Translational Cuts for Convex Minimization

James V. Burke
Department of Mathematics, GN–50, University of Washington, Seattle, WA 98195 USA

Allen A. Goldstein
Department of Mathematics, GN–50, University of Washington, Seattle, WA 98195 USA

Paul Tseng
Department of Mathematics, GN–50, University of Washington, Seattle, WA 98195 USA

Yinyu Ye
Department of Management Sciences, University of Iowa, Iowa City, IO 52242 USA

Abstract

We develop an iterative descent algorithm for minimizing the pointwise maximum of a finite collection of convex thrice-differentiable functions;

$$\min_x\{F(x) := \max_{i=1,\ldots,n} f_i(x)\}.$$

The proposed algorithm begins each iteration with a number R and an inexact 'analytic center', x_R, of the lower level set $\{x \in R^m : F(x) \leq R\}$; it then sets $R := (1-\alpha)F(x_R) + \alpha R$, with α an arbitrarily chosen constant in $(0, 1)$, and recomputes x_R accordingly. The resulting sequence of inexact analytic centers is a descent sequence for F and it is shown that the F value along this sequence comes within ϵ of $\min_x F(x)$ after at most

$$2(1-\alpha)^{-1}[n\ln(1/\epsilon) + \ln(\textstyle\prod_{i=1}^{n}(R_0 - f_i(x_0)))] + 3/2$$

iterations, where x_0 and R_0 are the initial values of x_R and R, respectively, and ϵ is the termination tolerance.

To recompute x_R after each update of R, we propose to use a global newton procedure of [3]. We show that, under a certain nondegeneracy assumption on F and assuming infinite precision arithmetic, the number of newton steps required to recompute x_R is at most a constant plus $\log_2 \log_2(1/\epsilon)$.

Keywords: Complexity, mini-max optimization, global Newton method, analytic center.

1 Introduction

Motivated by interior-point algorithms, several researchers recently studied ways of capturing the combinatorial structure of a convex polytope using a potential function (see [7], [8]). For example, it is shown in [8] that the max-potential of a convex polytope, where the polytope is given as

the intersection of halfspaces, is reduced by a factor of at least e, the base of the natural logarithm, when any of the hyperplanes defining the polytope is translated through the analytic center of the polytope [6]. In this paper we extend this result to convex sets that are given as the intersection of the lower level sets of a finite collection of convex thrice-differentiable functions; and, by using the extended result, we develop a translational-cuts algorithm for finding the minimum of the pointwise maximum of the given functions.

Let f_i , $i = 1, ..., n$, be given convex thrice-differentiable functions defined on the m-dimensional Euclidean space $\mathrm{R^m}$. Let F be the function on $\mathrm{R^m}$ defined to be the pointwise maximum of the f_i's, that is,

$$F(x) := \max\{f_i(x) : i = 1, ..., n\}.$$

We are interested in the problem of finding a minimum point of F, i.e., an $x^* \in \mathrm{R^m}$ satisfying $F(x^*) = R^*$, where for notational convenience we let

$$R^* := \min_x F(x).$$

This is a classical problem in optimization [2]. For the moment we make only the following assumption on F.

Assumption A The set of minimum points of F, i.e., $\{x : F(x) = R^*\}$, is nonempty and compact.

Additional assumptions will be introduced as needed.

As an immediate consequence of the compactness assumption on the minimum points of F, we have that for every $R > R^*$, the lower level set

$$lev_R := \{x : F(x) < R\}$$

is nonempty and bounded. Then, the 'logarithmic potential' function ϕ_R defined by

$$\phi_R(x) := \sum_{i=1}^{n} \ln(R - f_i(x)),$$

which is strictly concave and differentiable on lev_R, attains its maximum on lev_R.

Following the standard practice, we call any maximum point of ϕ_R the *analytic center* of lev_R. As R approaches R^*, the analytic center of lev_R approaches the set of minimum points of F. This motivates an iterative method for finding a minimum point of F whereby, at each iteration, an analytic center of lev_R is computed and R is moved some fraction of the distance towards R^*. However, it is impractical to compute an analytic center exactly and, to resolve this difficulty, we introduce below a notion of an inexact analytic center.

Let $g : [0, \infty) \mapsto [0, \infty)$ be any continuous function satisfying $g(\theta) > 0$ for all $\theta > 0$ and $g(0) = 0$. Relative to g, we say that a point x_R in lev_R is an *inexact analytic center* of lev_R if

$$||\nabla \phi_R(x_R)|| \leq \frac{\min\{R - F(x_R), 1\} g(R - R^*)}{R - R^*}. \tag{1}$$

Notice that, in contrast to previous notions of an inexact analytic center (see [6]), the preceding notion is based on the gradient of the potential function being 'small' (i.e., an inexact root of $\nabla \phi_R$). We have adopted this novel notion because the method which we will use to compute an inexact analytic center (for a fixed R) is based on reducing the norm of this gradient and the work is determined by the total reduction in norm. Suitable choices for g will be discussed shortly (see (3), (10)).

Below we formally describe our algorithm for finding a minimum point of F:

Translational-Cuts Algorithm.
0. Fix a parameter $\alpha \in (0, 1)$ and a termination tolerance $\epsilon > 0$. Start with any $R_0 > R^*$ and any inexact center of lev_{R_0}. Let $R = R_0$ and go to Step 1.
1. Given an $R > R^*$ and an inexact analytic center of lev_R, say x_R, check to see if the termination criterion

$$F(x_R) \leq R^* + \epsilon$$

is met. If yes, stop the algorithm; otherwise go to Step 2.
2. Set

$$R' := (1 - \alpha) F(x_R) + \alpha R$$

and apply the global newton method of [3] to find an inexact center of $lev_{R'}$, say x'_R, with x_R as the starting point for the method. Go to Step 3.
3. Set

$$x_R := x'_R, \qquad R := R'$$

and return to Step 1.

A few words on the parameter α are in order. This parameter controls the decrease in the size of the lower level set or, equivalently, the amount by which the inequalities $f_i(x) \leq R$, $i = 1, ..., n$, are translated (in going from R to R'). If $\alpha = 0$, then at least one inequality will be translated so to cut through the current inexact center x_R; if $\alpha = 1$, then no inequality will be translated; if $\alpha = 1/2$, then at least one inequality will translate halfway to x_R. The reason for introducing the parameter α is so that one can initiate the computation of the new inexact analytic center x'_R from the current inexact analytic center x_R.

The translational-cuts algorithm is most closely related to the 'large-step' non-parametric logarithmic barrier method of [4], applied to the following nonlinear programming formulation of $\min_x F(x)$:

$$(NLP) \qquad \min \gamma$$
$$\text{subject to } f_i(x) - \gamma \leq 0, \; i = 1, \ldots, n \; .$$

However, the method of [4] uses a different notion of inexact analytic center and requires a relative Lipschitz condition on $\nabla^2 f_i$ for all i. It also uses a different termination criterion and requires an objective value lower bound for initialization.

In the remainder of this paper, we analyze the complexity of the translational-cuts algorithm under additional assumptions on the f_i's (see Assumptions B–D). Notice that $F(x_R) < R$ and hence the value of R is monotonically decreasing in the algorithm. Also, although it appears that termination of the algorithm and the computation of an inexact center require knowledge of the optimal objective value R^*, we will show that, under mild assumptions on the f_i's, this difficulty can in fact be circumvented (see discussion at the end of Sections 2 and 3).

2 Iteration count for the translational-cuts algorithm

We will call each repetition of Steps 1-3 in the translational-cuts algorithm an iteration. In this section we estimate the number of iterations used in the algorithm.

By Assumption A, there exists a scalar $C_0 > 0$ such that

$$||x - x'|| \leq C_0 \quad \forall x, x' \in lev_{R_0}. \tag{2}$$

In the following lemma, we estimate the reduction in the potential after each iteration of the translational-cuts algorithm.

Lemma 2.1 *Fix an $\alpha \in (0, 1)$. For any $R \in (R^*, R_0]$ satisfying $g(R-R^*)/(R-R^*) \leq (1-\alpha)/(2C_0)$, any inexact analytic center x of lev_R and any inexact analytic center x' of $lev_{R'}$, where $R' = (1-\alpha)F(x) + \alpha R$, we have*

$$\phi_{R'}(x') \leq \phi_R(x) - (1-\alpha)/2.$$

Proof Since x is an inexact analytic center of lev_R, we have from (1) and the definition of ϕ_R that

$$\left\| \sum_{i=1}^{n} \frac{\nabla f_i(x)}{R - f_i(x)} \right\| \leq \frac{g(R - R^*)}{R - R^*}.$$

Let $\bar{x}'$ be an analytic center of $lev_{R'}$. For each $i = 1, \cdots, n$, let $\beta_i = (R-R')/(R-f_i(x))$ and let $\beta = \sum_{i=1}^{n} \beta_i$. Observe that each β_i satisfies $0 < \beta_i \leq 1 - \alpha$ with at least one

β_i equal to $1-\alpha$. Thus, $\beta \geq 1-\alpha$. Then, the convexity of the f_i's and the preceding relation yield:

$$\begin{aligned}
\sum_{i=1}^{n} \frac{R' - f_i(\bar{x}')}{R - f_i(x)} &= \sum_{i=1}^{n} \frac{R - f_i(\bar{x}') - \beta_i(R - f_i(x))}{R - f_i(x)} \\
&= \sum_{i=1}^{n} \frac{-f_i(\bar{x}') + f_i(x) + (1-\beta_i)(R - f_i(x))}{R - f_i(x)} \\
&= \sum_{i=1}^{n} \frac{-f_i(\bar{x}') + f_i(x)}{R - f_i(x)} + n - \beta \\
&\leq \sum_{i=1}^{n} \frac{-\nabla f_i(x)^T(\bar{x}' - x)}{R - f_i(x)} + n - \beta \\
&\leq \left\| \sum_{i=1}^{n} \frac{\nabla f_i(x)}{R - f_i(x)} \right\| \|\bar{x}' - x\| + n - \beta \\
&\leq \frac{g(R - R^*)}{R - R^*} \|\bar{x}' - x\| + n - \beta \\
&\leq \frac{g(R - R^*)}{R - R^*} C_0 + n - \beta \;\leq\; n - \beta/2,
\end{aligned}$$

where the fourth inequality follows from (2) and the last inequality follows from the inequality $g(R-R^*)C_0/(R-R^*) \leq (1-\alpha)/2 \leq \beta/2$. Dividing both sides by n and taking the logarithm, we obtain

$$\ln(1 - \frac{\beta}{2n}) \geq \ln\left(\frac{1}{n}\sum_{i=1}^{n} \frac{R' - f_i(\bar{x}')}{R - f_i(x)}\right)$$

$$\geq \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{R' - f_i(\bar{x}')}{R - f_i(x)}\right) = \frac{1}{n}(\phi_{R'}(\bar{x}') - \phi_R(x)),$$

where the second inequality uses the concavity of the logarithm function. Upon observing that the lefthand side is bounded above by $-\beta/(2n)$, which in turn is bounded above by $-(1-\alpha)/(2n)$, and then using $\phi_{R'}(\bar{x}') \geq \phi_{R'}(x')$ to bound the righthand side, the result follows. □

By using Lemma 2.1, we can now estimate the number of iterations required to reduce the objective value $F(x_R)$ to within ϵ of the optimal objective value R^*.

Theorem 2.2 *Assume that g is chosen so that*

$$g(\theta)/\theta \leq (1-\alpha)/(4C_0) \quad \forall \theta \in (0, R_0 - R^*]. \tag{3}$$

Then, the translational-cuts algorithm terminates in at most

$$\frac{2}{1-\alpha}[\phi_{R_0}(x_{R_0}) + n\ln(\frac{1}{\epsilon})] + \frac{3}{2}$$

iterations.

Proof Lemma 2.1 assures us that, after k iterations of the translational-cuts algorithm, we have

$$\phi_R(x_R) \leq \phi_{R_0}(x_{R_0}) - k(1-\alpha)/2.$$

Thus, for $k \geq 2[\phi_{R_0}(x_{R_0}) - n \ln \epsilon]/(1-\alpha) + 1/2$, we have

$$\phi_R(x_R) \leq n \ln \epsilon - \frac{1-\alpha}{4}. \tag{4}$$

Let $\bar{x}_R$ be any analytic center of lev_R. Then, using the concavity of ϕ_R and the fact that x_R is an inexact analytic center of lev_R yields

$$\begin{aligned} \phi_R(\bar{x}_R) &\leq \phi_R(x_R) + \nabla \phi_R(x_R)^T(\bar{x}_R - x_R) \\ &\leq n \ln \epsilon - \frac{1-\alpha}{4} + ||\nabla \phi_R(x_R)|| ||\bar{x}_R - x_R|| \\ &\leq n \ln \epsilon - \frac{1-\alpha}{4} + \frac{g(R-R^*)}{R-R^*} C_0 \\ &\leq n \ln \epsilon, \end{aligned}$$

where the last inequality follows from (2) and (3). Therefore,

$$\phi_R(x^*) \leq \phi_R(\bar{x}_R) \leq n \ln \epsilon,$$

where x^* is any minimum point of F. Dividing both sides by n yields

$$\ln \epsilon \geq \frac{\phi_R(x^*)}{n} = \sum_{i=1}^{n} \frac{\ln(R - f_i(x^*))}{n} \geq \min_i \ln(R - f_i(x^*)) = \ln(R - R^*)$$

and thus $R \leq R^* + \epsilon$. Since $F(x_R) < R$ trivially, this completes the proof. □

One possible choice for g that satisfies (3) is the linear function

$$g(\theta) = \theta(1-\alpha)/(4C_0).$$

With this choice, determining whether a point is an inexact analytic center does not require any knowledge of R^*.

Also, it can be seen from the analysis that a sufficient condition for terminating the translational-cuts algorithm is (4) and the iteration count given in Theorem 2.2 is that for satisfying this condition. Thus, we can instead use (4) as a termination condition. This condition has the advantage that it does not require knowledge of the optimal objective value R^*.

3 Work per iteration for the translational-cuts algorithm

In this section we estimate, under additional nondegeneracy and sharp-minimum assumptions (see Assumptions B–D), the work per iteration in the translational-cuts algorithm. This estimate, together with the iteration count obtained in the previous section, will then yield an overall complexity estimate for the translational-cuts algorithm.

The work per iteration is dominated by Step 2 in which the global newton method of [3] is applied to find an inexact analytic center of $lev_{R'}$, starting from x_R. Thus, our analysis will focus on estimating the work for this method.

We start our analysis by applying directly Claim 1 in [3] to obtain a preliminary estimate of the work in Step 2: Let R, R' and x_R be as in Step 2 and let

$$\begin{aligned}
S_{R'} &:= \{\ x \in lev_{R'}\ :\ ||\nabla\phi_{R'}(x)|| \leq ||\nabla\phi_{R'}(x_R)||\ \}, \\
\lambda_{R'} &:= \max_{x\in S_{R'}}\{\ \text{largest eigenvalue of } -\nabla^2\phi_{R'}(x)\ \}, \\
\mu_{R'} &:= \min_{x\in S_{R'}}\{\ \text{smallest eigenvalue of } -\nabla^2\phi_{R'}(x)\ \}, \\
L_{R'} &:= \max_{x\in S_{R'}} \|\nabla^3\phi_{R'}(x)\|.
\end{aligned}$$

According to Claim 1 in [3], the number of steps required by the method of [3] to find an approximate root of $\nabla\phi_{R'}$ (i.e., reducing $\|\nabla\phi_{R'}\|$ to less than $\beta_{R'} := (\mu_{R'})^2/4L_{R'}$), starting from an inexact analytic center x_R of lev_R, is at most

$$(\|\nabla\phi_{R'}(x_R)\| - \beta_{R'})/4\delta^2\beta_{R'}, \tag{5}$$

where δ is chosen from the interval $(0, 1/\sqrt{8}]$. Moreover, the number of steps required by the method of [3] to find an inexact analytic center x'_R of $lev_{R'}$ (i.e., x'_R satisfies $\|\nabla\phi_{R'}(x'_R)\| \leq \min\{R'-F(x'_R), 1\}g(R'-R^*)/(R'-R^*)$), starting from an approximate root of $\nabla\phi_{R'}$, is at most

$$\log_2\log_2\left(\frac{\mu_{R'}\lambda_{R'}(R'-R^*)}{L_{R'}\min\{R'-F(x'_R),1\}g(R'-R^*)}\right). \tag{6}$$

In the remainder of this section, we focus on estimating the four quantities:

$$\|\nabla\phi_{R'}(x_R)\|,\ \lambda_{R'},\ \mu_{R'},\ L_{R'} \tag{7}$$

in terms of the problem parameters. We will estimate the latter three quantities by replacing the set $S_{R'}$ in their definition with the following set

$$T_{R'} := \left\{\ x \in lev_{R'}\ :\ ||\nabla\phi_{R'}(x)|| \leq \frac{2n\sqrt{n}B_1}{\alpha(R'-R^*)}\ \right\}, \tag{8}$$

where B_1 is any positive scalar satisfying

$$||\nabla f_i(x)|| \leq B_1 \quad \forall x \in lev_{R_0},\ i = 1, ..., n \tag{9}$$

(B_1 exists by Assumption A).

Our first order of business is to show that $S_{R'} \subset T_{R'}$. To show this, we make the following simplifying assumption, namely, $-\nabla^2\phi_R$ is positive definite uniformly over

lev_R and over all $R \in (R^*, R_0]$. This assumption will also be needed for bounding $\mu_{R'}$ from below.

Assumption B There exists a scalar $\mu > 0$ such that

$$(\text{smallest eigenvalue of } -\nabla^2 \phi_R(x)) \geq \mu \quad \forall x \in lev_R, \ \forall R \in (R^*, R_0].$$

Assumption B is actually quite mild. For example, Assumption B holds when any one of the f_i's is strongly convex. Alternatively, in view of Assumption A, we can enforce that Assumption B holds by choosing a sufficiently large number L so that the box

$$\{ \ x \in \Re^m \mid x \leq (L, ..., L) \ \}$$

contains the set of minimum points of F, and then set

$$f_{n+j}(x_1, ..., x_m) := x_j - L - R_0, \quad j = 1, ..., m.$$

It is straightforward to verify that $\max_{i=1,m+n} f_i(x)$ has the same set of minimum points as $F(x)$ (so that Assumption A and Assumptions C and D, which are to come, hold for $f_1, ..., f_{n+m}$ whenever they hold for $f_1, ..., f_n$) and that Assumption B holds with $\phi_R(x)$ replaced by $\sum_{i=1}^{n+m} \ln(R - f_i(x))$.

We show below that $S_{R'} \subset T_{R'}$ through a sequence of three lemmas.

Lemma 3.1 *For any $R \in (R^*, R_0]$ and any analytic center $\bar{x}_R$ of lev_R, we have*

$$R - R^* \leq n(R - F(\bar{x}_R)).$$

Proof By definition of an analytic center, we have

$$0 = \nabla \phi_R(\bar{x}_R) = \sum_{i=1}^{n} \frac{\nabla f_i(\bar{x}_R)}{R - f_i(\bar{x}_R)},$$

so that

$$\sum_{i=1}^{n} \frac{R - f_i(\bar{x}_R) - \nabla f_i(\bar{x}_R)(x - \bar{x}_R)}{R - f_i(\bar{x}_R)} = n \quad \forall x \in lev_R.$$

For each i and each $x \in lev_R$, we have from the convexity of f_i that $R \geq f_i(x) \geq f_i(\bar{x}_R) + \nabla f_i(\bar{x}_R)(x - \bar{x}_R)$ and hence every term in the above sum is nonnegative. Thus, we conclude that

$$\frac{R - f_i(\bar{x}_R) - \nabla f_i(\bar{x}_R)(x - \bar{x}_R)}{R - f_i(\bar{x}_R)} \leq n \quad \forall x \in lev_R, \ i = 1, ..., n.$$

[This relation can also be inferred from observing that $\bar{x}_R$ is an analytic center of the polytope $\{x : \nabla f_i(\bar{x}_R)(x - \bar{x}_R) \leq R - f_i(\bar{x}_R), \ i = 1, ..., n\}$ and invoking a containing-ellipsoid property for this polytope.] Let x^* be any minimum point of F. Since

$x^* \in lev_R$, we obtain, upon using the convexity of the f_i's and invoking the above relation with $x = x^*$, that

$$R - f_i(x^*) \le R - f_i(\bar{x}_R) - \nabla f_i(\bar{x}_R)(x^* - \bar{x}_R) \le n(R - f_i(\bar{x}_R)), \quad i = 1, ..., n.$$

Thus,

$$R - R^* \le n(R - f_i(\bar{x}_R)), \quad i = 1, ..., n,$$

and the result readily follows. □

Lemma 3.2 *Let Assumption B hold. Suppose that g is chosen so that*

$$g(\theta)/\theta \le \mu/2B_1 \qquad \forall \theta \in (0, R_0 - R^*]. \tag{10}$$

Then, for any $R \in (R^*, R_0]$ and any inexact analytic center x_R of lev_R, we have

$$R - R^* \le 2n(R - F(x_R)).$$

Proof Let $\bar{x}_R$ be an analytic center of lev_R. By using $\nabla \phi_R(\bar{x}_R) = 0$ and Assumption B and the concavity of $\phi_R(.)$, we have

$$(\mu/2)\|\bar{x}_R - x_R\|^2 \le \phi_R(\bar{x}_R) - \phi_R(x_R) \le \nabla \phi_R(x_R)(\bar{x}_R - x_R),$$

which implies

$$\|\bar{x}_R - x_R\| \le \frac{2}{\mu}||\nabla \phi_R(x_R)||.$$

Also, for each $i \in \{1, ..., n\}$, we have

$$R - f_i(\bar{x}_R) - (R - f_i(x_R)) = \int_0^1 -\nabla f_i(x_R + t(\bar{x}_R - x_R))^T(\bar{x}_R - x_R)dt.$$

Upon combining the above two relations, we obtain

$$\begin{aligned} \left|\frac{R - f_i(\bar{x}_R)}{R - f_i(x_R)} - 1\right| &= \frac{|\int_0^1 \nabla f_i(x_R + t(\bar{x}_R - x_R))^T(\bar{x}_R - x_R)dt|}{R - f_i(x_R)} \\ &\le \frac{\int_0^1 \|\nabla f_i(x_R + t(\bar{x}_R - x_R))\|\|\bar{x}_R - x_R\|dt}{R - f_i(x_R)} \\ &\le \frac{\int_0^1 B_1\|\bar{x}_R - x_R\|dt}{R - f_i(x_R)} \\ &\le \frac{2B_1\|\nabla \phi_R(x_R)\|}{\mu(R - f_i(x_R))} \\ &\le \frac{2B_1 g(R - R^*)(R - F(x_R))}{\mu(R - f_i(x_R))(R - R^*)} \\ &\le \frac{2B_1 g(R - R^*)}{\mu(R - R^*)} \le 1, \end{aligned}$$

where the second inequality follows from (9), the fourth inequality follows from (1), the fifth inequality follows from the relation $0 < R - F(x_R) \leq R - f_i(x_R)$, and the last inequality follows from (10). This inequality together with Lemma 3.1 yields

$$2(R - f_i(x_R)) \geq R - f_i(\bar{x}_R) \geq \frac{R - R^*}{n}, \ i = 1, ..., n,$$

and the result readily follows. □

Lemma 3.3 *Let Assumption B hold and suppose that g satisfies (10). Then, for any $\alpha \in (0,1)$, any $R \in (R^*, R_0]$, and any inexact analytic center x_R of lev_R, we have*

$$||\nabla \phi_{R'}(x_R)|| \leq \frac{2n\sqrt{n}B_1}{\alpha(R' - R^*)},$$

where $R' := (1-\alpha)F(x_R) + \alpha R$.

Proof From the definition of R', we have $R' - F(x_R) = \alpha(R - F(x_R))$ and so Lemma 3.2 yields

$$R' - F(x_R) \geq \frac{\alpha}{2n}(R - R^*).$$

Then, using the fact $x_R \in lev_{R_0}$ and (9), we conclude

$$\begin{aligned} \|\nabla \phi_{R'}(x_R)\| &= \left\| \sum_{i=1}^{n} \frac{\nabla f_i(x_R)}{R' - f_i(x_R)} \right\| \leq \frac{\sqrt{n}B_1}{R' - F(x_R)} \\ &\leq \frac{2n\sqrt{n}B_1}{\alpha(R - R^*)} \leq \frac{2n\sqrt{n}B_1}{\alpha(R' - R^*)}, \end{aligned}$$

where the last inequality follows from $R \geq R'$. □

Lemma 3.3 shows that $S_{R'} \subset T_{R'}$, provided that g is chosen to satisfy (10). We will now estimate the quantities (7) by replacing $S_{R'}$ in their definition with the larger set $T_{R'}$. To do this, we need to make the following assumption on F.

Assumption C The function F has a *strongly unique* minimum x^*, that is, $F(x^*) = R^*$ and there exists a scalar $\sigma > 0$ such that

$$F(x) \geq R^* + \sigma\|x - x^*\| \quad \forall x.$$

The notion of a strongly unique minimum (or 'sharp minimum' [5]) was first extensively studied in [1], where its connection to the convergence behavior of algorithms was reviewed. Notice that Assumption C superceeds Assumption A.

Let

$$I^* := \{i \in \{1, ..., n\} \ : \ f_i(x^*) = R^*\}. \tag{11}$$

A key part of our analysis lies in showing that (see (19))

$$\liminf_{R\downarrow 0} \min_{x\in T_R} \left\{ \frac{R - F(x)}{R - f_i(x)} \right\} > 0 \quad \forall i \in I^* \ .$$

To show this, we need to make one further assumption on F.

Assumption D For any set of nonnegative scalars λ_i, $i \in I^*$, satisfying

$$\sum_{i\in I^*} \lambda_i \nabla f_i(x^*) = 0 \ \text{ and } \sum_{i\in I^*} \lambda_i = 1 \ , \tag{12}$$

there holds $\lambda_i > 0$ for all $i \in I^*$.

It can be seen that Assumption C implies $|I^*| \geq m+1$ and that Assumption D implies $|I^*| \leq m+1$. (To see the former, note that if $|I^*| \leq m$, there would exist a nonzero $d \in \mathrm{R^m}$ satisfying $\nabla f_i(x^*)^T d \leq 0$ for all $i \in I^*$, implying $\lim_{\theta\downarrow 0}(F(x^* + \theta d) - R^*)/\theta \leq 0$ and thus violating Assumption C.) Thus, Assumptions C and D together imply $|I^*| = m+1$. The condition specified in Assumption D is a standard nondegeneracy condition used in the convergence analysis of algorithms for mini-max optimization. In fact, this condition corresponds to two separate conditions often employed in nonlinear programming: *linear independence of the active constraint gradients* and *strict complementary slackness*. To see this, recall that the problem $\min_x F(x)$ is equivalent to the nonlinear program (NLP). Then, Assumption D is equivalent to the assumption that the gradients of the constraint functions in (NLP) that are active at x^*, namely

$$\left\{ \begin{pmatrix} \nabla f_i(x^*) \\ -1 \end{pmatrix} : i \in I^* \right\} ,$$

are linearly independent and that there exists a set of strictly complementary Lagrange multipliers at x^*, i. e., there exist scalars $\lambda_i > 0, \ i \in I^*$, satisfying (12).

Lemma 3.4 *Suppose that Assumptions B through D hold. Fix any $\alpha \in (0,1)$. Then there exist scalars $C_1 > 0$ and $C_2 > 0$ satisfying*

$$R - F(x) \ \geq \ C_1(R - R^*) \qquad \forall x \in T_R, \ \forall R \in (R^*, R_0) \tag{13}$$

$$R - F(x) \ \geq \ C_2(R - f_i(x)) \qquad \forall x \in T_R, \ \forall R \in (R^*, R_0), \ \forall i \in I^*, \tag{14}$$

where T_R and I^* are given by (8) and (11), respectively.

Proof We will argue (13) by contradiction. Suppose that (13) does not hold for any $C_1 > 0$. Then, there would exist a sequence $\{(x^k, R^k)\}$ satisfying

$$F(x^k) < R^k < R_0 \text{ and } ||\nabla \phi_{R^k}(x^k)|| \leq \frac{2n\sqrt{n}B_1}{\alpha(R^k - R^*)} \qquad \forall k, \tag{15}$$

and yet

$$\frac{R^k - F(x^k)}{R^k - R^*} \to 0. \tag{16}$$

These two relations together imply

$$(R^k - F(x^k)) \to 0 \tag{17}$$

and

$$\frac{||\nabla\phi_{R^k}(x^k)||}{\sum_{i=1}^n 1/(R^k - f_i(x^k))} \leq \frac{||\nabla\phi_{R^k}(x^k)||}{1/(R^k - F(x^k))} \leq \frac{2n\sqrt{n}B_1(R^k - F(x^k))}{\alpha(R^k - R^*)} \to 0 \tag{18}$$

as $k \to 0$. Let $((\lambda_1, ..., \lambda_n), R^\infty, x^\infty)$ be any cluster point of the sequence

$$\left\{ \frac{((1/(R^k - f_1(x^k)), ..., 1/(R^k - f_n(x^k)))}{\sum_{i=1}^n 1/(R^k - f_i(x^k))}, R^k, x^k \right\},$$

and let $I^\infty := \{i \in \{1, ..., n\} : f_i(x^\infty) = F(x^\infty)\}$. The relation (18) implies that

$$\sum_{i=1}^n \lambda_i \nabla f_i(x^\infty) = 0, \qquad \sum_{i=1}^n \lambda_i = 1, \qquad \lambda_i \geq 0,\ i = 1, ..., n.$$

Since $R^\infty \geq F(x^\infty) > f_i(x^\infty)$ for all $i \notin I^\infty$, we know from (17) that $\lambda_i = 0$ for all $i \notin I^\infty$. Therefore, x^∞ satisfies the sufficient conditions for a minimum point of F. Since x^* is the unique minimum point of F, this implies $x^\infty = x^*$ and $I^\infty = I^*$, so (17) yields $R^\infty = F(x^\infty) = R^*$ and Assumption D yields $\lambda_i > 0$ for all $i \in I^*$. Thus, we conclude that there exist positive scalars η_i, $i \in I^*$, such that

$$\frac{R^k - F(x^k)}{R^k - f_i(x^k)} > \eta_i \quad \forall i \in I^*,\ \forall k. \tag{19}$$

Fix any k. By using the concavity of ϕ_{R^k}, relation (15), and Assumption C, we have

$$\begin{aligned} \phi_{R^k}(x^*) &\leq \phi_{R^k}(x^k) + \langle \nabla\phi_{R^k}(x^k), x^* - x^k \rangle \\ &\leq \phi_{R^k}(x^k) + ||\nabla\phi_{R^k}(x^k)||||x^* - x^k|| \\ &\leq \phi_{R^k}(x^k) + \frac{2n\sqrt{n}B_1}{\alpha(R^k - R^*)}||x^* - x^k|| \\ &\leq \phi_{R^k}(x^k) + \frac{2n\sqrt{n}B_1}{\alpha\sigma}. \end{aligned}$$

Upon combining this with (19), we conclude that

$$|I^*| \ln(R^k - R^*) \quad + \quad \sum_{i \notin I^*} \ln(R^k - f_i(x^*))$$

$$
\begin{aligned}
&= \phi_{R^k}(x^*) \\
&\le \phi_{R^k}(x^k) + \frac{2n\sqrt{n}B_1}{\alpha\sigma} \\
&= \sum_{i \in I^*} \ln(R^k - f_i(x^k)) \\
&\qquad + \sum_{i \notin I^*} \ln(R^k - f_i(x^k)) + \frac{2n\sqrt{n}B_1}{\alpha\sigma} \\
&\le |I^*| \ln(R^k - F(x^k)) - \sum_{i \in I^*} \ln(\eta_i) \\
&\qquad + \sum_{i \notin I^*} \ln(R^k - f_i(x^k)) + \frac{2n\sqrt{n}B_1}{\alpha\sigma}.
\end{aligned}
$$

Also, by continuity of the f_i's, there exists a scalar $\theta > 0$ such that

$$
R^k - f_i(x^k) \ge \theta \quad \forall i \notin I^* \ , \ \forall k.
$$

Upon combining the above two relations and using the fact that $\ln(\cdot)$ is an increasing function, we obtain that

$$
\begin{aligned}
& |I^*| \ln(R^k - R^*) + (n - |I^*|) \ln(\theta) \\
\le\ & |I^*| \ln(R^k - F(x^k)) - \sum_{i \in I^*} \ln(\eta_i) + \sum_{i \notin I^*} \ln(B_2) + \frac{2n\sqrt{n}B_1}{\alpha\sigma},
\end{aligned}
$$

where B_2 is any positive scalar satisfying

$$
R_0 - f_i(x) \le B_2 \qquad \forall i \in I^*,\ \forall x \in lev_{R_0}
$$

(such a B_2 exists by the compactness of lev_{R_0}). Dividing both sides by $|I^*|$ and then taking the exponential yields the inequality

$$
R^k - R^* \le C'(R^k - F(x^k)) \quad \forall k,
$$

where C' is some suitable positive scalar. This contradicts (16) and thus (13) must hold for some $C_1 > 0$.

Finally, we show that there exists a scalar $C_2 > 0$ such that (14) holds. By the preceding argument, there exists a scalar $C_1 > 0$ such that (13) holds. Consider any $R \in (R^*, R_0)$ and any $x \in T_R$. For each $i \in I^*$, we have from $f_i(x^*) = R^*$ and the convexity of f_i that

$$
f_i(x) \ge R^* - B_1||x - x^*||,
$$

so Assumption C yields

$$
\begin{aligned}
R - f_i(x) &\le R - R^* + B_1||x - x^*|| \\
&\le R - R^* + \frac{B_1}{\sigma}(R - R^*).
\end{aligned}
$$

Combining this with (13) and we obtain

$$R - F(x) \geq \frac{C_1}{1 + B_1/\sigma}(R - f_i(x)) \qquad \forall i \in I^*.$$

Set $C_2 := C_1/(1 + B_1/\sigma)$. □

Armed with Lemmas 3.3 and 3.4, we can now proceed to bound, in terms of the problem parameters, the four quantities (7) used in the complexity estimates (5) and (6). In what follows, we assume that Assumptions B–D hold and that g is chosen to satisfy (10).

First, we bound from above $\lambda_{R'}$. For each R, straightforward calculation shows that

$$-\nabla^2 \phi_R(x) = \sum_{i=1}^n \frac{\nabla f_i(x) \nabla f_i(x)^T}{(R - f_i(x))^2} + \sum_{i=1}^n \frac{\nabla^2 f_i(x)}{R - f_i(x)}.$$

Since lev_{R_0} is bounded (by Assumption A), it follows that the largest eigenvalue of $-\nabla^2 \phi_R(x)$ is bounded above by some constant divided by $(R - F(x))^2$ for all $x \in T_R$ and all $R \in (R^*, R_0)$. By (13) in Lemma 3.4, the quantity $R - F(x)$ is bounded below by a constant times $R - R^*$; hence we conclude that (also using the fact $S_{R'} \subset T_{R'}$)

$$\lambda_{R'} \leq C_3/(R' - R^*)^2,$$

for some scalar $C_3 > 0$.

Next, we bound from below $\mu_{R'}$. By Assumption C, the smallest eigenvalue of

$$\sum_{i \in I^*} \nabla f_i(x) \nabla f_i(x)^T$$

is bounded below by some positive constant, uniformly over all x in some neighborhood of x^*. (Otherwise there would exist a $d \in \mathrm{R}^{\mathrm{m}}$ satisfying $\nabla f_i(x^*)^T d = 0$ for all $i \in I^*$, implying $\lim_{\theta \downarrow 0} (F(x^* + \theta d) - R^*)/\theta = 0$ and thus violating Assumption C.) This observation together with (14) in Lemma 3.4 implies that the smallest eigenvalue of

$$\sum_{i \in I^*} \frac{\nabla f_i(x) \nabla f_i(x)^T}{(R - f_i(x))^2}$$

is bounded below by some positive constant divided by $(R - F(x))^2$, uniformly over all x in T_R and all R sufficiently close to R^*. Since $F(x) \geq R^*$, we can replace $F(x)$ in the preceding bound by R^*. Summarizing the above results, we see that the smallest eigenvalue of $-\nabla^2 \phi_R$ is bounded below by some positive scalar C_4 divided by $(R - R^*)^2$, uniformly over T_R and over all R sufficiently close to R^*. In view of Assumption B (and by taking C_4 sufficiently small if necessary), we can extend this bound to hold over T_R and over all $R < R_0$. Since $S_{R'} \subset T_{R'}$, this yields

$$\mu_{R'} \geq C_4/(R' - R^*)^2.$$

Finally, we determine a suitable value for $L_{R'}$. For each R, straightforward calculation shows that

$$\|\nabla^3\phi_R(x)\| \le 3\sum_{i=1}^n \frac{\|\nabla f_i(x)\|\|\nabla^2 f_i(x)\|}{(R-f_i(x))^2} + \sum_{i=1}^n \frac{\|\nabla^3 f_i(x)\|}{R-f_i(x)} + 2\sum_{i=1}^n \frac{\|\nabla^2 f_i(x)\|^3}{(R-f_i(x))^3}.$$

Since lev_{R_0} is bounded, this implies that $\|\nabla^3\phi_R(x)\|$ is bounded above by some constant divided by $(R-F(x))^3$, uniformly over all $x \in T_R$ and all $R < R_0$. By (13) in Lemma 3.4, we can in turn bound $R - F(x)$ from below by $C_1(R-R^*)$. Since $S_{R'} \subset T_{R'}$, this implies that we can take

$$L_{R'} := C_5/(R'-R^*)^3$$

for some suitable scalar C_5.

Thus

$$\beta_{R'} = \frac{(\mu_{R'})^2}{4L_{R'}} \ge \frac{(C_4/(R'-R^*)^2)^2}{4(C_5/(R'-R^*)^3)} = \frac{(C_4)^2}{4C_5(R'-R^*)},$$

which together with Lemma 3.3 yields that the quantity in (5) is bounded above by some positive constant independent of $R'-R^*$. Similarly, using the trivial observation $\mu_{R'} \le \lambda_{R'}$, we have that the quantity in (6) is bounded above by

$$\log_2\log_2\left(\frac{(C_3)^2}{C_5\min\{R'-F(x'_R),1\}g(R'-R^*)}\right).$$

Since x'_R is an inexact analytic of $lev_{R'}$, we can apply (13) with x and R replaced by x'_R and $R'' := (1-\alpha)F(x'_R)+\alpha R'$, respectively, to obtain

$$R'-F(x'_R) \ge R''-F(x'_R) \ge C_1(R''-R^*) \ge \alpha C_1(R'-R^*).$$

Combining the preceding observations and we obtain the following key complexity estimate for Step 2 of the translational-cuts algorithm.

Theorem 3.5 *Let Assumptions B–D hold and assume that g is chosen to satisfy (10). Then, the total number of steps required by the method of [3] to find an inexact analytic center of $lev_{R'}$, starting from an inexact analytic center x_R of lev_R (and with $R' := (1-\alpha)F(x_R)+\alpha R$), is at most a constant plus $\log_2\log_2(1/(R'-R^*)g(R'-R^*))$.*

Suppose furthermore that g is an increasing function. Then, since the termination criterion in Step 1 is not met each time we visit Step 2, we know that $R'-R^* = (1-\alpha)(F(x_R)-R^*)+\alpha(R-R^*) \ge F(x_R)-R^* > \epsilon$; hence the complexity estimate in Theorem 3.5 is in turn bounded above by some constant plus $\log_2\log_2(1/\epsilon g(\epsilon))$.

Finally, we note that a similar estimate can be made if (4) is used as the termination criterion. In particular, observe that

$$\phi_R(x_R) \le (n-1)\ln(B_2) + \ln(R-F(x_R))$$

for any $R \in (R^*, R_0)$ and any x_R in lev_R, where B_2 is defined as in Lemma 3.4. Thus, if in addition (4) is not satisfied (so that $\phi_R(x_R) > n \ln \epsilon - (1-\alpha)/4$), we would have

$$R - R^* \geq R - F(x_R) \geq \frac{\epsilon^n}{e^{(1-\alpha)/4}(B_2)^{n-1}}$$

and so $R' - R^*$ (which is bounded below by $\alpha(R - R^*)$) is bounded below by some scalar C_6 times ϵ^n. Thus, in this case (and assuming that g is an increasing function), the total number of steps required by the method of [3] to find an inexact analytic center of $lev_{R'}$, starting from an inexact analytic center x_R of lev_R, is at most a constant plus $\log_2 \log_2(1/\epsilon^n g(C_6\epsilon^n))$.

References

[1] L. Cromme, *Strong uniqueness*, Numerische Mathematik, 29 (1978), 179–193.

[2] V.F. Demyanov and V.N. Malozemov, *Introduction to Minimax*, John Wiley and Sons, New York, 1974.

[3] A.A. Goldstein, *A global Newton method*, Applied Geometry and Discrete Mathematics, 4 (1991), 301-307.

[4] D.D. Hertog, C. Roos, and T. Terlaky, *A large-step analytic center method for a class of smooth convex programming problems*, SIAM Journal on Optimization, 2 (1992), 55-70.

[5] B.T. Polyak, *Sharp minima*, Institute of Control Sciences Lecture Notes, Moscow (1979); Presented at the IIASA Workshop on Generalized Lagrangians and Their Applications, Laxenburg (1979).

[6] G. Sonnevend, *An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming*, System Modelling and Optimization: Proceedings of the 12th IFIP–Conference, Budapest (1985); Eds. A. Prekopa, J. Szelezsan and B. Strazicky, Lecture Notes in Control and Information Sciences, 84 (1986), 866–876.

[7] P.M. Vaidya, *A new algorithm for minimizing convex functions over convex sets*, in Proceedings of 30th Annual IEEE Symposium on the Foundations of Computer Science (IEEE Computer Society Press, Los Alamitos, 1989), 338-343.

[8] Y. Ye, *A Combinatorial property of the analytic centers of polytopes*, Manuscript, University of Iowa, Iowa City (1989).