Stat 6341.001: Numerical Linear Algebra and Statistical Computing, Fall 2026 Topic 11

Newton's Method and Equality Constraints

Lecturer: Yun Wei *Scribe: Raghvendra Omer*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the instructor.*

1 Newton's method for unconstrained minimization

Consider the problem

$$\min_{x \in \operatorname{dom} f} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable. At a current point x, the second-order Taylor approximation is

$$f(x+v) \approx f(x) + \nabla f(x)^\top v + \frac{1}{2} v^\top \nabla^2 f(x) v.$$

Minimizing this quadratic approximation over v gives the Newton direction Δx_{nt} by solving

$$\nabla^2 f(x) \Delta x_{\text{nt}} = -\nabla f(x).$$

When the Hessian is nonsingular, this is equivalently

$$\Delta x_{\text{nt}} = -[\nabla^2 f(x)]^{-1} \nabla f(x).$$

The Newton decrement measures the size of the Newton step in the local quadratic metric:

$$\lambda^2 = \nabla f(x)^\top [\nabla^2 f(x)]^{-1} \nabla f(x) = \Delta x_{\text{nt}}^\top \nabla^2 f(x) \Delta x_{\text{nt}}.$$

1.1 Algorithm

Given a starting point $x \in \operatorname{dom} f$, tolerance $\epsilon > 0$.

repeat

1. Compute the Newton step and decrement:

$$\Delta x_{\text{nt}} := -\nabla^2 f(x)^{-1} \nabla f(x); \qquad \lambda^2 := \nabla f(x)^\top \nabla^2 f(x)^{-1} \nabla f(x).$$

2. **Stopping criterion.** Quit if $\lambda^2/2 \le \epsilon$.

3. **Line search.** Choose step size t by backtracking line search.

4. **Update.**

$$x := x + t \Delta x_{\text{nt}}.$$

1.2 Backtracking line search

Suppose Δx is a descent direction for f at $x \in \operatorname{dom} f$ $\alpha \in (0, \frac{1}{2})$ and $\beta \in (0, 1)$.

1. Initialize $t = 1$.
2. While
$$f(x + t\Delta x) \leq f(x) + \alpha t \nabla f(x)^\top \Delta x$$
is not satisfied, reduce the step size by setting $t \leftarrow \beta t$.
3. Return the first step size satisfying the condition.

The right-hand side is smaller than $f(x)$ because $\nabla f(x)^\top \Delta x < 0$. Thus, the condition requires a sufficient decrease in the objective value.

2 Convergence for Newton's method

The convergence argument uses the following assumptions:

1. f is twice continuously differentiable.
2. f is strongly convex with parameter $m > 0$; in particular,
$$\nabla^2 f(x) \succeq mI, \qquad x \in \mathcal{S}.$$
$$\nabla^2 f(x) \preceq MI, \qquad x \in \mathcal{S},$$
3. The Hessian is bounded above and is Lipschitz continuous. For constants $M > 0$ and $L > 0$,
$$\left\|\nabla^2 f(x) - \nabla^2 f(y)\right\|_2 \leq L \left\|x - y\right\|_2, \qquad x, y \in \mathcal{S}.$$

Here, $\|A\|_2$ denotes the spectral norm, which is the largest singular value of A.

$$\text{We will show } \exists \ \eta, \gamma \text{ s.t.} \quad 0 < \eta \leq \frac{m^2}{L}, \quad \gamma > 0,$$
$$\text{s.t. the following hold:}$$

- If $\|\nabla f(x^{(k)})\|_2 \geq \eta$, then
$$f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma.$$
- If $\|\nabla f(x^{(k)})\|_2 < \eta$, then the backtracking line search selects $t^{(k)} = 1$ and
$$\frac{L}{2m^2}\|\nabla f(x^{(k+1)})\|_2 \leq \left(\frac{L}{2m^2}\|\nabla f(x^{(k)})\|_2\right)^2.$$

Suppose that for some iteration k, $\left\|\nabla f(x^{(k)})\right\|_2 < \eta$, then $\left\|\nabla f(x^{(k)})\right\|_2 \leq m^2/L$

$$\frac{L}{2m^2}\left\|\nabla f(x^{(k)})\right\|_2 \leq \frac{1}{2} < 1.$$

$$\frac{L}{2m^2}\left\|\nabla f(x^{(k+1)})\right\|_2 \leq \left(\frac{L}{2m^2}\|\nabla f(x^{(k)})\|_2\right)^2 \leq \frac{L}{2m^2}\left\|\nabla f(x^{(k)})\right\|_2$$

$$\left\|\nabla f(x^{(k+1)})\right\|_2 \le \left\|\nabla f(x^{(k)})\right\|_2 < \eta$$

$$\boxed{\text{So } \left\|\nabla f(x^{(s)})\right\|_2 < \eta, \forall s \ge k.}$$

Note

$$\frac{L}{2m^2}\left\|\nabla f(x^{(k+1)})\right\|_2 \le (\frac{1}{2})^2 < 1$$

for all $\ell \ge k$,

$$\begin{aligned}
\frac{L}{2m^2}\left\|\nabla f(x^{(\ell)})\right\|_2 &\le \left(\frac{L}{2m^2}\left\|\nabla f(x^{(\ell-1)})\right\|_2\right)^2 \\
&\le \left(\frac{L}{2m^2}\left\|\nabla f(x^{(\ell-2)})\right\|_2\right)^{2^2} \\
&\le \cdots \\
&\le \left(\frac{L}{2m^2}\left\|\nabla f(x^{(k)})\right\|_2\right)^{2^{\ell-k}} \\
&\le \left(\frac{1}{2}\right)^{2^{\ell-k}}.
\end{aligned}$$

Recall the strongly convex objective-gap bound

$$f(x) - p^\star \le \frac{1}{2m}\left\|\nabla f(x)\right\|_2^2,$$

where $p^\star = \min_x f(x)$. Therefore,

$$\begin{aligned}
f(x^{(\ell)}) - p^\star &\le \frac{1}{2m}\|\nabla f(x^{(\ell)})\|_2^2 \\
&\le \frac{1}{2m}\left(\frac{2m^2}{L}\left(\frac{1}{2}\right)^{2^{\ell-k}}\right)^2 \\
&= \frac{2m^3}{L^2}\left(\frac{1}{2}\right)^{2^{\ell-k+1}}.
\end{aligned}$$

Let $x^{(0)}$ is the initial point then the number of iterations in Stage 1 is at most

$$\frac{f(x^{(0)}) - p^\star}{\gamma}.$$

For Stage 2, we want

$$f(x^{(\ell)}) - p^\star \le \epsilon.$$

To achieve an objective gap no larger than ε, it is sufficient that

$$\frac{2m^3}{L^2}\left(\frac{1}{2}\right)^{2^{\ell-k+1}} \le \epsilon.$$

$$\left(\frac{1}{2}\right)^{2^{\ell-k+1}} \le \frac{\epsilon}{\epsilon_0}, \qquad \text{where } \epsilon_0 = \frac{2m^3}{L^2}.$$

$$2^{-2^{\ell-k+1}} \leq \frac{\epsilon}{\epsilon_0}.$$

Taking logarithms base 2,

$$-2^{\ell-k+1} \leq \log_2\left(\frac{\epsilon}{\epsilon_0}\right).$$

Therefore,

$$2^{\ell-k+1} \geq \log_2\left(\frac{\epsilon_0}{\epsilon}\right).$$

Taking logarithms base 2 again,

$$\boxed{\ell - k + 1 \geq \log_2 \log_2\left(\frac{\epsilon_0}{\epsilon}\right)}$$

Combining the two stages gives the below iteration-count estimate:

$$\frac{f(x^{(0)}) - p^\star}{\gamma} + \log_2\left[\log_2\left(\frac{\varepsilon_0}{\varepsilon}\right)\right],$$

The number of iterations in the second stage is bounded by a term of order $\log_2 \log_2(\varepsilon_0/\varepsilon)$, which is very small (almost constant) compared with the first-stage bound. The main contribution is coming from the first stage.

3 Newton's method with equality constraints

Now consider the equality-constrained problem

$$\min_x f(x) \qquad \text{subject to} \qquad Ax = b,$$

where $A \in \mathbb{R}^{p\times n}$. Suppose the current iterate x is feasible, meaning $Ax = b$.

For a step v, the second-order approximation to the objective is

$$f(x+v) \approx f(x) + \nabla f(x)^\top v + \frac{1}{2} v^\top \nabla^2 f(x) v := \hat{f}(x+v).$$

Feasibility of the new point requires

$$A(x+v) = b.$$

Since $Ax = b$, the step must satisfy $Av = 0$. Therefore, the local quadratic problem is

$$\begin{aligned} \text{minimize}_v \quad & \nabla f(x)^\top v + \frac{1}{2} v^\top \nabla^2 f(x) v \\ \text{subject to} \quad & Av = 0. \end{aligned}$$

The constant term $f(x)$ can be omitted because it does not affect the minimizer.

Recall:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^T P x + q^T x + r \\ \text{subject to} \quad & Ax = b \end{aligned}$$

KKT system:

$$Ax^\star = b$$

$$Px^\star + q + A^T\nu^\star = 0$$

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

Optimize over Δx_{nt}:

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

Newton decrement

$$\lambda(x) = \left(\Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}}\right)^{1/2}$$

$$f(x) - p^\star \approx f(x) - \inf_{\substack{v \\ A(x+v)=b}} \left\{\hat{f}(x+v)\right\} = \frac{\lambda(x)^2}{2},$$

3.1 Algorithm for Newton's method with equality constraints

Given a starting point $x \in \operatorname{dom} f$ satisfying $Ax = b$ and a tolerance $\varepsilon > 0$, repeat the following steps.

1. Compute the constrained Newton step Δx_{nt} by solving the KKT system. Compute the squared Newton decrement
$$\lambda(x)^2 = \Delta x_{\mathrm{nt}}^\top \nabla^2 f(x) \Delta x_{\mathrm{nt}}.$$
2. **Stopping criterion:** stop if
$$\frac{\lambda(x)^2}{2} \le \varepsilon.$$
3. **Line search:** choose a step size t using backtracking line search
4. **Update:** set
$$x \leftarrow x + t\Delta x_{\mathrm{nt}}.$$

Since the constrained Newton direction satisfies $A\Delta x_{\mathrm{nt}} = 0$, each update preserves $Ax = b$.

3.2 Feasible descent directions

A vector $v \in \mathbb{R}^n$ is a feasible direction for the equality constraint if

$$Av = 0.$$

It is a feasible descent direction at x if

$$Av = 0 \qquad \text{and} \qquad f(x+tv) < f(x) \quad \text{for all sufficiently small } t > 0.$$

Reference

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK. Available at: `https://web.stanford.edu/~boyd/cvxbook/`