CS5321
Numerical Optimization

17 Penalty and Augmented Lagrangian Methods

Outline

- Both active set methods and interior point methods require a feasible initial point.
- Penalty methods need not a feasible initial point.
  1. Quadratic penalty method
  2. Nonsmooth exact penalty method
  3. Augmented Lagrangian methods

1. Quadratic penalty function

- For $\min_x f(x)$ s.t. $c_i(x) = 0, i \in \mathbf{E}$

  the quadratic penalty function is

$$Q(x, \mu) = f(x) + \frac{\mu}{2} \sum_{i \in \mathbf{E}} c_i^2(x)$$

  - μ is the penalty parameter

- For $\min_x f(x)$ s.t. $c_i(x) = 0, i \in \mathbf{E}, c_i(x) \geq 0, i \in \mathbf{I}$

  the quadratic penalty function is

$$Q(x, \mu) = f(x) + \frac{\mu}{2} \sum_{i \in \mathbf{E}} c_i^2(x) + \frac{\mu}{2} \sum_{i \in \mathbf{I}} ([c_i^2(x)]^-)^2$$

Quadratic penalty method

1. Given μ_0, $\{\tau_k | \tau_k \to 0, \tau_k > 0\}$, starting point x_0
2. For $k = 0,1,2,\ldots$
   a) Find a solution x_k of Q (:, μ_k) s.t. $\|\nabla_x Q(:, \mu_k)\| \le \tau_k$.
   b) If converged, stop
   c) Choose $\mu_{k+1} > \mu_k$ and another starting x_k.

- Theorem 17.1: If x_k is a global exact solution to step 2(a), and $\mu_k \to \infty$, x_k converges to the global solution x^* of the original problem.

The Hessian matrix

- Let $A(x)^{\mathrm{T}}=[\nabla c_i(x)]_{i\in\mathbf{E}}$. The Hessian of Q is

$$\nabla^2_{xx}Q(x,\mu_k) = \nabla^2 f(x) + \sum_{i\in\mathbf{E}} \mu_k c_i(x)\nabla^2 c_i(x) + \mu_k A(x)^T A(x)$$

  - Step 2(a) needs to solve $\nabla^2_{xx}Q(x,\mu_k)p = -\nabla_x Q(x,\mu)$
  - $A^T A$ only has rank m ($m<n$). As μ_k increases, the system becomes ill-conditioned
- Solve a larger system with a better condition

$$\begin{bmatrix} \nabla^2 f + \sum_{i\in\mathbf{E}} \mu_k c_i \nabla^2 c_i & A(x)^T \\ A(x) & -(1/u_k)I \end{bmatrix} \begin{bmatrix} p \\ \zeta \end{bmatrix} = \begin{bmatrix} -\nabla_x Q \\ 0 \end{bmatrix}$$

2. Nonsmooth Penalty function

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

- $[y]^- = \max\{0, -y\}$, which is not differentiable.

- But the functions inside are differentiable.
  - Approximate it by linear functions

$$\begin{aligned} q(p,\mu) &= f(x) + \nabla f(x)^T p + \frac{1}{2} p^T W p + \\ &\quad \mu \sum_{i\in\mathbf{E}} |c_i(x) + \nabla c_i(x)^T p| + \\ &\quad \mu \sum_{i\in\mathbf{I}} [c_i(x) + \nabla c_i(x)^T p]^- \end{aligned}$$

Smoothed object 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 object function can be rewritten as

$$\begin{aligned}
\min_{p,r,s,t} \quad & \nabla f(x)^T p + \frac{1}{2} p^T W p + \mu \sum_{i\in\mathbf{E}} (r_i + s_i) + \mu \sum_{i\in\mathbf{I}} t_i \\
\text{s.t.} \quad & \nabla c_i(x)^T p + c_i(x) = r_i + s_i \qquad i \in \mathbf{E} \\
& \nabla c_i(x)^T p + c_i(x) \geq -t_i \qquad i \in \mathbf{I} \\
& r, s, t \geq 0
\end{aligned}$$

3. Augmented Lagrangian

- For $\min_x f(x)$ s.t. $c_i(x) = 0, i \in \mathbf{E}$, define the augmented Lagrangian function

$$L_A(x, \lambda, \mu) = f(x) - \sum_{i \in \mathbf{E}} \lambda_i c_i(x) + \frac{\mu}{2} \sum_{i \in \mathbf{E}} c_i^2(x)$$

- Theorem 17.5: If x^* is a solution of the original problem, and $\nabla c_i(x)$ are linearly independent, and the second order optimality conditions are satisfied, then there is a μ^*, such that for all $\mu \geq \mu^*$, x^* is a local minimizer of $L_A(x,\lambda,\mu)$

Lagrangian multiplier

- The gradient of $L_A(x,\lambda,\mu)$ is

$$\nabla L_A(x_k, \lambda_k, \mu_k) = \nabla f(x_k) - \sum_{i \in \mathbf{E}} [(\lambda_i)_k - \mu_k c_i(x)] \nabla c_i(x_k)$$

  - In the Lagrangian function property, the Lagrangian multiplier $\lambda_i^* = (\lambda_i)_k - \mu_k c_i(x)$
  - Thus, $c_i(x) = -[\lambda_i^* - (\lambda_i)_k]/\mu_k$.
  - To satisfy $c_i(x)=0$, either $\mu \to \infty$ or $(\lambda_i)_k \to \lambda_i^*$
    - Previous penalty methods only use $\mu \to \infty$ to make $c_i(x)=0$
  - Parameter λ can be updated as

$$(\lambda_i)_{k+1} = (\lambda_i)_k - \mu_k c_i(x_k)$$

Inequality constrains

- For inequality constraints, add slack variables

$$c_i(x) - s_i = 0, s_i \geq 0, \text{ for all } i \in \mathbf{I}$$

- Bounded constrained Lagrangian (BCL)

$$L_A(x, \lambda, \mu) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x) + \frac{\mu}{2} \sum_{i=1}^{m} c_i^2(x)$$

$$\min_x L_A(x, \lambda, \mu) \text{ s.t. } l \leq x \leq u$$

  - How to solve this will be discussed in chap 18. (gradient projection method)