Dynamics of non-homogeneous systems. Proceedings of ISA RAS, 1999, v. 2, pp. 63–67.

Evtushenko Yu.G., Golikov A.I.

Computing Centre of Russian Academy of Sciences
Moscow, Russia

THE AUGMENTED LAGRANGIAN FUNCTION FOR THE LINEAR PROGRAMMING PROBLEM[1]

Revised version 26 May 2003

1. The augmented Lagrangian function is used to solve the linear programming (LP) problem. We define the interval of the penalty coefficient that provides a means of finding a solution of the original problem via a single minimization of a quadratic function on the positive orthant.

Consider the primal and dual linear programming problems:

$$f_* = \min_{x \in X} c^\top x, \qquad X = \{x \in \mathbb{R}^n_+ : b - Ax = 0_m\}, \tag{1}$$

$$\max_{u \in U} b^\top u, \qquad U = \{u \in \mathbb{R}^m : c - A^\top u \geq 0_n\}. \tag{2}$$

Here and further, A denotes an $m \times n$ matrix with rank m, $m < n$; vectors $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n_+$, $u \in \mathbb{R}^m$, 0_i is the i-dimensional null vector.

Let

$$H(x, u, \varepsilon) = c^\top x + u^\top (b - Ax) + \frac{\|b - Ax\|^2}{2\varepsilon}$$

denote the augmented Lagrangian function, where $\varepsilon > 0$ is the penalty coefficient and $\|\cdot\|$ is the Euclidean norm. Consider the auxiliary minimization problem

$$\min_{x \in \mathbb{R}^n_+} H(x, \tilde{u}, \varepsilon). \tag{3}$$

Here the Lagrange multiplier $\tilde{u}$ is fixed. If $\tilde{u} \equiv 0_m$, then the function $H(x, 0, \varepsilon)$ is the well-known quadratic penalty function. Therefore, the results obtained bellow are applicable to the traditional exterior method based on the quadratic penalty function.

We assume that the primal LP problem (1) may have a unique solution x_* which may be degenerate. Let $x^L_* > 0_\ell$ denote a set of positive components of vector x_*. For a nondegenerate solution x_* we have $\ell = m$. We denote by I^L_* the set of these positive components of vector x_*. If x_* is the unique degenerate solution, then the dual LP problem (2) has multiple solutions. From the solution set U_* of the dual LP problem (2) we can choose a vector u_* which is the nearest to some fixed vector $\tilde{u}$. Therefore, the vector u_* is the unique solution of the following quadratic problem:

$$\min \left\{ \frac{\|u - \tilde{u}\|^2}{2} : c - A^\top u \geq 0_n, \quad b^\top u \geq f_* \right\}. \tag{4}$$

[1]Research was supported by Grants N° 99-01-01186 and N° 96-15-96124 from the Russian Foundation for Basic Research.

The pair $[x_*, u_*]$ satisfies the Kuhn–Tucker conditions for problem (1) and these conditions can be written as follows:

$$\begin{array}{lcll} v_*^L & = & c^L - B_L^\top u_* = 0_\ell, & x_*^L > 0_\ell; \\ v_*^S & = & c^S - B_S^\top u_* = 0_s, & x_*^S > 0_s; \\ v_*^N & = & c^N - N^\top u_* > 0_r, & x_*^N = 0_r; \\ B_L x_*^L & = & b. & \end{array} \tag{5}$$

Here the matrix $A = [B \vdots N]$ corresponds to partitioning the vector $v_* = c - A^\top u_*$ into zero $\left[[v_*^L]^\top, [v_*^S]^\top\right] = [v_*^B]^\top = 0_k$ and positive $v_*^N > 0_r$ components, where $k = \ell + s \le m$, $r = n - k$. Using this partition, the vector x_* can be represented as $[x_*]^\top = \left[x_*^B]^\top, [x_*^N]^\top\right]$. The matrix B corresponds to partitioning the vector x_*^B into positive $x_*^L > 0_\ell$ and zero $x_*^S = 0_s$ components, i.e., $B = [B_L \vdots B_S]$. Since the solution x_* is unique, the matrix B consists of $k \le m$ linearly independent columns, i.e., its rank is k.

Let us introduce a vector $\eta \in \mathbb{R}^k$ as $\eta = (B^\top B)^{-1}(c^B - B^\top \tilde{u})$. Notice that if $k = m$, then $\eta = B^{-1}(u_* - \tilde{u})$.

Lemma 1. *The solution of problem* (4) *is given by the vector*

$$u_* = \tilde{u} + B\eta = \tilde{u} + B_L \eta^L + B_S \eta^S,$$

where $\eta^S \le 0_s$.

Proof. Since the vector u_* is a unique solution of problem (4), there exist Lagrange multipliers $\bar{x} \ge 0_n$, $\bar{t} \ge 0$, such that the triple $[u_*, \bar{x}, \bar{t}]$ satisfies the Kuhn–Tucker conditions for problem (4):

$$\begin{array}{c} D(\bar{x})v_* = 0_n, \qquad v_* = c - A^\top u_* \ge 0_n, \qquad \bar{x} \ge 0_n, \\ u_* - \tilde{u} - \bar{t}b + A\bar{x} = 0_m, \\ b^\top u_* = f_*, \qquad \bar{t} \ge 0. \end{array} \tag{6}$$

Since the pair $[x_*, u_*]$ satisfies the Kuhn–Tucker condition (5), it is possible to rewrite (6) as

$$\begin{array}{lcll} v_*^B & = & c^B - B^\top u_* = 0_k, & \bar{x}^B \ge 0_k, \\ v_*^N & = & c^N - N^\top u_* > 0_r, & \bar{x}^N = 0_r, \end{array} \tag{7}$$

$$\begin{array}{c} u_* - \tilde{u} - \bar{t}b + B\bar{x}^B = 0_m, \\ b^\top u_* = f_*, \qquad \bar{t} \ge 0. \end{array} \tag{8}$$

Solving the system of equations (7) and (8), we have

$$\begin{array}{c} \bar{x}^B = \bar{t}(B^\top B)^{-1}B^\top b - \eta \ge 0_k, \\ u_* = \tilde{u} + \bar{t}Mb + B\eta. \end{array} \tag{9}$$

Here we denote the orthogonal projector by $M = I - B(B^\top B)^{-1}B^\top$ and, since the vector b belongs to the column space of matrix B^L, $Mb = 0_n$ holds. The matrix B consists of linearly independent columns, hence there exists only one vector $\eta = \begin{bmatrix} \eta^L \\ \eta^S \end{bmatrix}$ such that $u_* - \tilde{u} = B\eta = $ $= B_L\eta^L + B_S\eta^S$. We show that $\eta^S \le 0_s$. From the condition $Bx_*^B = b$ we obtain

$$x_*^B = (B^\top B)^{-1}B^\top b = \begin{bmatrix} x_*^L \\ x_*^S \end{bmatrix} \begin{array}{l} > 0_\ell \\ = 0_s \end{array}.$$

From (9) and the last formula, we have

$$\bar{x}^B = \begin{bmatrix} \bar{x}^L \\ \bar{x}^S \end{bmatrix} = \begin{bmatrix} \bar{t}x_*^L - \eta^L \\ -\eta^S \end{bmatrix} \geq 0_k. \tag{10}$$

This concludes the proof of the lemma. $\square$

Further, we introduce the following penalty coefficient:

$$\varepsilon_* = \begin{cases} \min\limits_{i\in I_*^L,\,(\eta^L)^i>0} \dfrac{(x_*^L)^i}{(\eta^L)^i}; \\ +\infty, \qquad \text{if } (\eta^L)^i \leq 0 \text{ for all } i \in J_*^L. \end{cases} \tag{11}$$

Theorem 1. *Let x_* be a unique (possibly degenerate) solution of the LP problem* (1)*. Then for each $0 < \varepsilon \leq \varepsilon_*$, where ε_* is defined by* (11)*, we have:*

1) *the problem* (3) *has a unique solution $x(\varepsilon)$ with components: $x^L(\varepsilon) = x_*^L - \varepsilon\eta^L \geq 0_\ell$, $x^S(\varepsilon) = -\varepsilon\eta^S \geq 0_s$ and $x^N(\varepsilon) = 0_r$;*

2) *the solution of the dual LP problem* (2) *is given by*

$$u_* = \bar{u} + \frac{[b - Ax(\varepsilon)]}{\varepsilon}. \tag{12}$$

Proof. The necessary and sufficient conditions for a minimum of problem (3) can be formulated as follows:

$$\varepsilon(c^L - B_L^\top\tilde{u}) - B_L^\top(b - B_Lx^L(\varepsilon) - B_Sx^S(\varepsilon)) = 0_\ell, \qquad x^L(\varepsilon) \geq 0_\ell, \tag{13}$$
$$\varepsilon(c^S - B_S^\top\tilde{u}) - B_S^\top(b - B_Lx^L(\varepsilon) - B_Sx^S(\varepsilon)) = 0_s, \qquad x^S(\varepsilon) \geq 0_s, \tag{14}$$
$$\varepsilon(c^N - N^\top\tilde{u}) - N^\top(b - B_Lx^L(\varepsilon) - B_Sx^S(\varepsilon)) = 0_r, \qquad x^N(\varepsilon) \geq 0_r. \tag{15}$$

Since the $m \times k$ matrix $B = [B_L \vdots B_S]$ has rank k, it is easy to show that the system of equations (13), (14) has a unique solution

$$x^B(\varepsilon) = \begin{bmatrix} x^L(\varepsilon) \\ x^S(\varepsilon) \end{bmatrix} = \begin{bmatrix} x_*^L - \varepsilon\eta^L \\ -\varepsilon\eta^S \end{bmatrix} \geq 0_k. \tag{16}$$

Substituting $x^B(\varepsilon)$ into (15) yields $\varepsilon(c^N - N^\top\tilde{u}) - N^\top(b - B_Lx_*^L + \varepsilon B_L\eta^L + \varepsilon B_S\eta^S) = \varepsilon(c^N - N^\top(\tilde{u} + B\eta)) = \varepsilon(c^N - N^\top u_*) = \varepsilon v_*^N > 0_r$. The last inequality holds for any positive ε.

Inequality (16) holds due to the lemma's assertion ($\eta^S \leq 0_s$) and the existence of the interval $(0, \varepsilon]$ mentioned in the condition of the theorem. Hence, the system (13) – (15) is consistent and statement **1)** is proved.

Proof of assertion **2)** follows from substituting the solution $x(\varepsilon)$ into the expression (12). $\square$

Remark 1. *Note the important special case where the problem* (1) *has a nondegenerate solution x_*. In this case the indices of nonzero components of vectors $x(\varepsilon)$ and x_* coincide for each $0 < \varepsilon < \varepsilon_*$. Since $x_*^B = x^B(\varepsilon) + \eta$, we obtain a means of solving the problem* (1) *via a single minimization of $H(x, \tilde{u}, \varepsilon)$ on the positive orthant. Hence, $H(x, \tilde{u}, \varepsilon)$ could be called "the exact auxiliary function" if we slightly extend the notion defined in* [1].

Remark 2. *Formulas* (10) *and* (16) *show that there exists a certain connection between the solution $x(\varepsilon)$ of problem* (3) *and the optimal Lagrange multiplier $\bar{x}$ for problem* (4)*. Expression*

(10) *is valid for* $\bar{t} \geq \dfrac{1}{\varepsilon_*}$. *If we set* $\bar{t} = \dfrac{1}{\varepsilon}$, *then from* (10) *and* (16) *we obtain that for each* $0 < \varepsilon \leq \varepsilon_*$ *the equality* $x^B(\varepsilon) = \varepsilon \bar{x}^B$ *holds.*

$\boxed{\mathbf{2.}}$ It is well known (see, e.g., [2]), that problem (3) is dual to the following problem:

$$\max_{u \in U} \left[b^\top u - \varepsilon \frac{\|u - \tilde{u}\|}{2} \right], \qquad U = \{u \in \mathbb{R}^m : c - A^\top u \geq 0_n\}. \tag{17}$$

We denote by $u(\varepsilon)$ the solution of this problem. Then the solutions $x(\varepsilon)$ and $u(\varepsilon)$ are connected by

$$u(\varepsilon) = \tilde{u} + \frac{b - Ax(\varepsilon)}{\varepsilon}.$$

The following theorem is a direct consequence of Theorem 1 due to the above duality.

Theorem 2. *Let the conditions of theorem* 1 *hold. Then a unique solution of problem* (17) *does not depend on* ε *and coincides with the vector* u_*, *that is a solution of the problem* (2).

Remark 3. *Paper* [3] *was probably the first where the coincidence of the solution of problem* (17) *and the solution of problem* (2) *was investigated (in the case* $\tilde{u} \equiv 0_m$*). Various estimates of value* ε_*, *using the Lagrange multiplier* t, *for problem* (4) *were obtained in* [4, 5, 6] *for the same case. The result concerning the finiteness of the augmented Lagrangian method (problem* (3)*) was first obtained in* [7] *under the assumption of solvability of the LP problem. This question was thoroughly investigated in* [8] *with the help of the sensitivity function.*

References

1. *Evtushenko Yu.G. and Zhadan V.G.* Exact Auxiliary Functions in Optimization Problems // U.S.S.R. Comput. Maths. Math. Phys. 1990. Vol. 30. N° 1. P. 31–42.

2. *Polyak B.T.* Introduction in Optimization. Moscow: Nauka, 1983 (in Russian).

3. *Uzawa H.* Iterative Method for Concave Programming // Studies in Linear and Nonlinear Programming. Arrow K.J., Hurwicz L. and Uzawa H. (eds.). 1958.

4. *Eremin I.I. and Astafiev N.N.* Introduction in Theory of Linear and Convex Programming. Moscow: Nauka, 1976 (in Russian).

5. *Mangasarian O.L. and Meyer R.R.* Nonlinear Perturbation of Linear Programs // SIAM J. Control and Optimization. 1979. Vol. 17. N° 6. P. 745–752.

6. *Mangasarian O.L.* Normal Solutions of Linear Programs // Math. Program. Study. 1984. Vol. 22. P. 206–216.

7. *Poljak B.T. and Tretjakov N.V.* An Iterative Method for Linear Programming and Its Economic Interpretation // Matecon. 1974. Vol. 10. P. 81–100.

8. *Golshtein E.G. and Tretjakov N.V.* The Modified Lagrange Functions. Moscow: Nauka, 1989 (in Russian).