Lemma 3.7 Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$. Then for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\|\nabla \boldsymbol{f}(\boldsymbol{y}) - \nabla \boldsymbol{f}(\boldsymbol{x}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})\|_2 \leq \frac{M}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2,$$

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \nabla \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle - \frac{1}{2}\langle \nabla^2 \boldsymbol{f}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle| \leq \frac{M}{6}\|\boldsymbol{y} - \boldsymbol{x}\|_2^3.$$

Lemma 3.8 Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, with $\|\nabla^2 \boldsymbol{f}(\boldsymbol{x}) - \nabla^2 \boldsymbol{f}(\boldsymbol{y})\|_2 \leq M\|\boldsymbol{x} - \boldsymbol{y}\|_2$. Then

$$\nabla^2 \boldsymbol{f}(\boldsymbol{x}) - M\|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I} \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{y}) \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{x}) + M\|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I}.$$

Proof:

Since $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, $\|\nabla^2 \boldsymbol{f}(\boldsymbol{y}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x})\|_2 \leq M\|\boldsymbol{y} - \boldsymbol{x}\|_2$. This means that the eigenvalues of the symmetric matrix $\nabla^2 \boldsymbol{f}(\boldsymbol{y}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x})$ satisfy:

$$|\lambda_i(\nabla^2 \boldsymbol{f}(\boldsymbol{y}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x}))| \leq M\|\boldsymbol{y} - \boldsymbol{x}\|_2, \quad i = 1, 2, \ldots, n.$$

Therefore,

$$-M\|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I} \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{y}) - \nabla^2 \boldsymbol{f}(\boldsymbol{x}) \preceq M\|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I}.$$

∎

3.1 Exercises

1. Prove Lemma 3.7.

4 Optimality Conditions and Algorithms for Minimizing Functions

4.1 General Minimization Problem and Terminologies

Definition 4.1 We define the *general minimization problem* as follows

$$\begin{cases} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & f_j(\boldsymbol{x}) \ \& \ 0, \quad j = 1, 2, \ldots, m \\ & \boldsymbol{x} \in S, \end{cases} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $f_j : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, 2, \ldots, m)$, the symbol & could be $=$, $\geq$, or $\leq$, and $S \subseteq \mathbb{R}^n$.

Definition 4.2 The *feasible set* Q of (1) is

$$Q = \{\boldsymbol{x} \in S \mid f_j(\boldsymbol{x}) \ \& \ 0, \ (j = 1, 2, \ldots, m)\}.$$

In the following, we assume $S \equiv \mathbb{R}^n$.

- If $Q \equiv \mathbb{R}^n$, (1) is a *unconstrained optimization problem.*
- If $Q \subsetneq \mathbb{R}^n$, (1) is a *constrained optimization problem.*
- If all functionals $f(\boldsymbol{x}), f_j(\boldsymbol{x})$ are differentiable, (1) is a *smooth optimization problem.*
- If one of functionals $f(\boldsymbol{x})$, $f_j(\boldsymbol{x})$ is non-differentiable, (1) is a *non-smooth optimization problem.*
- If all constraints are linear $f_j(\boldsymbol{x}) = \langle \boldsymbol{a}_j, \boldsymbol{x}\rangle + b_j$ $(j = 1, 2, \ldots, m)$, (1) is a *linear constrained optimization problem.*

- In addition, if $f(\boldsymbol{x})$ is linear, (1) is a *linear programming problem.*
- In addition, if $f(\boldsymbol{x})$ is quadratic, (1) is a *quadratic programming problem.*

- If $f(\boldsymbol{x})$, $f_j(\boldsymbol{x})$ $(j = 1, 2, \ldots, m)$ are quadratic, (1) is a *quadratically constrained quadratic programming problem.*

Definition 4.3 $\boldsymbol{x}^*$ is called a *global optimal solution* of (1) if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in Q$. Moreover, $f(\boldsymbol{x}^*)$ is called the *global optimal value.* $\boldsymbol{x}^*$ is called a *local optimal solution* of (1) if there exists an open ball $B(\boldsymbol{x}^*, \varepsilon) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}^*\|_2 < \varepsilon\}$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in B(\boldsymbol{x}^*, \varepsilon) \cap Q$. Moreover, $f(\boldsymbol{x}^*)$ is called a *local optimal value.*

4.2 Complexity Bound for a Global Optimization Problem on the Unit Box

Consider one of the simplest problems in optimization, that is, minimizing a function on the n-dimensional box.

$$\begin{cases} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in B_n := \{\boldsymbol{x} \in \mathbb{R}^n \mid 0 \leq [\boldsymbol{x}]_i \leq 1, \ i = 1, 2, \ldots, n\}. \end{cases} \tag{2}$$

To be coherent, we use the ℓ_∞-norm:

$$\|\boldsymbol{x}\|_\infty = \max_{1 \leq i \leq n} |[\boldsymbol{x}]_i|.$$

Let us also assume that $f(\boldsymbol{x})$ is *Lipschitz continuous on B_n*:

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_\infty, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in B_n.$$

Let us define a very simple method to solve (2), the **uniform grid method**.

Given a positive integer $p > 0$,

1. Form $(p+1)^n$ points
$$\boldsymbol{x}_{i_1, i_2, \ldots, i_n} = \left(\frac{i_1}{p}, \frac{i_2}{p}, \ldots, \frac{i_n}{p}\right)^T$$
where $(i_1, i_2, \ldots, i_n) \in \{0, 1, \ldots, p\}^n$.

2. Among all points $\boldsymbol{x}_{i_1, i_2, \ldots, i_n}$, find a point $\bar{\boldsymbol{x}}$ which has the minimal value for the objective function.

3. Return the pair $(\bar{\boldsymbol{x}}, f(\bar{\boldsymbol{x}}))$ as the result.

Theorem 4.4 Let $f(\boldsymbol{x}^*)$ be the global optimal value for (2). Then the uniform grid method yields

$$f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq \frac{L}{2p}.$$

Proof:

Let $\boldsymbol{x}^*$ be a global optimal solution. Then there are coordinates $(i_1, i_2, \ldots, i_n)$ such that $\boldsymbol{x} := \boldsymbol{x}_{i_1, i_2, \ldots, i_n} \leq \boldsymbol{x}^* \leq \boldsymbol{x}_{i_1+1, i_2+1, \ldots, i_n+1} =: \boldsymbol{y}$. Observe that $[\boldsymbol{y}]_i - [\boldsymbol{x}]_i = 1/p$ for $i = 1, 2, \ldots, n$ and $[\boldsymbol{x}^*]_i \in [[\boldsymbol{x}]_i, [\boldsymbol{y}]_i]$ $(i = 1, 2, \ldots, n)$.

Consider $\hat{\boldsymbol{x}} = (\boldsymbol{x} + \boldsymbol{y})/2$ and form a new point $\tilde{\boldsymbol{x}}$ as:

$$[\tilde{\boldsymbol{x}}]_i := \begin{cases} [\boldsymbol{y}]_i, & \text{if } [\boldsymbol{x}^*]_i \geq [\hat{\boldsymbol{x}}]_i \\ [\boldsymbol{x}]_i, & \text{otherwise.} \end{cases}$$

It is clear that $|[\tilde{\boldsymbol{x}}]_i - [\boldsymbol{x}^*]_i| \leq 1/(2p)$ for $i = 1, 2, \ldots, n$. Then $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}^*\|_\infty = \max_{1 \leq i \leq n} |[\tilde{\boldsymbol{x}}]_i - [\boldsymbol{x}^*]_i| \leq 1/(2p)$. Since $\tilde{\boldsymbol{x}}$ belongs to the grid,

$$f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq f(\tilde{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq L\|\tilde{\boldsymbol{x}} - \boldsymbol{x}^*\|_\infty \leq L/(2p).$$

∎

Let us define our goal

$$\boxed{\text{Find } \boldsymbol{x} \in B_n \text{ such that } f(\boldsymbol{x}) - f(\boldsymbol{x}^*) < \varepsilon.}$$

Corollary 4.5 The number of iterations necessary for the problem (2) to achieve the above goal using the uniform grid method is at most

$$\left(\left\lfloor \frac{L}{2\varepsilon} \right\rfloor + 2 \right)^n.$$

Proof:
Take $p = \lfloor L/(2\varepsilon) \rfloor + 1$. Then, $p > L/(2\varepsilon)$ and from the previous theorem, $f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) \leq L/(2p) < \varepsilon$. Observe that we constructed $(p+1)^n$ points. ∎

Consider the class of problems $\mathcal{P}$ defined as follows:

Model:	$\min_{\boldsymbol{x} \in B_n} f(\boldsymbol{x})$,
	$f(\boldsymbol{x})$ is ℓ_∞-Lipschitz continuous on B_n.
Oracle:	Only function values are available
Approximate solution:	Find $\bar{\boldsymbol{x}} \in B_n$ such that $f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) < \varepsilon$

Theorem 4.6 For $\varepsilon < \frac{L}{2}$, the number of iterations necessary for the class of problems $\mathcal{P}$ using any method which uses only function evaluations is always at least $(\lfloor \frac{L}{2\varepsilon} \rfloor)^n$.

Proof:
Let $p = \lfloor \frac{L}{2\varepsilon} \rfloor$ (which is ≥ 1 from the hypothesis).

Suppose that there is a method which requires $N < p^n$ calls of the oracle to solve the problem in $\mathcal{P}$.

Then, there is a point $\hat{\boldsymbol{x}} \in B_n = \{\boldsymbol{x} \in \mathbb{R}^n \mid 0 \leq [\boldsymbol{x}]_i \leq 1, \ i = 1, 2, \ldots, n\}$ where there is no test points in the <u>interior</u> of $B := \{\boldsymbol{x} \mid \hat{\boldsymbol{x}} \leq \boldsymbol{x} \leq \hat{\boldsymbol{x}} + \boldsymbol{e}/p\}$ where $\boldsymbol{e} = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$.

Let $\boldsymbol{x}^* := \hat{\boldsymbol{x}} + \boldsymbol{e}/(2p)$ and consider the function $\bar{f}(\boldsymbol{x}) := \min\{0, L\|\boldsymbol{x} - \boldsymbol{x}^*\|_\infty - \varepsilon\}$. Clearly, $\bar{f}$ is ℓ_∞-Lipschitz continuous with constant L and its global minimum is $-\varepsilon$. Moreover, $\bar{f}(\boldsymbol{x})$ is non-zero valued only inside the box $B' := \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\|_\infty \leq \varepsilon/L\}$.

Since $2p \leq L/\varepsilon$, $B' \subseteq \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\|_\infty \leq 1/(2p)\} = B$.

Therefore, $\bar{f}(\boldsymbol{x})$ is equal to zero to all test points of our method and the accuracy of the method is ε.

If the number of calls of the oracle is less than p^n, the accuracy can not be better than ε. ∎

Theorem 4.6 supports the claim that the *general optimization problem is unsolvable.*

Example 4.7 Consider a problem defined by the following parameters. $L = 2$, $n = 10$, and $\varepsilon = 0.01$.