CS 450: Numerical Anlaysis

Lecture 15

Chapter 6 Numerical Optimization

Secant Updating Methods and Basics of Numerical Optimization

Edgar Solomonik

Department of Computer Science
University of Illinois at Urbana-Champaign

March 7, 2018

Secant Updating Methods

In solving a nonlinear equation, seek approximate Jacobian $\boxed{J_f(x_k)}$ for each x_k

- Find $\underline{B_{k+1}} = B_k + \delta B_k \approx J_f(x_{k+1})$, so as to approximate *secant equation*

$$B_{k+1}\underbrace{(x_{k+1} - x_k)}_{\delta x} = \underbrace{f(x_{k+1}) - f(x_k)}_{\delta f}$$

· under determined

· $(B + \delta B_{k+1})\,\delta x = \delta f \Rightarrow \underbrace{\delta B_{k+1}}\,\delta x = \underbrace{\delta f - B_k \delta x}$

- *Broyden's method* is given by minimizing $||\delta B_k||_F$:

rank-1 perturbation $\quad \delta B_k = \dfrac{\delta f - B_k \delta x}{||\delta x||^2}\delta x^T = | \cdot \square |$

Sherman-Morrison formula

$B_{k+1}^{-1} \leftarrow B_k^{-1} + \dots \quad O(n^2)$ computation

Newton-Like Methods

- Can dampen step-size to improve reliability of Newton or Broyden iteration:

$$x_{k+1} = x_k + \alpha_k s$$

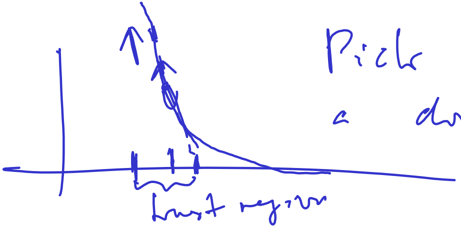
pick α_k so that $\|f(x_{k+1})\| \le \|f(x_k)\|$

pick α_k so that $\|f(x_{k+1})\|$ is minimized

$\alpha_k = 1 \Rightarrow$ Newton's method

- *Trust region methods* provide general step-size control:

trust region is region including x_k where we trust approximation of derivatives at x_k



Pick next guess to be in a direction within the trust region that minimizes $\|f(x_{k+1})\|$

Numerical Optimization

- Our focus will be on *continuous* rather than *combinatorial* optimization:

$$\min_{\boldsymbol{x}} \underline{f(\boldsymbol{x})} \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = \mathbf{0} \quad \text{and} \quad \boldsymbol{h}(\boldsymbol{x}) \leq \mathbf{0}$$

assume f is differentiable

if $g = h = 0$ for all x, then problem is unconstrained

otherwise constrained

- We consider linear, quadratic, and general nonlinear optimization problems:

linear programming if f, g, h are linear (affine)

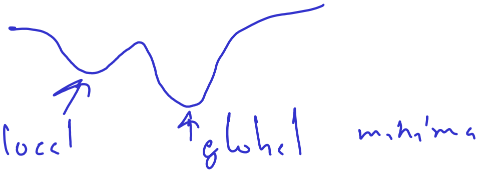
nonlinear programming which include

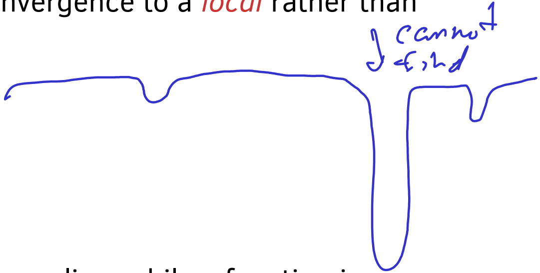
quadratic programming, f is quadratic

g, h are linear

Local Minima and Convexity

- Without knowledge of the analytical form of the function, numerical optimization methods at best achieve convergence to a *local* rather than *global* minimum:





- A set is *convex* if it includes all points on any line, while a function is (strictly) convex if its (unique) local minimum is always a global minimum:

if $x, y \in S$, $\forall \alpha [0,1]$

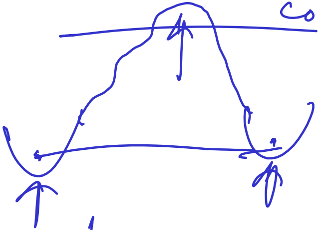
$$\alpha x + (1-\alpha) y \in S$$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

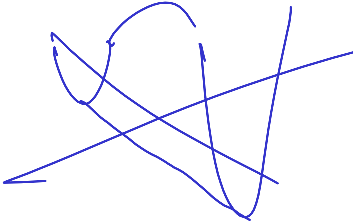
Convexity of Functions

concavity

ok



local minima



Existence of Local Minima

- *Level sets* are all points for which f has a given value, sublevel sets are all points for which the value of f is less than a given value:

$$L(z) = \{x : f(x) = z\} \quad \text{level set}$$

$$S(z) = \{x : f(x) \leq z\} \quad \text{sublevel set}$$

- If there exists a closed and bounded sublevel set in the domain of feasible points, then f has has a global minimum in that set:

given set $S(z)$ for some z

if $S(z)$ is contiguous and closed and bounded, then a global minimum exists in $S(z)$

Optimality Conditions

- If x is an interior point in the feasible domain and is a local minima, $\nabla f(\boldsymbol{x}) = \mathbf{0}$:

  satisfy constraints

  not on constraint boundary

  if $\nabla f(x) \neq 0$, $\exists i$, $(\nabla f(x))_i > 0$

  then $f(x - \delta x) < f(x)$

- Critical points x satisfy $\nabla f(\boldsymbol{x}) = 0$ and can be minima, maxima, saddle points:

  if $n = 1$, check $f''(x)$

Hessian Matrix

- To ascertain whether an interior point $\boldsymbol{x}$ for which $\nabla f(\boldsymbol{x}) = 0$ is a local minima, consider the Hessian matrix

$$\boldsymbol{H}_f(\boldsymbol{x}) = \boldsymbol{J}_{\nabla f}(\boldsymbol{x}) = \begin{bmatrix} \frac{d^2 f(\boldsymbol{x})}{dx_1^2} & \cdots & \frac{d^2 f(\boldsymbol{x})}{dx_1 dx_n} \\ \vdots & \ddots & \vdots \\ \frac{d^2 f(\boldsymbol{x})}{dx_n dx_1} & \cdots & \frac{d^2 f(\boldsymbol{x})}{dx_n dx_n} \end{bmatrix}$$

symmetric

$$f(x^* + \delta x) = f(x^*) + \underbrace{J_f(x^*)\,\delta x}_{0} + \overbrace{\frac{1}{2}\delta x^T H_f(x^*)\,\delta x}$$

$+ \ldots$

- If $\boldsymbol{x}^*$ is a minima of f, then $\boldsymbol{H}_f(\boldsymbol{x}^*)$ is positive semi-definite:

minimizer

$H_f(x^*)$ is positive definite

$\forall s,\ s^T H_f(x^*)\, s \geq 0,$

if for some δx, $\frac{1}{2}\delta x^T H_f(x^*)\,\delta x < 0$

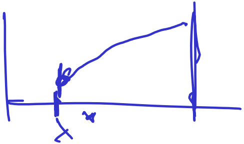
then $f(x+\delta x) < f(x)$

Optimality on Feasible Region Border

$h(x) \le 0$ ignored

- In equality-constrained optimization $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$, minimizers $\boldsymbol{x}^*$ are often found on the border of the feasible region (set of points satisfying constraints), in which case we must ensure any direction of decrease of f from $\boldsymbol{x}^*$ leads to an infeasible point, which gives us the condition:

Lagrange multipliers



$$\exists \boldsymbol{\lambda} \in \mathbb{R}^n, \quad -\nabla f(\boldsymbol{x}^*) = \boldsymbol{J}_{\boldsymbol{g}}^T(\boldsymbol{x}^*)\boldsymbol{\lambda}$$

directions of improvement all are constrained

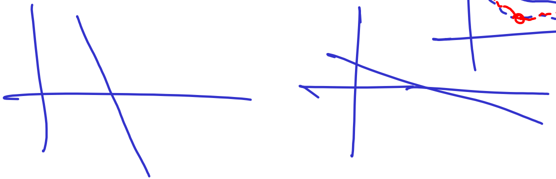
- Seek critical points in the Lagrangian function $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x})$, described by the nonlinear equation,

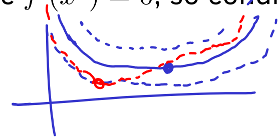
$$\nabla \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla f(\boldsymbol{x}) + \boldsymbol{J}_{\boldsymbol{g}}^T(\boldsymbol{x})\boldsymbol{\lambda} \\ \boldsymbol{g}(\boldsymbol{x}) \end{bmatrix} = \boldsymbol{0}$$

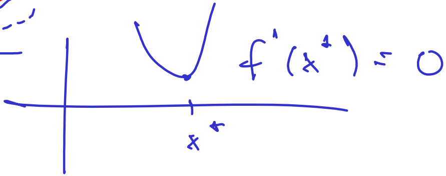
unconstrained with $\mathcal{L}$

Sensitivity and Conditioning

- The condition number of solving a nonlinear equations is $\boxed{1/f'(x^*)}$, however for a minimizer x^*, we have $f'(x^*) = 0$, so conditioning of optimization is inherently bad:







- To analyze worst case error, consider how far we have to move from a root x^* to perturb the function value by ϵ:

$$\epsilon = f(x^* + h) - f(x^*) = \cancel{f(x^*)} + \underbrace{f'(x^*)}_{0}h + f''(x^*)\frac{h^2}{2} + O(h^3)$$

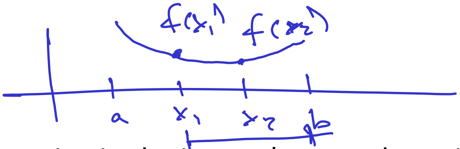
Taylor expansion

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

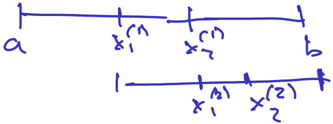
Golden Section Search

Golden ratio $= \frac{1+\sqrt{5}}{2}$

- Given bracket $[a, b]$ with a unique minimum (f is *unimodal* on the interval), if we consider points $f(x_1), f(x_2)$, $a < x_1 < x_2 < b$, we can discard subinterval $[a, x_1]$ or $[x_2, b]$:



- Since one point remains in the interval, we seek to pick x_1 and x_2 so they can be reused in the next iteration:



want $x_2^{(i)} = x_1^{(i+1)}$

$x_2 = (b-a) - x_1$

- We must ensure that the scaled distance of x_2 from the start of the interval $[x_1, 1]$ is the same as the distance of x_1 from 0, so $\frac{x_2 - x_1}{1 - x_1} = x_1$:

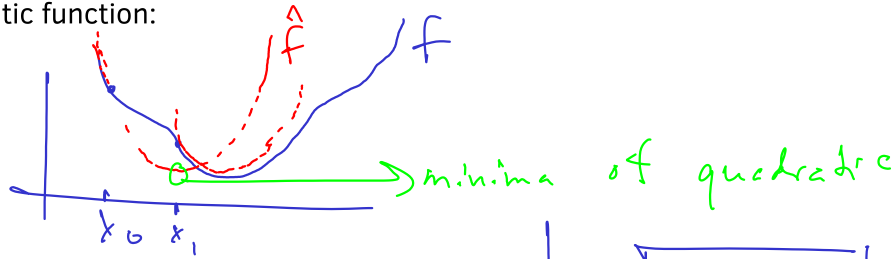
$[a, b] = [0, 1]$

$x_2 = 1 - x_1 \Rightarrow 1 - 2x_1 = x_1(1 - x_1)$

$x_1 = \frac{3 - \sqrt{5}}{2}$

Newton's Method for Optimization

- At each iteration, approximate function by quadratic and find minimum of quadratic function:



$h =$

- The new approximate guess will be given by $x_{k+1} - x_k = \boxed{-f'(x_k)/f''(x_k)}$:

$$f(x_k+h) \approx \hat{f}(x_k+h) = f(x_k) + f'(x_k)h + \frac{1}{2} f''(x_k) h^2$$

$$0 = \frac{\partial \hat{f}(x_k+h)}{\partial h} = f'(x_k) + f''(x_k) h \Rightarrow$$