

1 Optimization

Definition. Multiple Variable Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

1. We say that $a \in \mathbb{R}^n$ is a **local maximum** of f if there exists a neighborhood $U \subseteq \mathbb{R}^n$ containing a such that $f(x) \leq f(a)$ for all $x \in U$ (i.e. There exists an open ball $B_\epsilon(a)$ such that $f(a) > f(x)$ for all $x \in B_\epsilon(a)$)
2. We say that $a \in \mathbb{R}^n$ is a **local minimum** of f if there exists a neighborhood $U \subseteq \mathbb{R}^n$ containing a such that $f(x) \geq f(a)$ for all $x \in U$

Definition. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable, we say that $c \in \mathbb{R}^n$ is a critical point of f if $\nabla f(c) = 0$. If c is a **critical point**, we say that $f(c)$ is a **critical value**. All points which are not critical are termed **regular points**

Remark. If $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$ we say that $c \in \mathbb{R}^n$ is a critical point if $Df(c)$ does not have maximal rank. Also critical points may be entire curves / planes in higher dimensions

Definition. The column rank of A is the dimension of the column space of A , while the row rank of A is the dimension of the row space of A . Column rank and row rank are always equal. A matrix is said to have **full rank** if its rank equals the largest possible for a matrix of the same dimensions, which is the lesser of the number of rows and columns.

Proposition. Single Variable If $f : [a, b] \rightarrow \mathbb{R}$ is differentiable and c is an interior point which is either a local maximum or local minimum, then necessarily $f'(c) = 0$

Proposition. Multiple Variable (Fermat's Theorem) Let $U \subseteq \mathbb{R}^n$. If $f : U \rightarrow \mathbb{R}$ is differentiable and $c \in U$ (note $c \in \text{int}U$) is either a local maximum or minimum of f , then $\nabla f(c) = 0$

Remark. The theorem implies that all local maxima and minima of a differentiable function occur at critical points. It is a method to find local maxima and minima of differentiable functions on *open* sets by showing that every local extremum of the function is a stationary point (the function derivative is zero in that point). Note that critical points would be necessary but not sufficient for extrema to occur. There are cases where critical point does not yield extrema. For example, $f(x, y) = y^2 - x^2$ has critical point at $(0, 0)$, which is not an extrema

Definition. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is C^2 and c is a critical point of f , then we say that c is a **degenerate critical point** if $\text{rank } H(c) < n$

Remark. When doing 2nd Derivative Test, we first evaluate if the Hessian matrix yield degenerate critical point. Subsequent discussion depends on the critical point being non-degenerate. If the critical point is degenerate, then the behavior of f in a neighborhood of c , even in a very qualitative sense and in a very small neighborhood, may depend on higher-order terms in the Taylor series. Remember that we ignored Taylor polynomial for terms with order larger than 3 when discussing about the 2nd Derivative Test

Proposition. If $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ and c be a critical point

1. If $\det H(c) < 0$ then c is a saddle point
2. If $\det H(c) > 0$
 - (a) If $\partial_{11}f > 0$ then c is a minimum
 - (b) If $\partial_{11}f < 0$ then c is a maximum
3. If $\det H(c) = 0$ then this is inconclusive

Definition. Given a square matrix $A \in \mathbb{R}^{n \times n}$ and a vector $x \in \mathbb{R}^n$, $x^T A x$ is called a **quadratic form**

1. A symmetric matrix $A \in S^n$ is **positive definite** if for all non-zero vectors $x \in \mathbb{R}^n$, $x^T A x > 0$. Or simply all eigenvalues are positive.
2. A symmetric matrix $A \in S^n$ is **positive semidefinite** if for all non-zero vectors $x \in \mathbb{R}^n$, $x^T A x \geq 0$. Or simply all eigenvalues are non-negative.

Remark. One important property of positive definite and negative definite matrices is that they are always full rank, and hence, invertible. Also, if A is positive definite then $-A$ is negative definite and vice versa.

Definition. Sylvester's Criterion a matrix A is positive definite iff the determinants associated with all upper-left submatrices of A are positive.

Remark. Used to determine if the Hessian matrix is positive definite. To see if the matrix is negative definite, we simply evaluate negatives of upper-left submatrices $-A$

Theorem. For every symmetric real matrix A there exists a real orthogonal matrix Q such that $D = Q^T A Q$ is a diagonal matrix. Every symmetric matrix is thus, up to choice of an orthonormal basis, a diagonal matrix.

Proposition. 2nd Derivative Test let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be C^2 in the neighborhood of a critical point a . Assume $\nabla f(a) = 0$. Form Hessian matrix H at point a

1. If H is positive definite (i.e. All eigenvalues positive), then a is (strictly) local minimum
2. If H is negative definite (i.e. All eigenvalues negative), then a is (strictly) local maximum
3. If H is neither (i.e. A mix of positive and negative eigenvalues), but $\det(H) \neq 0$ at point a , then a is a saddle point.
4. If $\det(H) = 0$ test is inconclusive

Proof. Since $\nabla f(a) = 0$, the second-order Taylor polynomial tells us that the function is approximately,

$$f(x) = f(a) + \frac{1}{2}(x-a)^T H(a)(x-a)$$

If H is positive definite or all its eigenvalues are positive, i.e. $(x-a)^T H(a)(x-a) > 0$, then

$$f(x) > f(a)$$

we see that a is a local minimum

Remark. There are cases where all eigenvalues of $H(a)$ is nonnegative but a is not a local min.

□

Proposition. For a matrix $A \in \mathbb{R}^{n \times n}$, λ is an eigenvalue if there exists a non-zero $v \in \mathbb{R}^n$ such that

$$Av = \lambda v \iff (A - \lambda I)v = 0$$

which is a linear system of equations with $(A - \lambda I)$ as coefficients. The system has one solution if and only if $(A - \lambda I)$ is invertible, i.e. $\det(A - \lambda I) \neq 0$. Since the zero vector is a solution and we know that $v \neq 0$, we must have

$$\det(A - \lambda I) = 0$$

From this we can ensure that diagonal matrices has eigenvalues as entries on its diagonals

Theorem. Constrained Optimization by method of Lagrange Multipliers Let $f : G : \mathbb{R}^n \rightarrow \mathbb{R}$ be C^1 functions, and set $S = G^{-1}(0)$, i.e. S provides $x \in \mathbb{R}^n$ such that $G(x) = 0$. If the restriction $f : S \rightarrow \mathbb{R}$ has a maximum or a minimum at a point $c \in S$ and $\nabla G(c) \neq 0$ then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(c) = \lambda \nabla G(c)$$

Proof. Intuition: $\nabla f(x)$ is the direction of f at point x . As long as we are not moving perpendicular to $\nabla f(x)$, we are either moving uphill or downhill. f achieves its local extrema at c whenever we are only restricted to move in directions perpendicular to $\nabla f(c)$. This is made possible because we have to abide by the constraint $G(x) = 0$; that is, we can only move along the tangents to the constraint surface G . Note that ∇G is always perpendicular to the tangent space, direction of possible movements because G is a constant function. Together, at local extrema c , $\nabla f(c)$ and $\nabla G(c)$ both become perpendicular to the tangent space of the constraint surface. Therefore $\nabla f(c)$ and $\nabla G(c)$ are parallel and hence proportional to each other by an arbitrary constant λ , the Lagrange Multiplier □

Remark. We can use 2nd Derivative Test wherever possible, i.e. for interior of sets, for finding critical points. For determining critical points on the boundary, we apply the method of lagrange multiplier. We can then infer that absolute extrema of f exists by arguing with

Extreme Value Theorem. Since absolute extrema are also themselves local extrema, they must have been in one of the critical points determined previously. Then, we simply evaluate f at those critical points. The largest value corresponds to where maximum occurs. Also note that when $\nabla G(c) = 0$, we consider c as a critical point because the constraint surface is not smooth. We would still evaluate $f(c)$ and compare to other critical points

Proposition. *For multiple constraints, we need additional multipliers. More precisely, if $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is given by $G(x) = (G_1(x), \dots, G_m(x))$, we set $S = G^{-1}(0)$, and we tasked with optimizing $f : S \rightarrow \mathbb{R}$, then if $c \in S$ is a maximum or minimum there exists $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ such that*

$$\nabla f(c) = \sum_{i=1}^m \lambda_i \nabla G_i(c)$$

Remark. If $\nabla G_i(c)$ happens to be linear dependent of each other, it's valid to reduce the expression to just linearly independent $G_i(c)$.