Intro to Nonlinear Optimization

- We now relax the proportionality and additivity assumptions of LP
- What are the challenges of nonlinear programs (NLP's)?
  - Objectives and constraints can use any function:

$$\max\ f(x)$$
$$\text{subject to}$$
$$G_1(X) \le b_1$$
$$G_2(X) = b_2$$

G(X) means a matrix of constraints

  - Feasible region is not guaranteed to be convex
  - Optima may not occur at extreme points
  - May be many "local" optima; may not be possible to determine the "global" optimum
  - No general-purpose algorithm suits all problems

Example: Nonlinear Warehouse Location

- Warehouse location
  - Suppose we want to locate a set of warehouses
  - Let i = warehouses, j = markets
  - Data:
    - C_i = capacity of warehouse i
    - R_j = demand in market j
    - (a_j, b_j) location of market j in (x,y) coordinates
  - Variables
    - (x_i, y_i) = location of warehouse
    - d_{ij} = distance from warehouse i to market j
    - w_{ij} = units shipped from warehouse i to market j

Warehouse Location Formulation

- One formulation is:

$$\min z = \sum_{i,j} w_{ij} * d_{ij}$$

subject to

$$\sum_j w_{ij} \le C_i \text{ for all } i \text{ (capacity constraints)}$$

$$\sum_i w_{ij} \ge R_j \text{ for all j (demand constraints)}$$

$$d_{ij} = \sqrt{(x_i - a_i)^2 + (y_i - b_i)^2} \text{ for all } i,j \text{ (distance constraints)}$$

$$w_{ij} \ge 0 \text{ for all } i,j$$

- The objective function and the distance constraints are nonlinear

Min Cost Network Congestion Problem

- Here's something that looks like a MCNFP, but with a nonlinear twist:

$$\min z = \sum_{i,j \in ARCS(i,j)} \left(\frac{x_{ij}}{U_{ij} - x_{ij}} \right)$$

subject to

$$\left[\sum_{j \in ARCS(i,j)} x_{ij} \right] - \left[\sum_{j \in ARCS(j,i)} x_{ji} \right] = SD_i \text{ for all nodes } i$$

$$0 \le x_{ij} \le U_{ij} \text{ for all } ARCS(i, j)$$

- What does the nonlinear objective function do?

Attacking a Nonlinear Problem

- So, nonlinear problems can be nasty
- Need to consider:
  - The form of the objective function; how pathological is it?
  - The form of the feasible region; in particular, is it convex? If not, then you'll have to search local optima
- Some modeling advice
  - In general, life is easier if you can restrict the nonlinearity of the problem to the objective function
  - Spreadsheet solvers allow you to define arbitrary nonlinear problems, and they do give solutions - but BE CAREFUL OF THE SOLUTION!
  - There is no "one size fits all" approach to NLPs; various heuristics (simulated annealing, genetic algorithms) sound cool, but they *still heuristics*

Convex Sets and Convex Functions

- Recall that a set ***S*** is convex if:
  - x_1 and x_2 are elements of S, then (1-***c***)x_1 + cx_2 is also in ***S***, for 0 <= ***c*** <= 1
  - Knowing whether the feasible region is convex helps analyze a nonlinear problem
- Another important part of nonlinear optimization are convex *functions*
- A function ***f*** is convex on a convex set ***S*** if, for any x_1 and x_2 in ***S***:

$$f\left(cx_1+\left[1-c\right]x_2\right)\leq cf\left(x_1\right)+\left[1-c\right]f\left(x_2\right)$$

  - Note that x_1 and x_2 can be either scalars or vectors

Convexity, Concavity

- A function is *concave* if the reverse inequality holds:

$$f\left(cx_1 + [1-c]x_2\right) \geq cf\left(x_1\right) + [1-c]f\left(x_2\right)$$

- Winston (p. 631) shows the difference in 2-d; essentially, a function ***f*** is concave if ***-f*** is convex
- Why do we care about this?
- BECAUSE:
  - If the feasible region of a maximization NLP is convex and the objective function is concave, any local optimum is also the global optimum (Theorem 1, p. 632)
  - If the feasible region of a minimization NLP is convex and the objective function is convex, any local optimum is also the global optimum (Theorem 1', p. 632)

Proving a Function is Convex/Concave

- For functions of a single variable, we use calculus:
  - If ***f''(x)*** >= 0 for all ***x*** in a convex set ***S***, ***f*** is convex
  - If ***f''(x)*** <= 0 for all ***x*** in a convex set S, ***f*** is concave
- For multivariate functions ***f(X)***, this is a bit more difficult
- There are some rules:
  - (1) A linear combination of convex (concave) functions is convex (concave):
    $$g(X) = \sum_i c_i f_i(X)$$
  - (2) If ***f*** is a concave function and > 0 on ***S***, then the following function is convex:
    $$g(X) = 1/f(X)$$
  - (3) If ***f*** is a nondecreasing, univariate convex function, and ***h*** is a convex function, then the following is convex:
    $$g(X) = f\left[h(X)\right]$$

Proving Concavity/Convexity; Hessians

- Rules (cont'd)
  - If ***f*** is a convex multivariate function, then the following, where ***A*** is a matrix and ***b*** a vector, is also convex:

$$g(X) = f\left[AX + b\right]$$

  - If ***f*** has continuous second derivatives on ***S***, and its Hessian matrix is *positive semidefinite* for all points in ***S***, then ***f*** is *convex*
  - If ***f*** has continuous second derivatives on ***S***, and its Hessian matrix is *negative semidefinite* for all points in ***S***, then ***f*** is *concave*
- So:
  - What's a Hessian (if it's not a German mercenary)?
  - What's it mean to be positive or negative semidefinite?

Hessian Matrices

- If the function has continuous second derivatives on ***S***, we can analyze a thing called the "Hessian"
- This is the multivariate analog of a second derivative; the Hessian ***H(x)*** of a function $\boldsymbol{f(x_1, x_2, \ldots x_n)}$ is:

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

Positive and Negative Semidefiniteness

- A Hessian matrix is positive semidefinite if, for any ***x**** in a set ***S***:

$$x\,H(x^*)x \le 0 \text{ for all } x \in E^n \text{ (the set of } n\text{-dim real vectors)}$$

- A Hessian matrix is negative semidefinite if, if, for any ***x**** in a set ***S***:

$$x\,H(x^*)x \ge 0 \text{ for all } x \in E^n \text{ (the set of } n\text{-dim real vectors)}$$

- So, semidefiniteness determines whether the function is convex or concave
- NOTE: since ***f*** is concave if ***-f*** is convex, I will only talk about *convexity* from now on

Testing for Positive Semidefiniteness

- **Exclusionary rules:** suppose ***H*** is an n-dimensional Hessian matrix with elements h_{ij}. Then:
  - If any diagonal element is < 0, ***H*** isn't positive semidefinite
  - If a diagonal element h_{ii} = 0, then row ***i*** and column ***i*** must also be 0, or else ***H*** is not positive semidefinite
- **Principal minors:**
  - A "principal minor" of an ***n*** x ***n*** matrix is the ***i*** x **i** matrix you get from deleting ***n-i*** rows and columns of ***H***
  - If the determinants of all principal minors of ***H*** are all >= 0, then ***H*** is positive semidefinite
- Now we have some tests; let's try some examples

Examples of Convexity/Concavity Testing

- First example:

$$f(x_1, x_2) = 2x_1 + 6x_2 - 2x_1^2 - 3x_2^2 + 4x_1 x_2$$

$$H(x) = \begin{bmatrix} -4 & 4 \\ 4 & -6 \end{bmatrix}$$

- This matrix has 3 principal minors; the entire matrix, and the two diagonal elements
- The determinants of the principal minors are -4, -6, and (-4 x -6) - (4 x 4) = 8
- It's not positive semidefinite; however, ***- H*** is! Therefore, it's *negative* semidefinite, and the function's *concave*

Another Example

- Here's where a function is positive semidefinite only in a particular region:

$$f(x_1, x_2) = x_1^3 + 2x_2^2$$

$$H(x) = \begin{bmatrix} 6x_1 & 0 \\ 0 & 4 \end{bmatrix}$$

- The determinants of the principal minors are $6\boldsymbol{x_1}$, 4, and $24\boldsymbol{x_1}$
- This function is positive semidefinite, and convex, only if $\boldsymbol{x_1} >= 0$

Some Miscellanea About Definiteness

- If ***H*** is an ***n*** x ***n*** matrix:
  - The "characteristic equation" of ***H*** is $|H - \lambda I| = 0$
  - The λ's are called the "eigenvalues" of ***H***
  - If they are all >= 0, ***H*** is positive semidefinite; if they are all <= 0, ***H*** is negative semidefinite
- This is another way to test, if you can compute the eigenvalues easily
- A couple of good references:
  - *Linear Algebra and Its Applications* (Gilbert Strang)
  - *Calculus* (K. G. Binmore; this is very good, and unique, book)

Multivariate Unconstrained Optimization

- This is Sec. 12-5 of Winston
- The basic problem is to:

$$\text{max or min} \quad f(x_1, x_2, \ldots x_n)$$
$$\text{subject to} \; (x_1, x_2, \ldots x_n) \in R^n$$

- We're assuming that f has continuous first and second partial derivatives
- For an univariate function, we know candidate critical points occur where ***f'(x) = 0***
- The same argument applies to multivariate functions

Gradients

- The gradient of a function f is a vector of the first partial derivatives:

$$\nabla f(x_1, x_2, \ldots x_n) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right)$$

- We're looking for points where the gradient vector is 0
- Example:

$$f(x_1, x_2) = x_1^{\,3} - 3x_1 x_2^{\,2} + x_2^{\,4}$$

$$\nabla f(x_1, x_2) = \left(3x_1^{\,2} - 3x_2^{\,2}, -6x_1 x_2 + 4x_2^{\,3}\right)$$

$$\text{Set } 3x_1^{\,2} - 3x_2^{\,2} = 0$$

$$\text{Set } -6x_1 x_2 + 4x_2^{\,3} = 0$$

Critical Points

- When we look at these equations, we find that:
  - Either $\boldsymbol{x_1 = x_2}$ or $\boldsymbol{x_1 = -x_2}$ (first equation)
  - Either $\boldsymbol{x_2 = 0}$ or $\boldsymbol{x_2^2 = 1.5x_1}$ (second equation)
  - So the 3 possible points are (0,0), (1.5,1.5), and (1.5,-1.5)
  - Think about the above … do you see the argument?
- Now we have to test these points using the Hessian

$$H(x) = \begin{bmatrix} 6x_1 & -6x_2 \\ -6x_2 & -6x_1 + 12{x_2}^2 \end{bmatrix}$$

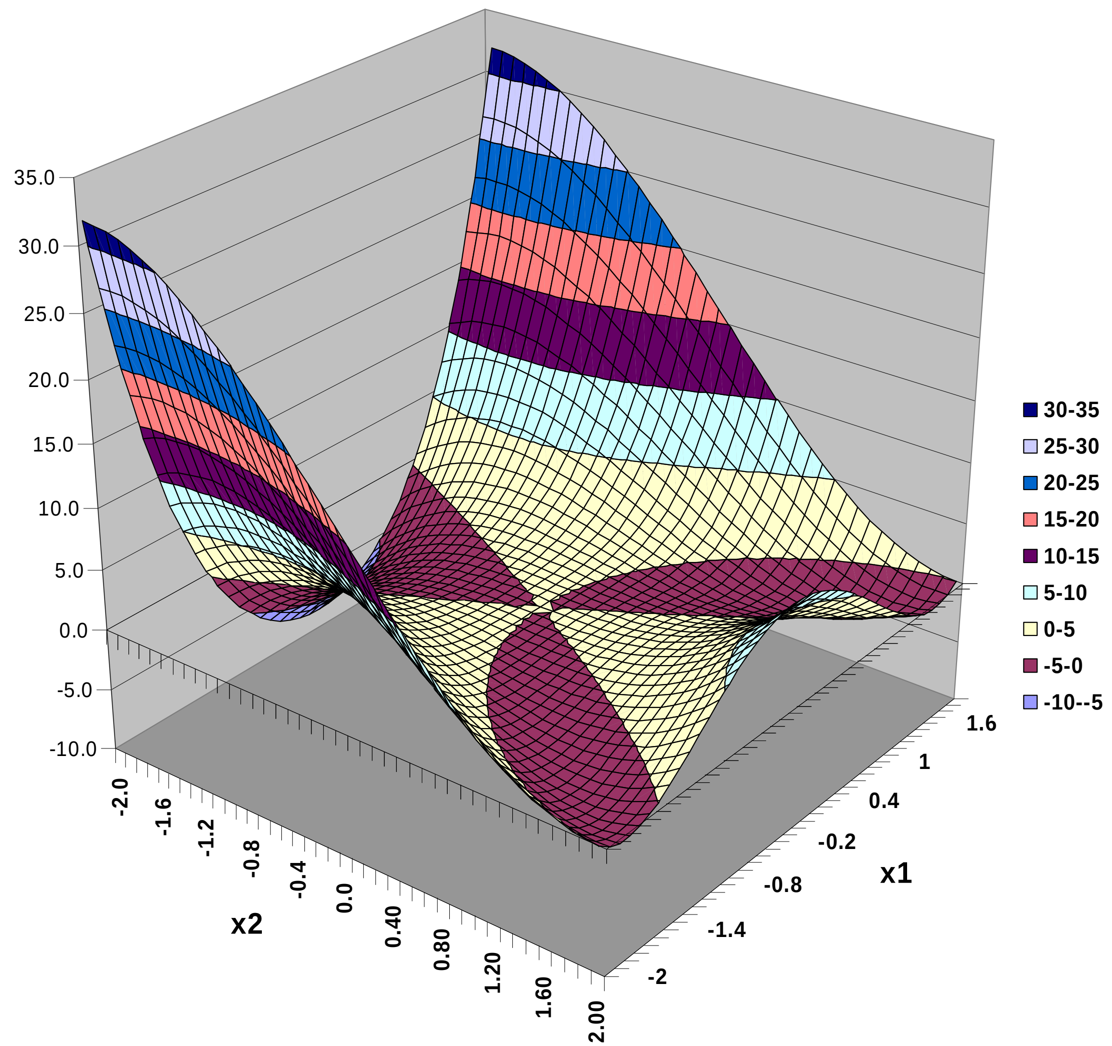
Rules for Testing with a Hessian

- Here's a slightly different summary of Winston
- Suppose x* is a critical point
  - **If the determinant of H(x*) = 0**, the test is inconclusive (useless)
  - **If the determinant of H(x*) > 0**, and all the principal minors are > 0, then x* is a *local minimum*
  - **If the determinant of H(x*) < 0**, the signs of the "even" principal minors are > 0, and the signs of the "odd" principal minors are < 0, then x* is a local maximum
  - **If the determinant of H(x*) <> 0** and the other tests fail, x* is a "saddle point"

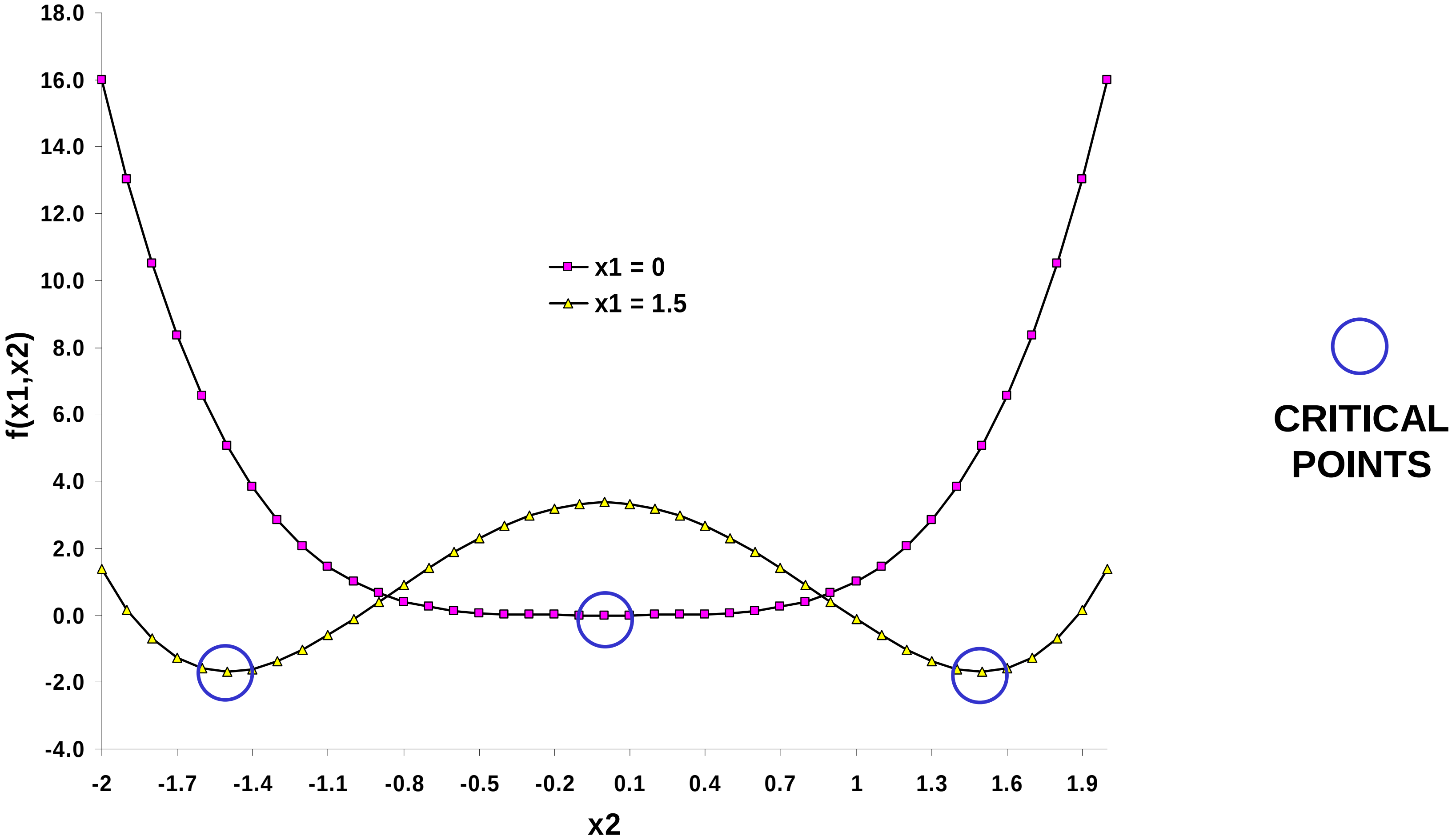
Testing the Points in the Example

- At (0,0):
  - ***H(x)*** = 0
  - The test is useless
- At (1.5, 1.5):
  - The determinant of the entire matrix is 81 > 0
  - The diagonal elements are 9 and 27, both > 0
  - This point is a local *minimum*
- At (1.5, -1.5):
  - As before, the determinate of the entire matrix is 81 > 0
  - The diagonals are 9 and 27, both > 0
  - This point is also a local *minimum*

The Function in 3-D



A 2-D Slice



18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-2.0
-4.0
f(x1,x2)
x1 = 0
x1 = 1.5
-2
-1.7
-1.4
-1.1
-0.8
-0.5
-0.2
0.1
0.4
0.7
1
1.3
1.6
1.9
x2
CRITICAL POINTS

What Happens If We Try This in MPL?

- Here's the MPL code for this problem
  - Note there's no constraints
  - The "OPTIONS" statement tells MPL it's nonlinear

```
INDEX
   i := 1..2;

OPTIONS
   ModelType=nonlinear;

FREE VARIABLES
   x[i];

MODEL
  min z =  x[1]^3 - 3*x[1]*(x[2]^2) + x[2]^4;

END
```

- CONOPT reports the point (0,0) is the min

Now, Add Some Constraints

- Let's look around for more critical points; add

```
SUBJECT TO
   x[1] > 0.1;
   x[2] > 0.1;
```

- Now CONOPT says the min is (1.5, 1.5)
- Try for the third critical point:

```
SUBJECT TO
   x[1] > 0.1;
   x[2] < -0.1;
```

- CONOPT says the min is (1.5, -1.5)

Last Experiment

- What happens for:

```
SUBJECT TO
   x[1] < -0.1;
   x[2] < 0;
```

- CONOPT reports problem is *unbounded*
  - Why didn't it tell us this for the unconstrained case?
  - Looking at the function, setting $x_2 = 0$ allows x_1^3 to go to positive or negative infinity
  - Is the solver screwed up?

Moral(s)

- This is the overarching lesson with nonlinear optimization
  - If the objective or constraints are nonconvex, you will get local optima
  - You should figure this out *before* you start
  - You have to have some way of finding multiple local optima; putting in bounds as in the example is a cheap, fast way
- Commercial nonlinear solvers generally work as follows:
  - They find an initial feasible point
  - They solve a local linear approximation of the problem to find an improving direction and a “step size”
  - They step along the improving direction, maintaining feasibility
  - They then repeat the procedure until they find a local optimum
  - The responsibility to check the solution is ***YOURS***

Nonlinear Problems w/ Equality Constraints

- We now begin to introduce constraints to nonlinear problems
- The general form is:

$$
\begin{aligned}
&\text{max or min} \quad f(x_1, x_2, \ldots x_n) \\
&\text{subject to} \\
&g_1(x_1, x_2, \ldots x_n) = b_1 \\
&g_2(x_1, x_2, \ldots x_n) = b_2 \\
&\vdots \\
&g_{\mathrm{m}}(x_1, x_2, \ldots x_n) = b_m
\end{aligned}
$$

From Calculus: Lagrange Multipliers

- Consider the following problem:
  - $$\max \;\; z = 4xy$$
    $$\text{subject to}$$
    $$\frac{x^2}{9} + \frac{y^2}{16} = 1$$
  - Translation: find the largest rectangle that can be inscribed in an ellipse with major and minor axes of 4 and 3, respectively
- Way back in calculus, we formed the following function:

$$L(x, y, \lambda) = 4xy - \lambda\left(\frac{x^2}{9} + \frac{y^2}{16} - 1\right)$$

- The new function is the *Langrangean*, and the new variable is a *Lagrange multiplier*

Some Arguments

- We now maximize the (unconstrained) Lagrangean function:

$$\max z = L(x, y, \lambda) = 4xy - \lambda\left(\frac{x^2}{9} + \frac{y^2}{16} - 1\right)$$

- What is this, and why does it work?
- Some functional arguments:
  - The term we have added to the objective is essentially a penalty term
  - Any solution that does not have points on the ellipse penalizes the objective (depending on what the Lagrange multiplier value is)
  - Does this look similar to complementary slackness?

Lagrange's Theorem

- Here's what makes this go:
  - Let functions ***f*** and g have continuous first partial derivatives
  - Also, let f have an extremum at the point ($x_1^*, x_2^*, \ldots, x_n^*$) on the constraint function $g(x_1, x_2, \ldots, x_n) = c$
  - If $g(x_1^*, x_2^*, \ldots, x_n^*)$ <> 0, then there is a real number, λ, such that:

$$\nabla f\left(x_1^*, x_2^*, \cdots, x_n^*\right) = \lambda \nabla g\left(x_1^*, x_2^*, \cdots, x_n^*\right)$$

- This theorem says that the objective function and constraint gradients are parallel at the optimal point
- Consequently, the constraint is tangent to the objective at the point

The Method of Langrange Multipliers

- Convert the problem to an unconstrained one
  - Form the Lagrangean function
  - Each equality constraint requires a separate Lagrange multiplier
- Find the critical points of the Lagrangean
  - Take the partial derivative with respect to each variable
  - Set the resulting equations to 0; solve for critical points
- Test each critical point to determine the optimum

Back to the Example

- The Lagrangean function was:

$$L(x, y, \lambda) = 4xy - \lambda\left(\frac{x^2}{9} + \frac{y^2}{16} - 1\right)$$

- The partials (set equal to 0) are:

$$\frac{\partial L}{\partial x} = 4y - 2\lambda\frac{x}{9} = 18y - x\lambda = 0$$
$$\frac{\partial L}{\partial y} = 4x - 2\lambda\frac{y}{16} = 32x - y\lambda = 0$$
$$\frac{\partial L}{\partial \lambda} = -\frac{x^2}{9} - \frac{y^2}{16} + 1 = 16x^2 + 9y^2 - 144 = 0$$

The Finale

- I won't show the algebra here, but you would:
  - Solve for λ in the first equation
  - Substitute that into the second equation, so you are left with an equation in ***x*** and ***y***
  - Substitute *that* into the third equation, eliminate ***x***, solve for ***y***
  - Solve for ***x***; don't bother to compute λ
- This only has one critical point
  - (***x*,y****) = $\left(\frac{3}{\sqrt{2}}, 2\sqrt{2}\right)$
  - When we evaluate this in the original ***f***, we get the area = 24

So What Happens if Algebra Doesn't Work?

- It may not be possible to solve the equations algebraically
  - There are various numerical techniques available
  - Covering them is beyond the scope of this course
  - If you only have one Lagrange multiplier, you can just do some sort of line search (like bisection)
- What does MPL do with this?
  - **First try:** CONOPT says "locally infeasible"
  - **Second try:** change the constraint to <=; CONOPT says optimum is (0,0)
  - **Third try:** add the constraints x > 1 and y > 1, CONOPT finds the optimum
- MORAL: use constraints to help find a starting point!