CSE6243: Advanced Machine Learning **Fall 2024**

Lecture 2: Optimization: convex set and function

Lecturer: Bo Dai *Scribes: Jack Fratto, Brandon Ho*

Note: *LaTeX template courtesy of UC Berkeley EECS Department.*

2.1 Recap

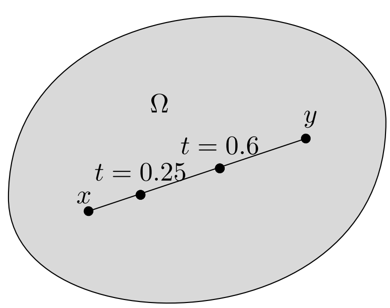
1. optimization
2. convex optimization
   (a) local optimum are also global optimum
   (b) the standard form for an optimization problem is defined as:

$$\begin{aligned} &\min \quad f(x) \\ &x \in \Omega \\ &\text{s.t.} \quad g(x) \geq 0 \end{aligned} \tag{2.1}$$

2.2 New Content

2.2.1 Convex Sets

We say a set is convex if the line segment between any two points x and $y \in \Omega$ lies within Ω for any $0 \leq t \leq 1$:



Convex set

$$\begin{aligned} &x, y \in \Omega \\ &t \in [0, 1] \\ &tx + (1-t)y \in \Omega \end{aligned} \tag{2.2}$$

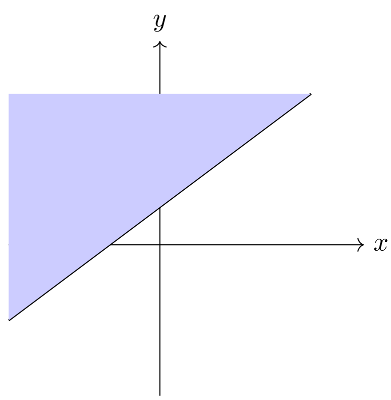
Example - l_2 convex ball

l_2 - Ball where $B(r) \subseteq \mathbb{R}^d$, such that the magnitude of any vector x is less than r. formally: $\{x : ||x||_2 \leq r\}$:
Then any two points x, y within the convex set of B can be defined as

$$\begin{aligned} & t \in [0,1] \\ & x \in B(r) \\ & y \in B(r) \\ & ||tx + (1-t)y||_2 \leq r \\ & ||tx||_2 + (1-t)||y||_2 \leq r \end{aligned} \tag{2.3}$$

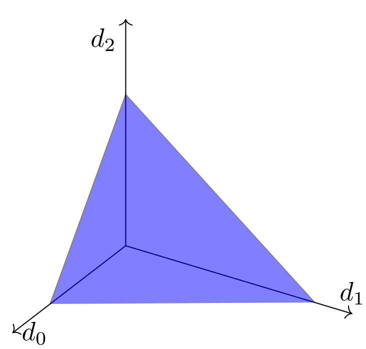
Example - Halfspace in $d = 2$

$$\{Ax + b \leq 0\} \in \mathbb{R}^d \tag{2.4}$$



Example - simplex in $d = 3$

$$\Delta = \{P : P \geq 0, \sum_{i=1}^{d} P_i = 1\} \tag{2.5}$$



Operations which preserve convexity

1. Intersection

$$\Omega = \Omega_1 \cap \Omega_2 \tag{2.6}$$

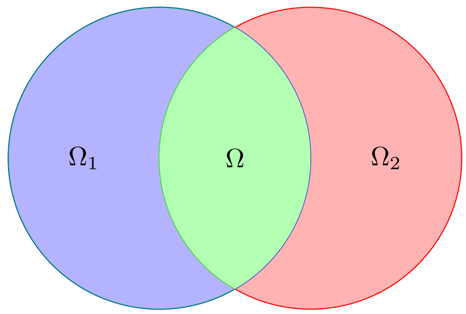


Figure 2.2: Intersection of two convex sets (circles). The intersection is also convex.

2. Affine (i.e. Translate, Rotate, Scale, Shear, Mirror)

$$\begin{aligned} C &= A\Omega + b \\ &= \{Ax + b, x \in \Omega\} \end{aligned} \tag{2.7}$$

Any time you have a convex set and apply a linear operation to the set, the result remains a convex set

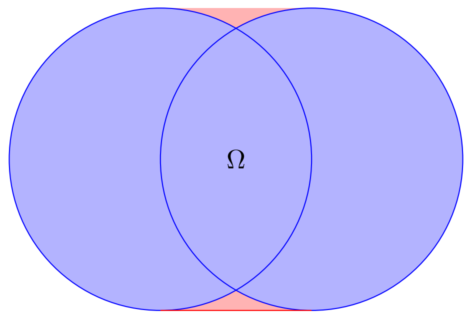
3. Linear Fraction

$$\begin{aligned} C = &\left\{ f(x) = \frac{A^T x + b}{Cx + c_0}, x \in \Omega \right\} \\ &\text{s.t.} \quad Cx + c_0 > 0 \\ &A \in \mathbb{R}^{d \times p}, x \in \mathbb{R}^d, b \in \mathbb{R}^{p \times 1} \\ &Ax + b \in \mathbb{R}^{p \times 1} \\ &C \in \mathbb{R}, c_0 \in \mathbb{R} \end{aligned} \tag{2.8}$$

Operations which DO NOT preserve convexity

1. Union

$$\Omega = \Omega_1 \cup \Omega_2 \tag{2.9}$$



$x \notin \Omega$

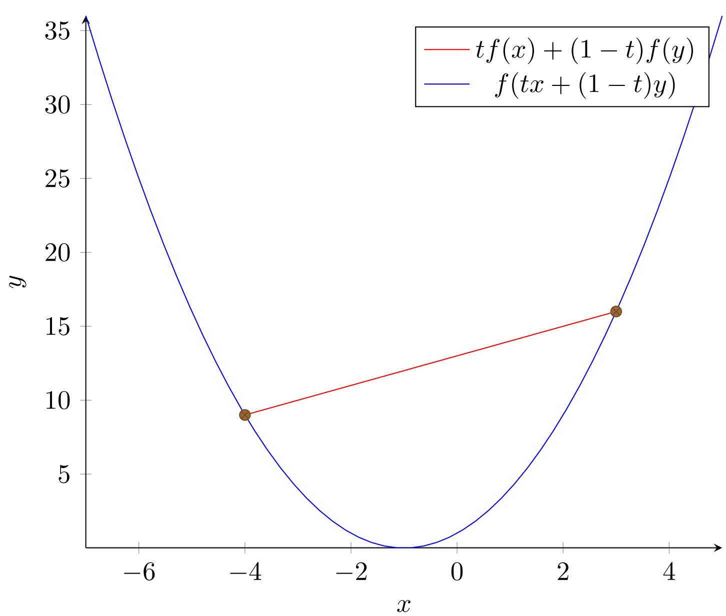
Figure 2.3: Union of two convex sets (circles). The union is **not** necessarily convex as shown in red

2.2.2 Convex Functions

Definition (Zeroth Order)

Given a domain $\Omega \in \mathbb{R}^d$, a function f is convex if and only if $\forall x, y \in \Omega$,

$$(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{2.10}$$



Theorem: First Order Condition

Given a domain $\Omega \in \mathbb{R}^d$, a function f is convex if and only if $\forall x, y \in \Omega$,

$$f(x) \geq f(y) + \nabla f(y)^T (x - y) \tag{2.11}$$

where ∇f is the gradient of f

Proof

if ($\Rightarrow$)
suppose

$$x_0 = \lambda x + (1 - \lambda) y \tag{2.12}$$

and we have

$$\begin{cases} f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0) \\ f(y) \geq f(x_0) + \nabla f(x_0)^T (y - x_0) \end{cases} \tag{2.13}$$

then summing the two equations in (2.13) and multiplying by $(1 - \lambda)$:

$$\lambda f(x) + (1 - \lambda) f(y) \geq \cancelto{f(x_0)}{\lambda f(x_0) + (1 - \lambda) f(x_0)} + \cancelto{0}{\nabla f(x_0)^T (\lambda(x - x_0) + (1 - \lambda)(y - x_0))} \tag{2.14}$$

Then we're left with:

$$f(x_0) \leq \lambda f(x) + (1 - \lambda) f(y) \tag{2.15}$$

only if ($\Leftarrow$)
Dividing the equation from (2.10) we have:

$$\frac{f((1 - t)y + tx)}{t} \leq f(x) - f(y) + \frac{f(y)}{t} \tag{2.16}$$

which reduces to:

$$f(x) \geq f(y) + \frac{f((1 - t)y + tx) - f(y)}{t} \tag{2.17}$$

as we take the limit as t approaches 0, this becomes:

$$f(x) \geq f(y) + \nabla f(y)^T (x - y) \tag{2.18}$$

□

Theorem: Second Order Condition

Given a domain $\Omega \in \mathbb{R}^d$, a function f is convex if and only if $\forall x \in \Omega$,

$$\nabla^2 f(x) \geq 0$$

or $\forall h, x \in \Omega$

$$h^T \nabla^2 f(x) h \geq 0$$

where $\nabla^2 f$ is the hessian of f

Proof

if (⇒)
We know by definition that any matrix $x^T A x$ is positive semi-definite, then:

$$f(x+h) = f(x) + h^T \nabla f(x) + \underbrace{\frac{1}{2} h^T \nabla^2 f(z) h}_{\geq 0} \tag{2.19}$$

then by definition:

$$f(x+h) \geq f(x) + h^T \nabla f(x) \tag{2.20}$$

End of Lecture