J. Japan Statist. Soc.
Vol. 37 No. 1 2007 105–121

MAXIMIZATION OF CORRELATION UNDER A QUADRATIC CONSTRAINT

Akihiro Hashimoto*, Hisao Miyano** and Masaaki Taguri**

An algebraic method is suggested to search for the optimal solution that maximizes a correlation criterion under a quadratic constraint. First it is shown that the problem formulated in a sample space can be reformulated in a parameter space, and then some properties of a matrix which specifies the quadratic constraint are provided along with its geometrical interpretation; the solution can be obtained by solving a nonlinear equation derived from the singular value decomposition of the matrix. Numerical results based on artificial data and entrance examination data are given to examine how our solution differs from the least squares solution under a quadratic constraint.

Key words and phrases: Canonical correlation, correlation coefficient, nonlinear optimization, quadratic constraint, singular value decomposition.

1. Introduction

In this paper we discuss the interrelation between a dependent variable y and a set of p concomitant variables $\boldsymbol{z} = (z_1, z_2, \ldots, z_p)'$, where there exists some restriction on parameters. More precisely, we find the linear combination of concomitant variables that has maximum correlation with the dependent variable under a given quadratic constraint on the coefficients of the linear combination.

This kind of problem arises in some practical situations. For example, a university wants to assess the abilities of candidates in a particular field by the scores of several subjects in the entrance examination. The weighted totals of these subjects are sometimes used in such a case. The university may determine the weights so as to attain the maximum correlation between the abilities and the weighted totals. However, it may be better that the weights are not so much changed from the currently used weights.

This is one of the simplest cases in canonical correlation analysis, but it has a quadratic constraint. This problem may also be considered as a prediction problem in regression analysis. Rao (1973) classified problems into the following three cases according to the criterion to be adopted;

(i) Minimizing the mean squared error for predictor,
(ii) Maximizing the correlation between a dependent variable and its predictor,
(iii) Maximizing the expected performance in selection.

If none of the restrictions is there, the solutions for these three cases are all given by the conditional mean of y given $\boldsymbol{z}$ (see Rao (1973), section 4g.1).

Received December 22, 2005. Revised March 16, 2006. Accepted July 10, 2006.

*Niigata College of Nursing, 240 Shin-nan, Joetsu, Niigata 943-0147, Japan.
**The National Center for University Entrance Examinations, 2-19-23 Komaba, Meguro-ku, Tokyo 153-8501, Japan. Email: taguri@rd.dnc.ac.jp

For case (i), the least squares case, Golub and Van Loan (1996) investigated the optimization problem with a quadratic constraint, and gave the algorithm for solving it based on the Lagrangian method. Although it is easily seen that the optimal solutions for cases (i) and (ii) coincide with each other in a linear constraint case, how about in a quadratic constraint case?

For case (i), the optimization problems with quadratic or cone constraints have been addressed in the literature, including Golub and Van Loan mentioned above, Vavasis (1991), Lawlor (1991), Faraut and Koranyi (1994), Shapiro (1997), Kojima (1998), and Vandenberghe and Boyd (1996). In this paper we investigate the optimization problem for case (ii), and show the solution of (ii) is, in general, different from that of (i) under a quadratic constraint. Note that the correlation coefficient between a dependent variable and its predictor is invariant for the scale transformation of these variables. This property must be quite preferable in many practical situations.

In Section 2, we first state the original problem in terms of a sample space. It is then translated to the problem in a parameter space, because the given constraint is on parameters. In Section 3, our problem is reformulated so as to make its handling easier, and some properties of a matrix which specifies the quadratic constraint are investigated in Section 4. Although these sections are preliminary, essential ideas for the solution are presented here. The main result is given in Section 5, where we also give its geometrical interpretation which is quite helpful for intuitive comprehension of our algorithm. The algorithm for calculating the solutions and numerical examples are given in Section 6, which should make clear the difference of cases (i) and (ii) mentioned above. In the final Section 7, a brief summary and some extensions of our problem are given as concluding remarks. The Appendix completes the validation of the algorithm in Section 6.

2. Formulation of the problem and the solutions in trivial cases

Suppose that a dataset $(y_i, \boldsymbol{z}_i)$ is given for $i = 1, 2, \ldots, n$, where y_i and $\boldsymbol{z}_i$ are the i-th observed values of a dependent variable y and a p-dimensional concomitant variable $\boldsymbol{z}$ $(n \geq p)$, respectively. The observation vector $\boldsymbol{y}$ and the design matrix Z are defined by $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ and $Z = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n)'$, where all columns of Z are assumed to be linearly independent. Let us now denote a linear combination of the concomitant variables by $\boldsymbol{w}'\boldsymbol{z}$, where $\boldsymbol{w}$ is a p-dimensional coefficient (weight) column vector. The sample correlation coefficient R_0 between the dependent variable y and the composed variable $\boldsymbol{w}'\boldsymbol{z}$ is then given by

$$R_0 = \frac{\boldsymbol{y}'(I-Q)Z\boldsymbol{w}}{\sqrt{\boldsymbol{y}'(I-Q)\boldsymbol{y}}\sqrt{\boldsymbol{w}'Z'(I-Q)Z\boldsymbol{w}}}, \tag{2.1}$$

where I is an $n \times n$ identity matrix. Q is given by $Q = \frac{1}{n}\boldsymbol{1}\boldsymbol{1}'$, and $\boldsymbol{1} = (1, 1, \ldots, 1)'$. Note that the matrix Q and so $I - Q$ is idempotent and symmetric; that is, $I - Q$ is a projector.

The primary objective is to maximize the absolute value of R_0, however we consider the case that the following quadratic constraint is imposed on the coefficient vector $\boldsymbol{w}$;

$$(\boldsymbol{w}-\boldsymbol{w}_0)'W^{-1}(\boldsymbol{w}-\boldsymbol{w}_0) \le r^2, \tag{2.2}$$

where W, $\boldsymbol{w}_0$ and r are given $p\times p$ positive definite matrix, p dimensional column vector and positive constant, respectively. Our problem is then as follows:

PROBLEM 0. *Find* $\boldsymbol{w}$ *which maximizes* $|R_0|$ *subject to* (2.2).

In Problem 0 the objective function $|R_0|$ is regarded as a function of $\boldsymbol{w}$, which is a p-dimentional vector in the parameter space. Also the constraint (2.2) determines some region in the parameter space. On the other hand, $(I-Q)\boldsymbol{y}$ and $(I-Q)Z\boldsymbol{w}$, which appear in the definition of R_0, are vectors in the sample space.

Let us therefore consider to express Problem 0 by using the vectors in the parameter space. The singular value decomposition of the matrix $(I-Q)Z$ is given by $(I-Q)Z = U\Lambda V$, where U and V are orthogonal matrices with size n and p, respectively. Λ is an $n\times p$ matrix and has the form $\Lambda = [\Delta\ \ \mathbf{0}]'$, where Δ is the diagonal matrix of nonzero singular values δ_i $(i=1,\ldots,p)$ of $(I-Q)Z$, and $\mathbf{0}$ is a $p\times(n-p)$ zero matrix.

We now partition U as $U=[U_1\ \ U_2]$, where U_1 is $n\times p$ and U_2 is $n\times(n-p)$, and put $T=\Delta V$. Using these notations, the numerator and the denominator of R_0 can be written as

$$\begin{cases} \boldsymbol{y}'(I-Q)Z\boldsymbol{w} = \boldsymbol{y}'U_1\Delta V\boldsymbol{w} = (U_1'\boldsymbol{y})'(T\boldsymbol{w}) = \boldsymbol{b}'\boldsymbol{x}, \\ \boldsymbol{w}'Z'(I-Q)Z\boldsymbol{w} = \boldsymbol{w}'(U_1T)'(U_1T)\boldsymbol{w} = (T\boldsymbol{w})'(T\boldsymbol{w}) = \boldsymbol{x}'\boldsymbol{x}, \end{cases}$$

where $\boldsymbol{b}=U_1'\boldsymbol{y}$ and $\boldsymbol{x}=T\boldsymbol{w}$.

R_0 is then given by

$$\begin{aligned} R_0 &= \frac{\sqrt{\boldsymbol{b}'\boldsymbol{b}}}{\sqrt{\boldsymbol{y}'(I-Q)\boldsymbol{y}}}\frac{\boldsymbol{b}'\boldsymbol{x}}{\sqrt{\boldsymbol{b}'\boldsymbol{b}}\sqrt{\boldsymbol{x}'\boldsymbol{x}}} \\ &= \frac{\sqrt{\boldsymbol{b}'\boldsymbol{b}}}{\sqrt{\boldsymbol{y}'(I-Q)\boldsymbol{y}}}\frac{\boldsymbol{b}'\boldsymbol{x}}{\|\boldsymbol{b}\|\|\boldsymbol{x}\|}. \end{aligned} \tag{2.3}$$

Since $\sqrt{\boldsymbol{b}'\boldsymbol{b}}/\sqrt{\boldsymbol{y}'(I-Q)\boldsymbol{y}}$ does not depend on $\boldsymbol{w}$, maximizing $|R_0|$ with respect to $\boldsymbol{w}$ is equivalent to maximizing $|R_1| = |\boldsymbol{b}'\boldsymbol{x}|/(\|\boldsymbol{b}\|\|\boldsymbol{x}\|)$ with respect to $\boldsymbol{x}$.

As for the constraint (2.2), put $\boldsymbol{a}=T\boldsymbol{w}_0$ and $\Sigma = TWT'$. Since T is nonsingular from our assumption, the constraint is give by

$$\begin{aligned} (\boldsymbol{w}-\boldsymbol{w}_0)'W^{-1}(\boldsymbol{w}-\boldsymbol{w}_0) &= (\boldsymbol{x}-\boldsymbol{a})'(T^{-1})'W^{-1}T^{-1}(\boldsymbol{x}-\boldsymbol{a}) \\ &= (\boldsymbol{x}-\boldsymbol{a})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{a}) \le r^2. \end{aligned} \tag{2.4}$$

Note that Σ is positive definite. Thus, Problem 0 is now translated to the following Problem 1:

Problem 1. *Find $\boldsymbol{x}$ which maximizes $|R_1|$ subject to* (2.4).

The optimal solution for this problem is trivial or does not exist in some cases, which we investigate here. First, let us consider the case that the origin of the parameter space is an inner point of the restriction region (2.4), that is, $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} < r^2$ holds. The optimal solution $\boldsymbol{x}^*$ is then given by $\boldsymbol{x}^* = c\boldsymbol{b}$, and the optimal value of $|R_1|$ is equal to 1, where the constant c must satisfy $(c\boldsymbol{b} - \boldsymbol{a})'\Sigma^{-1}(c\boldsymbol{b} - \boldsymbol{a}) \le r^2$. Solving this inequality for c, we have

$$\frac{\boldsymbol{b}'\Sigma^{-1}\boldsymbol{a} - \sqrt{D}}{\boldsymbol{b}'\Sigma^{-1}\boldsymbol{b}} \le c \le \frac{\boldsymbol{b}'\Sigma^{-1}\boldsymbol{a} + \sqrt{D}}{\boldsymbol{b}'\Sigma^{-1}\boldsymbol{b}}, \tag{2.5}$$

where $D = (\boldsymbol{b}'\Sigma^{-1}\boldsymbol{a})^2 - (\boldsymbol{b}'\Sigma^{-1}\boldsymbol{b})(\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)$.

Second, we consider the case that the origin is a boundary point of the region (2.4), that is, $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} = r^2$ holds. If the vector $\boldsymbol{b}$ is not included in the tangent space of the surface $(\boldsymbol{x} - \boldsymbol{a})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{a}) = r^2$ at the origin, the same discussion mentioned above holds. The optimal solution is given by $\boldsymbol{x}^* = c\boldsymbol{b}$ and $|R_1| = 1$, where c must satisfy (2.5). We then examine the case that $\boldsymbol{b}$ is on the tangent plane of the surface $(\boldsymbol{x}-\boldsymbol{a})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{a}) = r^2$ at the origin, which is given by $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{x} = 0$. Consider the vector $\boldsymbol{x} = c(k\boldsymbol{b} - \boldsymbol{a})$, where c and k are some constants. For any fixed k, we can determine the value of c so as to satisfy (2.4). Using the relations $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} = r^2$ and $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{b} = 0$, we have

$$-\frac{r^2}{k^2\boldsymbol{b}'\Sigma^{-1}\boldsymbol{b} + r^2} \le c \le 0.$$

The value of $|R_1|$, which depends on k in this case, is given by

$$|R_1| = \frac{|\boldsymbol{b}'(k\boldsymbol{b} - \boldsymbol{a})|}{\sqrt{\boldsymbol{b}'\boldsymbol{b}}\sqrt{(k\boldsymbol{b} - \boldsymbol{a})'(k\boldsymbol{b} - \boldsymbol{a})}},$$

and $\lim_{k\to\pm\infty}|R_1| = 1$ holds. This means that the value of $|R_1|$ can be as large as one likes, hence Problem 1 has no solution.

Let us now summarize the results obtained above as the following Theorem 1:

Theorem 1.

(i) *If either $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} < r^2$ is satisfied or both $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} = r^2$ and $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{b} \ne 0$ are satisfied, then the optimal solution $\boldsymbol{x}^*$ for Problem* 1 *is given by $\boldsymbol{x}^* = c\boldsymbol{b}$, where c is any constant that satisfies* (2.5). *The corresponding optimal value of $|R_1|$ is $|R_1| = 1$.*

(ii) *If $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} = r^2$ and $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{b} = 0$ are satisfied, then there is no optimal solution for Problem* 1.

This theorem states that our problem is rather trivial in the case of $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} \le r^2$. Hence, in the following sections, we mainly focus on the case that $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$ holds.

3. Some preparatory considerations

As described in Section 1, Golub and Van Loan (1996) solved the problem of least squares under a quadratic constraint by using the Lagrangian method. Considering the wide applicability of the method to constrained optimization problems, Problem 1 may be solved by the Lagrangian method. However, since the objective function in Problem 1 is rather complicated, the Lagrangian method will fail to give a simple algorithm for our problem. We therefore apply an algebraic method that is completely different from the Lagrangian method; more precisely, we first solve the following problem named Problem 2, and then obtain the solution for Problem 1 by using the solution for Problem 2. The validation of this procedure is given by Theorem 2 below.

Problem 2. *Find $\boldsymbol{x} \in R^p$, a p-dimensional real space, that maximizes the objective function*

$$|R(\boldsymbol{x})| = \frac{|\boldsymbol{x}'\boldsymbol{b}|}{\|\boldsymbol{x}\|\|\boldsymbol{b}\|}, \tag{3.1}$$

subject to the constraint $\boldsymbol{x}'P\boldsymbol{x} \le 0$, where $P = \Sigma^{-1} - (\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)^{-1}\Sigma^{-1}\boldsymbol{a}\boldsymbol{a}'\Sigma^{-1}$ and $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$.

Theorem 2. *If $\tilde{\boldsymbol{x}}$ is the solution for Problem 2, then the solution $\boldsymbol{x}^*$ for Problem 1 is given by*

$$\boldsymbol{x}^* = \frac{\tilde{\boldsymbol{x}}'\Sigma^{-1}\boldsymbol{a}}{\tilde{\boldsymbol{x}}'\Sigma^{-1}\tilde{\boldsymbol{x}}}\tilde{\boldsymbol{x}}. \tag{3.2}$$

Proof. Let M, N be the sets of points defined by

$$M = \{\boldsymbol{x} \mid (\boldsymbol{x} - \boldsymbol{a})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{a}) \le r^2, \boldsymbol{x} \in R^p\},$$

and

$$N = \{\boldsymbol{x} \mid t\boldsymbol{x} \in M, t \in R^1\}.$$

First we will show that $\boldsymbol{x}$ is a member of N if and only if $\boldsymbol{x}$ satisfies the condition $\boldsymbol{x}'P\boldsymbol{x} \le 0$. From the definition of N, it is clear that $\boldsymbol{x} \in N$ if and only if there exists $t \in R^1$ that satisifies the condition

$$(t\boldsymbol{x} - \boldsymbol{a})'\Sigma^{-1}(t\boldsymbol{x} - \boldsymbol{a}) \le r^2.$$

Since this condition is equivalent to the condition

$$\boldsymbol{x}'\Sigma^{-1}\boldsymbol{a}\boldsymbol{a}'\Sigma^{-1}\boldsymbol{x} - (\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x} = -(\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)\boldsymbol{x}'P\boldsymbol{x} \ge 0,$$

we conclude that $\boldsymbol{x} \in N$ if and only if $\boldsymbol{x}$ satifies $\boldsymbol{x}'P\boldsymbol{x} \le 0$, since $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$ is assumed.

Next we prove the relation $\boldsymbol{x}^* = \frac{\tilde{\boldsymbol{x}}'\Sigma^{-1}\boldsymbol{a}}{\tilde{\boldsymbol{x}}'\Sigma^{-1}\tilde{\boldsymbol{x}}}\tilde{\boldsymbol{x}}$. Since $N \supset M$, it is clear that the inequality $|R(\boldsymbol{x}^*)| \leq |R(\tilde{\boldsymbol{x}})|$ is satisfied for $\boldsymbol{x}^*$ and $\tilde{\boldsymbol{x}}$. Also it can be shown that $\boldsymbol{x}^*$ defined by $\boldsymbol{x}^* = \frac{\tilde{\boldsymbol{x}}'\Sigma^{-1}\boldsymbol{a}}{\tilde{\boldsymbol{x}}'\Sigma^{-1}\boldsymbol{x}}\tilde{\boldsymbol{x}} \in N$, belongs to M, since

$$\begin{aligned}(\boldsymbol{x}^* - \boldsymbol{a})'\Sigma^{-1}(\boldsymbol{x}^* - \boldsymbol{a}) - r^2 &= \boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2 - \frac{(\tilde{\boldsymbol{x}}'\Sigma^{-1}\boldsymbol{a})^2}{\tilde{\boldsymbol{x}}'\Sigma^{-1}\tilde{\boldsymbol{x}}} \\ &= \frac{\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2}{\tilde{\boldsymbol{x}}'\Sigma^{-1}\tilde{\boldsymbol{x}}}\tilde{\boldsymbol{x}}'P\tilde{\boldsymbol{x}} \leq 0.\end{aligned}$$

Hence we conclude that $\boldsymbol{x}^*$ is the solution for Problem 1, and $|R(\boldsymbol{x}^*)| = |R(\tilde{\boldsymbol{x}})|$.

4. Some algebraic properties of the matrix P

The matrix P defined in the previous section plays a significant role in the development of our algorithm for solving Problem 2. We here summarize its algebraic properties that will be used to validate our algorithm.

THEOREM 3. *Let Σ^{-1} be a $p \times p$ positive definite matrix, $\boldsymbol{a} \in R^p$ be a nonzero vector, and r be a nonzero real number.*

(i) *If $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$, then the matrix P, defined by*

$$P = \Sigma^{-1} - (\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)^{-1}\Sigma^{-1}\boldsymbol{a}\boldsymbol{a}'\Sigma^{-1}, \tag{4.1}$$

has $p-1$ positive and one negative eigenvalues.

(ii) *If $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} < r^2$, then all eigenvalues of P are positive.*

PROOF. Let Σ^{-1} and P be matrices with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ and $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_p$, respectively. Then, the second part of the theorem is trivial, since if $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} < r^2$, then $-(\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)^{-1}\Sigma^{-1}\boldsymbol{a}\boldsymbol{a}'\Sigma^{-1}$ is nonnegative. If $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$, the eigenvalues $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_p$ satisfy the following inequalities (Golub and Van Loan (1996), p. 442);

$$\lambda_i \geq \gamma_i \geq \lambda_{i+1}, \qquad i = 1, 2, \ldots, p-1, \tag{4.2}$$

$$\lambda_p \geq \gamma_p \geq \lambda_p - (\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)^{-1}\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a}. \tag{4.3}$$

Hence we get that γ_i is positive for $i = 1, 2, \ldots, p-1$.

Let us now show that γ_p is negative when $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$: Let Σ_e be the $(p+1) \times (p+1)$ matrix defined by

$$\Sigma_e^{-1} = \begin{pmatrix} \boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2 & -\boldsymbol{a}'\Sigma^{-1} \\ -\Sigma^{-1}\boldsymbol{a} & \Sigma^{-1} \end{pmatrix}. \tag{4.4}$$

Then the determinant of Σ_e^{-1} can be written as

$$|\Sigma_e^{-1}| = (\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)|P| = (\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)\prod_{i=1}^{p}\gamma_i, \tag{4.5}$$

or

$$|\Sigma_e^{-1}| = |\Sigma^{-1}|(\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2 - \boldsymbol{a}'\Sigma^{-1}\Sigma\Sigma^{-1}\boldsymbol{a}) = -r^2|\Sigma^{-1}|. \tag{4.6}$$

From the latter equation, it is clear that $|\Sigma_e^{-1}|$ is negative. Hence we obtain that γ_p is negative when $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$.

COROLLARY 1.
(i) *If* $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} \neq r^2$, *then*

$$\boldsymbol{a}'\Sigma^{-1}\boldsymbol{v}_p = -\gamma_p\boldsymbol{a}'\boldsymbol{v}_p\frac{\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2}{r^2}, \tag{4.7}$$

where $\boldsymbol{v}_p$ *is the eigenvector of the matrix* P *associated with its smallest eigenvalue* γ_p.
(ii) *If* $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$, *then* $\boldsymbol{a}'\boldsymbol{v}_p$ *is not zero.*

PROOF. From the definition of the matrix P, we have

$$\begin{aligned} \boldsymbol{a}'\Sigma^{-1}\boldsymbol{v}_p &= \boldsymbol{a}'\{P + (\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)^{-1}\Sigma^{-1}\boldsymbol{a}\boldsymbol{a}'\Sigma^{-1}\}\boldsymbol{v}_p \\ &= \gamma_p\boldsymbol{a}'\boldsymbol{v}_p + \{(\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)^{-1}\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a}\}\boldsymbol{a}'\Sigma^{-1}\boldsymbol{v}_p, \end{aligned} \tag{4.8}$$

and hence the first result follows.

The second result can be obtained by showing a contradiction; that is, if $\boldsymbol{a}'\boldsymbol{v}_p = 0$, then we get $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{v}_p = 0$ from (4.8), but this contradicts the assumption that the matrix Σ is positive definite, because by (i) it can be shown that $\boldsymbol{v}_p'\Sigma^{-1}\boldsymbol{v}_p = \boldsymbol{v}_p'P\boldsymbol{v}_p = \gamma_p\boldsymbol{v}_p'\boldsymbol{v}_p < 0$.

COROLLARY 2. *If* $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} \neq r^2$, *then the sign of* $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{v}_p$ *is equal to that of* $\boldsymbol{a}'\boldsymbol{v}_p$.

The result means that if $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$, we can assume $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{v}_p > 0$ without any loss of generality, since we can take $-\boldsymbol{v}_p$ as $\boldsymbol{v}_p$ if $\boldsymbol{a}'\boldsymbol{v}_p < 0$.

LEMMA 1. *Let* $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$, α *be a scalar*, $\boldsymbol{x}$ *be a nonzero vector that satisfies* $\boldsymbol{x}'P\boldsymbol{x} = 0$, *and* $g(\alpha)$ *be the function of* α *defined by*

$$g(\alpha) = (\boldsymbol{x} + \alpha P\boldsymbol{x})'P(\boldsymbol{x} + \alpha P\boldsymbol{x}). \tag{4.9}$$

Then $g(\alpha) > 0$ *for any* α *in* $(0, -\frac{2}{\gamma_p})$, *and there exists a negative constant* θ *for which* $g(\alpha) < 0$ *for any* α *in* $(\theta, 0)$.

PROOF. First we note that the function $g(\alpha)$ can be rewritten as $g(\alpha) = \alpha\boldsymbol{x}'P(2I + \alpha P)P\boldsymbol{x}$, and its derivative $g'(\alpha)$ is given by $g'(\alpha) = 2\boldsymbol{x}'P(I + \alpha P)P\boldsymbol{x}$. Hence it is clear that there exists an interval $(\theta, 0)$ in which $g(\alpha)$ is always negative, since $g(\alpha)$ is continuous, $g(0) = 0$, and $g'(0) > 0$.

On the other hand, γ_p is negative in the case of $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$, so all the eigenvalues of the matrix $(2I + \alpha P)$ are positive for any α in $(0, -\frac{2}{\gamma_p})$. Hence we obtain $g(\alpha) > 0$ for any α in this interval.

LEMMA 2. *Let* α *be a scalar*, $\boldsymbol{x}$ *be a vector that satisfies* $\boldsymbol{x}'P\boldsymbol{x} > 0$, *and* $f(\alpha)$ *be the function of* α *defined by*

$$f(\alpha) = \boldsymbol{x}'(I + \alpha P)^{-1}P(I + \alpha P)^{-1}\boldsymbol{x}. \tag{4.10}$$

If $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$, *then the equation* $f(\alpha) = 0$ *has a unique root in* $(0, -\frac{1}{\gamma_p})$.

PROOF. Let T be an orthogonal matrix whose columns are the eigenvectors of the matrix P. Note that $P = T\Gamma T'$ and $I + \alpha P = T(I + \alpha\Gamma)T'$, where Γ is the diagonal matrix with the i-th diagonal element γ_i, and $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{p-1} > 0 > \gamma_p$ are the eigenvalues of P. Then setting $\boldsymbol{c} = (c_1, c_2, \ldots, c_p)' = T'\boldsymbol{x}$, $f(\alpha)$ can be rewritten as

$$f(\alpha) = \sum_{i=1}^{p} \frac{\gamma_i}{(1+\alpha\gamma_i)^2} c_i^2. \tag{4.11}$$

The result directly follows from the fact that $f(0) = \boldsymbol{x}'P\boldsymbol{x} > 0$, $\lim_{\alpha \to -1/\gamma_p} f(\alpha) = -\infty$, and

$$f'(\alpha) = -2\sum_{i=1}^{p} \frac{\gamma_i^2}{(1+\alpha\gamma_i)^3} c_i^2 < 0, \qquad \text{for any} \quad \alpha \in \left(0, -\frac{1}{\gamma_p}\right). \tag{4.12}$$

5. Main result and its geometrical interpretation

Based on the results given in the previous section, here we present the algebraic solution for Problem 2, and develop an algorithm for computing $\boldsymbol{x}^*$, that maximizes $|R(\boldsymbol{x})|$ under the quadratic constraint on $\boldsymbol{x}$.

5.1. Main result

When the vector $\boldsymbol{b}$ given in Section 2 is in the region defined by the constraint $\boldsymbol{x}'P\boldsymbol{x} \leq 0$, the solution $\tilde{\boldsymbol{x}}$ of Problem 2 is trivial; i.e., $\tilde{\boldsymbol{x}} = \boldsymbol{b}$. Hence our main concern is in the case of $\boldsymbol{b}$ for which $\boldsymbol{b}'P\boldsymbol{b} > 0$. The following lemma and theorem give the solution of Problem 2 for this case.

LEMMA 3. *Let* $\tilde{\boldsymbol{x}}$ *be the solution of Problem* 2, *and* $\boldsymbol{b}$ *satisfy the condition* $\boldsymbol{b}'P\boldsymbol{b} > 0$. *Then* $\tilde{\boldsymbol{x}}$ *satisfies* $\tilde{\boldsymbol{x}}'P\tilde{\boldsymbol{x}} = 0$.

PROOF. We give a proof by showing a contradiction. Let $\boldsymbol{x}_0 (\|\boldsymbol{x}_0\| = 1)$ be the solution for Problem 2 satisfying $\boldsymbol{x}_0'P\boldsymbol{x}_0 < 0$, and r_0 be the value of $|R(\boldsymbol{x})|$ at $\boldsymbol{x}_0$; that is, $r_0 = |\boldsymbol{x}_0'\boldsymbol{b}|/\|\boldsymbol{b}\|$. Without loss of generality, we assume that $\|\boldsymbol{b}\| = 1$, and $\boldsymbol{x}_0'\boldsymbol{b} \geq 0$. If $r_0 = 0$, it contradicts the optimality of $\boldsymbol{x}_0$, because we can always take the vector such that $\tilde{\boldsymbol{x}}'P\tilde{\boldsymbol{x}} = 0$ and $|R(\tilde{\boldsymbol{x}})| > 0$. Then for any positive r_0, we can find the vector $\boldsymbol{x}(t) = (1-t)\boldsymbol{x}_0 + r_0 t\boldsymbol{b}$ where t is an appropriate value in $(0, 1)$, satisfying $\boldsymbol{x}'(t)P\boldsymbol{x}(t) = 0$. It is always possible to find such a t because $f(t) = \boldsymbol{x}'(t)P\boldsymbol{x}(t)$ is an at most second order polynomial of t with $f(0) = \boldsymbol{x}_0'P\boldsymbol{x}_0 < 0$ and $f(1) = r_0^2\boldsymbol{b}'P\boldsymbol{b} > 0$.

Since $\boldsymbol{x}'(t)\boldsymbol{x}(t) = (1-t)^2(1-r_0^2) + r_0^2$, we have

$$R^2(\boldsymbol{x}(t)) = \frac{r_0^2}{(1-t)^2(1-r_0^2) + r_0^2}.$$

Then, noting that $r_0^2 < (1-t)^2(1-r_0^2) + r_0^2 < 1$ for any t in $(0, 1)$, we obtain $R^2(\boldsymbol{x}(t)) > r_0^2$, which contradicts the assumption of $\boldsymbol{x}_0$ being optimal.

THEOREM 4. *Let* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *satisfy the conditions* $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$ *and* $\boldsymbol{b}'P\boldsymbol{b} > 0$. *If* $\boldsymbol{b}'\boldsymbol{v}_p \neq 0$, *then the solution* $\tilde{\boldsymbol{x}}$ *for Problem* 2 *is given by*

$$\tilde{\boldsymbol{x}} = (I + \alpha P)^{-1}\boldsymbol{b}, \tag{5.1}$$

where α *is a unique constant that satisfies*

$$\boldsymbol{b}'(I + \alpha P)^{-1}\boldsymbol{b} - \boldsymbol{b}'(I + \alpha P)^{-2}\boldsymbol{b} = 0, \qquad 0 < \alpha < -\frac{1}{\gamma_p}, \tag{5.2}$$

and the maximum of $|R(\boldsymbol{x})|$ *is* $\|(I+\alpha P)^{-1}\boldsymbol{b}\|/\|\boldsymbol{b}\|$. *If* $\boldsymbol{b}'\boldsymbol{v}_p = 0$, *then the solution* $\tilde{\boldsymbol{x}}$ *is given by*

$$\tilde{\boldsymbol{x}} = \left(I - \frac{1}{\gamma_p}P\right)^{+}\boldsymbol{b} + \sqrt{\frac{\boldsymbol{b}'\left(I - \frac{1}{\gamma_p}P\right)^{+} P \left(I - \frac{1}{\gamma_p}P\right)^{+}\boldsymbol{b}}{-\gamma_p}}\,\boldsymbol{v}_p, \tag{5.3}$$

where $(I - \frac{1}{\gamma_p}P)^{+}$ *is the Moore-Penrose generalized inverse of* $(I - \frac{1}{\gamma_p}P)$.

PROOF. The proof of the latter case is given in the Appendix. Without loss of generality, let $\boldsymbol{a}$ and $\boldsymbol{b}$ satisfy $\boldsymbol{a}'\boldsymbol{v}_p > 0$ and $\boldsymbol{b}'\boldsymbol{v}_p > 0$, since the sign of $\boldsymbol{b}$ does not change $|R(\boldsymbol{x})|$ (see also Corollary 2). Then, from the above lemma, $\tilde{\boldsymbol{x}}$ satisfies $\tilde{\boldsymbol{x}}'P\tilde{\boldsymbol{x}} = 0$; that is, $P\tilde{\boldsymbol{x}}$ is orthogonal to $\tilde{\boldsymbol{x}}$, and there must be a constant α for which $\boldsymbol{b} = \tilde{\boldsymbol{x}} + \alpha P\tilde{\boldsymbol{x}}$. Hence, the result can be proved by showing that there is a unique constant that satisfies (5.2).

First, from the condition $\boldsymbol{b}'P\boldsymbol{b} > 0$, α must satisfy the inequality given by

$$(\tilde{\boldsymbol{x}} + \alpha P\tilde{\boldsymbol{x}})'P(\tilde{\boldsymbol{x}} + \alpha P\tilde{\boldsymbol{x}}) > 0. \tag{5.4}$$

From Lemma 1, this inequality is satisfied when α is at least in $(0, -\frac{1}{\gamma_p})$. Note that $I + \alpha P$ is not singular for any α in $(0, -\frac{1}{\gamma_p})$.

Also, from the condition $\tilde{\boldsymbol{x}}'P\tilde{\boldsymbol{x}} = 0$, α must satisfy the equation

$$\alpha\boldsymbol{b}'(I + \alpha P)^{-1}P(I + \alpha P)^{-1}\boldsymbol{b} = \boldsymbol{b}'(I + \alpha P)^{-1}\boldsymbol{b} - \boldsymbol{b}'(I + \alpha P)^{-2}\boldsymbol{b} = 0. \tag{5.5}$$

From Lemma 2, we can assert that this equation has a unique root in $(0, -\frac{1}{\gamma_p})$.

Formally this result is very similar to that given by Golub and Van Loan (1996). However, we may assert that essentially it is very different from their result, since we have derived it without assuming that P is positive.

5.2. Geometrical interpretations

As a means to getting an intuitive, or pictorial insight into our method, we here provide its geometrical interpretations. First, under the assumption $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$ we define the following three sets of points in R^p:

$$M = \{\boldsymbol{x} \mid (\boldsymbol{x} - \boldsymbol{a})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{a}) \leq r^2, \boldsymbol{x} \in R^p\}, \tag{5.6}$$

$$N = \{\boldsymbol{x} \mid \boldsymbol{x}'P\boldsymbol{x} \leq 0, \boldsymbol{x} \in R^p\}, \tag{5.7}$$

$$\partial N = \{\boldsymbol{x} \mid \boldsymbol{x}'P\boldsymbol{x} = 0, \boldsymbol{x} \in R^p\}. \tag{5.8}$$

Proposition 1.

(i) *M is an ellipsoid that does not include the origin of R^p, and N is an ellipsoidal cone defined by M; i.e., $N = \{\boldsymbol{x} \mid t\boldsymbol{x} \in M, t \in R\}$.*

(ii) *For every $\boldsymbol{x} \in \partial N$, $P\boldsymbol{x}$ is normal to N at $\boldsymbol{x}$.*

Proposition 2. *Let $\boldsymbol{x}$ be a point on ∂N. Then the point $\boldsymbol{y}_\alpha$ specified by $\boldsymbol{y}_\alpha = \boldsymbol{x} + \alpha P\boldsymbol{x}$ $(0 < \alpha < -2/\gamma_p)$ is not included in N.*

Proof. From Lemma 1, $2I+\alpha P$ is positive definite for any $\alpha \in (0, -2/\gamma_p)$. So we have

$$\boldsymbol{y}'_\alpha P\boldsymbol{y}_\alpha = \boldsymbol{x}'(I+\alpha P)P(I+\alpha P)\boldsymbol{x} = \alpha(P\boldsymbol{x})'(2I+\alpha P)(P\boldsymbol{x}) > 0.$$

Hence the result follows.

Proposition 3. *Let $\boldsymbol{x}$ be a point on ∂N. Then $\boldsymbol{y}_{-1/\gamma_i} = \boldsymbol{x} - 1/\gamma_i P\boldsymbol{x}$ is orthogonal to the i-th eigenvector $\boldsymbol{v}_i$ of P.*

Proof. $\boldsymbol{y}'_{-1/\gamma_i}\boldsymbol{v}_i = \boldsymbol{x}'\boldsymbol{v}_i - 1/\gamma_i\boldsymbol{x}'P\boldsymbol{v}_i = \boldsymbol{x}'\boldsymbol{v}_i - 1/\gamma_i\boldsymbol{x}'(\gamma_i\boldsymbol{v}_i) = 0.$

Noting that the direction of the pth principal axis of N is given by the eigenvector $\boldsymbol{v}_p$, these propositions provide a simple graphical representation of our method. Figure 1 illustrates three typical cases which explain how the solution for Problem 2 depends on a quadratic constraint and a vector $\boldsymbol{b}$. In Fig. 1, $\boldsymbol{b}_1$ is the case of the vector $\boldsymbol{b}$ being inside of the ellipsoidal cone N; i.e., $\boldsymbol{b}'P\boldsymbol{b} \le 0$. Apparently this case yields the trivial solution $\tilde{\boldsymbol{x}} = \boldsymbol{b}$. $\boldsymbol{b}_2$ is the case of the vector $\boldsymbol{b}$ being outside of the cone N, and making an acute angle with the principal axis of N. $\boldsymbol{b}_3$ is the case of $\boldsymbol{b}'\boldsymbol{v}_p = 0$. These last two cases are the cases mentioned in Theorem 4.

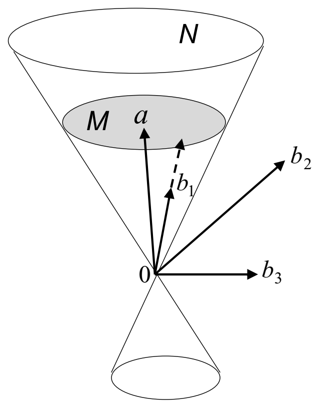

Figure 1. Illustration of three typical cases of vector $\boldsymbol{b}$.

6. Algorithm and numerical examples

As shown in Theorem 1, if the condition $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} \le r^2$ holds, then our optimization problem becomes rather trivial. In this section we hence restrict our probelm to the case that $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} > r^2$ is satisfied, and give a computational algorithm for this case. Then some artificial data and entrance examination data are analyzed to see how our optimal solution differs from the optimal solution for least squares case.

6.1. Algorithm

The results in preceding sections yield the following algorithm to compute the optimal solution:

Step 1. Calculate the matrix $P = \Sigma^{-1} - (\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} - r^2)^{-1}\Sigma^{-1}\boldsymbol{a}\boldsymbol{a}'\Sigma^{-1}$, and obtain the smallest eigenvalue γ_p and the corresponding eigenvector $\boldsymbol{v}_p$.

Step 2. If $\boldsymbol{b}'P\boldsymbol{b} \le 0$, then compute $\boldsymbol{x}^* = \frac{\boldsymbol{b}'\Sigma^{-1}\boldsymbol{a}}{\boldsymbol{b}'\Sigma^{-1}\boldsymbol{b}}\boldsymbol{b}$; otherwise go to next step.

Step 3. If $\boldsymbol{b}'\boldsymbol{v}_p \ne 0$, then compute $\boldsymbol{x}^* = \frac{\tilde{\boldsymbol{x}}'\Sigma^{-1}\boldsymbol{a}}{\tilde{\boldsymbol{x}}'\Sigma^{-1}\tilde{\boldsymbol{x}}}\tilde{\boldsymbol{x}}$, where $\tilde{\boldsymbol{x}} = (I + \alpha P)^{-1}\boldsymbol{b}$, and α is a uniquely determined constant that satisfies (5.2). If $\boldsymbol{b}'\boldsymbol{v}_p = 0$, then compute the optimal solution $\tilde{\boldsymbol{x}}$ by using (5.3).

Figure 2 illustrates the computational flow of our algorithm, which includes the case of $\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a} \le r^2$.

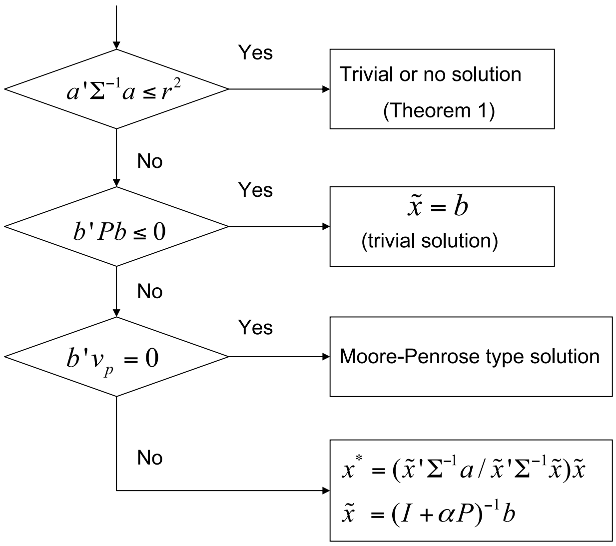

Figure 2. Computational flow.

6.2. Numerical examples

Artificial data. Our solution $\boldsymbol{x}^*$ usually differs from the least squares solution $\boldsymbol{x}_{LS}$. Since the difference between these two solutions depends on the parameters $\boldsymbol{a}$, Σ, and $\boldsymbol{b}$, our computation was carried out by changing these parameter values systematically. Let R^* and R_{LS} be the correlation coefficients corresponding to $\boldsymbol{x}^*$ and $\boldsymbol{x}_{LS}$, respectively. These values might depend on the mutual relation of $\boldsymbol{a}$ and $\boldsymbol{b}$, so we fixed $\boldsymbol{a} = (0, 0, 10)'$ and changed the value of $\boldsymbol{b} = t(\cos\theta_1 \cos\theta_2, \sin\theta_1 \cos\theta_2, \sin\theta_2)'$, where $t = 1, 10, 20$, and $\theta_i = 0, \frac{\pi}{12}, \frac{\pi}{6}, \frac{\pi}{4}$ for $i = 1, 2$ (see Fig. 3). The values of R^* and R_{LS} also depend on Σ and r, so we fixed $r = 1$ and changed the value of $\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, where $\sigma_i = 1, 4, 7$ for $i = 1, 2, 3$. The numerical examination was carried out for all combinations of these values; that is, the total number of cases examined was 1296 ($= 3 \times 4^2 \times 3^3$). For each case we calculated $\boldsymbol{x}^*$ and $\boldsymbol{x}_{LS}$ respectively by using our method and that of Golub and Van Loan (1996), and evaluated the difference of these two solutions by computing R^* and R_{LS}. Table 1 summarizes part of the obtained results that may be enough to see how the parameters affect the difference between $\boldsymbol{x}^*$ and $\boldsymbol{x}_{LS}$. In the table, d is the relative difference defined by $d = \{(R^*)^2 - (R_{LS})^2\}/(R^*)^2$.

As for the difference of R^* and R_{LS}, we obtain the following findings:

(1) The difference is large for $t = 1$; that is, when the length of $\boldsymbol{b}$ is short, R^* is rather large compared with R_{LS}.

(2) The difference is especially large for the cases $(\sigma_2, \sigma_3) = (7, 1)$ and $(\theta_1, \theta_2) = (\frac{\pi}{4}, \frac{\pi}{12})$, or $(\sigma_2, \sigma_3) = (1, 7)$ and $(\theta_1, \theta_2) = (\frac{\pi}{12}, \frac{\pi}{12})$; that is, when the direction of $\boldsymbol{b}$ projected onto the 1–2 plane is not so different from that of the major axis of the ellipsoid M, R^* is large compared with R_{LS}.

(3) For any fixed value of σ_i's, the difference is monotone decreasing of θ_2 in most cases; that is, when the vector $\boldsymbol{b}$ is far from the ellipsoidal cone N, the difference of R^* and R_{LS} is large.

For the value of R^*, the following findings are obtained:

(4) The value of R^* is large for large values of θ_2, when the vector $\boldsymbol{b}$ is close to the ellipsoidal cone N.

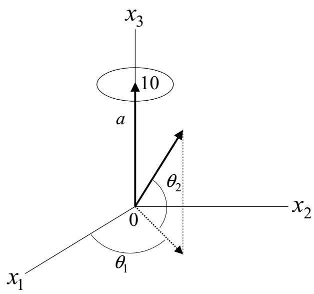

Figure 3. Generation of artificial data.

Table 1. Summary of results for artificial data.

$\Sigma^{\dagger}$	θ_1	θ_2	t	R^*	R_{LS}	d	Σ	θ_1	θ_2	t	R^*	R_{LS}	d
171	$\frac{\pi}{12}$	$\frac{\pi}{12}$	1	0.404	0.274	0.538	771	$\frac{\pi}{12}$	$\frac{\pi}{12}$	1	0.442	0.287	0.576
			10	0.404	0.377	0.127				10	0.442	0.438	0.015
			20	0.404	0.402	0.009				20	0.442	0.433	0.039
	$\frac{\pi}{12}$	$\frac{\pi}{4}$	1	0.790	0.716	0.180		$\frac{\pi}{12}$	$\frac{\pi}{4}$	1	0.814	0.724	0.209
			10	0.790	0.787	0.008				10	0.814	0.805	0.021
			20	0.790	0.784	0.016				20	0.814	0.769	0.107
	$\frac{\pi}{4}$	$\frac{\pi}{12}$	1	0.613	0.307	0.748		$\frac{\pi}{4}$	$\frac{\pi}{12}$	1	0.677	0.372	0.698
			10	0.613	0.585	0.087				10	0.677	0.677	0.000
			20	0.613	0.613	0.000				20	0.677	0.651	0.075
	$\frac{\pi}{4}$	$\frac{\pi}{4}$	1	0.887	0.733	0.317		$\frac{\pi}{4}$	$\frac{\pi}{4}$	1	0.907	0.768	0.282
			10	0.887	0.881	0.012				10	0.907	0.900	0.015
			20	0.887	0.879	0.016				20	0.907	0.866	0.089
111	$\frac{\pi}{12}$	$\frac{\pi}{12}$	1	0.354	0.269	0.421	711	$\frac{\pi}{12}$	$\frac{\pi}{12}$	1	0.390	0.274	0.506
			10	0.354	0.339	0.082				10	0.390	0.387	0.020
			20	0.354	0.353	0.006				20	0.390	0.383	0.036
	$\frac{\pi}{12}$	$\frac{\pi}{4}$	1	0.774	0.713	0.152		$\frac{\pi}{12}$	$\frac{\pi}{4}$	1	0.798	0.717	0.194
			10	0.774	0.771	0.007				10	0.798	0.790	0.022
			20	0.774	0.770	0.012				20	0.798	0.753	0.110
	$\frac{\pi}{4}$	$\frac{\pi}{12}$	1	0.354	0.269	0.421		$\frac{\pi}{4}$	$\frac{\pi}{12}$	1	0.390	0.274	0.507
			10	0.354	0.339	0.082				10	0.390	0.386	0.020
			20	0.354	0.353	0.006				20	0.390	0.383	0.036
	$\frac{\pi}{4}$	$\frac{\pi}{4}$	1	0.774	0.713	0.152		$\frac{\pi}{4}$	$\frac{\pi}{4}$	1	0.798	0.717	0.194
			10	0.774	0.771	0.007				10	0.798	0.790	0.022
			20	0.774	0.770	0.012				20	0.798	0.753	0.110
117	$\frac{\pi}{12}$	$\frac{\pi}{12}$	1	0.749	0.340	0.794	717	$\frac{\pi}{12}$	$\frac{\pi}{12}$	1	0.839	0.454	0.707
			10	0.749	0.740	0.025				10	0.839	0.837	0.006
			20	0.749	0.749	0.001				20	0.839	0.793	0.107
	$\frac{\pi}{12}$	$\frac{\pi}{4}$	1	0.973	0.750	0.405		$\frac{\pi}{12}$	$\frac{\pi}{4}$	1	0.990	0.812	0.326
			10	0.973	0.967	0.012				10	0.990	0.986	0.008
			20	0.973	0.971	0.004				20	0.990	0.951	0.077
	$\frac{\pi}{4}$	$\frac{\pi}{12}$	1	0.613	0.307	0.748		$\frac{\pi}{4}$	$\frac{\pi}{12}$	1	0.677	0.372	0.698
			10	0.613	0.585	0.087				10	0.677	0.677	0.000
			20	0.613	0.613	0.000				20	0.677	0.651	0.075
	$\frac{\pi}{4}$	$\frac{\pi}{4}$	1	0.887	0.733	0.317		$\frac{\pi}{4}$	$\frac{\pi}{4}$	1	0.907	0.768	0.282
			10	0.887	0.881	0.012				10	0.907	0.900	0.015
			20	0.887	0.879	0.016				20	0.907	0.866	0.089

† $\Sigma = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$, and three-digit numbers represent the values of σ_i's; for example, "171" means $\sigma_1 = 1$, $\sigma_2 = 7$, and $\sigma_3 = 1$.

(5) If $\sigma_3 < \sigma_2$ holds, R^* is monotone increasing with respect to θ_1; or equivalently if $\sigma_3 > \sigma_2$ holds, R^* is monotone decreasing with respect to θ_1.

From the above examinations, the difference between R^* and R_{LS} becomes especially large when the length of $\boldsymbol{b}$ is small ($t = 1$), the value of σ_i's is large,

Table 2. Summary of results for entrance examination data.

$1/\sigma$	Japanese	Social Sci.	Mathematics	Natural Sci.	English	R^*/R_{LS}
20	0.176	0.195	0.226	0.219	0.171	0.499
	0.208	0.209	0.242	0.221	0.214	0.487
16	0.170	0.193	0.231	0.224	0.163	0.504
	0.210	0.211	0.252	0.226	0.217	0.489
14	0.165	0.191	0.234	0.227	0.157	0.508
	0.212	0.212	0.260	0.230	0.219	0.490
12	0.158	0.189	0.238	0.231	0.149	0.512
	0.214	0.215	0.270	0.235	0.222	0.492
10	0.148	0.186	0.242	0.237	0.137	0.519
	0.216	0.217	0.284	0.242	0.226	0.494
8	0.133	0.180	0.246	0.245	0.119	0.528
	0.220	0.222	0.305	0.253	0.232	0.497
6	0.106	0.166	0.246	0.255	0.088	0.542
	0.226	0.229	0.340	0.271	0.242	0.502
4	0.049	0.120	0.220	0.267	0.032	0.567
	0.236	0.243	0.410	0.309	0.259	0.510

Table 3. Comparison of optimal weights and least squares weights. Weights are normalized to be $\sum_{i=1}^{5} w_i = 1$.

w^*

$1/\sigma$	w_1^*	w_2^*	w_3^*	w_4^*	w_5^*
20	0.178	0.198	0.229	0.222	0.173
16	0.173	0.197	0.235	0.228	0.166
14	0.169	0.196	0.240	0.233	0.161
12	0.164	0.196	0.247	0.239	0.154
10	0.156	0.196	0.255	0.249	0.144
8	0.144	0.195	0.267	0.265	0.129
6	0.123	0.193	0.286	0.296	0.102
4	0.071	0.174	0.320	0.388	0.047

w^{LS}

$1/\sigma$	w_1^{LS}	w_2^{LS}	w_3^{LS}	w_4^{LS}	w_5^{LS}
20	0.190	0.191	0.221	0.202	0.196
16	0.188	0.189	0.226	0.203	0.194
14	0.187	0.187	0.229	0.203	0.193
12	0.185	0.186	0.234	0.203	0.192
10	0.182	0.183	0.240	0.204	0.191
8	0.179	0.180	0.248	0.205	0.188
6	0.173	0.175	0.260	0.207	0.185
4	0.162	0.167	0.281	0.212	0.178

and θ_2 is small. For example, in the case of $(\sigma_1, \sigma_2, \sigma_3) = (1, 1, 7), \theta_1 = \theta_2 = \frac{\pi}{12}$, and $t = 1$, the values of R^* and R_{LS} are 0.749 and 0.340, respectively. So the relative difference is $d = \frac{(0.749)^2-(0.340)^2}{(0.749)^2} = 0.794$.

Entrance examination data. As a typical application of our method, we analyzed entrance examination data of 198 candidates, which were obtained from a university in Japan. The data consisted of two kinds of tests; one which was administered by the university, and the other that was administered by The National Center for University Entrance Examinations (NCUEE). For these examination data, we investigated how well we could predict the scores of the university examination from those of the NCUEE on five subjects (Japanese, Social Science, Mathematics, Natural Science, and English). Numerical computations were carried out under the assumptions that $\boldsymbol{w}_0 = (0.2, 0.2, \ldots, 0.2)'$, and $(\boldsymbol{w} - \boldsymbol{w}_0)'\Sigma^{-1}(\boldsymbol{w} - \boldsymbol{w}_0) \leq 1$, where $\Sigma = \sigma I, \sigma = (\frac{1}{20}, \frac{1}{16}, \frac{1}{14}, \frac{1}{12}, \frac{1}{10}, \frac{1}{8}, \frac{1}{6}, \frac{1}{4})$. The obtained results are shown in Table 2, where the upper and the lower numerals in each cell are corresponding to the optimal solutions and least squares solutions, respectively. From the last column of the table, the difference of R^* and R_{LS} is monotone increasing with respect to σ. When $\sigma = \frac{1}{4}$, the relative difference d, defined before, is fairly large as $d = \frac{(0.567)^2-(0.510)^2}{(0.567)^2} = 0.191$. Table 3 shows how w_i^*'s or w_i^{LS}'s are changed according to σ's. From this, the change of w_i^*'s is larger than that of w_i^{LS}'s, which is remarkable in the part of σ's being large.

7. Concluding remarks

In this paper, we addressed one of the simplest canonical correlation problems in which the correlation between an observed variable and its linear predictor was optimized under a given quadratic constraint. We derived some algebraic properties of the optimal solution, together with a computational algorithm for it. We also confirmed the effectiveness of the algorithm by analyzing a set of systematically generated artificial data, and especially examined how our optimal solutions differ from least squares solutions.

Regarding natural extensions of our algorithm, we can consider at least two extensions. One is the extension to the algorithm that is applicable to the correlation maximization problem under quadratic and linear constraints. For this extension, we are expecting that the problem with quadratic and linear constraints may be transformed to that with a quadratic constraint. The other extension is to canonical correlation analysis. Although we have not considerd this extension in detail, we believe that our current approach will be helpful to see possibilities of the extension, since our algorithm is just a special case of it.

Appendix: Derivation of (5.3)

We give the solution $\tilde{\boldsymbol{x}}$ for Problem 2 under the condition $\boldsymbol{b}'\boldsymbol{v}_p = 0$. As shown in the proof of Theorem 4, the solution $\tilde{\boldsymbol{x}}$ must satisfy the following equations;

$$\tilde{\boldsymbol{x}}'P\tilde{\boldsymbol{x}} = 0,$$

and

$$\tilde{\boldsymbol{x}} + \alpha P \tilde{\boldsymbol{x}} = \boldsymbol{b}.$$

Multiplying both sides of the latter by $\boldsymbol{v}_p'$, we obtain $\alpha = -1/\gamma_p$, since $P\boldsymbol{v}_p = \gamma_p \boldsymbol{v}_p$ and $\boldsymbol{v}_p' \tilde{\boldsymbol{x}} \neq 0$. Hence the solution $\tilde{\boldsymbol{x}}$ satisfies the equation

$$\left(I - \frac{1}{\gamma_p} P\right) \tilde{\boldsymbol{x}} = \boldsymbol{b}.$$

Apparently the matrix $I - \frac{1}{\gamma_p} P$ has the eigenvalues $\delta_j = 1 - \frac{\gamma_j}{\gamma_p}$, $j = 1, 2, \ldots, p$, and $\boldsymbol{v}_j$ is an eigenvector of $I - \frac{1}{\gamma_p} P$ corresponding to the eigenvalue δ_j. Then, the Moore-Penrose generalized inverse $(I - \frac{1}{\gamma_p} P)^+$ of $I - \frac{1}{\gamma_p} P$ can be defined by

$$\left(I - \frac{1}{\gamma_p} P\right)^+ = \sum_{j=1}^{p-1} \frac{1}{\delta_j} \boldsymbol{v}_j \boldsymbol{v}_j'.$$

The solution $\tilde{\boldsymbol{x}}$ is given by

$$\tilde{\boldsymbol{x}} = \left(I - \frac{1}{\gamma_p} P\right)^+ \boldsymbol{b} + \eta \boldsymbol{v}_p,$$

where η is a constant that satisfies the condition $\tilde{\boldsymbol{x}}' P \tilde{\boldsymbol{x}} = 0$; that is, η satisfies the equation given by

$$\gamma_p \eta^2 + \boldsymbol{b}' \left(I - \frac{1}{\gamma_p} P\right)^+ P \left(I - \frac{1}{\gamma_p} P\right)^+ \boldsymbol{b} = 0,$$

since $(I - \frac{1}{\gamma_p} P)^+ \boldsymbol{v}_p = 0$. Hence, the solution $\tilde{\boldsymbol{x}}$ is given by

$$\tilde{\boldsymbol{x}} = \left(I - \frac{1}{\gamma_p} P\right)^+ \boldsymbol{b} + \sqrt{\frac{\boldsymbol{b}' \left(I - \frac{1}{\gamma_p} P\right)^+ P \left(I - \frac{1}{\gamma_p} P\right)^+ \boldsymbol{b}}{-\gamma_p}} \boldsymbol{v}_p.$$

Acknowledgements

The authors wish to thank Prof. Mamoru Hoshi of the Electro-Communication University for his helpful comments and encouragement.

References

Faraut, J. and Koranyi, A. (1994). *Analysis on Symmetric Cones*, Oxford University Press, New York.

Golub, G. and Van Loan, C. (1996). *Matrix Computations*, 3rd ed., Johns Hopkins University Press, London.

Kojima, M. (1998). Semidefinite programming relaxation and global optimization, *Proceedings of 10th RAMP Symposium in Kyoto*, *Japan*, 195–204.

Lawlor, G. R. (1991). A sufficient criterion for a cone to be area-minimizing, *AMS Memoirs*, 446.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Shapiro, A. (1997). On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints, *SIAM Journal on Optimization*, **7**, 508–518.

Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming, *SIAM Review*, **38**, 49–95.

Vavasis, S. A. (1991). *Nonlinear Optimization*, Oxford University Press, New York.