A NUMERICAL METHOD TO SOLVE A QUADRATIC CONSTRAINED 
MAXIMIZATION

ALIREZA ESNA ASHARI∗, RAMINE NIKOUKHAH†, AND STEPHEN L. CAMPBELL ‡

Abstract. The problem of maximizing a quadratic function subject to an ellipsoidal constraint is considered. The algorithm produces an approximation of the optimal solution in a finite number of iterations. In particular, the method can be used to solve the ill-conditioned problems in which the solution consists of two parts from two orthogonal subspaces. Without restrictive assumptions, the solution generated by the method satisfies the first and second order necessary conditions for a maximizer of the objective function. Numerical experiments clearly demonstrate that the method is successful.

Key words. constrained optimization, quadratic optimization, lagrangian method

1. Introduction. In many identification and fault detection problems uncertainties are represented by unknown parameters belonging to predefined ellipsoids. Basically there are two ways of treating initial state uncertainty, and system dynamic disturbances. One way to treat information about these variables is to use stochastic modeling and consider the initial condition and the inputs to be random variables and stochastic processes. This type of approaches does not directly define a set of behaviors for the system, but associates to each trajectory a probability.

An alternative approach is to consider input noises and initial states unknown except for the fact that they belong to given bounded sets [1]. Therefore, the information of system states in normal or faulty modes is described by a family of sets. Especially, for the purpose of fault detection, it is enough to test the membership of the measurements [6] in these sets. The model selection methods based on this approach are "Robust" in the sense that the detection of the fault is guaranteed for all the predefined set of uncertainties. In these types of approaches, we may face the problem of finding the worst case uncertainty for a given input. This application is a major motivation for studying a group of constrained quadratic optimization problems

\[ J = \max_{\mu} \|a + B\mu\|^2, \quad s.t. \quad \|\mu\|^2 \leq \Delta, \]  

(1.1)

where \(a \in \mathbb{R}^n\) and \(B \in \mathbb{R}^{n \times m}\) are given vector and matrix respectively, of proper dimensions and \(\mu \in \mathbb{R}^m\) is an unknown vector to be calculated. \(\|\cdot\|\) is the Euclidean norm in \(\mathbb{R}^n\). This type of maximization is usually the inner maximization of the Min-Max problem which we may face in signal design problems for model selection. Studying the properties of this problem helps us to understand some unusual attributes of those types of problems [3].

In the literature, similar quadratic problems have been considered and solved approximately. As an example, consider the following minimization,

\[ J = \min_{\mu} g^T \mu + \frac{1}{2} \mu^T C \mu, \quad s.t. \quad \|\mu\|^2 \leq \Delta. \]  

(1.2)

Problem (1.2) has attracted much attention, as it is used in computing trust regions (see [5, 2]). Several algorithms have been proposed to solve (1.2) based on the assumption that

∗INRIA, Rocquencourt BP 105, 78153 Le Chesnay Cedex, France (Alireza.esna_ashari@inria.fr).
†INRIA, Rocquencourt BP 105, 78153 Le Chesnay Cedex, France (Ramine.nikoukhah@inria.fr).
‡Department of Mathematics and Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC 27695-8205, USA (slc@math.ncsu.edu). Research supported in part by NSF Grant DMS-0907832.
it has a solution on the boundary $\|\mu\| = \Delta$. It is the same situation that occurs in (1.1). Note that (1.2) is more general than (1.1). Both of the problems (1.1) and (1.2) look simple. However, there are some difficulties in some special cases of these problems.

In (1.1) the solution is on the boundary of the set of uncertainties $\mu$. Assuming $\Delta = 1$, the hard case is when we cannot find any $\lambda \geq 0$ such that $\lambda I - B^T B$ is positive definite and the solution satisfies

$$\|(\lambda I - B^T B)^{-1} B^T a\|^2 = 1,$$

where $I$ represents the identity matrix. The problem turns out to be similar to the hard case of problem (1.2). The solution of (1.2) is considered in [7, 5]. In [5] it is explained how in a similar hard case this leads to numerical difficulties. An approximate solution to this problem is provided in [5] which gives a nearly optimal solution which is near enough to the optimal solution (see also [4, 7, 2]).

In this paper, knowing that the hard case occurs, we develop an algorithm for problem (1.1) that, unlike the approach of [5], gives an exact solution to (1.1). Considering the structure of the problem, the proposed algorithm uses the bisection method which is also capable of solving the hard cases. Based on the complete solution to (1.1) we propose a recursive algorithm to find the solution in the hard case. Also we consider the solution of a dynamic counterpart optimization, which can be solved by using a dynamic programming approach and designing a filter to calculate the bound on uncertainties in the forward solution of a Riccati equation.

2. Main results.

2.1. Static Maximization. In order to propose an efficient approach to solve the optimization problem, we need to understand the structure of the problem. Problem (1.2) has been already investigated in [7]. In this paper we study the structure of (1.1) in a different way which is more compatible with the algorithm given in Section 3 and the utilized bisec-

THEOREM 2.1. Consider the following optimization problem

$$J = \max_{\mu} \|a + B\mu\|^2, \quad \text{s.t.} \quad \|D\mu\|^2 \leq 1,$$

where $B$ and $D$ are matrices, not necessarily square, and $a$ is a vector with proper dimensions. $D$ is assumed to be full column rank. Let $\bar{\lambda}$ denote the maximum eigenvalue of the matrix pair $\{B^T B, D^T D\}$ for which $\bar{\lambda} D^T D - B^T B$ is singular and denote the optimum value of $\mu$ by $\mu^*$. Also, consider the condition

$$\text{Left} - \text{Kernel}(\bar{\lambda} D^T D - B^T B) \subseteq \text{Left} - \text{Kernel}(B^T a).$$

There are two possibilities for the solution of $J$:

1. If (2.2) is not satisfied, or if (2.2) is satisfied but

$$\|(\bar{\lambda} D^T D - B^T B)^{\dagger} B^T a\| > 1,$$

then

$$\mu^* = (\bar{\lambda}^* D^T D - B^T B)^{-1} B^T a.$$
\( \lambda^* \) is the value of \( \lambda > \bar{\lambda} \) for which \( \|D\mu^*\|^2 = 1 \) and \((\dagger)\) represents the pseudo-inverse.

2. if not, then

\[
\mu^* = (\bar{\lambda}D^TD - B^TB)^\dagger B^Ta + \Psi \zeta, \tag{2.4}
\]

where \( \Psi \in \text{Right–Kernel} (\bar{\lambda}D^TD - B^TB) \) is a full column rank matrix and \( \zeta \) is calculated such that

\[
\|\zeta\| = \sqrt{1 - \| (\bar{\lambda}D^TD - B^TB)^\dagger B^Ta \|^2}. \tag{2.5}
\]

**Proof.** First of all we should note that the solution occurs on the boundary where \( \|D\mu\| = 1 \). Using the Lagrangian method, the Lagrangian is

\[
J = \|a + B\mu\|^2 - \lambda\|D\mu\|^2. \tag{2.6}
\]

Taking the derivative of \( J \) with respect to \( \mu \) gives

\[
(B^TB - \lambda D^TD)\mu = -B^Ta. \tag{2.7}
\]

In order to have a maximum, it is necessary that

\[
\lambda \geq \bar{\lambda}. \tag{2.8}
\]

Assuming that \( \lambda > \bar{\lambda} \), we may have

\[
\mu = (\lambda D^TD - B^TB)^{-1}B^Ta. \tag{2.9}
\]

We define

\[
f(\lambda) = \|D\mu\|^2. \tag{2.10}
\]

Here, \( f(\lambda) \) is a monotonically decreasing function of \( \lambda \) on \((\bar{\lambda}, \infty)\). It is easy to show by calculating the derivative of \( f(\lambda) \) with respect to \( \lambda \) that

\[
\frac{df(\lambda)}{d\lambda} = -2a^TB(\lambda D^TD - B^TB)^{-1}
\]

\[
(D^TD)(\lambda D^TD - B^TB)^{-1}D^TD(\lambda D^TD - B^TB)^{-1}B^Ta, \tag{2.11}
\]

which is negative over the range of \( \lambda \) in (2.8). On the other hand

\[
\lim_{\lambda \to \infty} f(\lambda) = 0, \tag{2.12}
\]

and

\[
\lim_{\lambda \to \lambda} f(\lambda) = \infty. \tag{2.13}
\]

So, for some value of \( \lambda \), \( f(\lambda) \) equals one. If (2.2) is satisfied, the same situation occurs. In this case

\[
\lim f(\bar{\lambda}) = (\bar{\lambda}D^TD - B^TB)^\dagger B^Ta. \tag{2.14}
\]
Here, if \( f(\bar{\lambda}) > 1 \), it equals 1 for a greater value of \( \lambda \) and \( \mu^* \) is calculated from (2.3). If \( f(\bar{\lambda}) < 1 \), we can make it equal to one as in (2.4).

**Corollary 2.2.** \( f(\lambda) = \|D\mu\|^2 \) is monotonically decreasing function of \( \lambda \) on \((\bar{\lambda}, \infty)\).

The result of corollary (2.2) is useful for us to propose the algorithm in the next section.

To simplify our notation we define

\[
\Pi = (\bar{\lambda}D^T D - B^T B)^\dagger B^T a. \tag{2.15}
\]

Now, we are ready to give a solution for the second case of Theorem 2.1. We need to define vector \( \zeta \) in such a case. Let us put \( \mu \) from (2.4) in the original problem to obtain the following new optimization problem

\[
C = \max \|a + B\Pi + B\Psi \zeta\|^2 \quad s.t. \|D\Pi + D\Psi \zeta\|^2 = 1. \tag{2.16}
\]

Again we use Lagrangian method, defining \( J_2 \) as

\[
J_2 = \|a + B\Pi + B\Psi \zeta\|^2 - \rho \|D\Pi + D\Psi \zeta\|^2 \tag{2.17}
\]

where \( \rho \) is a Lagrange multiplier. The first derivative of (2.17) is put equal to zero to find the stationary points of the optimization problem

\[
\frac{\partial J_2}{\partial \zeta} = 0, \tag{2.18}
\]

that is,

\[
(\Psi^T B^T B \Psi - \rho \Psi^T D^T D \Psi) \zeta + \Psi^T (B^T a + B^T B\Pi - \rho D^T D\Pi) = 0 \tag{2.19}
\]

In order for the solution to be a maximum of the problem, we need to check the second derivative of \( J_2 \):

\[
\frac{\partial^2 J_2}{\partial \zeta^2} \leq 0, \tag{2.20}
\]

that is

\[
\Psi^T (B^T B - \rho D^T D) \Psi \leq 0. \tag{2.21}
\]

A sufficient condition for satisfying (2.21) is

\[
\bar{\lambda} \leq \rho. \tag{2.22}
\]

If \( \rho \neq \bar{\lambda} \), then the solution \( \zeta \) is obtained as follows,

\[
\zeta = (\rho \Psi^T D^T D \Psi - \Psi^T B^T B \Psi)^{-1} \Psi^T (B^T a + B^T B\Pi - \rho D^T D\Pi). \tag{2.23}
\]

However, considering that the solution of (2.1) occurs on \( \|D\mu\| = 1 \) and \( \rho = \lambda \) and \( C = J \) it is concluded that

\[
\rho = \bar{\lambda}. \tag{2.24}
\]

But (2.19) is already satisfied for \( \rho = \bar{\lambda} \) for any selection of \( \zeta \). Thus we have freedom in the selection of this vector.
3. An algorithm to find the solution. There is a Lagrange multiplier, \( \lambda \), which should be determined to solve the problem. It is found using a bisection method where

\[
\lambda \geq \bar{\lambda}.
\]

Here, the algorithm to find the multiplier and the final solution of (2.1) is summarized. To solve the problem (1.1), we choose \( D \) as the identity matrix. However, in the hard case, the solution comes from a new optimization problem of the general form (2.1), but has less dimensions. This leads to a recursive algorithm. Note that \( B^T B - \lambda D^T D \) may have more than one zero eigenvalue.

1. Select a large enough upper bound for \( \lambda \), and call it \( \lambda_{max} \). The lower limit is set to \( \lambda_{min} = \bar{\lambda} \). These two variables may change during the algorithm.
2. Select \( \lambda \) as \( \lambda = (\lambda_{max} + \lambda_{min})/2 \). If |\( \lambda - \bar{\lambda} \) < \( \epsilon \) for a small positive value of \( \epsilon \), solve the problem for \( \bar{\lambda} \).
3. If \( \lambda = \bar{\lambda} \), go to step 5. Otherwise calculate \( \mu = (\lambda D^T D - B^T B)^{-1} B^T a \) for the selected \( \lambda \).
4. Let \( k = ||D\mu||^2 \). The algorithm stops when \( k = 1 \). In this case, \( \mu = (\lambda D^T D - B^T B)^{-1} B^T a \) is the solution. If \( k > 1 \), set \( \lambda_{max} = \lambda \), and if \( k < 1 \), set \( \lambda_{min} = \lambda \), and go to step 2.
5. If \( \lambda = \bar{\lambda} \), calculate \( (\bar{\lambda} D^T D - B^T B)^{-1} B^T a \) and \( \Psi \) as the basis of the right-kernel of \( \bar{\lambda} D^T D - B^T B \).
6. If \( \zeta \) is a scalar, solve a second order equation \( ||D\Psi\zeta||^2 = 1 - ||D\Pi||^2 \) to find \( \zeta \). In this case the algorithm terminates at this point. Otherwise go to the next step.
7. If \( \zeta \) is a vector, then there are multiple solutions. In order to determine a solution, introduce a new quadratic cost function by defining \( \bar{a} \) and \( \bar{B} \) to get a new problem

\[
J = \max_{\zeta} ||\bar{a} + \bar{B}\zeta||^2, \quad s.t. \, ||D\Psi\zeta||^2 = 1 - ||D\Pi||^2. \quad (3.1)
\]

\( \bar{a} \) and \( \bar{B} \) are free to meet any other design considerations.
8. Call this algorithm from the first step to solve (3.1). Note that (3.1) is a problem of type (2.1), but now the dimensions of \( \zeta \) is less than the dimensions of \( \mu \).

4. Examples. Consider the following system:

\[
a = \begin{pmatrix} -0.0068 \\ 4.3380 \\ -0.9731 \end{pmatrix}, \quad B = \begin{pmatrix} -3.4010 & 1.5781 & 0.0812 \\ -0.2067 & -0.4676 & 0.6879 \\ -1.2059 & -2.7120 & 2.1410 \end{pmatrix}.
\]

The set of eigenvalues of the matrix \( B^T B \) is \{0.0637, 14.0637, 14.0637\}. So we have \( \bar{\lambda} = 14.0637 \). This is an example of the hard case where the \( \lambda \) determined by the algorithm converges to \( \bar{\lambda} \) and we need to solve the problem for \( \lambda = \bar{\lambda} \). We calculate

\[
\bar{\lambda} I - B^T B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}, \quad B^T a = \begin{pmatrix} 0.3 \\ 0.6 \\ 0.9 \end{pmatrix}.
\]

In this case \( \zeta \) is a vector of dimension 2 and

\[
\Psi = \begin{pmatrix} -0.8729 & 0.4082 \\ -0.2182 & -0.8165 \\ 0.4364 & 0.4082 \end{pmatrix}, \quad \Pi = \begin{pmatrix} 0.0214 \\ 0.0429 \\ 0.0043 \end{pmatrix}.
\]
A choice for $\zeta$ and $\mu$ is

$$
\zeta = \begin{pmatrix} 0.1598 \\ 0.9839 \end{pmatrix}, \quad \mu = \begin{pmatrix} 0.2836 \\ -0.7954 \\ 0.5357 \end{pmatrix}.
$$

The optimal value of the cost function is $J^* = 5.8240$ for any selection of $\zeta$.

5. Conclusions. The problem of maximizing a quadratic function subject to an ellipsoidal constraint is considered and a method is given to solve the problem. In particular, the method can be used to solve the ill-conditioned problems in which the solution consists of two parts from two orthogonal subspaces. Numerical experiments clearly demonstrate that the method is successful.

REFERENCES


