Lecture 3.

Optimal Methods
Minimization over Simple Sets

– Optimal Methods.

– Convex Sets.

– Constrained Minimization Problem.

– Gradient Mapping.

– Minimization Methods over a simple set.

Optimal Methods

Problem:

$$\min_{x \in R^n} f(x), \quad f \in \mathcal{S}^{1,1}_{\mu,L}(R^n).$$

We allow the value $\mu = 0$ ($\mathcal{S}^{1,1}_{0,L}(R^n) \equiv \mathcal{F}^{1,1}_L(R^n)$).

Gradient Method:

$$f \in \mathcal{F}^{1,1}_L(R^n) \;\Rightarrow\; f(x_k) - f^* \le \frac{2L\|x_0 - x^*\|^2}{k+4},$$

$$f \in \mathcal{S}^{1,1}_{\mu,L}(R^n) \;\Rightarrow\; f(x_k) - f^* \le \frac{L\|x_0 - x^*\|^2}{2\left(\frac{L+\mu}{L-\mu}\right)^{2k}}.$$

It is not optimal!

Note:

1. The gradient method forms a relaxation sequence:

$$f(x_{k+1}) \le f(x_k).$$

2. Optimal methods never rely on that. Relaxation is too expensive for optimality.

Estimate sequence

Definition. A pair of sequences $\{\phi_k(x)\}_{k=0}^{\infty}$ and $\{\lambda_k\}_{k=0}^{\infty}$, $\lambda_k \geq 0$ is called an *estimate sequence* of function $f(x)$ if $\lambda_k \to 0$ and for any $x \in R^n$ and $k \geq 0$ we have:

$$\phi_k(x) \leq (1 - \lambda_k) f(x) + \lambda_k \phi_0(x). \qquad (3.1)$$

Lemma 3.1 *If for some sequence* $\{x_k\}$ *we have*

$$f(x_k) \leq \phi_k^* \equiv \min_{x \in R^n} \phi_k(x) \qquad (3.2)$$

then

$$f(x_k) - f^* \leq \lambda_k [\phi_0(x^*) - f^*] \to 0.$$

Proof. Indeed,

$$\begin{aligned} f(x_k) &\leq \phi_k^* = \min_{x \in R^n} \phi_k(x) \\ &\leq \min_{x \in R^n} [(1 - \lambda_k) f(x) + \lambda_k \phi_0(x)] \\ &\leq (1 - \lambda_k) f(x^*) + \lambda_k \phi_0(x^*). \end{aligned}$$

□

Thus, the rate of convergence of $\{\lambda_k\}$ defines the rate of convergence of the sequence $\{x_k\}$.

Questions:

1. How to form the estimate sequence?

2. How we can ensure (3.2)?

Lemma 3.2 *Let us assume that:*

1. $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$.

2. $\phi_0(x)$ *is an arbitrary function on* R^n.

3. $\{y_k\}_{k=0}^{\infty}$ *is an arbitrary sequence in* R^n.

4. $\{\alpha_k\}_{k=0}^{\infty}: \ \alpha_k \in (0,1), \quad \sum\limits_{k=0}^{\infty} \alpha_k = \infty.$

5. $\lambda_0 = 1$.

Then the pair $\{\phi_k(x)\}_{k=0}^{\infty}$, $\{\lambda_k\}_{k=0}^{\infty}$:

$$\left.\begin{aligned} \lambda_{k+1} &= (1-\alpha_k)\lambda_k, \\ \phi_{k+1}(x) &= (1-\alpha_k)\phi_k(x) + \alpha_k[f(y_k) \\ &\quad + \langle f'(y_k), x - y_k\rangle + \tfrac{\mu}{2} \parallel x - y_k \parallel^2], \end{aligned}\right\} \quad (3.3)$$

is an estimate sequence.

Proof. Indeed,

$$\phi_0(x) \le (1-\lambda_0)f(x) + \lambda_0\phi_0(x) \equiv \phi_0(x).$$

Further, let (3.1) holds for some $k \ge 0$. Then

$$\begin{aligned} \phi_{k+1}(x) \le\ & (1-\alpha_k)\phi_k(x) + \alpha_k f(x) \\ =\ & (1-(1-\alpha_k)\lambda_k)f(x) \\ & +(1-\alpha_k)(\phi_k(x) - (1-\lambda_k)f(x)) \\ \le\ & (1-(1-\alpha_k)\lambda_k)f(x) + (1-\alpha_k)\lambda_k\phi_0(x) \\ =\ & (1-\lambda_{k+1})f(x) + \lambda_{k+1}\phi_0(x). \end{aligned}$$

□

Variation of ϕ_k^*

Lemma 3.3 *Let* $\phi_0(x) = \phi_0^* + \frac{\gamma_0}{2} \parallel x - v_0 \parallel^2$. *Then the process (3.3) forms* $\phi_k(x) \equiv \phi_k^* + \frac{\gamma_k}{2} \parallel x - v_k \parallel^2$, *where the sequences* $\{\gamma_k\}$, $\{v_k\}$ *and* $\{\phi_k^*\}$ *are defined as follows:*

$$\begin{aligned}
\gamma_{k+1} &= (1-\alpha_k)\gamma_k + \alpha_k\mu, \\
v_{k+1} &= \tfrac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k f'(y_k)], \\
\phi_{k+1}^* &= (1-\alpha_k)\phi_k^* + \alpha_k f(y_k) - \tfrac{\alpha_k^2}{2\gamma_{k+1}} \parallel f'(y_k) \parallel^2 \\
&\quad + \tfrac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}} \left(\tfrac{\mu}{2} \parallel y_k - v_k \parallel^2 + \langle f'(y_k), v_k - y_k \rangle \right).
\end{aligned}$$

Proof. Note that $\phi_0''(x) = \gamma_0 I_n$. Therefore

$$\begin{aligned}
\phi_{k+1}''(x) &= (1-\alpha_k)\phi_k''(x) + \alpha_k \mu I_n \\
&= ((1-\alpha_k)\gamma_k + \alpha_k\mu) I_n \equiv \gamma_{k+1} I_n.
\end{aligned}$$

Further,

$$\begin{aligned}
&\phi_{k+1}(x) = (1-\alpha_k)\left(\phi_k^* + \tfrac{\gamma_k}{2} \parallel x - v_k \parallel^2\right) \\
&+\alpha_k[f(y_k) + \langle f'(y_k), x - y_k \rangle + \tfrac{\mu}{2} \parallel x - y_k \parallel^2].
\end{aligned}$$

Therefore the equation $\phi_{k+1}'(x) = 0$ is as follows:

$$(1-\alpha_k)\gamma_k(x - v_k) + \alpha_k f'(y_k) + \alpha_k\mu(x - y_k) = 0,$$

And we get the equation for v_{k+1}.

Finally, let us compute ϕ^*_{k+1}. We have:

$$\begin{aligned} &\phi^*_{k+1} + \tfrac{\gamma_{k+1}}{2} \parallel v_{k+1} - y_k \parallel^2 \\ &= (1-\alpha_k)\left(\phi^*_k + \tfrac{\gamma_k}{2} \parallel y_k - v_k \parallel^2\right) + \alpha_k f(y_k). \end{aligned} \qquad (3.4)$$

Note that

$$v_{k+1} - y_k = \tfrac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k(v_k - y_k) - \alpha_k f'(y_k)].$$

Therefore

$$\begin{aligned} &\tfrac{\gamma_{k+1}}{2} \parallel v_{k+1} - y_k \parallel^2 = \tfrac{1}{2\gamma_{k+1}}[(1-\alpha_k)^2\gamma_k^2 \parallel v_k - y_k \parallel^2 \\ &-2\alpha_k(1-\alpha_k)\gamma_k\langle f'(y_k), v_k - y_k\rangle + \alpha_k^2 \parallel f'(y_k) \parallel^2]. \end{aligned}$$

It remains to substitute this relation in (3.4).

Note that the coefficient for $\parallel y_k - v_k \parallel^2$ is as follows:

$$\begin{aligned} &(1-\alpha_k)\tfrac{\gamma_k}{2} - \tfrac{1}{2\gamma_{k+1}}(1-\alpha_k)^2\gamma_k^2 \\ &= (1-\alpha_k)\tfrac{\gamma_k}{2}\left(1 - \tfrac{(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\right) = (1-\alpha_k)\tfrac{\gamma_k}{2} \cdot \tfrac{\alpha_k \mu}{\gamma_{k+1}}. \end{aligned}$$

□

Let we have x_k: $\phi_k^* \geq f(x_k)$. Then

$$\phi_{k+1}^* \geq \; (1-\alpha_k)f(x_k) + \alpha_k f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}} \parallel f'(y_k) \parallel^2$$
$$+\frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle f'(y_k), v_k - y_k\rangle.$$

Since $f(x_k) \geq f(y_k) + \langle f'(y_k), x_k - y_k\rangle$, we get:

$$\phi_{k+1}^* \geq \; f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}} \parallel f'(y_k) \parallel^2$$
$$+(1-\alpha_k)\langle f'(y_k), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k\rangle.$$

We want to have $\phi_{k+1}^* \geq f(x_{k+1})$. Note that:

1. By a gradient step $x_{k+1} = y_k - h_k f'(y_k)$ we can guarantee that

$$f(y_k) - \frac{1}{2L} \parallel f'(y_k) \parallel^2 \geq f(x_{k+1})$$

(for example, $h_k = \frac{1}{L}$). This gives the following equation for α_k:

$$L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu \quad (= \gamma_{k+1}).$$

2. We can kill the second term choosing y_k from the equation

$$\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k = 0.$$

That is

$$y_k = \frac{\alpha_k\gamma_k v_k + \gamma_{k+1}x_k}{\gamma_k + \alpha_k\mu}.$$

General scheme (3.5)

0. Choose $x_0 \in R^n$ and $\gamma_0 > 0$.

   Set $v_0 = x_0$.

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

   a). Compute $\alpha_k \in (0, 1)$ from the equation

   $$L\alpha_k^2 = (1 - \alpha_k)\gamma_k + \alpha_k \mu.$$

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

   b). Choose

   $$y_k = \frac{\alpha_k \gamma_k v_k + \gamma_{k+1} x_k}{\gamma_k + \alpha_k \mu}.$$

   Compute $f(y_k)$ and $f'(y_k)$.

   c). Find $x_{k+1} = y_k - h_k f'(y_k)$ such that

   $$f(x_{k+1}) \leq f(y_k) - \frac{1}{2L} \parallel f'(y_k) \parallel^2$$

   d). Set

   $$v_{k+1} = \frac{1}{\gamma_{k+1}}[(1 - \alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k f'(y_k)].$$

Remark:

In Step 1c) of the scheme we can choose any x_{k+1} such that

$$f(x_{k+1}) \leq f(y_k) - \frac{\omega}{2} \parallel f'(y_k) \parallel^2 .$$

Then the constant $\frac{1}{\omega}$ should replace L in the equation of Step 1a).

Theorem 3.1 *The scheme (3.5) generates a sequence $\{x_k\}_{k=0}^{\infty}$ such that*

$$f(x_k) - f^* \leq \lambda_k \left[f(x_0) - f^* + \frac{\gamma_0}{2} \parallel x_0 - x^* \parallel^2 \right],$$

where $\lambda_0 = 1$ and

$$\lambda_k = \Pi_{i=0}^{k-1}(1 - \alpha_i).$$

Proof.. Indeed, let us choose

$$\phi_0(x) = f(x_0) + \frac{\gamma_0}{2} \parallel x - v_0 \parallel^2 .$$

Then

$$f(x_0) = \phi_0^*$$

and we get $f(x_k) \leq \phi_k^*$ by construction of the scheme.

It remains to use Lemma 3.1. □

Lemma 3.4 *If we take $\gamma_0 \geq \mu$, then*

$$\lambda_k \leq \min \left\{ \left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{(2\sqrt{L} + k\sqrt{\gamma_0})^2} \right\}.$$

Proof. Indeed, if $\gamma_k \geq \mu$ then

$$L\alpha_k^2 = (1 - \alpha_k)\gamma_k + \alpha_k \mu \geq \mu.$$

Hence, $\alpha_k \geq \sqrt{\frac{\mu}{L}}$ and we get also

$$\gamma_{k+1} = L\alpha_k^2 \geq \mu.$$

Further, let us prove that $\gamma_k \geq \gamma_0 \lambda_k$. Indeed, since

$$\gamma_0 = \gamma_0 \lambda_0,$$

we can use induction:

$$\gamma_{k+1} \geq (1 - \alpha_k)\gamma_k \geq (1 - \alpha_k)\gamma_0 \lambda_k = \gamma_0 \lambda_{k+1}.$$

Therefore

$$L\alpha_k^2 = \gamma_{k+1} \geq \gamma_0 \lambda_{k+1}.$$

Denote $a_k = \frac{1}{\sqrt{\lambda_k}}$. Since $\{\lambda_k\}$ decrease, we have:

$$a_{k+1} - a_k = \frac{\sqrt{\lambda_k} - \sqrt{\lambda_{k+1}}}{\sqrt{\lambda_k \lambda_{k+1}}} = \frac{\lambda_k - \lambda_{k+1}}{\sqrt{\lambda_k \lambda_{k+1}}(\sqrt{\lambda_k} + \sqrt{\lambda_{k+1}})}$$

$$\geq \frac{\lambda_k - \lambda_{k+1}}{2\lambda_k \sqrt{\lambda_{k+1}}} = \frac{\lambda_k - (1 - \alpha_k)\lambda_k}{2\lambda_k \sqrt{\lambda_{k+1}}} = \frac{\alpha_k}{2\sqrt{\lambda_{k+1}}} \geq \frac{1}{2}\sqrt{\frac{\gamma_0}{L}}.$$

Thus, $a_k \geq 1 + \frac{k}{2}\sqrt{\frac{\gamma_0}{L}}$. □

Theorem 3.2 *Let us take in (3.5)* $\gamma_0 = L$*. Then this scheme generates a sequence* $\{x_k\}_{k=0}^{\infty}$ *such that*

$$f(x_k) - f^* \leq L \min\left\{\left(1 - \sqrt{\tfrac{\mu}{L}}\right)^k, \tfrac{4}{(k+2)^2}\right\} \parallel x_0 - x^* \parallel^2 .$$

This means that it is optimal *for the class* $\mathcal{S}^{1,1}_{\mu,L}(R^n)$ *with* $\mu \geq 0$*.*

Proof. We get the above inequality using

$$f(x_0) - f^* \leq \tfrac{L}{2} \parallel x_0 - x^* \parallel^2$$

and the previous results.

Further, from the lower complexity bounds for the class $\mathcal{S}^{1,1}_{\mu,L}(R^n)$, $\mu > 0$, we have:

$$f(x_k) - f^* \geq \tfrac{\mu}{2}\left(\tfrac{\sqrt{Q}-1}{\sqrt{Q}+1}\right)^{2k} R^2 \geq \tfrac{\mu}{2}\exp\left(-\tfrac{4k}{\sqrt{Q}-1}\right) R^2$$

where $Q = L/\mu$ and $R = \parallel x_0 - x^* \parallel$. Therefore, the worst case estimate for finding x_k: $f(x_k) - f^* \leq \epsilon$ cannot be better than

$$k \geq \tfrac{\sqrt{Q}-1}{4}\left[\ln\tfrac{1}{\epsilon} + \ln\tfrac{\mu}{2} + 2\ln R\right].$$

For our scheme we have:

$$f(x_k) - f^* \leq LR^2\left(1 - \sqrt{\tfrac{\mu}{L}}\right)^k \leq LR^2\exp\left(-\tfrac{k}{\sqrt{Q}}\right).$$

Therefore we guarantee that

$$k \leq \sqrt{Q}\left[\ln\tfrac{1}{\epsilon} + \ln L + 2\ln R\right].$$

Thus, the main term in this estimate, $\sqrt{Q}\ln\frac{1}{\epsilon}$, is proportional to the lower bound.

The same reasoning can be used for $\mathcal{S}^{1,1}_{0,L}(R^n)$. □

Constant Step Scheme (3.6)

0. Choose $x_0 \in R^n$ and $\gamma_0 > 0$.

   Set $v_0 = x_0$.

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

   a). Compute $\alpha_k \in (0, 1)$ from the equation

   $$L\alpha_k^2 = (1 - \alpha_k)\gamma_k + \alpha_k \mu.$$

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

   b). Choose

   $$y_k = \frac{\alpha_k \gamma_k v_k + \gamma_{k+1} x_k}{\gamma_k + \alpha_k \mu}.$$

   Compute $f(y_k)$ and $f'(y_k)$.

   c). Set

   $$x_{k+1} = y_k - \frac{1}{L} f'(y_k).$$

   $$v_{k+1} = \frac{1}{\gamma_{k+1}}[(1 - \alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k f'(y_k)].$$

Let us simplify this scheme.

Note that:

$$y_k = \frac{\alpha_k\gamma_k v_k + \gamma_{k+1}x_k}{\gamma_k + \alpha_k\mu}, \quad x_{k+1} = y_k - \tfrac{1}{L}f'(y_k),$$

$$v_{k+1} = \frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k v_k + \alpha_k\mu y_k - \alpha_k f'(y_k)].$$

Therefore

$$\begin{aligned} v_{k+1} = & \tfrac{1}{\gamma_{k+1}}\{\tfrac{(1-\alpha_k)}{\alpha_k}[(\gamma_k + \alpha_k\mu)y_k - \gamma_{k+1}x_k] \\ & +\alpha_k\mu y_k - \alpha_k f'(y_k)\} \\ = & \tfrac{1}{\gamma_{k+1}}\left\{\tfrac{(1-\alpha_k)\gamma_k}{\alpha_k}y_k + \mu y_k\right\} - \tfrac{1-\alpha_k}{\alpha_k}x_k - \tfrac{\alpha_k}{\gamma_{k+1}}f'(y_k) \\ = & x_k + \tfrac{1}{\alpha_k}(y_k - x_k) - \tfrac{1}{\alpha_k L}f'(y_k) \\ = & x_k + \tfrac{1}{\alpha_k}(x_{k+1} - x_k). \end{aligned}$$

Hence,

$$\begin{aligned} y_{k+1} = & \tfrac{1}{\gamma_{k+1}+\alpha_{k+1}\mu}(\alpha_{k+1}\gamma_{k+1}v_{k+1} + \gamma_{k+2}x_{k+1}) \\ = & x_{k+1} + \tfrac{\alpha_{k+1}\gamma_{k+1}(v_{k+1}-x_{k+1})}{\gamma_{k+1}+\alpha_{k+1}\mu} = x_{k+1} + \beta_k(x_{k+1} - x_k). \end{aligned}$$

where

$$\beta_k = \frac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1} + \alpha_{k+1}\mu)}.$$

Thus, we managed to eliminate $\{v_k\}$.

What can we say about β_k?

We have:

$$\alpha_k^2 L = (1-\alpha_k)\gamma_k + \mu\alpha_k \equiv \gamma_{k+1}.$$

Therefore

$$\beta_k = \frac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}\mu)}$$

$$= \frac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}^2 L - (1-\alpha_{k+1})\gamma_{k+1})}$$

$$= \frac{\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}L)} = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}.$$

Note also that

$$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + q\alpha_{k+1},$$

where $q = \mu/L$, and

$$\alpha_0^2 L = (1-\alpha_0)\gamma_0 + \mu\alpha_0.$$

Constant Step Scheme (3.7)

0. Choose $x_0 \in R^n$ and $\alpha_0 \in (0,1)$. Set

$$y_0 = x_0, \quad q = \mu/L.$$

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

   a). Compute $f(y_k)$ and $f'(y_k)$. Set

$$x_{k+1} = y_k - \frac{1}{L} f'(y_k).$$

   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

$$\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2 + \alpha_{k+1}},$$

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

Theorem 3.3 *If in (3.7) we take*

$$\alpha_0 \geq \sqrt{\frac{\mu}{L}}, \tag{3.8}$$

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

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

Remarks.

1. Condition (3.8) is equivalent to $\gamma_0 \geq \mu$.

2. If $\alpha_0 = \sqrt{\frac{\mu}{L}}$ then

$$\alpha_k = \sqrt{\frac{\mu}{L}}, \quad \beta_k = \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}$$

for all $k \geq 0$.

Convex sets

Problem:

$$\min_{x \in Q} \; f(x).$$

We work with differentiable convex functions:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y),$$

$$\forall x, y \in R^n, \;\; \alpha \in [0,1].$$

What is the natural domain of convex function?

Definition. A set Q is called *convex* if for any x, $y \in Q$ and $\alpha \in [0,1]$ we have:

$$\alpha x + (1-\alpha) y \in Q.$$

Terminology:

Segment: $[x, y] = \{z = \alpha x + (1-\alpha) y, \;\; \alpha \in [0,1]\}$.

Convex combination of two points:

$$\alpha x + (1-\alpha) y$$

for some $\alpha \in [0,1]$.

Lemma 3.5 *If $f(x)$ is a convex function, then for any $\alpha \in R$ its* sublevel set

$$\mathcal{L}_f(\beta) = \{x \in R^n \mid \ f(x) \leq \beta\}$$

is either convex or empty.

Proof. Indeed, let x and y belong to $\mathcal{L}_f(\beta)$. Then

$$f(x) \leq \beta, \quad f(y) \leq \beta.$$

Therefore

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) \leq \beta.$$

$\square$

Lemma 3.6 *Let $f(x)$ be a convex function. Then its* epigraph

$$\mathcal{E}_f = \{(x,\tau) \in R^{n+1} \mid \ f(x) \leq \tau\}$$

is a convex set.

Proof. Indeed, let

$$z_1 = (x_1, \tau_1) \in \mathcal{E}_f, \quad z_2 = (x_2, \tau_2) \in \mathcal{E}_f.$$

Then for any $\alpha \in [0,1]$ we have:

$$\begin{aligned} z_\alpha &\equiv \alpha z_1 + (1-\alpha) z_2 \\ &= (\alpha x_1 + (1-\alpha)x_2, \alpha\tau_1 + (1-\alpha)\tau_2), \\ f(\alpha x_1 + (1-\alpha)x_2) &\leq \alpha f(x_1) + (1-\alpha) f(x_2) \\ &\leq \alpha\tau_1 + (1-\alpha)\tau_2. \end{aligned}$$

Thus, $z_\alpha \in \mathcal{E}_f$. $\square$

Properties of Convex Sets

Theorem 3.4 *Let $Q_1 \subseteq R^n$ and $Q_2 \subseteq R^m$ be convex sets and $\mathcal{A}(x)$ be a linear operator:*

$$\mathcal{A}(x) = Ax + b : \ R^n \ \to \ R^m.$$

Then all of the following sets are convex:

1. *Intersection ($m = n$):*

$$Q_1 \cap Q_2 = \{x \in R^n \mid \ x \in Q_1, \ x \in Q_2\}.$$

2. *Sum ($m = n$):*

$$Q_1 + Q_2 = \{z = x + y \mid \ x \in Q_1, \ y \in Q_2\}.$$

3. *Direct sum:*

$$Q_1 \times Q_2 = \{(x, y) \in R^{n+m} \mid \ x \in Q_1, \ y \in Q_2\}.$$

4. *Conic hull:*

$$\mathcal{K}(Q_1) = \{z \in R^n \mid \ z = \beta x, \ x \in Q_1, \beta \geq 0\}.$$

5. *Convex hull*

$$\mathrm{Conv}\,(Q_1, Q_2) = \{ \ z \in R^n \mid z = \alpha x + (1 - \alpha) y, \quad x \in Q_1, \ y \in Q_2, \ \alpha \in [0, 1]\}.$$

6. *Affine image:*

$$\mathcal{A}(Q_1) = \{y \in R^m \mid \ y = \mathcal{A}(x), \ x \in Q_1\}.$$

7. *Inverse affine image:*

$$\mathcal{A}^{-1}(Q_2) = \{x \in R^n \mid \ y = \mathcal{A}(x), \ y \in Q_2\}.$$

Proof.

1. If $x_1 \in Q_1 \cap Q_2$ and $x_1 \in Q_1 \cap Q_2$, then

$$[x_1, x_2] \subset Q_1, \quad [x_1, x_2] \subset Q_2.$$

Therefore $[x_1, x_2] \subset Q_1 \cap Q_2$.

2. If $z_1 = x_1 + x_2$, $x_1 \in Q_1$, $x_2 \in Q_2$ and $z_2 = y_1 + y_2$, $y_1 \in Q_1$, $y_2 \in Q_2$, then

$$\begin{aligned}\alpha z_1 + (1-\alpha) z_2 = & \ [\alpha x_1 + (1-\alpha) y_1]_1 \\ & +[\alpha x_2 + (1-\alpha) y_2]_2,\end{aligned}$$

where $[\cdot]_1 \in Q_1$ and $[\cdot]_2 \in Q_2$.

3. If $z_1 = (x_1, x_2)$, $x_1 \in Q_1$, $x_2 \in Q_2$ and $z_2 = (y_1, y_2)$, $y_1 \in Q_1$, $y_2 \in Q_2$, then

$$\begin{aligned}\alpha z_1 + (1-\alpha) z_2 = (& \ [\alpha x_1 + (1-\alpha) y_1]_1, \\ & \ [\alpha x_2 + (1-\alpha) y_2]_2),\end{aligned}$$

where $[\cdot]_1 \in Q_1$ and $[\cdot]_2 \in Q_2$.

4. If $z_1 = \beta_1 x_1$, $x_1 \in Q_1$, $\beta_1 \geq 0$, and $z_2 = \beta_2 x_2$, $x_2 \in Q_1$, $\beta_2 \geq 0$, then for any $\alpha \in [0, 1]$ we have:

$$\begin{aligned}\alpha z_1 + (1-\alpha) z_2 & = \alpha \beta_1 x_1 + (1-\alpha) \beta_2 x_2 \\ & = \gamma(\bar{\alpha} x_1 + (1-\bar{\alpha}) x_2,\end{aligned}$$

where

$$\gamma = \alpha \beta_1 + (1-\alpha) \beta_2, \quad \bar{\alpha} = \alpha \beta_1 / \gamma.$$

5. If $z_1 = \beta_1 x_1 + (1-\beta_1)x_2$, $x_1 \in Q_1$, $x_2 \in Q_2$, $\beta_1 \in [0,1]$, and $z_2 = \beta_2 y_1 + (1-\beta_2)y_2$, $y_1 \in Q_1$, $y_2 \in Q_2$, $\beta_2 \in [0,1]$, then for any $\alpha \in [0,1]$ we have:

$$\begin{aligned} \alpha z_1 + (1-\alpha)z_2 &= \alpha(\beta_1 x_1 + (1-\beta_1)x_2) \\ &\quad +(1-\alpha)(\beta_2 y_1 + (1-\beta_2)y_2) \\ &= \bar{\alpha}(\bar{\beta}_1 x_1 + (1-\bar{\beta}_1)y_1) \\ &\quad +(1-\bar{\alpha})(\bar{\beta}_2 x_2 + (1-\bar{\beta}_2)y_2), \end{aligned}$$

where $\bar{\alpha} = \alpha\beta_1 + (1-\alpha)\beta_2$ and

$$\bar{\beta}_1 = \alpha\beta_1/\bar{\alpha}, \quad \bar{\beta}_2 = \alpha(1-\beta_1)/(1-\bar{\alpha}).$$

6. If y_1, $y_2 \in \mathcal{A}(Q_1)$ then

$$y_1 = Ax_1 + b, \quad y_2 = Ax_2 + b, \quad x_1, x_2 \in Q_1.$$

For $y(\alpha) = \alpha y_1 + (1-\alpha)y_2$, $0 \le \alpha \le 1$, we have:

$$\begin{aligned} y(\alpha) &= \alpha(Ax_1 + b) + (1-\alpha)(Ax_2 + b) \\ &= A(\alpha x_1 + (1-\alpha)x_2) + b. \end{aligned}$$

Thus, $y(\alpha) \in \mathcal{A}(Q_1)$.

7. If x_1, $x_2 \in \mathcal{A}^{-1}(Q_2)$ then

$$y_1 = Ax_1 + b, \quad y_2 = Ax_2 + b, \quad y_1, y_2 \in Q_2.$$

For $x(\alpha) = \alpha x_1 + (1-\alpha)x_2$, $0 \le \alpha \le 1$, we have:

$$\begin{aligned} \mathcal{A}(x(\alpha)) &= A(\alpha x_1 + (1-\alpha)x_2) + b \\ &= \alpha(Ax_1 + b) + (1-\alpha)(Ax_2 + b) \\ &= \alpha y_1 + (1-\alpha)y_2 \in Q_2. \qquad \square \end{aligned}$$

Examples.

1. Half-space:

$$\{x \in R^n \mid \langle a, x \rangle \leq \beta\}$$

is convex (since a linear function is convex).

2. Polytope:

$$\{x \in R^n \mid \langle a_i, x \rangle \leq \beta_i, \; i = 1, \ldots, m\}$$

is convex (as an intersection of convex sets).

3. Ellipsoid. Let $A = A^T \geq 0$. Then the set

$$\{x \in R^n \mid \langle Ax, x \rangle \leq r^2\}$$

is convex (since $\langle Ax, x \rangle$ is a convex function).

And many others.

Optimality Conditions

Problem:

$$\min_{x \in Q} f(x), \quad f \in \mathcal{F}^1(R^n), \tag{3.9}$$

where Q is a closed convex set.

Example

$$\min_{x \geq 0} x.$$

Here $x \in R$, $Q = \{x \geq 0\}$, $f(x) = x$.

Note that

$$x^* = 0, \quad f'(x^*) = 1 > 0.$$

Thus, $f'(x^*) \neq 0$.

Theorem 3.5 *Let $f \in \mathcal{F}^1(R^n)$ and Q be a closed convex set.*

The point x^* is a solution of (3.9) iff

$$\langle f'(x^*), x - x^* \rangle \geq 0 \tag{3.10}$$

for all $x \in Q$.

Proof. Indeed, if (3.10) is true, then

$$f(x) \geq f(x^*) + \langle f'(x^*), x - x^* \rangle \geq f(x^*)$$

for all $x \in Q$.

Let x^* be a solution to (3.9). Assume that

$$\exists x \in Q : \quad \langle f'(x^*), x - x^* \rangle < 0.$$

Consider the function

$$\phi(\alpha) = f(x^* + \alpha(x - x^*)), \quad \alpha \in [0, 1].$$

Note that

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

Therefore, for small enough α we have:

$$f(x^* + \alpha(x - x^*)) = \phi(\alpha) < \phi(0) = f(x^*).$$

That is a contradiction. □

Theorem 3.6 *Let $f \in \mathcal{S}^1_\mu(R^n)$ and Q be a closed convex set.*

Then the solution x^* of the problem (3.9) exists and unique.

Proof. Let $x_0 \in Q$. Consider the set

$$\bar{Q} = \{x \in Q \mid \ f(x) \leq f(x_0)\}.$$

Note that the problem (3.9) is equivalent to the following:

$$\min\{f(x) \mid \ x \in \bar{Q}\}. \tag{3.11}$$

But $\bar{Q}$ is bounded: $\forall x \in \bar{Q}$

$$\begin{aligned} f(x_0) \ &\geq f(x) \\ &\geq f(x_0) + \langle f'(x_0), x - x_0\rangle + \tfrac{\mu}{2} \parallel x - x_0 \parallel^2 . \end{aligned}$$

Hence, $\parallel x - x_0 \parallel \leq \frac{2}{\mu} \parallel f'(x_0) \parallel$.

Thus, the solution x^* of (3.11) ($\equiv$ (3.9)) exists.

If x_1^* is also a solution to (3.9), then

$$\begin{aligned} f^* \ &= f(x_1^*) \\ &\geq f(x^*) + \langle f'(x^*), x_1^* - x^*\rangle + \tfrac{\mu}{2} \parallel x_1^* - x^* \parallel^2 \\ &\geq f^* + \tfrac{\mu}{2} \parallel x_1^* - x^* \parallel^2 \end{aligned}$$

(we have used Theorem 3.5).

Therefore $x_1^* = x^*$. $\square$

Gradient Mapping

Properties of the gradient:

Let $f \in \mathcal{F}_L^{1,1}(R^n)$. Then

- $f(x - \frac{1}{L} f'(x)) \leq f(x) - \frac{1}{2L} \parallel f'(x) \parallel^2$.
- $\langle f'(x), x - x^* \rangle \geq \frac{1}{L} \parallel f'(x) \parallel^2$.

What could replace it for (3.9)?

Let us fix $\gamma > 0$. Denote

$$\begin{aligned} x_Q(\gamma, x_0) &= \arg\min_{x \in Q} f(x_0) + \langle \ f'(x_0), x - x_0 \rangle \\ &\qquad + \tfrac{\gamma}{2} \parallel x - x_0 \parallel^2, \\ g_Q(\gamma, x_0) &= \gamma(x_0 - x_Q(\gamma, x_0)) \end{aligned}$$

We call $g_Q(\gamma, x)$ the *gradient mapping* of f on Q.

Note: 1. If $Q \equiv R^n$ then

$$\begin{aligned} x_Q(\gamma, x_0) &= x_0 - \tfrac{1}{\gamma} f'(x_0), \\ g_Q(\gamma, x_0) &= f'(x_0). \end{aligned}$$

2. We can use $x_0 \notin Q$.

Theorem 3.7 *Let* $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$*,* $\gamma \geq L$ *and* $x_0 \in R^n$*. Then for any* $x \in Q$ *we have:*

$$\begin{aligned} f(x) \geq\ & f(x_Q(\gamma, x_0)) + \langle g_Q(\gamma, x_0), x - x_0\rangle \\ & + \tfrac{1}{2\gamma} \parallel g_Q(\gamma, x_0) \parallel^2 + \tfrac{\mu}{2} \parallel x - x_0 \parallel^2 . \end{aligned} \tag{3.12}$$

Proof. Denote $x_Q = x_Q(\gamma, x_0)$, $g_Q = g_Q(\gamma, x_0)$,

$$\phi(x) = f(x_0) + \langle f'(x_0), x - x_0\rangle + \tfrac{\gamma}{2} \parallel x - x_0 \parallel^2 .$$

Then $\phi'(x) = f'(x_0) + \gamma(x - x_0)$, and for any $x \in Q$ we have:

$$\begin{aligned} 0 &\leq \langle \phi'(x_Q), x - x_Q\rangle \\ &= \langle f'(x_0) - g_Q, x - x_Q\rangle. \end{aligned}$$

Hence,

$$\begin{aligned} & f(x) - \tfrac{\mu}{2} \parallel x - x_0 \parallel^2 \geq f(x_0) + \langle f'(x_0), x - x_0\rangle \\ & = f(x_0) + \langle f'(x_0), x_Q - x_0\rangle + \langle f'(x_0), x - x_Q\rangle \\ & \geq f(x_0) + \langle f'(x_0), x_Q - x_0\rangle + \langle g_Q, x - x_Q\rangle \\ & = \phi(x_Q) - \tfrac{\gamma}{2} \parallel x_Q - x_0 \parallel^2 + \langle g_Q, x - x_Q\rangle \\ & = \phi(x_Q) - \tfrac{1}{2\gamma} \parallel g_Q \parallel^2 + \langle g_Q, x - x_Q\rangle \\ & = \phi(x_Q) + \tfrac{1}{2\gamma} \parallel g_Q \parallel^2 + \langle g_Q, x - x_0\rangle \end{aligned}$$

and $\phi(x_Q) \geq f(x_Q)$ since $\gamma \geq L$. □

Corollary 3.1 *Let* $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$, $\gamma \geq L$ *and* $x_0 \in R^n$. *Then*

$$f(x_Q(\gamma, x_0)) \leq f(x_0) - \tfrac{1}{2\gamma} \parallel g_Q(\gamma, x_0) \parallel^2, \qquad (3.13)$$

$$\begin{aligned} \langle g_Q(\gamma, x_0), x_0 - x^* \rangle \geq \; & \tfrac{1}{2\gamma} \parallel g_Q(\gamma, x_0) \parallel^2 \\ & + \tfrac{\mu}{2} \parallel x^* - x_0 \parallel^2 . \end{aligned} \qquad (3.14)$$

Proof. Indeed, using (3.12) with $x = x_0$, we get (3.13).

Using (3.12) with $x = x^*$, we get (3.14) since

$$f(x_Q(\gamma, x_0)) \geq f(x^*).$$

$\square$

Gradient Method

Problem:

$$\min_{x \in Q} f(x), \quad f \in \mathcal{S}^{1,1}_{\mu,L}(R^n), \quad \mu > 0,$$

where Q is a closed convex set.

Scheme:

$$x_0 \in Q,$$

$$x_{k+1} = x_k - h g_Q(L, x_k), \quad k = 0, \ldots\ .$$

Theorem 3.8 *If we choose* $h = \frac{1}{L}$*, then*

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

Proof. Denote $r_k = \| x_k - x^* \|$, $g_Q = g_Q(L, x_k)$. Then

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

□

Note: If $h = \frac{1}{L}$, then

$$x_{k+1} = x_k - \frac{1}{L} g_Q(L, x_k) = x_Q(L, x_k).$$

Optimal Methods

Estimate sequence:

$$\begin{aligned} x_0 &\in Q, \\ \phi_0(x) &= f(x_0) + \tfrac{\gamma_0}{2} \parallel x - x_0 \parallel^2, \\ \phi_{k+1}(x) &= (1-\alpha_k)\phi_k(x) \\ &+\alpha_k[f(x_Q(\gamma, y_k)) + \langle g_Q(L, y_k), x - y_k\rangle \\ &+\tfrac{1}{2L} \parallel g_Q(L, y_k) \parallel^2 + \tfrac{\mu}{2} \parallel x - y_k \parallel^2], \\ \phi_k(x) &\equiv \phi_k^* + \tfrac{\gamma_k}{2} \parallel x - v_k \parallel^2 . \end{aligned}$$

Similarly we get the following updating rules:

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

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

$$\phi_{k+1}^* = (1-\alpha_k)\phi_k^* + \alpha_k f(x_Q(L, y_k))$$

$$+\left(\tfrac{\alpha_k}{2L} - \tfrac{\alpha_k^2}{2\gamma_{k+1}}\right) \parallel g_Q(L, y_k) \parallel^2$$

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

Further, assuming $\phi_k^* \geq f(x_k)$ and using the inequality

$$\begin{aligned} f(x_k) \geq\ & f(x_Q(L, y_k)) + \langle g_Q(L, y_k), x_k - y_k\rangle \\ & + \tfrac{1}{2L} \parallel g_Q(L, y_k) \parallel^2 + \tfrac{\mu}{2} \parallel x_k - y_k \parallel^2, \end{aligned}$$

we come to the following lower bound:

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

Thus, again we can choose

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

Constant Step Scheme (3.15)

0. Choose $x_0 \in Q$ and $\alpha_0 \in (0, 1)$. Set

$$y_0 = x_0, \quad q = \mu/L.$$

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

   a). Compute $f(y_k)$ and $f'(y_k)$. Set

   $$x_{k+1} = x_Q(L, y_k).$$

   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

   $$\beta_k = \frac{\alpha_k(1 - \alpha_k)}{\alpha_k^2 + \alpha_{k+1}},$$

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

Note:
1. This method has the optimal rate of convergence.
2. Only x_k belong to Q.