Minimization Problem with Smooth Components

Yu. Nesterov
Presenter: Lei Tang

Department of CSE
Arizona State University

Dec. 7th, 2008

Outline

- MiniMax problem
- Gradient Mapping for MiniMax problem ;
- The complexity of gradient and optimal method;
- Optimization with functional constraint (General constrained optimization problem)
- Constrained Minimization Problem

MiniMax Problem

- Objective function is composed with several components.
- The simplest problem of that type is *minimax* problem.
- We'll focus on smooth minimax problem:

$$\min_{x \in Q} \left[f(x) = \max_{1 \leq i \leq m} f_i(x) \right]$$

where $f_i \in \mathcal{S}^{1,1}_{\mu,L}(R^n), \quad i = 1, \cdots, m$ and Q is a closed convex set.

- $f(x)$: the max-type function composed by the components $f_i(x)$.

- In general, $f(x)$ is not differentiable.
- We use $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$ to denote all the $f_i \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$.

MiniMax Problem

- Objective function is composed with several components.
- The simplest problem of that type is *minimax* problem.
- We'll focus on smooth minimax problem:

$$\min_{x \in Q} \left[f(x) = \max_{1 \le i \le m} f_i(x) \right]$$

where $f_i \in \mathcal{S}_{\mu,L}^{1,1}(R^n), \quad i = 1, \cdots, m$ and Q is a closed convex set.

- $f(x)$: the max-type function composed by the components $f_i(x)$.

- In general, $f(x)$ is not differentiable.
- We use $f \in \mathcal{S}_{\mu,L}^{1,1}(R^n)$ to denote all the $f_i \in \mathcal{S}_{\mu,L}^{1,1}(R^n)$.

Connection with General Minimization Problem

General Minimization Problem

$$\begin{aligned} \min \quad & f_0(x) & (1)\\ s.t. \quad & f_i(x) \leq 0, \quad i = 1, \cdots, m & (2)\\ & x \in Q & (3) \end{aligned}$$

parametric max-type function

$$f(t; x) = \max\{f_0(x) - t; f_i(x)\}$$

Will be showed later:

- the optimal value of $f_0(x)$ corresponds to the root t of $f(t; x) = 0$;
- minimax problem is used as a subroutine to solve (1);

Linear approximation

Linearization

max-type function $f(x) = \max_{1 \le i \le m} f_i(x)$

linearization of $f(x)$ $f(\bar{x}; x) = \max_{1 \le i \le m} [f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x} \rangle]$

Essentially, linearization over each component.

Properties

- $f(\bar{x}; x) + \frac{\mu}{2}||x - \bar{x}||^2 \le f(x) \le f(\bar{x}; x) + \frac{L}{2}||x - \bar{x}||^2$;
- $x^* \in Q \Leftrightarrow f(x^*; x) \ge f(x^*; x^*) = f(x^*)$.
- $f(x) \ge f(x^*) + \frac{\mu}{2}||x - x^*||^2$
- the solution x^* exists and unique.

Lemma 2.3.1

$f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2 \leq f(x) \leq f(\bar{x};x) + \frac{L}{2}||x-\bar{x}||^2$

- $f_i \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$
- For strongly convex function, we have

$$\begin{aligned} f_i(x) &\geq f_i(\bar{x}) + \langle f_i'(\bar{x}, x-\bar{x})\rangle + \frac{\mu}{2}||x-\bar{x}||^2 \\ &= f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2 \end{aligned}$$

Take the max on both sides: $f(x) \geq f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2$

- For Lipshitz continuous function, it follows

$$\begin{aligned} f_i(x) &\leq f_i(\bar{x}) + \langle f_i'(\bar{x}, x-\bar{x})\rangle + \frac{L}{2}||x-\bar{x}||^2 \\ &= f(\bar{x};x) + \frac{L}{2}||x-\bar{x}||^2 \end{aligned}$$

- max operation keeps the property as smooth strongly convex function.

Lemma 2.3.1

$f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2 \le f(x) \le f(\bar{x};x) + \frac{L}{2}||x-\bar{x}||^2$

- $f_i \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$
- For strongly convex function, we have

$$\begin{aligned} f_i(x) &\ge f_i(\bar{x}) + \langle f_i'(\bar{x}, x-\bar{x})\rangle + \frac{\mu}{2}||x-\bar{x}||^2 \\ &= f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2 \end{aligned}$$

  Take the max on both sides: $f(x) \ge f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2$

- For Lipshitz continuous function, it follows

$$\begin{aligned} f_i(x) &\le f_i(\bar{x}) + \langle f_i'(\bar{x}, x-\bar{x})\rangle + \frac{L}{2}||x-\bar{x}||^2 \\ &= f(\bar{x};x) + \frac{L}{2}||x-\bar{x}||^2 \end{aligned}$$

- max operation keeps the property as smooth strongly convex function.

Lemma 2.3.1

$f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2 \leq f(x) \leq f(\bar{x};x) + \frac{L}{2}||x-\bar{x}||^2$

- $f_i \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$
- For strongly convex function, we have

$$\begin{aligned} f_i(x) &\geq f_i(\bar{x}) + \langle f_i'(\bar{x}, x-\bar{x})\rangle + \frac{\mu}{2}||x-\bar{x}||^2 \\ &= f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2 \end{aligned}$$

  Take the max on both sides: $f(x) \geq f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2$

- For Lipshitz continuous function, it follows

$$\begin{aligned} f_i(x) &\leq f_i(\bar{x}) + \langle f_i'(\bar{x}, x-\bar{x})\rangle + \frac{L}{2}||x-\bar{x}||^2 \\ &= f(\bar{x};x) + \frac{L}{2}||x-\bar{x}||^2 \end{aligned}$$

- max operation keeps the property as smooth strongly convex function.

Theorem 2.3.1: $x^* \in Q \Leftrightarrow f(x^*; x) \geq f(x^*; x^*) = f(x^*)$

$\Leftarrow$ As $f(x) \geq f(\bar{x}; x) + \frac{\mu}{2}||x - \bar{x}||^2$, we have

$$f(x) \geq f(x^*; x) + \frac{\mu}{2}||x - x^*||^2 \geq f(x^*; x^*) + 0 = f(x^*)$$

$\Rightarrow$ Prove by contradiction: if $f(x^*; x) < f(x^*)$, then for $1 \leq i \leq m$

$$f_i(x^*) + \langle f'(\bar{x}; x^*), x - x^* \rangle < f(x^*) = \max_{1 \leq i \leq m} f_i(x^*)$$

Define $\phi_i(\alpha) = f_i(x^* + \alpha(x - x^*)), \quad \alpha \in [0, 1]$
So either $\phi_i(0) \equiv f_i(x^*) < f(x^*)$ or

$$\phi_i(0) = f(x^*), \quad \phi_i'(0) = \langle f_i'(x^*), x - x^* \rangle < 0$$

So small enough α,

$$f_i(x^* + \alpha(x - x^*)) = \phi_i(\alpha) < f(x^*) \quad \forall 1 \leq i \leq m$$

contradiction!

Linearization achieves its minimum at x^*

Theorem 2.3.1: $x^* \in Q \Leftrightarrow f(x^*; x) \geq f(x^*; x^*) = f(x^*)$

$\Leftarrow$ As $f(x) \geq f(\bar{x}; x) + \frac{\mu}{2}||x - \bar{x}||^2$, we have

$$f(x) \geq f(x^*; x) + \frac{\mu}{2}||x - x^*||^2 \geq f(x^*; x^*) + 0 = f(x^*)$$

$\Rightarrow$ Prove by contradiction: if $f(x^*; x) < f(x^*)$, then for $1 \leq i \leq m$

$$f_i(x^*) + \langle f'(\bar{x}; x^*), x - x^* \rangle < f(x^*) = \max_{1 \leq i \leq m} f_i(x^*)$$

Define $\phi_i(\alpha) = f_i(x^* + \alpha(x - x^*)), \quad \alpha \in [0, 1]$
So either $\phi_i(0) \equiv f_i(x^*) < f(x^*)$ or

$$\phi_i(0) = f(x^*), \quad \phi_i'(0) = \langle f_i'(x^*), x - x^* \rangle < 0$$

So small enough α,

$$f_i(x^* + \alpha(x - x^*)) = \phi_i(\alpha) < f(x^*) \quad \forall 1 \leq i \leq m$$

contradiction!

Linearization achieves its minimum at x^*

Theorem 2.3.1: $x^* \in Q \Leftrightarrow f(x^*; x) \geq f(x^*; x^*) = f(x^*)$

$\Leftarrow$ As $f(x) \geq f(\bar{x}; x) + \frac{\mu}{2}||x - \bar{x}||^2$, we have

$$f(x) \geq f(x^*; x) + \frac{\mu}{2}||x - x^*||^2 \geq f(x^*; x^*) + 0 = f(x^*)$$

$\Rightarrow$ Prove by contradiction: if $f(x^*; x) < f(x^*)$, then for $1 \leq i \leq m$

$$f_i(x^*) + \langle f'(\bar{x}; x^*), x - x^* \rangle < f(x^*) = \max_{1 \leq i \leq m} f_i(x^*)$$

Define $\phi_i(\alpha) = f_i(x^* + \alpha(x - x^*)), \quad \alpha \in [0, 1]$

So either $\phi_i(0) \equiv f_i(x^*) < f(x^*)$ or

$$\phi_i(0) = f(x^*), \quad \phi_i'(0) = \langle f_i'(x^*), x - x^* \rangle < 0$$

So small enough α,

$$f_i(x^* + \alpha(x - x^*)) = \phi_i(\alpha) < f(x^*) \quad \forall 1 \leq i \leq m$$

contradiction!

Linearization achieves its minimum at x^*

Theorem 2.3.1: $x^* \in Q \Leftrightarrow f(x^*;x) \geq f(x^*;x^*) = f(x^*)$

$\Leftarrow$ As $f(x) \geq f(\bar{x};x) + \frac{\mu}{2}||x-\bar{x}||^2$, we have

$$f(x) \geq f(x^*;x) + \frac{\mu}{2}||x-x^*||^2 \geq f(x^*;x^*) + 0 = f(x^*)$$

$\Rightarrow$ Prove by contradiction: if $f(x^*;x) < f(x^*)$, then for $1 \leq i \leq m$

$$f_i(x^*) + \langle f'(\bar{x};x^*), x - x^* \rangle < f(x^*) = \max_{1 \leq i \leq m} f_i(x^*)$$

Define $\phi_i(\alpha) = f_i(x^* + \alpha(x - x^*)), \quad \alpha \in [0,1]$
So either $\phi_i(0) \equiv f_i(x^*) < f(x^*)$ or

$$\phi_i(0) = f(x^*), \quad \phi_i'(0) = \langle f_i'(x^*), x - x^* \rangle < 0$$

So small enough α,

$$f_i(x^* + \alpha(x - x^*)) = \phi_i(\alpha) < f(x^*) \quad \forall 1 \leq i \leq m$$

contradiction!

Linearization achieves its minimum at x^*

Corollary 2.3.1

$$f(x) \geq f(x^*) + \frac{\mu}{2}||x - x^*||^2$$

$$\begin{aligned} f(x) &\geq f(\bar{x}; x) + \frac{\mu}{2}||x - \bar{x}||^2 \\ &\geq f(x^*; x) + \frac{\mu}{2}||x - x^*||^2 \\ &\geq f(x^*; x^*) + \frac{\mu}{2}||x - x^*||^2 \\ &= f(x^*) + \frac{\mu}{2}||x - x^*||^2 \end{aligned}$$

So if x^* exists, it must be unique.

Corollary 2.3.1

$$f(x) \geq f(x^*) + \frac{\mu}{2}||x - x^*||^2$$

$$\begin{aligned} f(x) &\geq f(\bar{x}; x) + \frac{\mu}{2}||x - \bar{x}||^2 \\ &\geq f(x^*; x) + \frac{\mu}{2}||x - x^*||^2 \\ &\geq f(x^*; x^*) + \frac{\mu}{2}||x - x^*||^2 \\ &= f(x^*) + \frac{\mu}{2}||x - x^*||^2 \end{aligned}$$

So if x^* exists, it must be unique.

Theorem 3.2

Let a max-type function $f(x) \in \mathcal{S}_\mu^1(R^n)$, $\mu > 0$, and Q be a closed convex set. Then the solution x^* exists and unique.

- Let $\bar{x} \in Q$, consider the set $\bar{Q} = \{x \in Q | f(x) \leq f(\bar{x})\}$.
- Transform to a problem as

$$\min\{f(x) | x \in \bar{Q}\}$$

- Need to show $\bar{Q}$ is bounded.

$$\begin{aligned} & f(\bar{x}) \geq f_i(x) \geq f_i(\bar{x}) + \langle f'(\bar{x}), x - \bar{x}\rangle + \frac{\mu}{2}||x - \bar{x}||^2 \\ \Longrightarrow \quad & \frac{\mu}{2}||x - \bar{x}||^2 \leq ||f'(\bar{x})|| \cdot ||x - \bar{x}|| + f(\bar{x}) - f_i(\bar{x}) \end{aligned}$$

- So the solution x^* exists and is unique

Quick Summary

MiniMax, though generally not smooth, share all the properties as minimizing smooth strongly convex functions over simple convex set.

Linearization

max-type function	$f(x) = \max_{1 \le i \le m} f_i(x)$
linearization of $f(x)$	$f(\bar{x}; x) = \max_{1 \le i \le m} [f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x} \rangle]$

Essentially, linearization over each component.

Properties

- $f(\bar{x}; x) + \frac{\mu}{2}||x - \bar{x}||^2 \le f(x) \le f(\bar{x}; x) + \frac{L}{2}||x - \bar{x}||^2$;
- $x^* \in Q \Leftrightarrow f(x^*; x) \ge f(x^*; x^*) = f(x^*)$.
- $f(x) \ge f(x^*) + \frac{\mu}{2}||x - x^*||^2$
- the solution x^* exists and unique.

Road Map

- MiniMax problem
- Gradient Mapping for MiniMax problem;
- The complexity of gradient and optimal method;
- Optimization with functional constraint (General constrained optimization problem)
- Constrained Minimization Problem

As expected, share most of the properties as minimization over simple convex set.

Gradient Mapping

Similar as the case on minimization with convex set, we can define gradient mapping as follows:

$$f_\gamma(\bar{x}; x) = f(\bar{x}; x) + \frac{\gamma}{2}||x - \bar{x}||^2 \quad \text{(quadratic approximation)} \tag{4}$$

$$f^*(\bar{x}; \gamma) = \min_{x \in Q} f_\gamma(\bar{x}; x) \tag{5}$$

$$x_f(\bar{x}; \gamma) = \operatorname*{argmin}_{x \in Q} f_\gamma(\bar{x}; x) \tag{6}$$

$$g_f(\bar{x}; \gamma) = \gamma(\bar{x} - x_f(\bar{x}; \gamma)) \quad \text{(gradient mapping)} \tag{7}$$

The only difference is the linearization part $f(\bar{x}; x)$.

- When $m = 1$(only one component), the same as minimization over simple convex set;
- the linearization point $\bar{x}$ does not necessarily belong to Q;
- $f_\gamma(\bar{x}; x)$ is a max-type function composed with components:

$$f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x} \rangle + \frac{\gamma}{2}||x - \bar{x}||^2 \in \mathcal{S}^{1,1}_{\gamma,\gamma}(R^n), \quad i = 1, \cdots, m \tag{8}$$

Gradient Mapping

Similar as the case on minimization with convex set, we can define gradient mapping as follows:

$$
\begin{aligned}
f_\gamma(\bar{x}; x) &= f(\bar{x}; x) + \frac{\gamma}{2}||x - \bar{x}||^2 \qquad \text{(quadratic approximation)} && (4)\\
f^*(\bar{x}; \gamma) &= \min_{x \in Q} f_\gamma(\bar{x}; x) && (5)\\
x_f(\bar{x}; \gamma) &= \operatorname*{argmin}_{x \in Q} f_\gamma(\bar{x}; x) && (6)\\
g_f(\bar{x}; \gamma) &= \gamma\left(\bar{x} - x_f(\bar{x}; \gamma)\right) \qquad \text{(gradient mapping)} && (7)
\end{aligned}
$$

The only difference is the linearization part $f(\bar{x}; x)$.

- When $m = 1$(only one component), the same as minimization over simple convex set;
- the linearization point $\bar{x}$ does not necessarily belong to Q;
- $f_\gamma(\bar{x}; x)$ is a max-type function composed with components:

$$
f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x}\rangle + \frac{\gamma}{2}||x - \bar{x}||^2 \in \mathcal{S}^{1,1}_{\gamma,\gamma}(R^n), \quad i = 1, \cdots, m \tag{8}
$$

Linearization and gradient mapping

$f(x)$ is bounded by the linearization (plus quadratic term), Could we somehow bound the linearization part with gradient mapping?

Theorem 2.3.3

Let $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$, then for all $x \in Q$

$$f(\bar{x};x) \geq f^*(\bar{x};\gamma) + \langle g_f(\bar{x};\gamma), x - \bar{x}\rangle + \frac{1}{2\gamma}||g_f(\bar{x};\gamma)||^2 \quad (9)$$

$$\begin{aligned}
f(\bar{x};x) &= f_\gamma(\bar{x};x) - \frac{\gamma}{2}||x-\bar{x}||^2 && (10)\\
&\geq \underbrace{f_\gamma(\bar{x};x_f) + \frac{\gamma}{2}(||x-x_f||^2 - ||x-\bar{x}||^2)}_{f_\gamma(\bar{x};x)\in\mathcal{S}^{1,1}_{\gamma,\gamma}(R^n)} && (11)\\
&= f^*(\bar{x};\gamma) + \frac{\gamma}{2}\langle(\bar{x}-x_f, 2x - x_f - \bar{x}\rangle && (12)\\
&= f^*(\bar{x};\gamma) + \frac{\gamma}{2}\langle(\bar{x}-x_f, 2(x-\bar{x}) + (\bar{x}-x_f)\rangle && (13)\\
&= f^*(\bar{x};\gamma) + \langle g_f, x-\bar{x}\rangle + \frac{1}{2\gamma}||g_f||^2 && (14)
\end{aligned}$$

Linearization and gradient mapping

$f(x)$ is bounded by the linearization (plus quadratic term), Could we somehow bound the linearization part with gradient mapping?

Theorem 2.3.3

Let $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$, then for all $x \in Q$

$$f(\bar{x};x) \geq f^*(\bar{x};\gamma) + \langle g_f(\bar{x};\gamma), x - \bar{x}\rangle + \frac{1}{2\gamma}||g_f(\bar{x};\gamma)||^2 \tag{9}$$

$$\begin{aligned}
f(\bar{x};x) &= f_\gamma(\bar{x};x) - \frac{\gamma}{2}||x-\bar{x}||^2 && (10)\\
&\geq \underbrace{f_\gamma(\bar{x};x_f) + \frac{\gamma}{2}(||x-x_f||^2 - ||x-\bar{x}||^2)}_{f_\gamma(\bar{x};x)\in\mathcal{S}^{1,1}_{\gamma,\gamma}(R^n)} && (11)\\
&= f^*(\bar{x};\gamma) + \frac{\gamma}{2}\langle(\bar{x}-x_f, 2x - x_f - \bar{x}\rangle && (12)\\
&= f^*(\bar{x};\gamma) + \frac{\gamma}{2}\langle(\bar{x}-x_f, 2(x-\bar{x}) + (\bar{x}-x_f)\rangle && (13)\\
&= f^*(\bar{x};\gamma) + \langle g_f, x-\bar{x}\rangle + \frac{1}{2\gamma}||g_f||^2 && (14)
\end{aligned}$$

Properties with respect to gradient mapping

Since $f(\bar{x}; x) \geq f^*(\bar{x}; \gamma) + \langle g_f(\bar{x}; \gamma), x - \bar{x} \rangle + \frac{1}{2\gamma} ||g_f(\bar{x}; \gamma)||^2$

Corollary 2.3.2 *Let $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$ and $\gamma \geq L$. Then:*

1. For any $x \in Q$ and $\bar{x} \in R^n$ we have:

$$f(x) \geq f(x_f(\bar{x}; \gamma)) + \langle g_f(\bar{x}; \gamma), x - \bar{x} \rangle + \frac{1}{2\gamma} \| g_f(\bar{x}; \gamma) \|^2 + \frac{\mu}{2} \| x - \bar{x} \|^2 . \tag{2.3.7}$$

2. If $\bar{x} \in Q$ then

$$f(x_f(\bar{x}; \gamma)) \leq f(\bar{x}) - \frac{1}{2\gamma} \| g_f(\bar{x}; \gamma) \|^2, \tag{2.3.8}$$

3. For any $\bar{x} \in R^n$ we have:

$$\langle g_f(\bar{x}; \gamma), \bar{x} - x^* \rangle \geq \frac{1}{2\gamma} \| g_f(\bar{x}; \gamma) \|^2 + \frac{\mu}{2} \| x^* - \bar{x} \|^2 . \tag{2.3.9}$$

Proof:
Assumption $\gamma \geq L$ implies that $f^*(\bar{x}; \gamma) \geq f(x_f(\bar{x}; \gamma))$. Therefore (2.3.7) follows from (2.3.6) since

$$f(x) \geq f(\bar{x}; x) + \frac{\mu}{2} \| x - \bar{x} \|^2$$

for all $x \in R^n$ (see Lemma 2.3.1).

Using (2.3.7) with $x = \bar{x}$, we get (2.3.8), and using (2.3.7) with $x = x^*$, we get (2.3.9) since $f(x_f(\bar{x}; \gamma)) - f(x^*) \geq 0$. □

Variance with respect to γ

Lemma 2.3.2 *For any γ_1, $\gamma_2 > 0$ and $\bar{x} \in R^n$ we have:*

$$f^*(\bar{x};\gamma_2) \geq f^*(\bar{x};\gamma_1) + \frac{\gamma_2 - \gamma_1}{2\gamma_1\gamma_2} \parallel g_f(\bar{x};\gamma_1) \parallel^2 .$$

Proof:
Denote $x_i = x_f(\bar{x};\gamma_i)$, $g_i = g_f(\bar{x};\gamma_i)$, $i = 1, 2$. In view of (2.3.6), we have:

$$f(\bar{x};x) + \frac{\gamma_2}{2} \parallel x - \bar{x} \parallel^2 \geq f^*(\bar{x};\gamma_1) + \langle g_1, x - \bar{x}\rangle + \frac{1}{2\gamma_1} \parallel g_1 \parallel^2 + \frac{\gamma_2}{2} \parallel x - \bar{x} \parallel^2 \qquad (2.3.10)$$

for all $x \in Q$. In particular, for $x = x_2$ we obtain:

$$\begin{aligned} f^*(\bar{x};\gamma_2) = {} & f(\bar{x};x_2) + \tfrac{\gamma_2}{2} \parallel x_2 - \bar{x} \parallel^2 \\ \geq {} & f^*(\bar{x};\gamma_1) + \langle g_1, x_2 - \bar{x}\rangle + \tfrac{1}{2\gamma_1} \parallel g_1 \parallel^2 + \tfrac{\gamma_2}{2} \parallel x_2 - \bar{x} \parallel^2 \\ = {} & f^*(\bar{x};\gamma_1) + \tfrac{1}{2\gamma_1} \parallel g_1 \parallel^2 - \tfrac{1}{\gamma_2}\langle g_1, g_2\rangle + \tfrac{1}{2\gamma_2} \parallel g_2 \parallel^2 \\ \geq {} & f^*(\bar{x};\gamma_1) + \tfrac{1}{2\gamma_1} \parallel g_1 \parallel^2 - \tfrac{1}{2\gamma_2} \parallel g_1 \parallel^2 . \end{aligned}$$

Road Map

- MiniMax problem
- Gradient Mapping for MiniMax problem;
- The complexity of gradient and optimal method;
- Optimization with functional constraint (General constrained optimization problem)
- Constrained Minimization Problem

Gradient Method: Comparison

General Scheme for Gradient Method:

$$\begin{aligned} x_0 &\in Q, \\ x_{k+1} &= x_k - h g_f(x_k; L), \quad k = 0, \cdots \end{aligned}$$

On minimization over Minimax Problem (Same as over simple set)

If we choose $h \leq \frac{1}{L}$ in General Scheme for Gradient Method, then

$$\|x_k - x^*\|^2 \leq (1 - \mu h)^k \|x_0 - x^*\|^2. \tag{15}$$

If $h = \frac{1}{L}$

$$\|x_k - x^*\|^2 \leq \left(1 - \frac{\mu}{L}\right)^k \|x_0 - x^*\|^2 \tag{16}$$

the gradient method has the same rate of convergence as in the smooth case.

Let $r_k = \|x_k - x^*\|, g = g_f(x_k; L)$, (As $2\langle g, x_k - x*\rangle \geq \frac{1}{\gamma}\|g\|^2 + \mu\|x_k - x^*\|^2$)

$$\begin{aligned} r_{k+1}^2 &= \|x_k - x^* - h_{g_f}\|^2 \\ &= r_k^2 - 2h\langle g_f, x_k - x^*\rangle + h^2\|g_f\|^2 \\ &\leq (1 - h\mu)r_k^2 + h(h - \frac{1}{L})\|g_f\|^2 \leq (1 - \frac{\mu}{L})r_k^2. \end{aligned}$$

Minimization Method - Optimal Method

- **Step 1: define the estimate sequence** Assume that we have $x_0 \in Q$. Define

$$\phi_0(x) = \phi_0^* + \frac{\gamma_0}{2}\|x - v_0\|^2, \quad (17)$$

$$\phi_k(x) = (1-\alpha_k)\phi_k + \alpha_k\left[f(x_Q) + \langle g_Q, x - y_k\rangle + \frac{1}{2\gamma}\|g_Q\|^2 + \frac{\mu}{2}\|x - y_k\|^2\right], \quad (18)$$

where $x_Q = x_Q(y_k; L)$ and $g_Q = g_Q(y_k; L)$.

- **Step 2: rewrite the sequence** $\{\phi_k(x)\}$ For $k \geq 0$, we have

$$\phi_k(x) = \phi_k^* + \frac{\gamma_k}{2}\|x - v_k\|^2, \quad (19)$$

where the following recursive rules are defined for γ_k, v_k, and ϕ_k^* as

$$\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k\mu, \quad (20)$$

$$v_{k+1} = \frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k g_Q], \quad (21)$$

$$\phi_{k+1}^* = (1-\alpha_k)\phi_k^* + \alpha_k f(x_Q) + \left(\frac{\alpha_k}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|g_Q\|^2 + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|y_k - v_k\|^2 + \langle g_Q, v_k - y_k\rangle\right). \quad (22)$$

- **Step 3: ensure** $\phi_k^* \geq f(x_k)$ Using the inequality

$$f(x_k) \geq f(x_Q) + \langle g_Q, x_k - y_k \rangle + \frac{1}{2\gamma}\|g_Q\|^2 + \frac{\mu}{2}\|x_k - y_k\|^2, \quad (23)$$

we come to the following lower bound

$$\begin{aligned} \phi_{k+1}^* \quad &\geq \quad (1-\alpha_k)f(x_k) + \alpha_k f(x_Q) + \left(\frac{\alpha_k}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|g_Q\|^2 \\ &\quad + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|y_k - v_k\|^2 + \langle g_Q, v_k - y_k \rangle\right) \\ &\geq \quad f(x_Q) + \left(\frac{1}{2L} - \frac{\alpha_k^2}{2\gamma_{k+1}}\right)\|g_Q\|^2 + (1-\alpha_k)\left\langle g_Q, \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k \right\rangle. \end{aligned}$$

Therefore, we choose

$$\begin{aligned} x_{k+1} &= x_Q, \\ L\alpha_k^2 &= (1-\alpha_k)\gamma_k + \alpha_k\mu = \gamma_{k+1}, \\ y_k &= \frac{1}{\gamma_k + \alpha_k\mu}(\alpha_k\gamma_k v_k + \gamma_{k+1}x_k). \end{aligned}$$

Constant Step Scheme 3 for Simple Set

1. Choose $x_0 \in Q$ and $\alpha_0 \in (0,1)$. Set $y_0 = x_0, q = \mu/L$.
2. kth iteration ($k \geq 0$).
   - Compute $f(y_k)$ and $f'(y_k)$. Set
     $$x_{k+1} = x_Q. \tag{24}$$
   - Compute $\alpha_{k+1} \in (0,1)$ from the equation
     $$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + q\alpha_{k+1},$$
     and set
     $$\beta_k = \frac{\alpha_k(1-\alpha)}{\alpha_k^2 + \alpha_{k+1}}, \quad y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k). \tag{25}$$

Note that only $\{x_k\}$ are feasible for Q, while $\{y_k\}$ can not be guaranteed to be feasible.

Completely identical to unconstrained case. The convergent rate is exactly the same as unconstrained case.

Convergence Rate

Theorem 2.3.5 *Let the max-type function* f *belong to* $\mathcal{S}^{1,1}_{\mu,L}(R^n)$. *If in (2.3.12) we take* $\alpha_0 \geq \sqrt{\frac{\mu}{L}}$, *then*

$$f(x_k) - f^* \leq \left[f(x_0) - f^* + \frac{\gamma_0}{2} \parallel x_0 - x^* \parallel^2 \right] \times \min \left\{ \left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{(2\sqrt{L} + k\sqrt{\gamma_0})^2} \right\},$$

where $\gamma_0 = \frac{\alpha_0(\alpha_0 L - \mu)}{1-\alpha_0}$.

$$\textbf{Scheme for } f \in \mathcal{S}^{1,1}_{\mu,L}(R^n) \qquad (2.3.13)$$

0. Choose $x_0 \in Q$. Set $y_0 = x_0$, $\beta = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$.

1. kth iteration ($k \geq 0$). Compute $\{f_i(y_k)\}$ and $\{f_i'(y_k)\}$. Set

$$x_{k+1} = x_f(y_k; L), \qquad y_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k).$$

Theorem 2.3.6 *For this scheme we have:*

$$f(x_k) - f^* \leq 2\left(1 - \sqrt{\frac{\mu}{L}}\right)^k (f(x_0) - f^*). \qquad (2.3.14)$$

Proof:
Scheme (2.3.13) corresponds to $\alpha_0 = \sqrt{\frac{\mu}{L}}$. Then $\gamma_0 = \mu$ and we get (2.3.14) since $f(x_0) \geq f^* + \frac{\mu}{2} \parallel x_0 - x^* \parallel^2$ in view of Corollary 2.3.1. □

Optimization with Functional Constraints

Problem 2.3.16

$$\min \quad f_0(x) \tag{26}$$
$$s.t. \quad f_i(x) \leq 0, \quad i = 1, \cdots, m \tag{27}$$
$$x \in Q \tag{28}$$

parametric max-type function

$$f(t; x) = \max\{f_0(x) - t; f_i(x)\}$$

$f(t; \cdot)$ are strongly convex in x. For any t, $x^*(t)$ exists and unique.

$$f^*(t) = \min_{x \in Q} f(t; x)$$

We'll try to get close to the solution using a process based on the approximate values of the function $f^* t(x)$ (aka. *sequential quadratic programming*)

Lemma 2.3.4

Lemma 2.3.4 *Let t^* be the optimal value of the problem (2.3.16). Then*

$$f^*(t) \leq 0 \quad \textit{for all } t \geq t^*,$$

$$f^*(t) > 0 \quad \textit{for all } t < t^*.$$

Proof:
Let x^* be the solution to (2.3.16). If $t \geq t^*$ then

$$f^*(t) \leq f(t; x^*) = \max\{f_0(x^*) - t; f_i(x^*)\} \leq \max\{t^* - t; f_i(x^*)\} \leq 0.$$

Suppose that $t < t^*$ and $f^*(t) \leq 0$. Then there exists $y \in Q$ such that

$$f_0(y) \leq t < t^*, \quad f_i(y) \leq 0, \; i = 1, \ldots, m.$$

Thus, t^* cannot be the optimal value of (2.3.16). □

Note that, as t increases, $f^*(t)$ decreases in a sense.
Hence, the smallest root of the function $f^*(t)$ corresponds to the optimal value of the problem of functional constraint.
Our goal is to form a process of finding the root.

Lemma 2.3.5 *For any $\Delta \geq 0$ we have:*

$$f^*(t) - \Delta \leq f^*(t+\Delta) \leq f^*(t).$$

Proof:
Indeed,

$$\begin{aligned} f^*(t+\Delta) &= \min_{x \in Q} \max_{1 \leq i \leq m} \{f_0(x) - t - \Delta; f_i(x)\} \\ &\leq \min_{x \in Q} \max_{1 \leq i \leq m} \{f_0(x) - t; f_i(x)\} = f^*(t), \\ f^*(t+\Delta) &= \min_{x \in Q} \max_{1 \leq i \leq m} \{f_0(x) - t; f_i(x) + \Delta\} - \Delta \\ &\geq \min_{x \in Q} \max_{1 \leq i \leq m} \{f_0(x) - t; f_i(x)\} - \Delta = f^*(t) - \Delta. \quad \square \end{aligned}$$

- Thus, $f^*(t)$ decreases in t and is Lipshitz continuous with the constant equal to 1.
- Keep in mind that here the property is satisfied for any max-type function like $f_\mu(\bar{x}; x)$ and $f_L(\bar{x}; x)$.

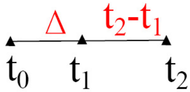
Lemma 2.3.6

For any $t_1 < t_2$ and $\Delta \geq 0$, we have

$$f^*(t_1 - \Delta) \geq f^*(t_1) + \Delta \frac{f^*(t_1) - f^*(t_2)}{t_2 - t_1} = f^*(t_1) - \Delta \frac{f^*(t_2) - f^*(t_1)}{t_2 - t_1} \quad (29)$$

Let $t_0 = t_1 - \Delta$, $\alpha = \frac{\Delta}{t_2 - t_0}$, then $t_1 = (1 - \alpha) t_0 + \alpha t_2$, so

$$(29) :\equiv f^*(t_1) \leq (1 - \alpha) f^*(t_0) + \alpha f^*(t_0)$$



Let $x_\alpha = (1 - \alpha) x^*(t_0) + \alpha x^*(t_2)$. We have:

$$f^*(t_1) \leq \max_{1 \leq i \leq m} \{f_0(x_\alpha) - t_1; f_i(x_\alpha)\}$$

$$\leq \max_{1 \leq i \leq m} \{(1 - \alpha)(f_0(x^*(t_0)) - t_0) + \alpha(f_0(x^*(t_2)) - t_2); (1 - \alpha) f_i(x^*(t_0)) + \alpha f_i(x^*(t_2))\}$$

$$\leq (1 - \alpha) \max_{1 \leq i \leq m} \{f_0(x^*(t_0)) - t_0; f_i(x^*(t_0))\} + \alpha \max_{1 \leq i \leq m} \{f_0(x^*(t_2)) - t_2; f_i(x^*(t_2))\}$$

$$= (1 - \alpha) f^*(t_0) + \alpha f^*(t_2),$$

Lemma 2.3.5

For any $\Delta > 0$, we have:

$$f^*(t) - \Delta \leq f^*(t + \Delta) \leq f^*(t)$$

Lemma 2.3.6

For any $t_1 < t_2$ and $\Delta \geq 0$, we have

$$f^*(t_1 - \Delta) \geq f^*(t_1) + \Delta \frac{f^*(t_1) - f^*(t_2)}{t_2 - t_1} = f^*(t_1) - \Delta \frac{f^*(t_2) - f^*(t_1)}{t_2 - t_1}$$

- Both Lemmas are valid for any parametric max-type functions.

Linearization and Gradient Mapping

Linearization of $f(t;x)$:

$$f(t;\bar{x};x) = \max_{1\le i\le m}\{f_0(x) + \langle f_0'(\bar{x}), x-\bar{x}\rangle - t;\quad f_i(x) + \langle f_i'(\bar{x}), x-\bar{x}\rangle\}$$

$$f_\gamma(t;\bar{x};x) = f(t;\bar{x};x) + \frac{\gamma}{2}\|x-\bar{x}\|^2 \quad (30)$$

$$f^*(t;\bar{x};\gamma) = \min_{x\in Q} f_\gamma(t;\bar{x};x) \quad (31)$$

$$x_f(t;\bar{x};\gamma) = \operatorname{argmin}_{x\in Q} f_\gamma(t;\bar{x};\gamma) \quad (32)$$

$$g_f(t;\bar{x};\gamma) = \gamma(\bar{x} - x_f(t;\bar{x};\gamma)) \quad (33)$$

g_f is the *constrained gradient mapping*; $\bar{x}$ is not necessarily in Q.

Bounds for the Linearization

$$f_\gamma(t;\bar{x};x) = f(t;\bar{x};x) + \frac{\gamma}{2}\|x-\bar{x}\|^2$$

- $f_\gamma(t;\bar{x};x)$ is itself a max-type function;
- $f_\gamma(t;\bar{x};x) \in \mathcal{S}^{1,1}_{\gamma,\gamma}(R^n)$. So for any t, the constrained gradient mapping is well defined;
- $f_\mu(t;\bar{x};x) \le f(t;x) \le f_L(t;\bar{x};x)$, as $f(t;x) \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$; Hence

$$f^*_\mu(t;\bar{x};x) \le f^*(t) \le f^*_L(t;\bar{x};x)$$

- For any $\bar{x} \in R^n$, $\gamma > 0$, $\Delta \ge 0$ and $t_1 < t_2$, we have

$$f^*(t_1-\Delta;\bar{x};\gamma) \ge f^*(t_1;\bar{x};\gamma) + \frac{\Delta}{t_2-t_1}(f^*(t_1;\bar{x};\gamma) - f^*(t_2;\bar{x};\gamma))$$

- $f^*(t;\bar{x};\mu) \ge f^*(t;\bar{x};L) - \frac{L-\mu}{2\mu L}\|g_f(t;\bar{x};L)\|^2$

Bounds for the Linearization

$$f_\gamma(t;\bar{x};x) = f(t;\bar{x};x) + \frac{\gamma}{2}\|x-\bar{x}\|^2$$

- $f_\gamma(t;\bar{x};x)$ is itself a max-type function;
- $f_\gamma(t;\bar{x};x) \in \mathcal{S}^{1,1}_{\gamma,\gamma}(R^n)$. So for any t, the constrained gradient mapping is well defined;
- $f_\mu(t;\bar{x};x) \le f(t;x) \le f_L(t;\bar{x};x)$, as $f(t;x) \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$; Hence

$$f^*_\mu(t;\bar{x};x) \le f^*(t) \le f^*_L(t;\bar{x};x)$$

- For any $\bar{x} \in R^n$, $\gamma > 0$, $\Delta \ge 0$ and $t_1 < t_2$, we have

$$f^*(t_1-\Delta;\bar{x};\gamma) \ge f^*(t_1;\bar{x};\gamma) + \frac{\Delta}{t_2-t_1}(f^*(t_1;\bar{x};\gamma) - f^*(t_2;\bar{x};\gamma))$$

- $f^*(t;\bar{x};\mu) \ge f^*(t;\bar{x};L) - \frac{L-\mu}{2\mu L}\|g_f(t;\bar{x};L)\|^2$

Bounds for the Linearization

$$f_\gamma(t;\bar{x};x) = f(t;\bar{x};x) + \frac{\gamma}{2}\|x-\bar{x}\|^2$$

- $f_\gamma(t;\bar{x};x)$ is itself a max-type function;
- $f_\gamma(t;\bar{x};x) \in \mathcal{S}^{1,1}_{\gamma,\gamma}(R^n)$. So for any t, the constrained gradient mapping is well defined;
- $f_\mu(t;\bar{x};x) \le f(t;x) \le f_L(t;\bar{x};x)$, as $f(t;x) \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$; Hence

$$f^*_\mu(t;\bar{x};x) \le f^*(t) \le f^*_L(t;\bar{x};x)$$

- For any $\bar{x} \in R^n$, $\gamma > 0$, $\Delta \ge 0$ and $t_1 < t_2$, we have

$$f^*(t_1-\Delta;\bar{x};\gamma) \ge f^*(t_1;\bar{x};\gamma) + \frac{\Delta}{t_2-t_1}(f^*(t_1;\bar{x};\gamma) - f^*(t_2;\bar{x};\gamma))$$

- $f^*(t;\bar{x};\mu) \ge f^*(t;\bar{x};L) - \frac{L-\mu}{2\mu L}\|g_f(t;\bar{x};L)\|^2$

Root of $f^*(t; \bar{x}; \mu)$

- We are interested in finding the root of the function $f^*(t)$. We focus on the approximation of $f_\gamma(t; \bar{x}; \gamma)$.

$$t^*(\bar{x}; t) = root_t(f^*(t; \bar{x}; \mu))$$

- The root of the lower-bound quadratic approximation.
- Notice that the notation is a little confusing here $t^*(\bar{x}; t)$ actually depends on $\bar{x}$, not t.

Lemma 2.3.7 *Let $\bar{x} \in R^n$ and $\bar{t} < t^*$ are such that*

$$f^*(\bar{t}; \bar{x}; \mu) \geq (1 - \kappa) f^*(\bar{t}; \bar{x}; L)$$

for some $\kappa \in (0, 1)$. Then $\bar{t} < t^*(\bar{x}, \bar{t}) \leq t^*$. Moreover, for any $t < \bar{t}$ and $x \in R^n$ we have:

$$f^*(t; x; L) \geq 2(1 - \kappa) f^*(\bar{t}; \bar{x}; L) \sqrt{\frac{\bar{t} - t}{t^*(\bar{x}, \bar{t}) - \bar{t}}}.$$

Root of $f^*(t; \bar{x}; \mu)$

- We are interested in finding the root of the function $f^*(t)$. We focus on the approximation of $f_\gamma(t; \bar{x}; \gamma)$.

$$t^*(\bar{x}; t) = root_t(f^*(t; \bar{x}; \mu))$$

- The root of the lower-bound quadratic approximation.
- Notice that the notation is a little confusing here $t^*(\bar{x}; t)$ actually depends on $\bar{x}$, not t.

Lemma 2.3.7 *Let $\bar{x} \in R^n$ and $\bar{t} < t^*$ are such that*

$$f^*(\bar{t}; \bar{x}; \mu) \geq (1 - \kappa) f^*(\bar{t}; \bar{x}; L)$$

for some $\kappa \in (0, 1)$. Then $\bar{t} < t^*(\bar{x}, \bar{t}) \leq t^*$. Moreover, for any $t < \bar{t}$ and $x \in R^n$ we have:

$$f^*(t; x; L) \geq 2(1 - \kappa) f^*(\bar{t}; \bar{x}; L) \sqrt{\frac{\bar{t} - t}{t^*(\bar{x}, \bar{t}) - \bar{t}}}.$$

Lemma 2.3.7

Lemma 2.3.7 *Let $\bar{x} \in R^n$ and $\bar{t} < t^*$ are such that*

$$f^*(\bar{t}; \bar{x}; \mu) \geq (1-\kappa) f^*(\bar{t}; \bar{x}; L)$$

for some $\kappa \in (0,1)$. Then $\bar{t} < t^*(\bar{x}, \bar{t}) \leq t^*$. Moreover, for any $t < \bar{t}$ and $x \in R^n$ we have:

$$f^*(t; x; L) \geq 2(1-\kappa) f^*(\bar{t}; \bar{x}; L) \sqrt{\frac{\bar{t} - t}{t^*(\bar{x}, \bar{t}) - \bar{t}}}.$$

Proof:

Since $\bar{t} < t^*$, we have:

$$0 < f^*(\bar{t}) \leq f^*(\bar{t}; \bar{x}; L) \leq \frac{1}{1-\kappa} f^*(\bar{t}; \bar{x}; \mu).$$

Thus, $f^*(\bar{t}; \bar{x}; \mu) > 0$ and therefore $t^*(\bar{x}, \bar{t}) > \bar{t}$ since $f^*(t; \bar{x}; \mu)$ decreases in t.

Denote $\Delta = \bar{t} - t$, Then,

$$\begin{aligned} f^*(t;x;L) &\geq f^*(t) \geq f^*(t;\bar{x};\mu) && (34) \\ &\geq f^*(\bar{t};\bar{x};\mu) + \frac{\Delta}{t^*(\bar{x},t) - \bar{t}} \left[f^*(\bar{t};\bar{x};\mu) - \underbrace{f^*(t^*(\bar{x},t),\bar{x},\mu)}_{=0} \right] \\ &\geq (1-\kappa)\left(1 + \frac{\Delta}{t^*(\bar{x};t) - \bar{t}}\right) f^*(\bar{t};\bar{x};L) && (35) \\ &\geq (1-\kappa) 2\sqrt{\frac{\Delta}{t^*(\bar{x};t) - \bar{t}}} \; f^*(\bar{t};\bar{x};L) && (36) \\ &= 2(1-\kappa) f^*(\bar{t};\bar{x};L) \sqrt{\frac{\bar{t} - t}{t^*(\bar{x};t) - \bar{t}}} && (37) \end{aligned}$$

Constrained Minimization Scheme (2.3.22)

0. Choose $x_0 \in Q$ and $t_0 < t^*$. Choose $\kappa \in (0, \frac{1}{2})$ and the accuracy $\epsilon > 0$.

1. kth iteration ($k \geq 0$).

   a). Generate the sequence $\{x_{k,j}\}$ by the minimax method (2.3.13) as applied to the max-type function $f(t_k; x)$ with the starting point $x_{k,0} = x_k$. If

$$f^*(t_k; x_{k,j}; \mu) \geq (1-\kappa) f^*(t_k; x_{k,j}; L)$$

   then stop the internal process and set $j(k) = j$,

$$\begin{aligned} j^*(k) &= \arg \min_{0 \leq j \leq j(k)} f^*(t_k; x_{k,j}; L), \\ x_{k+1} &= x_f(t_k; x_{k,j^*(k)}; L). \end{aligned}$$

   Global Stop: Terminate the whole process if at some iteration of the internal scheme we have $f^*(t_k; x_{k,j}; L) \leq \epsilon$.

   b). Set $t_{k+1} = t^*(x_{k,j(k)}, t_k)$. □

Comment on the Scheme

Essentially two steps:

- Given t, find x until the lower bound $f(t;\bar{x};\mu)$ and the upper bound $f(t;\bar{x};L)$ of $f(t,\bar{x})$ is not too distant; Then pick the minimum one during the internal process;
- Given x, update t via finding the root of the lower bound;
  QCQP:

of the function

$$f^*(t;\bar{x};\mu) = \min_{x \in Q} f_\mu(t;\bar{x};x),$$

where $f_\mu(t;\bar{x};x)$ is a max-type function composed with the components

$$f_0(\bar{x}) + \langle f_0'(\bar{x}), x - \bar{x}\rangle + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2 - t,$$

$$f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x}\rangle + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2, \; i = 1, \dots, m.$$

In view of Lemma 2.3.4, it is the optimal value of the following minimization problem:

$$\begin{aligned} &\min \, [f_0(\bar{x}) + \langle f_0'(\bar{x}), x - \bar{x}\rangle + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2], \\ \text{s.t.} \quad & f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x}\rangle + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2 \le 0, \; i = 1, \dots, m, \\ & x \in Q. \end{aligned}$$

- The master process is continued until the upper bound function is close enough to 0 ($< \epsilon$)
- We start from a $t_0 < t^*$, and increases t gradually.

Following

- Here, we only focus on analytical complexity of this method.
- The total cost is of the order

$$\ln \frac{t_0 - t^*}{\epsilon} \sqrt{\frac{L}{\mu}} \ln \sqrt{\frac{L}{\mu}}$$

- This value differs from the lower bound for the unconstrained minimization problem by a factor of $\ln \frac{L}{\mu}$. (Not quite sure)
- Thus, the scheme is suboptimal for constrained optimization problems. But we cannot say more since the specific lower complexity bounds for constrained minimization are not known.
- We'll estimate the complexity of the master process;
- Then estimate the complexity for the internal process (given t, estimate an x);
- Finally, we get the total complexity.

Following

- Here, we only focus on analytical complexity of this method.
- The total cost is of the order

$$\ln \frac{t_0 - t^*}{\epsilon} \sqrt{\frac{L}{\mu}} \ln \sqrt{\frac{L}{\mu}}$$

- This value differs from the lower bound for the unconstrained minimization problem by a factor of $\ln \frac{L}{\mu}$. (Not quite sure)
- Thus, the scheme is suboptimal for constrained optimization problems. But we cannot say more since the specific lower complexity bounds for constrained minimization are not known.
- We'll estimate the complexity of the master process;
- Then estimate the complexity for the internal process (given t, estimate an x);
- Finally, we get the total complexity.

Lemma 2.3.8: complexity of master process

Lemma 2.3.8

$$f^*(t_k; x_{k+1}; L) \leq \frac{t^* - t_0}{1-\kappa}\left[\frac{1}{2(1-\kappa)}\right]^k$$

$$Let\ \beta = \frac{1}{2(1-\kappa)}(< 1\ as\ \kappa < 0.5) \qquad \delta_k = \frac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1} - t_k}}$$

Lemma 2.3.7

For $t < \bar{t} < t^*(\bar{x}, \bar{t}) \leq t^*$, we have

$$f^*(t; \bar{x}; L) \geq 2(1-\kappa) f^*(\bar{t}; \bar{x}; L)\sqrt{\frac{\bar{t} - t}{t^*(\bar{x}; t) - \bar{t}}} \tag{38}$$

Let $t = t_{k-1}, \bar{t} = t_k, t^*(\bar{x}; t) = t_{k+1}(As\ t_{k+1} = t^*(x_{k,j(k)}, t_k))$, we have

$$2(1-\kappa)\frac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1} - t_k}} \leq \frac{f^*(t_{k-1}; x_{k-1,j(k-1)}; L)}{\sqrt{t_k - t_{k-1}}} \implies \delta_k \leq \beta\delta_{k-1} \tag{39}$$

$$f^*(t_k; x_{k,j(k)}; L) = \delta_k\sqrt{t_{k+1} - t_k} \leq \beta^k \delta_0 \sqrt{t_{k+1} - t_k} \tag{40}$$

$$= \beta^k f^*(t_0; x_{x_0,j(0)}; L)\sqrt{\frac{t_{k+1} - t_k}{t_1 - t_0}} \tag{41}$$

Lemma 2.3.8: complexity of master process

Lemma 2.3.8

$$f^*(t_k; x_{k+1}; L) \leq \frac{t^* - t_0}{1-\kappa}\left[\frac{1}{2(1-\kappa)}\right]^k$$

$$Let\ \beta = \frac{1}{2(1-\kappa)} (< 1\ as\ \kappa < 0.5) \qquad \delta_k = \frac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1} - t_k}}$$

Lemma 2.3.5

For any $\Delta > 0$, we have:

$$f^*(t) - \Delta \leq f^*(t + \Delta) \leq f^*(t)$$

Let $t_1 = t_0 + \Delta$, we have $t_1 - t_0 \geq f^*(t_0; x_{0,j(0)}; \mu)$. So

$$f^*(t_k; x_{k,j(k)}; L) = \beta^k f^*(t_0; x_{x_0,j(0)}; L)\sqrt{\frac{t_{k+1} - t_k}{t_1 - t_0}} \quad (42)$$

$$\leq \beta^k f^*(t_0; x_{x_0,j(0)}; L)\sqrt{\frac{t_{k+1} - t_k}{f^*(t_0; x_{0,j(0)}; \mu)}} \quad (43)$$

$$\leq \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0; x_{0,j(0)}; \mu)(t_{k+1} - t_k)} \leq \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0)(t_0 - t^*)} \quad (44)$$

$$\leq \frac{t^* - t_0}{1-\kappa}\left[\frac{1}{2(1-\kappa)}\right]^k \quad (As\ \ f^*(t_0) \leq t^* - t_0) \quad (45)$$

Lemma 2.3.8: complexity of master process

Master Process

$$f^*(t_k; x_{k+1}; L) \leq \frac{t^* - t_0}{1-\kappa} \left[\frac{1}{2(1-\kappa)}\right]^k < \epsilon \Longrightarrow N(\epsilon) = \frac{1}{\ln[2(1-\kappa)]} \ln \frac{t^* - t_0}{(1-\kappa)\epsilon}$$

Complexity of Internal Process

Lemma 2.3.10 *For all k, $0 \le k \le N$, we have:*

$$j(k) \le 1 + \sqrt{\frac{L}{\mu}} \cdot \ln \frac{2(L-\mu)\Delta_k}{\kappa\mu\Delta_{k+1}}.$$

Corollary 2.3.3

$$\sum_{k=0}^{N} j(k) \le (N+1)\left[1 + \sqrt{\tfrac{L}{\mu}} \cdot \ln \tfrac{2(L-\mu)}{\kappa\mu}\right] + \sqrt{\tfrac{L}{\mu}} \cdot \ln \tfrac{\Delta_0}{\Delta_{N+1}}.$$

Lemma 2.3.11

$$j^* \le 1 + \sqrt{\frac{L}{\mu}} \cdot \ln \frac{2(L-\mu)\Delta_{N+1}}{\kappa\mu\epsilon}.$$

Corollary 2.3.4

$$j^* + \sum_{k=0}^{N} j(k) \le (N+2)\left[1 + \sqrt{\frac{L}{\mu}} \cdot \ln \frac{2(L-\mu)}{\kappa\mu}\right] + \sqrt{\frac{L}{\mu}} \cdot \ln \frac{\Delta_0}{\epsilon}.$$

As $N(\epsilon) = \frac{1}{\ln[2(1-\kappa)]} \ln \frac{t^* - t_0}{(1-\kappa)\epsilon}$, the total cost is:

$$\left[\tfrac{1}{\ln[2(1-\kappa)]} \ln \tfrac{t_0 - t^*}{(1-\kappa)\epsilon} + 2\right] \cdot \left[1 + \sqrt{\tfrac{L}{\mu}} \cdot \ln \tfrac{2(L-\mu)}{\kappa\mu}\right]$$
$$+ \sqrt{\tfrac{L}{\mu}} \cdot \ln\left(\tfrac{1}{\epsilon} \cdot \max_{1 \le i \le m}\{f_0(x_0) - t_0; f_i(x_0)\}\right). \quad (2.3.26)$$