MAT 653: Statistical Simulation

Instructor: Dr. Wei Li Scribe: Peichen Yu

Sep 30th, 2021

Line Search by interpolation

Goal :Find α satisfies the sufficient reduction condition without being too small.

Let

$$\phi(\alpha) = f(x_k + \alpha p_k), (\alpha > 0),$$

The Armijo condition is

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \quad \phi'(0) = \nabla f(x_k)^T p_k,$$

Suppose our initial guess for next α is $\alpha_0 > 0$.

If Armijo condition if satisfied, then done.

If Armijo condition is not satisfied, then we know $\phi(\alpha)$ may be minimized furthor on $[0, \alpha_0]$

The interpolation idea is to approx ϕ by quadratic approx $\phi_q(\cdot)$, such that $\phi_q(\cdot) = a\alpha^2 + b\alpha + c$, satisfying

$$\begin{aligned}
\phi_q(0) &= \phi(0) \\
\phi_q(\alpha_0) &= \phi(\alpha_0) \\
\phi_q'(\alpha_0) &= \phi'(\alpha_0)
\end{aligned}$$

Solve for a, b, c in ternms of $\phi(0), \phi(\alpha_0)$ and $\phi'(\alpha_0)$.

The minimizer of $\phi_q(\cdot)$ over α is given by

$$\alpha_{min} = -\frac{b}{2a} = 0 - \frac{1}{2} \frac{(0 - \alpha_0)\phi'(\alpha_0)}{\left(\phi'(\alpha_0) - \frac{\phi(0) - \phi(\alpha_0)}{0 - \alpha_0}\right)}$$

$\alpha_1 \leftarrow \alpha_{min}$.

If α_1 satisfy Armijo condition, then done.

If not, then we interpolate a cubic function at $\phi(0), \phi'(0), \phi(\alpha_0)$ and $\phi(\alpha_1)$ obtaining $\phi_c(\alpha) = a\alpha^3 + b\alpha^2 + \alpha\phi'(0) + \phi(0)$

where

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\alpha_0^2 \alpha_1^2 (\alpha_1 - \alpha_0)} \begin{bmatrix} \alpha_0^2 & -\alpha_1^2 \\ -\alpha_0^3 & \alpha_1^3 \end{bmatrix} \begin{bmatrix} \phi(\alpha_1) - \phi(0) - \phi'(0)\alpha_1 \\ \phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0 \end{bmatrix}$$

The minimizer α_2 of $\phi_c(\cdot)$ turns out lies in the interval $[0, \alpha_1]$ and is given by $\alpha_2 = \frac{-b+\sqrt{b^2-3a\phi'(0)}}{3a}$

If α_2 satisfy Armijo condition, then done.

If not, continue the process using a cubic interpolate of $\phi(0), \phi'(0)$ and two more recent values of ϕ, until α is found to satisfy Amijo condition.

If any α_i is either too close to it's predecessor α_{i-1} or too much closer to 0, then we simply set $\alpha_i = \frac{\alpha_{i-1}}{2}$.

Modification of Hession matrix for Newton Method

By eigen-decomposition, we write for the square symmetric matrix $\nabla^2 f(x_k) = V_k D_k V_k^T$. By definition, $\nabla^2 f(x_k) V_k = V_k D_k$ where ($D_k = \delta_j$diagonal matrix consists of eigenvalues of the $\nabla^2 f(x_k)$).

If $\delta_j < \epsilon$, then set $\delta_j \leftarrow 2\epsilon$, call the new $D_k \to \widetilde{D}_k$; redefine Hessian to be $V_k \widetilde{D} V_K^T$. Note that $\nabla^2 f(x_k)^{-1} = V_k \widetilde{D}_k^{-1} V_k^T$.

No linear-LS-Problems

$f(x) = \frac{1}{2}\sum_{j=1}^m r_j^2(x), f : R^n \to R, r_j$:residual,

$$r = \begin{bmatrix} r_1(x) \\ r_2(x) \\ \vdots \\ r_m(x) \end{bmatrix}$$

then $f(x) = \frac{1}{2}||r(x)||^2$.

Let $J(x)$ denote the Jacobian of $r(x)$:

$$J(x) := \begin{bmatrix} (\nabla r_1(x))^T \\ (\nabla r_2(x))^T \\ \vdots \\ (\nabla r_m(x))^T \end{bmatrix}$$

We can show that

$$\nabla f(x) = \sum_{j=1}^m r_j(x) \nabla r_j(x) = J(x)^T r(x)$$

$$\nabla^2 f(x) = \sum_{j=1}^m \nabla r_j(x) \nabla r_j(x)^T + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x) = J(x)^T J(x) + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x)$$

Use Newton's Method: solve p_k from $\nabla^2 f(x_k) p_k = -\nabla f(x_k)$

Solve non-linear equatins

$r : R^n \to R^n, r(x) = 0$

$r(x_k + p) \approx r(x_k) + J(x_k)p$ (Taylor expansion for multiple equation and $J(x) = \nabla r(x)$ is the Jacobian of r–a $n \times n$ matrix)

$p = -J^{-1}(x_k) r(x_k)$ if $J(x_k)$ is invertible, then $x_{k-1} \leftarrow x_k + p_k$.

L2-regularzation

Example: Linear LS problem

$minf(x) = ||Ax - b||^2 = x^T A^T Ax - 2b^T Ax + b^T b$

Example: non-linear LS problem (logistic regression):

N data points: $(x_i y_i)_{i=1}^N, 0 \leq y_i \leq 1, f(x_i) \approx y_i$ logistic function:$\delta(t) = \frac{1}{1+e^{-t}}$, $y_i \approx \delta(a + bx_i)$, $y_i \approx \delta(\alpha + x_i^T \beta)$ if x is a vector.

LS problem: $min_{\alpha,\beta} \sum_{i=1}^N (\delta(\alpha + x_i^T \beta) - y_i)^2$

$$\widetilde{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$\widetilde{x_i} = \begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

$$\begin{aligned} \nabla RSS(\widetilde{\beta}) &= \frac{d}{d_{\widetilde{\beta}}} \sum_{i=1}^N (\delta(\widetilde{x}_i^T \widetilde{\beta}) - y_i)^2 \\ &= 2\sum_{i=1}^N (\delta(\widetilde{x}_i^T \widetilde{\beta}) - y_i)\delta(\widetilde{x}_i^T \widetilde{\beta})(1 - \delta(\widetilde{x}_i^T \widetilde{\beta}))\widetilde{x}_i \end{aligned}$$

Regularized logistic regression

$\sum_{i=1}^N (\delta(\alpha + x_i^T \beta) - y_i)^2 + \lambda ||\beta||^2$

Note: Regulizer is a simple convex function that is often added to a non-convex objective function slightly convexifying it and helping numerical optimazation technique avoid some poor solution in some flat area.