ON FINITELY CONVERGENT ITERATIVE METHODS FOR THE CONVEX FEASIBILITY PROBLEM

ALFREDO N. IUSEM AND LEONARDO MOLEDO

Abstract. Iterative algorithms for the Convex Feasibility Problem can be modified so that at iteration k the original convex sets are perturbed with a parameter ε_k which tends to zero as k increases. We establish conditions on such algorithms which guarantee existence of a sequence of perturbation parameters which make them finitely convergent when applied to a convex feasibility problem whose feasible set has non empty interior.

1. Introduction

The Convex Feasibility Problem (CFP) consists of finding a point x in the intersection Q of m closed convex sets $Q^1,\dots,Q^m \subset \mathbb{R}^n$. Without loss of generality we may assume that Q^j is expressed as

$$Q^j = \{x \in \mathbb{R}^n : g_j(x) \leq 0\}$$

where $g_j : \mathbb{R}^n \to \mathbb{R}$ is a convex function. In this formulation CFP consists of finding a solution to the system of inequalities

$$g_j(x) \leq 0 \qquad (1 \leq j \leq m) \tag{1}$$

Systems of linear equations or inequalities are particular cases of CFP. For the linear case, iterative algorithms have been used for a long time, beginning with Jacobi's and Gauss-Seidel's methods. More recently, interest in iterative methods has been aroused by applications, such as Computerized Tomography [7] and

Recebido em 25/06/86. Revisão em 14/05/87.

Transportation Theory [9], where the system of constraints is very large and presents no detectable structure. For these applications, it is convenient to have algorithms which satisfy the following conditions:

1) The system of constraints is not modified at all during the whole procedure.

2) In each step of the algorithm only one constraint is used.

3) In addition to the current iterate, only one or perhaps just a few previously computed n-vectors are used.

Algorithms having these features have been called "Row-action methods" by Censor [1]. Iterative algorithms for the CFP can be found, for instance, in [2], [4], [5], [6], [10]. We are interested here in iterative algorithms for the CFP which are finitely convergent when the set Q has non empty interior. In such case, the functions g_j $(1 \le j \le m)$ can be chosen in such a way that there exists a real number ε such that the system

$$g_j(x) + \varepsilon \lesssim 0 \qquad (1 \le j \le m) \tag{2}$$

is feasible. For instance, if $d(x,A)$ denotes the distance from $x \in \mathbb{R}^n$ to a set $A \subset \mathbb{R}^n$, and ∂A denotes the boundary of A, we may take

$$g_j(x) = \begin{cases} d(x,Q^j) & \text{if } x \notin Q^j \\ -d(x,\partial Q^j) & \text{if } x \in Q^j \end{cases} \qquad (1 \le j \le m)$$

It is easy to show that the g_j's so defined are convex. If ε is known beforehand, any convergent iterative algorithm for the CFP can be transformed into a finitely convergent one, by applying it to (2) instead of (1). But of course, the interesting case is that in which ε is not known. A natural option then is to try a sequence of positive numbers $\{\varepsilon_k\}$ decreasing to zero, so that at iteration k we consider the system:

$$g_j(x) + \varepsilon_k \lesssim 0 \qquad (1 \le j \le m) \tag{3}$$

instead of (1). But then, finite convergence is not guaranteed "a priori". We study in this paper conditions upon the algorithms and the sequence of perturbation parameters $\{\varepsilon_k\}$ so that finite convergence is achieved, while preserving the "row-action" nature of the original algorithms. This condition is important if the perturbed methods are to be used in the previously mentioned applications. There are other finitely convergent algorithms for CPF's satisfying a Slater conditions, like Shor's algorithm [12], which, however, are not "row-action" methods. We prove that if an algorithm generates a sequence x^k that is Fejẽr monotone with respect to Q (i.e. the distance from x^{k+1} to any point in Q is less than or equal to the distance from x^k to that point), the rate of convergence is at least linear (i.e. the ratio between the distances from x^{k+1} and x^k to Q is less than a constant less than one) and this rate satisfies a regularity condition (as a function of Q) then a "diagonal perturbation" of such algorithm is finitely convergent, when Q has non empty interior.

The procedure analyzed in this paper was used in [3], [8] to modify previously known "row-action" methods obtaining finitely convergent ones.

2. Convergence Results

Let $\mathcal{B}$ be the set of functions $g:\mathbb{R}^n \to \mathbb{R}^m$ convex componentwise (ie. $g(x) = (g_1(x),\ldots,g_m(x))$ with $g_i:\mathbb{R}^n \to \mathbb{R}$ convex $(1 \le i \le m)$). For $g \in \mathcal{B}$, let $C(g) = \{x \in \mathbb{R}^n : g_i(x) \lesssim 0$ $(1 \le i \le m)\}$. Let $\mathcal{B}^1 = \{g \in \mathcal{B} : C(g) \neq \emptyset\}$ and $\mathcal{B}^2 = \{g \in \mathcal{B} :$ int $C(g) \neq \emptyset\}$ where "int" denotes interior. Given $g \in \mathcal{B}$, let $g+\varepsilon$ denote the function with components $g_1+\varepsilon,\ldots,g_m+\varepsilon$. $(\varepsilon > 0)$. Observe that if $g \in \mathcal{B}^2$ then there exists $\tilde{\varepsilon} > 0$ such that $g+\varepsilon \in \mathcal{B}^2$ for $\varepsilon \in (0,\tilde{\varepsilon})$. We consider algorithms for the CFP of the form: $x^0 \in \mathbb{R}^n$ arbitrary, $x^{k+1} = F(g,x^k)$, where $F : \mathcal{B}\times\mathbb{R}^n \to \mathbb{R}^n$, and look for conditions on F such that there exists a sequence of positive real numbers $\{\varepsilon_k\}$ so that the algorithm: $x^0 \in \mathbb{R}^n$ arbitrary, $x^{k+1} = F(g+\varepsilon_k, x^k)$, converges finitely for any $g \in B^2$ (i.e. there exists k such that $x^k \in C(g)$].

Let $\|\cdot\|$ denote the Euclidean norm in $\mathbb{R}^n$, and for $S \subset \mathbb{R}^n$, $x \in \mathbb{R}^n$ $\mathrm{dist}(x,S) = \inf_{y \in S} \|x-y\|$.

We make the following assumptions on the algorithmic operator F

(a) If $g \in B^1$, then $\|F(g,x)-y\| \le \|x-y\|$ $\forall\, x \in \mathbb{R}^n$, $y \in C(g)$ (Fejér monotonicity).

(b) $\exists\, \sigma: B^2 \to (0,1)$ such that $\mathrm{dist}(F(g,x), C(g)) \le \sigma(g)\, \mathrm{dist}(x, C(g))$ $\forall\, g \in B^2$, $x \in \mathbb{R}^n$ (linear rate of convergence)

Given $g \in B$ and a sequence $\{\varepsilon_k\}$ of positive real numbers decreasing to zero, consider the sequence

$$x^0 \in \mathbb{R}^n \text{ arbitrary}$$

$$x^{k+1} = F(g+\varepsilon_k, x^k) \tag{4}$$

Let $\lambda_k = \sigma(g+\varepsilon_k)$, $d_k = \mathrm{dist}(x^k, C(g+\varepsilon_k))$.

Lemma 1: If $g \in B^2$, and (a) and (b) hold, then there exists K such that for $k \ge K$, $d_{k+1} \le \lambda_k d_k$.

Proof: Since $g \in B^2$, there exists K so that $C(g+\varepsilon_k) \neq \emptyset$ for $k \ge K$. Since $\{\varepsilon_k\}$ is decreasing $C(g+\varepsilon_{k+1}) \supset C(g+\varepsilon_k)$. So, using condition (b): $d_{k+1} \le \mathrm{dist}(x^{k+1}, C(g+\varepsilon_k)) =$ $= \mathrm{dist}(F(g+\varepsilon_k, x^k)\, C(g+\varepsilon_k)] \le \lambda_k d_k$.

Corollary 1. If $g \in B^2$ and K is as in Lemma 1 then, for $k > K$, $d_k \le \left(\prod_{j=K}^{k-1} \lambda_j\right) d_K$.

Let $<\ ,\ >$ denote the Euclidean inner product in $\mathbb{R}^n$ and $\partial g_i(x)$ the set of subgradients of g_i at x, i.e. the set of vectors $t \in \mathbb{R}^n$ such that $\langle t, y-x\rangle \le g_i(y) - g_i(x)$ $\forall\, y \in \mathbb{R}^n$. Since g_i is convex, $\partial g_i(x)$ is non empty. See, e.g. [11, sections 23, 24 and 25].

Lemma 2: If $g \in B^2$, (a) and (b) hold, $k > K$ and $x^k \notin C(g)$ then there exists $\rho > 0$ (depending only on g and x^0) such that $\varepsilon_k \le \rho d_k$.

Proof: Since $g+\varepsilon_k \in B^1$, take $z^k \in C(g+\varepsilon_k)$ such that $d_k = \|x^k - z^k\|$ So:

$$g_i(z^k) + \varepsilon_k \le 0 \qquad (1 \le i \le m) \tag{5}$$

since $x^k \notin C(g)$, for some j:

$$g_j(x^k) > 0 \tag{6}$$

From (5) and (6)

$$\varepsilon_k \le g_j(x^k) - g_j(z^k) \le \langle t_j^k, x^k - z^k\rangle \tag{7}$$

where $t_j^k \in \partial g_j(x^k)$. Take now a fixed $y \in C(g+\varepsilon_K)$. So $y \in C(g+\varepsilon_k)$ for $k > K$. From condition a) $\|x^{k+1}-y\| = \|F(g+\varepsilon_k, x^k)-y\| \le \|x^k-y\|$. So $\|x^{k+1}-y\| \le \|x^K-y\|$.

Let $U = \{x \in \mathbb{R}^n : \|x-y\| \le \|x^K-y\|\}$ and $\rho = \max\{\|t\| : t \in \partial g_i(x)$ for some i and $x \in U\}$. Since U is compact, ρ exists provided that the effective domain of each g_i is the whole $\mathbb{R}^n$. ρ depends only on g and x^K, i.e., on g and x^0. Since $x^k \in U$ for $k > K$, from (7) get $\varepsilon_k \le \|t_j^k\| \|x^k - z^k\| \le \rho d_k$.

Corollary 2. If $g \in B^2$, $k > K$ and $x^k \notin C(g)$ then

$$\varepsilon_k \le \rho d_K \left(\prod_{j=K}^{k-1} \lambda_j\right) \tag{8}$$

Proof: Follows from Corollary 1 and Lemma 2.

If the algorithm (4) is not finitely convergent for a $g \in B^2$ then (8) holds for any $k > K$. We will impose additional conditions on σ so that a sequence $\{\varepsilon_k\}$ can be chosen violating (8) for large enough k. Hence convergence must indeed be finite. Consider first condition:

(c_1) $\exists\ \hat{\varepsilon}$ (depending on g) such that $\sigma(g+\varepsilon) = \sigma(g)$ for $\varepsilon \in (0,\hat{\varepsilon})$.

Observe that condition (c_1) is implied by condition:

(c_1') $\forall\ g \in B^2$, $\sigma(g+\varepsilon)$ is continuous as a function of ε at $\varepsilon = 0$, because assuming (c_1'), since $\sigma(g) < 1$, given any $\mu \in (\sigma(g),1)$ there exists $\hat{\varepsilon}$ so that $\sigma(g+\varepsilon) \le \mu$ for $\varepsilon \in (0,\hat{\varepsilon})$ and we may redefine $\sigma(g+\varepsilon) = \mu$, for $\varepsilon \in [0,\hat{\varepsilon})$, satisfying b).

Theorem 1. If $g \in B^2$, (a), (b) and (c_1) are satisfied and $\sum_{k=1}^{\infty} \varepsilon_k = \infty$ then algorithm (4) converges finitely to a point in $C(g)$.

Proof. Consider $K' > K$ so that $\varepsilon_k < \hat{\varepsilon}$ for $k \ge K'$. If convergence is not finite, then, from Corollary 2, for all $k \ge K'$.

$$\varepsilon_k \le \rho d_{K'} \left(\prod_{j=K'}^{k-1} \lambda_j\right) = \rho d_{K'}\, \sigma(g)^{k-K'} \Longrightarrow \sum_{k=K'}^{\infty} \varepsilon_k \le \frac{\rho d_{K'}}{1-\sigma(g)}$$

in contradiction with the hypothesis $\sum_{k=1}^{\infty} \varepsilon_k = \infty$.

So, if F satisfies (c_1), we may take $\varepsilon_k = \frac{1}{k}$ and get finite convergence for any $g \in B^2$. We can weaken hypothesis (c_1), but

then we need to impose even slower convergence to 0 for the sequence $\{\varepsilon_k\}$. Assume:

(c_2) $\exists\ \hat{\varepsilon} > 0$ so that $\sigma(g+\varepsilon) \le 1-\varepsilon$ for $\varepsilon \in (0,\hat{\varepsilon})$ and choose $\{\varepsilon_k\}$ so that:

$$\varepsilon_k < 1, \quad \lim_{k\to\infty} \frac{\prod_{j=1}^{k} (1-\varepsilon_j)}{\varepsilon_k} = 0. \tag{9}$$

Condition (9) is implied by:

$$\lim_{k\to\infty} \left(\sum_{j=1}^{k} \varepsilon_j + \ln \varepsilon_k\right) = \infty \tag{10}$$

$\varepsilon_k = \frac{1}{k}$ does not satisfy either (9) or (10) but $\varepsilon_k = \frac{1}{\sqrt{k+1}}$, for instance, satisfies both of them.

Theorem 2. If $g \in B^2$, (a), (b) and (c_2) are satisfied and (9) holds then algorithm (4) converges finitely to a point in $C(g)$.

Proof: As in the proof of Theorem , take a convenient K' and apply Corollary 2 for $k \ge K'$. Use condition (c_2), assuming convergence is not finite, and get

$$\varepsilon_k \le \rho d_{K'} \left(\prod_{j=K'}^{k-1} \lambda_j\right) \le \rho d_{K'} \prod_{j=K'}^{k-1} (1-\varepsilon_j) \Longrightarrow 0 < \frac{1}{\rho d_{K'}} \le \frac{\prod_{j=K'}^{k-1} (1-\varepsilon_j)}{\varepsilon_k}$$

for all $k \ge K'$, in contradiction with (9).

References

[1] Censor, Y., *Row action methods for huge and sparse systems and their applications*. SIAM Rev. 23: 444-464 (1981).

[2] Censor, Y., Lent, A., *Ciclic subgradient projections*. Math. Prog. 24: 233-235 (1982).

[3] De Pierro, A., Iusem, A., *A Finitely Convergent Cyclic Subgradient Projections Method* (to be published in Applied Math. and Optimization).

[4] De Pierro, A., Iusem, A., *A parallel projection method of finding a common point of a family of convex sets*, Pesquisa Operacional 5: 1-20 (1985).

[5] Eremin, I., *The relaxation method for solving systems of inequalities with convex functions on the left hand side*. Soviet Math. Doklady 6: 219-222 (1965).

[6] Gubin, L.G., Polyak, B.T., Raik, E.V., *The method of projections for finding the common point of convex sets*. USSR Comp. Math. Math. Phys. 7: 1-24 (1967).

[7] Herman, G.T., Lent, A., *A family of iterative quadratic optimization algorithms for pairs of inequalities, with application in diagnostic radiology*. Math. Prog. Studies 9: 15-29 (1978).

[8] Iusem, A., Moledo, L., *A Finitely Convergent Method of Simultaneous Subgradient Projections for the Convex Feasibility Problem*. Matemática Aplicada e Computacional 5,2: 169-184 (1986).

[9] Lamond, B., Stewart, N.F., *Bregman's balancing method*. Transp. Res. 15 B: 239-248 (1981).

[10] Motzkin, S., Schoenberg, I.J., *The relaxation method for linear inequalities*. Canadian J. Math. 6: 393-404 (1954).

[11] Rockafellar, R.T., *Convex Analysis*. Princeton University Press (1970).

[12] Shor, N.Z., *Cut-off Method with Space Extension in Convex Programming Problems*. Kibernetika 13, 1: 94,95 (1977).

Instituto de Matemática Pura e Aplicada
Estrada Dona Castorina 110
22.460 Rio de Janeiro-RJ
Brasil

Instituto de Investigaciones Económicas
Universidad de Buenos Aires
Córdoba 2122. 1120 Buenos Aires
Argentina