Mathematical Programming Study 3 (1975) 26–34.
North-Holland Publishing Company

# ON IMPROVING RELAXATION METHODS BY MODIFIED GRADIENT TECHNIQUES*

P.M. CAMERINI, L. FRATTA and F. MAFFIOLI

*Istituto di Elettrotecnica ed Elettronica, Politecnico di Milano, Milano, Italy*

Received 14 November 1974
Revised manuscript received 4 August 1975

Relaxation methods have been recently shown to be very effective for some large scale linear problems. The aim of this paper is to show that these procedures can be considerably improved by following a modified gradient step direction.

## 1. Introduction

Given the constants $c_k \in \mathbf{R}$, $\mu_k \in \mathbf{R}^n$ $(k = 1, \ldots, K)$, $n$ and $K$ being two positive integers, and a variable vector $\pi \in \mathbf{R}^n$, find

P. 1
$$\max_{\pi} w(\pi) = \min_{k} \{c_k + \pi \cdot \mu_k\} = c_{k(\pi)} + \pi \cdot \mu_{k(\pi)}.$$

It has been shown [9, 16] that whenever $K$ is large and an efficient method exists to find $w(\pi)$ for each value of $\pi$ (this happens for instance for some combinatorial optimization problems [1, 9, 16]) a method related to gradient techniques provides a very effective attack on P. 1. Specifically one can adopt the following iterative scheme:

$$\begin{cases} \pi^0 = 0, \\ \pi^{m+1} = \pi^m + t_m s^m, \end{cases} \tag{1}$$

$\{t_m\}$ being a sequence of scalars, and $s^m$ the gradient of $w(\pi^m)$ such that

$$s^m = \nabla w(\pi^m) = \mu_{k(\pi^m)}. \tag{2}$$

* Partially supported by the Centro di Telecomunicazioni Spaziali of CNR (Italy). A provisional version of this work was presented at the International Conference on Operation Research (Eger, August 1974).

Since $k(\pi)$ is in general a multivalued function, there exist some points where the gradient is undefined. At such points a set of subgradients can be defined and any of them can be used as $s^m$ [16]. (Note that this iteration scheme is not applied as in conventional gradient procedures, but rather, as in relaxation methods, in order to come closer and closer to the optimal region; the objective function need not be improved at each step.)

Computational experience with some large scale problems (assignment, traveling salesman, multicommodity max flow) [1, 9, 10, 16] has shown that the above relaxation method works better than classical gradient procedures or column generation techniques [7, 15].

The aim of this paper is to prove that the efficiency of the relaxation technique is further improved by selecting the modified gradient direction

$$s^m = \mu^m + \beta_m s^{m-1}, \tag{3}$$

where $\mu^m = \mu_{k(\pi^m)}$ and $\beta_m$ is a suitable scalar ($s^{m-1} = 0$ for $m = 0$).

Note that (3) is in fact equivalent to a weighted sum of all preceding gradient directions, which has been successfully used by Crowder [5] in order to avoid some possible troublesome effects due to the "subgradient's alternating components". In [5], two kinds of modified gradient directions are proposed and, although their general properties have not been investigated, computational experience shows their advantage over use of the simple gradient direction.

In the following section we give a rationale for choosing the iteration scheme (1) and (3), leading to a policy for determining $\beta_m$ and $t_m$ at each step. We also show that the properties given in [9] for (2) can be extended to (3) and that (3) is always at least as good a direction as (2). In Section 3, some computational experiments are discussed which confirm the theoretical results.

## 2. Properties of modified gradient directions

**Lemma 1.** [9]. *Let $\bar{\pi}$ and $\pi^m$ be such that $\bar{w} = w(\bar{\pi}) \geqslant w(\pi^m) = w_m$. Then*

$$(\bar{\pi} - \pi^m) \cdot \mu^m \geqslant \bar{w} - w_m \geqslant 0. \tag{4}$$

*Let, for all m,*

$$0 < t_m \leqslant \frac{\bar{w} - w_m}{\|s^m\|^2} \quad \text{and} \quad \beta_m \geqslant 0. \tag{5}$$

*Then*

$$(\bar{\pi} - \pi^m) \cdot s^m \geqslant (\bar{\pi} - \pi^m) \cdot \mu^m \tag{6}$$

*for all m.*

**Theorem 1.** *Let*

$$\beta_m = \begin{cases} -\gamma_m \dfrac{s^{m-1} \cdot \mu^m}{\|s^{m-1}\|^2} & \text{if } s^{m-1} \cdot \mu^m < 0, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

*with*

$$0 \leqslant \gamma_m \leqslant 2. \tag{8}$$

*Then*

$$\frac{(\bar{\pi} - \pi^m) \cdot s^m}{\|s^m\|} \geqslant \frac{(\bar{\pi} - \pi^m) \cdot \mu^m}{\|\mu^m\|}. \tag{9}$$

**Theorem 2.**

$$\|\bar{\pi} - \pi^m\| > \|\bar{\pi} - \pi^{m+1}\|. \tag{10}$$

**Proofs.** Lemma 2, and Theorems 1 and 2 are proved in the Appendix.

Lemma 1 guarantees that the direction of $\mu^m$ forms an acute angle with the direction leading from $\pi^m$ to the optimum $\bar{\pi}$, while Lemma 2 extends this property to $s^m$. Theorem 1 shows that by a proper choice of $\beta_m$, $s^m$ is always at least as good a direction as $\mu^m$.

Fig. 1 attempts to illustrate such a behaviour in a two-dimensional case.

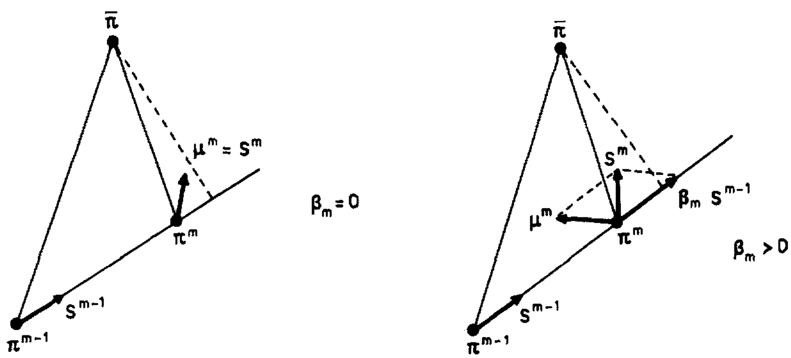


Fig. 1.

Theorem 2 guarantees that a point closer and closer to the optimum is obtained at each iteration, and that the following convergence property holds.

If (5) holds and, for some $\varepsilon > 0$,

$$t_m \geqslant \varepsilon \frac{\bar{w} - w_m}{\|s^m\|^2} \quad \text{for all } m, \tag{11}$$

where $\bar{w} < \max_\pi w(\pi)$, then the sequence $\{\pi^m\}$ either includes a point $\pi^l \in P_{\bar{w}}$ or converges to a point on the boundary of $P_{\bar{w}}$, where $P_{\bar{w}}$ denotes the polyhedron of feasible solutions to $\bar{w} \leqslant c_k + \pi \cdot \mu_k$ for all $k$. In fact it can be shown exactly as in [9] and [14] that $\{\pi^m\}$ is Féjer-monotone relative to $P_{\bar{w}}$ and hence converges. From (1), (5) and (11) it follows that if no $\pi^l$ in $\{\pi^m\}$ exists such that $\pi^l \in P_{\bar{w}}$, then $w(\lim_{m\to\infty} \pi^m) = \bar{w}$ and the limit point is on the boundary of $P_{\bar{w}}$.

Similar results were proved in [9] for the iteration scheme (1) and (2) with a condition on $t_m$ less restrictive than (5), namely

$$0 < t_m < \frac{2(\bar{w} - w_m)}{\|\mu^m\|^2}. \tag{12}$$

However, as it is represented in Fig. 2, the best choice for $t_m$ would be that yielding the nearest position to $\bar{\pi}$ in both directions ($H$ and $H'$). Following Lemmas 1 and 2, an estimate for this step is given either by letting $t_m$ be equal to half the upper limit in (12) or equal to the upper limit in (5).

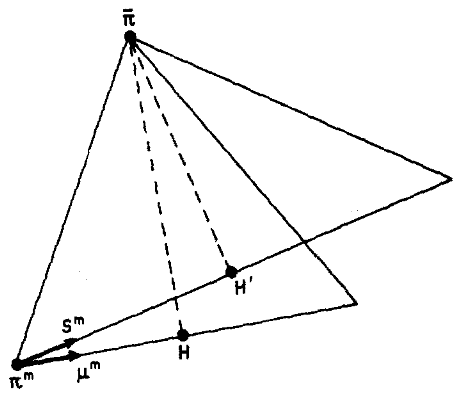


Fig. 2.

As a final remark, note that the policy (7) tends to avoid "zig-zag" behaviour of the sequence $\{\pi^m\}$, since, when the actual gradient direction forms an obtuse angle with the preceding step direction, $\beta_m$ is set greater than zero, thus favouring the "subgradient's persistent components" [5].

## 3. Computational results

For choosing $t_m$ and $s^m$ three policies have been tested.

(a) $s^m \equiv \mu^m$ and $t_m = 1$ [9].

(b) $s^m \equiv \mu^m$ and $t_m = (w^* - w_m)/\|s^m\|^2$, where $w^*$ is a good estimate of $\max_\pi w(\pi)$.

(c) $s^m = \mu^m + \beta_m s^{m-1}$, where

$$\beta_m = \begin{cases} -\gamma \dfrac{s^{m-1} \cdot \mu^m}{\|s^{m-1}\|^2} & \text{if } s^{m-1} \cdot \mu^m < 0, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$t_m = \frac{w^* - w_m}{\|s^m\|^2}.$$

Choosing $\gamma = 1$ would amount to using a direction orthogonal to $s^{m-1}$. Better computational results have been obtained choosing $\gamma = 1.5$. This value of $\gamma$ may be heuristically justified by the following considerations. Let us define an improvement ratio $\eta = \cos\delta_s / \cos\delta_\mu$, where $\delta_s$ and $\delta_\mu$ are respectively the angles which the vector $s^m$ and the vector $\mu^m$ form with the optimum direction $\bar{\pi} - \pi^m$. It is easy to verify that, when $s^{m-1} \cdot \mu^m < 0$,

$$\eta = \frac{1 + \gamma\rho\cos\alpha}{[1 - \gamma(2 - \gamma)\cos^2\alpha]^{1/2}},$$

where $\alpha$ is the (acute) angle between $-s^{m-1}$ and $\mu^m$, and $\rho = \cos\varphi / \cos\psi$, $\varphi$ and $\psi$ being respectively the angles which $s^{m-1}$ and $\mu^m$ form with $\bar{\pi} - \pi^m$. The maximum value of $\eta$ is

$$\bar{\eta} = \left[\frac{1 + 2\rho\cos\alpha + \rho^2}{1 - \cos^2\alpha}\right]^{1/2} \tag{13}$$

which is obtained by the following value of $\gamma$:

$$\bar{\gamma} = \frac{\rho + \cos\alpha}{\cos\alpha\,(1 + \rho\cos\alpha)}. \tag{14}$$

A simple heuristic estimate of $\rho$ is $\rho = 1$ (which amounts to assuming that on the average, when $s^{m-1} \cdot \mu^m < 0$, $s^{m-1}$ and $\mu^m$ are equivalent directions with respect to $\bar{\pi} - \pi^m$). From (13) and (14), two estimates for $\bar{\eta}$ and $\bar{\gamma}$ follow, namely

$$\hat{\eta} = \left[\frac{2}{1 - \cos\alpha}\right]^{1/2}, \qquad \hat{\gamma} = 1 / \cos\alpha.$$

Therefore a policy for choosing $\gamma_m$ is

$$\gamma_m = -\frac{\|s^{m-1}\| \, \|\mu^m\|}{s^{m-1} \cdot \mu^m}, \tag{15}$$

and we may note that if we assume $\frac{1}{4}\pi$ as the mean value of $\alpha$, we obtain $\hat{\gamma} = \sqrt{2}$ and $\hat{\eta} \simeq 2.61$. This value of $\hat{\gamma}$ agrees fairly well with the value of 1.5, suggested by our computational experience.

In order to experiment with the three above policies, the shortest hamiltonian path (SHP) problem (traveling salesman problem) has been solved for the graphs listed in Table 1 by utilizing the heuristically guided algorithm presented in [1, 3]. For any state $v_i$ of the search, a lower bound to the length of a SHP spanning the set $N_i$ of nodes not yet connected is obtained by solving a problem of the form $\max_\pi w_i(\pi)$, $\pi$ being a $|N_i|$- dimensional vector. For all successors $v_j$ of the state $v_i$ the corresponding problems of the form $\max_\pi w_j(\pi)$ have to be solved on the same sub-graph.

Table 1

| | Example | Number of nodes | SHP length |
|---|---|---|---|
| 1 | Croes [4] | 20 | 246 |
| 2 | Random not euclidean (35)[a] | 20 | 25 |
| 3 | Random euclidean (350)[a] | 20 | 1078 |
| 4 | Karg & Thompson [11] | 33 | 10861 |
| 5 | Dantzig et al. [6] | 42 | 699 |
| 6 | Held & Karp [8] | 48 | 11461 |
| 7 | Random euclidean (1000)[a] | 48 | 5394 |
| 8 | Karg & Thompson [11] | 57 | 12955 |
| 9 | Random euclidean (1000)[a] | 67 | 5456 |
| 10 | Join 1 & 6[b] | 67 | 16689 |
| 11 | Join 4 & 10[b] | 100 | 28298 |

[a] Random (not) euclidean ($x$) is a randomly generated graph with (not) euclidean distances between nodes not greater than $x$.

[b] Join $x$ & $y$ is a graph obtained by joining graphs $x$ and $y$ by means of Lin's procedure [12].

As a consequence, these problems lead to almost the same performance of the relaxation method. Let $M$ be the mean number of iterations of (1) and $\Delta_w$ the mean relative increment of $w_j(\pi)$ over the set $\{v_j\}$ of all successors of the state $v_i$. For each $v_i$, we assume as a performance measure of the method the ratio $\sigma = \Delta_w / M$. Some of these values for the three policies and for different $n = |N_i| - 1$ are reported in Fig. 3. The corresponding three

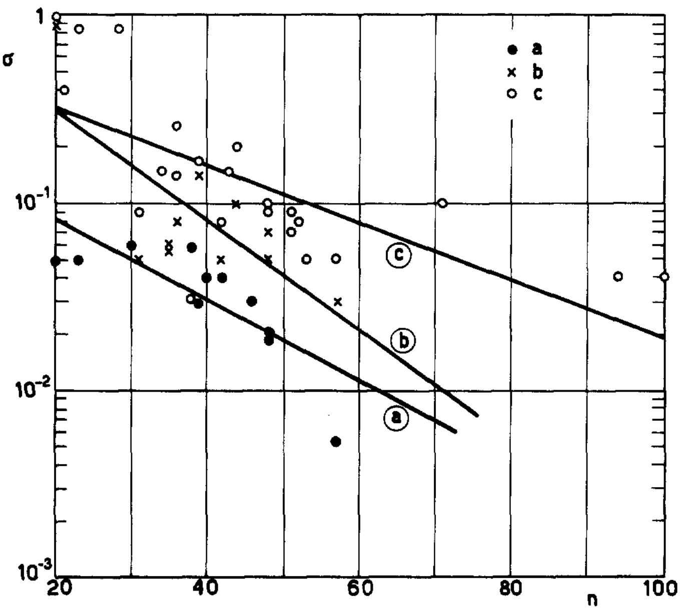


Fig. 3.

minimum mean square regression lines are also represented. Even if the number of samples is not sufficiently large for a satisfactory statistical analysis, one can see that for any $n$ (except $n = 38$) $\sigma$ steadily increases when passing from policy (a), to policy (b), to policy (c), in accordance with the previous theoretical results.

## 4. Conclusions

Relaxation methods, recently revived, and applied to some large scale linear problems have been shown here to be considerably improved by a suitable choice of the direction of search, which turns out to be given by a modified gradient vector. More computational experience will be obtained by applying these methods to other problems such as those mentioned in [2, 5, 13, 16] and in testing the performance of policy (15) for choosing $\gamma_m$.

## Appendix

**Proof of Lemma 2.** The proof is by induction on $m$, since (6) is valid for $m = 0$ with an equal sign. Assume therefore (6) is valid for $m$. Hence from (5) and Lemma 1

$$t_m \|s^m\|^2 \leqslant (\bar{\pi} - \pi^m) \cdot s^m.$$

Since $\beta_{m+1} \geqslant 0$, we may write

$$\beta_{m+1}[(\bar{\pi} - \pi^m) \cdot s^m - t_m \|s^m\|^2] \geqslant 0,$$

i.e., from (1)

$$\beta_{m+1}(\bar{\pi} - \pi^{m+1}) \cdot s^m \geqslant 0. \tag{16}$$

Then Lemma 2 follows from (16), Lemma 1 and (3).

**Proof of Theorem 1.** The proof is trivial when $\beta_m = 0$. When $\beta_m > 0$,

$$\|s^m\|^2 - \|\mu^m\|^2 = \beta_m^2 \|s^{m-1}\|^2 + 2\beta_m(s^{m-1} \cdot \mu^m) \leqslant 0$$

provided (8) holds. Then

$$\|s^m\| \leqslant \|\mu^m\|$$

and from Lemma 2, the theorem follows.

**Proof of Theorem 2.** From (5)

$$t_m \|s^m\|^2 \leqslant \bar{w} - w_m < 2(\bar{w} - w_m).$$

From Lemmas 1 and 2,

$$t_m \|s^m\|^2 < 2(\bar{\pi} - \pi^m) \cdot s^m.$$

This may be written, since $t_m > 0$, as

$$\|\pi - \pi^m\|^2 + t_m^2 \|s^m\|^2 - 2t_m(\pi - \pi^m) \cdot s^m < \|\bar{\pi} - \pi^m\|^2.$$

## References


[1] P.M. Camerini, L. Fratta and F. Maffioli, "A heuristically guided algorithm for the traveling salesman problem", *Journal of the Institution of Computer Science* 4 (1973) 31–35.

[2] P.M. Camerini and F. Maffioli, "Bounds for 3-matroid intersection problems", *Information Processing Letters* 3 (1975) 81–83.

[3] P.M. Camerini, L. Fratta and F. Maffioli, "Traveling salesman problem: heuristically guided search and modified gradient techniques", to appear.

[4] G.A. Croes, "A method for solving traveling salesman problems", *Operations Research* 6 (1958) 791–812.

[5] H. Crowder, "Computational improvements for subgradient optimization", IBM Research Rept. RC 4907 (No. 21841). Thomas J. Watson Research Center (June 1974).

[6] G.B. Dantzig, D.R. Fulkerson and S.M. Johnson, "Solution of a large scale traveling salesman problem", *Operations Research* 2 (1954) 393–410.

[7] J.B. Dantzig, *Linear programming and extensions* (Princeton University Press, Princeton, N.J., 1963) ch. 23.

[8] M. Held and R.M. Karp, "A dynamic programming approach to sequencing problems", *SIAM Journal on Applied Mathematics* 10 (1962) 195–210.

[9] M. Held and R.M. Karp, "The traveling salesman problem and minimum spanning trees: part II", *Mathematical Programming* 1 (1971) 6–25.

[10] M. Held, R.M. Karp and P. Wolfe, "Large scale optimization and the relaxation methods", in: *Proceedings of the 25th Conference of the ACM*, August 1972, pp. 507–509.

[11] L.L. Karg and G.L. Thompson, "A heuristic approach to solving traveling salesman problems", *Management Science* 10 (1964) 225–248.

[12] S. Lin, "Computer solution of the traveling salesman problem", *The Bell System Technical Journal* 44 (1965) 2245–2269.

[13] F. Maffioli, "Shortest spanning hypertrees", in: *Symposium on optimization problems in engineering and economics*, Naples, December 1974.

[14] T. Motzkin and I.J. Schoenberg, "The relaxation method for linear inequalities", *Canadian Journal of Mathematics* 6 (1954) 393–404.

[15] J.F. Shapiro, "A decomposition algorithm for integer programming problems with many columns", in: *Proceedings of the 25th Conference of the ACM*, August 1972, pp. 528–533.

[16] P. Wolfe, M. Held and H. Crowder, "Validation of subgradient optimization", *Mathematical Programming* 6 (1974) 62–88.