On the Solution of a Quadratic Programming Problem

Stefan M. Stefanov

Department of Informatics, Neofit Rilski South-Western University,
Blagoevgrad, BULGARIA; E-mail: stefm@swu.bg

Abstract

In this paper, the problem of minimizing/maximizing the quadratic form $\langle A\mathbf{x}, \mathbf{x}\rangle$ provided that the variables $\mathbf{x}$ belong to the unit sphere of $\mathbf{R}^n$ is considered. A nonstandard approach, that uses the Cauchy-Schwarz inequality and induced matrix norms, for solving this problem is suggested.

AMS Subject Classification. 90C20, 52A40.

Key words and phrases. Quadratic programming, vector norms, induced matrix norms, Cauchy-Schwarz inequality, eigenvalues and eigenvectors.

1 Introduction. Statement of the problem

Consider the following

Problem. Let A be a $n \times n$ symmetric matrix with real components (linear continuous self-conjugate operator), $\mathbf{x} \in \mathbf{R}^n$. Find

$$\min \langle A\mathbf{x}, \mathbf{x}\rangle \tag{1}$$

subject to

$$\langle \mathbf{x}, \mathbf{x}\rangle \equiv \|\mathbf{x}\|^2 = 1, \tag{2}$$

where $\langle ., . \rangle$ denotes the inner (scalar) product of the Hilbert space $\mathbf{R}^n$, $\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^n x_i y_i$ for each $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$.

According to a Weierstrass theorem, this problem has an optimal solution because the objective function is continuous and the feasible region is a compact set in $\mathbf{R}^n$.

A traditional way for solving this problem is that with the use of the Lagrange multipliers. It can be obtained that

$$\min_{\|\mathbf{x}\|^2=1} \langle A\mathbf{x}, \mathbf{x}\rangle = \lambda_1^*, \tag{3}$$

where λ_1^* is the least eigenvalue of A and the solution to this problem is the eigenvector $\mathbf{x}^*$ of A associated with λ_1^*.

Indeed, the Lagrangian function for problem (1) – (2) is

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \lambda_0 \langle A\mathbf{x}, \mathbf{x} \rangle + \lambda_1 \langle \mathbf{x}, \mathbf{x} \rangle.$$

According to the Lagrange multipliers theorem ([3], [5]), the necessary condition for the minimum solution is

$$\lambda_0 A\mathbf{x} + \lambda_1 \mathbf{x} = \mathbf{0},$$

where $\langle \mathbf{x}, \mathbf{x} \rangle = 1$, and the Lagrange multipliers λ_0 and λ_1 cannot be simultaneously equal to zero. If we assume that $\lambda_0 = 0$, it would follow that $\lambda_1 \mathbf{x} = \mathbf{0}$, where λ_1 must be different from zero because $\lambda_0 = 0$ by the assumption. Hence, $\mathbf{x} = \mathbf{0}$. However, $\mathbf{x} = \mathbf{0}$ is not a feasible value because $\langle \mathbf{0}, \mathbf{0} \rangle \neq 1$. Therefore, $\lambda_0 \neq 0$, and $\lambda_0 A\mathbf{x} + \lambda_1 \mathbf{x} = \mathbf{0}$. Dividing both sides of this vector equality by $\lambda_0 \neq 0$, we get $A\mathbf{x} + \frac{\lambda_1}{\lambda_0}\mathbf{x} = \mathbf{0}$, or $A\mathbf{x} = -\frac{\lambda_1}{\lambda_0}\mathbf{x}$. Hence,

$$A\mathbf{x} = \lambda_1^* \mathbf{x}, \tag{4}$$

where $\lambda_1^* \stackrel{\text{def}}{=} -\frac{\lambda_1}{\lambda_0}$, which means that the optimal solution $\mathbf{x}^*$ of problem (1) – (2) is an eigenvector of matrix A, and λ_1^* is the corresponding eigenvalue. Multiplying both sides of (4) by $\mathbf{x}$, we obtain

$$\min_{\|\mathbf{x}\|^2=1} \langle A\mathbf{x}, \mathbf{x} \rangle = \lambda_1^* \langle \mathbf{x}, \mathbf{x} \rangle = \lambda_1^*,$$

where we have used (2). This means that the optimal solution to problem (1) – (2) is the eigenvector $\mathbf{x}^*$ associated with the least eigenvalue λ_1^* of matrix A, and λ_1^* is the minimal value of the quadratic function $\langle A\mathbf{x}, \mathbf{x} \rangle$ subject to $\langle \mathbf{x}, \mathbf{x} \rangle = 1$.

In this paper, we suggest another approach for solving problem (1) – (2) which uses characteristics of this problem.

2 Preliminaries

As it is known, a linear operator is continuous if and only if it is bounded, and the concepts of "symmetry" and "self-conjugacy" are equivalent for bounded operators.

Let V be a vector space over the field $F(\mathbf{R}$ or $\mathbf{C})$.

Definition 1 ([1]) A function $\|.\| : V \to \mathbf{R}$ is called a *vector norm* if for all $\mathbf{x}, \mathbf{y} \in V$,

1) $\|\mathbf{x}\| \geq 0$ — Nonnegative

$\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$ — Positive, Nonsingular

2) $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$ for all scalars $c \in F$ — Homogeneous

3) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ — Triangle inequality.

Let M_n be the totality of all $n \times n$ square matrices over F, L be the space of continuous linear operators in F^n. We can use vector norms on F^n (on F^{n^2} for M_n) defined in Definition 1 as norms on M_n because M_n is itself a vector space of dimension n^2. However, the natural multiplication operation of matrices (of linear continuous operators) allows to introduce a new concept which turns out to be useful.

Definition 2 ([1], [2]) A function $\|.\| : M_n \to \mathbf{R}$ ($\|.\| : L \to \mathbf{R}$) is called a *matrix norm* (a *linear operator norm*) if for all $A, B \in M_n$ (if for all $A, B \in L$) it satisfies the following axioms:

1) $\|A\| \geq 0$ — Nonnegative
   $\|A\| = 0$ if and only if $A = 0$ — Positive, Nonsingular
2) $\|cA\| = |c|\|A\|$ for all scalars $c \in F$ — Homogeneous
3) $\|A + B\| \leq \|A\| + \|B\|$ — Triangle inequality
4) $\|AB\| \leq \|A\|\|B\|$ — Submultiplicative.

Since the linear operators in $\mathbf{R}^n$ can be identified with matrices, we are able to consider both cases simultaneously and it is sufficient to deal with matrix norms only. Moreover, a linear operator $\mathcal{A}$ in $\mathbf{R}^n$ is self-conjugate if and only if the matrix A associated with $\mathcal{A}$ is symmetric.

Some of vector norms turn out to be matrix norms in the vector space M_n in the sense of Definition 2, other vector norms are not matrix norms.

Definition 3 ([1], [4]) Let $\|.\|$ be a vector norm on $\mathbf{C}^n$. Define *matrix norm* $\|.\|$ on M_n by

$$\|A\| := \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|. \tag{5}$$

Remark 1 *We can use "max" rather than "sup" in Definition 3, according to Weierstrass theorem, because* $\|A\mathbf{x}\|$ *is a continuous function of* $\mathbf{x}$ *and the unit sphere* $S_{\|.\|} \equiv \{\mathbf{x} \in V : \|\mathbf{x}\| = 1\}$ *of the vector norm is a compact set in the finite-dimensional vector space* V.

Theorem 1 ([1]) *The function* $\|.\|$ *defined in Definition 3 is a matrix norm on* M_n, $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$ *for all* $A \in M_n$ *and for all* $\mathbf{x} \in \mathbf{C}^n$, *and* $\|I\| = 1$, *where* I *is the identity.*

Definition 4 The matrix norm from Definition 3 is said to be *induced* by the vector norm $\|.\|$. Sometimes it is called *operator norm* or the *least upper bound norm* associated with the vector norm $\|.\|$.

Remark 2 *The inequality* $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$ *in the statement of Theorem 1 shows that the induced matrix norm is compatible with the corresponding vector norm* $\|.\|$.

It is worthy to note another property of the induced matrix norms – "minimality", according to which each induced matrix norm $\|.\|_m$ is minimal, that is, it is such that $\|A\|_m \leq \|A\|$ for all $A \in M_n$ and for any other matrix norm $\|.\|$ on M_n.

For example, matrix norm induced by the Euclidean (Hilbert, l_2-) vector norm

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \tag{6}$$

is the so-called *spectral norm*: $\|A\|_2 = \sqrt{\lambda_1^*}$, where λ_1^* is the largest eigenvalue of A^*A, A^* is the matrix (operator) conjugate to A.

Remark 3 *Since* A^*A *is a symmetric and positive definite matrix (symmetric and positive operator) then all its eigenvalues are real and nonnegative.*

There is only one minimal norm in the class of unitarily invariant matrix norms (that is, matrix norms $\|.\|$ such that $\|A\| = \|UAV\|$ for all $A \in M_n$ and for all unitary matrices $U, V \in M_n$) and it is namely the spectral norm.

The conjugate space to the Hilbert space $\mathbf{R}^n$ with l_2- norm is $\mathbf{R}^n$ itself with l_2- norm. More generally, the conjugate space H^* to a Hilbert space H is the same Hilbert space H, that is, $H^* = H$.

3 Main result

According to the Cauchy-Schwarz inequality $|\langle A\mathbf{x},\mathbf{x}\rangle| \le \|A\mathbf{x}\|\|\mathbf{x}\|$ we have

$$-\|A\mathbf{x}\|\|\mathbf{x}\| \ \le\ \langle A\mathbf{x},\mathbf{x}\rangle \ \le \|A\mathbf{x}\|\|\mathbf{x}\|$$

where

a) the left inequality is satisfied with equality if and only if $A\mathbf{x} = \alpha\mathbf{x}$,

b) the right inequality is satisfied with equality if and only if $A\mathbf{x} = \beta\mathbf{x}$

for some $\alpha, \beta \in F$.

By the assumption, $\|\mathbf{x}\| \overset{\text{def}}{=} \sqrt{\langle \mathbf{x},\mathbf{x}\rangle} = 1$. Therefore,

$$\min_{\|\mathbf{x}\|=1} (-\|A\mathbf{x}\|\|\mathbf{x}\|) \le \min_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle,$$

$$\max_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle \ \le \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|\|\mathbf{x}\|.$$

It follows that

$$-\max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| \le \min_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle,$$

$$\max_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle \ \le \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|,$$

where we have used that $\min\{-f(\mathbf{x}), \mathbf{x} \in X\} = -\max\{f(\mathbf{x}), \mathbf{x} \in X\}$ for each objective function f and for each feasible region X. In accordance with Definition 3 we get

$$-\|A\| \le \min_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle,$$

$$\max_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle \ \le \|A\|.$$

Concerning Euclidean vector norm of $\mathbf{x}$ we have

$$-\sqrt{\lambda_1^*} \le \min_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle,$$

$$\max_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle \ \le \sqrt{\lambda_1^*}$$

where λ_1^* is the largest eigenvalue among eigenvalues $\lambda_j^*,\ j = 1,\dots,n$ of A^TA. However, A is symmetric matrix (operator). Therefore, $A^T = A, A^TA = A^2$, hence, $\lambda_j^* = \lambda_j^2,\ j = 1,\dots,n$, where $\lambda_j, j = 1,\dots,n$ are the eigenvalues of A. In particular, $\lambda_1^* = \lambda_1^2$. Therefore,

$$-|\lambda_1| \le \min_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle, \tag{7}$$

$$\max_{\|\mathbf{x}\|=1} \langle A\mathbf{x},\mathbf{x}\rangle \ \le |\lambda_1|, \tag{8}$$

where λ_1 is the eigenvalue of A with the largest absolute value.

According to a) and b), the left and the right inequalities are satisfied as equalities with $A\mathbf{x} = \lambda_m\mathbf{x}$ and $A\mathbf{x} = \lambda_M\mathbf{x}$, respectively, that is, for the eigenvectors $\mathbf{x}_m$ and $\mathbf{x}_M$ associated with the eigenvalues λ_m and λ_M of A, respectively, and the minimum and the maximum value of $\langle A\mathbf{x},\mathbf{x}\rangle$ with $\|\mathbf{x}\| = 1$ are λ_m and λ_M, respectively. Indeed,

$$\min_{\|\mathbf{x}\|=1} \langle A\mathbf{x}, \mathbf{x}\rangle = \langle \lambda_m \mathbf{x}_m, \mathbf{x}_m\rangle = \lambda_m \langle \mathbf{x}_m, \mathbf{x}_m\rangle = \lambda_m, \tag{9}$$

and

$$\max_{\|\mathbf{x}\|=1} \langle A\mathbf{x}, \mathbf{x}\rangle = \langle \lambda_M \mathbf{x}_M, \mathbf{x}_M\rangle = \lambda_M \langle \mathbf{x}_M, \mathbf{x}_M\rangle = \lambda_M. \tag{10}$$

Since λ_m and λ_M are eigenvalues of A and since (9), (10) hold then λ_m is the least eigenvalue of A and λ_M is the largest eigenvalue of A. If we assume the contrary, that there exist a smaller eigenvalue α of A than λ_m and a larger eigenvalue β of A than λ_M it would follow that

$$\min_{\|\mathbf{x}\|=1} \langle A\mathbf{x}, \mathbf{x}\rangle = \langle \alpha \mathbf{x}, \mathbf{x}\rangle = \alpha \langle \mathbf{x}, \mathbf{x}\rangle = \alpha < \lambda_m,$$

and

$$\max_{\|\mathbf{x}\|=1} \langle A\mathbf{x}, \mathbf{x}\rangle = \langle \beta \mathbf{x}, \mathbf{x}\rangle = \beta \langle \mathbf{x}, \mathbf{x}\rangle = \beta > \lambda_M$$

which is in contradiction with (9) and (10).

Thus,

$$\min_{\|\mathbf{x}\|=1} \langle A\mathbf{x}, \mathbf{x}\rangle = \lambda_m \tag{11}$$

and the solution to the original problem $(1) - (2)$ is the eigenvector $\mathbf{x}_m$ of A associated with λ_m, and at the same time we obtained that

$$\max_{\|\mathbf{x}\|=1} \langle A\mathbf{x}, \mathbf{x}\rangle = \lambda_M \tag{12}$$

and the solution to this problem is the eigenvector $\mathbf{x}_M$ of A associated with λ_M, where λ_m and λ_M were defined above.

4 Conclusions

As it is reasonable to expect, both approaches – Lagrange multipliers theorem on the one hand, and the Cauchy-Schwarz inequality along with the properties of induced matrix norms, on the other hand – for solving the problem under consideration give the same result.

References

[1] R. Horn, C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[2] V.C.L. Hutson, J.S. Pym, *Applications of Functional Analysis and Operator Theory*, Academic Press, New York, 1980.

[3] O.L. Mangasarian, *Nonlinear Programming*, 2nd ed., SIAM, Philadelphia,1994.

[4] A. Ralston, P. Rabinowitz, *A First Course in Numerical Analysis*, 2nd ed., McGraw-Hill, New York, 1978.

[5] S.M. Stefanov, *Separable Programming: Theory and Methods*, Kluwer Academic Publishers, Dordrecht-Boston-London, 2001.