Math 408A
Unconstrained Optimization

Optimality Conditions

Unconstrained Optimization

Coercivity

The Directional Derivative

First-Order Optimality Conditions

Second-Order Optimality Conditions

Nonlinear Programming

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_j(x) \leq 0,\ j = 1, 2, \dots, s \\ & f_j(x) = 0,\ j = s = 1, \dots, m\ . \end{array}$$

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\overline{x} \in \Omega$ is said to be a <u>global</u> solution to the problem

$$\min\{f_0(x) : x \in \Omega\}$$

  if $f_0(\overline{x}) \leq f_0(x)$ for all $x \in \Omega$.

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\overline{x} \in \Omega$ is said to be a global solution to the problem

  $$\min\{f_0(x) : x \in \Omega\}$$

  if $f_0(\overline{x}) \le f_0(x)$ for all $x \in \Omega$.
- If in fact $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$, then $\overline{x}$ is said to be a strict global solution.

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\overline{x} \in \Omega$ is said to be a global solution to the problem
$$\min\{f_0(x) : x \in \Omega\}$$
if $f_0(\overline{x}) \leq f_0(x)$ for all $x \in \Omega$.
- If in fact $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$, then $\overline{x}$ is said to be a strict global solution.
- $\overline{x} \in \Omega$ is said to be a local solution to the problem $\min\{f_0(x) : x \in \Omega\}$ if there is an $\epsilon > 0$ such that
$$f_0(\overline{x}) \leq f_0(x) \quad \text{for all } x \in \Omega \text{ satisfying} \quad \|\overline{x} - x\| \leq \epsilon.$$

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\overline{x} \in \Omega$ is said to be a global solution to the problem
$$\min\{f_0(x) : x \in \Omega\}$$
if $f_0(\overline{x}) \le f_0(x)$ for all $x \in \Omega$.
- If in fact $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$, then $\overline{x}$ is said to be a strict global solution.
- $\overline{x} \in \Omega$ is said to be a local solution to the problem $\min\{f_0(x) : x \in \Omega\}$ if there is an $\epsilon > 0$ such that
$$f_0(\overline{x}) \le f_0(x) \quad \text{for all } x \in \Omega \text{ satisfying } \quad \|\overline{x} - x\| \le \epsilon.$$
- If $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$ with $\|x - \overline{x}\| \le \epsilon$, then $\overline{x}$ is called a strict local solution.

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\overline{x} \in \Omega$ is said to be a global solution to the problem

$$\min\{f_0(x) : x \in \Omega\}$$

  if $f_0(\overline{x}) \leq f_0(x)$ for all $x \in \Omega$.
- If in fact $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$, then $\overline{x}$ is said to be a strict global solution.
- $\overline{x} \in \Omega$ is said to be a local solution to the problem $\min\{f_0(x) : x \in \Omega\}$ if there is an $\epsilon > 0$ such that

$$f_0(\overline{x}) \leq f_0(x) \quad \text{for all } x \in \Omega \text{ satisfying} \quad \|\overline{x} - x\| \leq \epsilon.$$

- If $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$ with $\|x - \overline{x}\| \leq \epsilon$, then $\overline{x}$ is called a strict local solution.
- The solution $\overline{x}$ is said to be isolated if $\overline{x}$ is the only local solution in the set $\{x \in \Omega : \|x - \overline{x}\| \leq \epsilon\}$.

Optimality Conditions

These definitions, although sensible, are not practical since they require one to check infinitely many points to determine either local or global optimality.

Optimality Conditions

These definitions, although sensible, are not practical since they require one to check infinitely many points to determine either local or global optimality.

An optimality condition is a simple practical test for optimality.

Optimality Conditions

These definitions, although sensible, are not practical since they require one to check infinitely many points to determine either local or global optimality.

An optimality condition is a simple practical test for optimality.

Theorem: *If* f_0 *is differentiable at* $\overline{x}$ *and* $\overline{x}$ *is a local solution to the problem* $\min\{f_0(x) : x \in \mathbb{R}^n\}$, *then* $\nabla f_0(\overline{x}) = 0$.

Optimality Conditions

These definitions, although sensible, are not practical since they require one to check infinitely many points to determine either local or global optimality.

An optimality condition is a simple practical test for optimality.

Theorem: *If* f_0 *is differentiable at* $\overline{x}$ *and* $\overline{x}$ *is a local solution to the problem* $\min\{f_0(x) : x \in \mathbb{R}^n\}$, *then* $\nabla f_0(\overline{x}) = 0$.

The condition $\nabla f_0(\bar{x}) = 0$ is necessary, but clearly not sufficient for optimality.

Optimality Conditions

Optimality conditions play a key role in both the design of our algorithms and our tests for termination.

Optimality Conditions

Optimality conditions play a key role in both the design of our algorithms and our tests for termination.

We design our algorithms to locate points that satisfy some testable, or constructive, optimality conditions and then terminate when the procedure "nearly" solves these conditions.

Optimality Conditions

Optimality conditions play a key role in both the design of our algorithms and our tests for termination.

We design our algorithms to locate points that satisfy some testable, or constructive, optimality conditions and then terminate when the procedure "nearly" solves these conditions.

We begin by reviewing the optimality conditions for unconstrained optimization problems.

Unconstrained Optimization

$$\mathcal{P} \qquad \min_{x \in \mathbb{R}^n} f(x)$$

Given an initial guess x^0 at a solution to $\mathcal{P}$, we wish to develop strategies for updating x^0 to a new point x^1 such that

$$f(x^1) < f(x^0).$$

Unconstrained Optimization

$$\mathcal{P} \qquad \min_{x\in\mathbb{R}^n} f(x)$$

Given an initial guess x^0 at a solution to $\mathcal{P}$, we wish to develop strategies for updating x^0 to a new point x^1 such that

$$f(x^1) \;<\; f(x^0).$$

In this way we can generate a sequence $\{x^\nu\}$ of approximate solutions satisfying

$$f(x^{\nu+1}) \;<\; f(x^\nu) \qquad \nu = 1, 2, \ldots \;.$$

Unconstrained Optimization

$$\mathcal{P} \qquad \min_{x\in\mathbb{R}^n} f(x)$$

Given an initial guess x^0 at a solution to $\mathcal{P}$, we wish to develop strategies for updating x^0 to a new point x^1 such that

$$f(x^1) \;<\; f(x^0).$$

In this way we can generate a sequence $\{x^\nu\}$ of approximate solutions satisfying

$$f(x^{\nu+1}) \;<\; f(x^\nu) \qquad \nu = 1, 2, \ldots \;.$$

Such methods are called descent methods.

Unconstrained Optimization

$$\mathcal{P} \qquad \min_{x\in\mathbb{R}^n} f(x)$$

Given an initial guess x^0 at a solution to $\mathcal{P}$, we wish to develop strategies for updating x^0 to a new point x^1 such that

$$f(x^1) \ < \ f(x^0).$$

In this way we can generate a sequence $\{x^\nu\}$ of approximate solutions satisfying

$$f(x^{\nu+1}) \ < \ f(x^\nu) \qquad \nu = 1, 2, \ldots \ .$$

Such methods are called descent methods.
But before developing these methods, we revisit the question of the existence of solutions and how to identify them.

Existence

$$\mathcal{P} \qquad \min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ where f is continuous on $\mathbb{R}^n$.

Existence

$$\mathcal{P} \qquad \min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ where f is continuous on $\mathbb{R}^n$.

f may have no minimum value, or if it does, it may not be attained.

Existence

$$\mathcal{P} \qquad \min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ where f is continuous on $\mathbb{R}^n$.

f may have no minimum value, or if it does, it may not be attained.

Weierstrass Extreme Value Theorem
Every continuous function on a compact set attains its extreme values on that set.

Coercivity and Existence

Definition: *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is said to be coercive if for every sequence* $\{x^\nu\} \subset \mathbb{R}^n$ *for which* $\|x^\nu\| \to \infty$ *it must be the case that* $f(x^\nu) \to \infty$ *as well.*

Coercivity and Existence

Definition: *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is said to be coercive if for every sequence* $\{x^\nu\} \subset \mathbb{R}^n$ *for which* $\|x^\nu\| \to \infty$ *it must be the case that* $f(x^\nu) \to \infty$ *as well.*

Theorem: *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be continuous on all of* $\mathbb{R}^n$*. The function* f *is coercive if and only if for every* $\alpha \in \mathbb{R}$ *the set* $\{x \mid f(x) \le \alpha\}$ *is compact.*

Coercivity and Existence

Definition: *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is said to be coercive if for every sequence* $\{x^\nu\} \subset \mathbb{R}^n$ *for which* $\|x^\nu\| \to \infty$ *it must be the case that* $f(x^\nu) \to \infty$ *as well.*

Theorem: *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be continuous on all of* $\mathbb{R}^n$. *The function* f *is coercive if and only if for every* $\alpha \in \mathbb{R}$ *the set* $\{x \mid f(x) \leq \alpha\}$ *is compact.*

Theorem: *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be continuous on all of* $\mathbb{R}^n$. *If* f *is coercive, then* f *has at least one global minimizer.*

Back to Optimality Conditions

The Directional Derivative:

Let $f : \mathbb{R}^n \to \mathbb{R}$. Define

$$f'(x; d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}.$$

Back to Optimality Conditions

The Directional Derivative:
Let $f : \mathbb{R}^n \to \mathbb{R}$. Define

$$f'(x;d) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t}.$$

When this limit exists, we say that f is directionally differentiable at x in the direction d with directional derivative $f'(x;d)$.

Back to Optimality Conditions

The Directional Derivative:
Let $f : \mathbb{R}^n \to \mathbb{R}$. Define

$$f'(x; d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}.$$

When this limit exists, we say that f is directionally differentiable at x in the direction d with directional derivative $f'(x; d)$.

If $f'(x; d) < 0$, then there must be a $\bar{t} > 0$ such that

$$\frac{f(x + td) - f(x)}{t} < 0 \quad \text{whenever} \quad 0 < t < \bar{t} \ .$$

Back to Optimality Conditions

The Directional Derivative:
Let $f : \mathbb{R}^n \to \mathbb{R}$. Define

$$f'(x; d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}.$$

When this limit exists, we say that f is directionally differentiable at x in the direction d with directional derivative $f'(x; d)$.

If $f'(x; d) < 0$, then there must be a $\bar{t} > 0$ such that

$$\frac{f(x + td) - f(x)}{t} < 0 \quad \text{whenever} \quad 0 < t < \bar{t} \ .$$

Any direction d for which $f'(x; d) < 0$ is called a direction of strict descent for f at x.

First-Order Optimality Conditions

If there is a direction d such that $f'(x;d)$ exists with $f'(x;d) < 0$, then x cannot be a local solution to the problem $\min_{x\in\mathbb{R}^n} f(x)$.

First-Order Optimality Conditions

If there is a direction d such that $f'(x;d)$ exists with $f'(x;d) < 0$, then x cannot be a local solution to the problem $\min_{x\in\mathbb{R}^n} f(x)$.

Equivalently, if x is a local to the problem $\min_{x\in\mathbb{R}^n} f(x)$, then $f'(x;d) \geq 0$ whenever $f'(x;d)$ exists.

First-Order Optimality Conditions

If there is a direction d such that $f'(x;d)$ exists with $f'(x;d) < 0$, then x cannot be a local solution to the problem $\min_{x \in \mathbb{R}^n} f(x)$.

Equivalently, if x is a local to the problem $\min_{x \in \mathbb{R}^n} f(x)$, then $f'(x;d) \geq 0$ whenever $f'(x;d)$ exists.

Lemma:[Basic First-Order Optimality Result]
Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\bar{x} \in \mathbb{R}^n$ be a local solution to the problem $\min_{x \in \mathbb{R}^n} f(x)$. *Then*

$$f'(x;d) \geq 0$$

for every direction $d \in \mathbb{R}^n$ for which $f'(x;d)$ exists.

First-Order Optimality Conditions

If there is a direction d such that $f'(x;d)$ exists with $f'(x;d) < 0$, then x cannot be a local solution to the problem $\min_{x\in\mathbb{R}^n} f(x)$.

Equivalently, if x is a local to the problem $\min_{x\in\mathbb{R}^n} f(x)$, then $f'(x;d) \geq 0$ whenever $f'(x;d)$ exists.

Lemma:[Basic First-Order Optimality Result]
Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\bar{x} \in \mathbb{R}^n$ be a local solution to the problem $\min_{x\in\mathbb{R}^n} f(x)$. *Then*

$$f'(x;d) \geq 0$$

for every direction $d \in \mathbb{R}^n$ for which $f'(x;d)$ exists.

Theorem: *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a point $\bar{x} \in \mathbb{R}^n$. If $\bar{x}$ is a local minimum of f, then $\nabla f(\bar{x}) = 0$.*

First-order Optimality Conditions

Theorem: *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be differentiable at a point* $\overline{x} \in \mathbb{R}^n$. *If* $\overline{x}$ *is a local minimum of* f, *then* $\nabla f(\overline{x}) = 0$.

First-order Optimality Conditions

Theorem: *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be differentiable at a point* $\overline{x} \in \mathbb{R}^n$*. If* $\overline{x}$ *is a local minimum of* f*, then* $\nabla f(\overline{x}) = 0$*.*

Proof: By the basic lemma we have

$$0 \leq f'(\bar{x}; d) = \nabla f(\bar{x})^T d \quad \text{for all} \quad d \in \mathbb{R}^n \ .$$

Taking $d = -\nabla f(\bar{x})$ we get

$$0 \leq -\nabla f(\bar{x})^T \nabla f(\bar{x}) = -\|\nabla f(\bar{x})\|^2 \leq 0.$$

Therefore, $\nabla f(\bar{x}) = 0$.

Stationary and Critical Points

When $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, any point $x \in \mathbb{R}^n$ satisfying $\nabla f(x) = 0$ is said to be a stationary (critical point) of f.
These are candidate points for optimality (minima or maxima).

Stationary and Critical Points

When $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, any point $x \in \mathbb{R}^n$ satisfying $\nabla f(x) = 0$ is said to be a stationary (critical point) of f.
These are candidate points for optimality (minima or maxima).

Theorem: *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable on all of $\mathbb{R}^n$. If f is coercive, then f has at least one global minimizer. These global minimizers can be found from among the critical points of f.*

Second-Order Optimality Conditions

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable at the point $\overline{x} \in \mathbb{R}^n$.

Second-Order Optimality Conditions

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable at the point $\overline{x} \in \mathbb{R}^n$.

1. (Necessity) If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive semi-definite.

Second-Order Optimality Conditions

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable at the point $\overline{x} \in \mathbb{R}^n$.

1. (Necessity) If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive semi-definite.
2. (Sufficiency) If $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive definite, then there is an $\alpha > 0$ such that $f(x) \geq f(\overline{x}) + \alpha\|x - \overline{x}\|^2$ for all x near $\overline{x}$.

Second-Order Optimality Conditions

Since $\nabla^2 f(x)$ is symmetric, it has an orthonormal basis of eigenvectors of $\nabla^2 f(x)$, $v^1, v^2, \ldots, v^n$ such that

$$\nabla^2 f(x) = V^T \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \ldots & \ldots & \lambda_n \end{bmatrix} V$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $\nabla^2 f(x)$ and the columns of V are the corresponding orthonormal eigenvectors vectors.

Second-Order Optimality Conditions

Since $\nabla^2 f(x)$ is symmetric, it has an orthonormal basis of eigenvectors of $\nabla^2 f(x)$, $v^1, v^2, \ldots, v^n$ such that

$$\nabla^2 f(x) = V^T \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \ldots & \ldots & \lambda_n \end{bmatrix} V$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $\nabla^2 f(x)$ and the columns of V are the corresponding orthonormal eigenvectors vectors.
$\nabla^2 f(x)$ is positive semi-definite if and only if

$$\lambda_i \geq 0, \ \ i = 1, 2, \ldots, n,$$

and it is positive definite if and only if

$$\lambda_i > 0, \ \ i = 1, 2, \ldots, n.$$

Second-Order Optimality Conditions

In particular, if $\nabla^2 f(x)$ is positive definite, then

$$d^T \nabla^2 f(x) d \geq \lambda_{\min} \|d\|^2 \quad \text{for all } d \in \mathbb{R}^n,$$

where $\lambda_{\min}$ is the smallest eigenvalue of $\nabla^2 f(x)$.

Second-Order Optimality Conditions

(Necessity) If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive semi-definite.

Second-Order Optimality Conditions

(Necessity) If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive semi-definite.

Proof: Use of the second-order Taylor expansion

$$f(x) = f(\overline{x}) + \nabla f(\overline{x})^T(x - \overline{x}) + \frac{1}{2}(x - \overline{x})^T \nabla^2 f(\overline{x})(x - \overline{x}) + o(\|x - \overline{x}\|^2).$$

Second-Order Optimality Conditions

(Necessity) If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive semi-definite.

Proof: Use of the second-order Taylor expansion

$$f(x) = f(\overline{x}) + \nabla f(\overline{x})^T(x - \overline{x}) + \frac{1}{2}(x - \overline{x})^T \nabla^2 f(\overline{x})(x - \overline{x}) + o(\|x - \overline{x}\|^2).$$

Given $d \in \mathbb{R}^n$ and $t > 0$ plug in $x := \overline{x} + td$ to get

$$0 \leq \frac{f(\bar{x} + td) - f(\bar{x})}{t^2} = \frac{1}{2} d^T \nabla^2 f(\overline{x}) d + \frac{o(t^2)}{t^2}$$

since $\nabla f(\overline{x}) = 0$.

Second-Order Optimality Conditions

(Necessity) If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive semi-definite.

Proof: Use of the second-order Taylor expansion

$$f(x) = f(\overline{x}) + \nabla f(\overline{x})^T(x - \overline{x}) + \frac{1}{2}(x - \overline{x})^T \nabla^2 f(\overline{x})(x - \overline{x}) + o(\|x - \overline{x}\|^2).$$

Given $d \in \mathbb{R}^n$ and $t > 0$ plug in $x := \overline{x} + td$ to get

$$0 \le \frac{f(\bar{x} + td) - f(\bar{x})}{t^2} = \frac{1}{2} d^T \nabla^2 f(\overline{x}) d + \frac{o(t^2)}{t^2}$$

since $\nabla f(\overline{x}) = 0$. Taking the limit as $t \to 0$ gives

$$0 \le d^T \nabla^2 f(\overline{x}) d.$$

Second-Order Optimality Conditions

(Necessity) If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive semi-definite.

Proof: Use of the second-order Taylor expansion

$$f(x) = f(\overline{x}) + \nabla f(\overline{x})^T(x - \overline{x}) + \frac{1}{2}(x - \overline{x})^T \nabla^2 f(\overline{x})(x - \overline{x}) + o(\|x - \overline{x}\|^2).$$

Given $d \in \mathbb{R}^n$ and $t > 0$ plug in $x := \overline{x} + td$ to get

$$0 \leq \frac{f(\bar{x} + td) - f(\bar{x})}{t^2} = \frac{1}{2} d^T \nabla^2 f(\overline{x}) d + \frac{o(t^2)}{t^2}$$

since $\nabla f(\overline{x}) = 0$. Taking the limit as $t \to 0$ gives

$$0 \leq d^T \nabla^2 f(\overline{x}) d.$$

Since d was chosen arbitrarily, $\nabla^2 f(\overline{x})$ is positive semi-definite.

Second-Order Optimality Conditions

(Sufficiency) If $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive definite, then there is an $\alpha > 0$ such that $f(x) \geq f(\overline{x}) + \alpha\|x - \overline{x}\|^2$ for all x near $\overline{x}$.

Second-Order Optimality Conditions

(Sufficiency) If $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive definite, then there is an $\alpha > 0$ such that $f(x) \geq f(\overline{x}) + \alpha \|x - \overline{x}\|^2$ for all x near $\overline{x}$.

Proof: Use the fact that $\nabla f(\bar{x}) = 0$ to write the 2nd-order Taylor expansion of f at $\bar{x}$ as

$$f(x) = f(\overline{x}) + \frac{1}{2}(x - \overline{x})^T \nabla^2 f(\overline{x})(x - \overline{x}) + o(\|x - \overline{x}\|^2).$$

Second-Order Optimality Conditions

(Sufficiency) If $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive definite, then there is an $\alpha > 0$ such that $f(x) \geq f(\overline{x}) + \alpha\|x - \overline{x}\|^2$ for all x near $\overline{x}$.

Proof: Use the fact that $\nabla f(\bar{x}) = 0$ to write the 2nd-order Taylor expansion of f at $\bar{x}$ as

$$f(x) = f(\overline{x}) + \frac{1}{2}(x - \overline{x})^T \nabla^2 f(\overline{x})(x - \overline{x}) + o(\|x - \overline{x}\|^2).$$

Then divide through by $\|x - \bar{x}\|^2$ to get

$$\frac{f(x) - f(\overline{x})}{\|x - \overline{x}\|^2} = \frac{1}{2}\frac{(x - \overline{x})^T}{\|x - \overline{x}\|}\nabla^2 f(\overline{x})\frac{(x - \overline{x})}{\|x - \overline{x}\|} + \frac{o(\|x - \overline{x}\|^2)}{\|x - \overline{x}\|^2}.$$

Second-Order Optimality Conditions

$$\frac{f(x) - f(\bar{x})}{\|x - \bar{x}\|^2} = \frac{1}{2} \frac{(x - \bar{x})^T}{\|x - \bar{x}\|} \nabla^2 f(\bar{x}) \frac{(x - \bar{x})}{\|x - \bar{x}\|} + \frac{o(\|x - \bar{x}\|^2)}{\|x - \bar{x}\|^2}.$$

Second-Order Optimality Conditions

$$\frac{f(x)-f(\bar{x})}{\|x-\bar{x}\|^2} = \frac{1}{2}\frac{(x-\bar{x})^T}{\|x-\bar{x}\|}\nabla^2 f(\bar{x})\frac{(x-\bar{x})}{\|x-\bar{x}\|} + \frac{o(\|x-\bar{x}\|^2)}{\|x-\bar{x}\|^2}.$$

If $\lambda_{\min} > 0$ is the smallest eigenvalue of $\nabla^2 f(\bar{x})$, choose $\epsilon > 0$ so that

$$\left|\frac{o(\|x-\bar{x}\|^2)}{\|x-\bar{x}\|^2}\right| \le \frac{\lambda_{\min}}{4}$$

whenever $\|x-\bar{x}\| < \epsilon$.

Second-Order Optimality Conditions

$$\frac{f(x)-f(\bar{x})}{\|x-\bar{x}\|^2} = \frac{1}{2}\frac{(x-\bar{x})^T}{\|x-\bar{x}\|}\nabla^2 f(\bar{x})\frac{(x-\bar{x})}{\|x-\bar{x}\|} + \frac{o(\|x-\bar{x}\|^2)}{\|x-\bar{x}\|^2}.$$

If $\lambda_{\min} > 0$ is the smallest eigenvalue of $\nabla^2 f(\bar{x})$, choose $\epsilon > 0$ so that

$$\left|\frac{o(\|x-\bar{x}\|^2)}{\|x-\bar{x}\|^2}\right| \leq \frac{\lambda_{\min}}{4}$$

whenever $\|x-\bar{x}\| < \epsilon$.
Then whenever $\|x-\bar{x}\| < \epsilon$ we have

$$\begin{aligned} \frac{f(x)-f(\bar{x})}{\|x-\bar{x}\|^2} &\geq \tfrac{1}{2}\lambda_{\min} + \frac{o(\|x-\bar{x}\|^2)}{\|x-\bar{x}\|^2} \\ &\geq \tfrac{1}{4}\lambda_{\min}. \end{aligned}$$

2nd-Order Optimality Conditions

Then whenever $\|x - \bar{x}\| < \epsilon$ we have

$$\begin{aligned} \frac{f(x)-f(\bar{x})}{\|x-\bar{x}\|^2} &\geq \frac{1}{2}\lambda_{\min} + \frac{o(\|x-\bar{x}\|^2)}{\|x-\bar{x}\|^2} \\ &\geq \frac{1}{4}\lambda_{\min}. \end{aligned}$$

2nd-Order Optimality Conditions

Then whenever $\|x - \overline{x}\| < \epsilon$ we have

$$\begin{aligned} \frac{f(x)-f(\overline{x})}{\|x-\overline{x}\|^2} &\geq \frac{1}{2}\lambda_{\min} + \frac{o(\|x-\overline{x}\|^2)}{\|x-\overline{x}\|^2} \\ &\geq \frac{1}{4}\lambda_{\min}. \end{aligned}$$

Consequently, if we set $\alpha = \frac{1}{4}\lambda_{\min}$, then

$$f(x) \geq f(\overline{x}) + \alpha\|x - \overline{x}\|^2$$

whenever $\|x - \overline{x}\| < \epsilon$.