Chapter II. Convex Differentiable Optimization

Lecture 6: Minimization problems with smooth components

Yurii Nesterov
EPLouvain (UCL)

Outline

- MiniMax Problem.
- Gradient Mapping for MiniMax.
- Gradient Method for MiniMax.
- Optimal Methods for MiniMax.
- Problem with functional constraints.
- Augmented Function's Approach.
- Methods for Constrained Minimization.

MiniMax Problem

Problem:

$$\min_{x \in Q} \left[f(x) = \max_{1 \le i \le m} f_i(x) \right], \quad f_i \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n), \tag{1}$$

where Q is a closed convex set.

In general, $f(\cdot)$ is not differentiable.

Motivation:

1. Multi-criterion Optimization.
2. Game Theory.
3. Constrained Optimization.

Simple inequalities

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

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

the *linearization* of function $f(\cdot)$ at $\bar{x}$.

Lemma 1: For any $x \in \mathbb{R}^n$ we have:

$$f(x) \quad \ge \quad f(\bar{x}; x) + \tfrac{\mu}{2} \| \, x - \bar{x} \, \|^2, \tag{2}$$

$$f(x) \quad \le \quad f(\bar{x}; x) + \tfrac{L}{2} \| \, x - \bar{x} \, \|^2 \, . \tag{3}$$

Proof: Indeed,

$$f_i(x) \quad \ge \quad f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x} \rangle + \tfrac{\mu}{2} \| \, x - \bar{x} \, \|^2 \, .$$

Taking the maximum of the RHS, we get (**??**).

For (**??**) we use inequality

$$f_i(x) \quad \le \quad f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x} \rangle + \tfrac{L}{2} \| \, x - \bar{x} \, \|^2,$$

for $i = 1, \dots, m$. $\square$

Optimality Conditions

Theorem 1: A point $x^* \in Q$ is a solution to (**??**) if and only if for any $x \in Q$ we have:

$$f(x^*; x) \quad \geq \quad f(x^*; x^*) \quad (\equiv f(x^*)). \tag{4}$$

Proof: Indeed, if (**??**) is true, then for all $x \in Q$

$$f(x) \quad \geq \quad f(x^*; x) \ \geq \ f(x^*; x^*) \ = \ f(x^*).$$

Let x^* be a solution to (**??**). Assume that $\exists x \in Q : \quad f(x^*; x) < f(x^*)$.

Consider the functions $\phi_i(\alpha) = f_i(x^* + \alpha(x - x^*))$, $\alpha \in [0, 1]$.

Note that $f_i(x^*) + \langle f_i'(x^*), x - x^* \rangle < f(x^*) = \max\limits_{1 \leq i \leq m} f_i(x^*)$.

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

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

Therefore, for small enough α we have:

$$f_i(x^* + \alpha(x - x^*)) = \phi_i(\alpha) < f(x^*).$$

This is a contradiction. $\square$

Corollary 1

Let x^* be a minimum of max-type function $f(x)$ on the set Q.

If all components of f belong to $\mathcal{S}^1_\mu(\mathbb{R}^n)$, then

$$f(x) \quad \geq \quad f(x^*) + \tfrac{\mu}{2} \parallel x - x^* \parallel^2$$

for all $x \in Q$.

Proof: Indeed, in view of (**??**) and Theorem 1, for any $x \in Q$ we have:

$$\begin{aligned} f(x) \quad &\geq \quad f(x^*; x) + \tfrac{\mu}{2} \parallel x - x^* \parallel^2 \\ &\geq \quad f(x^*; x^*) + \tfrac{\mu}{2} \parallel x - x^* \parallel^2 \\ &= \quad f(x^*) + \tfrac{\mu}{2} \parallel x - x^* \parallel^2 . \end{aligned}$$

$\square$

Theorem 2

Let $f_i \in \mathcal{S}^1_\mu(\mathbb{R}^n)$, $\mu > 0$, and Q be a closed convex set.

Then the solution x^* of problem (**??**) exists and unique.

Proof: Let $\bar{x} \in Q$. Consider the set $\bar{Q} = \{x \in Q \mid \; f(x) \le f(\bar{x})\}$.

Note that the problem (**??**) is equivalent to the following:

$$\min\{f(x) \mid \; x \in \bar{Q}\}. \tag{5}$$

However, the set $\bar{Q}$ is bounded: $\forall x \in \bar{Q}$

$$f(\bar{x}) \;\ge\; f_i(x) \;\ge\; f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x}\rangle + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2 .$$

Hence, $\frac{\mu}{2} \parallel x - \bar{x} \parallel^2 \le \parallel f_i'(\bar{x}) \parallel \cdot \parallel x - \bar{x} \parallel + f(\bar{x}) - f_i(\bar{x})$.

Thus, the solution x^* of (**??**) ($\equiv$ (**??**)) exists.

If x_1^* is also a solution to (**??**), then

$$\begin{aligned} f^* \;&=\; f(x_1^*) \;\ge\; f(x^*; x_1^*) + \tfrac{\mu}{2} \parallel x_1^* - x^* \parallel^2 \\ &\ge\; f^* + \tfrac{\mu}{2} \parallel x_1^* - x^* \parallel^2 \end{aligned}$$

(we have used (**??**)). Therefore $x_1^* = x^*$. $\square$

Gradient Mapping

Let us fix $\gamma > 0$. For max-type function $f(\cdot)$ denote

$$\begin{aligned} f_\gamma(\bar{x};x) &= f(\bar{x};x) + \frac{\gamma}{2} \parallel x - \bar{x} \parallel^2, \\ f^*(\bar{x};\gamma) &= \min_{x \in Q} f_\gamma(\bar{x};x), \\ x_f(\bar{x};\gamma) &= \arg\min_{x \in Q} f_\gamma(\bar{x};x), \\ g_f(\bar{x};\gamma) &= \gamma(\bar{x} - x_f(\bar{x};\gamma)). \end{aligned}$$

We call $g_f(x;\gamma)$ the *Gradient Mapping* of max-type function f on Q.

Note: 1. If $m = 1$ then that is the standard gradient mapping from Lecture 5.

2. We can use $\bar{x} \notin Q$.

3. Function $f_\gamma(\bar{x};x)$ is a max-type function itself with the components

$$f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x} \rangle + \frac{\gamma}{2} \parallel x - \bar{x} \parallel^2 \quad \in \quad \mathcal{S}^{1,1}_{\gamma,\gamma}(\mathbb{R}^n), \; i = 1, \dots, m.$$

Therefore the Gradient Mapping is well-defined (Theorem 2).

Main Inequality

For a max-type function f, we write $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ if all its components belong to this class.

Theorem 3: Let $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$. Then for all $x \in Q$ we have:

$$f(\bar{x};x) \quad \geq \quad f^*(\bar{x};\gamma) + \frac{1}{2\gamma} \parallel g_f(\bar{x};\gamma) \parallel^2 + \langle g_f(\bar{x};\gamma), x - \bar{x}\rangle. \qquad (6)$$

Proof: Denote $x_f = x_f(\bar{x};\gamma)$, $g_f = g_f(\bar{x};\gamma)$. Note that $f_\gamma(\bar{x};x) \in \mathcal{S}^{1,1}_{\gamma,\gamma}(\mathbb{R}^n)$ and it is a max-type function.

Since $x_f = \arg\min\limits_{x\in Q} f_\gamma(\bar{x};x)$, in view of Corollary 1 and Theorem 1,

$$\begin{aligned}
f(\bar{x};x) &= f_\gamma(\bar{x};x) - \tfrac{\gamma}{2} \parallel x - \bar{x} \parallel^2 \\
&\geq f_\gamma(\bar{x};x_f) + \tfrac{\gamma}{2}(\parallel x - x_f \parallel^2 - \parallel x - \bar{x} \parallel^2) \\
&= f^*(\bar{x};\gamma) + \tfrac{\gamma}{2}\langle \bar{x} - x_f, 2x - x_f - \bar{x}\rangle \\
&= f^*(\bar{x};\gamma) + \tfrac{\gamma}{2}\langle \bar{x} - x_f, 2(x - \bar{x}) + \bar{x} - x_f\rangle \\
&= f^*(\bar{x};\gamma) + \langle g_f, x - \bar{x}\rangle + \tfrac{1}{2\gamma} \parallel g_f \parallel^2. \quad \square
\end{aligned}$$

Corollary 2

Let $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ and $\gamma \geq L$. Denote $x_f = x_f(\bar{x};\gamma)$ and $g_f = g_f(\bar{x};\gamma)$. Then:

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

$$f(x) \quad \geq \quad f(x_f) + \tfrac{1}{2\gamma} \parallel g_f \parallel^2 + \langle g_f, x - \bar{x} \rangle + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2 . \qquad (7)$$

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

$$f(x_f) \quad \leq \quad f(\bar{x}) - \tfrac{1}{2\gamma} \parallel g_f \parallel^2, \qquad (8)$$

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

$$\langle g_f, \bar{x} - x^* \rangle \quad \geq \quad \tfrac{1}{2\gamma} \parallel g_f \parallel^2 + \tfrac{\mu}{2} \parallel x^* - \bar{x} \parallel^2 + \tfrac{\mu}{2} \parallel x^* - x_f \parallel^2 . \qquad (9)$$

Proof: Assumption $\gamma \geq L$ implies $f^*(\bar{x};\gamma) \geq f(x_f)$.

Therefore (**??**) follows from (**??**) since

$$f(x) \quad \geq \quad f(\bar{x};x) + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2$$

for all $x \in \mathbb{R}^n$ (see Lemma 1). Using (**??**) with $x = \bar{x}$, we get (**??**).

(**??**) with $x = x^*$ gives (**??**) since $f(x_f) - f(x^*) \geq \frac{\mu}{2} \parallel x^* - x_f \parallel^2$. □

Lemma 2

For any $\gamma_1,\ \gamma_2 > 0$ and $\bar{x} \in \mathbb{R}^n$ we have:

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

Proof: Denote $x_i = x_f(\bar{x};\gamma_i)$, $g_i = g_f(\bar{x};\gamma_i)$. In view of (**??**), for all $x \in Q$ we have

$$\begin{aligned} f(\bar{x};x) + \tfrac{\gamma_2}{2} \parallel x - \bar{x} \parallel^2 \quad &\geq \quad f^*(\bar{x};\gamma_1) + \langle g_1, x - \bar{x}\rangle \\ &\quad + \tfrac{1}{2\gamma_1} \parallel g_1 \parallel^2 + \tfrac{\gamma_2}{2} \parallel x - \bar{x} \parallel^2 . \end{aligned} \tag{10}$$

In particular, for $x = x_2$ we obtain:

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

□

Gradient Method

Problem: $\min\limits_{x\in Q}\left[f(x) = \max\limits_{1\le i\le m} f_i(x)\right], \;\; f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n), \;\; \mu > 0,$

where Q is a closed convex set.

Scheme: $x_0 \in Q, \; h > 0, \;\; x_{k+1} = x_k - hg_f(x_k; L), \; k = 0, \dots .$

Theorem 4: If we choose $h \le \frac{1}{L}$, then

$$\| x_k - x^* \|^2 \le \left(\frac{1-h\mu}{1+h\mu}\right)^k \| x_0 - x^* \|^2.$$

Proof: Denote $r_k = \| x_k - x^* \|$, $g = g_f(x_k; L)$. Then, in view of (**??**) we have:

$$\begin{aligned} r_{k+1}^2 &= \| x_k - x^* - hg_Q \|^2 \;=\; r_k^2 - 2h\langle g, x_k - x^* \rangle + h^2 \| g \|^2 \\ &\le (1-h\mu)r_k^2 - h\mu r_{k+1}^2 + h\left(h - \tfrac{1}{L}\right) \| g \|^2 \\ &\le (1-h\mu)r_k^2 - h\mu r_{k+1}^2. \end{aligned}$$

□

NB: With $h = \frac{1}{L}$ we have $x_{k+1} = x_k - \frac{1}{L} g_f(x_k; L) = x_f(x_k; L)$.

Optimal Methods

Estimating sequences: $x_0 \in Q, \;\; \phi_0(x) = f(x_0) + \frac{\gamma_0}{2} \parallel x - x_0 \parallel^2,$

$$\begin{aligned} \phi_{k+1}(x) &= (1-\alpha_k)\phi_k(x) + \alpha_k [\boxed{\; f(x_f(y_k; L)) + \frac{1}{2L} \parallel g_f(y_k; L) \parallel^2 } \\ &\quad + \langle g_f(y_k; L), x - y_k \rangle + \frac{\mu}{2} \parallel x - y_k \parallel^2], \\ \phi_k(x) &\equiv \phi_k^* + \frac{\gamma_k}{2} \parallel x - v_k \parallel^2 . \end{aligned}$$

Similarly we get the following updating rules:

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

$$v_{k+1} = \frac{1}{\gamma_{k+1}} [(1-\alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k g_f(y_k; L)],$$

$$\phi_{k+1}^* = (1-\alpha_k)\phi_k + \alpha_k \left(\boxed{ f(x_f(y_k; L)) + \frac{1}{2L} \parallel g_f(y_k; L) \parallel^2 } \right)$$

$$+ \frac{\alpha_k^2}{2\gamma_{k+1}} \parallel g_f(y_k; L) \parallel^2 + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}} \left(\frac{\mu}{2} \parallel y_k - v_k \parallel^2 + \langle g_f(y_k; L), v_k - y_k \rangle \right).$$

Note: In Lecture 5 $\boxed{\cdot}$ was just $f(y_k)$.

Updating rules

Assuming $\phi_k^* \geq f(x_k)$ and using the inequality (**??**) with $x = x_k$ and $\bar{x} = y_k$, we get

$$f(x_k) \quad \geq \quad f(x_f(y_k; L)) + \langle g_f(y_k; L), x_k - y_k \rangle + \tfrac{1}{2L} \parallel g_f(y_k; L) \parallel^2 .$$

We come to the following lower bound:

$$\begin{aligned} \phi_{k+1}^* \quad & \geq \quad (1-\alpha_k) f(x_k) + \alpha_k f(x_f(y_k; L)) + \left(\tfrac{\alpha_k}{2L} - \tfrac{\alpha_k^2}{2\gamma_{k+1}} \right) \parallel g_f(y_k; L) \parallel^2 \\ & \qquad + \tfrac{\alpha_k (1-\alpha_k)\gamma_k}{\gamma_{k+1}} \langle g_f(y_k; L), v_k - y_k \rangle \\ & \\ & \geq \quad f(x_f(y_k; L)) + \left(\tfrac{1}{2L} - \tfrac{\alpha_k^2}{2\gamma_{k+1}} \right) \parallel g_f(y_k; L) \parallel^2 \\ & \qquad + (1-\alpha_k) \langle g_f(y_k; L), \tfrac{\alpha_k \gamma_k}{\gamma_{k+1}} (v_k - y_k) + x_k - y_k \rangle. \end{aligned}$$

Thus, again we can choose

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

Constant Step Scheme (CSS)

0. Choose $x_0 \in Q$ and $\alpha_0 \in (0,1)$. Set $y_0 = x_0$, $q = \mu/L$.

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

   a). Compute $\{f_i(y_k)\}$ and $\{f_i'(y_k)\}$. Set $x_{k+1} = x_f(y_k; L)$.

   b). 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

$$\begin{aligned} \beta_k &= \tfrac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}, \\ y_{k+1} &= x_{k+1} + \beta_k(x_{k+1} - x_k). \end{aligned}$$

Theorem 5

Let the max-type function f belong to $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$.

If in the Constant Step Scheme we take $\alpha_0 \geq \sqrt{\frac{\mu}{L}}$, then

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

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

Scheme for $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ (MaxCSS)

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

$$x_{k+1} = \quad x_f(y_k; L),$$

$$y_{k+1} = \quad x_{k+1} + \beta(x_{k+1} - x_k).$$

Theorem 6: For this scheme we have:

$$f(x_k) - f^* \quad \leq \quad 2\left(1 - \sqrt{\frac{\mu}{L}}\right)^k (f(x_0) - f^*). \tag{11}$$

Proof: This scheme corresponds to $\alpha_0 = \sqrt{\frac{\mu}{L}}$.

Then $\gamma_0 = \mu$ and we get (**??**) since

$$f(x_0) \quad \geq \quad f^* + \frac{\mu}{2} \| x_0 - x^* \|^2$$

in view of Corollary 1. □

Auxiliary Problem

Problem

$$\min_{x \in Q} \left\{ \max_{1 \le i \le m} [f_i(x_0) + \langle f_i'(x_0), x - x_0 \rangle] + \frac{\gamma}{2} \parallel x - x_0 \parallel^2 \right\}$$

is equivalent to the following:

$$\begin{aligned} &\min_{x \in Q, t \in \mathbb{R}} \{ t + \tfrac{\gamma}{2} \parallel x - x_0 \parallel^2 : \\ &f_i(x_0) + \langle f_i'(x_0), x - x_0 \rangle \le t, \ i = 1, \dots, m \}. \end{aligned} \tag{12}$$

Note: If Q is a polytope then

1. (**??**) is a Quadratic Programming Problem.

2. There are finite methods for solving (**??**).

3. This problem can be also solved by Interior Point Methods (then we can consider more complicated Q).

Optimization with Functional Constraints

Problem:

$$\min_{x \in Q} \{ \; f_0(x) : \; f_i(x) \leq 0, \; i = 1, \ldots, m \}, \tag{13}$$

where $f_i \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n), \quad i = 0, \ldots, m,$ with $\mu > 0$, and Q is a closed convex set.

Consider the *parametric* max-type function

$$\begin{aligned} f(t;x) &= \max\{f_0(x) - t; \; f_i(x), i = 1 \ldots m\}, \\ f^*(t) &= \min_{x \in Q} \; f(t;x). \end{aligned}$$

The solution of this problem $x^*(t)$ exists and it is unique (Theorem 2).

We will try to get close to the solution of (**??**) using a process based on *approximate values* of function $f^*(t)$.

Note:

1. This approach is called *Sequential Quadratic Programming*.
2. It can be applied also to nonconvex problems.

Properties of function $f^*(t)$

Lemma 3: Let t^* be the optimal value of the problem (**??**). Then

$$f^*(t) \leq 0 \quad \forall t \geq t^* \quad \text{and} \quad f^*(t) > 0 \quad \forall t < t^*.$$

Proof: Let x^* be a solution to (**??**). If $t \geq t^*$, then

$$f^*(t) \leq f(t; x^*) = \max_{1 \leq i \leq m} \{f_0(x^*) - t; f_i(x^*)\} = \max_{1 \leq i \leq m} \{t^* - t; f_i(x^*)\} \leq 0.$$

Suppose that $t < t^*$ and $f^*(t) \leq 0$. Then

$$\exists y \in Q : \quad f_0(y) \leq t, \quad f_i(y) \leq 0.$$

Thus, $t^* > t$ is not the optimal value of (**??**). □

Lemma 4: For any $\Delta \geq 0$ we have: $f^*(t) \geq f^*(t + \Delta) \geq f^*(t) - \Delta$.

Proof: Indeed,

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

Lemma 5

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

$$f^*(t_1 - \Delta) \quad \geq \quad f^*(t_1) + \Delta \frac{f^*(t_1) - f^*(t_2)}{t_2 - t_1}. \tag{14}$$

Proof: Denote $t_0 = t_1 - \Delta$ and $\alpha = \frac{\Delta}{t_2 - t_0} \equiv \frac{\Delta}{t_2 - t_1 + \Delta} \in [0, 1]$.

Then $t_1 = (1-\alpha) t_0 + \alpha t_2$ and (**??**) can be written as

$$f^*(t_1) \quad \leq \quad (1-\alpha) f^*(t_0) + \alpha f^*(t_2). \tag{15}$$

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

$$\begin{aligned}
f^*(t_1) \quad & \leq \max_{1 \leq i \leq m} \{ \quad f_0(x_\alpha) - t_1; \quad f_i(x_\alpha) \ \} \\
& \leq \max_{1 \leq i \leq m} \{ \quad (1-\alpha)(f_0(x^*(t_0)) - t_0) + \alpha(f_0(x^*(t_2)) - t_2); \\
& \qquad\qquad\quad (1-\alpha) f_i(x^*(t_0)) + \alpha f_i(x^*(t_2)) \} \\
& \leq (1-\alpha) \quad \max_{1 \leq i \leq m} \{ f_0(x^*(t_0)) - t_0; f_i(x^*(t_0)) \} \\
& \qquad\quad + \alpha \quad \max_{1 \leq i \leq m} \{ f_0(x^*(t_2)) - t_2; f_i(x^*(t_2)) \} \\
& = (1-\alpha) \quad f^*(t_0) + \alpha f^*(t_2),
\end{aligned}$$

and we get (**??**). □

Note: Lemmas 4 and 5 are valid for *any* parametric max-type functions.

Gradient Mapping

The *linearization* of the parametric max-type function $f(t;\cdot)$ is

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

Let us fix $\gamma > 0$. Denote $f_\gamma(t;\bar{x};x) = f(t;\bar{x};x) + \frac{\gamma}{2} \parallel x-\bar{x} \parallel^2$, and

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

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

We call $g_f(t;\bar{x};\gamma)$ the *Constrained* Gradient Mapping of problem (**??**).

Note: 1. We can use $\bar{x} \notin Q$.

2. $f_\gamma(t;\bar{x};\cdot)$ is a max-type function with components

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

$$f_i(\bar{x}) + \langle f_i'(\bar{x}), x-\bar{x}\rangle - t + \frac{\gamma}{2} \parallel x-\bar{x} \parallel^2,\ i = 1,\ldots.m.$$

Therefore $f_\gamma(t;\bar{x};x) \in \mathcal{S}^{1,1}_{\gamma,\gamma}(\mathbb{R}^n)$ and the Gradient Mapping is well defined (Theorem 2).

Main Properties

$f(t;x) \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, we have $\quad f_\mu(t;\bar{x};x) \leq f(t;x) \leq f_L(t;\bar{x};x)$.

Therefore $f^*(t;\bar{x};\mu) \leq f^*(t) \leq f^*(t;\bar{x};L)$.

Using Lemma 5, we obtain the following result:

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

$$f^*(t_1 - \Delta;\bar{x};\gamma) \geq f^*(t_1;\bar{x};\gamma) + \tfrac{\Delta}{t_2 - t_1}(f^*(t_1;\bar{x};\gamma) - f^*(t_2;\bar{x};\gamma)). \tag{16}$$

Important cases: $\quad \gamma = L, \quad$ and $\quad \gamma = \mu$.

Using Lemma 2, we obtain:

$$f^*(t;\bar{x};\mu) \quad \geq \quad f^*(t;\bar{x};L) - \tfrac{L-\mu}{2\mu L} \parallel g_f(t;\bar{x};L) \parallel^2 . \tag{17}$$

Efficiency of the step in t

Denote $\quad t^*(\bar{x}, t) = \mathrm{root}_t(f^*(t; \bar{x}; \mu))$.

Lemma 6: Let $\bar{x} \in \mathbb{R}^n$ and $\bar{t} < t^*$ are such that for some $\kappa \in (0,1)$

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

Then 1. $\bar{t} < t^*(\bar{x}, \bar{t}) \leq t^*$. 2. For any $t < \bar{t}$ and $x \in \mathbb{R}^n$ we have:

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

Proof: 1. Since $\bar{t} < t^*$, we have:

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

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

2. Denote $\Delta = \bar{t} - t$. Then, in view of (**??**), we have:

$$f^*(t; x; L) \;\geq\; f^*(t) \geq f^*(t; \bar{x}; \mu) \quad \geq\; f^*(\bar{t}; \bar{x}; \mu) + \tfrac{\Delta}{t^*(\bar{x}, \bar{t}) - \bar{t}} f^*(\bar{t}; \bar{x}; \mu)$$

$$\geq\; (1-\kappa)\left(1 + \tfrac{\Delta}{t^*(\bar{x}, \bar{t}) - \bar{t}}\right) f^*(\bar{t}; \bar{x}; L) \;\geq\; 2(1-\kappa) f^*(\bar{t}; \bar{x}; L) \sqrt{\tfrac{\Delta}{t^*(\bar{x}, \bar{t}) - \bar{t}}}.$$

$\square$

Constrained Minimization Scheme (CMS)

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

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

   a). Generate the sequence $\{x_{k,j}\}$ by MaxCSS as applied to $f(t_k; x)$ with $x_{k,0} = x_k$.

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

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

   Global Stop: Terminate if $f^*(t_k; x_{k,j}; L) \leq \epsilon$.

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

NB: We are interested in the *analytical complexity* of CMS. Therefore:

1. Complexity of computing $t^*(x, t)$ and $f^*(t; x, \gamma)$ is not important now.

2. We need to estimate the rate of convergence of the main process in CMS (the *Master Process*).

3. We need to estimate the complexity of Step 1a).

Master Process

Lemma 7: $f^*(t_k; x_{k+1}; L) \le \frac{t_0 - t^*}{1-\kappa} \left[\frac{1}{2(1-\kappa)}\right]^k.$

Proof: Denote $\delta_k = \frac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1}-t_k}}$, and $\beta = \frac{1}{2(1-\kappa)} \quad (< 1)$.

Since $t_{k+1} = t^*(x_{k,j(k)}, t_k)$, in view of Lemma 6, for $k \ge 1$ we have:

$$2(1-\kappa)\frac{f^*(t_k; x_{k,j(k)}; L)}{\sqrt{t_{k+1}-t_k}} \quad \le \quad \frac{f^*(t_{k-1}; x_{k-1,j(k-1)}; L))}{\sqrt{t_k - t_{k-1}}}.$$

Thus, $\delta_k \le \beta \delta_{k-1}$ and we obtain

$$\begin{aligned} f^*(t_k; x_{k,j(k)}; L) &= \delta_k \sqrt{t_{k+1}-t_k} \le \beta^k \delta_0 \sqrt{t_{k+1}-t_k} \\ &= \beta^k f^*(t_0; x_{0,j(0)}; L)\sqrt{\frac{t_{k+1}-t_k}{t_1 - t_0}}. \end{aligned}$$

Further, in view of Lemma 4, $t_1 - t_0 \ge f^*(t_0; x_{0,j(0)}; \mu)$. Therefore

$$\begin{aligned} f^*(t_k; x_{k,j(k)}; L) &\le \beta^k f^*(t_0; x_{0,j(0)}; L)\sqrt{\frac{t_{k+1}-t_k}{f^*(t_0; x_{0,j(0)}; \mu)}} \\ &\le \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0; x_{0,j(0)}; \mu)(t_{k+1}-t_k)} \\ &\le \frac{\beta^k}{1-\kappa}\sqrt{f^*(t_0)(t_0 - t^*)}. \end{aligned}$$

It remains to note that $f^*(t_0) \le t_0 - t^*$ (Lemma 4) and
$f^*(t_k; x_{k+1}; L) \equiv f^*(t_k; x_{k,j^*(k)}; L) \le f^*(t_k; x_{k,j(k)}; L)$. $\square$

Complexity Analysis

Let $f^*(t_k; x_{k,j}; L) \leq \epsilon$. Then for $x_* = x_f(t_k; x_{k,j}; L)$ we have:

$$f(t_k; x_*) \;=\; \max_{1 \leq i \leq m} \{ f_0(x_*) - t_k; f_i(x_*) \} \;\leq\; f^*(t_k; x_{k,j}; L) \;\leq\; \epsilon.$$

Since $t_k \leq t^*$, we conclude that

$$f_0(x_*) \;\leq t^* + \epsilon, \quad f_i(x_*) \;\leq\; \epsilon,\; i = 1, \ldots, m. \tag{18}$$

In view of Lemma 7, we can get (**??**) at most in

$$N(\epsilon) \;=\; \tfrac{1}{\ln[2(1-\kappa)]} \ln \tfrac{t_0 - t^*}{(1-\kappa)\epsilon}$$

full iterations of the master process.

(The last iteration of the process is terminated by the Global Stop rule).

Note:

κ is an absolute constant (for example, $\kappa = \frac{1}{4}$).

Internal Process

Let the sequence $\{x_{k,j}\}$ be generated by MaxCSS with starting point $x_{k,0} = x_k$.

In view of Theorem 6, denoting $\sigma = \sqrt{\frac{\mu}{L}}$, we have:

$$
\begin{aligned}
f(t_k; x_{k,j}) - f^*(t_k) &\le 2\left(1 - \sqrt{\tfrac{\mu}{L}}\right)^j (f(t_k; x_k) - f^*(t_k)) \\
&\le 2e^{-\sigma \cdot j}(f(t_k; x_k) - f^*(t_k)) \\
&\le 2e^{-\sigma \cdot j} f(t_k; x_k).
\end{aligned}
$$

Denote by N the number of full iterations of CMS $(N \le N(\epsilon))$.

Thus, $j(k)$ is defined for all k, $0 \le k \le N$.

Note that $t_k = t^*(x_{k-1,j(k-1)}, t_{k-1}) > t_{k-1}$. Therefore

$$f(t_k; x_k) \le f(t_{k-1}; x_k) \le f^*(t_{k-1}; x_{k-1,j^*(k-1)}, L).$$

Denote $\Delta_0 = f(t_0; x_0)$, $\Delta_k = f^*(t_{k-1}; x_{k-1,j^*(k-1)}, L)$, $k \ge 1$.

Then, for all $k \ge 0$ we have: $f(t_k; x_k) - f^*(t_k) \le \Delta_k$.

Lemma 8

For all k, $0 \leq k \leq N$, the internal process is terminated when the following condition is satisfied:

$$f(t_k; x_{k,j}) - f^*(t_k) \quad \leq \quad \tfrac{\mu\kappa}{L-\mu} \cdot f^*(t_k; x_{k,j}; L). \tag{19}$$

Proof: Assume that (**??**) is satisfied. Then, in view of (**??**), we have:

$$\begin{aligned} \tfrac{1}{2L} \parallel g_f(t_k; x_{k,j}; L \parallel^2 \quad &\leq \quad f(t_k; x_{k,j}) - f(t_k; x_f(t_k; x_{k,j}; L)) \\ &\leq \quad f(t_k; x_{k,j}) - f^*(t_k). \end{aligned}$$

Therefore, using (**??**), we obtain:

$$\begin{aligned} f^*(t_k; x_{k,j}; \mu) \quad &\geq \quad f^*(t_k; x_{k,j}; L) - \tfrac{L-\mu}{2\mu L} \parallel g_f(t_k; x_{k,j}; L \parallel^2 \\ &\geq \quad f^*(t_k; x_{k,j}; L) - \tfrac{L-\mu}{\mu}(f(t_k; x_{k,j}) - f^*(t_k)) \\ &\overset{(\mathbf{??})}{\geq} \quad (1-\kappa) f^*(t_k; x_{k,j}; L). \end{aligned}$$

This is the termination criterion of the internal process in Step 1a) of CMS. $\square$

Lemma 9

For all k, $0 \leq k \leq N$, we have: $j(k) \leq 1 + \sqrt{\frac{L}{\mu}} \cdot \ln \frac{2(L-\mu)\Delta_k}{\kappa\mu\Delta_{k+1}}$.

Proof: Assume that $j(k) - 1 > \frac{1}{\sigma} \ln \frac{2(L-\mu)\Delta_k}{\kappa\mu\Delta_{k+1}}$, where $\sigma = \sqrt{\frac{\mu}{L}}$.

Recall that $\Delta_{k+1} = \min\limits_{0 \leq j \leq j(k)} f^*(t_k; x_{k,j}; L)$.

Note that the stopping criterion of the internal process did not work for $j = j(k) - 1$.

Therefore, in view of Lemma 8,

$$\begin{array}{rcl} f^*(t_k; x_{k,j}; L) & \leq & \frac{L-\mu}{\mu\kappa}(f(t_k; x_{k,j}) - f^*(t_k)) \\ & \leq & 2\frac{L-\mu}{\mu\kappa} e^{-\sigma \cdot j} \Delta_k \ < \Delta_{k+1}. \end{array}$$

This is a contradiction. □

Corollary 3:

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

□

Last Iteration

Let j^* be the last step in the internal minimization process of $(N+1)$st iteration of the Master Process.

Lemma 10: $j^* \leq 1 + \sqrt{\frac{L}{\mu}} \cdot \ln \frac{2(L-\mu)\Delta_{N+1}}{\kappa\mu\epsilon}$.

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

Note that for $j = j^* - 1$ we have:

$$\begin{aligned} \epsilon &\leq f^*(t_{N+1}; x_{N+1,j}; L) \\ &\leq \tfrac{L-\mu}{\mu\kappa}\left(f(t_{N+1}; x_{N+1,j}) - f^*(t_{N+1})\right) \\ &\leq 2\tfrac{L-\mu}{\mu\kappa} e^{-\sigma \cdot j} \Delta_{N+1} \;<\; \epsilon. \end{aligned}$$

This is a contradiction. □

Corollary 4:

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

Total Complexity

Thus, we have proved the following upper bound for the total number of internal iterations in CMS:

$$\left[\frac{1}{\ln[2(1-\kappa)]}\ln\frac{t_0-t^*}{(1-\kappa)\epsilon}+2\right]\cdot\left[1+\sqrt{\frac{L}{\mu}}\cdot\ln\frac{2(L-\mu)}{\kappa\mu}\right]$$

$$+\sqrt{\frac{L}{\mu}}\cdot\ln\left(\frac{1}{\epsilon}\cdot\max_{1\le i\le m}\{f_0(x_0)-t_0;f_i(x_0)\}\right).$$

The principal term in this estimate is of the order $\boxed{\ln\frac{t_0-t^*}{\epsilon}\cdot\sqrt{\frac{L}{\mu}}\cdot\ln\frac{L}{\mu}}$.

It differs from the *lower bound* for the unconstrained minimization problem by a factor of $\ln\frac{L}{\mu}$.

Conclusion: CMS is *suboptimal*.

What is $t^*(\bar{x}, t)$?

Recall that $t^*(\bar{x}, t)$ is the root 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 by components

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

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

In view of Lemma 3, it is the optimal value of the following minimization problem:

$$\min_{x \in Q} \; f_0(\bar{x}) + \langle f_0'(\bar{x}), x - \bar{x} \rangle + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2,$$

$$\text{s.t.} \quad f_i(\bar{x}) + \langle f_i'(\bar{x}), x - \bar{x} \rangle + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2 \leq 0,$$

$$i = 1, \ldots, m.$$

This problem can be solved, for example, by Interior-Point Methods.