Optimization Techniques

MATHEMATICAL METHODS OF OPTIMIZATION

Nonlinear Programming

Optimization Techniques

INTRODUCTION TO NONLINEAR PROGRAMMING

- Many business problems can be modeled only with nonlinear functions.
- Problems that fit the general linear programming format but contain nonlinear functions are termed nonlinear programming (NLP) problems.
- There is no single method that can solve general nonlinear programmming
- Solution methods are more complex than linear programming methods.
- Often difficult, if not impossible, to determine optimal solution.
- Solution techniques generally involve searching a solution surface for high or low points requiring the use of advanced mathematics.

Optimization Techniques

Optimal Value of a Single Nonlinear Function Basic Model

Profit function, Z, with volume independent of price:

$Z = vp - c_f - vc_v$

where v = sales volume

p = price

c_f = fixed cost

c_v = variable cost

Add volume/price relationship:

$v = 1{,}500 - 24.6p$

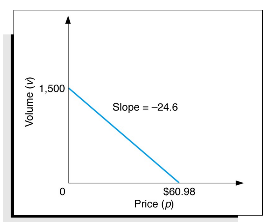


Figure 11.1
Linear Relationship of Volume to Price

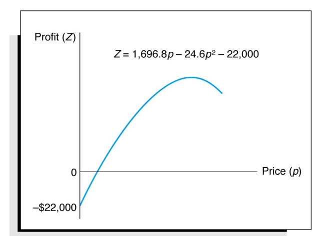
Optimization Techniques

Optimal Value of a Single Nonlinear Function Expanding the Basic Model to a Nonlinear Model

With fixed cost (c_f = \$10,000) and variable cost (c_v = \$8):

$Z = 1{,}696.8p - 24.6p^2 - 22{,}000$

Figure 11.2
The Nonlinear Profit Function

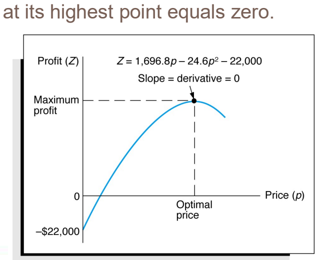


Optimization Techniques

Optimal Value of a Single Nonlinear Function Maximum Point on a Curve

- The slope of a curve at any point is equal to the derivative of the curve's function.
- The slope of a curve at its highest point equals zero.

Figure 11.3
Maximum profit for the profit function



Optimization Techniques

Optimal Value of a Single Nonlinear Function Solution Using Calculus

nonlinearity!

$Z = 1{,}696.8 \cdot p - 24.6 \cdot p^2 - 22{,}000$

$dZ/dp = 1{,}696.8 - 49.2p$

$p = \$34.49$

$v = 1{,}500 - 24.6p$

v = 651.6 pairs of jeans

Z = \$7,259.45

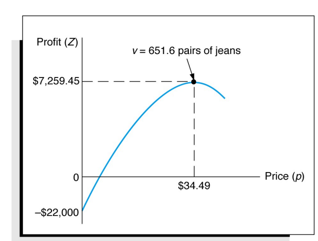


Figure 11.4
Maximum Profit, Optimal Price, and Optimal Volume

Optimization Techniques

Constrained Optimization in Nonlinear Problems Definition

- If a nonlinear problem contains one or more constraints it becomes a constrained optimization model or a *nonlinear programming* model.
- A nonlinear programming model has the same general form as the linear programming model except that the objective function *and/or* the constraint(s) are nonlinear.
- Solution procedures are much more complex and no guaranteed procedure exists.

Optimization Techniques

Constrained Optimization in Nonlinear Problems Graphical Interpretation (1 of 3)

- Effect of adding constraints to nonlinear problem:

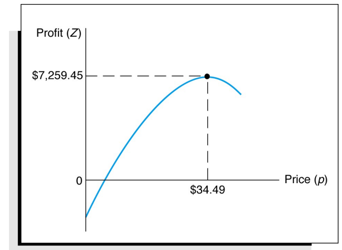


Figure 11.5
Nonlinear Profit Curve for the Profit Analysis Model

Optimization Techniques

Constrained Optimization in Nonlinear Problems Graphical Interpretation (2 of 3)

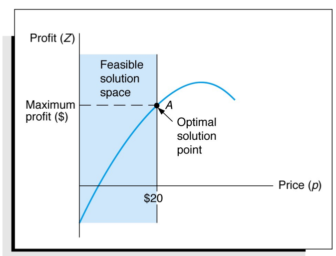


Figure 11.6
A Constrained Optimization Model

Optimization Techniques

Constrained Optimization in Nonlinear Problems Graphical Interpretation (3 of 3)

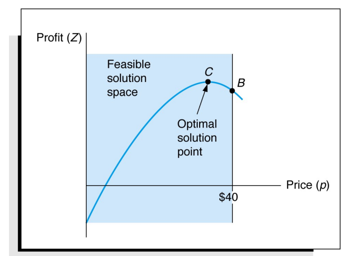


Figure 11.7
A Constrained Optimization Model with a Solution Point Not on the Constraint Boundary

Optimization Techniques

Constrained Optimization in Nonlinear Problems Characteristics

- Unlike linear programming, solution is often not on the boundary of the feasible solution space.
- Cannot simply look at points on the solution space boundary but must consider other points on the surface of the objective function.
- This greatly complicates solution approaches.
- Solution techniques can be very complex.

Optimization Techniques

INTRODUCTION TO NONLINEAR PROGRAMMING

- We optimized one-variable nonlinear functions using the 1st and 2nd derivatives.
- We will use the same concept here extended to functions with more than one variable.

Optimization Techniques

MULTIVARIABLE UNCONSTRAINED OPTIMIZATION

- For functions with one variable, we use the 1st and 2nd derivatives.
- For functions with multiple variables, we use identical information that is the gradient and the Hessian.
- The gradient is the first derivative with respect to all variables whereas the Hessian is the equivalent of the second derivative

Optimization Techniques

THE GRADIENT

- Review of the gradient (∇):

For a function "f", of variables $x_1, x_2, \ldots, x_n$:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Example: $f = 15x_1 + 2(x_2)^3 - 3x_1(x_3)^2$

$$\nabla f = \begin{bmatrix} 15 - 3(x_3)^2 & 6(x_2)^2 & -6x_1x_3 \end{bmatrix}$$

Optimization Techniques

THE HESSIAN

- The Hessian (∇^2) of $f(x_1, x_2, \ldots, x_n)$ is:

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Optimization Techniques

HESSIAN EXAMPLE

- Example (from previously):

$$f = 15x_1 + 2(x_2)^3 - 3x_1(x_3)^2$$

$$\nabla f = \begin{bmatrix} 15 - 3(x_3)^2 & 6(x_2)^2 & -6x_1x_3 \end{bmatrix}$$

$$\nabla^2 f = \begin{bmatrix} 0 & 0 & -6x_3 \\ 0 & 12x_2 & 0 \\ -6x_3 & 0 & -6x_1 \end{bmatrix}$$

Optimization Techniques

UNCONSTRAINED OPTIMIZATION

The optimization procedure for multivariable functions is:

1. Solve for the gradient of the function equal to zero to obtain candidate points.
2. Obtain the Hessian of the function and evaluate it at each of the candidate points
   - If the result is "positive definite" (defined later) then the point is a local minimum.
   - If the result is "negative definite" (defined later) then the point is a local maximum.

Optimization Techniques

POSITIVE/NEGATIVE DEFINITE

- A matrix is "positive definite" if all of the eigenvalues of the matrix are positive (> 0)
- A matrix is "negative definite" if all of the eigenvalues of the matrix are negative (< 0)

POSITIVE/NEGATIVE SEMI-DEFINITE

- A matrix is "positive semi-definite" if all of the eigenvalues are non-negative (≥ 0)
- A matrix is "negative semi-definite" if all of the eigenvalues are non-positive (≤ 0)

EIGEN VALUE AND EIGEN VECTOR DEFINITIONS

Definition 1: A nonzero vector x is an *eigenvector* (or *characteristic vector*) of a square matrix A if there exists a scalar λ such that Ax = λx. Then λ is an *eigenvalue* (or *characteristic value*) of A.

Note: The zero vector can not be an eigenvector even though A0 = λ0. But λ = 0 can be an eigenvalue.

Example:

$$Show\ x=\begin{bmatrix}2\\1\end{bmatrix} is\,an\,eigenvector\ for\ A=\begin{bmatrix}2 & -4\\3 & -6\end{bmatrix}$$

$$Solution: Ax=\begin{bmatrix}2 & -4\\3 & -6\end{bmatrix}\begin{bmatrix}2\\1\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

$$But\ \ for\ \ \lambda=0,\ \ \lambda x=0\begin{bmatrix}2\\1\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

$$Thus, x\,is\,an\,eigenvector\,of\ A, and\ \lambda=0\ is\,an\,eigenvalue.$$

GEOMETRIC INTERPRETATION OF EIGENVALUES AND EIGENVECTORS

An n × n matrix A multiplied by n × 1 vector x results in another n × 1 vector y=Ax. Thus A can be considered as a transformation matrix.

In general, a matrix acts on a vector by changing both its magnitude and its direction. However, a matrix may act on certain vectors by changing only their magnitude, and leaving their direction unchanged (or possibly reversing it). These vectors are the eigenvectors of the matrix.

A matrix acts on an eigenvector by multiplying its magnitude by a factor, which is positive if its direction is unchanged and negative if its direction is reversed. This factor is the eigenvalue associated with that eigenvector.

EIGENVALUES

Let x be an eigenvector of the matrix A. Then there must exist an eigenvalue λ such that Ax = λx or, equivalently,

Ax - λx = 0 or

(A – λI)x = 0

If we define a new matrix B = A – λI, then

Bx = 0

If B has an inverse then x = $B^{-1}0$ = 0. But an eigenvector cannot be zero.

Thus, it follows that x will be an eigenvector of A if and only if B does not have an inverse, or equivalently det(B)=0, or

det(A – λI) = 0

This is called the characteristic equation of A. Its roots determine the eigenvalues of A.

EIGENVALUES: EXAMPLES

Example 1: Find the eigenvalues of $A=\begin{bmatrix}2 & -12\\1 & -5\end{bmatrix}$

$$|\lambda I-A|=\begin{vmatrix}\lambda-2 & 12\\-1 & \lambda+5\end{vmatrix}=(\lambda-2)(\lambda+5)+12$$

$$=\lambda^2+3\lambda+2=(\lambda+1)(\lambda+2)$$

two eigenvalues: −1, − 2

Note: The roots of the characteristic equation can be repeated. That is, $\lambda_1 = \lambda_2 = \ldots = \lambda_k$. If that happens, the eigenvalue is said to be of multiplicity k.

Example 2: Find the eigenvalues of $A=\begin{bmatrix}2 & 1 & 0\\0 & 2 & 0\\0 & 0 & 2\end{bmatrix}$

$$|\lambda I-A|=\begin{vmatrix}\lambda-2 & -1 & 0\\0 & \lambda-2 & 0\\0 & 0 & \lambda-2\end{vmatrix}=(\lambda-2)^3=0$$

λ = 2 is an eigenvector of multiplicity 3.

EIGENVECTORS

To each distinct eigenvalue of a matrix A there will correspond at least one eigenvector which can be found by solving the appropriate set of homogenous equations. If λ_i is an eigenvalue then the corresponding eigenvector x_i is the solution of $(A-\lambda_i I)x_i = 0$

Example 1 (cont.):

$$\lambda=-1:(-1)I-A=\begin{bmatrix}-3 & 12\\-1 & 4\end{bmatrix}\Rightarrow\begin{bmatrix}1 & -4\\0 & 0\end{bmatrix}$$

$$x_1-4x_2=0\Rightarrow x_1=4t,\ x_2=t$$

$$\mathbf{x}_1=\begin{bmatrix}x_1\\x_2\end{bmatrix}=t\begin{bmatrix}4\\1\end{bmatrix},\ t\neq 0$$

$$\lambda=-2:(-2)I-A=\begin{bmatrix}-4 & 12\\-1 & 3\end{bmatrix}\Rightarrow\begin{bmatrix}1 & -3\\0 & 0\end{bmatrix}$$

$$\mathbf{x}_2=\begin{bmatrix}x_1\\x_2\end{bmatrix}=s\begin{bmatrix}3\\1\end{bmatrix},\ s\neq 0$$

Optimization Techniques

EIGENVECTORS

Example 2 (cont.): Find the eigenvectors of $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Recall that λ = 2 is an eigenvector of multiplicity 3.

Solve the homogeneous linear system represented by

$$(2I - A)\mathbf{x} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$x_1 = s, x_3 = t$

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} s \\ 0 \\ t \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, The eigenvectors of λ = 2 are of the form

s and *t* not both zero.

Optimization Techniques

PROPERTIES OF EIGENVALUES AND EIGENVECTORS

Definition: The trace of a matrix A, designated by tr(A), is the sum of the elements on the main diagonal.

Property 1: The sum of the eigenvalues of a matrix equals the trace of the matrix.

Property 2: A matrix is singular if and only if it has a zero eigenvalue (A square matrix that does not have a matrix inverse. A matrix is singular iff its determinant is 0).

Property 3: The eigenvalues of an upper (or lower) triangular matrix are the elements on the main diagonal.

Property 4: If λ is an eigenvalue of A and A is invertible, then 1/λ is an eigenvalue of matrix A^{-1}.

Optimization Techniques

PROPERTIES OF EIGENVALUES AND EIGENVECTORS

Property 5: If λ is an eigenvalue of A then kλ is an eigenvalue of kA where k is any arbitrary scalar.

Property 6: If λ is an eigenvalue of A then λ^k is an eigenvalue of A^k for any positive integer k.

Property 8: If λ is an eigenvalue of A then λ is an eigenvalue of A^T.

Property 9: The product of the eigenvalues (counting multiplicity) of a matrix equals the determinant of the matrix.

Optimization Techniques

LINEARLY INDEPENDENT EIGENVECTORS

Theorem: Eigenvectors corresponding to distinct (that is, different) eigenvalues are linearly independent.

Theorem: If λ is an eigenvalue of multiplicity k of an n × n matrix A then the number of linearly independent eigenvectors of A associated with λ is given by m = n - r(A- λI). Furthermore, 1 ≤ m ≤ k.

Example 2 (cont.): The eigenvectors of λ = 2 are of the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} s \\ 0 \\ t \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$ *s* and *t* not both zero.

λ = 2 has *two linearly independent eigenvectors*

Optimization Techniques

EXAMPLE MATRIX

Given the matrix A:

$$A = \begin{bmatrix} 2 & 4 & 5 \\ -5 & -7 & -1 \\ 1 & 1 & 2 \end{bmatrix}$$

The eigenvalues of A are:

$\lambda_1 = -3.702$ $\lambda_2 = -2$ $\lambda_3 = 2.702$

This matrix is negative definite

Optimization Techniques

UNCONSTRAINED NLP EXAMPLE

Consider the problem:

Minimize $f(x_1, x_2, x_3) = (x_1)^2 + x_1(1 - x_2) + (x_2)^2 - x_2x_3 + (x_3)^2 + x_3$

First, we find the gradient with respect to x_i:

$$\nabla f = \begin{bmatrix} 2x_1 + 1 - x_2 \\ -x_1 + 2x_2 - x_3 \\ -x_2 + 2x_3 + 1 \end{bmatrix}$$

Optimization Techniques

UNCONSTRAINED NLP EXAMPLE

Next, we set the gradient equal to zero:

$$\nabla f = 0 \quad \Rightarrow \quad \begin{bmatrix} 2x_1 + 1 - x_2 \\ -x_1 + 2_2 - x_3 \\ -x_2 + 2x_3 + 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

So, we have a system of 3 equations and 3 unknowns. When we solve, we get:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Optimization Techniques

UNCONSTRAINED NLP EXAMPLE

So we have only one candidate point to check.

Find the Hessian:

$$\nabla^2 f = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Optimization Techniques

UNCONSTRAINED NLP EXAMPLE

The eigenvalues of this matrix are:

$$\lambda_1 = 3.414 \qquad \lambda_2 = 0.586 \qquad \lambda_3 = 2$$

All of the eigenvalues are > 0, so the Hessian is positive definite.

So, the point $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$ is a minimum

Optimization Techniques

UNCONSTRAINED NLP EXAMPLE

Unlike in Linear Programming, unless we know the shape of the function being minimized or can determine whether it is convex, we cannot tell whether this point is the global minimum or if there are function values smaller than it.

Optimization Techniques

9.2: Taylor Series

Brook Taylor was an accomplished musician and painter. He did research in a variety of areas, but is most famous for his development of ideas regarding infinite series.

Brook Taylor
1685 - 1731

Greg Kelly, Hanford High School, Richland, Washington

Optimization Techniques

Suppose we wanted to find a fourth degree polynomial of the form:

$$P(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$$

that approximates the behavior of $f(x) = \ln(x+1)$ at $x = 0$

If we make $P(0) = f(0)$ and the first, second, third and fourth derivatives the same, then we would have a pretty good approximation.

→

Optimization

$$P(x)=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4 \qquad f(x)=\ln(x+1)$$

$f(x)=\ln(x+1)$ $\qquad P(x)=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4$

$f(0)=\ln(1)=0$ $\qquad P(0)=a_0 \rightarrow \boxed{a_0=0}$

$f'(x)=\frac{1}{1+x}$ $\qquad P'(x)=a_1+2a_2x+3a_3x^2+4a_4x^3$

$f'(0)=\frac{1}{1}=1$ $\qquad P'(0)=a_1 \rightarrow \boxed{a_1=1}$

$f''(x)=-\frac{1}{(1+x)^2}$ $\qquad P''(x)=2a_2+6a_3x+12a_4x^2$

$f''(0)=-\frac{1}{1}=-1$ $\qquad P''(0)=2a_2 \rightarrow \boxed{a_2=-\frac{1}{2}}$

Optimization

$$P(x)=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4 \qquad f(x)=\ln(x+1)$$

$f''(x)=-\frac{1}{(1+x)^2}$ $\qquad P''(x)=2a_2+6a_3x+12a_4x^2$

$f''(0)=-\frac{1}{1}=-1$ $\qquad P''(0)=2a_2 \rightarrow \boxed{a_2=-\frac{1}{2}}$

$f'''(x)=2\cdot\frac{1}{(1+x)^3}$ $\qquad P'''(x)=6a_3+24a_4x$

$f'''(0)=2$ $\qquad P'''(0)=6a_3 \rightarrow \boxed{a_3=\frac{2}{6}}$

$f^{(4)}(x)=-6\frac{1}{(1+x)^4}$ $\qquad P^{(4)}(x)=24a_4$

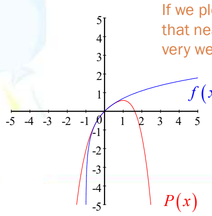
$f^{(4)}(0)=-6$ $\qquad P^{(4)}(0)=24a_4 \rightarrow \boxed{a_4=-\frac{6}{24}}$

Optimization

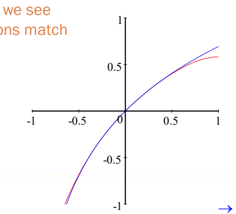
$$P(x)=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4 \qquad f(x)=\ln(x+1)$$

$$P(x)=0+1x-\frac{1}{2}x^2+\frac{2}{6}x^3-\frac{6}{24}x^4$$

$$P(x)=0+x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4} \qquad f(x)=\ln(x+1)$$



If we plot both functions, we see that near zero the functions match very well!



Optimization Techniques

Our polynomial: $\quad 0+1x-\frac{1}{2}x^2+\frac{2}{6}x^3-\frac{6}{24}x^4$

has the form: $\quad f(0)+f'(0)x+\frac{f''(0)}{2}x^2+\frac{f'''(0)}{6}x^3+\frac{f^{(4)}(0)}{24}x^4$

or: $\quad \frac{f(0)}{0!}+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\frac{f^{(4)}(0)}{4!}x^4$

This pattern occurs no matter what the original function was!

Optimization

Maclaurin Series:

(generated by f at $x=0$)

$$P(x)=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\cdots$$

If we want to center the series (and it's graph) at some point other than zero, we get the Taylor Series:

Taylor Series:

(generated by f at $x=a$)

$$P(x)=f(a)+f'(a)(x-a)+\frac{f''(a)}{2!}(x-a)^2+\frac{f'''(a)}{3!}(x-a)^3+\cdots$$

Optimization Techniques

example: $y=\cos x$

$f(x)=\cos x \quad f(0)=1 \qquad f'''(x)=\sin x \quad f'''(0)=0$

$f'(x)=-\sin x \quad f'(0)=0 \qquad f^{(4)}(x)=\cos x \quad f^{(4)}(0)=1$

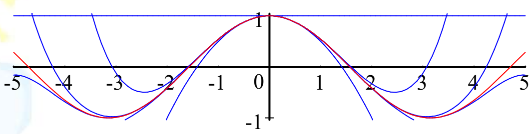
$f''(x)=-\cos x \quad f''(0)=-1$

$$P(x)=1+0x-\frac{1x^2}{2!}+\frac{0x^3}{3!}+\frac{1x^4}{4!}+\frac{0x^5}{5!}-\frac{1x^6}{6!}+\cdots$$

$$P(x)=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\frac{x^8}{8!}-\frac{x^{10}}{10!}\cdots$$

Optimization Techniques

$y = \cos x$ $\qquad P(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{x^{10}}{10!} \cdots$



The more terms we add, the better our approximation.

Hint: On the TI-89, the factorial symbol is: ◆ ÷

Optimization Techniques

example: $y = \cos(2x)$

Rather than start from scratch, we can use the function that we already know:

$$P(x) = 1 - \frac{(2x)^2}{2!} + \frac{(2x)^4}{4!} - \frac{(2x)^6}{6!} + \frac{(2x)^8}{8!} - \frac{(2x)^{10}}{10!} \cdots$$

Optimization Techniques

example: $y = \cos(x)$ at $x = \frac{\pi}{2}$

$f(x) = \cos x \quad f\left(\frac{\pi}{2}\right) = 0$

$f'(x) = -\sin x \quad f'\left(\frac{\pi}{2}\right) = -1$

$f''(x) = -\cos x \quad f''\left(\frac{\pi}{2}\right) = 0$

$f'''(x) = \sin x \quad f'''\left(\frac{\pi}{2}\right) = 1$

$f^{(4)}(x) = \cos x \quad f^{(4)}\left(\frac{\pi}{2}\right) = 0$

$$P(x) = 0 - 1\left(x - \frac{\pi}{2}\right) + \frac{0}{2!}\left(x - \frac{\pi}{2}\right)^2 + \frac{1}{3!}\left(x - \frac{\pi}{2}\right)^3 + \cdots$$

$$\boxed{P(x) = -\left(x - \frac{\pi}{2}\right) + \frac{\left(x - \frac{\pi}{2}\right)^3}{3!} - \frac{\left(x - \frac{\pi}{2}\right)^5}{5!} + \cdots}$$

Optimization Techniques

There are some Maclaurin series that occur often enough that they should be memorized. They are on your formula sheet.

Optimization Techniques

When referring to Taylor polynomials, we can talk about **number of terms**, **order** or **degree**.

$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!}$ This is a polynomial in **3 terms**.

It is a **4th order** Taylor polynomial, because it was found using the 4th derivative.

It is also a **4th degree** polynomial, because x is raised to the 4th power.

The **3rd order** polynomial for $\cos x$ is $1 - \frac{x^2}{2!}$, but it is **degree 2**.

A recent AP exam required the student to know the difference between *order* and *degree*.

Optimization Techniques

The TI-89 finds Taylor Polynomials:

taylor (expression, variable, order, [point])

F3 9

taylor $(\cos(x), x, 6)$ $\qquad \frac{-x^6}{720} + \frac{x^4}{24} - \frac{x^2}{2} + 1$

taylor $(\cos(2x), x, 6)$ $\qquad \frac{-4x^6}{45} + \frac{2x^4}{3} - 2x^2 + 1$

taylor $(\cos(x), x, 5, \pi/2)$ $\qquad \frac{-(2x-\pi)^5}{3840} + \frac{(2x-\pi)^3}{48} - \frac{2x-\pi}{2}$

π

Optimization Techniques — Summary

- A first order approximation of $f(x)$**around point** $\bar{x}$ is given by
- $f_1(x) = f(x') + f'(x') * (x - x')$
- *Taylor Approximation around a vector x*
- $f_1(\mathbf{x}) = f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T (\mathbf{x} - \mathbf{x}^k)$
- $f1(\mathbf{x}) = f(\mathbf{x}^k) + \sum_{j=1}^{n} \frac{\partial f(\mathbf{x}^k)}{\partial x_j}(x_j - x_j^k)$
- Determine first order Taylor approximation of the function
  - $f(x_1, x_2) = x_1^4 + x_1^2 + 2x_2^2 - 2x_1x_2$ around a point (x_1, x_2)=(1,1)

Optimization Techniques — Summary

- A second order approximation of $f(x)$**around point** $\bar{x}$ for single variable is given by
- $f_2(x) = f(x') + f'(x')(x - x') + \frac{1}{2} f''(x')(x - x')^2$
- *A second order Taylor Approximation around a vector x*
- $f_2(\mathbf{x}) = f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T (\mathbf{x} - \mathbf{x}^k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^k)^T H(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k)$
- $f1(\mathbf{x}) = f(\mathbf{x}^k) + \sum_{j=1}^{n} \frac{\partial f(\mathbf{x}^k)}{\partial x_j} * (x_j - x_j^k) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f(\mathbf{x}^k)}{\partial x_i \partial x_j}(x_i - x_i^k)(x_j - x_j^k)$

$f(x_1, x_2) = x_1^4 + x_1^2 + 2x_2^2 - 2x_1x_2$ around a point (x_1, x_2)=(1,1)

Optimization Techniques — METHOD OF SOLUTION

- In the previous example, when we set the gradient equal to zero, we had a system of 3 linear equations & 3 unknowns.
- For other problems, these equations could be nonlinear.
- Thus, the problem can become trying to solve a system of nonlinear equations, which can be very difficult.

Optimization Techniques — METHOD OF SOLUTION

- To avoid this difficulty, NLP problems are usually solved numerically.
- We will now look at examples of numerical methods used to find the optimum point for single-variable NLP problems. These and other methods may be found in any numerical methods reference.

Optimization Techniques — NEWTON'S METHOD

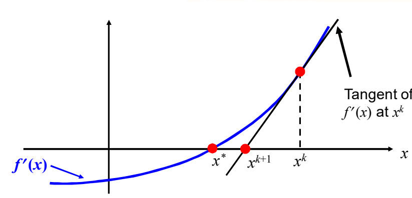
When solving the equation $f'(x) = 0$ to find a minimum or maximum, one can use the iteration step:

$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)}$$

where k is the current iteration.

Iteration is continued until $|x^{k+1} - x^k| < \varepsilon$ where ε is some specified tolerance.

Optimization Techniques — NEWTON'S METHOD DIAGRAM



Newton's Method approximates $f'(x)$ as a straight line at x^k and obtains a new point (x^{k+1}), which is used to approximate the function at the next iteration. This is carried on until the new point is sufficiently close to x^*.

NEWTON'S METHOD COMMENTS

- One must ensure that $f(x^{k+1}) < f(x^k)$ for finding a minimum and $f(x^{k+1}) > f(x^k)$ for finding a maximum.
- Disadvantages:
  - The initial guess is very important – if it is not close enough to the solution, the method may not converge
  - Both the first and second derivatives must be calculated

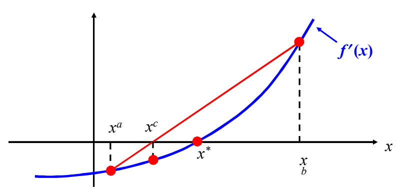
REGULA-FALSI METHOD

This method requires two points, x^a & x^b that bracket the solution to the equation $f'(x) = 0$.

$$x^c = x^b - \frac{f'(x^b)\cdot(x^b - x^a)}{f'(x^b) - f'(x^a)}$$

where x^c will be between x^a & x^b. The next interval will be x^c and either x^a or x^b, whichever has the sign opposite of x^c.

REGULA-FALSI DIAGRAM



The Regula-Falsi method approximates the function $f'(x)$ as a straight line and interpolates to find the root.

REGULA-FALSI COMMENTS

- This method requires initial knowledge of two points bounding the solution
- However, it does not require the calculation of the second derivative
- The Regula-Falsi Method requires slightly more iterations to converge than the Newton's Method

MULTIVARIABLE OPTIMIZATION

- Now we will consider unconstrained multivariable optimization
- Nearly all multivariable optimization methods do the following:
  1. Choose a search direction $\mathbf{d}^k$
  2. Minimize along that direction to find a new point:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$$

where k is the current iteration number and α^k is a positive scalar called the step size.

THE STEP SIZE

- The step size, α^k, is calculated in the following way:
- We want to minimize the function $f(\mathbf{x}^{k+1}) = f(\mathbf{x}^k + \alpha^k \mathbf{d}^k)$ where the only variable is α^k because $\mathbf{x}^k$ & $\mathbf{d}^k$ are known.
- We set $\frac{df(\mathbf{x}^k + \alpha^k \mathbf{d}^k)}{d\alpha^k} = 0$ and solve for α^k using a single-variable solution method such as the ones shown previously.

Optimization Techniques

STEEPEST DESCENT METHOD

- This method is very simple – it uses the gradient (for maximization) or the negative gradient (for minimization) as the search direction:

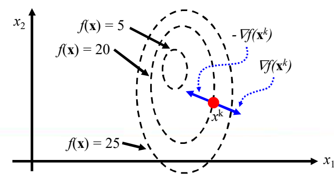
$$\mathbf{d}^k = \begin{Bmatrix} + \\ - \end{Bmatrix} \nabla f(\mathbf{x}^k) \text{ for } \begin{Bmatrix} \max \\ \min \end{Bmatrix}$$

So, $\mathbf{x}^{k+1} = \mathbf{x}^k \begin{Bmatrix} + \\ - \end{Bmatrix} \alpha^k \nabla f(\mathbf{x}^k)$

Optimization Techniques

STEEPEST DESCENT METHOD

- Because the gradient is the rate of change of the function at that point, using the gradient (or negative gradient) as the search direction helps reduce the number of iterations needed



Optimization Techniques

STEEPEST DESCENT METHOD STEPS

So the steps of the Steepest Descent Method are:

1. Choose an initial point $\mathbf{x}^0$
2. Calculate the gradient $\nabla f(\mathbf{x}^k)$ where k is the iteration number
3. Calculate the search vector: $\mathbf{d}^k = \pm \nabla f(\mathbf{x}^k)$
4. Calculate the next $\mathbf{x}$: $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
   Use a single-variable optimization method to determine α^k.

Optimization Techniques

STEEPEST DESCENT METHOD STEPS

5. To determine convergence, either use some given tolerance ε_1 and evaluate:

   $$\left| f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \right| < \varepsilon_1$$

   for convergence

Or, use another tolerance ε_2 and evaluate:

for convergence $\left\| \nabla f(\mathbf{x}^k) \right\| < \varepsilon_2$

Optimization Techniques

CONVERGENCE

- These two criteria can be used for any of the multivariable optimization methods discussed here

Recall: The norm of a vector, $\|\mathbf{x}\|$ is given by:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \cdot \mathbf{x}} = \sqrt{(x_1)^2 + (x_2)^2 + \cdots + (x_n)^2}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Let's solve the earlier problem with the Steepest Descent Method:

Minimize $f(x_1, x_2, x_3) = (x_1)^2 + x_1(1 - x_2) + (x_2)^2 - x_2x_3 + (x_3)^2 + x_3$

Let's pick

$$\mathbf{x}^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Optimization Techniques
STEEPEST DESCENT EXAMPLE

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 + (1 - x_2) & -x_1 + 2x_2 - x_3 & -x_2 + 2x_3 + 1 \end{bmatrix}$$

$$\mathbf{d}^0 = -\nabla f(\mathbf{x}^0) = -\begin{bmatrix} 2(0) + 1 - 0 & -0 + 0 - 0 & -0 + 0 + 1 \end{bmatrix}$$
$$= -\begin{bmatrix} 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & -1 \end{bmatrix}$$

$$\mathbf{x}^1 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} + \alpha^0 \cdot \begin{bmatrix} -1 & 0 & -1 \end{bmatrix}$$

Now, we need to determine α^0

Optimization Techniques
STEEPEST DESCENT EXAMPLE

$$f(\mathbf{x}^1) = (\alpha^0)^2 + (-\alpha^0)(1) + 0 - 0 + (\alpha^0)^2 + (-\alpha^0)$$
$$= 2(\alpha^0)^2 - 2(\alpha^0)$$

$$\frac{df(\mathbf{x}^1)}{d\alpha^0} = 4(\alpha^0) - 2$$

Now, set equal to zero and solve:

$$4(\alpha^0) = 2 \Rightarrow \alpha^0 = 2/4 = 1/2$$

Optimization Techniques
STEEPEST DESCENT EXAMPLE

So,

$$\mathbf{x}^1 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} + \alpha^0 \cdot \begin{bmatrix} -1 & 0 & -1 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$
$$\therefore \mathbf{x}^1 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

Optimization Techniques
STEEPEST DESCENT EXAMPLE

Take the negative gradient to find the next search direction:

$$\mathbf{d}^1 = -\nabla f(\mathbf{x}^1) = -\begin{bmatrix} -1 + 1 + 0 & \frac{1}{2} + 0 + \frac{1}{2} & 0 - 1 + 1 \end{bmatrix}$$
$$\therefore \mathbf{d}^1 = \begin{bmatrix} 0 & -1 & 0 \end{bmatrix}$$

Optimization Techniques
STEEPEST DESCENT EXAMPLE

Update the iteration formula:

$$\mathbf{x}^2 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} + \alpha^1 \cdot \begin{bmatrix} 0 & -1 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} -\frac{1}{2} & -\alpha^1 & -\frac{1}{2} \end{bmatrix}$$

Optimization Techniques
STEEPEST DESCENT EXAMPLE

Insert into the original function & take the derivative so that we can find α^1:

$$f(\mathbf{x}^2) = \frac{1}{4} + \left(-\frac{1}{2}\right)(1 + \alpha^1) + (\alpha^1)^2 - (\alpha^1)\left(\frac{1}{2}\right) + \frac{1}{4} - \frac{1}{2}$$
$$= (\alpha^1)^2 - \alpha^1 - 1/2$$

$$\frac{df(\mathbf{x}^1)}{d\alpha^1} = 2(\alpha^1) - 1$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Now we can set the derivative equal to zero and solve for α^1:

$$2(\alpha^1) = 1 \quad \Rightarrow \quad \alpha^1 = \tfrac{1}{2}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Now, calculate $\mathbf{x}^2$:

$$\mathbf{x}^2 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} + \alpha^1 \cdot \begin{bmatrix} 0 & -1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} + \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$\therefore \mathbf{x}^2 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

$$\mathbf{d}^2 = -\nabla f(\mathbf{x}^2) = -\begin{bmatrix} -1+1+\frac{1}{2} & \frac{1}{2}-1+\frac{1}{2} & \frac{1}{2}-1+1 \end{bmatrix}$$

$$\therefore \mathbf{d}^2 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

So,

$$\mathbf{x}^3 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \alpha^2 \cdot \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2}(\alpha^2+1) & -\frac{1}{2} & -\frac{1}{2}(\alpha^2+1) \end{bmatrix}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Find α^2: $\quad f(\mathbf{x}^3) = \frac{1}{2}(\alpha^2+1)^2 - \frac{3}{2}(\alpha^2+1) + \frac{1}{4}$

$$\frac{df(\mathbf{x}^3)}{d\alpha^2} = (\alpha^2+1) - \frac{3}{2}$$

Set the derivative equal to zero and solve:

$$(\alpha^2+1) = \frac{3}{2} \quad \Rightarrow \quad \alpha^2 = \tfrac{1}{2}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Calculate $\mathbf{x}^3$:

$$\mathbf{x}^3 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \alpha^2 \cdot \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \begin{bmatrix} -\frac{1}{4} & 0 & -\frac{1}{4} \end{bmatrix}$$

$$\therefore \mathbf{x}^3 = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4} \end{bmatrix}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Find the next search direction:

$$\mathbf{d}^3 = -\nabla f(\mathbf{x}^3) = -\begin{bmatrix} 0 & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$\mathbf{x}^4 = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4} \end{bmatrix} + \alpha^3 \cdot \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2}(\alpha^3+1) & -\frac{3}{4} \end{bmatrix}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Find α^3:

$$f(\mathbf{x}^4) = \frac{1}{4}(\alpha^3+1)^2 - \frac{3}{2}(\alpha^3) - \frac{3}{2}$$

$$\frac{df(\mathbf{x}^4)}{d\alpha^3} = \frac{1}{2}(\alpha^3+1) - \frac{9}{8} = 0$$

$$\Rightarrow \alpha^3 = \frac{5}{4}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

So, $\mathbf{x}^4$ becomes:

$$\mathbf{x}^4 = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4} \end{bmatrix} + \begin{bmatrix} 0 & -\frac{5}{8} & 0 \end{bmatrix}$$

$$\therefore \mathbf{x}^4 = \begin{bmatrix} -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \end{bmatrix}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

The next search direction:

$$\mathbf{d}^4 = -\nabla f(\mathbf{x}^4) = -\begin{bmatrix} \frac{5}{8} & -\frac{3}{4} & \frac{5}{8} \end{bmatrix} = \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$

$$\mathbf{x}^5 = \begin{bmatrix} -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \end{bmatrix} + \alpha^4 \cdot \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$
$$= \begin{bmatrix} -\frac{1}{4}(3+\frac{5}{2}\alpha^4) & -\frac{3}{4}(\frac{3}{2}-\alpha^4) & -\frac{1}{4}(3+\frac{5}{2}\alpha^4) \end{bmatrix}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Find α^4:

$$f(\mathbf{x}^5) = \frac{73}{32}(\alpha^4)^2 - \frac{43}{32}\alpha^4 - \frac{51}{64}$$

$$\frac{df(\mathbf{x}^5)}{d\alpha^4} = \frac{73}{16}\alpha^4 - \frac{43}{32} = 0$$

$$\Rightarrow \alpha^4 = {}^{43}\!/_{146}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Update $\mathbf{x}^5$:

$$\mathbf{x}^5 = \begin{bmatrix} -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \end{bmatrix} + \frac{43}{146} \cdot \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$

$$\therefore \mathbf{x}^5 = \begin{bmatrix} -\frac{1091}{1168} & -\frac{66}{73} & -\frac{1091}{1168} \end{bmatrix}$$

Optimization Techniques

STEEPEST DESCENT EXAMPLE

Let's check to see if the convergence criteria is satisfied

Evaluate $||\nabla f(\mathbf{x}^5)||$:

$$\nabla f(\mathbf{x}^5) = \begin{bmatrix} \frac{21}{584} & \frac{35}{584} & \frac{21}{584} \end{bmatrix}$$

$$\left\|\nabla f(\mathbf{x}^5)\right\| = \sqrt{\left({}^{21}\!/_{584}\right)^2 + \left({}^{35}\!/_{584}\right)^2 + \left({}^{21}\!/_{584}\right)^2} = 0.0786$$

STEEPEST DESCENT EXAMPLE

So, $||\nabla f(\mathbf{x}^5)|| = 0.0786$, which is very small and we can take it to be close enough to zero for our example

Notice that the answer of

$$\mathbf{x} = \begin{bmatrix} -\frac{1091}{1168} & -\frac{66}{73} & -\frac{1091}{1168} \end{bmatrix}$$

is very close to the value of $\mathbf{x}^* = \begin{bmatrix} -1 & -1 & -1 \end{bmatrix}$ that we obtained analytically

CONSTRAINED NONLINEAR OPTIMIZATION

- Previously in this chapter, we solved NLP problems that only had objective functions, with no constraints.
- Now we will look at methods on how to solve problems that include constraints.

NLP WITH EQUALITY CONSTRAINTS

- First, we will look at problems that only contain equality constraints:

Minimize $f(\mathbf{x})$ $\quad\quad$ $\mathbf{x} = [x_1 \; x_2 \; \ldots \; x_n]$

Subject to: $h_i(\mathbf{x}) = b_i$ $\quad\quad$ $i = 1, 2, \ldots, m$

ILLUSTRATION

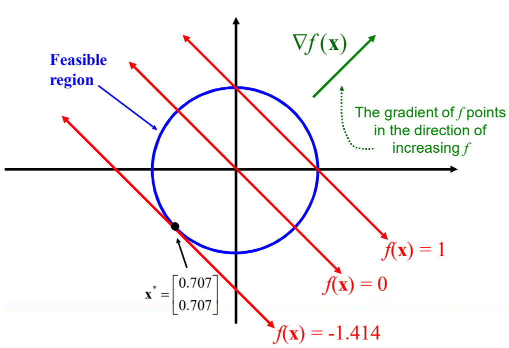
Consider the problem:

Minimize $x_1 + x_2$

Subject to: $(x_1)^2 + (x_2)^2 - 1 = 0$

The feasible region is a circle with a radius of one. The possible objective function curves are lines with a slope of -1. The minimum will be the point where the lowest line still touches the circle.

GRAPH OF ILLUSTRATION

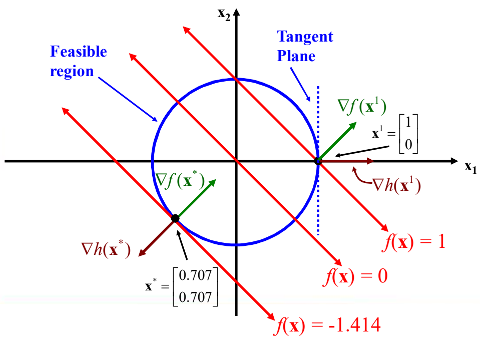


MORE ON THE GRAPH

- Since the objective function lines are straight parallel lines, the gradient of f is a straight line pointing toward the direction of increasing f, which is to the upper right
- The gradient of h will be pointing out from the circle and so its direction will depend on the point at which the gradient is evaluated.

Optimization Techniques

FURTHER DETAILS



Optimization Techniques

CONCLUSIONS

- At the optimum point, $\nabla f(\mathbf{x})$ is collinear to $\nabla h(\mathbf{x})$
- As we can see at point $\mathbf{x}^1$, $\nabla f(\mathbf{x})$ is not perpendicular to $\nabla h(\mathbf{x})$ and we can move (down) to improve the objective function
- We can say that at a max or min, $\nabla f(\mathbf{x})$ must be collinear to $\nabla h(\mathbf{x})$
  - Otherwise, we could improve the objective function by changing position

Optimization Techniques

FIRST ORDER NECESSARY CONDITIONS

So, in order for a point to be a minimum (or maximum), it must satisfy the following equation:

$$\nabla f(\mathbf{x}^*) + \lambda^* \cdot \nabla h(\mathbf{x}^*) = 0$$

This equation means that $\nabla f(\mathbf{x}^*)$ and $\nabla h(\mathbf{x}^*)$ must be in exactly opposite directions at a minimum or maximum point

Optimization Techniques

THE LAGRANGIAN FUNCTION

To help in using this fact, we introduce the Lagrangian Function, $L(\mathbf{x},\lambda)$:

$$L(\mathbf{x},\lambda) = f(\mathbf{x}) + \lambda \cdot h(\mathbf{x})$$

Review: The notation $\nabla_{\mathbf{x}} f(x,y)$ means the gradient of f with respect to $\mathbf{x}$.

So,

$$\nabla_{\mathbf{x}} L(\mathbf{x},\lambda) = \nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda \cdot \nabla_{\mathbf{x}} h(\mathbf{x})$$

Optimization Techniques

FIRST ORDER NECESSARY CONDITIONS

- So, using the new notation to express the First Order Necessary Conditions (FONC), if $\mathbf{x}^*$ is a minimum (or maximum) then

$$\nabla_{\mathbf{x}} L(\mathbf{x},\lambda)\big|_{(\mathbf{x}^*,\lambda^*)} = 0$$

and $h(\mathbf{x}^*) = 0$ to ensure feasibility.

Optimization Techniques

FIRST ORDER NECESSARY CONDITIONS

- Another way to think about it is that the one Lagrangian function includes all information about our problem
- So, we can treat the Lagrangian as an unconstrained optimization problem with variables $x_1, x_2, \ldots, x_n$ and $\lambda_1, \lambda_2, \ldots, \lambda_m$.

We can solve it by solving the equations

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{0} \quad \& \quad \frac{\partial L}{\partial \boldsymbol{\lambda}} = \mathbf{0}$$

USING THE FONC

Optimization Techniques

Using the FONC for the previous example,

$$L(\mathbf{x},\lambda) = f(\mathbf{x}) + \lambda \cdot h(\mathbf{x})$$
$$= x_1 + x_2 + \lambda \cdot \left((x_1)^2 + (x_2)^2 - 1\right)$$

And the first FONC equation is:

$$\nabla_x L(\mathbf{x},\lambda) = \begin{bmatrix} \partial L / \partial x_1 \\ \partial L / \partial x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

FONC EXAMPLE

Optimization Techniques

This becomes: $\frac{\partial L}{\partial x_1} = 1 + 2\lambda x_1 = 0$

& $\frac{\partial L}{\partial x_2} = 1 + 2\lambda x_2 = 0$

The feasibility equation is:

$$(x_1)^2 + (x_2)^2 - 1 = 0$$

or, $\frac{\partial L}{\partial \lambda} = (x_1)^2 + (x_2)^2 - 1 = 0$

FONC EXAMPLE

Optimization Techniques

So, we have three equations and three unknowns. When they are solved simultaneously, we obtain

$$x_1 = x_2 = \pm 0.707 \quad \& \quad \lambda = \mp 0.707$$

We can see from the graph that positive x_1 & x_2 corresponds to a maximum while negative x_1 & x_2 corresponds to the minimum.

FONC OBSERVATIONS

Optimization Techniques

- If you go back to the LP Chapter and look at the mathematical definition of the KKT conditions, you may notice that they look just like our FONC that we just used
- This is because it is the same concept
- We simply used a slightly different derivation this time but obtained the same result

LIMITATIONS OF FONC

Optimization Techniques

- The FONC do not guarantee that the solution(s) will be minimums/maximums.
- As in the case of unconstrained optimization, they only provide us with candidate points that need to be verified by the second order conditions.
- Only if the problem is convex do the FONC guarantee the solutions will be extreme points.

SECOND ORDER NECESSARY CONDITIONS (SONC)

Optimization Techniques

For $\nabla_x^2 L(\mathbf{x},\lambda)$ where

$$\nabla_x^2 L(\mathbf{x},\lambda) = \nabla^2 f(\mathbf{x}) + \lambda \cdot \nabla^2 h(\mathbf{x})$$

and for $\mathbf{y}$ where $\mathbf{J}_h(\mathbf{x}^*) \cdot \mathbf{y} = \begin{bmatrix} \partial h_1 / \partial \mathbf{x} \\ \vdots \\ \partial h_m / \partial \mathbf{x} \end{bmatrix}_{(\mathbf{x}^*)} \cdot \mathbf{y} = \mathbf{0}$

If $\mathbf{x}^*$ is a local minimum, then $\mathbf{y}^T \cdot \nabla_x^2 L(\mathbf{x}^*,\lambda) \cdot \mathbf{y} \geq 0$

Optimization Techniques

SECOND ORDER SUFFICIENT CONDITIONS (SOSC)

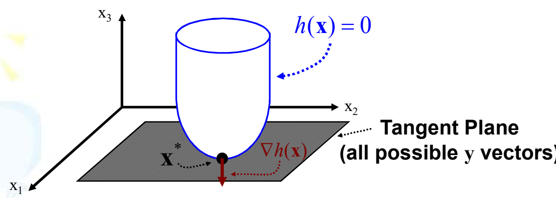
$$\mathbf{y}^T \cdot \nabla_x^2 L(\mathbf{x}^*, \lambda) \cdot \mathbf{y} > 0$$

- **y** can be thought of as being a tangent plane as in the graphical example shown previously
  - $\mathbf{J}_h$ is just the gradients of each $h(\mathbf{x})$ equation and we saw in the example that the tangent plane must be perpendicular to $\nabla h(\mathbf{x})$ and that is why

$$\mathbf{J}_h(x) \cdot \mathbf{y} = \mathbf{0}$$

Optimization Techniques

THE Y VECTOR



The tangent plane is the location of all $\mathbf{y}$ vectors and intersects with $\mathbf{x}^*$

It must be orthogonal (perpendicular) to $\nabla h(\mathbf{x})$

Optimization Techniques

MAXIMIZATION PROBLEMS

- The previous definitions of the SONC & SOSC were for minimization problems
- For maximization problems, the sense of the inequality sign will be reversed

For maximization problems:

SONC:

$$\mathbf{y}^T \cdot \nabla_x^2 L(\mathbf{x}^*, \lambda) \cdot \mathbf{y} \le 0$$

SOSC:

$$\mathbf{y}^T \cdot \nabla_x^2 L(\mathbf{x}^*, \lambda) \cdot \mathbf{y} < 0$$

Optimization Techniques

NECESSARY & SUFFICIENT

- The necessary conditions are required for a point to be an extremum but even if they are satisfied, they do not guarantee that the point is an extremum.
- If the sufficient conditions are true, then the point is guaranteed to be an extremum. But if they are not satisfied, this does not mean that the point is not an extremum.

Optimization Techniques

PROCEDURE

1. Solve the FONC to obtain candidate points.
2. Test the candidate points with the SONC
   - Eliminate any points that do not satisfy the SONC
3. Test the remaining points with the SOSC
   - The points that satisfy them are min/max's
   - For the points that do not satisfy, we cannot say whether they are extreme points or not

Optimization Techniques

PROBLEMS WITH INEQUALITY CONSTRAINTS

We will consider problems such as:

Minimize $f(\mathbf{x})$

Subject to: $h_i(\mathbf{x}) = 0 \quad i = 1, \ldots, m$

& $g_j(\mathbf{x}) \le 0 \quad j = 1, \ldots, p$

An inequality constraint, $g_j(\mathbf{x}) \le 0$ is called "active" at $\mathbf{x}^*$ if $g_j(\mathbf{x}^*) = 0$.

Let the set $I(\mathbf{x}^*)$ contain all the indices of the active constraints at $\mathbf{x}^*$:

$$g_j(\mathbf{x}^*) = 0 \quad \text{for all } j \text{ in set } I(\mathbf{x}^*)$$

Optimization Techniques

LAGRANGIAN FOR EQUALITY & INEQUALITY CONSTRAINT PROBLEMS

The Lagrangian is written:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \cdot h_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j \cdot g_j(\mathbf{x})$$

We use λ's for the equalities & μ's for the inequalities.

Optimization Techniques

FONC FOR EQUALITY & INEQUALITY CONSTRAINTS

For the general Lagrangian, the FONC become

$$\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j^* \cdot \nabla g_j(\mathbf{x}^*) = \mathbf{0}$$

and the complementary slackness condition:

$$\mu_j^* \geq 0, \quad \mu_j^* \cdot \left[g_j(\mathbf{x}^*)\right] = 0, \quad j = 1, \ldots p$$

Optimization Techniques

SONC FOR EQUALITY & INEQUALITY CONSTRAINTS

The SONC (for a minimization problem) are:

$$\mathbf{y}^T \cdot \nabla_{\mathbf{x}}^2 L(\mathbf{x}^*, \lambda^*, \mu^*) \cdot \mathbf{y} \geq \mathbf{0}$$

where $\mathbf{J}(\mathbf{x}^*) \cdot \mathbf{y} = \mathbf{0}$ as before.

This time, $\mathbf{J}(\mathbf{x}^*)$ is the matrix of the gradients of all the equality constraints and only the inequality constraints that are active at $\mathbf{x}^*$.

Optimization Techniques

SOSC FOR EQUALITY & INEQUALITY CONSTRAINTS

- The SOSC for a minimization problem with equality & inequality constraints are:

$$\mathbf{y}^T \cdot \nabla_x^2 L(\mathbf{x}^*, \lambda^*, \mu^*) \cdot \mathbf{y} > \mathbf{0}$$

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

- Solve the problem:

Minimize $f(\mathbf{x}) = (x_1 - 1)^2 + (x_2)^2$

Subject to: $h(\mathbf{x}) = (x_1)^2 + (x_2)^2 + x_1 + x_2 = 0$

$g(\mathbf{x}) = x_1 - (x_2)^2 \leq 0$

The Lagrangian for this problem is:

$$L(\mathbf{x}, \lambda, \mu) = (x_1 - 1)^2 + (x_2)^2 + \lambda \cdot \left\{(x_1)^2 + (x_2)^2 + x_1 + x_2\right\} + \mu \cdot \left\{x_1 - (x_2)^2\right\}$$

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

- The first order necessary conditions:

$$\frac{\partial L}{\partial x_1} = 2(x_1 - 1) + 2 \cdot \lambda \cdot x_1 + \lambda + \mu = 0$$

$$\frac{\partial L}{\partial x_2} = 2x_2 + 2 \cdot \lambda \cdot x_2 + \lambda - 2 \cdot \mu \cdot x_2 = 0$$

$$\frac{\partial L}{\partial \lambda} = (x_1)^2 + (x_2)^2 + x_1 + x_2 = 0$$

$$\mu \cdot \left(x_1 - (x_2)^2\right) = 0$$

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

Solving the 4 FONC equations, we get 2 solutions:

1) $\mathbf{x}^{(1)} = \begin{bmatrix} 0.2056 \\ -0.4534 \end{bmatrix}$ $\lambda = 0.45$ & $\mu = 0.9537$

and 2) $\mathbf{x}^{(2)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\lambda = 0$ & $\mu = -2$

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

Now try the SONC at the 1st solution:

Both $h(\mathbf{x})$ & $g(\mathbf{x})$ are active at this point (they both equal zero). So, the Jacobian is the gradient of both functions evaluated at $\mathbf{x}^{(1)}$:

$$\mathbf{J}\left(\mathbf{x}^{(1)}\right) = \begin{bmatrix} 2x_1+1 & 2x_2+1 \\ -1 & 2x_2 \end{bmatrix}_{\mathbf{x}^{(1)}} = \begin{bmatrix} 1.411 & 0.0932 \\ -1 & -0.9068 \end{bmatrix}$$

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

The only solution to the equation:

$$\mathbf{J}(\mathbf{x}^{(1)}) \cdot \mathbf{y} = \mathbf{0}$$

is: $\mathbf{y} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

And the Hessian of the Lagrangian is:

$$\nabla_{\mathbf{x}}^2 L = \begin{bmatrix} 2+2\lambda & 0 \\ 0 & 2+2\lambda-2\mu \end{bmatrix}_{\mathbf{x}^{(1)}} = \begin{bmatrix} 2.9 & 0 \\ 0 & 0.993 \end{bmatrix}$$

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

So, the SONC equation is:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2.9 & 0 \\ 0 & 0.993 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \geq \mathbf{0}$$

This inequality is true, so the SONC is satisfied for $\mathbf{x}^{(1)}$ and it is still a candidate point.

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

The SOSC equation is:

$$\mathbf{y}^T \cdot \nabla_{\mathbf{x}}^2 L(\mathbf{x}^*, \lambda^*, \mu^*) \cdot \mathbf{y} > \mathbf{0}$$

And we just calculated the left-hand side of the equation to be the zero matrix. So, in our case for $\mathbf{x}^2$:

$$\mathbf{y}^T \cdot \nabla_{\mathbf{x}}^2 L\big|_{\mathbf{x}^{(1)}} \cdot \mathbf{y} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \not> \mathbf{0}$$

So, the SOSC are <u>not</u> satisfied.

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

For the second solution:

Again, both $h(\mathbf{x})$ & $g(\mathbf{x})$ are active at this point. So, the Jacobian is:

$$\mathbf{J}\left(\mathbf{x}^{(2)}\right) = \begin{bmatrix} 2x_1+1 & 2x_2+1 \\ -1 & 2x_2 \end{bmatrix}_{\mathbf{x}^{(2)}} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$$

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

The only solution to the equation:

$$\mathbf{J}(\mathbf{x}^{(2)}) \cdot \mathbf{y} = \mathbf{0}$$

is: $\mathbf{y} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

And the Hessian of the Lagrangian is:

$$\nabla_{\mathbf{x}}^2 L = \begin{bmatrix} 2+2\lambda & 0 \\ 0 & 2+2\lambda-2\mu \end{bmatrix}\Bigg|_{\mathbf{x}^{(2)}} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

So, the SONC equation is:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \geq \mathbf{0}$$

This inequality is true, so the SONC is satisfied for $\mathbf{x}^{(2)}$ and it is still a candidate point

Optimization Techniques

GENERALIZED LAGRANGIAN EXAMPLE

The SOSC equation is:

$$\mathbf{y}^T \cdot \nabla_{\mathbf{x}}^2 L(\mathbf{x}^*, \lambda^*, \mu^*) \cdot \mathbf{y} > \mathbf{0}$$

And we just calculated the left-hand side of the equation to be the zero matrix. So, in our case for $\mathbf{x}^2$:

$$\mathbf{y}^T \cdot \nabla_{\mathbf{x}}^2 L\big|_{\mathbf{x}^{(2)}} \cdot \mathbf{y} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \not> \mathbf{0}$$

So, the SOSC are <u>not</u> satisfied.

Optimization Techniques

EXAMPLE CONCLUSIONS

- So, we can say that both $\mathbf{x}^{(1)}$ & $\mathbf{x}^{(2)}$ may be local minimums, but we cannot be sure because the SOSC are not satisfied for either point.

Optimization Techniques

NUMERICAL METHODS

- As you can see from this example, the most difficult step is to solve a system of nonlinear equations to obtain the candidate points.
- Instead of taking gradients of functions, automated NLP solvers use various methods to change a general NLP into an easier optimization problem.

Optimization Techniques

CONCLUSIONS

- So, by varying the initial values, we can get both of the candidate points we found previously
- However, the NLP solver tells us that they are both local minimum points

Optimization Techniques

REFERENCES

Material for this chapter has been taken from:

- Optimization of Chemical Processes 2nd Ed.; Edgar, Thomas; David Himmelblau; & Leon Lasdon.