Math 5600 5/14/14

- The "big O" notation

$$f(h) = O(h^p) \text{ as } h \to 0 \text{ if}$$

$$\lim_{h \to 0} \frac{f(h)}{h^p} = C \neq 0$$

- $f(h)$ is usually an error that depends on h.

- Examples: $f(h) = h^3 + h^4 = O(h^3)$

ignore higher order terms

$$\lim_{h \to 0} \frac{h^3 + h^4}{h^3} = \lim_{h \to 0} 1 + h = 1$$

- h^3 could be multiplied by any constant, it would still be $O(h^3)$

- There are several variants of this definition.

- Instead of h^p we could have some other function of h.

- there is a "little o" notation

$$f(h) = o(h^p) \text{ if } \lim \frac{f(h)}{h^p} = 0$$

- There is also a notion of big O as the variable goes to infinity.

- In that case the variable is often an integer and denoted by n

- $f(n) = O(n^k)$ as $n \to \infty$

$$\text{if } \lim_{n\to\infty} \frac{f(n)}{n^k} = C \neq 0$$

- In that case we ignore lower order terms

- Example

$$\frac{n^3}{3} + n^2 + n + 5 = O(n^3) \text{ as } n \to \infty$$

- The effort to solve linear systems by standard techniques is $O(n^3)$

- Taylor Series $f(x+h) = \sum_{j=0}^{n} \frac{f^{(j)}(x)}{j!} h^j + O(h^{n+1})$

- Let's review some parts of multivariate Calculus.

$$f: \mathbb{R}^n \to \mathbb{R}^m$$

- $\mathbb{R}^n$ space of vectors with n components

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix}$$

- The <u>Jacobian</u> of f is the matrix

$$\nabla f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = J$$

- This is a linear function from

$$\mathbb{R}^n \to \mathbb{R}^m$$

$$L(x) = J x$$

x in $\mathbb{R}^n$

$L(x)$ in $\mathbb{R}^m$

- if $m = 1$ we have a scalar valued function. The Jacobian is then the gradient, the vector of partial derivatives.
- It should be a row, a $1 \times n$ matrix
- Confusingly the gradient is usually written as a column.
- The matrix of second order partial derivatives of a scalar valued function ($m=1$)

$$\left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right] \quad i,j = 1, \cdots, n$$

is the <u>Hessian</u> of f.

- It's a square symmetric matrix (since mixed partials commute.)
- All of this is relevant for minimization
- Taylor Series

$$f(x+h) = f(x) + h f'(x) + \frac{h^2}{2} f''(x) + O(h^3)$$

- when is $f(x)$ a local minimum?
- we must have $f'(x) = 0$
- if in addition we also have $f''(x) > 0$ then locally we can only increase f, and so we have a minimum.
- The same idea works in several variables.

$f: \mathbb{R}^n \to \mathbb{R}$ $\qquad$ x, h in $\mathbb{R}^n$

Taylor series

$$f(x+h) = f(x) + h^T \nabla f(x) + \frac{1}{2} h^T \nabla^2 f(x) h + O(\|h\|)^2$$

$$\|x\| = \sqrt{\sum x_i^2}$$

- clearly, to have a minimum we must have $\nabla f(x) = 0$
- Let $J = \nabla^2 f(x)$
- if in addition

$$h^T J h > 0 \quad \text{for all } h \neq 0$$

then we can be sure to have a minimum.

- A symmetric matrix A is positive definite if

$$x^T A x > 0 \quad \text{for all } x \neq 0$$

- Positive Definiteness of Matrices is the natural generalization of positiveness of numbers.

- $\nabla f(x) = 0$ is a nonlinear system

- Newton's Method from Calculus

- start with a single equation

$$f(x) = 0 \qquad f: \mathbb{R} \to \mathbb{R}$$

- we want a sequence

$$x_0, \; x_1, \; x_2, \dots$$

that converges to a solution

- suppose we are given x_k. What's x_{k+1}

$$f(x_{k+1}) = f(x_k) + (x_{k+1} - x_k) f'(x_k) + O((x_{k+1} - x_k)^2)$$

$$f(x_k) + (x_{k+1} - x_k) f'(x_k) = 0$$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$$x_0 \text{ given}$$

Newton's Method

- Example

$$f(x) = x^2 - 2 = 0$$

$$f'(x) = 2x$$

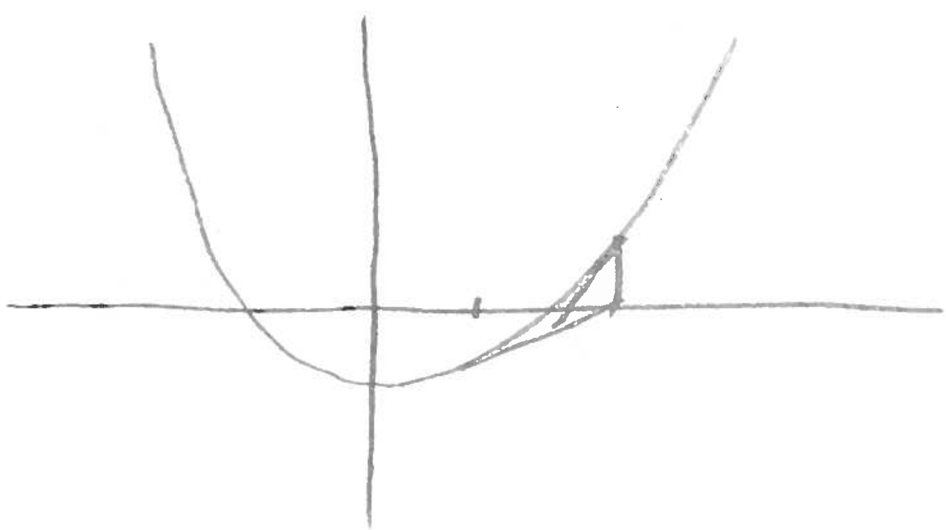
$$x_0 = 1 \qquad x_{k+1} = x_k - \frac{x_k^2 - 2}{2x_k}$$

$x_0 = 1$

$x_1 = 1.5$

$x_2 = 1.417$

$x_3 = 1.4142\ldots$ converges very quickly.



- How does this work in several variables?

$$F(x) = 0 \qquad F: \mathbb{R}^n \to \mathbb{R}^n$$

- subscripts denote components, superscripts exponents
- use superscripts in parentheses

$$x^{(0)}, x^{(1)}, x^{(2)}, \ldots$$

$$F(x^{(k+1)}) \approx F(x^{(k)}) + J(x^{(k)}) \left(x^{(k+1)} - x^{(k)}\right) = 0$$

$$x^{(k+1)} = x^{(k)} - \left(J(x^{(k)})\right)^{-1} F(x^{(k)})$$

- we don't usually compute the inverse

$x^{(0)}$ given

For $k = 0, 1, 2, \ldots$ until satisfied

$$\left\{ \begin{array}{ll} \text{solve} & J(x^{(k)})\, s^{(k)} = -F(x^{(k)}) \\ & x^{(k+1)} = x^{(k)} + s^{(k)} \end{array} \right.$$

- stop when $\|x^{(k+1)} - x^{(k)}\| \leq \varepsilon$ for ε sufficiently small

- Things get more complicated when we have an overdetermined system.

- $$F : \mathbb{R}^m \to \mathbb{R}^n \qquad n > m$$

can't solve $F(x) = 0$ since we have more equations than unknowns.

- solve
$$\|F(x)\|^2 = \min$$
instead
$$\nabla \|F(x)\|^2 = 0$$

- nonlinear system, apply Newton's Method.