Lecture 21 : Minimizing a Function Step by Step

Optimization

Taylor Series (Basic Facts)

$$F(x+\Delta x) \approx F(x) + \Delta x \frac{dF}{dx} + \frac{1}{2}(\Delta x)^2 \frac{d^2F}{dx^2}$$ for one variable x

$$F(x+\Delta x) \approx F(x) + (\Delta x)^T \nabla F(x) + \frac{1}{2}(\Delta x)^T H (\Delta x)$$ for $x = (x_1, \cdots, x_n)$

Parallel version (Think of f as ∇F) → gradient of $F(x)$

$f = (f_1(x), \cdots, f_n(x))$ n functions

$x = (x_1, \cdots, x_n)$ n equations

$$f(x+\Delta x) = f(x) + J\,\Delta x$$

Solve $f = 0$

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

$$\boxed{x_{k+1} = x_k - J(x_k)^{-1} f(x_k)}$$ Newton's method (1)

Straightforward example

$$f(x) = x^2 - 9 = 0$$

$$x_{k+1} = x_k - \frac{1}{2x_k}(x_k^2 - 9) = \frac{1}{2}x_k + \frac{9}{2}\frac{1}{x_k}$$

Looking at error to $x^* = 3$ (solution)

$$(x_{k+1} - 3) = \frac{1}{x_k}\left[\frac{9}{2} + \frac{1}{2}x_k^2 - 3x_k\right]$$

$$= \frac{1}{2x_k}(x_k - 3)^2$$

$\therefore$ If starting point x_k is close to x^*, it achieves quadratic convergence, the error is being squared each step.

minimize $F(x)$ $\approx$ solving $\underset{f}{\underset{=}{\nabla F}} = 0$

① Steepest descent (Linear rate of convergence)

$F(x)$ is a function of n variables

$x = (x_1, \cdots, x_n)$

$$x_{k+1} = x_k - s_k \nabla F$$

stepsize s / learning rate

② Newton's method (Fast convergence rate, but expensive to calculate H^{-1})

$$x_{k+1} = x_k - H^{-1}(\nabla F)$$

→ equivalent to (1)

Jacobian of f is Hessian

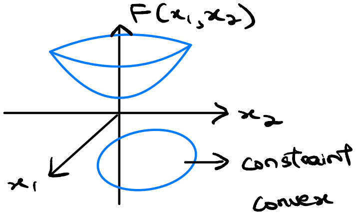
$\boxed{f = \nabla F}$

Convex set $F(x)$

K is a convex set — If x and y are in K, so is this the line from x to y

F is a convex function — The set of points on and above the graph of F is convex

Convex minimization : $\min\limits_{x \text{ in convex set}} F(x)$



Facts: ① $f(x)$ is convex if $\frac{d^2 f}{dx^2} \geq 0$

② $F(x_1, \ldots, x_n)$ is convex if its second derivative matrix $H(x)$ is positive semidefinite at all x.