# Projected Gradient Descent

Cong Ma

University of Chicago, Winter 2026

# Constrained optimization problems

$$\begin{array}{ll} \text{minimize}_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{C} \end{array}$$

- $f(\cdot)$: convex function
- $\mathcal{C} \subseteq \mathbb{R}^n$: closed convex set

# Stationary Points for Constrained Problems

Let $C \subseteq \mathbb{R}^d$ be a closed, convex set and let $f : C \to \mathbb{R}$ be continuously differentiable.

We consider the constrained problem

$$\text{(P)} \qquad \min_{x \in C} \; f(x).$$

**Definition 1 (Stationary point)**

A point $x^\star \in C$ is called a stationary point of (P) if

$$\nabla f(x^\star)^\top (x - x^\star) \;\geq\; 0, \qquad \forall x \in C.$$

- Interpretation: no feasible descent direction.

# Stationarity Is Necessary for Local Optimality

Let $C \subseteq \mathbb{R}^d$ be closed and convex, and let $f : C \to \mathbb{R}$ be continuously differentiable. Consider

$$\text{(P)} \qquad \min_{x \in C} f(x).$$

**Theorem 2 (Stationarity as a necessary condition)**

*If $x^\star$ is a local minimum of (P), then $x^\star$ is a stationary point of (P), i.e.*

$$\nabla f(x^\star)^\top (x - x^\star) \;\geq\; 0, \qquad \forall x \in C.$$

# For Convex Problems, Stationary $\Leftrightarrow$ Optimal

Let $C \subseteq \mathbb{R}^d$ be closed and convex, and let $f : C \to \mathbb{R}$ be continuously differentiable and convex. Consider

$$\text{(P)} \qquad \min_{x \in C} f(x).$$

**Theorem 3 (Stationarity is necessary and sufficient)**

*A point $x^\star \in C$ is stationary for (P), i.e.*

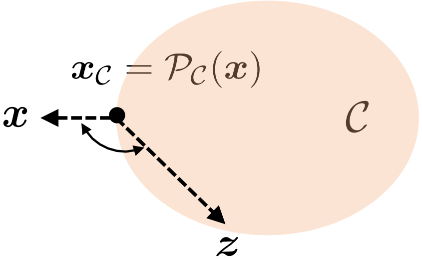
$$\nabla f(x^\star)^\top (x - x^\star) \geq 0 \quad \forall x \in C,$$

*if and only if $x^\star$ is a global optimal solution of (P).*

# Projection onto Closed and Convex Set

Let $C \subseteq \mathbb{R}^d$ be nonempty, closed and convex. Given $x \in \mathbb{R}^d$, define the projection

$$x_C := P_C(x) \in \arg\min_{z \in C} \|z - x\|_2^2.$$

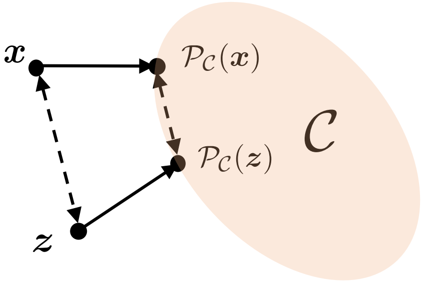



**Fact 4 (Projection theorem)**

*Let $\mathcal{C}$ be closed & convex. Then $\boldsymbol{x}_{\mathcal{C}}$ is the projection of $\boldsymbol{x}$ onto $\mathcal{C}$ iff*

$$(\boldsymbol{x} - \boldsymbol{x}_{\mathcal{C}})^\top (\boldsymbol{z} - \boldsymbol{x}_{\mathcal{C}}) \le 0, \qquad \forall \boldsymbol{z} \in \mathcal{C}$$

# Aside: nonexpansiveness of projection operator




**Fact 5 (Nonexpansivness of projection)**

*For any* $\boldsymbol{x}$ *and* $\boldsymbol{z}$, *one has* $\|\mathcal{P}_{\mathcal{C}}(\boldsymbol{x}) - \mathcal{P}_{\mathcal{C}}(\boldsymbol{z})\|_2 \leq \|\boldsymbol{x} - \boldsymbol{z}\|_2$

# Projection is Firmly Nonexpansive

Let $C \subseteq \mathbb{R}^d$ be nonempty, closed, and convex. Denote the projection by

$$P_C(v) := \arg\min_{x \in C} \|x - v\|_2.$$

**Theorem 6**

*For any $v, w \in \mathbb{R}^d$:*

1. ***(Firm nonexpansiveness)***

$$\left(P_C(v) - P_C(w)\right)^\top (v - w) \;\geq\; \|P_C(v) - P_C(w)\|_2^2. \tag{9.9}$$

2. ***(Nonexpansiveness)***

$$\|P_C(v) - P_C(w)\|_2 \;\leq\; \|v - w\|_2. \tag{9.10}$$

# Key tool: Projection theorem (Theorem 9.8)

Recall: $y = P_C(x)$ iff

$$(x - y)^\top (z - y) \leq 0, \qquad \forall z \in C. \tag{9.11}$$

We will apply (9.11) twice, with the choices:

$$(x, y, z) = (v,\, P_C(v),\, P_C(w)) \quad \text{and} \quad (x, y, z) = (w,\, P_C(w),\, P_C(v)).$$

# Step 1: two inequalities from (9.11)

Substitute $x = v$, $y = P_C(v)$, $z = P_C(w)$ into (9.11):

$$\big(v - P_C(v)\big)^\top \big(P_C(w) - P_C(v)\big) \leq 0. \tag{9.12}$$

Substitute $x = w$, $y = P_C(w)$, $z = P_C(v)$ into (9.11):

$$\big(w - P_C(w)\big)^\top \big(P_C(v) - P_C(w)\big) \leq 0. \tag{9.13}$$

# Step 2: add them and rearrange to get (9.9)

Add (9.12) and (9.13). Noting that

$$(P_C(w) - P_C(v)) = -(P_C(v) - P_C(w)),$$

we obtain

$$\big(P_C(v) - P_C(w)\big)^\top (v - w) \;\geq\; \|P_C(v) - P_C(w)\|_2^2,$$

which is exactly (9.9). $\square$

# Step 3: deduce nonexpansiveness (9.10) from (9.9)

If $P_C(v) = P_C(w)$, then (9.10) is trivial.

Otherwise, apply Cauchy–Schwarz:

$$\big(P_C(v) - P_C(w)\big)^\top (v - w) \le \|P_C(v) - P_C(w)\|_2 \, \|v - w\|_2.$$

Combine with (9.9):

$$\|P_C(v) - P_C(w)\|_2^2 \le \|P_C(v) - P_C(w)\|_2 \, \|v - w\|_2.$$

Divide by $\|P_C(v) - P_C(w)\|_2 > 0$ to get

$$\|P_C(v) - P_C(w)\|_2 \le \|v - w\|_2,$$

which is (9.10). $\square$

# Stationarity $\Leftrightarrow$ Projected-Gradient Fixed Point

Let $C \subseteq \mathbb{R}^d$ be closed and convex, and let $f : C \to \mathbb{R}$ be continuously differentiable. Fix any stepsize $s > 0$.

Consider the constrained problem

$$\text{(P)} \qquad \min_{x \in C} f(x).$$

**Theorem 7**

*A point $x^\star \in C$ is a stationary point of (P), i.e.*

$$\nabla f(x^\star)^\top (x - x^\star) \geq 0 \quad \forall x \in C,$$

*if and only if*

$$x^\star = P_C\big(x^\star - s\nabla f(x^\star)\big). \tag{9.14}$$

# Proof: apply the projection theorem to $x^\star - s\nabla f(x^\star)$

Recall the projection theorem (Theorem 9.8):

$$y = P_C(x) \quad \iff \quad (x-y)^\top (z-y) \le 0, \;\; \forall z \in C.$$

Apply it with

$$x = x^\star - s\nabla f(x^\star), \qquad y = x^\star.$$

Then

$$x^\star = P_C\big(x^\star - s\nabla f(x^\star)\big)$$

holds iff for all $z \in C$,

$$\big((x^\star - s\nabla f(x^\star)) - x^\star\big)^\top (z - x^\star) \le 0.$$

# Simplify: fixed point condition $\Leftrightarrow$ stationarity

The inequality becomes

$$(-s\nabla f(x^\star))^\top (z - x^\star) \le 0 \qquad \forall z \in C.$$

Since $s > 0$, divide by $s$ and multiply by $-1$:

$$\nabla f(x^\star)^\top (z - x^\star) \ge 0 \qquad \forall z \in C.$$

This is exactly the stationarity condition. $\square$

# Remark: the condition does *not* really depend on $s$

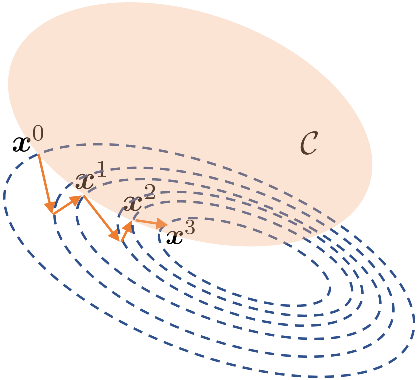
Although (9.14) is written with a stepsize $s > 0$, the equivalence shows:

$$x^\star \text{ stationary} \quad \Longleftrightarrow \quad x^\star = P_C(x^\star - s\nabla f(x^\star)) \ \text{ for } \textit{any } s > 0.$$

- This is the fixed-point form behind the projected gradient method:
$$x^{t+1} = P_C\big(x^t - s\nabla f(x^t)\big).$$
- Stationary points are exactly the fixed points of this iteration.

# Projected gradient methods

# Projected gradient descent



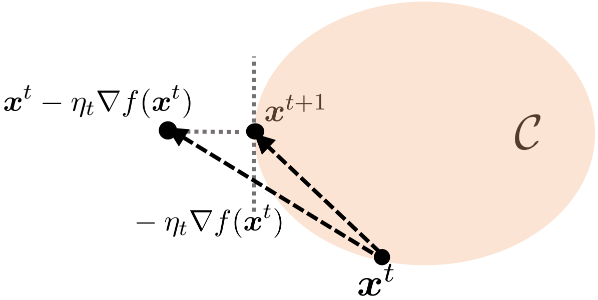

works well if projection onto $\mathcal{C}$ can be computed efficiently

**for** $t = 0, 1, \cdots$ :

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

where $\mathcal{P}_{\mathcal{C}}(x) := \arg\min_{z \in \mathcal{C}} \|x - z\|_2^2$ is $\underbrace{\text{Euclidean projection}}_{\text{quadratic minimization}}$ onto $\mathcal{C}$

# Descent direction




From the above figure, we know

$$-\nabla f(\boldsymbol{x}^t)^\top (\boldsymbol{x}^{t+1} - \boldsymbol{x}^t) \geq 0$$

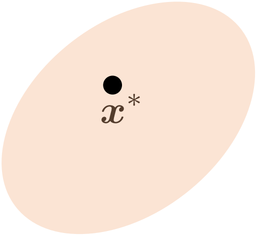
$\boldsymbol{x}^{t+1} - \boldsymbol{x}^t$ is positively correlated with the steepest descent direction

# Strongly convex and smooth problems

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{C} \end{aligned}$$

- $f(\cdot)$: $\mu$-strongly convex and $L$-smooth
- $\mathcal{C} \subseteq \mathbb{R}^n$: closed and convex

# Convergence for strongly convex and smooth problems




Let's start with the simple case when $x^*$ lies in the interior of $\mathcal{C}$ (so that $\nabla f(x^*) = \mathbf{0}$)

# Convergence for strongly convex and smooth problems

**Theorem 8**

*Suppose* $\boldsymbol{x}^* \in \mathrm{int}(\mathcal{C})$, *and let* $f$ *be* $\mu$*-strongly convex and* $L$*-smooth. If* $\eta_t = \frac{2}{\mu+L}$, *then*

$$\|\boldsymbol{x}^t - \boldsymbol{x}^*\|_2 \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^t \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2$$

*where* $\kappa = L/\mu$ *is condition number*

- the same convergence rate as for the unconstrained case

# Proof of Theorem 8

We have shown for the unconstrained case with $\nabla f(\boldsymbol{x}^*) = 0$ that

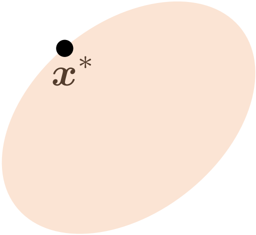
$$\|\boldsymbol{x}^t - \eta_t \nabla f(\boldsymbol{x}^t) - \boldsymbol{x}^*\|_2 \leq \frac{\kappa - 1}{\kappa + 1} \|\boldsymbol{x}^t - \boldsymbol{x}^*\|_2$$

From the nonexpansiveness of $\mathcal{P}_{\mathcal{C}}$, we know

$$\begin{aligned}
\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^*\|_2 &= \|\mathcal{P}_{\mathcal{C}}(\boldsymbol{x}^t - \eta_t \nabla f(\boldsymbol{x}^t)) - \mathcal{P}_{\mathcal{C}}(\boldsymbol{x}^*)\|_2 \\
&\leq \|\boldsymbol{x}^t - \eta_t \nabla f(\boldsymbol{x}^t) - \boldsymbol{x}^*\|_2 \\
&\leq \frac{\kappa - 1}{\kappa + 1} \|\boldsymbol{x}^t - \boldsymbol{x}^*\|_2
\end{aligned}$$

Apply it recursively to conclude the proof

# Convergence for strongly convex and smooth problems




What happens if we don't know whether $\boldsymbol{x}^* \in \text{int}(\mathcal{C})$?

- main issue: $\nabla f(\boldsymbol{x}^*)$ may not be $\mathbf{0}$ (so prior analysis might fail)

# Convergence for strongly convex and smooth problems

**Theorem 9 (projected GD for strongly convex and smooth problems)**

*Let $f$ be $\mu$-strongly convex and $L$-smooth. If $\eta_t \equiv \eta = \frac{1}{L}$, then*

$$\|\boldsymbol{x}^t - \boldsymbol{x}^*\|_2^2 \leq \left(1 - \frac{\mu}{L}\right)^t \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2^2$$

- slightly weaker convergence guarantees than Theorem 8

# Proof of Theorem 9: one-step contraction

Set $\eta = \frac{1}{L}$ and

$$x^+ := P_C\big(x - \eta \nabla f(x)\big), \qquad g_C(x) := \frac{1}{\eta}(x - x^+) = L(x - x^+).$$

Then the update is $x^{t+1} = x^+ = (x^t)^+$.

**Goal:** show the one-step contraction

$$\|x^+ - x^\star\|_2^2 \ \le\ \left(1 - \frac{\mu}{L}\right) \|x - x^\star\|_2^2,$$

then apply it recursively with $x = x^t$.

**Key inequality (from projection):**

$$\langle \nabla f(x) - g_C(x),\, x^\star - x^+ \rangle \ \le\ 0. \qquad (\clubsuit)$$

(This is Fact 4 applied to $x - \eta \nabla f(x)$ and $x^+ = P_C(\cdot)$.)

# Two ingredients + one line of algebra

**(1) Smoothness (descent lemma)** with $\eta = \frac{1}{L}$:

$$f(x^+) \leq f(x) + \langle \nabla f(x), x^+ - x \rangle + \frac{L}{2}\|x^+ - x\|_2^2 = f(x) - \frac{1}{2L}\|g_C(x)\|_2^2. \tag{A}$$

**(2) Strong convexity**:

$$f(x) - f(x^\star) \leq \langle \nabla f(x), x - x^\star \rangle - \frac{\mu}{2}\|x - x^\star\|_2^2. \tag{B}$$

Combine (A)+(B) and use $f(x^+) \geq f(x^\star)$:

$$0 \leq f(x^+) - f(x^\star) \leq \langle \nabla f(x), x^+ - x^\star \rangle + \frac{1}{2L}\|g_C(x)\|_2^2 - \frac{\mu}{2}\|x - x^\star\|_2^2. \tag{C}$$

Now apply (♣) to replace $\langle \nabla f(x), x^+ - x^\star \rangle$ by $\langle g_C(x), x^+ - x^\star \rangle$ (projection only helps).

# Finish: regularity $\Rightarrow$ contraction

From (C) and (♣):

$$\langle g_C(x), x^+ - x^\star \rangle \geq \frac{\mu}{2}\|x - x^\star\|_2^2 - \frac{1}{2L}\|g_C(x)\|_2^2.$$

Since $x^+ = x - \frac{1}{L}g_C(x)$, we have

$$\langle g_C(x), x^+ - x^\star \rangle = \langle g_C(x), x - x^\star \rangle - \frac{1}{L}\|g_C(x)\|_2^2.$$

Therefore the regularity inequality holds:

$$\boxed{\langle g_C(x), x - x^\star \rangle \geq \frac{\mu}{2}\|x - x^\star\|_2^2 + \frac{1}{2L}\|g_C(x)\|_2^2} \qquad \text{(R)}$$

Finally,

$$\begin{aligned}
\|x^+ - x^\star\|_2^2 &= \left\|x - x^\star - \frac{1}{L} g_C(x)\right\|_2^2 \\
&= \|x - x^\star\|_2^2 - \frac{2}{L}\langle g_C(x), x - x^\star\rangle + \frac{1}{L^2}\|g_C(x)\|_2^2 \\
&\le \|x - x^\star\|_2^2 - \frac{\mu}{L}\|x - x^\star\|_2^2 = \left(1 - \frac{\mu}{L}\right)\|x - x^\star\|_2^2,
\end{aligned}$$

where we used (R). Apply recursively to conclude. □

# Convex and smooth problems

$$\begin{array}{ll} \text{minimize}_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{C} \end{array}$$

- $f(\cdot)$: convex and $L$-smooth
- $\mathcal{C} \subseteq \mathbb{R}^n$: closed and convex

# Convergence for convex and smooth problems

**Theorem 10 (projected GD for convex and smooth problems)**

*Let $f$ be convex and $L$-smooth. If $\eta_t \equiv \eta = \frac{1}{L}$, then*

$$f(\boldsymbol{x}^t) - f(\boldsymbol{x}^*) \leq \frac{3L\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2^2 + f(\boldsymbol{x}^0) - f(\boldsymbol{x}^*)}{t+1}$$

- similar convergence rate as for the unconstrained case
- cannot replace $f(\boldsymbol{x}^0) - f(\boldsymbol{x}^*)$ with $\frac{1}{2}L\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2^2$ since in general $\nabla f(\boldsymbol{x}^*) \neq 0$

# Proof of Theorem 10

We first recall our main steps when handling the unconstrained case

**Step 1:** show cost improvement

$$f(\boldsymbol{x}^{t+1}) \leq f(\boldsymbol{x}^t) - \frac{1}{2L}\|\nabla f(\boldsymbol{x}^t)\|_2^2$$

**Step 2:** connect $\|\nabla f(\boldsymbol{x}^t)\|_2$ with $f(\boldsymbol{x}^t)$

$$\|\nabla f(\boldsymbol{x}^t)\|_2 \geq \frac{f(\boldsymbol{x}^t) - f(\boldsymbol{x}^*)}{\|\boldsymbol{x}^t - \boldsymbol{x}^*\|_2} \geq \frac{f(\boldsymbol{x}^t) - f(\boldsymbol{x}^*)}{\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2}$$

**Step 3:** let $\Delta_t := f(\boldsymbol{x}^t) - f(\boldsymbol{x}^*)$ to get

$$\Delta_{t+1} - \Delta_t \leq -\frac{\Delta_t^2}{2L\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2^2}$$

and complete the proof by induction

# Proof of Theorem 10 (cont.)

We then modify these steps for the constrained case. As before, set $\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x}^t) = L(\boldsymbol{x}^t - \boldsymbol{x}^{t+1})$, which generalizes $\nabla f(\boldsymbol{x}^t)$ in constrained case

**Step 1:** show cost improvement

$$f(\boldsymbol{x}^{t+1}) \leq f(\boldsymbol{x}^t) - \frac{1}{2L}\|\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x}^t)\|_2^2$$

**Step 2:** connect $\|\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x}^t)\|_2$ with $f(\boldsymbol{x}^t)$

$$\|\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x}^t)\|_2 \geq \frac{f(\boldsymbol{x}^{t+1}) - f(\boldsymbol{x}^*)}{\|\boldsymbol{x}^t - \boldsymbol{x}^*\|_2} \geq \frac{f(\boldsymbol{x}^{t+1}) - f(\boldsymbol{x}^*)}{\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2}$$

**Step 3:** let $\Delta_t := f(\boldsymbol{x}^t) - f(\boldsymbol{x}^*)$ to get

$$\Delta_{t+1} - \Delta_t \leq -\frac{\Delta_{t+1}^2}{2L\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2^2}$$

and complete the proof by induction

# Proof of Theorem 10 (cont.)

**Main pillar:** generalize smoothness condition (under convexity) as follows

**Lemma 11**

*Suppose* $f$ *is convex and* $L$*-smooth. For any* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$, *let* $\boldsymbol{x}^+ = \mathcal{P}_{\mathcal{C}}(\boldsymbol{x} - \frac{1}{L}\nabla f(\boldsymbol{x}))$ *and* $g_{\mathcal{C}}(\boldsymbol{x}) = L(\boldsymbol{x} - \boldsymbol{x}^+)$. *Then*

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}^+) + g_{\mathcal{C}}(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2L}\|g_{\mathcal{C}}(\boldsymbol{x})\|_2^2$$

# Proof of Theorem 10 (cont.)

**Step 1:** set $\boldsymbol{x} = \boldsymbol{y} = \boldsymbol{x}^t$ in Lemma 11 to reach

$$f(\boldsymbol{x}^t) \geq f(\boldsymbol{x}^{t+1}) + \frac{1}{2L}\|g_{\mathcal{C}}(\boldsymbol{x}^t)\|_2^2$$

as desired

**Step 2:** set $\boldsymbol{x} = \boldsymbol{x}^t$ and $\boldsymbol{y} = \boldsymbol{x}^*$ in Lemma 11 to get

$$\begin{aligned} 0 \geq f(\boldsymbol{x}^*) - f(\boldsymbol{x}^{t+1}) &\geq g_{\mathcal{C}}(\boldsymbol{x}^t)^\top(\boldsymbol{x}^* - \boldsymbol{x}^t) + \frac{1}{2L}\|g_{\mathcal{C}}(\boldsymbol{x}^t)\|_2^2 \\ &\geq g_{\mathcal{C}}(\boldsymbol{x}^t)^\top(\boldsymbol{x}^* - \boldsymbol{x}^t) \end{aligned}$$

which together with Cauchy-Schwarz yields

$$\|\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x}^t)\|_2 \geq \frac{f(\boldsymbol{x}^{t+1}) - f(\boldsymbol{x}^*)}{\|\boldsymbol{x}^t - \boldsymbol{x}^*\|_2} \tag{1}$$

# Proof of Theorem 10 (cont.)

It also follows from our analysis for the strongly convex case that (by taking $\mu = 0$ in Theorem 9)

$$\|\boldsymbol{x}^t - \boldsymbol{x}^*\|_2 \leq \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2$$

which combined with (1) reveals

$$\|\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x}^t)\|_2 \geq \frac{f(\boldsymbol{x}^{t+1}) - f(\boldsymbol{x}^*)}{\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2}$$

**Step 3:** letting $\Delta_t = f(\boldsymbol{x}^t) - f(\boldsymbol{x}^*)$, the previous bounds together give

$$\Delta_{t+1} - \Delta_t \leq -\frac{\Delta_{t+1}^2}{2L\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2^2}$$

Use induction to finish the proof (which we omit here)

# Proof of Lemma 11

$$f(\boldsymbol{y}) - f(\boldsymbol{x}^+) = f(\boldsymbol{y}) - f(\boldsymbol{x}) - \big(f(\boldsymbol{x}^+) - f(\boldsymbol{x})\big)$$

$$\geq \underbrace{\nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x})}_{\text{convexity}} - \underbrace{\left(\nabla f(\boldsymbol{x})^\top (\boldsymbol{x}^+ - \boldsymbol{x}) + \frac{L}{2}\|\boldsymbol{x}^+ - \boldsymbol{x}\|_2^2\right)}_{\text{smoothness}}$$

$$= \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}^+) - \frac{L}{2}\|\boldsymbol{x}^+ - \boldsymbol{x}\|_2^2$$

$$\geq \boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}^+) - \frac{L}{2}\|\boldsymbol{x}^+ - \boldsymbol{x}\|_2^2 \qquad \text{(by (??))}$$

$$= \boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x})^\top \underbrace{(\boldsymbol{x} - \boldsymbol{x}^+)}_{=\frac{1}{L}\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x})} - \frac{L}{2}\|\underbrace{\boldsymbol{x}^+ - \boldsymbol{x}}_{=-\frac{1}{L}\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x})}\|_2^2$$

$$= \boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2L}\|\boldsymbol{g}_{\mathcal{C}}(\boldsymbol{x})\|_2^2$$

# Summary

- Frank-Wolfe: projection-free

| | stepsize rule | convergence rate | iteration complexity |
|---|---|---|---|
| convex & smooth problems | $\eta_t \asymp \frac{1}{t}$ | $O\left(\frac{1}{t}\right)$ | $O\left(\frac{1}{\varepsilon}\right)$ |

- projected gradient descent

| | stepsize rule | convergence rate | iteration complexity |
|---|---|---|---|
| convex & smooth problems | $\eta_t = \frac{1}{L}$ | $O\left(\frac{1}{t}\right)$ | $O\left(\frac{1}{\varepsilon}\right)$ |
| strongly convex & smooth problems | $\eta_t = \frac{1}{L}$ | $O\left(\left(1 - \frac{1}{\kappa}\right)^t\right)$ | $O\left(\kappa \log \frac{1}{\varepsilon}\right)$ |