Convex Optimization

TTIC 31070 / CMSC 35470 / BUSF 36903 / STAT 31015

Prof. Nati Srebro

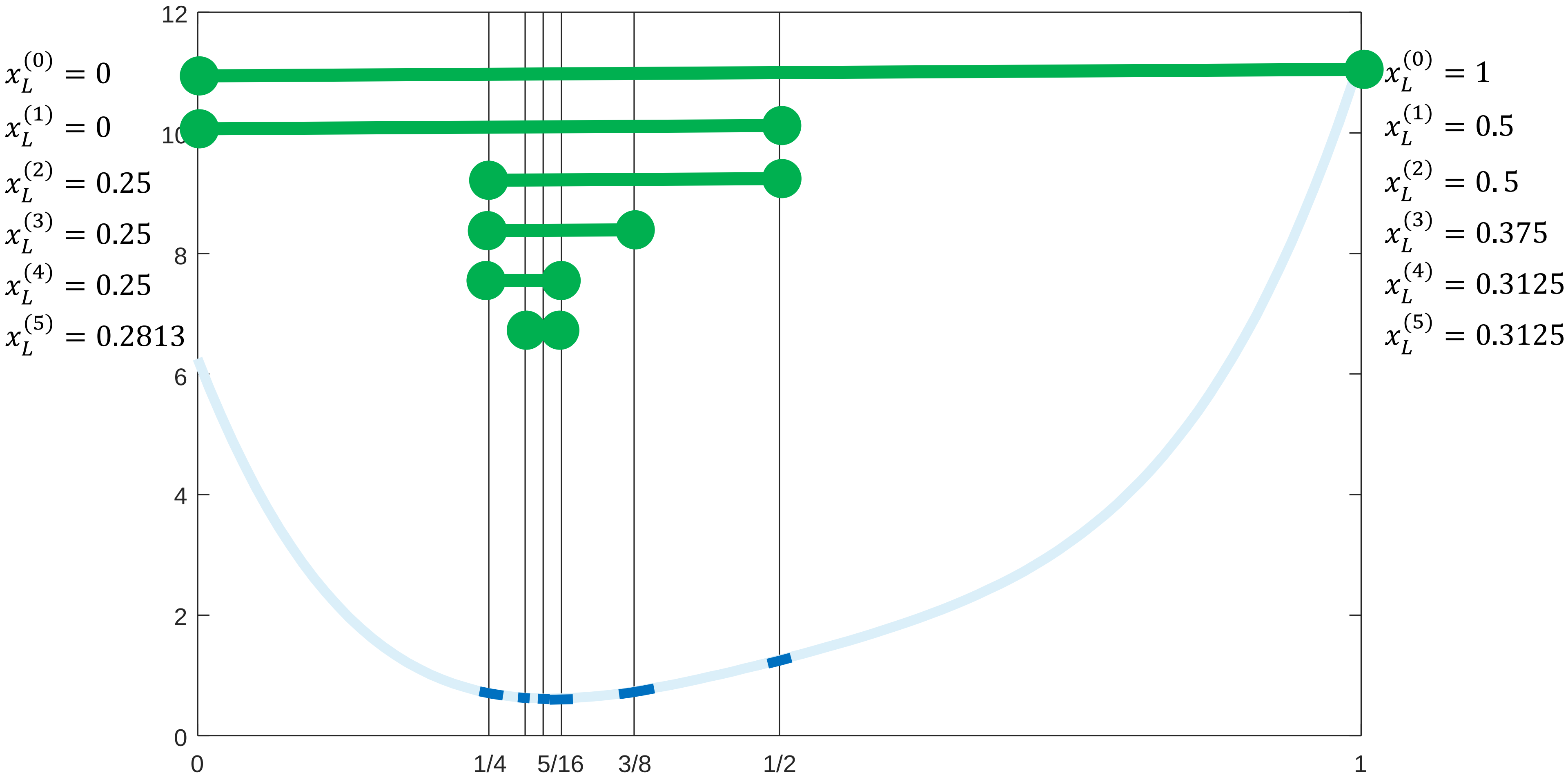
Lecture 2:

Convexity

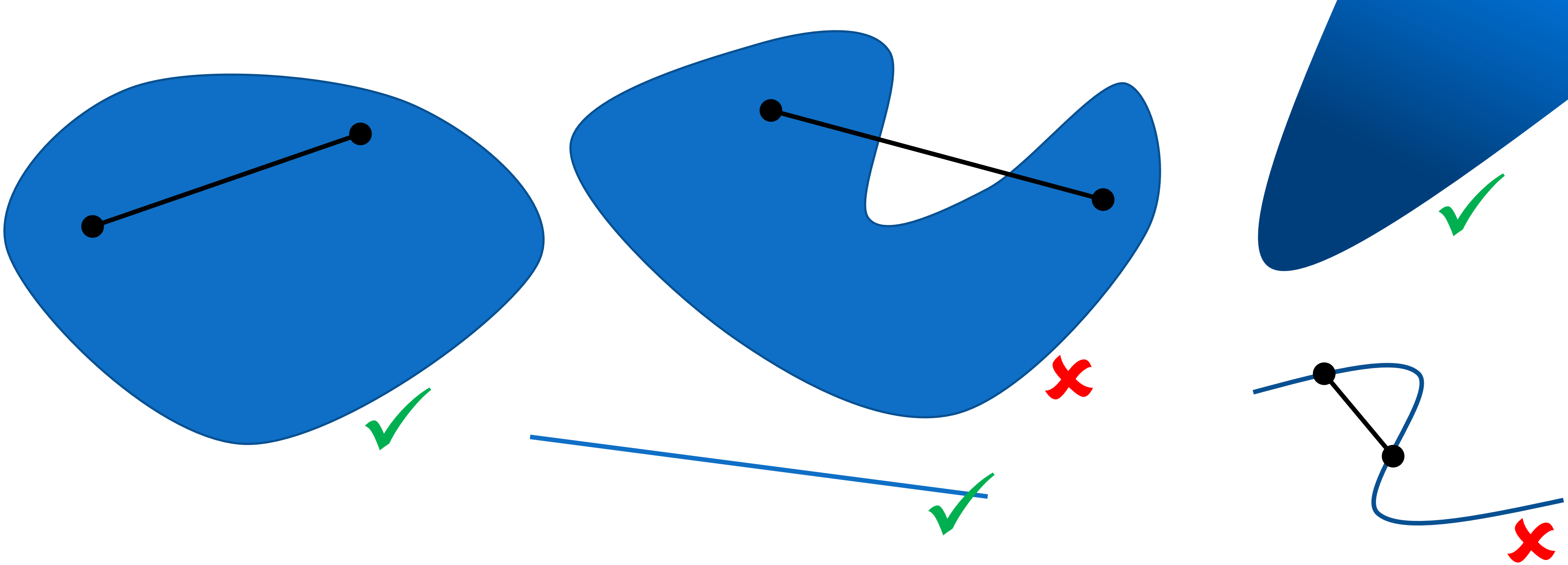
Center of Mass Algorithm

Suggested reading: Boyd and Vandenbergh 2.1-2.3,2.5,3.1-3.2 (Convexity)
or: Bubeck 1.2-1.3 (Convexity and Sub-Gradients), 2.1 (Center of Mass)
Optional reading on Center of Mass: Nemirovskii Information Based Complexity 2.1-2.2

Bisection Search



Return $x^{(5)} = 0.296875$



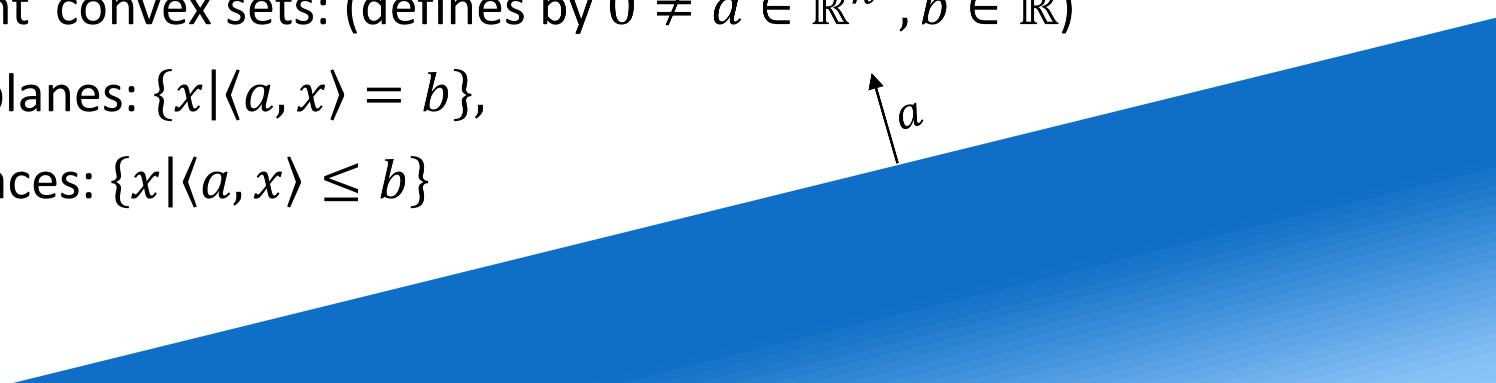
<u>Definition</u>: A set $C \subseteq \mathbb{R}^n$ is **convex** iff

$$\forall_{x,y\in C}\, \forall_{0<\alpha<1} \alpha x + (1-\alpha)y \in C$$

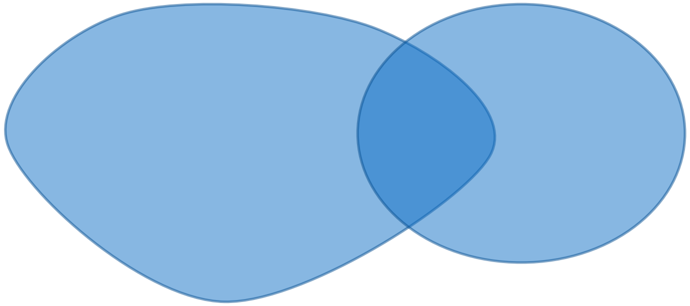
- Scaling, rotation, translation of convex set are convex
- C convex ➔ $AC + b = \{\, Ax + b \mid x \in C \,\}$ convex for any $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$
- Two important convex sets: (defines by $0 \neq a \in \mathbb{R}^{n^*}, b \in \mathbb{R}$)

Hyperplanes: $\{x | \langle a, x\rangle = b\}$,

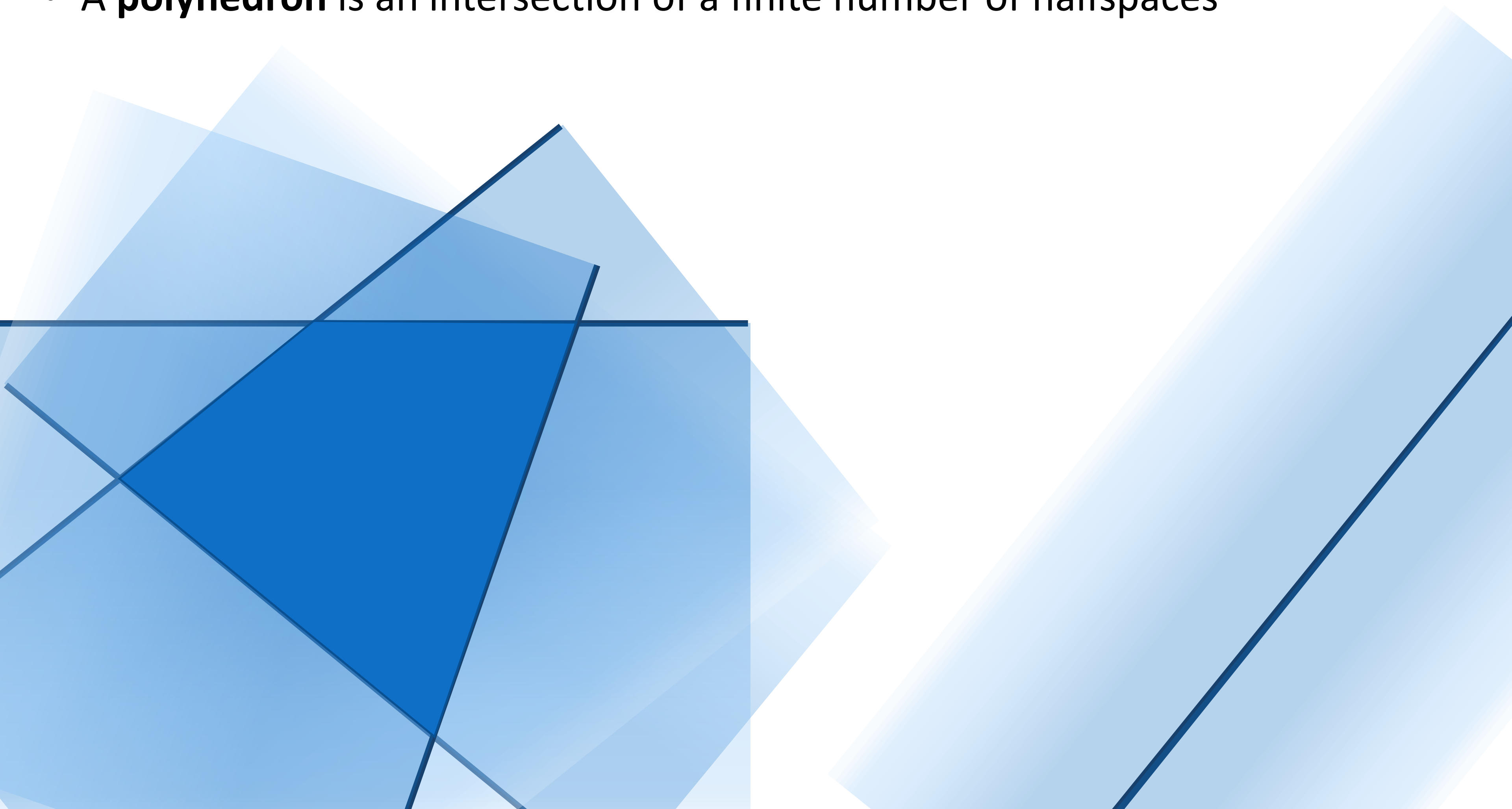
Halfspaces: $\{x | \langle a, x\rangle \leq b\}$



- If C_1, C_2 are convex ➔ $C_1 \cap C_2$ convex
  ($C_1 \cup C_2$ could be non-convex)



- A **polyhedron** is an intersection of a finite number of halfspaces



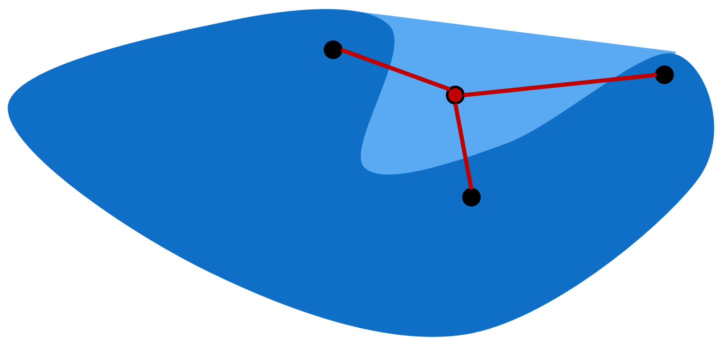
Definition: the **convex hull** of a set $S \subseteq \mathbb{R}^n$ is the intersection of all convex sets containing it

$$conv(S) = \cap \{C \mid S \subseteq C, C \text{ convex}\}$$

and is thus the smallest convex set containing S.

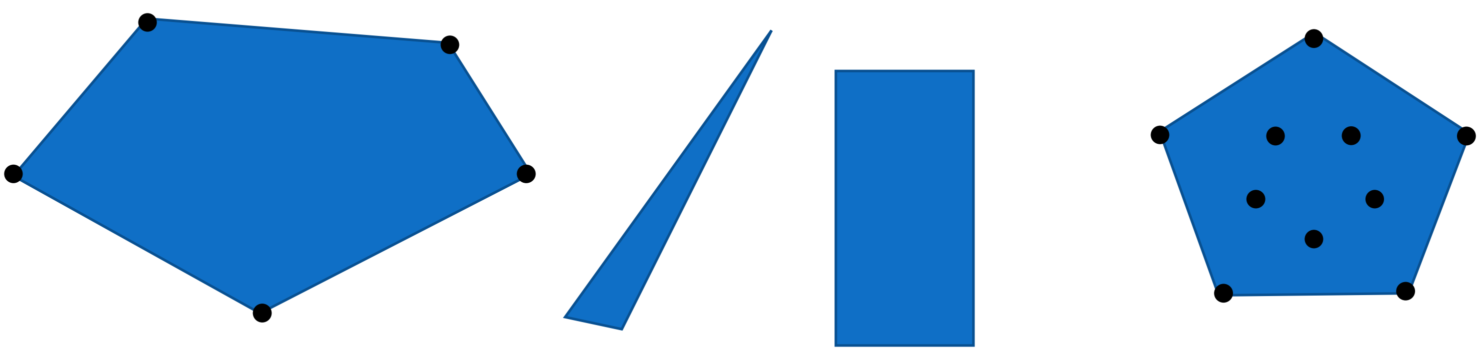
Equivalently: the set of all finite convex combinations of points from S:

$$conv(S) = \left\{\sum_{i=1}^{k} \alpha_i \, x_i \middle| x_i \in S, \sum_{i=1}^{k} \alpha_i = 1, \alpha_i > 0 \right\}$$

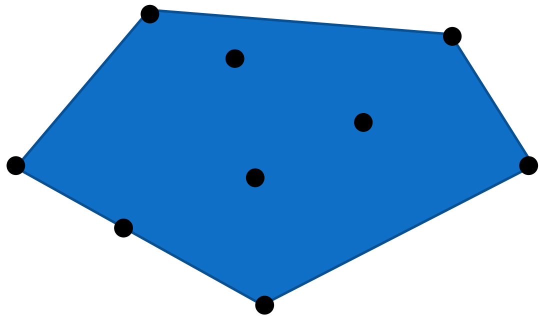


A **polytope** is a convex hull of a finite number of points

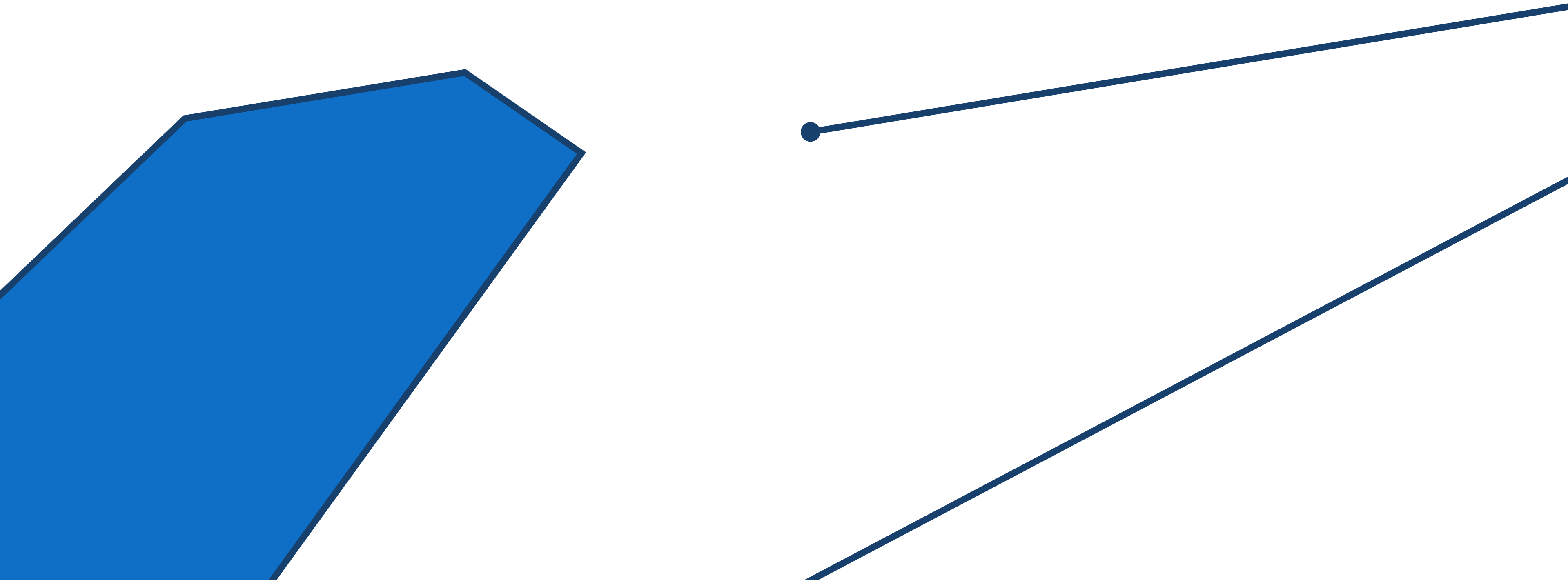
$$conv(x_1, \dots, x_m) = \{\sum_{i=1}^{m} \alpha_i \, x_i \,|\, \sum_{i=1}^{m} \alpha_i = 1, \alpha_i > 0 \}$$



- Is every polytope a polyhedron?

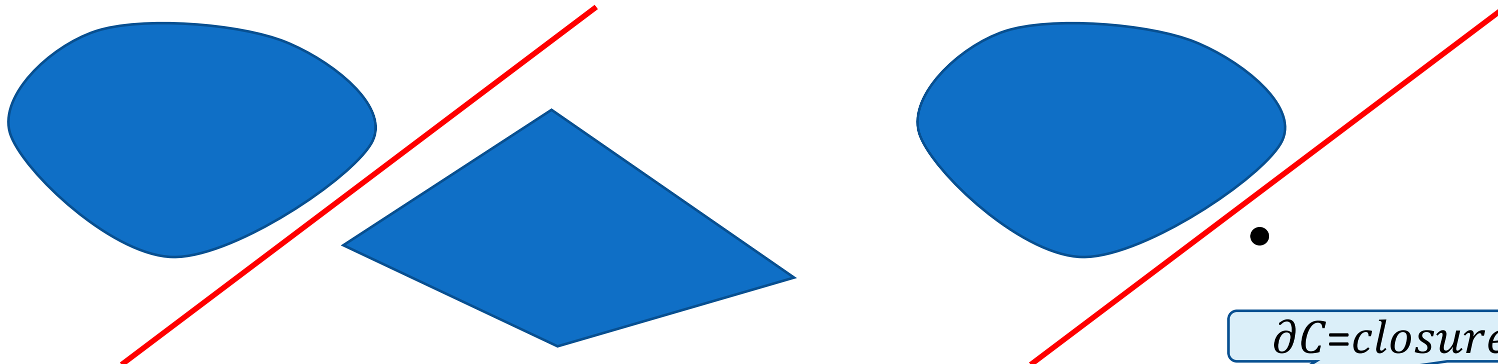


- Is every polyhedron a polytope?



Theorem: Suppose $C, D \subseteq \mathbb{R}^n$ are convex and disjoint (i.e. $C \cap D = \emptyset$), then there exists a **separating hyperplane** $\{x | \langle a, x \rangle = b\}$ s.t.

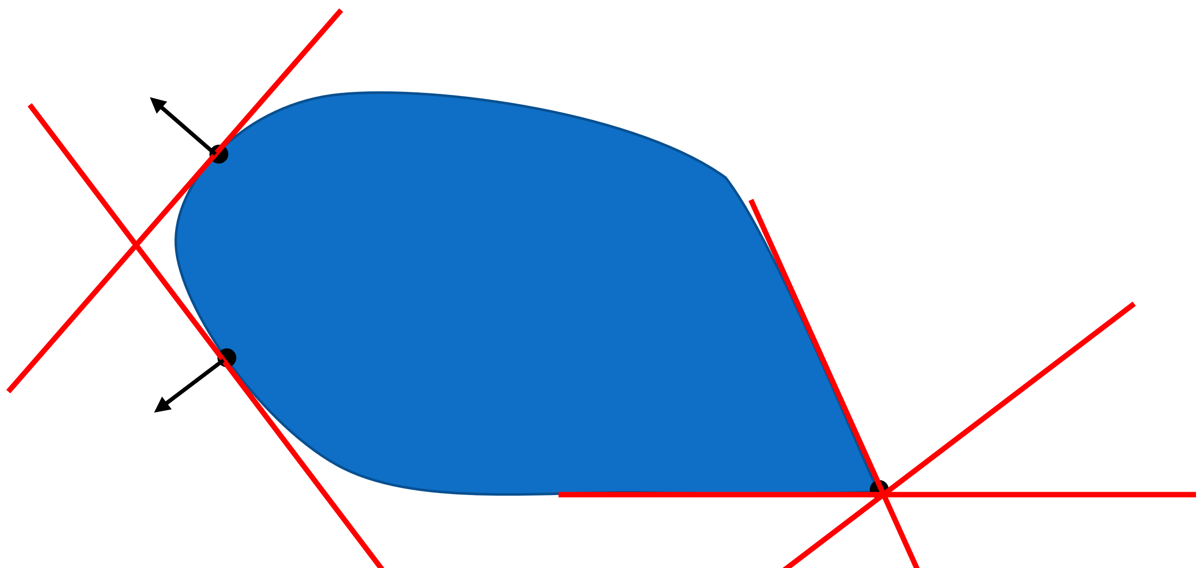
$$C \subseteq \{x | \langle a, x \rangle \leq b\} \qquad D \subseteq \{x | \langle a, x \rangle \geq b\}$$



∂C=closure(C)−int(C)

Theorem: For every convex set $C \subset \mathbb{R}^n$ and point $x_0 \in \partial C$ on its boundary, there exists a **supporting hyperplane** s.t.

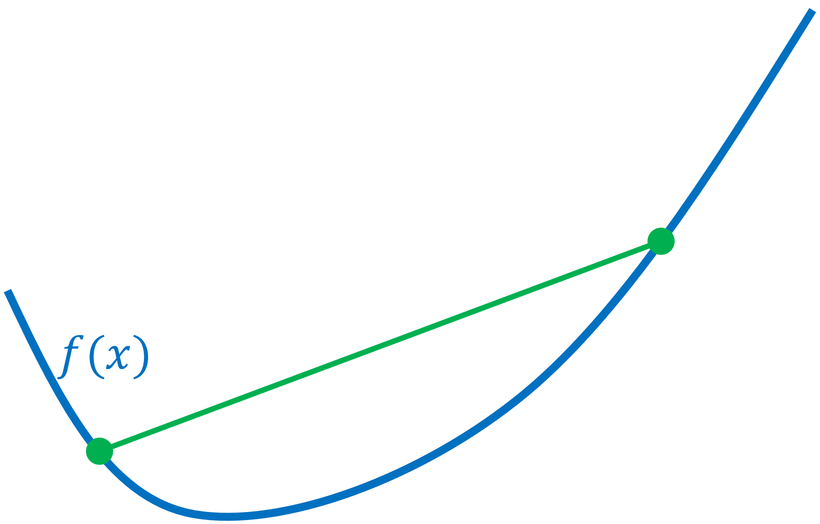
$$x_0 \in \{x | \langle a, x \rangle = b\} \qquad C \subseteq \{x | \langle a, x \rangle \leq b\}$$

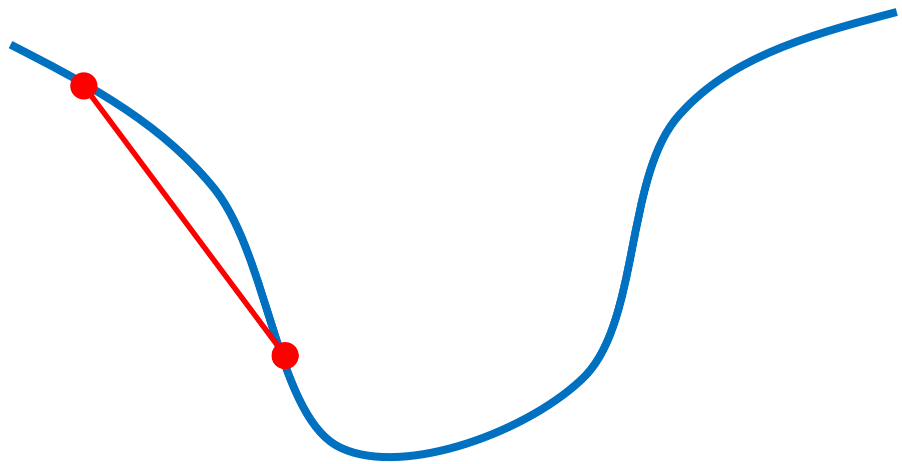


Definition: A function $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is **convex** iff

$$\forall_{x,y \in dom(f), 0<\alpha<1}\ f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

(implies that $\alpha x + (1-\alpha)y \in dom(f)$, i.e. $dom(f)$ is convex)





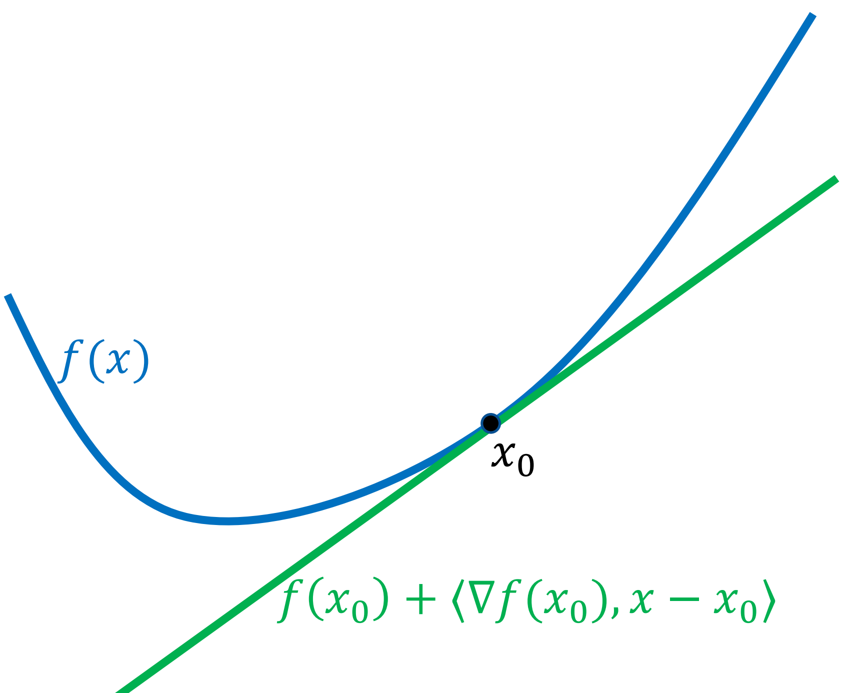
Definition: A function $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is **convex** iff

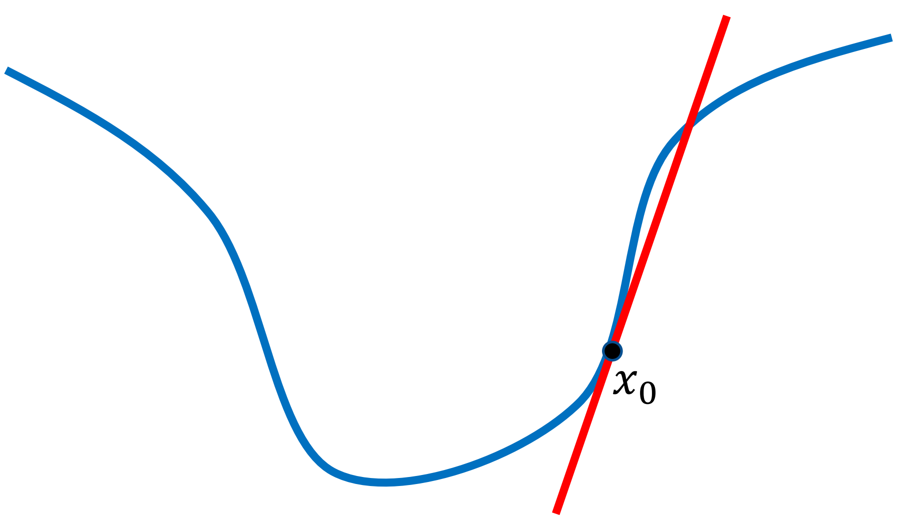
$$\forall_{x,y \in dom(f), 0<\alpha<1}\ f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

(implies that $\alpha x + (1-\alpha)y \in dom(f)$, i.e. $dom(f)$ is convex)

First order characterization: If f is differentiable, it is convex iff $dom(f)$ is convex and for all $x, x_0 \in dom(f)$:

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$$





Definition: A function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is **convex** iff

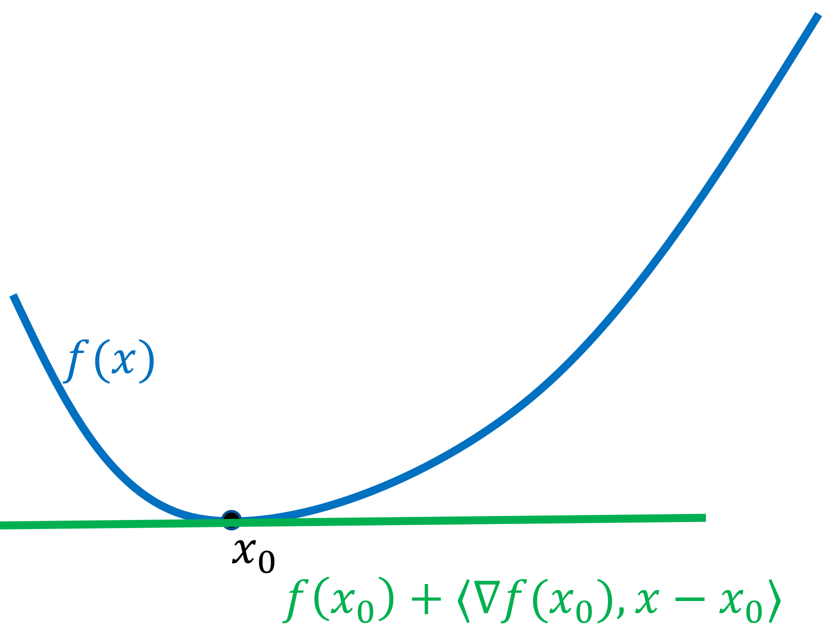
$$\forall_{x,y \in dom(f), 0<\alpha<1}\ f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

(implies that $\alpha x + (1-\alpha)y \in dom(f)$, i.e. $dom(f)$ is convex)

First order characterization: If f is differentiable, it is convex iff $dom(f)$ is convex and for all $x, x_0 \in dom(f)$:

$$f(x) \ge f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$$

Corollary: For diff. convex f, $\nabla f(x_0) = 0$ ➔ x_0 is an optimum

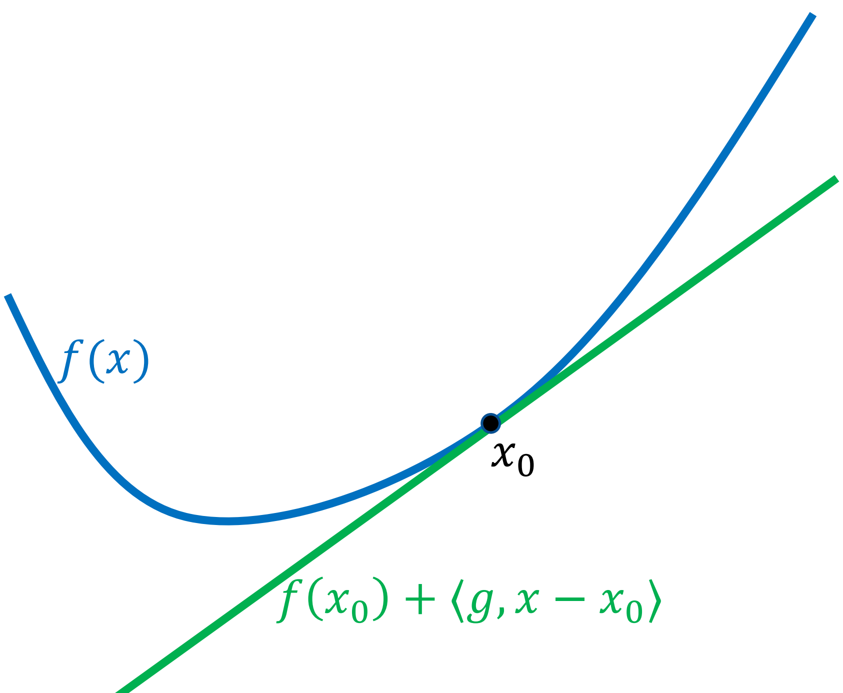


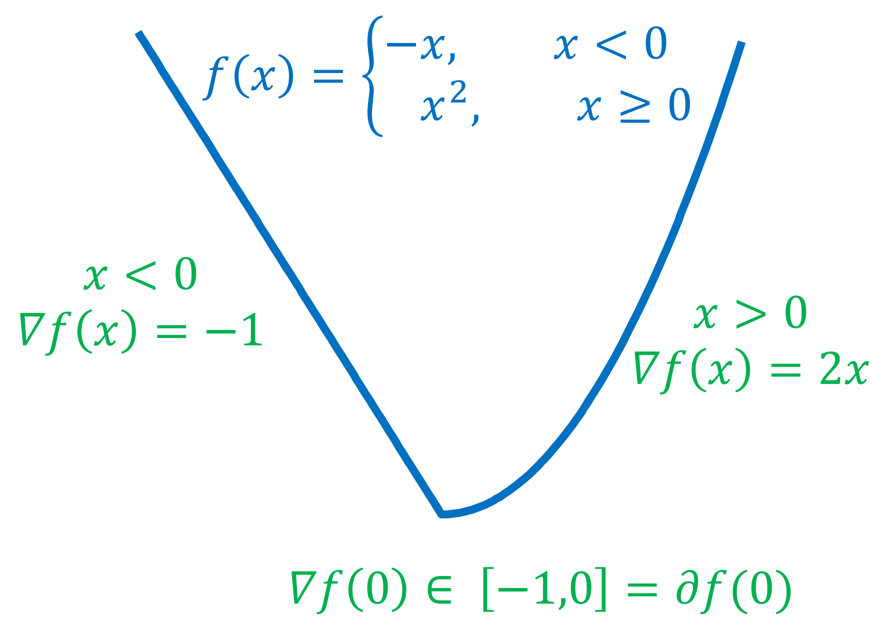
<u>Definition</u>: $g \in \mathbb{R}^{n^*}$ is a **subgradient** of f at $x_0 \in dom(f)$ iff

$$\forall_x \; f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

<u>Claim</u>: If $f(x)$ is convex and differentiable at $x_0 \in int(dom(f))$, then its unique subgradient at x_0 is its gradient $\nabla f(x_0)$

- At non-differentiable points, there might be multiple sub-gradients. We denote subgradients (even if not unique) $\nabla f(x)$ and the set of all subgradients $\partial f(x)$





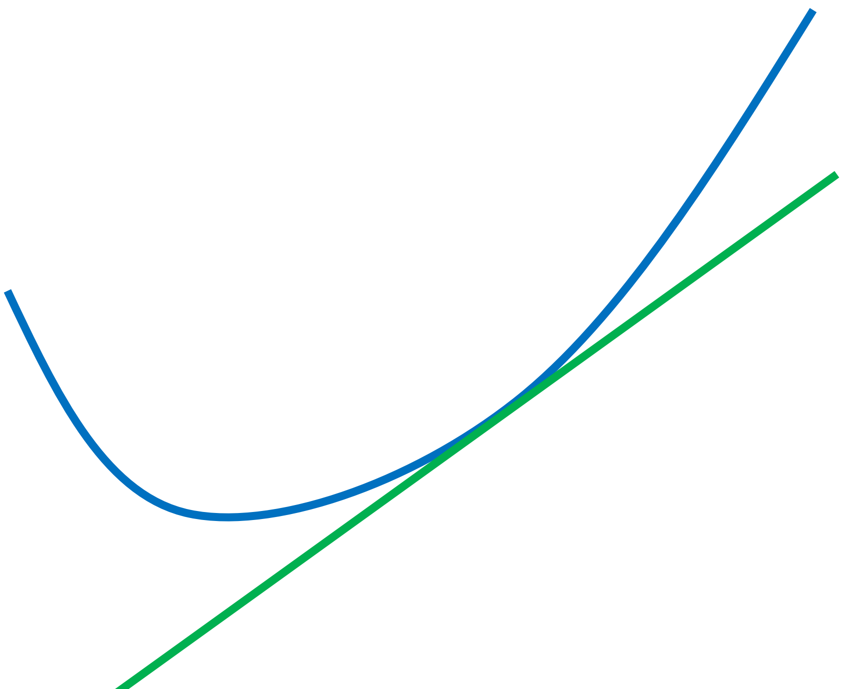
Definition: $g \in \mathbb{R}^{n^*}$ is a **subgradient** of f at $x_0 \in dom(f)$ iff

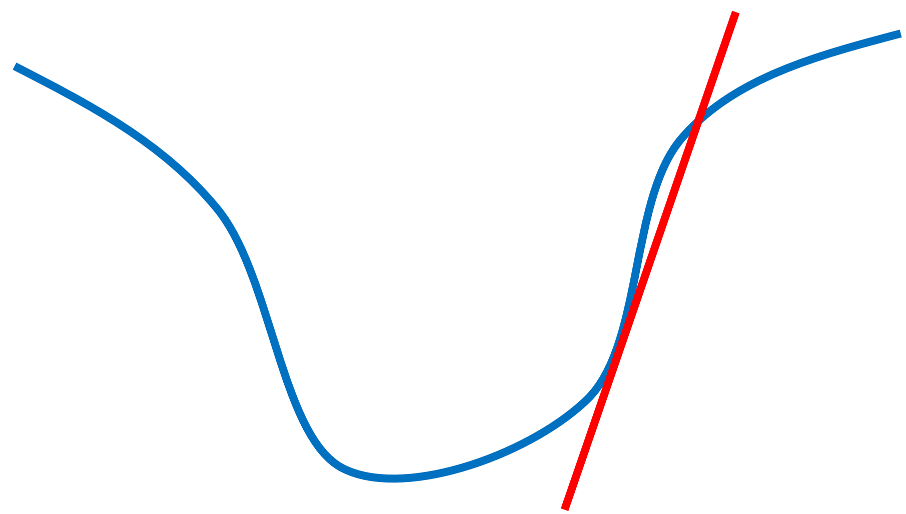
$$\forall_x\, f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

Claim: If $f(x)$ is convex and differentiable at $x_0 \in int(dom(f))$, then its unique subgradient at x_0 is its gradient $\nabla f(x_0)$

- At non-differentiable points, there might be multiple sub-gradients. We denote subgradients (even if not unique) $\nabla f(x)$ and the set of all subgradients $\partial f(x)$

Claim: A function is convex if and only if it has (at least one) subgradient at each point





Definition: $g \in \mathbb{R}^{n^*}$ is a **subgradient** of f at $x_0 \in dom(f)$ iff

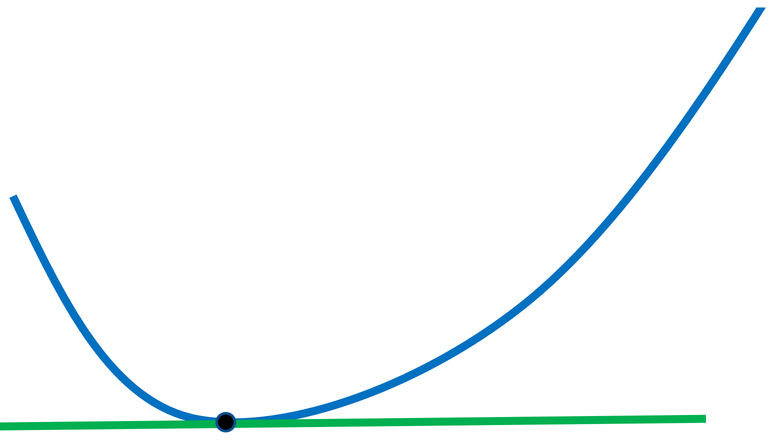
$$\forall_x\ f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

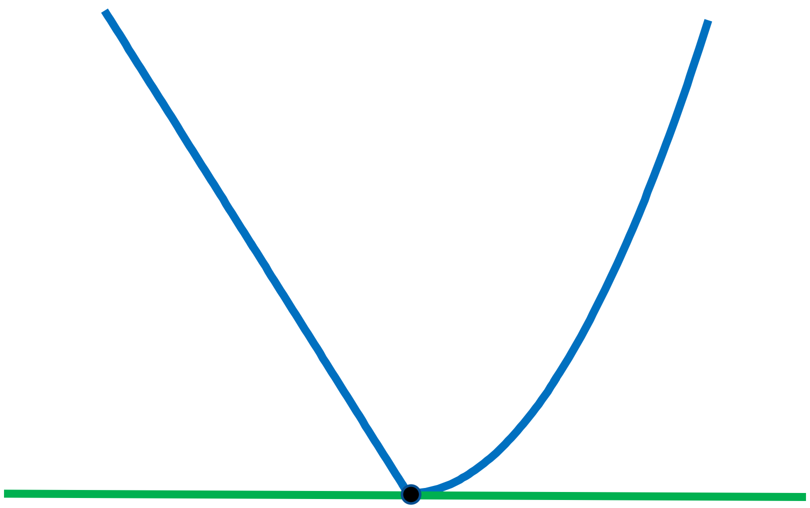
Claim: If $f(x)$ is convex and differentiable at $x_0 \in int(dom(f))$, then its unique subgradient at x_0 is its gradient $\nabla f(x_0)$

- At non-differentiable points, there might be multiple sub-gradients. We denote subgradients (even if not unique) $\nabla f(x)$ and the set of all subgradients $\partial f(x)$

Claim: A function is convex if and only if it has (at least one) subgradient at each point

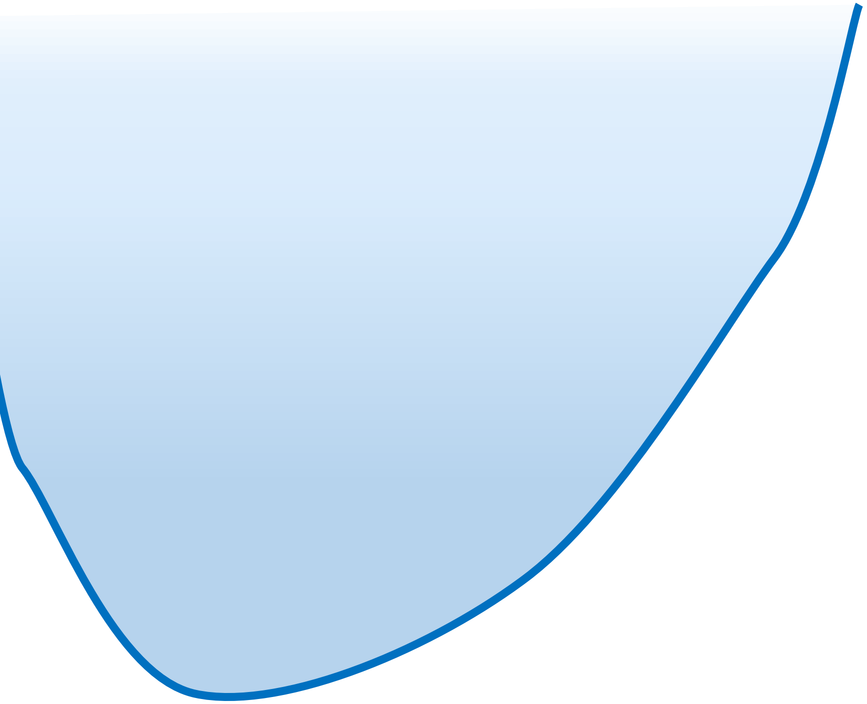
Corollary: For convex f, x_0 is an optimum $\Leftrightarrow$ $0 \in \partial f(x_0)$

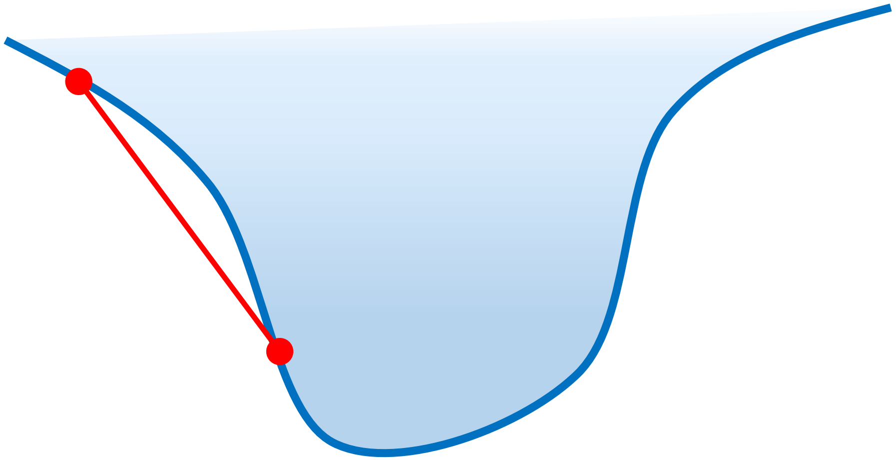




<u>Definition</u>: The epigraph of f is $epi(f) = \{(x,t) | f(x) \leq t\} \subseteq \mathbb{R}^{n+1}$

Claim: f is convex $\Leftrightarrow$ $epi(f)$ is a convex set

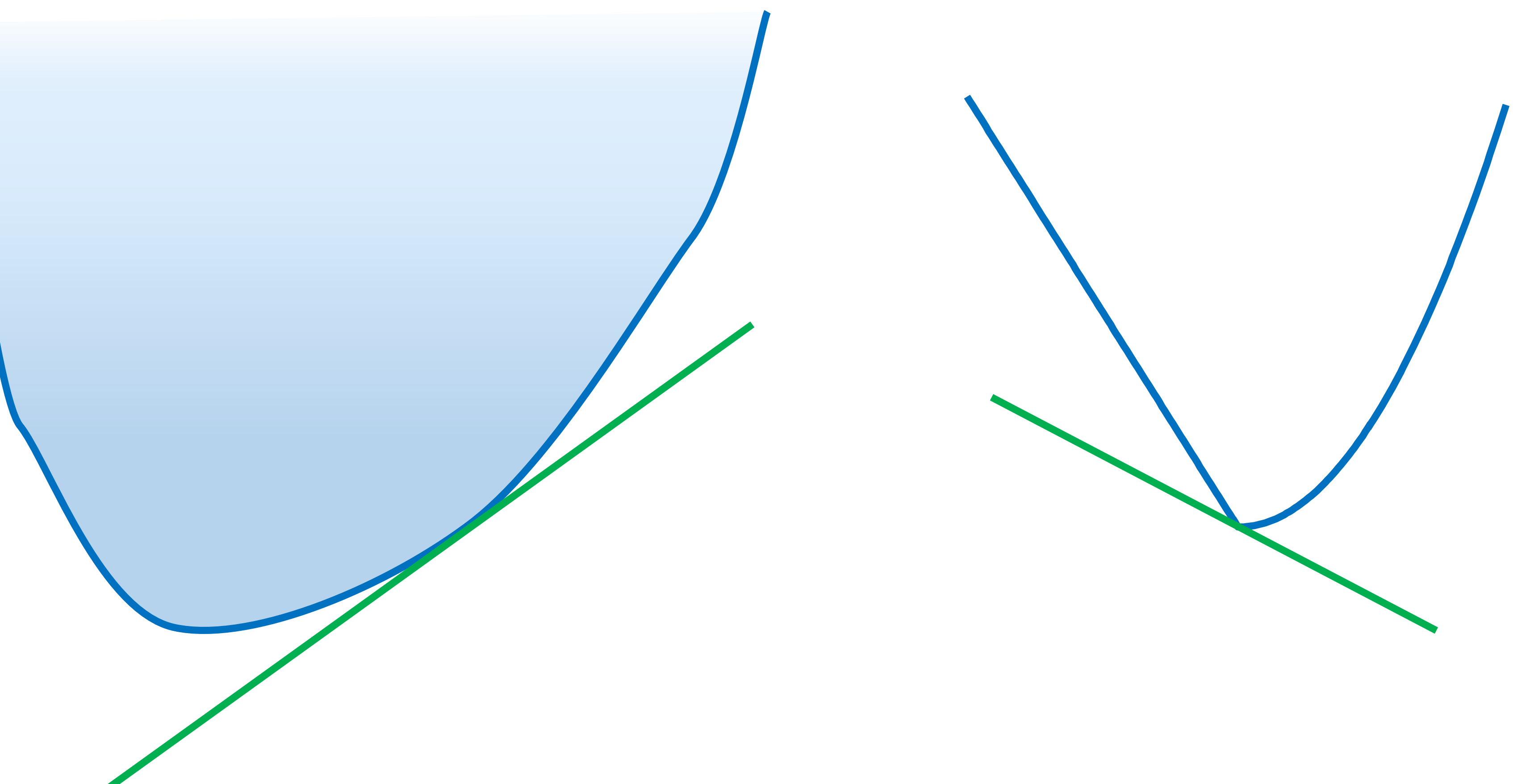




Definition: The epigraph of f is $epi(f) = \{(x,t) | f(x) \leq t\} \subseteq \mathbb{R}^{n+1}$

Claim: f is convex $\Leftrightarrow$ $epi(f)$ is a convex set

Subgradients of f define supporting hyperplanes of $epi(f)$



<u>Definition</u>: The **epigraph** of f is $epi(f) = \{(x,t) | f(x) \le t\} \subseteq \mathbb{R}^{n+1}$

Claim: f is convex ⇔ $epi(f)$ is a convex set

Subgradients of f define supporting hyperplanes of $epi(f)$

<u>Definition</u>: For $\alpha \in \mathbb{R}$ the $\boldsymbol{\alpha}$**-sublevel** set of f is

$$S_\alpha = \{x | f(x) \le \alpha\} \subseteq \mathbb{R}^n$$

<u>Claim</u>: f is convex ➔ S_α are convex

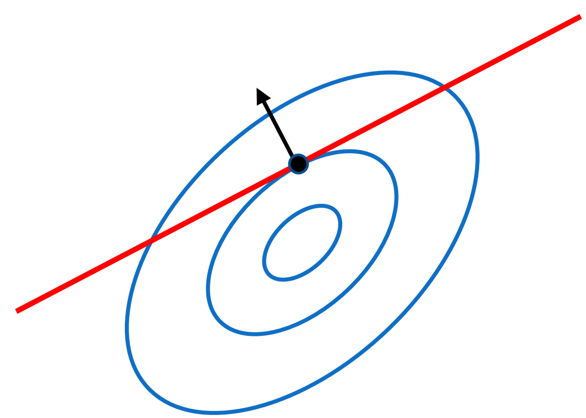
Subgradients of f define supporting hyperplanes of S_α:

Denote $\alpha = f(x_0)$. For $x \in S_\alpha$ we have:

$$f(x_0) = \alpha \ge f(x) \ge f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$$

➔ $\langle \nabla f(x_0), x \rangle \le \langle \nabla f(x_0), x_0 \rangle$

➔ $S_\alpha \subseteq \{x | \langle \nabla f(x_0), x \rangle \le \underbrace{b}_{b = \langle \nabla f(x_0), x_0 \rangle}\}$

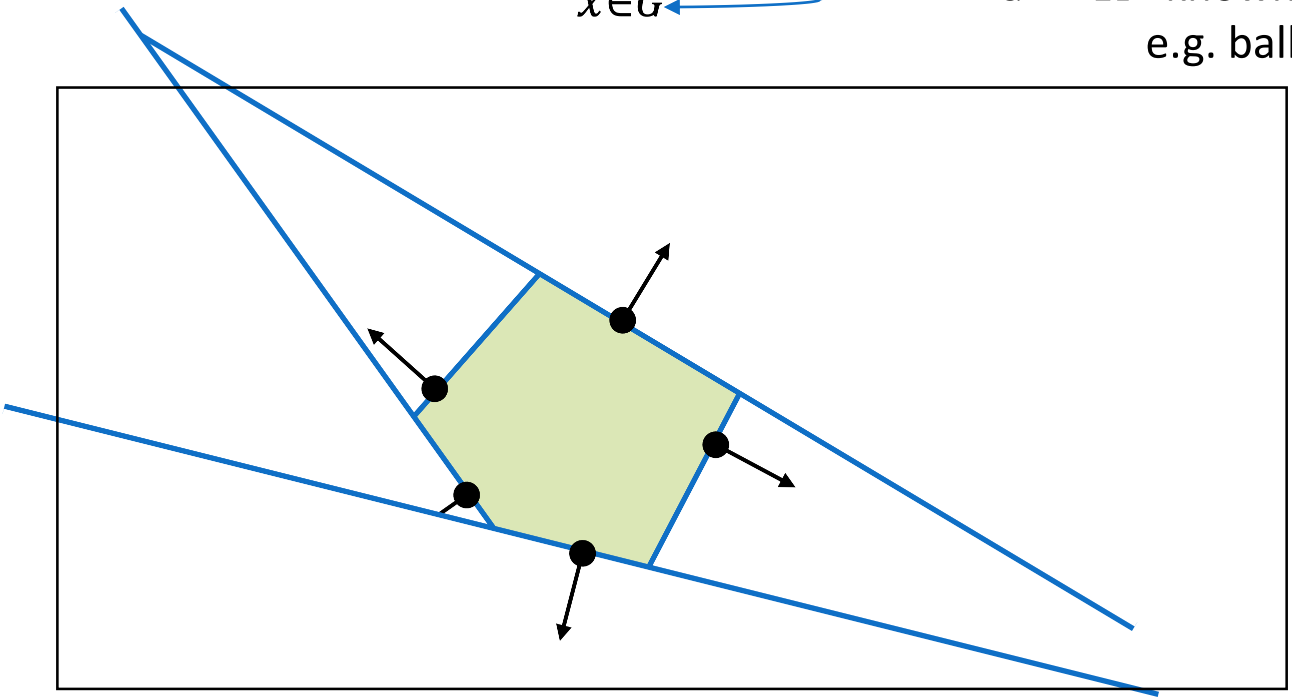


If $\nabla f(x_0) \neq 0$, this is a supporting hyperplane of S_α

Cutting Plane Method

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

$G \subset \mathbb{R}^n$ known and bounded,
e.g. ball or box



Init: $G^{(0)} = G$

Iter: Pick $x^{(k)} \in G^{(k)}$

$a^{(k)} = \nabla f\left(x^{(k)}\right), b^{(k)} = \langle a^{(k)}, x^{(k)} \rangle$

$G^{(k+1)} \leftarrow G^{(k)} \cap \{ x \mid \langle a^{(k)}, x \rangle < b^{(k)} \}$

Which point should we pick?

- We want to ensure that, no matter what direction the (sub)gradient is, we will shrink $G^{(k)}$ significantly

Theorem (Grünbaum 1960): Let G be a bounded convex set with center of mass

$$c = \frac{\int_G x dx}{\int_G dx}$$

then for any hyperplane $\{x | \langle a, x \rangle = \langle a, c \rangle\}$ passing through c we have:

$$vol(G \cap \{x | \langle a, x \rangle \leq \langle a, c \rangle\}) \leq \left(1 - {}^{1}\!/_{e}\right) vol(G)$$

Conclusion: if at each iteration we pick the center of mass of $G^{(k)}$:

$$vol\left(G^{(k)}\right) \leq 0.633^k vol(G)$$

Furthermore, it is possible to show that if $|f(x)| \leq B$ then

$$\min_{i=1..k} f\left(x^{(i)}\right) \leq f(x^*) + 2B \cdot 0.633^{k/n}$$

Intuition: need to shrink to an ϵ-neighborhood around x^*, i.e. until $vol\left(G^{(k)}\right) \propto \epsilon^n$

Claim: if we select $x^{(k)}$=center of mass of $G^{(k)}$, then $\min_{i=1..k} f\left(x^{(i)}\right) \le f(x^*) + 2B \cdot 0.633^k$

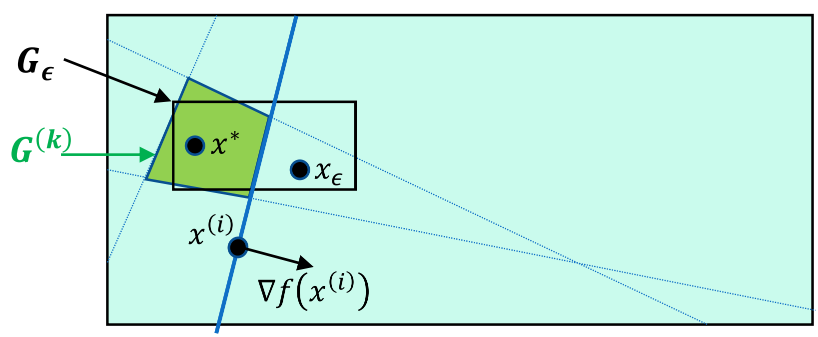
Proof:

- For simplicity we assume G is closed (otherwise, use its closure). Since G, f are bounded, an optimum is attained. Denote some such optimum x^*
- Because we only remove points with $f(x) \ge f\left(x^{(i)}\right)$, unless some $x^{(i)}$ is optimal, we have $x^* \in G^{(k)}$ (if it is optimal, then $f\left(x^{(i)}\right) = f(x^*)$ and we are done)
- Define the ϵ-shrinking of G towards x^*:
$$G_\epsilon = \{(1-\epsilon)x^* + \epsilon x | x \in G\}$$
- n-dim volume shrinks as ϵ^n and so $vol(G_\epsilon) = \epsilon^n vol(G)$.
$$vol\left(G^{(k)}\right) < 0.633^k \cdot vol(G) = 0.633^k \epsilon^{-n} \cdot vol(G_\epsilon)$$
- Set $\epsilon = 0.633^{k/n}$ so that $vol\left(G^{(k)}\right) < vol(G_\epsilon)$, that is after k iterations we are left with a set smaller then G_ϵ. This means there exists a point $x_\epsilon \in G_\epsilon$ but $x_\epsilon \notin G^{(k)}$. Let j be the iteration in which it was lost, i.e. s.t. $x_\epsilon \in G^{(j)}$ but $x_\epsilon \notin G^{(j+1)}$.

x_ϵ was removed at iter j — convexity — $f(x) - f(x^*) \le 2B$

$$f\left(x^{(j)}\right) < f(x_\epsilon) = f\left((1-\epsilon)x^* + \epsilon x\right) \le (1-\epsilon)f(x^*) + \epsilon f(x) \le f(x^*) + 2\epsilon B$$

def of G_ϵ, for some $x \in G$



Details of proof not required material

Center of Mass Method

(Levin  and Newman  1965)

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

Requirements: $G \subset \mathbb{R}^n$ is a known bounded convex set
evaluation and (sub)gradient access $x \mapsto f(x), \nabla f(x)$

Init: $G^{(0)} = G$

Iter: $x^{(k)} \leftarrow$ center of mass of $G^{(k)}$

obtain subgradient $\nabla f\left(x^{(k)}\right)$

$a^{(k)} = \nabla f\left(x^{(k)}\right), b^{(k)} = \langle a^{(k)}, x^{(k)} \rangle$

$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{ x \,\middle|\, \langle a^{(k)}, x \rangle < b^{(k)} \right\}$

Return: $\tilde{x}^{(k)} = \arg\min_{i=0..k} f\left(x^{(i)}\right)$

Assumption: f is convex and bounded on G: $\forall_{x \in G} |f(x)| \leq B$

Guarantee: after $k = O\left(n \log \frac{1}{\epsilon}\right)$ iterations, $f\left(\tilde{x}^{(k)}\right) \leq f(x^*) + \epsilon$