R. Conti & A. Ruberti, editors, Springer-Verlag, 1973.
(series: "Lecture Notes in Computer Sciences")

PENALTY METHODS AND AUGMENTED LAGRANGIANS IN NONLINEAR PROGRAMMING

R. Tyrrell Rockafellar
Dept. of Mathematics, University of Washington
Seattle, Washington 98195 U.S.A.

The usual penalty methods for solving nonlinear programming problems are subject to numerical instabilities, because the derivatives of the penalty functions increase without bound near the solution as computation proceeds. In recent years, the idea has arisen that such instabilities might be circumvented by an approach involving a Lagrangian function containing additional, penalty-like terms. Most of the work in this direction has been for problems with equality constraints. Here some new results of the author for the inequality case are described, along with references to the current literature. The proofs of these results will appear elsewhere.

Equality Constraints

Let $f_o, f_1, \ldots, f_m$ be real-valued functions on a subset X of a linear topological space, and consider the problem

(1) minimize $f_o(x)$ over $\{x \in X | f_i(x) = 0 \text{ for } i=1,\ldots,m\}$.

The augmented Lagrangian for this problem, as first introduced in 1958 by Arrow and Solow [2], is

$$L(x,y,r) = f_o(x) + \sum_{i=1}^{m} [rf_i(x)^2 + y_i f_i(x)], \tag{2}$$

where $r \geq 0$ is a penalty parameter and $y = (y_1, \ldots, y_m) \in R^m$. In fact, this is just the ordinary Lagrangian function for the altered problem in which the objective function f_o is replaced by $f_o + rf_1{}^2 + \ldots + rf_m{}^2$, with which it agrees for all points satisfying the constraints.

The motivation behind the introduction of the quadratic terms is that they may lead to a representation of a local optimal solution in terms of a local unconstrained minimum. If $\bar{x}$ is a local optimal solution to (1) with corresponding Lagrange multipliers $\bar{y}_i$, as furnished by classical theory, the function

$$L_o(x,\bar{y}) = f_o(x) + \sum_{i=1}^{m} \bar{y}_i f_i(x)$$

*This work was supported in part by the Air Force Office of Scientific Research under grant AF-AFOSR-72-2269.

has a stationary point at $\bar{x}$ which is a local minimum relative to the manifold of feasible soltuions. However, this stationary point need not be a local minimum in the unconstrained sense, and L may even have negative second derivatives at $\bar{x}$ in certain directions normal to the feasible manifold. The hope is that by adding the terms $rf_i(x)^2$, the latter possibility can be countered, at least for r large enough. It is not difficult to show this is true if $\bar{x}$ satisfies second-order sufficient conditions for optimality (cf. [1]).

The augmented Lagrangian gives rise to a basic class of algorithms having the following form:

(3) Given (y^k, r^k), minimize $L(x, y^k, r^k)$ (partially ?) in $x \in X$ to get x^k. Then, by some rule, modify (y^k, r^k) to get (y^{k+1}, r^{k+1}).

Typical exterior penalty methods correspond to the case where $y^{k+1} = y^k = 0$ and $r^{k+1} = \alpha r^k$ (α = some factor > 1). In 1968, Hestenes [10] and Powell [19] independently drew attention to potential advantages of the case

$$y^{k+1} = y^k + 2r^k \nabla_y L(x^k, y^k, r^k), \quad r^{k+1} \geq r^k. \tag{4}$$

The same type of algorithm was subsequently proposed also by Haarhoff and Buys [9] and investigated by Buys in his thesis [4]. Some discussion may also be found in the book of Luenberger [13]. Recently Bertsekas [3] has obtained definitive results in the case where an ε-bound on the gradient is used as the stopping criterion for the minimization at each stage. These results confirm that the convergence is essentially superlinear when $r^k \to \infty$. Various numerical experiments involving modifications of the Hestenes-Powell algorithm still in the pattern of (3) have been carried out by Miele and his associates [15], [16], [17], [18]; see also Tripathi and Narendra [26]. Some infinite-dimensional applications have been considered by Rupp [24], [25].

An algorithm of Fletcher [6] (see also [7], [8]) may, in one form, be considered also as a "continuous" version of (3) in which certain functions of x are substituted for y and r in $L(x,y,r)$; one then has a single function to be minimized. The original work of Arrow and Solow [2] also concerned, in effect, a "continuous" version of (3) in which x and y values were modified simultaneously in locating a saddle point of L.

Inequality Constraints.

For the inequality-constrained problem,

(P) minimize $f_o(x)$ over $\{x \in X \mid f_i(x) \leq 0,\ i = 1,\dots,m\}$,

it is not immediately apparent what form the augmented Lagrangian should have, but the natural generalization turns out to be

$$L(x,y,r) = f_o(x) + \sum_{i=1}^{m} \lambda(f_i(x),y_i,r), \tag{5}$$

where

$$\lambda(f_i(x),y_i,r) = \begin{cases} rf_i(x)^2 + y_i f_i(x) & \text{if } f_i(x) \geq -y_i/2r, \\ -y_i^2/4r & \text{if } f_i(x) \leq -y_i/2r. \end{cases} \tag{6}$$

In dealing with (5), the multipliers y_i are <u>not</u> constrained to be nonnegative, in contrast with the ordinary Kuhn-Tucker theory. This Lagrangian was introduced by the author in 1970 [20] and studied in a series of papers [21], [22], [23], the main results of which will be indicated below. It has also been treated by Buys [3] and Arrow, Gould and Howe [1]. Related approaches to the inequality-constrained problem may be found in papers of Wierzbicki [27], [28], [29], Fletcher [7], Kort and Bertsekas [11], Lill [12], and Mangasarian [14].

To relate the augmented Lagrangian to penalty approaches, it should be noted that by taking $y = 0$ one obtains the standard "quadratic" penalty function. Observe also that the classical Lagrangian for problems with inequalities can be viewed as a limiting case:

$$\lim_{r \downarrow 0} L(x,y,r) = L_o(x,y) = \begin{cases} f_o(x) + \sum_{i=1}^{m} y_i f_i(x) & \text{if } y \geq 0, \\ -\infty & \text{if } y \ngeq 0. \end{cases} \tag{7}$$

The following properties of (5)-(6) can be verified [21], [23]: $L(x,y,r)$ is always concave in (y,r), and it is continuously differentiable (once) in x if every f_i is differentiable. Furthermore, it is convex in x if (X and) every f_i is convex; the latter is referred to as the <u>convex case.</u> Higher-order differentiability is not inherited by L from the functions f_i along the "transition surfaces" corresponding to formula (6). However, as will be seen from Theorem 4 below, most of the interest in connection with algorithms and their convergence centers on the local properties of L in a neighborhood of a point $(\bar{x},\bar{y},\bar{r})$ such that $\bar{x}$ is a local optimal solution to (P), $\bar{y}$ is a corresponding multiplier vector in the classical sense of Kuhn and Tucker, and $\bar{r} > 0$. If the multipliers $\bar{y}_i$ satisfy the complementary slackness conditions, as usually has to be assumed in a close analysis of convergence, it is clear that none of the "transition surfaces" will pass through $(\bar{x},\bar{y},\bar{r})$, and hence L will be two or three times continuously differentiable in some neighborhood of $(\bar{x},\bar{y},\bar{r})$, if every f_i has this order of differentiability. (Certain related Lagrangians recently proposed by Mangasarian [14] inherit higher-order differentiability everywhere, but they are not concave in (y,r).)

The class of algorithms (3) described above for the equality case may also be studied in the inequality case. In particular, rule (4) gives an immediate generalization of the Hestenes-Powell algorithm. We have shown in [22] that in the finite-dimensional convex case, this algorithm always converges globally if, say, an optimal solution $\bar{x}$ exists along with a Kuhn-Tucker vector $\bar{y}$. This is true even if the minimization in obtaining x^k is only approximate in a certain sense. The multiplier vectors y^k converge to some particular Kuhn-Tucker vector $\bar{y}$, even though the problem may possess more than one such vector. For convex and nonconvex problems, results on local rates of convergence in the equality case are applicable if the multipliers at the locally optimal solution in question satisfy complementary slackness conditions.

Dual Problem.

The main theoretical properties of the augmented Lagrangian, fundamental to all applications, can be described in terms of a certain dual problem corresponding to the global saddle point problem for L. To shorten the presentation here, we henceforth make the simplifying assumption that X is compact and the functions f_i are continuous. It must be emphasized that this assumption is not required, and that the more general setting is in fact the one treated in [21], [22], [23]. It should also be clear that our focus on inequality constraints involves no real restriction. Mixtures of equations and inequalities can be handled in much the same way.

The dual problem which we associate with (P) in terms of the augmented Lagrangian L is

(D) maximize $g(y,r)$ over all $y \in R^m$ and $r > 0$, where

$$g(y,r) = \min_{x \in X} L(x,y,r) \qquad \text{(finite)}.$$

Note that constraint $y \geq 0$ is not present in this problem. Nor does the condition $r > 0$ represent a true constraint, since, as is easily seen, $g(y,r)$ is nondecreasing as a function of r for every y. Thus the dual problem is one of unconstrained maximization. Further, $g(y,r)$ is concave in (y,r), and in the convex case it is continuously differentiable, regardless of the differentiability of f_i [21].

THEOREM 1[23]. $\min(P) = \sup(D) = \lim g(y^k, r^k)$, *where* $(y^k, r^k)_{k=1}^{\infty}$ *denotes an arbitrary sequence with* y^k *bounded and* $r^k \to \infty$.

THEOREM 2[23]. *Let* $(y^k, r^k)_{k=1}^{\infty}$ *denote any sequence with* $\sup(D) = \lim g(y^k, r^k)$ *and* y^k *bounded (but not necessarily with* $r^k \to \infty$*). Let* x^k *minimize* $L(x, y^k, r^k)$ *over* X *to within* ε^k, *where* $\varepsilon^k \downarrow 0$. *Then all cluster points of the sequence* x^k *are optimal solutions to* (P).

If $y^k \equiv 0$, Theorem 1 asserts the familiar fact in the theory of penalty functions that

$$\min(P) = \lim_{r^k \to \infty} \min_{x \in X} [f_o(x) + r^k \sum_{i=1}^{m} \max^2\{0, f_i(x)\}]. \tag{8}$$

More generally, it suggests a larger class of penalty-like methods in which still $r^k \to \infty$, but y^k is allowed to vary. Perhaps, through a good rule for choosing y^k, such a method could yield improved convergence and thereby reduce some of the numerical instabilities associated with having $r^k \to \infty$. Theorem 2 even holds out the attractive possibility of algorithms in which both y^k and r^k remain bounded. The fundamental question here is whether a bounded maximizing sequence (y^k, r^k) exists at all for (D). In other words, under what circumstances can it be said that the dual problem has an optimal solution $(\bar{y}, \bar{r})$?

It is elementary from Theorem 1 and the definition of the dual that a necessary and sufficient condition for $(\bar{y}, \bar{r})$ to be an optimal solution to (D) and $\bar{x}$ to be a (globally) optimal solution to (P) is that $(\bar{x}, \bar{y}, \bar{r})$ be a (global) saddle point of L. The following theorems on saddle points therefore show that our question about the existence of bounded maximizing sequences (y^k, r^k) has an affirmative answer for "most" problems.

THEOREM 3 [21]. *In the convex case,* $(\bar{x}, \bar{y}, \bar{r})$ *is a saddle point of* L *if and only if* $(\bar{x}, \bar{y})$ *is a saddle point of the classical Lagrangian* L_o *in* (7).

THEOREM 4 [23]. *Suppose that* $\bar{x} \in \text{int } X \subset R^n$, *and that each* f_i *is differentiable of class* C^2 *near* $\bar{x}$.

(a) *If* $(\bar{x}, \bar{y}, \bar{r})$ *is a global saddle point of* L, *then* $(\bar{x}, \bar{y})$ *satisfies the second-order necessary conditions* [5, p.25] *for local optimality in* (P), *and* $\bar{x}$ *is globally optimal.*

(b) If $(\bar{x},\bar{y})$ satisfies the second-order sufficient conditions [5,p.30] for local optimality in (P) and $\bar{x}$ is uniquely globally optimal, then $(\bar{x},\bar{y},\bar{r})$ is a global saddle point of L for all $\bar{r}$ sufficiently large.

Part (b) of Theorem 4 stengthens a local result of Arrow, Gould and Howe [1] involving assumptions of complementary slackness and the superfluous constraint $y \geq 0$. A corresponding local result has also been furnished by Mangasarian [14] for his different family of Lagrangians. It is shown in [23] that the existence of a dual optimal solution $(\bar{y},\bar{r})$ depends precisely on whether (P) has a second-order stability property with respect to the ordinary class of perturbations.

REFERENCES

1. K. J. Arrow, F. J. Gould and S. M. Howe, "A general saddle point result for constrained optimization", Institute of Statistics Mimeo Series No. 774, Univ. of N. Carolina (Chapel Hill), 1971.

2. K. J. Arrow and R. M. Solow, "Gradient methods for constrained maxima, with weakened assumptions", in Studies in Linear and Nonlinear Programming, K. Arrow, L. Hurwicz and H. Uzawa (editors), Stanford Univ. Press, 1958.

3. D. P. Bertsekas, "Combined primal-dual and penalty methods for constrained minimization", SIAM J. Control, to appear.

4. J. D. Buys, "Dual algorithms for constrained optimization", Thesis, Leiden, 1972.

5. A. V. Fiacco and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Optimization Techniques, Wiley, 1968.

6. R. Fletcher, "A class of methods for nonlinear programming with termination and convergence properties", in Integer and Nonlinear Programming, J. Abadie (editor), North-Holland, 1970.

7. R. Fletcher, "A class of methods for non-linear programming III: Rates of convergence", in Numerical Methods for Non-linear Optimization, F. A. Lootsma (editor), Academic Press, 1973.

8. R. Fletcher and S. Lill, "A class of methods for nonlinear programming, II: computational experience", in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter (editors), Academic Press, 1971.

9. P. C. Haarhoff and J. D. Buys, "A new method for the optimization of a nonlinear function subject to nonlinear constraints", Computer J. 13 (1970), 178-184.

10. M. R. Hestenes, "Multiplier and gradient methods", J. Opt. Theory Appl. 4(1969), 303-320.

11. B. W. Kort and D. P. Bertsekas, "A new penalty function method for constrained minimization", Proc. of IEEE Decision and Control

Conference, New Orleans, Dec. 1972.

12. S. A. Lill, "Generalization of an exact method for solving equality constrained problems to deal with inequality constraints", in Numerical Methods for Nonlinear Optimization, F. A. Lootsma (editor), Academic Press, 1973.

13. D. G. Luenberger, Introduction to linear and nonlinear programming, Addison-Wesley, 1973, 320-322.

14. O. L. Mangasarian, "Unconstrained Lagrangians in nonlinear programming", Computer Sciences Tech. Report #174, Univ. of Wisconsin, Madison, 1973.

15. A. Miele, E. E. Cragg, R. R. Iver and A. V. Levy, "Use of the augmented penalty function in mathematical programming, part I", J. Opt. Theory Appl. 8(1971), 115-130.

16. A. Miele, E. E. Cragg and A. V. Levy, "Use of the augmented penalty function in mathematical programming problems, part II", J. Opt. Theory Appl. 8(1971, 131-153.

17. A. Miele, P. E. Moseley and E. E. Cragg, "A modification of the method of multipliers for mathematical programming problems", in Techniques of Optimization, A. V. Balakrishnan (editor), Academic Press, 1972.

18. A. Miele, P. E. Moseley, A. V. Levy and G. M. Coggins, "On the method of multipliers for mathematical programming problems", J. Opt. Theory Appl. 10(1972), 1-33.

19. M. J. D. Powell, "A method for nonlinear optimiztion in minimization problems", in Optimization, R. Fletcher (editor), Academic Press, 1969.

20. R. T. Rockafellar, "New applications of duality in convex programming", written version of talk at 7th International Symposium on Math. Programming (the Hague, 1970) and elsewhere, published in the Proc. of the 4th Conference on Probability (Braşov, Romania, 1971).

21. R. T. Rockafellar, "A dual approach to solving nonlinear programming problems by unconstrained optimization", Math. Prog., to appear.

22. R. T. Rockafellar, "The multiplier method of Hestenes and Powell applied to convex programming", J. Opt. Theory Appl., to appear.

23. R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming", SIAM J. Control, to appear.

24. R. D. Rupp, "A method for solving a quadratic optimal control problem", J. Opt. Theory Appl. 9(1972), 238-250.

25. R. D. Rupp, "Approximation of the classical isoperimetric problem", J. Opt. Theory Appl. 9(1972), 251-264.

26. S. S. Tripathi and K. S. Narendra, "Constrained optimization problems using multiplier methods", J. Opt. Theory Appl. 9(1972), 59-70.

27. A. P. Wierzbicki, "Convergence properties of a penalty shifting algorithm for nonlinear programming porblems with inequality constraints", Archiwum Automatiki i Telemechaniki (1970).

28. A. P. Wierzbicki, "A penalty function shifting method in constrained static optimization and its convergence properties", Archiwum Automatyki i Telemechaniki 16(1971), 395-416.

29. A. P. Wierzbicki and A. Hatko, "Computational methods in Hilbert space for optimal control problems with delays", these proceedings.