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

Equality Constrained Minimization Problems and Newton's Method

Lecturer: Yun Wei *Scribe: Matthew Bidwell*

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 Introduction

The previous lecture considered convex minimization with linear equality constraints. We first review its optimality conditions, then eliminate the constraints by parameterizing the feasible set. This reduces the problem to unconstrained minimization. We next introduce Newton's method for unconstrained minimization, interpret its search direction using Taylor approximations, and derive the Newton decrement and a stopping criterion.

2 Review: equality-constrained minimization

Consider

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b, \end{aligned} \tag{1}$$

where f is convex and twice continuously differentiable, $A \in \mathbb{R}^{p\times n}$, and $\operatorname{rank}(A) = p$. Thus the equality constraints have no redundant rows. At a primal optimum x^*, the KKT conditions are

$$\nabla f(x^*) + A^T\nu^* = 0, \qquad Ax^* = b, \tag{2}$$

where $\nu^* \in \mathbb{R}^p$ is dual optimal.

2.1 Quadratic objective

For the quadratic objective

$$f(x) = \frac{1}{2}x^T Px + q^T x + r,$$

the gradient is $Px + q$. Consequently, (2) becomes the linear system

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}. \tag{3}$$

The block matrix has size $(n+p)\times(n+p)$. If it is invertible, the system has a unique solution. If the feasible convex quadratic problem has no solution to (3), its objective is unbounded below.

3 Eliminating equality constraints

3.1 Parameterizing the feasible set (Section 10.1.2)

Choose any particular solution $\widehat{x}$ of $Ax = b$, and let $F \in \mathbb{R}^{n\times(n-p)}$ have columns forming a basis of the null space of A. In particular,

$$AF = 0, \qquad \operatorname{range}(F) = \operatorname{null}(A).$$

Every point can be written as

$$\{x : Ax = b\} = \{Fz + \widehat{x} : z \in \mathbb{R}^{n-p}\}. \tag{4}$$

To verify, take $x = Fz + \widehat{x}$. Then

$$Ax = A(Fz + \widehat{x}) = AFz + A\widehat{x} = 0 + b = b.$$

Define $\widetilde{f}(z) = f(Fz + \widehat{x})$. Problem (1) is equivalent to the unconstrained problem

$$\underset{z \in \mathbb{R}^{n-p}}{\text{minimize}} \quad \widetilde{f}(z).$$

After finding a minimizer z^*, we recover a minimizer of the original problem as $x^* = Fz^* + \widehat{x}$.

3.2 Optimal allocation with a resource constraint (Example 10.1)

Consider the separable objective and resource constraint

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} f_i(x_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = b. \end{aligned}$$

Here $A = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$ has rank one. Eliminating the last variable gives

$$x_n = b - \sum_{i=1}^{n-1} x_i,$$

and hence the unconstrained function in $n-1$ variables is

$$\sum_{i=1}^{n-1} f_i(x_i) + f_n\left(b - \sum_{i=1}^{n-1} x_i\right).$$

Now the constraint is removed.
This substitution can be written in the format of (4). Let e_n be the last canonical basis vector in $\mathbb{R}^n$, and choose

$$\widehat{x} = be_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b \end{bmatrix}, \qquad F = \begin{bmatrix} I_{n-1} \\ -\mathbf{1}^T \end{bmatrix} \in \mathbb{R}^{n \times (n-1)},$$

where $\mathbf{1} \in \mathbb{R}^{n-1}$ is the vector of ones. We have $A\widehat{x} = b$, $AF = 0$, and the columns of F are linearly independent, so they form a basis of $\text{null}(A)$. Therefore

$$Fz + \widehat{x} = \begin{bmatrix} z \\ b - \sum_{i=1}^{n-1} z_i \end{bmatrix}, \qquad z \in \mathbb{R}^{n-1},$$

which recovers precisely the elimination formula above.

4 Newton's method for unconstrained minimization (Section 9.5)

Now consider minimizing $f(x)$ without equality constraints. At a current point $x \in \text{dom}(f)$, the Newton step is

$$\Delta x_{\text{nt}} = -[\nabla^2 f(x)]^{-1} \nabla f(x). \tag{5}$$

Equivalently, it solves the linear system $\nabla^2 f(x) \Delta x_{\text{nt}} = -\nabla f(x)$.

4.1 The Newton step is a descent direction

For a direction v, the directional derivative at x is

$$\frac{d}{dt} f(x+tv)\bigg|_{t=0} = \langle \nabla f(x), v \rangle = \nabla f(x)^T v.$$

A negative value means that v is a descent direction. Substituting (5) yields

$$\frac{d}{dt} f(x+t\Delta x_{\mathrm{nt}})\bigg|_{t=0} = -\nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x) \leq 0.$$

4.2 One-dimensional intuition

In one dimension, we seek a zero of $f'(x)$. Approximating f' by its tangent line at x_k gives

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

The next iterate is the horizontal-axis intercept of that tangent line. Repeating the construction gives successive Newton iterates.

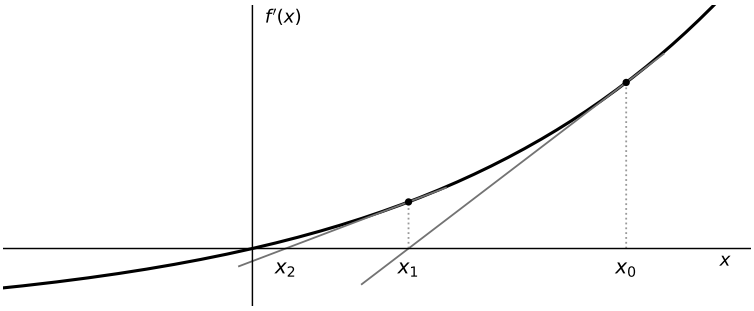


Figure 1: Illustration of the tangent-line construction applied to a general f'..

Geometrically, Newton's method uses a tangent line to approximate $f'(x)$ and find where $f'(x) = 0$. The tangent line has the form

$$g'(x) = ax + b.$$

This means that we are approximating $f(x)$ using a quadratic function of the form

$$g(x) = \frac{1}{2}ax^2 + bx + c.$$

The derivative of this quadratic is $g'(x) = ax + b$, so finding the zero of the tangent-line approximation of $f'(x)$ is equivalent to finding the minimum of the quadratic approximation of $f(x)$ when $a > 0$.

4.3 First interpretation: minimizing a quadratic approximation

The second-order Taylor approximation about x is

$$\widehat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v. \tag{6}$$

Treat the right-hand side as a function $g(v)$. Its gradient is

$$\nabla g(v) = \nabla f(x) + \nabla^2 f(x) v.$$

Because the Hessian of g is positive definite, the condition $\nabla g(v) = 0$ is necessary and sufficient for its unique minimum. Solving gives

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

Thus the Newton step minimizes the local quadratic model.

4.4 Second interpretation: linearizing the optimality condition

For unconstrained convex minimization, the first-order optimality condition is $\nabla f(x + v) = 0$. The first order Taylor expansion gives

$$\nabla f(x + v) \approx \nabla f(x) + \nabla^2 f(x)v.$$

Setting this approximation equal to zero again gives $v = \Delta x_{\text{nt}}$. Unlike the negative gradient direction, the Newton direction uses curvature information through the Hessian. If $\nabla^2 f(x) = I$, then $\Delta x_{\text{nt}} = -\nabla f(x)$, simplifying to the standard gradient descent direction.

5 Newton decrement and stopping criterion

Definition 5.1. The Newton decrement at x is

$$\lambda(x) = \left(\nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x)\right)^{1/2}. \tag{7}$$

It measures the improvement predicted by the quadratic approximation. Since (6) is minimized at the Newton step,

$$\begin{aligned} f(x) - p^* = f(x) - \inf_y \widehat{f}(y) &= f(x) - \widehat{f}(x + \Delta x_{\text{nt}}) \\ &= -\nabla f(x)^T \Delta x_{\text{nt}} - \frac{1}{2}\Delta x_{\text{nt}}^T \nabla^2 f(x) \Delta x_{\text{nt}} \\ &= \nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x) - \frac{1}{2}\nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x) \\ &= \frac{1}{2}\lambda(x)^2. \end{aligned}$$

In simplifying the quadratic term, we use symmetry of the Hessian and its inverse.

5.1 Newton's method algorithm

Given a starting point $x \in \text{dom}(f)$ and tolerance $\epsilon > 0$, repeat:

1. Compute the Newton step and squared decrement:

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

2. Stop if $\lambda(x)^2/2 \leq \epsilon$.
3. Choose a step size t by backtracking line search.
4. Update $x \leftarrow x + t\Delta x_{\text{nt}}$.

6 Convergence discussion

Newton's method can converge much faster than gradient descent because it uses curvature information in addition to the gradient. Further convergence analysis will be discussed in the next lecture.

7 References

[1] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. Sec 10.1 and Sec 9.5