SIAM J. OPTIM.
Vol. 20, No. 2, pp. 935–947

OPTIMIZING CONDITION NUMBERS*

PIERRE MARÉCHAL† AND JANE J. YE‡

Abstract. In this paper we study the problem of minimizing condition numbers over a compact convex subset of the cone of symmetric positive semidefinite $n \times n$ matrices. We show that the condition number is a Clarke regular strongly pseudoconvex function. We prove that a global solution of the problem can be approximated by an exact or an inexact solution of a nonsmooth convex program. This asymptotic analysis provides a valuable tool for designing an implementable algorithm for solving the problem of minimizing condition numbers.

Key words. condition numbers, strongly pseudoconvex functions, quasi-convex functions, nonsmooth analysis, exact and inexact approximations

AMS subject classifications. 90C26, 90C30

DOI. 10.1137/080740544

1. Introduction.

We consider optimization problems of the form

$$(\mathscr{P}) \quad \left| \begin{array}{ll} \text{minimize} & \kappa(A) \\ \text{subject to} & A \in \Omega, \end{array} \right.$$

in which Ω is a compact convex subset of S_n^+, the cone of symmetric positive semidefinite $n \times n$ matrices, and $\kappa(A)$ denotes the condition number of A. On denoting $\lambda_1(A), \ldots, \lambda_n(A)$ the decreasingly ordered eigenvalues of A, the function κ considered here is defined by

$$\kappa(A) = \begin{cases} \lambda_1(A)/\lambda_n(A) & \text{if } \lambda_n(A) > 0, \\ \infty & \text{if } \lambda_n(A) = 0 \text{ and } \lambda_1(A) > 0, \\ 0 & \text{if } A = 0. \end{cases}$$

The reason for choosing the above extension of $\kappa(A)$ in the cases where $\lambda_n(A) = 0$ will appear clearly in section 3 below. Notice that, with such an extension, κ reaches its global minimum value at $A = 0$. In order to avoid this trivial case, we assume throughout that the set Ω does not contain the null matrix. From the definition of $\kappa(A)$, it is clear that if the constraint set Ω contains a positive definite matrix, then a minimizer for problem $(\mathscr{P})$ (as well as problem $(\mathscr{P}_p)$ to be defined below) must belong to S_n^{++}, the cone of symmetric positive definite $n \times n$ matrices.

Problems such as $(\mathscr{P})$ arise in several applications. The following example, which can be found in [3], is one of them.

Example 1.1. The Markovitz model for portfolio selection consists in selecting a portfolio $x \in \mathbb{R}_+^n$ (where $\mathbb{R}_+^n$ denotes the nonnegative orthant of $\mathbb{R}^n$) which is a solution to an optimization problem of the form

$$(\mathscr{M}) \quad \left| \begin{array}{rl} \text{minimize} & \langle x, Qx \rangle \\ \text{s.t.} & x \in \Delta_n,\ \langle c, x \rangle \geq b, \end{array} \right.$$

*Received by the editors November 12, 2008; accepted for publication (in revised form) May 14, 2009; published electronically July 2, 2009.
http://www.siam.org/journals/siopt/20-2/74054.html
†Institut de Mathématiques de Toulouse, Université Paul Sabatier, 31062 Toulouse cedex 9, France (pierre.marechal@math.univ-toulouse.fr).
‡Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8P 5C2, Canada (janeye@uvic.ca). The research of this author was partially supported by NSERC.

in which Q is a covariance matrix, $\Delta_n = \{x \in \mathbb{R}^n_+ \,|\, \sum_j x_j \leq 1\}$, $c \in \mathbb{R}^n$, and $b \in \mathbb{R}$. In the above problem, Q is in fact a parameter. Statistical considerations provide either an estimate $\hat{Q}$ as well as an upper bound η for $\|\hat{Q} - Q\|_\infty$ or a polytope of the form $\mathrm{co}\,\{Q_1, \ldots, Q_m\}$, where $\mathrm{co}\,C$ denotes the convex hull of the set C, outside of which the *true* Q is very unlikely to lie. In both cases, Q is constrained to belong to a polytope P. The analysis of the sensitivity of the solution $\bar{x}$ to $(\mathscr{M})$, together with the fact that statistical estimates of Q tend to underestimate the smallest eigenvalues of Q and to overestimate its largest eigenvalues, suggest that Q should be calibrated by means of the optimization problem [3, section 3.4.3.3]:

$$\left|\begin{array}{rl} \text{Minimize} & \kappa(Q) \\ \text{s.t.} & Q \in S_n^+ \cap P, \end{array}\right.$$

in which P is either of the previously mentioned polytopes.

It is well-known that the condition number function $A \mapsto \kappa(A)$ is Lipschitz continuous near any positive definite matrix A. However, the minimization of κ cannot be performed by using classical nonlinear programming algorithms. The fundamental difficulty lies in that κ is both nonconvex and not everywhere differentiable. For nonsmooth convex problems, there are some effective numerical algorithms such as bundle algorithms (see, e.g., Hiriart-Urruty and Lemaréchal [4] and Mäkelä [8] for a survey of earlier works, and Kiwiel [6] and the references within for more recent results). It is well-known that these algorithms are effective only for nonsmooth convex optimization problems because of the global nature of convexity but not for nonsmooth nonconvex optimization problems since, in general, first order information no longer provides a lower approximation to the objective function. The consequence is that bundle algorithms are much more complicated in the nonconvex case. For an explanation of the difficulty of extending nonsmooth convex algorithms to the nonconvex case and an extensive discussion on several classes of algorithms for nonsmooth, nonconvex optimization problems, the reader is referred to the excellent book of Kiwiel [5] as well as the recent paper of Burke, Lewis, and Overton on a gradient sampling algorithm [2].

On the other hand, it is easy to show that κ is a quasi-convex function and hence some existing algorithms for quasi-convex programming (see [14] and the references within) may be used. In fact in this paper we will show that κ is not only a quasi-convex function but also a (nonsmooth) pseudoconvex function! One of the consequences of this interesting fact is that the nonsmooth Lagrange multiplier rule for problem $(\mathscr{P})$ is not only a necessary but also a sufficient optimality condition. To the best of our knowledge, the algorithms for nonsmooth quasi-convex programming are mostly conceptual and not at all easy to implement with the exception of the level function method in [14], and even using the level function method one needs to solve a sequence of nonsmooth convex problems.

Our approach to problem $(\mathscr{P})$ is based on the observation that $\kappa(A)$ is the pointwise limit of the function $\kappa_p(A) := \lambda_1^{(p+1)/p}(A)/\lambda_n(A)$ as $p \to \infty$, and that the latter is expected to be easier to minimize, since κ_p^p, the pth power of κ_p, turns out to be convex and hence the effective bundle algorithms for nonsmooth nonconvex optimization problems may be used. For convenience, we consider the (lower semicontinuous extensions) of κ_p defined by

$$\kappa_p^p(A) = \begin{cases} \lambda_1^{p+1}(A)/\lambda_n^p(A) & \text{if } \lambda_n(A) > 0, \\ \delta_{\{0\}}(A) & \text{if } \lambda_n(A) = 0, \end{cases}$$

in which $\delta_C(A)$ denotes as usual the *indicator function* of the set C. Recall that $\delta_C(A) = 0$ if $A \in C$ and $\delta_C(A) = \infty$ if $A \notin C$.

Both κ and κ_p are quasi-convex. Recall that a function $f\colon \mathbb{R}^n \to \bar{\mathbb{R}}$ is said to be quasi-convex if it has convex level sets or, equivalently, if

$$\forall x, y \in \mathbb{R}^n \ \ \forall \alpha \in (0,1), \quad f((1-\alpha)x + \alpha y) \le \max\{f(x), f(y)\}.$$

Recall also that the level set of level $\alpha \in \mathbb{R}$ of a function $f\colon \mathbb{R}^n \to \bar{\mathbb{R}}$ is the set $\operatorname{lev}_\alpha(f) := \left\{ x \in \mathbb{R}^n \;\middle|\; f(x) \le \alpha \right\}$. For every $\alpha \ge 1$,

$$\operatorname{lev}_\alpha(\kappa) = \{0\} \cup \left\{ A \in S_n^{++} \;\middle|\; \lambda_1(A) - \alpha\lambda_n(A) \le 0 \right\}$$

and

$$\operatorname{lev}_\alpha(\kappa_p) = \{0\} \cup \left\{ A \in S_n^{++} \;\middle|\; \lambda_1(A)^{(p+1)/p} - \alpha\lambda_n(A) \le 0 \right\}.$$

The convexity of $\operatorname{lev}_\alpha(\kappa)$ and $\operatorname{lev}_\alpha(\kappa_p)$ is an immediate consequence of the following proposition.

PROPOSITION 1.1. *The functions $A \mapsto \lambda_1(A)$, $A \mapsto \lambda_n(A)$, and $A \mapsto \lambda_1(A)^{(p+1)/p}$ are, respectively, convex, concave, and convex on S_n^+.*

Proof. For every symmetric positive semidefinite matrix A,

$$\lambda_1(A) = \max_{\|x\|=1} \langle x, Ax \rangle \quad \text{and} \quad \lambda_n(A) = \min_{\|x\|=1} \langle x, Ax \rangle.$$

Thus $A \mapsto \lambda_1(A)$ and $A \mapsto \lambda_n(A)$ are, respectively, convex and concave. The convexity of $A \mapsto \lambda_1(A)^{(p+1)/p}$ results from the well-known fact that the postcomposition of a convex function by a convex increasing function is convex. ∎

The pointwise convergence of κ_p to κ suggests that one may tackle problem $(\mathscr{P})$ via a sequence of approximate problems in which the objective function κ is replaced by κ_p. We shall therefore also consider the following problem:

$$(\mathscr{P}_p) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa_p^p(A) \\ \text{s.t.} & A \in \Omega. \end{array} \right.$$

In section 3, we shall prove that κ_p^p is convexs so that problem $(\mathscr{P}_p)$ is in fact equivalent to the convex problem of minimizing κ_p^p over Ω. In section 4, we shall compute the subdifferentials of all three functions κ, κ_p, and κ_p^p in order to obtain information on the asymptotic behavior of minimizers of κ_p as p goes to infinity. This asymptotic behavior is then considered in section 5.

2. Preliminaries.

2.1. Nonsmooth analysis tools.

DEFINITION 2.1. *Let E be a Banach space, S be a subset of E, and $x_0 \in S$. Let $f\colon S \to \mathbb{R}$ be Lipschitz near x_0. We define the directional derivative of f at x_0 in direction v to be the number*

$$f'(x_0; v) = \lim_{t \downarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t},$$

provided it exists. We define the Clarke directional derivative of f at x_0 in direction v to be the number

$$f^\circ(x_0; v) = \limsup_{\substack{x \to x_0 \\ t \downarrow 0}} \frac{f(x + tv) - f(x)}{t}.$$

The function f is said to be Clarke regular at x_0 (or merely regular at x_0) if, for every $v \in E$, $f'(x_0; v)$ exists and $f'(x_0; v) = f^\circ(x_0; v)$.

DEFINITION 2.2. *Let E be a Banach space with dual E^*, S be a subset of E, and $x_0 \in S$. Let $f\colon S \to \mathbb{R}$ be Lipschitz near x_0. The Clarke subdifferential of f at x_0 is the weak* compact convex subset of E defined by*

$$\partial f(x_0) = \left\{ \xi \in E^* \;\middle|\; \forall v \in E,\ \langle \xi, v\rangle \le f^\circ(x_0; v) \right\}.$$

We shall need the following quotient rule from [1].

PROPOSITION 2.1. *Let $f_1, f_2\colon E \to \bar{\mathbb{R}}$, where E is a Banach space, be Lipschitz near x. Assume that $f_1(x) \ge 0$, $f_2(x) > 0$, and f_1 and $-f_2$ are Clarke regular at x. Then f_1/f_2 is Clarke regular at x and*

$$\partial \left(\frac{f_1}{f_2} \right)(x) = \frac{f_2(x)\partial f_1(x) - f_1(x)\partial f_2(x)}{f_2^2(x)}.$$

We shall also need the following regularity result and chain rule, which we shall prove for convenience.

PROPOSITION 2.2. *Let E be a Banach space, S be a subset of E, and $x_0 \in \operatorname{int} S$. Let $f = g \circ h$, where $h\colon S \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$. Assume that g is continuously differentiable at $h(x_0)$ and h is Lipschitz near x_0. Then*

$$\partial f(x_0) = g'(h(x_0))\partial h(x_0).$$

Moreover, if g is continuously differentiable in a neighborhood of $h(x_0)$ and h is Clarke regular at x_0, then f is also Clarke regular at x_0.

Proof. The first statement is a special case of [1, Theorem 2.3.9(ii)]. Let us prove the second statement. For every $v \in E$ and every $t > 0$ small enough, we have, by the mean value theorem,

$$\begin{aligned} \frac{f(x+tv) - f(x)}{t} &= \frac{g\big(h(x+tv)\big) - g\big(h(x)\big)}{t} \\ &= \frac{g'(\xi)\big(h(x+tv) - h(x)\big)}{t}, \end{aligned}$$

with $\xi \in [h(x), h(x+tv)]$. Taking the limit as $t \downarrow 0$ yields

$$f'(x; v) = g'\big(h(x)\big)h'(x; v).$$

On the other hand, for every x' near x, for every $v \in E$, and every $t > 0$ small enough,

$$\begin{aligned} \frac{f(x'+tv) - f(x')}{t} &= \frac{g\big(h(x'+tv)\big) - g\big(h(x')\big)}{t} \\ &= \frac{g'(\xi')\big(h(x'+tv) - h(x')\big)}{t}, \end{aligned}$$

with $\xi' \in [h(x'), h(x'+tv)]$. Taking the lim sup as $x' \to x$ and $t \downarrow 0$ yields

$$f^\circ(x; v) = g'\big(h(x)\big)h^\circ(x; v) = g'\big(h(x)\big)h'(x; v)$$

by regularity of h. This proves f is regular at x and the chain rule formula. ∎

In general, the Clarke regularity of a function f does not imply regularity of its negative. For example, $f(x) = |x|$ is regular at $x_0 = 0$ but its negative $f(x) = -|x|$ is not regular at the same point. However, the following holds.

LEMMA 2.1. *Let E be a Banach space, S be a subset of E, and $x_0 \in S$. Let $f\colon S \to \mathbb{R}$ be such that $-f$ is Clarke regular at x_0, and let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be continuously differentiable and nondecreasing at $f(x_0)$. Then $-\varphi \circ f$ is Clarke regular at x_0 and*

$$\partial(-\varphi \circ f)(x_0) = -\varphi'(f(x_0))\partial f(x_0) = \varphi'(f(x_0))\partial(-f)(x_0). \tag{1}$$

Proof. Since f is Lipschitz near x_0, applying twice the chain rule (see Proposition 2.2) yields (1). By the mean value theorem applied to the function φ,

$$-\varphi(f(x_0 + tv)) + \varphi(f(x_0)) = \varphi'(u)\big(-f(x_0 + tv) + f(x_0)\big)$$

for some $u \in [f(x_0), f(x_0 + tv)]$. Therefore,

$$\begin{aligned}\frac{-\varphi(f(x_0 + tv)) + \varphi(f(x_0))}{t} &= \varphi'(u)\frac{-f(x_0 + tv) + f(x_0)}{t}\\ &\to \varphi'(f(x_0))(-f)'(x_0; v) \quad \text{as} \quad t \downarrow 0,\end{aligned}$$

where the regularity of $-f$ ensures existence of the limit. Therefore,

$$(-\varphi \circ f)' = \varphi'(f(x_0))(-f)'(x_0; v). \tag{2}$$

Finally,

$$\begin{aligned}(-\varphi \circ f)^\circ(x_0; v) &= \max_{s \in \partial(-\varphi\circ f)(x_0)} \langle s, v\rangle\\ &= \max_{s' \in \partial(-f)(x_0)} \varphi'(f(x_0))\langle s', v\rangle\\ &= \varphi'(f(x_0)) \max_{s' \in \partial(-f)(x_0)} \langle s', v\rangle\\ &= \varphi'(f(x_0))(-f)^\circ(x_0; v)\\ &= \varphi'(f(x_0))(-f)'(x_0; v),\end{aligned}$$

in which the last equality is due to the Clarke regularity of $-f$. The conclusion then follows from (2). ◻

2.2. Convex tools.

DEFINITION 2.3. *Let E be an n-dimensional Euclidean space, $f\colon E \to \bar{\mathbb{R}}$ be a proper convex function, and x be a point such that $f(x)$ is finite. A vector $\xi \in E$ is an ε-subgradient of f at x if, for all $y \in E$,*

$$f(y) \geq f(x) + \big\langle \xi, y - x\big\rangle - \varepsilon.$$

The set of all subgradients of f at x is denoted by $\partial_\varepsilon f(x)$ and is called the ε-subdifferential of f at x. When $\varepsilon = 0$, $\partial f(x) := \partial_0 f(x)$ is the set of subgradients of f at x in the sense of convex analysis and coincides with the Clarke subdifferential of f at x for a convex function f. The ε-normal set to a closed convex set C at x is defined as the ε-subdifferential of the indicator function $\delta(\cdot|C)$ of C at x:

$$N_{C,\varepsilon}(x) := \partial_\varepsilon\delta(\cdot|C)(x) = \big\{\xi \in E \;\big|\; \langle \xi, y - x\rangle \leq \varepsilon \textit{ for all } y \in C\big\}.$$

When $\varepsilon = 0$, $N_C(x) := N_{C,0}(x)$ *is the normal cone of* C *at* x *in the sense of convex analysis.*

The proof of the following result can be found in [4].

PROPOSITION 2.3. *Let* E *be an* n*-dimensional Euclidean space,* $f\colon E \to \bar{\mathbb{R}}$ *be a proper convex function, and* x *be a point such that* $f(x)$ *is finite. Then the following conditions hold.*

(i) $0 \in \partial_\varepsilon f(x)$ *if and only if* $f(x) \leq f(y) + \varepsilon$ *for all* $y \in E$; *see* [4, *Volume* II, *Theorem* XI.1.1.5].

(ii) *If* $\operatorname{ri}\operatorname{dom} f \cap \operatorname{ri} C \neq \emptyset$ *and* $x \in C$, *then*

$$\partial_\varepsilon(f + \delta(\cdot|C))(x) = \bigcup_{\alpha\in[0,\varepsilon]} \partial_\alpha f(x) + N_{C,\varepsilon-\alpha}(x),$$

where $\operatorname{dom} f := \{x : f(x) \neq +\infty\}$ *is the domain of* f *and* $\operatorname{ri} C$ *denotes the set of relative interior points of* C; *see* [4, *Volume* II, *Theorem* XI.3.1.1].

(iii) $\partial_\varepsilon f(x) \subset \bigcup\{\partial f(y) + B(0,\sqrt{\varepsilon})|y \in B(x,\sqrt{\varepsilon})\}$, *where* $B(0,\sqrt{\varepsilon})$ *denotes the open ball centered at* 0 *with radius* $\sqrt{\varepsilon}$; *see* [4, *Volume* II, *Theorem* XI.4.2.1].

3. Convexity of κ_p^p. The recession cone of a convex set $C \subset \mathbb{R}^n$ is defined as the set of vectors y such that $C + \{y\} \subset C$. We denote it by 0^+C. Recall that if C is a closed convex set containing the origin, then

$$0^+C = \bigcap_{\beta>0} \beta C$$

(see [13, Corollary 8.3.2]). Let $\Gamma(\mathbb{R}^n)$ denote the set of all convex subsets of $\mathbb{R}^n$. We define on $\mathbb{R}_+ \times \Gamma(\mathbb{R}^n)$ the binary operation

$$(r, C) \longmapsto r \cdot C := \begin{cases} rC & \text{if } r > 0, \\ 0^+C & \text{if } r = 0. \end{cases}$$

A set-valued mapping $\sigma\colon \mathbb{R} \to 2^{\mathbb{R}^n}$ is said to be increasing whenever $r_1 \geq r_2$ implies $\sigma(r_1) \supset \sigma(r_2)$. The proof of the following lemma can be found in [12].

LEMMA 3.1. *The set-valued mapping* $r \mapsto r \cdot C$ *is increasing on* $\mathbb{R}_+$ *if and only if* $C \subset \mathbb{R}^n$ *is a convex set containing the origin. Consequently, if* $g\colon \mathbb{R}^m \to \bar{\mathbb{R}}$ *is concave and nonnegative on its domain, then the set*

$$\bigcup_{y\in\operatorname{dom} g} (g(y) \cdot C \times \{y\})$$

is a convex subset of $\mathbb{R}^n \times \mathbb{R}^m$.

PROPOSITION 3.1. *Let* $f\colon \mathbb{R}^n \to [0, \infty]$ *be quasi-convex, lower semicontinuous at* 0 *and positively homogeneous of degree* $p \geq 1$. *Then* f *is convex.*

Proof. If f is identically equal to ∞, there is nothing to prove. Assuming that there exists $x_0 \neq 0$ such that $f(x_0) < \infty$, let us prove that $f(0) = 0$. Since f is lower semicontinuous, one has

$$f(0) = f(\lim_{t\downarrow 0} tx_0) \leq \lim_{t\downarrow 0} f(tx_0) = \lim_{t\downarrow 0} t^p f(x_0) = 0.$$

Since f takes its values in $[0,\infty]$, one must have $f(0)=0$. Next, let us prove that $\operatorname{lev}_0(f) = 0^+(\operatorname{lev}_1(f))$. One has

$$\begin{aligned}
0^+(\operatorname{lev}_1(f)) &= 0^+\{x \in \mathbb{R}^n \,|\, f(x) \le 1\} \\
&= \bigcap_{\beta>0} \{\beta x \in \mathbb{R}^n \,|\, f(x) \le 1\} \\
&= \bigcap_{\beta>0} \{x' \in \mathbb{R}^n \,|\, f(x'/\beta) \le 1\} \\
&= \bigcap_{\beta>0} \operatorname{lev}_{\beta^p}(f) \;=\; \operatorname{lev}_0(f).
\end{aligned}$$

Consequently, the formula $\operatorname{lev}_r(f) = r^{1/p} \cdot \operatorname{lev}_1(f)$ holds for every $r \ge 0$. Finally, let us prove that the epigraph of f is convex. One has

$$\begin{aligned}
\operatorname{epi} f &= \{(x,r) \in \mathbb{R}^n \times \mathbb{R}_+ \,|\, f(x) \le r\} \\
&= \bigcup_{r\in\mathbb{R}_+} (\operatorname{lev}_r(f) \times \{r\}) \\
&= \bigcup_{r\in\mathbb{R}_+} \left(r^{1/p} \cdot \operatorname{lev}_1(f) \times \{r\}\right).
\end{aligned}$$

The conclusion then follows from Lemma 3.1, with $m=1$, $g(r)=r^{1/p}$, and $C = \operatorname{lev}_1(f)$. ☐

COROLLARY 3.1. *The function κ_p^p is convex for $p \ge 0$.*

Proof. It is easy to check that κ_p^p has convex level sets and that it is positively homogeneous of degree 1. Since it is lower semicontinuous on S_n^+ (by construction), the conclusion follows. ☐

It is worth noticing that, roughly speaking, κ_p^p is the restriction of the function $(A,B) \mapsto \lambda_1^{p+1}(A)/\lambda_n^p(B)$ to the linear manifold $\{A=B\}$, and that the latter function turns out to be jointly convex. The proof of the latter fact relies on the binary operation $\triangle$ introduced in [11], which we now outline. Let $f,g\colon E \to \bar{\mathbb{R}}$, where E is a Euclidean space. Assume that f is closed proper convex, with $f(0) \le 0$, and that g is closed proper concave and nonnegative on its effective domain $\operatorname{dom} g := \{x \in E \,|\, g(x) > -\infty\}$. Let $f0^+$ denote the recession function of f. Recall that $f0^+(x) = \lim_{t\downarrow 0} tf(x/t)$ (see [13]). The function

$$(f\triangle g)(x,y) := \begin{cases} g(y)f(x/g(y)) & \text{if } g(y) > 0, \\ f0^+(x) & \text{if } g(y) = 0, \\ \infty & \text{if } g(y) < 0 \end{cases}$$

is then closed proper convex on $E \times E$ (see [11, Theorem 2.1]). Now let f and g be defined on S_n by

$$f(A) = \begin{cases} \lambda_1^{p+1}(A) & \text{if } \lambda_1(A) \ge 0, \\ \infty & \text{if } \lambda_1(A) < 0 \end{cases}$$

and

$$g(B) = \begin{cases} \lambda_n(B) & \text{if } \lambda_n(B) \ge 0, \\ \infty & \text{if } \lambda_n(B) < 0. \end{cases}$$

The functions f and g are, respectively, closed proper convex and closed proper concave on S_n. Furthermore, $f(0) = 0$ and g is nonnegative on its domain $\operatorname{dom} g = S_n^+$. It then results from the above-mentioned construction that

$$(f\triangle g)(A,B) = \begin{cases} \lambda_1^{p+1}(A)/\lambda_n^p(B) & \text{if } \lambda_n(B) > 0 \text{ and } \lambda_1(A) \geq 0,\\ f0^+(A) & \text{if } \lambda_n(B) = 0 \text{ and } \lambda_1(A) \geq 0,\\ \infty & \text{if } \lambda_n(B) < 0 \text{ or } \lambda_1(A) < 0 \end{cases}$$

(with $f0^+(A) = \delta_{\{0\}}(A)$ in the case where $p > 0$ and $f0^+(A) = \lambda_1(A)$ in the case where $p = 0$) is closed proper and convex.

4. Subdifferentials. We shall use the following result, which is due to Cox, Overton, and Lewis (see [7], Corollary 10).

PROPOSITION 4.1. *The Clarke subdifferential of* λ_k *is given by*

$$\partial\lambda_k(A) = \operatorname{co}\{xx^\top \,|\, x \in \mathbb{R}^n,\ \|x\| = 1,\ Ax = \lambda_k(A)x\}.$$

We are now ready to obtain formulas for the subdifferentials of κ, κ_p^p, and κ_p.

PROPOSITION 4.2. *Assume* $A \in S_n^{++}$. *Then* κ, κ_p^p, *and* κ_p *are Clarke regular at* A, *and their Clarke subdifferentials at* A *are given, respectively, by*

$$\partial\kappa(A) = \lambda_1(A)^{-1}\kappa(A)\Big(\partial\lambda_1(A) - \kappa(A)\partial\lambda_n(A)\Big),$$

$$\partial\kappa_p^p(A) = \kappa(A)^p\Big((p+1)\partial\lambda_1(A) - p\kappa(A)\partial\lambda_n(A)\Big),$$

$$\partial\kappa_p(A) = \lambda_1(A)^{(1-p)/p}\kappa(A)\left(\frac{p+1}{p}\partial\lambda_1(A) - \kappa(A)\partial\lambda_n(A)\right).$$

Proof. Since $A \in S_n^{++}$,

$$\kappa(A) = \lambda_1(A)/\lambda_n(A) \quad \text{and} \quad \kappa_p(A)^p = \lambda_1(A)^{p+1}/\lambda_n(A)^p.$$

Regularity of κ at A follows from the fact that λ_1 and $-\lambda_n$ are convex, and the formula for $\partial\kappa(A)$ follows straightforwardly from Proposition 2.1, since both λ_1 and λ_n are locally Lipschitz (as convex or concave functions).

Now Proposition 2.2 implies that λ_1^{p+1} is regular at A and that

$$\partial\lambda_1^{p+1}(A) = (p+1)\lambda_1(A)^p\partial\lambda_1(A). \tag{3}$$

Next, Lemma 2.1 implies Clarke regularity of $-\lambda_n^p$, and the formula

$$\partial(-\lambda_n^p)(A) = -p\lambda_n(A)^{p-1}\partial\lambda_n(A) = p\lambda_n(A)^{p-1}\partial(-\lambda_n)(A). \tag{4}$$

Now (3) and (4) yield the desired formula for $\partial\kappa_p^p(A)$ via Proposition 2.1.

Finally, regularity of κ_p^p and the chain rule (Proposition 2.2) give rise to the regularity of κ_p as well as the claimed formula for $\partial\kappa_p(A)$. ∎

5. Convergence of approximate solutions. In this section, we show that, denoting by $\bar{A}_p$ a solution to problem $(\mathscr{P}_p)$, we can extract a sequence $(\bar{A}_{p_k})$ (with

$p_k \to \infty$ as $k \to \infty$) which converges to a global solution of problem ($\mathscr{P}$). We call *stationary point of* ($\mathscr{P}$) any matrix $\bar{A}$ satisfying

$$0 \in \partial\kappa(\bar{A}) + N_\Omega(\bar{A}),$$

in which $N_\Omega(\bar{A})$ is the normal cone to Ω at $\bar{A}$.

5.1. Exact approximation.

DEFINITION 5.1. *Let E be a Banach space, Ω be a subset of E, and $f\colon E \to \bar{\mathbb{R}}$ be lower semicontinuous and Lipschitz near a point $\bar{x} \in \Omega$. We say that f is strongly pseudoconvex at $\bar{x}$ on Ω if for every $\xi \in \partial f(\bar{x})$ and every $x \in \Omega$,*

$$\langle \xi, x - \bar{x} \rangle \geq 0 \implies f(x) \geq f(\bar{x}).$$

We say that f is strongly pseudoconvex on Ω if f is strongly pseudoconvex at every $\bar{x}$ on Ω.

We emphasize that our notion of strong pseudoconvexity is indeed stronger than the standard notion of pseudoconvexity, since, in the latter, it is assumed that

$$\forall x \in \Omega, \quad f^\circ(\bar{x}; x - \bar{x}) \geq 0 \implies f(x) \geq f(\bar{x})$$

and it is known that $f^\circ(\bar{x}; x - \bar{x}) = \max\{\langle \xi, x - \bar{x} \rangle : \xi \in \partial f(\bar{x})\}$.

PROPOSITION 5.1. *Let E be a Banach space, Ω be a closed convex subset of E, and $f\colon E \to \bar{\mathbb{R}}$ be lower semicontinuous and Lipschitz near $\bar{x}$ and pseudoconvex at $\bar{x}$ on Ω. A necessary and sufficient condition for $\bar{x}$ to be a global minimizer of f on Ω is that it satisfies the stationary condition*

$$0 \in \partial f(\bar{x}) + N_\Omega(\bar{x}).$$

Proof. The necessity is contained in [9, Chapter 5, Proposition 5.3]. Let us prove the sufficiency. Let $\xi \in \partial f(\bar{x})$ be such that $-\xi \in N_\Omega(\bar{x})$. By definition of the normal cone, for all $x \in \Omega$, $\langle \xi, x - \bar{x} \rangle \geq 0$. Therefore, $\max\{\langle \xi, x - \bar{x} \rangle | \xi \in \partial f(\bar{x})\} \geq 0$ for all $x \in \Omega$. The pseudoconvexity of f at $\bar{x}$ on Ω then implies that, for all $x \in \Omega$, $f(x) \geq f(\bar{x})$. ∎

PROPOSITION 5.2. *The function κ is strongly pseudoconvex on S_n^{++}.*

Proof. Let $\bar{A} \in S_n^{++}$. We shall prove that, for every $V \in \partial\kappa(\bar{A})$, the condition $\langle V, A - \bar{A} \rangle \geq 0$ implies that $\kappa(A) \geq \kappa(\bar{A})$. By Proposition 4.2, every $V \in \partial\kappa(\bar{A})$ is of the form

$$\lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(V_1 - \kappa(\bar{A})V_n\big),$$

with $V_1 \in \partial\lambda_1(\bar{A})$ and $V_n \in \partial\lambda_n(\bar{A})$. Since λ_1 and $-\lambda_n$ are convex, we have

$$\lambda_1(A) - \lambda_1(\bar{A}) \geq \langle V_1, A - \bar{A} \rangle \quad \text{and} \quad -\lambda_n(A) + \lambda_n(\bar{A}) \geq \langle -V_n, A - \bar{A} \rangle.$$

It follows that

$$\begin{aligned}\lambda_1(A) - \kappa(\bar{A})\lambda_n(A) &= \lambda_1(A) - \lambda_1(\bar{A}) + \kappa(\bar{A})\big(-\lambda_n(A) + \lambda_n(\bar{A})\big) \\ &\geq \big\langle V_1 - \kappa(\bar{A})V_n \, , \, A - \bar{A} \big\rangle \\ &= \lambda_1(\bar{A})\kappa(\bar{A})^{-1} \big\langle V \, , \, A - \bar{A} \big\rangle .\end{aligned}$$

Therefore, if $\langle V, A - \bar{A} \rangle \geq 0$, then $\kappa(A) \geq \kappa(\bar{A})$. ∎

The following corollary is an immediate consequence of Propositions 4.2, 5.1, and 5.2.

COROLLARY 5.1. *A feasible solution $\bar{A}$ of problem $(\mathscr{P})$ is a global solution if and only if*

$$0 \in \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\Big(\partial\lambda_1(\bar{A}) - \kappa(\bar{A})\partial\lambda_n(\bar{A})\Big) + N_\Omega(\bar{A}). \tag{5}$$

Note that since $\lambda_1(\bar{A})^{-1}\kappa(\bar{A}) > 0$ and $N_\Omega(\bar{A})$ is a cone, inclusion (5) is equivalent to

$$0 \in \partial\lambda_1(\bar{A}) - \kappa(\bar{A})\partial\lambda_n(\bar{A}) + N_\Omega(\bar{A}).$$

As is pointed out by one of the referees, the above necessary and sufficient optimality condition can be derived by transforming problem $(\mathscr{P})$ into the following equivalent convex optimization problem:

$$(\mathscr{P}') \quad \left| \begin{array}{rl} \text{Minimize} & \lambda_1(A) - \lambda_n(A)\kappa(\bar{A}) \\ \text{s.t.} & A \in \Omega. \end{array} \right.$$

Note that the above equivalent problem cannot be used to solve the problem since it involves the unknown optimal solution $\bar{A}$.

THEOREM 5.1. *Let $(p_k)_{k\in\mathbb{N}^\star} \subset [1,\infty)$ be a sequence which tends to infinity, and, for every $k \in \mathbb{N}^\star$, let $\bar{A}_{p_k}$ be a solution to problem $(\mathscr{P}_{p_k})$. Then every cluster point $\bar{A}$ of the sequence $(\bar{A}_{p_k})$ (there is at least one) is a global solution of problem $(\mathscr{P})$.*

Proof. By Proposition 5.1, $\bar{A}_{p_k}$ satisfies $0 \in \partial\kappa_p(\bar{A}_{p_k}) + N_\Omega(\bar{A}_{p_k})$ in which

$$\begin{aligned} &\partial\kappa_p(\bar{A}_{p_k}) \\ &= \lambda_1(\bar{A}_{p_k})^{(1-p_k)/p_k}\kappa(\bar{A}_{p_k})\left(\frac{p_k+1}{p_k}\partial\lambda_1(\bar{A}_{p_k}) - \kappa(\bar{A}_{p_k})\partial\lambda_n(\bar{A}_{p_k})\right) \end{aligned} \tag{6}$$

by Proposition 4.2. Taking a subsequence if necessary, we can assume that $\bar{A}_{p_k}$ converges to some matrix $\bar{A} \in \Omega$. Recall indeed that Ω is assumed to be compact. We now wish to show that

$$0 \in \partial\kappa(\bar{A}) + N_\Omega(\bar{A}).$$

By (6), there exist $V_1^{(k)} \in \partial\lambda_1(\bar{A}_{p_k})$ and $V_n^{(k)} \in \partial\lambda_n(\bar{A}_{p_k})$, and $V^{(k)} \in N_\Omega(\bar{A}_{p_k})$ such that

$$0 \in \lambda_1(\bar{A}_{p_k})^{(1-p_k)/p_k}\kappa(\bar{A}_{p_k})\left(\frac{p_k+1}{p_k}V_1^{(k)} - \kappa(\bar{A}_{p_k})V_n^{(k)}\right) + N_\Omega(\bar{A}_{p_k}). \tag{7}$$

From Proposition 4.1, we see that the sequences $V_1^{(k)}$ and $V_n^{(k)}$ are contained in the compact set $\operatorname{co}\{xx^\top \,|\, \|x\| = 1\}$. Therefore, by taking a subsequence if necessary, we can assume that

$$V_1^{(k)} \to \bar{V}_1 \quad \text{and} \quad V_n^{(k)} \to \bar{V}_n \quad \text{with} \quad \bar{V}_1 \in \partial\lambda_1(\bar{A}) \quad \text{and} \quad \bar{V}_n \in \partial\lambda_n(\bar{A}),$$

where we used the closedness of the multifunctions $\partial\lambda_1$ and $\partial\lambda_n$. Since the multifunction N_Ω is closed, we can pass to the limit in (7) and obtain

$$0 \in \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\bar{V}_1 - \kappa(\bar{A})\bar{V}_n\big) + N_\Omega(\bar{A}) \subset \partial\kappa(\bar{A}) + N_\Omega(\bar{A}).$$

By Corollary 5.1, the stationary point $\bar{A}$ is a global minimizer of κ over Ω. ∎

We emphasize that the solution set of $(\mathscr{P})$, which we denote by $S(\mathscr{P})$, is the intersection of the convex compact set Ω with the closed convex cone

$$\mathrm{lev}_{V(\mathscr{P})}(\kappa),$$

in which $V(\mathscr{P})$ is the optimal value of problem $(\mathscr{P})$. Therefore, depending on the shape of Ω, uniqueness will not be guaranteed in general. However, we can give more precise information on the limiting solution of the approximating sequence.

For every $A \in \Omega$, let $I_A := \{t > 0 | tA \in \Omega\}$. Clearly, I_A is a compact interval.

PROPOSITION 5.3. *For $p \in \mathbb{R}_+$ and every solution $\bar{A}_p$ to $(\mathscr{P}_p)$, $I_{\bar{A}_p}$ is of the form $[1, \nu]$ for some $\nu \geq 1$. In other words, $\bar{A}_p$ belongs to the set*

$$\Omega_0 := \left\{ A \in \Omega \;\middle|\; \min I_A = 1 \right\}.$$

Moreover, Ω_0 is compact. Consequently, the limit $\bar{A}$ provided by Theorem 5.1 *also belongs to Ω_0.*

Proof. The first assertion is an immediate consequence of the positive homogeneity of degree $(p + 1)/p$ of κ_p. Now we need only to prove that Ω_0 is closed. It is obvious that $\min I_A$ is equal to the optimal value function $V(A)$ for the following optimization problem:

$$(P_A) \quad \left| \begin{array}{ll} \text{Minimize} & t \\ \text{s.t.} & tA \in \Omega. \end{array} \right.$$

Although one can use the sensitivity analysis for the value function [10, Theorem 9] to prove that the value function $V(A)$ is locally Lipschitz on S_n^+, we give a direct proof here. Let $\bar{t}$ be a solution of the problem P_A. Then since $\bar{t}A \in \Omega$, we have

$$\bar{t} + d_\Omega(\bar{t}A) \leq t + d_\Omega(tA).$$

Therefore, $\bar{t}$ also minimizes the function $t \to t + d_\Omega(tA)$. That is, $V(A) = \min_t\{t + d_\Omega(tA)\}$ which is a locally Lipschitz function since $(t, A) \to d_\Omega(tA)$ is locally Lipschitz. The continuity of the function $\min I_A$ shows in particular that Ω_0 is closed and completes the proof. ◻

5.2. Inexact approximation. In this subsection, we denote by $\bar{A}_p^\varepsilon$ any ε-solution to problem $(\mathscr{P}_p)$.

THEOREM 5.2. *Let $(p_k)_{k \in \mathbb{N}^\star} \subset [1, \infty)$ be a sequence which tends to infinity. Let ε_k be a decreasing sequence of positive numbers which tends to zero and $\bar{A}_{p_k}^{\varepsilon_k}$ an ε_k-solution to problem $(\mathscr{P}_{p_k})$. For convenience, let $\bar{A}_k := \bar{A}_{p_k}^{\varepsilon_k}$. Then every cluster point $\bar{A}$ of the sequence $(\bar{A}_k)$ (there is at least one) is a global solution of problem $(\mathscr{P})$.*

Proof. By Proposition 2.3(i),

$$0 \in \partial_{\epsilon_k}\left(\kappa_{p_k}^{p_k} + \delta(\cdot|\Omega)\right)(\bar{A}_k).$$

By Proposition 2.3(ii), there exists $\alpha_k \in [0, \varepsilon_k]$ such that

$$0 \in \partial_{\alpha_k}\kappa_{p_k}^{p_k}(\bar{A}_k) + N_{\Omega, \varepsilon_k - \alpha_k}(\bar{A}_k).$$

By Proposition 2.3(iii), there exist

$$B_k \in B(\bar{A}_k, \sqrt{\alpha_k}) \quad \text{and} \quad C_k \in B(\bar{A}_k, \sqrt{\varepsilon_k - \alpha_k})$$

such that

$$\begin{aligned} \partial_{\alpha_k}\kappa_{p_k}^{p_k}(\bar{A}_k) &\subset \partial\kappa_{p_k}^{p_k}(B_k)+B(0,\sqrt{\alpha_k}) \\ \text{and}\quad N_{\Omega,\varepsilon_k-\alpha_k}(\bar{A}_k) &\subset N_\Omega(C_k)+B(0,\sqrt{\varepsilon_k-\alpha_k}). \end{aligned}$$

Consequently,

$$0\in\partial\kappa_{p_k}^{p_k}(B_k)+N_\Omega(C_k)+B(0,\sqrt{\alpha_k}+\sqrt{\varepsilon_k-\alpha_k}). \tag{8}$$

Now observe that, for every $A\in S_n^{++}$, by the chain rule (Proposition 2.2)

$$\partial\kappa_p(A)=\frac{\lambda_n(A)^{p-1}\lambda_1(A)^{\frac{1}{p}-p}}{p}\partial\kappa_p^p(A).$$

Therefore, inclusion (8) can be rewritten as

$$0\in\partial\kappa_{p_k}(B_k)+N_\Omega(C_k)+B(0,\nu_k),$$

with

$$\begin{aligned} \nu_k &:= \frac{\lambda_n(\bar{A}_k)^{p_k-1}\lambda_1(\bar{A}_k)^{\frac{1}{p_k}-p_k}}{p_k}\big(\sqrt{\alpha_k}+\sqrt{\varepsilon_k-\alpha_k}\big) \\ &= \frac{1}{p_k}\left(\frac{\lambda_n(\bar{A}_k)}{\lambda_1(\bar{A}_k)}\right)^{p_k}\frac{\lambda_1(\bar{A}_k)^{\frac{1}{p_k}}}{\lambda_n(\bar{A}_k)}\big(\sqrt{\alpha_k}+\sqrt{\varepsilon_k-\alpha_k}\big). \end{aligned}$$

Now, taking a subsequence if necessary, we can assume that $B_k\to\bar{A}$ and $C_k\to\bar{A}$ as $k\to\infty$. Since we also have that $\nu_k\to 0$ as $k\to\infty$, we can use the same argument as in the last part of the proof of Theorem 5.1 to conclude that

$$0\in\partial\kappa(\bar{A})+N_\Omega(\bar{A}).$$

By the strong pseudoconvexity of κ on S_n^{++}, we then deduce that $\bar{A}$ is globally optimal. ▯

6. Conclusion. The problem of minimizing condition numbers $(\mathscr{P})$ is a nonsmooth and nonconvex optimization problem. In this paper we provide the nonsmooth analysis of condition numbers. In particular we show that the condition number is a Clarke regular pseudoconvex function and we provide an exact formula for the Clarke subdifferential of the condition number. As a nonsmooth and nonconvex optimization problem, $(\mathscr{P})$ is a difficult problem to solve. We consider a nonsmooth convex problem $(\mathscr{P}_p)$ and show that as p goes to infinity, any cluster point of a sequence of exact or inexact solutions to problem $(\mathscr{P}_p)$ is a global solution of $(\mathscr{P})$. It is known that, using the bundle methods (see, e.g., [4]), an exact or an inexact solution of a nonsmooth problem such as $(\mathscr{P}_p)$ can be solved. Hence the asymptotic analysis given in this paper provides a basis to design an implementable algorithm for solving the nonsmooth and nonconvex problem of minimizing condition numbers. The actual design of an algorithm will be left as a future work.

Acknowledgment. We are indebted to the referees for their extremely careful reading of this paper and their valuable suggestions for improving the presentation of this paper. We also wish to thank Huifu Xu and Jérôme Malick for interesting discussions on the topic of the paper.

REFERENCES

[1] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.
[2] J.V. Burke, A.S. Lewis, and M.L. Overton, *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization*, SIAM J. Optim., 15 (2005), pp. 751–779.
[3] V. Guigues, *Inférence Statistique pour l'Optimisation Stochastique*, Ph.D. thesis, Université Joseph Fourier, Grenoble, France, 2005.
[4] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, I *and* II, Springer-Verlag, Berlin, 1993.
[5] K.C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Math. 1133, Springer-Verlag, Berlin, 1985.
[6] K.C. Kiwiel, *A proximal-projection bundle method for Lagrangian relaxation, including semidefinite programming*, SIAM I. Optim., 17 (2006), pp. 1015–1034.
[7] A.S. Lewis, *Nonsmooth analysis of eigenvalues*, Math. Program., 84 (1999), pp. 1–24.
[8] M.M. Mäkelä, *Survey of bundle methods for nonsmooth optimization*, Optim. Methods Softw., 17 (2002), pp. 1–29.
[9] B.S. Mordukhovitch, *Variational Analysis and Generalized Differentiation*, I *and* II, Springer-Verlag, Berlin, 2006.
[10] B.S. Mordukhovitch, N.M. Nam, and N.D. Yen, *Subgradients of marginal functions in parametric mathematical programming*, Math. Program. Ser. B, 116 (2009), pp. 369–396.
[11] P. Maréchal, *On a functional operation generating convex functions. Part I: duality*, J. Optim. Theory Appl., 126 (2005), pp. 175–189.
[12] P. Maréchal, *On a class of convex sets and functions*, Set-Valued Anal., 13 (2005), pp. 197–212.
[13] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[14] H. Xu, *Level function method for quasiconvex programming*, J. Optim. Theory Appl., 108, (2001), pp. 407–437.