Equality constrained minimization

- equality constrained minimization
- eliminating equality constraints
- Newton's method with equality constraints
- infeasible start Newton method

Equality constrained minimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

- f is convex, twice continuously differentiable
- $A \in R^{p \times n}$ with rank $A = p$
- assume p^* is finite and attained

optimality conditions: x^* is optimal iff $\exists \nu^*$ such that

$$\nabla f(x^*) + A^T \nu^* = 0, \quad Ax^* = b$$

Equality constrained minimization: quadratic

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TPx + q^Tx + r \\ \text{subject to} & Ax = b \end{array}$$

optimality conditions: x^* is optimal iff $\exists \nu^*$ such that

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

- coefficient matrix is called **KKT matrix**
- KKT matrix is non-singular
  - if $Ax = 0, x \neq 0 \implies x^TPx > 0$
  - if $P + A^T \succ 0$

Newton step

Newton step Δx_{nt} at feasible x is given by solution v of

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

interpretations

- Δx_{nt} solves second-order approximation of f at x

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b \end{array}$$

- Δx_{nt} solves first-order approximation of optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x) v + A^T w = 0, \quad A(x+v) = b$$

Newton decrement

Newton decrement

$$\lambda(x) = (\Delta x_{nt} \nabla^2 f(x) \Delta x_{nt})^{1/2} = \|\nabla^2 f(x)^{1/2} \Delta x_{nt}\|$$

which gives an estimate of $f(x) - p^*$ using quadratic approximation

$$f(x) - \inf_{Ay=b} \hat{f}(y) = \frac{1}{2}\lambda(x)^2$$

Newton direction with feasible start is a descent direction:

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

Newton's method with equality constraints: feasible start

---

given a starting point $x \in \mathrm{dom}\, f$ with $Ax = b$, tolerance $\epsilon > 0$
repeat

1. Compute the Newton step and decrement Δx_{nt}, $\lambda(x)$
2. Stopping criterion: quit if $\lambda^2/2 \leq \epsilon$
3. Line search: choose a step size $t > 0$ by backtracking line search
4. Update: $x := x + t\Delta x$

Starting point is feasible, and $f(x^{(k+1)}) < f(x^{(k)})$

Eliminating equality constraints

Find a matrix $F \in R^{n\times(n-p)}$ and a $\hat{x}$ such that

$$\{x \mid Ax = b\} = \{Fz + \hat{x} \mid z \in R^{n-p}\}$$

Reduced problem:

$$\text{minimize} \quad \tilde{f}(z) = f(Fz + \hat{x})$$

Remark:

- F is any matrix whose range is the nullspace of A: $AF = 0$
- Newton method with equality constraints: iterates are

$$x^{(k+1)} = Fz^{(k)} + \hat{x}$$

Hence convergence is the same as unconstrained Newton's method

Newton step with infeasible start

Linearizing optimality conditions at infeasible x:

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

Newton step with infeasible start: primal-dual interpretation

Optimality condition: $r(y) = 0$ with

$$y = (x, \nu), \qquad r(y) = (\nabla f(x) + A^T \nu, Ax - b)$$

Linearizing $r(y) = 0$

$$r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$$

which leads to

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ \Delta \nu_{nt} \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^T \nu \\ Ax - b \end{bmatrix}$$

Note that $w = \nu + \Delta \nu_{nt}$

Infeasible start Newton method

---

given a starting point $x \in \mathbf{dom}\, f$, ν, tolerance $\epsilon > 0$
repeat

1. Compute primal and dual Newton steps Δx_{nt}, $\Delta \nu_{nt}$
2. Line search on $\|r\|_2$: choose a step size $t > 0$ by backtracking line search
3. Update: $x := x + t\Delta x_{nt}$, $\nu := \nu + t\Delta \nu_{nt}$

until $Ax = b$ and $\|r(x, \nu)\|_2 \le \epsilon$

Remark:

- not a descent method: $f(x^{(k+1)}) > f(x^{(k)})$ is possible
- directional derivative of $\|r(y)\|_2$ in direction $\Delta y = (\Delta x_{nt}, \Delta \nu_{nt})$ is

$$\frac{d}{dt}\|r(y + t\Delta y)\|_2 \bigg|_{t=0} = -\|r(y)\|_2$$