A METHOD FOR THE MINIMIZATION OF A QUADRATIC CONVEX FUNCTION
OVER THE SIMPLEX

Francisco Facchinei, Roma
Stefano Lucidi, Roma

Abstract: The minimization of a quadratic function over the simplex occurs in many fields, such as non-smooth optimization, economy, statistics and management science. We propose an algorithm based on a differentiable exact penalty function that, under mild assumptions, finds the solution in a finite number of steps. Numerical results show that this approach is not affected by the ill conditioning, the degeneracy and the dimension of the problem.


1. Introduction

We are concerned with the minimization of a quadratic, convex function over the simplex:

\[ \minimize f(x) = \frac{1}{2} x' Q x + c' x \quad \text{s. t.} \quad e' x = 1, \quad x \geq 0 \tag{QP} \]

where \( Q \) is an \( n \times n \) positive semidefinite matrix, \( c \in \mathbb{R}^n \) and \( e \) is a \( n \)-dimensional vector of ones.

Problem (QP) abounds with applications, and has been intensively studied in recent years, see, e.g., /1/, /2/ and references therein.

In this paper we propose an algorithm for the solution of Problem (QP) based on the unconstrained minimization of a continuously differentiable exact penalty function that fully exploits the particular structure of the problem. The algorithm introduced conciliates many of the advantages of iterative methods and finite methods for the solution of quadratic programming problems. In fact on one hand it is capable of dropping and adding a large number of constraints to the active set at each iteration and it operates on input data only, thus preserving any sparsity pattern of the data, and, on the other hand, it can be shown to find the solution in a finite number of steps under a mild assumption weaker than strict convexity of the objective function. Reported numerical results on strictly convex problems indicate that the method is practically insensitive to the ill conditioning of the objective function and to the degeneracy of the solution; furthermore, the number of iterations does not appear to depend on the dimension of the problem.

In the sequel we shall employ the following notation. Let \( M \) be an \( n \times n \) matrix with columns \( M_i, \ i = 1, \ldots, n \), and let \( K \) be an index set such that \( K \subseteq \{1, \ldots, n\} \), we denote by \( M_K \) the submatrix of \( M \) consisting of columns \( M_i, \ i \in K \). Let \( w \) be an \( n \) vector, we denote by \( w_K \) the subvector with components \( w_i, \ i \in K \). We indicate by \( E \) the \( n \times n \) indentity matrix and by \( \| \cdot \| \) the euclidean norm.

2. The penalty function

We first describe the penalty function employed and report its salient features. A justification of its form along with a formal proof of the properties reported below can be found in /1/.

Given two positive constants \( \alpha \) and \( \beta \) we define the following set:

\[ \beta := \{ x \in \mathbb{R}^n : x \geq \alpha e, \quad (e' x - 1)^2 \leq \beta \}, \]
which is a compact perturbation of the feasible set. On the interior of \( \mathcal{B} \), we can define the positive functions
\[
a_i(x) = \alpha + x_i, \quad i = 1, \ldots, n;
\]
and the matrix \( A(x) := \text{diag}(a_i(x)) \). Furthermore we introduce two multipliers functions given by:
\[
u(x) = \frac{e'X^2(Qx + c)}{(e'x - 1)^2 + x'e'}, \quad \bar{v}(x) = Qx + c + e\nu(x).
\]
Now we can define the penalty function, which depends on a positive parameter \( \varepsilon \):
\[
P(x; \varepsilon) = f(x) + v(x)'r(x; \varepsilon) + u(x)(e'x - 1) + \frac{1}{\varepsilon} A(x)^{-1}r(x; \varepsilon) + \frac{1}{\varepsilon} \frac{(e'x - 1)^2}{b(x)} ,
\]
\[
r_i(z; \varepsilon) = \max(-x_i, -\varepsilon a_i(z)v_i(z)).
\]
We note that function (2.2) includes barrier terms \( 1/a_i(x) \) and \( 1/b(x) \) that cause the penalty function to go to infinity on the boundary of \( \mathcal{B} \) for any positive value of \( \varepsilon \).

Now we state some properties of the multiplier functions and of the penalty function.

**Proposition 2.1.** For any \( \varepsilon > 0 \)

(i) if \( (x, \bar{u}, \bar{v}) \) is a K-T triple for Problem (P), then \( u(x) = \bar{u} \) and \( v(x) = \bar{v} \).

(ii) the gradient of \( u \) and the Jacobian of \( v \) are given by:
\[
\nabla u(x) = -\frac{QXx + 2X(Qx + c)}{(e'x - 1)^2 + x'e'} + 2\frac{e'X^2(Qx + c) ((e'x - 1)e + x)}{((e'x - 1)^2 + x'e^2)} ,
\]
\[
\nabla v(x) = Q + \nabla u(x)e'.
\]

(iii) \( P \) is continuously differentiable on the interior of \( \mathcal{B} \) and its gradient is:
\[
\nabla P(x; \varepsilon) = \left[ \nabla v(x) - \frac{2}{\varepsilon} A(x)^{-1} r(x; \varepsilon) \right] r(x; \varepsilon) + \left[ \nabla u(x) + \frac{2}{\varepsilon} \frac{1}{b(x)} - \frac{2}{\varepsilon} \frac{(e'x - 1)^2}{b(x)} \right] (e'x - 1),
\]
where \( R(x; \varepsilon) := \text{diag}(r_i(x; \varepsilon)) \).

(iv) for any \( \varepsilon > 0 \), if \( z \) is a solution of Problem (QP) then \( \nabla P(z; \varepsilon) = 0 \). \( \square \)

The main properties that make \( P \) suitable for the solution of Problem (QP) are reported in the next theorem.

**Theorem 2.2.** Let \( \Gamma \) be any given constant. Then there exists an \( \varepsilon > 0 \) such that, for any \( \varepsilon \in (0, \varepsilon] \)

(i) if \( x_\varepsilon \in \mathcal{B} \) is a stationary point of \( P(x; \varepsilon) \) and \( P(x_\varepsilon; \varepsilon) \leq \Gamma \), \( x_\varepsilon \) is a solution of Problem (QP);

(ii) any solution of Problem (QP) is a global minimum point of \( P \) on \( \mathcal{B} \). \( \square \)

Theorem 2.2 essentially says that there exists a one-to-one correspondence between solutions of Problem (QP) and stationary points of \( P \). This property enables us to solve Problem (QP) by using any unconstrained minimization algorithm to locate a stationary point of \( P \). We shall show in the next section how we can exploit these considerations to construct an efficient iterative algorithm for the solution of Problem (QP). We conclude this section by proving some additional results we shall need in the sequel. We recall that a function is said to be a \( C^{0,1} \) function on an open set \( \mathcal{A} \) if it is continuously differentiable on \( \mathcal{A} \) and its gradient is locally lipschitz on \( \mathcal{A} \). For a \( C^{0,1} \) function \( F \) we can define a generalized Hessian at \( x \), denoted by \( \partial^2 F(x) \), as the convex hull of the set:
\[
\{ H : \exists x_i \rightarrow x \text{ with } F \text{ twice differentiable at } x_i, \text{ and } \nabla^2 F(x_i) \rightarrow H \}.
\]
We refer the reader to \( /3 / \) for a study on the properties of this class of functions. Before stating the next proposition we define some index sets. Let \((z, \bar{u}, \bar{v})\) be a K-T triple for Problem (QP), then:
\[
J_A := \{ i : z_i = 0 \}, \quad J_N := \{ i : z_i > 0 \}, \quad J_{AA} := \{ i : z_i = 0, \bar{v}_i > 0 \}, \quad J_{AO} := \{ i : z_i = 0, \bar{v}_i = 0 \}.
\]