Inequality constrained minimization

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \cdots, m \\
& Ax = b
\end{aligned}
$$

- f is convex, twice continuously differentiable
- $A \in R^{p \times n}$ with rank $A = p$
- assume p^* is finite and attained
- assume the problem is **strictly feasible**; hence strong duality holds

Logarithmic barrier

Reformulate the problem using indicator function:

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

where $I_-(u) = 0$ if $u \leq 0$ and $= \infty$ otherwise.

Approximation using logarithmic barrier

$$\begin{array}{ll} \text{minimize} & f_0(x) - \frac{1}{t} \sum_{i=1}^{m} \log(-f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

which is

- an equality constrained problem
- smooth approximation of the indicator function
- approximation improves as $t \to \infty$

Logarithmic barrier function

Define

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x))$$

which is

- convex
- twice continuously differentiable, with derivatives

$$\nabla\phi(x) = \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x)$$

$$\nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$

Central path

For $t > 0$, define $x^*(t)$ to be the solution of

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

Central path is: $\{x^*(t) \mid t > 0\}$

Dual points on central path

- optimality condition on $x^*(t)$: $\exists w$ such that

$$t\nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) + A^T w = 0, \quad Ax = b$$

- so, $x^*(t)$ minimizes the Lagangian of the **original problem**

$$L(x, \lambda^*(t), \nu^*(t)) = f_0(x) + \sum_{i=1}^{m} \lambda_i^*(t) f_i(x) + \nu^*(t)^T (Ax - b)$$

  with $\lambda_i^*(t) = \frac{1}{-t f_i(x^*(t))}$ and $\nu^*(t) = \frac{w}{t}$.

- $(\lambda^*(t), \nu^*(t))$ is **dual feasible**. Hence

$$\begin{aligned} p^* &\geq g(\lambda^*(t), \nu^*(t)) \\ &= L(x^*(t), \lambda^*(t), \nu^*(t)) \\ &= f_0(x^*(t)) - m/t \end{aligned}$$

  That is $f_0(x^*(t)) \to p^*$ as $t \to \infty$

Interpretation via KKT conditions

$x = x^*(t), \lambda = \lambda^*(t), \nu = \nu^*(t)$ satisfy

- primal feasible: $f_i(x) \le 0, \ \forall i = 1, \cdots, m, \quad Ax = b$
- dual feasible: $\lambda \ge 0$
- gradient of Lagrangian w.r.t. x vanishes

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T \nu = 0$$

- **approximate complementary slackness**:

$$-\lambda_i f_i(x) = 1/t, \quad \forall i = 1, \cdots, m$$

Barrier method

given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$
repeat

1. Centering step: compute $x^*(t)$ that minimizes $tf_0(x) + \phi(x)$ subject to $Ax = b$
2. Update: $x := x^*(t)$
3. Stopping criterion: quit if $m/t < \epsilon$
4. Increase t: $t := \mu t$

- stopping criterion $\implies$ $f_0(x) - p^* \leq \epsilon$
- centering step done using **Newton's method staring at current x**
- large μ means fewer outer iterations, more inner iterations; typically $\mu = 10 - 20$

Feasibility and phase I methods

Feasibility problem: find x such that

$$f_i(x) \leq 0, i = 1, \cdots, m, \quad Ax = b$$

Phase I problem: find x that is strictly feasible

Method: solve the following problem (min. over x and s)

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & f_i(x) \leq s, \quad i = 1, \cdots, m \\ & Ax = b \end{array}$$

Denote the optimal value of this problem $\bar{p}^*$:

- $\bar{p}^* < 0$: the corresponding x strictly feasible
- $\bar{p}^* > 0$: problem infeasible
- $\bar{p}^* = 0$:
  - $\bar{p}^*$ attained: feasible, but not strictly feasible
  - $\bar{p}^*$ not attained: not feasible

Complexity analysis via self-concordance

Assumptions:

- $tf_0(x) + \phi(x)$ is self-concordant with closed sublevel sets
- sublevel sets of f_0 on the feasible set are bounded

holds for LP, QP, QCQP.

Complexity analysis

Number of centering steps is exactly:

$$\frac{\log(m/(\epsilon t^{(0)})}{\log \mu}$$

Number of Newton iterations per centering is bounded above by:

$$\frac{f(x) - p^*}{\gamma} + c$$

where $\gamma = \alpha\beta(1-2\alpha)^2/(20-8\alpha)$ and constant c depends only on the tolerance ϵ_{nt}:

$$c = \log_2 \log_2(1/\epsilon_{nt})$$

Number of Newton iterations per centering step

Estimate to the number of Newton iterations for computing $x^*(\mu t)$, starting from $x^*(t)$
Denote $x = x^*(t)$, $x^+ = x^* \mu t$, $\lambda = \lambda^*(t)$, $\nu = \nu^*(t)$

$$
\begin{aligned}
&\mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+) \\
&= \mu t f_0(x) - \mu t f_0(x^+) + \sum_{i=1}^{m} \log(-\mu t \lambda_i f_i(x^+)) - m \log \mu \\
&\leq \mu t f_0(x) - \mu t f_0(x^+) - \mu \sum_{i=1}^{m} \lambda_i f_i(x^+) - m - m \log \mu \\
&\leq \mu t f_0(x) - \mu t f_0(x^+) - m - m \log \mu \\
&= m(\mu - 1 - \log \mu)
\end{aligned}
$$

The total number of Newton iterations

The total number of Newton iterations (excluding the first centering step) is upper bounded:

$$N = \frac{\log(m/(\epsilon t^{(0)})}{\log \mu} \left(\frac{m(\mu - 1 - \log \mu)}{\gamma} + c \right)$$

- tradeoff in choosing μ
- if choosing $\mu = 1 + 1/\sqrt{m}$, then $N = O\left(\sqrt{m}\ \log(m/(t^{(0)}\epsilon)\right)$
- in practice, often fix μ $(= 10, \cdots, 20)$

Generalized inequality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \cdots, m \\ & Ax = b \end{array}$$

- f_0 convex, $f_i : R^n \to R^{k_i}$ convex w.r.t. proper cone $K_i \in R^{k_i}$
- f_i twice continuously differentiable
- $A \in R^{p \times n}$ with rank $A = p$
- assume p^* is finite and attained
- assume the problem is **strictly feasible**; hence strong duality holds

Karush-Kuhn-Tucker (KKT) conditions

If strong duality holds, x is primal optimal, (λ, ν) is dual optimal, and f_i, h_i are differentiable, then the following four conditions (called **KKT conditions**) must hold

1. primal constraints: $f_i(x) \preceq_{K_i} 0, \ i = 1, \cdots, m, \quad Ax = b$
2. dual constraints: $\lambda_i \succeq_{K_i^*} 0, \quad i = 1, \cdots, m$
3. complementary slackness: $\lambda_i^T f_i(x) = 0, \quad i = 1, \cdots, m$
4. gradient of Lagrangian w.r.t. x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i^T Df_i(x) + A^T \nu = 0$$

Generalized logarithm for a proper cone

Definition (generalized logarithm)

$\psi : R^q \to R$ is a **generalized logarithm** for proper cone $K \subseteq R^q$ if

- $\mathbf{dom}\,\psi = \mathbf{int}\,K$, and $\nabla^2\psi(y) \prec 0$ for all $y \in \mathbf{int}\,K$.
- $\exists\theta > 0$ such that for all $y \succ_K 0$ and all $s > 0$

$$\psi(sy) = \psi(y) + \theta \log s$$

  That is, ψ behaves like a log along any ray in K

Properties: if $y \succ_K 0$, then

- $\nabla\psi(y) \succ_{K^*} 0$, which means that ψ is K-increasing
- $y^T\nabla\psi(y) = \theta$ (derived by taking the derivative of $\psi(sy)$ w.r.t. s)

Generalized logarithm for a proper cone: examples

Examples of ψ:

- **nonnegative orthant**: $K = R^n_+$, $\psi(y) = \sum_{i=1}^n \log y_i$ $(\theta = n)$

$$\nabla\psi(y) = (1/y_1, \cdots, 1/y_n), \quad y^T\nabla\psi(y) = n$$

- **positive semidefinite cone**: $K = S^n_+$, $\psi(Y) = \log\det Y$ $(\theta = n)$

$$\nabla\psi(Y) = Y^{-1}, \quad \mathrm{tr}(Y\nabla\psi(Y)) = n$$

- **second-order cone**: $K = \{y \in R^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \le y_{n+1}\}$

$$\psi(y) = \log(y_{n+1}^2 - y_1^2 - \cdots - y_n^2) \quad (\theta = 2)$$

$$\nabla\psi(y) = \frac{2}{y_{n+1}^2 - y_1^2 - \cdots - y_n^2}\begin{bmatrix} -y_1 \\ \cdots \\ -y_n \\ y_{n+1} \end{bmatrix}, \quad y^T\nabla\psi(y) = 2$$

Logarithmic barrier and central path

Logarithm barrier: define

$$\phi(x) = -\sum_{i=1}^{m} \psi(-f_i(x))$$

where

- ψ_i generalized log for proper cone K_i
- ϕ convex, twice continuously differentiable

Central path: $\{x^*(t) \mid t > 0\}$ where $x^*(t)$ is the solution of

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

Dual points on central path

- optimality condition on $x^*(t)$: $\exists w \in R^p$ such that

$$t\nabla f_0(x) + \sum_{i=1}^{m} Df_i(x)^T \nabla\psi_i(-f_i(x)) + A^T w = 0, \quad Ax = b$$

  where $Df_i(x) \in R^{k_i \times n}$ is the derivative matrix of f_i

- so, $x^*(t)$ minimizes the Lagangian $L(x, \lambda^*(t), \nu^*(t))$ with

$$\lambda_i^*(t) = (1/t)\nabla\psi_i(-f_i(x^*(t))), \quad \nu^*(t) = w/t$$

- properties of $\psi \implies \lambda_i^*(t) \succeq_{K_i^*} 0$ with duality gap

$$f_0(x^*(t)) - g(\lambda^*(t), \nu^*(t)) = (1/t)\sum_{i=1}^{m} \theta_i$$

  therefore $f_0(x^*(t)) \to p^*$ as $t \to \infty$

Semidefinite program (SDP)

Primal SDP

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & F(x) = x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \end{aligned}$$

where $F_i, G \in S^k$

Dual SDP

$$\begin{aligned} \max \quad & \mathbf{tr}(GZ) \\ \text{s.t.} \quad & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \cdots, n \\ & Z \succeq 0 \end{aligned}$$

where $Z \in S^k$.

Strong duality if primal SDP is strictly feasible, i.e. $\exists x$ with $x_1 F_1 + \cdots + x_n F_n + G \prec 0$

Semidefinite program (SDP): barrier method

- Logarithmic barrier: $\phi(x) = -\log\det(-F(x))$
- central path: $x^*(t)$ minimizes $tc^Tx - \log\det(-F(x))$; therefore

$$tc_i - \mathbf{tr}(F_iF(x^*(t))^{-1}) = 0, \quad i = 1, \cdots, n$$

- $Z^*(t) = -(1/t)F(x^*(t))^{-1}$ is dual feasible
- duality gap: $c^Tx^*(t) - \mathbf{tr}(GZ^*(t)) = p/t$