KEYPOINT SUMMARY

I. LINEAR ALGEBRA BASICS

1. A nonempty set $\mathcal{S} \in \mathbb{R}^n$ is a **vector subspace** if
   (a) $\mathbf{0} \in \mathcal{S}$, and
   (b) $\boldsymbol{v_1} + \boldsymbol{v_2} \in \mathcal{S}$, $\lambda \boldsymbol{v_1} \in \mathcal{S}$ for all $\boldsymbol{v_1}, \boldsymbol{v_2} \in \mathcal{S}$.

2. Let $\mathcal{S}$ be a vector subspace of $\mathbb{R}^n$. A group of vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}$ are said to **span** $\mathcal{S}$ if every vector in $\mathcal{S}$ is a linear combination of $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}$.

   A group of vectors are called a **basis** of $\mathcal{S}$ if they are *linearly independent* and *span* $\mathcal{S}$.

   If $\mathcal{S}$ has a basis consisting of k vectors, it is said to be k-*dimensional* and denoted as dim $S = k$.

   - Any two bases of the same vector contains the *same* number of vectors.
   - Any combination of n *linearly independent* vectors in $\mathbb{R}^n$ forms a basis for $\mathbb{R}^n$.
   - Let $\mathcal{S}$ be a subspace of a finite-dimensional vector space $\mathcal{R}$. If $\dim \mathcal{S} = \dim \mathcal{R}$, then $\mathcal{S} = \mathcal{R}$.

3. Given an $m \times n$ matrix $\boldsymbol{A}$. The subspace *spanned* by the columns/rows of $\boldsymbol{A}$ is called the **column/row space** of $\boldsymbol{A}$.

   The dimension of the column/row space of $\boldsymbol{A}$ is called the **column/row rank** of $\boldsymbol{A}$.

   - The column and row space are not the same if $m \neq n$.
   - The column and row rank of a matrix are aways equal.

   The solutions of the system $\boldsymbol{Ax} = \mathbf{0}$ form a vector space called the **null space (kernel)** of $\boldsymbol{A}$.

4. **Fundamental Theorem of Linear Algebra**: For any $m \times n$ matrix $\boldsymbol{A}$
$$\dim Ker(\boldsymbol{A}) + rank(\boldsymbol{A}) = n$$

5. Equivalent conditions for an $n \times n$ matrix $\boldsymbol{A}$ to be **invertible**:
   (a) $\boldsymbol{A}$ has n row-leading 1s in *reduced row echelon form*;
   (b) The equation $\boldsymbol{Ax} = \mathbf{0}$ has only *trivial solution* $\boldsymbol{x} = \mathbf{0}$;
   (c) The columns of $\boldsymbol{A}$ are *linearly independent*;
   (d) The columns of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$;
   (e) The rank of $\boldsymbol{A}$ is n;
   (f) The kernel of $\boldsymbol{A}$ has 0 dimension;
   (g) The eigenvalues of $\boldsymbol{A}$ does not contain 0;
   (h) The *determinant* of $\boldsymbol{A}$ is not 0.

6. Vectors $\boldsymbol{v_1}, \ldots, \boldsymbol{v_k}$ are **orthogonal** to each other if $\boldsymbol{v_i} \cdot \boldsymbol{v_j} = 0$ for $i \neq j$, and are **orthonormal** to each other if they further satisfies $\|\boldsymbol{v_1}\| = \cdots = \|\boldsymbol{v_k}\|$.

   A matrix $\boldsymbol{O}$ is an **orthogonal** matrix if $\boldsymbol{O}^T\boldsymbol{O} = I$. The columns (or rows) of an orthogonal matrix are orthonormal to each other.

II. EIGENVALUES AND MATRIX DIAGONALIZATION

1. Let $\boldsymbol{A}$ be an $n \times n$ square matrix, if there exists $\lambda \in \mathbb{R}$ and nonzero $\boldsymbol{v} \in \mathbb{R}^n$ such that
$$\boldsymbol{Av} = \lambda \boldsymbol{v}$$
   λ is an **eigenvalue** of $\boldsymbol{A}$, and $\boldsymbol{v}$ is an **eigenvector** of $\boldsymbol{A}$ corresponding to eigenvalue λ.

   The eigenvalues of $\boldsymbol{A}$ are the *roots* of the **characteristic polynomial** of $\boldsymbol{A}$:
$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0$$
   The set of all eigenvectors corresponding to an eigenvalue λ forms the **eigenspace** of λ.

   - Eigenvalues associated with different eigenvalues are *linearly independent*.
   - Suppose $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$. Then,
     - $\sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii}$ (trace of $\boldsymbol{A}$);
     - $\Pi_{i=1}^n \lambda_i = |\boldsymbol{A}|$.

2. The **algebraic multiplicity** of an eigenvalue λ_i is its exponent in the characteristic polynomial. The **geometric multiplicity** of an eigenvalue λ_i is the dimension of its eigenspace.

   - Geometric multiplicity can *never exceed* the algebraic multiplicity.
   - If for every eigenvalue of matrix $\boldsymbol{A}$, the geometric multiplicity equals the algebraic multiplicity, then $\boldsymbol{A}$ is *diagonalizable*.

3. A matrix $\boldsymbol{A}$ is **diagonalizable** if there exist matrices $\boldsymbol{P}$ and $\boldsymbol{D}$ where $\boldsymbol{D}$ is a *diagonal matrix* such that
$$\boldsymbol{A} = \boldsymbol{PDP}^{-1}$$
   - An $n \times n$ matrix $\boldsymbol{A}$ is diagonalizable *if and only if* the dimension of its eigenspace is equal to n, i.e. the eigenvectors of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$.
   - Sufficient condition for diagonalization: $\boldsymbol{A}$ is diagonalizable if it has n *distinct* eigenvalues. However, the converse might be false.
   - The kernel of $\boldsymbol{A}$ is the *eigenspace* corresponding to eigenvalue 0. Thus the rank of of $\boldsymbol{A}$ is the number of its non-zero eigenvalues (counted by *multiplicity*).

III. SYMMETRIC MATRICES AND QUADRATIC FORMS

1. Matrix $\boldsymbol{A}$ is **symmetric** if $\boldsymbol{A} = \boldsymbol{A}^T$. For any symmetric matrix $\boldsymbol{A}$,
   (a) $\boldsymbol{A}$ has only *real* eigenvalues;
   (b) $\boldsymbol{A}$'s eigenvectors corresponding to distinct eigenvalues are *orthogonal*;
   (c) $\boldsymbol{A}$ is diagonalizable with orthogonal matrices, i.e. there exists a *real diagonal* matrix $\boldsymbol{D}$ and an $orthogonal$ matrix $\boldsymbol{O}$ such that $\boldsymbol{A} = \boldsymbol{ODO}^{-1}$.

2. Suppose $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}$ are linearly independent and span vector subspace $\mathcal{S}$. A set of *orthogonal* vectors $\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_k}$ can be found that span the same vector subspace $\mathcal{S}$ following the **Gram-Schmidt Orthogonalization Process**:

   - $\boldsymbol{w_1} = \boldsymbol{v_1}$
   - $\boldsymbol{w_2} = \boldsymbol{v_2} - \dfrac{\boldsymbol{w_1} \cdot \boldsymbol{v_2}}{\boldsymbol{w_1} \cdot \boldsymbol{w_1}} \boldsymbol{w_1}$
   - $\boldsymbol{w_3} = \boldsymbol{v_3} - \dfrac{\boldsymbol{w_1} \cdot \boldsymbol{v_3}}{\boldsymbol{w_1} \cdot \boldsymbol{w_1}} \boldsymbol{w_1} - \dfrac{\boldsymbol{w_2} \cdot \boldsymbol{v_3}}{\boldsymbol{w_2} \cdot \boldsymbol{w_2}} \boldsymbol{w_2}$, and so on.

3. Let $\boldsymbol{A}$ be an $n \times n$ *symmetric* matrix, then $\boldsymbol{A}$ is:

   (a) **positive definite** if $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$,

   (b) **negative definite** if $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} < 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$,

   (c) **positive semidefinite** if $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$,

   (d) **negative semidefinite** if $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \leq 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$,

   (e) **indefinite** if $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0$ for some $\boldsymbol{x}$ and < 0 for some other $\boldsymbol{x}$.

   - Definiteness of quadratic forms can be defined with non-symmetric matrices, but we do not consider this case. The following theorems only hold for symmetric matrices.
   - If a matrix is positive or negative definite, then it must be *nonsingular*.

4. Let $\boldsymbol{A}$ be a symmetric matrix. Then $\boldsymbol{A}$ is:

   (a) positive definite iff all its eigenvalues are > 0;

   (b) negative definite iff all its eigenvalues are < 0;

   (c) positive semidefinite iff all its eigenvalues are ≥ 0;

   (d) negative semidefinite iff all its eigenvalues are ≤ 0;

   (e) indefinite iff $\boldsymbol{A}$ has both positive eigenvalues and negative eigenvalues.

5. Let $\boldsymbol{A}$ be a symmetric matrix, and D_k be its **leading principal minor** of order k, Δ_k for any of its **principal minors** of order k. Then $\boldsymbol{A}$ is:

   (a) positive definite iff all $D_k > 0$;

   (b) negative definite iff all $(-1)^k D_k > 0$;

   (c) positive semidefinite iff all $\Delta_k \geq 0$;

   (d) negative semidefinite iff all $(-1)^k \Delta_k \geq 0$;

   (e) indefinite if $D_k \neq 0$ but does not fit into any of the above patterns.

6. Let $\boldsymbol{A}$ be a symmetric matrix. The following statements are equivalent:

   (a) $\boldsymbol{A}$ is positive definite;

   (b) There exists nonsingular matrix $\boldsymbol{B}$: $\boldsymbol{A} = \boldsymbol{B}^T \boldsymbol{B}$;

   (c) There exists nonsingular matrix $\boldsymbol{Q}$: $\boldsymbol{Q}^T \boldsymbol{A} \boldsymbol{Q} = \boldsymbol{I}$.

IV. SINGULAR VALUE DECOMPOSITION

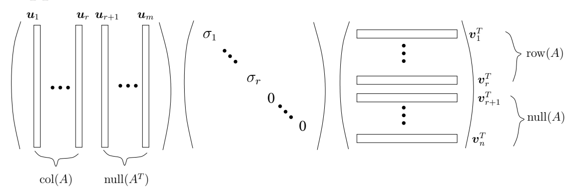
1. **Singular Value Decomposition**: Any $m \times n$ matrix $\boldsymbol{A}$ can be decomposed as

$$\underset{m\times n}{\boldsymbol{A}} = \underset{m\times m}{\boldsymbol{U}} \underset{m\times n}{\boldsymbol{\Sigma}} \underset{n\times n}{\boldsymbol{V}^T}$$

   where $\boldsymbol{\Sigma}$ is a *diagonal* matrix, $\boldsymbol{U}$ and $\boldsymbol{V}$ are *orthogonal* matrices. The diagonal entries of $\boldsymbol{\Sigma}$ is called the **singular values** of $\boldsymbol{A}$.

   - The two facts below are used to compute an SVD:
   - $\boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{V} \boldsymbol{\Sigma}^T \boldsymbol{\Sigma} \boldsymbol{V}^T$
   - $\boldsymbol{A} \boldsymbol{V} = \boldsymbol{U} \boldsymbol{\Sigma}$

2. Suppose $m \times n$ matrix $\boldsymbol{A}$ has SVD of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$:



   Then the following properties hold:

   (a) $rank(\boldsymbol{A}) = rank(\boldsymbol{\Sigma}) = r$;

   (b) First r columns of $\boldsymbol{U}$ are a basis of the column space;

   (c) First r columns of $\boldsymbol{V}$ are a basis of the row space;

   (d) Last $n - r$ columns of $\boldsymbol{V}$ are a basis of $\mathcal{N}(\boldsymbol{A})$;

   (e) Last ast $m - r$ columns of $\boldsymbol{U}$ are a basis of $\mathcal{N}(\boldsymbol{A^T})$.

V. COMPACT SETS AND EXTREME VALUE THEOREM

1. A subset $\mathcal{S} \subset \mathbb{R}^n$ is an **open set** if for every $\boldsymbol{x} \in \mathcal{S}$, there exists an open ball with $\epsilon > 0$ such that $B(\boldsymbol{x}, \epsilon) \subset \mathcal{S}$. A subset $\mathcal{C} \subset \mathbb{R}^n$ is a **closed set** if the complement of $\mathcal{C}$ is open.
   - All balls $B(\boldsymbol{x}, \epsilon)$ are open sets.
   - $\emptyset$ and $\mathbb{R}^n$ are both open and closed.
2. A *closed* and *bounded* subset of $\mathbb{R}^n$ is a **compact set**.
3. **Extreme Value Theorem**: Any *continuous* function defined on a *compact set* has a maximum and a minimum.

VI. CALCULUS OF MULTIPLE VARIABLES

1. A vector-to-vector function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ is defined as:

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$$

   It is a bundle of m functions from $\mathbb{R}^n$ to $\mathbb{R}$.

2. Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, and $\boldsymbol{x}^* \in \mathbb{R}^n$. $\boldsymbol{f}$ is **differentiable** at $\boldsymbol{x}^*$ if there exists a **Jacobian matrix** $\boldsymbol{J}_{m\times n}$ such that

$$\boldsymbol{f}(\boldsymbol{x}^* + \boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{x}^*) = \boldsymbol{J}\boldsymbol{h} + \boldsymbol{r}(\boldsymbol{h})$$

   where $\boldsymbol{h} \in \mathbb{R}^n$ and $\lim_{\|\boldsymbol{h}\| \to 0} \frac{\|\boldsymbol{r}(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0$.
   $\boldsymbol{J}$ is the **derivative** of $\boldsymbol{f}$ at $\boldsymbol{x}^*$ and denoted as $D\boldsymbol{f}(\boldsymbol{x}^*)$. Suppose function $\boldsymbol{f}$ takes the form

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$$

Then the Jocabian matrix of $\boldsymbol{f}$ is calculated as

$$D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

3. The **gradient** of function $f(x_1, \ldots, x_n)$ is defined as

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

The **directional derivative** of f along $\boldsymbol{v}$ is defined as

$$D_v f(\boldsymbol{x}) = (\nabla f(\boldsymbol{x})) \cdot \boldsymbol{v}$$

where $\boldsymbol{v}$ is a unit vector.

- $\nabla f(\boldsymbol{x})$ is a *vector* while $D_v f(\boldsymbol{x})$ is a *scalar*.
- The vector ∇f points to the direction along which f *increases most rapidly*.
- $D_v f(\boldsymbol{x})$ represents the *instantaneous rate of change* of f moving through $\boldsymbol{x}$ with a velocity specified by $\boldsymbol{v}$.

4. Let $\boldsymbol{A}$ be a constant matrix, $\boldsymbol{a}$ be a constant vector. Let $\boldsymbol{u}, \boldsymbol{v}$ be vector functions and $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x})$, $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x})$.

- $\dfrac{\partial \boldsymbol{a} \cdot \boldsymbol{x}}{\partial \boldsymbol{x}} = \dfrac{\partial \boldsymbol{a}^T \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}^T$
- $\dfrac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}$, $\dfrac{\partial \boldsymbol{x}^T \boldsymbol{A}}{\partial \boldsymbol{x}} = \boldsymbol{A}^T$
- $\dfrac{\partial (\boldsymbol{u} + \boldsymbol{v})}{\partial \boldsymbol{x}} = \dfrac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}} + \dfrac{\partial \boldsymbol{v}}{\partial \boldsymbol{x}}$
- $\dfrac{\partial (\boldsymbol{u} \cdot \boldsymbol{v})}{\partial \boldsymbol{x}} = \boldsymbol{u}^T \dfrac{\partial \boldsymbol{v}}{\partial \boldsymbol{x}} + \boldsymbol{v}^T \dfrac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}}$
- $\dfrac{\partial \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{u}))}{\partial \boldsymbol{x}} = \dfrac{\partial \boldsymbol{f}(\boldsymbol{g})}{\partial \boldsymbol{g}} \dfrac{\partial \boldsymbol{g}(\boldsymbol{u})}{\partial \boldsymbol{u}} \dfrac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}}$
- $\dfrac{\partial \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{x}^T (\boldsymbol{A} + \boldsymbol{A}^T) \underbrace{= 2\boldsymbol{x}^T \boldsymbol{A}}_{\boldsymbol{A} \text{ is symmetric}}$

VII. IMPLICIT FUNCTION THEOREM

1. **Implicit Function Theorem I**: Let $h(x_1, \ldots, x_k, y)$ be a $\mathcal{C}^1$ function around the point $\boldsymbol{x}^* = (x_1^*, \ldots, x_k^*, y^*)$. Suppose $h(\boldsymbol{x}^*, y^*) = c$. If it satisfies

$$\frac{\partial h}{\partial y}(x_1^*, \ldots, x_k^*, y^*) \neq 0$$

Then there exits a $\mathcal{C}^1$ function $y = g(\boldsymbol{x})$ defined on an open ball $B(\boldsymbol{x}^*, \epsilon)$ so that:

(a) $y^* = g(\boldsymbol{x}^*)$;

(b) $g(\boldsymbol{x}, g(\boldsymbol{x})) = c$ for all $(\boldsymbol{x}) \in B(\boldsymbol{x}^*, \epsilon)$;

(c) $\left.\dfrac{\partial y}{\partial x_i}\right|_{\boldsymbol{x}^*} = - \left.\dfrac{\frac{\partial G}{\partial x_i}}{\frac{\partial G}{\partial y}}\right|_{(\boldsymbol{x}^*, y^*)}$

2. **Implicit Function Theorem II**: Let $\boldsymbol{f} : \mathbb{R}^{m+n} \to \mathbb{R}^m$ be $\mathcal{C}^1$ where $\boldsymbol{f}(\boldsymbol{y}, \boldsymbol{x}) = (f_1(\boldsymbol{y}, \boldsymbol{x}), \ldots, f_m(\boldsymbol{y}, \boldsymbol{x}))$.
Suppose $\boldsymbol{f}(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{c}$, or be written as

$$\begin{gathered} f_1(y_1, \ldots, y_m, x_1, \ldots, x_n) = c_1 \\ \vdots \\ f_m(y_1, \ldots, y_m, x_1, \ldots, x_n) = c_m \end{gathered}$$

If the Jacobian matrix of $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}$ is *nonsingular*, i.e.

$$\left| \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}} \right| = \begin{vmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{vmatrix} \neq 0$$

Then there exists a $\mathcal{C}^1$ function $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$ defined on an open ball $B(\boldsymbol{x}^*, \epsilon)$ such that:

(a) $\boldsymbol{y}^* = \boldsymbol{g}(\mathbf{x}^*)$;

(b) $\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{x}) = \boldsymbol{c}$ for all $\boldsymbol{x} \in B$;

(c) $\left.\dfrac{\partial \boldsymbol{y}}{\partial x_k}\right|_{(\boldsymbol{y}^*, \boldsymbol{x}^*)}$ can be computed as below:

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_k} \\ \vdots \\ \frac{\partial y_m}{\partial x_k} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1}{\partial x_k} \\ \vdots \\ \frac{\partial f_m}{\partial x_k} \end{bmatrix}$$

3. **Inverse Function Theorem**: Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ (In order for a smooth map from $\mathbb{R}^n$ to $\mathbb{R}^m$ to be invertible, it is required $m = n$) be a $\mathcal{C}^1$ function with $\boldsymbol{f}(\boldsymbol{x}^*) = \boldsymbol{y}^*$. If $D\boldsymbol{f}(\mathbf{x}^*)$ is *nonsingular*, then there exists an open ball $B(\boldsymbol{x}^*, \epsilon)$ and an open set $\mathcal{R}$ about $\boldsymbol{y}^*$ such that $\boldsymbol{f}$ is a *one-to-one onto* map: $\mathcal{B} \to \mathcal{R}$, and the inverse map $\boldsymbol{f}^{-1} : \mathcal{R} \to \mathcal{B}$ is also $\mathcal{C}^1$ and further

$$(D\boldsymbol{f}^{-1})(\boldsymbol{y}^*) = (D\boldsymbol{f}(\boldsymbol{x}^*))^{-1}$$

VIII. CONVEXITY AND CONCAVITY

1. A set $\mathcal{S}$ is called a **convex set** if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$ and $t \in [0, 1]$, $t\boldsymbol{y} + (1-t)\boldsymbol{x} \in \mathcal{S}$.

- The whole space $\mathbb{R}^n$ is convex.
- Every vector subspace is also convex.
- The *intersection* of convex sets is also a convex set.
- The set $\{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \leq 0 \text{ for } i = 1, 2, \ldots, n\}$ is convex if g_i is *quasiconvex*.
- The set $\{\boldsymbol{x} \in \mathbb{R}^n : h_i(\boldsymbol{x}) = 0 \text{ for } i = 1, 2, \ldots, n\}$ is convex if h_i is *affine*.

2. Let f be a function defined on a convex set $\mathcal{S}$. $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$ and $t \in [0, 1]$,
f is **convex** if $f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \leq tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{y})$;
f is **concave** if $f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \geq tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{y})$.
f is **strictly convex/concave** if the strict inequalities hold whenever $\boldsymbol{x} \neq \boldsymbol{y}$ and $t \in (0, 1)$.

- The sum of convex/concave functions are also convex/concave.

- In maximization problems, if the objective function is convex *in its parameters*, the value function is also convex.
- In minimization problems, if the objective function is concave *in its parameters*, the value function is also concave.
- If f is a $\mathcal{C}^2$ function, let $\mathbf{H}$ be its Hessian matrix, then:
  - f is convex $\Leftrightarrow$ $\mathbf{H}$ is positive semidefinite for $\forall \boldsymbol{x}$;
  - f is concave $\Leftrightarrow$ $\mathbf{H}$ is negative semidefinite for $\forall \boldsymbol{x}$;
  - $\mathbf{H}$ is positive definite for $\forall \boldsymbol{x} \Rightarrow f$ is strictly convex;
  - $\mathbf{H}$ is negative definite for $\forall \boldsymbol{x} \Rightarrow f$ is strictly concave.

3. Let f be a function defined on a convex set $\mathcal{S}$. $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$ and $t \in [0, 1]$,

   f is **quasiconvex** if $f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \leq \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}$;

   f is **quasiconcave** if $f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \geq \min\{f(\boldsymbol{x}), f(\boldsymbol{y})\}$.

   f is **strictly quasiconvex/quasiconcave** if the strict inequality hold whenever $\boldsymbol{x} \neq \boldsymbol{y}$ and $t \in (0, 1)$.

   - A (strictly) convex/concave function is also (strictly) quasiconvex/quasiconcave.
   - If f is (strictly) quasiconvex/quasiconcave, and g is strictly increasing, then $g \circ f$ is also (strictly) quasiconvex/quasiconcave.
   - Convex sets and quasiconvex/quasiconcave functions:
     - **Upper level set**: $\mathcal{C}_a^+ = \{\boldsymbol{x} \in \mathcal{S} : f(\boldsymbol{x}) \geq a\}$
     - **Lower level set**: $\mathcal{C}_a^- = \{\boldsymbol{x} \in \mathcal{S} : f(\boldsymbol{x}) \leq a\}$
     - f is quasiconcave if *every* upper level set of f is convex;
     - f is quasiconvex if *every* lower level set of f is convex.
   - If f is a $\mathcal{C}^2$ function, let $\mathbf{H}$ be the Hessian for all $\boldsymbol{x}$, then
     - f is quasiconvex $\Rightarrow$ $\mathbf{H}$ is positive semidefinite on $\mathcal{N}(\nabla f)$;
     - f is quasiconcave $\Rightarrow$ $\mathbf{H}$ is negative semidefinite on $\mathcal{N}(\nabla f)$;
     - $\mathbf{H}$ is positive definite on $\mathcal{N}(\nabla f) \Rightarrow f$ is quasiconvex;
     - $\mathbf{H}$ is negative definite on $\mathcal{N}(\nabla f) \Rightarrow f$ is quasiconcave;
     - In practice, let $\boldsymbol{W}$ be a matrix whose columns are the basis of $\mathcal{N}(\nabla f)$. Then $\mathbf{H}$ is negative (semi) definite on $\mathcal{N}(\nabla f)$ if and only if $\boldsymbol{W}^T\mathbf{H}\boldsymbol{W}$ is a negative (semi) definite matrix.
   - If f is a $\mathcal{C}^2$ function, define the r-th order **Bordered Hessian** of f as:

     $$\bar{\mathbf{H}} = \begin{bmatrix} 0 & f_1 & f_2 & \dots & f_r \\ f_1 & f_{11} & f_{12} & \dots & f_{1r} \\ f_2 & f_{21} & f_{22} & \dots & f_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_r & f_{r1} & f_{r2} & \dots & f_{rr} \end{bmatrix}$$

     Let D_r be the determinant of its r-th order bordered Hessian, then
     - f is quasiconvex $\Rightarrow D_k \leq 0$ for $\forall \boldsymbol{x} \in \mathcal{S}$;
     - f is quasiconcave $\Rightarrow (-1)^k D_k \geq 0$ for $\forall \boldsymbol{x} \in \mathcal{S}$;
     - $D_k < 0$ for $\forall \boldsymbol{x} \in \mathcal{S} \Rightarrow f$ is quasiconvex;
     - $(-1)^k D_k > 0$ for $\forall \boldsymbol{x} \in \mathcal{S} \Rightarrow f$ is quasiconcave.

IX. OPTIMIZATION

1. **Existence of Solution**: Consider the optimization problem: $\max f(\boldsymbol{x})$ s.t. $g_i(\boldsymbol{x}) \leq 0,\ h_i(\boldsymbol{x}) = 0$. The solution *exists* if:
   (a) The feasible set is *non-empty*;
   (b) The objective function is continuous;
   (c) The constraint functions are continuous;
   (d) The constraints are all *weak inequalities*;
   (e) The feasible set is *bounded.*

   Sometimes solution still exists even if the feasible set is not bounded if it can be proved that unbounded values are not optimal.

2. **Uniqueness of Solution**: Let $f : \mathcal{S} \to \mathbb{R}$. If
   (a) $\mathcal{S}$ is a *convex set*, and
   (b) f is *strictly quasiconcave*;

   Then the *global maximum* of f on $\mathcal{S}$ (if exists) is unique.

3. **Unconstrained Optimization**:

   Let f be a $\mathcal{C}^2$ function. Consider the problem:

   $$\max_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$$

   - Necessary conditions for interior optimum:

     $$\nabla f(\boldsymbol{x}^*) = 0$$

   - Sufficient conditions for local optimum: Let $\mathbf{H}$ be the Hessian matrix of f at $\boldsymbol{x}^*$.
     - $\boldsymbol{x}^*$ is a local maximizer $\Rightarrow$ $\mathbf{H}(\boldsymbol{x}^*)$ is negative semidefinite;
     - $\mathbf{H}(\boldsymbol{x}^*)$ is negative definite $\Rightarrow$ $\boldsymbol{x}^*$ is a local maximizer;
   - Sufficient conditions for global optimum:
     - f is *concave* $\Rightarrow$ $\boldsymbol{x}^*$ is a global maximizer.

4. **Constrained Optimization with Equality Constraints**:

   Let $f, h_1, \dots, h_k$ be $\mathcal{C}^2$ functions on $\mathbb{R}^n$. Consider the problem:

   $$\max_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) \text{ s.t. } h_j(\boldsymbol{x}) = c_j$$

   Let $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{j=1}^k \mu_j (h_j(\boldsymbol{x}) - c_j)$.

   - Necessary conditions for interior optimum:

     $$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = 0$$

     $$\forall j,\ h_j(\boldsymbol{x}^*) = c_j$$

   - Constraint Qualification: The rows of the Jacobian $\boldsymbol{h}'(\boldsymbol{x}^*)$ are *linearly independent.*
   - Sufficient conditions for local optimum: Let $\boldsymbol{\mathcal{H}}_{\boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ be the Hessian of $\mathcal{L}$ with respect to $\boldsymbol{x}$ at $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$.
     - $\boldsymbol{x}^*$ is a local maximizer $\Rightarrow$ $\boldsymbol{\mathcal{H}}_{\boldsymbol{x}}$ is negative semidefinite on $\mathcal{N}(\boldsymbol{h}'(\boldsymbol{x}^*))$;
     - $\boldsymbol{\mathcal{H}}_{\boldsymbol{x}}$ is negative definite on $\mathcal{N}(\boldsymbol{h}'(\boldsymbol{x}^*)) \Rightarrow \boldsymbol{x}^*$ is a local maximizer.

- Border the $n \times n$ Hessian $\boldsymbol{\mathcal{H}}_{\boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ with the $k \times n$ matrix $\boldsymbol{h}'(\boldsymbol{x}^*)$:

$$\begin{bmatrix} 0 & \cdots & 0 & | & \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & | & \frac{\partial h_k}{\partial x_1} & \cdots & \frac{\partial h_k}{\partial x_n} \\ -- & -- & -- & -- & -- & -- & -- \\ \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_k}{\partial x_1} & | & \frac{\partial^2 \mathcal{L}}{\partial x_1^2} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\ \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_k}{\partial x_n} & | & \frac{\partial^2 \mathcal{L}}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial x_n^2} \end{bmatrix}$$

Let D_i be the i-th leading principal minor. If $(-1)^{k+i} D_i > 0$, then $\boldsymbol{\mathcal{H}}_{\boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ is negative definite on $\mathcal{N}(\boldsymbol{h}'(\boldsymbol{x}^*))$.

- Sufficient conditions for global optimum:
  - $\mathcal{L}$ is *concave* $\Rightarrow \boldsymbol{x}^*$ is a global maximizer s.t. $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{c}$.
  - In particular, f if concave and $\mu_j^* h_j$ is convex $\forall j \Rightarrow \boldsymbol{x}^*$ is a global maximizer s.t. $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{c}$.

5. **Constrained Optimization with Inequality Constraints**:

Let f, $g_1, \ldots, g_m$, $h_1, \ldots, h_k$ be $\mathcal{C}^2$ functions on $\mathbb{R}^n$. Consider the problem:

$$\max_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) \text{ s.t. } g_i(\boldsymbol{x}) \leq b_i,\ h_j(\boldsymbol{x}) = c_i$$

Define the Lagrangian:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i (g_i(\boldsymbol{x}) - b_i) - \sum_{j=1}^{k} \mu_j (h_j(\boldsymbol{x}) - c_j)$$

- Necessary conditions for interior optimum:

$$\begin{aligned} &\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0 \\ &\forall i,\ g_i(\boldsymbol{x}^*) \leq b_i, \lambda_i \geq 0, \lambda_i (g_i(\boldsymbol{x}^*) - b_i) = 0 \\ &\forall j,\ h_j(\boldsymbol{x}^*) = c_j \end{aligned}$$

- Sufficient conditions for local optimum:

  Suppose that $g_1, \ldots, g_e$ are *binding* at $\boldsymbol{x}^*$, and $g_{e+1}, \ldots, g_m$ are non-binding. Let $\boldsymbol{g}_E = (g_1, \ldots, g_e)$. Let $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{g}_E'(\boldsymbol{x}^*) \\ \boldsymbol{h}'(\boldsymbol{x}^*) \end{bmatrix}$.

  Let $\boldsymbol{\mathcal{H}}_{\boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ be the Hessian of $\mathcal{L}$ with respect to $\boldsymbol{x}$ at $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$.

  - $\boldsymbol{x}^*$ is a local maximizer $\Rightarrow \boldsymbol{\mathcal{H}}_{\boldsymbol{x}}$ is negative semidefinite on $\mathcal{N}(\boldsymbol{C})$;
  - $\boldsymbol{\mathcal{H}}_{\boldsymbol{x}}$ is negative definite on $\mathcal{N}(\boldsymbol{C}) \Rightarrow \boldsymbol{x}^*$ is a local maximizer.

- Sufficient conditions for global optimum: $\boldsymbol{x}^*$ is a global maximizer if
  (a) The feasible set is convex;
  (b) f is concave, or
  (c) f is quasiconcave and $\nabla f(\boldsymbol{x}^*) \neq 0$.

6. **Envelope Theorem (Unconstrained Version)**:

$$V(\boldsymbol{a}) = \max_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}, \boldsymbol{a})$$

Assume that:
(a) $\boldsymbol{x}^*(\boldsymbol{a}_0)$ is the *unique* global maximum of $f(\boldsymbol{x}, \boldsymbol{a}_0)$;
(b) $f(\boldsymbol{x}^*, \boldsymbol{a})$ is continuously differentiable in $\boldsymbol{a}$ at $(\boldsymbol{x}^*, \boldsymbol{a}_0)$;

Then $V(\boldsymbol{a})$ is differentiable at $\boldsymbol{a}_0$ and

$$\nabla_{\boldsymbol{a}} V(\boldsymbol{a}_0) = \nabla_{\boldsymbol{a}} f(\boldsymbol{x}^*, \boldsymbol{a}_0)$$

- Differentiability of $\boldsymbol{x}^*$, or differentiability of f with respect to $\boldsymbol{x}$ are not required.
- This theorem also works for constrained optimization problems as long as no constraints depend on parameter $\boldsymbol{a}$.

Envelope Theorem (Constrained Version):

$$V(\boldsymbol{a}) = \max_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}, \boldsymbol{a}) \text{ s.t. } g_i(\boldsymbol{x}, \boldsymbol{a}) \leq b_i, h_j(\boldsymbol{x}, \boldsymbol{a}) = c_i$$

Assume that:
(a) f, g_i, h_j all continuously differentiable with respect to both $\boldsymbol{x}$ and $\boldsymbol{a}$;
(b) $\boldsymbol{x}^*(\boldsymbol{a}_0)$ satisfies the *constraint qualification* and is the *unique* global constrained maximum;
(c) The set of binding inequalities remain *unchanged* at optimality for near $\boldsymbol{a}_0$;

Then $V(\boldsymbol{a})$ is differentiable at $\boldsymbol{a}_0$ and

$$\nabla_{\boldsymbol{a}} V(\boldsymbol{a}_0) = \nabla_{\boldsymbol{a}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{a}_0)$$