distance matrix, in *Recent Advances in Nonsmooth Optimization* (Eds.

Columns which determine F	$r^{(0)}$	NQP	PM$r^{(k)}$	NQP	$\sum \theta_i^*$
1,2,5,6	2	5	3	6	542.77356
1,3,4,5	2	12			633.15784
1,2,3,6,8,10	3	4	5	5	305.48170
1,2,4,5,6,8	3	6	4	4	564.46331
1–6	3	7	4	4	535.36227
1–8	5	7	6	6	641.83848
1–10	6	9	8	11	690.78040
1–12	8	3	10	9	747.48921
1–14	10	6	12	9	671.27506
1–16	11	9	14	10	663.46204
1–18	13	7	15	16	747.50574
1–20	15	5	18	21	820.34265

Table 6.3: Results for the educational testing problem from the l_1SQP–projection method of Section 7.3.
PM$r^{(k)}$:rank r updated from the projection method.

Columns which		PM	l_1SQP		Pl_1SQP			l_1SQPP	
determine F	r^*	TNII	$r^{(0)}$	NQP	TNII	$r^{(0)}$	NQP	$r^{(0)}$	TNQP
1,2,5,6	3	197	2	14	4	3	11	2	11
1,3,4,5	2	224	2	12	2	2	12	2	12
1,2,3,6,8,10	5	580	3	9	11	4	8	3	9
1,2,4,5,6,8	4	4994	3	13	4	4	13	3	10
1–6	4	1351	3	14	6	4	10	3	11
1–8	6	1948	5	29	13	5	14	5	13
1–10	8	2918	6	34	15	7	21	6	20
1–12	9	2403	8	29	23	9	9	8	12
1–14	12	3196	10	36	25	10	34	10	15
1–16	14	5215	11	42	22	11	44	11	19
1–18	15	14043	13	27	20	12	27	13	23
1–20	18	8255	15	39	29	14	39	15	26

Table 6.4: Comparing the four methods.
Pl_1SQP: the projection–l_1SQP method.
l_1SQPP: the l_1SQP–projection method.
TNQP : total number of NQP.

D. Du, L. Qi and R. Womersley), World Scientific Publishing Co. Pte. Ltd., Singapore, pp. 1–17.

[2] Bentler, P. M. [1972]. A lower-bound method for the dimension-free measurement of internal consistency, *Social Sci. Res., 1,pp. 343-357.*

[3] Cheney, E. W. and Goldstein, A. [1959]. Proximity maps for convex sets, *Proc. Amer. Math. Soc., 10, pp. 448-450.*

[4] Dykstra, R. L. [1983]. An algorithm for restricted least squares regression, *J. Amer. Stat. Assoc. 78, pp. 839–842.*

[5] Fletcher, R. [1981]. A nonlinear programming problem in statistics (educational testing), *SIAM J. Sci. Stat. Comput., 2, pp. 257–267.*

[6] Fletcher, R. [1985]. Semi–definite matrix constraints in optimization, *SIAM J. Control and Optimization, 23, pp. 493–513.*

[7] Glunt, W. [1995]. An alternating projections methd for certain linear problems in a Hilbert space, *IMA Journal of Numerical Analysis, 15, pp.291-305.*

[8] Golub, G. H. and Van Loan, C. F. [1989]. *Matrix Computations*, Johns Hopkins Universty Press, Baltimore, MD.

[9] Guttman, L. [1945]. A basis for analyzing test–retest reliability, *Psychometrika, 10, pp. 255–282.*

[10] Hald, J. and Madsen, K. [1981]. Combined LP and quasi–Newton methods for minmax optimization, *Math. Programming, 20, pp. 49–62.*

[11] Higham, N. [1988]. Computing a nearest symmetric positive semi–definite matrix, *Linear Alg. and Appl., 103, pp. 103–118.*

[12] Von Neumann, J [1950]. *Functional Operators II. The Geometry of Orthogonal Spaces,* Annals of Math. Studies No. 22, Princeton Univ. Press.

[13] Woodhouse, B. [1976]. Lower bounds for the reliability of a test, M.Sc. Thesis, Dept of Statistics, University of Wales, Aberystwyth.

[14] Woodhouse, B. and Jackson, P. H. [1977]. Lower bounds for the reliability of the total score on a test composed of non–homogeneous items: II. A search procedure to locate the greatest lower bound, *Psychometrika 42, pp. 579–591.*