Penalty Function Methods*

A.R. Conn
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada

Research Report CS-81-31
October 1981

* This work was supported in part by NSERC grant number A8639.

Penalty Function Methods

A.R. Conn

Department of Computer Science,
University of Waterloo,
Waterloo, Ontario, CANADA N2L 3G1.

This work was supported in part by NSERC grant number A8639.

Historically (around 1960!), there were no effective numerical algorithms available for nonlinear, constrained optimization, but good techniques for unconstrained optimization were starting to be developed. For example, Rosen's (1960) gradient projection method (perhaps the first effective technique for linear constraints) was published then, and Davidson's (1959) variable metric method was formulated in 1959, was re-presented in 1963 (Fletcher and Powell, 1963), and is still the basis of many current unconstrained algorithms. Therefore the original motivation for penalty functions was that they changed the extremely difficult nonlinear, constrained optimization problem into a sequence of (it was hoped) easier unconstrained problems for which reasonable algorithms existed. The questions I wish to pursue here are : why are we still interested in penalty functions, should we continue to be interested in them, and (finally) what is their future ?

The major immediate advantage of penalty functions, as we have already remarked, is that they represent the constrained problem in terms of unconstrained problems. However, this is of more significance than just the historical point concerning the availability of superior algorithms for the unconstrained calculation. Typically, the reformulation of the problem enables one to move away from nonlinear constraint boundaries, especially when one is not in the vicinity of a stationary point. In addition, to my mind, a considerable advantage is the intuitive global interpretation of the meaning of both the penalty function and the corresponding iterates of any related algorithm. Any particular penalty function, in effect, incorporates two aims, that of minimizing the objective function and that of satisfying the constraints. Consequently, if one is so far from a solution that one can make large gains in the objective function value, then it is of less importance to satisfy the constraints accurately. However, when one is close to a solution, then feasibility is likely to become a more dominant issue. Penalty functions are merely a consistent method of measuring progress in these two, often naturally conflicting, aims.

A major disadvantage is the choice of penalty parameters. In other words, the choice to give to the weights that fix the relative importance of the objective function and the various constraint functions.

It is, perhaps, worth pointing out that several aspects of penalty function methods can be closely related to techniques that are used in augmented Lagrangian, successive quadratic programming, and reduced gradient methods. In some respects penalty functions provide a more natural interpretation of these techniques (see for example Coleman and Conn, 1981a; Gill, Murray and Wright, 1981; Han, 1977; Powerll, 1978; Tapia, 1977). With these relationships in mind, it is certainly legitimate to state that some of the best algorithms available are, in effect, penalty techniques.

A particular subclass of penalty functions are the so-called exact penalty functions, that is those that give rise to a finite sequence of unconstrained problems, which may be differentiable (see the paper by Bertsekas in this volume) or non-differentiable. The non-differentiable forms are typified by the penalty function of Zangwill (1967) and Pietrzykowski (1969), and have recently enjoyed a broad popularity. Essentially, in spite of their piecewise nature, these functions are simpler and more closely related to the original problem and its Lagrangian function than either the non-exact penalty functions or the differentialble exact penalty functions. Therefore they are useful even in the special cases of linear programming (Conn, 1976) and quadratic programming (Conn and Sinclair, 1975). I predict that, with the possible exception of trajectory methods (Murray and Wright, 1978), exact penalty functions will continue to hold more promise than the non-exact penalty functions, as long as one is not unduly concerned with the intermediate generation of infeasible points. However, in practice, one often requires even intermediate points to be feasible (see Gutterman 1982). For such problems those penalty methods referred to as barrier methods are the only viable penalty techniques, and the trajectory barrier methods are then particularly promising.

Trajectory analysis is certainly an approach that deserves more attention.

One consequence of this interest in exact penalty functions is that I expect there to be much more effort than there is at present on special line searches, trust regions, and other step size selecting mechanisms for piecewise differentiable functions. Some initial work in this area has been done by Murray and Overton (1979) and Charalambous and Conn (1978). It has also been suggested at this meeting (Davidon, 1982) that conic approximations may be particularly useful for line searches because of their applicability to functions with poles (this remark also applies to differentiable penalty functions). However, conic approximations appear to me to be more useful when the positions of the poles are not known, which is not the case here. In particular, some of the methods mentioned above rely strongly for their success on an a priori knowledge of the location of the neighbourhoods of the non-differentiable points.

It is difficult to foresee any significant advances in the local convergence properties of current algorithms applied to small dense problems that does not presuppose equivalent advances in unconstrained optimization. The reason for this comment is that, in using the term local, I am assuming that one is in the neighbourhood of a solution, and, in particular, that one has identified correctly the active constraints. Therefore the essential problem at hand is to find a stationary point of a Lagrangian, or, equivalently, to solve an unconstrained problem in a reduced space. In fact, the current state of local convergence of penalty function methods is analogous to the corresponding state of unconstrained optimization with two exceptions. There is still no published quasi-Newton method that updates only a projected Hessian matrix and that maintains a superlinear convergence rate; such methods are considered by Coleman and Conn (1981b). Secondly, whenever we have preferred to choose our search directions from a model that is distinct from the line search merit function (as is common in most of the successive quadratic programming methods; see Bartholomew-Biggs, 1982), there has been difficulty in guaranteeing that a stepsize of unity is eventually acceptable (see, for example, Chamberlain, Lemaréchal, Pedersen and Powell, 1980). At present a unit stepsize is an essential ingredient in any algorithm that claims to mimic efficient unconstrained algorithms asymptotically.

Thus, one current difficulty is to identify when the region of asymptotic convergence has been reached. I suggest that more effort should be spent on studying the earlier stages of iterative methods. An isolated example of genuine non-asymptotic analysis is given by Overton (1981).

A related question concerns the type of penalty transformation that is used. Lemarechal (1982) considered two distinct extreme kinds of algorithm, namely a) when the set of all generalised gradients is used explicitly, and b) when only one element of the set is used. He then refers to the method of bundles, which can be regarded as a compromise of b) that is closer to a). There is also the possibility of a compromise of a) closer to b), which suggests a penalty transformation that involves only a few constraint functions.

For example, consider the problem

$$\begin{aligned}&\text{minimize } f(x), \quad x \in R^n\\ &\text{subject to } \quad c_i(x) \geq 0, \; i \in I,\end{aligned}$$

and consider replacing the usual exact penalty function

$$p(x,\mu) = f(x) - \frac{1}{\mu} \sum_{i \in I} \min\,[0, c_i(x)]$$

by the exact penalty function

$$\hat{p}(x,\mu) = f(x) - \frac{1}{\mu} \min\,[0, \min_i c_i(x)] \;,$$

i.e. we replace the sum of the infeasibilities in the penalty term by the greatest infeasibility. We have in mind the possibility of "guessing" j such that $c_j(x) = \min\{c_i(x)\}$, and then defining

$$\tilde{p}(x,\mu) = f(x) - \frac{1}{\mu} \; c_j(x) \; .$$

From time to time we expect to update our guess.Assuming one uses a projection-like method for determining a descent direction for $\tilde{p}$, the optimal number of active constraints to include in the projection is an open question. It is also uncertain how often one should update j.

The main advantage of this approach would be economies of computation far from the solution when the original problem contains many constraints. A disadvantage is that one has to correct for an inappropriate choice of j. It is of interest to note that, assuming j always corresponds to a violated constraint, there does exist a positively weighted exact penalty function of the form

$$\bar{p}(x,\mu) = f(x) - \frac{1}{\mu} \sum_{i \in I} w_i(x) \min\left[0, c_i(x)\right] ,$$

such that the sequence generated by $\tilde{p}$ produces a monotonically decreasing sequence of values of $\bar{p}$.

With reference to the identification of the correct active set, there are some known difficulties that have been largely ignored. For example, serious problems can be caused by degeneracy (due to redundant active constraints), near-degeneracy, spurious constraints that are nearly active (all these difficulties occur often in nonlinear ℓ_1 data fitting), and near-zero multipliers. Near-zero multipliers are discussed by Gill, Murray, Saunders and Wright (1979).

Furthermore these problems are not necessarily confined to neighbourhoods of local optima. In fact, a related problem that requires closer examination is false stationary points; an example is given in Coleman and Conn (1980).

Another algorithmic detail that has been largely ignored is the question of the choice of penalty parameters. Although, in the case of exact penalty functions, their threshold values are known in terms of the optimal Lagrange multipliers (see Luenberger, 1970, for example), there is little understanding of which values are ideal. At present, there is no global algorithm with an exact penalty merit function for which sophisticated updating techniques are applied to the parameters, and yet the resulting algorithm is robust and superlinearly convergent.

As in all areas of optimization, scaling is a significant problem for which few useful results are known.

Since some recent advances can be related to minimizing a quadratic function subject to quadratic constraints (Coleman and Conn, 1981), perhaps such models will be more generally applied. This is the simplest approach that allows curved constraints. Further, it would enable, for example, the practising model builder to replace 2n upper and lower bounds by one elliptical constraint, which is clearly more reasonable if one is using a penalty function algorithm, instead of a feasible direction algorithm.

It is heartening to note that the experience of nonlinear programmers with piecewise problems, that are motivated by exact penalty functions, has influenced the algorithmic and theoretical development of many other piecewise differentiable problems, including linear ones. (For examples, other than those already mentioned, see Bartels, 1980; Bartels and Conn, 1981; Bartels, Conn and Charalambous, 1978; Bartels, Conn and Sinclair, 1978; Calamai and Conn, 1980, 1981; Fletcher and Watson, 1981; Murray and Overton, 1980, 1981; Watson, 1981).

I anticipate that these fruitful offshoots will continue to grow, and, hopefully, will include broader aspects of non-smooth optimization.

I also feel that advances in large sparse nonlinearly constrained optimization will involve penalty function techniques. For example, MINOS-augmented (Murtagh and Saunders, 1980) is essentially a method based on Robinson's (1972) algorithm and the method of MINOS (Murtagh and Saunders, 1978). However, it is possible to replace the objective function of Robinson's algorithm by any exact penalty function, which poses some interesting challenges. For example, the usual dense techniques involve precise projections, orthogonal matrix decompositions, careful multiplier estimates, and the precise determination of quadratic minima, but sparse techniques may have to be less precise. In addition, for many large scale problems, the work of the matrix calculations of an iteration is very much greater than the cost of evaluating the objective function. For a fuller discussion of large sparse optimization problems see Gill, Murray and Wright (1981).

I am sure that much has been omitted in this brief introduction. One topic that comes to mind is the impact of computer hardware on the subject; for example, parallel processing may alter entirely the choice of a suitable penalty function. Another topic is whether penalty functions have any significant contributions to make to discrete optimization or global optimization.

I would like to emphasize the possibility of using penalty functions for purely linear problems, quadratic problems, and any other problems with special structure. It can be very helpful, provided one uses a penalty function that does not destroy the advantages attainable from the original structure.

Thus, I hope that in this paper I have made it clear that we should, indeed, still be interested in penalty functions, that they have certain inherent desirable properties, and that there is still research to be done.

REFERENCES

R.H. Bartels, A Penalty Linear Programming Method Using Reduced Gradient Basis-Exchange Techniques. Lin. Alg. and Applic. 29, pp. 17-32, 1980.

R.H. Bartels and A.R. Conn, An Approach to Nonlinear ℓ_1 Data Fitting, Proceedings 3rd Workshop on Numerical Analysis, Cocoyoc, Mexico, Springer-Verlag, 1981 (to appear).

R.H. Bartels, A.R. Conn and C. Charalambous, On Cline's Direct Method for Solving Overdetermined Linear Systems in the ℓ_∞ Sense, SIAM J. Num. Anal. 15, pp. 255-270, 1978.

R.H. Bartels, A.R. Conn and J.W. Sinclair, Minimization Techniques for Piecewise Differentiable Functions: The ℓ_1 Solution to an Overdetermined Linear System, SIAM J. Numer. Anal. 15, pp. 224-241, 1978.

M.C. Bartholomew-Biggs, Recursive Quadratic Programming Methods, paper 4.2 in Proceedings of NATO ARC, Cambridge, 1981 (Ed. M.J.D. Powell), Academic Press, 1982.

D.P. Bertsekas, Augmented Lagrangians and Differentiable Exact Penalty Functions, paper 4.3 in Proceedings of NATO ARI, Cambridge, 1981 (Ed. M.J.D. Powell), Academic Press, 1982.

P.H. Calamai and A.R. Conn, A Second-Order Method for Solving the Continuous Multifacility Location Problem, Proceedings of 1981 Dundee Biennial Conference on Numerical Analysis: Ed. G.A. Watson, Springer-Verlag (to appear).

P.H. Calamai and A.R. Conn, A Stable Algorithm for Solving the Multi-facility Location Problem Involving Euclidean Distances, SIAM J. Scient. and Stat. Comp. 1, pp. 512-526 1980.

R.M. Chamberlain, C. Lemarechal, H.C. Pedersen and M.J.D. Powell, The Watchdog Technique for Forcing Convergence in Algorithms for Constrained Optimization, Rept. DAMTP 80/NA 1, University of Cambridge, U.K. 1980.

T.F. Coleman and A.R. Conn, Nonlinear Programming VIA an Exact Penalty Function: Global Analysis, Math. Program. 1981 (to appear).

T.F. Coleman and A.R. Conn, A Successive Quadratic Programming Method, that is quasi-Newton but updates only projected Hessians (to appear) 1981.

T.F. Coleman and A.R. Conn, Second-Order Conditions for an Exact Penalty Function, Math. Program. 19, pp. 178-185, 1980.

C. Charalambous and A.R. Conn, An Efficient Method to Solve the Minimax Problem Directly, SIAM J. Numer. Anal. 15, pp. 162-187, 1978.

A.R. Conn, Linear Programming via a Nondifferentiable Penalty Function, SIAM J. Numer. Anal. 13, pp. 145-154, 1976.

A.R. Conn and J.W. Sinclair, Quadratic Programming via a Non-differentiable Penalty Function, Tech. Rept. 75/15, Dept. of Combinatorics and Optimization, University of Waterloo, Canada, 1975.

W.C. Davidon, Variable Metric Method for Minimization, Argonne National Laboratory Report ANL - 5990 (Rev.) 1959.

W.C. Davidon, Discussion Contribution 1.20 in Proceedings of NATO ARI, Cambridge, 1981 (Ed. M.J.D. Powell), Academic Press, 1982.

R. Fletcher and M.J.D. Powell, A Rapidly Convergent Descent Method for Minimization, Comp. J. 6, pp. 163-168, 1963.

R. Fletcher and G.A. Watson, First and Second Order Conditions for a Class of Non-differentiable Problems, Math. Programming 20, 1981.

P.E. Gill and W. Murray, The Computation of Lagrange-Multiplier Estimates for Constrained Minimization. Report NAC 77 N.P.L. England.

M.M. Gutterman, Discussion Contribution 3.32 in Proceedings of NATO ARI, Cambridge, 1981 (Ed. M.J.D. Powell), Academic Press, 1982.

S.P. Han, A Globally Convergent Method for Non-linear Programming, J.O.T.A. 22, pp. 297-309, 1977.

C. Lemarechal, Paper 2.3 in Proceedings of NATO ARI, Cambridge, 1981 (Ed. M.J.D. Powell), Academic Press, 1982.

D.G. Luenberger, Control Problems with Kinks, IEEE Transaction on Automatic Control, AC-15, pp. 570-574, 1970.

W. Murray and M.L. Overton, Steplength Algorithms for Minimizing a Class of Nondifferentiable Functions, Computing 23, pp. 309-331, 1979.

W. Murray and M.L. Overton, A Projected Lagrangian Algorithm for Nonlinear Minimax Optimization, SIAM J. Scient. Stat. Comp. 1, pp. 345-370, 1980.

W. Murray and M.L. Overton, A Projected Lagrangian Algorithm for Nonlinear ℓ_1 Optimization, Report SOL. 80-4, Stanford, 1981.

W. Murray and M. Wright, Projected Lagrangian Methods Based on the Trajectories of Penalty and Barrier Functions, SOL 78-23, Stanford, 1978.

B.A. Murtagh and M.A. Saunders, Large Scale Linearly Constrained Optimization, Math. Program. 15, pp. 41-72, 1978.

B.A. Murtagh and M.A. Saunders, A Projected Lagrangian Algorithm and its Implementation for Sparse Nonlinear Constraints, Mathematical Programming Study, on constrained minimization (in press).

M.L. Overton, A Quadratically Convergent Method for Minimizing a Sum of Euclidean Norms, Technical Report No. 030 - Computer Science Department, Courant Institute, New York University 1981.

T. Pietrzykowski, An Exact Potential Method for Constrained Maxima, SIAM J. Num. Anal. 6, pp. 299-304, 1969.

M.J.D. Powell, Algorithms for Nonlinear Constraints that Use Lagrangian Functions, Math. Program. 14, pp. 224-248, 1978.

S.M. Robinson, A Quadratically-Converent Algorithm for General Nonlinear Programming Problems, Math. Program. 3, pp. 145-156, 1972.

J.B. Rosen, The Gradient Projection Method for Non-linear Programming Part I: Linear Constraints, J. SIAM, Vol. 8, pp. 181-217, 1960.

R.A. Tapia, Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization, J.O.T.A. 22, pp.135-194, 1977.

G.A. Watson, An Algorithm for Linear II Approximation of Continuous Functions. I.M.A.J. Numerical Analysis Vol. pp.157 - 167, 1981.

W.I. Zangwill, Non-linear Programming via Penalty Functions, Manag. Sci., 13, pp. 344-358, 1967.