

ASEN 5070  
Statistical Orbit determination I

Fall 2012



Professor George H. Born  
Professor Jeffrey S. Parker

Lecture 4: Coding and Linear Algebra Review



University of Colorado  
Boulder

# Announcements

- ▶ Homework 1 due today
- ▶ Homework 2 due in 7 days
- ▶ I will most likely not be available during my Monday office hours. Definitely use the TAs – I hear they're bored.



# Quiz Results

**Question 1 (1 point)**

If you have two standard six-sided dice and you roll each of them once, what is the probability that neither die lands with a "1" up? (both dice have one and only one face with a "1" on it).

1/36

11/36

25/36

5/6



# Quiz Results

**Question 1 (1 point)**

If you have two standard six-sided dice and you roll each of them once, what is the probability that neither die lands with a "1" up? (both dice have one and only one face with a "1" on it).

1/36

11/36

25/36

5/6



# Quiz Results

## Question 2 (1 point)

If a number is randomly selected from a normal gaussian distribution with a mean of zero and a standard deviation of 1, i.e.,  $N(0,1)$ , what is the probability that the number will be between -1 and +1?

- 0%
- The probability is somewhere between 0% and 50%
- The probability is somewhere between 50% and 100%
- 100%



# Quiz Results

## Question 2 (1 point)

If a number is randomly selected from a normal gaussian distribution with a mean of zero and a standard deviation of 1, i.e.,  $N(0,1)$ , what is the probability that the number will be between -1 and +1?

- 0%
- The probability is somewhere between 0% and 50%
- The probability is somewhere between 50% and 100%
- 100%



# Quiz Results

**Question 3 (1 point)**

What is the expected value for a number drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 1?

0

1

2

3



# Quiz Results

**Question 3 (1 point)**

What is the expected value for a number drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 1?

0

1

2

3



# Quiz Results

**Question 4 (1 point)**



All square matrices are invertible.

- True
- False

# Quiz Results

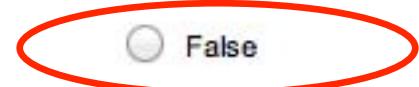
**Question 4 (1 point)**



All square matrices are invertible.

True

False

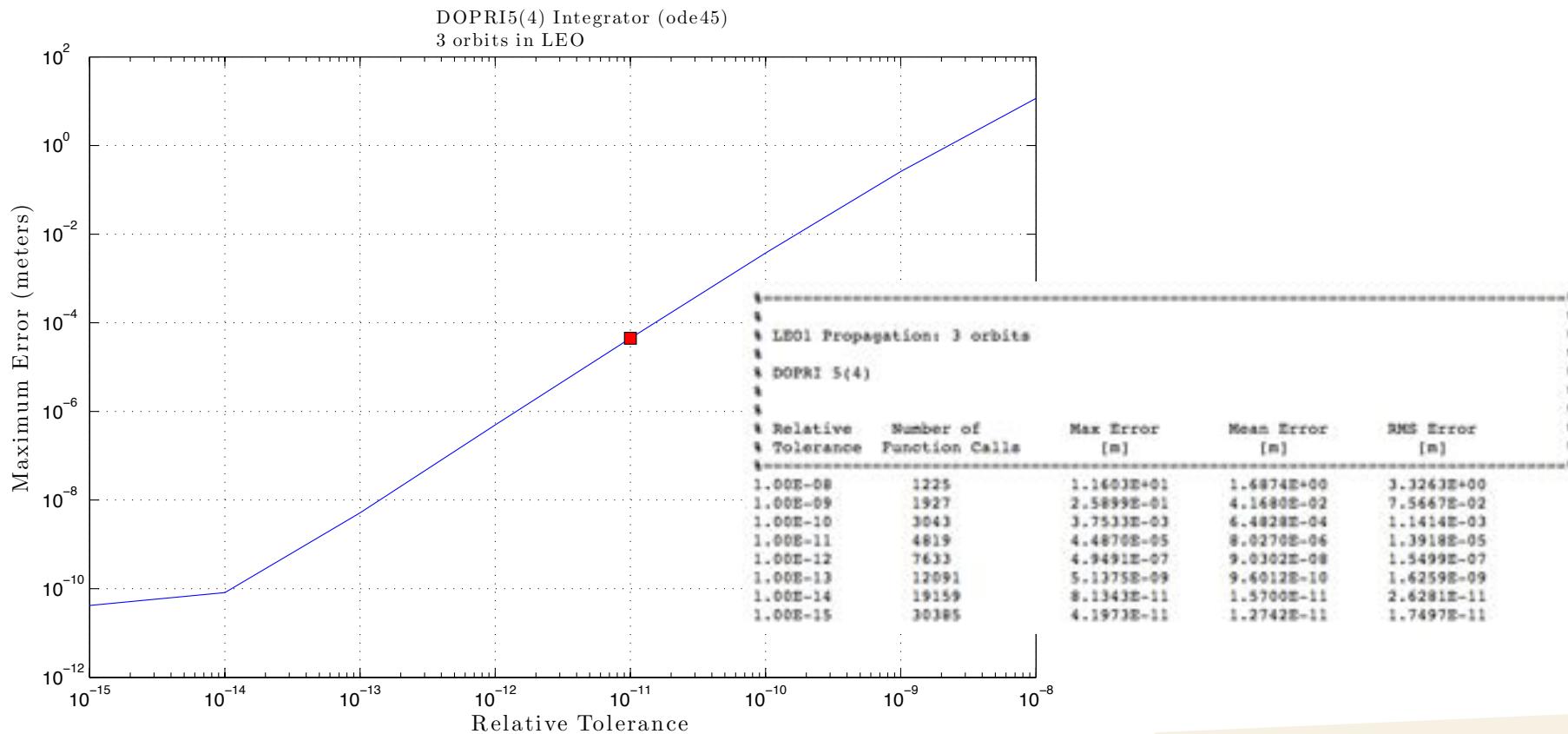


# Today's Lecture

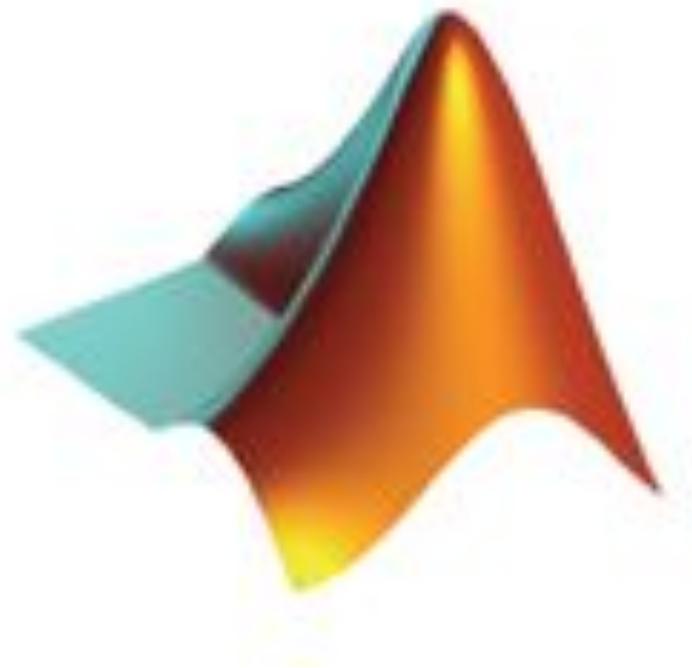
- ▶ Coding hints and tricks
  - MATLAB: ways to speed up your code
  - Python: intro
- ▶ Review of Linear Algebra
- ▶ Review of Statistics Tuesday

# First: 1 slide on integration tolerances

- ▶ ode45's default tolerance:  $1e-6$
- ▶ What should you set it to be?



# Matlab Overview



# Generic Tips: Pre-allocate

## Inefficient MATLAB

```
x = 1:1:100;  
for i = 1:length(x)  
    y(i) = sqrt(x(i));  
end
```

## Better MATLAB

```
x = 1:1:100;  
y = zeros(size(x));  
for i = 1:length(x)  
    y(i) = sqrt(x(i));  
end
```



- ▶ Learn to use help function/online resources
- ▶ % Comment your code!!!
- ▶ Name functions and outputs descriptively



# Symbolic Toolbox

MATLAB

```
syms x y z

radius = sqrt( x^2 + y^2 + z^2 );

vec = [ x; y; z ];

drdx = diff( radius, x );
drdz = diff( radius, z );

radius = int( drdx, 'x' );
radius = int( drdz, 'z' );

drdy = jacobian( radius, vec );
```



# reshape() Command

Use to convert STM from matrix to a vector so it can be numerically integrated

## MATLAB: reshape() Example

```
>> x = [ 1, 2; 3, 4 ];  
  
>> y = reshape( x, 4,1 );  
  
y =  
    1  
    2  
    3  
    4
```





- ▶ Python is a free, open source programming language that runs on nearly everything.
- ▶ Resources:
  - <http://www.python.org/>
  - Python 3 Tutorial:  
<http://docs.python.org/py3k/tutorial/>
  - IDEs:
    - Xcode
    - Emacs / xemacs / vi

- ▶ Python is high-level
  - No memory management requirements by the user.  
Hurrah!
- ▶ Python is object oriented.
  - Matlab is a little, but Python is a lot.
- ▶ Python can do anything that Matlab can do,  
but differently.
  - Some things are easier, some harder.

# Python

- ▶ Working on some example tutorials that parallel the Matlab tutorials.



# Questions

- ▶ Questions on Coding?
- ▶ Quick Break
- ▶ Next topics:
  - Review of Linear Algebra
  - Review of Statistics



# Notation

The following matrix notation, definitions, and theorems are used extensively in this class.

- A matrix  $\mathbf{A}$  will have elements denoted by  $a_{ij}$ , where  $i$  refers to the row and  $j$  to the column.
- $\mathbf{A}^T$  will denote the transpose of  $\mathbf{A}$ .
- $\mathbf{A}^{-1}$  will denote the inverse of  $\mathbf{A}$ .
- $|\mathbf{A}|$  will denote the determinant of  $\mathbf{A}$ .
- The dimension of a matrix is the number of its rows by the number of its columns.
- An  $n \times m$  matrix  $\mathbf{A}$  will have  $n$  rows and  $m$  columns.
- If  $m = 1$ , the matrix will be called an  $n \times 1$  vector.



# Matrix Multiplication

- Given

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

and

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

- then

$$\begin{aligned}
 \mathbf{AB} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix} \quad (\text{B.1.2})
 \end{aligned}$$

provided the elements of  $\mathbf{A}$  and  $\mathbf{B}$  are conformable.



# Fundamentals

- For  $\mathbf{A} + \mathbf{B}$  to be defined,  $\mathbf{A}$  and  $\mathbf{B}$  must have the same dimension.
- The transpose of  $\mathbf{A}^T$  equals  $\mathbf{A}$ ; that is,  $(\mathbf{A}^T)^T = \mathbf{A}$ .
- The inverse of  $\mathbf{A}^{-1}$  is  $\mathbf{A}$ ; that is,  $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ .
- The transpose and inverse symbols may be permuted; that is,  $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$ .
- $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$ .
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$  if  $\mathbf{A}$  and  $\mathbf{B}$  are each nonsingular.



# Fundamentals

- $(AB)C = A(BC)$ , the associative law holds.
- In general  $AB \neq BA$ , the commutative law does not hold.
- From  $AB = 0$  we cannot in general conclude that at least one of  $A$  or  $B = 0$ .
- From  $AB = AC$  we cannot in general conclude that  $B = C$ .
- If  $\mathbf{X}$  and  $\mathbf{Y}$  are vectors and if  $\mathbf{A}$  is a nonsingular matrix and if the equation  $\mathbf{Y} = \mathbf{AX}$  holds, then  $\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}$ .



# Matrix Rank

- The *rank* of a matrix is the dimension of its largest square nonsingular sub-matrix; that is, one whose determinant is nonzero.



# Matrix Rank

- The *rank* of a matrix is the dimension of its largest square nonsingular submatrix; that is, one whose determinant is nonzero.

Example: What is the rank of the following matrices?

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 2 & 2 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix}$$



# Matrix Rank

- The *rank* of a matrix is the dimension of its largest square nonsingular submatrix; that is, one whose determinant is nonzero.
- The rank of the product  $\mathbf{AB}$  of the two matrices  $\mathbf{A}$  and  $\mathbf{B}$  is less than or equal to the rank of  $\mathbf{A}$  and is less than or equal to the rank of  $\mathbf{B}$ .
- If  $\mathbf{A}$  is an  $n \times n$  matrix and if  $|\mathbf{A}| = 0$ , then the rank of  $\mathbf{A}$  is less than  $n$ .
- If the rank of  $\mathbf{A}$  is  $m \leq n$ , then the number of linearly independent rows is  $m$ ; also, the number of linearly independent columns is  $m$  ( $\mathbf{A}$  is  $n \times n$ ).
- The rank of  $\mathbf{AA}^T$  equals the rank of  $\mathbf{A}^T\mathbf{A}$ , equals the rank of  $\mathbf{A}$ , equals the rank of  $\mathbf{A}^T$ .



# Quadratic Forms

- The rank of the quadratic form  $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$  is defined as the rank of the matrix  $\mathbf{A}$  where  $\mathbf{Y}$  is a vector and  $\mathbf{Y} \neq 0$ .
- The quadratic form  $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$  is said to be *positive definite* if and only if  $\mathbf{Y}^T \mathbf{A} \mathbf{Y} > 0$  for all vectors  $\mathbf{Y}$  where  $\mathbf{Y} \neq 0$ .
- A quadratic form  $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$  is said to be *positive semidefinite* if and only if  $\mathbf{Y}^T \mathbf{A} \mathbf{Y} \geq 0$  for all  $\mathbf{Y}$ , and  $\mathbf{Y}^T \mathbf{A} \mathbf{Y} = 0$  for some vector  $\mathbf{Y} \neq 0$ .



# Quadratic Forms

- A quadratic form  $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$  that may be either positive definite or positive semidefinite is called *nonnegative definite*.
- The matrix  $\mathbf{A}$  of a quadratic form  $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$  is said to be positive definite (semidefinite) when the quadratic form is positive definite (semidefinite).
- If  $\mathbf{A}$  is an  $m \times n$  matrix of rank  $n < m$ , then  $\mathbf{A}^T \mathbf{A}$  is positive definite and  $\mathbf{A} \mathbf{A}^T$  is positive semidefinite.
- If  $\mathbf{A}$  and  $\mathbf{B}$  are symmetric conformable matrices,  $\mathbf{A}$  is said to be greater than  $\mathbf{B}$  if  $\mathbf{A} - \mathbf{B}$  is nonnegative definite.



# Triangle Matrices

- A triangular matrix has non-zero elements on the diagonal and above (upper triangular) or below (lower triangular).
- Example of a  $3 \times 3$  upper triangular matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

- A unitary triangular matrix has diagonal elements equal to 1.



# Matrix Square Root

- The square root of an  $n \times n$  matrix,  $P$ , is defined as  $P=AA$ , where  $A = \sqrt{P}$  is the square root of  $P$ .
- A symmetric positive semidefinite matrix has a unique symmetric positive semidefinite square root.
- If  $S^T S = P$ , where  $P$  is symmetric positive semidefinite and  $S$  is upper triangular, then an orthogonal matrix  $Q$  exists so that

$$S = QA$$

$$\sqrt{AB} \neq \sqrt{A}\sqrt{B}$$



# Determinants

- For each square matrix  $\mathbf{A}$ , there is a uniquely defined scalar called the *determinant* of  $\mathbf{A}$  and denoted by  $|\mathbf{A}|$ .
- The determinant of a diagonal matrix is equal to the product of the diagonal elements.
- If  $\mathbf{A}$  and  $\mathbf{B}$  are  $n \times n$  matrices, then  $|\mathbf{AB}| = |\mathbf{BA}| = |\mathbf{A}||\mathbf{B}|$ .
- If  $\mathbf{A}$  is singular if and only if  $|\mathbf{A}| = 0$ .



# Determinants

- If  $\mathbf{C}$  is an  $n \times n$  matrix such that  $\mathbf{C}^T \mathbf{C} = \mathbf{I}$ , then  $\mathbf{C}$  is said to be an *orthogonal* matrix, and  $\mathbf{C}^T = \mathbf{C}^{-1}$ .
- If  $\mathbf{C}$  is an orthogonal matrix, then  $|\mathbf{C}| = \pm 1$ .
- The determinant of a positive definite matrix is positive.
- The determinant of a triangular matrix is equal to the product of the diagonal elements.
- The determinant of a matrix is equal to the product of its eigenvalues.



# Matrix Trace

- The *trace* of a matrix  $\mathbf{A}$ , which will be written  $\text{tr}(\mathbf{A})$ , is equal to the sum of the diagonal elements of  $\mathbf{A}$ ; that is,

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}. \quad (\text{B.5.1})$$

- $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ .
- $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA})$ ; that is, the trace of the product of matrices is invariant under any cyclic permutation of the matrices.
- Note that the trace is defined only for a square matrix.
- If  $\mathbf{C}$  is an orthogonal matrix,  $\text{tr}(\mathbf{C}^T \mathbf{AC}) = \text{tr}(\mathbf{A})$ .



# Eigenvalues and Eigenvectors

- A *characteristic root (eigenvalue)* of a  $p \times p$  matrix  $\mathbf{A}$  is a scalar  $\lambda$  such that  $\mathbf{AX} = \lambda\mathbf{X}$  for some vector  $\mathbf{X} \neq 0$ .
- $\mathbf{X}$  is called the *characteristic vector (eigenvector) of the matrix  $\mathbf{A}$* .
- The eigenvalue of a matrix  $\mathbf{A}$  can be defined as a scalar  $\lambda$  such that  $|\mathbf{A} - \lambda\mathbf{I}| = 0$ .
- $|\mathbf{A} - \lambda\mathbf{I}|$  is a  $p$ th degree polynomial in  $\lambda$ .
- This polynomial is called the *characteristic polynomial*, and its roots are the eigenvalues of the matrix  $\mathbf{A}$ .



# Eigenvalues and Eigenvectors

- The number of nonzero eigenvalues of a matrix  $\mathbf{A}$  is equal to the rank of  $\mathbf{A}$ .
- The trace of  $\mathbf{A}$  is equal to the sum of its eigenvalues.
- The eigenvalues of a symmetric matrix are real.
- The eigenvalues of a positive definite matrix  $\mathbf{A}$  are positive; the eigenvalues of a positive semidefinite matrix are nonnegative.



# Eigenvalues and Eigenvectors

Example:

$$\mathbf{A} = \begin{bmatrix} 4 & 6 & 8 \\ 6 & 16 & 10 \\ 8 & 10 & 25 \end{bmatrix}$$

- The normalized eigenvectors of  $\mathbf{A}$  are:

$$\begin{bmatrix} 0.9486 & 0.0684 & 0.3091 \\ -0.2298 & 0.8204 & 0.5235 \\ -0.2178 & -0.5676 & 0.7940 \end{bmatrix}$$

- The eigenvalues of  $\mathbf{A}$  are:  $D(0.7096, 9.5818, 34.7086)$ 
  - $\mathbf{A}$  is a positive definite symmetric matrix.
  - Rank of  $\mathbf{A} = 3$ : Three nonzero eigenvalues.
  - The eigenvalues of a symmetric matrix are real.
  - The sum of the eigenvalues = the trace of  $\mathbf{A}$ , i.e. 45.



# Derivatives

- Let  $\mathbf{X}$  be an  $n \times 1$  vector and let  $Z$  be a scalar that is a function of  $\mathbf{X}$ . The derivative of  $Z$  with respect to the vector  $\mathbf{X}$ , which will be written  $\partial Z / \partial \mathbf{X}$ , will mean the  $1 \times n$  row vector\*

$$\frac{\partial Z}{\partial \mathbf{X}} = \left[ \frac{\partial Z}{\partial x_1} \quad \frac{\partial Z}{\partial x_2} \quad \dots \quad \frac{\partial Z}{\partial x_n} \right]. \quad (\text{B.7.1})$$

---

\*Generally this partial derivative would be defined as a column vector. However, it is defined as a row vector here because we have defined  $\tilde{H} = \frac{\partial G(\mathbf{X})}{\partial \mathbf{X}}$  as a row vector in the text.

# Derivatives

- If  $\mathbf{A}$  and  $\mathbf{B}$  are  $m \times 1$  vectors, which are a function of the  $n \times 1$  vector  $\mathbf{X}$ , and we define

$$\frac{\partial(\mathbf{A}^T \mathbf{B})}{\partial \mathbf{X}}$$

to be a row vector as in Eq. (B.7.1), then

$$\frac{\partial(\mathbf{A}^T \mathbf{B}) / \partial \mathbf{X}}{} = \mathbf{B}^T \frac{\partial \mathbf{A}}{\partial \mathbf{X}} + \mathbf{A}^T \frac{\partial \mathbf{B}}{\partial \mathbf{X}} \quad (\text{B.7.3})$$

- where  $\frac{\partial \mathbf{A}}{\partial \mathbf{X}}$  is an  $m \times n$  matrix whose  $ij$  element is  $\frac{\partial A_i}{\partial X_j}$
- and  $\frac{\partial(\mathbf{A}^T \mathbf{B})}{\partial \mathbf{X}}$  is a  $1 \times n$  row vector.
- The derivative of a matrix product with respect to a scalar is given by

$$\frac{d}{dt} \{ \mathbf{A}(t) \mathbf{B}(t) \} = \frac{d\mathbf{A}(t)}{dt} \mathbf{B}(t) + \mathbf{A}(t) \frac{d\mathbf{B}(t)}{dt}. \quad (\text{B.7.7})$$



# Maxima and Minima

The goal is to find an extrema (maximum or minimum) of a function  $f(x_1, x_2, \dots, x_n)$  that depends on  $n$  independent variables  $x_1 \dots x_n$ .

- Assuming that first and second order derivatives are continuous, an extrema of  $f(x_1 \dots x_n)$  occurs only at points where

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0$$



# Maxima and Minima

- The extrema will be a maximum if the Hessian matrix  $\mathbf{H}$  is negative definite, where

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n} \end{bmatrix} n \times n$$

- The extrema will be a minimum if  $\mathbf{H}$  is positive definite.



# Maxima and Minima

If there are constraints imposed on the independent variables, i.e., these variables are no longer independent, where the constraints are given by

$$g_l = (x_1, \dots, x_n) = 0, l = 1 \dots m \text{ where } m < n$$

We may adjoin the constraints to the original function  $f(x_1 \dots x_n)$  using a set of constant unknown Lagrange multipliers,  $\lambda_l, l = 1 \dots m$ , to obtain the Lagrangian function

$$L(x_1 \dots x_n, \lambda_1 \dots \lambda_m) = f + \sum_{l=1}^m \lambda_l g_l$$



# Maxima and Minima

- The extrema of  $f$  is now given by

$$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial x_2} \dots \dots \frac{\partial L}{\partial x_n} = 0$$

$$g_l(x_1 \dots x_n) = 0, \quad l = 1 \dots m$$

- These  $n + m$  equations must be solved simultaneously for  $x_1 \dots x_n, \lambda_1 \dots \lambda_m$  in order to obtain the extrema



# Maxima and Minima

- The extrema will be a maximum if the Hessian matrix of  $L$  is negative definite and a minimum if the Hessian is positive definite, where

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 L}{\partial x_1^2} & \cdots & \cdots & \frac{\partial^2 L}{\partial x_1 \partial x_n} \\ \frac{\partial^2 L}{\partial x_2 \partial x_1} & \cdots & \cdots & \frac{\partial^2 L}{\partial x_2 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 L}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 L}{\partial x_n^2} \end{bmatrix}_{n \times n}$$



# Maxima and Minima

Example: Find the extrema of  $f(x_1, x_2) = x_1^2 + x_2^2$

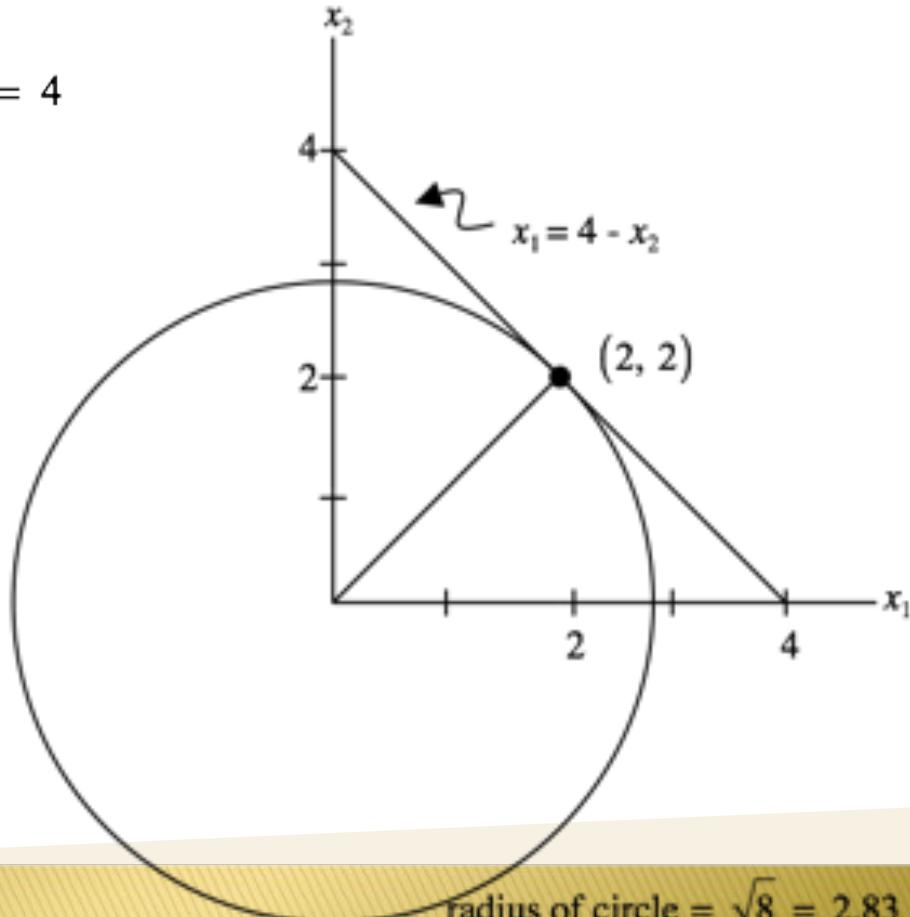
with the constraint  $x_1 + x_2 = 4$

$$L = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 4)$$

$$\frac{\partial L}{\partial x_1} = 2x_1 + \lambda = 0$$

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

$$x_1 = x_2 = 2, \lambda = 4$$



$$\text{radius of circle} = \sqrt{8} = 2.83$$



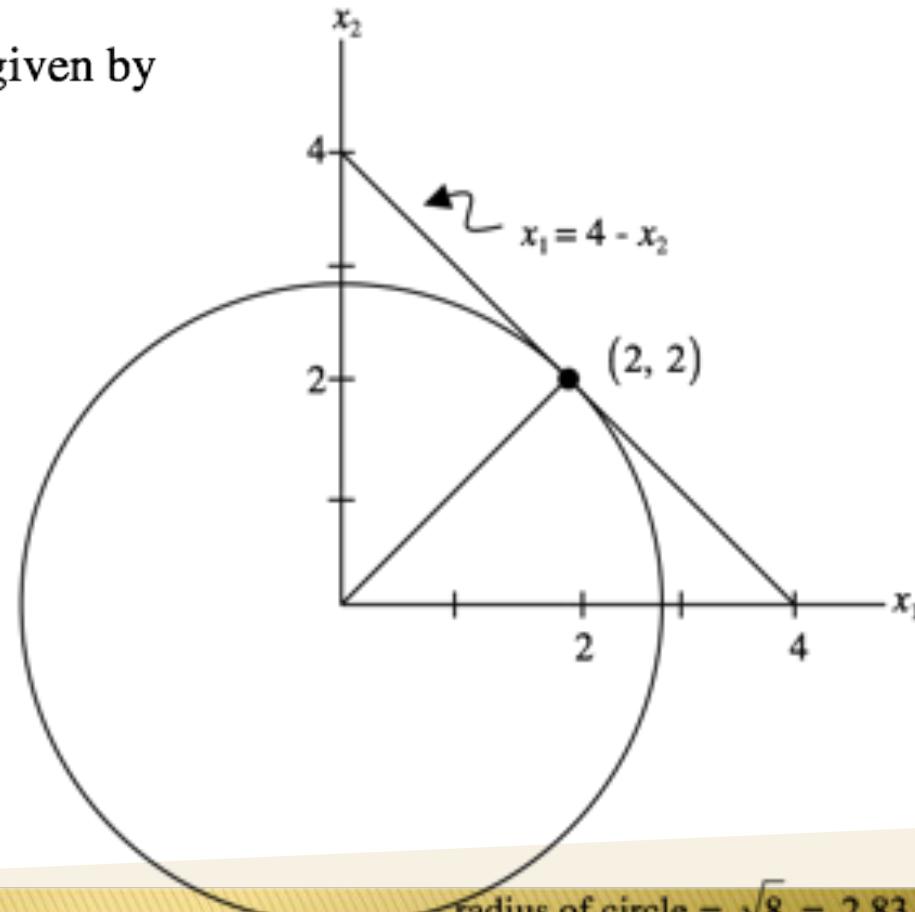
# Maxima and Minima

Example:

The Hessian matrix for this example is given by

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \frac{\partial^2 L}{\partial x_2 \partial x_1} & \frac{\partial^2 L}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

- $\mathbf{H}$  is PD so this extrema is a minimum.



radius of circle =  $\sqrt{8} = 2.83$



University of Colorado  
Boulder

# Matrix Inversion Theorems

**Theorem 1:** Let  $\mathbf{A}$  and  $\mathbf{B}$  be  $n \times n$  positive definite (PD) matrices. If  $\mathbf{A}^{-1} + \mathbf{B}^{-1}$  is PD, then  $\mathbf{A} + \mathbf{B}$  is PD and

$$\begin{aligned}
 (\mathbf{A} + \mathbf{B})^{-1} &= \mathbf{B}^{-1} (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} \mathbf{A}^{-1} \\
 &= \mathbf{A}^{-1} (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} \mathbf{B}^{-1}. \tag{B.9.1}
 \end{aligned}$$

**Theorem 2:** Let  $\mathbf{A}$  and  $\mathbf{B}$  be  $n \times n$  PD matrices. If  $\mathbf{A} + \mathbf{B}$  is PD, then  $\mathbf{I} + \mathbf{AB}^{-1}$  and  $\mathbf{I} + \mathbf{BA}^{-1}$  are PD and

$$\begin{aligned}
 (\mathbf{A} + \mathbf{B})^{-1} &= \mathbf{B}^{-1} - \mathbf{B}^{-1} (\mathbf{I} + \mathbf{AB}^{-1})^{-1} \mathbf{AB}^{-1} \\
 &= \mathbf{A}^{-1} - \mathbf{A}^{-1} (\mathbf{I} + \mathbf{BA}^{-1})^{-1} \mathbf{BA}^{-1}. \tag{B.9.2}
 \end{aligned}$$



# Matrix Inversion Theorems

**Theorem 3:** If  $\mathbf{A}$  and  $\mathbf{B}$  are PD matrices of order  $n$  and  $m$ , respectively, and if  $\mathbf{C}$  is of order  $n \times m$ , then

$$(\mathbf{C}^T \mathbf{A}^{-1} \mathbf{C} + \mathbf{B}^{-1})^{-1} \mathbf{C}^T \mathbf{A}^{-1} = \mathbf{B} \mathbf{C}^T (\mathbf{A} + \mathbf{C} \mathbf{B} \mathbf{C}^T)^{-1} \quad (\text{B.9.3})$$

provided the inverse exists.

**Theorem 4:** The Schur Identity or insideout rule. If  $\mathbf{A}$  is a PD  $n \times n$  matrix and if  $\mathbf{B}$  and  $\mathbf{C}$  are any conformable matrices such that  $\mathbf{BC}$  is  $n \times n$ , then

$$(\mathbf{A} + \mathbf{B}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (\text{B.9.4})$$



# Final Thoughts

- ▶ Homework 1 due today
- ▶ Homework 2 due in 7 days
- ▶ Next quiz active Monday at 1pm.

