



## A characterization of the optimal set of linear programs based on the augmented lagrangian

Mustafa C. Pinar

To cite this article: Mustafa C. Pinar (1999) A characterization of the optimal set of linear programs based on the augmented lagrangian, Journal of Information and Optimization Sciences, 20:2, 299-308, DOI: [10.1080/02522667.1999.10699419](https://doi.org/10.1080/02522667.1999.10699419)

To link to this article: <https://doi.org/10.1080/02522667.1999.10699419>



Published online: 18 Jun 2013.



Submit your article to this journal [↗](#)



Article views: 11

## A characterization of the optimal set of linear programs based on the augmented lagrangian

Mustafa C. Pinar

*Department of Industrial Engineering*

*Bilkent University*

*TR-06533 Ankara*

*Turkey*

---

### ABSTRACT

It is proved that in a certain neighborhood of the optimal set of multipliers, the set of minimizers of the augmented lagrangian function generates a new characterization of the optimal solution set of the linear program.

---

### 1. INTRODUCTION

The purpose of this note is to give a novel characterization of the optimal set of linear programs based on augmented lagrangians. The new characterization allows one to check optimality of the iterates of the proximal point (or, augmented lagrangian) method applied to linear programs in a novel way.

We consider the primal linear programming problem

$$\begin{aligned} [P] \quad & \text{minimize} && c^T x \\ & \text{s.t.} && Ax = b \\ & && x \geq 0 \end{aligned}$$

where  $x \in \mathbf{R}^n$ ,  $A$  is a  $m \times n$  matrix,  $b \in \mathbf{R}^m$  and  $c \in \mathbf{R}^n$ , and the dual problem to [P]:

$$\begin{aligned} [D] \quad & \text{maximize} && -b^T y \\ & \text{s.t.} && A^T y + c \geq 0 \end{aligned}$$

---

*Journal of Information & Optimization Sciences*

Vol. 20 (1999), No. 2, pp. 299-308

© Analytic Publishing Co.

0252-2667/99 \$2.00+.25

where  $y \in \mathbf{R}^m$ . For convenience in exposition, we apply the augmented Lagrangian algorithm (also known as the method of multipliers) to the dual. The method of multipliers applied to [D] consists of the unconstrained minimization phase:

$$y(t+1) = \arg \min_y \left\{ b^T y + \frac{1}{2\mu} \sum_{j=1}^n \max\{0, p_j(t) + \mu(-c_j - a_j^T y)\}^2 \right\} \quad (1)$$

followed by multiplier updates

$$p_j(t+1) = \max\{0, \mu(-c_j - a_j^T y(t+1))\}, \quad \text{for all } j = 1, \dots, n, \quad (2)$$

where  $\mu$  is a positive scalar and  $t$  is the iteration index. It is well-known that the above iteration yields a primal-dual optimal pair to the linear program after a finite number of unconstrained minimization phases. A finite Newton-type method to carry out the unconstrained minimization (1) is given in [11]. Also well-known is the fact that the method of multipliers is “dual” to the proximal minimization algorithm since the dual to the minimization problem (1) is:

$$\begin{aligned} \text{minimize} \quad & c^T p + \frac{1}{2\mu} \|p - p(t)\|_2^2 \\ \text{s.t.} \quad & Ap = b \\ & p \geq 0. \end{aligned}$$

It can be shown that  $p(t+1)$  obtained from the multiplier iteration (2) is the unique optimal solution to the above problem.

There is an extensive literature on the method of multipliers. A detailed treatment can be found in the books by D. P. Bertsekas, and D. P. Bertsekas and J. N. Tsitsiklis [2, 3]. The method of multipliers was originally proposed for nonlinear programs. The origins go back to papers by M. R. Hestenes and M. J. D. Powell [9, 10]. T. R. Rockafellar [6, 7] and D. P. Bertsekas [2] also made very important contributions to the subject. For a bibliography that covers the developments until 1982, see the monograph by Bertsekas [2]. Two recent applications of the method to linear programs are given in the papers by O. Güler [4] and S. J. Wright [5]. It is also possible to devise methods of multipliers based on non-quadratic functions (a.k.a. *D*-functions or Bregman functions) such as the entropy function; see [16, 18, 17, 19] for recent studies.

In the main result of the present paper, using a result by Bertsekas (Proposition 3 of [1]), it is shown (c.f. Theorem 3) that the optimal solution of the linear program can be generated using information from the set of minimizers of the augmented Lagrangian for multiplier vectors contained in a certain neighborhood of the optimal set of multipliers. This result yields a new, easily implementable termination criterion for the method of multipliers applied to linear programs. It also gives a new sufficient condition for the optimal solution to be unique. The present study is related to our previous work on quadratic penalty functions [12] and the joint work on the Huber approximation of  $l_1$  problems [13, 14]. The main difference between the approach of these papers and the present is that in [12, 13, 14] continuation with respect to a scalar parameter is studied whereas in the present paper we view the multiplier vector as the continuation parameter itself.

## 2. STRUCTURE OF THE SET OF MINIMIZERS OF THE AUGMENTED LAGRANGIAN FUNCTION

In this section we examine the properties and structure of the set of minimizers of  $F$ .

For a fixed  $\mu$  and  $p$  we can cast the augmented Lagrangian function in the following quadratic form:

$$F(y, p, \mu) \equiv b^T y + \frac{1}{2\mu} r^T(y, p) W(y, p) r(y, p), \quad (3)$$

where  $r(y, p, \mu) \equiv p + \mu(-A^T y - c)$ , and  $W$  is a diagonal  $m \times m$  matrix with entries:

$$W_{ii}(y, p, \mu) = \begin{cases} 1 & \text{if } r_i(y, p, \mu) > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

We sometimes drop the arguments  $y, p$  and  $\mu$  of  $W, F$  and  $r$  for notational convenience when there is no confusion. In the sequel we refer to  $W$  as a "partition matrix" by analogy to the partitioning of  $\mathbf{R}^m$  by the hyperplanes  $r_j$ .

We use  $a_i$  to denote column  $i$  of  $A$ . We denote by  $X$  and  $Y$  the optimal solution sets of  $[P]$  and  $[D]$ , respectively. The following is well known; see Proposition 4.1(a) of [3].

**THEOREM 1.** *If [D] has a finite optimal value there exists a finite point that minimizes  $F(\cdot, p)$  for any  $p$  and positive scalar  $\mu$ .*

The gradient of the function  $F$  with respect to  $y$  is given by

$$F'(y, p, \mu) = -AW(y, p, \mu)r(y, p, \mu) + b. \quad (5)$$

We denote by  $U(p, \mu)$  the set of minimizers of  $F$  for fixed  $p$  and  $\mu$ . Since  $p(t+1)$  is the unique optimal solution to the dual program to (1), and  $p(t+1)$  can be written as

$$p(t+1) = [p(t) - \mu(A^T y + c)]_+$$

where  $[z]_+$  denotes the vector that is obtained from  $z$  by setting to zero its negative components, we have the following properties of the solution set  $U(p, \mu)$ .

**LEMMA 2.1.** *For  $y \in U(p, \mu)$ , if  $r_i(y, p, \mu) > 0$  then  $r_i(z, p, \mu)$  is constant for  $z \in U(p, \mu)$ . Furthermore,  $W(z, p, \mu)$  is constant for all  $z \in U(p, \mu)$ .*

Following the lemma we let  $W(U(p, \mu)) = W(y, p, \mu)$ ,  $y \in U(p, \mu)$  as the partition matrix corresponding to the solution set. Now, we can use the previous result to characterize the solution set  $U(p, \mu)$ .

**COROLLARY 2.1.**  *$U(p, \mu)$  is a convex set which is contained in the set  $C_w(p, \mu) \equiv \text{cl}\{y \mid W(y, p, \mu) = W\}$  where  $W = W(U(p, \mu))$ .*

**PROOF.** This follows from the linearity of the problem and the previous lemma.  $\square$

Now, define the set of indices  $\sigma_w = \{i \in [1, \dots, m] \mid W_{ii} = 0\}$  and the set  $\mathcal{D}_w(p, \mu) = \{y \in \mathbf{R}^m \mid (p + \mu(-A^T y - c))_i \leq 0 \quad \forall i \in \sigma_w\}$ .

**COROLLARY 2.2.** *Let  $y \in U(p, \mu)$ , and  $W = W(U(p, \mu))$ . Let  $\mathcal{N}_w$  be the orthogonal complement of  $\mathcal{V}_w = \text{span}\{a_i \mid W_{ii} = 1\}$ . Then*

$$U(p, \mu) = (y + \mathcal{N}_w) \cap \mathcal{D}_w(p, \mu).$$

**PROOF.** It follows from (5) that  $F'(y+u, p) = 0$  if  $u \in \mathcal{N}_w$  and  $y+u \in \mathcal{D}_w(p, \mu)$ . Thus

$$U(p, \mu) \supseteq (y + \mathcal{N}_w) \cap \mathcal{D}_w(p, \mu).$$

If  $z \in U(p, \mu)$  then  $r_i(y, p, \mu) = r_i(z, p, \mu)$  for  $W_{ii} = 1$ , and hence  $z - y \in \mathcal{N}_w$ . Therefore, Corollary 2.1 implies

$$U(p, \mu) \subseteq (y + \mathcal{N}_w) \cap \mathcal{D}_w(p, \mu)$$

which proves the result.  $\square$

An important consequence of the previous characterization of  $U(p, \mu)$  is a sufficient condition for the uniqueness of  $y \in U(p, \mu)$ .

**COROLLARY 2.3.** *Assume  $A$  has full rank. Let  $W = W(U(p, \mu))$ . Then,  $y \in U(p, \mu)$  is unique if  $\text{span}\{\alpha_i | W_{ii} = 1\} = \mathbf{R}^m$ .*

This condition is not necessary for uniqueness of solution as the following example demonstrates:

*Example 1.* Consider a linear program of the form  $[D]$  with the following data:

$$A = \begin{pmatrix} 1 & 3 & 5 & 6 & 7 & 1 \\ 0 & 4 & 8 & 1 & 2 & -1 \end{pmatrix},$$

$b = (23/9, 0)^T$ , and  $c = (0, 3, 4, 14, 15, 4)^T$ . For  $\mu = 1$  and  $p = 0$ , the unique minimizer of  $F$  occurs at  $y = (-23/9, 13/9)^T$  with  $r(y, p) = (23/9, -10/9, -25/9, -1/9, 0, 0)^T$  where  $\text{span}\{\alpha_i | W_{ii} = 1\} = \mathbf{R}^1$ .

### 3. NEW CHARACTERIZATION OF OPTIMAL SOLUTIONS

Now, assume  $y \in U(p, \mu)$ , and  $W = W(U(p, \mu))$ . Then  $y$  satisfies the following identity:

$$b - AWr(y, p, \mu) = 0. \quad (6)$$

Now, consider the multiplier iteration (2). This iteration can be recast as follows:

$$p(t+1) = Wr(y(t+1), p(t)). \quad (7)$$

**LEMMA 3.1.** *Let  $p \in \mathbf{R}^n$  with  $\mu > 0$  with  $W = W(U(p, \mu))$  and  $y \in U(p, \mu)$ . Let  $p' = Wr(y, p, \mu)$  and  $\mu' > 0$ . Suppose  $W = W(U(p, \mu)) = W(U(p', \mu'))$ . Then*

$$U(p', \mu') = \mathcal{S}_w \cap \mathcal{D}_w(p', \mu') \quad (8)$$

where  $\mathcal{S}_w$  is the set of solutions to

$$AWA^T y' = -AWc. \quad (9)$$

PROOF. Suppose that  $W(U(p, \mu)) = W(U(p', \mu'))$ . Let  $y \in U(p, \mu)$  and  $y' \in U(p', \mu')$ . This implies that

$$b - AW(p + \mu(-c - A^T y)) = 0, \quad (10)$$

and

$$b - AW(p' + \mu'(-c - A^T y')) = 0. \quad (11)$$

After straightforward algebraic manipulation it is easy to see that  $y'$  satisfies the following linear system of equations:

$$AWA^T y' = -AWc.$$

Conversely, let  $y'$  be a solution to the system (9). If  $W(y', p', \mu') = W(y, p, \mu)$ , then we have the following:

$$\begin{aligned} b - AW(p' + \mu'(-c - A^T y')) & \\ &= b - AWp' + \mu'AWc - \mu'AWc \\ &= b - AWp + \mu AWc + \mu AWA^T y' \\ &= b - AW(p + \mu(-c - A^T y)) \\ &= 0. \end{aligned}$$

Since  $WA^T y'$  is constant regardless of the choice of  $y'$  that solves (9),  $W(A^T y' + c)$  is constant. This completes the proof.  $\square$

*Remark 1.* Lemma 3.1 has an immediate consequence of practical value. If for a given  $p$  and  $\mu$ , a minimizer  $y$  of  $F$  is at hand, then the new multiplier vector  $p'$  is computed by the iteration (2), and by computing a solution  $y'$  to (9), one can check whether the projected point  $y'$  satisfies  $W(y', p', \mu') = W(y, p, \mu)$ . As a result of Lemma 3.1,  $y'$  is a minimizer of  $F(\cdot, p', \mu')$  if  $W(y', p', \mu') = W(y, p, \mu)$ .

Now, define the dual functional  $g(x)$ :

$$g(x) = \begin{cases} c^T x & \text{if } Ax = b, \text{ and } x \geq 0 \\ +\infty & \text{otherwise.} \end{cases} \quad (12)$$

Let  $\gamma > 0$  be a scalar such that

$$g(w) \geq \min_x g(x) + \gamma d(w, X), \quad \forall w \in \mathbf{R}^n, \quad (13)$$

where  $d(w, X) = \min_{x \in X} \|w - x\|$  and  $\|\cdot\|$  denotes Euclidean norm. The existence of such a scalar  $\gamma$  is guaranteed by Lemma 1 of [1]. Now, we are in a position to quote the following result from [1] (see Proposition 3 of [1]).

**THEOREM 2.** *Let  $p$  be any vector in  $\mathbf{R}^n$  and let  $y \in U(p, \mu)$  with  $W = W(U(p, \mu))$ . Let  $\gamma > 0$  be such that (13) holds. If*

$$d(p, X) \leq \gamma \mu,$$

*then the vector  $p'$  obtained from the multiplier iteration*

$$p' = Wr(y, p, \mu)$$

*belongs to  $X$  and is in fact the orthogonal projection of  $p$  on  $X$ .*

A more general form of this result is given in Proposition 4.1(d) pp. 233-242 of [3]. The result states that when  $p$  is sufficiently near the optimal set  $X$  of  $[P]$  ( $\mu$  is sufficiently large) one multiplier iteration suffices to produce an optimal multiplier vector. Similar results were established by Ferris in [15]. Ferris terms inequality (13) the *sharp minimum* property. Now, the main result can be given. This states that the optimal solution set  $Y$  of  $[D]$  can be described entirely using information from the set of minimizers of  $F$  when  $p$  is sufficiently near the optimal set  $X$ . In particular, the optimal set  $Y$  is expressed as the portion of the solution set of a linear system restricted to a particular polyhedral subset of the  $m$ -dimensional Euclidean space.

**THEOREM 3.** *Let  $p$  be any vector in  $\mathbf{R}^n$  and let  $y \in U(p, \mu)$  with  $W = W(U(p, \mu))$ . Let  $\gamma > 0$  such that (13) holds. Suppose that*

$$d(p, X) \leq \gamma \mu.$$

*If the vector  $p'$  is obtained from the multiplier iteration*

$$p' = Wr(y, p)$$

*and  $\mu'$  is any positive scalar then*

1.  $W(U(p', \mu')) = W, \quad \text{and} \quad Y \equiv U(p', \mu'),$
2.  $Y = S_w \cap D_w(p', \mu').$



PROOF. Let  $y' \in Y$ . Since  $p' \in X$  from the previous theorem, and using the complementary slackness theorem of linear programming we have:

$$(A^T y' + c)^T p' = 0, \quad \text{and} \quad p' \geq 0. \quad (14)$$

Let  $r = p + \mu(-c - A^T y)$ . If  $r_i > 0$  then  $W_{ii} = 1$ . This implies that  $p'_i > 0$  and  $(A^T y' + c)_i = 0$  (by complementarity). Now, let  $W_{ii}(y', p') = 1$  since  $p'_i + \mu'(-A^T y' - c)_i > 0$ . If  $r_i \leq 0$ , then  $W_{ii} = 0$  and  $p'_i = 0$ . By complementarity and feasibility, we have  $(A^T y' + c)_i \geq 0$  which implies that  $p'_i + \mu'(-A^T y' - c)_i \leq 0$ . Hence,  $W_{ii}(y', p', \mu') = 0$  for any  $\mu' > 0$ . We have thus constructed a point  $y'$  with  $W(y', p', \mu') = W$ . Now, we have

$$\begin{aligned} b - AW(p' + \mu(-c - A^T y')) \\ = b - AWr - \mu AW(-c - A^T y'). \end{aligned}$$

But, the term in the right-hand side is zero since  $b - AWr = 0$  ( $y \in U(p, \mu)$  and  $WW = W$ ), and  $W(-c - A^T y') = 0$  by construction. Hence, we have shown that  $Y \subseteq U(p', \mu')$  and  $W(U(p', \mu')) = W$  using Lemma 2.1.

To prove  $U(p') \subseteq Y$ , let  $y' \in Y$  and  $y \in U(p)$ . Then, by complementarity  $W(A^T y' + c) = 0$ . Furthermore, since  $W = W(y, p, \mu) = W(y', p', \mu')$  and  $p' = Wr(y, p)$  we have

$$AWA^T y' = -AWc. \quad (15)$$

But,  $WA^T y' = WA^T y$  for any  $y'$  that solves (15). Now, pick any  $y^* \in U(p')$ . Since  $WA^T y^* = WA^T y$  we have  $W(A^T y^* + c) = 0$ . For the indices  $i$  such that  $W_{ii} = 0$ , we clearly have  $(A^T y^* + c)_i \geq 0$ . Hence,  $y^*$  is feasible in  $[D]$  and complementary to  $p'$ . This implies that  $U(p') \subseteq Y$ . Therefore,  $Y \equiv U(p')$ . This proves 1.

Part 2 now follows directly from Lemma 3.1.  $\square$

*Remark 2.* From the previous theorem we deduce that whenever  $\mu$  is sufficiently large, not only  $p'$  obtained by the multiplier operation is optimal, but also any projected point  $y'$  where  $y'$  solves (9), may be optimal for  $[D]$ . Theorem 3 guarantees that  $y'$  will be partially complementary to  $p'$  (and partially feasible in  $[D]$ ) since  $W(A^T y' + c) = 0$ . Hence it suffices to check the feasibility of  $y'$  and complementarity to  $p'$  to decide optimality. To decide optimality, one usually has to

solve the minimization problem (1) with  $p'$ , obtain a minimizer  $y'$  and check the optimality conditions. In this respect, the above theorem gives a way to short circuit the optimality test by possibly avoiding another round of unconstrained minimization.

*Remark 3.* Notice that in Theorem 3 the set  $\mathcal{D}_w(p', \mu')$  where  $p' \in X$  coincides with the set  $\mathcal{D}_w = \{y \in \mathbf{R}^m \mid (A^T y + c)_i \geq 0 \quad \forall i \in \sigma_w\}$ . This implies that one could reiterate part 2 of the theorem using  $\mathcal{D}_w$ .

*Remark 4.* In part 1 of the theorem we have shown that  $Y \equiv U(p', \mu')$  for any  $\mu' > 0$  using the fact that  $W(U(p, \mu)) = W(U(p', \mu'))$ . The result  $Y \equiv U(p', \mu')$  is well-known; see Theorem 3.5 of [8]. The novelty of our Theorem 3 lies in the relation  $W(U(p, \mu)) = W(U(p', \mu'))$  which leads to the new characterization of  $Y$  based on Lemma 3.1.

The new characterization of  $Y$  yields a new sufficient condition for the optimal solution to  $[D]$  to be unique.

**COROLLARY 3.1.** *Assume  $A$  has full rank. Let  $p$  be any vector in  $\mathbf{R}^n$  and let  $y \in U(p, \mu)$  with  $W = W(U(p, \mu))$ . Let  $\gamma > 0$  be such that (13) holds. If*

$$d(p, X) \leq \gamma \mu$$

*then  $Y$  is a singleton if  $S_w$  is a singleton.*

#### REFERENCES

1. D. P. Bertsekas (1976), Newton's method for linear optimal control problems, in *Proceedings of the Symposium on Large Scale Systems* (Udine, 1976), pp. 353-359.
2. D. P. Bertsekas (1982), *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York.
3. D. P. Bertsekas and J. N. Tsitsiklis (1989), *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, New Jersey.
4. O. Güler (1992), Augmented lagrangian algorithms for linear programming, *J. Optim. Theory and Applics.*, Vol. 75, pp. 445-470.
5. S. J. Wright (1990), Implementing proximal point methods for linear programming, *J. Optim. Theory and Applics.*, Vol. 65, pp. 531-551.
6. T. R. Rockafellar (1976), Augmented lagrangians and applications of proximal point algorithm in convex programming, *Math. O.R.*, Vol. 1, pp. 97-116.
7. T. R. Rockafellar (1973), The multiplier method of Hestenes and Powell applied to convex programming, *J. Optim. Theory and Applics.*, Vol. 12, pp. 555-562.

8. T. R. Rockafellar (1973), A dual approach to solving nonlinear programming problems by unconstrained optimization, *Math. Programming*, Vol. 5, pp. 354-373.
9. M. R. Hestenes (1969), Multiplier and gradient methods, *J. Optim. Theory and Applics.*, Vol. 4, pp. 303-320.
10. M. J. D. Powell (1969), A method for nonlinear constraints in minimization problems, in *Optimization*, Edited by R. Fletcher, Academic Press, New York, pp. 283-298.
11. M. C. Pinar (1996), Minimization of the quadratic augmented lagrangian function in linear programming, Technical Report, Department of Industrial Engineering, Bilkent University, 06533, Bilkent, Ankara, Turkey.
12. M. C. Pinar (1994), Piecewise linear pathways to the optimal solution set in linear programming, Technical Report 94-01, Institute of Mathematical Modelling, Technical University of Denmark, revised August 1995 (to appear in *J. Optim. Theory and Applics.*).
13. K. Madsen, H. B. Nielsen and M. C. Pinar (1994), New characterizations of  $l_1$  solutions to overdetermined systems of linear equations, *O.R. Letters*, Vol. 16, pp. 159-166.
14. K. Madsen, H. B. Nielsen and M. C. Pinar (1996), A new finite continuation algorithm for linear programming, *SIAM J. on Optimization*, Vol. 6, August 1996.
15. M. C. Ferris (1991), Finite termination of the proximal point algorithm, *Math. Programming*, Vol. 50, pp. 359-366.
16. M. Teboulle (1992), Entropic proximal mappings with applications to nonlinear programming, *Math. O.R.*, Vol. 17, pp. 670-690.
17. J. Eckstein (1993), Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Math. O.R.*, Vol. 18, pp. 202-226.
18. Y. Censor and S. A. Zenios (1992), The proximal minimization algorithm with  $D$ -functions, *J. Optim. Theory and Applics.*, Vol. 73, pp. 451-466.
19. P. Tseng and D. P. Bertsekas (1993), On the convergence of the exponential multiplier method for convex programming, *Math. Programming*, Vol. 60, pp. 1-19.

*Received July, 1997*