## Tools for Teaching

**Intermodular Description Sheet**: UMAP Unit 522

**Title**: UNCONSTRAINED OPTIMIZATION

**Authors**
- Joan R. Hundhausen
- Robert A. Walsh
  Dept. of Mathematics
  Colorado School of Mines
  Golden, CO 80401

**Math Field**: Second-semester calculus

**Applications Field**: Gradient searches, optimization

**Target Audience**: Students in courses that introduce or apply the notion of gradient.

**Abstract**: This unit introduces Gradient Search Procedures, with examples and applications. Students are introduced to the use of computational algorithms, basic optimization theory, and how to find successive approximations to extreme points.

**Prerequisites**: Acquaintance with elementary partial differentiation, chain rules, Taylor series, gradients, and vector dot products.

**Related Units**: The Gradient and Some of Its Applications (Unit 431)

© 1986 COMAP, Inc.
All rights reserved.
Unconstrained Optimization

Joan R. Hundhausen
Department of Mathematics
Colorado School of Mines
Golden, Colorado 80401

Robert A. Walsh
1203 1/2 Washington Ave. #4
Golden Colorado 80401

Table of Contents

1. INTRODUCTION ................................................. 1
2. GRADIENT SEARCH PROCEDURE .......................... 2
3. A CLOSER LOOK AT GRADIENT SEARCH ................. 3
   3.1 The Objective Function Must Decrease ............. 4
   3.2 The Directions of Successive Gradients .......... 5
4. EXAMPLES OF THE GRADIENT SEARCH PROCEDURE ... 6
   4.1 A Function Whose Level Curves are Ellipses .... 7
   4.2 The Gradient Corridor ......................... 11
   4.3 An Example Possessing Circular Level Curves . 13
5. SPEEDING THINGS UP! ........................................ 14
6. A MORE COMPLICATED OBJECTIVE FUNCTION AND
   SOME PRACTICAL ADVICE ................................. 16
   6.1 The Importance of a Good Start .................. 16
   6.2 Slow Convergence: Discussion .................. 19
   6.3 Other Considerations ............................... 21
7. CONCLUDING REMARKS ................................. 21
8. REFERENCES ................................................... 22
9. OPTIONAL ADDITIONAL EXERCISES ......................... 22
10. SPECIFIC APPLICATIONS .................................. 23
11. MODEL EXAM ................................................ 26
12. HINTS AND ANSWERS TO EXERCISES ..................... 26
13. ANSWERS TO MODEL EXAM ................................ 29
    Acknowledgement ........................................... 30
MODULES AND MONOGRAPHS IN UNDERGRADUATE
MATHEMATICS AND ITS APPLICATIONS PROJECT (UMAP)

The goal of UMAP was to develop, through a community of users and developers, a system of instructional modules in undergraduate mathematics and its applications to be used to supplement existing courses and from which complete courses may eventually be built.

The Project was guided by a National Advisory Board of mathematicians, scientists, and educators. UMAP was funded by a grant from the National Science Foundation and is now supported by the Consortium for Mathematics and Its Applications, Inc. (COMAP), a nonprofit corporation engaged in research and development in mathematics education.

COMAP STAFF

- Solomon A. Garfunkel
- Laurie W. Aragon
- Philip A. McGaw
- Kathy Hollowell
- Mary A. Keating
- Brian Sterling
- Ann Marie Morgan

Executive Director, COMAP
Business Development Manager
Production Manager
Production Artist
Editorial Assistant
Distribution Manager
Administrative Assistant

UMAP ADVISORY BOARD

- Steven J. Brams
- Layron Clarkson
- Donald A. Larson
- R. Duncan Luce
- Frederick Mosteller
- George M. Miller
- Walter Sears
- Arnold A. Strassenburg
- Alfred B. Wilcox

New York University
Texas Southern University
SUNY at Buffalo
Harvard University
Harvard University
Nassau Community College
University of Michigan Press
SUNY at Stony Brook
Mathematical Association of America

The Project would like to thank Judith Elkins of Sweet Briar College, Peter F. Ash of Kennedy-King College, and Carroll O. Wilde of the Naval Postgraduate School for their reviews, and all others who assisted in the production of this unit.

This unit was field-tested and/or student reviewed in preliminary form by Michael B. Ward of Bucknell University and Richard A. Alo of Lamar University and has been revised on the basis of data received from these sites.

This material was prepared with the partial support of National Science Foundation Grants No. SED-8007731 and No. SPS-8304192. Recommendations expressed are those of the author and do not necessarily reflect the views of the NSF or the copyright holder.
1. Introduction

Consider the equation of a surface \( z = f(x, y) \) in 3-dimensional space. In general, when we wish to find a local extreme point \((x^*, y^*)\), (maximum, minimum, or saddle point) of a differentiable surface, \( z = f(x, y) \), we set \( \nabla f(x, y) = \vec{0} \) and try to obtain all possible solutions for the unknowns \( x \) and \( y \). If a candidate for an extreme point is extracted from the system

\[
\begin{align*}
\nabla f(x, y) = \vec{0} \quad \rightarrow \quad \begin{cases} 
    f_x(x, y) &= 0 \\
    f_y(x, y) &= 0
\end{cases} 
\end{align*}
\]

(and only too often the equations are non-linear and the calculations are rather involved) then we subject the candidate \((x^*, y^*)\) to a further test involving second partial derivatives in order to determine whether the local extreme point represents a maximum, minimum, or saddle point of \( z = f(x, y) \).

In this module we explore a technique for approximating solutions of System (A) when it is difficult to solve by the usual arithmetic approaches. In particular, we shall describe and illustrate the bare outlines of such a procedure, called Gradient Search. The Gradient Search procedure usually only approximates an extreme point, but is very useful, because often for practical reasons we wish to achieve only a reasonable approximation to the point \((x^*, y^*)\).

You should recall that the Gradient vector points locally in the direction of the greatest rate of increase of the surface, while the negative of the Gradient vector points in the direction of greatest rate of decrease of the surface. Realizing this, you have already some intuitive notion about what direction should be followed (at least locally) in seeking the extreme point \((x^*, y^*)\). The material that follows will provide both validity and precision to your intuition!

Keeping these above intentions as a guide, and considering the constraints on the length of this presentation, we have chosen to exemplify the technique of Gradient Search by treating several types of carefully selected surfaces \( z = f(x, y) \). These surfaces are not necessarily related to any specific area of application for the method of Gradient Search, but were selected because they exhibit certain properties of the method with great clarity.

In Section 10 we describe a collection of specific topics for application of an elementary optimization technique such as the Gradient Search. Each of those topics in itself could be the subject of extensive exploration and discussion of various advantages and
disadvantages (or possible modifications) of the Gradient Search Procedure. Indeed, in some cases, one would quickly be led to consider more sophisticated techniques of optimization. Pursuit of any of these areas of application could be the subject of an additional project for the class, or for independent study.

You may wish to peruse Section 10 to get some idea of the specific applications before you proceed with the main body of this module; such a preliminary reading may provide motivation for this study as well as incentive for further exploration.

2. Gradient Search Procedure

We first choose to present a condensed "recipe" (algorithm) for the procedure, assuming that we are searching for a point of minimization \((x^*, y^*)\), with detailed explanations and examples to follow in subsequent sections. (Modifications of the procedure when we seek a maximum point \((x^*, y^*)\) will be mentioned shortly.)

The function to be minimized, \(z = f(x, y)\), which represents a surface \(S\) in 3-dimensional space, is called the Objective Function; \(f(x, y)\) has a local minimum at \((x^*, y^*)\) if there is some open region \(R\) containing \((x^*, y^*)\) such that \(f(x, y) \leq f(x^*, y^*)\) for all other \((x, y)\) in \(R\). Local minima correspond to the low points on the surface \(S\); see Fig. 1.

To find an approximate local minimization point \((x^*, y^*)\) for \(f(x, y)\):

1. Select a point \((x_0, y_0)\) arbitrarily, or at a minimum approximately known or suspected, from, say, physical reasons.

2. Compute \(\nabla f(x_0, y_0)\) and \(f(x_0, y_0)\)

3. Form the function of a single variable \(t\), as follows:

\[
\psi(t) = f(x_0 + f_x(x_0, y_0)t, y_0 + f_y(x_0, y_0)t); \ t < 0.
\]

This function \(\psi(t)\) is called the Davidon function. (Davidon was a corporate research worker; for further information, you might check the text by Wilde and Beightler, listed in Section 8 of this module.) Find its minimum.

4. Call \(t^*\) the value for which the Davidon function is minimized.

5. Using that value of \(t^*\), form the new coordinates

\[
x_1 = x_0 + f_x(x_0, y_0)t^* \quad y_1 = y_0 + f_y(x_0, y_0)t^*.
\]
6. To see whether appropriate progress toward the minimum has been achieved, check whether \( f(x_{i1}, y_{i1}) \) is significantly less than \( f(x_{0}, y_{0}) \). If so, go to 1 and repeat all steps with \((x_{i1}, y_{i1})\) replacing \((x_{0}, y_{0})\). If not, stop.

![Diagram of a 3D surface with points labeled as \((x^*, y^*)\) and \((x^{**}, y^{**})\).](image)

Figure 1. The objective function \( z = f(x, y) \) represents a 3-D surface \( S \), with low points (local minima) at \((x^*, y^*)\) and \((x^{**}, y^{**})\).

**Brief Comments on the Above...**

In Step 3, the actual minimum is usually difficult to find; in practice, we numerically step a small negative distance \( t \) a few times, and if the objective function is decreased each time, use the final \( t \) (as \( t^* \)), save \( t^* \) and proceed.
In Step 6, you must decide the tolerance, a priori. A decision as to whether the process should be repeated will usually involve consideration of the effort (and expense) necessary to carry out all of the steps of the procedure as compared to the possible improvement of the objective function. Also, the physical interpretation of the problem may indicate that even a small decrease in the objective function (say .10) would be considered a significant improvement and further refinement undesirable. Along this line, the actual cost of obtaining an improvement in the objective function might be prohibitive. Suppose, for example, that in a practical application of this procedure, the objective function \( z \) depended upon not only two but perhaps fifteen variables and that it represented a "cost function" measured in units of tens of thousands of dollars per month; in this case, the actual cost of obtaining an improvement (decrease) in the objective function, even though that cost be quite large, might not be prohibitive! The decision on whether to proceed with another iteration of the algorithm or not is often a difficult one for the practitioner; he weighs many factors, such as cost of computer time, man-hours, possible advantages to be gained, known capability of a competing firm, production deadlines, etc. you can name some others which might influence his or her decision.

In general, the Gradient Search never converges to the exact \((x^*, y^*)\) after \( n \) steps and moreover converges very slowly to \((x^*, y^*)\) after the first few cycles (steps 1-6). Certain methods are available for "accelerating" convergence; one of these will be illustrated in detail in Section 5.

In searching for a maximum point, replace each usage of the word "minimum" in the above outline to "maximum"; also the inequalities in Steps 3 and 6 should be reversed. Even these relatively simple modifications may be avoided by realizing that a maximum point for the function \(f(x, y)\) is a minimum point for the function \(-f(x, y)\). (I prove this!)

3. A Closer Look at Gradient Search

In this section we establish two important theoretical aspects of the Gradient Search Procedure that will be useful as you follow through the examples and comments presented in later sections. The first indicates the way the objective function changes with \( t \), while the second shows the geometric relationship between successive directions of the gradient.
3.1 The Objective Function Must Decrease

If the parameter $t$ is appropriately chosen small and negative, say $\hat{t}$, the objective function must decrease as we move in the direction of the Gradient.

Proof:
Since $z = f(x, y)$ is assumed to be a differentiable function, we can expand $f(x, y)$ in a Taylor series about the point $(x_0, y_0)$:

$$f(x_1, y_1) = f(x_0, y_0) + f_x(x_0, y_0) (x_1-x_0)$$

$$+ f_y(x_0, y_0) (y_1-y_0)$$

$$+ \text{terms containing } (x_1-x_0)^2, (y_1-y_0)^2$$

$$\text{or } (x_1-x_0) (y_1-y_0).$$

By definition, from step 5, $x_1-x_0 = f_x(x_0, y_0) \hat{t}$ and $y_1-y_0 = f_y(x_0, y_0) \hat{t}$, and replacement of these expressions in the above formula yields

$$f(x_1, y_1) - f(x_0, y_0)$$

$$= |f_x(x_0, y_0)|^2 \hat{t} + |f_y(x_0, y_0)|^2 \hat{t}$$

$$+ \text{terms containing } (\hat{t})^2 \text{ and higher derivatives of } f(x, y)$$

$$\text{evaluated at } (x_0, y_0).$$

$$\vdots$$

$$\vdots$$

$$= \| \nabla f(x_0, y_0) \|^2 \hat{t} + \text{terms containing } (\hat{t})^2.$$

Now if $|\hat{t}|$ is small and the higher derivatives of $f$ are small, then the terms containing $\hat{t}$ will dominate those containing $(\hat{t})^2$, and thus $f(x_1, y_1) - f(x_0, y_0) < 0$ for small negative $\hat{t}$. (Note that for positive $t$, we have $f(x_1, y_1) - f(x_0, y_0) > 0$.)

Note:
The concept of the directional derivative at a particular point, denoted by $df/ds$, is also useful here to justify the decrease in the objective function. Recall that

$$\frac{df}{ds} = \nabla f \cdot \vec{u}$$
where \( \vec{u} \) is a unit vector in the s-direction. We can see that \( \frac{df}{ds} \) will be negative (i.e., the function \( f \) will decrease) if the direction \( s \) is chosen opposite to that of the gradient at a particular point. (The choice of a negative parameter \( t \) in the gradient search procedure has the effect of reversing the direction of the gradient).

However, we prefer the Taylor series approach presented above because it better illustrates the involvement of the parameter \( t \), which is central to the Gradient Search Procedure.

We emphasize that the Gradient Direction is normal to a level curve of the surface. Recall also that a move in the direction opposite to that of the Gradient vector at a point is a move in the direction of the steepest descent on the surface at that point.

### 3.2 The Directions of Successive Gradients

If \( f^* \) yields the exact minimum of the Davidson function \( \psi(t) \) in step 3, then

\[
\nabla f(x_0, y_0) \cdot \nabla f(x_1, y_1) = 0;
\]

i.e., successive gradients are perpendicular to each other. (See Fig. 2.)

![Diagram](image)

Figure 2. Solid arrows mark successive gradient directions starting from the point \((x_0, y_0)\); broken arrows indicate the path followed by the Gradient Search Procedure.
Proof:
The Davidon function in step 3 can be written as \( \psi(t) = f[u(t), v(t)] \), where \( u(t) \equiv x_0 + f_x(x_0, y_0)t \) and \( v(t) \equiv y_0 + f_y(x_0, y_0)t \). We note also (from Step 5) that \( u(t^*) = x_1 \) and \( v(t^*) = y_1 \); so the condition that \( t^* \) be a critical point for \( \psi(t) \) becomes

\[
\frac{d\psi}{dt} \bigg|_{t=t^*} = \frac{d}{dt} f[u(t), v(t)] \bigg|_{t=t^*} = 0.
\]

By the chain rule, the latter condition can be written

\[(B) \quad \frac{\partial}{\partial u} f(u, v) \frac{du}{dt} + \frac{\partial}{\partial v} f(u, v) \frac{dv}{dt} \bigg|_{t=t^*} = 0.
\]

But \( \frac{du}{dt} = f_x(x_0, y_0) \) and \( \frac{dv}{dt} = f_y(x_0, y_0) \); also

\[
\frac{\partial}{\partial u} f(u, v) \bigg|_{t=t^*} = \frac{\partial f}{\partial x} (x, y) \bigg|_{x=x(t^*) = x_1, \ y=y(t^*) = y_1} = f_x(x_1, y_1),
\]

and

\[
\frac{\partial}{\partial v} f(u, v) \bigg|_{t=t^*} = \frac{\partial f}{\partial y} (x, y) \bigg|_{x=x(t^*) = x_1, \ y=y(t^*) = y_1} = f_y(x_1, y_1)
\]

so that \((B)\) becomes

\[
\frac{\partial}{\partial u} f(u, v) f_x(x_0, y_0) + \frac{\partial}{\partial v} f(u, v) f_y(x_0, y_0) = 0 \quad \text{or} \quad \nabla f(x_0, y_0) \cdot \nabla f(x_1, y_1) = 0.
\]

4. Examples of the Gradient Search Procedure

4.1 A Function Whose Level Curves Are Ellipses

As a first example, we now apply the Gradient Search to the minimization of \( z = (1/2)(x^2 + 2y^2) \). The objective function is non-negative for all \( x \) and \( y \), and it is easy to see that the point of minimization is \((x^*, y^*) = (0, 0)\). We shall, however, use this somewhat simple example because it illustrates the computational aspects while providing some valuable geometric insight into the procedure. (See Fig. 3.)
Figure 3a: The surface $S: z = \frac{1}{2} (x^2 + 2y^2)$.

Figure 3b: Some level curves for the function $z = \frac{1}{2} (x^2 + 2y^2)$.

We begin with the (somewhat crude) approximation $(x_0, y_0) = (1, 1)$. While you might surmise that one iteration of the Gradient Search would enable us to reach $(x^*, y^*)$, we shall see that this is not the case.

Example 1.
Using $z = (1/2)(x^2 + 2y^2)$, let us follow the recipe which was presented in Section 2.

1. $(x_0, y_0) = (1, 1)$.

2. $\nabla f(x_0, y_0) = x\hat{i} + 2y\hat{j}\bigg|_{(1,1)} = 1\hat{i} + 2\hat{j}$;  
   $f(x_0, y_0) = \frac{3}{2}$.

3. $\psi(t) = \frac{1}{2} [(1+t)^2 + 2(1+2t)^2]$. The condition $\dot{\psi}(t) = 0$ yields the equation $(1+t) + 4(1+2t) = 0$ or $5 + 9t = 0$, and $t^* = -\frac{5}{9}$.

4. $t^* = -5/9$, in this case the exact minimizing point for $\psi(t)$. 

8
5. New coordinates are \((x_1, y_1) = \left(1 + 1 \left(-\frac{5}{9}\right), 1 + 2 \left(-\frac{5}{9}\right)\right)\)
\[= \left(\frac{4}{9}, -\frac{1}{9}\right).\]

6. \(f(x_1, y_1) = (1/2) (x_1^2 + 2y_1^2) = 1/9 < 3/2 = f(x_0, y_0)\).

Having achieved a decrease in the objective function, we repeat the process, with \((x_1, y_1) = (4/9, -1/9)\) replacing \((x_0, y_0)\).

At this point, we suggest that to gain confidence, you might perform Steps 2-6 of the algorithm yourself as an exercise. Use the starting point (4/9, -1/9), and you can check your results with ours at each step. (In fact, we suggest you use this approach throughout the module; instead of interrupting the flow of the text with many short and relatively simple exercises, we leave it to your discretion to practice and duplicate some of the computations.)

2'. \(\nabla f(x_1, y_1) = \left(\frac{4}{9}, -\frac{2}{9}\right)\), \(f(x_1, y_1) = \frac{1}{9}\).

3'. \(\psi(t) = \frac{1}{2} \left[ \left(\frac{4}{9} + \frac{4}{9}t\right)^2 + 2 \left(-\frac{1}{9} - \frac{2}{9}t\right)^2 \right] \leq \frac{1}{2(81)} \left[10(1+t)^2 + 2(1+2t)^2\right].\)

The condition \(\dot{\psi}(t) = 0\) becomes \(\frac{1}{2(81)} \left[32(1+t) + 8(1+2t)\right] = 0\)

or \(48t + 40 = 0\),

\(t^* = -\frac{5}{6}\),

where again this value of \(t\) represents an exact point of minimization for \(\psi(t)\).

4'. \(t^* = -\frac{5}{6}\).

5'. New coordinates are
\((x_2, y_2) = \left[\frac{4}{9} + \frac{4}{9} \left(-\frac{5}{9}\right), -\frac{1}{9} - \frac{2}{9} \left(-\frac{5}{9}\right)\right] = \left(\frac{2}{27}, \frac{2}{27}\right).\)

6'. \(f(x_2, y_2) = \frac{1}{2} (x_2^2 + 2y_2^2) = \frac{2}{243} < \frac{1}{9} = f(x_1, y_1).\)
We interrupt the algorithm to note that \((x_2, y_2)\) is a multiple of \((x_0, y_0)\). To compare the locations of the approximating points \((x_0, y_0)\), \((x_1, y_1)\) and \((x_2, y_2)\), refer to Fig. 4. Upon the next iteration, you will find that \((x_2, y_2) = (2/27)(4/9, -1/9)\) is a multiple of \((x_1, y_1)\). In fact, it can be shown that \((x_{m+2}, y_{m+2})\) is a multiple of \((x_m, y_m)\), \(m \geq 0\). (See Exercise 2.)

![Diagram](image)

Figure 4. Successive approximating points for Example 1.

Proceeding with one more iteration of the algorithm, we find

\[2^* \quad \nabla f(x_2, y_2) = \frac{2}{27} \hat{i} + \frac{4}{27} \hat{j}, \quad f(x_2, y_2) = \frac{2}{243}.\]

\[3^* \quad \psi(t) = \frac{1}{2} \left[ \left( \frac{2}{27} + \frac{2}{27}t \right)^2 + 2\left( \frac{2}{27} + \frac{4}{27}t \right)^2 \right] \]

\[= -\frac{2}{(27)^2} \left[ (1+t)^2 + 2(1+2t)^2 \right].\]

The condition \(\psi'(t) = 0\) becomes \(1 + t + 4(1+2t) = 0\)

\[5 + 9t = 0\]

\[t = -\frac{5}{9}.\]
4''. \( t^* = -\frac{5}{9} \).

5''. New coordinates are

\[
(x_3, y_3) = \left[ \left( \frac{2}{27} - \frac{2}{27} \left( \frac{5}{9} \right), \left( \frac{2}{27} - \frac{4}{27} \left( \frac{5}{9} \right) \right) \right] \\
= \frac{2}{27} \left[ \left( 1 - \frac{5}{9}, 1 - \frac{10}{9} \right) \right] = \frac{2}{27} \left( \frac{4}{9} - \frac{1}{9} \right),
\]

which is a multiple of \((x_1, y_1)\). Also, you may show that \( f(x_3, y_3) < f(x_2, y_2) \).

### 4.2 The Gradient Corridor

![Diagram](attachment:image.png)

Figure 5. Similar triangles formed by the approximating points.

We shall now take advantage of the pattern that has evolved, and anticipate the result of Exercise 2 to make a geometric construction of the path followed by the points \((x_0, y_0), (x_1, y_1), \ldots, (x_m, y_m)\) as they approach the true minimum value for \( f(x, y) \), viz., \((0,0)\). You will note that the result of Exercise 2 shows that \( x_{m+2}/x_m = y_{m+2}/y_m = p > 0 \), which is equivalent to the similarity of the two triangles in Fig. 5. Thus we see that the odd-subscripted points are collinear; apparently the single line which they define, which is directed along the hypotenuse of each triangle, leads directly to the true minimum. (The same could be said of the even-subscripted points.) Let us find the equation of the straight line joining odd-subscripted points, and then do the same for the even-subscripted points.
The straight line through the points \((x_1, y_1)\) and \((x_2, y_2)\) has equation

\[
\frac{y-y_1}{y_2-y_1} = \frac{x-x_1}{x_2-x_1} \quad \text{or} \quad y = \frac{1}{4} x.
\]

Similarly, the line through \((x_0, y_0)\) and \((x_2, y_2)\) is \(y = x\). Fig. 6 depicts these lines graphed in the \(xy\) plane. The shaded space between these lines is called the "Gradient Corridor" and you can trace the zig-zag pattern by the successive approximating points \((x_{mp}, y_{mp})\) as we approach the minimum value \((0,0)\). Note also the perpendicularity of the paths connecting the successive points!

"...the convergence...is assured only if we are sufficiently close to an extreme point..."

![Diagram](image)

Figure 6  The lines \(y = x\) and \(y = -(1/4)x\) which bound the "gradient corridor" for Example 1.

**Exercise**

1. Calculate the slope of the path connecting \((x_1, y_1)\) to \((x_2, y_2)\) and the slope of the path connecting \((x_2, y_2)\) to \((x_3, y_3)\) and show that the paths are perpendicular. Why is this true?

Example 1 is particularly instructive because of the geometric insight it contributes; you can indeed predict the path followed by the approximating points as they "home in" towards the true minimum. This straight-line corridor pattern is characteristic of any objective function of the form \(f(x,y) = x^2 + \mu y^2\), whose level curves are ellipses. \((\mu > 0)\).

For level curves having the general equation \(f(x,y) = C\), the Gradient Corridor will be meaningful only for points very close to a local maximum or minimum point \((x_0, y_0)\). Again we emphasize that the convergence of the Gradient Method along the "nearly" straight bounding lines of the Gradient Corridor is assured only if we are sufficiently close to an extreme point.
Exercise

2. Show that, for functions of the type \( z = f(x,y) = x^2 + \mu y^2 \) (take \( \mu > 1 \)), the Gradient Corridor lies between \( 2 \) straight lines; i.e., show that the approximating point \((x_{m+2}, y_{m+2})\) is a multiple of the approximating point \((x_m, y_m)\), for \( m \geq 0 \) in the sense that \( x_{m+2} = -\beta x_m, y_{m+2} = \gamma y_m \), where \( \beta > 0 \). (Hint: Start with the point \((x_m, y_m)\); use the procedure as illustrated in this example successively to find \((x_{m+1}, y_{m+1})\) and \((x_{m+2}, y_{m+2})\); further guidance will be found in the ANSWERS section.)

A few more examples and exercises will increase your understanding of the use of the Gradient Search Procedure and illustrate both its power and its limitations under specific circumstances.

4.3 An Example Possessing Circular Level Curves

The previous example illustrated the optimization of a function whose level curves were ellipses. (See Fig. 3b.) When the level curves are circular, the Gradient Search Procedure is even more effective. In fact, a single application of the procedure, starting from any initial point, leads to the exact location of the optimum:

Example 2.
Find the maximum value of the objective function \( z = f(x,y) = -x^2 - 2x + 2y \). By completing the square, you can verify that \( z = 5 - (x+1)^2 - (y-1)^2 \), so its maximum value is 5 which occurs at the point \((-2, 1)\). Because of the nature of this function (its level curves are circles), you will find that only one application of the gradient search procedure leads directly to the optimum point.

Exercise

3. Following the procedure outlined in Section 2 start with \((x_0, y_0) = (4, -4)\) and show that
a) \( \nabla f(4, -4) = \langle -12, 10 \rangle \); (use the direction vector \(-6\hat{i} + 5\hat{j}\));
b) The Davidson function \( \psi(t) = (4-6t)^2 - (4+5t)^2 - 4(4-6t) + 2(-4+5t) = -61t^2 + 122t + 56;\)
c) \( t^* = 1;\)
d) \((x_1, y_1) = (-2, 1) = (x^*, y^*)\), the maximum point for \( z \).

Here the “gradient corridor” in Exercise 3 degenerates to a single straight line; (see Fig. 7). It is only fair to mention that this rarely happens in practice!
“Gradient Search Procedure can be both necessary and effective...”

Figure 7. “The gradient corridor” reduces to a straight line when level curves are circular.

**Question:**
What would happen were we to attempt to continue the Gradient Search Procedure using the “approximating point” (-2,1)?

In summary, we note that an objective function with circular or elliptical level curves can be optimized quite simply by using the analytical tools of the calculus; no approximation method such as the Gradient Search Procedure needs to be considered. What we are trying to do is to illustrate the application of the method with simple examples and to provide you with some geometric insight. In the cases of more complicated objective functions, for which the Gradient Search Procedure can be both necessary and effective, the level curves are neither circular nor elliptical (see Section 6), and it is unlikely that any simple transformation can make them so. But having observed and understood the action of the Procedure for these simple cases, you will be better able to appreciate how it operates in the more complicated cases. This remark also applies to the method we illustrate in the next Section.

5. Speeding Things Up!

Having noted the progress of the iterations down the gradient corridor in Example 1 as well as the dramatic leap to the
optimum in Example 2, you might well ask if it wouldn’t be more efficient to compute a few approximation points \((x', y')\), use them to determine the boundary lines of the gradient corridor, and then slide down the corridor to reach the optimum. In the next example we show that your conjecture is well-founded! The process to be illustrated is called “acceleration to the optimum point.”

**Example 3.**

Consider the objective function \(z = f(x, y) = (1/2) \left( x^2 + 2xy + 2y^2 \right) - 3x - 2y + 6\), and begin the search at the point \((2, 1)\). Following the usual procedure, we have

\[
\nabla f(x, y) = (x + y - 3) \hat{i} + (x + 2y - 2) \hat{j};
\]

\[
\nabla f(2, 1) = 2 \hat{j}
\]

\[(x_1, y_1) = (2-0t, 1+2t),\]

\[
\psi(t) = \left(1/2\right) [2^2 + 2(2)(1+2t) + 2(1+2t)^2] - 6 - 2(1+2t) + 6
\]

\[= 4t^2 + 4t + 3\]

\[
\psi'(t) = 8t + 4; t^* = -1/2
\]

\[(x_1, y_1) = (2, 0).\]

\[
\nabla f(2, 0) = (2-3t) \hat{i} + (2-2t) \hat{j}
\]

\[= -1 \hat{i}\]

\[(x_2, y_2) = (2-t, 0)\]

\[
\psi(t) = \left(1/2\right) [(2-t)^2] - 3(2-t) + 6
\]

\[= (1/2)t^2 + t + 2\]

\[
\psi'(t) = t + 1; t^* = -1
\]

\[(x_2, y_2) = (3, 0).\]

Let us now find the direction of one of the bounding lines of the gradient corridor, and we shall optimize along that direction rather than compute another gradient direction. The vector directed between points \((x_0, y_0)\) and \((x_2, y_2)\) is \((3-2) \hat{i} + (0-1) \hat{j} = \hat{i} - \hat{j}\), but since the objective function can be expected to decrease with negative \(t\) (Section 3.1) we form the vector \(t(-\hat{i} + \hat{j})\) to indicate the new direction for optimization:

\[(x_3, y_3) = (3-t, 0+t);\]

\[
\psi(t) = \frac{1}{2} \left[ (3-t)^2 + 2(3-t)t + 2t^2 \right] - 3(3-t) - 2t + 6 = \frac{t^2}{2} + t + \frac{3}{2}.
\]
\[ \psi(t) = t + 1; \quad t^* = -1. \]

\((x_2, y_2) = (4, -1).\]

Exercise

4. Using the methods of calculus, verify that \((4, -1)\) is indeed the minimum point for the objective function of Example 3.

6. A More Complicated Objective Function and Some Practical Advice

Our final example will illustrate some difficulties which might accompany application of the Gradient Search Procedure to more complicated objective functions. As we work through it, we will make several digressions; these will serve to deepen your understanding of the method and to expand your horizons with respect to the concept of optimization in general.

6.1 The Importance of a Good Start

Example 4.

Let \( z = f(x, y) = x^2 + y^3 - 3xy. \) It is known that this function has a saddle point at \((0, 0)\) as well as one local minimum at \((9/4, 3/2)\) where \( f(3, y) = -1.6875. \) Suppose we are seeking this local minimum. (See Fig. 8.) First we shall illustrate what can happen if the first approximation to the local minimum point is not chosen sufficiently close to that minimum point itself.

Now \( \nabla f(x, y) = (2x - 3y)i + (3y^2 - 3x)j. \) Suppose \((x_0, y_0) = (0, -1); \)
then \( \nabla f(0, -1) = 3i + 3j, \) which has the direction of \( i + j. \) Then

\[
(x_1, y_1) = (t, -1 + t); \\
\psi(t) = t^2 + (t - 1)^2 - 3t(t - 1); \\
\psi'(t) = 3t^2 + 3(t - 1)^2 - 3(2t - 1); \\
\psi''(t) = 6t + 6; \\
\psi''(t) = 6t + 10. \\
\psi'(t) = 0 \text{ yields } t_1 = \frac{10 - \sqrt{100 - 72}}{6} > 0, \\
t_2 = \frac{10 + \sqrt{100 - 72}}{6} > 0.
\]

The second derivative test indicates that the first of these is a maximum for \( \psi(t), \) the second is a minimum for \( \psi(t), \) but since neither of these values for \( t \) is negative we can only conclude that
\( \psi(t) \) has no local minima for \( t < 0 \). We cannot hope to "home in" on the point \( (9/4, 3/2) \) from this starting location.

Figure 8. A quantitative rendition of the level curves of the function
\[ z = x^2 + \psi^3 \] considered in Example 4, with functional values labeled on the curves. From the poor starting point \((-1,0)\), the gradient heads down the saddle, never to approach the local minimum point; from the better starting point \((1,1)\), the successive gradient paths do approach the local minimum. Note especially the near-elliptical shape of the level curves in the vicinity of the minimum, and that the successive gradients are perpendicular to level curves.

What would happen if we were to persist in following the negative gradient direction in seeking a minimum for \( f(x,y) \)? Suppose a small step (say \( t = -1 \)) were to be taken in that direction, from the starting point \((0,-1)\) to the point \((1, y_1) = (-1, -2)\). Then we would see that while \( f(0, y_0) = -1, f(x, y_1) = -13 \). The choice \( t = -2 \) would yield a still larger negative value \( f(x, y) \), and we begin to realize that the gradient search procedure is not at all effective here, in locating the minimum point \((9/4, 3/2)\). What is happening is that the successive approximating points are following a precipitous path down the saddle! (See Fig. 8.)
We now discuss this "pitfall" briefly. In the previous examples, you have seen how the Gradient Search Procedure is both effective and efficient when applied to functions whose level curves are elliptical. Here we point out that, in the vicinity of an extremum, a well-behaved function (continuous first and second derivatives) will possess level curves which "resemble" ellipses; i.e., the local behavior of even a very complicated but well-behaved function can be expected to resemble that of the simpler functions treated in this module. Thus if we have prior knowledge of the approximate location of a local maximum or minimum, the Gradient Search Procedure will improve our educated guess if we start "close enough" to the point. However, if we are not "close enough" to the local max or min, we cannot always expect the Gradient Search to lead us nearer to this particular local optimum point.

We now try a better choice for the first approximation.

Let \((x_0, y_0) = (1,1)\); then \(\nabla f(1,1) = -\mathbf{i}\), and \(f(1,1) = -1\).

Then \((x_1, y_1) = (1-t, 1)\).

\[
\psi(t) = (1-t)^2 + 1 - 3(1-t);
\]

\[
\psi'(t) = 2(1-t), t^* = -\frac{1}{2}.
\]

Then \((x_1, y_1) = \left(\frac{3}{2}, 1\right)\); and \(f\left(\frac{3}{2}, 1\right) = -\frac{5}{4} < -1\),

so that some progress toward the minimum has been achieved. Another iteration of the gradient search procedure will be instructive.

\[\nabla f(x_1, y_1) = -\mathbf{i}, \text{ which has the direction of } -\mathbf{i}.\]

Then \((x_2, y_2) = \left(\frac{3}{2}, 1-t\right)\).

\[
\psi(t) = \frac{9}{4} + (1-t)^2 - \frac{9}{2}(1-t)
\]

\[
\psi'(t) = -3(1-t)^2 + \frac{9}{2}
\]

\[
\psi'(t) = 0 \text{ yields 2 roots: } t_1 = 1 + \sqrt{\frac{3}{2}}, t_2 = 1 - \sqrt{\frac{3}{2}}
\]

only one of which is negative.

Since \(\psi'(t) = \psi'(1-t)\) and \(\psi'(1-(\sqrt{3}/2)) = \psi'(\sqrt{3}/2) > 0\), \(\psi(t)\) does not have a local minimum value, and we conclude that \(t^* = 1 - \sqrt{3}/2\).
Then
\[(x_2,y_2) = \left( \frac{3}{2}, \sqrt{\frac{3}{2}} \right) \text{, and } f \left( \frac{3}{2}, \sqrt{\frac{3}{2}} \right) = \frac{9}{4} - 3 \sqrt{\frac{3}{2}} = -1.424.\]

The next iteration yields
\[\nabla f(x_2,y_2) = 3\left(1 - \sqrt{\frac{3}{2}}\right)i, \text{ or the direction of } -\hat{i}.\]

Then \((x_3,y_3) = \left( \frac{3}{2} - t, \sqrt{\frac{3}{2}} \right).\)
\[\psi(t) = \left( \frac{3}{2} - t \right)^2 + \left( \sqrt{\frac{3}{2}} \right)^2 - 3 \left( \frac{3}{2} - t \right) \sqrt{\frac{3}{2}}.\]
\[\psi'(t) = -2\left( \frac{3}{2} - t \right) + 3 \sqrt{\frac{3}{2}}; \text{ setting } \psi'(t) = 0,\]

we find the root
\[t^* = \frac{3}{2} - \frac{3}{2} \sqrt{\frac{3}{2}}.\]

Since \(\psi''(t^*) = 2, \) we do have a local minimum for \(\psi(t)\).

Then \((x_3,y_3) = \left( \frac{3}{2} \sqrt{\frac{3}{2}}, \sqrt{\frac{3}{2}} \right) = \left( \frac{3}{4} \sqrt{6}, \frac{\sqrt{6}}{2} \right) = (1.82, 1.2),\)

and \(f(x_3,y_3) = -1.53.\)

Note that progress toward the true minimum is quite regular.
(Recall that the true minimum value of \(f(x,y)\) is \(-1.6875\), achieved at \((x,y) = (9/4, 3/2).\) Such is not always the case. This is an opportune moment for further discussion of slow convergence — a deficiency which can sometimes be remedied.

6.2 Slow Convergence; Discussion

The Gradient Search Procedure, while yielding substantial initial progress toward the desired minimum, often appears to converge very slowly. Alas, this fact is true in general for gradient search patterns. Perhaps some acceleration procedure could be used to advantage. In the Model Exam, we ask you to perform the type of acceleration illustrated in Example 3. However, we note here that other acceleration schemes might be more efficient. Also
these could be used in combination with the Gradient Search Procedure to achieve a more efficient optimization algorithm. We mention the Newton-Raphson procedure as one example, and refer you to the text "Foundations of Optimization" by D. Wilde and C. Beightler.

In actual practice, the applied mathematician is often faced with decisions which are rooted in the realities of economics. If an objective function represents the cost (or the profit) related to some industrial operation, sometimes even a relatively small improvement can be measured in terms of thousands of dollars. Perhaps even a single application of the Gradient Search Procedure (or some other sophisticated optimization procedure) will yield satisfactory (if not optimal) results.

Sometimes, preliminary scaling of a problem may be necessary to achieve a desirable configuration of near-ellipses for level curves. The advantage of the scaling procedure can be easily illustrated by attention to a simple example such as $z = x^2/10^6 + y^2/4$, whose level curves are extremely narrow ellipses. The change of variable $x = 10^3x'$ eliminates the severe distortion; now $z = (x')^2 + y^2/4$ (see Figure 9). Without going into detail, we remark that an extreme discrepancy between magnitude of numerical coefficients in a problem can lead to computational difficulties, such as loss of accuracy due to round-off errors. We also note that in many practical cases it is far from obvious just how to scale appropriately in the vicinity of an extreme point.

In summary, we suggest giving some close attention to two preliminary aspects of a practical problem before applying the gradient search or any other optimization procedure.

a) Try to scale the variables:

b) If at all possible, choose an initial approximating point which is close to the optimum.

Figure 9. The effects of scaling the variables.
6.3 Other Considerations

Whether or not further iterations or further refinements of an optimization procedure should be pursued depends on many factors. In general, a prime consideration would be the amount of computer usage required to generate the successive approximating points. Naturally we have tried to keep the hand computations in this module to a minimum. However, it is not difficult to imagine formidable computations arising in the formulation of the Davidon function $\psi(t)$, its derivative $\psi'(t)$, and/or the determination of those negative roots $t^*$ which are crucial in the process. Generally the equation $\psi'(t) = 0$ is not linear nor even quadratic (unlike our examples) and a numerical procedure for its solution must be initiated. This could be both costly and time-consuming in terms of computer usage; in fact, a routine to compute the gradient itself could be expensive.

In practice the solution may not even be attempted, but instead, a small negative value for $t$ might be selected and tried, based upon the assurance that the objective function must show some improvement when the optimum is pursued in the direction of the negative gradient, (Section 3.1). (Successive gradients cannot, then, be expected to be perpendicular; see Section 3.2). Both intuition and experience serve as guides in these matters; familiarity with the physical characteristics of a problem and an awareness of the approximate location of the optimum can be extremely valuable.

7. Concluding Remarks

The Gradient Search Procedure as outlined here does extend to $n$-dimensional problems. Our intention is that the geometric insight you have gained from an examination of the two-dimensional case will guide your understanding and appreciation of the possibilities of optimization in more general circumstances. The acceleration procedures, however, become relatively more complicated in higher dimensions. For descriptions of both PARTAN and conjugate gradient procedures, we again refer you to the wealth of existing literature on the subject of optimization. (See Section 8, References). Constrained optimization, in which other conditions are placed upon the independent variables, is also a large area we have left untouched.

Finally, we would like to point out that our treatment in this module is strictly in the spirit of an introduction to the topic, and
hope that you will be motivated to explore further. We have mentioned briefly some of the difficulties, and some deciding factors in using the procedures outlined. The important topics of numerical precision and round-off have not been discussed.

Some would call the whole discipline of optimization an art form; others would quickly add that it is fraught with perils and pitfalls. Even though its methods are sometimes time-consuming and expensive, we often have no choice. Each "optimizer" must do the best possible, and will evolve a "bag of tricks" of his own, as he becomes more deeply involved in the subject.

8. References

The following text is well worth reading and you should derive from it some excellent incentive for further exploration of the subject of optimization.


Also our earlier UMAP module "The Gradient and Some of Its Applications" (Unit 431) would provide helpful background information on vector notation, the Gradient itself, level curves, and the chain rule.

Other recommended reference texts:


9. Optional Additional Exercises

Exercises:

5. For Examples 1, 2, and 3, choose an alternate initial approximating point \((x_0, y_0)\) and perform a few iterations of the Gradient Search Procedure. Discuss the "progress" toward the
optimum point. You may also wish to implement the acceleration procedure.

6. In Section 3.1, we presented a rather general proof concerning the (local) decrease of the objective function for negative $t$. Prove directly that for objective functions whose level curves are elliptical in the vicinity of a minimum point, application of the Gradient Search Procedure does effect a decrease. As in Exercise 2, take $z = f(x, y) = x^2 + \mu y^2; \mu > 0, \mu \neq 1$. Use successive points $(x_m, y_m)$ and $(x_{m+2}, y_{m+2})$ which lie along one edge of the Gradient Corridor and show that $f(x_{m+2}, y_{m+2}) < f(x_m, y_m)$. The hints given for Exercise 2 should be helpful.

7. Write a computer program to perform the computations for one of the examples or exercises which we have illustrated in previous sections. Could your program be easily modified to solve similar problems? (Optional)

10. Specific Applications

In this section, we outline a few applications which should be accessible to students of elementary calculus. It is not difficult to list examples which would quickly lead us into the more complex aspects of the field of Operations Research; since that is indeed more than we intend with this module, we present some examples for which you can find correct answers by a direct application of methods of calculus, and suggest that alternatively, you then apply to them the details of the Gradient Search Procedure which you have learned.

1. Consider the problem of solving two non-linear equations for a root $(x^*, y^*)$:

   \begin{align*}
   g(x, y) &= 0, \\
   h(x, y) &= 0.
   \end{align*}

   Following a "function norm" method, we could form the objective function

   $$f(x, y) = |g(x, y)|^2 + |h(x, y)|^2,$$

   and proceed to minimize $f(x, y)$ using the Gradient Search. (The reader should give intuitive reasons why this technique might or might not be reasonable for approximating $(x^*, y^*)$.)
You can see that, depending on the nature of the functions $g(x,y)$ and $h(x,y)$, the formation of the gradient of $f(x,y)$ and the Davidon function could be rather formidable.

**Example:**
Solve the (linear, this time) equations

\[
\begin{align*}
x + y &= 2 \\
2x - y &= 1
\end{align*}
\]

using the “function norm” method; i.e., find $(x^*,y^*)$.

**Solution:**
Set $f(x,y) = (x+y)^2 + (2x-y-1)^2$ and start the Gradient Search at the point $(x_0,y_0) = (0,0)$.

**Example:**
Solve the equations

\[
\begin{align*}
x^2 + y^2 &= 4 \\
x y &= 1
\end{align*}
\]

using the “function norm” method; what are the four choices for $(x^*,y^*)$?

**Solution:**
Set $f(x,y) = (x^2 + y^2 - 4)^2 + (xy-1)^2$ and start the Gradient Search at the point $(x_0,y_0) = (2,0)$.

2. Find the shortest distance from the point $(\bar{x}, \bar{y}, \bar{z})$ to the surface $z = h(x,y)$.

**Example:**
Let the surface $h(x,y) = xy$ and the point $(x,y,z) = (1,0,1)$.

**Solution:**
To find the shortest distance from the point to the surface we form the objective function

\[
f(x,y) = (x-1)^2 + (y \cdot 0)^2 + (\bar{z} - h(x,y))^2
\]

\[
= (x-1)^2 + (y-0)^2 + (1-xy)^2
\]

(which is the square of the usual formula for the distance between two points, with $z$ replaced by $h(x,y)$) and begin the Gradient Search at the point $(x_0,y_0) = (1,0)$. 

24
3. The “Least Squares” Procedure. Suppose we desire to fit a set of data points to a suitable straight line. Find constants \( m \) and \( b \) in the linear equation \( y = mx + b \) such that the \( y \) distance from the graph of \( y = mx + b \) to the pre-assigned data points is minimized locally in a precise way.

**Example:**
Suppose the data points are given by \((0, 1)\), \((2, 3)\), \((3, 6)\); then determine “best” values of \( m \) and \( b \) by minimizing the objective function

\[ f(m, b) = (b - 1)^2 + (2m + b - 3)^2 + (3m + b - 6)^2. \]

**Hint:**
Look up the Least Squares Procedure and find the exact answer by methods of calculus; then compute \( \nabla f(m, b) \) using the starting point \((m_0, b_0) = (1, 2)\). Is the Gradient Search Procedure a good way to find these constants \( m \) and \( b \)? Discuss this.

4. In the text by Hadley and Whitten (see References) you will find an interesting example of an objective function called a “penalty,” or “cost” function. To illustrate, let

\[ z = \frac{1}{3} x^2 + \frac{1}{3} (a-y)^2 + \frac{1}{3} (y+y)^2, \]

where \( z \) is the average annual cost, \( x \) is the quantity of stock ordered at any one time, \( y \) is the number of back-orders that accumulate before an order for more stock is placed; and \( a, b, c, \ldots \) are parameters (constants) of the system. The problem of minimizing cost is to determine \( x > 0 \) and \( y > 0 \) that minimize \( z \). We suggest that you refer to the above-mentioned text for further discussion of problems of this type, and have a computer handy for the calculations!

Finally, in the text by Cooper and Steinberg (see References) you will find a chapter entitled “Search Techniques.” It will be very instructive for you to get an overview of the various search techniques and their merits and disadvantages. There the Gradient Search Procedure is called the “Method of Steepest Descent.” An exercise (number 9) at the end of the Chapter suggests fitting some realistic chemical data to an exponential curve by using several different search techniques, among them the Gradient Search Procedure.
11. Model Exam

a) Find a local maximum for the function \( z = f(x,y) = 3xy - x^2 - y^3 \); begin at the point (3,2), and calculate 2 additional approximating points. (Do not use a calculator here.)

b) Use the direction of the gradient corridor (as formed by the line joining (3,2) = (x_0, y_0) and your second calculated approximating point (x_2, y_2)) to accelerate the progress toward the optimum point.

[Note of encouragement: You will find that if you perform the calculations carefully and do not introduce round-off error by using a calculator, the procedure will "home-in" to the exact point you are seeking!]

12. Hints and Answers to Exercises

Exercises:

1. The slopes of paths joining these pairs of points (in the order mentioned) are \( -1/2 \) and \( 2 \), respectively. The reason for the perpendicularity of the paths lies in part 1 of Section 3.

2. Note that \( \nabla f(x_m, y_m) = 2x_m x + 2y_m y \), and we can use the direction \( x_m x + y_m y \); then

\[
\begin{align*}
    x_{m+1} &= x_m + \lambda x_m = x_m (1 + \mu), \\
    y_{m+1} &= y_m + \mu y_m = y_m (1 + \mu).
\end{align*}
\]

The value of \( \mu \) which minimizes the Davidon function is

\[
    \mu^* = \left[ \frac{x_m^2 + \mu^2 y_m^2}{x_m^2 + \mu^2 y_m^2} \right] \text{ so that the new coordinates become}
\]

\[
\begin{align*}
    x_{m+1} &= x_m \left[ \frac{\mu^2 y_m^2 - \mu^2 y_m^2}{x_m^2 + \mu^2 y_m^2} \right] = y_m \mu^2 \alpha_m, \\
    y_{m+1} &= y_m \left[ \frac{y_m^2 - \mu^2 y_m^2}{x_m^2 + \mu^2 y_m^2} \right] = -x_m \alpha_m.
\end{align*}
\]
Then, finally, \( \psi(\mu) = -\frac{\alpha}{2} \mu^2 - \frac{9}{4} \mu^3 - \mu^4 \)
\(- \frac{9}{4} \mu^3 - \mu^4 \)

which has a maximum at \( \mu^* = \frac{3}{2} \).

Then \((x,y) = \left( \frac{9}{4}, \frac{3}{2} \right) \), which is the optimum point.

(Note that \( \mu^* = \frac{3}{2}, \lambda^* = \frac{1}{2}, \lambda^* = \frac{1}{2(2^2 - \sqrt{3})} \).)

a positive value as would be expected in searching for a maximum. You can now understand that the optimum point would probably not be reached exactly if a calculator had been used, due to round-off errors.)

Acknowledgement

We would like to acknowledge the helpful suggestions regarding applications which were offered by Professors Carroll O. Wilde, Frank D. Faulkner, and Guillermo Owen, all of the Naval Postgraduate School.