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 9, 2018

Secant Updating Methods

In solving a nonlinear equation, seek approximate Jacobian $\boldsymbol{J_f}(\boldsymbol{x}_k)$ for each $\boldsymbol{x}_k$

- Find $\boldsymbol{B}_{k+1} = \boldsymbol{B}_k + \boldsymbol{\delta B}_k \approx \boldsymbol{J_f}(\boldsymbol{x}_{k+1})$, so as to approximate *secant equation*

$$\boldsymbol{B}_{k+1}(\underbrace{\boldsymbol{x}_{k+1} - \boldsymbol{x}_k}_{\boldsymbol{\delta x}}) = \underbrace{\boldsymbol{f}(\boldsymbol{x}_{k+1}) - \boldsymbol{f}(\boldsymbol{x}_k)}_{\boldsymbol{\delta f}}$$

  Generally, the secant equation is underdetermined, as we usually need n finite-difference formulas to determine $\boldsymbol{J_f}(\boldsymbol{x}_k)$, so the secant updating methods find only approximate $\boldsymbol{B}_{k+1}$, usually as a modification of $\boldsymbol{B}_k$.

- *Broyden's method* is given by minimizing $||\boldsymbol{\delta B}_k||_F$:

$$\boldsymbol{\delta B}_k = \frac{\boldsymbol{\delta f} - \boldsymbol{B}_k \boldsymbol{\delta x}}{||\boldsymbol{\delta x}||^2} \boldsymbol{\delta x}^T$$

  Note that $\boldsymbol{\delta B}_k$ is rank-1. Consequently, we can use the Sherman-Morrison formula to update $\boldsymbol{B}_{k+1}^{-1}$ with $O(n^2)$ work. Various other variants exist.

Newton-Like Methods

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

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{s}_k \quad \textit{where} \quad \alpha_k \leq 1$$

  can pick α_k *so to ensure* $||\boldsymbol{f}(\boldsymbol{x}_{k+1})|| < ||\boldsymbol{f}(\boldsymbol{x}_k)||$ *or by doing a line-search to minimize* $||\boldsymbol{f}(\boldsymbol{x}_k + \alpha_k \boldsymbol{s}_k)||$.

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

  Establish/maintain/update region within which step is expected to be accurate. Pick each step to stay within trust region while minimizing $||\boldsymbol{f}(\boldsymbol{x}_{k+1})||$. *Observe that the Newton-like generally seek to progress to a minima of* $||\boldsymbol{f}(\boldsymbol{x}_{k+1})||$, *and indeed much of the theory of these methods targets optimization.*

Numerical Optimization

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

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

  where f is assumed to be differentiable. Without the constraints, i.e. if $\boldsymbol{g} = 0$ and $\boldsymbol{h} = 0$, the optimization problem is referred to as unconstrained.

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

  If f, $\boldsymbol{g}$, and $\boldsymbol{h}$ are affine (linear and constant terms only) then we have linear programming problem. If f is quadratic while $\boldsymbol{g}$ and $\boldsymbol{h}$ are linear, then we have a quadratic programming problem, for which specialized methods exist. Generally, we have a nonlinear programming problem.

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:

  If the input domain is infinite, we may never be able to find a starting point near the 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:

  Set S is convex if

$$\forall \boldsymbol{x}, \boldsymbol{y} \in S, \alpha \in [0,1], \alpha x + (1-\alpha)y \in S.$$

  Function f is convex if

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}).$$

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:

$$L(z) = \{\boldsymbol{x} : f(\boldsymbol{x}) = z\}$$
$$S(z) = \{\boldsymbol{x} : f(\boldsymbol{x}) \leq z\}$$

- 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:

  Need a value z such that $S(z)$ has finite size, is contiguous, and includes its own boundary.

Optimality Conditions

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

  If $\nabla f(\boldsymbol{x})_i < 0$ an infinitesimal increment to x_i improves the solution, while if $\nabla f(\boldsymbol{x})_i > 0$ an infinitesimal decrement to x_i improves the solution.

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

  For scalar function f, can distinguish the three by considering sign of $f''(x)$.

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

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

The Hessian matrix is always symmetric.

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

If $\boldsymbol{H}_f(\boldsymbol{x}^*)$ is not positive semi-definite, there exists normalized vector $\boldsymbol{s}$ such that $\boldsymbol{s}^T \boldsymbol{H}_f(\boldsymbol{x}^*)\boldsymbol{s} < 0$, which means that for a sufficiently small α, $\hat{\boldsymbol{x}} = \boldsymbol{x}^* + \alpha \boldsymbol{s}$ will have be a better solution, $f(\hat{\boldsymbol{x}}) < f(\boldsymbol{x}^*)$, since the gradient is zero at $\boldsymbol{x}^*$ and decreases for an infinitesimal perturbation of $\boldsymbol{x}^*$ in the direction of $\boldsymbol{s}$.

Optimality on Feasible Region Border

- 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:

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

  $\boldsymbol{\lambda}$ *are referred to as the Lagrange multipliers.*

- 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,

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

  Seeking $\boldsymbol{\lambda}$ that maximizes the global minimum of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ defines the dual optimization problem.

Sensitivity and Conditioning

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

  Consider perturbation of function values for a function that changes slowly near the minimum.

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

$$\epsilon = f(x^* + h) - f(x^*) = \underbrace{f'(x^*)h}_{0} + \frac{1}{2}f''(x^*)h^2 + O(h^3)$$

  so the function value changes by $\frac{1}{2}f''(x^*)h^2$, which implies we need $h = O(\sqrt{\epsilon})$, i.e. a perturbation to the function value in the kth significant digit, could result in the solution changing in the $k/2$th significant digit.

Golden Section Search

- 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 $[x_2, b]$:

  If $f(x_1) < f(x_2)$ *consider only* $[a, x_2]$, *otherwise consider* $[x_1, b]$.
- 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:

  For example, when $f(x_1) > f(x_2)$, x_2 *is inside* $[x_1, b]$ *and we would like* x_2 *to serve as the* x_1 *for the next iteration.*
- 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 $\frac{x_2 - x_1}{1 - x_1} = x_1$:

  We pick $x_2 = 1 - x_1$, *which gives* $1 - 2x_1 = x_1(1 - x_1)$, *a quadratic equation* $x_1^2 - 3x_1 + 1 = 0$ *with solution* $x_1 = (3 - \sqrt{5})/2$.

Newton's Method for Optimization

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

  Pick quadratic function $\hat{f}$ as first three terms of Taylor expansion of f about x_k, matching value and first two derivatives of f at x_k.

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

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

since the function is quadratic, we can find its unique critical point to find its minima,

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