Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Communications to the Editor—Exponential Forecasting: Some New Variations

Garth P. McCormick,

MANAGEMENT SCIENCE
Vol. 15, No. 5, January, 1969
Printed in U.S.A.

COMMUNICATIONS TO THE EDITOR

Exponential Forecasting: Some New Variations*

Exponential forecasting or smoothing models are presented which take into account all combinations of trend and cyclical effects in additive and multiplicative form. Although other authors have hinted that the proposed models existed, no one has ever presented them in comprehensive form. This paper presents the nine possible models in graphical form and summarizes them by one summary formula.

Introduction

Exponential Forecasting[1] is a technique for predicting sales of individual items by extrapolating sales time series. Its use is appropriate only in those cases where it is impractical to introduce information concerning the market, the industry, and the economy. Products that lend themselves to exponential forecasting are found in the market of replacement parts, hardware items, kitchen wares, and so forth.

The only input to an exponential forecasting system, once the system is operational, are the forecasting formula and the parameters of the particular formula. A forecast can then be obtained for each period, and this forecast is based on the historical time series, and on the actual sales of the most recent period. The latter must be supplied to the system each period.

The system described above was briefly mentioned by Magee [1], was discussed in more detail by Brown [2, 3], and eloquently introduced by Holt et al [4]. Magee only discussed a basic model which could not handle seasonal or trend effects. Brown and Holt *et al* introduced the notion of seasonal and trend effects, but never completely presented all possible seasonal and trend effect combinations. However, it is imperative that the appropriate model be used, once it has been decided to use exponential forecasting.

In this paper the basic exponential approach has been extended to several new types of stochastic processes. The ease of these extensions and their flexibility in handling different forecasting problems illustrates several attractive features of this forecasting approach.

A total of nine possible models will be presented. Graphical illustrations implicitly suggest which model is the most appropriate to use on the basis of historical sales.

The Exponential Models

The nine possible models mentioned above consist of three sets of three models. The first set, portrayed in Figure 1, consists of the basic exponential model, Model 1-A, the basic model with an additive trend effect, Model 1-B, and the basic model with a multiplicative trend effect, Model 1-C. The second set of three models, Models 2-A, 2-B, and 2-C, presented in Figure 2, consists of the three models in Figure 1 with an additive seasonal effect superimposed on them. The third set of three models, Models 3-A, 3-B, and 3-C, presented in Figure 3, again consists of the three models in Figure 1, but now with a multiplicative seasonal effect superimposed on them. Holt *et al* only discuss Models 1-A, 1-B, and 3-B, because they felt that it is more generally

* Received June 1967 and revised December 1967.

[1] Exponential forecasting is also known as exponential smoothing, adaptive smoothing, or adaptive forecasting.

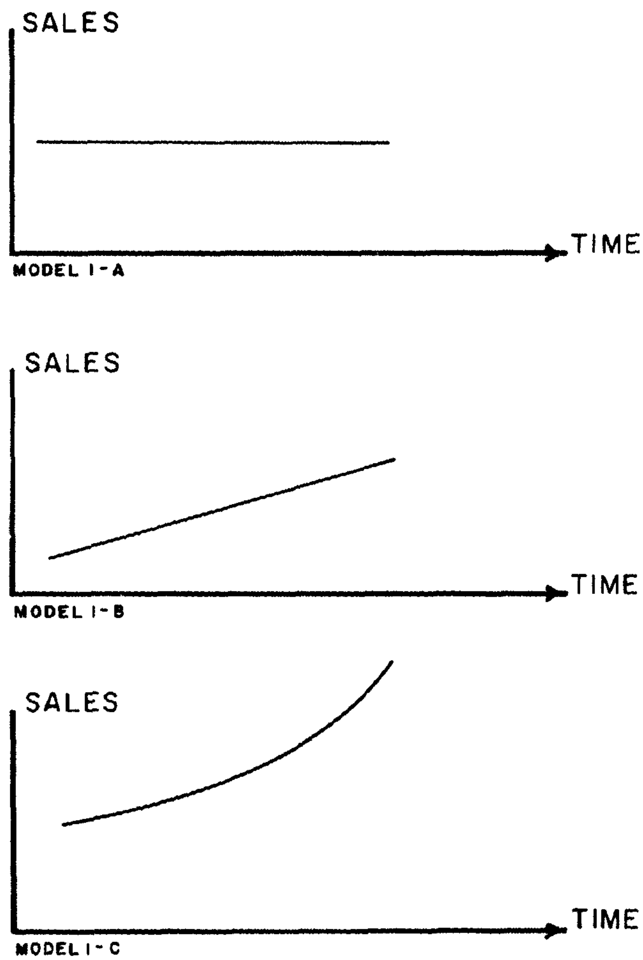

FIGURE 1

useful to work with these models. However, some of the nine models presented above, appear as probable, and Model 1-C even appears more probable in real-life applications than Model 1-B. Of the nine models presented, only Model 3-A appears unrealistic to the author.

The mathematical justification of exponential forecasting can be found in the references and will, therefore, not be discussed in this paper. However, all the models portrayed graphically above, will be presented in mathematical form below.

The nine models can be summarized by the formula:

$$\bar{S}_t = w_e d_{s,t} + (1 - w_e) d_{T,t}, \tag{1}$$

where $\bar{S}_t$ is the basic sales forecast component in period t, w_e is a weight, $0 \leqq w_e \leqq 1$, and $d_{s,t}$ and $d_{T,t}$ are given in Table 1. Each entry in the 3×3 matrix of the table is directly related to the model number shown in Figures 1, 2, and 3. For instance, the model with no trend effect but with an additive seasonal effect is Model A-2, obtained from row A and column 2.

The symbols denoted by $d_{s,t}$ and $d_{T,t}$ will now be discussed. The first one, S_t, is the actual sales in period t. The additive trend component, A_t, is derived from the

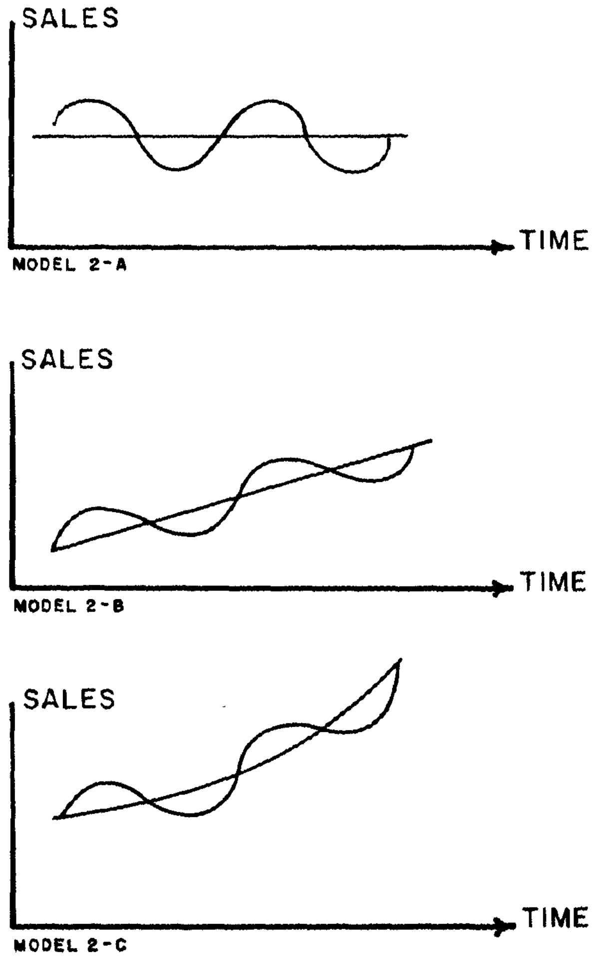

FIGURE 2

formula:

$$A_t = w_A(\bar{S}_t - \bar{S}_{t-1}) + (1 - w_A)A_{t-1}, \tag{2}$$

and represents the incremental increase in the basic sales forecast component for time period, t. w_A is a weight, $0 \leqq w_A \leqq 1$.

The multiplicative trend component, B_t, is derived from the formula:

$$B_t = w_B\bar{S}_t/\bar{S}_{t-1} + (1 - w_B)B_{t-1}, \tag{3}$$

and represents the incremental increase in the basic sales forecast for time period, t. w_B is a weight, $0 \leqq w_B \leqq 1$.

The additive seasonal component, G_t, is derived from the formula:

$$G_t = w_G(\bar{S}_t - S_t) + (1 - w_G)G_{t-L}, \tag{4}$$

and represents the parameter which normalizes actual sales in the basic forecasting equation. w_G is a weight, $0 \leqq w_G \leqq 1$.

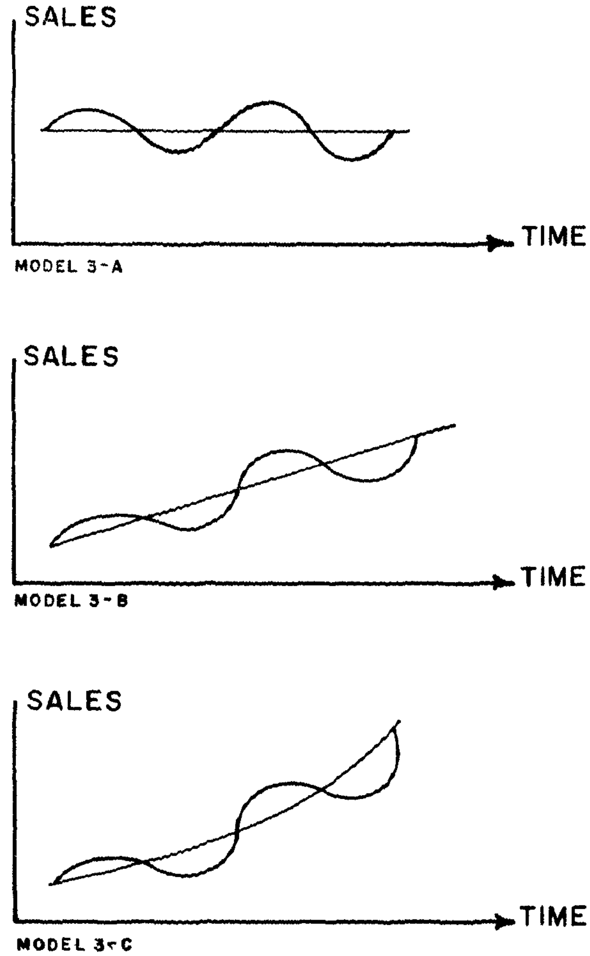

FIGURE 3

TABLE 1
Symbols Used in Summary Formula

		1	2	3
		no seasonal effect	additive seasonal effect	multiplicative seasonal effect
A. no trend effect	$d_{S,t}$	S_t	$S_t + G_{t-L}$	S_t/F_{t-L}
	$d_{T,t}$	$\bar{S}_{t-1}$	$\bar{S}_{t-1}$	$\bar{S}_{t-1}$
B. additive trend effect	$d_{S,t}$	S_t	$S_t + G_{t-L}$	S_t/F_{t-L}
	$d_{T,t}$	$\bar{S}_{t-1} + A_{t-1}$	$\bar{S}_{t-1} + A_{t-1}$	$\bar{S}_{t-1} + A_{t-1}$
C. multiplicative trend effect	$d_{S,t}$	S_t	$S_t + G_{t-L}$	S_t/F_{t-L}
	$d_{T,t}$	$\bar{S}_{t-1}B_{t-1}$	$\bar{S}_{t-1}B_{t-1}$	$\bar{S}_{t-1}B_{t-1}$

The multiplicative seasonal component, F_t, is derived from the formula:

$$F_t = w_F S_t/\bar{S}_t + (1 - w_F)F_{t-L}, \tag{5}$$

and represents the parameter which normalizes actual sales in the basic forecasting equation. w_F is a weight, $0 \leqq w_F \leqq 1$.

TABLE 2

Sales Forecasts for the Nine Models

	1	2	3
	no seasonal effect	additive seasonal effect	multiplicative seasonal effect
A. no trend effect	$\bar{S}_t$	$\bar{S}_t + G_{t-L+T}$	$\bar{S}_t F_{t-L+T}$
B. additive effect	$\bar{S}_t + TA_t$	$\bar{S}_t + TA_t + G_{t-L+T}$	$(\bar{S}_t + TA_t)F_{t-L+T}$
C. multiplicative trend	$\bar{S}_t B_t^T$	$\bar{S}_t B_t^T + G_{t-L+T}$	$\bar{S}_t F_{t-L+T} B_t^T$

Actual sales forecasts for period $t + T$ can now be obtained from the general formula:

$$S_{t,T} = g_t \tag{6}$$

where the g_t's for each model are the entries in Table 2.

Conclusions

In multi-item demand forecasting situations it may be necessary to prepare sales forecasts for tens of thousands of items every month. The models presented in this paper, together with the appropriate computing machine, will perform this task accurately and efficiently. The graphical representations of each proposed model even allow the selection of the appropriate model provided some information on historical sales or, in case of a new product, estimates of future sales behavior are known.

C. Carl Pegels
State University of New York at Buffalo

References

1. Magee, John F., *Production Planning and Inventory Control*, New York: McGraw-Hill, 1958, Chapter 6.
2. Brown, Robert G., *Statistical Forecasting for Inventory Control*, New York: McGraw-Hill, 1959, Chapter 2.
3. Brown, Robert G., *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963, Chapters 7, 8, and 9.
4. Holt, Charles C., Franco Modigliani, John F. Muth, and Herbert A. Simon, *Planning Production Inventories and Work Force*, Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1960, Chapter 14.

Anti-Zig-Zagging by Bending*

An algorithm for solving the optimization problem: minimize $f(x)$ subject to $x_j \geqq 0, j = 1, \cdots, n$ is presented. The algorithm is a simple modification of Cauchy's steepest descent method. The main idea is to move from any point x^k along a feasible piece-wise linear path defined by the constraint set and $\nabla f(x^k)$, taking x^{k+1} to be the first local minimizing point along that path. Convergence to a constrained stationary point is proved.

1. Introduction

The programming problem (in E^n)

$$\text{minimize } f(x) \tag{A}$$

$$\text{subject to } x_j \geqq 0, \qquad j = 1, \cdots, n \tag{1}$$

can be attacked by a variety of algorithms. The most straightforward way to approach

* Received November 1967 and revised June 1968.

this problem (although not the most efficient necessarily) is to apply a modification of Cauchy's steepest descent method [1]. Such a modification (in a more general setting) has been discussed in several places [3], [4], [7], [8], and [10]. It can be summarized as follows.

CM1 (Cauchy Modification Number 1)

Assume $f(x)$ is a continuously differentiable function. Let x^0 be a starting point of the process, and assume it is feasible (*i.e.*, satisfies (1)). Denote ∇f^0 as the n by 1 vector of first partial derivatives of f evaluated at x^0. Let x^0 be the direction vector at iteration 0 computed in the following way. Let

$$s_j^0 = 0 \quad \text{if} \quad x_j^0 = 0 \quad \text{and} \quad -\partial f(x^0)/\partial x_j \leqq 0, \tag{2}$$

$$s_j^0 = -\partial f(x^0)/\partial x_j \quad \text{otherwise.} \tag{3}$$

The next point in the iteration, x^1 is obtained as

$$x^1 = x^0 + t^0 s^0 \tag{4}$$

where $t^0 = \min(t_1, t_2)$ where t_1 is the maximum of the problem:

$$\text{maximize} \quad t \qquad \text{subject to} \quad x^0 + ts^0 \geqq 0, \tag{5}$$

and t_2 is the smallest local minimum to the problem:

$$\text{minimize} \quad f(x^0 + ts^0) \qquad \text{subject to} \quad t \geqq 0. \tag{6}$$

That such a t^0 can be obtained is a basic assumption of this paper. A procedure for obtaining t^0 comes in the category of assuming the existence of a "step-size" function and is discussed more thoroughly in [5] and in [9]. A detailed discussion of this is beyond the scope of this paper.

At iteration k, the same procedure described for iteration 0 is used, with x^k replacing x^0.

Several points need to be made about CM1. First, it is a feasible procedure. That is, if x^0 satisfies (1), then because of (5), x^k satisfies (1) for all k. Second, the algorithm will stop at x^k (i.e., $x^k = x^{k+1} = \cdots$) only if $s^k = 0$. This is a consequence of lemma 1.

We define x^k to be a *constrained stationary point* if s^k is given by (2) and (3) is the zero vector.

Lemma 1 [Stopping of the algorithm at a constrained stationary point]

If: (i) f is a continuously differentiable function,
(ii) at x^k, s^k is given by (2) and (3),
(iii) (4), (5), and (6) describing CM1 are used to obtain x^{k+1},
(iv) $s^k \neq 0$,

then: (a) $x^k \neq x^{k+1}$,
(b) $f(x^{k+1}) < f(x^k)$.

Proof: We need only show that $t^k \neq 0$, then both parts (a) and (b) follow.

Because of (2) and (3), there is an $\varepsilon_1 > 0$ such that for $0 \leqq t < \varepsilon_1$, $(x^k + ts^k) \geqq 0$. There is thus no difficulty in making small feasible movements from x^k along s^k.

Now if $t^k = 0$, for some $\varepsilon > 0$, $\varepsilon < \varepsilon_1$, that means $f(x^k) \leqq f(x^k + ts^k)$ for all $\varepsilon \geqq t > 0$. Rearranging, taking the limit yields

$$0 \leqq \nabla^T f^k s^k. \tag{7}$$

But from (2) and (3), this can happen only if $s^k = 0$. Thus t^k cannot equal zero.

Q.E.D.

This brings us to the main part of this paper. Although lemma 1 shows us that if the algorithm ceases after a *finite* number of steps, the final point must be a constrained stationary point, we would like to show that if the procedure is an infinite one, every limit point is a constrained stationary point.

This result, for the problem *without* constraints was proved by Curry [2]. An example showing that the algorithm CM1 can converge to a point which is *not* a constrained stationary point to problem (A) was given by Wolfe in [7]. The reason for this is the phenomenon of "zig-zagging" which can occur. Essentially this means that after a finite number of iterations, all t^k's are determined from (5). Putting it another way, locally minimizing $f(x)$ along s^k starting from x^k does not occur after a finite number of iterations.

Thus, the usual convergence lemmas do not apply. Zoutendijk [10] discussed this phenomenon and proposed "anti-zig-zagging" techniques to handle it. His techniques either require "remembering" the behaviour of previous iterations, or modifying the selection of the s^k vector by setting s_j^k to zero if x_j^k is less than ε and if $-\partial f^k/\partial x_j \leqq 0$, (i.e. by a perturbation of (2)). Neither of these is completely satisfactory since the conclusions of the desired convergence theorem have to be modified.

The purpose of this paper is to modify CM1 so as to avoid the problem of zig-zagging. The algorithm is autonomous, that is, each move is made independently of all previous iterations, and the vector s^k at each iteration is chosen as in (2) and (3). The revised method, called CM2 is presented in section 2 with the motivation for it. The proof of convergence to a constrained stationary point is given in section 3. Discussion of the extension of the method to general linearly constrained programming problems is in section 4.

2. Cauchy Modification Number Two (CM2)

It is important to observe (intuitively) that zig-zagging can only occur if, at some limit point (call it $\bar{x}$) of CM1, a variable, say x_1, has the two properties that $\bar{x}_1 = 0$, and $\partial f(\bar{x})/\partial x_1 = 0$. Otherwise, in a neighborhood of that point either x_1^k will remain zero ($\partial f(\bar{x})/\partial x_1 > 0$), or it will be increasing away from zero ($\partial f(\bar{x})/\partial x_1 < 0$). Also, if $\bar{x}$ is not a constrained stationary point then by definition for some variable, say x_2, $\partial f(\bar{x})/\partial x_2 \neq 0$, and $\bar{x}_2 > 0$ or if $\bar{x}_2 = 0$, then $\partial f(\bar{x})/\partial x_2 < 0$. In the limit then, the ratio of the magnitude of $\partial f(x^k)/\partial x_2$ to that of $\partial f(x^k)/\partial x_1$ will be infinitely large. It is overcoming this possibility which makes the following modified algorithm converge to a constrained stationary point.

CM2 At iteration k, s^k is chosen exactly as given by CM1, eqs. (2) and (3). The difference between the two algorithms is that if motion along the ray $x^k + t^k s^k$ ceases because of encountering a boundary of the constraints given by (1), CM2 continues from the boundary point along the same vector except that the appropriate component is set to zero. Minimization continues along this "bent" vector. The precise statement of the algorithm is as follows.

Define the vector $x^k(t)$ (for $t \geqq 0$) as

$$x_j^k(t) = \max\,[0, x_j^k - t\partial f(x^k)/\partial x_j], \qquad j = 1, \cdots, n. \tag{8}$$

Obviously $x^k(t)$ is a continuous function of t, and $x^k(0) = x^k$. The CM2 algorithm is simply at iteration k to find the value t^k such that t^k is the smallest local minimizing point of the problem

$$\text{(9)} \qquad \text{minimize } f[x^k(t)] \quad \text{subject to } t \geqq 0.$$

Then $x^{k+1} = x^k(t^k)$.

3. Convergence of CM2

We note for any t, the function $f[x^k(t)]$ has a right hand derivative at t and it is given by

$$\text{(10)} \qquad df[x^k(t)]/dt^+ = \nabla^T f[x^k(t)]u^k(t)$$

where

$$\text{(11)} \qquad u_j^k(t) = 0 \quad \text{if} \quad x_j^k(t) = 0 \quad \text{and} \quad -\partial f(x^k)/\partial x_j < 0,$$

$$\text{(12)} \qquad u_j^k(t) = -\partial f(x^k)/\partial x_j \quad \text{otherwise}, \quad j = 1, \cdots, n.$$

Thus t^k solves (9) only if

$$df[x^k(t^k)]/dt^+ \geqq 0.$$

Lemma 2. Assume $\bar{x}$ is not a constrained stationary point. Then if $t > 0$ is sufficiently small, for x sufficiently close to $\bar{x}$ (x assumed feasible) we must have

$$\text{(13)} \qquad f[x(t)] < f(\bar{x}).$$

Proof:[1] The counter assertion is for any $\varepsilon < 0$ there is a value t where $0 < t < \varepsilon$ and a feasible sequence $\{y^i\} \to \bar{x}$ such that $f[y^i(t)] \geqq f(\bar{x})$. By the continuity of f in x, eq. (8), $f[\bar{x}(t)] - f(\bar{x}) \geqq 0$. But dividing by t and taking the limit,

$$df[\bar{x}(0)]/dt^+ \geqq 0,$$

which is true only if $\bar{x}$ is a constrained stationary point. *Q.E.D.*

We now need to prove that if $\bar{x}$ is not a constrained stationary point, when x^k is close to $\bar{x}$, t^k is bounded away from zero, i.e., that consecutive points do not get arbitrarily close together as k gets large.

Lemma 3. For convenience let $\{x^k\}$ denote a subsequence converging to $\bar{x}$. If $\bar{x}$ is not a constrained stationary point, then

$$\liminf_{k\to\infty} t^k > 0.$$

Proof: In order for t^k to satisfy (9), $df[x^k(t^k)]/dt^+ \geqq 0$ must follow. We have (using (10), (11), and (12)),

$$\text{(14)} \qquad df[x^k(t^k)]/dt^+ = \Sigma_{j \notin B_k} - (\partial f[x^k(t^k)]/\partial x_j)(\partial x_j)(\partial f[x^k(0)]/\partial x_j)$$

where

$$B_k \equiv \{j \mid x_j^k(t_k) = 0 \quad \text{and} \quad -\partial f[x^k(0)]/\partial x_j < 0\},$$

or equals zero if the complement of B_k is empty. Now if some subsequence of t^k's goes

[1] The form of this proof and the following ones is similar to those contained in Wolfe [6] pp. 30–33.

to zero, for that subsequence, $x^k(t^k) \to \bar{x}$, and (14) yields

$$\lim_{k\to\infty} df[x(t^k)]/dt^+ = \Sigma_{j\epsilon\bar{N}} - (\partial f(\bar{x})/\partial x_j)^2 < 0$$

where $\bar{N} \equiv \{j \mid \bar{x}_j > 0 \text{ and } \partial f(\bar{x})/\partial x_j \neq 0, \text{ or } \bar{x}_j = 0 \text{ and } \partial f(\bar{x})/\partial x_j < 0\}$. The set $\bar{N}$ is not empty since $\bar{x}$ is assumed not to be a constrained stationary point. But this contradicts our assumption at the start of this proof. Hence no subsequence of t^k's can go to zero. *Q.E.D.*

The importance of lemma 3 is that if $\{x^k\}$ converges to $\bar{x}$ then the sequence $x^k(t^k)$ does not get close to $\bar{x}$ when $\bar{x}$ is not a constrained stationary point.

Theorem 1 [Convergence to a Constrained Stationary Point]

If: (i) f is a continuously differentiable function of x,

(ii) at x^k, (8) and (9) describing CM2 are used to obtain x^{k+1},

then: (a) every limit point of $\{x^k\}$ is a constrained stationary point.

Proof: If $t^k = 0$ for some k, the theorem is trivial (see proof of lemma 1). Otherwise let $\bar{x}$ be a limit point of $\{x^k\}$ and let $\{x^k\}$ denote also a subsequence converging to it. Assume $\bar{x}$ is not a constrained stationary point. Let $t\# = \lim \inf_{k\to\infty} t^k$ if this value is finite. Otherwise let $t\# = 1$. By Lemma 3, $t\# > 0$. Let t^* be such that $t\#/2 > t^* > 0$ and t^* sufficiently small so that lemma 2 applies.

Then, for k large enough,

$$\begin{aligned} f[x^k(t^k)] &< f[x^k(t^*)] && \text{(using (9) and the construction of } t^*) \\ &< f(\bar{x}) && \text{(lemma 2)} \\ &< f[x^k(t^k)] && \text{(monotonicity of } \{f(x^k)\}). \end{aligned}$$

This contradiction proves the theorem. *Q.E.D.*

4. Generalizations to Linearly Constrained Mathematical Programming Problems

The basic idea of this paper to prevent zig-zagging behaviour for problem (A) can of course be generalized to handle problems of the form

(B) $$\text{minimize} \quad f(x)$$

(15) $$\text{subject to} \quad Ax - b \geqq 0$$

where A is a $m \times n$ matrix. Using whatever method of vector generation is prescribed (such as that used in [8], [3]) minimization proceeds from the current point x^k along s^k until a constraint of the form (15) is encountered or until $f(x)$ is locally minimized. If the former occurs, say at $x^k + t^k s^k = x^k(t^k)$, a new vector is computed using the old gradient information at point x^k, not recomputing it at $x^k(t^k)$. Motion again proceeds, this time from $x^k(t^k)$. This pattern is continued until motion ceases at a point where $f(x)$ is locally minimized along some ray. Then and only then is the new gradient information used to generate a direction vector.

This method, besides avoiding zig-zagging, has the added advantage of requiring fewer computations of the objective function gradient. This will probably result in more efficient computer codes.

The extension of this algorithm to problem (B), using a particular method of direction vector generation, convergence theorems, and computational results will appear in a later paper.

GARTH P. MCCORMICK
Research Analysis Corporation

References

1. CAUCHY, A., "Méthode Générale pour la Résolution des Systèmes D'Equations Simultanées," *Comp. Rend. Acad. Sci.*, Vol. 25, pp. 536–538 (1847).
2. CURRY, H. B., "The Method of Steepest Descent for Non-linear Minimization Problems," *Quart. Appl. Math.*, Vol. 2, No. 3, pp. 250–261 (1944).
3. FAURE, P. AND P. HUARD, "Résolution des Programmes Mathématiques a Fonction Non-linéaire par la Méthode du Gradient Reduit," *Revue Francaise de Recherche Opérationelle*, Vol. 9, pp. 167–205 (1965).
4. HADLEY, G., *Nonlinear and Dynamic Programming*, Addison-Wesley Publishing Co., Reading, Mass. (1964).
5. TOPKIS, D. M. AND A. F. VEINOTT, JR., "On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming," Technical Report No. 6, Stanford University, Dept. of Industrial Engineering, Stanford, California, 5 Aug. (1966).
6. WOLFE, P., "Foundations of Nonlinear Programming: Notes on Linear Programming and Extensions—Part 65, "RAND Memo RM-4669-PR, Santa Monica, California (1965).
7. ——, "On the Convergence of Gradient Methods Under Constraint," IBM Research Paper, RC-1752, Jan. (1967).
8. ——, "Methods of Nonlinear Programming," *Recent Advances in Mathematical Programming*, R. L. Graves and P. Wolfe, eds.), McGraw-Hill Book Co., pp. 67–86 (1963).
9. ZANGWILL, W. I., "Convergence Conditions for Nonlinear Programming Algorithms," Working Paper No. 197, Center for Research in Management Science, University of California, Berkeley, California, Nov. (1966).
10. ZOUTENDIJK, G., *Methods of Feasible Directions*, Elsevier Publishing Company, Amsterdam and New York (1960).