Yugoslav Journal of Operations Research
1 (1991), Number 1, 71–77

A PROJECTION METHOD FOR LINEARLY CONSTRAINED PROBLEMS WHICH ONLY USES FUNCTION VALUES

Nada I. ĐURANOVIĆ-MILIČIĆ

Department of Mathematics, Faculty of Technology and Metallurgy, University of Belgrade, Karnegijeva 4, 11000 Belgrade, Yugoslavia

Abstract. In this paper we define an iterative algorithm which uses only function values for finding an optimal solution to the problem $\min\{\varphi(x) \mid x \in X\}$, where X is a convex polytope. It is shown that using this algorithm one can reduce the initial problem to a finite number of subproblems of the type $\min\{\varphi(x) \mid x \in C\}$, where C is a linear manifold. It is also shown that each cluster point of the sequence generated by the algorithm presents an optimal point to the considered optimization problem.

Key words and phrases: optimization algorithm, linear manifold, projected gradient, projected Hessian

1. INTRODUCTION

We consider the following minimization problem:

$$\min\{\varphi(x) \mid x \in X\}, \qquad X = \{x \in R^n \mid A^{\mathrm{T}} x \geq b\}, \tag{1}$$

where $A^{\mathrm{T}} = (a_j^{\mathrm{T}})$, $j \in \{1, \dots, m\}$ is an $m \times n$ matrix, b an m-vector.

We use in the sequel the following notation. We define the index set:

$$I(x) = \{j \in \{1, \dots, m\} \mid a_j^{\mathrm{T}} x = b_j\} = \{i_1, \dots, i_q\}.$$

We shall assume the following:

A1. $\varphi : R^n \to R$ is a twice continuously differentiable function with positive definite Hessian matrix; we denote by $g(x)$ and $G(x)$ the gradient and Hessian at x, respectively;

A2. $X \neq \varnothing$ and there exists a point $\bar{x} \in X$ such that the set $L = \{x \in X \mid \varphi(x) \leq \varphi(\bar{x})\}$ is bounded;

A3. the vectors a_j, $j \in I(x) = I$ are linearly independent for each index set I.

The algorithm presented in this paper is based on McCormick's optimization algorithm (see [1]) defined for minimizing a positive definite quadratic form on R^n.

In solving the linearly constrained problem (1) we use McCormick's algorithm for finding an optimal point of the function φ on a linear manifold by means of an active set strategy.

2. THE PROJECTION ALGORITHM FOR LINEARLY CONSTRAINED PROBLEMS WHICH ONLY USES FUNCTION VALUES

Given $\varepsilon_1, \varepsilon_2, \gamma$; $0 < \varepsilon_2 < \varepsilon_1 < 1$; $0 < \gamma < 1$.

STEP 0. Let x_0^1 be a given point in X. Because of assumption A2 it follows that the level set $L = \{x \in X \mid \varphi(x) \le \varphi(x_0^1)\}$ is a bounded set. Set $I = I(x_0^1)$. Let $\operatorname{card} I = q$, and let s_{i-1}, $i = 1, \dots, n-q$ be $n-q$ vectors with one in the i-th position and zeros elsewhere spanning the linear space R^{n-q}. If $q = 0$, set $Z_I = E$, where E is an $n \times n$ identity matrix. Otherwise find a matrix Z_I which is a basis matrix for the null space of A_I^T, where $A_I^T = (a_i^T)$, $i \in I$. Set $k \longleftarrow 1$.

STEP 1. Set $i \longleftarrow 1$.

STEP 2. Compute α_i^k satisfying the condition

$$\varphi(x_{i-1}^k + \alpha_i^k Z_I s_{i-1}) = \min\{\varphi(x_{i-1}^k + \alpha Z_I s_{i-1}) \mid x_{i-1}^k + \alpha Z_I s_{i-1} \in X\}.$$

Denote by $\bar{\alpha}_i^k$ the optimal step along the ray $Z_I s_{i-1}$ and by α_i^{*k} the step to the nearest inactive constraint. Select $\alpha_i^k = \min(\bar{\alpha}_i^k, \alpha_i^{*k})$. Set $x_i^k = x_{i-1}^k + \alpha_i^k Z_I s_{i-1}$. If $I(x_i^k) \supset I(x_{i-1}^k)$, go to step 3; otherwise go to step 4.

STEP 3. Set $x_0^k \longleftarrow x_i^k$, $k \longleftarrow k+1$, $q \longleftarrow \operatorname{card} I(x_i^k)$, $I \longleftarrow I(x_i^k)$, determine Z_I and go to step 1.

STEP 4. If $i = n - q$, go to step 5; otherwise set $i \longleftarrow i + 1$ and goto step 2.

STEP 5. Calculate

$$G_p^{i,i} = \frac{-2(\varphi_i^k - \varphi_{i-1}^k)}{(\alpha_i^k)^2}, \qquad i = 1, \dots, n-q,$$

where $\varphi_i^k = \varphi(x_i^k)$ and $G_p^{i,i}$ denotes a diagonal element of the approximation of the projected Hessian $G_p^k = Z_I^T G^k Z_I$ and go to step 5a.

STEP 5a. Set $i \longleftarrow 1$.

STEP 5b. Set $j \longleftarrow 1$.

STEP 5c. Set $x^k \longleftarrow x_{n-q}^k$, set $j \longleftarrow i + 1$. Let $s_{i,j} = e_i + e_j \in R^{n-q}$, where e_i is a unit coordinate $n - q$ vector. Let $\alpha_{i,j}$ be the scalar that minimizes φ on the subspace corresponding to I along $Z_I s_{i,j}$ starting from x_k. Calculate $x_{i,j} = x^k + \alpha_{i,j} Z_I s_{i,j}$. If $I(x_{i,j}) \supset I(x^k)$, set $x_0^k \longleftarrow x_{i,j}$, $k \longleftarrow k+1$, $q \longleftarrow \operatorname{card} I(x_{i,j})$, $I \longleftarrow I(x_{i,j})$, determine Z_I and go to step 1; otherwise set $x^{k+1} \longleftarrow x_{i,j}$, $\varphi^k = \varphi(x^k)$, $\varphi^{k+1} = \varphi(x^{k+1})$ and calculate

$$G_p^{i,j} = \frac{2(-\varphi^{k+1} + \varphi^k) - (\alpha_{i,j})^2 G_p^{i,i} - (\alpha_{i,j})^2 G_p^{j,j}}{2(\alpha_{i,j})^2}$$

(approximation of an off-diagonal element of the projected Hessian G_p^k) and go to step 5d.

STEP 5d. If $j = n - q$, go to step 5e; otherwise set $j \longleftarrow j + 1$ and go to step 5c.

STEP 5e. If $i < n - q$, set $i \longleftarrow i + 1$ and go to step 5c; otherwise calculate

$$(g_p^k)_i = \sum_{j=i+1}^{n-q} G_p^{i,j} \left(\sum_{l=1}^{n-q} \alpha_l^k s_{l-1} \right)_j$$

(the approximation of the projected gradient at the point x_k.) Calculate $\|g_p^k\|$ and go to step 6.

STEP 6. Set $p \longleftarrow 0$.

STEP 6a. If $\|g_p^k\| \le \varepsilon_1 \gamma^p$, go to step 6b; otherwise go to step 6f.

STEP 6b. Calculate the Lagrange multiplier estimate λ_k. If all $\lambda_i^k \ge 0$, go to step 6c; otherwise find λ_s^k — the most negative Lagrange multiplier estimate, and go to step 6d.

STEP 6c. If $\gamma^p \varepsilon_1 \le \varepsilon_2$, go to step 9 (set $x \longleftarrow x^k$), otherwise go to step 6f.

STEP 6d. If λ_s^k satisfies $\lambda_s^k \le -\gamma^p \varepsilon_1$, go to step 7; otherwise go to step 6e.

STEP 6e. If $\gamma^p \varepsilon_1 \le \varepsilon_2$, go to step 8; otherwise go to step 6f.

STEP 6f. Calculate the Newton vector

$$\bar{s}_k = -{G_p^k}^{-1} g_p^k \in R^{n-q}.$$

Let α^{k+1} be the scalar that minimizes φ on the subspace along $Z_I \bar{s}_k$ starting from x^k. Find $x^{k+1} = x^k + \alpha^{k+1} Z_I \bar{s}_k$. If $I(x^{k+1}) \supset I(x^k)$, set $x_0^k \longleftarrow x^{k+1}$, $k \longleftarrow k + 1$, $q \longleftarrow \operatorname{card} I(x^{k+1})$, $I \longleftarrow I(x^{k+1})$, determine Z_I and go to step 1; otherwise calculate the new approximation of the gradient g_p^{k+1}, set $x^k \longleftarrow x^{k+1}$, $k \longleftarrow k + 1$, $p \longleftarrow p + 1$ and go to step 6a.

STEP 7. Set $I \longleftarrow I \setminus \{i_s\}$. Set $x_0^k \longleftarrow x^k$, $k \longleftarrow k + 1$, $q \longleftarrow q - 1$, determine Z_I and go to step 1.

STEP 8. Calculate $p(\varepsilon_2) = \varepsilon_2 \sum_{s \in I_s} (A_I^+)^{\mathrm{T}} e_s$, where I_s is the set of indices of constraints with near-zero multipliers, i.e. $I_s = \{i \mid -\varepsilon_2 < \lambda_i^k \le 0\}$ and A_I^+ is the pseudoinverse of the matrix A_I. Calculate $x(\varepsilon_2) = x^k + p(\varepsilon_2)$ and the corresponding Lagrange multiplier estimate λ^{ε_2}. Let $(\lambda^{\varepsilon_2} - \lambda^k)_s = \min\{(\lambda^{\varepsilon_2} - \lambda^k)_i, i \in I_s\}$. If $(\lambda^{\varepsilon_2} - \lambda^k)_s < 0$, go to step 7; otherwise set $x \longleftarrow x^k$ and go to step 9.

STEP 9. $x_{\mathrm{opt}} \longleftarrow x$; STOP.

3. THE PROOF OF CONVERGENCE

If φ is a quadratic form, the Algorithm computes the exact projected gradient and projected Hessian. In the general case we must assume that the obtained approximations are "sufficiently accurate". For the sake of simplicity, in the proofs that follow we shall assume $g_p^k = g_p(x^k)$, $G_p^k = G_p(x^k)$.

At first, we shall show in the following lemma that we may delete any constraint for which λ_i^k is negative.

LEMMA. *Let x^k belong to the manifold defined by the index set $I = I(x^k)$ and suppose $g_p(x^k) = 0$. If we delete from the active set of constraints the constraint corresponding to a negative λ_s^k, then the vector $\hat{s}_k = -Z_I(G_p^k)^{-1}g_p^k$ in the subspace corresponding to the new index set $\bar{I} = I \backslash \{i_s\}$ is a feasible descent direction because the conditions*

$$a_s^{\mathrm{T}}(x^k + \alpha \hat{s}_k) > b_s, \quad \alpha > 0, \quad \textit{and} \quad g_k^{\mathrm{T}} \hat{s}_k < 0$$

are satisfied.

Proof. Since $g_p(x^k) = 0$, it follows that $g(x^k) = \sum_{j \in I} \lambda_j^k a_j$. Consequently,

$$g_p(x^k) = Z_{I\backslash\{i_s\}}^{\mathrm{T}} g(x^k) = Z_{I\backslash\{i_s\}}^{\mathrm{T}} \sum_{i \in I} \lambda_i^k a_i = \lambda_s^k Z_{I\backslash\{i_s\}}^{\mathrm{T}} a_s \neq 0 \tag{2}$$

since $\lambda_s^k < 0$ and the vectors a_i, $i \in I$ are, by assumption, linearly independent.

Now we have, using the notation $\bar{I} = I \setminus \{i_s\}$, $g_k = g(x^k)$, $g_k^{\mathrm{T}} \hat{s}_k = -g_k^{\mathrm{T}} Z_{\bar{I}}(G_p^k)^{-1} g_p^k = -(Z_{\bar{I}}^{\mathrm{T}} g_k)^{\mathrm{T}} (G_p^k)^{-1} (Z_{\bar{I}}^{\mathrm{T}} g_k) = -(g_p^k)^{\mathrm{T}} (G_p^k)^{-1} g_p^k < 0$ since G_p^k is positive definite and $g_p^k \neq 0$ by (2). Finally,

$$a_s^{\mathrm{T}} \hat{s}_k = -a_s^{\mathrm{T}} Z_{\bar{I}} (G_p^k)^{-1} Z_{\bar{I}}^{\mathrm{T}} \sum_{i \in I} \lambda_i^k a_i = -\lambda_s^k (Z_{\bar{I}}^{\mathrm{T}} a_s)^{\mathrm{T}} (G_p^k)^{-1} (Z_{\bar{I}}^{\mathrm{T}} a_s) > 0 \tag{3}$$

because $(G_p^k)^{-1}$ is positive definite, $\lambda_s^k < 0$ and $Z_{\bar{I}}^{\mathrm{T}} a_s \neq 0$ by linear independence of vectors a_i, $i \in I$. Since $a_s^{\mathrm{T}} x^k = b_s$, we obtain, using (3)

$$a_s^{\mathrm{T}}(x^k + \alpha \hat{s}_k) = a_s^{\mathrm{T}} x^k + \alpha a_s^{\mathrm{T}} \hat{s}_k > b_s, \quad \alpha > 0,$$

and that is what we had to prove.

THEOREM. *Let the above* (A1–A3) *assumptions be satisfied. Then the presented algorithm reduces the initial problem* (1) *to a finite number of subproblems of the type* $\min\{\varphi(x) \mid x \in C\}$*, where C is a linear manifold and each cluster point of the sequence generated by the algorithm presents an optimal point to the problem* (1).

Proof. At a k-th iteration one of the two following cases can be realized.

CASE 1. x^k is a near optimal point of φ on the subspace corresponding to the index set $I = I(x^k)$; that is, we obtain

$$\|g_p(x^k)\| \leq \varepsilon_1 \gamma^p$$

at step 6a.

After some finite number of steps we shall get

$$\|g_p(x^k)\| \leq \varepsilon_2, \qquad \varepsilon_1\gamma^p \leq \varepsilon_2 \qquad \text{and all} \quad \lambda_i^k \geq 0$$

(steps 6a, 6b, 6c); that is, we find an optimal point $x^k = x_{\text{opt}}$ (step 9) to the problem (1).

If $\|g_p(x^k)\| \leq \varepsilon_1\gamma^p$ and λ_s^k — the most negative Lagrange multiplier estimate satisfies the condition

$$\lambda_s^k \leq -\gamma^p\varepsilon_1$$

(steps 6a through 6d), we delete (at step 7) the constraint corresponding to the index i_s and repeat the procedure of minimization on the new so defined subspace.

Or, after some finite number of steps (steps 6a, 6d, 6e, 8) we shall obtain

$$\|g_p(x^k)\| \leq \varepsilon_2, \qquad \lambda_s^k > -\varepsilon_2, \qquad \gamma^p\varepsilon_1 \leq \varepsilon_2,$$

and we proceede to step 8 to check optimality conditions on an ε_2-active manifold, that is to check the sign of

$$(\lambda^{\varepsilon_2} - \lambda^k)_i, \qquad i \in I_s,$$

where I_s denotes the index set of the near-zero multipliers. If $(\lambda^{\varepsilon_2} - \lambda^k)_s \geq 0$, we have found an optimal point $x^k = x_{\text{opt}}$ (step 9); otherwise we delete a constraint (at step 7) and repeat the procedure at step 1.

During the above described procedure we must stay in the feasible region X. Our step-size is

$$\alpha^k = \min\{\bar{\alpha}^k, \alpha^{*k}\},$$

where $\bar{\alpha}^k$ denotes the optimal step along the given direction and α^{*k} the step to the nearest inactive constraint. If $\alpha^k = \alpha^{*k}$, that is, if we reach a new (inactive) constraint, the process of minimization on the subspace corresponding to the index set $I(x^k)$ is interrupted and we restart the process using the last obtained point on the boundary as a starting point at the next iteration (step 3, step 5c, step 6f).

In this way, either we find a near optimal point of φ on the subspace corresponding to $I(x^k)$, or the process is interrupted if $\alpha^k = \alpha^{*k}$. In both cases the last obtained point is used as a starting point (step 1 or step 7) and the minimization process is repeated.

From Lemma and Algorithm it follows that the sequence of function values $\{\varphi(x^k)\}$ is monotonically decreasing*.

Case 2. x^k is not a near optimal point of φ on the subspace corresponding to the index set $I = I(x^k)$; that is, we obtain

$$\|g_p(x^k)\| > \varepsilon_1\gamma^p$$

at step 6a, 6c. In that case we proceed to look for the minimum of φ on the subspace corresponding to $I(x^k)$ (step 6f) using the projected Newton vector $\bar{s}_k =$

---

* See the second part of the proof.

$-Z_I(G_p^k)^{-1}g_p x^k$ until we either find an optimal point on that subspace: $\|g_p(x^k)\| \leq \varepsilon_2$ at step 6e (or a global optimal point at step 6c) or we attach a new constraint for $\alpha^k = \alpha^{*k}$. In both cases the last obtained point is used as a starting point and the minimization process is repeated at step 1.

In both cases we also obtain decreasing values of φ since we make a positive step $\alpha^k > 0$ along a descent feasible direction $\hat{s}_k = -Z_I(G_p^k)^{-1}g_p(x^k)$ (by assumption of positive definiteness of G_p^k). (If $\alpha^k = \alpha^{*k}$, we have $\alpha^{*k} = \min_j\{(b_j - a_j^{\mathrm{T}}x^k)/(a_j^{\mathrm{T}}s_k), a_j\hat{s}_k < 0, j \notin I\}$. Since $b_j - a_j^{\mathrm{T}}x^k < 0$ and $a_j^{\mathrm{T}}\hat{s}_k < 0$, it follows that $\alpha^{*k} > 0$).

From the above consideration of Case 1 and Case 2 it follows that either we find an optimal point on a subspace (including a global minimum of φ on X) or we add a new constraint to the set of active constraints. Since, by assumption, the vectors a_i, $i \in I(x_k)$ are linearly independent, such an extension of set of active constraints must be finished after a finite number of iterations, which can not be greater then n, where n is the dimension of the problem (1).

Since the sequence $\{\varphi(x^k)\}$ is monotonically decreasing**, all the sets $I(x^k)$ are different.

Hence, this algorithm constructs an optimal point of φ on a subspace (including a global optimal point of φ on X) after a finite number of active constraint set changes.

Since all index sets $I(x^k) \subset \{1, \dots, m\}$, it follows that their number is finite (the number of subsets of a set consisiting of m elements — 2^m).

Therefore the number of problems $\min\{\varphi(x) \mid x \in C\}$, where C is a manifold corresponding to the set of active constraints, must be finite.

Let us prove the second part of the theorem.

By Taylor's theorem we have

$$\varphi(x^{k+1}) = \varphi(x^k) - \frac{\alpha_{\mathrm{opt}}^2}{2}\hat{s}_k^{\mathrm{T}} G_p^k(\eta_p^k)\hat{s}_k,$$
$$\eta_p^k = x^k + \theta_k(x^{k+1} - x^k), \qquad \theta_k \in (0,1).$$

Therefore it follows that $\varphi(x^{k+1}) < \varphi(x^k)$ since G_p^k is by assumption positive definite. Hence $\{x^k\} \subset L$, where L is by assumption a compact set. Consequently there exists a subsequence $\{x^{k_j}\}$ such that $x^{k_j} \to x^* \in L$ as $j \to \infty$. By continiuity of φ we have $\varphi(x^{k_j}) \to \varphi(x*)$ as $j \to \infty$.

From the algorithm we have $g_p(x^k) \to 0$ as $k \to \infty$. By continuity of g_p it follows that $g_p(x^{k_j}) \to g_p(x^*) = 0$ as $j \to \infty$. Since from the algorithm we have $\lambda(x^*) \geq 0$, it follows that x^* is an optimal point to the problem (1).

** See the second part of the proof.

4. CONCLUSIONS

The advantages of the presented algorithm are:

1. this algorithm uses the algorithm presented in [**1**, pp. 176–177] for unconstrained minimization on a subspace; so it requires only evaluation of function values;
2. since the direction vector is the Newton vector $\hat{s}_k = -Z_I(G_p^k)^{-1}g_p(x^k)$ it is natural to expect a good rate of convergence of the sequence of points generated by this algorithm to an optimal point of the problem (1);
3. the algorithm is simpler if φ is a quadratic convex function;
4. there exist suitable methods for evaluating the matrix Z and the Lagrangian vector (see [**2**] and [**3**]);
5. By Definition 4.1 and Definition 4.2 given in [**2**, p. 132] it follows that the presented algorithm is an ideal algorithm and that the sequence generated by this algorithm is ideal, too.

The main disadvantage of the presented algorithm is that the algorithm demands the calculation of an optimal step along the given direction and that the information built up about the approximation of the projected Hessian is lost every time a new constrained is encountered.

REFERENCES

[**1**] Fiacco A. V., and McCormick, G. P., *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.

[**2**] Gill P. E., and Murray W., *Numerical Methods for Constrained Optimization*, Academic Press, 1974.

[**3**] Gill P. E., Murray W., and Wright R. M., *Practical Optimization*, Academic Press, 1984.

[**4**] McCormick G. P., *Nonlinear Programming, Theory, Algorithms and Applications*, Wiley, New York, 1983.