Optimization Theory and Applications

Kun Zhu (zhukun@nuaa.edu.cn)

October 24, 2018

Notations

- X is a set, $x \in X$ means x is an element of X, otherwise $x \notin X$
- Also, we use $\{x1, x2, x3, ...\}$ to represent a set
- Alternatively, we can use $\{x : x \in R, x > 5\}$
- X, Y are sets, $X \subset Y$ denotes X is a subset of Y
- The notation $f : X- > Y$ means f is a function from the set X into the set Y
- The symbol $\triangleq$ means "equals by definition"

Vector and Matrix

- We define a column n-vector by

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

- A row n-vector is denoted as $[a_1, a_2, \ldots, a_n]$
- Transpose of a vector $\mathbf{a}^T$, e.g., $\mathbf{a}^T = [a_1, a_2, \ldots, a_n]$

Vector and Matrix

- The sum of two vectors $\mathbf{a}$, $\mathbf{b}$ is
  $\mathbf{a} + \mathbf{b} = [a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n]^T$
- The operation of ***addition*** of vectors has the following properties
  - Commutative: $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$
  - Associative: $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$

Vector and Matrix

- We define an operation of ***multiplication*** of a vector $\mathbf{a} \in \mathbb{R}^n$ by a real scalar $\alpha \in \mathbb{R}$ as $\alpha\mathbf{a} = [\alpha a_1, \alpha a_2, \ldots, \alpha a_n]^T$
- The operation has the following properties
  - Distributive for any real scalars α and β: $\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b}$ $(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a}$
  - Associative: $\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a}$

Vector and Matrix

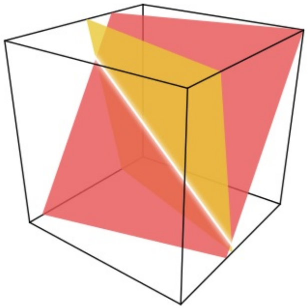
- A set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is said to be ***linearly independent*** if the equality $\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \ldots + \alpha_k\mathbf{a}_k = 0$ implies that all coefficients α_i are equal to zero.
- A set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly dependent if it is not ***linearly independent***
- A vector $\mathbf{a}$ is said to be a ***linear combination*** of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ if there are scalars $\alpha_1, \ldots, \alpha_k$ such that $\mathbf{a} = \alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \ldots + \alpha_k\mathbf{a}_k$
- **Proposition:** A set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly dependent if and only if one of the vectors from the set is a linear combination of the remaining vectors

Vector and Matrix

- A subset $\mathbb{V}$ of $\mathbb{R}^n$ is called a ***subspace*** of $\mathbb{R}^n$ if $\mathbb{V}$ is closed under the operations of vector addition and scalar multiplication
    - That is if $\mathbf{a}$ and $\mathbf{b}$ are vectors in $\mathbb{V}$, then the vectors $\mathbf{a} + \mathbf{b}$ and $\alpha\mathbf{a}$ are also in $\mathbb{V}$ for every scalar α
- Every subspace contains the zero vector $\mathbf{0}$
- Question: Any example of subspace?

Vector and Matrix

- In $\mathbb{R}^3$, the intersection of two-dimensional subspaces is one-dimensional



Vector and Matrix

- Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ be arbitrary vectors in $\mathbb{R}^n$. The set of all their linear combinations is called the ***span*** of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ (a subspace spanned by these vectors) and is denoted by

  span$[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k] = \{\sum_{i=1}^{k} \alpha_i \mathbf{a}_i : \alpha_1, \ldots, \alpha_k \in \mathbb{R}\}$
- Given a vector $\mathbf{a}$, the subspace span$[\mathbf{a}]$ is composed of the vectors $\alpha\mathbf{a}$, where α is an arbitrary real number $\alpha \in R$
- If $\mathbf{a}$ is a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$, then span$[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k, \mathbf{a}]$ = span$[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]$
- The span of any set of vectors is a subspace

Vector and Matrix

- Given a subspace $\mathbb{V}$, a ***basis*** of the subspace $\mathbb{V}$ is any set of linearly independent vectors $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\} \subset \mathbb{V}$ such that $\mathbb{V} = span[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]$
- Two factors:
    - $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\} \subset \mathbb{V}$ are Linearly independent
    - $\mathbb{V}$ can be spanned by $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}$
- Question: Is the basis of a subspace unique?

Vector and Matrix

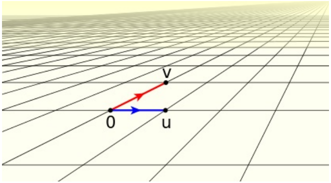
- All bases of a subspace $\mathbb{V}$ contain the same number of vectors. This number is called the ***dimension*** of $\mathbb{V}$, denoted dim $\mathbb{V}$
- **Proposition**: if $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}$ is a basis of $\mathbb{V}$, then any vector $\mathbf{a}$ of $\mathbb{V}$ can be represented uniquely as $\mathbf{a} = \alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \ldots + \alpha_k\mathbf{a}_k$, where $\alpha_i \in R$

Vector and Matrix

- The standard basis for $\mathbb{R}^n$ is the set of vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \ldots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

- The vectors $\mathbf{u}$ and $\mathbf{v}$ are a basis for the two-dimensional subspace of $\mathbb{R}^3$



Vector and Matrix

- A $m \times n$ matrix is denoted by

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- The transpose of matrix $\mathbf{A}$ is a $n \times m$ matrix denoted by

$$\boldsymbol{A}^\top = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

Rank of Matrix

- The maximal number of linearly independent columns of $\mathbf{A}$ is called the ***rank*** of the matrix $\mathbf{A}$
  - The rank $\mathbf{A}$ is the dimension of span$[\mathbf{a}_1, \ldots, \mathbf{a}_n]$
- **Proposition**: The rank of a matrix $\mathbf{A}$ is invariant under the following operations:
  - Multiplication of the columns of $\mathbf{A}$ by nonzero scalars
  - Interchange of the columns
  - Addition to a given column a linear combination of other columns

Linear Equations

- Suppose we are given m equations in n unknowns of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

- We can represent the set of equations as a vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n = \mathbf{b}$, where

$$\boldsymbol{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Liner Equations

- Associated with the system of equations is the matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$
- We can also represent the above system of equations as $\mathbf{A}\mathbf{x} = \mathbf{b}$
- **Theorem**: The system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution if and only if rank$\mathbf{A}$ = rank$[\mathbf{A}, \mathbf{b}]$
- When will the solution be unique?

Inner Product and Norms

- For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we define the ***Euclidean inner product*** by

$$< \mathbf{x}, \mathbf{y} >= \sum_{i=1}^{n} x_i y_i = \mathbf{x}^T \mathbf{y}$$

- The inner product is a ***real-valued function*** with following properties:
  - Positivity: $< \mathbf{x}, \mathbf{x} >\geq 0$, $< \mathbf{x}, \mathbf{x} >= 0$ if and only if $\mathbf{x} = 0$
  - Symmetry: $< \mathbf{x}, \mathbf{y} >=< \mathbf{y}, \mathbf{x} >$
  - Additivity: $< \mathbf{x} + \mathbf{y}, \mathbf{z} >=< \mathbf{x}, \mathbf{z} > + < \mathbf{y}, \mathbf{z} >$
  - Homogeneity: $< r\mathbf{x}, \mathbf{y} >= r < \mathbf{x}, \mathbf{y} >$, for every $r \in \mathbb{R}$
- The vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be orthogonal if $< \mathbf{x}, \mathbf{y} >= 0$

Inner Product and Norms

- The Euclidean norm of a vector $\mathbf{x}$ is defined as
$$||\mathbf{x}|| = \sqrt{< \mathbf{x}, \mathbf{x} >} = \sqrt{\mathbf{x}^T \mathbf{x}}$$

- **Theorem Cauchy-Schwarz Inequality**: For any two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, the Cauchy-Schwarz inequality
$$| < \mathbf{x}, \mathbf{y} > | \leq \|\mathbf{x}\| \|\mathbf{y}\|$$
holds. Furthermore, equality holds if and only if $\mathbf{x} = \alpha \mathbf{y}$ for some $\alpha \in \mathbb{R}$

Inner Product and Norms

- The Euclidean norm of a vector $\|\mathbf{x}\|$ has the following properties:
  - Positivity: $\|\mathbf{x}\| \geq 0$, $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$
  - Homogeneity: $\|r\mathbf{x}\| = |r|\|\mathbf{x}\|$, for $r \in \mathbb{R}$
  - Triangle inequality: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Linear Transformations

- A function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is called a ***linear transformation*** if
  - $\mathcal{L}(\alpha \mathbf{x}) = \alpha \mathcal{L}(\mathbf{x})$, for every $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$
  - $\mathcal{L}(\mathbf{x} + \mathbf{y}) = \mathcal{L}(\mathbf{x}) + \mathcal{L}(\mathbf{y})$, for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- Examples

Eigenvalues and Eigenvectors

- Let $\mathbf{A}$ be an $n \times n$ real square matrix. A scalar λ (possible complex) and a nonzero vector $\mathbf{v}$ satisfying the equation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ are the ***eigenvalue*** and ***eigenvector*** of $\mathbf{A}$
- For λ to be an eigenvalue it is necessary and sufficient for the matrix $\lambda\mathbf{I} - \mathbf{A}$ to be singular, that is det$[\lambda\mathbf{I} - \mathbf{A}] = 0$, where $\mathbf{I}$ is the $n \times n$ identity matrix
  - det$[\lambda\mathbf{I} - \mathbf{A}] = \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0 = 0$
- Can be easily obtained in Matlab: $[v, d] = eig[\mathbf{A}]$

Eigenvalues and Eigenvectors

- **Theorem**: Suppose the characteristic equation $\det[\lambda \mathbf{I} - \mathbf{A}] = 0$ has n distinct roots $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then there exists n linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ such that
$$\mathbb{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, 2, \ldots, n$$

Eigenvalues and Eigenvectors

- With the set of linearly independent set of eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, the matrix $\mathbf{A}$ can be diagnosed
    - i.e., $a_{ij} = 0$ for all $i \neq j$
- Let $\mathbf{T} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]^{-1}$, Then

$$
\begin{aligned}
\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} &= \boldsymbol{T}\boldsymbol{A}[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n] \\
&= \boldsymbol{T}[\boldsymbol{A}\boldsymbol{v}_1, \boldsymbol{A}\boldsymbol{v}_2, \ldots, \boldsymbol{A}\boldsymbol{v}_n] \\
&= \boldsymbol{T}[\lambda_1\boldsymbol{v}_1, \lambda_2\boldsymbol{v}_2, \ldots, \lambda_n\boldsymbol{v}_n] \\
&= \boldsymbol{T}\boldsymbol{T}^{-1}\begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix},
\end{aligned}
$$

Quadratic Forms

- A ***quadratic form*** $f : \mathbb{R}^n \to \mathbb{R}$ is a function $f(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{x}$, where $\mathbf{Q}$ is an $n \times n$ real matrix, and is symmetric: $\mathbf{Q} = \mathbf{Q}^T$
- A quadratic form $\mathbf{x}^T\mathbf{Q}\mathbf{x}$ is ***positive definite*** if $\mathbf{x}^T\mathbf{Q}\mathbf{x} > 0$ for all nonzero vectors $\mathbf{x}$
- It is ***positive semidefinite*** if $\mathbf{x}^T\mathbf{Q}\mathbf{x} \geq 0$ for all $\mathbf{x}$
- It is ***negative definite*** if $\mathbf{x}^T\mathbf{Q}\mathbf{x} < 0$ for all $\mathbf{x}$
- It is ***negative semidefinite*** if $\mathbf{x}^T\mathbf{Q}\mathbf{x} \leq 0$ for all $\mathbf{x}$
- Examples of Quadratic forms

Quadratic Forms

- Quadratic forms have wide applications in optimization
  - Quadratic programming

    minimize $\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{c}^T\mathbf{x}$

    subject to $\mathbf{A}\mathbf{x} \leq b$
  - Linear quadratic optimal control

    Minimize: $J(u, x_0, t_0, t_f) = \int_{t_0}^{t_f} \left[x^T(t)Q(t)x(t) + u^T(t)R(t)u(t)\right] dt + x(t_f)^T S x(t_f)$

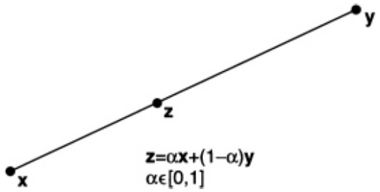
    s.t. $\dot{x} = A(t)x + B(t)u(t)$

Quadratic Forms

- Determining whether a quadratic form is positive definite
- **Theorem:** A quadratic form $\mathbf{x}^T\mathbf{Q}\mathbf{x}$, and $\mathbf{Q} = \mathbf{Q}^T$ is positive definite if and only if all the leading principal minors of $\mathbf{Q}$ are positive
- Or, the eigenvalues of $\mathbf{Q}$ are all positive

Concepts From Geometry

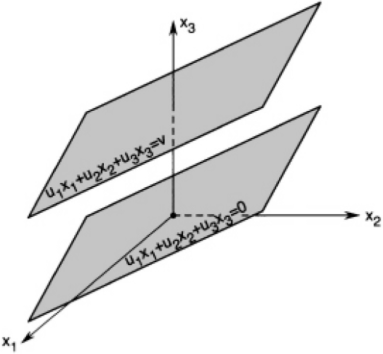
- The ***line segment*** between two points $\mathbf{x}$ and $\mathbf{y}$ is the set of points on the straight line joining points $\mathbf{x}$ and $\mathbf{y}$



- Question: How to define a line?

Hyperplane

- Let $u_1, u_2, \ldots, u_n, v \in \mathbb{R}$, where at least one of the u_i is nonzero. The set of all points $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ that satisfy the linear equation
$$u_1x_1 + u_2x_2 + \ldots + u_nx_n = v$$
is called a ***hyperplane*** of the space $\mathbb{R}^n$



Hyperplane

- Question: Is hyperplane a subspace of $\mathbb{R}^n$?
- Examples of hyperplane:
  - For $n = 2$, the hyperplane is a straight line
  - For $n = 3$, the hyperplane is an ordinary plane
- The hyperplane has dimension $n - 1$

Hyperplane

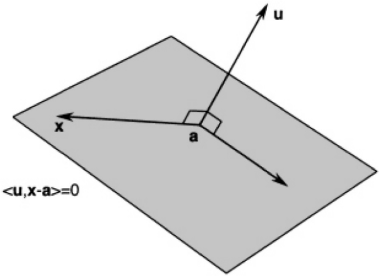
- The hyperplane H divides $\mathbb{R}^n$ into two ***half-spaces***
- One contains points satisfying inequality $u_1x_1 + u_2x_2 + \ldots + u_nx_n \geq v$, denoted by
$$H_+ = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{u}^T\mathbf{x} \geq v\}$$
- One contains points satisfying inequality $u_1x_1 + u_2x_2 + \ldots + u_nx_n \leq v$, denoted by
$$H_- = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{u}^T\mathbf{x} \leq v\}$$
- H_+ is called *positive half-space*, H_- is called *negative half-space*

Hyperplane

- Let $\mathbf{a} = [a_1, \ldots, a_n]^T$ be an arbitrary point of the hyperplane H. Thus, $\mathbf{u}^T\mathbf{a} - v = 0$, and

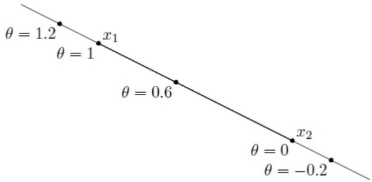
$$\begin{aligned} \boldsymbol{u}^\top \boldsymbol{x} - v &= \boldsymbol{u}^\top \boldsymbol{x} - v - (\boldsymbol{u}^\top \boldsymbol{a} - v) \\ &= \boldsymbol{u}^\top (\boldsymbol{x} - \boldsymbol{a}) \\ &= u_1(x_1 - a_1) + u_2(x_2 - a_2) + \cdots + u_n(x_n - a_n) = 0. \end{aligned}$$

- H contains points for which the vector $\mathbf{u}$ and $\mathbf{x} - \mathbf{a}$ are orthogonal



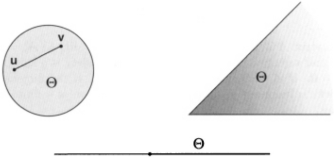
Affine Set

- A set $C \in \mathbb{R}^n$ is ***affine*** if the line through any two distinct points in C lies in C
    - For any $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$, $\theta x_1 + (1-\theta)x_2 \in C$
    - C contains the linear combination of any two points in C, provided the coefficients sum to one
- The idea can be generalized to more than two points
    - A point of the form $\theta_1 x_1 + \ldots + \theta_k x_k$, where $\theta_1 + \ldots + \theta_k = 1$, is an affine combination of points $x_1, \ldots, x_k$

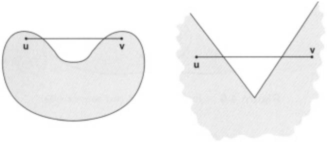


Convex Set

- Definition: a set $\Theta \in \mathbb{R}^n$ is ***convex*** if for all $\mathbf{u}, \mathbf{v} \in \Theta$, the line segment between $\mathbf{u}$ and $\mathbf{v}$ is in Θ. Mathematically, Θ is convex if and only if $\alpha\mathbf{u} + (1-\alpha)\mathbf{v} \in \Theta$ for all $\mathbf{u}, \mathbf{v} \in \Theta$ and $\alpha \in [0, 1]$
- Examples of convex set



- Examples of nonconvex set



Convex Set

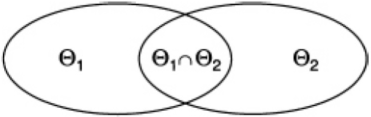
- Is the following set a convex set?

Convex Set

- Examples of convex set
  - The empty set
  - A set consisting of a single point
  - A line or a line segment
  - A subspace
  - A hyperplane
  - A half-space
  - $\mathbb{R}^n$

Convex Set

- **Theorem**: Convex subsets of $\mathbb{R}^n$ has following properties:
  - If Θ is a convex set and β is a real number, then the set
    $$\beta\Theta = \{\mathbf{x} : \mathbf{x} = \beta\mathbf{v}, \mathbf{v} \in \Theta\}$$
    is also convex
  - If Θ_1 and Θ_2 are convex sets, then the set
    $$\Theta_1 + \Theta_2 = \{\mathbf{x} : \mathbf{x} = \mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 \in \Theta_1, \mathbf{v}_2 \in \Theta_2\}$$
    is also convex
  - The intersection of any collection of convex set is convex



Convex Hull

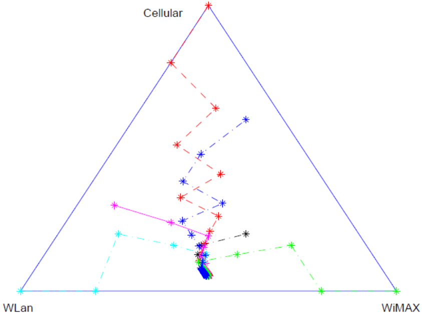
- We call a point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \geq 0$, a ***convex combination*** of the points $x_1, \ldots, x_k$
- A convex combination of points can be thought of as mixture or weighted average of the points, with θ_i as the fraction of x_i in the mixture
- The convex hull of a set C, denoted **conv** C, is the set of all convex combinations of points in C:
  conv $C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C,\ \theta_i \geq 0,\ i = 1, \ldots, k, \theta_1 + \cdots + \theta_k = 1\}$

Convex Hull

- Question: the convex hull for the following sets
  - x_1, x_2 in $\mathbb{R}^2$
  - x_1, x_2, and x_3 in $\mathbb{R}^2$
  - x_1, x_2, x_3, and x_4 in $\mathbb{R}^3$

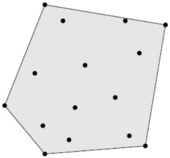
Convex Hull

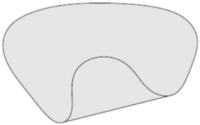
- How to draw the following figure?



Convex Hull

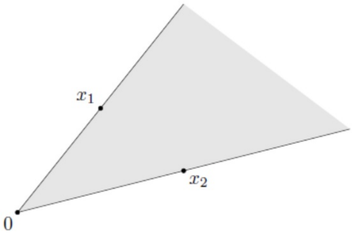
- The convex hull is always convex
- The convex hull is the smallest convex set that contains C
  - If B is any convex set that contains C, then **conv** $C \subset B$
- Convex hull is useful to "convex" a non-convex set





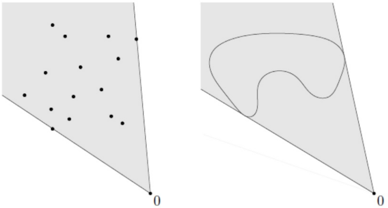
Cone

- A set C is a ***cone***, if for every $x \in C$ and $\theta \geq 0$, we have $\theta x \in C$
- A set C is a ***convex cone*** if it is convex and a cone
    - For any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$, we have $\theta_1 x_1 + \theta_2 x_2 \in C$



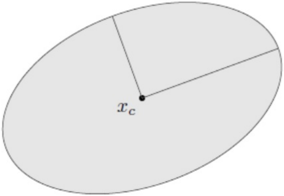
Cone

- A point of the form $\theta_1 x_1 + \ldots + \theta_k x_k$ with $\theta_1, \ldots, \theta_k \geq 0$ is called a ***conic combination*** of $x_1, \ldots, x_k$
- A set C is a convex cone if and only if it contains all conic combinations of its elements
- The ***conic hull*** of a set C is the set of all conic combinations of points in C
$$\{\theta_1 x_1 + \ldots + \theta_k x_k | x_i \in C, \theta_i \geq 0, i = 1, \ldots, k\}$$



Euclidean Balls and Ellipsoids

- An ***Euclidean ball*** in $\mathbb{R}^n$ has the form
  $B(\mathbf{x}_c, r) = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r\}$, where $r > 0$
- The vector $\mathbf{x}_c$ is the center and the scalar r is the radius
- An ***ellipsoid*** have the form
  $\varepsilon = \{\mathbf{x} | (\mathbf{x} - \mathbf{x}_c)^T P^{-1} (\mathbf{x} - \mathbf{x}_c) \leq 1\}$,
  where $P = P^T \succ 0$, i.e., P is symmetric and positive definite



Euclidean Balls and Ellipsoids

- An example of application

With column-wise model, the channel uncertainty is modeled by

$$g_{lm}(n) = \{\bar{g}_{lm}(n) + \triangle g_{lm}(n) : |\triangle g_{lm}(n)| \le \varepsilon_{lm}(n)\},$$

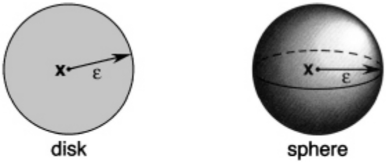
where $\varepsilon_{lm}(n)$ is the column-wise uncertainty bound. With ellipsoidal model, the channel uncertainty is described by

$$g_{lm}(n) = \left\{\bar{g}_{lm}(n) + \triangle g_{lm}(n) : \|\triangle g_{lm}(n)\|_{\mathbf{w}_{lm}(n)} \le \varepsilon_m(n)\right\},$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector and $\mathbf{w}_{lm}(n)$ is a vector of positive weights. Similarly, the uncertainty models for $g_{jk}(n)$, $h_{mk}(n)$, and $h_{km}(n)$ can be obtained.

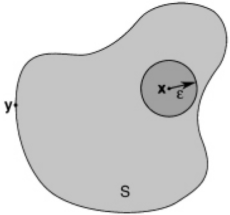
Neighborhoods

- A ***neighborhood*** of a point $\mathbf{x} \in \mathbb{R}^n$ is the set
$$\{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\| < \varepsilon\},$$
where ε is some positive number.
- The neighborhood is also called a *ball* with radius ε and center $\mathbf{x}$
- Examples of neighborhood of a point in $\mathbb{R}^2$ and $\mathbb{R}^3$



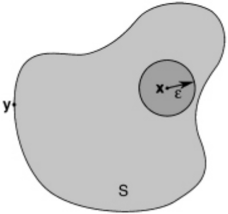
Neighborhoods

- Definition: A point $x \in S$ is said to be an ***interior point*** of the set S if it contains some neighborhood of x, and all points within the neighborhood are also in S
- The set of all interior points of S is called the ***interior*** of S



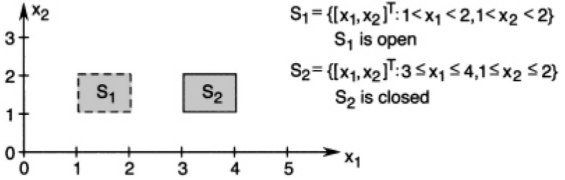
Neighborhoods

- A point x is a ***boundary point*** of the set S if every neighborhood of x contains a point in S and a point not in S
- The set of all boundary points of S is called the ***boundary*** of S
- Question: Is it necessary for a boundary point of S to be an element of S?



Neighborhoods

- A set S is ***open*** if it contains a neighborhood of each of its points (each of its points is an interior point), or equivalently, if S contains no boundary points
- A set S is ***closed*** if it contains its boundary
- A set that is contained in a ball of finite radius is ***bounded***
- A set is compact if it is both ***closed*** and ***bounded***

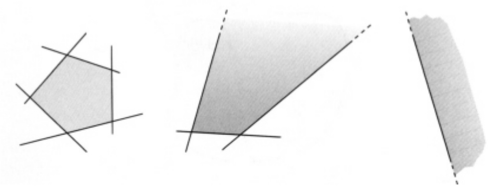


Neighborhoods

- Compact set are important in optimization
- **Theorem:** Let $f : \Omega \to \mathbb{R}$ be a continuous function, where $\Omega \subset \mathbb{R}^n$ is a compact set. Then there exists a point $\mathbf{x}_0 \in \Omega$ such that $f(\mathbf{x}_0) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \Omega$. In other words, f achieves its minimum on Ω
- You will frequently meet the assumption "the set is convex and compact"

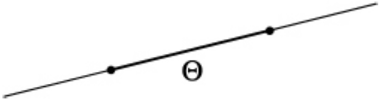
Polytopes and Polyhedra

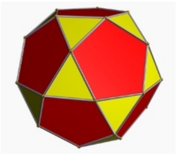
- Let Θ be a convex set, and suppose $\mathbf{y}$ is a boundary point of Θ. A hyperplane passing through $\mathbf{y}$ is called a ***supporting hyperplane*** of the set Θ if the entire set Θ lies completely in one of the two half-spaces into which this hyperplane divides the space $\mathbb{R}^n$
- A set that can be expressed as the intersection of a finite number half-spaces is called a ***convex polytope***



Polytopes and Polyhedra

- A nonempty bounded polytope is called a ***polyhedron***
- Examples of Polyhedrons





Elements of Calculus: Sequences and Limits

- A ***sequence*** of real numbers is a function whose domain is the set of natural numbers $1, 2, \ldots, k, \ldots$ and whose range is contained in $\mathbb{R}$

- A number $x^* \in \mathbb{R}$ is called the ***limit*** of the sequence $\{x_k\}$ if for any positive ε there is a number K such that for all $k > K$, $|x_k - x^*| < \varepsilon$, and we denote $x^* = \lim_{k\to\infty} x_k$

- A sequence that has a limit is called a ***convergent sequence***

- The definition can be generalized to $\mathbb{R}^n$. That is $\forall\varepsilon, \exists K$, such that $\forall k > K$, $\|\mathbf{x}_k - \mathbf{x}^*\| < \varepsilon$, then we denote $\mathbf{x}^* = \lim_{k\to\infty} \mathbf{x}_k$

Derivatives

- Given $f : \mathbb{R} \to \mathbb{R}$
- The derivative of f is a function $f' : \mathbb{R} \to \mathbb{R}$ given by

  $$f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$$

  if the limit exits
- Also written $\frac{df}{dx}$
- If the derivative exists, we say that f is differentiable
- If f' is continuous, we say that f is continuously differentiable

Derivatives

- Given $f : \mathbb{R}^n \to \mathbb{R}$
- The gradient of f is a function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ given by

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

- At each $\mathbf{x}$, $\nabla f(\mathbf{x})$ is a vector in $\mathbb{R}^n$
- Note that for $f : \mathbb{R}^n \to \mathbb{R}$, we have $\nabla f(\mathbf{x}) = Df(\mathbf{x})^T$

Derivatives

- Given $f : \mathbb{R}^n \to \mathbb{R}^m, f = [f_1, \ldots, f_m]^T$
- The derivative of f is a function $Df : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ given by

$$Df(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \ldots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$$

- Sometimes called Jacobian
- At each $\mathbf{x}$, $Df(\mathbf{x})$ is an $m \times n$ matrix
- If Df is continuous, we say that f is continuously differentiable

Derivatives

- If the derivative of ∇f exits, we say that f is twice differentiable
- Write the second derivative as $D^2 f$ (or F), and call it the Hessian of f

$$\boldsymbol{F} = D^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Derivatives

- Example: Find the first and second derivatives for $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$
- The first derivative is

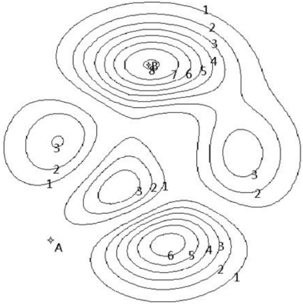
$$Df(\boldsymbol{x}) = (\nabla f(\boldsymbol{x}))^{\top} = \left[\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x})\right] = [5 + x_2 - 2x_1, 8 + x_1 - 4x_2]$$

- The second derivative is

$$\boldsymbol{F}(\boldsymbol{x}) = D^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}$$

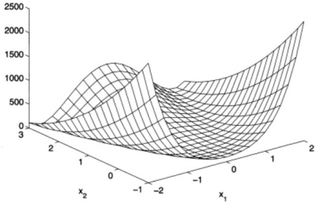
Level Sets

- The ***level set*** of a function $f : \mathbb{R}^n \to \mathbb{R}$ at level c is the set of points
$$S = \{\mathbf{x} : f(\mathbf{x}) = c\}$$
- Examples of level sets
  - for $f : \mathbb{R} \to \mathbb{R}$, the level sets are usually set of points
  - for $f : \mathbb{R}^2 \to \mathbb{R}$, the level sets are usually curves
  - for $f : \mathbb{R}^3 \to \mathbb{R}$, the level sets are usually surfaces

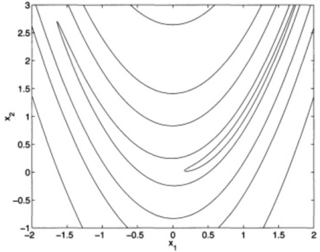


Level Sets Examples

- The Rosenbrock's function $f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$

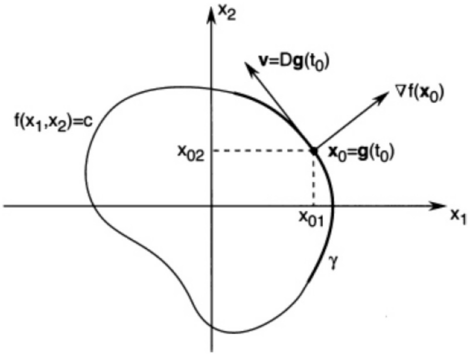


- The level sets of f at 0.7, 7, 70, 200, and 700



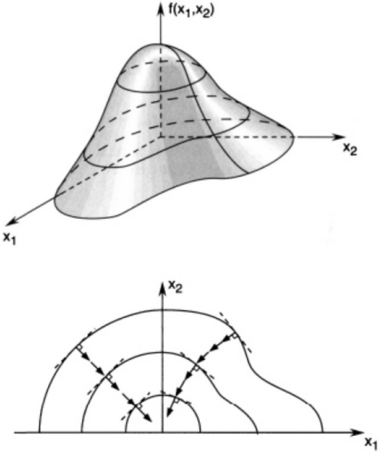
Level Sets and Gradient

- The vector $\nabla f(\mathbf{x}_0)$ is orthogonal to the tangent vector of the level set determined by $f(\mathbf{x}) = f(\mathbf{x_0})$ at $\mathbf{x}_0$
- $\nabla f(\mathbf{x}_0)$ is orthogonal to the level set at $\mathbf{x}_0$



Level Sets and Gradient

- $\nabla f(\mathbf{x}_0)$ is the direction of maximum rate of increase of f at $\mathbf{x}_0$



Level Sets and Gradient

- The values of the function are represented in black and white, black representing higher values, and its corresponding gradient is represented by blue arrows.