2. Basic Mathematical Analysis and Differential Calculus

2.1 Supremum & Infimum vs Maximum & Minimum

This week starts with some Analysis and Differential Calculus knowledge required to go through Optimization.

Not a lot of knowledge of mathematical analysis and differential calculus is needed.

Later actually, a concept of *Lipschitz Continuous*, *Smoothness* are needed later in Algorithms, but too far ahead.

✓ Bounded Above

Let a set $A \subset R$ be bounded above if $\exists b \in R$ such that $\forall a \in A, a \leq b$.

b is called an *upper bound* of A.

✓ Least Upper Bound = Supremum

Let $s \in R$ be the supremum of $A \in R$ if

1) s is an upper bound of A

2) If u is any upper bound of A, $s \leq u$.

✓ Maximum

A real number m is a maximum of a set A if

1) m is an upper bound of A

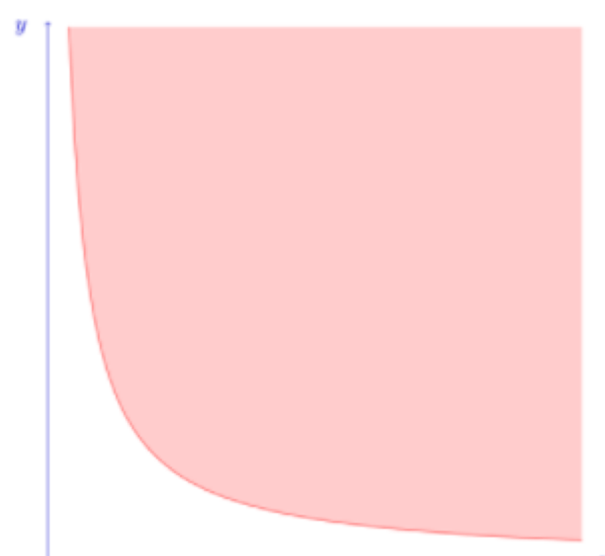
2) $m \in A$

- Why important in This course?

Ex) $\min_{x,y \in R} x \ s.t.$

$$\begin{pmatrix} x & 1 \\ 1 & y \end{pmatrix} \succeq 0$$

Thus, use *inf* instead of *min*!



2.2 Basic Topology

✓ Metric Space

A function $d : X \times X \to \mathbb{R}$ is a metric on the set X if for all $x, y, z \in X$,

1) For $x \neq y$, $d(x, y) > 0$ and $d(x, x) = 0$: Non-negativity

"distance between distinct points is positive"

2) $d(x, y) = d(y, x)$: Symmetry

"direction doesn't matter"

3) $d(x, z) \leq d(x, y) + d(y, z)$: Triangle Inequality

"Adding a midpoint increases the distance"

Then, call (X, d) a *metric space*.

✓ Epsilon Ball

Let (M, d) be a metric space.

The epsilon-ball centered at $a \in M$ "$B_\epsilon(a)$ is set of points distance *less* than ϵ from a where $\epsilon > 0$.

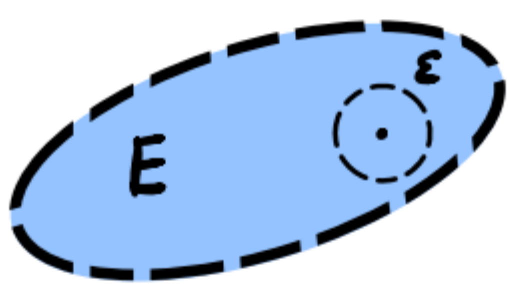
$B_\epsilon(a)$ = {$x|d(x, a) < \epsilon$}.

Without any mention about the metric, think of "d" as the familiar l_2 norm. In $\mathbb{R}$, it is simply $|x - a|$.

✓ Open Set

O is an Open set if $\forall o \in O$, $\exists\epsilon > 0$ s.t. $B_\epsilon(o) \subseteq O$

ex) Open, bounded set in $\mathbb{R}$ works like (a, b).



Take any element x in (a, b). Then, I can "choose" ϵ as $\frac{\min(x-a, b-x)}{2}$ using the density of $\mathbb{R}$.

✓ Limit Point

A point x is a limit point of a set F if every epsilon ball about x intersects F at some point other than x.

ex) $F = [a, b)$ then the set of limit points of F is $[a, b]$.

✓ Closed

F is closed if it contains all of its limit points.

ex) Closed, bounded set in $\mathbb{R}$ works like $[a, b]$.

$*$ Practice $*$

ϕ is *Open and Closed*. $\mathbb{R}$ is open and closed.

$[a, b)$ is neither open nor closed.

✓ Closure

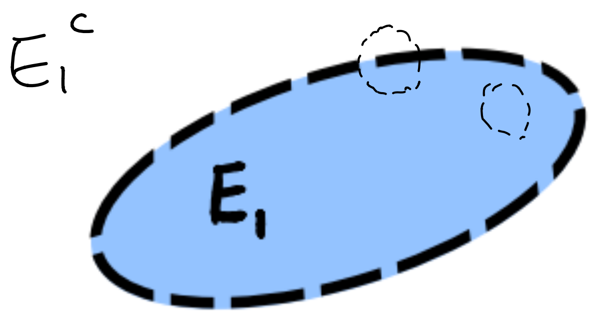
Let E be a proper, nonempty subset of $\mathbb{R}$ and L be the set of all limit points of E. The closure of E,

$\bar{E}$ = $E \cup L$.

ex) $E = [a, b)$ then $\bar{E} = [a, b]$

✓ Boundary

Let E be a proper, nonempty subset of $\mathbb{R}$. $\partial(E)$, the boundary of E is set of points "a" where $\forall B_\epsilon(a)$ contains points in E and E^c.



✓ Interior

The interior of a set E, $E^o := \{x \in E | \exists B_\epsilon(x) \subseteq E\}$

- Why important in This course?

The concept of *Relative Interior* is used in *Slater's condition*, a theory to satisfy $d^* = p^*$, the strong duality.

2.3 Differential Calculus

✓ Differentiation in 1 dimensional space

f is differentiable at $a \in \mathbb{R}$ if f is defined on some open set U containing a and $\frac{d}{dx} f(x)|_{x=a}$ = $f'(a)$ = $lim_{h \to 0} \frac{f(a+h)-f(a)}{h}$

$\frac{f(a+h)-f(a)}{h}$ is called *average rate of change* and $f'(a)$ is called *instantaneous rate of change*.

✓ Gradient of Mutivariable functions

The gradient of a function $f : R^n \to R$ at a point x where f is differentiable is defined as a column vector containing **first order partial derivative** of f w.r.t. $x_1, x_2, \ldots, x_n$.

$\nabla f(x)$ = $[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}]^T$

ex) $h(x) = 2x_1 + 3x_2 + x_3$ then $\nabla f(x) = [2, 5, 1]^T$

✓ Hessian of Mutivariable functions

$(H_{f(x)})_{ij}$ = $\frac{\partial^2 f}{\partial x_i \partial x_j}$. Note that $(H_{f(x)}) \in R^{n \times n}$ and *symmetric*!

ex) $f(x) = 4x_1^2 + 5x_1x_2 + x_2^2$ then find the Hessian of f

✓ Jacobian of *vector valued* functions

$f : R^n \to R^m$. a Jacobian of f is a matrix collecting all the first order derivatives. Jacobian of $f(x) = [f_1(x) \ldots f_m(x)]^T$

$(J_{f(x)})_{ij} = \frac{\partial f_i}{\partial x_j}$

ex) $f(x) = \begin{pmatrix} 3x_1x_2 + 2x_1 \\ 4x_2 \\ 8x_1^2 \end{pmatrix}$. Find the Jacobian of f.

✓ Taylor Series Expansion of $f : R \to R$.

Let $f \in C^\infty(a, b)$ and $x_0 \in (a, b)$. The taylor series of f centered at x_0 :

$T(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$

ex) $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, $x \in \mathbb{R}$

✓ Second Order Approximation of $R^n \to R$ at x.

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T H_{f(x)} (y-x)$$

Vector Derivative Practice)

$\min_{x \in \mathbb{R}^n} \|Ax-b\|_2^2$ s.t. $Gx=h$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $G \in \mathbb{R}^{p \times n}$, $h \in \mathbb{R}^p$, $\text{rank}(A)=n$

" p equality constraints "

$$\begin{aligned}
\mathcal{L}(x, \gamma) &= \|Ax-b\|_2^2 + \gamma^T (Gx-h) \\
&= (Ax-b)^T(Ax-b) + \gamma^T(Gx-h) \\
&= (x^TA^T - b^T)(Ax-b) + \gamma^T(Gx-h) \\
&= x^TA^TAx + (\gamma^TG - 2b^TA)x - \gamma^Th + b^Tb \\
&= x^TA^TAx + (G^T\gamma - 2A^Tb)^T x - \gamma^Th + b^Tb
\end{aligned}$$

Using convexity of the Lagrangian w.r.t. x,

$$\nabla_x \mathcal{L}(x, \gamma) = 2A^TAx + (G^T\gamma - 2A^Tb)\, x \underset{\text{let}}{=} 0$$

$$x^* = (A^TA)^{-1}\left(A^Tb + G^T\left(G(A^TA)^{-1}G^T\right)^{-1}\left(h - G(A^TA)^{-1}A^Tb\right)\right).$$