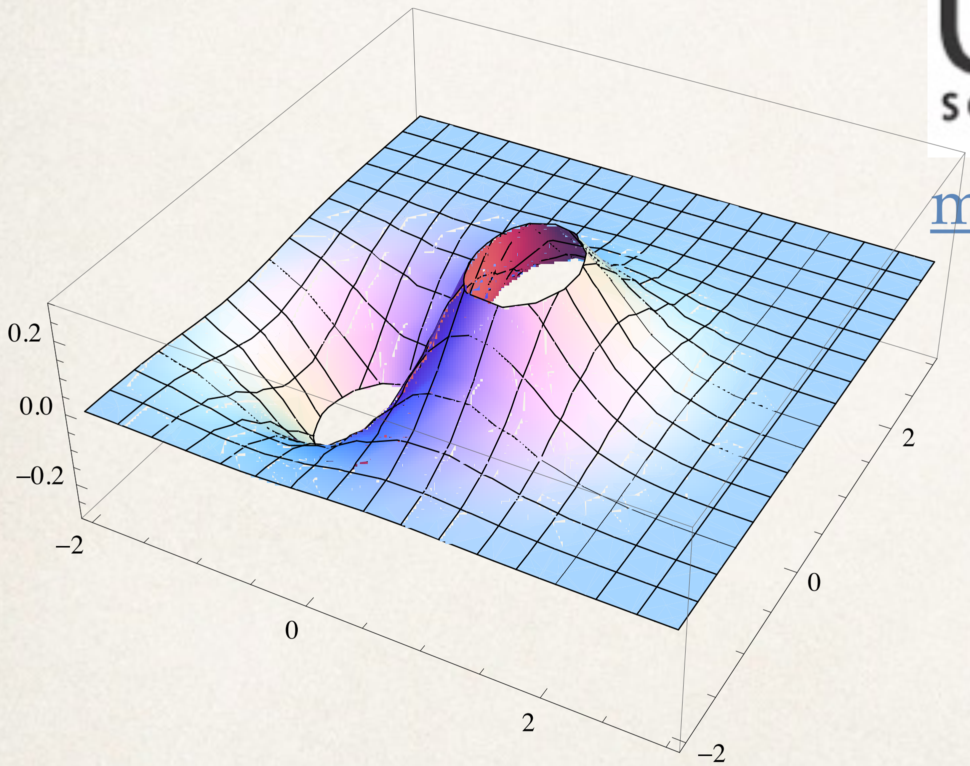
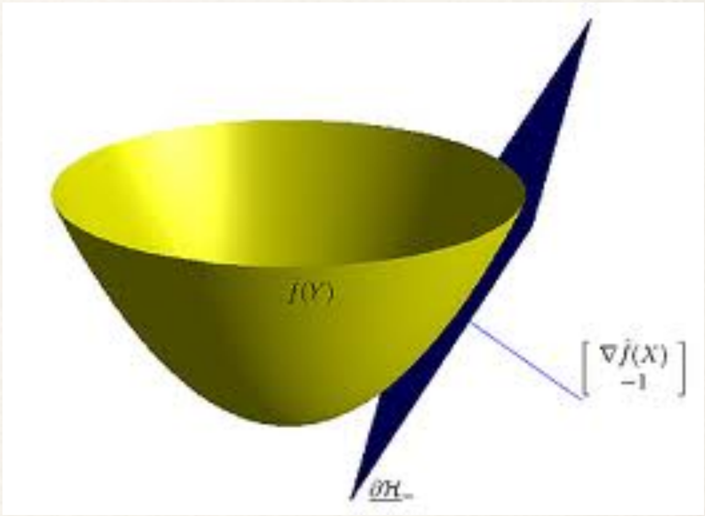
Optimization for CS

Sridhar Mahadevan

UMASS CS
SCHOOL OF COMPUTER SCIENCE

mahadeva@cs.umass.edu





Linear Programming

revenue functional

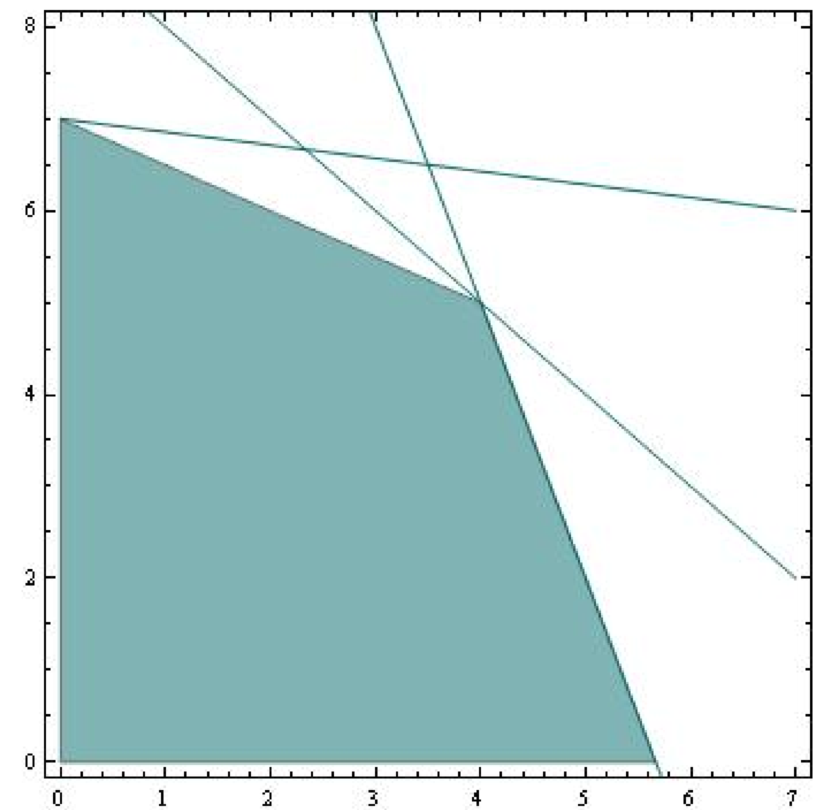
$$\max_{x,y} \quad p_1 x + p_2 y$$

such that

$$x + 2y \leqslant 14$$

$$3x + y \leqslant 17$$

$$x, y \geqslant 0$$



Example 1.3.1. Your small bakery can produce only two products: frosted cookies and cakes. A batch of frosted cookies uses up 1 pound of flour and 3 pounds of sugar. A batch of cake uses up 2 pounds of flour and 1 pound of sugar. Each day your suppliers bring you 14 pounds of flour and 17 pounds of sugar. Your optimization problem is to look at the market price of cookies and cakes and decide what to produce.

Operations preserving convexity of functions

- Nonnegative weighted sums: 3f + 2g is convex if f and g are
- Composition with affine functions: f(Ax + b) is convex if f is
- Pointwise maximum and supremum: f(x) = max{g(x), h(x), q(x)}
  - Maximum eigenvalue of symmetric matrix $\lambda_{\max}(X) = \sup_{\|y\|_2=1} y^T X y$
- Minimization: distance to convex set $h(x) = \inf_{y \in C} \|x - y\|$
- Composition

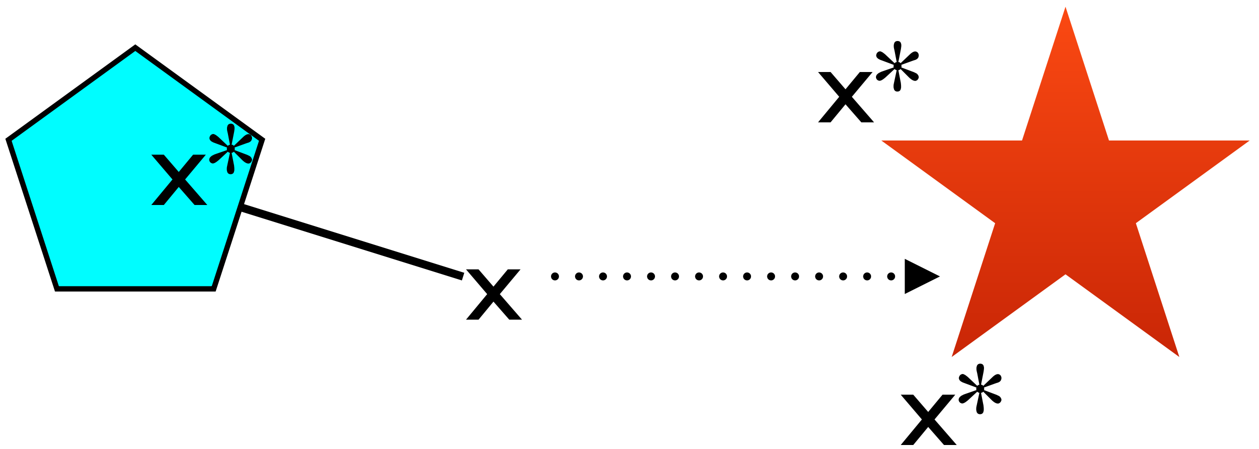
Matrix completion:Netflix problem

- One way to formulate the matrix completion problem:
  - Minimize $|X|_*$ such that $X_{i,j} = M_{i,j}$ (over observed samples)
  - Trace norm $|X|_* =$ sum of singular values of X
- Why is this a convex optimization problem?
  - Show maximum singular value can be written as

$$\sigma_{\max} = \sup_{\|x\|=1, \|y\|=1} x' A y$$

Projections

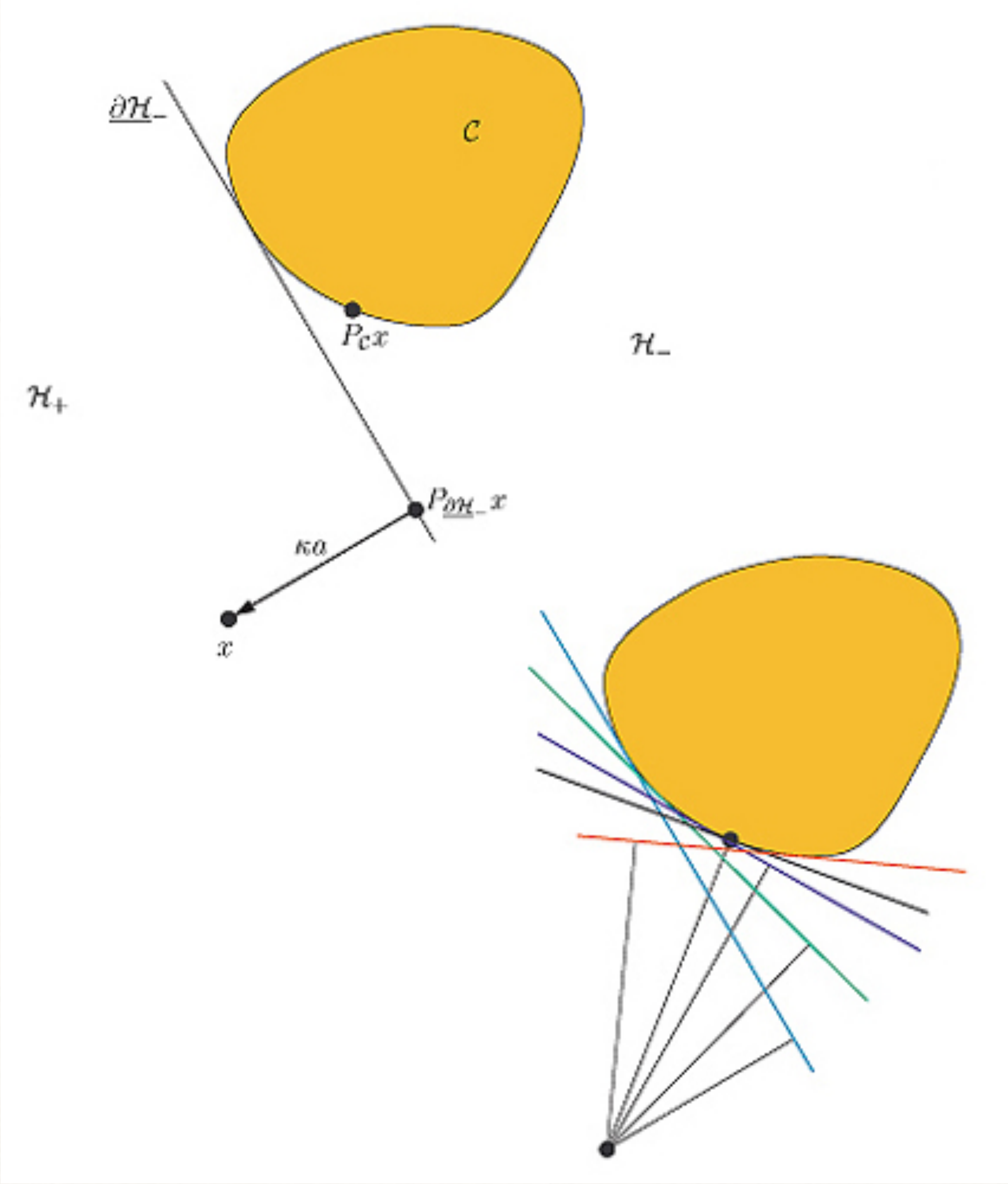
- The concept of a projection is fundamental to many optimization methods



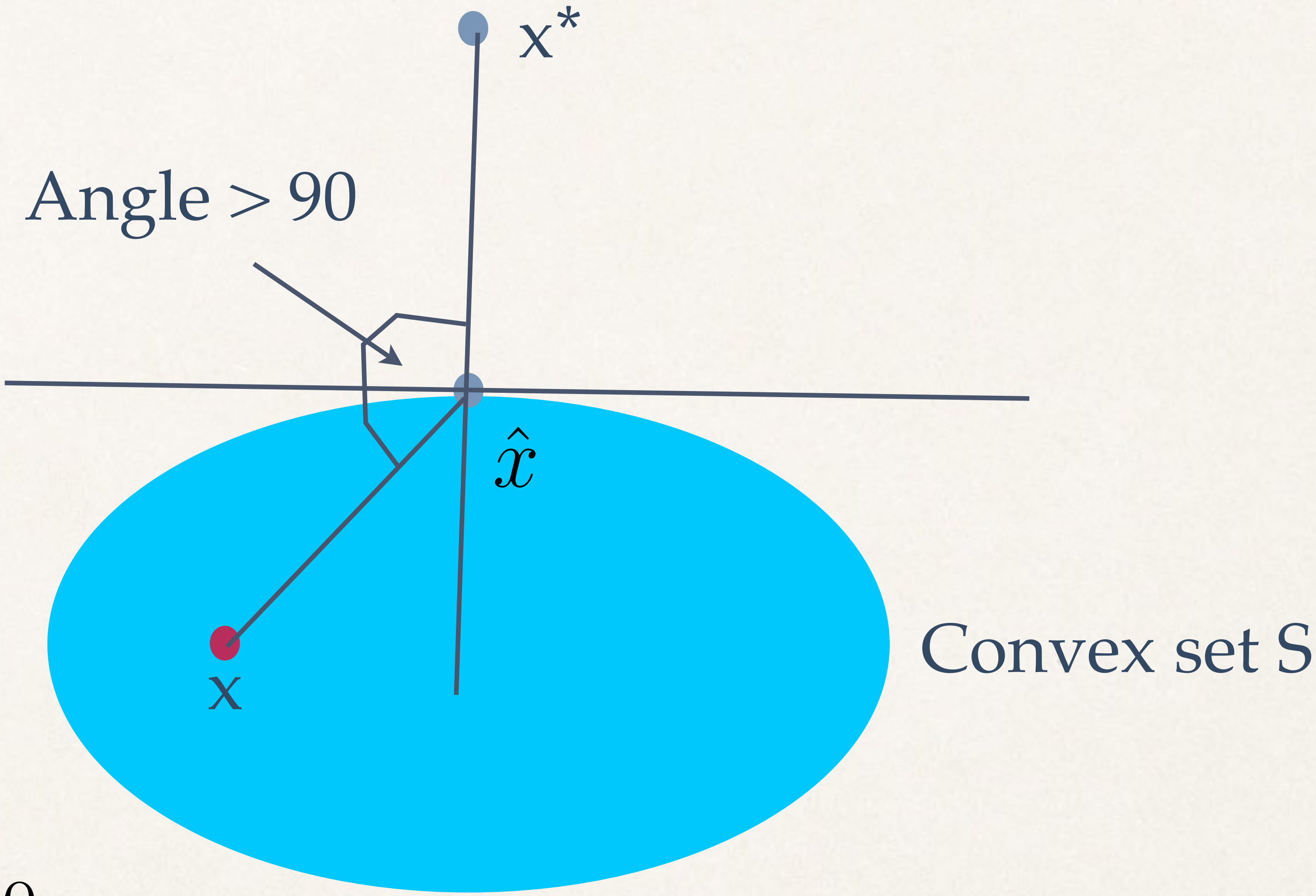
$$x^* = \Pi_C(x) = \text{argmin}_{u \in C} \|u - x\|_2^2$$

Geometric Hahn-Banach Theorem

Hyperplanes separate points from convex sets



Separating Hyperplane Theorem



$$(x^* - \hat{x})'(x - \hat{x}) \leq 0$$

Proof of Geometric HB Theorem

Consider the optimization problem

$$\hat{x} = \text{argmin}_{x \in S} \|x - x^*\|$$

Since S is a closed bounded set, according to Weierstrass' theorem, this optimization problem is well-defined. Let x be any point in S. Since S is convex,

$$(1-\lambda)\hat{x} + \lambda x \in S$$

It follows that

$$\begin{aligned} \|\hat{x} - x^*\|^2 &\leq \|\hat{x} + \lambda(x - \hat{x}) - x^*\|^2 \\ &= \|\hat{x} - x^*\|^2 + 2\lambda(\hat{x} - x^*)'(x - \hat{x}) + \lambda^2\|x - \hat{x}\|^2 \end{aligned}$$

which implies

$$2\lambda(\hat{x} - x^*)'(x - \hat{x}) + \lambda^2\|x - \hat{x}\|^2 \geq 0$$

Consequently

$$(\hat{x} - x^*)'(x - \hat{x}) \geq 0 \Rightarrow (x^* - \hat{x})'(x - \hat{x}) \leq 0$$

Proximal Mapping

- The proximal mapping of a convex function is defined as

$$\mathrm{prox}_h(x) = \mathrm{argmin}_u \left(h(u) + \frac{1}{2}\|u - x\|_2^2 \right)$$

- Examples:

$$h(x) = 0, \mathrm{prox}_h(x) = x$$

$$h(x) = I_C(x), \mathrm{prox}_h(x) = P_C(x) = \mathrm{argmin}_{u \in C}\|u - x\|_2^2$$

Prox Operator

Foundations and Trends® in Optimization
Vol. 1, No. 3 (2013) 123–231

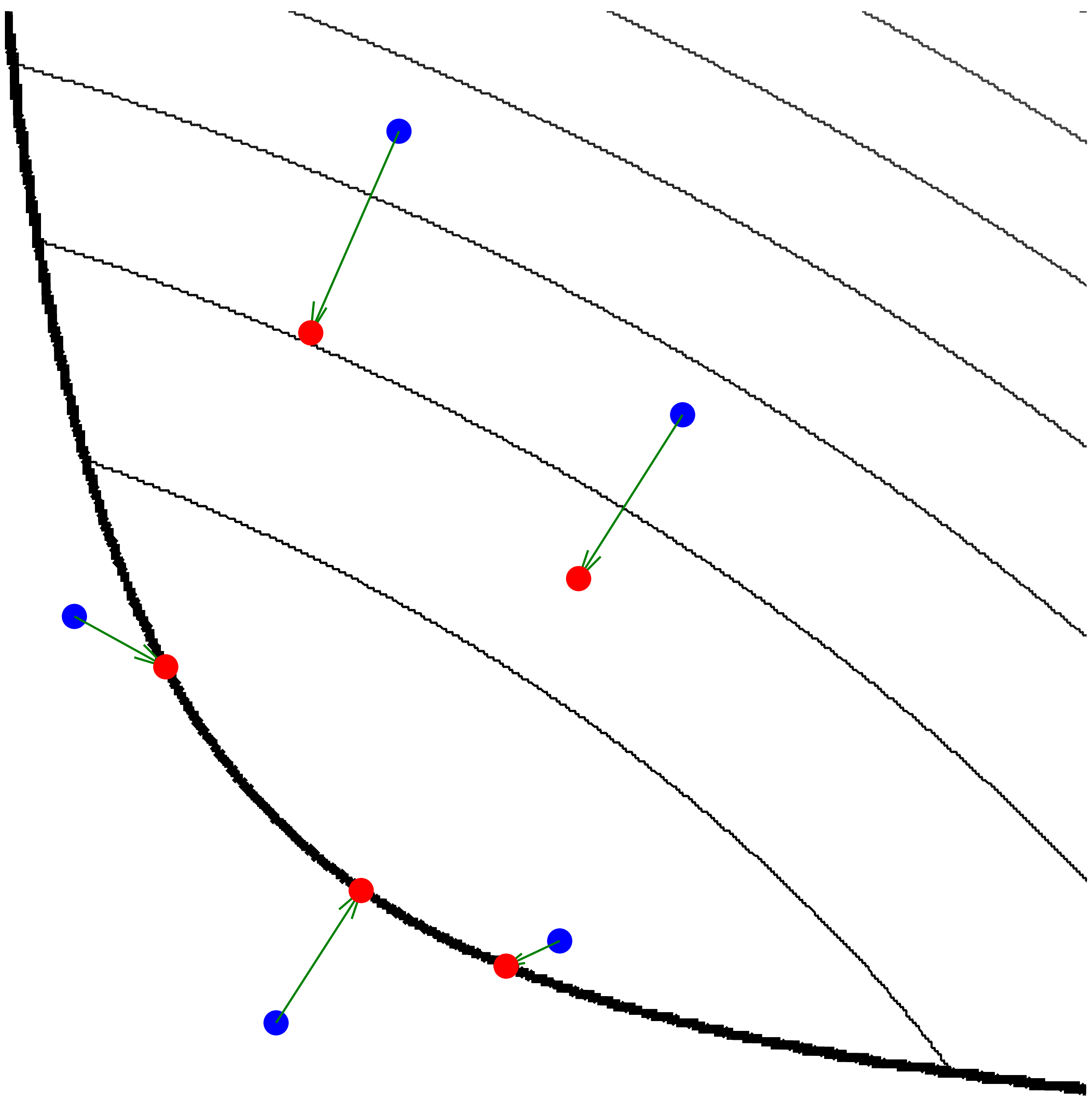
DOI: xxx

now
the essence of knowledge

Proximal Algorithms

Neal Parikh
Department of Computer Science
Stanford University
npparikh@cs.stanford.edu

Stephen Boyd
Department of Electrical Engineering
Stanford University
boyd@stanford.edu



Gradient Descent as proximal mapping

Answer:

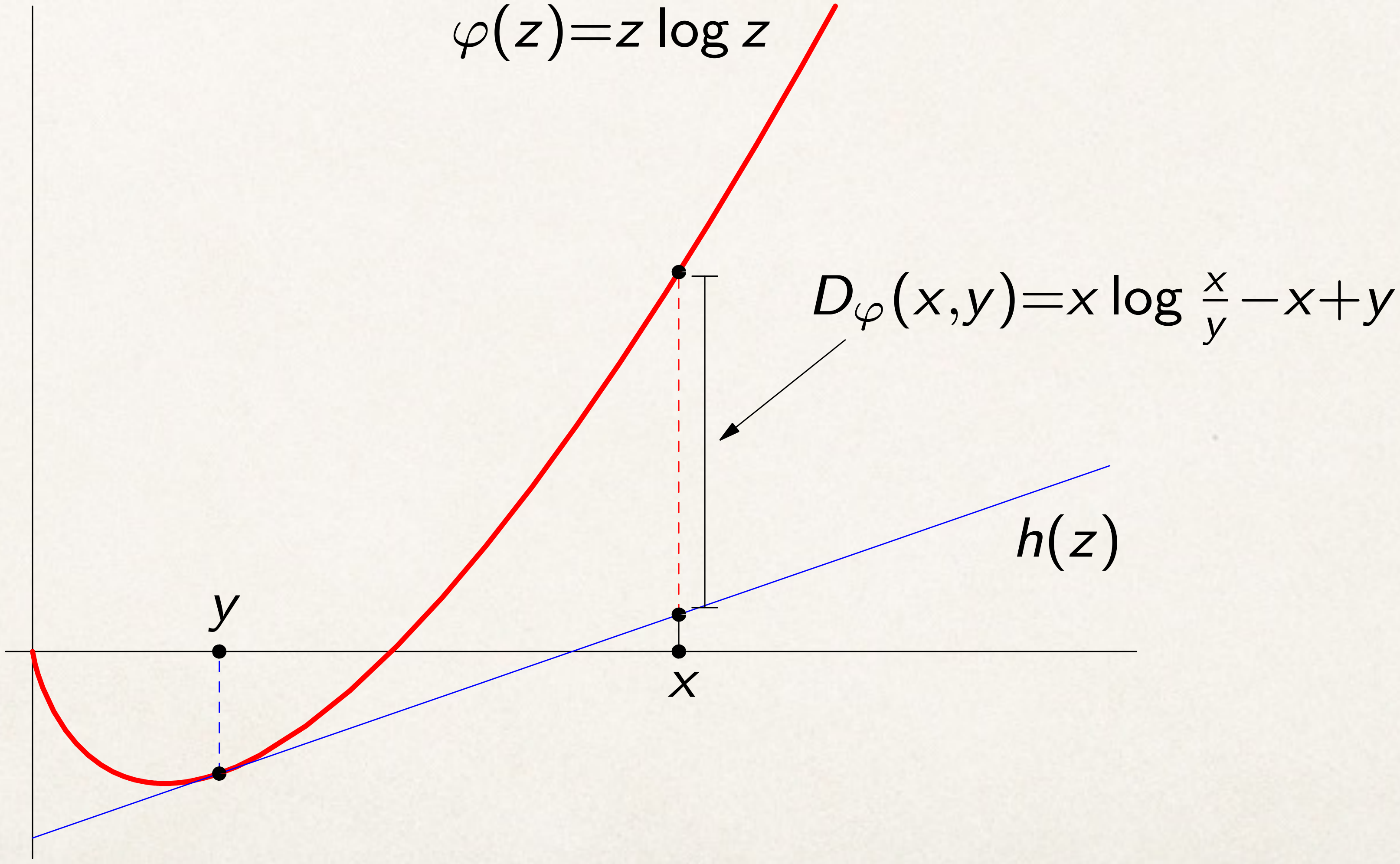
$$w_{t+1} \leftarrow w_t - \alpha_t \nabla f(w_t)$$

Question?

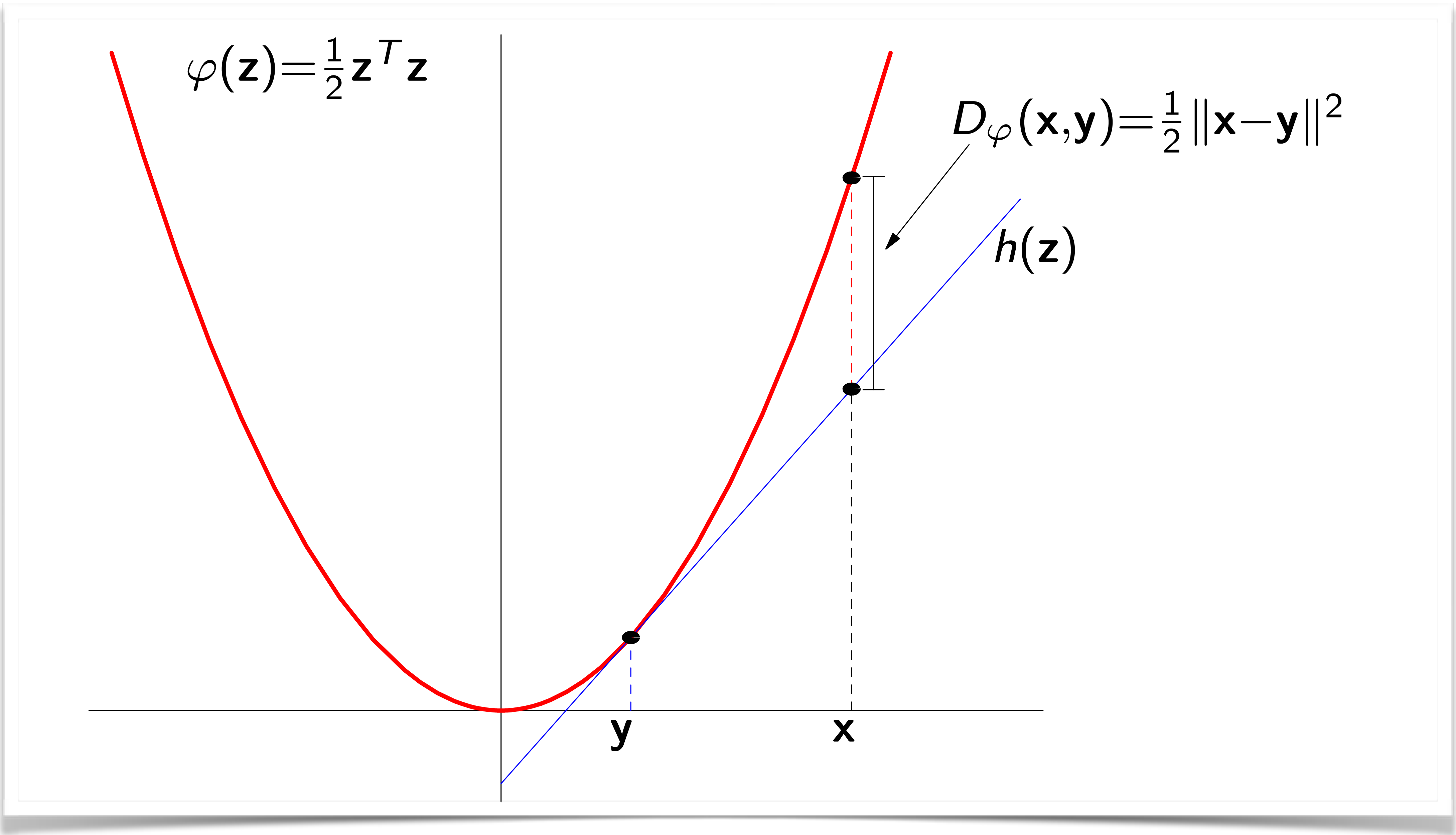
$$w_{k+1} = \min_u(\langle \nabla f(w_k), u \rangle + \frac{1}{2\alpha} \|u - w_k\|^2)$$

Bregman Divergence

$$D_\varphi(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - (\mathbf{x} - \mathbf{y})^T \nabla\varphi(\mathbf{y})$$



Euclidean Distance



Euclidean Distance

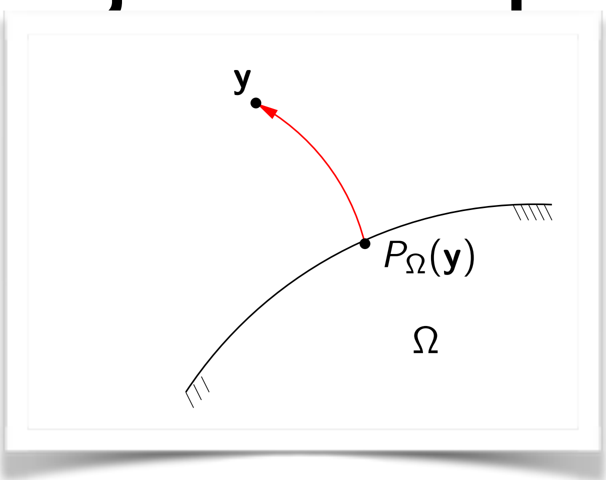
$$D_\phi(x, y) = \frac{1}{2}x^T x - \frac{1}{2}y^T y - y^T(x - y)$$

$$D_\phi(x, y) = \frac{1}{2}x^T x + \frac{1}{2}y^T y - y^T x$$

$$D_\phi(x, y) = \frac{1}{2}\|x - y\|_2^2$$

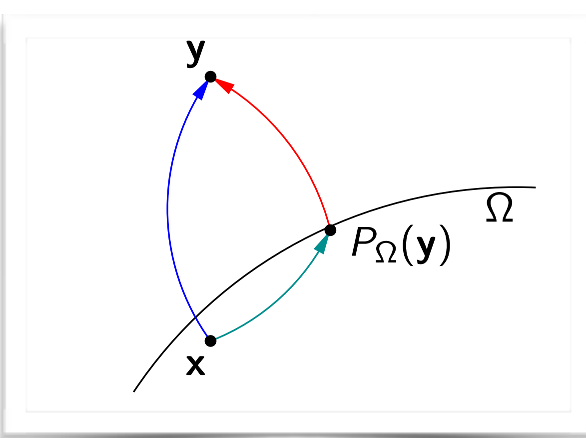
Generalized Projections

- Bregman Divergence leads to a generalized projection operation



$$P_\Omega(y) = \text{argmin}_{w \in \Omega} D_\phi(w, y)$$

- Generalized Pythagorean theorem

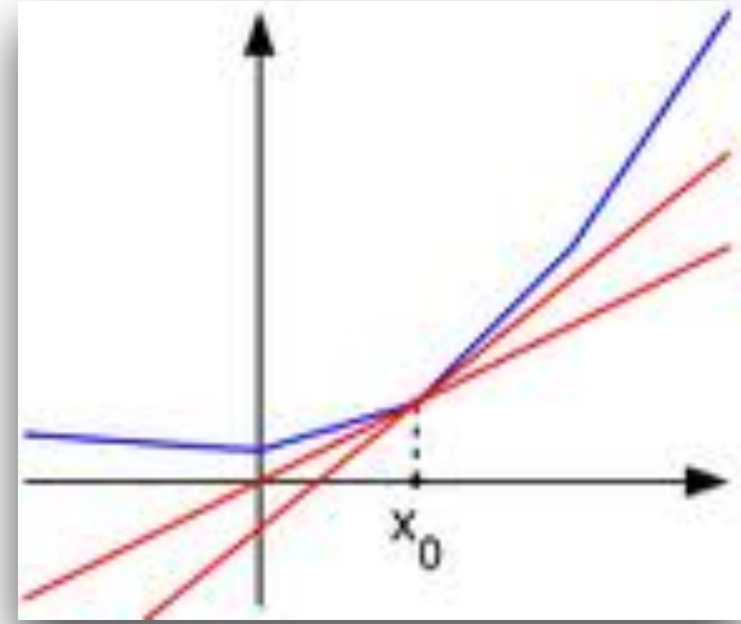


$$D_\phi(x, y) \geq D_\phi(x, P_\Omega(y)) + D_\phi(P_\Omega(y), y)$$

Subgradients and Subdifferentials

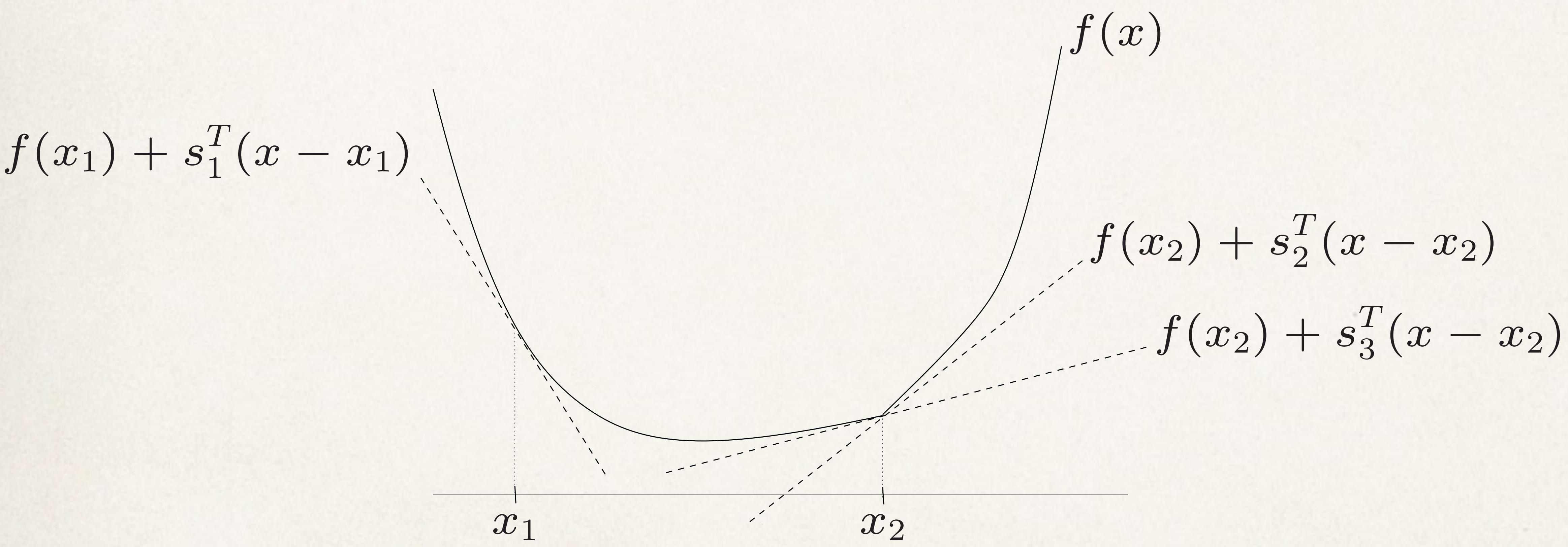
- Convex functions may not be differentiable
- Subgradient of a convex function:

$$f(y) \geq f(x) + \langle s, y - x \rangle$$

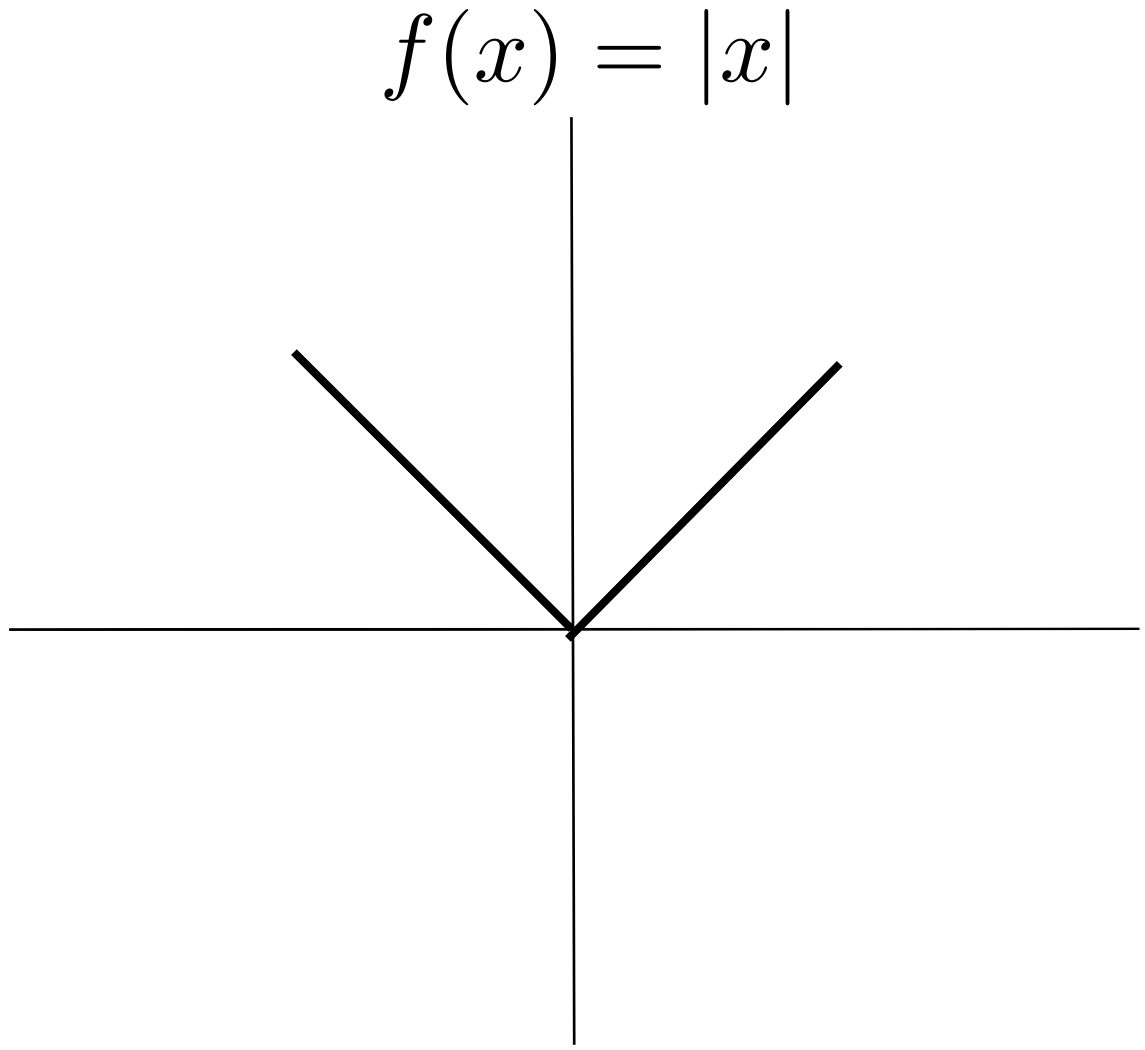


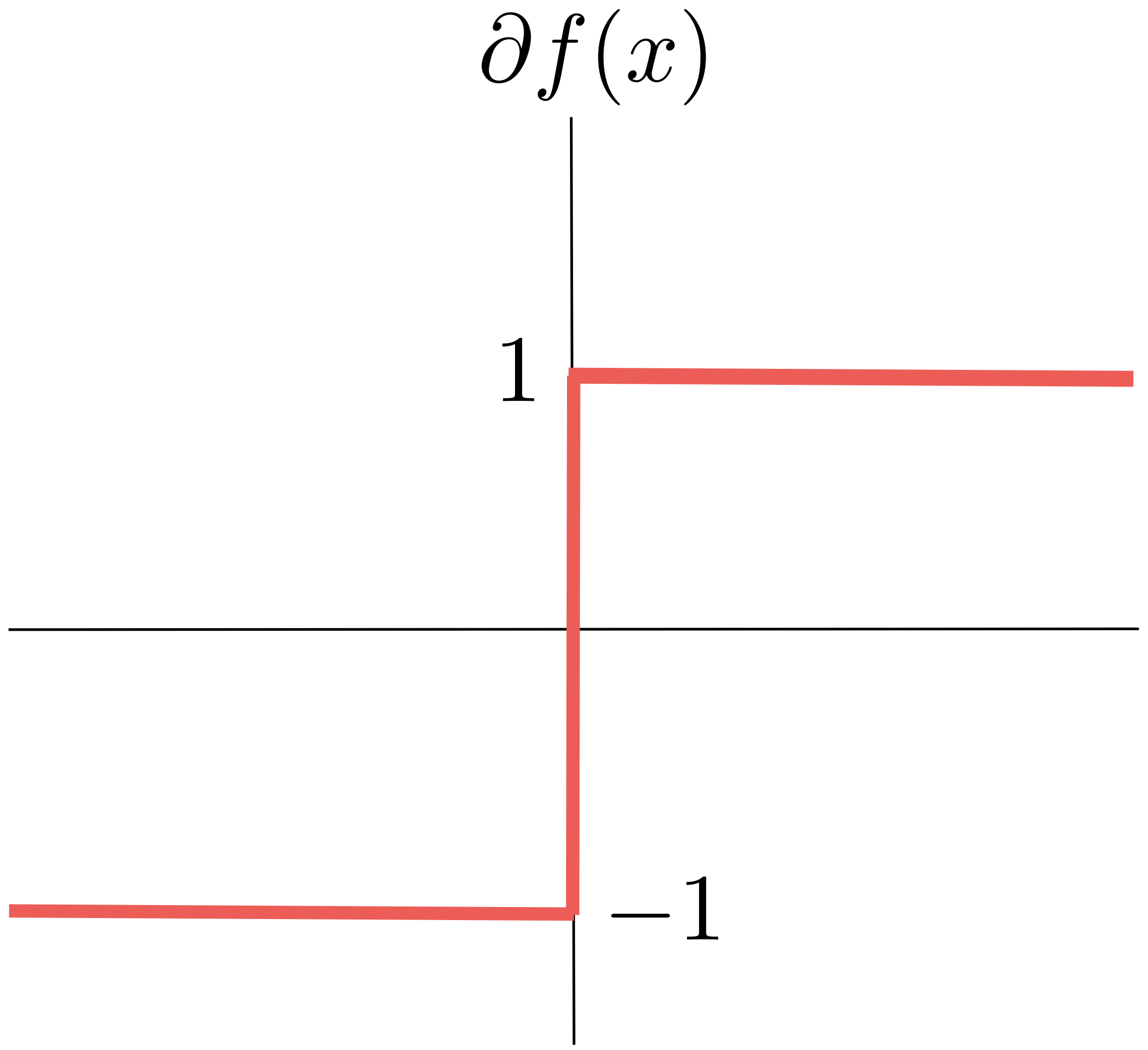
- Subdifferential: set of all subgradients

Subgradient of a function



Subdifferential of |x|

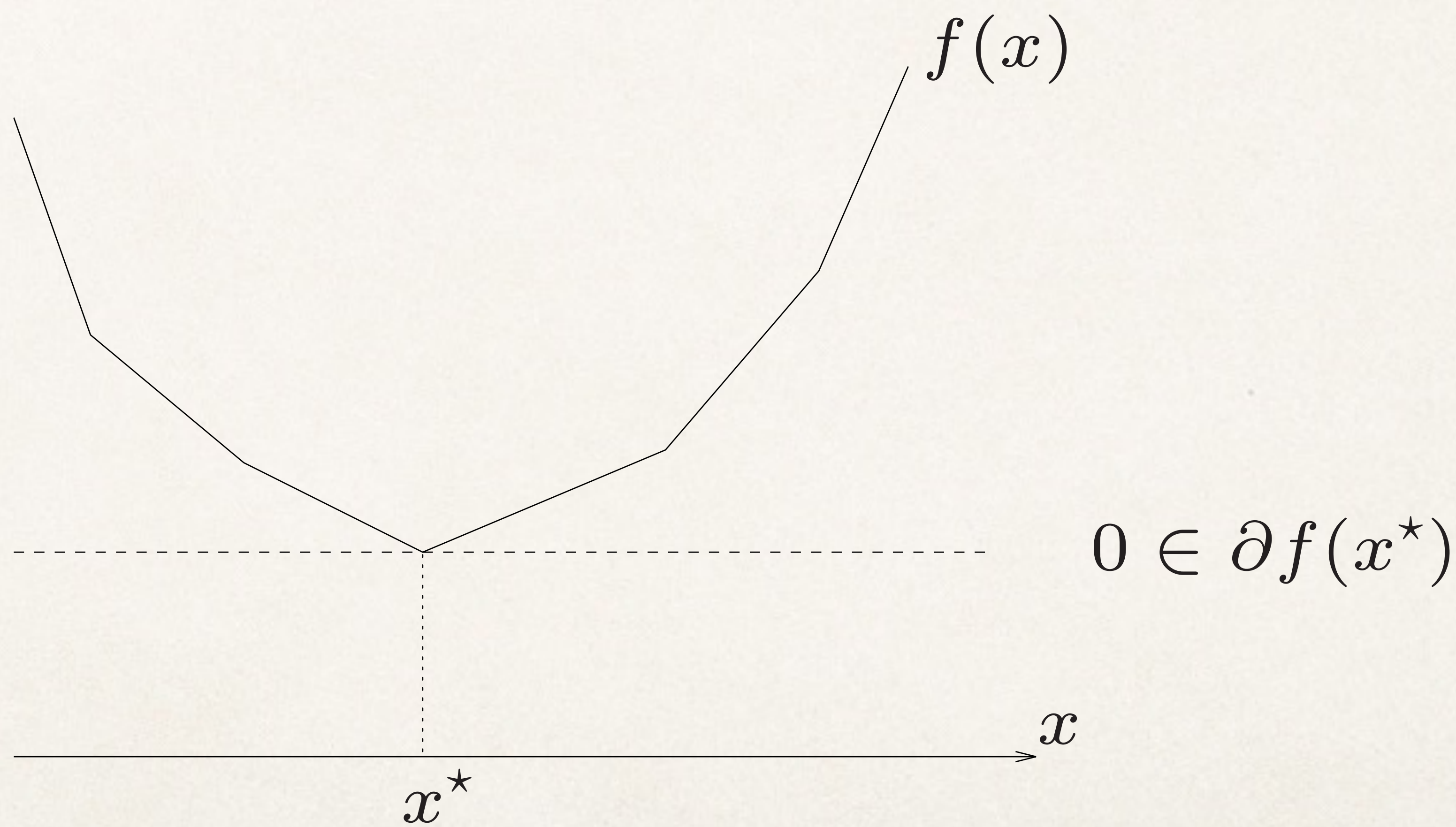




$$f(x) = |x| \geq f(0) + v^t(x - 0) \Rightarrow |v| \leq 1$$

Optimality Conditions

$$0 \in \partial f(x^\star)$$



$$f(y) \geq f(x^*) + 0^T(y - x^*)$$

Euclidean Norm

$$f(x) = \|x\|_2$$

$$\partial f(x) = \frac{1}{\|x\|_2} x \text{ if } x \neq 0$$

$$\partial f(x) = \{g \mid \|g\|_2 \leq 1\} \text{ if } x = 0$$

Projections and Proximal Mapping

proximal mapping of indicator function I_C is Euclidean projection on C

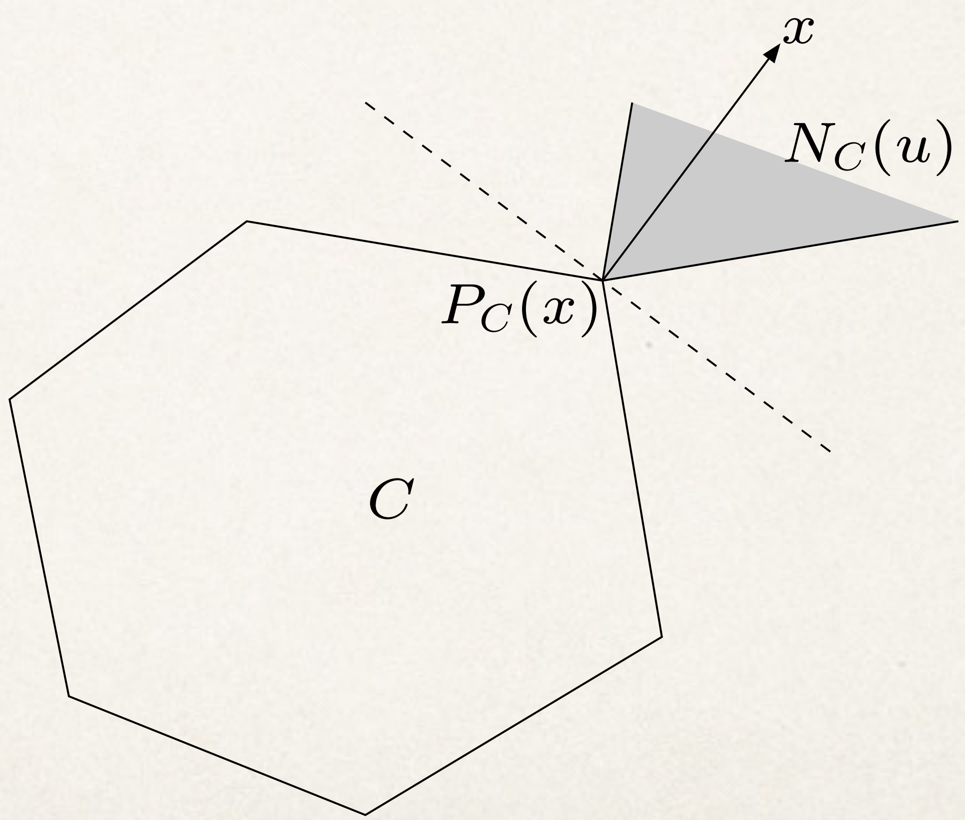
$$\mathbf{prox}_{I_C}(x) = \underset{u \in C}{\operatorname{argmin}} \|u - x\|_2^2 = P_C(x)$$

subgradient characterization

$$u = P_C(x)$$

$$\Updownarrow$$

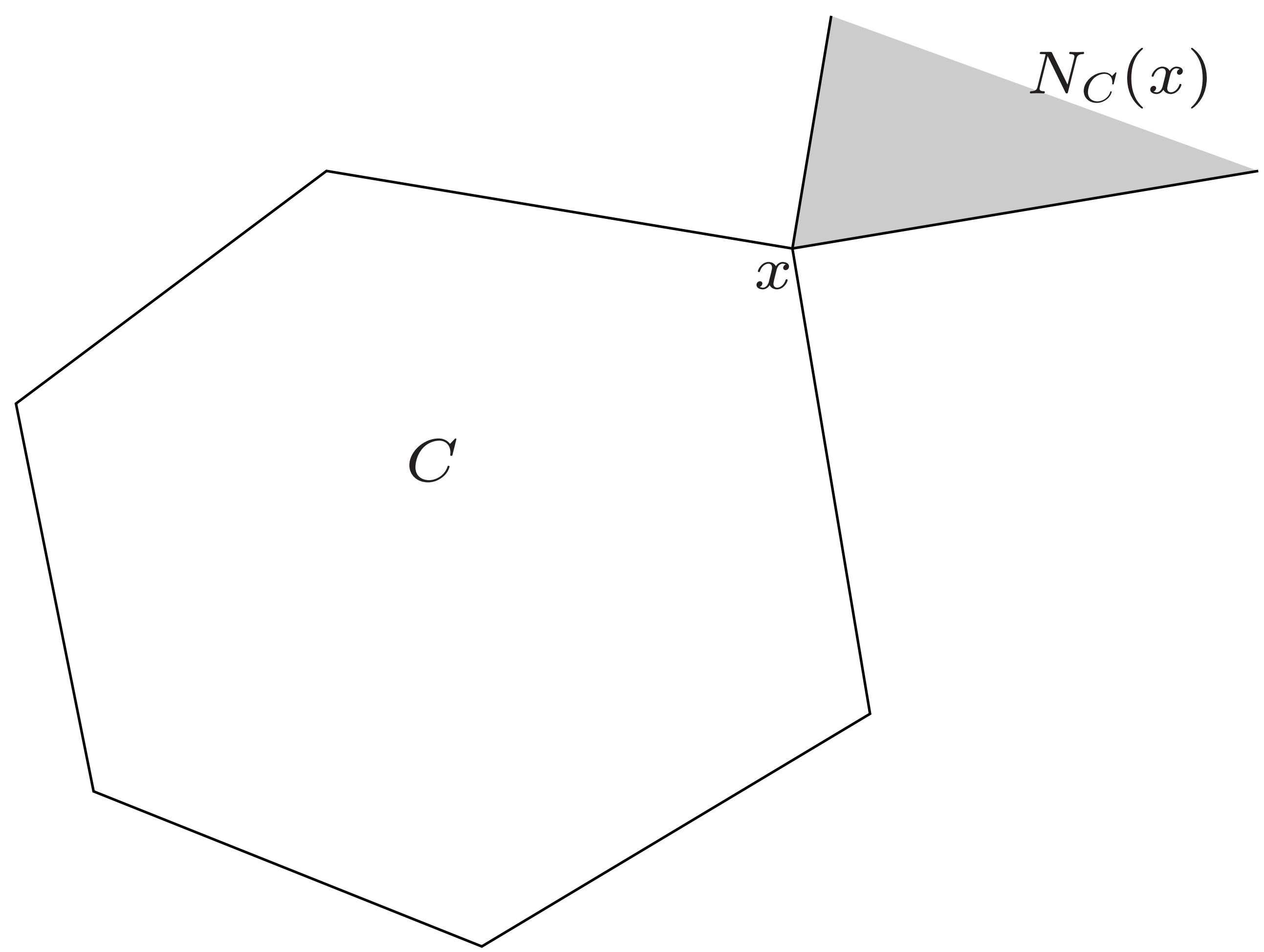
$$(x-u)^T(z-u) \leq 0 \quad \forall z \in C$$



Indicator Function

$$\partial I_C(x) = \{s \mid s^T(y-x) \le 0, \text{ for all } y \in C\}$$

this is known as the *normal cone* to C at x (notation: $N_C(x)$)



Monotonicity of Subdifferentials

if $s \in \partial f(x)$ and $\hat{s} \in \partial f(\hat{x})$, then

$$(\hat{s} - s)^T (x - \hat{x}) \geq 0$$

this property is called *monotonicity* of the (multivalued) mapping ∂f

proof: add left and righthand sides of the two inequalities

$$f(x) \geq f(\hat{x}) + \hat{s}^T (x - \hat{x})$$
$$f(\hat{x}) \geq f(x) + s^T (\hat{x} - \hat{x})$$

Proximal Mappings

if h is convex and closed, then

$$\mathbf{prox}_h(x) = \underset{u}{\operatorname{argmin}} \left(h(u) + \frac{1}{2}\|u - x\|_2^2 \right)$$

exists and is unique for all x

subgradient characterization

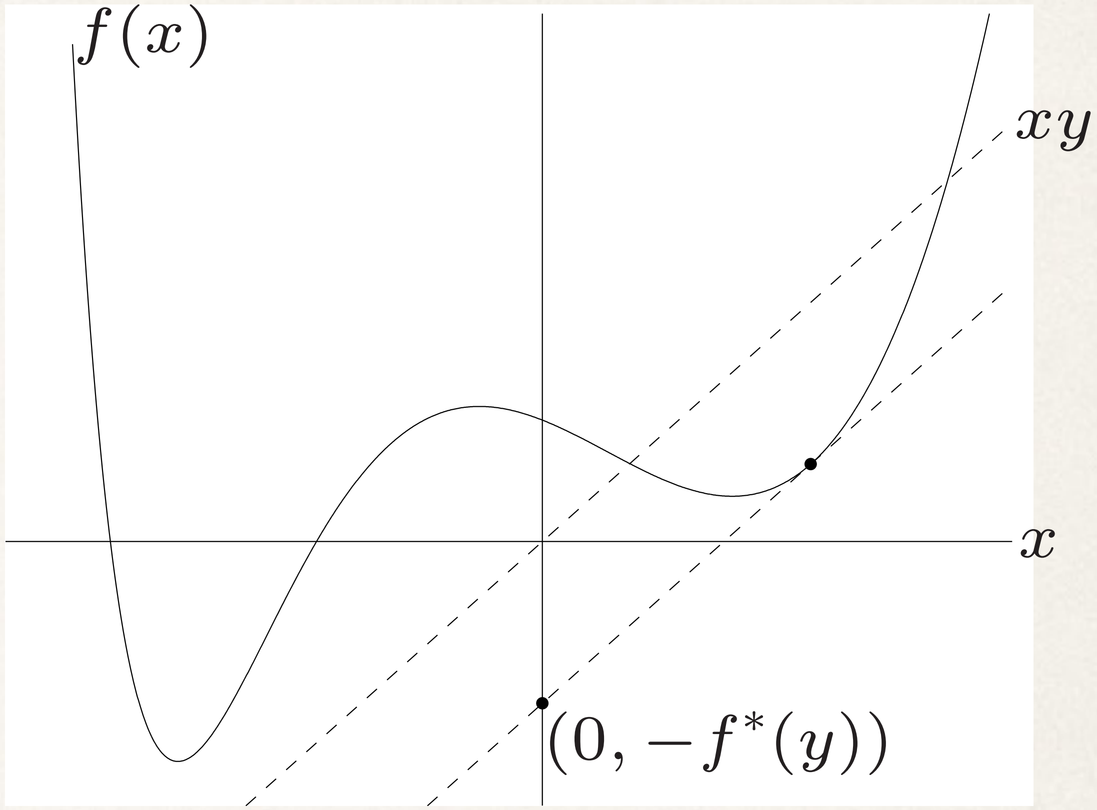
from optimality conditions of minimization in the definition:

$$\begin{aligned} u = \mathbf{prox}_h(x) \quad &\Longleftrightarrow \quad x - u \in \partial h(u) \\ &\Longleftrightarrow \quad h(z) \geq h(u) + (x-u)^T(z-u) \quad \forall z \end{aligned}$$

The conjugate function

the **conjugate** of a function f is

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$$

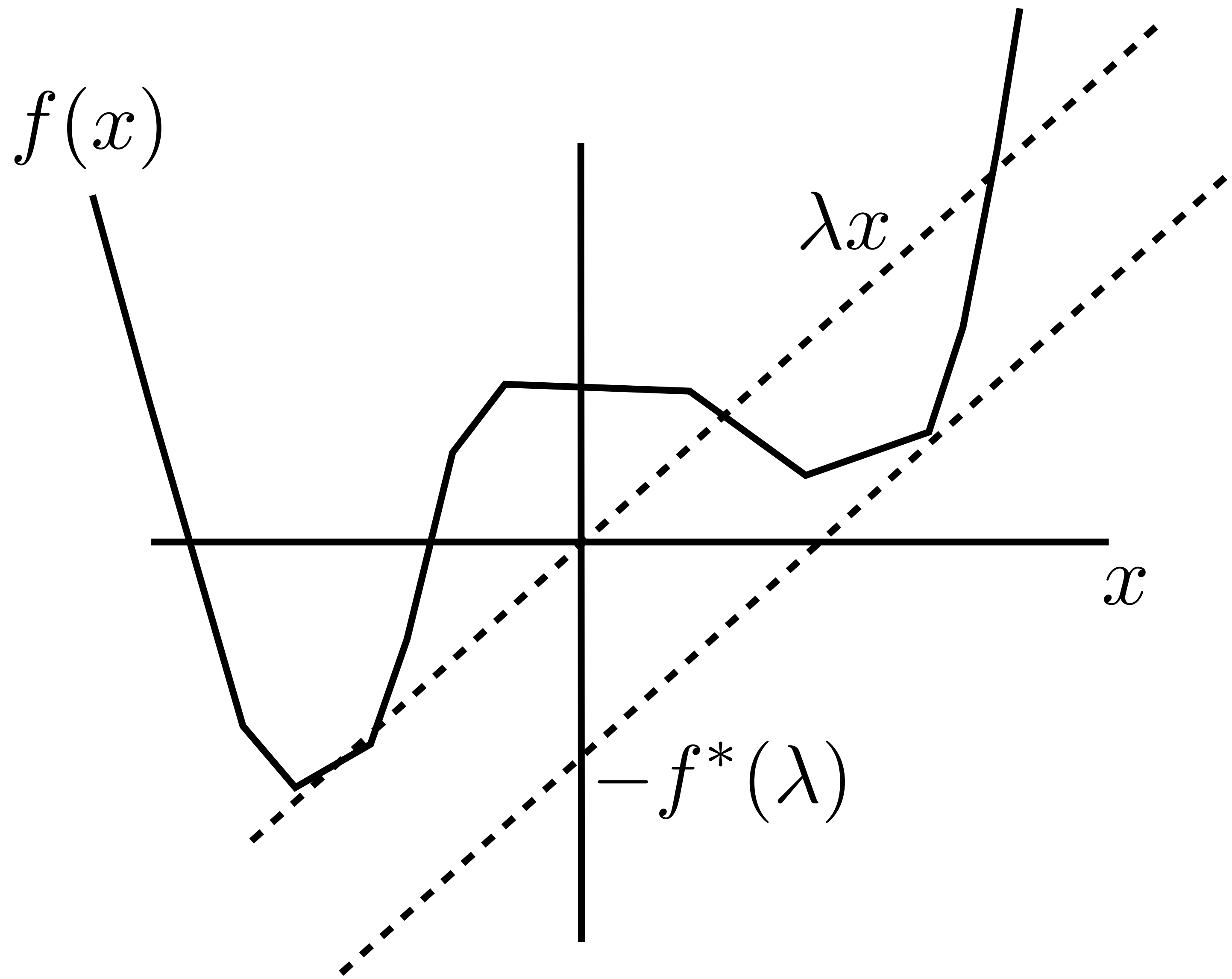


f^* is closed and convex (even if f is not)

Fenchel's inequality

$$f(x) + f^*(y) \geq x^T y \quad \forall x, y$$

Example



$$\boxed{f^*(\lambda) = \sup_x(\langle x, \lambda \rangle - f(x))}$$

Examples

negative logarithm $f(x) = -\log x$

$$\begin{aligned} f^*(y) &= \sup_{x>0} (xy + \log x) \\ &= \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

quadratic function $f(x) = (1/2)x^T Q x$ with $Q \in \mathbf{S}^n_{++}$

$$\begin{aligned} f^*(y) &= \sup_x (y^T x - (1/2)x^T Q x) \\ &= \frac{1}{2} y^T Q^{-1} y \end{aligned}$$

Examples

indicator function

$$I_C^*(y) = \sup \left(y^T x - I_C(x)\right) = \sup_{x \in C} y^T x$$

this is known as the *support function* of C

norm $f(x) = \|x\|$

$$f^*(y) = \sup_x \left(y^T x - \|x\|\right) = \begin{cases} 0 & \|y\|_* \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

$i.e.$, the indicator function of the norm ball of the *dual norm*

$$\|y\|_* = \sup_{\|x\| \leq 1} y^T x$$

Conjugate Function

- A core concept in convex analysis is the notion of conjugate functions

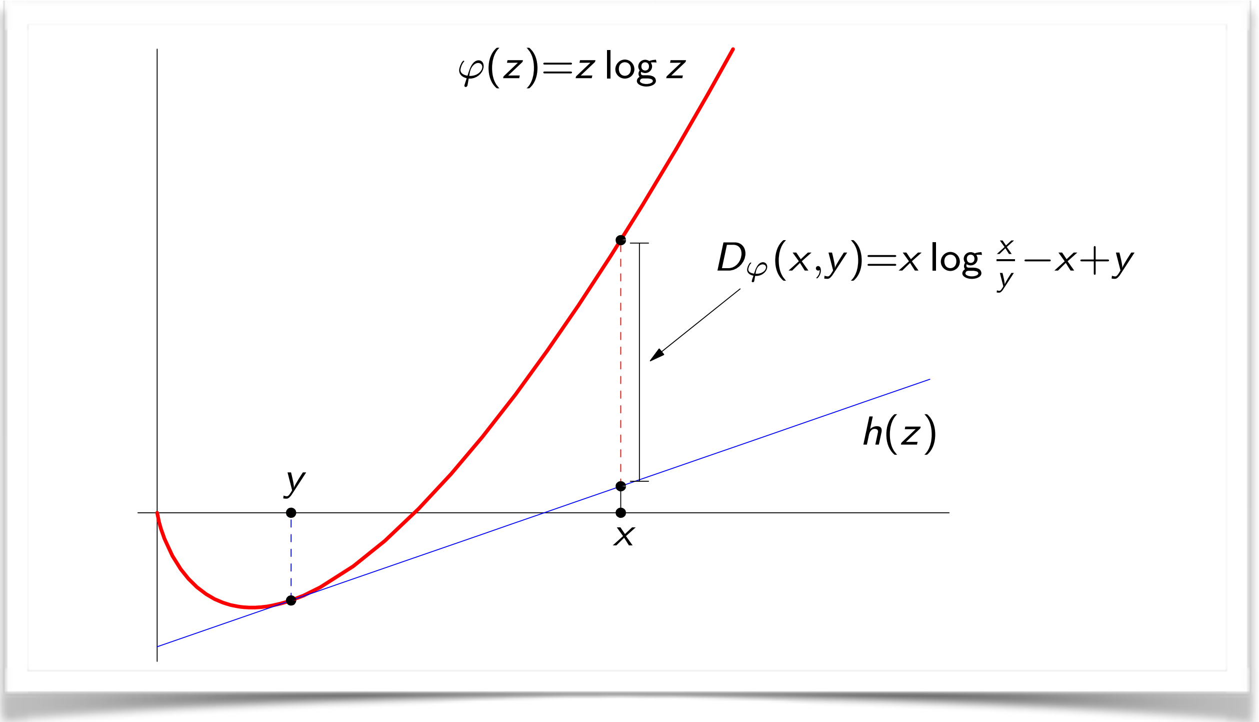
$$f^*(\lambda) = \sup_x(\langle x, \lambda \rangle - f(x))$$

- The conjugate function is always convex, even if the original function is not

- Legendre Transform: conjugate of differentiable function $\lambda = \nabla f(x)$

Conjugate Functions

$$\phi^*(z) = \log \sum_i e^{z_i}$$



Subgradient of Conjugate Function

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (x^T y - f(x))$$

weak subgradient rule

if $\hat{x}$ maximizes $x^T \hat{y} - f(x)$ over $x \in \mathbf{dom}\, f$, then $\hat{x} \in \partial f^*(\hat{y})$

$$\begin{aligned} f^*(y) = \sup_{x \in \mathbf{dom}\, f} (x^T y - f(x)) \quad &\geq \quad \hat{x}^T y - f(\hat{x}) \\ &= \quad \hat{x}^T \hat{y} - f(\hat{x}) + \hat{x}^T (y - \hat{y}) \\ &= \quad f^*(\hat{y}) + \hat{x}^T (y - \hat{y}) \end{aligned}$$

Quiz 1 released Thursday

- Covers Lectures 1 through 5
- True/false questions
- Simple numerical problems
- Review chapters 1 through 3 of Boyd