COMP 414/514: Optimization – Algorithms, Complexity and Approximations

Lecture 3

Overview

- In the last lecture, we:
  - Introduced some notions on smooth optimization
  - Introduced gradient descent and what we can say about its convergence rate

- In this lecture, we will:
  - Discuss briefly **smooth continuous optimization**
  - Introduce the important class of **convex optimization**
  - Discuss about **convergence rates** and some **lower bounds** on such rates

“What does convexity bring onto the table?”

Convex functions

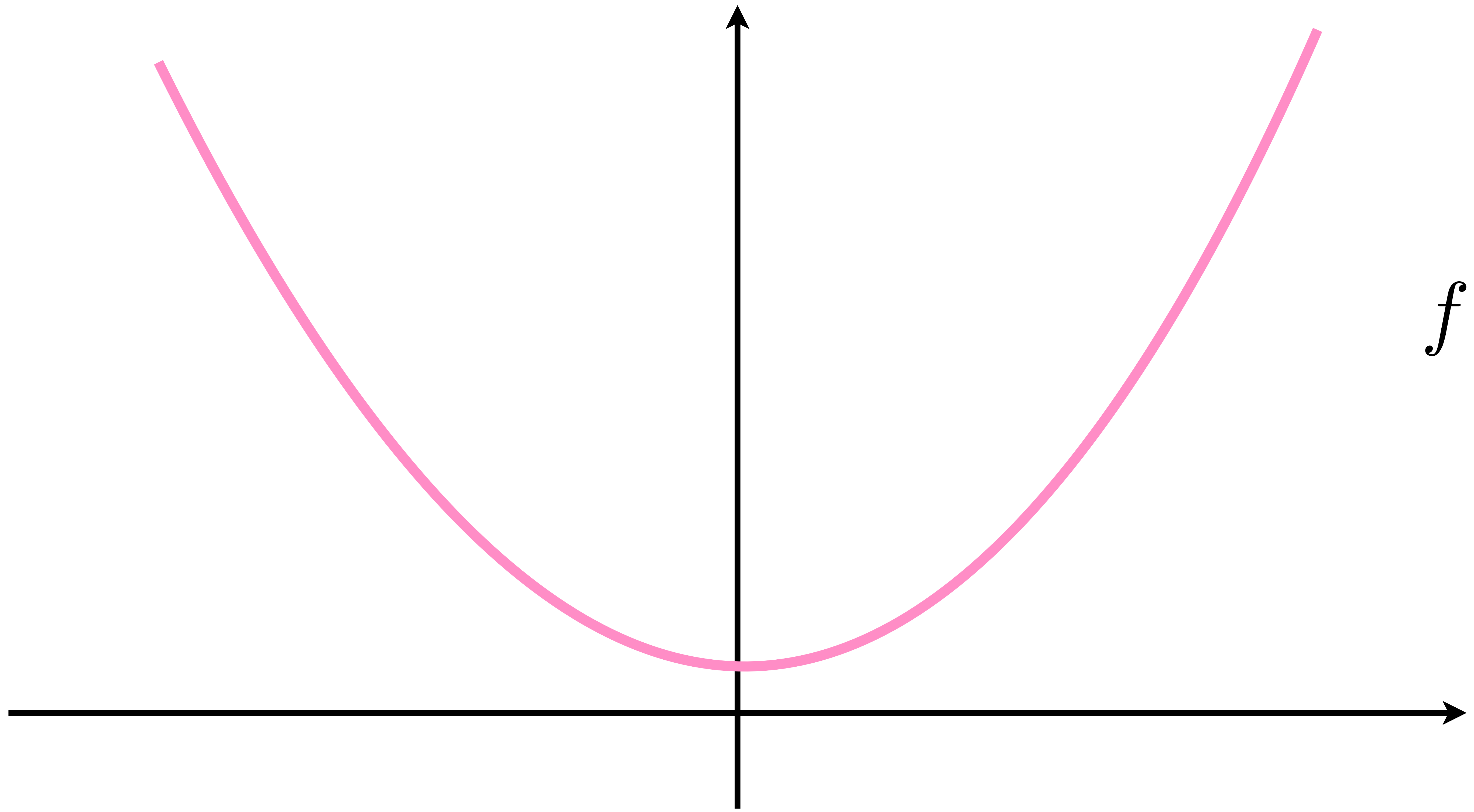
- General definition:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \ \forall a \in [0,1]$$

Convex functions

- General definition:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \ \forall a \in [0,1]$$

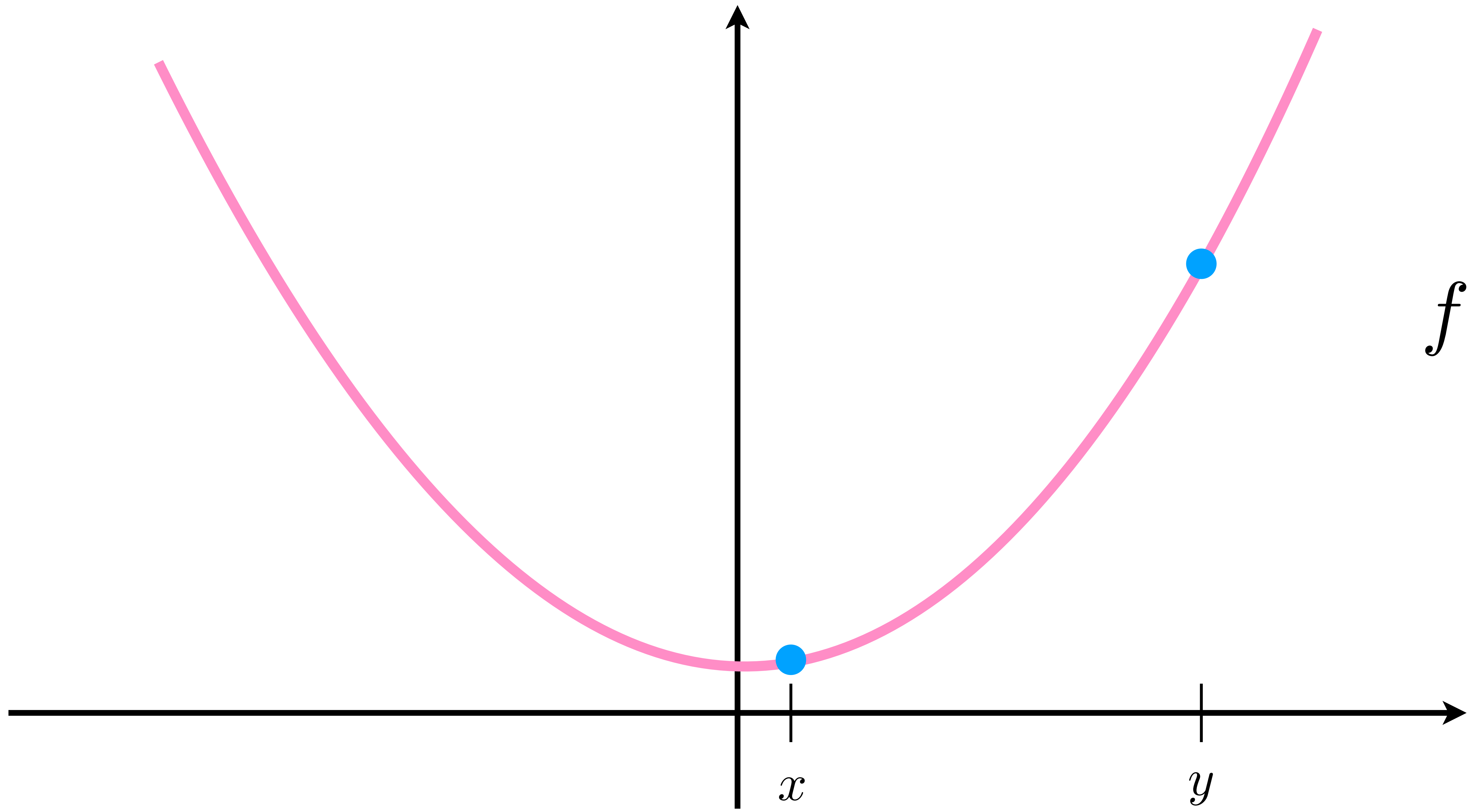


$$f(x) = x^2 + \beta$$

Convex functions

- General definition:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \ \forall a \in [0,1]$$

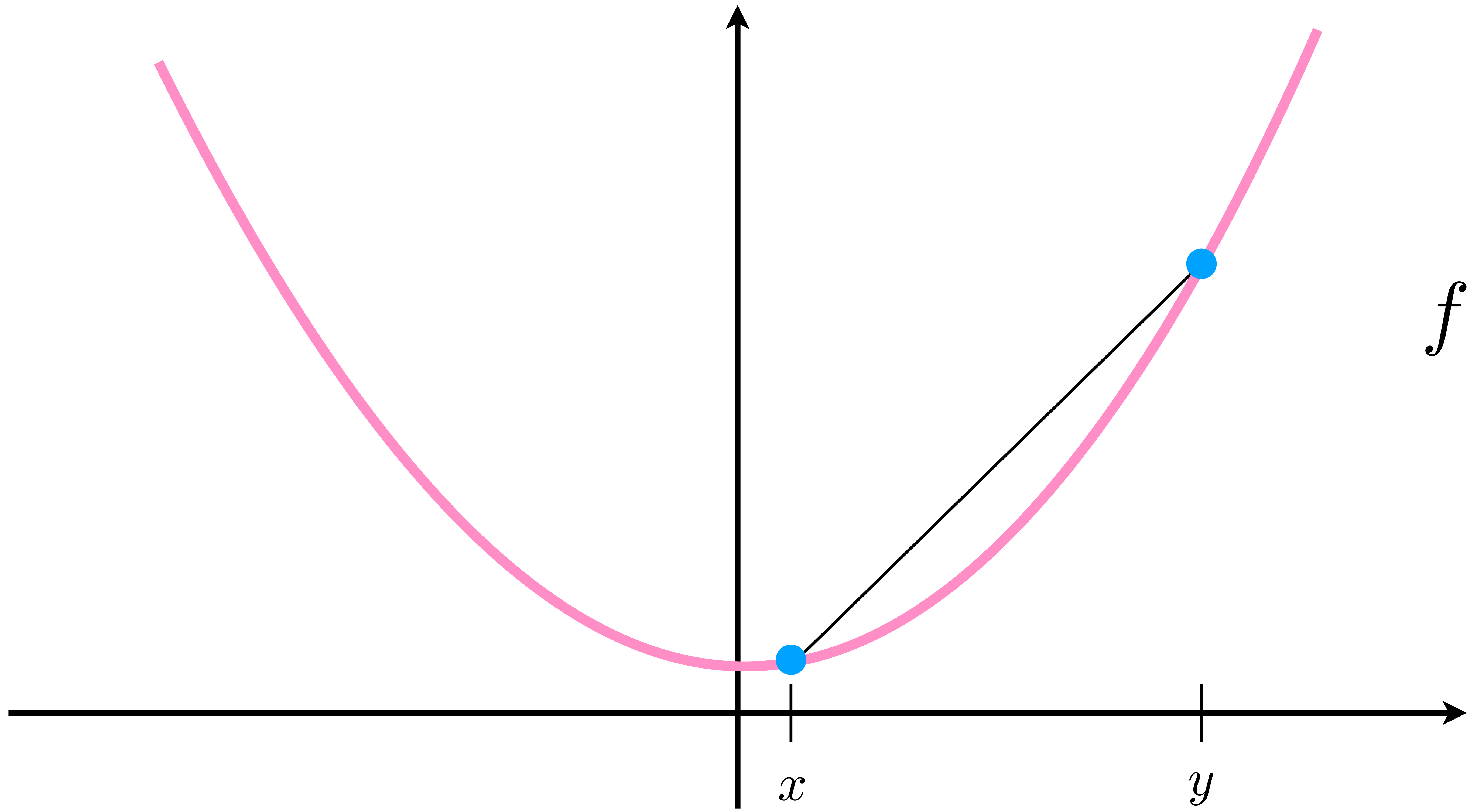


$$f(x) = x^2 + \beta$$

Convex functions

– General definition:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \;\; \forall a \in [0,1]$$

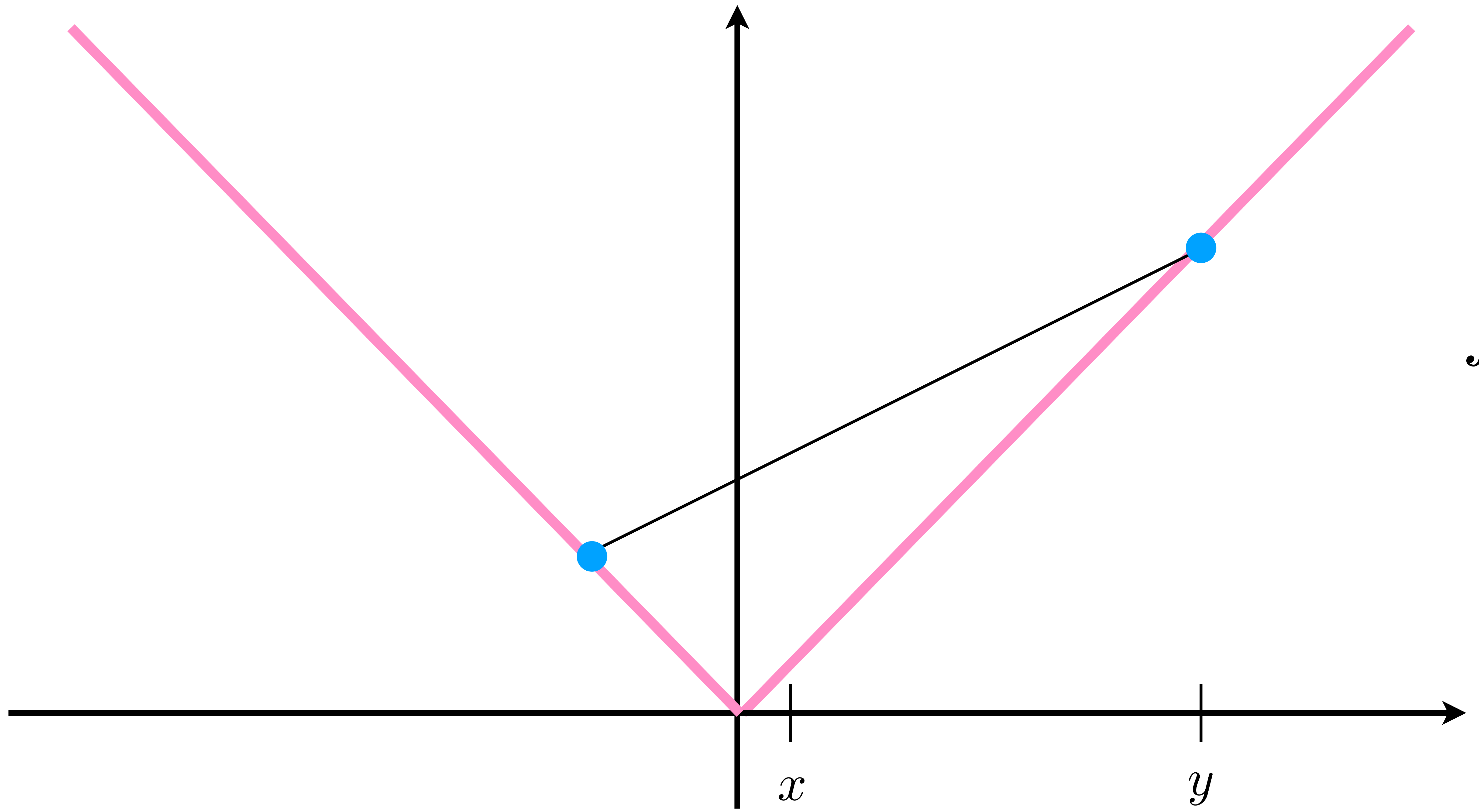


$$f(x) = x^2 + \beta$$

Convex functions

– General definition:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \;\; \forall a \in [0,1]$$



$$f(x) = |x|$$

Convex functions

- Examples:

Function	Example	Attributes
ℓ_p vector norms, $p \geq 1$	$\|\mathbf{x}\|_2, \|\mathbf{x}\|_1, \|\mathbf{x}\|_\infty$	convex
ℓ_p matrix norms, $p \geq 1$	$\|\mathbf{X}\|_* = \sum_{i=1}^{\text{rank}(\mathbf{X})} \sigma_i$	convex
Square root function	$\sqrt{x}$	concave, nondecreasing
Maximum of functions	$\max\{x_1, \ldots, x_n\}$	convex, nondecreasing
Minimum of functions	$\min\{x_1, \ldots, x_n\}$	concave, nondecreasing
Sum of convex functions	$\sum_{i=1}^{n} f_i$, f_i convex	convex
Logarithmic functions	$\log(\det(\mathbf{X}))$	concave, assumes $\mathbf{X} \succ 0$
Affine/linear functions	$\sum_{i=1}^{n} X_{ii}$	both convex and concave
Eigenvalue functions	$\lambda_{max}(\mathbf{X})$	convex, assumes $\mathbf{X} = \mathbf{X}^T$

Convex functions

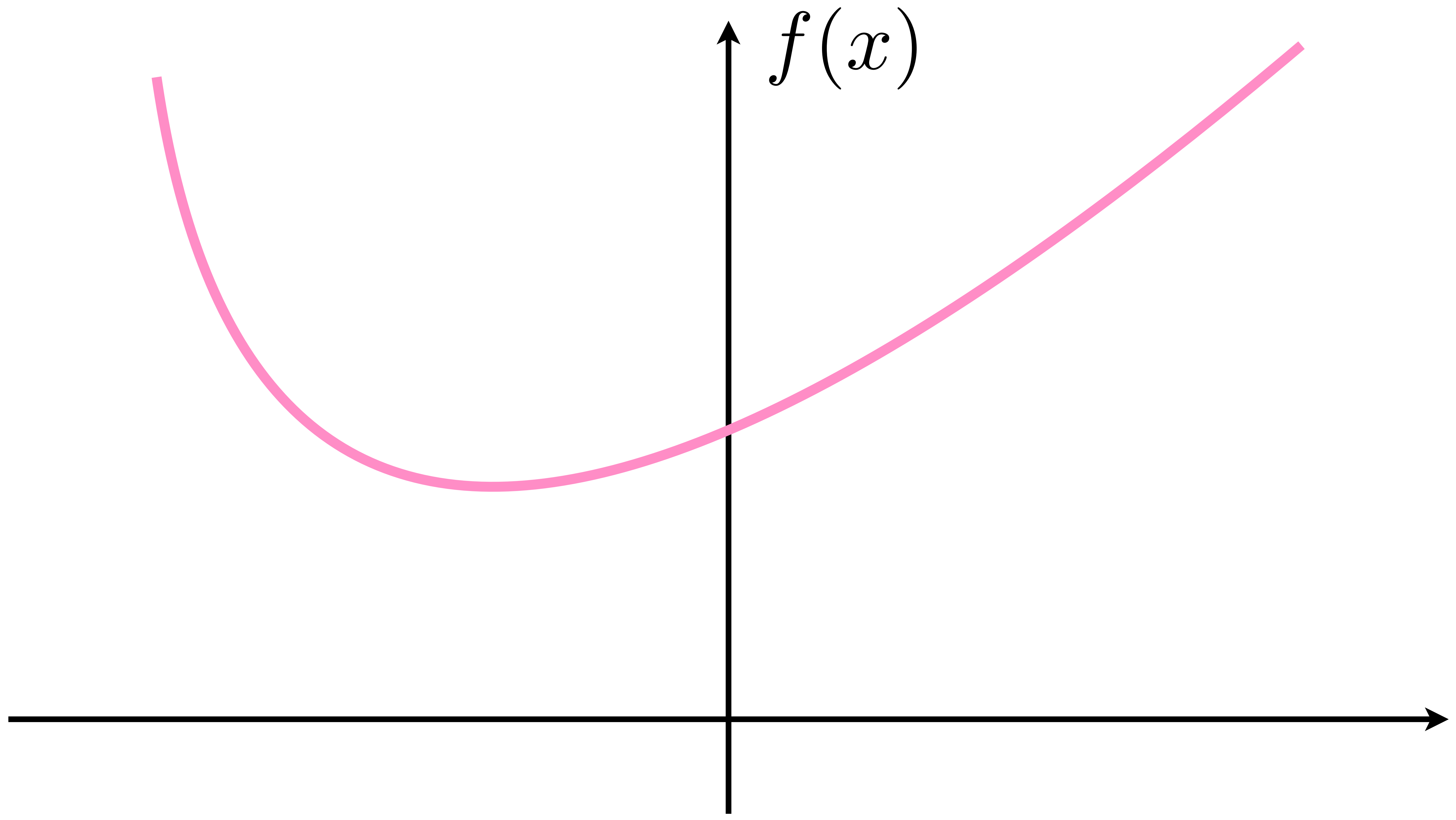
– Alternative (more practical) definitions of convexity

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \, , \quad \forall x, y$$

Convex functions

- Alternative (more practical) definitions of convexity

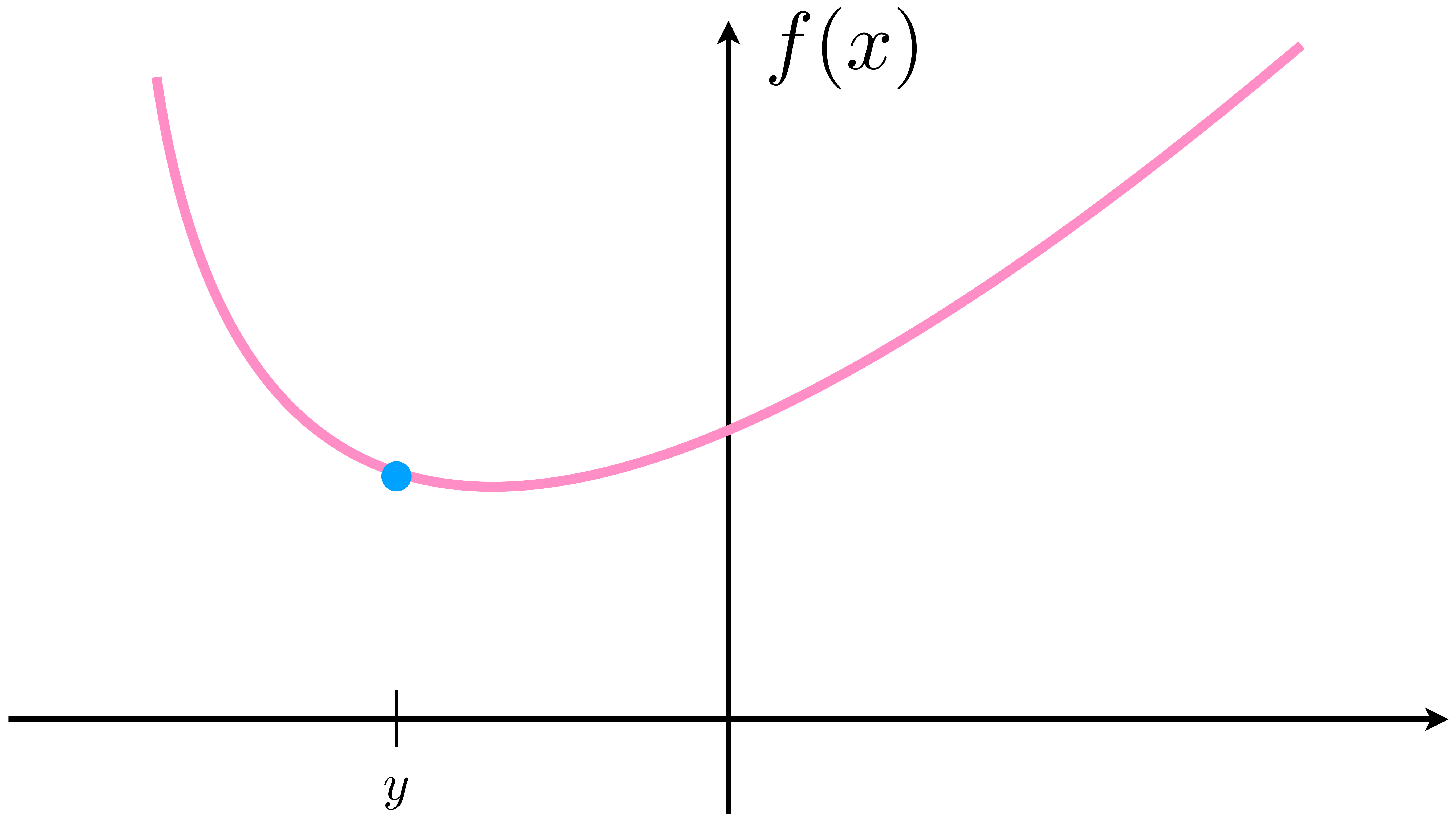
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \,, \quad \forall x, y$$



Convex functions

– Alternative (more practical) definitions of convexity

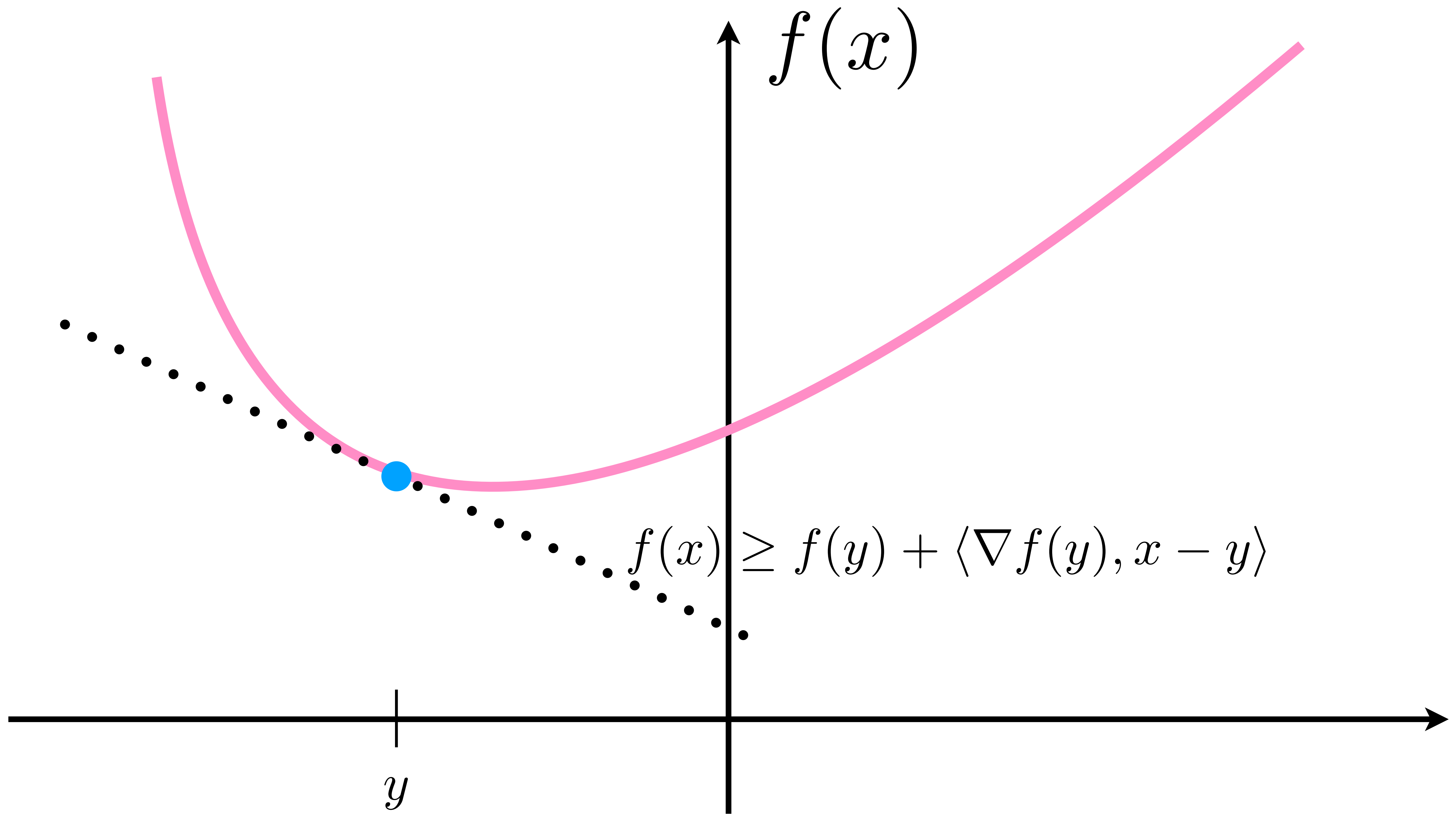
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \quad \forall x, y$$



Convex functions

– Alternative (more practical) definitions of convexity

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \quad \forall x, y$$



Convex functions

– Alternative (more practical) definitions of convexity

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \,, \quad \forall x, y$$

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0, \;\; \forall x, y$$

$$\nabla^2 f(x) \succeq 0, \;\; \forall x$$

(Assuming the function is twice differentiable)

Convex functions

– Alternative (more practical) definitions of convexity

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \quad \forall x, y$$

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0, \ \forall x, y$$

$$\nabla^2 f(x) \succeq 0, \ \forall x$$

(Assuming the function is twice differentiable)

Any interpretations?

Convex functions

– Key consequences of convexity

“Any stationary point is a global minimum”

Proof:

Convex functions

– Key consequences of convexity

"Any stationary point is a global minimum"

Proof: Assume a stationary point $x^\star$. This implies $\nabla f(x^\star) = 0$

Convex functions

– Key consequences of convexity

“Any stationary point is a global minimum”

Proof: Assume a stationary point $x^\star$. This implies $\nabla f(x^\star) = 0$

By convexity:

$$f(x) \geq f(x^\star) + \langle \nabla f(x^\star), x - x^\star \rangle = f(x^\star), \ \forall x$$

– This is what makes convex optimization preferable.

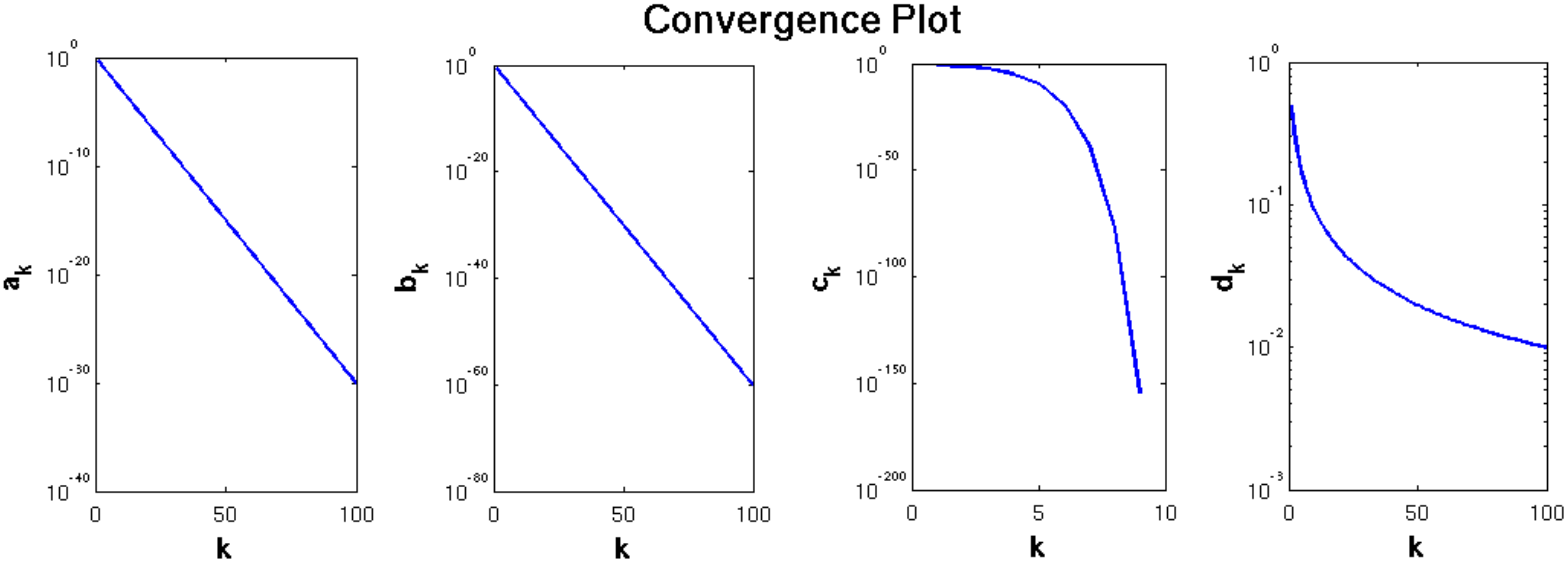
(Any local solution is actually global –
this does not mean that convex optimization is necessarily tractable!)

Does convexity improve guarantees?

Whiteboard

Convergence rates 101

(Source: Wikipedia)



$O(\log 1/\varepsilon)$

$q^k,\ q \in (0,1)$

$O\left(\log\log(1/\varepsilon)\right)$

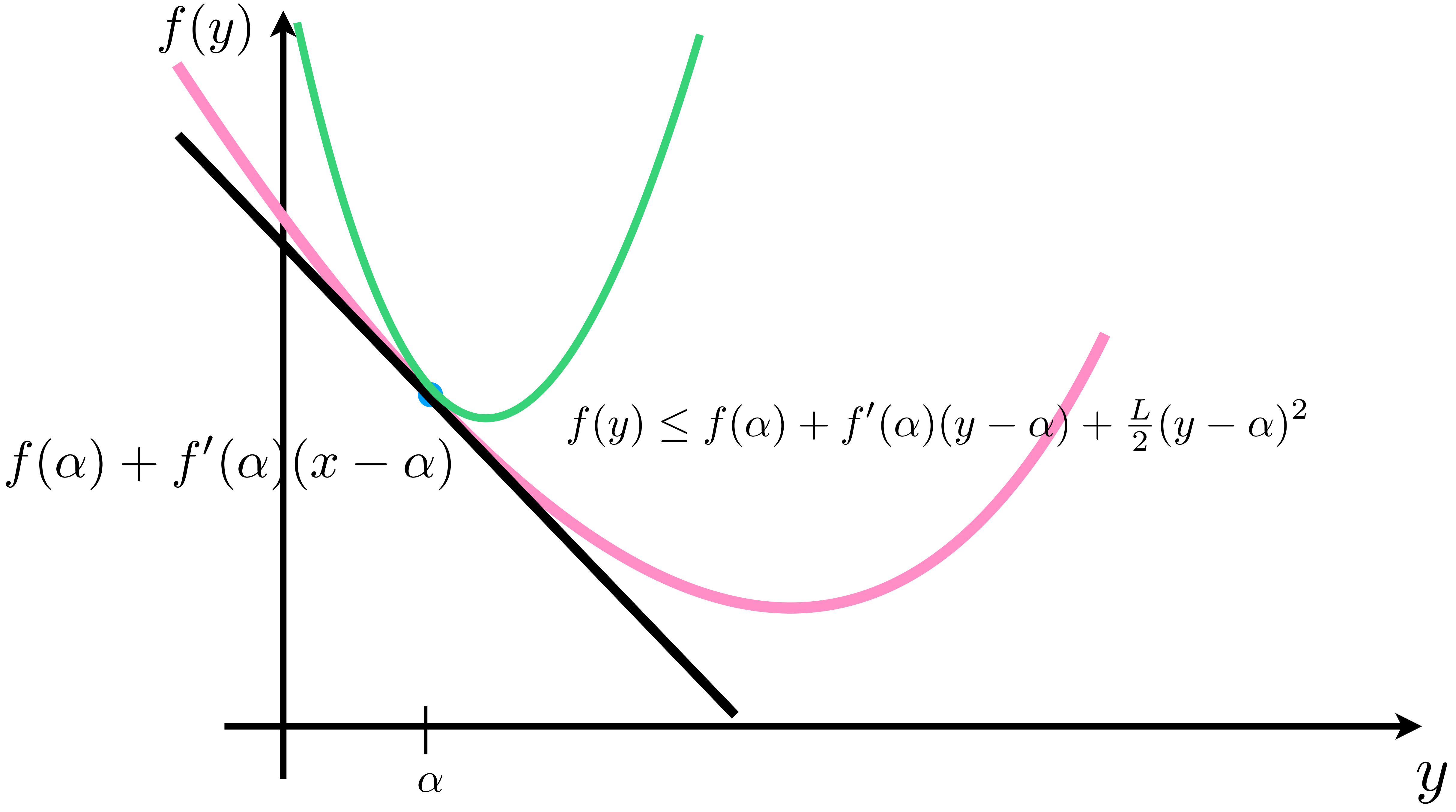
$O(1/\varepsilon^2),\ O(1/\varepsilon),\ O(1/\sqrt{\varepsilon})$

$O(1/k^2), O(1/k), O(\sqrt(k))$

Can we achieve a better performance?

- Strong convexity: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|x - y\|_2^2, \ \forall x, y$
  - Strong convexity parameter: $\mu > 0$

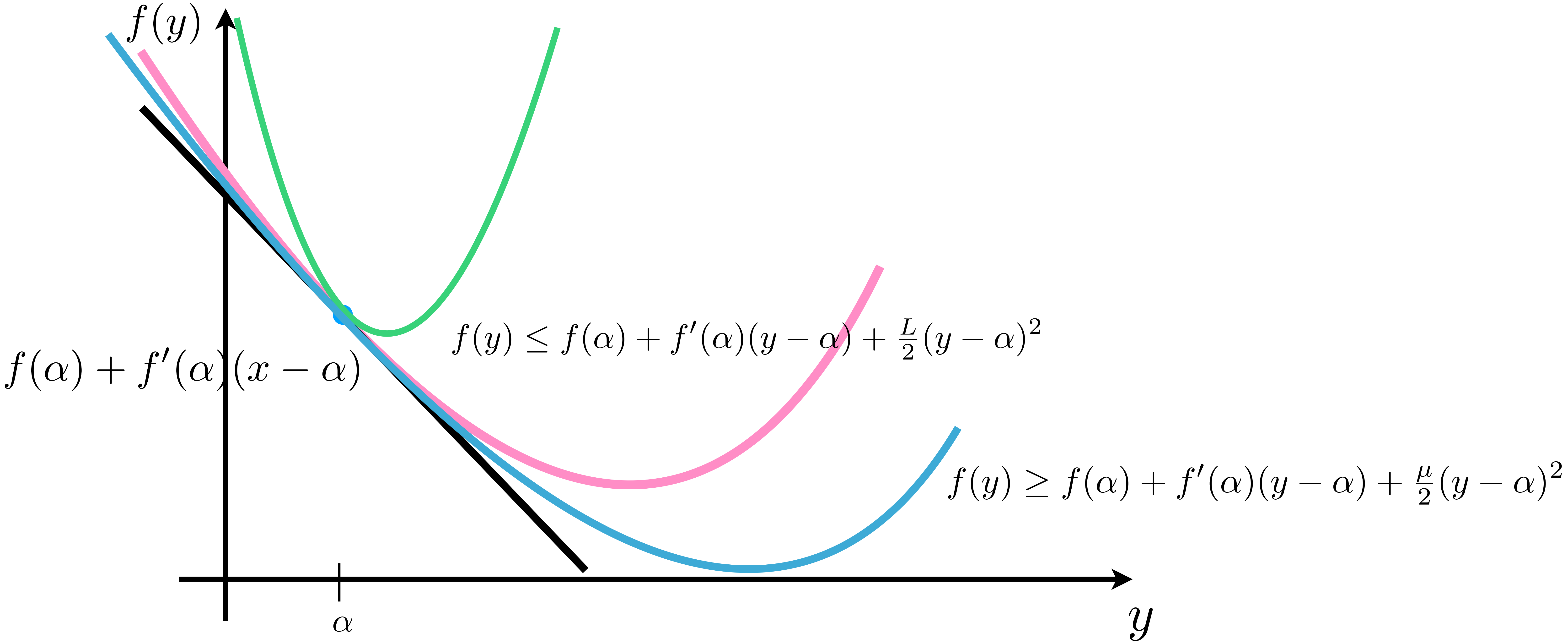
"Strong convexity implies that f should be steep enough to make progress"



Can we achieve a better performance?

- Strong convexity: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2, \ \forall x, y$
  - Strong convexity parameter: $\mu > 0$

"Strong convexity implies that f should be steep enough to make progress"



Strong convexity

– Equivalent characterizations: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2, \ \forall x, y$

$$\begin{aligned} \langle \nabla f(x) - \nabla f(y), x - y \rangle &\geq \mu \|x - y\|_2^2 \\ f(y) &\leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|_2^2 \\ \langle \nabla f(x) - \nabla f(y), x - y \rangle &\leq \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|_2^2 \\ \vdots \qquad & \quad \vdots \end{aligned}$$

Strong convexity

– Equivalent characterizations: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2, \ \forall x, y$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|_2^2$$

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|_2^2$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|_2^2$$

$$\vdots \qquad \vdots$$

– Another important one:

Interpretation?

$$\nabla^2 f(x) \succeq \mu I$$

Strong convexity

– Equivalent characterizations: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2, \ \forall x, y$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|_2^2$$

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|_2^2$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|_2^2$$

$$\vdots \qquad \vdots$$

– Another important one:

Interpretation?

$$\nabla^2 f(x) \succeq \mu I$$

– What if we also have Lipschitz continuous gradients?

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\mu L}{\mu + L}\|x - y\|_2^2 + \frac{1}{\mu + L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

Lipschitz conditions

- Equivalent characterizations: $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2, \quad \forall x, y$

(some require convexity – to be defined later)

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|x - y\|_2^2$$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L\|x - y\|_2^2$$

$$\vdots \quad \vdots$$

- Another important one:

$$\nabla^2 f(x) \preceq LI$$

Interpretation?

Strong convexity

– Equivalent characterizations: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2, \ \forall x, y$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|_2^2$$

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|_2^2$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|_2^2$$

$$\vdots \qquad \vdots$$

– Another important one:

$$\nabla^2 f(x) \succeq \mu I$$

– What if we also have Lipschitz continuous gradients?

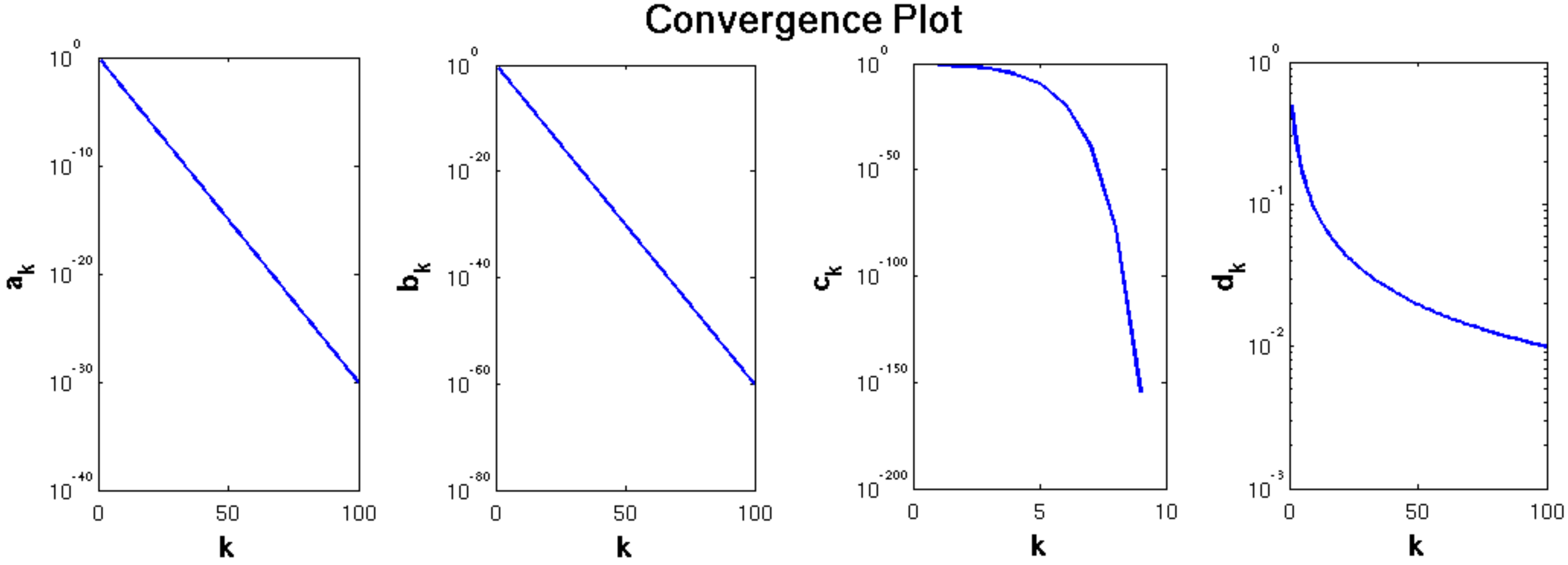
$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\mu L}{\mu + L}\|x - y\|_2^2 + \frac{1}{\mu + L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

What is the gain?

Whiteboard

Convergence rates 101

(Source: Wikipedia)



$O(\log 1/\varepsilon)$

$q^k,\ q \in (0,1)$

$O\left(\log\log(1/\varepsilon)\right)$

$O(1/\varepsilon^2),\ O(1/\varepsilon),\ O(1/\sqrt{\varepsilon})$

$O(1/k^2), O(1/k), O(\sqrt(k))$

What should be our expectations: Lower bounds

- For objectives with Lipschitz continuous gradients:

$$f(x_t) - f(x^\star) \geq \frac{3L\|x_0 - x^\star\|_2^2}{32(t+1)^2}$$

(Under these assumptions, and using only gradients, we cannot achieve better than $O\left(\frac{1}{t^2}\right)$)

What should be our expectations: Lower bounds

- For objectives with Lipschitz continuous gradients:

$$f(x_t) - f(x^\star) \geq \frac{3L\|x_0 - x^\star\|_2^2}{32(t+1)^2}$$

(Under these assumptions, and using only gradients, we cannot achieve better than $O\left(\frac{1}{t^2}\right)$)

- In addition, for objectives that are strongly convex:

$$\|x_t - x^\star\|_2^2 \geq \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2t} \|x_0 - x^\star\|_2^2 \qquad \kappa := \frac{L}{\mu}$$

(The case we described has near optimal exponent, but does not involve the square root of κ)

What should be our expectations: Lower bounds

– For objectives with Lipschitz continuous gradients:

$$f(x_t) - f(x^\star) \geq \frac{3L\|x_0 - x^\star\|_2^2}{32(t+1)^2}$$

(Under these assumptions, and using only gradients, we cannot achieve better than $O\left(\frac{1}{t^2}\right)$)

– In addition, for objectives that are strongly convex:

$$\|x_t - x^\star\|_2^2 \geq \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2t} \|x_0 - x^\star\|_2^2 \qquad \kappa := \frac{L}{\mu}$$

(The case we described has near optimal exponent, but does not involve the square root of κ)

– In future lectures: acceleration techniques that achieves these rates

Convex optimization

Demo

Are there other, more powerful, global assumptions?

- Remember, our analysis is based on: $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2, \quad \forall x, y$

(and on $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2$ when we talk about convex functions)

Are there other, more powerful, global assumptions?

- Remember, our analysis is based on: $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2, \quad \forall x, y$

(and on $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2$ when we talk about convex functions)

- Polyak−Lojasiewicz (PL) inequality

$$\frac{1}{2}\|\nabla f(x)\|_2^2 \geq \xi(f(x) - f(x^\star)), \quad \forall x, \quad \text{for some } \xi > 0$$

(Any thoughts about what this implies?)

Are there other, more powerful, global assumptions?

- Remember, our analysis is based on: $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2, \quad \forall x, y$

  (and on $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2$ when we talk about convex functions)

- Polyak–Lojasiewicz (PL) inequality

$$\frac{1}{2}\|\nabla f(x)\|_2^2 \geq \xi(f(x) - f(x^\star)), \quad \forall x, \quad \text{for some } \xi > 0$$

(Any thoughts about what this implies?)

- Using PL inequality + Lipschitz gradient continuity:

$$f(x_t) - f(x^\star) \leq \left(1 - \frac{\xi}{L}\right)^t (f(x_0) - f(x^\star)) \qquad \text{Whiteboard}$$

Are there other, more powerful, global assumptions?

- Remember, our analysis is based on: $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2, \quad \forall x, y$

  (and on $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2$ when we talk about convex functions)

- Polyak–Lojasiewicz (PL) inequality

$$\frac{1}{2}\|\nabla f(x)\|_2^2 \geq \xi(f(x) - f(x^\star)), \quad \forall x, \quad \text{for some } \xi > 0$$

(Any thoughts about what this implies?)

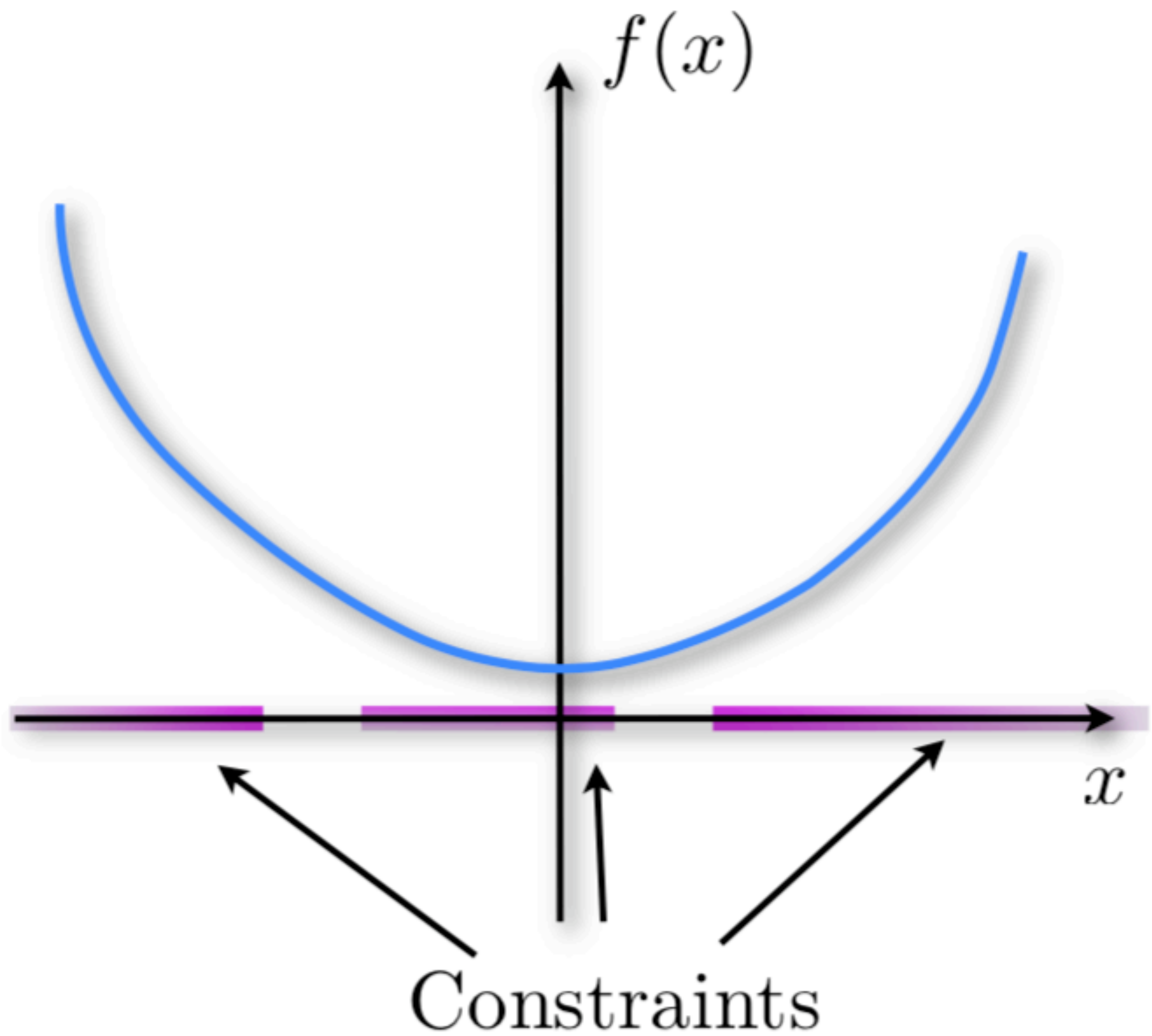
- Using PL inequality + Lipschitz gradient continuity:

$$f(x_t) - f(x^\star) \leq \left(1 - \frac{\xi}{L}\right)^t (f(x_0) - f(x^\star))$$

Whiteboard

- Does not use **convexity**: holds for invex functions (stationary = global)

Convex optimization is not only about the objective



- Back to the first slide:

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{s.t.} \quad & x \in \mathcal{C}
\end{aligned}
$$

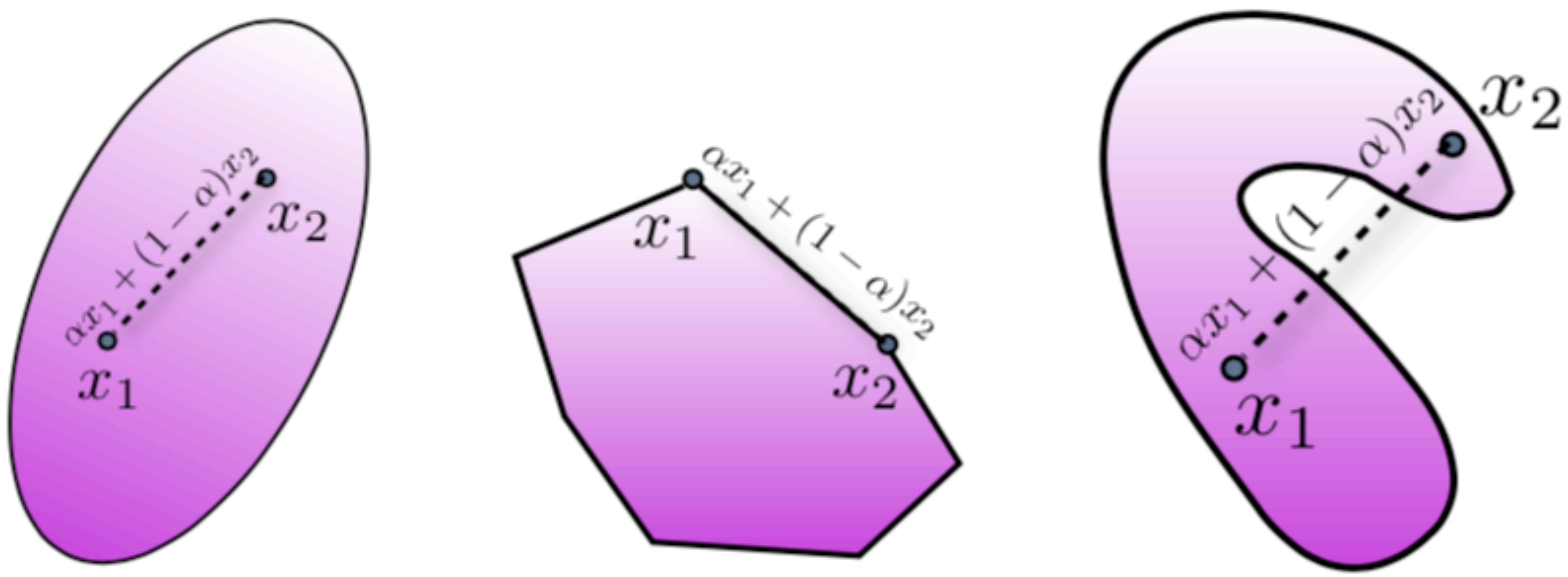
(We will worry about this in the lectures to follow!)

Convex sets

$\mathcal{C} \subseteq \mathbb{R}^p$ is convex if $\forall x_1, x_2 \in \mathcal{C},$ it holds $\forall \alpha \in [0,1] : \alpha x_1 + (1-\alpha) x_2 \in \mathcal{C}$

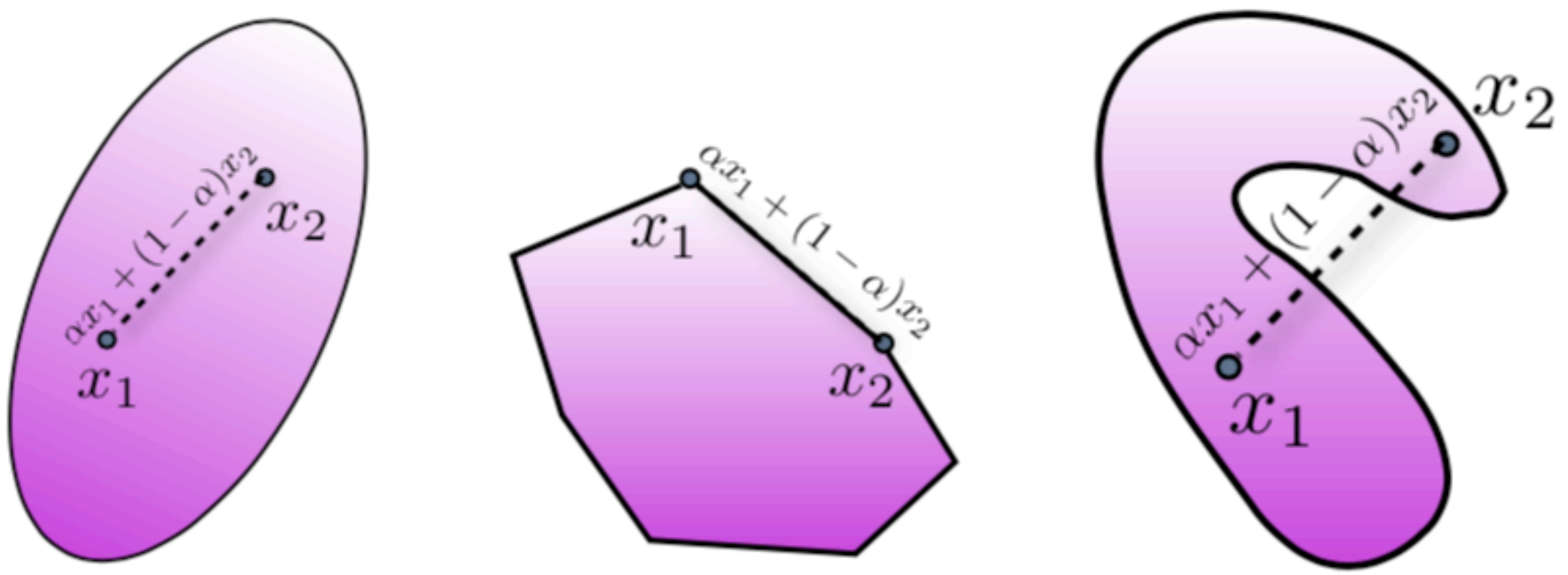
Convex sets

$\mathcal{C} \subseteq \mathbb{R}^p$ is convex if $\forall x_1, x_2 \in \mathcal{C}$, it holds $\forall \alpha \in [0,1] : \alpha x_1 + (1-\alpha) x_2 \in \mathcal{C}$



Convex sets

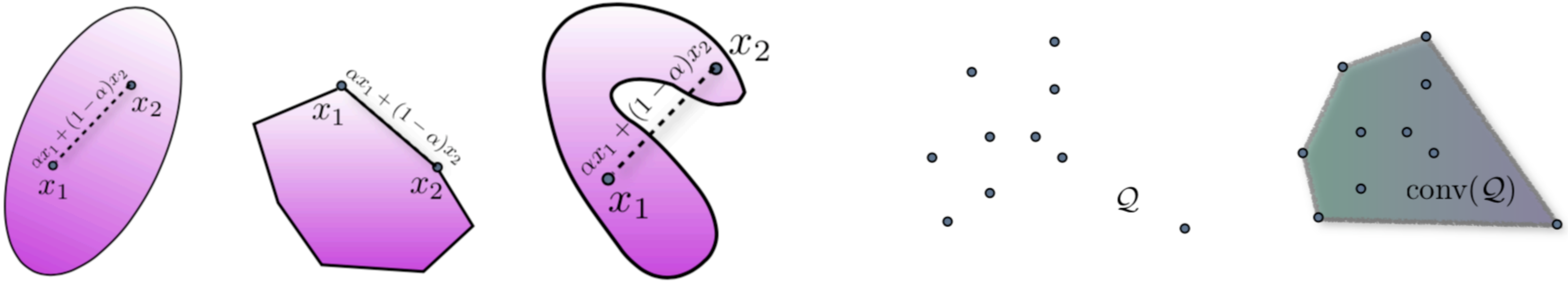
$\mathcal{C} \subseteq \mathbb{R}^p$ is convex if $\forall x_1, x_2 \in \mathcal{C}$, it holds $\forall \alpha \in [0,1] : \alpha x_1 + (1-\alpha) x_2 \in \mathcal{C}$



– Convex hull of points: $\mathrm{conv}(\mathcal{V}) = \left\{ \sum_{i=1}^{|\mathcal{V}|} \alpha_i x_i \; : \; \sum_{i=1}^{|\mathcal{V}|} \alpha_i = 1, \; \alpha_i \geq 0, x_i \in \mathcal{V} \right\}$

Convex sets

$\mathcal{C} \subseteq \mathbb{R}^p$ is convex if $\forall x_1, x_2 \in \mathcal{C}$, it holds $\forall \alpha \in [0,1] : \alpha x_1 + (1-\alpha) x_2 \in \mathcal{C}$



– Convex hull of points: $\mathrm{conv}(\mathcal{V}) = \left\{ \sum_{i=1}^{|\mathcal{V}|} \alpha_i x_i \ : \ \sum_{i=1}^{|\mathcal{V}|} \alpha_i = 1,\ \alpha_i \geq 0, x_i \in \mathcal{V} \right\}$

Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary
and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary
and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

Projections onto convex sets

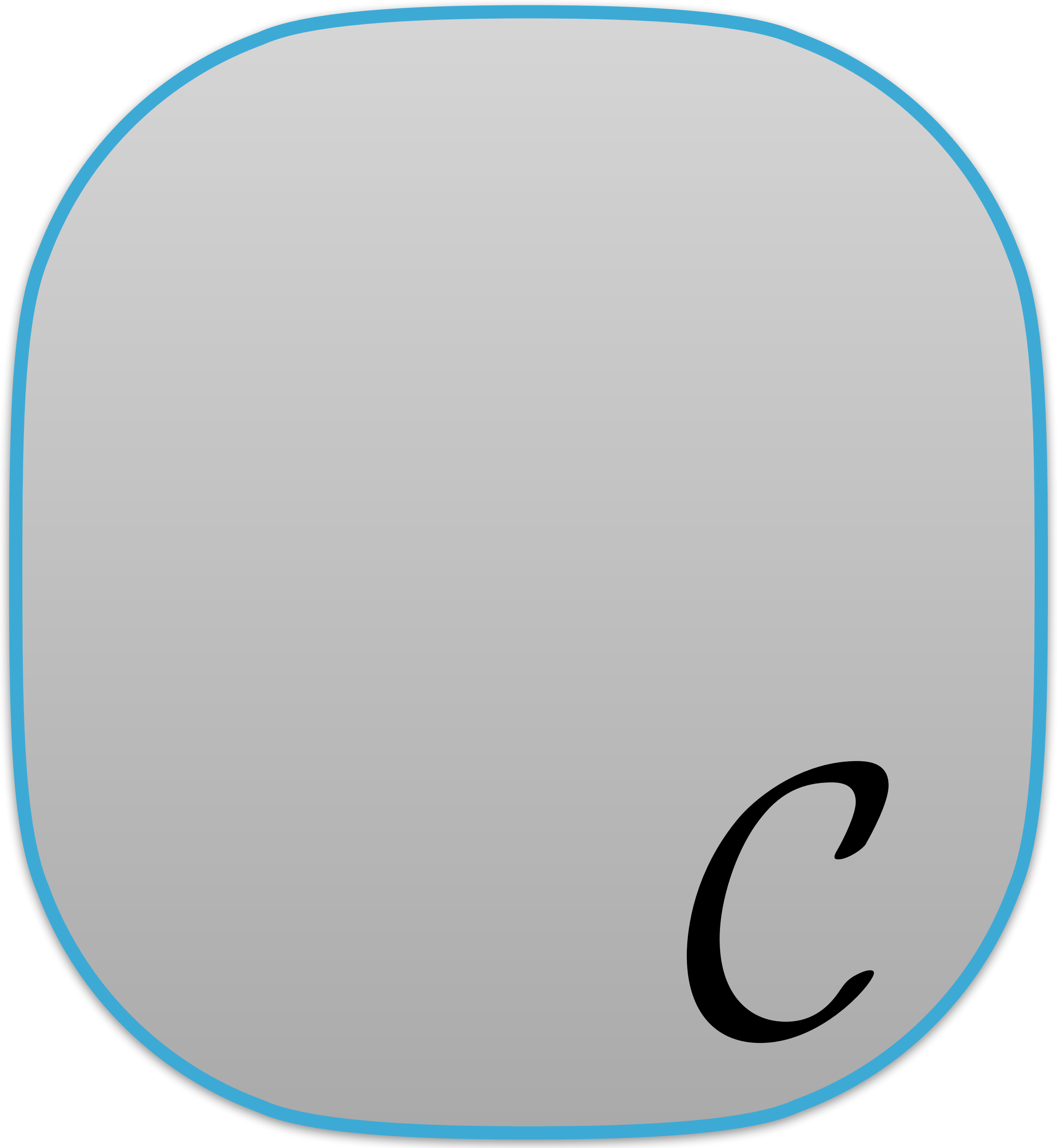
$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

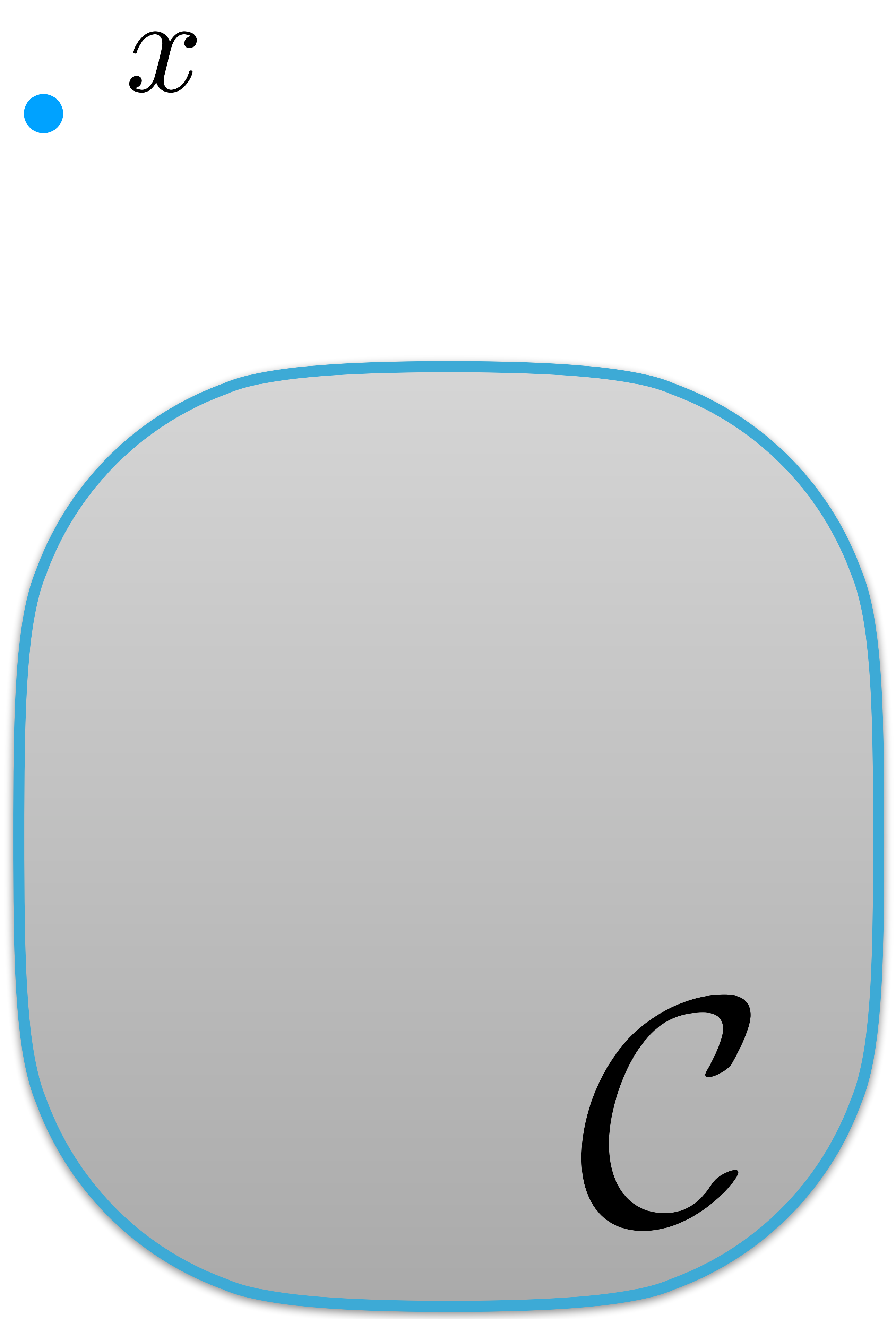
$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

- Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \le \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

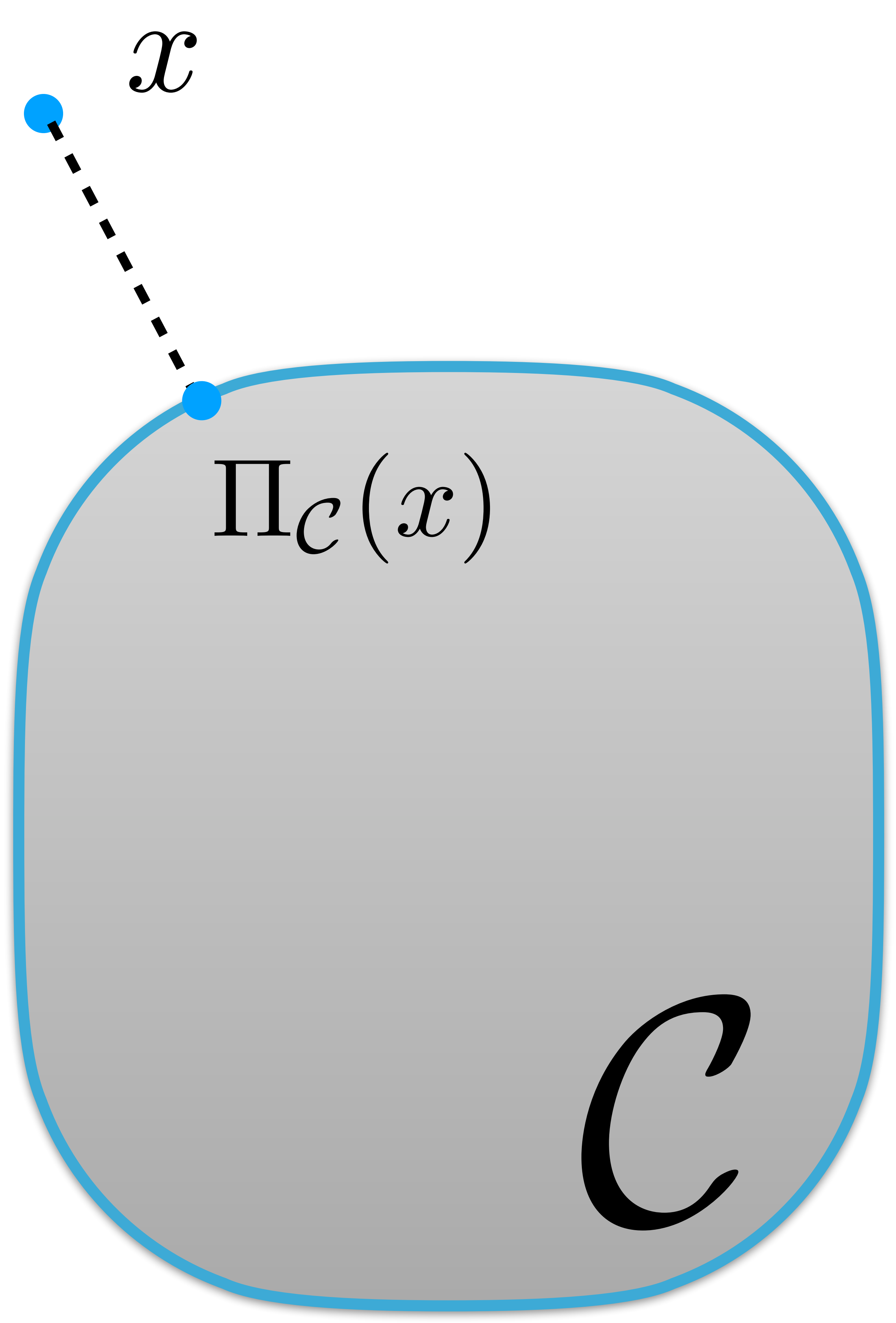
$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary
and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

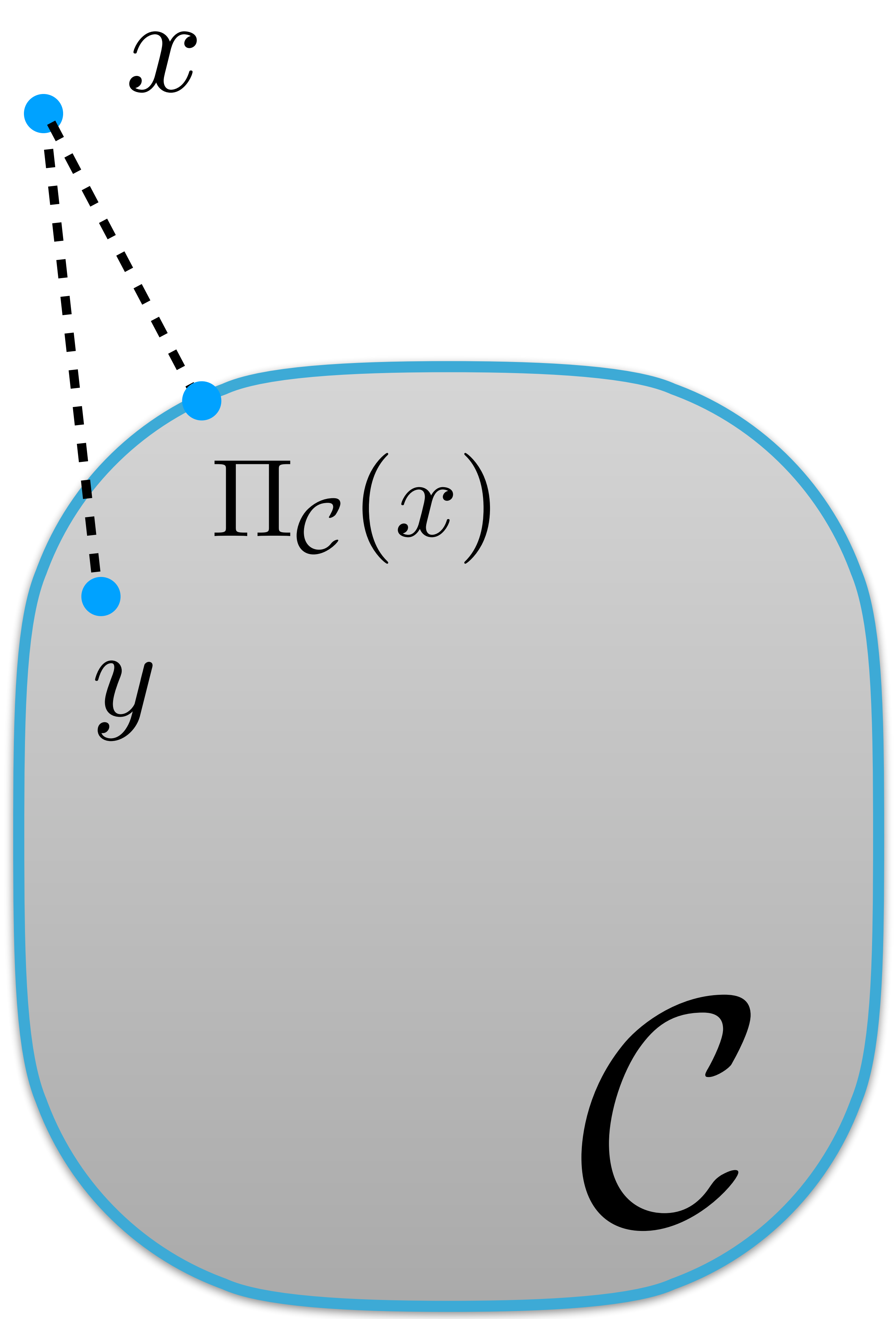
$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary
and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$

Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

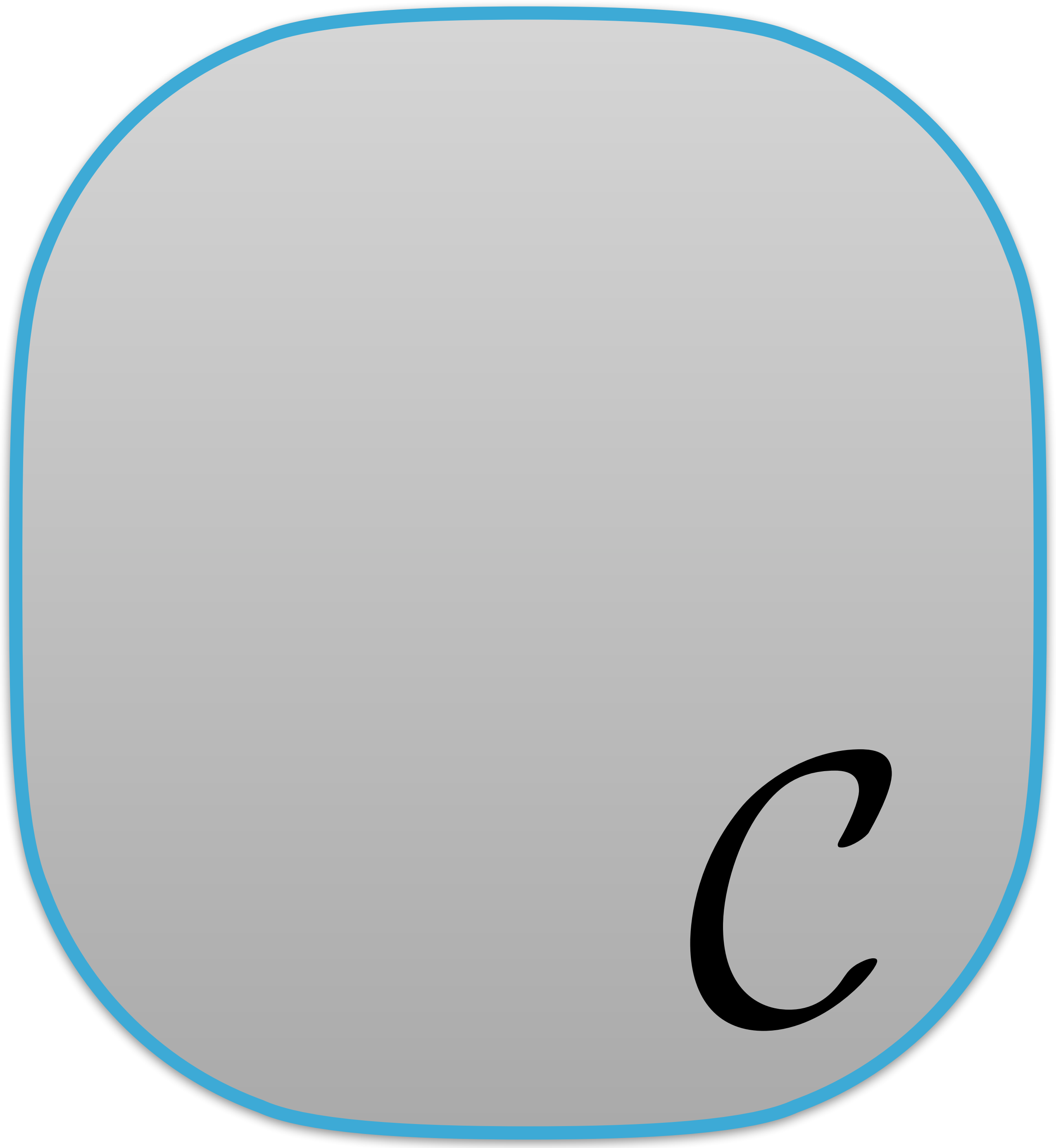
(The use of Euclidean norm is arbitrary
and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

- Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

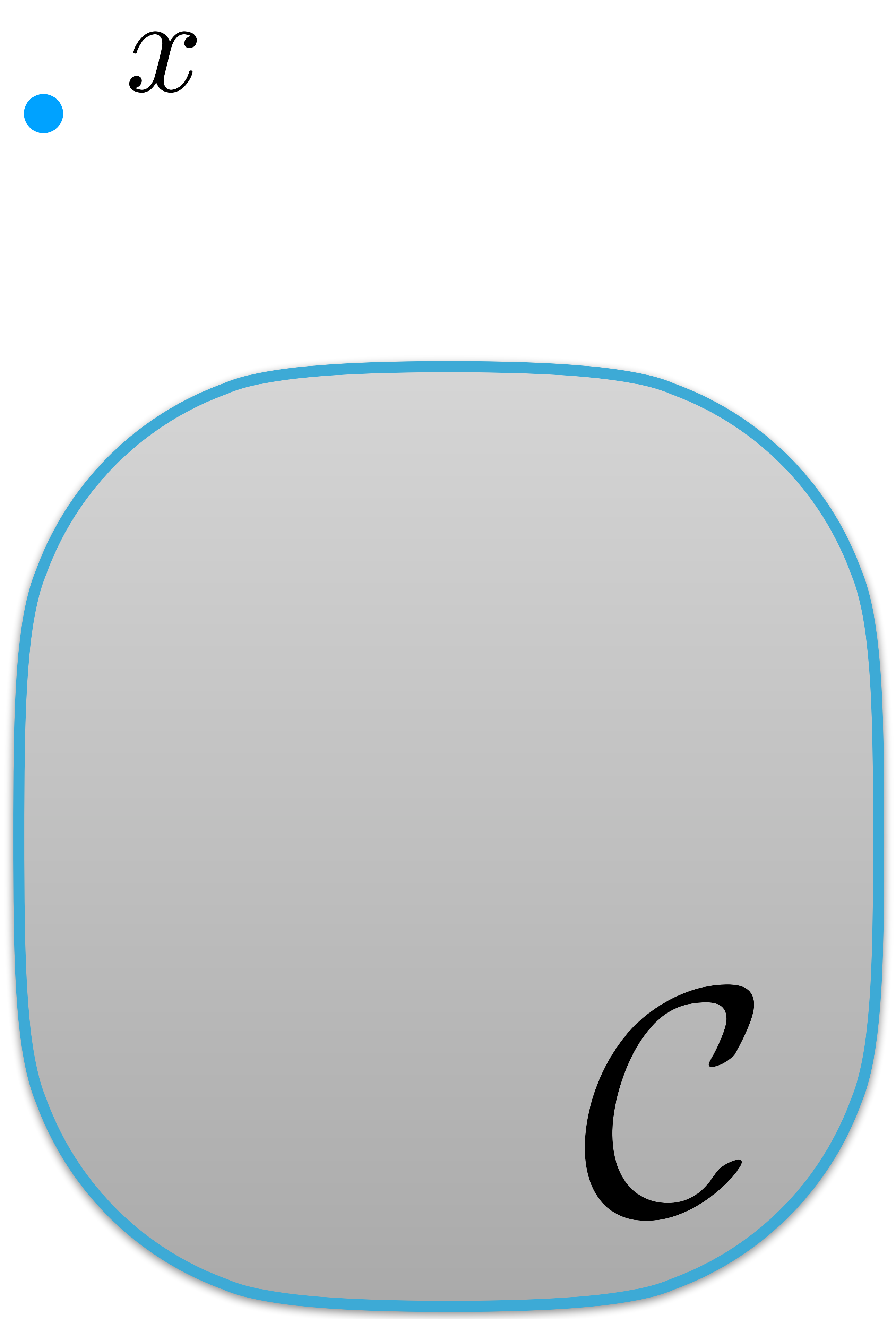
(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

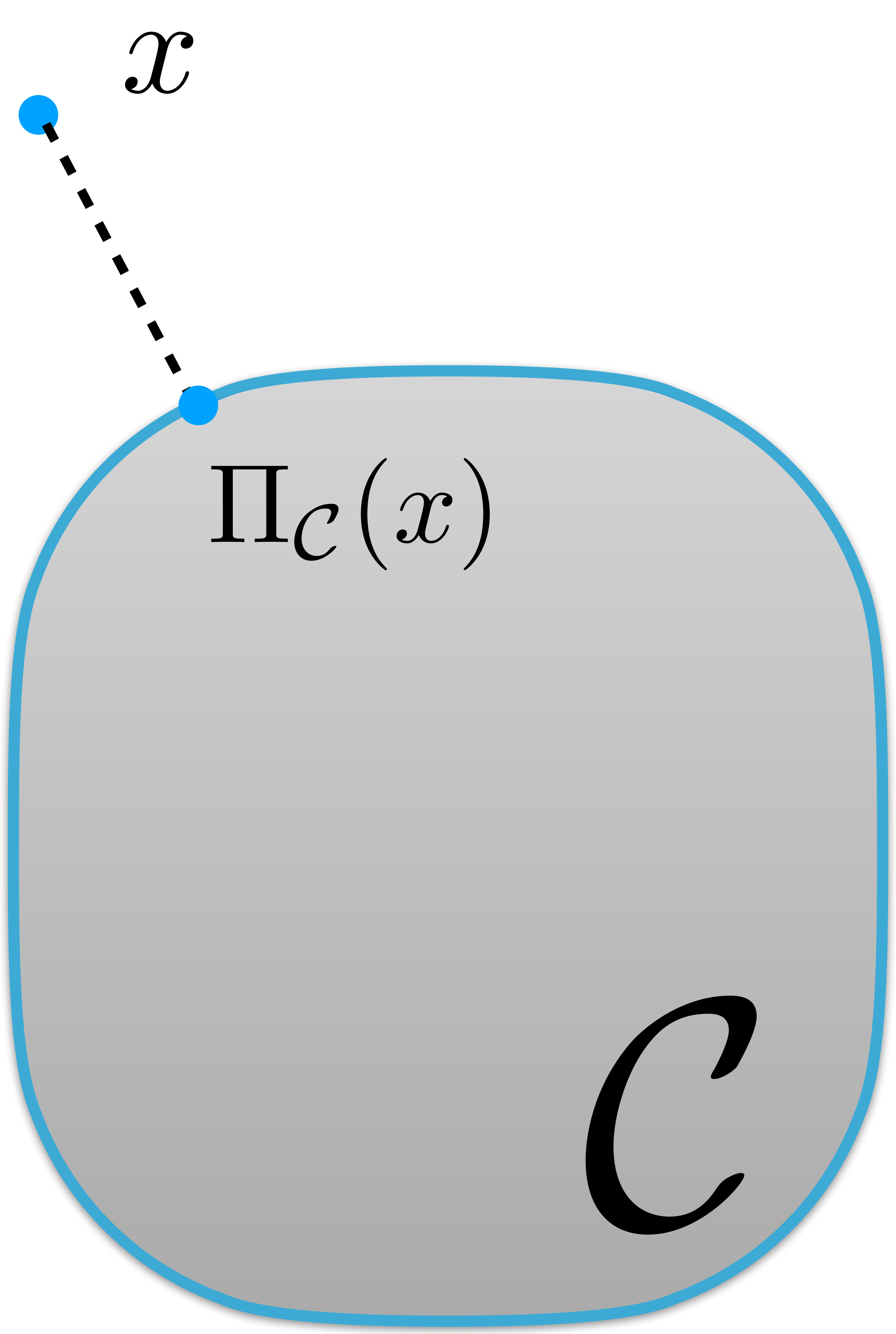
(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

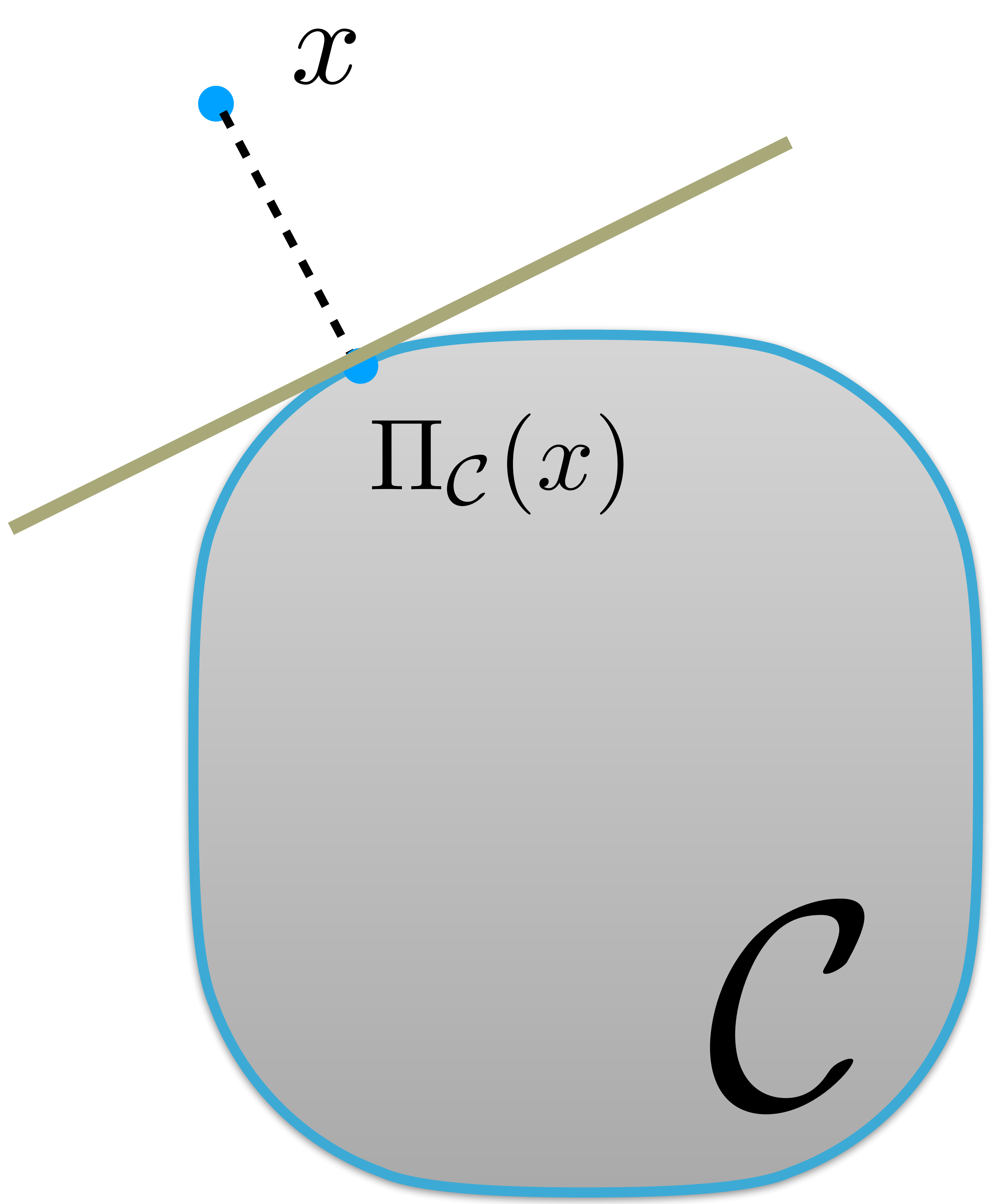
(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

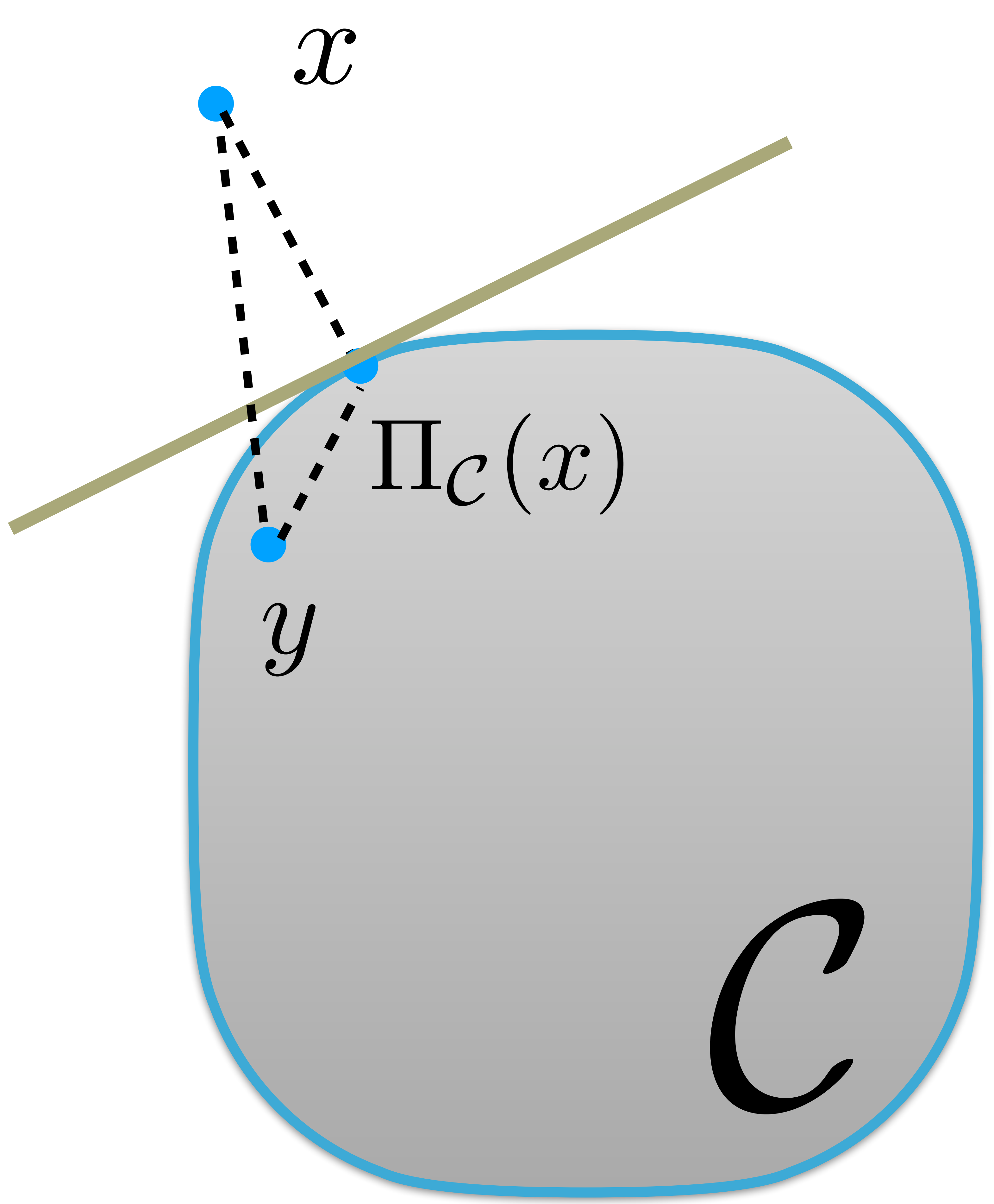
(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

- Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

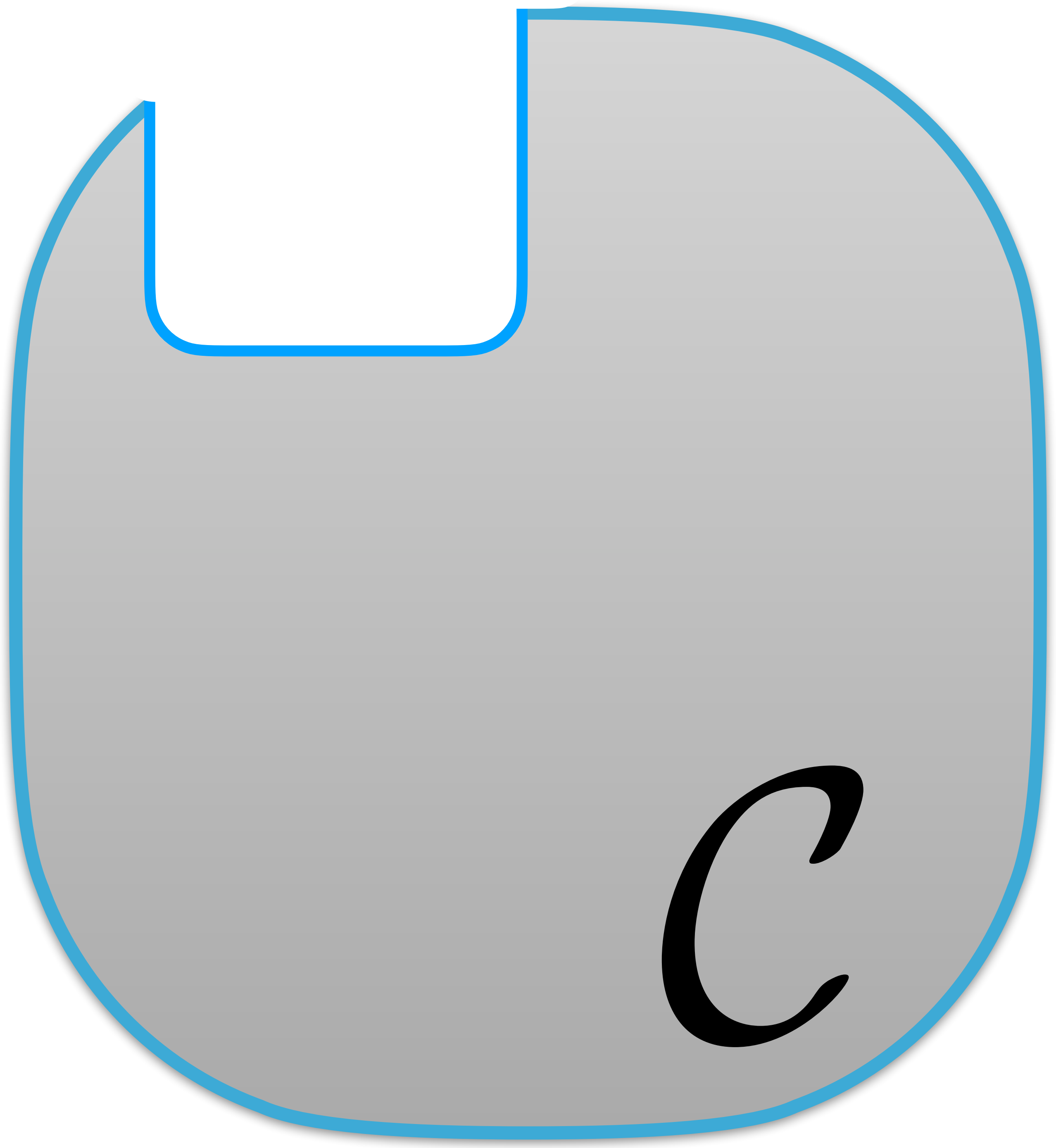
(The use of Euclidean norm is arbitrary
and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \; \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

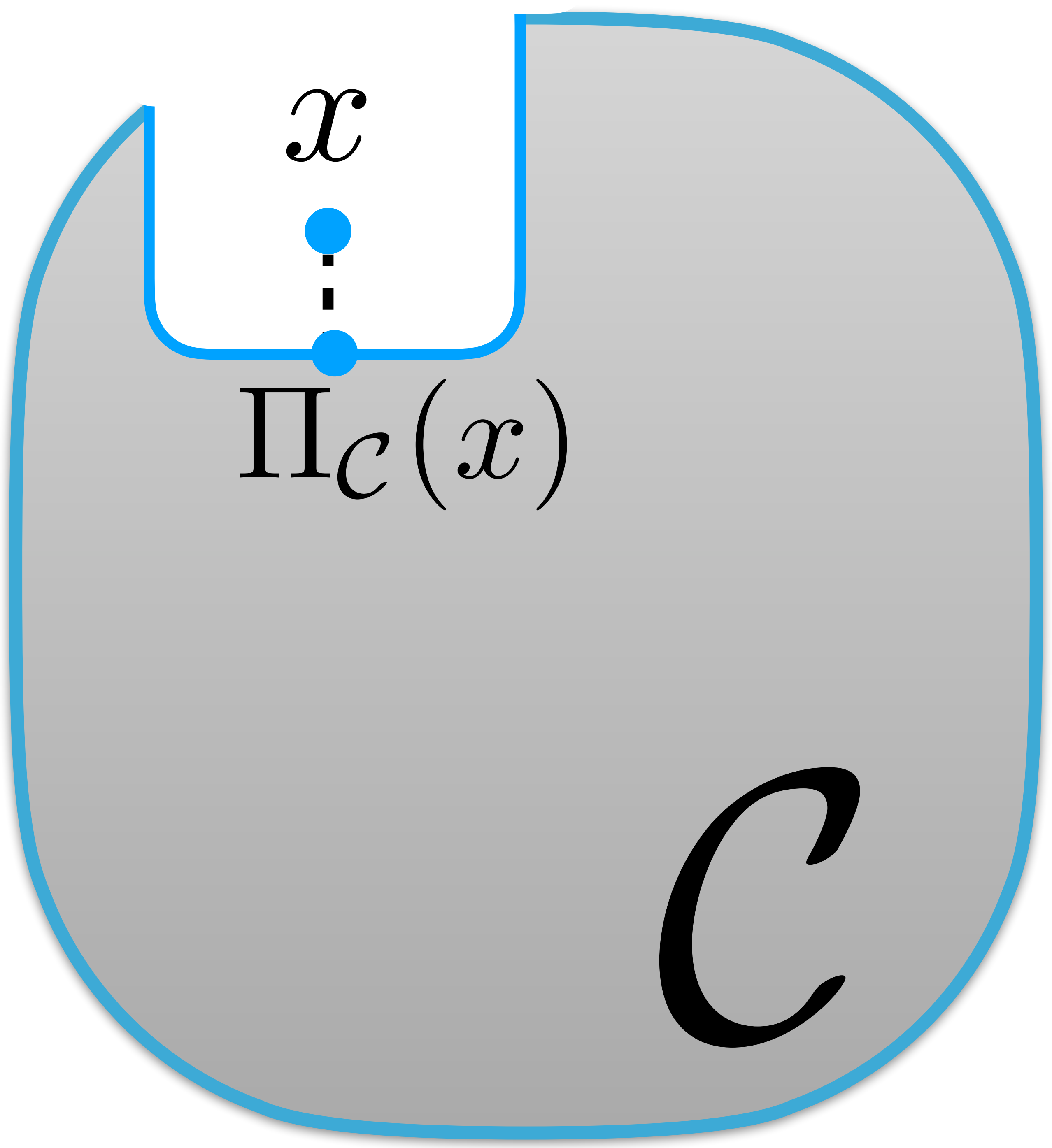
(The use of Euclidean norm is arbitrary
and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

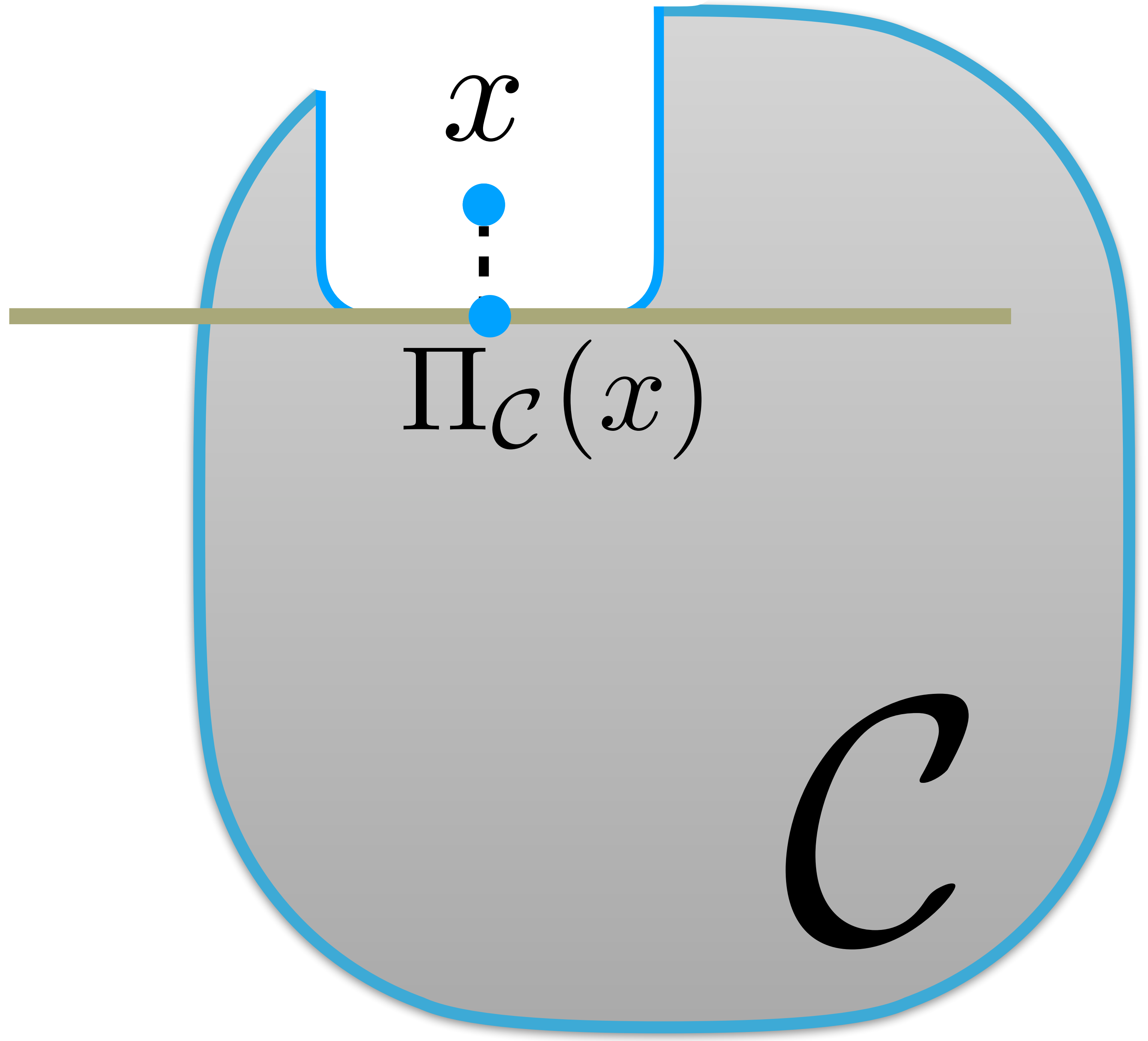
(The use of Euclidean norm is arbitrary
and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

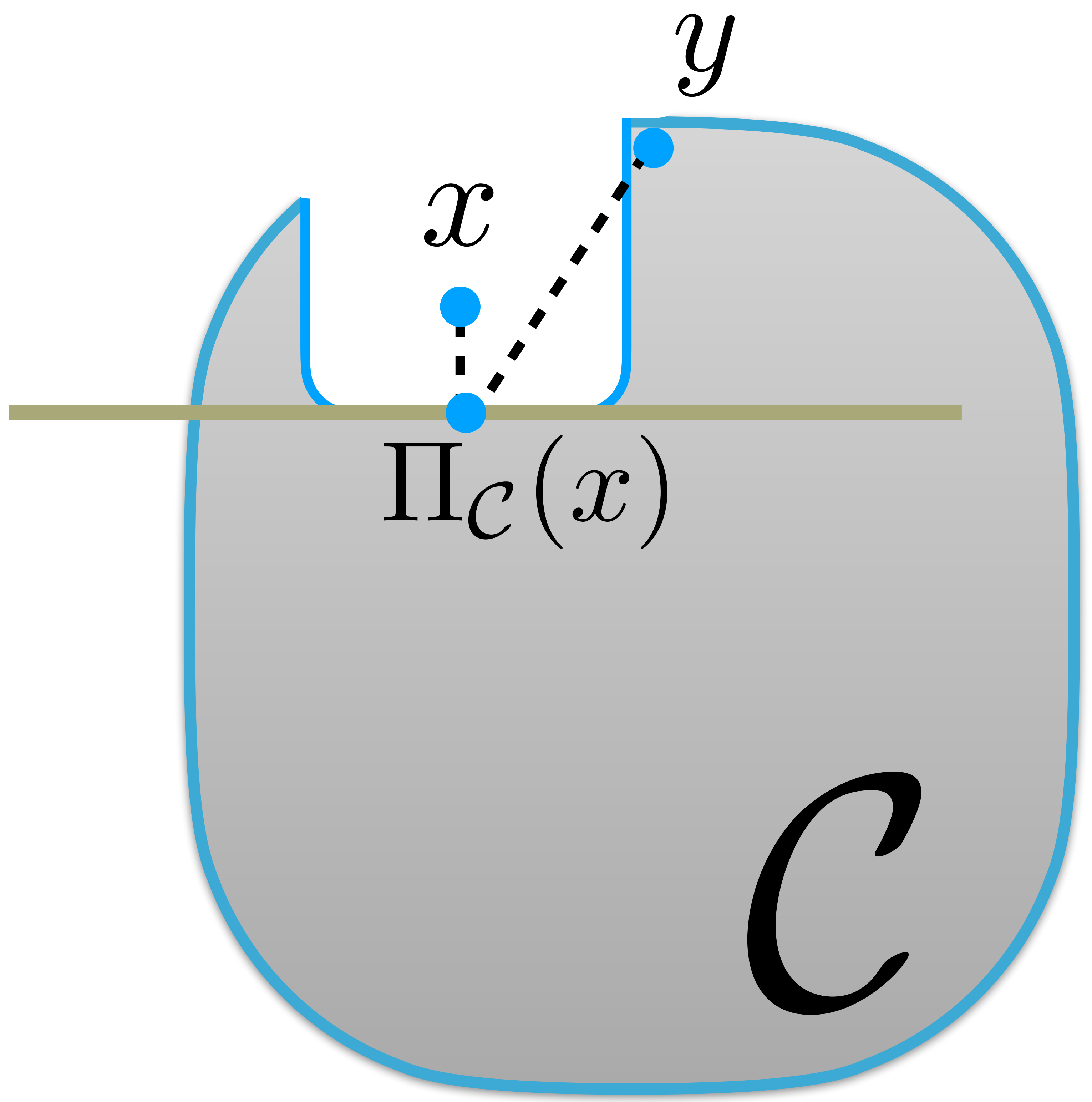
(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \; \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary
and often depends on the application)

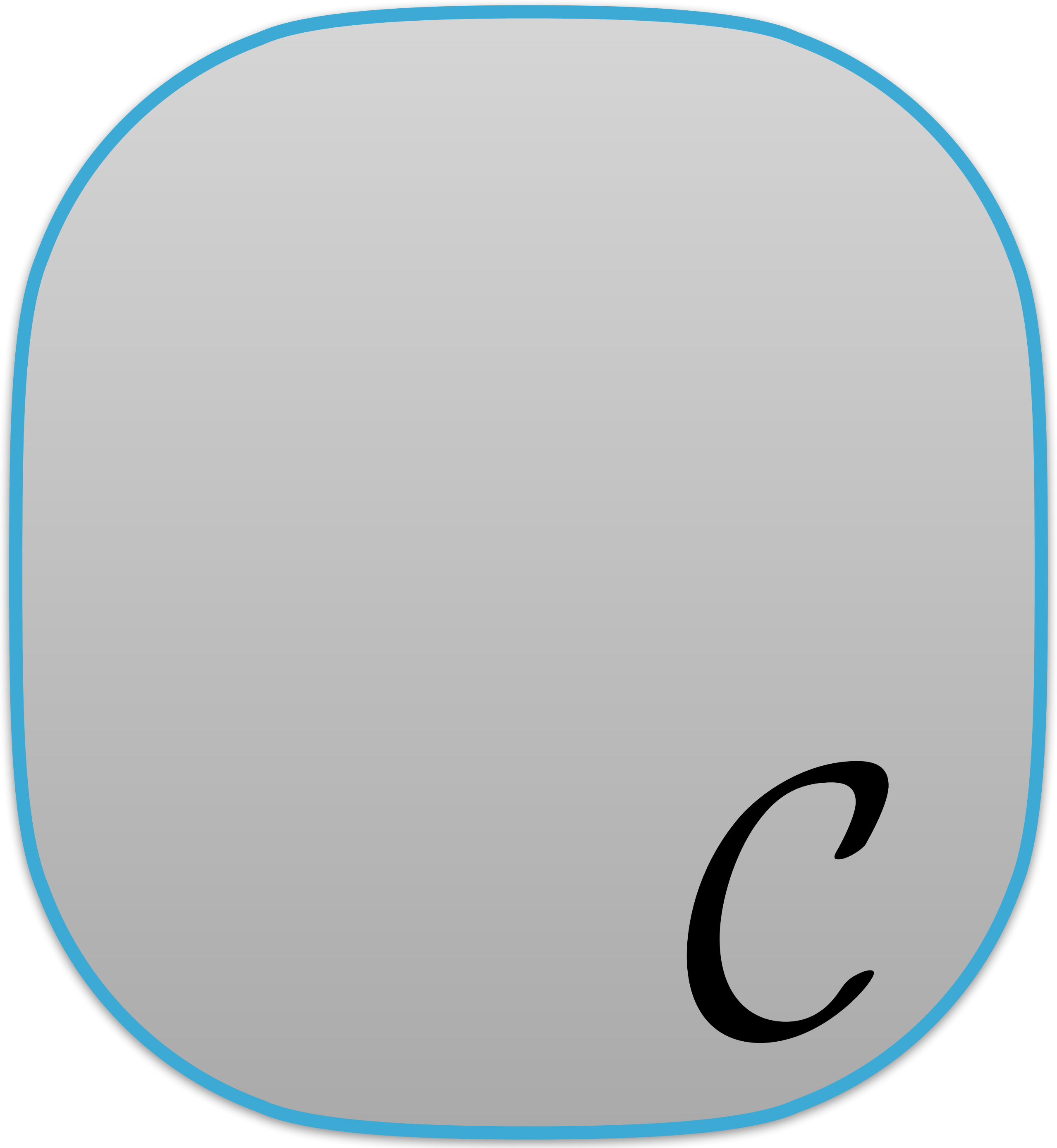
$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

- Key properties of convex sets

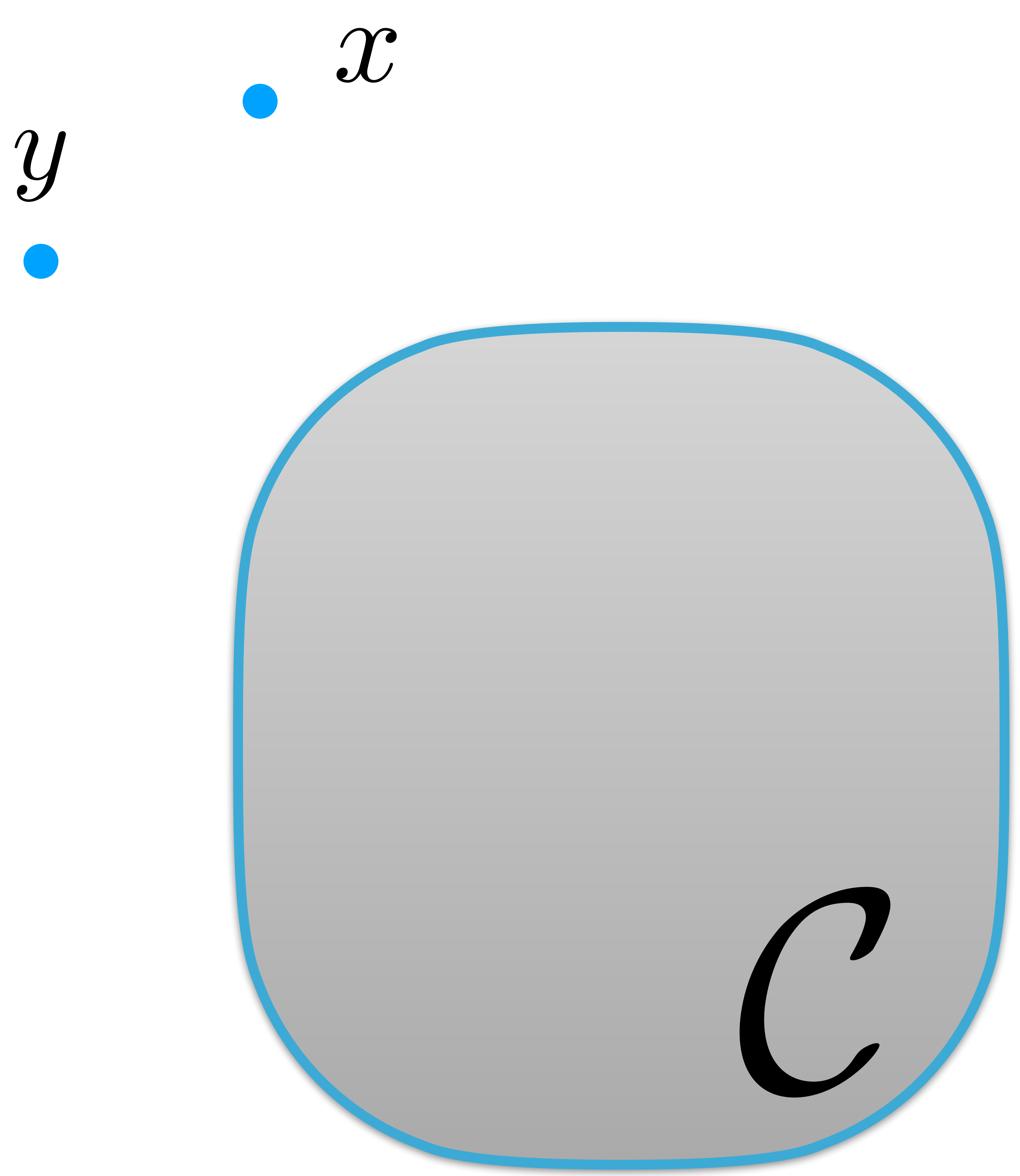
$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$

$$\|\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(y)\|_2 \leq \|x - y\|_2, \ \forall x, y$$



Projections onto convex sets



$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary and often depends on the application)

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$

$$\|\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(y)\|_2 \leq \|x - y\|_2, \ \forall x, y$$

Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary
and often depends on the application)

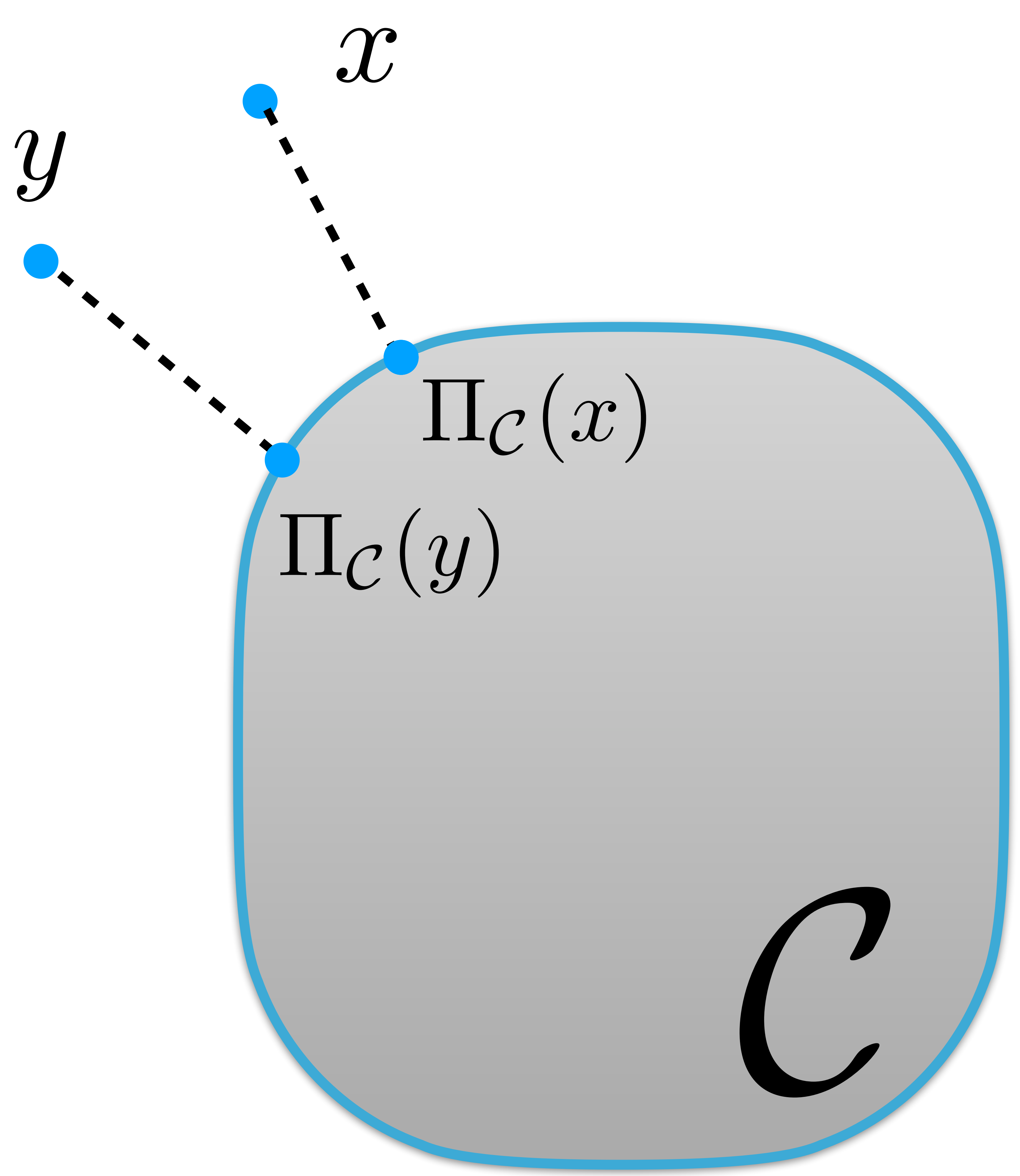
$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$

$$\|\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(y)\|_2 \leq \|x - y\|_2, \ \forall x, y$$



Projections onto convex sets

$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_2^2$$

(The use of Euclidean norm is arbitrary
and often depends on the application)

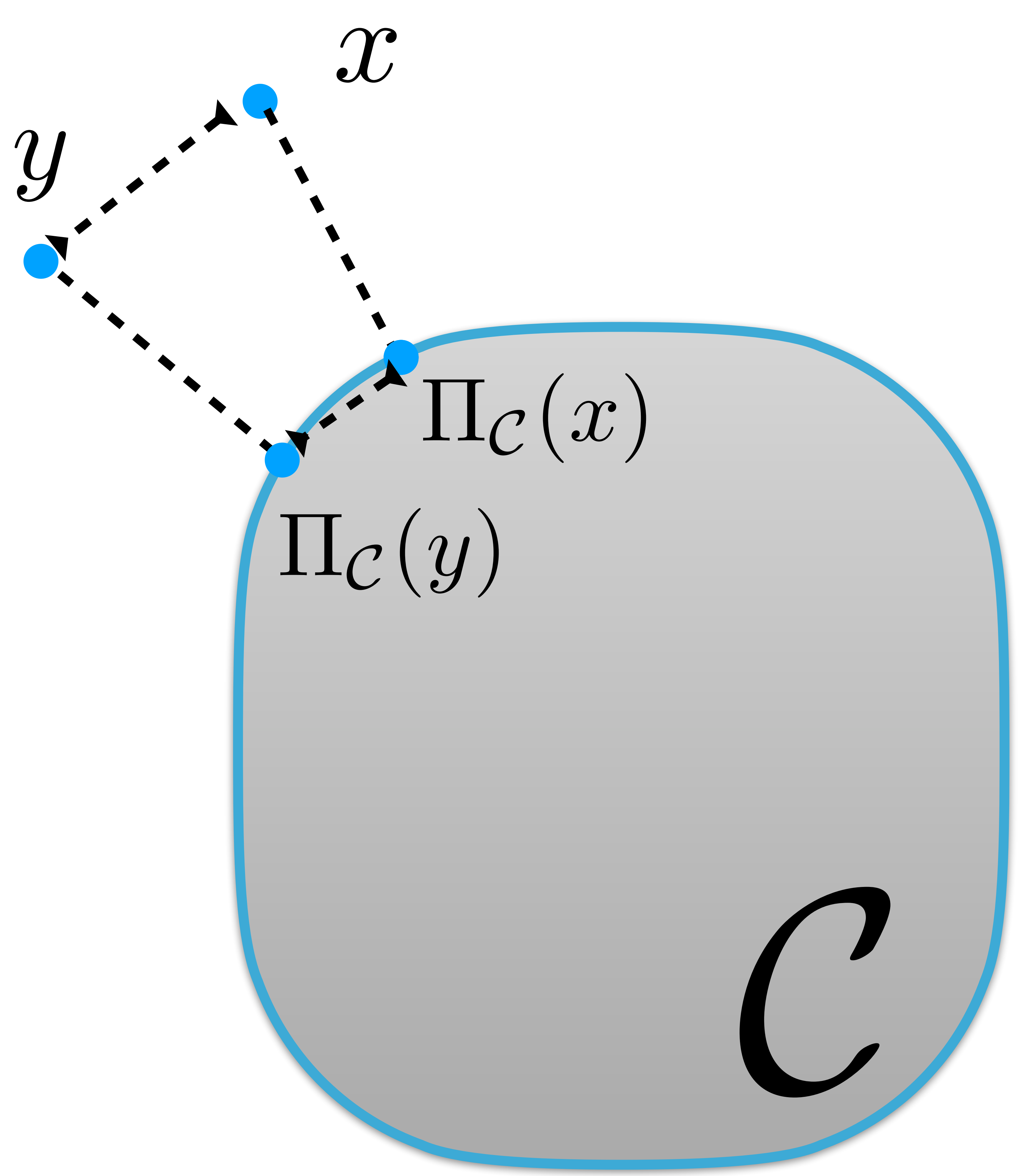
$$\Pi_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|_1$$

– Key properties of convex sets

$$\|x - \Pi_{\mathcal{C}}(x)\|_2^2 \leq \|x - y\|_2^2, \ \forall y \in \mathcal{C}, \forall x$$

$$\langle \Pi_{\mathcal{C}}(x) - y, \Pi_{\mathcal{C}}(x) - x \rangle \leq 0, \forall y \in \mathcal{C}, \forall x$$

$$\|\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(y)\|_2 \leq \|x - y\|_2, \ \forall x, y$$



Projected gradient descent

$$x_{t+1} = \Pi_{\mathcal{C}}\left(x_t - \eta \nabla f(x_t)\right)$$

Projected gradient descent

$$x_{t+1} = \Pi_{\mathcal{C}}\left(x_t - \eta \nabla f(x_t)\right)$$

- A two-step procedure:
  1. $\widetilde{x} = x_t - \eta \nabla f(x_t)$
  2. $x_{t+1} = \Pi_{\mathcal{C}}\left(\widetilde{x}\right)$

Projected gradient descent

$$x_{t+1} = \Pi_{\mathcal{C}} \left(x_t - \eta \nabla f(x_t) \right)$$

- A two-step procedure:
  1. $\widetilde{x} = x_t - \eta \nabla f(x_t)$
  2. $x_{t+1} = \Pi_{\mathcal{C}} \left(\widetilde{x} \right)$

Demo

Projected gradient descent

$$x_{t+1} = \Pi_{\mathcal{C}}\left(x_t - \eta \nabla f(x_t)\right)$$

- What about its convergence guarantees? Do we lose much by projecting?

Projected gradient descent

$$x_{t+1} = \Pi_{\mathcal{C}} \left(x_t - \eta \nabla f(x_t) \right)$$

– What about its convergence guarantees? Do we lose much by projecting?

Whiteboard

But wait; didn't we consider proj. GD before?

$$x_{t+1} = \Pi_{\mathcal{C}}\left(x_t - \eta \nabla f(x_t)\right)$$

But wait; didn't we consider proj. GD before?

$$x_{t+1} = \Pi_{\mathcal{C}} \left(x_t - \eta \nabla f(x_t) \right)$$

- Yes, in the case of exact sparse linear regression:

$$\mathcal{C} = \{ x \in \mathbb{R}^p \;\; : \;\; \|x\|_0 \leq k \}$$

Is this set convex? If no or yes, why?

But wait; didn't we consider proj. GD before?

$$x_{t+1} = \Pi_{\mathcal{C}}\left(x_t - \eta \nabla f(x_t)\right)$$

- Yes, in the case of exact sparse linear regression:

$$\mathcal{C} = \{x \in \mathbb{R}^p \ : \ \|x\|_0 \leq k\}$$

Is this set convex? If no or yes, why?

- But we observed that, despite non-convexity, it works just fine..

(Thus, a different analysis is needed, depending on the problem at hand)

But, constrained optimization can be hard..

$$x_{t+1} = \Pi_{\mathcal{C}}\left(x_t - \eta \nabla f(x_t)\right)$$

– There are constrained problems where we need exponentially many bits even to describe the solution..

But, constrained optimization can be hard..

$$x_{t+1} = \Pi_{\mathcal{C}} \left(x_t - \eta \nabla f(x_t) \right)$$

– There are constrained problems where we need exponentially many bits even to describe the solution..

– We considered cases where the constraints are simple:
Operations Research is an area where multiple, difficult constraints appear

But, constrained optimization can be hard..

$$x_{t+1} = \Pi_{\mathcal{C}}\left(x_t - \eta \nabla f(x_t)\right)$$

– There are constrained problems where we need exponentially many bits even to describe the solution..

– We considered cases where the constraints are simple:
Operations Research is an area where multiple, difficult constraints appear

– Prof. Richard Tapia is teaching a course on constrained convex opt.

Convex optimization: Is it a technology?

- Yes, we know about it more than most other areas of optimization

Convex optimization: Is it a technology?

- Yes, we know about it more than most other areas of optimization

- Yes, there are off-the-shelf solvers available online

CVXOPT – https://cvxopt.org
CVXPy – http://www.cvxpy.org/
CVX – http://cvxr.com/cvx/
JuliaOpt – https://www.juliaopt.org/

Many optimizers in NN training can be applied to convex problems

TensorFlow – https://www.tensorflow.org/
PyTorch – https://pytorch.org/

Convex optimization: Is it a technology?

- Yes, we know about it more than most other areas of optimization
- Yes, there are off-the-shelf solvers available online

CVXOPT – https://cvxopt.org
CVXPy – http://www.cvxpy.org/
CVX – http://cvxr.com/cvx/
JuliaOpt – https://www.juliaopt.org/

Many optimizers in NN training can be applied to convex problems
TensorFlow – https://www.tensorflow.org/
PyTorch – https://pytorch.org/

- Why should we still care about convex optimization?

Several practical problems are actually convex
Many practical problems can be approximated by convex ones
If one doesn't understand convex opt., why even try understanding non-convex opt.?

Conclusion

- We have introduced the notion of convexity
- We studied some of the merits of convex optimization

Conclusion

- We have introduced the notion of convexity
- We studied some of the merits of convex optimization

Next lecture

- We will consider an important variant for convex optimization for large-scale computing: Frank-Wolfe (conditional gradient) algorithm