


Optimization
Techniques

MATHEMATICAL METHODS OF OPTIMIZATION

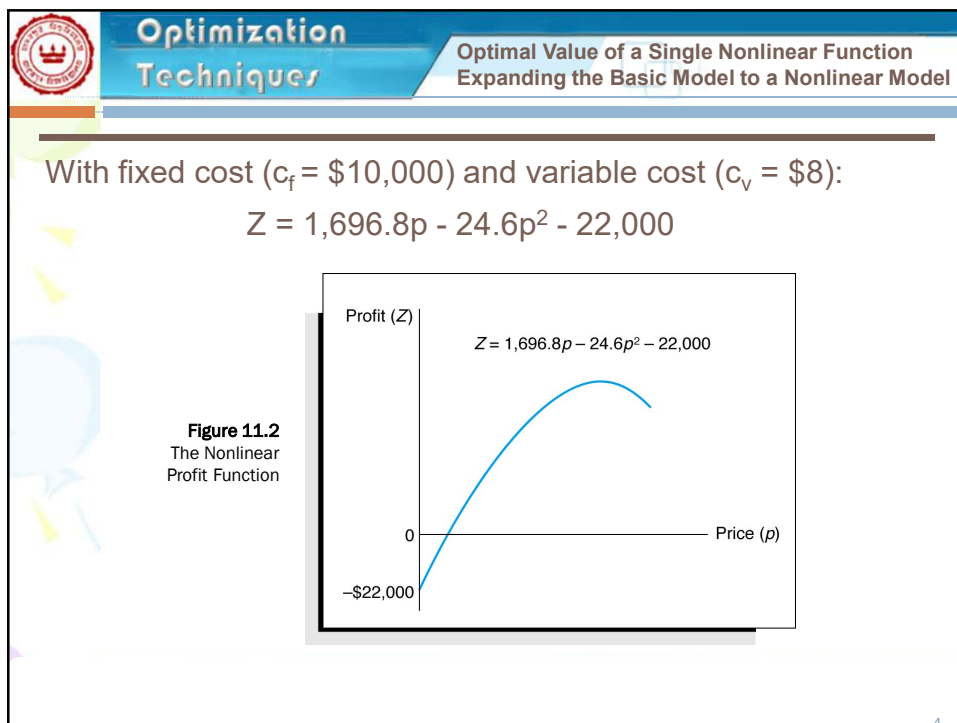
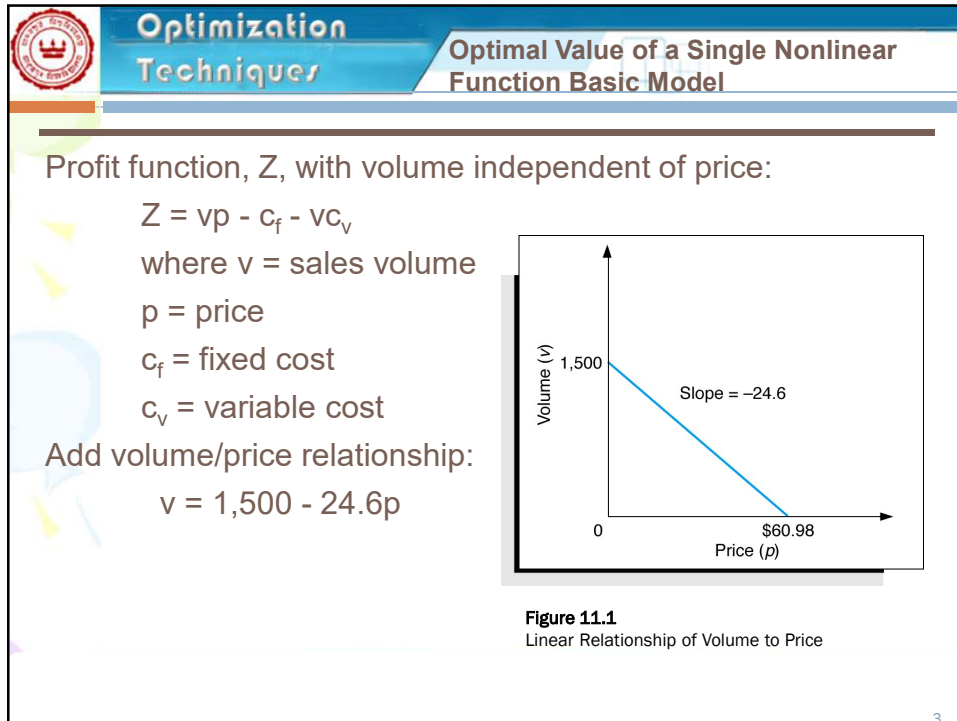
Nonlinear Programming

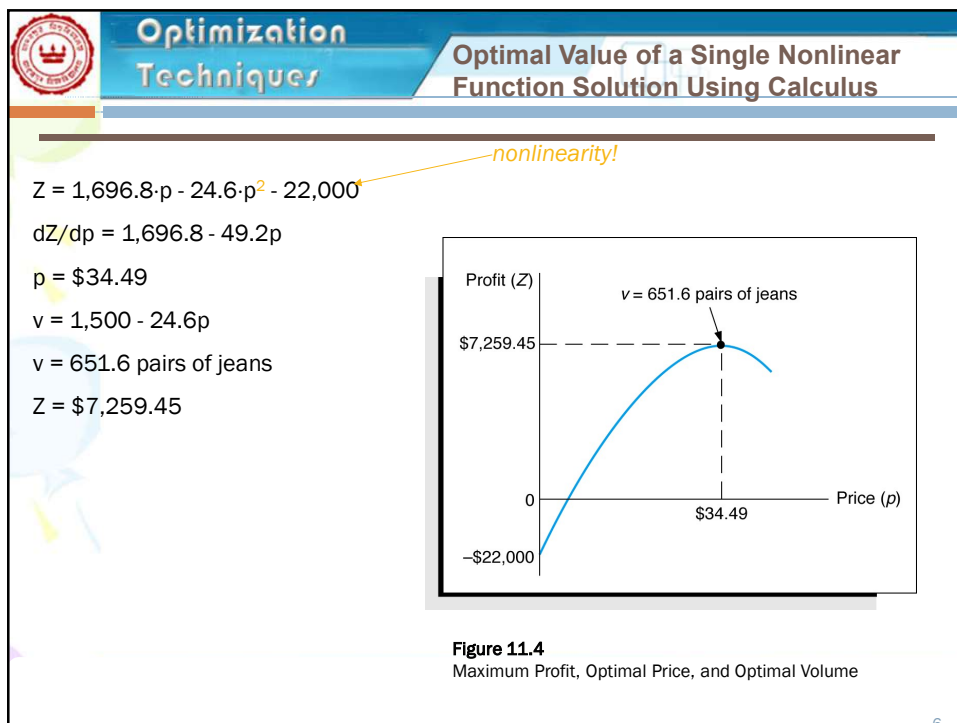
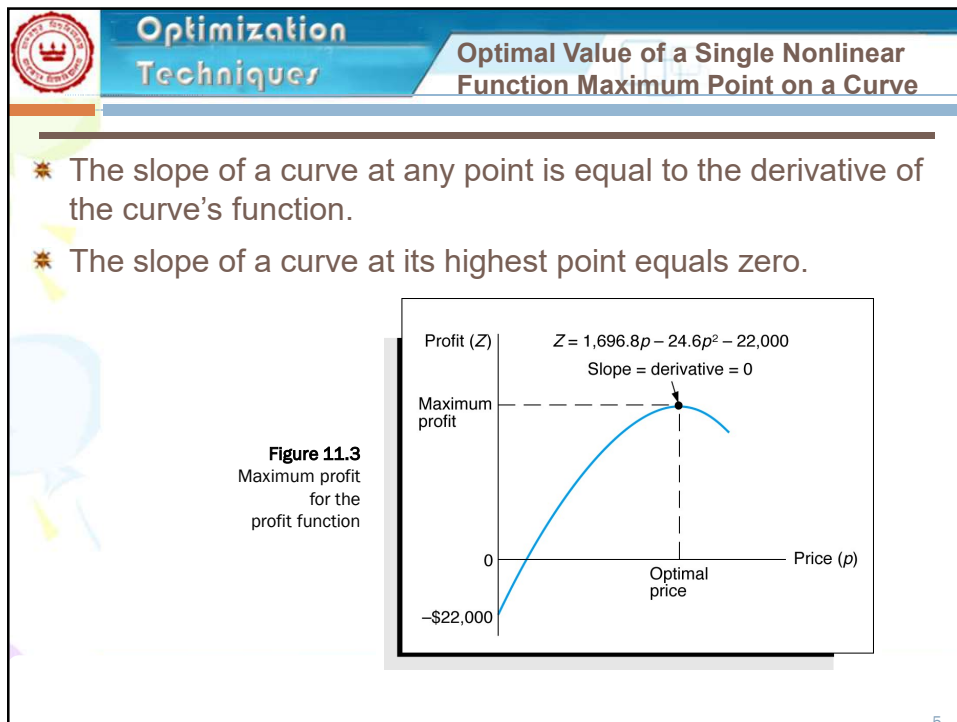



Optimization
Techniques

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 programming
- ✦ 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

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.

7



Optimization Techniques

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

✦ Effect of adding constraints to nonlinear problem:

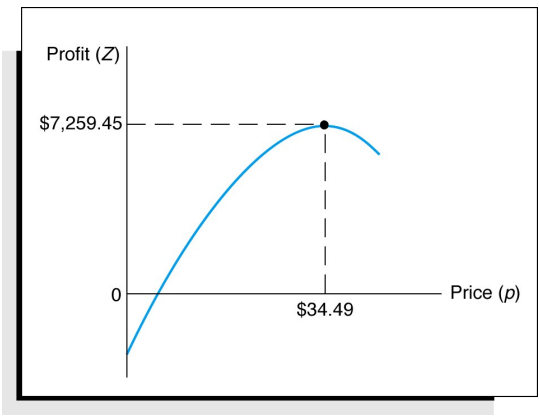
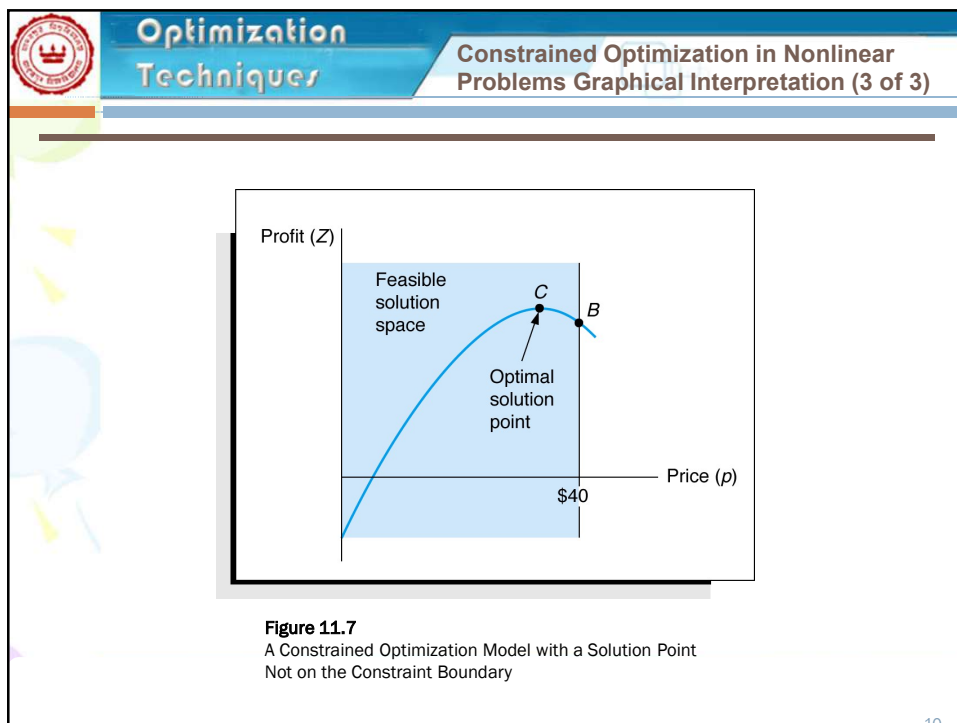
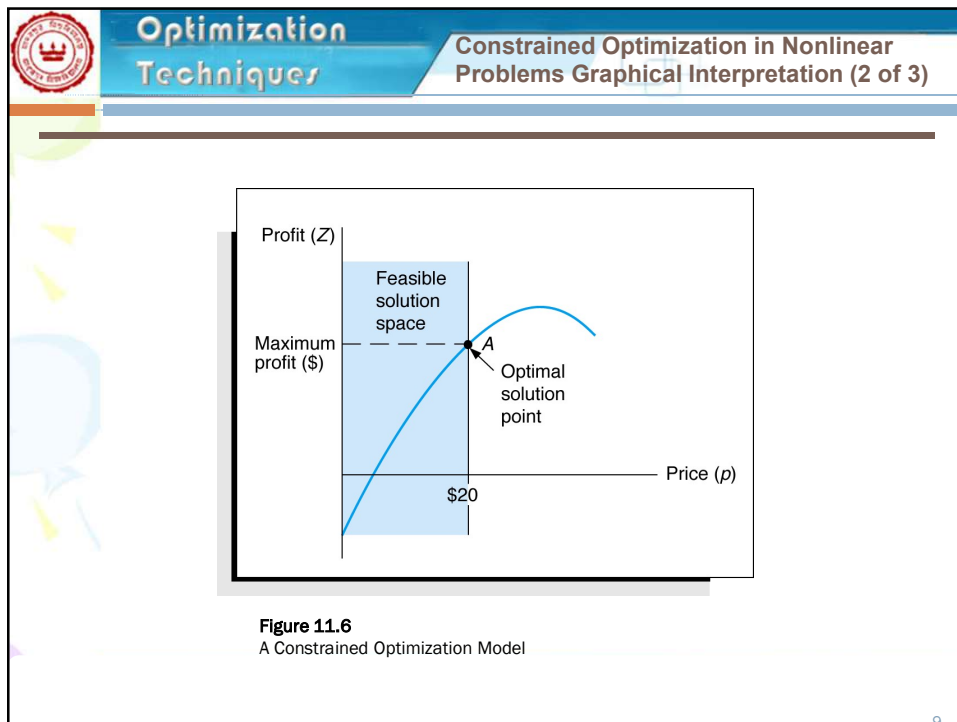



Figure 11.5
Nonlinear Profit Curve for the Profit Analysis Model

8






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.


11



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, \dots, x_n :

$$\nabla f = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_n} \right]$$

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

$$\nabla f = \left[15 - 3(x_3)^2 \quad 6(x_2)^2 \quad -6x_1x_3 \right]$$

Optimization Techniques **THE HESSIAN**

➤ The Hessian (∇^2) of $f(x_1, x_2, \dots, x_n)$ 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):

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

$$\nabla f = [15 - 3(x_3)^2 \quad 6(x_2)^2 \quad -6x_1x_3]$$

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




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)



Optimization Techniques

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)



Optimization Techniques

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 $Ax = \lambda x$. Then λ is an **eigenvalue** (or **characteristic value**) of A .

Note: The zero vector can not be an eigenvector even though $A0 = \lambda 0$. But $\lambda = 0$ 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.




Optimization Techniques

GEOMETRIC INTERPRETATION OF EIGENVALUES AND EIGENVECTORS

An $n \times n$ matrix A multiplied by $n \times 1$ vector x results in another $n \times 1$ 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.



Optimization Techniques

EIGENVALUES

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

$$Ax - \lambda x = 0 \quad \text{or}$$

$$(A - \lambda I)x = 0$$


If we define a new matrix $B = A - \lambda I$, 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

$$\det(A - \lambda I) = 0$$

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



Optimization Techniques

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, $\lambda_1 = \lambda_2 = \dots = \lambda_k$. If that happens, the eigenvalue is said to be of multiplicity k .

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

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

3. $\lambda = 2$ is an eigenvector of multiplicity



Optimization Techniques

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
Recall that $\lambda = 2$ is an eigenvector of multiplicity 3.


Solve the homogeneous linear system represented by

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

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

Let $x_1 = s, x_3 = t$. The eigenvectors of $\lambda = 2$ are of the form

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



Optimization Techniques

PROPERTIES OF EIGENVALUES AND EIGENVECTORS


Definition: The trace of a matrix A , designated by $\text{tr}(A)$, 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 $1/\lambda$ is an eigenvalue of matrix A^{-1} .



Optimization Techniques


PROPERTIES OF EIGENVALUES AND EIGENVECTORS

Property 5: If λ is an eigenvalue of A then $k\lambda$ 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 $n \times n$ matrix A then the number of linearly independent eigenvectors of A associated with λ is given by $m = n - r(A - \lambda I)$. Furthermore, $1 \leq m \leq k$.

Example 2 (cont.): The eigenvectors of $\lambda = 2$ are of the form

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

$\lambda = 2$ 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:

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

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:

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

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:

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

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:

$$P(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$$

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

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

$f(x) = \ln(x+1)$ $f(0) = \ln(1) = 0$	$P(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$ $P(0) = a_0 \rightarrow a_0 = 0$
$f'(x) = \frac{1}{1+x}$ $f'(0) = \frac{1}{1} = 1$	$P'(x) = a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3$ $P'(0) = a_1 \rightarrow a_1 = 1$
$f''(x) = -\frac{1}{(1+x)^2}$ $f''(0) = -\frac{1}{1} = -1$	$P''(x) = 2a_2 + 6a_3x + 12a_4x^2$ $P''(0) = 2a_2 \rightarrow a_2 = -\frac{1}{2}$

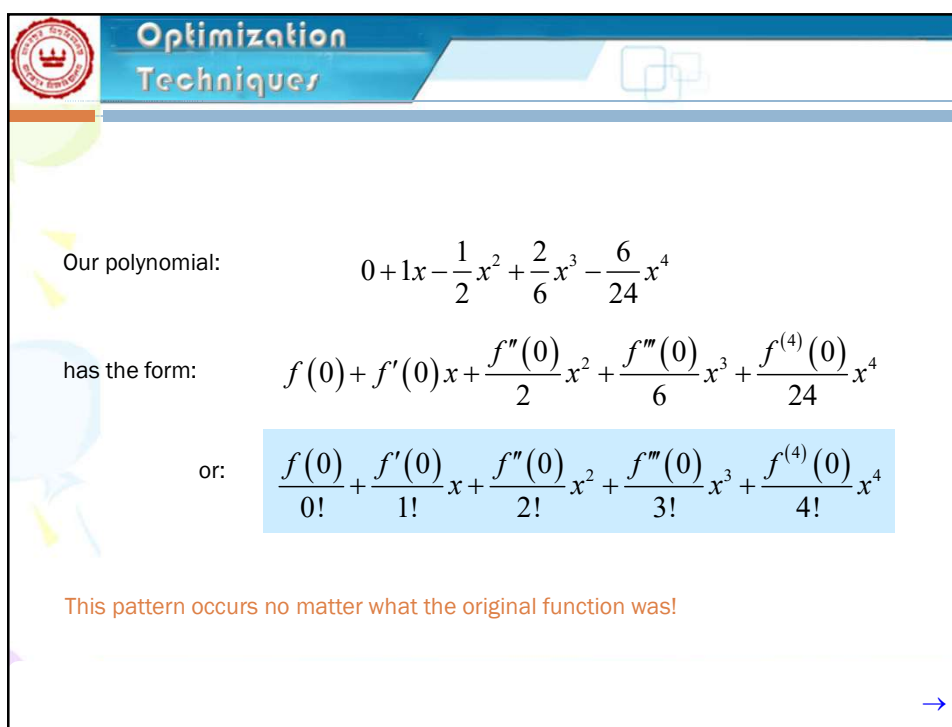
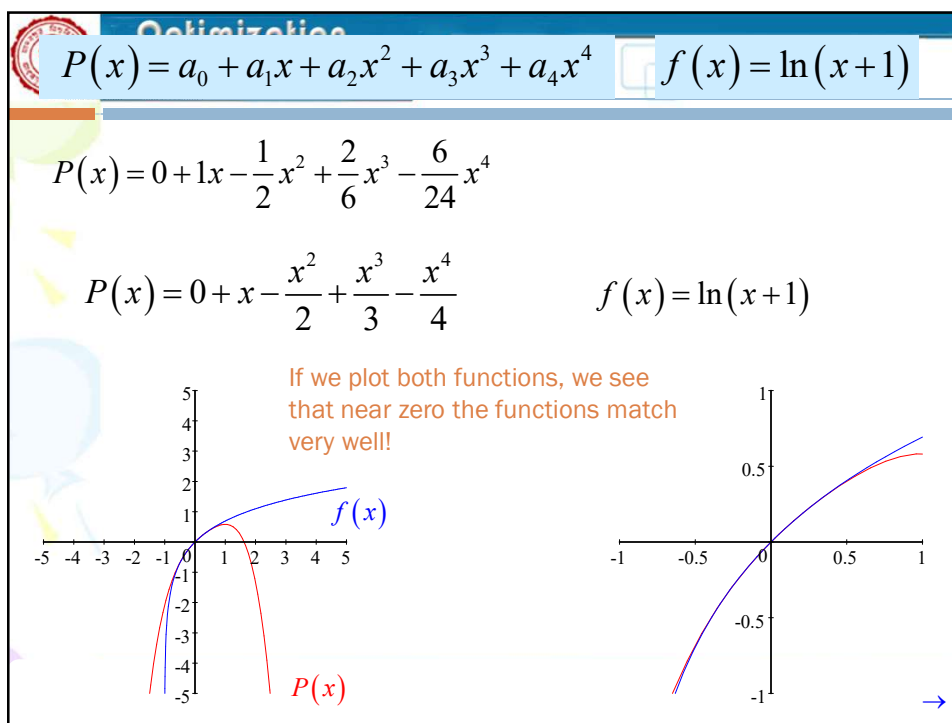
→

Optimization

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

$f''(x) = -\frac{1}{(1+x)^2}$ $f''(0) = -\frac{1}{1} = -1$	$P''(x) = 2a_2 + 6a_3x + 12a_4x^2$ $P''(0) = 2a_2 \rightarrow a_2 = -\frac{1}{2}$
$f'''(x) = 2 \cdot \frac{1}{(1+x)^3}$ $f'''(0) = 2$	$P'''(x) = 6a_3 + 24a_4x$ $P'''(0) = 6a_3 \rightarrow a_3 = \frac{2}{6}$
$f^{(4)}(x) = -6 \frac{1}{(1+x)^4}$ $f^{(4)}(0) = -6$	$P^{(4)}(x) = 24a_4$ $P^{(4)}(0) = 24a_4 \rightarrow a_4 = -\frac{6}{24}$

→



Optimization

Maclaurin Series:

(generated by f at $x = 0$)

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

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

Taylor Series:

(generated by f at $x = a$)

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

Optimization Techniques

example: $y = \cos x$

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

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

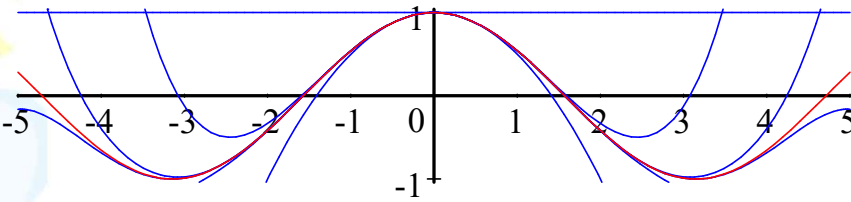
$$f''(x) = -\cos x \quad f''(0) = -1$$

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

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

Optimization Techniques

$y = \cos x$

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


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:

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

→

Optimization Techniques

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

$$f(x) = \cos x \quad f\left(\frac{\pi}{2}\right) = 0 \quad f'''(x) = \sin x \quad f'''\left(\frac{\pi}{2}\right) = 1$$

$$f'(x) = -\sin x \quad f'\left(\frac{\pi}{2}\right) = -1 \quad f^{(4)}(x) = \cos x \quad f^{(4)}\left(\frac{\pi}{2}\right) = 0$$

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

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

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

Optimization Techniques

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

Optimization

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 Tech

The TI-89 finds Taylor Polynomials:


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

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

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


taylor (cos(x), x, 5, $\pi/2$) $\frac{-(2x-\pi)^5}{3840} + \frac{(2x-\pi)^3}{48} - \frac{2x-\pi}{2}$ 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*
- $f_1(x) = f(x^k) + \nabla f(x^k)^T (x - x^k)$
- $f_1(x) = f(x^k) + \sum_{j=1}^n \frac{\partial f(x^k)}{\partial x_j} (x_j - x_j^k)$
- 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*
- $f_2(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T H(x^k) (x - x^k)$
- $f_1(x) = f(x^k) + \sum_{j=1}^n \frac{\partial f(x^k)}{\partial x_j} (x_j - x_j^k) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(x^k)}{\partial x_i \partial x_j} (x_i - x_i^k) (x_j - x_j^k)$


$f(x_1, x_2) = x_1^4 + x_1^2 + 2x_2^2 - 2x_1x_2$ 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:

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

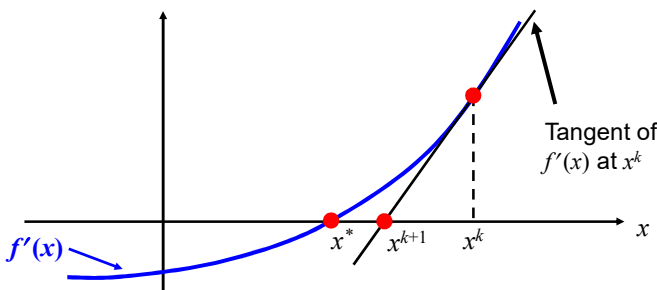
where k is the current iteration.

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^* .



Optimization Techniques

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



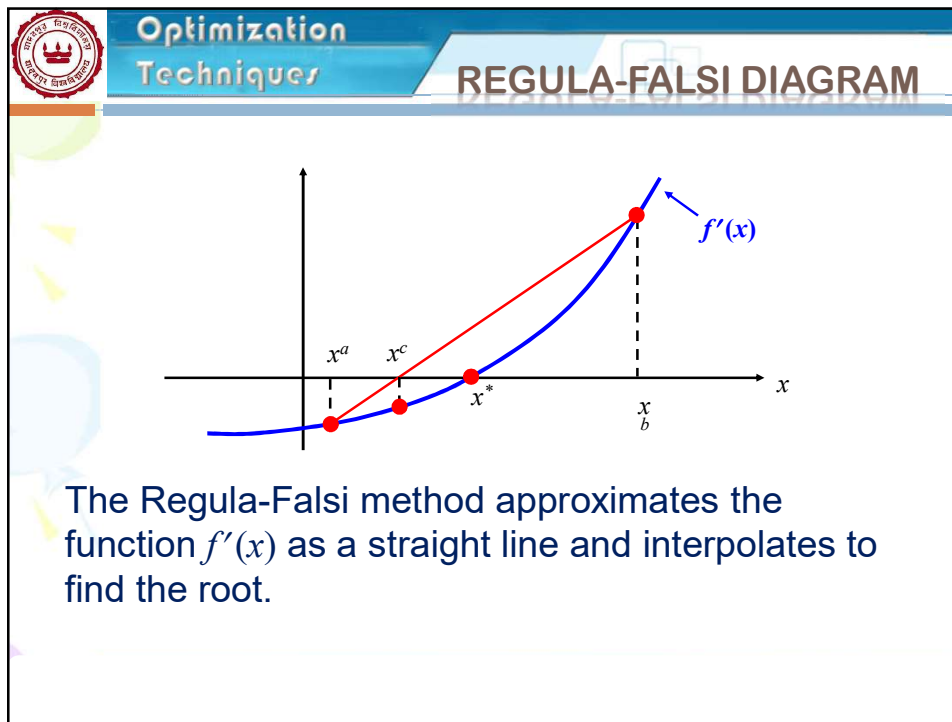
Optimization Techniques

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 .



Optimization Techniques

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