Electronic Transactions on Numerical Analysis.
Volume 31, pp. 25-29, 2008.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ON THE DECREASE OF A QUADRATIC FUNCTION ALONG THE PROJECTED-GRADIENT PATH*

ZDENĚK DOSTÁL†

Abstract. The Euclidean gradient projection is an efficient tool for the expansion of an active set in the active-set-based algorithms for the solution of bound-constrained quadratic programming problems. In this paper we examine the decrease of the convex cost function along the projected-gradient path and extend the earlier estimate given by Joachim Schöberl. The result is an important ingredient in the development of optimal algorithms for the solution of convex quadratic programming problems.

Key words. Bound-constrained quadratic programming, Euclidean gradient projection, rate of convergence.

AMS subject classifications. 65K05, 90C20, 49M10.

1. Introduction. While there are well known classical results concerning the rate of convergence of many algorithms for the solution of unconstrained quadratic programming problems in terms of bounds on the spectrum of the Hessian matrix [1], until recently there were no such results on the decrease of the cost function for the algorithms that were proposed to solve the problem

$$\min_{x \in \Omega} f(x), \tag{1.1}$$

where $\Omega = \{x : x \geq \ell\}$, $f(x) = \frac{1}{2}x^T Ax - x^T b$, ℓ and b are given column n-vectors and A is a $n \times n$ symmetric positive definite matrix. The standard results either provide bounds on the contraction of the gradient projection [2], or guarantee only some qualitative properties of convergence [2, 4, 5, 12, 13, 16]. For example, Luo and Tseng [14, 15] proved the linear rate of convergence of the cost function for the gradient projection method even for more general problems, but they did not make any attempt to specify the constants. Let us recall that the need for such estimates emerged in the development of scalable algorithms for the solution of the discretized variational inequalities. Indeed, the first result of this type is due to Schöberl [17], who found a bound on the R-linear convergence of the decrease of f for the gradient projection method and used the estimate to develop probably the first theoretically supported scalable algorithm for variational inequalities. Later he proposed a better proof which enabled him to improve the original estimate [11]. The result was exploited in the analysis of the rate of convergence of the active set based algorithms, which combined the conjugate gradient method with the fixed step length gradient projection and the proportioning algorithms [6, 11].

The estimates mentioned above share an unpleasant drawback, namely, they give a bound on the rate of convergence only for the step length $\alpha \in (0, \|A\|^{-1}]$, with the best bound for $\alpha = \|A\|^{-1}$, while the best results were observed experimentally for larger values of α, not supported by any estimate. The point of this note is to extend the estimate [11] providing a nontrivial bound also for $\alpha \in (\|A\|^{-1}, 2\|A\|^{-1}]$. Our proof is based on the analysis of the gradient path for the cost function which dominates $\|A\|^{-1} f$ and whose Hessian is the identity matrix I.

* Received January 2, 2008. Accepted May 15, 2008. Published online on September 23, 2008. Recommended by Martin H. Gutknecht.

†VSB-Technical University of Ostrava, Department of Applied Mathematics, 17. listopadu 15, CZ-70833 Ostrava, Czech Republic (zdenek.dostal@vsb.cz). This research has been supported by the grants GA CR 201/07/0294, AS CR 1ET400300415 and the Ministry of Education of the Czech Republic No. MSM6198910027.

ETNA
Kent State University
etna@mcs.kent.edu

2. The projected-gradient path of a function whose Hessian is the identity. We start our exposition by an analysis of the special quadratic function

$$F(x) := \frac{1}{2}x^T x - c^T x,$$

defined for $x \in \mathbb{R}^n$ along the projected-gradient path

$$P_\Omega\left(x - \alpha \nabla F(x)\right) = \max\{x - \alpha g, \ell\}, \quad g := \nabla F(x) = x - c,$$

where the $\max$ is assumed to be carried out componentwise. Alternatively, for $\alpha > 0$ and a fixed $x \in \mathbb{R}^n$, we can describe the projected-gradient path by means of the reduced gradient $\widetilde{g}$, which is defined componentwise by

$$\widetilde{g}_i(\alpha) := \min\{(x_i - \ell_i)/\alpha, g_i\}.$$

Thus,

$$P_\Omega\left(x - \alpha g\right) = x - \alpha \widetilde{g}(x)$$

and we can define

$$\begin{aligned} \psi(\alpha) &:= F\big(P_\Omega\left(x - \alpha g\right)\big) = F(x) + \varphi(\alpha), \\ \varphi(\alpha) &:= -\alpha g^T \widetilde{g}(\alpha) + \frac{\alpha^2}{2}\|\widetilde{g}(\alpha)\|^2, \end{aligned} \tag{2.1}$$

where $\|\cdot\|$ denotes the Euclidean norm. Let us first examine the one dimensional case.

LEMMA 2.1. *Let x, ℓ, c denote real numbers, with $x \geq \ell$. For $\alpha \in (0, 2)$, let F be defined as above for $n = 1$ and let φ be defined by (2.1). Then, for any $\alpha \in (0, 1]$,*

$$\varphi(\alpha) \geq \varphi(2 - \alpha). \tag{2.2}$$

Proof. First observe that, if $n = 1$ and $\alpha > 0$, then the above definitions reduce to

$$g = x - c, \quad \widetilde{g}(\alpha) = \min\{(x - \ell)/\alpha, g\}$$

and

$$\varphi(\alpha) = -\alpha \widetilde{g}(\alpha) g + \frac{\alpha^2}{2}\left(\widetilde{g}(\alpha)\right)^2.$$

Moreover, if $g = 0$ and $\alpha > 0$, then $\varphi(\alpha) = 0$ and, if $g \neq 0$,

$$\varphi(\alpha) = \begin{cases} \varphi_F(\alpha) & \text{for} \quad \alpha \leq (x - \ell)/g \quad \text{or} \quad g < 0, \\ \varphi_A(\alpha) & \text{for} \quad \alpha \leq (x - \ell)/g \quad \text{and} \quad g > 0, \end{cases}$$

where

$$\varphi_F(\alpha) := \left(-\alpha + \frac{\alpha^2}{2}\right) g^2 \quad \text{and} \quad \varphi_A(\alpha) := -(x - \ell)g + \frac{1}{2}(x - \ell)^2.$$

Thus, for any α,

$$\varphi_F(2 - \alpha) = \left(-(2 - \alpha) + \frac{(2 - \alpha)^2}{2}\right) g^2 = \varphi_F(\alpha), \tag{2.3}$$

ETNA
Kent State University
etna@mcs.kent.edu

and, if $g \leq 0$,

$$\varphi(\alpha) = \varphi_F(\alpha) = \varphi_F(2-\alpha) = \varphi(2-\alpha),$$

i.e., (2.2) holds true.

Let us now assume that $g > 0$ and denote $\overline{\alpha} = (x-\ell)/g$. Thus, if $\overline{\alpha} \in (0,1]$, then φ is nonincreasing on $(0,2)$ and (2.2) is satisfied for $\alpha \in (0,1]$. To finish the proof, notice that if $1 < \overline{\alpha}$, then

$$\begin{aligned} \varphi(\alpha) &= \varphi_F(\alpha), \quad \alpha \in (0,1], \\ \varphi(\alpha) &\leq \varphi_F(\alpha), \quad \alpha \in (1,2), \end{aligned}$$

so that we can use (2.3) to obtain, for $\alpha \in (0,1]$,

$$\varphi(2-\alpha) \leq \varphi_F(2-\alpha) = \varphi_F(\alpha) = \varphi(\alpha). \qquad \square$$

Now we are able to extract the information we need on the values of F along the projected-gradient path.

LEMMA 2.2. *Let $x, \ell, c \in \mathbb{R}^n$, with $x \geq \ell$. For $\alpha \in (0,2)$, let F be defined as above and let φ be defined by (2.1). Then, for any $\alpha \in (0,1]$,*

$$\psi(\alpha) \geq \psi(2-\alpha).$$

Proof. Let us define, for any $\xi \in \mathbb{R}$ and $\alpha > 0$,

$$F_i(\xi) = \frac{1}{2}\xi^2 - c_i\xi \qquad \text{and} \qquad \varphi_i(\alpha) = -\alpha g_i \widetilde{g}_i(\alpha) + \frac{\alpha^2}{2}\left(\widetilde{g}_i(\alpha)\right)^2.$$

Using the notation introduced above, we get

$$\psi(\alpha) = \sum_{i=1}^{n} \big(F_i(x_i) + \varphi_i(\alpha)\big).$$

To complete the proof, it is enough to apply Lemma 2.1. $\square$

3. Decrease of the cost function along the projected-gradient path. In order to use Lemma 2.2 in our analysis, let us assume that $x \in \Omega$ is arbitrary, but fixed, so that we can define, for each $\alpha \in \mathbb{R}$, a quadratic function

$$F_\alpha(y) = \alpha f(y) + \frac{1}{2}(y-x)^T(I - \alpha A)(y-x).$$

We shall assume that $\alpha\|A\| \leq 1$, so that, for any $y \in \mathbb{R}$,

$$F_\alpha(y) \geq \alpha f(y), \quad F_\alpha(x) = \alpha f(x) \quad \text{and} \quad \nabla F_\alpha(x) = \alpha \nabla f(x).$$

Moreover, the Hessian matrix of F_α is the identity, so that F has the form assumed in Lemma 2.2. We shall use some other relations from [11].

LEMMA 3.1. *Let $\widehat{x}$ denote the unique solution of (1.1), λ_1 be the smallest eigenvalue of A, $\alpha \in (0, \|A\|^{-1}]$, $x \in \Omega$ and $g = Ax - b$. Then*

$$\alpha f\big(P_\Omega(x - \alpha g)\big) - \alpha f(\widehat{x}) \leq F_\alpha\big(P_\Omega(x-\alpha g)\big) - \alpha f(\widehat{x})$$

and

$$F_\alpha\big(P_\Omega(x-\alpha g)\big) - \alpha f(\widehat{x}) \leq \alpha(1-\alpha\lambda_1)\big(f(x) - f(\widehat{x})\big). \tag{3.1}$$

Proof. Replace f by αf in the statement and the proof of Theorem 4.1 of [11]. □

Now we are ready to prove the main result.

THEOREM 3.2. *Let $\widehat{x}$ denote the unique solution of (1.1), λ_1 be the smallest eigenvalue of A, $x \in \Omega$, $g = Ax - b$, $\overline{\mu} = 2\|A\|^{-1}$ and $\alpha \in (0, \overline{\mu}]$. Then*

$$f\big(P_\Omega\,(x - \alpha g)\big) - f(\widehat{x}) \le \eta(\alpha)\big(f(x) - f(\widehat{x})\big),$$

where

$$\eta(\alpha) = \max\{1 - \alpha\lambda_1, 1 - (\overline{\mu} - \alpha)\lambda_1\}.$$

Proof. To begin, let us observe that, for $\alpha \in (0, \|A\|^{-1}]$, the statement reduces to Lemma 3.1. Moreover, it is enough to prove the statement for $\alpha \in (\|A\|^{-1}, \overline{\mu})$, as the case $\alpha = \overline{\mu}$, i.e.,

$$f\big(P_\Omega\,(x - \overline{\mu} g)\big) - f(\widehat{x}) \le f(x) - f(\widehat{x}),$$

follows by the continuity argument.

To prove the statement for $\alpha \in (\|A\|^{-1}, \overline{\mu})$, let us first assume that $\|A\| = 1$ and let $\alpha = 1 + \delta$, $\delta \in (0, 1)$. Then F_δ dominates δf and we can apply Lemma 2.2 to the function f to get

$$\delta f\big(P_\Omega\,(x - \alpha g)\big) \le F_\delta\big(P_\Omega\,(x - \alpha g)\big) \le F_\delta\big(P_\Omega\,(x - \delta g)\big).$$

Combining the latter inequality with (3.1), we get

$$\delta f\big(P_\Omega\,(x - \alpha g)\big) - \delta f(\widehat{x}) \le \delta(1 - \delta\lambda_1)\big((f(x) - f(\widehat{x})\big),$$

that is,

$$f\big(P_\Omega\,(x - \alpha g)\big) - f(\widehat{x}) \le (1 - \delta\lambda_1)\big((f(x) - f(\widehat{x})\big).$$

To complete the proof, it is enough to apply the last inequality to the function $\|A\|^{-1} f$. □

4. Comments and conclusions. Theorem 3.2 fills in a longstanding gap in our theory of *optimal* algorithms [6, 11] for the solution of bound-constrained quadratic programming problems. In particular, the result can be used in the analysis of these algorithms to obtain a bound on the rate of convergence in terms of bounds on the spectrum, for step lengths that are longer than allowed by the original theory. Moreover, the result also improves our understanding of the optimal algorithms for bound- and equality-constrained quadratic programming problems [7, 8].

We remark that these algorithms were the key ingredients, together with the theoretical results concerning the FETI, BETI, TFETI and TBETI domain decomposition methods, in the development of scalable algorithms for the solution of variational inequalities, discretized either by the finite element method [9, 10] or by the boundary element methods [3].

REFERENCES

[1] O. AXELSSON, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994.
[2] D. P. BERTSEKAS, *Nonlinear Optimization*, Athena Scientific, Belmont, 1999.
[3] J. BOUCHALA, Z. DOSTÁL, AND M. SADOWSKÁ, *Theoretically supported scalable BETI method for variational inequalities*, Computing, 82 (2008), pp. 53–75.
[4] P. H. CALAMAI AND J. J. MORÉ, *Projected gradient methods for linearly constrained problems*, Math. Programming, 39 (1987), pp. 93–116.

[5] Z. DOSTÁL, *Box constrained quadratic programming with proportioning and projections*, SIAM J. Optim., 7 (1997), 871–887.

[6] Z. DOSTÁL, *A proportioning based algorithm for bound constrained quadratic programming with the rate of convergence*, Numer. Algorithms, 34 (2003), pp. 293–302.

[7] Z. DOSTÁL, *Inexact semi-monotonic augmented Lagrangians with optimal feasibility convergence for convex bound and equality constrained quadratic programming*, SIAM J. Numer. Anal., 43 (2005), pp. 96–115.

[8] Z. DOSTÁL, *An optimal algorithm for bound and equality constrained quadratic programming problems with bounded spectrum*, Computing 78 (2006), pp. 311–328.

[9] Z. DOSTÁL AND D. HORÁK, *Theoretically supported scalable FETI for numerical solution of variational inequalities*, SIAM J. Numer. Anal., 45 (2007), pp. 500-513.

[10] Z. DOSTÁL, D. HORÁK, AND D. STEFANICA, *A scalable FETI-DP algorithm for a coercive variational inequality*, Appl. Numer. Math. 54 (2005), pp. 378–390.

[11] Z. DOSTÁL AND J. SCHÖBERL, *Minimizing quadratic functions subject to bound constraints with the rate of convergence and finite termination*, Comput. Optim. Appl., 30 (2005), pp. 23–44.

[12] A. FRIEDLANDER AND M. MARTÍNEZ, *On the maximization of a concave quadratic function with box constraints*, SIAM J. Optim., 4 (1994), pp. 177–192.

[13] M. HINTERMÜLLER, K. ITO, AND K. KUNISCH, *The primal–dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.

[14] Z.–Q. LUO AND P. TSENG *On the linear convergence of descent methods for convex essentially smooth minimization*, SIAM J. Control Optim., 30 (1992), pp. 408–425.

[15] Z.–Q. LUO AND P. TSENG, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Ann. Oper. Res., 46 (1993), pp. 157–178.

[16] J. J. MORÉ AND G. TORALDO, *On the solution of large quadratic programming problems with bound constraints*, SIAM J. Optim., 1 (1991), pp. 93–113.

[17] J. SCHÖBERL, *Solving the Signorini problem on the basis of domain decomposition techniques*, Computing, 60 (1998), pp. 323–344.