Unconstrained optimization

Paul Schrimpf

UBC
Economics 526

October 18, 2013

Section 1

Notation and definitions

Notation and definitions

- Unconstrained optimization problem

$$\max_{x \in U} F(x)$$

  - $x \in U \subseteq \mathbb{R}^n$
  - $F : U \to \mathbb{R}$.

Definition

$F^* = \max_{x \in U} F(x)$ is the **maximum** of F on U if $F(x) \leq F^*$ for all $x \in U$ and $F(x^*) = F^*$ for some $x^* \in U$

Definition

$x^* \in U$ is a **maximizer** of F on U if $F(x*) = \max_{x \in U} F(x)$. The set of all maximizers is denoted $\arg\max_{x \in U} F(x)$.

Definition

$x^* \in U$ is a **strict maximizer** of F on U if $F(x*) > F(x)$ for all $x \in U$ with $x \neq x*$.

Definition

F has a **local maximum** at x if $\exists \delta > 0$ such that $F(y) \leq F(x)$ for all $y \in N_\delta(x) \cap U$. Each such x is called a **local maximizer** of F. If $F(y) < F(x)$ for all $y \neq x$, $y \in N_\delta(x) \cap U$, then we say F has a **strict local maximum** at x.

Example

Here are some examples of functions from $\mathbb{R} \to \mathbb{R}$ and their maxima and minima.

1. $F(x) = x^2$ is minimized at $x = 0$ with minimum 0.
2. $F(x) = c$ has minimum and maximum c. Any x is a maximizer.
3. $F(x) = \cos(x)$ has maximum 1 and minimum -1. $2\pi n$ for $n \in \mathbb{Z}$ is a maximizer.
4. $F(x) = \cos(x) + x/2$ has no global maximizer or minimizer, but has many local ones.

Supremum vs maximum

- The supremum of F on U is $\sup_{x \in U} F(x) = S$ if $S \geq F(x)$ $\forall x \in U$, and for any $S' < S$, $\exists x$ s.t. $F(x) > S'$.
- When $\max_{x \in U} F(x)$ exists, $\sup_{x \in U} F(x) = \max_{x \in U} F(x)$
- Sometimes maximum does not exist, but supremum does e.g. $F(x) = x$, $U = (0, 1)$
- The infinum is to the minimum as the supremum is to the maximum

Section 2

First order conditions

First order conditions

Theorem

Let $U \subseteq \mathbb{R}^n$, $F : U \to \mathbb{R}$, and suppose F has a local maximum or minimum at x, F is differentiable at x, and $x \in \text{interior}(U)$. Then $DF_x = 0$.

Definition

Any point such that $DF_x = 0$ is call a **critical point** of F.

- F cannot have local minima or maxima (=local extrema) at interior non-critical points.
- F might have a local extrema its critical points, but it does not have to.
- e.g. $F(x) = x^3$, $F(x) = x_1^2 - x_2^2$

Section 3

Second order conditions

Second order conditions

- $F : \mathbb{R}^n \to \mathbb{R}$, x^* is a critical point
- $DF_{x^*} = 0$
- To check if x^* is a local maximum, need to look at $F(x^* + h)$ for small h
- Taylor expansion:

$$F(x^* + h) = F(x^*) + DF_{x^*} h + h^T D^2 F_{x*} h + r(x^*, h)$$

- $D^2 F_{x^*}$ is the matrix of second derivatives, called the **Hessian** of F.
- r is small, so ignore it, then $F(x^* + h) < F(x^*)$ if $h^T D^2 F_{x^*} h < 0$ for all h

Definition

Let A be a symmetric matrix, then A is

- **Negative definite** if $x^T A x < 0$ for all $x \neq 0$
- **Negative semi-definite** if $x^T A x \leq 0$ for all $x \neq 0$
- **Positive definite** if $x^T A x > 0$ for all $x \neq 0$
- **Positive semi-definite** if $x^T A x \geq 0$ for all $x \neq 0$
- **Indefinite** if $\exists$ x_1 s.t. $x_1^T A x_1 > 0$ and some other x_2 such that $x_2^T A x_2 < 0$.

Second order condition

Theorem

Let $F : U \to \mathbb{R}$ be twice continuously differentiable on U and let x^* be a critical point in the interior of U. If

1. *The Hessian, $D^2F_{x^*}$ is negative definite, then x^* is a strict local maximizer.*
2. *The Hessian, $D^2F_{x^*}$ is positive definite, then x^* is a strict local minimizer.*
3. *The Hessian, $D^2F_{x^*}$ is indefinite, x^* is neither a local min nor a local max.*
4. *The Hessian is positive or negative semi-definite, then x^* could be a local maximum, minimum, or neither.*

Proof.

$$
\begin{aligned}
F(x^*+h) - F(x^*) =& h^T D^2 F_{x^*} h + r(x^*, h) \\
=& h^T D^2 F_{x^*} h \pm h^T h \frac{r(x^*, h)}{\|h\|^2} \\
=& h^T \left(D^2 F_{x^*} \pm \frac{r(x^*, h)}{\|h\|^2} \right) h \\
\leq& h^T \left(D^2 F_{x^*} \pm \epsilon \right) h \\
\leq& t^2 (q^T D^2 F_{x*} q) + \epsilon t^2 \text{ where} tq = h, \|q\| = 1
\end{aligned}
$$

Pick $\epsilon < \inf_{\|q\|=1} |q^T D^2 F_{x*} q|$. The set $\{q : \|q\| = 1\}$ is compact so there is some q that achieves this minimum, and $q^T D^2 F_{x^*} q < 0$, so $\epsilon > 0$.

$$
\begin{aligned}
F(x^*+h) - F(x^*) <& t^2 \left((q^T D^2 F_{x*} q) + |(q^T D^2 F_{x*} q)| \right) \\
<& 0
\end{aligned}
$$

□

Example

$F : \mathbb{R} \to \mathbb{R}$, $F(x) = x^4$. The first order condition is $4x^3 = 0$, so $x^* = 0$ is the only critical point. The Hessian is $F''(x) = 12x^2 = 0$ at x^*. However, x^4 has a strict local minimum at 0.

Example

$F : \mathbb{R}^2 \to \mathbb{R}$, $F(x_1, x_2) = -x_1^2$. The first order condition is $DF_x = (-2x_1, 0) = 0$, so the $x_1^* = 0$, $x_2^* \in \mathbb{R}$ are all critical points. The Hessian is

$$D^2 F_x = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}$$

This is negative semi-definite because $h^T D^2 F_x h = -2h_1^2 \leq 0$. Also, graphing the function would make it clear that $x_1^* = 0$, $x_2^* \in \mathbb{R}$ are all (non-strict) local maxima.

Example

$F : \mathbb{R}^2 \to \mathbb{R}$, $F(x_1, x_2) = -x_1^2 + x_2^4$. The first order condition is $DF_x = (-2x_1, 4x_2^3) = 0$, so the $x^* = (0, 0)$ is a critical point. The Hessian is

$$D^2 F_x = \begin{pmatrix} -2 & 0 \\ 0 & 12x_2^2 \end{pmatrix}$$

This is negative semi-definite at 0 because $h^T D^2 F_0 h = -2h_1^2 \leq 0$. However, 0 is not a local maximum because $F(0, x_2) > F(0, 0)$ for any $x_2 \neq 0$. 0 is also not a local minimum because $F(x_1, 0) < F(0, 0)$ for all $x_1 \neq 0$.

Theorem

Let $F : U \to \mathbb{R}$ be twice continuously differentiable on U and let x^* in the interior of U be a local maximizer (or minimizer) of F. Then $DF_{x^*} = 0$ and $D^2 F_{x^*}$ is negative (or positive) semi-definite.

Proof.

Using the same notation and setup as in the proof of theorem 9,

$$
\begin{aligned}
0 > F(x^* + tq) - F(x^*) =& t^2 q^T D^2 F_{x^*} q + r(x, tq) \\
0 >& t^2 \left(q^T D^2 F_{x^*} q + \frac{r(x, tq)}{t^2} \right)
\end{aligned}
$$

Because $\lim_{t \to 0} \frac{r(x,tq)}{t^2} = 0$, for any $\epsilon > 0$ $\exists \delta > 0$ such that if $|t| < \delta$, then

$$
\begin{aligned}
0 >& t^2 \left(q^T D^2 F_{x^*} q + \frac{r(x, tq)}{t^2} \right) \geq t^2 \left(q^T D^2 F_{x^*} q - \epsilon \right) \\
t^2 \epsilon >& t^2 q^T D^2 F_{x^*} q
\end{aligned}
$$

Section 4

Definite matrices

Definite matrices

- Second order condition depends on checking whether x^TAx is always positive or negative for symmetric A
- 1 by 1: $x^TAx = ax^2$, negative definite if $a < 0$
- 2 by 2:

$$x^TAx = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= ax_1^2 + 2bx_1x_2 + cx_2^2 = \qquad a\left(x_1 + \frac{2b}{a}x_1x_2 + \frac{b^2}{a^2}x\right.$$

$$= a\left(x_1 + \frac{b}{a}x_2\right)^2 + \frac{ac - b^2}{a}x_2^2$$

Thus $x^TAx < 0$ for all $x \neq 0$ if $a < 0$ and $\det A = ac - b^2 > 0$

Definition

Let A by an n by n matrix. The k by k submatrix

$$A_k = \begin{pmatrix} a_{11} & \cdots a_{1k} & \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}$$

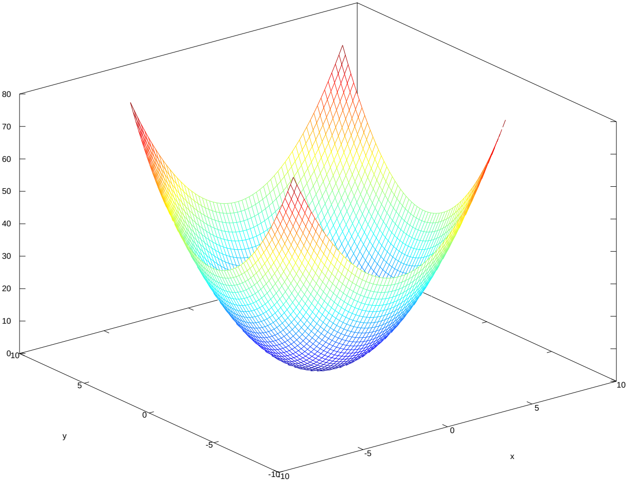
is the k**th leading principal submatrix** of A. The determinant of A_k is the kth order **leading principal minor** of A.

Theorem

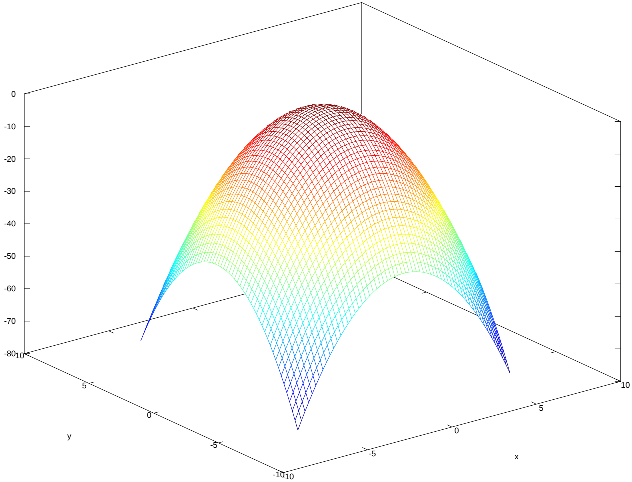
Let A be an n by n symmetric matrix. Then

1. *A is positive definite if and only if all n of its leading principal minors are strictly positive.*
2. *A is positive semi-definite if and only if all n of its leading principal minors are weakly positive.*
3. *A is negative definite if and only if all n of its leading principal minors alternate in sign as follows: $\det A_1 < 0$, $\det A_2 > 0$, $\det A_3 < 0$, etc.*
4. *A is negative semi-definite if and only if all n of its leading principal minors weakly alternate in sign as follows: $\det A_1 \leq 0$, $\det A_2 \geq 0$, $\det A_3 \leq 0$, etc*
5. *A is indefinite if and only if none of the five above cases hold, and $\det A_k \neq 0$ for at least one k.*

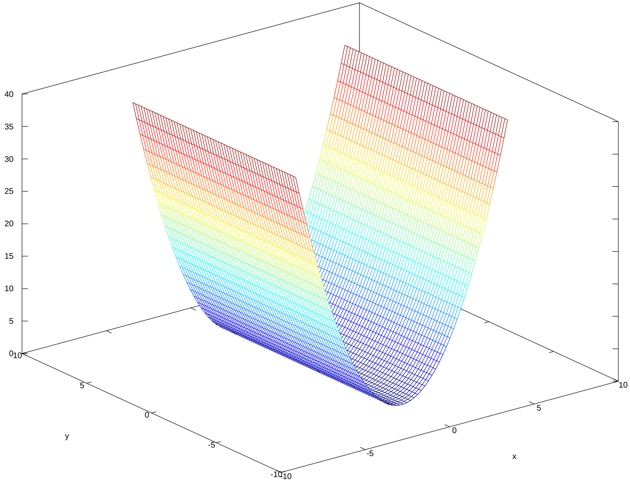
Positive definite



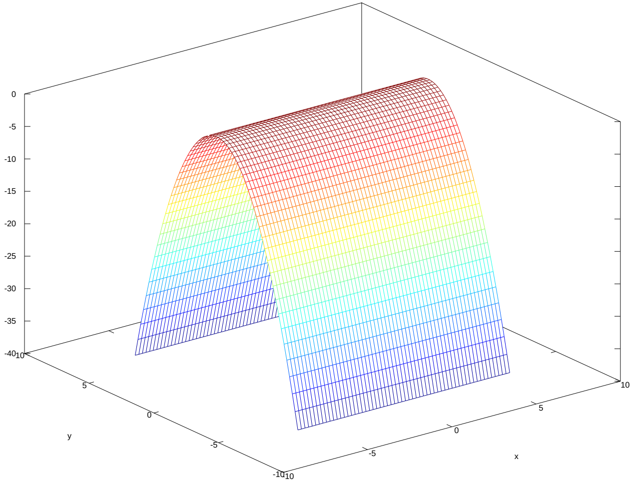
Negative definite



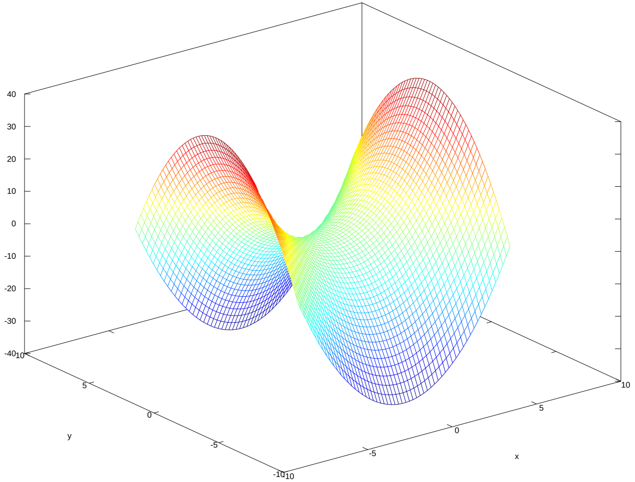
Positive semi-definite



Negative semi-definite



Indefinite



Eigenvalues and eigenvectors

Definition

If A is an n by n matrix, λ is a scalar, $v \in \mathbb{R}^n$ with $\|v\| = 1$, and

$$Av = \lambda v$$

then λ is a **eigenvalue** of A and v is an **eigenvector**.

Lemma

Let A be an n by n matrix and λ a scalar. Each of the following are equivalent.

1. *λ is a eigenvalue of A.*
2. *$A - \lambda I$ is singular.*
3. $\det(A - \lambda I) = 0$.

Proof.

We know that (2) and (3) from our results on systems of linear equations and matrices. Also, if $A - \lambda I$ is singular, then the null space of $A - \lambda I$ contains non-zero vectors. Choose $v \in \mathcal{N}(A - \lambda I)$ such that $v \neq 0$. Then $(A - \lambda I)v = 0$, so $A(v/\|v\|) = \lambda(v/\|v\|)$ as in the definition of eigenvalues. □

- $\chi_A : \mathbb{R} \to \mathbb{R}$ defined by $\chi_A(x) = \det(A - xI)$ is called the **characteristic polynomial** of A
  - Polynomial of order n.
  - n roots, possibly not real and not distinct
  - For symmetric nonsingular matrices, the roots are real and distinct

Lemma

Let A be an n by n matrix with n distinct real eigenvalues then the eigenvectors of A are linearly independent.

- So if A has n distinct eigenvalues, it has n linearly independent eigenvectors.
- Can make the eigenvectors orthogonal
- Make matrix V with columns composed of the orthogonal eigenvectors, then V is orthogonal, so $V^{-1} = V^T$.
- Definition of eigenvalues implies

$$AV = V\Lambda$$

where Λ is diagonal matrix of eigenvalues.

Theorem (Eigendecomposition)

Let A be an n by n non-singular symmetric matrix, then A has n distinct real eigenvalues

$$A = V\Lambda V^T$$

where Λ is the diagonal matrix consisting of the eigenvalues of A and the columns of V are the eigenvectors of A, and V is an orthogonal matrix.

Theorem

If A is an n by n symmetric non-singular matrix with eigenvalues $\lambda_1, ..., \lambda_n$, then

1. $\lambda_i > 0$ *for all* i, *iff* A *is positive definite,*
2. $\lambda_i \geq 0$ *for all* i, *iff* A *is positive semi-definite,*
3. $\lambda_i < 0$ *for all* i, *iff* A *is negative definite,*
4. $\lambda_i \leq 0$ *for all* i, *iff* A *is negative semi-definite,*
5. *if some* $\lambda_i > 0$ *and some* $\lambda_j < 0$, *then* A *is indefinite.*

Proof.

Let $A = V \Lambda V^T$ be the eigendecomposition of A. Let $x \neq 0 \in \mathbb{R}^n$. Then $x^T V = z^T \neq 0$ because V is nonsingular. Also,

$$x^T A x = x^T V \Lambda V^T x = z^T \Lambda z.$$

Since Λ is diagonal we have

$$z^T \Lambda z = z_1^2 \lambda_1 + ... + z_n^2 \lambda_n$$

Thus, $x^T A x = z^T \Lambda z > 0$ for all x iff $\lambda_i > 0$ for each i. The other parts of the theorem follow similarly. □

Section 5

Global maximum and minimum

Global maximum and minimum

- First and second order conditions give nice way of finding local maxima
- Global maximum: find all interior local maxima and compare them with each other and the value of F on the boundary
- Tedious if many local maxima or boundary of U is complicated
- Generally, no nice necessary and sufficient condition for global maximum
- One sufficient condition is that concave functions on convex sets have a unique global maximum

Definition

Let $f : U \to \mathbb{R}$. f is **convex** if for all $x, y \in U$ with $\ell(x, y) \subseteq U$ we have $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$ for all $t \in [0, 1]$.

Definition

Let $f : U \to \mathbb{R}$. f is **concave** if for all $x, y \in U$ with $\ell(x, y) \subseteq U$ we have $f(tx + (1-t)y) \geq tf(x) + (1-t)f(y)$ for all $t \in [0, 1]$.

Theorem

Let $F : U \to \mathbb{R}$ be twice continuously differentiable and $U \subseteq \mathbb{R}^n$ be convex. Then,

1. *The following three conditions are equivalent:*
   1. *F is a concave function on U*
   2. *$F(y) - F(x) \leq DF_x(y - x)$ for all $x, y \in U$,*
   3. *D^2F_x is negative semi-definite for all $x, y \in U$*
2. *If F is a concave function on U and $DF_{x^*} = 0$ for some $x^* \in U$, then x^* is the global maximizer of F on U.*