262B-Lecture 12

Date created: 2021.03.02
N. of Pages: 15

$\min \; f(x) \quad \text{s.t.} \quad x \in X$

Assume: X = Convex set, f = arbitrary

Assume: x_* satisfies FOC and

$$\Delta x^T \nabla^2 f(x_*) \Delta x > 0 \quad \forall \, \Delta x \quad \text{s.t.}$$

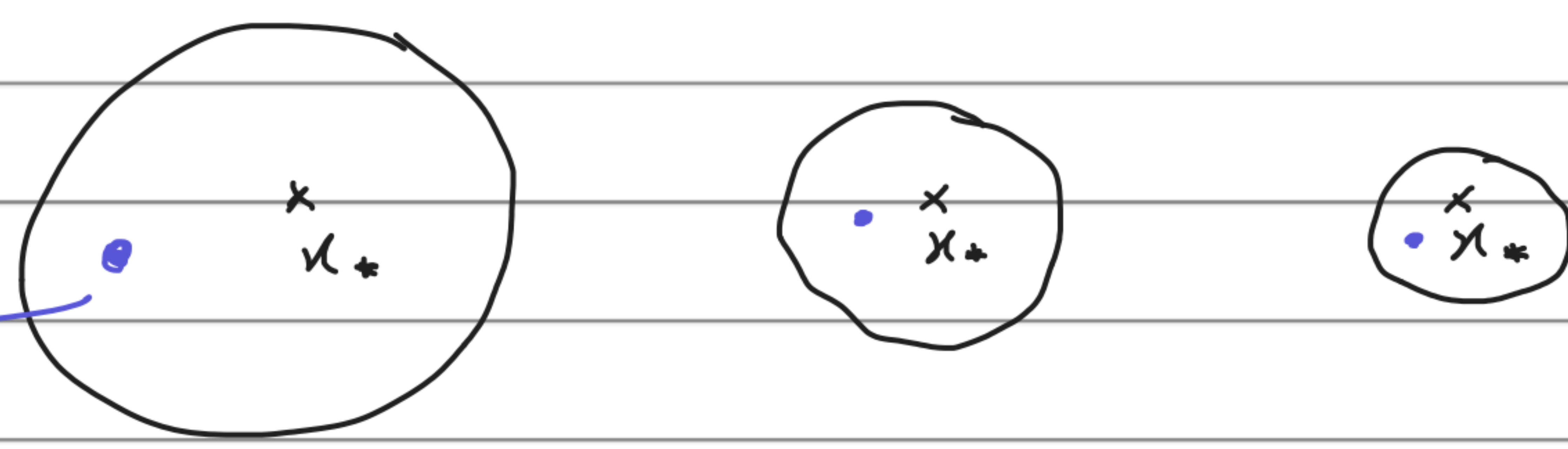
$$\Delta x \neq 0, \quad \Delta x \in T_X(x_*), \quad \nabla f(x_*)^T \Delta x = 0$$

$\Rightarrow$ x_* = local min

SOC sufficient

Proof: Assume the Conditions are satisfied but x_* is not a local min (proof by contradiction)

$\Rightarrow$ $\forall$ Ball around x_* $\rightarrow$ there is a better point inside.



better point

$$\Rightarrow \; \exists \; \{x^{(k)}\}_{k=1}^{\infty} \subseteq X \quad \text{s.t.} \quad x^{(k)} \longrightarrow x_*$$

and $f(x^{(k)}) < f(x_*)$

$$f(x^{(k)}) = f(x_*) + \nabla f(x_*)^T (x^{(k)} - x_*)$$
$$+ \frac{1}{2} (x^{(k)} - x_*)^T \nabla^2 f(x_*) (x^{(k)} - x_*) + o(\|x^{(k)} - x_*\|^2)$$

1

Consider $\left\{ \frac{x^{(k)} - x_*}{\|x^{(k)} - x_*} \right\}_{k=1}^{\infty}$ $\longrightarrow$ all point are on sphere (bounded)

$\Rightarrow$ It has a convergent subsequence : K

Define : $\lim\limits_{\substack{k \to \infty \\ k \in K}} \frac{x^{(k)} - x_*}{\|x^{(k)} - x_*\|} = \Delta x$

FOC : $\nabla f(x_*)^T (\underbrace{x^{(k)}} - x_*) \geq 0$

$\Rightarrow \nabla f(x_*)^T \left(\frac{x^{(k)} - x_*}{\|x^{(k)} - x_*\|} \right) \geq 0$

Take the limit over K $\Rightarrow$

$$\boxed{\nabla f(x_*)^T \Delta x \geq 0}$$

2

Also,

$$① \Rightarrow \nabla f(x_*)^T \frac{x^{(k)} - x_*}{\|x^{(k)} - x_*\|} + \frac{1}{2} \|x^{(k)} - x_*\|$$

$$\times \underbrace{\left(\frac{x^{(k)} - x_*}{\|x^{(k)} - x_*\|}\right)^T \nabla^2 f(x_*) \left(\frac{x^{(k)} - x_*}{\|x^{(k)} - x_*\|}\right)}_{A} +$$

$$\frac{O(\|x^{(k)} - x_*\|^2)}{\|x^{(k)} - x_*\|} < 0 \qquad (f(x^{(k)}) < f(x_*))$$

As $k \to \infty$, $k \in \mathcal{K}$: $\|x^{(k)} - x_*\| \to 0$

$A \to$ bounded number, $\frac{O(\|x^{(k)} - x_*\|^2)}{\|x^{(k)} - x_*\|} \to 0$

$$\Rightarrow \boxed{\nabla f(x_*)^T \Delta x \leq 0} \qquad ③$$

$$②, ③ \Rightarrow \nabla f(x_*)^T \Delta x = 0$$

$$① \Rightarrow \boxed{\nabla f(x_*)^T (x^{(k)} - x_*)} + \frac{1}{2} (x^{(k)} - x_*)^T \nabla^2 f(x_*)$$

$$\times (x^{(k)} - x_*) + O(\|x^{(k)} - x_*\|^2) < 0$$

FOC : $\nabla f(x_*)^T (x^{(k)} - x_*) \geq 0$

$$\Rightarrow \frac{1}{2}\left(\frac{x^{(k)} - x_*}{\|x^{(k)} - x_*\|}\right)^T \nabla^2 f(x_*) \left(\frac{x^{(k)} - x_*}{\|x^{(k)} - x_*\|}\right)$$

$$+ \frac{O(\|x^{(k)} - x_*\|^2)}{\|x^{(k)} - x_*\|^2} < 0 \longrightarrow \text{Take the limit } k \to \infty, \; k \in K$$

$$\Rightarrow \Delta x^T \nabla^2 f(x_*) \Delta x \leq 0$$

This violates the assumption:

$$\Delta x^T \nabla^2 f(x_*) \Delta x > 0$$

Algorithms:

1_ Feasible directions: Conditional gradient methods

$\min \; f(x) \longleftarrow$ arbitrary

s.t. $x \in X \longleftarrow$ Convex

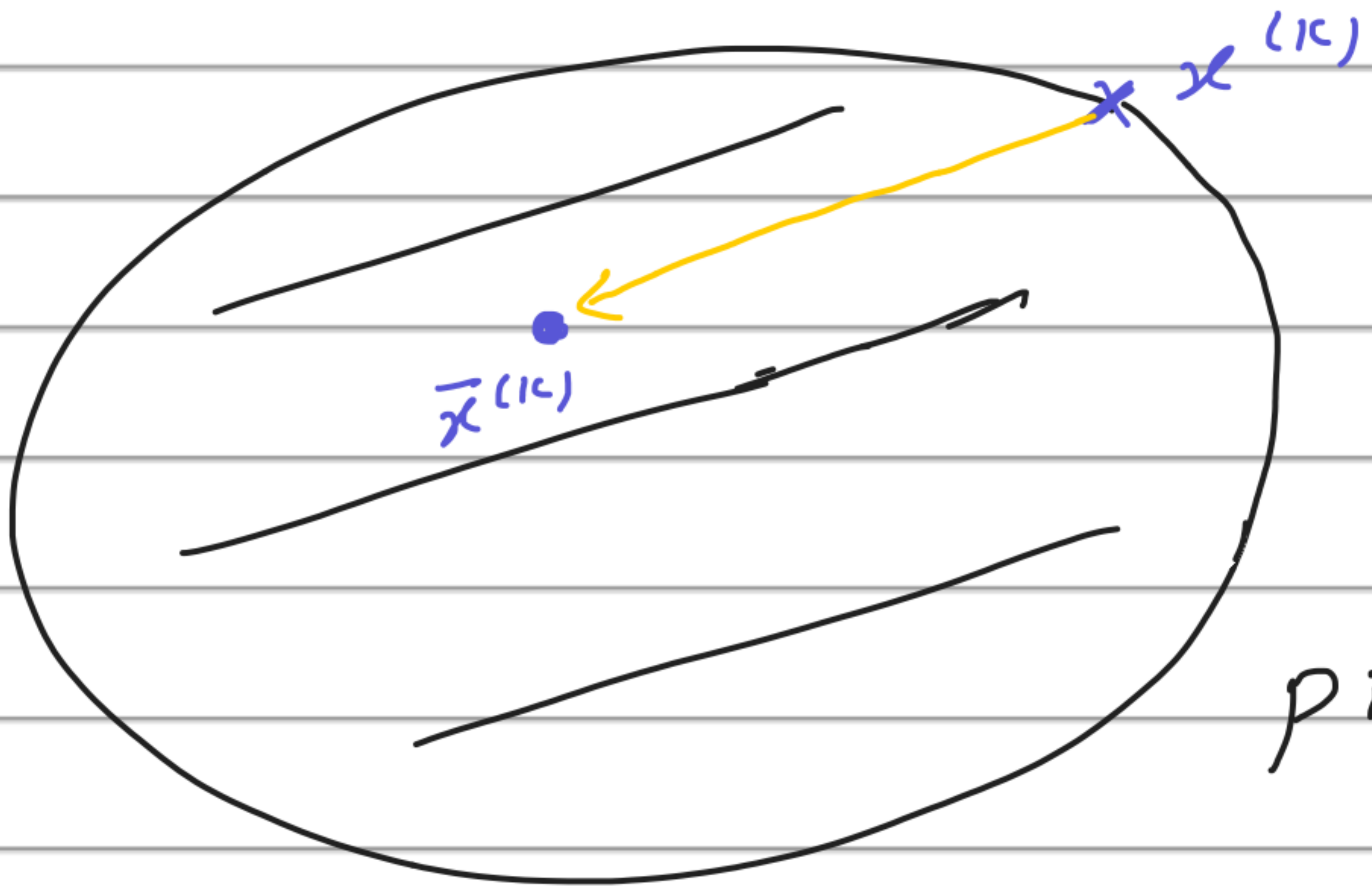
Alg.: $x^{(k+1)} = x^{(k)} + \alpha^{(k)} \underbrace{\Delta x^{(k)}}$

descent $\qquad \in F_X(x^{(k)})$

So, 1 - $\Delta x^{(k)} \in F_X(x^{(k)})$

2 - $\nabla f(x^{(k)})^T \Delta x^{(k)} < 0$

Choice : $\Delta x^{(k)} = \bar{x}^{(k)} - x^{(k)}$ where $\bar{x}^{(k)} \in X$



since $\bar{x}^{(k)} \in X \implies$

$\Delta x^{(k)} \in F_X(x^{(k)})$

pick $\alpha^{(k)} \in (0, 1]$

$$x^{(k)} + \alpha^{(k)}(\bar{x}^{(k)} - x^{(k)}) = \underbrace{(1-\alpha^{(k)})x^{(k)} + \alpha^{(k)}\bar{x}^{(k)}}$$

$x^{(k+1)} \in$ segment $\in X$

How to find $\alpha^{(k)}$:

1 - $\underbrace{\text{Limited}}$ Line search

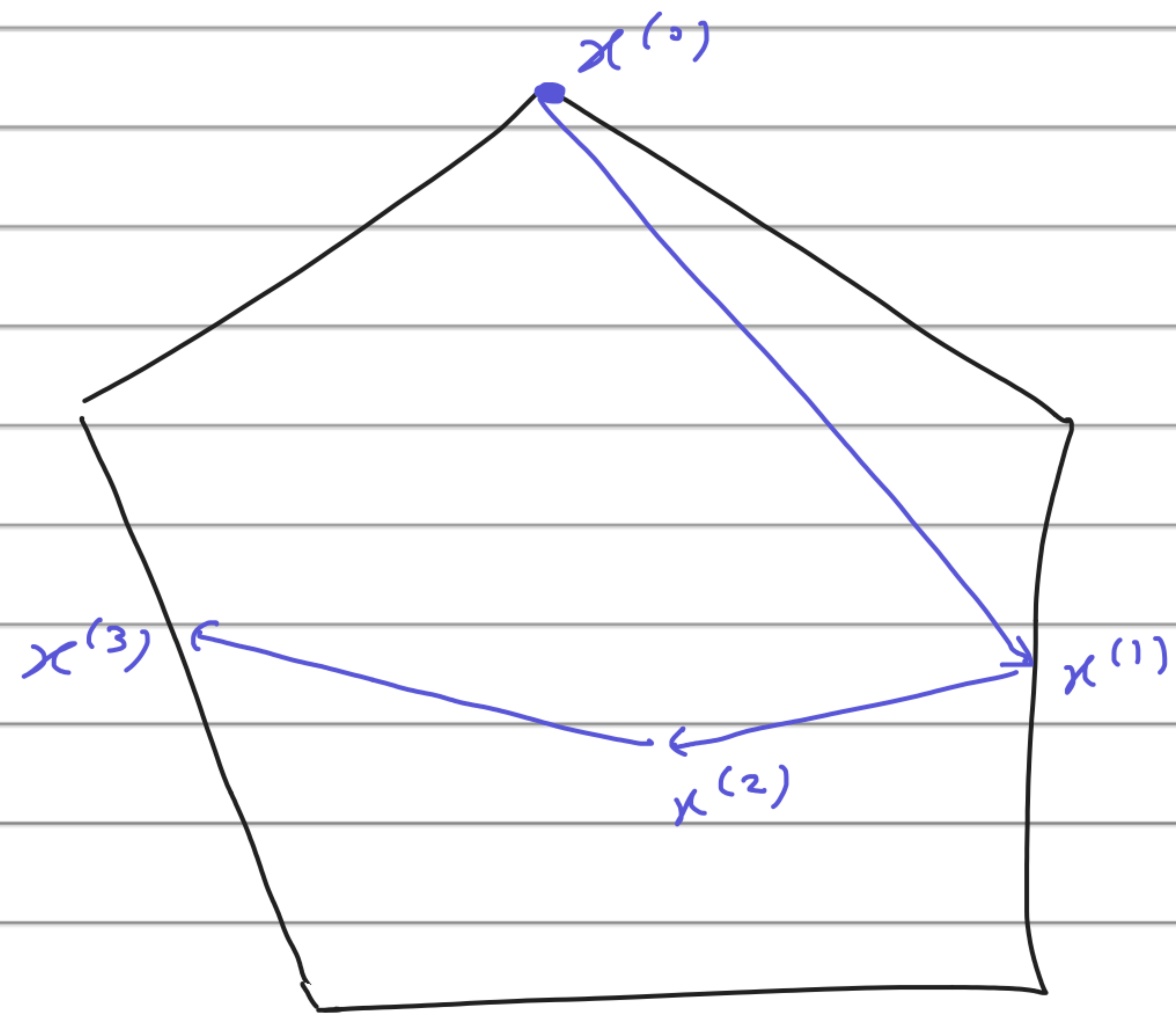
$$\alpha^{(k)} : \min_{\alpha} f(x^{(k)} + \alpha \Delta x^{(k)}) \quad \text{s.t. } 0 \le \alpha \le 1$$

2 - Constant stepsize $\alpha^{(k)} = 1$ (any constant)

3 - Armijo rule: $\alpha^{(k)}$: 1, β, β^2, $\vdots$

$0 < \sigma < 1$

Condition: $f(x^{(k+1)}) - f(x^{(k)}) < \sigma \, \alpha^{(k)} \, \nabla f(x^{(k)})^T \Delta x^{(k)}$



trapped inside the feasible set.

challenge: $\boxed{x^{(0)} \in X}$

$$\boxed{\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0 \quad i = 1, \dots, m \end{aligned}} \longrightarrow \boxed{\begin{aligned} \min_{x,y} \quad & f(x) + h(y) \\ & g_i(x) \le y_i, \quad y_i \ge 0 \\ & i = 1, \dots, m \end{aligned}}$$

hard to find a feasible point

$h(y)$: penalty that pushes y toward 0

(x, y) = convex set $\Leftarrow$ feasible point: x = arbitrary, y = large

Frank-wolfe : conditional gradient method

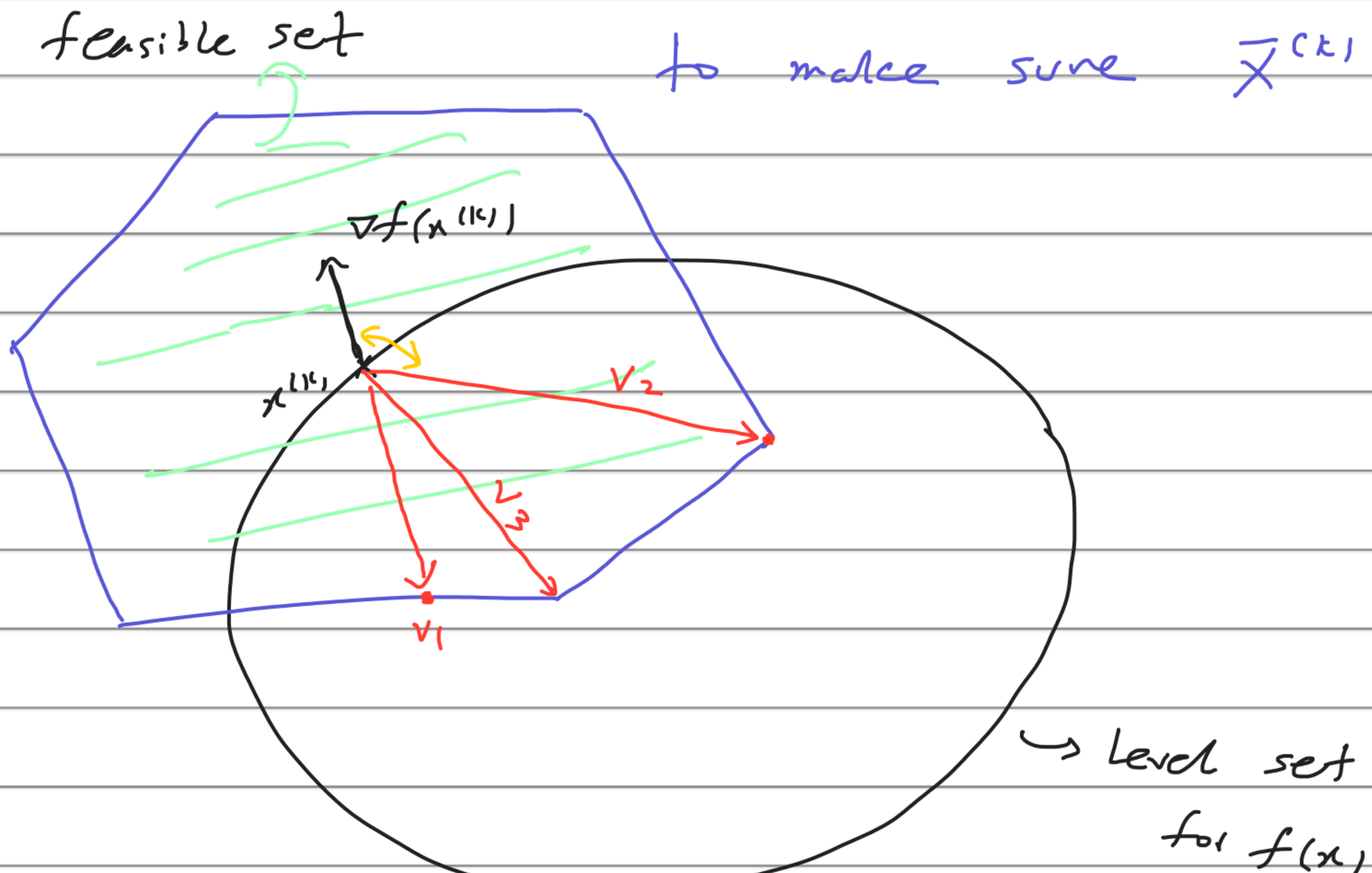
Find the most descent & feasible direction.

$$\min_{x} \ \nabla f(x^{(k)})^T \underbrace{(x - x^{(k)})}_{\Delta x^{(k)}}$$

s.t. $x \in X$ → solution : $\bar{x}^{(k)}$

assumption : X = Compact

to make sure $\bar{X}^{(k)}$ is not infinity



V_1 : direction of negative gradient, but small length

V_2 : long length but small angle with $\nabla f(x^{(k)})$

⇒ V_3 = best (has a trade-off between length & angle)

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} (\underbrace{\bar{x}^{(k)}} - x^{(k)})$$

Comes out of an optimization subproblem

$\Rightarrow$ The most expensive part of iteration update is solving for $\bar{x}^{(k)}$

Alg. :

$$\boxed{\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X \end{array}} \longrightarrow x^{(0)} \to x^{(1)} \to \cdots \to x^{(k)} \longrightarrow x^{(k+1)} \to \cdots$$

$$\boxed{\begin{array}{ll} \min & \nabla f(x^{(k)})^T (x - x^{(k)}) \\ \text{s.t.} & x \in X \end{array}}$$

$\Downarrow$

objective = Linear in $\underline{\underline{x}}$

assumption: minimization of $f(x)$ over X is hard but minimization of a Linear function over X is easy. $\Rightarrow$ X = nice structure.

subproblem = closed-form solution

Example: min $f(x)$ over simplex

$$X = \{ x \mid x \geq 0, \ \sum_{i=1}^{n} x_i = \text{given number } \underline{r} \}$$

$\bar{x}^{(k)}$:
$$\min \sum_{i=1}^{n} \frac{\partial f(x^{(k)})}{\partial x_i} (x_i - x_i^{(k)})$$
$$\text{s.t.} \sum_{i=1}^{n} x_i = r, \quad x_1, \ldots, x_n \geq 0$$

$$\Longrightarrow \bar{x}^{(k)} : \ \bar{x}_1^{(k)}, \ldots, \bar{x}_{j-1}^{(k)}, \bar{x}_j^{(k)}, \bar{x}_{j+1}^{(k)}, \ldots, \bar{x}_n^{(k)}$$
$$\qquad\qquad\quad \downarrow \qquad\quad \downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$
$$\qquad\qquad\quad 0 \qquad\quad 0 \qquad r \qquad 0 \qquad 0$$

$$j = \arg\min_{i=1,\ldots,n} \frac{\partial f(x^{(k)})}{\partial x_i}$$

subproblem has a closed-form solution.

Convergence: $x^{(k+1)} = x^{(k)} + \alpha^{(k)} (\underbrace{\bar{x}^{(k)}}_{\in X} - x^{(k)})$ (general method)

Need to make sure that $\bar{x}^{(k)} - x^{(k)}$

that is descent is not asymptotically orthogonal

to $\nabla f(x^{(k)}) \longrightarrow \{\bar{x}^{(k)} - x^{(k)}\}$ is gradient related

That means for any subsequence $\{x^{(k)}\}_{k\in K}$ that converges to a point not satisfying FOC, the corresponding sequence $\{\bar{x}^{(k)} - x^{(k)}\}_{k\in K}$ is:

- Bounded
- $\limsup\limits_{\substack{k\to\infty \\ k\in K}} \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) < 0$

Thm: If $\{\bar{x}^{(k)} - x^{(k)}\}$ is gradient related & $\alpha^{(k)}$ is designed based on limited line search or Armijo rule $\Rightarrow$ every limit point of $\{x^{(k)}\}$ $\underbrace{\text{satisfies FOC.}}_{\text{That point is called stationary point.}}$

Thm: Conditional gradient method generates gradient related directions.

suppose $\{x^{(k)}\}_{k\in K}$ Converges to a non-stationary point $\tilde{x}$. Need to prove:

1 - $\limsup_{\substack{k\to\infty \\ k\in K}} \|\bar{x}^{(k)} - x^{(k)}\| < \infty$

2 - $\limsup_{\substack{k\to\infty \\ k\in K}} \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) < 0$

1 is true since $\bar{x}^{(k)}, x^{(k)} \in X$ = Compact

For 2 : $\bar{x}^{(k)}$: $\min \nabla f(x^{(k)})^T(x - x^{(k)})$

s.t. $x \in X$

$\Rightarrow \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) \leq \nabla f(x^{(k)})^T(x - x^{(k)})$

$\forall x \in X$

Take Limit on $k\to\infty$, $k\in K$:

$\Rightarrow \limsup_{\substack{k\to\infty \\ k\in K}} \nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) \leq \nabla f(\tilde{x})^T(x - \tilde{x})$ $\forall x \in X$

since $\tilde{x} \neq$ a stationary point:

1

$\exists y : \nabla f(\tilde{x})^T(y - \tilde{x}) < 0$, $y \in X$

2

①, ② by choosing $x = y$ $\Longrightarrow$

$$\limsup_{\substack{k \to \infty \\ k \in K}} \nabla f(x^{(k)})^T (\bar{x}^{(k)} - x^{(k)}) < 0$$

Conditional gradient method: $\longrightarrow$ Find stationary points

Assume $\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \| \quad \forall x, y \in X$

Don't need to use Armijo rule or Line search,

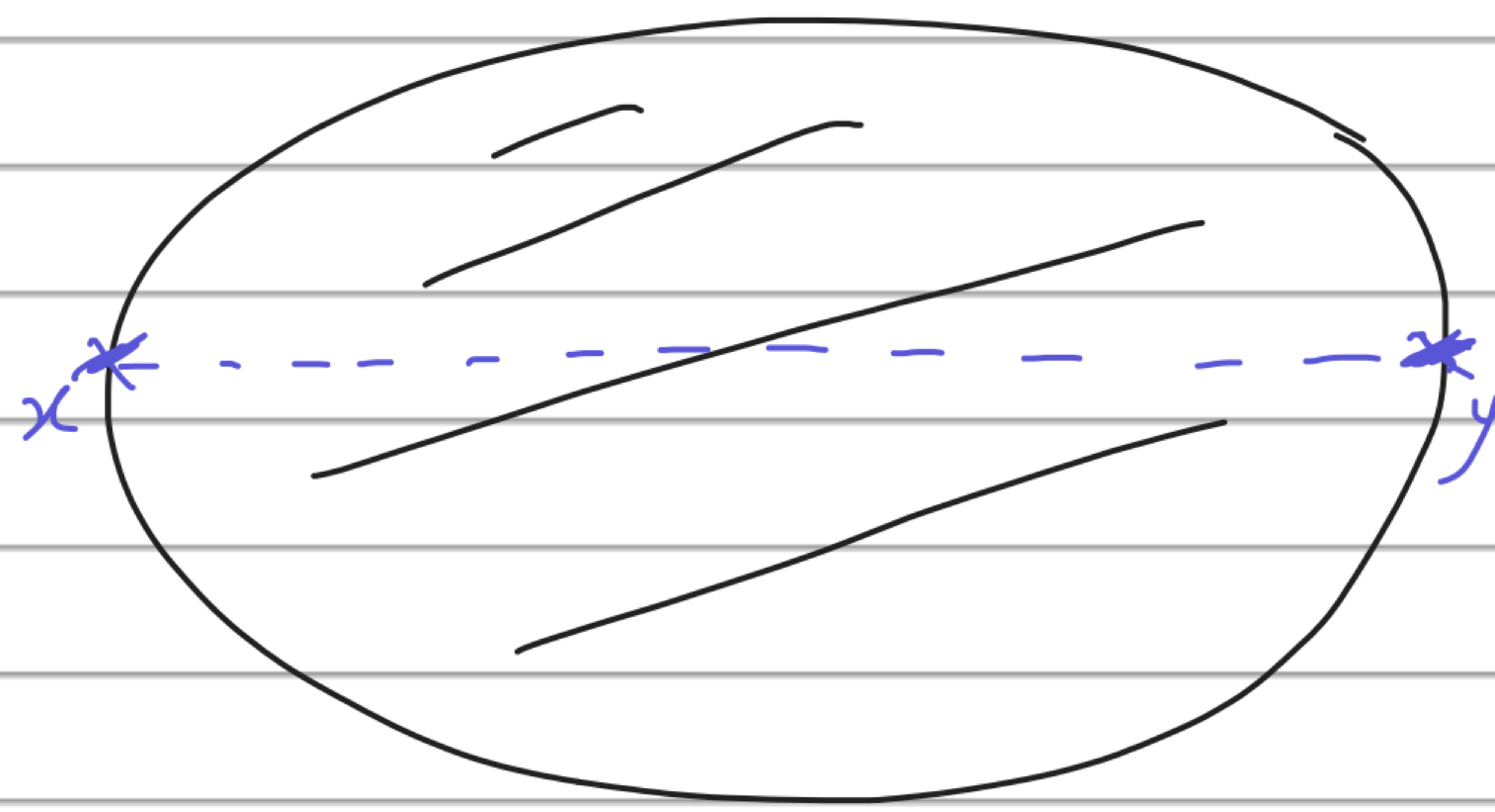
pick $\alpha^{(k)} = \min \left\{ 1, \dfrac{\nabla f(x^{(k)})^T (\bar{x}^{(k)} - x^{(k)})}{L \| \bar{x}^{(k)} - x^{(k)} \|^2} \right\}$ C

$\Rightarrow$ every limit point of $\{x^{(k)}\}$ is a stationary point (proof in HW 4)

Unconstrained & gradient method: $C = \frac{1}{L}$

Define diameter of set X:

$$D = \max_{x,y \in X} \|x - y\|$$



Thm: Assume $f(x)$ is convex & its gradient is Lipschitz continuous on X. Pick stepsize to be diminishing: $\alpha^{(k)} = \frac{2}{2+k}$

$\alpha^{(k)} \to 0$
$\sum \alpha^{(k)} = \infty$
$\sum (\alpha^{(k)})^2 <$

$$\Rightarrow \boxed{f(x^{(k)}) - f_* \le \frac{2LD^2}{k+2} \quad \forall k}$$

(*)

Proof: $f(x^{(k+1)}) = f(x^{(k)} + \alpha^{(k)}(\bar{x}^{(k)} - x^{(k)}))$

$$\le f(x^{(k)}) + \nabla f(x^{(k)})\left(\alpha^{(k)}(\bar{x}^{(k)} - x^{(k)})\right) + \frac{L}{2}\|\alpha^{(k)}(\bar{x}^{(k)} - x^{(k)})\|$$

(1)

quadratic over-estimator due to L.

$$\nabla f(x^{(k)})^T(\bar{x}^{(k)} - x^{(k)}) \le \nabla f(x^{(k)})^T(\underbrace{x_*}_{\text{arbitrary min}} - x^{(k)})$$

due to $\bar{x}^{(k)}$ being the most descent direction

(2)

$$\nabla f(x^{(k)})^T(x_* - x^{(k)}) \le \underbrace{f(x_*)}_{f_*} - f(x^{(k)})$$

due to convexity of $f(x)$

(3)

(1), (2), (3) $\Longrightarrow$

$$f(x^{(k+1)}) \le f(x^{(k)}) + \alpha^{(k)}(f_* - f(x^{(k)})) + \frac{L}{2}(\alpha^{(k)})^2 \underbrace{\|\bar{x}^{(k)} - x^{(k)}\|^2}_{\le D^2}$$

$$\Longrightarrow f(x^{(k+1)}) - f_* \le (1-\alpha^{(k)})(f(x^{(k)}) - f_*) + \frac{L}{2}(\alpha^{(k)})^2 D^2$$

pick : $\alpha^{(k)} = \frac{2}{2+k} \Longrightarrow$ induction : (*)

$$\Rightarrow \quad f(x^{(k)}) - f_* = O\left(\frac{1}{k}\right)$$

or iteration complexity: $O\left(\frac{1}{\varepsilon}\right)$