Curvilinear Projection Developments

Henry J. Kelley* and Leon Lefton†
Analytical Mechanics Associates, Inc., Jericho, N.Y.
and
Ivan L. Johnson Jr.‡
NASA Johnson Space Center, Houston, Texas

Gradient projection is a powerful algorithm for minimization of a function subject to constraints, at its best when the constraint functions are linear or nearly so. Constraint nonlinearities hamper projection computations, often requiring termination of a one-dimensional search in the projected negative gradient direction short of the one-dimensional minimum sought, on account of buildup of constraint violations. The constraints must then be restored before another projection cycle, at a certain computational expense. The restoration steps taken in the process of following nonlinear constraint surfaces can be used as a guide to the construction of a curve which more nearly follows the constraints than does the straight line in the projected gradient direction. This scheme, termed "curvilinear projection," was explored in earlier research. The study presently reported carries out some computational experiments using a related version of the technique. Some other details of projection computations which turn out to be practically important are taken up: rules for updating the variable metric in projection when early termination of the one-dimensional search on constraint violation occurs, and active-constraint logic for screening inequalities that makes use of the Kuhn-Tucker necessary conditions. Computational comparisons on simple problems are presented.

Introduction

ORIGINAL and variable-metric versions of gradient projection algorithms for constrained minimization of a function are reported in Refs. 1-5. The present paper presents some recent improvements, and further investigations of a curved-search feature explored in Refs. 7 and 8 which affords improved constraint following.

The resurgence of interest in projection, on the part of the present writers, came with a surprise in the results of a comparison involving a seemingly slight modification of the Kelley-Speyer projection algorithm.[3] The modification was a provision for early updating of the variable metric whenever a screening test is passed. A notable convergence improvement was realized, resulting in the projection algorithm, which had been carried along merely for comparison, outperforming a more complex algorithm utilizing linear and quadratic penalties.

The algorithm will first be reviewed in its original equality-constraint version, then the updating rule just mentioned taken up. The restoration of constraints and the handling of inequality constraints will be discussed. Attention will then turn to the use of search along a curve, proposed in Refs. 7 and 8 with the idea of staying closer to constraint surfaces. Some computational experiments will then be described.

Variable-Metric Projection

The projection version of the Davidon-Fletcher Powell (DFP) algorithm[6] described in the following is essentially the algorithm of Ref. 3; some details are different, however, and the differences are important computationally. The process begins with constraint restoration, usually requiring several cycles; then optimization cycles alternate with restorations, which sometimes require more than one cycle. The present section will deal with optimization cycles, the following one with restorations.

A function $f(x)$ (x an n-vector) is to be minimized subject to m equality constraints

$$g_j = 0 \quad j = 1,\ldots,m \tag{1}$$

The process of Ref. 3 employs the formulas

$$\Delta x = -\alpha H(f_x + g_x\lambda) \tag{2}$$

$$\lambda = -(g_x^T H g_x)^{-1} g_x^T H f_x \tag{3}$$

A one-dimensional search on the scalar α is then carried out to minimize the function $f + g\lambda$. Appropriate penalty terms arrest the one-dimensional search whenever equality constraint violations much exceed $c_j |\bar{g}_j|$ in magnitude.[5] Projection cycles employ a DFP H-matrix, separate from that used in restoration cycles, updated according to

$$H + \Delta H = H + \frac{\Delta x \Delta x^T}{\Delta x^T(\Delta f_x + \Delta g_x\lambda)} - \frac{H(\Delta f_x + \Delta g_x\lambda)(\Delta f_x + \Delta g_x\lambda)^T H}{(\Delta f_x + \Delta g_x\lambda)^T H(\Delta f_x + \Delta g_x\lambda)} \tag{4}$$

The update is performed only if

$$\Delta x^T(\Delta f_x + \Delta g_x\lambda) > 0 \tag{5}$$

which assures positive definiteness of the updated H. This represents a departure from earlier versions of the algorithm[3,5] in which termination of one-dimensional search on a minimum of $f + g\lambda$ was required before updating of H was permitted, i.e., updating was deferred until the vicinity of the constrained minimum had been reached.

Constraint Restoration Phase

The initial nulling out of constraint functions often proves more challenging than subsequent restorations in that the constraint violations to be dealt with are ordinarily larger in

Received July 27, 1977; presented at the AAS/AIAA Astrodynamics Conference, Jackson, Wyo., Sept. 7-9, 1977; revision received Sept. 7, 1977.

Index categories: Guidance and Control; Performance.

*Vice President. Associate Fellow AIAA.
†Senior Programmer/Analyst.
‡Aerospace Technologist, Mission Planning and Analysis Division.

magnitude. For this purpose, minimization of a function $\hat{f}$ is employed:

$$\hat{f}=\frac{1}{2}\sum_{j=1}^{m}k_j g_j^2+\frac{k_0}{2}(f_0-f)^2 h(f-f_0) \qquad (6)$$

This is a weighted sum of squares of the constraint functions plus a term intended to counter gross increases in f. The term corresponds to penalty-function treatment of an inequality $f_0-f\geq 0$. Here h is the heaviside unit step function. The k_j are determined from

$$k_j=\frac{\bar{k}}{m}\frac{\sum_{i=1}^{m}|g_{i_x}|^2}{|g_{j_x}|^2} \qquad j=1,2,\ldots,m \qquad (7)$$

where $\bar{k}$ is input. This choice would make equal the contribution of each equality constraint to the second directional derivative of Eq. (6) in its own gradient direction at $g_j=0$, if the constraints were linear. The constraint $f_0-f\geq 0$ is included quadratic-penaltywise in Eq. (6) only during the first restoration sequence, with a coefficient k_0 taken as 1/10 the smallest of the k_j calculated from Eq. (7). The constant f_0 is estimated as the initial value of $f+g\lambda$.

The metric employed in correction sequences may be denoted A (to distinguish it from H of the optimization cycles). It is adjusted approximately for changes in the k_j, one at a time, using

$$A+\Delta A=A-\left[\frac{\Delta k_j}{1+\Delta k_j g_{j_x}^T A g_{j_x}}\right]A g_{j_x} g_{j_x}^T A \qquad (8)$$

This correction, from Ref. 9, is based on the idea that A approximates $\hat{f}_{xx}^{-1}$. The metric to start the first correction sequence is obtained as $A+\Delta A$ from Eq. (8), using $A=I$ and $\Delta k_j=k_j-1$ [k_j from Eq. (7)]. If n or more updates are completed in this sequence, the emerging DFP metric is carried over to the next; if not, the initial metric is carried over. In either case, adjustments for any changes in the k_j are performed via Eq. (8) before use. Negative increments Δk_j are limited in magnitude to insure that the denominator of the fraction in parentheses does not nearly vanish.

The second and subsequent restoration sequences employ

$$\Delta x=-\alpha A g_x^T(g_x^T A g_x)^{-1}g \qquad (9)$$

together with a one-dimensional search vs α for a minimum of $\hat{f}$ given by Eq. (6), but with the last term deleted. This correction scheme, with $\alpha=1$ and without a search, was originally proposed by Rosen;[1] it effects restoration in a single step for linear g. The existence of the inverse in Eq. (9) [and in Eq. (3)] requires that the matrix g_x have rank m. This condition is met at the constrained minimum in the classical normal case in which the tangent-plane approximations to the constraints are well defined and distinct. Note that there is no guarantee that Eq. (9) is a direction of descent for $\hat{f}$, with general k_j values; thus the one-dimensional search may fail and reversion to DFP minimization of $\hat{f}$ become necessary.

The magnitude of constraint violation upon which optimization cycles are terminated short of a one-dimensional minimum is $c_j\bar{g}_j$, where $\bar{g}_j$ is a preconceived tolerance and c_j, usually $\gg 1$, is a factor adjusted with the aim of just permitting restoration with a single cycle of Eq. (9), to within the tolerance. Since the use of a single c-factor for all constraints met with only limited success, a c-vector was resorted to. The components were adjusted adaptively if somewhat heuristically in the following way: c_j is increased 10% if a single restoration proves successful; it is halved if two restoration cycles are required; and it is cut to one-quarter if there are additional cycles.

Treatment of Inequalities

It is of interest to determine a minimum subject to a mix of equality and inequality constraints, the latter expressed by

$$g_j\geq 0 \quad j=m+1,\ldots,m+p \qquad (10)$$

During the initial correction sequence, these are handled penalty-function fashion,[2] the function $\hat{f}$ to be minimized given by

$$\hat{f}=\frac{1}{2}\sum_{j=1}^{m}k_j g_j^2+\frac{1}{2}\sum_{j=m+1}^{m+p}k_j g_j^2 h(-g_j)$$
$$+\frac{k_0}{2}(f_0-f)^2 h(f-f_0) \qquad (11)$$

with the k_j determined as though all constraints were equalities:

$$k_j=\frac{\bar{k}}{(m+p)}\frac{\sum_{i=1}^{m+p}|g_{i_x}|^2}{|g_{j_x}|^2} \qquad j=1,2,\ldots,m+p \qquad (12)$$

The determination of the active constraint set for optimization and restoration cycles proceeds first by excluding those satisfied with a margin $g_j\geq\bar{g}_j$, where $\bar{g}_j>0$ is a preset threshold. Those candidate inequality constraints for which $g_j<\bar{g}_j$ are then screened further via the Kuhn-Tucker conditions $\lambda_j\leq 0$,[10,11] using Eq. (3) first with *all* the candidates included, then successively with Kuhn-Tucker violators dropped, as many times as necessary, until all $\lambda_j\leq 0$ or all candidates are screened out. Inactive constraints are treated in penalty-function approximation.

The Kuhn-Tucker conditions employed apply to the problem of minimizing a linear approximation to the function f subject to linearized constraints and to a quadratic constraint on step size. They become identical to the Kuhn-Tucker conditions for the original problem when evaluated at the constrained minimum sought.

The Kuhn-Tucker screening has generally been found to be worth the computational expense in reducing tendencies of constraints to switch between active and inactive status from cycle to cycle. The present effort has proceeded on the assumption that vector-matrix operations are cheap computationally in relation to the cost of gradient and function samples; this is realistic for the trajectory optimization applications of particular interest to the writers.

Curved Search

Constraint nonlinearities hamper projection computations a great deal in applications work, often requiring termination of a one-dimensional search short of the one-dimensional minimum sought on account of constraint-violation buildup. It is of interest to deflect the search away from the straight line in the negative projected gradient direction so as to follow approximately the nonlinear constraint intersection, as proposed in Refs. 7 and 8. An improved version of the curved-search technique is given in the following.

It is assumed that at least one projection cycle has already been completed (the first is done with a linear search) and that the derivative of $f+g\lambda$ with respect to the step-size parameter α has been reduced in magnitude by no more than half, that the constraints have been restored by one or more correction cycles, and that there has been no change in the active constraint set.

A curvilinear-projection cycle proceeds by

$$\Delta x=\xi\alpha+\eta\alpha^2 \qquad (13)$$

which replaces Eq. (11). Here

$$\xi \quad =-H(f_x+g_x\lambda) \qquad (14)$$

$$\eta = -(\Delta\tilde{x} - \xi\tilde{\alpha})/\tilde{\alpha}^2 \quad (15)$$

$$\tilde{\alpha} = -\sum_{i=1}^{n} \xi_i \Delta\tilde{x}_i \Big/ \sum_{i=1}^{n} \xi_i^2 \quad (16)$$

The vector $\Delta\tilde{x}$ is the difference between x from the beginning of the preceding projection cycle to the present restored point, the beginning of the next. The scalar $\tilde{\alpha}<0$ is such that the earlier point is regenerated when $\alpha=\tilde{\alpha}$ is introduced into Eq. (13). Thus, Eq. (13) generates a parabola in x space which passes through both restored points and is tangent to the projected gradient vector at the later one.

The curved-search sequencing presently in use provides for a possible curved-search on all optimization cycles except the first, which uses a linear search. Subsequent optimization cycles use a curved search provided the H-update test of Eq. (5) was met on the preceding cycle, none of the inequality constraints has changed status (from or to active), and the preceding one-dimensional search did not proceed more than halfway to a minimum (as measured in terms of the magnitude of the derivative of $f+g\lambda$ with respect to the step-size parameter α). Earlier exploratory computations were more cautious in the use of curved searches, and generally less successful. The curving steps do nothing beneficial for conjugacy in the subspace of the constraint intersection, but this is already a lost cause with DFP when full steps to one-dimensional minima are not being taken.

Test Problems

The test problems employed for experimentation were

$$f = x_1 + a_1 x_2^2 + a_2 x_1^3$$

$$g_1 = x_1 - b_1 x_2^2 - b_2 x_3^2 - b_3 x_2^4$$

$$g_2 = x_3 - c$$

The coefficients were

$$a_1 = -10^{-2}, \quad a_2 = 10^3$$

$$b_1 = 1, \quad b_2 = 10^2, \quad b_3 = 10^{-1}$$

$$c = 10^{-1}$$

The starting point for the numerical computations to be presented was

$$x_1 = 10, \quad x_2 = 5, \quad x_3 = 10$$

Test problem 1 had a single equality, $g_1=0$; 2 had two equalities, $g_1=0$, $g_2=0$; 3 one equality, $g_1=0$, and one inequality, $g_2\geq 0$; 4 two inequalities, $g_1\geq 0$, $g_2\geq 0$; 5 two inequalities with the second reversed, $g_1\geq 0$, $-g_2\geq 0$; and 6 one equality, $g_1=0$, and one inequality, also reversed, $-g_2\geq 0$.

Computational Comparisons

The results of Table 1 illustrate various features in the context of equality constraints. Linear vs curvilinear projection results are presented in Table 2 for a larger set of test problems. It is noted that the improvement provided by the curved-search feature is considerably smaller in problems which include inequalities.

Table 1 Number of gradient and function-sample computations required for various test problems

	Test problem 1	Test problem 2
Original Kelley-Speyer[3]	44, 691	72, 1022
Kelley-Speyer plus early H update	35, 480	40, 499
Improved restoration logic added	29, 338	31, 324
Curved search added	17, 218	24, 253

Table 2 Number of gradient and function-sample computations required for various test problems: linear vs curvilinear projection

Test problem	1	2	3	4	5	6
Kelley-Speyer improved, linear	29, 338	31, 324	27, 304	27, 310	27, 324	27, 329
Kelley-Speyer curvilinear	17, 218	24, 253	23, 259	24, 254	26, 296	25, 285

Concluding Remarks

Several developments and refinements of variable-metric projection have been presented, including a curved-search technique for nonlinear-constraint-surface following, improved means for control and correction of constraint violations, and screening criteria for active-constraint logic for use with inequalities. Projection schemes appear quite promising and worth further development and evaluation effort. Experience with a larger variety of problems applying the various features described is of interest.

Acknowledgment

This work was performed in part under Contract NAS 9-14818 with NASA Johnson Space Center, Houston, Texas.

References

[1] Rosen, J.B., "The Gradient-Projection Method for Nonlinear Programming. Part I: Linear Constraints," *SIAM Journal,* March 1960; "Part II: Nonlinear Constraints," *SIAM Journal,* Dec. 1961.

[2] Kelley, H.J., "Method of Gradients," *Optimization Techniques,* edited by G. Leitmann, Academic Press, New York, 1962.

[3] Kelley, H.J. and Speyer, J.L., "Accelerated Gradient Projection," *Lecture Notes in Mathematics 132,* Springer-Verlag, Berlin, 1970.

[4] Rosen, J.B. and Kreuser, J., "A Gradient-Projection Algorithm for Nonlinear Constraints," *Conference on Numerical Methods for Nonlinear Optimization,* University of Dundee, Scotland, June 28-July 1, 1971.

[5] Myers, G.E., "Numerical Experience with Accelerated Gradient Projection," *Conference on Numerical Methods for Nonlinear Optimization,* University of Dundee, Scotland, June 28-July 1, 1971.

[6] Fletcher, R. and Powell, M.J.D., "A Rapidly Convergent Descent Method for Minimization," *Computer Journal,* July 1963.

[7] Kelley, H.J. and Johnson, I.L., Jr., "Curvilinear Projection," *IFIP Colloquium on Optimization Techniques,* Santa Monica, Calif., Oct. 19-22, 1971.

[8] Falco, M. and Kelley, H.J., "Aircraft Symmetric Flight Optimization," *Control and Dynamic Systems,* Vol. 10, Academic Press, New York, 1973.

[9] Kelley, H.J., Denham, W.F., Johnson, I.L., Jr., and Wheatley, P.O., "An Accelerated Gradient Method for Parameter Optimization with Nonlinear Constraints," *Journal of Astronautical Sciences,* Vol. 13, July-Aug. 1966.

[10] Kuhn, H.W. and Tucker, A.W., "Nonlinear Programming," *Proceedings of the Second Berkeley Symposium,* University of California Press, 1951.

[11] Aoki, M., *Introduction to Optimization Techniques,* McMillan, New York, 1971.