THE NEWTON BRACKETING METHOD FOR THE MINIMIZATION OF CONVEX FUNCTIONS SUBJECT TO AFFINE CONSTRAINTS

ADI BEN-ISRAEL AND YURI LEVIN

Abstract. The Newton Bracketing method [9] for the minimization of convex functions $f : \mathbb{R}^n \to \mathbb{R}$ is extended to affinely constrained convex minimization problems. The results are illustrated for affinely constrained Fermat–Weber location problems.

1. Introduction

The *Newton Bracketing method* (*NB method* for short) is an iterative method for the minimization of convex functions $f : \mathbb{R}^n \to \mathbb{R}$, see [9]. An iteration of the NB method begins with an interval (or *bracket*) $[L, U]$ containing the sought minimum value of f. An iteration consists of one Newton iteration and results in a reduction of the bracket.

The NB method is valid for $n = 1$, and for $n > 1$ if f is well-conditioned. Its advantage over other methods of convex minimization is that the NB method has a natural *stopping rule*, namely the size $U - L$ of the bracket.

We recall that the *Fermat–Weber problem* is to determine the optimal location of a facility serving a given set of customers, where the objective function to be minimized is the sum of weighted distances between the facility and customers, see, e.g., [5], [6] and [11] for surveys of theory, applications and methods.

The NB method was applied in [9] to the Fermat–Weber problem, and in [10] to multi-facility location problems. These are natural applications, because in large scale location problems the objective is well-conditioned, and the NB method is valid, with fast convergence.

In this paper we propose an extension of the NB method to the *affinely constrained convex minimization problem*

$$\min f(\mathbf{x}) \tag{CP}$$
$$\text{s.t.} \quad A\mathbf{x} = \mathbf{b}, \tag{1}$$

Date: August 25, 2005.

Key words and phrases. Convex minimization, directional Newton method, affine constraints, Fermat–Weber location problem.

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, differentiable[1], and its restriction to $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$ is bounded below, with attained infimum.

As in [9] we illustrate our results for location problems, where it is often the case that there are affine constraints on the facility location: for example, a warehouse may have to be located along a given highway or railroad track, which can be locally approximated as a line in the plane. Such cases are instances of the *affinely constrained location problem*:

$$\min_{\mathbf{x}} \sum_{i=1}^{N} w_i \|\mathbf{x} - \mathbf{a}_i\| \tag{CL}$$

$$\text{s.t.} \quad A\mathbf{x} = \mathbf{b}, \tag{1}$$

where:

$\|\cdot\|$ denotes the *norm* used (Euclidean unless otherwise stated);
N is the *number of customers*;
$\mathbf{a}_i$ is the *location* (coordinates) of the i^{th} customer;
w_i is a *weight* (cost, demand) associated with the i^{th} customer;
$\mathbf{x}$ is the sought *location* of the facility serving the customers; and
$A\mathbf{x} = \mathbf{b}$, the *linear constraints* on the location $\mathbf{x}$.

Plan of this paper: The NB method is reviewed in Section 3. In Section 4 we present an extension of the NB method, called the *projected gradient NB method* (*PNB method* for short), for solving the affinely constrained convex minimization problem (CP). The PNB method is studied in Section 5, and applied in Section 6 to the linearly constrained location problem (CL).

In Section 7 we report numerical experience with the PNB method. The PNB method is suitable for large-scale location problems (CL), see Example 1, and has certain advantages over its unconstrained analog, the NB method. These advantages are discussed in Section 8. In particular, the PNB method is valid for line constraints, Theorem 2.

2. Notation and preliminaries

Let L be a linear subspace of $\mathbb{R}^n$, P_L the *orthogonal projection* on L. It is calculated by

$$P_L = \sum_{i=1}^{\ell} \mathbf{v}_i \mathbf{v}_i^T \,, \tag{2}$$

where $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_\ell\}$ is an orthonormal basis of L.

The equation (1) is assumed consistent, i.e., the manifold

$$S = \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\} \tag{3}$$

---

[1]This is assumed for convenience, as differentiability can be relaxed using standard results of convex analysis.

is nonempty. It can be written as

$$S = \mathbf{x}^0 + N(A), \tag{4}$$

where $\mathbf{x}^0$ is any point in S, and

$$N(A) = \{\mathbf{x} : A\mathbf{x} = \mathbf{0}\} \tag{5}$$

is the *null space* of A. The orthogonal projection $P_{N(A)}$ has the following explicit form, alternative to (2),

$$P_{N(A)} = I - A^\dagger A, \tag{6}$$

where $A^\dagger$ is the Moore–Penrose inverse of A. The orthogonal projection on the manifold S can be written as

$$P_S(\mathbf{x}) = A^\dagger \mathbf{b} + P_{N(A)}\mathbf{x}. \tag{7}$$

It is the unique minimizer of $\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in S\}$, where $\|\cdot\|$ is the *Euclidean norm.*

We have occasion to use the directional Newton iteration introduced in [8] for a function $f : \mathbb{R}^n \to \mathbb{R}$ and a direction $\mathbf{d} \in \mathbb{R}^n$, $\mathbf{d} \neq \mathbf{0}$,

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\langle \nabla f(\mathbf{x}), \mathbf{d}\rangle}\mathbf{d}, \tag{8a}$$

which for $n = 1$ reduces to the ordinary Newton iteration

$$x_+ := x - \frac{f(x)}{f'(x)}. \tag{8b}$$

A common choice of the direction $\mathbf{d}$ is the gradient $\nabla f(x)$, in which case (8a) becomes

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2}\nabla f(\mathbf{x}). \tag{8c}$$

3. The NB method

Consider the (unconstrained) *convex minimization* problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}), \tag{P}$$

where f is a differentiable convex function, bounded below, with attained infimum $f_{\min}$.

An iteration of the NB method begins with an approximate solution $\mathbf{x}$, and an interval $[L, U]$, called a *bracket*, containing the minimum value $f_{\min}$,

$$L \leq f_{\min} \leq U. \tag{9}$$

The upper bound is $U := f(\mathbf{x})$ where $\mathbf{x}$ is the current iterate. An initial lower bound L^0 is assumed known. At each iteration the bracket $[L, U]$ is reduced, either by lowering U or by raising L.

If the bracket is sufficiently small, say

$$U - L < \epsilon \tag{10}$$

then the current $\mathbf{x}$ is declared optimal, and computations stop.

For each non-terminal step, define $0 < \alpha < 1$ and select $M \in [L, U]$:

$$M := \alpha U + (1 - \alpha) L,\ \ 0 < \alpha < 1. \tag{11}$$

For a suitable direction $\mathbf{d}$, do one iteration of the directional Newton method (8a),

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x}) - M}{\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle} \mathbf{d}, \tag{12}$$

as if to solve

$$f(\mathbf{x}) = M. \tag{13}$$

The new value $f(\mathbf{x}_+)$ then allows narrowing the bracket $[L, U]$ to obtain a new bracket $[L_+, U_+]$, as follows:

$$\text{Case 1: if } f(\mathbf{x}_+) \ < \ f(\mathbf{x}) \ \text{ then } \ U_+ := f(\mathbf{x}_+), \tag{14a}$$

$$\text{Case 2: if } f(\mathbf{x}_+) \ \geq \ f(\mathbf{x}) \ \text{ then } \ L_+ := M\ ,\ \mathbf{x}_+ := \mathbf{x}. \tag{14b}$$

In either case the bracket is reduced, the *reduction ratio* is

$$\frac{U_+ - L_+}{U - L} = \begin{cases} \dfrac{f(\mathbf{x}_+) - L}{f(\mathbf{x}) - L} & \text{in Case 1,} \\ 1 - \alpha & \text{in Case 2.} \end{cases} \tag{15}$$

The NB method is *valid* for minimizing f if every iteration produces a bracket, i.e., if (9) holds throughout the iterations. To prove validity it suffices to show that the lower bound L_+ in (14b) is correct (the update in (14a) is clearly valid).

The NB method is valid in the case $n = 1$, see [9, Theorem 1]. It is valid for $n > 1$, using the directional Newton iteration (8c) in (12), if the level sets of f are not "too narrow", see [9, Theorems 2–5]. A typical sufficient condition is: let f be a quadratic function

$$f(\mathbf{x}) = \tfrac{1}{2}\, \mathbf{x}^T Q \mathbf{x} - c^T \mathbf{x} + \gamma\,, \tag{16}$$

where the matrix Q is positive definite with eigenvalues

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

Then the NB method is valid for minimizing a quadratic function f if the *condition number* of the matrix Q is sufficiently small,

$$\operatorname{cond}(Q) := \frac{\lambda_n}{\lambda_1} \leq \frac{1}{7 - \sqrt{48}} \approx 13.92820356, \tag{17}$$

see [9, Theorem 4].

4. The PNB method for solving linearly constrained convex minimization problems

The problem

$$\min f(\mathbf{x}) \tag{CP}$$
$$\text{s.t.} \quad A\mathbf{x} = \mathbf{b}, \tag{1}$$

is specified by the triple $\{f, A, \mathbf{b}\}$, where $S = \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$ is assumed nonempty and $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, differentiable and bounded below with an attained infimum $f_{\min}$ on S.

The NB method of § 3 is easily adapted to solve the problem (CP) by using the projected gradient direction

$$\mathbf{d} = P_{N(A)} \nabla f(x) \tag{18}$$

in the Newton iteration (12), which becomes

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x}) - M}{\|P_{N(A)} \nabla f(\mathbf{x})\|^2} P_{N(A)} \nabla f(\mathbf{x}). \tag{19}$$

This guarantees that all iterates lie in S if the initial $\mathbf{x}^0 \in S$. The NB method with iterations (19) is called the *Projected Gradient NB method*, or *PNB method* for short.

The method needs three parameters:

L^0, a lower bound on $f_{\min}$;
$\epsilon > 0$, a tolerance (used in the stopping rule (10)); and
$0 < \alpha < 1$, a convex weight, used in (11).

Given $\{f, A, \mathbf{b}, L^0, \epsilon, \alpha\}$, the algorithm is described as follows:

Algorithm 1 (The PNB method for (CP) problems)**.**

0	**initialize**	$k = 0$	
	solve	$A\mathbf{x} = \mathbf{b}$	to get a solution $\mathbf{x}^0$
	set	$U^0 = f(\mathbf{x}^0)$	
		L^0	(given lower bound on $f_{\min}$)
1	**if**	$U^k - L^k < \epsilon$	**then solution** $:= \mathbf{x}^k$, **stop**
2	**endif**	**select**	$M^k := \alpha\, U^k + (1-\alpha)\, L^k$
3		**do**	$\mathbf{x}^{k+1} := \mathbf{x}^k - \dfrac{f(\mathbf{x}^k) - M^k}{\|P_{N(A)} \nabla f(\mathbf{x}^k)\|^2} P_{N(A)} \nabla f(\mathbf{x}^k)$
4	**if**	$f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k)$	**set** $U^{k+1} := f(\mathbf{x}^{k+1})$, $L^{k+1} := L^k$
5		**else**	**set** $L^{k+1} := M^k$, $U^{k+1} := U^k$, $\mathbf{x}^{k+1} := \mathbf{x}^k$
	endif	$k := k + 1$	**go to** 1

5. Geometric interpretation

Given a point $\mathbf{x}^0$ and *direction* $\mathbf{d} \in \mathbb{R}^n$, consider the *directional Newton iteration* (8a)

$$\mathbf{x}^1(\mathbf{d}) := \mathbf{x}^0 - \frac{f(\mathbf{x}^0)}{\nabla f(\mathbf{x}^0) \cdot \mathbf{d}}\, \mathbf{d}\ ; \tag{20a}$$

the special case of $\mathbf{d} = \nabla f(\mathbf{x})$,

$$\mathbf{x}^1 := \mathbf{x}^0 - \frac{f(\mathbf{x}^0)}{\|\nabla f(\mathbf{x}^0)\|^2}\, \nabla f(\mathbf{x}^0)\ ; \tag{20b}$$

and given a subspace L, the projected gradient step,

$$\mathbf{x}^1_L := \mathbf{x}^0 - \frac{f(\mathbf{x}^0)}{\|P_L\, \nabla f(\mathbf{x}^0)\|^2}\, P_L\, \nabla f(\mathbf{x}^0)\ . \tag{20c}$$

The geometric interpretation of (20a)–(20c) is given next.

Theorem 1. Let $\mathbf{x}^0$ be a point where $f(\mathbf{x}^0) \neq 0$ and $\nabla f(\mathbf{x}^0) \neq \mathbf{0}$. Let $\mathbf{d}$ be an arbitrary nonzero vector in $\mathbb{R}^n$, and define $\mathbf{x}^1(\mathbf{d})$ by (20a).
(a) The set

$$X(\mathbf{d}) := \{\mathbf{x}^1(\mathbf{d}) :\ \mathbf{d} \in \mathbb{R}^n,\ \mathbf{d} \neq \mathbf{0}\}$$

is a hyperplane in $\mathbb{R}^n$, defined as the intersection of $\mathbb{R}^n$ and the tangent hyperplane (in $\mathbb{R}^{n+1}$) of the graph of f at $(\mathbf{x}^0,\ f(\mathbf{x}^0))$.
(b) The iterate (20b) is the orthogonal projection of $\mathbf{x}^0$ on $X(\mathbf{d})$.
(c) The step lengths of (20b) and (20c) are related by

$$\frac{\|\mathbf{x}^1 - \mathbf{x}^0\|}{\|\mathbf{x}^1_L - \mathbf{x}^0\|} = \frac{\|P_L\, \nabla f(\mathbf{x}^0)\|}{\|\nabla f(\mathbf{x}^0)\|}. \tag{21}$$

Proof. (a) We may assume, without loss of generality, that $\|\mathbf{d}\| = 1$. Since $\nabla f(\mathbf{x}^0) \neq \mathbf{0}$ it follows that the tangent hyperplane of the graph of f at $(\mathbf{x}^0,\ f(\mathbf{x}^0))$ is "not horizontal". Its intersection with $\mathbb{R}^n$ is the hyperplane

$$f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0) \cdot (\mathbf{x} - \mathbf{x}^0)\ =\ 0, \tag{22}$$

which does not contain $\mathbf{x}^0$ since $f(\mathbf{x}^0) \neq 0$. Therefore, any point $\mathbf{x}$ in the above intersection is of the form

$$\mathbf{x} = \mathbf{x}^0 + t\, \mathbf{d}\,, \tag{23}$$

where $\|\,\mathbf{d}\,\| = 1$ and $t \neq 0$. Substituting (23) in (22) we get

$$t := -\frac{f(\mathbf{x}^0)}{\langle \nabla f(\mathbf{x}^0), \mathbf{d}\rangle}\ . \tag{24}$$

(b) The absolute value of the step length (24) is shortest if $\mathbf{d}$ is along the gradient $\nabla f(\mathbf{x}^0)$.
(c) Follows by a comparison of (20b) and (20c). □

6. Constrained location problems

Given

a set of points $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N\} \subset \mathbb{R}^n$;
positive weights $\{w_1, w_2, \dots, w_N\}$; and
an affine set $S = \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\} \subset \mathbb{R}^n$, assumed nonempty;

the *constrained* (*Fermat–Weber*) *location problem* is:

$$\text{find a point} \quad \mathbf{x} \in S \tag{CL}$$

$$\text{minimizing } f(\mathbf{x}) = \sum_{i=1}^{N} w_i \|\mathbf{x} - \mathbf{a}_i\| \,, \tag{25}$$

the sum of the weighted Euclidean distances. The gradient of f

$$\nabla f(\mathbf{x}) = \sum_{i=1}^{N} w_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|} \tag{26}$$

exists for all $\mathbf{x} \notin \mathcal{A}$. A point $\mathbf{x}^* \in S$ is optimal iff $\partial f(\mathbf{x}^*) \subset N(A)^\perp$, which reduces to $P_{N(A)} \nabla f(\mathbf{x}^*) = \mathbf{0}$ if f is differentiable at $\mathbf{x}^*$.

In the unconstrained case ($N(A) = \mathbb{R}^n$) it follows from (26) that $\mathbf{x}^*$ is a convex combination of the points of $\mathcal{A}$

$$\mathbf{x}^* = \sum_{i=1}^{N} \lambda_i(\mathbf{x}^*) \, \mathbf{a}_i, \tag{27}$$

with weights

$$\lambda_i(\mathbf{x}) = \frac{w_i \|\mathbf{x} - \mathbf{a}_i\|^{-1}}{\sum\limits_{j=1}^{N} w_j \|\mathbf{x} - \mathbf{a}_j\|^{-1}}. \tag{28}$$

The *Weiszfeld Method* [12] for solving the unconstrained location problem is an iterative method with updates

$$\mathbf{x}_+ := \sum_{i=1}^{N} \lambda_i(\mathbf{x}) \, \mathbf{a}_i, \tag{29}$$

giving the *next iterate* $\mathbf{x}_+$ as a convex combination, with weights $\lambda_i(\mathbf{x})$ computed by (28) for the *current iterate* $\mathbf{x}$. Note that $\lambda_i(\mathbf{x})$ is undefined if $\mathbf{x} = \mathbf{a}_i$.

The Weiszfeld method [12] is the best–known method for solving the Fermat–Weber location problem, see the history in [11, § 1.3].

There is no obvious way to adapt the Weiszfeld method to solve the affinely constrained location problem (CL). In contrast, the PNB method applies naturally to (CL). The lower bound L^0 (needed in the initial bracket) can be taken as $L^0 = 0$, or better

$$L^0 = \|\mathbf{a}_i - \mathbf{a}_j\| \min\{w_i, w_j\}, \text{ for any two points in } \mathcal{A}.$$

α	0.5	0.61	0.8
No. of iterations in Example 1	35	31	25

Table 1. The number of iterations for the given values of α in Example 1.

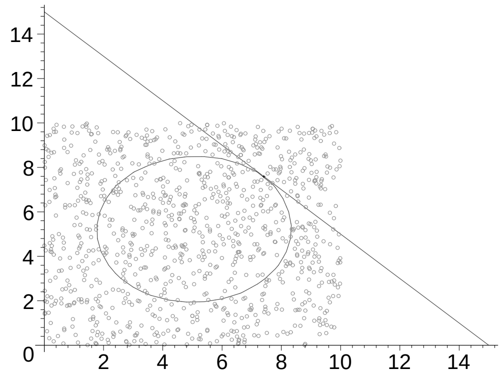

Figure 1. The location problem of Example 1 with 1000 random points in $(0, 10) \times (0, 10)$ and facility constrained to the given line.

The next example shows that the PNB method is well-suited for large-scale location problems.

Example 1. A problem (CL) with 1000 random points in $(0, 10) \times (0, 10)$, all weights $w_i = 1$, and $S = \{\mathbf{x} : x_1 + x_2 = 15\}$, was solved using the PNB method with $\mathbf{x}_0 = (0, 15)$, different values of α, and the stopping rule: $\epsilon = 10^{-6}$. Figure 1 shows the 1000 points, the line S, the level-set corresponding to the optimal value of the distance function (25), and the optimal solution at the intersection of the line S and the level-set.

The number of iterations depends on α. Table 1 shows 3 typical values.

A remarkable result in our numerical experience is that the number of iterations to solve a problem with 1000 points is only slightly higher than the number of iterations for a problem with say 10 points, see e.g. Table 2 below. This may be explained by the fact that the level sets of the function f become more circular as the number of points increases.

7. Numerical results

In the numerical experiments below, all the weights w_i in (25) were taken equal to 1.

Experiment 1. We generated 20 problems (CL) with 100 random points in $(0, 10) \times (0, 10)$, and the line $S = \{\mathbf{x} : x_1 + x_2 = 5\}$ as the feasible set.

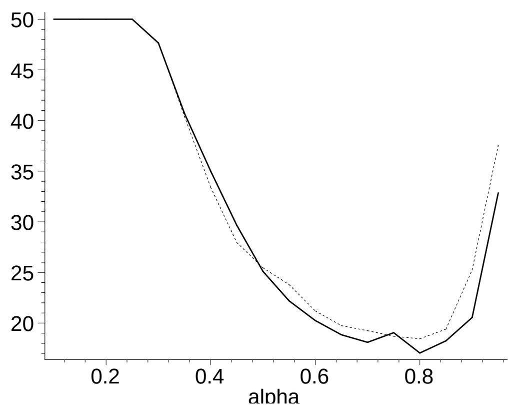

FIGURE 2. Results of Experiment 1: Average numbers of iterations of the PNB method (solid line) and the NB method (dashed line), depending on α.

The corresponding 20 unconstrained location problems (L) have the same points, but no constraints. We solved the constrained problems using the PNB method (Algorithm 1) and the unconstrained problems using the NB method, for different values of parameter α. The purposes of this experiment are:

1. comparison of the performance of the PNB and NB methods; and
2. determination of the optimal α in both methods for such location problems.

Figure 2 shows the average number of iterations for both methods, using the initial point $\mathbf{x}_0 = (10, -5)$, and the stopping rule: $\epsilon = 10^{-3}$ and at most 50 iterations.

Similar results were obtained for different choices of S, $\mathbf{x}_0$, and ϵ.

The optimal α (corresponding to the smallest number of iterations) in both methods is around $\alpha = .8$.

Experiment 2. Using $\alpha = 0.8$ (as determined in Experiment 1), we compare the performance of the PNB method (Algorithm 1) and the (unconstrained) NB method on 20 random location problems with N points, for different values of $N = 10, 50, 100, 250, 500, 750, 1000$.

The random points are generated in $(0, 10) \times (0, 10)$, the feasible set is the line $\{\mathbf{x} : x_1 + x_2 = 5\}$, and the initial iterate is $\mathbf{x}_0 = (10, -5)$.

Table 2 shows the average numbers of iterations in both methods, using the stopping rule: $\epsilon = 10^{-3}$ and at most 50 iterations.

Similar results were obtained for different choices of S, $\mathbf{x}_0$, and ϵ.

N	10	50	100	250	500	750	1000
PNB	15.40	17.70	17.30	19.70	18.80	20.30	20.80
NB	19.10	19.10	18.90	19.50	20.10	19.80	20.80

Table 2. Results of Experiment 2: Average numbers of iterations in PNB and NB methods, for 20 location problems, depending on the number of points N.

8. Comparison of the NB and PNB methods

The projected gradient NB method (abbreviated PNB) applied to a linearly constrained problem (CP), and the NB method for solving its unconstrained counterpart (P), require the same computational effort, notwithstanding the constraints. In addition, the PNB method is more reliable than the NB method, in that it is valid under weaker assumptions. These results are explained in §§ 8.1-8.3. Finally, the PNB method is always valid if the affine set S is a line, see § 8.4.

8.1. **Reliability.** The affine set S (4) consists of the points

$$\mathbf{x} = \mathbf{x}^0 + P_{N(A)}\mathbf{y}\,, \tag{30}$$

where $\mathbf{y} \in \mathbb{R}^n$ is arbitrary. Substituting (30) in a quadratic function (16), we get a quadratic function in $\mathbf{y}$

$$\phi(\mathbf{y}) = \tfrac{1}{2}\mathbf{y}^T\left(P_{N(A)}QP_{N(A)}\right)\mathbf{y} + \text{a linear expression in } \mathbf{y}\,. \tag{31}$$

It follows from the *inclusion principle*, see, e.g. [7, Corollary 4.3.16] that for any positive definite matrix Q and any compatible projection matrix P,

$$\operatorname{cond}(PQP) \le \operatorname{cond}(Q). \tag{32}$$

In particular, $\operatorname{cond}(P_{N(A)}QP_{N(A)}) \le \operatorname{cond}(Q)$, and therefore the sufficient condition (17) is more likely to hold in the linearly constrained case, showing that the projected gradient NB method is more reliable[2] than the NB method.

Example 2. To illustrate (32), we generated 20 random pairs of Q (positive definite $n \times n$ matrix) and P ($n \times n$ projection matrix of rank r), and computed the ratios of condition numbers $\operatorname{cond}(PQP)/\operatorname{cond}(Q)$ for given values of n, r. The average ratios are shown in Table 3, and the maximal (worst case) ratios are given in Table 4. Figure 3 illustrates the average ratios $\operatorname{cond}(PQP)/\operatorname{cond}(Q)$ for $n = 10$ and $r = 2, \ldots, 10$.

[2]I.e., converges under weaker assumptions.

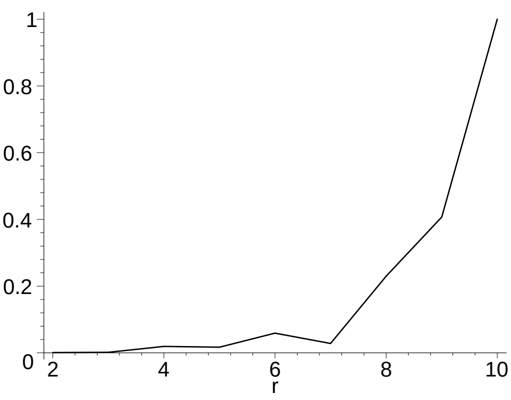

FIGURE 3. The averages of the ratios $\text{cond}(PQP)/\,\text{cond}(Q)$ for $n = 10$ and $r = 2, \ldots, 10$.

$n =$ size of Q	$r=$ rank of the projection matrix P							
	2	3	4	5	6	7	8	9
3	.013							
5	.007	.003	.329					
7	.0020	.0013	.064	.053	.337			
10	.0010	.0017	.0193	.0170	.0592	.0281	.2297	.4069

TABLE 3. The averages of the ratios $\text{cond}(PQP)/\,\text{cond}(Q)$ for 20 random pairs of Q ($n \times n$ positive definite matrix) and P (projection matrix of rank r), for the given values of n, r.

$n =$ size of Q	$r=$ rank of the projection matrix P							
	2	3	4	5	6	7	8	9
3	.0424							
5	.0457	.0196	.8515					
7	.0091	.0070	.1889	.2758	.8640			
10	.0046	.0066	.0854	.0644	.1761	.0816	.7608	.9517

TABLE 4. The maximal ratios of $\text{cond}(PQP)/\,\text{cond}(Q)$ in 20 random pairs of Q (positive definite $n \times n$ matrix) and P (projection matrix of rank r), for the given values of n, r.

8.2. **Convergence.** Part (c) of Theorem 1 relates the step lengths:

$\|\mathbf{x}^1 - \mathbf{x}^0\|$ of the NB method of Section 3, and

$\|\mathbf{x}^1_{N(A)} - \mathbf{x}^0\|$ of the PNB method of Section 4 for linearly constrained convex minimization (CP).

Let $\mathbf{x}^\infty$ be an optimal solution of (CP), where the gradient $\nabla f(\mathbf{x}^\infty)$ is perpendicular to the affine set S, i.e., $P_{N(A)}\nabla f(\mathbf{x}^\infty) = \mathbf{0}$. If $\mathbf{x}^\infty$ does not happen to be a solution of the unconstrained problem (P), then $\nabla f(\mathbf{x}^\infty) \neq \mathbf{0}$. As the iterates $\{\mathbf{x}^k\}$ of the projected gradient NB method converge to $\mathbf{x}^\infty$, the ratios

$$\frac{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|}{\|\mathbf{x}^{k+1}_{N(A)} - \mathbf{x}^k\|} = \frac{\|P_{N(A)}\,\nabla f(\mathbf{x}^k)\|}{\|\nabla f(\mathbf{x}^k)\|}$$

tend to zero, causing the PNB method to employ larger steps than the NB method, and resulting in more frequent occurrences of Case 2 (see (14b)), and faster convergence.

8.3. **Work per iteration.** The above results show that for comparable problems, the (unconstrained) NB method, and the PNB method, require about the same number of iterations, for the same stopping rule.

It is therefore important to measure the effort per iteration in these two methods. This, of course, depends on how we compute the projection. Using (2), an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ of the null space $N(A)$ is required, and the PNB method requires ℓ inner products per iteration, more than the NB method.

If the affine set S is a line in $\mathbb{R}^n$, the work per iteration is about the same in both methods. We show this for the case $n = 2$, i.e. a line in the plane (this is the case for location problems).

Example 3. The projection of any vector (u,v) on the null-space of a line $ax_1 + bx_2 = c$ in the plane is

$$\begin{cases} \dfrac{b^2(u - \frac{a}{b}\,v)}{a^2 + b^2}\left(1, -\frac{a}{b}\right), & \text{if } b \neq 0, \\ (0, v), & \text{if } b = 0,\ a \neq 0. \end{cases}$$

At each iteration we perform one directional Newton iteration for $f(\mathbf{x}) = M$ in the direction:

$$\mathbf{d} = \begin{cases} \nabla f(\mathbf{x}^k), & \text{(NB method)}, \\ P_{N(A)}\nabla f(\mathbf{x}^k), & \text{(PNB method)}. \end{cases}$$

Therefore both methods for location problems have about the same effort per iteration.

8.4. **The case of one-dimensional affine set.** Consider next the special case where the affine set S is one-dimensional, and let S be generated by a (nonzero) vector $\mathbf{v}$, i.e.,

$$S = \{\mathbf{x} = \mathbf{x}^0 + t\mathbf{v} : t \in \mathbb{R}\}. \tag{33}$$

If $P_{N(A)}\nabla f(\mathbf{x}) \neq \mathbf{0}$, then (19) can be written as

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x}) - M}{\langle \mathbf{v}, \nabla f(\mathbf{x})\rangle}\mathbf{v} \tag{34a}$$

$$= \mathbf{x} - \frac{f(\mathbf{x}) - M}{f'(\mathbf{x},\mathbf{v})}\mathbf{v}\,, \tag{34b}$$

where $f'(\mathbf{x},\mathbf{v})$ is the directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{v}$.

Denoting the restriction of f to the line S by

$$\phi(t) := f(\mathbf{x}^0 + t\mathbf{v}), \tag{35}$$

the iteration (34b) corresponds to the ordinary Newton iteration

$$t_+ := t - \frac{\phi(t) - M}{\phi'(t)}. \tag{36}$$

Since the NB method is valid for $n = 1$, we have:

Theorem 2. Let S be a line in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function, whose restriction to S is bounded below, with attained infimum. Then the projected gradient NB method is valid for solving

$$\min\{f(\mathbf{x}) : \mathbf{x} \in S\}. \tag{CP}$$

Acknowledgment

The research of the second author was supported by Natural Sciences and Engineering Research Council of Canada (grant number 261512-04).

References

[1] J. Brimberg, Further notes on convergence of the Weiszfeld algorithm, *YUJOR* **13** (2003), 199–206
[2] J. Brimberg, The Fermat–Weber location problem revisited, *Math. Prog.* **71** (1995), no. 1 (Ser. A), 71–76
[3] J. Brimberg, R. Chen and D. Chen, Accelerating convergence in the Fermat–Weber location problem, *Oper. Res. Lett.* **22**(1998), no. 4-5, 151–157
[4] J. Brimberg and R. F. Love, Local convexity results in a generalized Fermat–Weber problem, *Comput. Math. Appl.* **37** (1999), 87–97
[5] Z. Drezner, *Facility Location: A Survey of Applications and Methods*, Springer, New York, 1995
[6] Z. Drezner and H.W. Hamacher (Editors), *Facility Location: Applications and Theory*, Springer, New York, 2001
[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985
[8] Y. Levin and A. Ben-Israel, Directional Newton methods in n variables, *Math. of Computation* **71**(2002), 237–250
[9] Y. Levin and A. Ben-Israel, The Newton Bracketing method for convex minimization, *Comput. Optimiz. and Appl.* **21**(2002), 213–229
[10] Y. Levin and A. Ben-Israel, A heuristic method for large-scale multi-facility location problems, *Computers & Operations Research* **31**(2004), 257–272
[11] R. Love, J. Morris and G. Wesolowsky, Facilities Location: Models and Methods, *North-Holland*, 1988

[12] E. Weiszfeld, Sur le point par lequel la somme des distances de n points donnés est minimum, *Tohoku Math. J.* **43** (1937), 355–386

Adi Ben-Israel, RUTCOR–Rutgers Center for Operations Research, Rutgers University, 640 Bartholomew Rd, Piscataway, NJ 08854-8003, USA
E-mail address: benisrael@rbsmail.rutgers.edu

Yuri Levin, School of Business, Queen's University, Kingston, Ontario, Canada K7L 3N6
E-mail address: ylevin@business.queensu.ca