EQUALITY CONSTRAINED OPTIMIZATION

Equality constrained minimization

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

- f convex, twice continuously differentiable

Equality constrained minimization

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

- f convex, twice continuously differentiable

optimality conditions: $x^\star$ is optimal iff there exists a $\nu^\star$ such that

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

dual feasibility equations

primal feasibility equations

KKT

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$

2. dual constraints: $\lambda \succeq 0$

3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$

4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

equality constrained quadratic minimization

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b \end{array}$$

(with $P \in \mathbf{S}^n_+$)

optimality condition:

equality constrained quadratic minimization

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

(with $P \in \mathbf{S}^n_+$)

optimality condition:

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

- coefficient matrix is called KKT matrix

Approaches

Eliminating equality constraints

Solving Dual Problem

Newton's method with equality constraints

Eliminating equality constraints

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

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

- $\hat{x}$ is (any) particular solution
- range of $F \in \mathbf{R}^{n \times (n-p)}$ is nullspace of A

reduced or eliminated problem

$$\boxed{\text{minimize} \quad f(Fz + \hat{x})}$$

- an unconstrained problem with variable $z \in \mathbf{R}^{n-p}$
- from solution $z^\star$, obtain $x^\star$ and $\nu^\star$ as

$$x^\star = Fz^\star + \hat{x},$$

Solving equality constrained problems via the dual

solve the dual, and then recover the optimal primal variable x*,

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

$$g(\nu) = -b^T\nu + \inf_x (f(x) + \nu^T A x)$$

Solving equality constrained problems via the dual

solve the dual, and then recover the optimal primal variable x*,

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

$$\begin{aligned} g(\nu) &= -b^T\nu + \inf_x (f(x) + \nu^T Ax) \\ &= -b^T\nu - \sup_x \left((-A^T\nu)^T x - f(x)\right) \\ &= -b^T\nu - f^*(-A^T\nu), \end{aligned}$$

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$$

$$\boxed{\text{maximize} \quad -b^T\nu - f^*(-A^T\nu).}$$

Solving equality constrained problems via the dual

solve the dual, and then recover the optimal primal variable x*,

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

$$\begin{aligned} g(\nu) &= -b^T\nu + \inf_x (f(x) + \nu^T Ax) \\ &= -b^T\nu - \sup_x \left((-A^T\nu)^T x - f(x)\right) \\ &= -b^T\nu - f^*(-A^T\nu), \end{aligned}$$

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$$

$$\boxed{\text{maximize} \quad -b^T\nu - f^*(-A^T\nu).}$$

Slater's condition holds

$$g(\nu^\star) = p^\star$$

we find an optimal dual variable v*, we reconstruct an optimal primal solution x* from it.

Example

$$\begin{array}{ll} \text{minimize} & f(x) = -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b, \end{array}$$

Example

$$\begin{array}{ll} \text{minimize} & f(x) = -\sum_{i=1}^n \log x_i \\ \text{subject to} & Ax = b, \end{array}$$

implicit constraint $x \succ 0$.

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x)) \qquad f^*(y) = \sum_{i=1}^n (-1 - \log(-y_i)) = -n - \sum_{i=1}^n \log(-y_i)$$

$$\boxed{\text{maximize} \quad g(\nu) = -b^T\nu + n + \sum_{i=1}^n \log(A^T\nu)_i}$$

Example

$$\begin{array}{ll} \text{minimize} & f(x) = -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b, \end{array}$$

implicit constraint $x \succ 0$.

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x)) \qquad f^*(y) = \sum_{i=1}^{n} (-1 - \log(-y_i)) = -n - \sum_{i=1}^{n} \log(-y_i)$$

$$\boxed{\text{maximize} \quad g(\nu) = -b^T \nu + n + \sum_{i=1}^{n} \log(A^T \nu)_i}$$

solve the dual feasibility equation,

$$\nabla f(x) + A^T \nu = -\,\mathbf{diag}(1/x_1, \ldots, 1/x_n) + A^T \nu = 0,$$

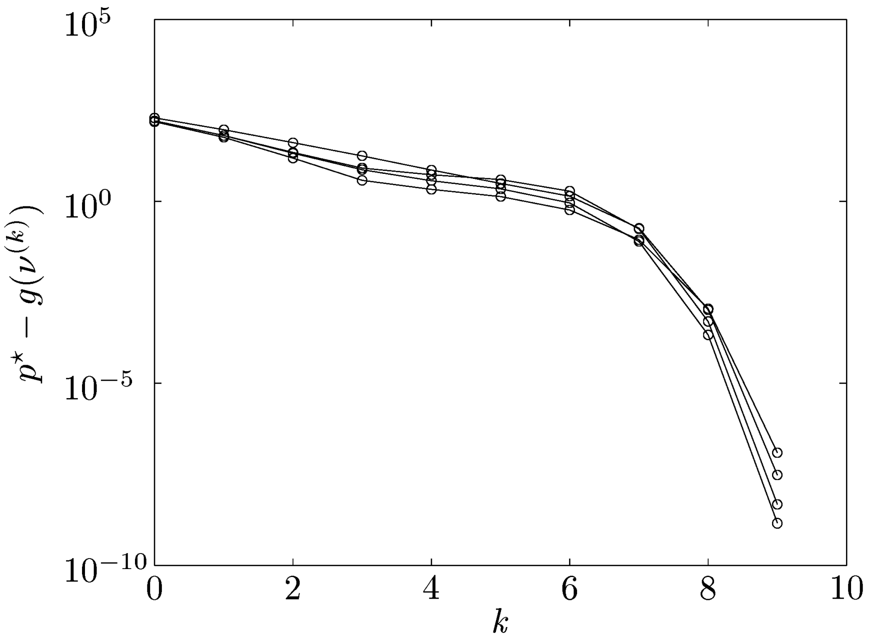
$$x_i(\nu) = 1/(A^T \nu)_i.$$

Example

Newton's method applied to the dual

$$\text{maximize} \quad g(\nu) = -b^T\nu + \sum_{i=1}^{n} \log(A^T\nu)_i + n$$

$p = 100$, $n = 500$



Newton's method with equality constraints

Newton's method without constraints, except for two differences:

✓The initial point must be feasible
✓the definition of Newton step is modified to take the equality constraints into account.

$$A\Delta x_{\mathrm{nt}} = 0.$$

Newton step

Definition via second-order approximation

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

at the feasible point x

Newton step

Definition via second-order approximation

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

at the feasible point x

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

Newton step

Definition via second-order approximation

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

at the feasible point x

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

quadratic minimization problem with equality constraints

✓can be solved analytically.

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

Another interpretation of the Newton step

$$Ax^\star = b, \qquad \nabla f(x^\star) + A^T \nu^\star = 0.$$

substitute $x + \Delta x_{\text{nt}}$ for $x^\star$ and w for $\nu^\star$

Another interpretation of the Newton step

$$Ax^\star = b, \qquad \nabla f(x^\star) + A^T \nu^\star = 0.$$

substitute $x + \Delta x_{\mathrm{nt}}$ for $x^\star$ and w for $\nu^\star$

replace the gradient term in the second equation by its linearized approximation near x

$$A(x + \Delta x_{\mathrm{nt}}) = b, \qquad \nabla f(x + \Delta x_{\mathrm{nt}}) + A^T w \approx \nabla f(x) + \nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w = 0.$$

Another interpretation of the Newton step

$$Ax^\star = b, \qquad \nabla f(x^\star) + A^T \nu^\star = 0.$$

substitute $x + \Delta x_{\mathrm{nt}}$ for $x^\star$ and w for $\nu^\star$

replace the gradient term in the second equation by its linearized approximation near x

$$A(x + \Delta x_{\mathrm{nt}}) = b, \qquad \nabla f(x + \Delta x_{\mathrm{nt}}) + A^T w \approx \nabla f(x) + \nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w = 0.$$

$$\boxed{A\Delta x_{\mathrm{nt}} = 0, \qquad \nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w = -\nabla f(x),}$$

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

Newton's method with equality constraints

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

Algorithm 10.1 *Newton's method for equality constrained minimization.*

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

Newton Decrement

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 step size t by backtracking line search.
4. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$.

Newton's method with equality constraints

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

Algorithm 10.1 *Newton's method for equality constrained minimization.*

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

repeat

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

Newton Decrement

- a feasible descent method: $x^{(k)}$ feasible and $f(x^{(k+1)}) < f(x^{(k)})$

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

Example

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b. \end{array}$$

$Ax^{(0)} = b,\ x^{(0)} \succ 0$

$p = 100,\ n = 500.$

Newton's method with equality constraints

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ w \end{bmatrix} = \begin{bmatrix} -g \\ 0 \end{bmatrix}$$

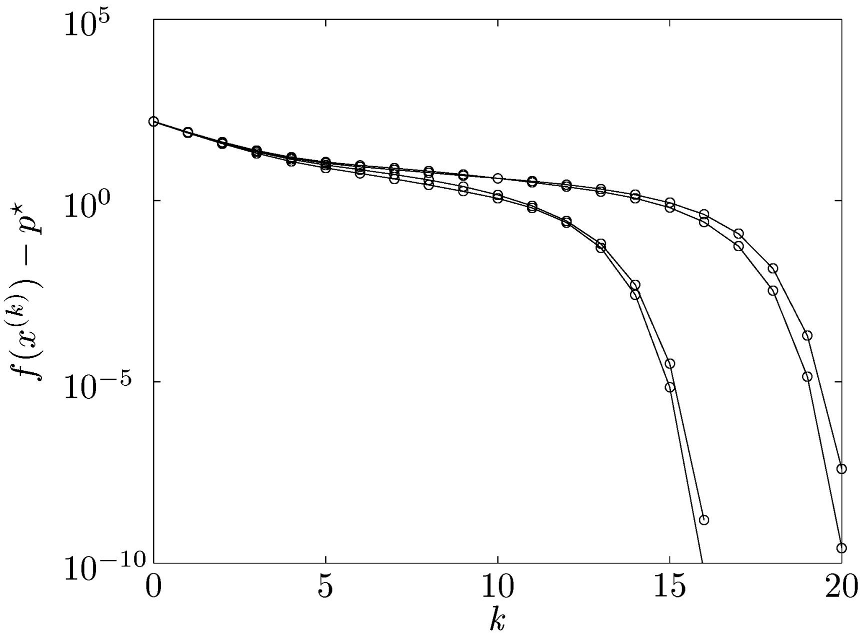
$$H = \mathbf{diag}(1/x_1^2, \ldots, 1/x_n^2), \text{ and } g = -(1/x_1, \ldots, 1/x_n)$$

Example

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b. \end{array}$$

$p = 100$, $n = 500$.

$Ax^{(0)} = b$, $x^{(0)} \succ 0$



Newton's method with equality constraints

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm nt} \\ w \end{bmatrix} = \begin{bmatrix} -g \\ 0 \end{bmatrix}$$

$H = \mathbf{diag}(1/x_1^2, \ldots, 1/x_n^2)$, and $g = -(1/x_1, \ldots, 1/x_n)$

Newton's method and elimination

minimize $f(x)$
subject to $Ax = b$

$\Longleftrightarrow$

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

- variables $z \in \mathbf{R}^{n-p}$
- $\hat{x}$ satisfies $A\hat{x} = b$; $\mathbf{rank}\, F = n - p$ and $AF = 0$

iterates in Newton's method for the equality constrained coincide with the iterates in Newton's method applied to the reduced problem

$$\Delta z_{\mathrm{nt}} = -\nabla^2 \tilde{f}(z)^{-1} \nabla \tilde{f}(z) \qquad \Delta x_{\mathrm{nt}} = F \Delta z_{\mathrm{nt}}$$

- Newton's method for $\tilde{f}$, started at $z^{(0)}$, generates iterates $z^{(k)}$

when started at $x^{(0)} = Fz^{(0)} + \hat{x}$, iterates are

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

Newton step at infeasible points

$$Ax^\star = b, \qquad \nabla f(x^\star) + A^T \nu^\star = 0.$$

✓Let x denote the current point, which we do not assume to be feasible
✓Our goal is to find a step Δx so that $x + \Delta x$ satisfies (at least approximately) the optimality conditions
✓The first-order approximation

$$\nabla f(x + \Delta x) \approx \nabla f(x) + \nabla^2 f(x) \Delta x$$

substitute $x + \Delta x$ for $x^\star$ and w for $\nu^\star$

Newton step at infeasible points

$$Ax^\star = b, \qquad \nabla f(x^\star) + A^T \nu^\star = 0.$$

✓Let x denote the current point, which we do not assume to be feasible
✓Our goal is to find a step Δx so that $x + \Delta x$ satisfies (at least approximately) the optimality conditions
✓The first-order approximation

$$\nabla f(x + \Delta x) \approx \nabla f(x) + \nabla^2 f(x) \Delta x$$

substitute $x + \Delta x$ for $x^\star$ and w for $\nu^\star$

$$A(x + \Delta x) = b, \qquad \nabla f(x) + \nabla^2 f(x) \Delta x + A^T w = 0.$$

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

Interpretation as primal-dual Newton step

✓Interpretation in terms of a primal-dual method
✓mean one in which we update both the primal variable, and the dual variable, in order to (approximately) satisfy the optimality conditions.

- write optimality condition as $r(y) = 0$, where

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

$r : \mathbf{R}^n \times \mathbf{R}^p \to \mathbf{R}^n \times \mathbf{R}^p$

Dual residual

Primal residual

- linearizing $r(y) = 0$ gives $r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$:

primal-dual step

$$\boxed{Dr(y)\Delta y_{\text{pd}} = -r(y).}$$

Interpretation as primal-dual Newton step

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

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

$$\Delta x_{\rm nt} = \Delta x_{\rm pd}, \qquad w = \nu + \Delta \nu_{\rm pd}$$

Residual norm reduction property

The Newton direction, at an infeasible point, is not necessarily a descent direction for f.

$$\frac{d}{dt} f(x + t\Delta x)\bigg|_{t=0} = \quad -\Delta x^T \nabla^2 f(x) \Delta x + (Ax - b)^T w,$$

- directional derivative of $\|r(y)\|_2$ in direction $\Delta y = (\Delta x_{\mathrm{nt}}, \Delta \nu_{\mathrm{nt}})$ is

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

Infeasible start Newton method

Algorithm 10.2 *Infeasible start Newton method.*

given starting point $x \in \mathbf{dom}\, f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.

repeat

1. Compute primal and dual Newton steps Δx_{nt}, $\Delta \nu_{\mathrm{nt}}$.
2. *Backtracking line search on* $\|r\|_2$.
   $t := 1.$
   while $\|r(x + t\Delta x_{\mathrm{nt}}, \nu + t\Delta\nu_{\mathrm{nt}})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2, \quad t := \beta t.$
3. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$, $\nu := \nu + t\Delta\nu_{\mathrm{nt}}$.

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

Example

infeasible start Newton method

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b. \end{array}$$

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm nt} \\ \Delta \nu_{\rm nt} \end{bmatrix} = - \begin{bmatrix} g + A^T \nu \\ Ax - b \end{bmatrix}$$

$H = \mathbf{diag}(1/x_1^2, \ldots, 1/x_n^2)$, and $g = -(1/x_1, \ldots, 1/x_n)$

the norm of the residual

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