- classical convergence analysis for newton method

  $\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \le L \|x-y\|_2$ affine transformation changes the norm

  we have to have a way to say that {third derivative is small} that's {affine invariant}

- self-concordance

  classical convergence analysis:

  depends on unknown constants

  bound is not affine-invariant

  Convergence analysis via self-concordance

  does not depend on unknown analysis

  gives affine-invariant bound

  applies to special class of convex functions ( self-concordant )

- Self-concordant functions

  $f: \mathbb{R} \to \mathbb{R}$ is self-concordant if $|f'''(x)| \le 2 f''(x)^{3/2}$ for all $x \in \mathrm{dom} f$

  $f: \mathbb{R}^n \to \mathbb{R}$ is self-concordant if $g(t) = f(x+tv)$ is self-concordant for all $x \in \mathrm{dom} f$, $v \in \mathbb{R}^n$

  eg on $\mathbb{R}$:

  linear and quadratic functions ( $f'''$ is $0$ )

  negative logarithm $f(x) = -\log(x)$

  $f(x) = x\log x - \log x$

  affine invariance:

  if $f: \mathbb{R} \to \mathbb{R}$ is self-concordant, the $\tilde{f}(y) = f(ay+b)$ is self-concordant

  $\tilde{f}'''(y) = a^3 f'''(ay+b)$ $\tilde{f}''(y) = a^2 f''(ay+b)$

  $|f'''(x)| \le 2 f''(x)^{3/2}$ ; the 3/2 power makes the a drop away.

- Self-concordant calculus

  preserved by positive scaling $a \ge 1$, and sum.

  preserved under affine composition

eg. log-barrier

$f(y) = -\sum \log(b_i - a_i^T x)$

$\log'(y) = -1/y \quad \log''(y) = 1/y^2 \quad \log'''(y) = -2/y^3$

$|2/y^3| \leq (1/y^2)^{\frac{3}{2}}$

$\therefore -\log(y)$ is s.c $\Rightarrow -\log(b_i - a_i^T x)$ is s.c $\Rightarrow -\sum \log(b_i - a_i^T x)$ is s.c

eg. minus logdet

$$\begin{aligned}\tilde{f}(t) &= -\log\det(X + tV) \\ &= -\log\det[X^{1/2}(I + t \cdot X^{-1/2} V X^{-1/2}) X^{1/2}] \\ &= -\log\det X - \log\det(I + t \cdot X^{-1/2} V X^{-1/2}) \\ &= -\log\det X - \sum \log(1 + t \cdot \lambda_i)\end{aligned}$$

$\lambda_i$ are eigenvalues of $X^{-1/2} V X^{-1/2}$

is self-concordant in $t$

- Convergence analysis via self-concordance

there exist a constant $\eta \in (0, 1/4]$, $\gamma > 0$ such that

1. if $\lambda(x) > \eta$

$$f(x^{(k+1)}) \leq f(x^{(k)}) - \gamma$$

2. if $\lambda(x) \leq \eta$

$$2\lambda(x^{(k+1)}) \leq [2\lambda(x^{(k)})]^2$$

( $\eta$ and $\gamma$ depend only on line search parameter $\alpha, \beta$ )

complexity bound: number of newton iteration bounded by

$$\frac{f(x^{(0)}) - p^*}{\gamma} + \log_2 \log_2 (1/\epsilon)$$

- implementation

main effort in each iteration. $\Delta x = -\nabla^2 f(x)^{-1} \nabla f(x)$

( solving a linear system $Hv = -g \quad H = \nabla^2 f(x) \quad g = -\nabla f(x)$ )

via cholesky factorization

$$H = LL^T \quad \Delta x = L^{-T} L^{-1} g$$

$(1/3)n^3$ flops for unstructured system

$\ll (1/3)n^3$ if $H$ is structured

eg. Sparse hessian

$$f(x) = \sum \exp(x_i - x_{i+1})$$

the hessian is tridiagonal

eg. Diagonal-plus-low-rank

$$f(x) = \underbrace{\sum_i \psi_i(x_i)}_{\text{hessian diagonal}} + \psi_0(Ax+b) \qquad H = D + A^T H_0 A$$

$A \in \mathbb{R}^{p\times n}$ with $p \ll n$, dense

{ factor $H_0 = \overbrace{L_0}^{p\times p} L_0^T \Rightarrow \left[D + \overbrace{(L_0^T A)}^{n\times p}{}^T (L_0^T A)\right] v = -g$

$$\begin{array}{c} n \\ p \end{array}\begin{bmatrix} D & (L_0^T A)^T \\ L_0^T A & -I \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} -g \\ 0 \end{bmatrix} \qquad \begin{cases} Dv + (L_0^T A)^T w = -g \\ w = L_0^T A v \end{cases}$$

$\Downarrow$

$$-I - \underbrace{(L_0^T A D^{-1} A^T L_0)}_{p\times p} w = -(L_0^T A D^{-1} g)$$

} complexity $2p^2 n$ (dominated by $L_0^T A D^{-1} A^T L_0$)

- Equality constrained minimization

$$\text{min. } f(x)$$
$$\text{s.t. } Ax = b$$

$f$ convex, twice continuously differentiable

$A \in \mathbb{R}^{p\times n}$ with $\operatorname{rank} A = p$

$p^*$ is finite and attained

optimality condition: $x^*$ is optimal iff there is a $v^*$ s.t.

$$\nabla f(x^*) + A^T v^* = 0 \qquad Ax^* = b$$

eg. equality constrained quadratic minimization $(P \in S_{++}^{n})$

$$\min. \ \tfrac{1}{2}x^TPx + q^Tx + r$$
$$\text{s.t.} \quad Ax = b$$

$$L(x, \nu) = \tfrac{1}{2}x^TPx + q^Tx + r + \nu^T(Ax - b)$$
$$\nabla_x L = Px + q + A^T\nu$$

$$\underbrace{\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}}_{\text{KKT matrix}} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \begin{matrix} \} \text{ dual feasibility} \\ \} \text{ primal feasibility} \end{matrix}$$

KKT matrix is nonsingular iff:

$$Ax = 0, \ x \neq 0 \Rightarrow x^TPx > 0$$

P is pd on the nullspace of A

equivalent to the nonsingularity:

$$P + A^TA \succ 0$$

· Eliminating equality constraints

represent $\{x \mid Ax = b\}$ as

$$\{x \mid Ax = b\} = \{Fz + \hat{x}\}$$

$\hat{x}$ is any particular solution

$F$ has nullspace of $A$ as columns

$$\Downarrow$$

$$\min. \ f(Fz + \hat{x})$$

$$\Downarrow$$

$$x^* = Fz^* + \hat{x} \qquad \nu^* = -(AA^T)^{-1}A\nabla f(x^*)$$

eg. optimal allocation with resource constraint

$$\min. \ f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n)$$
$$\text{s.t.} \quad x_1 + \cdots + x_n = b \qquad (\mathbf{1}^Tx = b)$$

$$\text{choose} \quad \hat{x} = b \cdot e_n \qquad F = \begin{bmatrix} I \\ -\mathbf{1}^T \end{bmatrix} \in \mathbb{R}^{n \times (n-1)}$$


$$\Downarrow$$

$$\min. \ \underbrace{f_1(x_1) + \cdots + f_{n-1}(x_{n-1})}_{\text{hessian diagonal}} + \underbrace{f_n(b - x_1 - \cdots - x_{n-1})}_{\text{hessian rank-one}}$$

- Newton step with equality constraint

newton step of $f$ at feasible $x$ is given by

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

interpretation 1.

$\Delta x_{nt}$ solves the second-order taylor approximation

$$\min_v \ \hat{f}(x+v) = f(x) + \nabla f(x)^T v + \tfrac{1}{2} v^T \nabla^2 f(x) v$$
$$\text{s.t.} \quad A(x+v) = b$$

$$\mathcal{L}(v,w) = f(x) + \nabla f(x)^T v + \tfrac{1}{2} v^T \nabla^2 f(x) v + w^T(Ax + Av - b)$$

$$\left.\begin{array}{l} \nabla_v \mathcal{L} = \nabla f(x) + \nabla^2 f(x) v + A^T w = 0 \\ Av = 0 \end{array}\right\} \begin{array}{l} \text{dual feasible} \\ \text{primal feasible} \end{array}$$

interpretation 2.

$\Delta x_{nt}$ solve the optimaly condition.

$$\begin{cases} \nabla f(x+v) + A^T w = 0 \\ A(x+v) = b \end{cases} \Rightarrow \begin{array}{l} \nabla f(x) + \nabla^2 f(x) v + A^T w = 0 \\ Av = 0 \end{array}$$

- Newton decrement

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

An estimate of decrease

$$f(x) - \inf_y \hat{f}(y) = -\tfrac{1}{2} \lambda(x)^2$$

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix} \quad \left.\begin{cases} \nabla^2 f(x) \Delta x + A^T w = -\nabla f(x) \\ A \Delta x = 0 \end{cases}\right\} \Delta x^T \nabla^2 f(x) \Delta x + \Delta x^T A^T w = -\Delta x^T \nabla f(x)$$

$$\Downarrow$$

$$\Delta x^T \nabla^2 f(x) \Delta x = -\nabla f(x)^T \Delta x$$

$$\hat{f}(x+\Delta x) = f(x) + \nabla f(x)^T \Delta x + \tfrac{1}{2} \nabla f(x)^T \nabla^2 f(x) \nabla f(x)$$
$$= f(x) - \tfrac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x$$
$$= f(x) - \lambda(x)^2 / 2$$

Directional derivative in Newton direction

$$\frac{d}{dt} f(x + t \cdot \Delta x_{nt}) = \nabla f(x)^T \Delta x = -\lambda(x)^2$$

- Newton's method with equality constraints

given starting point $x \in \text{dom} f$ with $Ax = b$

repeat {

Compute the newton step and newton decrement $\Delta x_{nt}$, $\lambda(x)$

Stopping criterion: quit if $\lambda^2/2 \le \epsilon$

Line search: chose step size $t$

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

}

Affine invariance

- Newton's method and elimination

$$\min \ \tilde{f}(z) = f(Fz + \hat{x})$$

- Newton step at infeasible point

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

} primal residual; revert to standard newton if $r_p = 0$