# 5 Algorithms

## 5.1 Gradient Descent Methods

Consider the following minimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

where $f$ is a differentiable function.

A general optimization algorithm is of the following form:
Choose initial point $x^0$ and construct a sequence $\{x^k\}$ by

$$x^{t+1} = x^t + \eta_t d^t, \;\; k = 0, 1, ...$$

What should we choose for $d^t$? What should we choose for $\eta_t$?
For the first question, we want $d^t$ to be a descent direction, that is

$$f'(x^t; d^t) = \langle \nabla f(x^t), d^t \rangle \leq 0$$

Note that

$$-\nabla f(x) = \arg \min_{d | \|d\| \leq 1} f'(x; d) = \arg \min_{d | \|d\| \leq 1} \langle \nabla f(x), d \rangle$$

By choosing $d^t = \nabla f(x^t)$, we get the greatest rate of function value improvement.
This is the gradient descent or steepest descent:

$$x^{t+1} = x^t - \eta_t \nabla f(x^t)$$

As for the second question, there are mainly three ways to select $\eta_t$.
**Fixed step size**: $\eta_t$ is constant.
**Exact line search**

$$\eta_t = \text{argmin}_{\eta \geq 0} f(x + \eta d^t)$$

**Backtracking line search:** Shrink the step size until it satisfy some conditions.
One popular condition is the Armijo's condition:
Choose $0 < \alpha \leq \frac{1}{2}, 0 < \beta < 1,$ initialize $\eta_t = 1;$ take $\eta_t := \beta \eta_t$ until

$$f\left(x^t - \eta_t \nabla f(x^t)\right) < f(x^t) - \frac{1}{2} \alpha \eta_t \|\nabla f(x^t)\|^2$$

### 5.1.1 Strongly Convex and $L$-smooth

Before proving convergence results, we need to introduce two notations of a function.

**Definition:(Strongly Convex)** A differentiable function $f$ is called $\mu$-*strongly convex* if

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}||y-x||^2, \text{ for all } x, y$$

**Definition:($L$-smooth)** A differentiable function $f$ is called $L$-*smooth* if

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}||y-x||^2, \text{ for all } x, y$$

We have the following characterization for the two notations.

**Proposition:(Characterization of $\mu$-strongly convex)** Given a differentiable function $f$, the following are equivalent:

1. $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}||y-x||^2, \text{ for all } x, y$
2. $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)||y-x||^2, \text{ for all } x, y, \ \lambda \in [0,1]$
3. $g(x) := f(x) - \frac{\mu}{2}||x||^2$ is convex.
4. $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \mu ||y-x||^2, \text{ for all } x, y$
5. $\nabla^2 f(x) - \mu I \succeq 0, \text{ for all } x$ (if $f$ is $C^2$).

*Proof.* We have seen that (1), (3), (4), (5) are equivalent.
Let's prove (2), (3) are equivalent.
(2)⇒(3) Multiply by $\lambda, (1-\lambda)$ respectively, we get

$$\lambda f(z) \leq \lambda^2 f(x) + \lambda(1-\lambda)f(y) - \frac{\mu}{2}\lambda^2(1-\lambda)||y-x||^2$$

$$(1-\lambda)f(z) \leq \lambda(1-\lambda)f(x) + (1-\lambda)^2 f(y) - \frac{\mu}{2}\lambda(1-\lambda)^2||y-x||^2$$

Summing up, we get

$$\begin{aligned} f(z) &\leq \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)||y-x||^2 \\ &= \lambda f(x) - \frac{\mu}{2}\lambda||x||^2 + (1-\lambda)f(y) - \frac{\mu}{2}(1-\lambda)||y||^2 + \frac{\mu}{2}||\lambda x + (1-\lambda)y||^2 \\ &= \lambda g(x) + (1-\lambda)g(y) + \frac{\mu}{2}||z||^2 \end{aligned}$$

(3)⇒(2) Since $g$ is convex, for $\lambda \in [0,1]$, we have

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y), \text{ for all } x, y$$

Hence,

$$
\begin{aligned}
&f(\lambda x + (1-\lambda)y) \\
&\le \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda||x||^2 - \frac{\mu}{2}(1-\lambda)||y||^2 + \frac{\mu}{2}||\lambda x + (1-\lambda)y||^2 \\
&= \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}(\lambda||x||^2 + (1-\lambda)||y||^2 - \lambda^2||x||^2 - 2\lambda(1-\lambda)\langle x, y\rangle - (1-\lambda)^2||y||^2) \\
&= \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)(||x||^2 - 2\langle x, y\rangle + ||y||^2) \\
&= \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}||y-x||^2
\end{aligned}
$$

□

**Proposition:(Characterization of $L$-smooth)** Given a differentiable convex function $f$, the following are equivalent:

1. $f(y) \le f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{2}||y-x||^2$, for all $x, y$
2. $f(\lambda x + (1-\lambda)y) \ge \lambda f(x) + (1-\lambda)f(y) - \frac{L}{2}\lambda(1-\lambda)||y-x||^2$, for all $x, y$, $\lambda \in [0,1]$
3. $h(x) := \frac{L}{2}||x||^2 - f(x)$ is convex.
4. $\langle \nabla f(y) - \nabla f(x), y-x\rangle \ge \frac{1}{L}||\nabla f(y) - \nabla f(x)||^2$, for all $x, y$
5. $||\nabla f(y) - \nabla f(x)|| \le L||y-x||$, for all $x, y$ ($L$-Lipschtiz gradient)
6. $LI - \nabla^2 f(x) \succeq 0$, for all $x$ (if $f$ is $C^2$).

*Proof.* The equivalence of (1), (2), (3), (6) is similar to that of strong convexity. We will show that (5)⇒(1)⇒(4)⇒(5) holds.
(5)⇒(1): Consider $g(t) = f(x+t(y-x))$. Then $g'(t) = \langle \nabla f(x+t(y-x)), (y-x)\rangle$. Then

$$
\begin{aligned}
&f(y) - f(x) - \langle \nabla f(x), y-x\rangle \\
&= g(1) - g(0) - \langle \nabla f(x), y-x\rangle \\
&= \int_0^1 \langle \nabla f(x+t(y-x)), y-x\rangle - \langle \nabla f(x), y-x\rangle dt \\
&= \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x\rangle dt \\
&\le \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\|\|y-x\| dt \\
&\le \int_0^1 Lt||y-x||^2 dt \\
&= \frac{L}{2}||y-x||^2
\end{aligned}
$$

(1)⇒(4): Consider the function $\phi_x(z) := f(z) - \langle \nabla f(x), z\rangle$.
$\phi_x$ is convex and $\nabla \phi_x(z) = \nabla f(z) - \nabla f(x)$.
Since, $f(z) \le f(y) + \langle \nabla f(y), z - y\rangle + \frac{L}{2}||z - y||^2$, we have

$$f(z) - \langle \nabla f(x), z\rangle \le f(y) - \langle \nabla f(x), y\rangle + \langle \nabla f(y) - \nabla f(x), z - y\rangle + \frac{L}{2}||z - y||^2$$

That is

$$\phi_x(z) \le \phi_x(y) + \langle \nabla \phi_x(y), z - y\rangle + \frac{L}{2}||z - y||^2$$

We minimized both sides over $z$. The left hand side is minimized at $z = x$. The right hand side is minimized at $z = -\frac{1}{L}\nabla \phi_x(y) + y$. Hence,

$$\begin{aligned} f(x) - \langle \nabla f(x), x\rangle = \phi_x(x) &\le \phi_x(y) + \langle \nabla \phi_x(y), -\frac{1}{L}\nabla \phi_x(y)\rangle + \frac{L}{2}\|\frac{1}{L}\nabla \phi_x(y)\|^2 \\ &= f(y) - \langle \nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2 \end{aligned}$$

So

$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle \ge \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2$$

Interchange the role of $x, y$, we get

$$f(x) - f(y) - \langle \nabla f(y), x - y\rangle \ge \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2$$

Adding the two inequalities, we get

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$$

(1)⇒(2): Let $z = \lambda x + (1 - \lambda)y$. Then

$$\lambda f(x) \le \lambda f(z) + \langle \nabla f(z), \lambda(x - z)\rangle + \frac{L}{2}\lambda\|x - z\|^2$$

$$(1 - \lambda)f(y) \le (1 - \lambda)f(z) + \langle \nabla f(z), (1 - \lambda)(y - z)\rangle + \frac{L}{2}(1 - \lambda)\|y - z\|^2$$

Then,

$$\lambda f(x) + (1 - \lambda)f(y) \le f(z) + \frac{L}{2}\lambda(1 - \lambda)\|y - x\|^2$$

(4)⇒(5): We have

$$\begin{aligned} \|\nabla f(y) - \nabla f(x)\|^2 &\le L\langle \nabla f(y) - \nabla f(x), y - x\rangle \\ &\le L\|\nabla f(y) - \nabla f(x)\|\|y - x\| \end{aligned}$$

□

### 5.1.2 Convergence of Gradient Descent Methods

We start of analysis of gradient descent method with $L$-smooth objective function.
We suppose the optimal value of $f$ is finite and is denoted by $f^*$. Also suppose $x^*$ is a optimal solution.

**Proposition:** Suppose $f$ is a convex $C^1$ function and is $L$-smooth. If the step size $\eta \leq \frac{1}{L}$, then the fixed size gradient descent satisfies

$$f(x^t) - f(x^*) \leq \frac{1}{2t\eta} \left\|x^0 - x^*\right\|^2$$

*Proof.* Let $x^+ := x - \eta \nabla f(x)$. Then using quadratic upper bound, we have,

$$f\left(x^+\right) \leq f(x) + \left(-\eta + \frac{L\eta^2}{2}\right) \|\nabla f(x)\|^2 \leq f(x) - \frac{\eta}{2}\|\nabla f(x)\|^2$$

Hence, the sequence generated by gradient descent method is descending. That is,

$$f(x^{t+1}) \leq f(x^t)$$

Since $f$ is convex, $f\left(x^*\right) \geq f(x) + \langle \nabla f(x), x^* - x\rangle$. Then

$$\begin{aligned}
f\left(x^+\right) &\leq f(x) - \frac{\eta}{2}\|\nabla f(x)\|^2 \\
&\leq f^* + \langle \nabla f(x), x - x^* \rangle - \frac{\eta}{2}\|\nabla f(x)\|^2 \\
&= f^* + \frac{1}{2\eta}\left(\|x - x^*\|^2 - \|x - x^* - \eta\nabla f(x)\|^2\right) \\
&= f^* + \frac{1}{2\eta}\left(\|x - x^*\|^2 - \left\|x^+ - x^*\right\|^2\right)
\end{aligned}$$

Summing the above, we get

$$\begin{aligned}
\sum_{i=1}^{t}\left(f\left(x^i\right) - f^*\right) &\leq \frac{1}{2\eta}\sum_{i=1}^{t}\left(\left\|x^{i-1} - x^*\right\|^2 - \left\|x^i - x^*\right\|^2\right) \\
&= \frac{1}{2\eta}\left(\left\|x^0 - x^*\right\|^2 - \left\|x^t - x^*\right\|^2\right) \\
&\leq \frac{1}{2\eta}\left\|x^0 - x^*\right\|^2
\end{aligned}$$

But $f(x^i)$ is decreasing, hence

$$f\left(x^t\right) - f^* \leq \frac{1}{t}\sum_{i=1}^{t}\left(f\left(x^i\right) - f^*\right) \leq \frac{1}{2t\eta}\left\|x^0 - x^*\right\|^2$$

□

Therefore in order to get $f(x^t) - f^* \leq \epsilon$, we need $O(\frac{1}{\epsilon})$ iterations.
We can get a similar result for the backtracking line search method.

In order to get faster convergence, more assumptions are needed.
**Lemma:** Suppose $f$ is $\mu$-strongly convex and $L$-smooth. Then

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \frac{1}{L+\mu} \|\nabla f(x) - \nabla f(y)\|^2 + \frac{\mu L}{L+\mu} \|x-y\|^2$$

*Proof.* Consider $\phi(x) = f(x) - \frac{\mu}{2}\|x\|^2$. $\nabla \phi(x) = \nabla f(x) - \mu x$. So

$$\begin{aligned}
\|\nabla \phi(x) - \nabla \phi(y)\|^2 &= \|\nabla f(x) - \nabla f(y) - \mu(x-y)\|^2 \\
&= \|\nabla f(x) - \nabla f(y)\|^2 - 2\mu\langle \nabla f(x) - \nabla f(y), x-y\rangle + \mu^2\|x-y\|^2 \\
&\leq (1 - \frac{2\mu}{L})\|\nabla f(x) - \nabla f(y)\|^2 + \mu^2\|x-y\|^2 \\
&\leq (1 - \frac{2\mu}{L})L^2\|x-y\|^2 + \mu^2\|x-y\|^2 \\
&= (L-\mu)^2\|x-y\|^2
\end{aligned}$$

Hence $\phi(x)$ is $L-\mu$-smooth.
Then, $\langle \nabla\phi(y) - \nabla\phi(x), y-x\rangle \geq \frac{1}{L-\mu}\|\nabla\phi(y) - \nabla\phi(x)\|^2$. Hence

$$\langle \nabla f(y) - \nabla f(x) - \mu(y-x), y-x\rangle \geq \frac{1}{L-\mu}\|\nabla f(y) - \nabla f(x) - \mu(y-x)\|^2$$

After expanding, we get out required inequlity. □

**Proposition:** Suppose $f$ is $\mu$-strongly convex and $L$-smooth. Then the constant step size gradient descent method with $\eta_t = \frac{2}{\mu+L}$ satisfies:

$$\|x^t - x^*\| \leq \left(\frac{K-1}{K+1}\right)^t \|x^0 - x^*\|$$

where $K = L/\mu$.

*Proof.*

$$\begin{aligned}
\|x^{t+1} - x^*\|^2 &= \|x^t - \eta\nabla f(x^t) - x^*\|^2 \\
&= \|x^t - x^*\|^2 - \langle x^t - x^*, 2\eta\nabla f(x^t)\rangle + \eta^2\|\nabla f(x^t)\|^2 \\
&\leq \|x^t - x^*\|^2 - \eta\frac{2}{L+\mu}\|\nabla f(x^t)\|^2 - \frac{2\eta\mu L}{L+\mu}\|x^t - x^*\|^2 + \eta^2\|\nabla f(x^t)\|^2 \\
&= (1 - \frac{2\eta\mu L}{L+\mu})\|x^t - x^*\|^2 \\
&= (\frac{L-\mu}{L+\mu})^2\|x^t - x^*\|^2
\end{aligned}$$

Hence

$$\|x^t - x^*\| \leq \left(\frac{K-1}{K+1}\right)^t \|x^0 - x^*\|$$

□

We can get a similar result with the backtracking gradient descent.

**Lemma:** Suppose $f$ is $\mu$-strongly convex and $L$-smooth. Then

$$2\mu(f(x) - f^*) \leq \|\nabla f(x)\|^2$$

*Proof.* Since $f$ is $\mu$-strongly convex,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|^2$$

By minimizing the right hand side with respect to $y$, we find the minimizer is $x - \frac{1}{\mu}\nabla f(x)$.
Therefore,

$$f(y) \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2$$

Since this holds for all $y$, we have

$$f^* \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2$$

□

**Proposition:** Suppose $f$ is $\mu$-strongly convex and $L$-smooth. Then the gradient descent method with backtracking line search satisfies:

$$f(x^t) - f^* \leq c^t(f(x^0) - f^*)$$

where $c = 1 - \min\{2\alpha\mu, 2\alpha\beta\mu/L\}$.

*Proof.* We first show that the step size is either $\eta_t = 1$ or satisfies $\eta_t \geq \beta/L$.
Let $x^+ := x - \eta\nabla f(x)$. If $0 \leq \eta \leq 1/L$. Then

$$\begin{aligned} f(x^+) &\leq f(x) - \eta\|\nabla f(x)\|^2 + \frac{L\eta^2}{2}\|\nabla f(x)\|^2 \\ &\leq f(x) - \frac{\eta}{2}\|\nabla f(x)\|^2 \\ &\leq f(x) - \alpha\eta\|\nabla f(x)\|^2 \end{aligned}$$

Let $\eta_t$ be the step size chosen at iteration $t$.
If the Armijo's condition is satisfied at the initialization, then $\eta_t = 1$.
Otherwise, $\eta_t/\beta$ does not satisfy the Armijo's condition.
So $\frac{\eta_t}{\beta} \geq \frac{1}{L}$. Hence $\eta_t \geq \frac{\beta}{L}$.
If $\eta_t = 1$, then

$$f(x^{t+1}) \leq f(x^t) - \alpha\|\nabla f(x^t)\|^2$$

If $\eta_t \geq \beta/L$, then

$$f(x^{t+1}) \leq f(x^t) - \alpha\eta_t\|\nabla f(x^t)\|^2 \leq f(x^t) - \alpha\beta/L\|\nabla f(x^t)\|^2$$

Therefore

$$f(x^{t+1}) - f^* \leq f(x^t) - f^* - \min\{\alpha, \alpha\beta/L\}\|\nabla f(x)\|^2$$

Since $2\mu(f(x^t) - f^*) \leq \|\nabla f(x^t)\|^2$,

$$f(x^{t+1}) - f^* \leq (1 - \min\{2\alpha\beta\mu, 2\alpha\beta\mu/L\})(f(x^t) - f^*)$$

Therefore,

$$f(x^{t+1}) - f^* \leq (1 - \min\{2\alpha\mu, 2\alpha\beta\mu/L\})^t(f(x^0) - f^*)$$

□

In order to get a $\epsilon$ accuracy, we need $O(\log(1/\epsilon))$ iterations.
Therefore, we get a linear convergence if the objective function is also strongly convex.

## 5.2 Projected Gradient Descent

Let's consider the problem:

$$\min_{x \in C} f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, $C$ is a closed convex set.
If we simply carry out a gradient descent, the iterate points may not be in $C$.
One simplest one to modify the gradient descent is to consider the projected version, which is called projected gradient descent:

$$x^{t+1} = P_C(x^t - \eta_t\nabla f(x^t))$$

where $P_C(\cdot)$ is the projection to $C$.
Recall the following results about projection to a closed convex set.

**Proposition:** Let $C$ be a nonempty convex set and let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex $C^1$ function. Then $x^* \in C$ minimizes $f$ over $C$ if and only if

$$\langle\nabla f(x^*), (z - x^*)\rangle \geq 0, \ \forall z \in C.$$

**Proposition:** $x^* = P_C(z)$ if and only if $\langle z - x^*, x - x^*\rangle \leq 0, \ \forall x \in C$.

### 5.2.1 Convergence for $L$-smooth objective

We will first show convergence result for $L$-smooth objective $f$.

**Lemma** Suppose $f$ is $L$-smooth. Then the projected gradient descent with fixed step size $\eta_t = \eta \leq \frac{1}{L}$ satisfies:

$$f(x^{t+1}) \leq f(x^t) - \frac{L}{2}\|x^{t+1} - x^t\|^2$$

*Proof.* We have

$$\langle x^t - x^{t+1}, x^t - \eta_t \nabla f(x^t) - x^{t+1}] \rangle \leq 0$$

That is

$$\langle \nabla f(x^t), x^{t+1} - x^t \rangle \leq -\frac{1}{\eta_t}\|x^{t+1} - x^t\|^2$$

Since $f$ is $L$-smooth,

$$\begin{aligned} f(x^{t+1}) &\leq f(x^t) + \langle \nabla f(x^t), x^{t+1} - x^t \rangle + \frac{L}{2}\|x^{t+1} - x^t\|^2 \\ &\leq f(x^t) + (-\frac{1}{\eta_t} + \frac{L}{2})\|x^{t+1} - x^t\}^2 \\ &\leq f(x^t) - \frac{L}{2}\|x^{t+1} - x^t\|^2 \end{aligned}$$

□

**Proposition:** Let $f$ be $L$-smooth. Then the projected gradient descent with fixed step size $\eta_t = \eta \leq \frac{1}{L}$ satisfies:

$$f(x^t) - f^* \leq \frac{L}{2t}\|x^0 - x^*\|^2$$

*Proof.* Since $x^{t+1} = P_C(x^t - \eta_t \nabla f(x^t))$, we have

$$\langle x^* - x^{t+1}, x^t - \eta_t \nabla f(x^t) - x^{t+1} \rangle \leq 0$$

That is

$$\langle \nabla f(x^t), x^{t+1} - x^* \rangle \leq \frac{1}{\eta_t}\langle x^{t+1} - x^*, x^t - x^{t+1} \rangle$$

Since $f$ is convex, we have

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

Since $f$ is $L$-smooth, then

$$\begin{aligned}
f(x^{t+1}) &\le f(x^t) + \langle \nabla f(x^t), x^{t+1} - x^t \rangle + \frac{L}{2}\|x^{t+1} - x^t\|^2 \\
&\le f(x^*) - \langle \nabla f(x^t), x^* - x^t \rangle + \langle \nabla f(x^t), x^{t+1} - x^t \rangle + \frac{L}{2}\|x^{t+1} - x^t\|^2 \\
&= f(x^*) + \langle \nabla f(x^t), x^{t+1} - x^* \rangle + \frac{L}{2}\|x^{t+1} - x^t\|^2 \\
&\le f(x^*) + \frac{1}{\eta_t}\langle x^{t+1} - x^*, x^t - x^{t+1} \rangle + \frac{L}{2}\|x^{t+1} - x^t\|^2 \\
&\le f(x^*) - L\langle x^{t+1} - x^*, x^{t+1} - x^t \rangle + \frac{L}{2}\|x^{t+1} - x^t\|^2 \\
&= f(x^*) + \frac{L}{2}(\|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2)
\end{aligned}$$

Summing up, we have

$$\begin{aligned}
\sum_{i=1}^{t} f(x^i) - f^* &\le \sum_{i=1}^{t} \frac{L}{2}(\|x^{i-1} - x^*\|^2 - \|x^i - x^*\|^2) \\
&= \frac{L}{2}(\|x^0 - x^*\|^2 - \|x^t - x^*\|^2) \\
&\le \frac{L}{2}(\|x^0 - x^*\|^2)
\end{aligned}$$

Since $f(x^t)$ is decreasing, we have

$$f(x^t) - f^* \le \frac{L}{2t}\|x^0 - x^*\|^2$$

□

### 5.2.2 Convergence rate under strong convexity

Let's now consider the projected gradient descent under the assumption that $f$ is $\mu$-strongly convex.
We denote $G_\eta(x) = P_C(x - \eta\nabla f(x))$. A optimal solution of the problem is in fact a fixed point of $G_\eta$.
If we can show that $G_\eta$ is a contraction, then $\{x^t\}$ generated by the projected gradient method converges linearly to an optimal solution.

**Proposition:** Suppose $f$ is $\mu$-strongly convex and $L$-smooth. Then $G_\eta$ satisfies

$$\|G_\eta(x) - G_\eta(y)\| \le \max\{|1 - \eta L|, |1 - \eta\mu|\}\|x - y\|, \forall x, y$$

and is a contraction for all $\eta \in (0, 2/L)$.

*Proof.* We first prove that $\|P_C(x) - P_C(y)\| \le \|x - y\|$ for all $x, y$.
By the projection property

$$\langle (x - P_C(x), z - P_C(x)\rangle \le 0,\ \forall z \in C$$

Put $z = P_C(y)$, then $\langle x - P_C(x), P_C(y) - P_C(x)\rangle \leq 0$.
Similarly, $\langle y - P_C(y), P_C(x) - P_C(y)\rangle \leq 0$. Hence

$$\langle y - x - (P_C(y) - P_C(x)), P_C(x) - P_C(y)\rangle \leq 0$$

$$\|P_C(x) - P_C(y)\|^2 \leq \langle x - y, P_C(x) - P_C(y)\rangle$$

By Cauchy-Schwarz, $\|P_C(x) - P_C(y)\| \leq \|x - y\|$.
Hence

$$\begin{aligned}
&\|G_\eta(x) - G_\eta(y)\|^2 \\
&= \|P_C(x - \eta\nabla f(x)) - P_C(y - \eta\nabla f(y)\|^2 \\
&\leq \|(x - \eta\nabla f(x) - (y - \eta\nabla f(y)\|^2 \\
&= \|x - y\|^2 - 2\eta\langle\nabla f(x) - \nabla f(y), x - y\rangle + \eta^2\|\nabla f(x) - \nabla f(y)\|^2 \\
&\leq \|x - y\|^2 - \frac{2\eta\mu L}{\mu + L}\|x - y\|^2 - \frac{2\eta}{\mu + L}\|\nabla f(x) - \nabla f(y)\|^2 + \eta^2\|\nabla f(x) - \nabla f(y)\|^2 \\
&= (1 - \frac{2\eta\mu L}{\mu + L})\|x - y\|^2 + \eta(\eta - \frac{2}{\mu + L})\|\nabla f(x) - \nabla f(y)\|^2 \\
&\leq (1 - \frac{2\eta\mu L}{\mu + L})\|x - y\|^2 + \eta\max\{L^2(\eta - \frac{2}{\mu + L}), \mu^2(\eta - \frac{2}{\mu + L})\}\|x - y\|^2 \\
&= \max\{(1 - \eta L)^2, (1 - \eta\mu)^2\}\|x - y\|^2
\end{aligned}$$

□

**Proposition:** Suppose $f$ is $\mu$-strongly convex and $L$-smooth. Then the projected gradient descent with fixed step size $\frac{2}{L+\mu}$ satisfies:

$$\|x^t - x^*\| \leq \big(\frac{L - \mu}{L + \mu}\big)^t\|x^0 - x^*\|$$

*Proof.* Since $\eta = \frac{2}{L+\mu}$, $\max\{(1 - \eta L), (1 - \eta\mu)\} = \frac{L-\mu}{L+\mu}$. Then

$$\|x_{t+1} - x^*\| = \|P_C(x_t) - P_C(x^*)\| \leq \frac{L - \mu}{L + \mu}\|x_t - x^*\|$$

□

Therefore, we achieve the same convergence rate as gradient descent methods for projected gradient descent.