HILL-CLIMBING: SOME REMARKS ON MULTIPLE OPTIMIZATION*

Bradford Howland**

Marvin Minsky**

Oliver G. Selfridge**

* The work reported in this paper was performed by Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology with the joint support of the U.S. Army, Navy and Air Force.

** Lincoln Laboratory, Massachusetts Institute of Technology.

J/ 33

INDEX

Summary

If we have a machine with 1000 knobs, how can we set them so as to minimize some output S of the machine, which represents, say, its error or departure from the behavior we want from it? We describe units, driven by S, which will each work a knob so that the whole system will tend toward an optimum. The units can be substantially identical, regardless of the actual structure of the machine. We exhibit three versions of these units, each with virtues and faults, and discuss their behavior, with concrete and synthetic illustrative experiments. There are divers aspects of their joint behaviors shown, and some caveats about their use, especially in large numbers. We have not finished testing these units in large assemblies, but it is probably unimportant that the component parts of each unit work accurately or reliably.

0.0 Introduction

One frequently wishes to adjust the controls of a machine, or the parameters of a system, so as to minimize or maximize some single output. This is a very general task, since the single output can be construed to represent the machine's departure from the behavior we desire of it. A general method of minimization would give us a powerful tool for solving any problem for which we could construct a usable measure of relative performance.

Occasionally one may be able to make a thorough "systems analysis" of the machine and obtain solutions for its "system equations" (if they can be found). But we are construing the problem as not inherently involving the actual system function relating inputs to output. Instead we consider the class of search processes which strive to find, through a sequence or iteration of local improvements, some best or "optimal" condition. It is intuitively convenient to think of this as a "hill-climbing" problem.

Some examples of the problems interesting us are:

1. To balance a bridge by simultaneously adjusting controls to minimize detector output.
2. To align a multi-stage superheterodyne receiver.
3. To vary the connections in a complicated net so as to approach some described behavior.
4. To adjust the parameters of a strategy function so as to improve the effectiveness of a search in which trials are expensive.

We are interested then in solving the minimization problem given the machine itself, rather than (as in servomechanism theory) given instead usable equations for the system; for often usable equations do not even exist.

The obvious approach is to supply the optimizer with some way of obtaining information about the lay of the land in its neighborhood, and also to supply it with some method for using this information to improve this position. Obviously the inductive method used must match, in some heuristic way, the hill or ensemble of hills to be encountered. We will therefore assume that the function has certain reasonable properties of continuity or smoothness.

Heuristically, the advantage of a "path" technique over a more global search is clear. Each new sample point is chosen in the neighborhood of the best result so far obtained, so that, in a sense, the machine is making good use of its prior experience. There is always the possibility that the machine will fail by becoming trapped on a local peak. See (7). In that case recourse must be taken to some more global sampling technique.

0.1 Background

Adaptive control techniques are becoming increasingly evident in the literature, and there is already far too vast an available bibliography for us to give here. Two examples must suffice. H.S. Tsien, Engineering Cybernetics, McGraw-Hill, New York, 1954, discusses the problem of "Optimizing Controls". R.R. Brown and B. Widrow have discussed "Adaptive Sampled-Data Systems" and simulation studies, chiefly in the Program Reports of the Computer Components and Systems Group, M.I.T.

But we feel that our approach and emphases are sufficiently different from those of these and other investigations to justify this study.

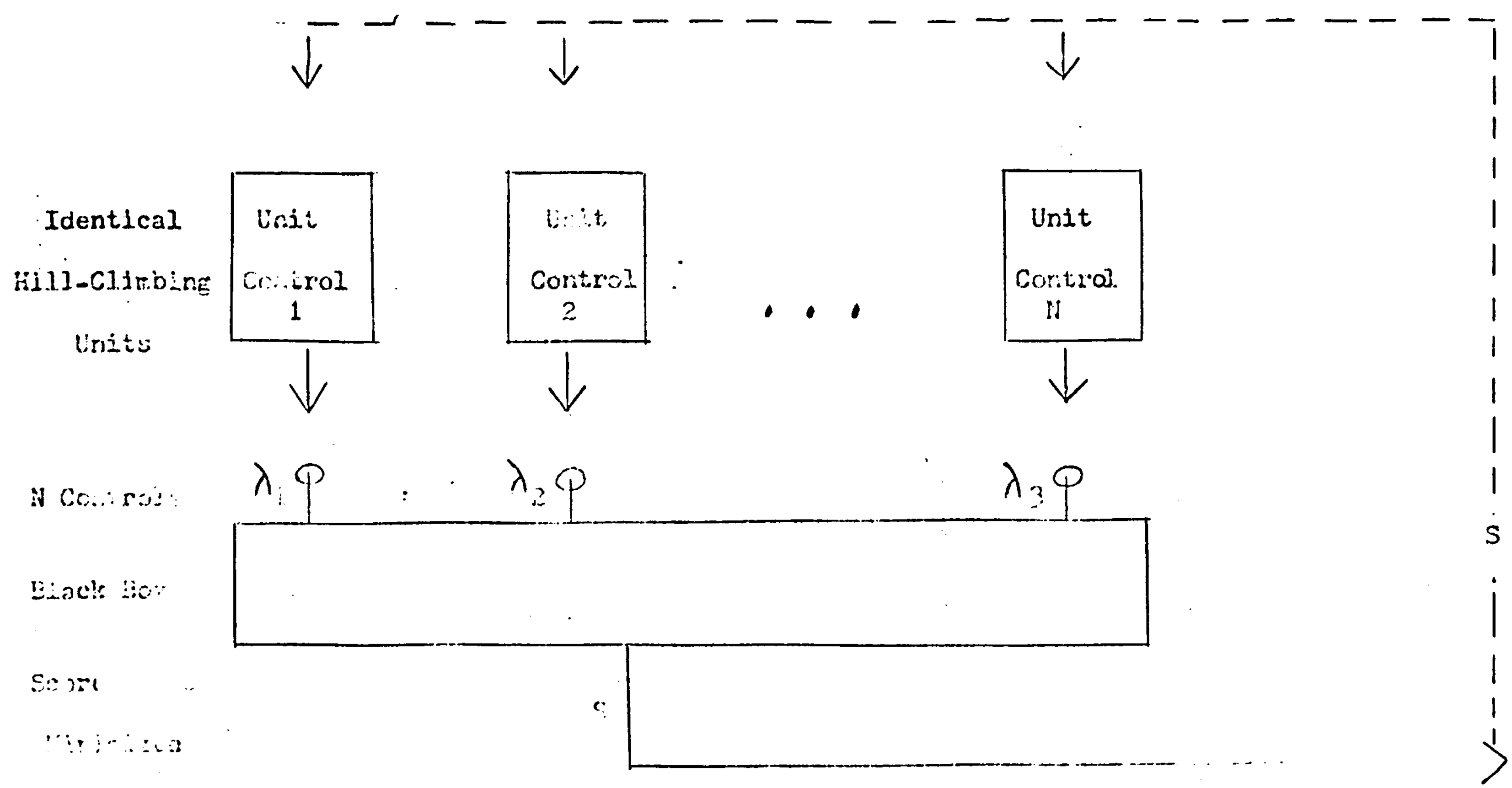


Fig. 1 Required Hill-Climbing Organization

1. The General Problem

We are given a function S(Score) of N variables $\{\lambda_i\}$ and must find its minimum (or maximum) by taking samples at various points. We assume henceforth that we wish to minimize S, and that it is reasonably "smooth".

The crux of our problem is that N is large, and hence so is the space to be searched. We should prefer a technique easily extendable to large N and, therefore prefer to treat all variables independently, similarly and, perhaps, simultaneously. Our technique, then, should have the general organization shown in fig. 1.

The notion of smoothness leads to the notion of a path of improvement in the N-dimensional space. We wish, then, to compute some analogue of $\frac{\partial S}{\partial \lambda_i}$, so that λ_i may be changed accordingly to improve (diminish) S; that is, so that

$$\frac{d\lambda_i}{dt} = -k \frac{\partial S}{\partial \lambda_i} \quad \text{in some sense.} \tag{1}$$

One may approximate $\frac{\partial S}{\partial \lambda_i}$ as

$$\frac{S(\lambda_i + \Delta \lambda_i) - S(\lambda_i - \Delta \lambda_i)}{2 \Delta \lambda_i} \quad ; \tag{2}$$

and even if the derivative $\frac{\partial S}{\partial \lambda_i}$ does not exist, if (2) is positive, the reduction of λ_i by $\Delta \lambda_i$ will improve S. Here we do not wish to go to the limit, for $\Delta \lambda_i$ is to be regarded rather as a real finite exemplary perturbation than an infinitesimal.

It seems natural to perform this calculation (2) sequentially and to generalize it to the form

$$\frac{\partial S}{\partial \lambda_i} = \frac{\overline{J_i(t)\, S\,(\lambda_i + J_i(t))}}{2 \left| \overline{J_i\,(t)^2} \right|} \quad , \tag{3}$$

where $J_i(t)$ is some 'jitter' or search function added to λ_i; since J_i should be directionally neutral, we have

$$\overline{J_i} = 0 \; . \tag{4}$$

Then (2) follows from (3) if $J_i(t) = \pm \Delta \lambda_i$. As it stands, (3) is not obviously useful unless one can restrict the averaging time and treat $\frac{\partial S}{\partial \lambda_i}$ as quasi-stationary; the point is to be able to use (3) in setting up the system (1). Of course the values of S will depend as well on the jitters of the other variables. Therefore the averaging must span sufficient intervals to make each equation (3) significant.

1 a.

This generalized gradient has the form of a correlation, not surprisingly. The use of a jitter function suggests a sequence of " experiments " to determine the gradient or local nature of the space; one uses correlations, or correlation-like calculations, to determine the overall or average trends in the outcomes of a series of experiments.

2. The Unit Hill-Climbers

The notions of section 1 we have incorporated in three different versions of unit hill-climbers. (See figs. 2, 7, 11). Naturally, they have different properties in detail, but they all satisfy the requirements of fig. 1.

To sense a control's effect on the score S, it must be moved. All three versions, Mods 1, 2, and 3, in some way jitter the controls λ_i with jitter functions J_i. We may state in general that, to avoid interaction:

$$\overline{J_i J_j} = 0 \quad \text{unless } i = j . \tag{5}$$

We shall illustrate tracks of hill-climbers with periodic J_i, but our arguments in the previous section suggest that random noise should do about as well (requiring, as before, that eqs. (5) and (6) hold). Indeed, one can argue that the random jitters can evade topographic anomalies that might trap any less general search patterns. The point will be further discussed in section 3.

2.1 Mod 1 Hill Climber

The most straightforward realization of equations 1, 2 and 3 is shown in fig. 2[1]. Here J_i is added to the control λ_i and the resulting S is correlated with J_i: first the product $\alpha J_i S$ is formed. The product is then subtracted from $\frac{dJ_i}{dt}$, so that the difference may be treated as a change in λ. If we wish to maximize S we replace the subtraction with an addition. The argument is clear when $\alpha << 1$, for then the averaging times are effectively long and the process may be considered quasi-stationary.[2]

Fig. 3a shows the score obtained experimentally as function of time for a 1-dimensional conical hill with the function $S = |\lambda|$, and with $J(t) = \sin(t - \phi)$. ϕ is a starting phase and there will be a family of tracks, depending on ϕ, shown in fig. 3a. A typical 2-dimensional trajectory is shown in fig. 4.

Analytically, from fig. 2, we have:

$$\lambda = J(t) - \alpha \int_{-\infty}^{t} J)\tau) S(\tau)\, d\tau . \qquad (6)$$

Let us suppose a conical hill, $S = \beta|\lambda|$; further that $J(t) = \pm 1$ with a period of 1. Solution [3] shows that λ approaches 0 at a linear rate $(e^{\alpha\beta} - 1)$, until $S = 2$; and the converges much more slowly to a steady state oscillation, whose peak values are $\lambda = \pm \frac{2}{1 + e^{-\alpha\beta}}$. The course of λ is shown in fig. 5. For small α and β,

$$e^{\beta\alpha} - 1 \approx \alpha\beta . \qquad (7)$$

---

(1) The second equivalent block diagram is included because it generalizes most easily, especially to Mod. The final block, 'S', may be considered a motor driven by

$$\frac{d\lambda_i}{dt} = \frac{dJ_i}{dt} - \alpha J_i S .$$

(2) But see sections 3 and 4.

(3) The simplest way to the solution is to write

$$\frac{d\lambda}{dt} = \frac{dJ}{dt} - \alpha J(t) S(t).$$

Here, and subsequently, we shall handle derivatives of discontinuous functions with (usually implied) delta function techniques, which are in all cases straightforward and unequivocal.

3 a.

3 b.

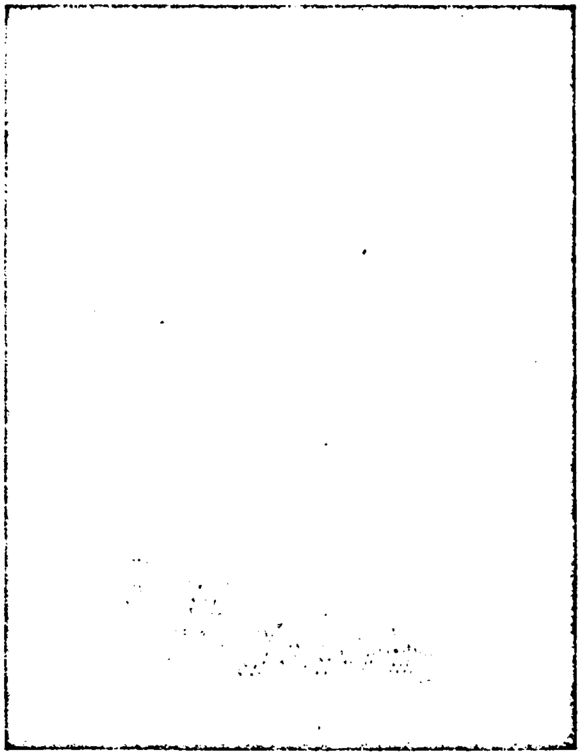


FIG. 3

a. SCORE vs. TIME OF UNIT MOD 1 HILL-CLIMBER ON CONICAL HILL, WITH SINUSOIDAL JITTER.

b. FAMILY OF TRACKS, SAME SITUATION, WITH 12 STARTING PHASES.

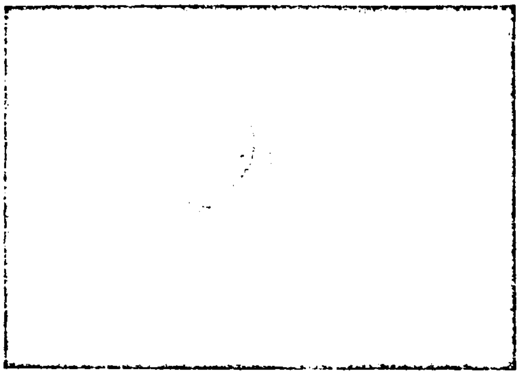


FIG. 4. TRAJECTORY OF 2-DIMENSIONAL HILL-CLIMBER WITH CONSTANT CIRCULAR JITTER CLIMBING CONICAL HILL.

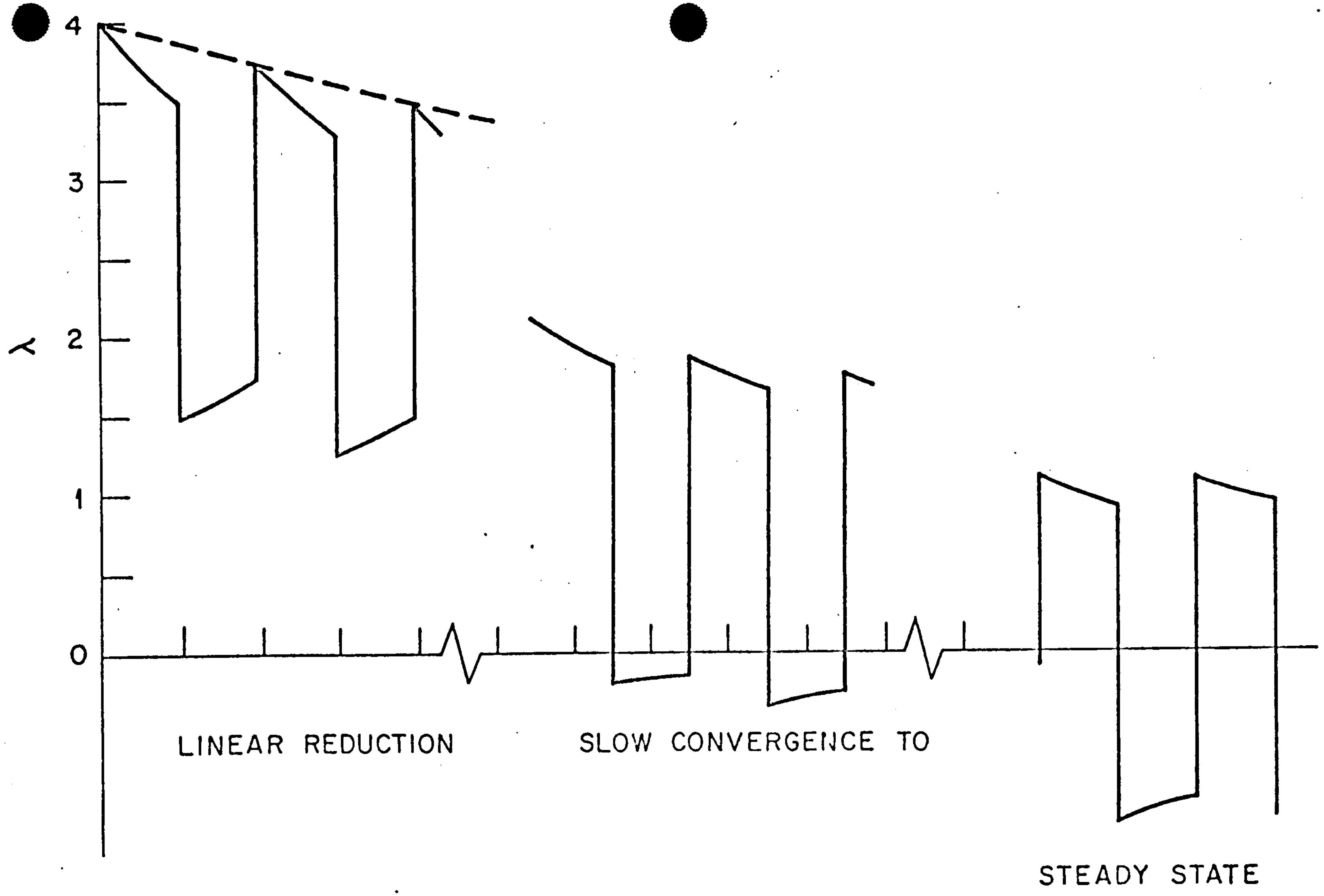


FIG. 5. BEHAVIOR OF MOD. 1 HILL-CLIMBERS FOR CONICAL HILL WITH SQUARE WAVE JITTER.

Thus for small slopes β and gains α, the rate of ascent is proportional to the slope and the gain, whenever the hill-climber is not too close to the summit. This is qualitatively true if J(t) is a more general function.

If we now suppose a parabolic hill $A = \beta\lambda^2$, eq. 7 leads to

$$\frac{d\lambda}{dt} = -\alpha\beta\lambda^2 \text{ except when } t = 0, 1, 2, \ldots . \quad (8)$$

The track of λ will be hyperbolic sections

$$\lambda = \frac{\lambda([t])}{+1 \pm \alpha\beta(t-[t])\lambda([t])} \quad (9)$$

(where [ ] means 'the integral part of'); the sections are joined by alternate jumps of ± 2 as in fig. 6, where we put $J = \pm 1$ as before, and $\alpha\beta = 0.1$. Eventually steady state obtains with peak values approximately $\lambda = \pm 1.05$.

On steep hills like this, and they are very cliff-like indeed for large λ, one must watch one's step, and the hill-climber can fall off to divergent behavior if it starts on the wrong foot. This problem is discussed more fully in section 4.

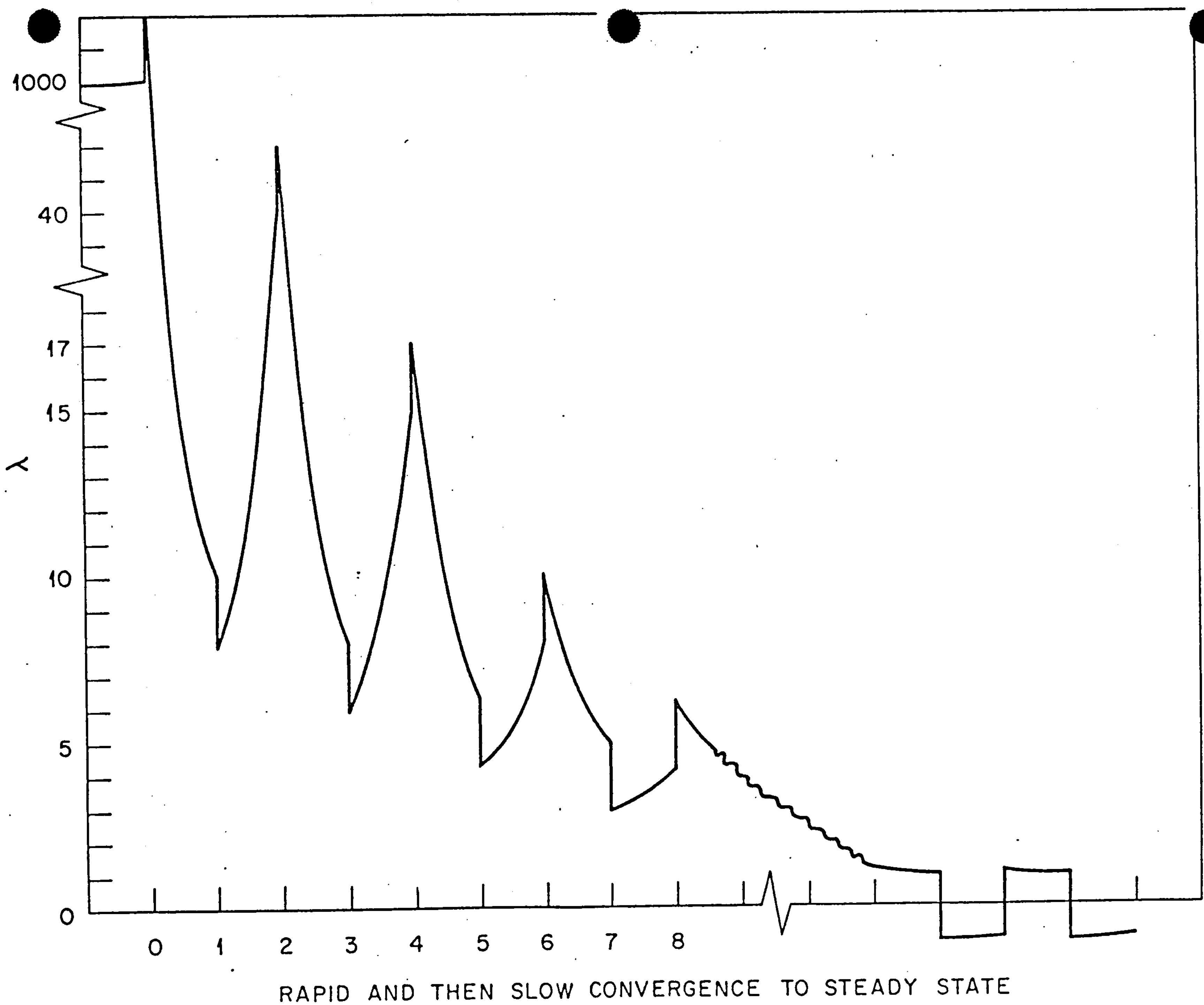
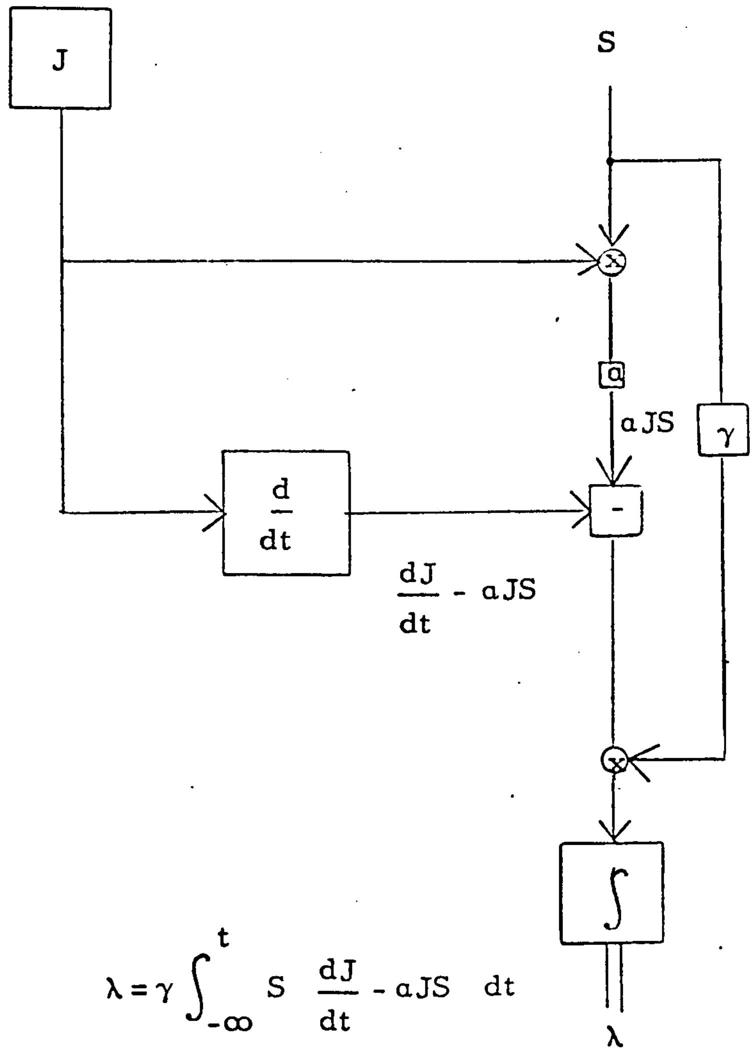


FIG. 6. BEHAVIOR OF MOD. 1 HILL CLIMBER FOR PARABOLIC HILL $S = \lambda^2$, WITH SQUARE WAVE JITTER.



$$\lambda = \gamma \int_{-\infty}^{t} S \quad \frac{dJ}{dt} - aJS \quad dt$$

FIG. 7 Unit Hill-Climber: MOD 2. Decaying Jitter

2.2 Mod 2 Hill-Climber

The most obvious in the first version is the jitter residual when the optimum is approached. We can cure this easily by reducing the jitter as improvement reduces S; figure 7 is a plausible way to do this. It is identical to Fig. 2 , except that the changes to λ_i are multiplied by S, so that the effect of the jitter reduces as S itself reduces. A typical track is shown in fig. 8, which shows the error when $S = \lambda$, and $J(t) = \sin(t - \phi)$. Also shown is a family of tracks, as before.

Fig. 9 shows some two-dimensional tracks. The two axes show λ_1 and λ_2, so that time is not exhibited explicitly. However, since $J_1(t) = \sin t = J_2(t+\frac{\pi}{2})$, the search pattern is periodic and circularly spiral.

Analytically, fig. 7 leads to

$$\frac{d\lambda}{dt} = \gamma S(t)\left[\frac{dJ}{dt} - \alpha J S\right]$$

$$= \gamma\beta\left|\lambda\right|\left[\frac{dJ}{dt} - \alpha\beta\left|\lambda\right| J\right], \text{ if } S = \beta\left|\lambda\right|. \quad (10)$$

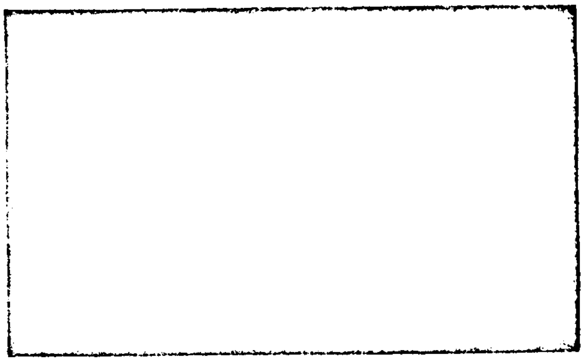
If, as before, $J = \pm 1$. then λ will trace out sections of hyperbolae

$$\lambda = \frac{\lambda([t])}{1 \pm \alpha\beta^2\gamma(t-[t])\lambda([t])} \quad (11)$$

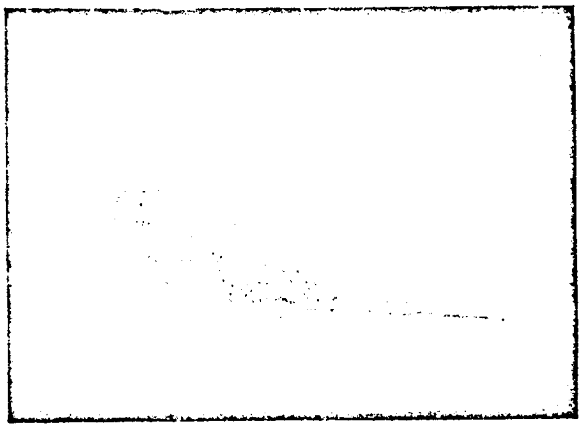
joined with jumps of size $2\gamma\beta\lambda$ (fig. 10). Here it is possible (as in 2.1) that λ might diverge in one half cycle of J to infinity; but see section 4 and fig. 9. On the other hand, for a conical hill if γ happens to be equal to $\frac{1}{2\beta}$, the first jump will land at the optimum.

For a parabolic hill, the behavior is essentially similar.

Note that unless the machine actually attains its optimum, that is $S \to 0$, there will always remain some residual jitter, which may or may not be objectionable.



8a.



8b.

FIG. 8 SCORE vs. TIME OF UNIT MOD 2 HILL-CLIMBER ON CONICAL HILL, WITH SINUSOIDAL JITTER.

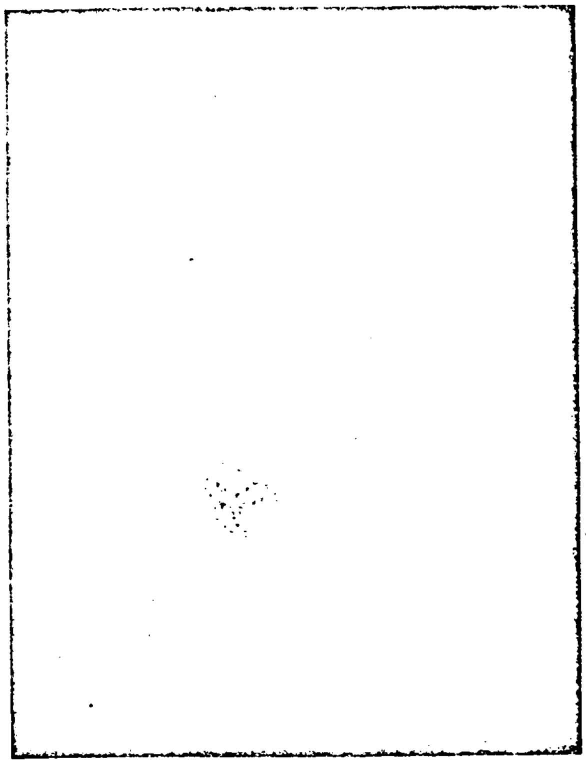


FIG. 9 TRAJECTORIES OF 2-DIMENSIONAL HILL-CLIMBER, MOD 2, CIRCULAR JITTER, CONICAL HILL.

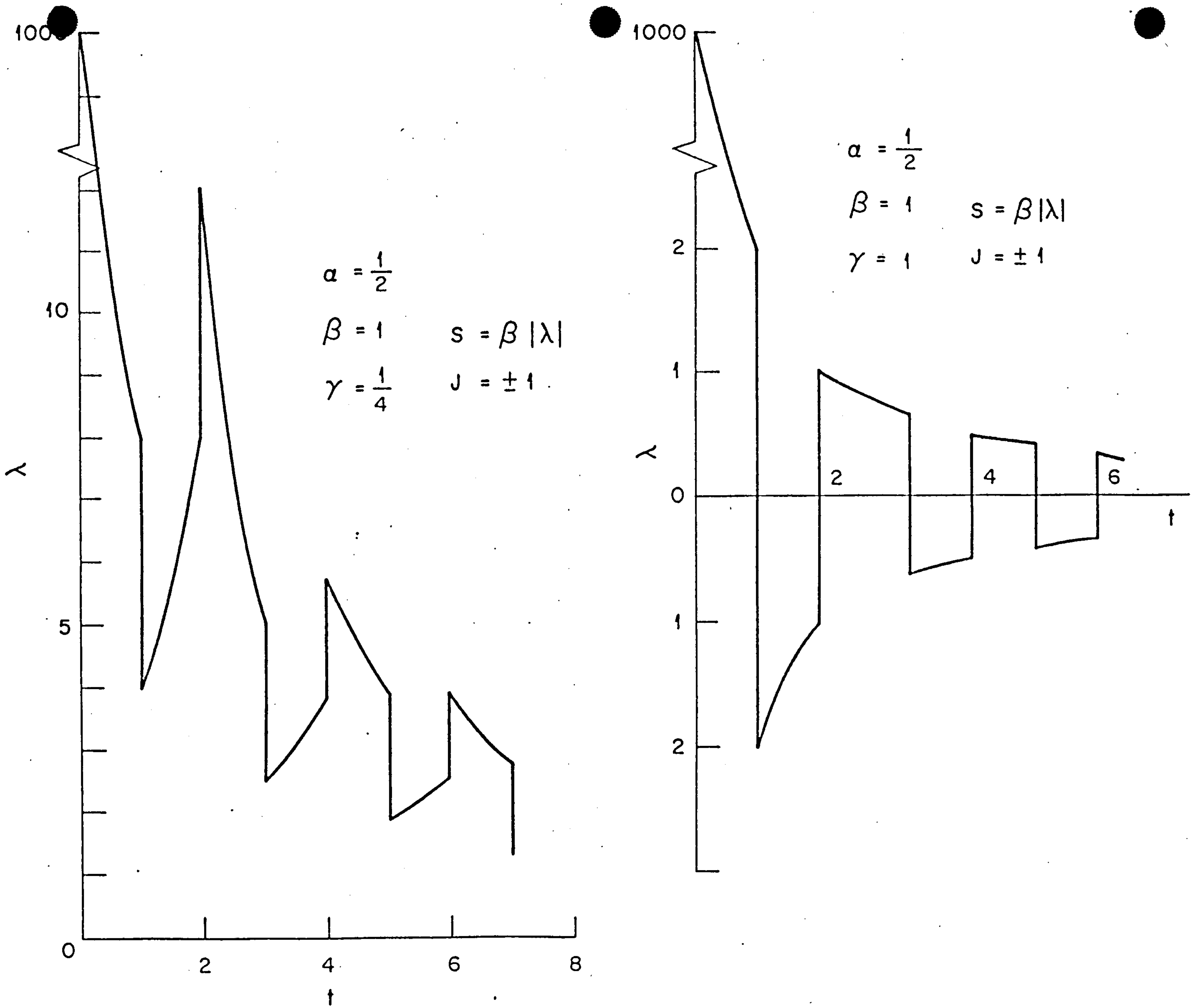
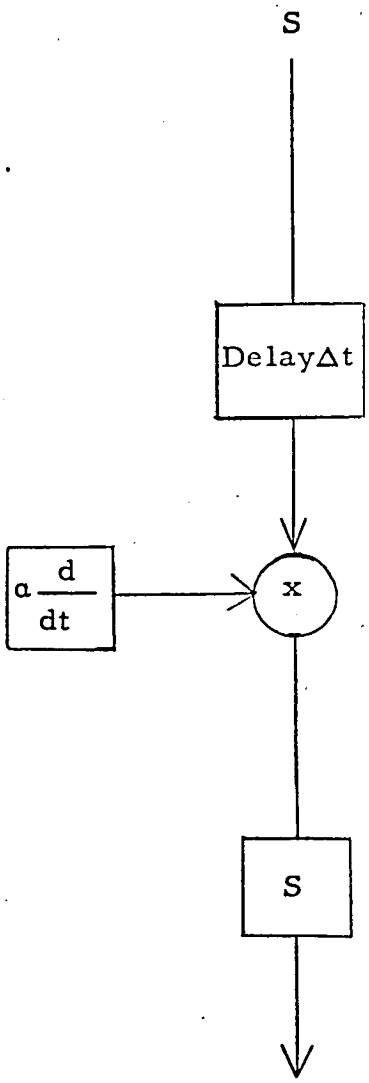


FIG. 10 BEHAVIOR OF MOD 2 HILL-CLIMBER



$$\lambda = a \int_{-\infty}^{t} \frac{dJ}{dt} S(\tau - \Delta t) d\tau$$

FIG. 11 Unit Hill-Climber: MOD 3

2.3 Mod 3 Hill-Climber

Another version leads to an impressively simple block diagram, shown in fig. 11; the addition of the jitter has atrophied completely. It illustrates the fact that increasing a number and then reducing it by the same percentage invariably result in a smaller number.

This version, however, cannot be considered from a merely quasi-stationary point of view. Note that if we eliminate the delay τ, λ is periodic and non-converging for J's with derivatives everywhere, and undefinable if J is discontinuous. In any real machine there is a built-in delay between λ and S; but even so, it may be so small that for speedy convergence one must insert a larger one.

Examples of the behavior of this hill-climber is shown in fig. 12.

Analytically, fig. 11 leads to:

$$\frac{d\lambda}{dt} = a\, J(t)\, S\,(t - \Delta t)\ . \qquad (12)$$

Using the same example as before, that is, $S = \beta|\lambda|$, put $J = \sum_{0}^{\infty} (\delta(t-2n) - \delta(tn+1))$.[1] λ will converge exponentially to the optimum, losing at each pair of jumps

$$4a^2\beta^2; \qquad \text{if } a\beta \leq \tfrac{1}{2}$$

or

$$4a\beta - 4a^2\beta^2; \qquad \text{if } \tfrac{1}{2} \leq a\beta \leq 1 \qquad (13)$$

as a fraction of its magnitude. If $a\beta > 1$, λ will diverge. This is a fundamental flaw in the otherwise rather elegant arrangement of fig. 11. But see section 4.

---

(1) That is J consists of delta functions of unit weight of alternating signs at integral values of t.

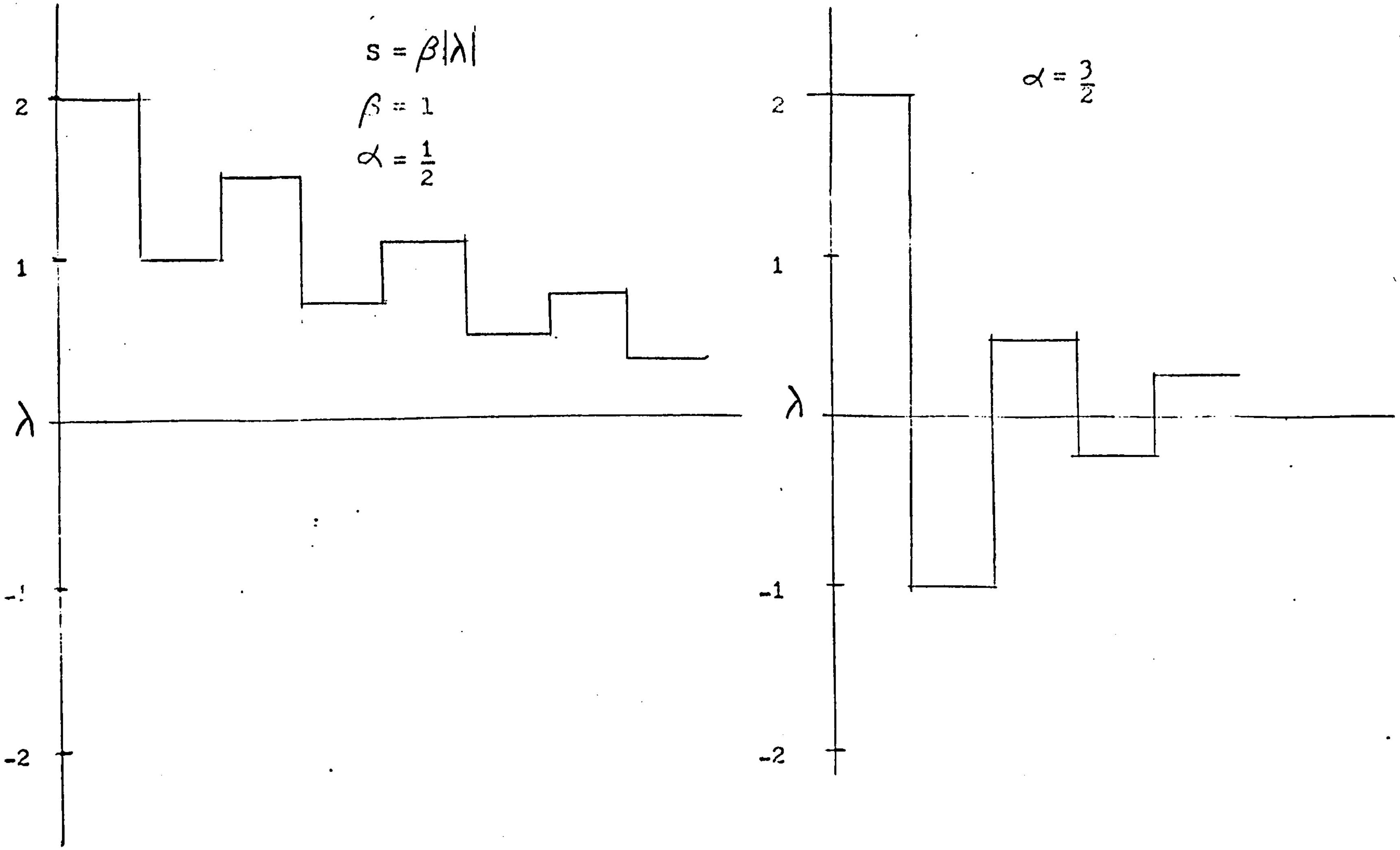
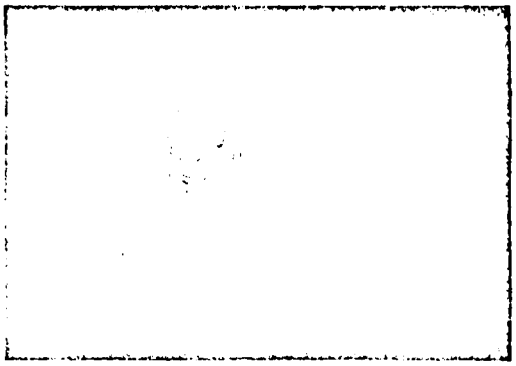
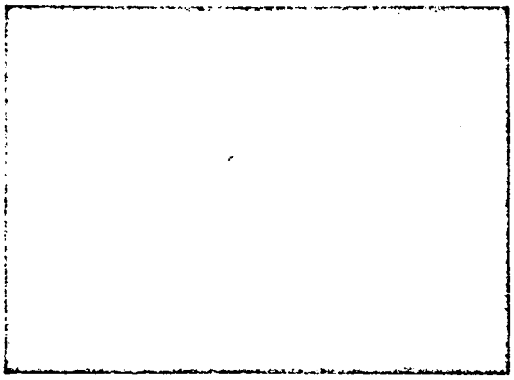


Fig. 12 Behavior of MOD 3 Hill-Climber

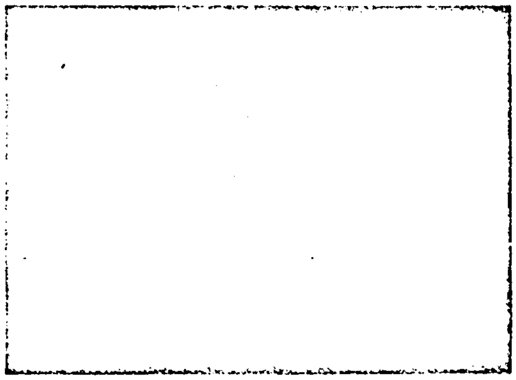
$\alpha = 0.05$



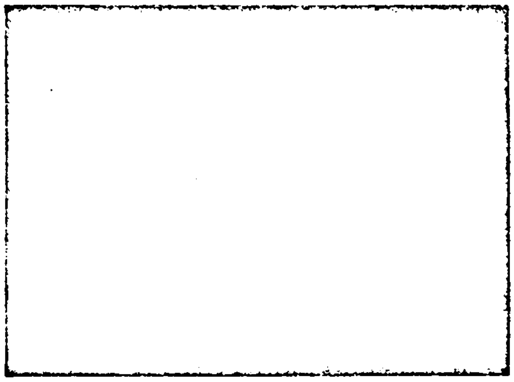
$\alpha = 0.1$



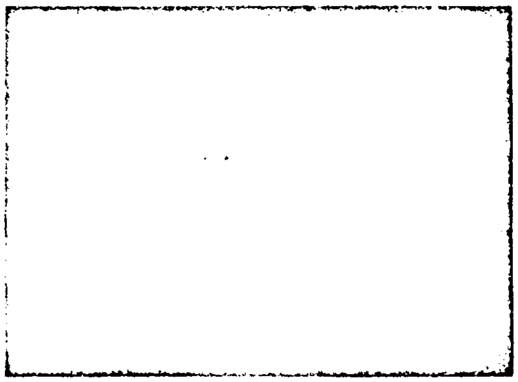
$\alpha = 0.15$



$\alpha = 0.25$



$\alpha = 0.35$



$\alpha = 0.6$

FIG. 13. MOD.1 2-DIMENSIONAL HILL-CLIMBERS CLIMBING CONICAL HILL WITH GAINS $\alpha = 0.05$, 0.1, 0.15, 0.25, 0.35, AND 0.6

3.1 Gains α of the Hill-Climbers

Our previous arguments have generally relied on quasi-stationary behavior. In fact, of course, gains may be usefully raised so that convergence is much more rapid. Fig. 13 shows a set of tracks of a Mod 1 Hill-Climber with different gains climbing a two-dimensional conical hill. Fig. 14 shows the same for a Mod 2 Hill-Climber and fig. 15 for a Mod 3 one.

It is clear that higher rather than lower gains cause faster convergence, at least in Mod 1 and Mod 2 units. But note the possibilities of divergence for steep parabolic hills (section 4). Furthermore, high gains must be carefully used in many-dimensional problems (section 5). In such problems, however, one might profitably use hierarchical gain control (section 6); but we have not checked this experimentally.

Note that pure conical hills are not the real problems we are eventually after. Rather must we expect complex and noisy contours, and turbulent interactions among the variables (but see section 8).

A crucial question - one might say the crucial next question - is what set of optimizing problems is best solved by our parallel hill-climbers and what by such sequential ones. We hope to answer this in part experimentally in the future. (But see the last paragraph of section 8).

$\alpha = 0.05$

$\alpha = 0.1$

$\alpha = 0.15$

$\alpha = 0.25$

FIG. 14. MOD. 2 2-DIMENSIONAL HILL-CLIMBERS ON CONICAL HILL WITH GAINS $\alpha = 0.05, 0.1, 0.15, 0.25$.

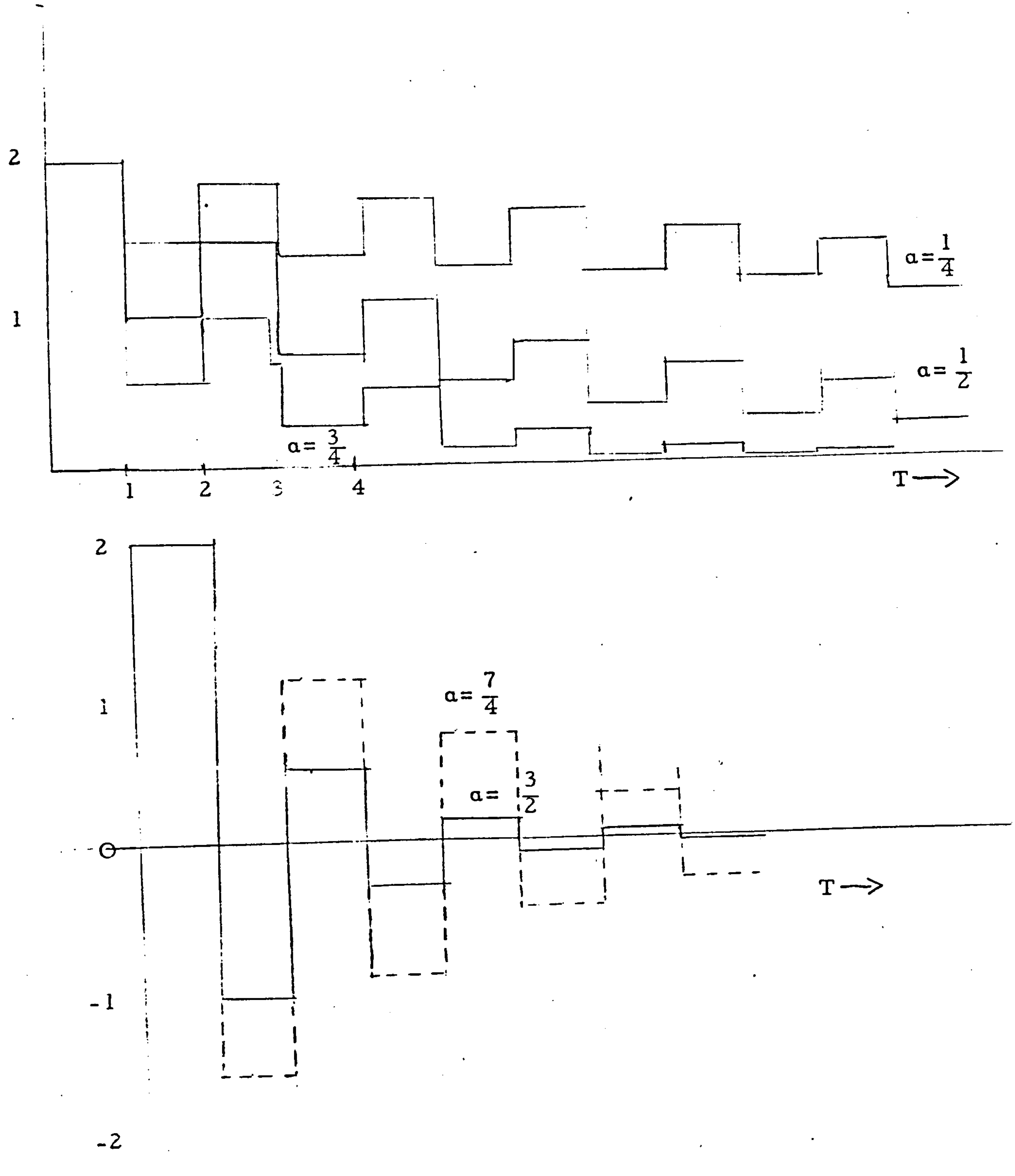


Fig. 15 MOD 3 1-Dimensional Hill-Climbers on Conical Hill with Gains $a = 0.25,\ 0.5,\ .75,\ 1.25,\ 1.5$.

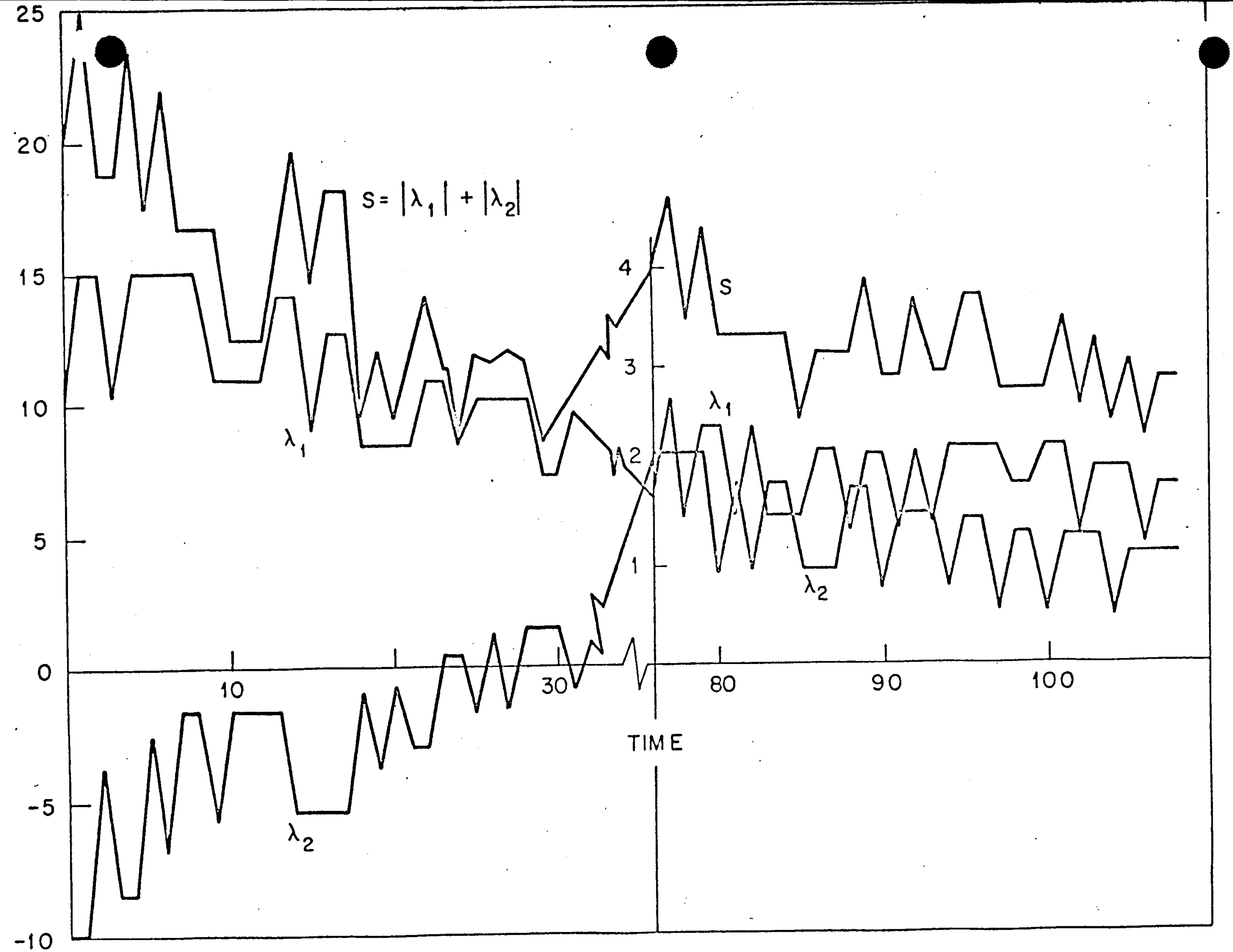


FIG. 16 BEHAVIOR OF MOD. 3 2-D HILL-CLIMBER WITH RANDOM BINARY JITTERS.

3.2 Jitter Functions $\{J_i\}$ of the Hill-Climbers

As equations (5) and (4) declare, the $\{J_i\}$ should be orthogonal and have mean 0. The choice of the $\{J_i\}$ is a central point in the optimization problem.

In fact (4) is not strictly necessary, though the mean $\overline{J_i}$ should exist. For Mod 1 hill-climbers, a biassed J_i does little harm beyond displacing the stable point by the amount $\overline{J_i}$. For Mod 2 hill-climbers, the convergence will be merely hindered but not prevented. For Mod 3 hill-climbers, the question is more complicated.

In general, though, equation 4 does not represent a very strait restriction.

The actual choice of generation of the $\{J_i\}$ must be largely governed by practical matters. A square wave is most sensitive --- in the sense that for a given maximum excursion it has the maximum average excursion --- but sine waves are often easier to generate. This is not invariably true, however, as the experimental example of section 5 illustrates.

An obvious choice is $\sin n(t \pm \frac{\pi}{4})$, i.e., the usual trigonometric sequence. Orthogonal square waves may be generated from counters or shift registers in many obvious ways. As mentioned in section 2, a sample of random noise should suffice[(1)], and it does --- a two dimensional Mod 3 exemplary track is shown in fig. 16. In general, though, random J_i for Mod 3 hill-climbers will threaten divergence unless the duty cycle of the J_i is kept low. In fig. 16, the duty cycle is 50 %.

By manipulating the jitter functions we can at will transform our parallel hill-climber into sequential ones. For example, first set $J_i = 0$ $i > 1$, and set $J_i = \pm 1$ until λ_1 has been optimized; and continue similarly with the others. It

(1) Using a different one for each J_i !

will be necessary to repeat the process until no further improvement is attainable.

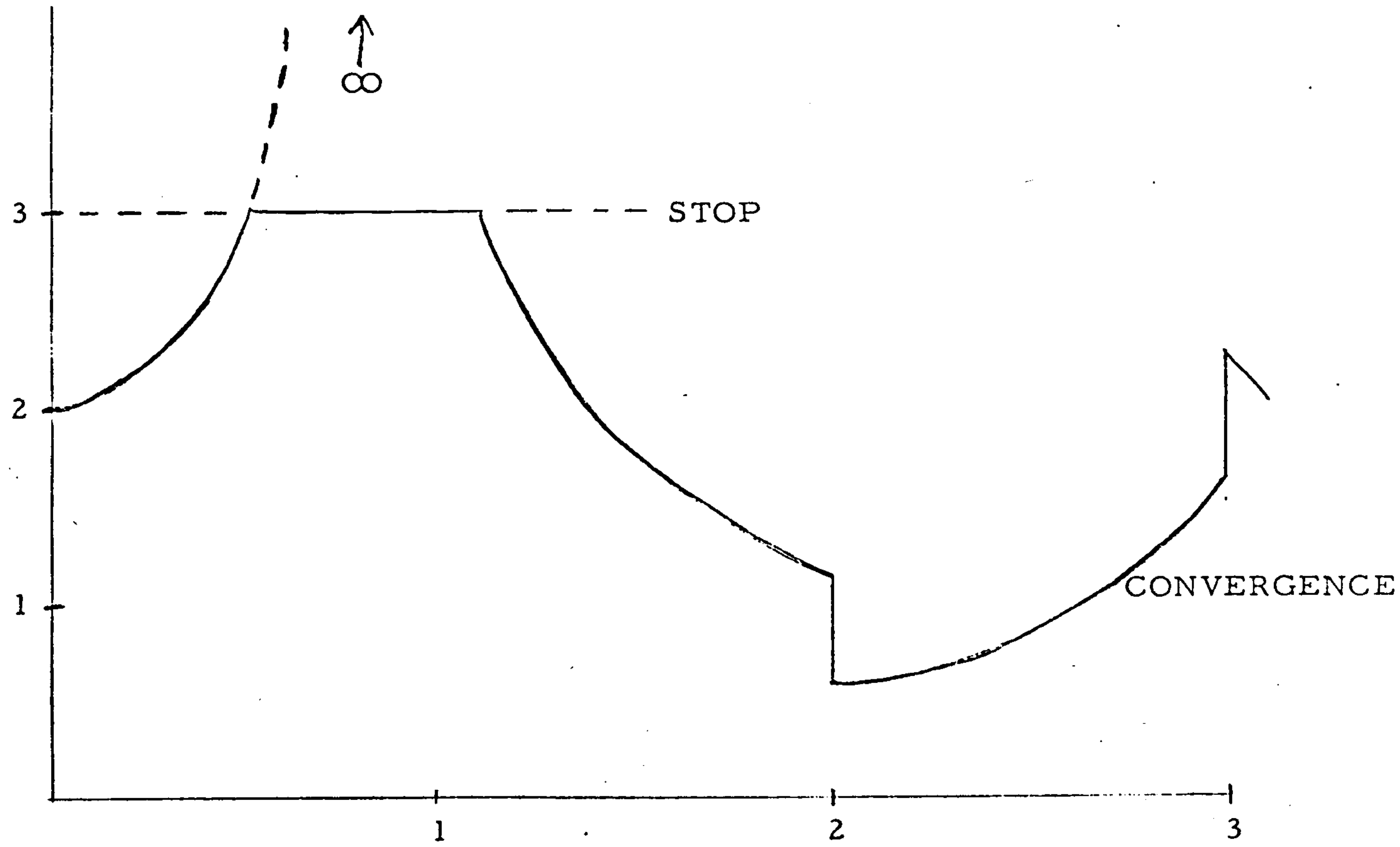


Fig. 17 MCD 1 Hill-Climber Restrained From Momentary Divergence by Stop

4. Divergence and Its Cure

The Hill-Climbers, Mod 1, 2 and 3, all have the disadvantage that on certain steep slopes they may be subject to behavioral singularities or divergence. In practice there are usually restraints of finitude, in which a runaway variable jams itself temporarily against a stop or upper bound of position or velocity, and recovers when its jitter next changes sign; such an example is illustrated in fig. 17.

For Mods 1 and 2, instead of relying on possibly unknown bounds to variables, therefore, one can add an artificial stop to a Mod 2 unit as shown in fig. 18. For large S this behaves like Mod 1 of fig. 2, and for small like Mod 2 of fig. 7. It is important to set the $\{M_i\}$ large enough so that they are rarely used.

For Mod 3 units, no such remedy will cure divergence, which is of a rather different kind, depending only on the gain and slope (fig. 19 and eq. (13)) and diverging not catastrophically but inexorably.

For Mod 3 units, therefore, we must be careful to set the gains a_i low enough to avoid divergence, or we must use special diagnostic techniques to find those a's that are too large (which is rather repugnant to our approach), or we must resort to hierarchical gain control (section 6). In any case, it is a serious restriction on the use of these Mod 3 units.

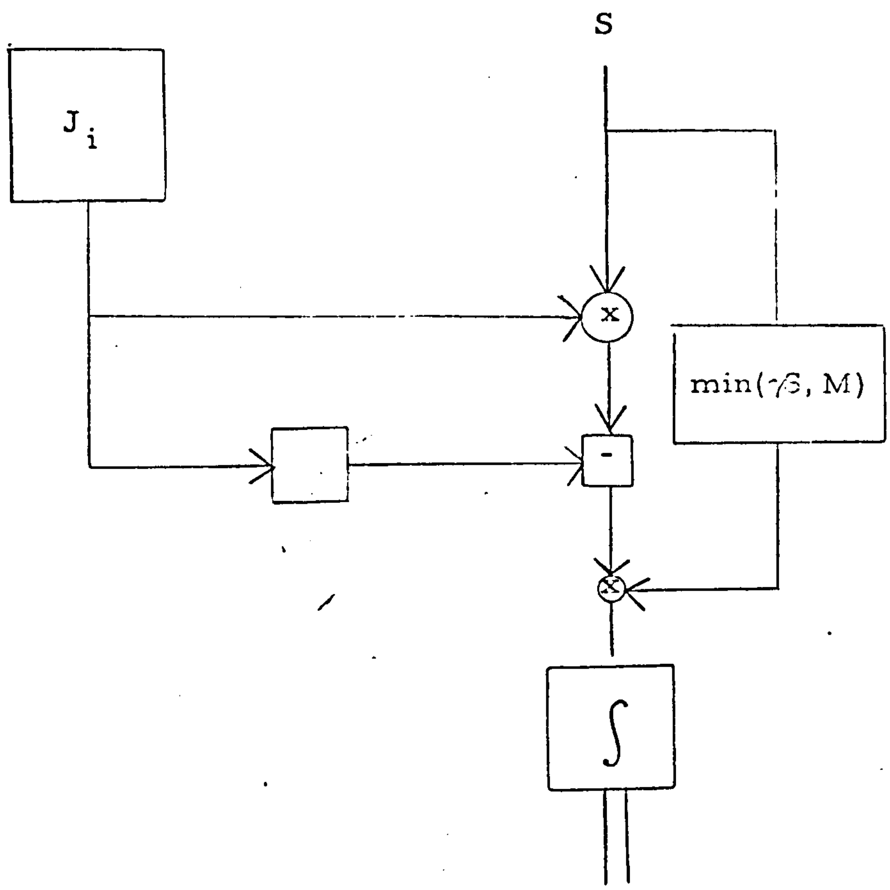


FIG. 18 MOD 2 Hill-Climber with Stop for Divergence

5. Performance of a Multitude

5.1 Speed of Convergence: The Argument

It is important to consider how the system will behave with increasing N, the number of variables or dimensions. For the volume of the space grows exponentially, and so must any uniform sampling method. Such an exponential growth, if it has to be reflected in amount of either hardware or computation time, would rapidly make optimization impractical.

Fortunately, the length of an improvement path need not grow nearly so quickly with dimension. And in general, one needs to take only N sample points to get an idea of the gradient in one's neighborhood. Let us regard, for a moment, the device to be optimized as though it were a communication channel.

In order to compare systems with different dimensions we must have some natural progression. We may suppose, for instance, that we are dealing with a sequence of machines where the i-th machine has i input channels and one output channel. They are all to be made of the same kind of hardware: we assume that their output channels have a certain uniform bandwidth (rate of response) W. Equivalently, we may put no restriction on the machines but instead assume that our detection and correlation devices are all the same and have a uniform input bandwidth limitation W. This seems like a natural restriction since our goal is to use more and more copies of the same equipment to solve more and more difficult optimization problems.

Now suppose that each of the N independent jitter functions $J_i(t)$ contributes roughly equal parts to the output, and that interaction terms are small. Then if we look for one of these signals in the output, regarding the rest of the output as noise, the $\frac{\text{signal + noise}}{\text{noise}}$ ratio is like

$$\frac{1 + N - 1}{N - 1} = \frac{N}{N - 1} \qquad (13)$$

Then the (maximum) rate of information, per dimension, is, following Shannon (1948), of the order of W $\log(\frac{N}{N-1})$ bits per second. Assuming that we need about one bit for each dimension (e.g., whether there is a rise or fall in that direction) we must wait until the channel has been able to transmit this one bit. This requires a time of the order of

$$\frac{1}{W \log N/N-1)} \text{ or about } \frac{N}{W}$$

seconds. Note that this is the total time since all of the N variable controls can operate simultaneously. Thus the averaging times seem to grow linearly with N, which agrees with the fact that one need only look at N independent sample points. Of course, this is merely a heuristic argument which suggests a lower bound for the time; each of our assumptions is optimistic.

Once we have found the way to move, the rate of ascent may then be expected to vary roughly with $N^{\frac{1}{2}}$. On the other hand the path lengths will increase roughly with $N^{\frac{1}{2}}$, so that we may still expect that the time needed for optimization will vary linearly with N, providing the topography becomes no less 'smooth'.

We cannot make this kind of general argument more precise without restricting ourselves to a particular ensemble of hills. But it does seem reasonable to interpret it to mean that if a certain kind of problem with a large number of parameters, can be solved by this kind of iterative improvement procedure, then addition of more variables of the same kind will not cause an inordinate increase in difficulty. This is of heuristic importance in that it can help us discriminate between techniques which will extend smoothly in practice, to large systems and those which will not.

6. Hierarchies

How can we improve our hill-climbing techniques? Efficiency is related to speed of ascent, but the definition must be relative to the ranges of hills to be climbed. We might define efficiency by converting the problem into one of tracking: measure the mean error, or mean square error, while the hill itself is carried through a smooth path representative of the hill ensemble. The available parameters $\{J_i\}$, α_i, etc., could then be controlled themselves by a second-level hill-climber. We might thus start on a hierarchy of optimizers.

However the hierarchy is assembled, it seems obvious that the upper layers must in some sense have longer time constants than the lower ones. Classically, loop instability arises from a controlled element's taking longer to respond than the controlling element to correct.

The most direct application might be gain control for a large assembly: the variables' gains α_i could be controlled jointly by one hill-climber, or severally by individual hill-climbers, trying to optimize, not S, but $\frac{dS}{dt}$. For the dependency of S on the λ's may change during the course of the optimizations; since the information about this is available in the estimates of the $\frac{dS}{d\lambda_i}$'s, it would seem reasonable to make use of it.

7. The Local Peak Problem

When the hill-climbers come to a halt, what assurance do we have that the system controlled really has been "optimized"? Very little, in general. All we are assured is that the system is (very probably) at a local extremum-- a point near which there is no improvement. It might be at a local peak, or it might have reached a level plateau. If we know in advance what the optimal value is, e.g., as in the bridge-balancing problem where one is searching for a zero value, then one can tell whether the peak is the right one or not. But what should be done in the general case?

The answer depends on the specific ensemble of hills in question. With large N, the volume of the spaces becomes too large to permit any kind of exhaustive search. If, then, the hills of the ensembles are characterized by really acutely isolated extrema, we can do almost nothing with our techniques. Such a situation can obviously obtain, for example, in cryptanalysis.

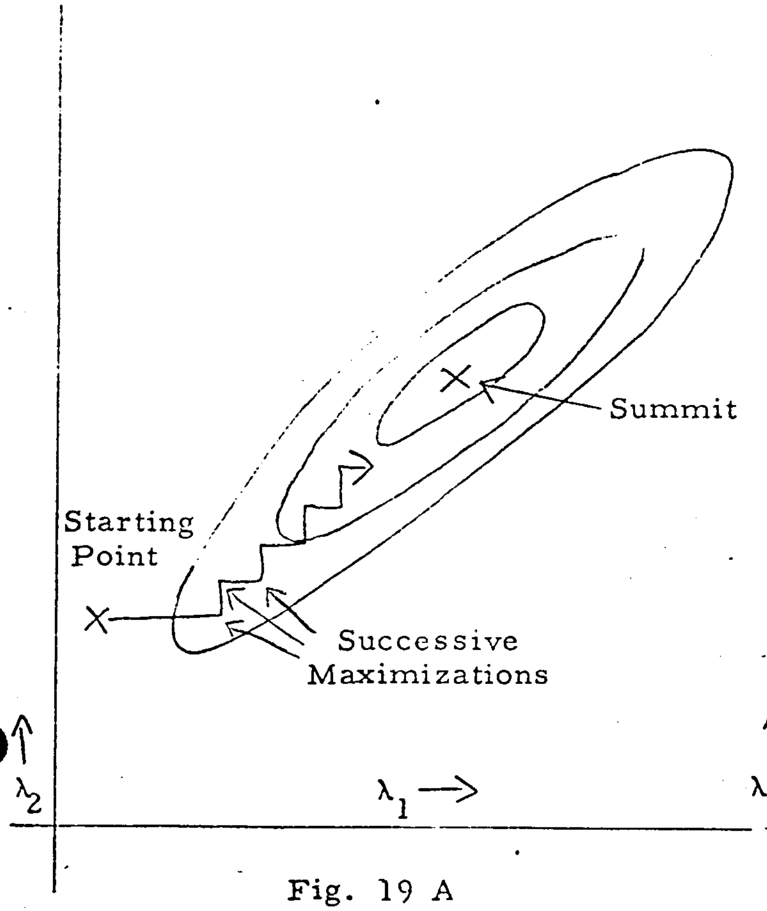


Fig. 19 A

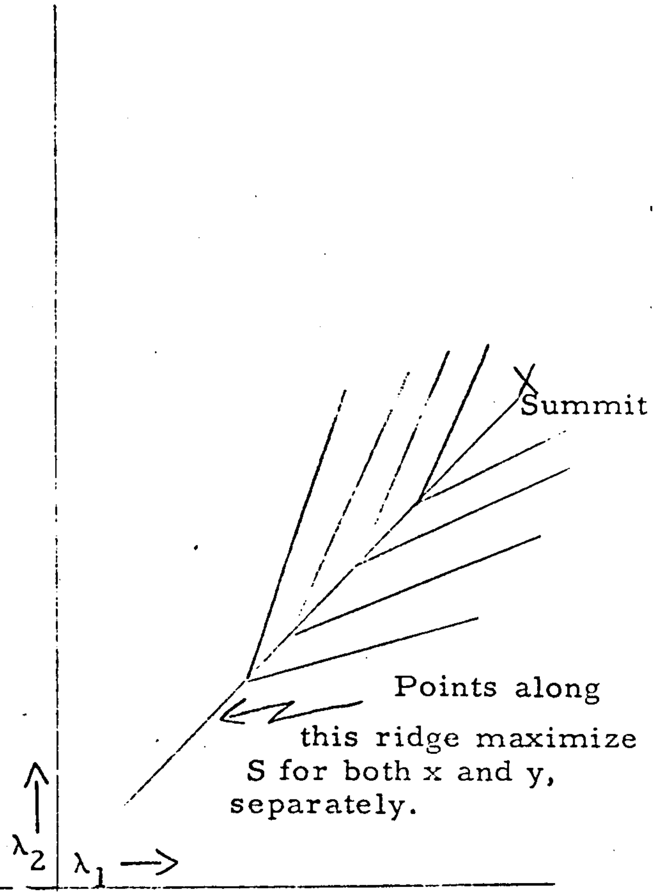


Fig. 19 B

Fig. 19 Score Contours for Sliding Balance - Real and Extreme

8. Interactions of Variables

In spite of our hopeful protestations of independence among the variables, they are in fact often coupled. Suppose that the influence of two variables is channeled through their sum, or, more generally, through

$$f(\lambda_1) + g(\lambda_2) \quad .$$

Then λ_1 and λ_2 will contentedly stabilize at any pair of values such that that expression is itself at an optimum.

That is not true if the influence of the variables is reflected in their products. Suppose that $S = \left| M - \lambda_1 \lambda_2 \right|$. Then the optimum will settle at $\lambda_1 = \lambda_2 = \pm \sqrt{M}$ [1]. On the other hand if $S = \left| M - \frac{1}{\lambda_1 \lambda_2} \right|$ the hill-climbers are unstable, one variable approaching 0 and the other flingling itself out to infinity, until it is restrained by some finite aspect of the system (see also section 4). This instability does not affect the arrival at the summit; rather the summit itself is a horizontal ridge along which the system slides.

Sliding Balance is a term used to describe the difficulty of alternately optimizing two variables whose inter-relation may be shown in fig. 19 A; it is commonly met in balancing certain electrical bridges. The solid lines are contours of S, and the dotted line is the trajectory followed by successive alternate optimizations. In fig. 19 B successive optimization never will succeed in moving up along the ridge. Parallel hill-climbers will have no difficulty with such a terrain, and illustrate one distinct advantage over sequential techniques.

---

(1) We are discussing hill-climber units with decaying jitter, i.e., Mod 2 or Mod 3. Residual jitter causes λ_1 and λ_2 to optimize at a value slightly less than M .

9. General Discussion

9.1 Applicability

We have been discussing the problem of optimizing controls, especially when there may be many of them, in order to minimize some single output. We have shown three examples of units which may be used and described some of their properties.

It is important to distinguish when it is appropriate to use them and when it is not. Some outputs must be optimized at the middle of their range - e.g., AFC controls, cruising speed controls, shipborne steering gear - so that error signals can show the existence size and sign of the error, and hence of the deviation needed to correct. These cases servo-mechanism theory is well equipped to handle.

But in many problems that is not the case, and the effect of deviating a control in one direction may be expected even to change sign as other controls are changed. Then the hill-climbers of this paper are applicable.

9.2 Precision and Accuracy

It will be clear that the hill-climbing capabilities of our units are in no wise dependent upon the precise realization of the operations shown in the block diagrams. They will in fact work well so long as the operations are approximated only very roughly. We have already (section 3.2) declared that little harm is done if J_i does not have mean precisely 0.

Generally we must merely require that actual operations preserve the monotonicity of the ideal ones. Detailed examination of the deterioration caused by unreliable or inprecise components is beyond our scope here. Our first experimental Mod 1 hill-climber unit worked not observably worse when the switching relay failed to switch rather more than 10 % of the time.

We believe that the success of complicated organizations of many interacting parts must rely on such a phenomenon.