CS5321
Numerical Optimization

15 Fundamentals of Algorithms for Nonlinear Constrained Optimization

Outline

$$\min_{x \in \mathbf{R}^n} f(x) \text{ subject to } \begin{cases} c_i(x) = 0 & i \in \mathbf{E} \\ c_i(x) \geq 0 & i \in \mathbf{I} \end{cases}$$

- **E**, **I** are index sets for equality and inequality constraints

- Dealing with constraints: equality and inequality
- Basic strategies
  - Freedom elimination algorithms
  - Merit functions, augmented Lagrangian, and filters
    - Second order correction and non-monotone techniques

Categorizing algorithms

- Freedom elimination algorithms
  - Algorithms for quadratic programming (chap 16)
    - Basic algorithms for many other methods
  - Sequential quadratic programming method (chap 18)
    - Active set methods
  - Interior point methods, barrier methods (chap 19)
- Merit functions and filters
  - Penalty and augmented Lagrangian methods (chap 17)
  - Filters, and non monotone methods (this chap)

Equality constraints

- Ex: min $f(x_1,x_2)$ s.t. $x_1+x_2=1$
  - This equals to min $f(x_1,1-x_1)$
- Linear equality constraints: $\min_x f(x)$ s.t. $Ax=b$
  - A is $m\times n$. When $m<n$, the solution is not unique.
  - x can be expressed as $x=x_0+Zv$, where
    - x_0 is a particular solution to $Ax=b$.
    - The columns of Z spans the null space of A.
    - Vector v is of length $n-m$.
  - The problem becomes $\min_v f(x_0+Zv)$

Elimination of variables

- How to find x_0 and Z?
  1. QR decomposition of A^{T}.

     $$A^T = Q \begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1 R$$

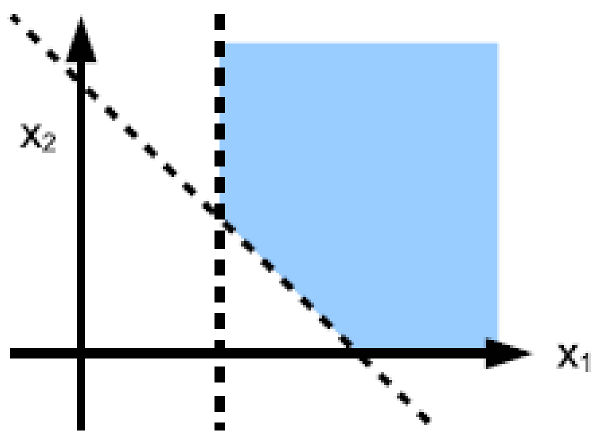
     - Q_1 is $n \times m$, Q_2 is $n \times (n-m)$, and R is $m \times m$.
     - $Z = Q_2$
  2. Solve $Ax=b$ and let the result be x_0.

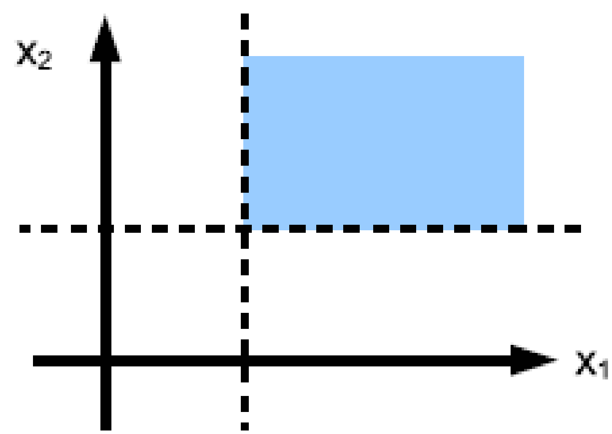
     $$b = Ax = R^T Q_1^T x, x_0 = Q_1 R^{-T} b$$

Inequality constraints

- Inequality constraints cannot be operated as equality ones
- Ex: $\begin{cases} x_1 & = 1 \\ x_1 + x_2 & = 2 \end{cases} \Leftrightarrow \begin{cases} x_1 & = 1 \\ x_2 & = 1 \end{cases}$

$$\begin{cases} x_1 & \geq 1 \\ x_1 + x_2 & \geq 2 \end{cases} \qquad \begin{cases} x_1 & \geq 1 \\ x_2 & \geq 1 \end{cases}$$





Active set

- An inequality constraint $c_i(x) \geq 0$ is *active* if $c_i(x)=0$
- *Active set* $\mathbf{A}(x) = \mathbf{E} \cup \{i \in \mathbf{I} \mid c_i(x)=0\}$
- Different active set results different solution. (example 15.1)
- Active set method: find the optimal active set
  - The *combinatorial difficulty*: search space is $2^{|\mathbf{I}|}$.
  - The simplex method for LP is an active set method.

Merit functions

- Change a constrained problem to an unconstrained one

$$\min f(x) \quad \text{s.t.} \begin{cases} c_i(x) = 0 & i \in \mathbf{E} \\ c_i(x) \geq 0 & i \in \mathbf{I} \end{cases}$$

- The $\ell 1$ penalty function

$$\phi_1(x,\mu) = f(x) + \mu \sum_{i \in \mathbf{E}} |c_i(x)| + \mu \sum_{i \in \mathbf{I}} [c_i(x)]^-$$

  - The function $[z]^- = \max\{0, -z\}$. $\mu > 0$ penalty parameter
  - It is an *exact* merit function: the optimal solution of ϕ_1 is the optimal solution of the constrained problem
- The $\ell 2$ function $\phi_2(x,\mu) = f(x) + \mu \|c(x)\|_2$

Augmented Lagrangian

- The Fletcher's augmented Lagrangian

$$\phi_F(x,\mu) = f(x) + \lambda(x)^T c(x) + \frac{1}{2}\mu \sum_{i \in \mathbf{E}} c_i(x)^2$$

  - $\lambda(x) = [A(x)A(x)^T]^{-1} A(x) \nabla f(x)$, $A(x)$ the Jacobian of $c(x)$
  - ϕ_F is exact and smooth
- The standard augmented Lagrangian

$$L_A(x,\lambda,\mu) = f(x) + \lambda^T c(x) + \frac{1}{2}\mu \|c(x)\|_2^2$$

  - Not exact

Filters

- Define $h(x) = \sum_{i \in \mathbf{E}} |c_i(x)| + \sum_{i \in \mathbf{I}} [c_i(x)]^-$
- Solves $\min_x f(x)$ and $\min_x h(x)$ simultaneously
  - Accept a new x^+ if $(f(x^+), h(x^+))$ is not *dominated* by the previous pair $(f(x), h(x))$
  - (a, b) dominates (c, d) if $a<c$ and $b<d$.
- Filter is a list of (f, h) pairs, in which no pair dominates any other.
  - Pairs with sufficient decrease are also rejected.

Maratos effect

- An example that merit function and filter fail

$$\min_{x_1, x_2} 2(x_1^2 + x_2^2 - 1) - x_1 \text{ s.t. } x_1^2 + x_2^2 - 1$$

  - The optimal solution is at (1,0). Let $x_k = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$

$$f(x_k) = -\cos\theta, h(x_k) = 0$$

  - For $p_k = \begin{pmatrix} \sin^2\theta \\ -\sin\theta\cos\theta \end{pmatrix}$, $x_{k+1} = x_k + p_k$
  - x_{k+1} yields quadratic convergence, but increases f and h.

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = \frac{2\sin^2(\theta/2)}{|2\sin(\theta/2)|^2} = \frac{1}{2} \qquad \begin{aligned} f(x_{k+1}) &= \sin^2\theta - \cos\theta \\ h(x_{k+1}) &= \sin^2\theta \end{aligned}$$

Second order correction

- Let $A_k = A(x_k)$ be the Jacobian of constraints $c(x_k)$.
  - Suppose A_k, $m \times n$, $m < n$, has full row rank.
- The linear approximation to $c(x)$ at $x = x_k + p_k$ is $A_k p^* + c(x_k + p_k)$.
  - One solution for $A_k p^* + c(x_k + p_k) = 0$ is

$$p^* = -A_k^T (A_k A_k^T)^{-1} c(x_k + p_k)$$

- If $x_{k+1} = x_k + p_k + p^*$, $\|c(x)\|$ can be further decreased.
  - With $A_k p_k + c(x_k) = 0$ and proper step length.

Non-monotone techniques

- To resolve the Maratos effect, try steps p_k that increase f and h.
- *Watchdog strategy*: the merit function is allowed to increase on t iterations.
  - Typically, t=5~8
  - If after t iterations, the merit function does not decrease sufficiently, rollback