Nonlinear Least Squares with Its Application to GPS Technology

Moody T. Chu

North Carolina State University

MA325 @ North Carolina State University

Take Home Message

- Many sophisticated modern applications are based on simple mathematical theory.
- Many complicated mathematical concepts are based on elementary geometry and calculus.
- The global position system (GPS) is one such an example.

Course Plan of This Module

1. Review of Calculus (2 hour)
2. Linear Least Squares Problems (2 hours)
   - Homework
3. Nonlinear Least Squares Problems (1 hours)
4. 2-D Set up (2 hours)
   - Homework
5. 3-D Set up (2 hours)
   - Project

Outline

Review of Calculus
- Optimization
- Constrained Optimization

Linear Least Squares
- Data Fitting
- Mathematics Behind
- Numerical Techniques

Nonlinear Least Squares
- Mathematical Setup
- Gauss-Newton Method

2-D GPS Setup
- Observation Along a Straight Line
- Measurement by Distance

3-D GPS Mechanism
- Satellite setup
- Navigation Equation

Outline

Review of Calculus
- Optimization
- Constrained Optimization

Linear Least Squares
- Data Fitting
- Mathematics Behind
- Numerical Techniques

Nonlinear Least Squares
- Mathematical Setup
- Gauss-Newton Method

2-D GPS Setup
- Observation Along a Straight Line
- Measurement by Distance

3-D GPS Mechanism
- Satellite setup
- Navigation Equation

Outline

Review of Calculus
- Optimization
- Constrained Optimization

Linear Least Squares
- Data Fitting
- Mathematics Behind
- Numerical Techniques

Nonlinear Least Squares
- Mathematical Setup
- Gauss-Newton Method

2-D GPS Setup
- Observation Along a Straight Line
- Measurement by Distance

3-D GPS Mechanism
- Satellite setup
- Navigation Equation

Outline

Review of Calculus
- Optimization
- Constrained Optimization

Linear Least Squares
- Data Fitting
- Mathematics Behind
- Numerical Techniques

Nonlinear Least Squares
- Mathematical Setup
- Gauss-Newton Method

2-D GPS Setup
- Observation Along a Straight Line
- Measurement by Distance

3-D GPS Mechanism
- Satellite setup
- Navigation Equation

Outline

Review of Calculus
Optimization
Constrained Optimization

Linear Least Squares
Data Fitting
Mathematics Behind
Numerical Techniques

Nonlinear Least Squares
Mathematical Setup
Gauss-Newton Method

2-D GPS Setup
Observation Along a Straight Line
Measurement by Distance

3-D GPS Mechanism
Satellite setup
Navigation Equation

Gradient

- Given a scalar function

$$f : \mathbb{R}^n \longrightarrow \mathbb{R},$$

  define the gradient of η by

$$\nabla f := \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right].$$

- Significance:
  - Points in the direction where the function $f(\mathbf{x})$ ascends most rapidly.
  - Attainable maximum rate of change is precisely $\|\nabla f(\mathbf{x})\|$.

First Order Optimality Condition

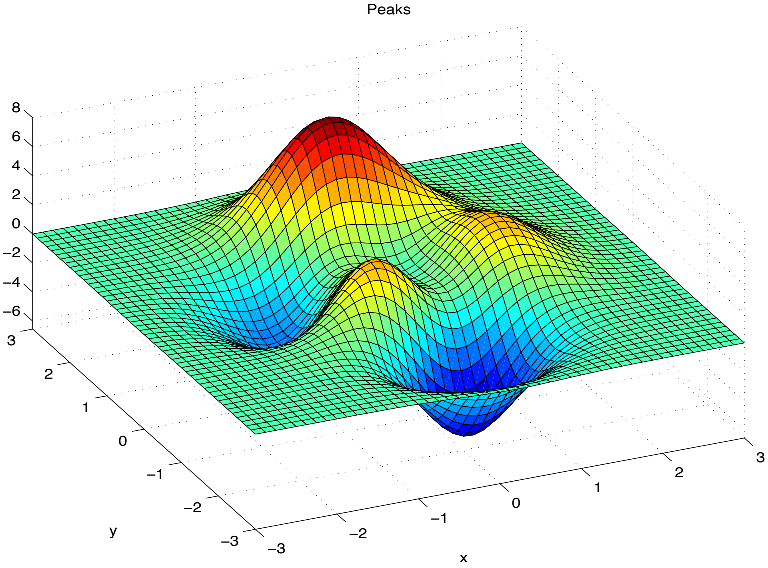
Suppose $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a smooth function over an open domain. Then

- The functional value $f(\mathbf{x})$ reaches an extreme value, either maximum or minimum, at a point $\mathbf{x} \in \mathbb{R}^n$ only if

$$\nabla f(\mathbf{x}) = 0.$$

- The extreme is only a relative (local) extreme.

$$f(x,y) = 3(1-x)^2 e^{-x^2-(y+1)^2} - 10\left(\frac{x}{5} - x^3 - y^5\right)e^{(-x^2-y^2)} - \frac{1}{3}e^{-(x+1)^2-y^2}$$



Second Order Optimality Condition

- Need a way to tell the concavity.
- In Calculus III, for the case $n = 2$, we have learned the basic rules:
  - Compute the second derivative, the so called Hessian matrix,

$$H_f(x, y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}.$$

  - If $f_{xx} f_{yy} - f_{xy} f_{yx} < 0$, then it is a saddle.
  - If $f_{xx} f_{yy} - f_{xy} f_{yx} > 0$ and $f_{xx} > 0$, then it is a minimum (cup).
  - If $f_{xx} f_{yy} - f_{xy} f_{yx} > 0$ and $f_{xx} < 0$, then it is a maximum (cap).
  - If any of these is zero, then go to graduate school.
- What is going on here?
- How to generalize this concept to more than two variables?

Symmetric and Positive Definite Matrix

- A matrix $A \in \mathbb{R}^{n \times n}$ is said to be *positive definite* if and only if

$$\mathbf{x}^\top A\mathbf{x} > 0 \quad \text{for all } \mathbf{x} \neq 0.$$

  - A is said to be *positive semi-definite* if “$>$" is replaced by “$\geq$".
- There are multiple equivalent conditions for determining whether a matrix A is symmetric and positive definite.
  - All eigenvalues of A are positive.
  - All *principal minors* have positive determinant.

Negative Definite Matrix

- How to define a negative definite matrix?
- What are some conditions for determining whether a matrix A is symmetric and negative definite?
- Where does the symmetry of a Hessian matrix come from?

The Real Second Order Optimality Condition

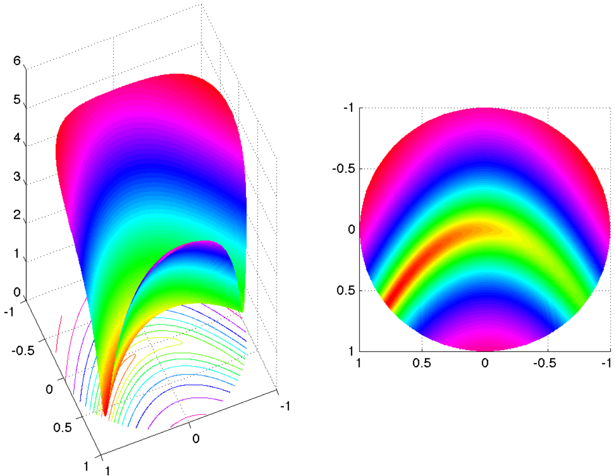
- If $\mathbf{x}$ is a critical point and is a local minimum for a smooth function f, then its Hessian $H_f(\mathbf{x})$ is necessarily positive semi-definite.
- If $\mathbf{x}$ is a critical point and if its Hessian $H_f(\mathbf{x})$ is positive definite, then $\mathbf{x}$ is a local minimum.
  - What is the difference?
- What can be said about a local maximum?

Constrained Optimization

- In real world, we cannot do whatever we want to do.
- Even we are interested in maximizing the gain or minimizing the loss, often we are subject to some constraints.
- The challenge is how to handle this type of constrained optimization?

Roserbrock Function

$$\begin{array}{ll} \min & 100(y-x^2)^2+(1-x)^2, \\ \text{subject to} & x^2+y^2 \le 1. \end{array}$$



Method of Lagrange Multiplier

Suppose that the optimization problem is

$$\begin{array}{cl} \text{min} & f(x, y), \\ \text{subject to} & g(x, y) = c. \end{array}$$

- Introduce a new variable λ, called a Lagrange multiplier.
- Define the Lagrange function, called Lagrangian, defined by

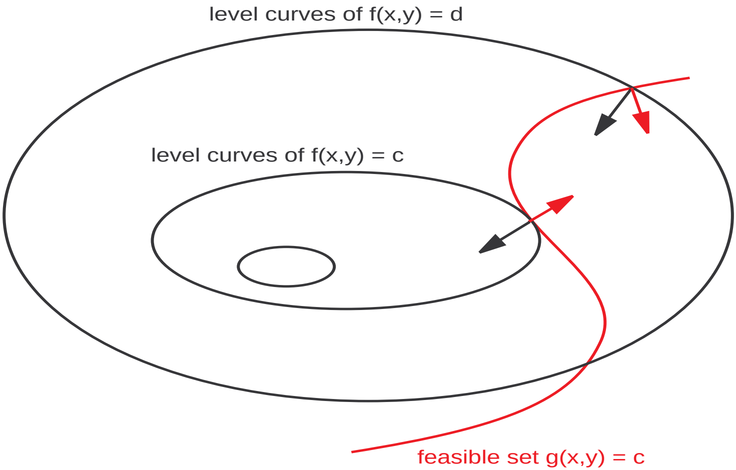
$$\Lambda(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - c) \tag{1}$$

- The critical point must satisfy

$$\nabla\Lambda = 0.$$

- Why?

Geometric Meaning



An Example

Find the dimensions of the box with largest volume if the total surface area is A.

- Setup:

$$\begin{array}{rl} \max & xyz, \\ \text{subject to} & 2xy + 2xz + 2yz = A. \end{array}$$

- Lagrangian:

$$\Lambda(x, y, z) = xyz - \lambda(xy + xz + yz - \frac{A}{2}).$$

- Necessary condition:

$$\begin{cases} yz & = & \lambda(y+z), \\ xz & = & \lambda(x+z), \\ xy & = & \lambda(x+y), \\ xy + xz + yz & = & \frac{A}{2}. \end{cases}$$

- Need to solve the above system of equations for (x, y, z, λ).
  - Hint: Multiply the first equation by x and the second equation by y. Make an argument from here.

Parameter Estimation

Parameter estimation is an important technique used for modeling in many areas of disciplines.

- To mimic a complicated physical phenomenon, we sometimes can create a model via a relationship such as

$$y = f(z; x_1, \ldots, x_n). \tag{2}$$

  - f is a prescribed model determined up to values of $x_1, \ldots, x_n$.
  - $x_1, \ldots, x_n$ are the parameters.
  - z is the control variable or input.
  - y is the expected response or output to z.
- For more sophisticated models, both input z and output y can be vectors.

Using Observations

- Perform m experiments and collected m observed quantities $(z_i, y_i), i = 1, \ldots, m$.
  - Typically $(m \geq n)$.
  - Why?
- Due to measurement errors (called noise), (z_i, y_i) may not satisfy (2) exactly.
- Seek to adjust the parameters $x_1, \ldots, x_n$ so that the expression

$$g(x_1, \ldots, x_n) := \sum_{i=1}^{m} \|y_i - f(z_i; x_1, \ldots, x_n)\|^2 \tag{3}$$

  is minimized.
- When the norm used in (3) is either the $2-$norm or the Frobenius norm, we say we have a *least squares problem* .

Polynomial Fitting

- Suppose an $(n-1)$-th degree polynomial

$$f(z; x_1, \ldots, x_n) = x_1 z^{n-1} + \ldots + x_{n-1} z + x_n. \quad (4)$$

is to fit m points in the plane.

- Ideally, want to solve the system

$$\begin{bmatrix} z_1^{n-1} & z_1^{n-2} & \ldots & z_1 & 1 \\ z_2^{n-1} & & & & \\ \vdots & & & & \\ z_m^{n-1} & z_m^{n-1} & \ldots & z_m & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad (5)$$

for the coefficients $(x_1, \ldots, x_n)$.

  - The system (5) is overdetermined, so generally there is no solution.

General Linear Least Squares Problem

- Want to solve the optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_2^2 \quad (6)$$

where $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$ are known quantities.

- “Linear" in the sense that the expected response y depends linearly on the parameters $\mathbf{x}$.

Normal Equation

- Can write the objective function as

$$g(\mathbf{x}) = \frac{1}{2}(A\mathbf{x} - \mathbf{b})^\top (A\mathbf{x} - \mathbf{b}).$$

- The first order condition becomes

$$\nabla g(\mathbf{x}) = A^\top A\mathbf{x} - A^\top \mathbf{b} = 0.$$

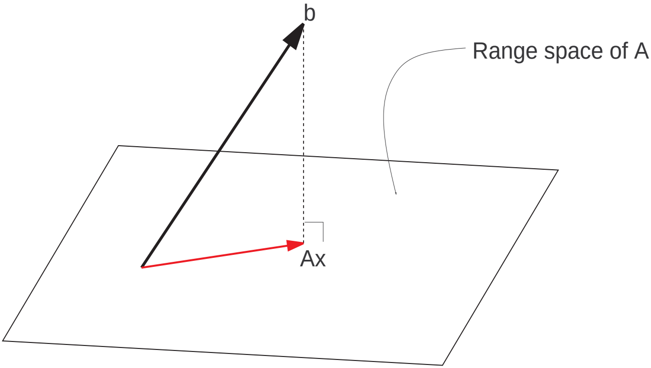
  - Prefer $A\mathbf{x} = \mathbf{b}$; now $A^\top A\mathbf{x} = A^\top \mathbf{b}$.

Geometry behind Linear Least Squares

- Let the columns of $A \in \mathbb{R}^{m \times n}$ be denoted as $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ where each $\mathbf{a}_i \in \mathbb{R}^m$.
- The product $A\mathbf{x}$ can be written as

$$A\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{a}_i,$$

  - $A\mathbf{x}$ is a linear combination of columns of A and hence is an element in the range space of A.
- Solving the equation $A\mathbf{x} = \mathbf{b}$ is equivalent to finding an appropriate combination of columns of A that makes up the vector $\mathbf{b}$.
  - A necessary condition for $A\mathbf{x} = \mathbf{b}$ to have a solution is that $\mathbf{b} \in R(A)$.
  - What to do when $\mathbf{b} \notin R(A)$?



- The best we can hope for is to find a combination so that the residual $\mathbf{b} - A\mathbf{x}$ is minimized.
- The residual $\mathbf{b} - A\mathbf{x}$ must be perpendicular to $R(A)$.
  - How to quantify this geometry?

Mathematical Setup

- From the assumed model $y = f(z; x_1, \ldots, x_n)$, define a residual

$$r_i = r_i(x_1, \ldots, x_n) := y_i - f(z_i; x_1, \ldots, x_n)$$

for each observed data (z_i, y_i), $i = 1, \ldots, m$.

- Intend to minimize the overall residual

$$g(x_1, \ldots, x_n) := \sum_{i=1}^{m} \|r_i\|_2^2.$$

- Rewrite the notion as an unconstrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} F(\mathbf{x})$$

where

$$F(\mathbf{x}) := \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|_2^2 \tag{7}$$

and

$$\mathbf{r}(\mathbf{x}) := [r_1(\mathbf{x}), \ldots, r_m(\mathbf{x})]^\top.$$

First Optimality Condition

- The necessary condition for $\mathbf{x}$ to be a critical point is that $\nabla F(\mathbf{x}) = 0$.
- We calculate the gradient of F to be

$$\nabla F(\mathbf{x}) = J(\mathbf{x})^T t(\mathbf{x}) \tag{8}$$

where

$$J(\mathbf{x}) := \frac{\partial \mathbf{r}}{\partial \mathbf{x}} := \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \vdots & & & \\ \frac{\partial r_m}{\partial x_1} & \frac{\partial r_m}{\partial x_2} & \cdots & \frac{\partial r_m}{\partial x_n} \end{bmatrix}$$

is the $m \times n$ Jacobian matrix of f.

Finding Critical Points

- Note that $\nabla F : \mathbb{R}^n \to \mathbb{R}^n$ in nonlinear in general.
- Need an algorithm to solve the equation $\nabla F(\mathbf{x}) = 0$.
  - The Newton-Ralphson method is generally too expensive.
  - Special techniques are available for this type of problems.
  - See **lsqnonlin** in MATLAB.

A MATLAB Demonstration

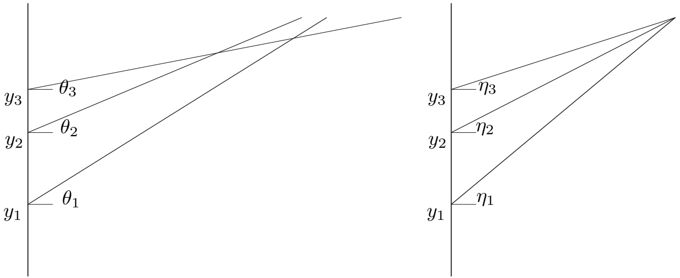
- Want to minimize the function

$$\sum_{k=1}^{10} \underbrace{\left(2 + 2k - e^{kx_1} - e^{kx_2}\right)}_{r_k(\mathbf{x})}^2 .$$

```
%%
function example_lsqnonlin

x0 = [0.3 0.4]                          % Starting guess
[x,resnorm] = lsqnonlin(@myfun,x0);    % Invoke optimizer

    function F = myfun(x)
        k = 1:10;
        F = 2 + 2*k-exp(k*x(1))-exp(k*x(2));
    end

end
```

Basic Ideas of GPS

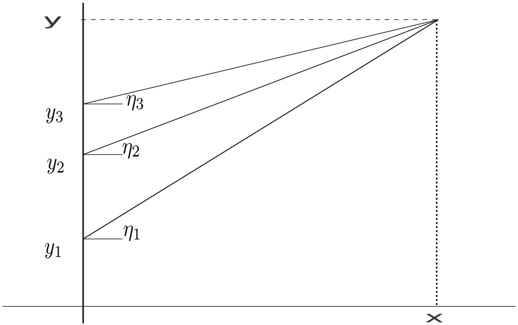
- Consider the scenario:
  - Three observers, located along a straight line, measure the angle of their line-of-sight to a certain object.
  - Assume that the angles are measured counterclockwise from the east (normal to the baseline of observers).
  - Due to various reasons, such as atmosphere turbulence, their observations are obscured.



Information Retrieval

- The three measured angles θ_i, $i = 1, 2, 3$, are more or less correct but carry some small uncertainties.
- It is desired to estimate the true position of the object.
- How to correct the problem?

Necessary Conditions on the True Solution



$$\frac{y - y_i}{x} = \tan \eta_i, \quad i = 1, 2, 3.$$

Build up Something Workable

- Get rid of y:

$$y = y_1 + x\tan\eta_1 = y_2 + x\tan\eta_2 = y_3 + x\tan\eta_3.$$

- Get rid of x:

$$\frac{y_2 - y_1}{\tan\eta_1 - \tan\eta_2} = \frac{y_3 - y_2}{\tan\eta_2 - \tan\eta_3}.$$

- The constraint:

$$(y_2 - y_1)(\tan\eta_2 - \tan\eta_3) = (y_3 - y_2)(\tan\eta_1 - \tan\eta_2).$$

  - Why is this significant?

Constrained Least Squares

$$\begin{array}{ll} \text{min} & f(\eta_1, \eta_2, \eta_3) := \sum_{i=1}^{3} (\theta_i - \eta_i)^2, \\ \text{subject to} & (y_2 - y_1)(\tan\eta_2 - \tan\eta_3) = (y_3 - y_2)(\tan\eta_1 - \tan\eta_2). \end{array}$$

- How to handle this type of optimization problem?
- **Homework:** Using Lagrange multiplier theory, show that the optimal angles are given by

$$\begin{aligned} \eta_1 &= \theta_1 + \omega(y_2 - y_3)\sec^2\eta_1, \\ \eta_2 &= \theta_2 + \omega(y_3 - y_1)\sec^2\eta_2, \\ \eta_3 &= \theta_3 + \omega(y_1 - y_2)\sec^2\eta_3, \end{aligned}$$

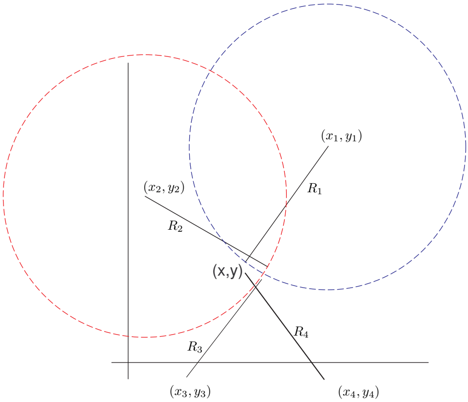
where ω is a constant that ensures the lines of sight define a single point of intersection.

Improvement of Techonology

- Must the measurement be done by angles?
  - What are pros and cons in doing measurement by angles?
- Must the observers be lined up?
  - What will happen if more observers ($m > 3$) are providing information?
- The newer technology allows us to measure long distances.
  - (L)ight (a)mplification by (s)timulated (e)mission of (r)adiation.
  - Electronic signals.
  - Satellite.
  - GPS.

Observation Around a Point

- Consider the scenario:
  - Four observers, located at known positions in the plane, measure the distance of their line-of-sight to a certain object.
  - For various reasons, the measurements do not add up.
  - Where is the correct position of (x, y)?



Setup

- Define the residuals:

$$r_i(x,y) \quad = \quad \sqrt{(x-x_i)^2+(y-y_i)^2} - R_i, \qquad i = 1,2,3,4.$$

- Want to minimize the overall residual

$$F(\mathbf{x}) := \frac{1}{2}\|\mathbf{r}(\mathbf{x})\|_2^2.$$

- Need to solve the first order optimality condition:

$$\begin{bmatrix} \frac{x-x_1}{S_1} & \frac{x-x_2}{S_2} & \frac{x-x_3}{S_3} & \frac{x-x_4}{S_4} \\ \frac{y-y_1}{S_1} & \frac{y-y_2}{S_2} & \frac{y-y_3}{S_3} & \frac{y-y_4}{S_4} \end{bmatrix} \begin{bmatrix} r_1(x,y) \\ r_2(x,y) \\ r_3(x,y) \\ r_4(x,y) \end{bmatrix} = 0.$$

  - $S_i := \sqrt{(x-x_i)^2+(y-y_i)^2}$.

A Blessed Curse

- Suppose
  - The observers are the satellites.
  - A signal traveling at speed c is sent between the satellite and the receiver.
  - The distance R_i is calculated by measuring the transmission time t_i. Ideally,
$$R_i = ct_i.$$
- The clock in the typical low-cost receiver, i.e., the GPS, has relatively poor precision. It carries an *unknown* latency d. So, in reality,
$$R_i = c(t_i - d)$$
  - d is part of the calculation.
- How precise the time measurement of the atomic clocks must be to keep the precision of distance to within 3 meter? 10 nanoseconds

Reformulation

- Define the residuals:

$$r_i(x, y, d) = \sqrt{(x - x_i)^2 + (y - y_i)^2} - c(t_i - d), \quad i = 1, 2, 3, 4.$$

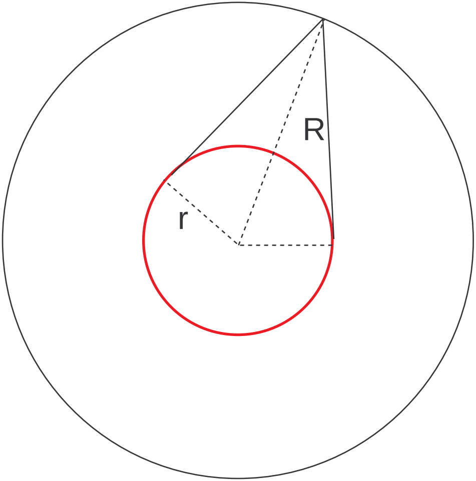
- Need to solve the first order optimality condition:

$$\begin{bmatrix} \frac{x-x_1}{S_1} & \frac{x-x_2}{S_2} & \frac{x-x_3}{S_3} & \frac{x-x_4}{S_4} \\ \frac{y-y_1}{S_1} & \frac{y-y_2}{S_2} & \frac{y-y_3}{S_3} & \frac{y-y_4}{S_4} \\ c & c & c & c \end{bmatrix} \begin{bmatrix} r_1(x, y, d) \\ r_2(x, y, d) \\ r_3(x, y, d) \\ r_4(x, y, d) \end{bmatrix} = 0.$$

  - How many solutions are there in the system?.
    - Why is this question important?

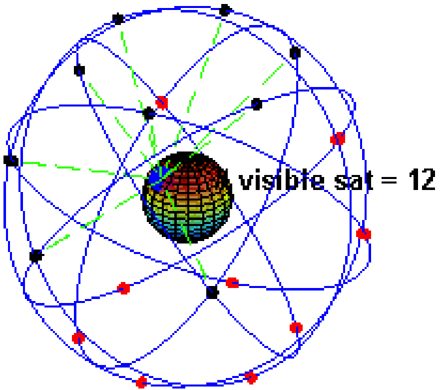
How Many Satellites?

- Suppose that
  - The object is moving around a circle (the earth) center at the origin with radius r.
  - The satellites are moving around the earth at a height of R from the center.
- Assume that
  - Satellites are programmed to automatically avoid collision.
  - The object can occur at arbitrary point on the circle.
- To fully cover any point on the earth at any given time by four satellites, how many satellites in total are needed in the orbit?


R
r

General GPS Description

- Currently, there are 24 satellites carrying atomic clocks.
- Orbit at an altitude of $20,200$ km.
- Four satellites in each of six planes, slanted at 55^o with respect to the poles, make two revolutions per day.
- At any time, from any point on earth, five to eight satellites are in the direct line of sight.
- Transmit synchronized signals from predetermined positions in space.
- The receivers (GPS) on earth will pick up the signals.
- Do the mathematics to determine the accurate (x, y, z) coordinates of the receiver.



– from Wikipedia

Navigation Equation

- Define the residuals:

$$r_i(x, y, z, d) = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} - c(t_i - d).$$

- Four equations in four unknowns.
  - Subtracting the first equation from the last three leads to three linear equations in (x, y, z).
  - By Gaussian elimination, a single solution (x, y, z) is found.
  - Obtain a quadratic equation in d upon substitution.
  - At most two real solutions can be found.

Other Concerns

- There are other technical issues when GPS is deployed.
  - Conditioning of the navigation equation.
  - Transmission speed might be less than the speed of light
    - Need to pass through 100 km ionosphere and 10 km troposphere while subjecting to electromagnetic fields.
    - Might encounter obstacles or atmospheric degradation.
- Can overcome the issues by adding more satellites.
  - No longer a square problem.
  - Need fast nonlinear least squares techniques.
- Who is maintaining the GBS?
  - civilian GPS (CPS) versus military GPS (PPS)
    - two frequencies + ionosphere correction.
  - DoD, $1.3B, US taxpayers' money.
  - Galileo, EU, €5.0B, 30 satellites by 2019.
  - GLONESS, Russian, 24 satellites.
  - BeiDou, China, 35 satellites by 2020.
  - QZSS, Japan, 7 satellites by 2023, high precision (6 cm)
  - NAVIC, India, 7 by 2018.