Journal of Convex Analysis
Volume 9 (2002), No. 2, 693–700

On the Distance Theorem in Quadratic Optimization

T. Zolezzi
Dip. di Matematica, Universita' di Genova,
via Dodecaneso 35, 16146 Genova, Italy
zolezzi@dima.unige.it

Received November 22, 2000
Revised manuscript received April 11, 2001

The optimization of convex quadratic forms on Banach spaces is considered. A suitable notion of conditioning under linear perturbations leads to the distance theorem in the free case, thereby extending to the optimization setting the classical Eckart-Young formula: the distance to ill-conditioning equals to the reciprocal of the condition number. Partial results are presented for the linearly constrained case.

Keywords: Conditioning, distance theorem, condition number theorem, convex optimization

1991 Mathematics Subject Classification: 49K40

1. Introduction

A suitable notion of conditioning of a matrix leads to the well-known distance theorem of Eckart-Young [6] (generalized by Gastinel [7] to arbitrary norms): the distance of the matrix to the set of all singular matrices is the reciprocal of its condition number. This remarkable distance theorem, sometimes referred to as the condition number theorem, has been extended to several problems: computation of eigenvalues and eigenvectors, of zeros of polynomials, pole assignment in linear control systems, QR decomposition, trigonometric equations. Abstract versions of the theorem are also available. See [4], [3] and their references.

Generally speaking, if a distance theorem is available for a given problem, then the closer the problem is to ill-conditioning, the more difficult the problem will be to solve, so that more precision, hence computation, is needed. For a problem far away to ill-conditioning, an efficient algorithm is expected to perform well with less precise data. Thus a distance theorem is important not only because it reveals the geometrical meaning of the (properly defined) condition number, but also because it is intimately related to the computational complexity analysis of the given problem.

Conditioning measures, similar to condition numbers, related to the reciprocal of the distance to infeasibility have been introduced in [10] for linear inequality systems, and related to the computational complexity of interior point methods in linear programming, see also [11]. An extension of the distance theorem for convex processes is obtained in [9].

In this paper we obtain the distance theorem for free optimization problems involving convex quadratic forms in Banach spaces. In Section 3 we apply the standard definition of absolute condition number (recalled in Section 2) with respect to natural perturbations acting on the variational problem, namely linear perturbations. The condition number is then defined directly as a measure of how sensitive the minimizer is to slight data pertur-

ISSN 0944-6532 / $ 2.50

bations. Then we prove that the distance theorem holds: the distance to ill-conditioning equals to the reciprocal of the condition number. Here the distance is measured through the uniform operator norm of the gradient of the quadratic form.

In Section 4 we partially extend the results of Section 3 to optimization problems for convex quadratic forms with linear constraints in Hilbert spaces. We adopt the classical setting of the standard theory of variational inequalities. We compute the condition number and show that its reciprocal is a lower bound of the distance to ill-conditioning.

More complete results about the distance theorem and extensions to more general classes of optimization problems are under investigation.

2. Basic Notions and Notations

The following is a standard definition of conditioning of a mathematical problem, see e.g. [4]. Let X, Y be linear normed spaces and consider nonempty subset $D \subset X, H \subset Y$. D is the set of data of the given problem, H is the set of solutions. Assume that the problem has a unique solution $m(p) \in H$ for each data $p \in D$. Fix $p^* \in D$ which defines the given problem. Then the (absolute) *condition number* of the given problem is defined as

$$\text{cond } (p^*) = \limsup_{p \to p^*} \|m(p) - m(p^*)\| / \|p - p^*\|. \tag{1}$$

Of course $0 \leq$ cond $(p^*) \leq +\infty$.

In the following it will be useful to think of p as a parameter which defines a perturbation of the fixed optimization problem corresponding to p^*. This problem is then called

well-conditioned iff cond $(p^*) < +\infty$,

ill-conditioned iff cond $(p^*) = +\infty$. The set of ill-conditioned problems will be denoted by IC.

Throughout the paper E denotes a real Banach space. The pairing between the dual space E^* and E is denoted by $< \cdot, \cdot >$. $\mathcal{L}(E, E^*)$ is the Banach space of all linear continuous operators between E and E^*. Given $x \in E^*, x \oplus x$ denotes the linear bounded map $L : E \to E^*$ defined by

$$< Lu, v >=< x, u >< x, v >; u, v \in E.$$

$S(E, E^*)$ denotes the set of those $A \in \mathcal{L}(E, E^*)$ such that A is symmetric and nonnegative. Thus for every $u, v \in E$ we have

$$< Au, v >=< Av, u >, < Au, u >\geq 0.$$

For a set G, cl G denotes its closure.

3. The Distance Theorem for Quadratic Forms

Let $A \in S(E, E^*)$ be given. Then the function $f : E \to R$ defined by

$$f(x) = \frac{1}{2} < Ax, x >, x \in E \tag{2}$$

is the convex quadratic form associated to A. By standard properties, A is uniquely determined by f. Denote by T the set of all quadratic forms f given by (2) as $A \in S(E, E^*)$. For each $p \in E^*$ and $f \in T$ consider

$$f_p(x) = \frac{1}{2} < Ax, x > - < p, x >, x \in E. \tag{3}$$

We are interested in characterizing well-conditioning of the problem, to minimize f on E, when the data are the linear continuous perturbations of f corresponding to $p \in E^*$ via (3). Thus, according to the previous setting, we have $D = E^*, p^* = 0$.

The problem of minimizing $f \in T$ on E is then called *well-conditioned* iff

$$\text{for every } p \in E^*, f_p \text{ has a unique global minimizer } m(p); \tag{4}$$

$$\limsup_{p \to 0} \|m(p)\|/\|p\| < +\infty, \tag{5}$$

since $m(0) = 0$. For short we simply say that f *is well-conditioned* and write

$$\text{cond } (f) = \limsup_{p \to 0} \|m(p)\|/\|p\|, \tag{6}$$

the condition number of f with respect to the perturbations defined by $f_p, p \in E^*$ (according to (1)).

Proposition 3.1. *Let $f \in T$ be with corresponding operator A. The following are equivalent:*

$$f \textit{ is well-conditioned };\tag{7}$$

$$\textit{there exists } \alpha > 0 \textit{ such that } < Ax, x > \geq \alpha \|x\|^2, x \in E. \tag{8}$$

Proof. Let (7) hold. Then for every $p \in E^*$ the unique minimizer $u = m(p)$ of f_p is characterized by $\nabla f_p(u) = 0$, hence $Au = p$, whence A is onto. By (6), $u = 0$ is the unique minimizer of $f(x) = \frac{1}{2} < Ax, x >$, hence $< Ax, x >$ is positive for each $x \neq 0$. It follows that $Ax = 0$ implies $x = 0$, thus A is one-to-one. By a known property (see [1, Lemma 4.123 p. 365]) (8) follows. Conversely, (8) implies (4) and $m(p) = A^{-1}p$, hence

$$\limsup_{p \to 0} \|A^{-1}(p)\|/\|p\| \leq \|A^{-1}\|, \tag{9}$$

yielding (7). □

Proposition 3.2. *Let $f \in T$ with corresponding operator A be well-conditioned. Then*

$$\textit{cond } f = \|A^{-1}\|. \tag{10}$$

Proof. Let $y_n \in E^*$ be such that

$$\|y_n\| = 1, \|A^{-1}y_n\| \to \|A^{-1}\|$$

then $x_n = y_n/n \to 0$ and

$$\|A^{-1}x_n\|/\|x_n\| = \|A^{-1}y_n\| \to \|A^{-1}\|,$$

hence (10). □

Given f, g in T with corresponding operators A, B we define the distance between the two forms f, g by the operator norm

$$\text{dist } (f, g) = \|A - B\|. \tag{11}$$

We denote by IC the set of those forms in T which are ill-conditioned.

The main result of this section is the following

Theorem 3.3. *Let $f \in T$ be well-conditioned. Then*

$$\textit{dist } (f, IC) = 1/ \textit{ cond } (f). \tag{12}$$

Proof. By Proposition 3.1, f corresponds to $A \in S(E, E^*)$ which fulfils (8). Let $B \in S(E, E^*)$ be such that

$$\|A - B\| < 1/\|A^{-1}\|.$$

Then by a standard result B is an isomorphism. By a known property (see [1, Lemma 4.123 p. 365]), it follows that B fulfils the coercivity condition (8). Consider

$$g(x) = \frac{1}{2} < Bx, x >, x \in E$$

then g is well-conditioned by Proposition 1 and

$$\text{dist } (f, g) < 1/ \text{ cond } (f)$$

by Proposition 3.2, hence

$$\text{dist } (f, IC) \geq 1/ \text{ cond } (f). \tag{13}$$

To show the opposite inequality, we shall exhibit a sequence $B_n \in S(E, E^*)$ such that if

$$f_n(x) = \frac{1}{2} < B_n x, x >, x \in E,$$

then each $f_n \in IC$ and

$$\|A - B_n\| \to 1/\|A^{-1}\|.$$

It suffices to find $D_n \in \mathcal{L}(E, E^*)$ such that for each n

$$A - D_n \in S(E, E^*); \tag{14}$$

$$A - D_n \text{ fails to be one-to-one ;} \tag{15}$$

$$\|D_n\| \to 1/\|A^{-1}\|; \tag{16}$$

then $B_n = A - D_n$ will do. As it is well known

$$\|A^{-1}\| = \sup \{< x, A^{-1}x >: x \in E^*, \|x\| = 1\},$$

then let $x_n \in E^*$ be such that

$$\|x_n\| = 1, < x_n, A^{-1}x_n > \to \|A^{-1}\| \tag{17}$$

and put

$$D_n = \alpha_n x_n \oplus x_n, \tag{18}$$

$\alpha_n > 0$ to be chosen later. By (18), $A - D_n \in \mathcal{L}(E, E^*)$, moreover $A - D_n$ is symmetric. For every $u \in E$

$$< (A - D_n)A^{-1}x_n, u >=< x_n, u > -\alpha_n < x_n \oplus x_n A^{-1}x_n, u >=$$

$$=< x_n, u > (1 - \alpha_n < x_n, A^{-1}x_n >) = 0 \tag{19}$$

provided

$$\alpha_n = 1/ < x_n, A^{-1}x_n >$$

which makes sense for all sufficiently large n, since then $< x_n, A^{-1}x_n >$ is positive by (17). Thus, setting for such n

$$D_n = x_n \oplus x_n / < x_n, A^{-1}x_n >,$$

we have that (15) is true by (19). For any $u \in E$

$$< (A - D_n)u, u >=< Au, u > - < x_n, u >^2 / < x_n, A^{-1}x_n >\geq 0$$

since by a well-known inequality

$$< Au, u >< x, A^{-1}x >\geq< x, u >^2$$

for all $x \in E^*$ and $u \in E$. Then (14) is proved. To end the proof we check (16). We have

$$\|D_n\| = \alpha_n \|x_n \oplus x_n\| = \|x_n\|^2 / < x_n, A^{-1}x_n >$$

hence (16) by (17). □

Even in the finite dimensional setting, Theorem 3.3 is not a particular case of the Eckart-Young Theorem [6], since here we deal with a proper subset of the space of all singular matrices (possibly increasing distances).

4. Variational (In)equalities

In this section we present some partial result dealing with variational inequalities with linear constraints in Hilbert spaces.

Firstly we characterize well-conditioning and compute the condition number of a variational inequality with respect to suitable linear perturbations. Then we obtain a lower bound of the distance to ill-conditioning , namely the reciprocal of the condition number, thereby partially extending the results of Section 3.

Throughout this section, E denotes a real Hilbert space with scalar product $< \cdot, \cdot >$. K is a fixed closed linear subspace of E. We consider E^* as identified with E and write $\mathcal{L}(E) = \mathcal{L}(E, E^*), S(E) = S(E, E^*)$. Given $A \in S(E)$, we consider $f : E \to R$ defined by (2) and the set T of all quadratic convex forms f as above. For each $p \in E$ and $f \in T$ we consider f_p defined by (3). Let $p^* = 0$. The problem (K, f), of minimizing f on K, is well-conditioned iff (4),(5) hold. The condition number of f is again defined by (6) and denoted by cond (f).

We want to emphasize that all quadratic forms involved are restrictions to the constraint K of forms originally defined on the whole space E, in order to deal with conditioning of a given form (2) over possibly varying subspaces K. Moreover, only the elements $p \in E$ define the perturbations f_p acting on f, i.e. the whole space K^* of linear perturbations is not used, according to the standard setting of the classical theory of variational inequalities, see e.g. [2, Section V.3]. This precludes using directly the results of the previous section.

We need the following

Lemma 4.1. *Let $B \in S(E)$ be given with $B(K)$ closed and $B : K \to B(K)$ an isomorphism. Then there exists $\omega > 0$ such that*

$$< Bu, u > \geq \omega \|u\|^2, u \in K.$$

Proof. Let $Z \in S(E)$ be the square root of B , see [8, Theorem 3.35 p. 281]. Then Z is one-to-one on K since B is. The conclusion will then follow by closedness of $Z(K)$ due to [5, Lemma 1 p. 487], since then for every $u \in K$ and some constant $\omega > 0$

$$< Bu, u > = < Z^2 u, u > = \|Zu\|^2 \geq \omega \|u\|^2.$$

So let $y \in D = \text{ cl } Z(K), u_n \in K$ and $y_n = Zu_n \to y$. Then

$$Zy_n = Bu_n \to Zy \in B(K)$$

since $B(K)$ is closed. Thus there exists $u \in K$ such that $Zy = Bu$, hence

$$Z(y - Zu) = 0, y - Zu \in D,$$

whence $y = Zu$ because Z is one-to-one on D, as we check now. If $v \in D$ and $Zv = 0$, there exists $v_n \in Z(K), v_n \to v$, then $v_n = Zw_n$ with $w_n \in K$. It follows that $Zv_n = Bw_n \to 0$, whence $w_n \to 0$ because B is an isomorphism. We conclude that $Zw_n \to 0$, whence $v = 0$, ending the proof.

The following is an extension of Proposition 3.1.

Proposition 4.2. *Let $A \in S(E)$ and $f \in T$ be the corresponding quadratic form. Then the following are equivalent:*

$$\textit{problem } (K, f) \textit{ is well-conditioned}\ ; \tag{20}$$

$$\textit{there exists } \alpha > 0 \ \textit{ such that } < Au, u > \geq \alpha \|u\|^2, u \in K. \tag{21}$$

Proof. Assume (20). If $x \in K$ and $Ax = 0$ then $f(x) = 0$ hence $x = 0$ by well-conditioning. If $y \in \text{ cl } A(K)$ then $Ax_n = p_n \to y$ for some sequence $x_n \in K$. It follows that x_n minimizes f_{p_n} on K since obviously

$$< Ax_n, x - x_n > = < p_n, x - x_n >, x \in K$$

and by (20)

$$\|x_n\| \leq \text{ (const.) } \|p_n\| \leq \text{ const.}$$

For a subsequence, $x_n \rightharpoonup x \in K$ hence $Ax_n \rightharpoonup Ax = y$ whence $y \in A(K)$. This proves that $A(K)$ is closed. It follows that the restriction of A to K is an isomorphism. Then (21) follows by Lemma 4.1. Conversely, (21) yields (20) by well-known results, see [2, chapter V]. $\square$

Given an isomorphism $A \in S(E)$, let us consider the scalar product on E defined by

$$(x, y) =< Ax, y > . \tag{22}$$

We shall need the best approximation operator

$$Q : E \to K$$

with respect to the scalar product (22), i.e. for every $x \in E, Qx$ is the unique closest point in K to x with respect to the (equivalent) norm $< Ay, y >^{1/2}$ on E.

Proposition 4.3. *Let $A \in S(E)$ be an isomorphism and $f \in T$ be the corresponding quadratic form. Then*

$$\textit{cond}\ (f) = \|QA^{-1}\|. \tag{23}$$

Proof. $Q \in \mathcal{L}(E)$ since K is a linear closed subspace of E equipped with the inner produt (22). For any $p \in E$ consider f_p given by (3). Its unique minimizer $u = m(p)$ on K fulfils

$$< Au - p, x - u >= 0, x \in K$$

(see [2, corollaire V.4 p. 80]), hence by (22)

$$(u - A^{-1}p, x - u) = 0, x \in K.$$

Then $m(p) = QA^{-1}p$ by the projection theorem ([2, corollaire V.4 p. 80]). The conclusion follows by (6). □

As in the free case of Section 3, the distance between $f, g \in T$ with corresponding operators $A, B \in S(E)$ is defined by the operator norm in $\mathcal{L}(E)$,

$$\text{dist}\ (f, g) = \|A - B\|.$$

Theorem 4.4. *If $A \in S(E)$ is an isomorphism and $f \in T$ is the corresponding quadratic form, then*

$$\textit{dist}\ (f, IC) \geq 1/\ \textit{cond}\ (f).$$

Proof. For a given operator $U \in L(E)$ write

$$\|U\|_K = \sup\ \{\|U(x)\| : x \in K, \|x\| \leq 1\}.$$

Let $B \in S(E)$ be such that

$$\|A - B\| < 1/\|QA^{-1}\|.$$

Then

$$\|A - B\|_K \leq \|A - B\| < 1/\|QA^{-1}\|. \tag{24}$$

Let L denote the restriction of A to K. Then $L : K \to A(K)$ is an isomorphism. We have

$$\|L^{-1}\| = \sup\ \{\|L^{-1}y\| : y \in A(K), \|y\| \leq 1\} = \sup\ \{\|QL^{-1}y\| : y \in A(K), \|y\| \leq 1\} \leq$$

$$\leq \sup\ \{\|QA^{-1}y\| : y \in E, \|y\| \leq 1\} = \|QA^{-1}\|,$$

hence by (24)

$$\|A - B\|_K < 1/\|L^{-1}\|. \tag{25}$$

By a standard result it follows that $B : K \to B(K)$ is an isomorphism provided we show that $B(K)$ is closed. To this aim, denote by C the restriction of B to K and by P the orthogonal projection of E onto the closed linear subspace $A(K)$. Consider

$$W = CL^{-1} : A(K) \to B(K)$$

and the identity operator $I : A(K) \to A(K)$. Then $PW \in \mathcal{L}[A(K)]$ and taking operator norms in $\mathcal{L}[A(K)]$ we get

$$\|I - PW\| = \|PLL^{-1} - PCL^{-1}\| \leq \|P\|\|L - C\|\|L^{-1}\| \leq \|L - C\|\|L^{-1}\| < 1$$

by (25). Then, again by a standard result, $PW : A(K) \to A(K)$ is an isomorphism. Now let $u_n \in K$ be such that $Bu_n \to y$. Then

$$PWAu_n = PCu_n = PBu_n \to Py$$

hence

$$Au_n \to (PW)^{-1}Py$$

whence $u_n \to u \in K$ say. Therefore $y = Bu \in B(K)$ which is closed. It follows that, by Lemma 4.1, B is coercive on K and, by Proposition 4.2, the corresponding quadratic form defines a well-conditioned problem. Summarizing, the ball of radius $1/\|QA^{-1}\|$ around f is contained in the set of well-conditioned problems, and this by Proposition 4.3 ends the proof.

Acknowledgements. Some results of this paper were presented at the FGI Conference on Optimization (Montpellier, September 2000). Work partially supported by MURST and Universita' di Genova.

References

[1] J. F. Bonnans, A. Shapiro: Perturbation Analysis of Optimization Problems, Springer (2000).

[2] H. Brezis: Analyse Fonctionelle, Masson (1983).

[3] J.-P. Dedieu: Approximate solutions of numerical problems, condition number analysis and the condition number theorem, Lectures in Appl. Math. 32, Amer. Math. Soc. (1996) 263–283.

[4] J. W. Demmel: On condition numbers and the distance to the nearest ill-posed problem, Numerische Math. 51 (1987) 251–289.

[5] N. Dunford, J. Schwartz: Linear Operators, part I, Interscience (1958).

[6] C. Eckart, G. Young: The approximation of one matrix by another of lower rank, Psychometrika 1 (1936) 211–218.

[7] W. Kahan: Numerical linear algebra, Canad. Math. Bull. 9 (1966) 757–801.

[8] T. Kato: Perturbation Theory for Linear Operators, Springer (1966).

[9] A. S. Lewis: Ill-conditioned convex processes and conic linear systems (preprint).

[10] J. Renegar: Linear programming, complexity theory and elementary functional analysis, Math. Programming 70 (1995) 279–351.

[11] J. Renegar: Incorporating condition measures into the complexity theory of linear programming, SIAM J. Optim. 5 (1995) 506–524.