INTRODUCTION TO NONLINEAR OPTIMIZATION

An optimization problem in continuous variables in which one or more of constraint or objective functions is not linear is called a nonlinear program.

If there's no constraint function, then it's called unconstrained nonlinear program. If the constraint functions are linear, then it's called linearly constrained nonlinear program.

Some features of non-linear programs:

- The presence of at least one nonlinear function
- One or more variables (all of them are continuous)
- Inequality constraints, equality constraints, or no constraints (For LP, you'd always have constraints.)
- Properties of functions (convexity, differentiability ...)
- Somehow complicated optimality criteria
- Convergent (but usually not finite) solution algorithms (In LP we have finite algorithms.)

General form of a nonlinear program:

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m \\
& h_j(x) = 0, \quad j = 1, \dots, p \\
& x \in X
\end{aligned}
$$

Clearly, one can consider this as a maximization problem as we have

$$\min_{x \in S} f(x) = -\max_{x \in S} (-f(x)).$$

Most of the time X is taken as $\mathbb{R}^n$ and not written in the formulation.

If $i=0, j=0$ $\Rightarrow$ Unconstrained non-linear opt.

If $i=0, j\neq 0$ $\Rightarrow$ Equality constrained nonlin. opt.

If $i\neq 0, j=0$ $\Rightarrow$ Inequality " " "

If $i\neq 0, j\neq 0$ $\Rightarrow$ mixed-constrained non lin. opt.

Examples:

1- Linear least squares problem (linear regression)

- to fit data to a function that is linear in the model parameters

Suppose that for a given set of inputs, a process is run several times, say m, and for each run i, the output is recorded:

Run	Input vector	Output
i	$(a_{i1}, \dots, a_{in})$	b_i

Assume a linear function is used to relate the parameters with the output:

$$b_i = a_{i1}x_1 + \dots + a_{in}x_n \quad \text{for } i=1,\dots,m.$$

In practice we have $m \gg n$. The problem is to determine the parameters x_j that best fit the data specified by an overdetermined system of equations.

For $i = 1, \ldots, m$, define

$$\varepsilon_i = b_i - (a_{i1}x_1 + \cdots + a_{in}x_n)$$

The linear least-squares problem is then

$$\text{minimize} \sum_{i=1}^{m} \varepsilon_i^2$$

The problem can be written in matrix form. Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$

$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Then, we have

$$\sum_{i=1}^{m} \varepsilon_i^2 = \varepsilon^T \varepsilon = (b - Ax)^T (b - Ax) = b^T b - 2b^T A x + x^T A^T A x,$$

which is a quadratic function of n variables.

This clearly is an example of unconstrained nonlinear program.

2- Location Problems

A simple example is as follows:

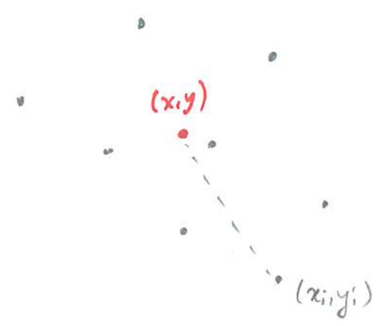
Assume a set of n "costumers" have known coordinates (x_i, y_i) for $i = 1, \ldots, n$.

Then, we seek a point in $(x, y) \in \mathbb{R}^2$ as the location of the "service facility"

such that the sum of the distances from (x, y) to each of (x_i, y_i) is minimum.

Using the Eucledean distance, the problem is:

$$\text{minimize} \sum_{i=1}^{n} \sqrt{(x - x_i)^2 + (y - y_i)^2}$$



There are many variations of location problems

- There could be a positive weight for each consumer

$$\min. \sum_{i=1}^{n} w_i \sqrt{(x-x_i)^2 + (y-y_i)^2}$$

If a "consumer" has priority or more important than the rest then w_i is larger.

- One can use different distance measures

e.g. Manhattan distance $\sum_{i=1}^{n} \{ |(x-x_i)| + |y-y_i| \}$

- Restrictions on admissible locations

e.g. certain regions on the plane are prohibited

- Higher dimensional problems.

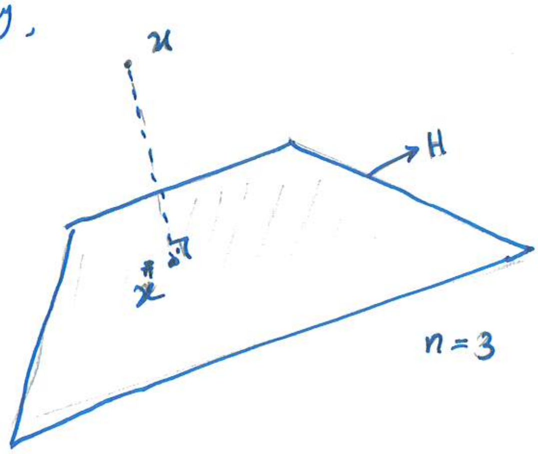
e.g. applications from electronics where given points may lie in 3-dim. space.

3- Projection problem

Suppose H is a hyperplane in $\mathbb{R}^n$ (solutions set of a linear equation $a^T x = b$ for $a \in \mathbb{R}^n \setminus \{0\}$, $b \in \mathbb{R}$). Let $u \in \mathbb{R}^n$ be a given point in $\mathbb{R}^n$ but not in H.

One seeks the point x^* on H that is closest (Euclidean sense) to u.

In lower dimensions this can be solved geometrically, in general it's more difficult. Hence, we solve the optimization problem given as



$$\text{minimize} \quad \|x-u\|^2$$
$$\text{subject to} \quad a^T x = b$$

There exists a unique optimal solution to this problem and we will see methods of finding this analytical solution

$$x^* = u + a\left(\frac{b - a^T u}{a^T a}\right).$$

4- Portfolio selection problem:

Assume that an investor can invest in n assets (stocks, bonds, etc.) He/she want's to decide on the percentage x_j of the total amount to be invested in the j^{th} security, $j=1,\dots,n$. Hence we need

$$e^T x = 1, \quad x \geq 0$$

to hold for $e = (1,\dots,1)^T \in \mathbb{R}^n$. Here, x is called a portfolio. The aim is to maximize the expected return, while to minimize the risk, which is taken as the variance of the return here. (Markowitz, 1952).

Note that expected return is

$$\mathbb{E}(x) = \sum_{j=1}^{n} r_j x_j = r^T x$$

where r_j is the expected return on the j^{th} security. (Assumed to be known or estimated). The variance of the return is given by a quadratic form

$$Var(x) = x^T D x$$

where D is the symmetric covariance matrix (again assumed to be known or estimated.).

Note that it's not possible to maximize the expected return and to minimize the risk simultaneously as they would be conflicting in general.

One way is to form an objective function in the following form:

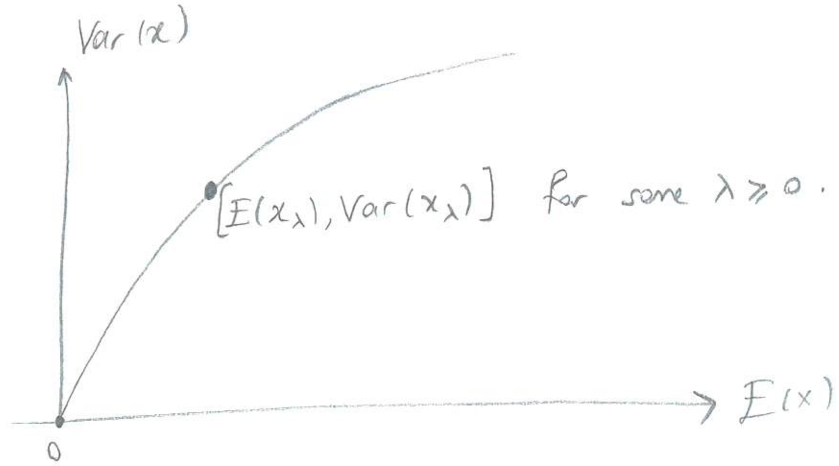
$$f_\lambda(x) = -\lambda E(x) + Var(x) = -\lambda r^T x + x^T D x$$

Here $\lambda \geq 0$ is a parameter to be fixed by the investor. Note that $\lambda = 0$ would yield the variance and as $\lambda \to \infty$ the dominating term becomes the expected return.

The problem is to

$$\begin{aligned} &\text{minimize} \quad f_\lambda(x) \\ &\text{subject to} \quad e^T x = 1 \\ &\qquad\qquad\quad x \geq 0. \end{aligned}$$

Each $\lambda \geq 0$ would yield a (possibly) different solution x_λ and x_λ is said to be an "efficient" portfolio for each λ. If one can solve the problem for each λ, then all the solutions constitute the "efficient frontier".



Mathematical Preliminaries

Notation:

$x \in \mathbb{R}^n$ is considered to be a column vector $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ througout the course.

$e_i \in \mathbb{R}^n$ is the unit vector where the ith component is 1, $i = 1, \dots, n$.

$e \in \mathbb{R}^n$ is the vector of ones.

$\mathbb{R}^n_+ = \{ x \in \mathbb{R}^n \mid x_i \geq 0, \ i = 1, \dots, n \}$

$\mathbb{R}^n_{++} = \{ x \in \mathbb{R}^n \mid x_i > 0, \ i = 1, \dots, n \}$

$[x, y] = \{ x + \alpha (y - x) \mid \alpha \in [0, 1] \}$ - closed line segment between x & y.

$(x, y) = \{ x + \alpha (y - x) \mid \alpha \in (0, 1) \}$ - open line segment

I_n : identity matrix of size $n \times n$

$0_{m \times n}$: zero matrix of size $m \times n$

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i$$

Definition: A <u>norm</u> $\| \cdot \|$ on $\mathbb{R}^n$ is a function $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ satisfying

i) $\| x \| \geq 0$ for all $x \in \mathbb{R}^n$ and $\| x \| = 0$ if and only if $x = 0$.

ii.) $\| \lambda x \| = |\lambda| . \| x \|$ for any $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ (positive homogeneity)

iii) $\| x + y \| \leq \| x \| + \| y \|$ for any $x, y \in \mathbb{R}^n$. (triangle inequality)

Examples:

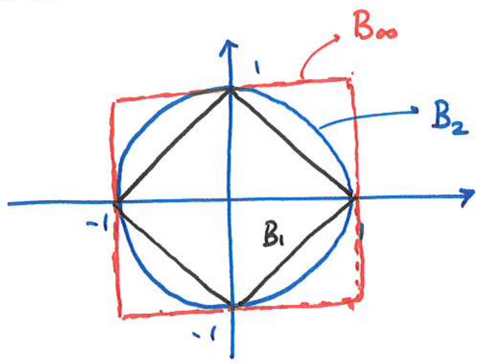
- Euclidean norm (ℓ_2-norm): $\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$

- ℓ_p-norm for $p \geq 1$: $\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$ (For $p \in (0,1)$ this is not a norm.)

$\|x\|_1 = \sum_{i=1}^{n} |x_i|$

- ℓ_∞-norm: $\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$ $\left(= \lim_{p\to\infty} \|x\|_p\right)$.

Let $B_p = \{ x \in \mathbb{R}^n \mid \|x\|_p \leq 1 \}$.



Cauchy-Schwarz Inequality

For any $x, y \in \mathbb{R}^n$ $|x^T y| \leq \|x\|_2 \|y\|_2$.

Equality is satisfied if and only if x and y are linearly dependent. ($x = k \cdot y$ for some $k \in \mathbb{R}$) or one of x, y is zero

There is also a concept of matrix norm:

Definition: A norm on $\mathbb{R}^{m\times n}$ is a function $\|\cdot\| : \mathbb{R}^{m \times n} \to \mathbb{R}$ satisfying

i) $\|A\| \geq 0$ for any $A \in \mathbb{R}^{m\times n}$ and $\|A\| = 0$ iff $A = 0 \in \mathbb{R}^{m \times n}$

ii) $\|\lambda A\| = |\lambda| \|A\|$ for any $A \in \mathbb{R}^{m\times n}$, $\lambda \in \mathbb{R}$

iii) $\|A+B\| \leq \|A\| + \|B\|$ for any $A, B \in \mathbb{R}^{m\times n}$

Example: Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ be a vector norm. Then

$\|A\| := \max_{\|x\| \leq 1} \|Ax\|$ defines a matrix norm.

If you start with ℓ_2-norm, then the induced norm is called "spectral norm":

$$\|A\|_2 = \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2 \leq 1} \|Ax\|_2$$

- An example which is not induced from a vector norm is the Frobenius norm:

$$\|A\|_F := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} \quad , \quad A \in \mathbb{R}^{m \times n}.$$

Linear Algebra Review

- Transpose of a matrix: $A_{m \times n} \Rightarrow A^T : n \times m$ matrix s.t. $(A^T)_{ij} = A_{ji}$

$(A+B)^T = A^T + B^T$

$(AB)^T = B^T A^T$

If $A = A^T \Rightarrow A$ is symmetric

(For $M \in \mathbb{R}^{m \times n}$, $M^T M \in \mathbb{R}^{n \times n}$ is symmetric.)

- Determinant of a square matrix:

(2×2) $\quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

(3×3) $\quad A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \Rightarrow \det(A) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$

Cofactor expansion $(n \times n)$: Pick any row or column, say we pick row i:

$$\det(A) = (-1)^{i+1} \cdot A_{i1} \cdot \det(\underbrace{A(i|1)}_{\substack{A \text{ without the} \\ i\text{th row \& first column}}}) + (-1)^{i+2} \cdot A_{i2} \cdot \det(A(i|2)) + \cdots$$

Properties: $\det A = \det A^T$

$\det A.B = \det A . \det B$

$\det A = 0 \iff A$ is singular (non-invertible)

$\det A^{-1} = \frac{1}{\det A}$

Inverse of a square matrix:

$A \in \mathbb{R}^{n\times n} \Rightarrow$ inverse is A^{-1} such that $A A^{-1} = A^{-1} A = I$ (identity)

Inverse of a matrix may not exist, if there's no inverse of it then the matrix is said to be singular.

Inverse of a 2×2 matrix: $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Properties: $(A^{-1})^{-1} = A$

$(A.B)^{-1} = B^{-1}.A^{-1}$ $\quad (A, B \in \mathbb{R}^{n\times n})$

$(A^T)^{-1} = (A^{-1})^T$

Trace of a square matrix:

$A \in \mathbb{R}^{n\times n} \Rightarrow \operatorname{tr} A = \sum_{i=1}^{n} A_{ii}$

Properties: $\operatorname{tr} A = \operatorname{tr} A^T$

$\operatorname{tr}(A+B) = \operatorname{tr} A + \operatorname{tr} B$ $\quad (B \in \mathbb{R}^{n\times n})$

$\operatorname{tr}(\underbrace{A.B}_{m\times m}) = \operatorname{tr}(\underbrace{B.A}_{n\times n})$ for $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times m}$

Also called "column space"

Range: let $A \in \mathbb{R}^{m\times n}$. Range of A is the set of all vectors that can be written as Ax for some $x \in \mathbb{R}^n$:

$$\mathcal{R}(A) \subseteq \mathbb{R}^m, \quad \mathcal{R}(A) = \{ y \in \mathbb{R}^m \mid y = Ax, \; x \in \mathbb{R}^n \}$$

- Columns of A are linearly independent if no column is in the range of the remaining columns.

In general, let $x^1, x^2, \ldots, x^k \in \mathbb{R}^n$ be vectors in $\mathbb{R}^n$. They are linearly independent if we have the following:

$$\lambda_1 x^1 + \lambda_2 x^2 + \ldots + \lambda_k x^k = 0 \in \mathbb{R}^n \text{ implies that } \lambda_1 = \lambda_2 = \ldots = \lambda_k = 0 \in \mathbb{R}.$$

Rank: Rank of $A \in \mathbb{R}^{m\times n}$ is the number of linearly independent columns

Properties: $\text{rank}(A) = \text{rank}(A^T)$

$\text{rank}(A) \leq \min(n, m)$

For $A \in \mathbb{R}^{n\times n}$, $\text{rank}(A) = n \iff \mathcal{R}(A) = \mathbb{R}^n \iff A$: non-singular

Orthogonality: $x, y \in \mathbb{R}^n$ are orthogonal if $x^T y = 0$. They are orthonormal if in addition $\|x\|_2 = \|y\|_2 = 1$.

- A matrix $U \in \mathbb{R}^{n\times n}$ is orthogonal if all of it's columns are orthonormal. This is true if and only if $U^T U = U U^T = I$.
- If U is orthogonal matrix, then its columns are linearly independent.

Nullspace: Let $A \in \mathbb{R}^{m \times n}$. Nullspace of A is set of all x s.t. $Ax = 0$,

$$\mathcal{N}(A) = \{ x \in \mathbb{R}^n \mid Ax = 0 \in \mathbb{R}^m \} \subseteq \mathbb{R}^n.$$

Clearly, $\mathcal{N}(A^T) = \{ y \in \mathbb{R}^m \mid A^T y = 0 \in \mathbb{R}^n \} \subseteq \mathbb{R}^m$

$\mathcal{R}(A)$ and $\mathcal{N}(A^T)$ are "orthogonal complements", that is

$$\mathcal{R}(A) \cup \mathcal{N}(A^T) = \mathbb{R}^m \quad \text{and} \quad \mathcal{R}(A) \cap \mathcal{N}(A^T) = \{0\}.$$

(Let $r \in \mathcal{R}(A)$, $n \in \mathcal{N}(A^T)$. We have $r^T n = 0$.)

Eigenvalues and Eigenvectors

For $A \in \mathbb{R}^{n \times n}$, a nonzero vector $v \in \mathbb{R}^n \setminus \{0\}$ is called an eigenvector of A if there exists a $\lambda \in \mathbb{C}$ such that

$$(*) \quad Av = \lambda v.$$

λ is called the eigenvalue corresponding to eigenvector v.

If A is a symmetric matrix, then all eigenvalues are real numbers.

Note that $(*)$ holds iff $(\lambda I - A) v = 0$. This equation have a solution whenever $\det(\lambda I - A) = 0$. (or $\det(A - \lambda I) = 0$)

This is called the characteristic equation of matrix A. One can compute the eigenvalues of A by solving the characteristic equation.

The eigenvalues of a symmetric $n \times n$ matrix A are denoted by

$$\lambda_{max}(A) = \lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A) = \lambda_{min}(A).$$

Spectral decomposition theorem: $A \in \mathbb{R}^{n \times n}$, symmetric.

There exists an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ and a diagonal matrix

$$D = \begin{bmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{bmatrix} = \operatorname{diag}(d_1, \ldots, d_n) \quad \text{such that}$$

$$U^T A U = D.$$

Moreover, the column vectors of ~~A~~ U are eigenvectors of A and d_i are the eigenvalues of A.

Note that trace is invariant under cyclic permutations, that is,

$$\operatorname{tr}(A.B.C) = \operatorname{tr}(B.C.A) = \operatorname{tr}(C.A.B) \quad (\neq \operatorname{tr}(ACB) \text{ in general}).$$

Then, $\operatorname{tr}(D) = \operatorname{tr}(U^T A U) = \operatorname{tr}(\underbrace{U U^T}_{I} A) = \operatorname{tr}(A)$ implies that

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i(A).$$

Similarly,
$$\begin{aligned} \det(D) = \det(U^T A U) &= \det U^T . \det A . \det U \\ &= \det U^T . \det U . \det A \\ &= \det(\underbrace{U^T U}_{I}) . \det A \quad = \det A \end{aligned}$$

implies that $\det(A) = \prod_{i=1}^{n} \lambda_i(A)$.

Basic Topological Concepts

Defn: $\bar{B}(c,r) = \{x \in \mathbb{R}^n \mid \|x-c\| \le r\}$, closed ball with center $c \in \mathbb{R}^n$, radius $r \in \mathbb{R}_{++}$

$B(c,r) = \{x \in \mathbb{R}^n \mid \|x-c\| < r\}$, open ball " " "

(Notation is different from the book!)

Defn: Given a set $U \subseteq \mathbb{R}^n$, a point $c \in U$ is an interior point of U if there exists $r > 0$ such that $B(c,r) \subseteq U$.

The set of all interior points is called the interior of U; $\text{int}(U)$.

e.g. $\text{int}\, \bar{B}(c,r) = B(c,r)$, $c \in \mathbb{R}^n$, $r > 0$.

$\text{int}\, \mathbb{R}^n_+ = \mathbb{R}^n_{++}$

Defn: An open set is a set that contains only interior points. That is, U is open if for every $x \in U$, $\exists\, r > 0$ s.t. $B(x,r) \subseteq U$.

Some facts: Union of any number of open sets is an open set.

Intersection of a finite number of open sets is an open set.

Intersection of infinitely many open sets may not be open: $\bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}$

Defn: A set $U \subseteq \mathbb{R}^n$ is closed if its complement $U^c = \mathbb{R}^n \setminus U$ is an open set.

Union of a finite number of closed sets is closed.

Intersection of any number of closed sets is closed.

X Union of infinitely many sets: $\bigcup_{n=1}^{\infty} \left[\frac{1}{n}, 2-\frac{1}{n}\right] = (0,2)$ not a closed set.

Alternative definition (equivalent) for closed sets:

$U \subseteq \mathbb{R}^n$ is a closed set if for every sequence of points $\{x_i\}_{i \geq 1}$ from U such that $x_i \to x^*$ as $i \to \infty$, it holds that $x^* \in U$.

Defn: <u>Closure</u> of a set $U \subseteq \mathbb{R}^n$ is the smallest closed set that contains U, denoted by $cl(U)$.

<u>Boundary</u> of a set $U \subseteq \mathbb{R}^n$ is

$$bd(U) := cl(U) \setminus int(U).$$

Defn: A set $U \subseteq \mathbb{R}^n$ is <u>bounded</u> if $\exists M > 0$ s.t. $U \subseteq B(0, M)$.
U is <u>compact</u> if it's closed and bounded.

e.g. $cl(B(c,r)) = \bar{B}(c,r)$, $cl\, \mathbb{R}^n_{++} = \mathbb{R}^n_+$

$bd(B(c,r)) = bd(\bar{B}(c,r)) = \{x \in \mathbb{R}^n \mid \|x - c\| = r\}$.

$bd(\mathbb{R}^n_{++}) = bd(\mathbb{R}^n_+) = \{x \in \mathbb{R}^n_+ \mid x_i = 0 \underbrace{\text{for some component}}_{\text{at least one}}\ i\}$

Differentiability

Let f be a function defined on a set $S \subseteq \mathbb{R}^n$, $f: S \to \mathbb{R}$.

Let $x \in int\, S$ and $d \in \mathbb{R}^n \setminus \{0\}$. Consider the limit

$$\lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t}.$$

If exists, this limit is called the <u>directional derivative</u> of f at x along the direction d. <u>Notation:</u> $f'(x;d)$

If one considers the directional derivative along the unit vector e_i, then it's called the i^{th} partial derivative and denoted by

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \downarrow 0} \frac{f(x+te_i)-f(x)}{t}$$

If all partial derivatives exist, then the column vector of all partial derivatives is the gradient of f,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}.$$

A function defined on an open set $U \subseteq \mathbb{R}^n$ is called <u>continuously differentiable</u> over U if all partial derivatives exist and continuous on U.

If a function is continuously differentiable, then we can compute directional derivatives by

$$f'(x;d) = \nabla f(x)^T d.$$

A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called <u>twice continuously differentiable</u> over U if all the second order partial derivatives, that is,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial \left(\frac{\partial f}{\partial x_j}\right)}{\partial x_i}(x) \qquad \text{for all } i,j \in \{1,\dots,n\}$$

exist and are continuous over U.

Note that second order partial derivatives are symmetric:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x) \quad \forall\, i \neq j.$$

<u>Hessian</u> of f at a point $x \in \mathcal{U}$ is the $n \times n$ matrix given by

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

Clearly, it's a symmetric matrix.

<u>Linear Approximation Thm:</u> Let $f : \mathcal{U} \to \mathbb{R}$ be twice cont'ly differentiable, $\mathcal{U} \subseteq \mathbb{R}^n$ open. Let $x \in \mathcal{U}$, $r > 0$ s.t. $B(x,r) \subseteq \mathcal{U}$. For any $y \in B(x,r)$, $\exists\, \xi \in [x,y]$ such that

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

<u>Quadratic Approximation Thm:</u> $f : \mathcal{U} \to \mathbb{R}$ twice conts. diff., $\mathcal{U} \subseteq \mathbb{R}^n$ open. Let $x \in \mathcal{U}$, $r > 0$ s.t. $B(x,r) \subseteq \mathcal{U}$. For any $y \in B(x,r)$, we have

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x) + o\left(\|y-x\|^2\right),$$

where $o(\cdot) : \mathbb{R}_+ \to \mathbb{R}$ is a function satisfying $\frac{o(t)}{t} \to 0$ as $t \downarrow 0$.