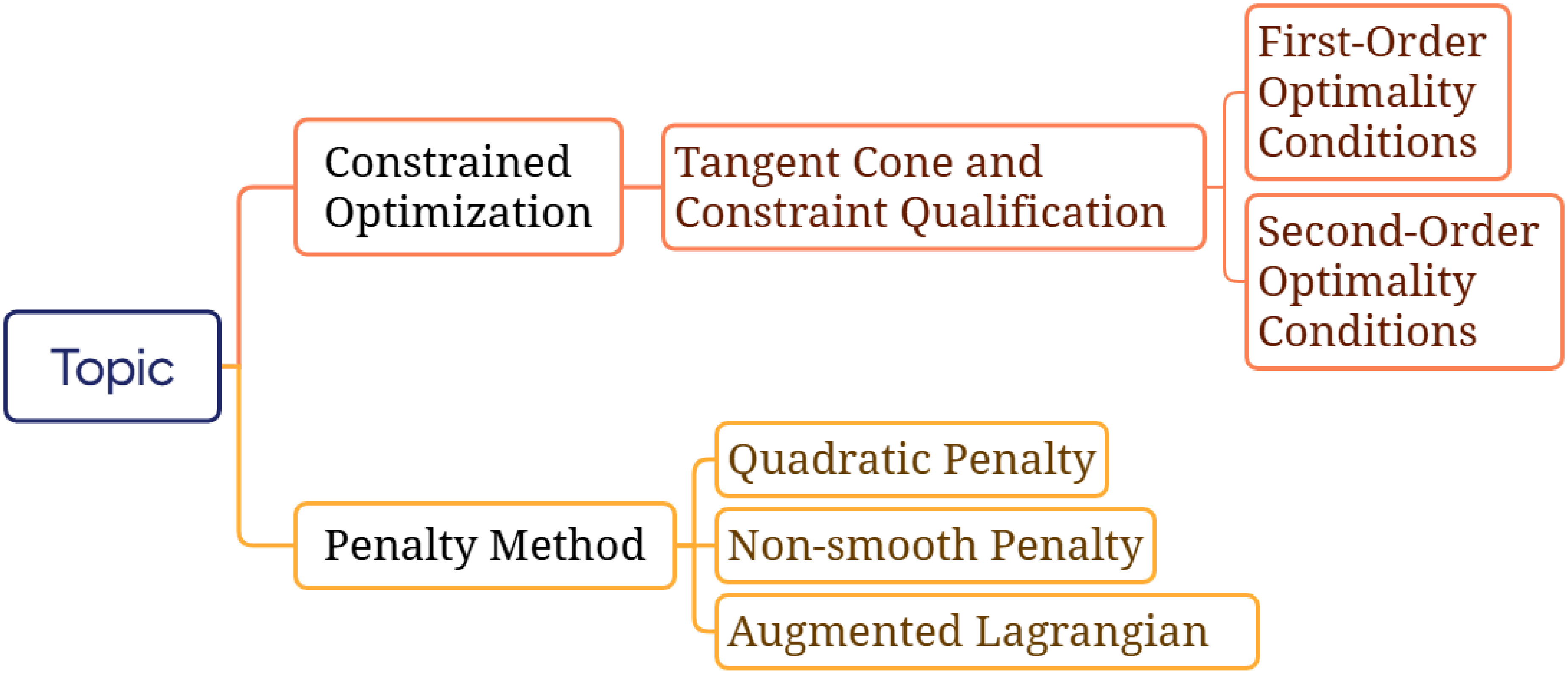
Constrained Optimization and Penalty Method

Abstract: In this topic, I will provide a brief introduction to constrained optimization and the penalty method, which transforms the original constrained problem into a single unconstrained problem. I will also discuss the linearized approximation of the feasible set, which helps derive the constraint conditions necessary for analyzing practical optimization methods. Additionally, unlike the quadratic penalty method, the penalty parameters in the augmented Lagrangian and exact penalty methods do not need to be infinite. However, the exact penalty method is non-smooth, which can be addressed by introducing artificial variables. Meanwhile, the augmented Lagrangian method also introduces additional slack variables when applied to constrained problems.

Outline


Topic
Constrained Optimization
Tangent Cone and Constraint Qualification
First-Order Optimality Conditions
Second-Order Optimality Conditions
Penalty Method
Quadratic Penalty
Non-smooth Penalty
Augmented Lagrangian

Motivation

- The optimal conditions of constrained optimization are more complex than unconstrained optimization.
- Constraint conditions provide a way to update the parameters and analysis the convergence of the optimization method.
- Penalty method can transform the original constrained problem into a single unconstrained problem.

Fundamental Conceptions

- Constrained Programming

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \boxed{\begin{array}{l} c_i(\mathbf{x}) = g_i(\mathbf{x}) \geq 0, i = 1, \cdots, m \\ c_i(\mathbf{x}) = h_j(\mathbf{x}) = 0, j = 1, \cdots, l \end{array}} \end{array} \tag{1}$$

Feasible Region S

- Active set: $\mathcal{A}$

$$\mathcal{A}(\mathbf{x}_0) = \{i | c_i(\mathbf{x}_0) = 0\} \tag{2}$$

- Feasible Direction $\mathbf{d}$:

$$F(\mathbf{x}_0, S) = \{\mathbf{d} | \mathbf{x}_0 + \lambda \mathbf{d} \in S, \exists \delta > 0, \forall \lambda \in (0, \delta)\} \tag{3}$$

Tangent Cone

Descent direction:

$$D(\mathbf{x}_0, f) = \{\mathbf{d} | \nabla f(\mathbf{x}_0)^T \mathbf{d} < 0\} \tag{4}$$

Local optimal solution $\bar{\mathbf{x}}$:

$$F(\bar{\mathbf{x}}, S) \cap D(\bar{\mathbf{x}}, f) = \emptyset \tag{5}$$

Tangent Cone $T(\mathbf{x}_0 | S)$:

$$T(\mathbf{x}_0 | S) = \{\mathbf{d} | \exists \tau_i \to 0, \{\mathbf{x}_i\} \subset S, \mathbf{x}_i \to \mathbf{x}_0, \text{s.t.} \mathbf{d}_i = \frac{\mathbf{x}_i - \mathbf{x}_0}{\tau_i} \to \mathbf{d}\} \tag{6}$$

Linear Approximation

Tangent Cone

Equality constraint: $\mathcal{E} = \{h_i(\mathbf{x}_0) = 0, i = 1, \dots, l\}$

$$\nabla h_i(\mathbf{x}_0)^T \mathbf{d} = \lim_{k\to\infty} \frac{h_i(\mathbf{x}_0 + \tau_k \mathbf{d}_k) - h_i(\mathbf{x}_0)}{\tau_k} = 0 \tag{7}$$

Linearized feasible directions L_h:

$$L_h := L(\mathbf{x}_0, h) = \{\mathbf{d} | \nabla h_j(\mathbf{x}_0)^T \mathbf{d} = 0, j = 1, \dots, l\} \tag{8}$$

Active set of Inequality constraint: $\mathcal{A}(\mathbf{x}_0) = \{i \in \{1, \dots, m\} | g_i(\mathbf{x}_0) = 0\}$

$$\nabla g_i(\mathbf{x}_0)^T \mathbf{d} = \lim_{k\to\infty} \frac{g_i(\mathbf{x}_0 + \tau_k \mathbf{d}_k) - g_i(\mathbf{x}_0)}{\tau_k} \geq 0, i \in \mathcal{A}(\mathbf{x}_0) \tag{9}$$

Linearized feasible directions L_g:

$$L_g := L(\mathbf{x}_0, g) = \{\mathbf{d} | \nabla g_i(\mathbf{x}_0)^T \mathbf{d} \geq 0, i \in \mathcal{A}(\mathbf{x}_0)\} \tag{10}$$

Constraint Qualification

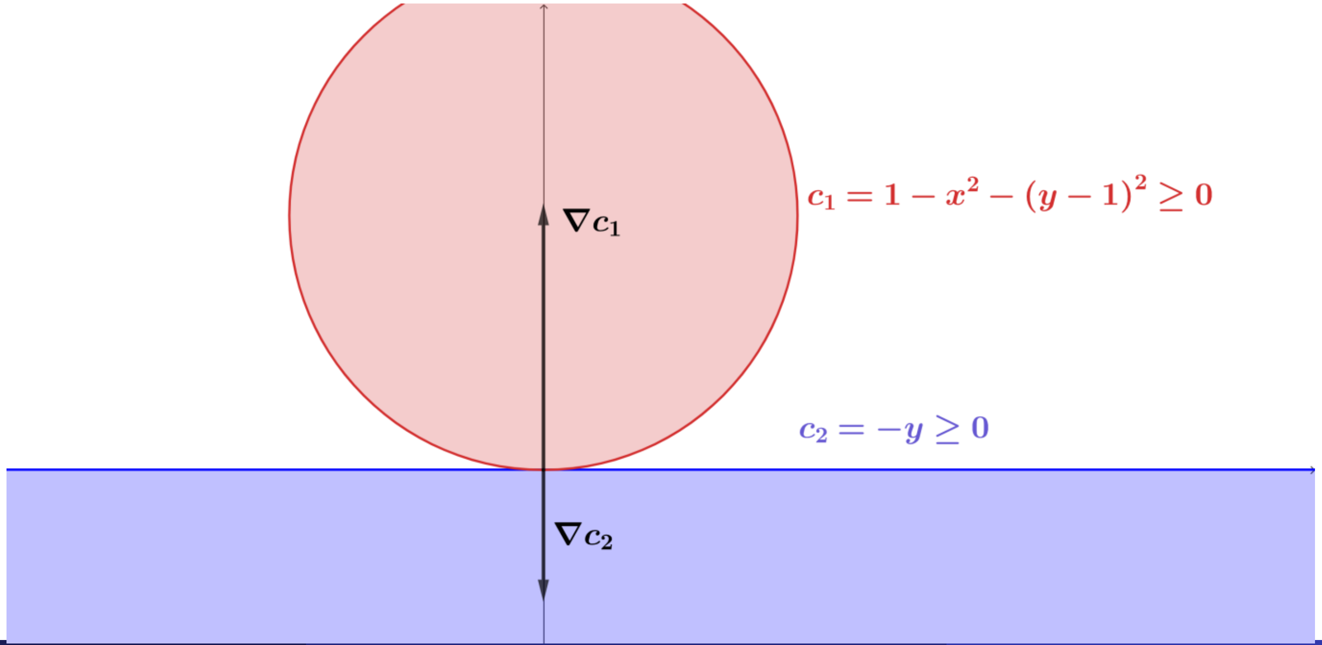
To ensure the equivalence of $T(\mathbf{x}_0|S)$ and $\mathrm{L}(\mathbf{x}_0, S)$

$\nabla h_i(\mathbf{x}_0), i = 1, \ldots, l$ & $\nabla g_j(\mathbf{x}_0), j \in \mathcal{A}(\mathbf{x}_0)$
are linearly independent(**LICQ**).

If $\mathbf{x}_0$ satisies LICQ, it means

$$L(\mathbf{x}_0, h) \cap L(\mathbf{x}_0, g) = T(\mathbf{x}_0|g) \cap T(\mathbf{x}_0|h) \tag{11}$$

e.g.



First-Order Optimality Conditions

If $\bar{\mathbf{x}}$ is **local optimal solution** and satisfies **LICQ**

First-Order Necessary Conditions(KKT)

$$\begin{cases} \text{Stationarity} & \nabla_{\mathrm{x}}\mathcal{L}(\overline{\mathbf{x}}, \lambda, \mu) = 0 \\ \text{Primal feasibility} & g_i(\overline{\mathbf{x}}) \geq 0, i = 1, \cdots, m; h_j(\overline{\mathbf{x}}) = 0, j = 1, \cdots, l. \\ \text{Complementary slackness} & \lambda_i g_i(\overline{\mathbf{x}}) = 0, i = 1, \cdots, m \\ \text{Dual feasibility} & \lambda_i \geq 0, i = 1, \cdots, m \end{cases} \tag{12}$$

where $$\mathcal{L}(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{l} \mu_j h_j(\mathbf{x})$$

To first-order approximation:
- Increase $d^T \nabla f(x^*) > 0$
- Keep the value $d^T \nabla f(x^*) = 0$

First-Order Optimality Conditions

First-Order Sufficient Conditions

For Convex Optimization:

$$
\begin{aligned}
\min \quad & f(\mathbf{x}) \\
\text{s.t.} \quad & c_i(\mathbf{x}) \leq 0, i = 1, \cdots, m \\
& A\mathbf{x} = \mathbf{b}.
\end{aligned} \tag{13}
$$

and exists *relative interior point* $\mathbf{x}_0$:

$$
\begin{aligned}
\text{s.t.} \quad & c_i(\mathbf{x_0}) < 0, i = 1, 2, \ldots, m; \\
& A\mathbf{x}_0 = \mathbf{b}
\end{aligned} \tag{14}
$$

KKT $\Leftrightarrow$ Local Optimal Solution

Second-Order Optimality Conditions

Critical Cone $C(\mathbf{x}^*, \lambda^*, \mu^*)$:

$$C(\mathbf{x}^*, \lambda^*, \mu^*) = \{\mathbf{d} \in L(\mathbf{x}^*, S) | \nabla g_i(\mathbf{x}^*)^T \mathbf{d} = 0, \forall i \in \mathcal{A}(\mathbf{x}^*), \lambda_i^* > 0\} \tag{15}$$

$$\mathbf{d}^T \nabla f(\mathbf{x}^*) = \sum_{i \in \mathcal{A}} \lambda_i^* \nabla g_i(\mathbf{x}^*)^T \mathbf{d} - \sum_{j=1}^{l} \mu_j^* \nabla h(\mathbf{x}^*)^T \mathbf{d} = 0 \tag{16}$$

The linearized feasible direction in this case cannot be determined as a descending or ascending direction based on the first-order optimality conditions.

Keep the active inequality and equality constraints invariant when we were to make small changes to the objective function.

We need second-order optimality conditions!

Second-Order Optimality Conditions

- Second-Order Necessary Conditions

LICQ and KKT conditions hold,

$$\mathbf{d}^T \nabla^2_{xx} \mathcal{L}(\mathbf{x}^*, \lambda^*, \mu^*)\mathbf{d} \geq 0, \quad \forall \mathbf{d} \in C(\mathbf{x}^*, \lambda^*, \mu^*). \tag{17}$$

- Second-Order Sufficient Conditions

KKT conditions hold and

$$\mathbf{d}^T \nabla^2_{xx} \mathcal{L}(\mathbf{x}^*, \lambda^*, \mu^*)\mathbf{d} > 0, \quad \forall \mathbf{d} \neq 0, \mathbf{d} \in C(\mathbf{x}^*, \lambda^*, \mu^*). \tag{18}$$

Penalty method

Constrained programming:

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & c_i(\mathbf{x}) = 0, i \in \mathcal{E} \\ & c_j(\mathbf{x}) \leq 0, j \in \mathcal{I}. \end{array} \tag{19}$$

Dilemmas:

- Global optimal solution $\mathbf{x}^*$: $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$
- $\mathbf{x_1} = \mathbf{x_0} + \Delta\mathbf{x} \notin S$

Penalty mothed:

original constrained problem ⇒ single unconstrained problem

Quadratic Penalty

- Case 1: equality-constrained problem

Quadratic penalty function

$$P_E(\mathbf{x}, \sigma) = f(\mathbf{x}) + \frac{1}{2}\sigma \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n \tag{20}$$

where $\sigma > 0$ is penalty parameter.

Algorithm:

1: Given $\sigma_1 > 0, \mathbf{x}_0, k = 1$, growth factor $\rho > 1$
2: while violate the convergence test do
3: $\quad \mathbf{x}^k = \underset{\mathbf{x}}{\operatorname{argmin}}\, P_E(\mathbf{x}, \sigma_k)$
4: $\quad$ choose $\sigma^{k+1} = \rho\sigma^k$.
5: $\quad k \rightarrow k + 1$
6: end.

Quadratic Penalty

KKT:

- Primal problem:

$$\nabla f(\mathbf{x}^*) + \sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i(\mathbf{x}^*) = 0$$

$$c_i(\mathbf{x}^*) = 0, \forall i \in \mathcal{E} \tag{21}$$

- Penalty problem:

$$\nabla f(\mathbf{x}) + \sum_{i \in \mathcal{E}} \sigma c_i(\mathbf{x}) \nabla c_i(\mathbf{x}) = 0 \tag{22}$$

We have:

$$\sigma c_i(\mathbf{x}) \approx \lambda_i^*, \forall i \in \mathcal{E} \tag{23}$$

means

$$c_i(\mathbf{x}) = 0 \Leftarrow \sigma \to \infty \tag{24}$$

Quadratic Penalty

- Case 2: inequality-constrained problem

$$P_I(\mathbf{x}, \sigma) = f(\mathbf{x}) + \frac{1}{2}\sigma \sum_{i \in \mathcal{I}} ([c_i(\mathbf{x})]^-)^2, \forall \mathbf{x} \in \mathbb{R}^n \tag{25}$$

first-order derivable respect to c_i

where

$$[c_i(\mathbf{x})]^- = \max\{c_i(\mathbf{x}), 0\} \tag{26}$$

Exact Penalty Methods

ℓ_1 penalty function

$$P(\mathbf{x}, \sigma) = f(\mathbf{x}) + \sigma \left[\sum_{i \in \mathcal{E}} |c_i(\mathbf{x})| + \sum_{i \in \mathcal{I}} [c_i(\mathbf{x})]^- \right] \tag{27}$$

Exactness:

If $\mathbf{x}^*$ is a strict local solution of the primal programming problem at which the KKT conditions are satisfied with Lagrange multipliers $\lambda_i^*, i \in \mathcal{E} \cup I$. Then $\mathbf{x}^*$ is a local minimizer of $P(\mathbf{x}, \sigma)$, where

$$\sigma^* = \|\lambda^*\|_\infty = \max_i |\lambda_i^*|$$

Exact Penalty Methods

More practical: linearizing this model as a subproblem

$$q(p;\mu) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T W p + \mu \left[\sum_{i \in \mathcal{E}} |c_i(x) + \nabla c_i(x)^T p| + \sum_{i \in \mathcal{I}} [c_i(x) + \nabla c_i(x)^T p]^- \right] \tag{28}$$

Smooth quadratic programming problem

$$\begin{aligned} \min_{p,r,s,t} \quad & f(x) + \tfrac{1}{2} p^T W p + \nabla f(x)^T p + \mu \textstyle\sum_{i \in \mathcal{E}} (r_i + s_i) + \mu \sum_{i \in \mathcal{I}} t_i \\ \text{s.t.} \quad & \nabla c_i(x)^T p + c_i(x) = r_i - s_i, \quad i \in \mathcal{E} \\ & \nabla c_i(x)^T p + c_i(x) \le t_i, \quad i \in \mathcal{I} \\ & r, s, t \ge 0. \end{aligned} \tag{29}$$

Augmented Lagrangian Method

- Case 1: equality-constrained problem

$$L_\sigma(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i \in \mathcal{E}} \lambda_i c_i(\mathbf{x}) + \frac{\sigma}{2} \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x}) \tag{30}$$

Given σ_k, λ^k, the minimal point x^{k+1} s.t.:

$$\nabla_x L_{\sigma_k}\left(\mathbf{x}^{k+1}, \lambda^k\right) = \nabla f\left(\mathbf{x}^{k+1}\right) + \sum_{i \in \mathcal{E}} \left(\lambda_i^k + \sigma_k c_i\left(\mathbf{x}^{k+1}\right)\right) \nabla c_i\left(\mathbf{x}^{k+1}\right) = 0 \tag{31}$$

Comparing to eq(21), obtain

$$\lambda_i^{k+1} \sim \lambda_i^* \approx \lambda_i^k + \sigma_k c_i(\mathbf{x}^{k+1}), \forall i \in \mathcal{E} \tag{32}$$

$\sigma_k \to \infty$ is not necessary.

$$c_i(\mathbf{x}^{k+1}) = \frac{\lambda_i^* - \lambda_i^k}{\sigma_k} \tag{33}$$

Augmented Lagrangian Method

Algorithm:

1: Given $x^0, \lambda^0, \sigma_0 > 0, \varepsilon > 0, \eta > 0, k = 0$

2: for $k = 0,1,\dots$ do

3: $\quad \mathbf{x}^{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}} L_{\sigma_k}(\mathbf{x}, \lambda^k)$, s.t. $\|\nabla_x L_{\sigma_k}(\mathbf{x}^{k+1}, \lambda^k)\| \le \eta_k$

4: $\quad$ if $\|c(x^{k+1})\| \le \varepsilon \,\&\, \eta_k \le \eta$ then

5: $\qquad$ return (x^{k+1}, λ_k)

6: $\qquad$ end do

7: $\quad$ end if

8: $\quad \lambda^{k+1} = \lambda^k + \sigma_k c(\mathbf{x}^{k+1}), \sigma_{k+1} = \rho\sigma_k$, decrease η_k .

9: end for

Augmented Lagrangian Method

- Case 2: general constrained problem

Introducing slack variables s_i:

$$\begin{aligned} \min_{\mathbf{x},s} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & c_i(\mathbf{x}) = 0, i \in \mathcal{E} \\ & c_j(\mathbf{x}) + s_j = 0, j \in \mathcal{I} \\ & s_i \geq 0, i \in \mathcal{I}. \end{aligned} \tag{34}$$

Preserving non-negative constraints

$$\begin{aligned} L_\sigma(\mathbf{x}, s, \lambda, \mu) = & f(\mathbf{x}) + \sum_{i\in\mathcal{E}} \lambda_i c_i(\mathbf{x}) + \sum_{i\in\mathcal{I}} \mu_i \left(c_i(x) + s_i\right) + \frac{\sigma}{2} p(\mathbf{x}, s), \\ & s_i \geqslant 0, i \in \mathcal{I}. \end{aligned} \tag{35}$$

Where $p(\mathbf{x}, s)$ is quadratic penalty function(constraint violating measurement)

$$p(\mathbf{x}, s) = \sum_{i\in\mathcal{E}} c_i^2(\mathbf{x}) + \sum_{i\in\mathcal{I}} (c_i(\mathbf{x}) + s_i)^2$$

$$v_k(\mathbf{x}^{k+1}) = \sqrt{\sum_{i\in\mathcal{E}} c_i^2(\mathbf{x}^{k+1}) + \sum_{i\in\mathcal{I}} (c_i(\mathbf{x}) + s_i^{k+1})^2} \tag{36}$$

Augmented Lagrangian Method

In step-k, given λ^k, μ^k and σ^k, to get x^{k+1}, s^{k+1}.

$$\min_{\mathbf{x}, s} \quad L_{\sigma_k}\left(\mathbf{x}, s, \lambda^k, \mu^k\right), \quad \text{s.t.} \quad s \geq 0 \tag{37}$$

Fix $\mathbf{x}$, subproblem about s:

$$\min_{s \geq 0} \quad \sum_{i \in \mathcal{I}} \mu_i (c_i(x) + s_i) + \frac{\sigma_k}{2} \sum_{i \in \mathcal{I}} (c_i(x) + s_i)^2 \tag{38}$$

Solution: $$s_i = \max\left\{-\frac{\mu_i}{\sigma_k} - c_i(x), 0\right\}$$

Simplify:

$$L_{\sigma_k}\left(x, \lambda^k, \mu^k\right) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{\sigma_k}{2} \sum_{i \in \mathcal{E}} c_i^2(x) +$$

derivable respect to $\mathbf{x}$

$$\frac{\sigma_k}{2} \sum_{i \in \mathcal{I}} \left(\max\left\{ \frac{\mu_i}{\sigma_k} + c_i(x), 0 \right\}^2 - \frac{\mu_i^2}{\sigma_k^2} \right) \tag{39}$$

Equivalently

$$\min_{\mathbf{x}} \quad L_{\sigma_k}\left(\mathbf{x}, s, \lambda^k, \mu^k\right) \tag{40}$$

Update Multipliers

KKT conditions:

$$0=\nabla f\left(x^*\right)+\sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i\left(x^*\right)+\sum_{i \in \mathcal{I}} \mu_i^* \nabla c_i\left(x^*\right), \quad \mu_i^* \geqslant 0, s_i^* \geqslant 0, i \in \mathcal{I} \quad (41)$$

For $\mathbf{x}^{k+1}, s^{k+1}$

$$\begin{aligned} & 0=\nabla f\left(x^{k+1}\right)+\textstyle\sum_{i \in \mathcal{E}}\left(\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right)\right) \nabla c_i\left(x^{k+1}\right)+ \\ & \textstyle\sum_{i \in \mathcal{I}}\left(\mu_i^k+\sigma_k\left(c_i\left(x^{k+1}\right)+s_i^{k+1}\right)\right) \nabla c_i\left(x^{k+1}\right) \\ & s_i^{k+1}=\max \left\{-\frac{\mu_i^k}{\sigma_k}-c_i\left(x^{k+1}\right), 0\right\}, \quad i \in \mathcal{I}. \end{aligned} \quad (42)$$

Obviously,

$$\begin{aligned} \lambda_i^{k+1} & =\lambda_i^k+\sigma_k c_i(\mathbf{x}^{k+1}), \quad i \in \mathcal{E}, \\ \mu_i^{k+1} & =\max\{\mu_i^k+\sigma_k c_i(\mathbf{x}^{k+1}), 0\}, \quad i \in \mathcal{I}. \end{aligned} \quad (43)$$

Augmented Lagrangian Method

Algorithm:

1: Given $x^0, \lambda^0, \sigma^0, \varepsilon > 0, \eta > 0, 0 < \alpha \le \beta \le 1$ and $\rho > 1$. let $\eta^0 = \frac{1}{\sigma^0}, \varepsilon_0 = 1/\sigma_0^\alpha$.

2: for $k = 0,1,2,\ldots$ do

3: $x^{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}} L_{\sigma_k}(\mathbf{x}, \lambda^k, \mu^k)$ s.t. $\left\|\nabla_x L_{\sigma_k}(\mathbf{x}^{k+1}, \lambda^k, \mu^k)\right\|_2 \le \eta_k$

4: if $v_k(x^{k+1}) \le \varepsilon_k$ then

5: if $v_k(x^{k+1}) \le \varepsilon$ & $\left\|\nabla_x L_{\sigma_k}(\mathbf{x}^{k+1}, \lambda^k, \mu^k)\right\|_2 \le \eta$

6: return $(x^{k+1}, \lambda^k, \mu^k)$ end do

7: end if

8: $\lambda_i^{k+1} = \lambda_i^k + \sigma_k c_i(x^{k+1})$; $\mu_j^{k+1} = \max\{\mu_j^k + \sigma_k(x^{k+1}), 0\}$,$\sigma_{k+1} = \sigma_k$

9: $\eta_{k+1} = \frac{\eta_k}{\sigma_{k+1}}, \varepsilon_{k+1} = \frac{\varepsilon_k}{\sigma_{k+1}^\beta}$.

10: else $\lambda^{k+1} = \lambda^k, \sigma_{k+1} = \rho\sigma_k, \eta_{k+1} = \frac{1}{\sigma_{k+1}}, \varepsilon_{k+1} = \frac{1}{\sigma_{k+1}^\alpha}$.

11: end if

12: end for

Conclusion

- Introduce the constraint qualification and give the optimality conditions of constrained optimization
- the penalty methods which convert the constraint problem to constraint problem and analysis the convergence using optimality conditions.
- Talk about a more practical penalty method: the augmented Lagrangian method for equality and inequality constrained case.
- Optimality conditions will give us the update step of parameters. For inequality case, we simplify this problem with slack variables.

Thank you and any questions