Unconstrained Optimization in Euclidean Space

Some Notation.

- Given any two vectors, $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, we write

$$\begin{array}{ll} x = y, & \text{if } x_i = y_i,\ i = 1, \ldots, n; \\ x \geq y, & \text{if } x_i \geq y_i,\ i = 1, \ldots, n; \\ x > y, & \text{if } x \geq y \text{ and } x \neq y; \\ x \gg y, & \text{if } x_i > y_i,\ i = 1, \ldots, n. \end{array}$$

- Note that
  - $x \geq y$ does not exclude the possibility that $x = y$, and
  - for $n > 1$, the vectors x and y may not be comparable under any of the categories above. For example, the vectors $x = (2, 1)$ and $y = (1, 2)$ in $\mathbb{R}^2$ do not satisfy $x \geq y$ or $y \geq x$.

- The *nonnegative* and *positive orthants* of $\mathbb{R}^n$, denoted by $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ respectively, are defined as

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$$

  and

$$\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n \mid x \gg 0\}.$$

Optimization Problems.

- First we introduce the notation used to represent abstract optimization problems and their solutions.

Definition. An *optimization problem (in $\mathbb{R}^n$)* is one where the values of a given function $f : \mathbb{R}^n \to \mathbb{R}$ are to be maximized or minimized over a given set $\mathcal{D} \subseteq \mathbb{R}^n$. The function f is called the *objective function*, and the set $\mathcal{D}$ the *constraint* or *feasible set*.

- Notationally, we represent a *maximation problem* by

$$\text{Maximize } f(x) \text{ subject to } x \in \mathcal{D},$$

  or, more compactly,

$$\max\{f(x) \mid x \in \mathcal{D}\}.$$

- Similarly, for a *minimization problem*, we write

$$\text{Minimize } f(x) \text{ subject to } x \in \mathcal{D},$$

  or

$$\min\{f(x) \mid x \in \mathcal{D}\}.$$

- A *solution* to a maximization problem is a point $x \in \mathcal{D}$ such that

$$f(x) \geq f(y) \text{ for all } y \in \mathcal{D}.$$

  We say that f *attains a maximum* on $\mathcal{D}$ at x, and refer to x as a *maximizer* of f on $\mathcal{D}$.

- Simlarly, a solution to a minimizaton problem is a point $x \in \mathcal{D}$ such that

$$f(x) \leq f(y) \text{ for all } y \in \mathcal{D}.$$

  We say that f *attains a minimum* on $\mathcal{D}$ at x, and refer to x as a *minimizer* of f on $\mathcal{D}$.

- The set of *attainable values* of f on $\mathcal{D}$, $f(\mathcal{D})$, is defined by

$$f(\mathcal{D}) = \{w \in \mathbb{R} \mid \ f(x) = w \text{ for some } x \in \mathcal{D}\}$$

  We also refer to $f(\mathcal{D})$ as the *image* of $\mathcal{D}$ under f. ▲

- The following examples show that for a given optimization problem a solution may not exist, and even if a solution exists it need no be unique.

Example 1.

1. Let $\mathcal{D} = \mathbb{R}_+$ and $f(x) = x$ for all $x \in \mathcal{D}$. Then $f(\mathcal{D}) = \mathbb{R}_+$ and $\sup f(\mathcal{D}) = +\infty$, so the problem $\max\{f(x) \mid x \in \mathcal{D}\}$ has no solution.
2. Let $\mathcal{D} = [0, 1]$ and $f(x) = x(1 - x)$ for all $x \in \mathcal{D}$. Then the problem of maximizing f on $\mathcal{D}$ has exactly one solution, $x = 1/2$.
3. Let $\mathcal{D} = [-1, 1]$ and $f(x) = x^2$ for all $x \in \mathcal{D}$. Then the problem of maximizing f on $\mathcal{D}$ has two solutions, $x = -1$ and $x = 1$. ⧫

- Now we introduce some terminology to refer to the sets of maximizers and minimizers.

Definition. The set of all maximizers of f on $\mathcal{D}$, denoted by $\arg\max\{f(x) \mid x \in \mathcal{D}\}$, is the set given by

$$\arg\max\{f(x) \mid x \in \mathcal{D}\} = \{x \in \mathcal{D} \mid f(x) \geq f(y) \text{ for all } y \in \mathcal{D}\}.$$

Similarly, the set of all minimizers of f on $\mathcal{D}$, denoted by $\arg\min\{f(x) \mid x \in \mathcal{D}\}$, is the set given by

$$\arg\min\{f(x) \mid x \in \mathcal{D}\} = \{x \in \mathcal{D} \mid f(x) \leq f(y) \text{ for all } y \in \mathcal{D}\}.$$ ▲

Example 2. In part (1) of the previous example, the set $\arg\max\{f(x) \mid x \in \mathcal{D}\}$ is $\emptyset$. In part (2) this set is $\{1/2\}$, and in part (3) this set is $\{-1, 1\}$. ⧫

- The following result allows us to apply all the results about maxima to minima.

Theorem 1. *Let $-f$ denote the function given by $-f(x)$ for all $x \in \mathcal{D}$. Then*

1. *x is a maximizer of f on $\mathcal{D}$ iff x is a minimizer of $-f$ on $\mathcal{D}$;*
2. *x is a minimizer of f on $\mathcal{D}$ iff x is a maximizer of $-f$ on $\mathcal{D}$.*

- The following theorem says a strictly increasing transformation of a function does not change the optimum points.

Theorem 2. *Let g be a strictly increasing function. Then*

1. *x is a maximizer of f on $\mathcal{D}$ iff x is a maximizer of $g \circ f$ on $\mathcal{D}$;*
2. *x is a minimizer of f on $\mathcal{D}$ iff x is a minimizer of $g \circ f$ on $\mathcal{D}$.*

Example 3.

1. Consider the model in consumer theory in which a single agent consumes n commodities in nonnegative quantities.
   - The agent's utility from consuming $x_i \geq 0$ units of commodity i ($i = 1, \ldots, n$) is given by $u(x_1, \ldots, x_n)$ where $u : \mathbb{R}^n \to \mathbb{R}$ is the agent's utility function.
   - The agent has an income $I \geq 0$, and faces the price vector $p = (p_1, \ldots, p_n)$, where $p_i \geq 0$ denotes the price per unit of the ith commodity.
   - The agent's budget set – the set of feasible or affordable consumption bundles, given income and prices – denoted by $\mathcal{B}(p, I)$, is given by

$$\mathcal{B}(p, I) = \{x \in \mathbb{R}^n_+ \mid p \cdot x \leq I\}.$$

1. - The agent's objective is to maximize the level of utility over the set of affordable commodity bundles, i.e. to solve

$$\text{Maximize } u(x) \text{ subject to } x \in \mathcal{B}(p, I).$$

2. An agent's expenditure minimization problem is dual to the utility maximization problem.
   - The agent's problem is to minimize the amount of income needed to give a utility level of at least $\overline{u}$, where $\overline{u}$ is some fixed utility level, given a price vector $p \in \mathbb{R}^n_+$.
   - In this problem

$$X(\overline{u}) = \{x \in \mathbb{R}^n_+ \mid u(x) \geq \overline{u}\}$$

   is the constraint set and

$$\text{Minimize } p \cdot x \text{ subject to } x \in X(\overline{u})$$

   is the objective is to solve. ⧫

The Weierstrass Theorem.

- We would like to know under what conditions on f and $\mathcal{D}$ we can guarantee the existence of optima.
- The idea behind the Weierstrass theorem is as follows.
  - If the constraint set $\mathcal{D}$ is a compact set and the function f is continuous, then the image $f(\mathcal{D})$ is also a compact set.
  - Since any closed and bounded set attains a maximum and a minimum, f must attain a maximum and minimum on $\mathcal{D}$.

Theorem 3 (Weierstrass Theorem)**.** *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be compact and let $f : \mathcal{D} \to \mathbb{R}$ be a continuous function on $\mathcal{D}$. Then f attains a maximum and a minimum on $\mathcal{D}$; that is, there exist points $z_1, z_2 \in \mathcal{D}$ such that $f(z_1) \geq f(x) \geq f(z_2)$ for all $x \in \mathcal{D}$.*

- The Weierstrass theorem only provides sufficient conditions for the existence of optima.
- As we shall see in the following examples we cannot say anything, in general, about what happens if these conditions are not satisfied.

Example 4.

1. Let $\mathcal{D} = \mathbb{R}$ and $f(x) = x^3$ for all $x \in \mathbb{R}$. Then f is continuous, but $\mathcal{D}$ is not compact (it is closed but not bounded). Since $f(\mathcal{D}) = \mathbb{R}$, we can see f attains neither a maximum nor a minimum on $\mathcal{D}$.

2. Let $\mathcal{D} = (0, 1)$ and $f(x) = x$ for all $x \in (0, 1)$. Then f is continuous, but $\mathcal{D}$ is not compact (it is bounded but not closed). The set $f(\mathcal{D})$ is the open interval $(0, 1)$ and so, again, f attains neither a maximum nor a minimum on $\mathcal{D}$

3. Let $\mathcal{D} = [-1, 1]$ and let f be given by

$$f(x) = \begin{cases} 0, & \text{if } x = -1 \text{ or } x = 1 \\ x, & \text{if } -1 < x < 1. \end{cases}$$

   Here $\mathcal{D}$ is compact, but f is not continuous at the points -1 and 1. In this case, $f(\mathcal{D})$ is the open interval $(-1, 1)$. Thus f fails to attain either a maximum or a minimum on $\mathcal{D}$.

4. Let $\mathcal{D} = \mathbb{R}_{++}$ and let $f : \mathcal{D} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0, & \text{otherwise.} \end{cases}$$

   Then $\mathcal{D}$ is not compact, and f is discontinuous at every point in $\mathbb{R}_{++}$. But, f does attain a maximum (at every rational number) and a minimum (at every irrational number). ⧫

- We now apply the Weierstrass theorem to the consumer's utility maximization problem.

Example 5. Again, consider the utility maximization problem

$$\text{Maximize } u(x) \text{ subject to } x \in \mathcal{B}(p, I) = \{x \in \mathbb{R}^n_+ \mid p \cdot x \leq I\}.$$

- Assume that the utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ is continuous on its domain.

- By the Wierstrass theorem, a solution to the utility maximization problem will always exist provided the budget set $\mathcal{B}(p, I)$ is compact.

- We "show" that compactness holds if $p \ll 0$.

- Notice that even if the agent spent all the income I on commodity j, consumption of this commodity cannot exceed I/p_j. Let

$$\psi = \max \left\{ \frac{I}{p_1}, \ldots, \frac{I}{p_n} \right\}.$$

  Then, for any $x \in \mathcal{B}(p, I)$, $x_i \leq \psi$ for all $i \in \{1, \ldots, n\}$, and we have $0 \leq x \leq (\psi, \ldots, \psi)$. Thus $\mathcal{B}(p, I)$ is bounded.

- A longer argument shows that $\mathcal{B}(p, I)$ is also closed (try showing this!). Thus, the budget set is compact and so a solution to the utility maximization problem exists. ♦

Unconstrained Optima.

- We now look at identifying *necessary* conditions that the derivative of f must satisfy at an optimum.

- We begin by looking at what are called unconstrained or interior optima. That is, optima at which the constraints are not binding.

Definition. Let f be a function on $\mathcal{D} \subseteq \mathbb{R}^n$.

- A point $x \in \mathcal{D}$ is an *unconstrained (global) maximizer* of f on $\mathcal{D}$ if $x \in \operatorname{int} \mathcal{D}$ and $f(x) \geq f(y)$ for all $y \in \mathcal{D}$.

- A point $x \in \mathcal{D}$ is a *local maximizer* of f on $\mathcal{D}$ if there is some $r > 0$ such that $f(x) \geq f(y)$ for all $y \in \mathcal{D} \cap B_r(x)$.

- A point $x \in \mathcal{D}$ is an *unconstrained local maximizer* of f on $\mathcal{D}$ if there is some $r > 0$ such that $B_r(x) \subseteq \mathcal{D}$ and $f(x) \geq f(y)$ for all $y \in B_r(x)$. ▲

- Similar definitions apply to local and global minimizers of f on $\mathcal{D}$.

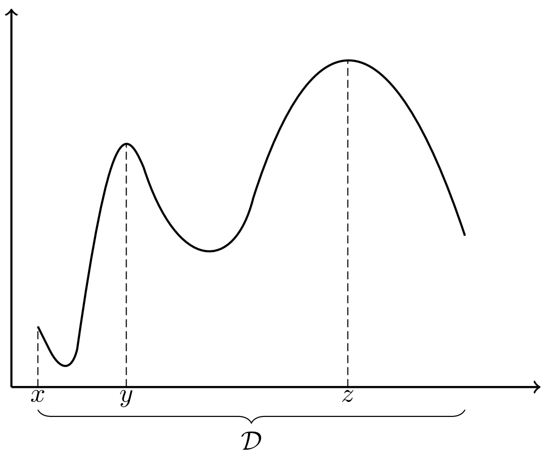

Figure 1: The points x, y, and z are local maximizers of f on $\mathcal{D}$; points y and z are unconstrained local maximizers of f on $\mathcal{D}$; while the point z is a (global) unconstrained maximizer of f on $\mathcal{D}$.

First Order Conditions.

- Our first result states that the derivative of f must be zero at every unconstrained local maximum or minimum.
- The result is easy to see in the one-dimensional case. Suppose x is, say, a local maximum and $f'(x) \neq 0$.
  - Then if $f'(x) > 0$, it would be possible to increase the value of f by moving to the right a small amount.
  - If $f'(x) < 0$, the value of f could be increased by moving slightly to the left.
  - At an *unconstrained* maximizer movement in either direction is feasible.
  - But then the definition of a local maximum would be violated.

Theorem 4. *Let $x \in \operatorname{int} \mathcal{D} \subseteq \mathbb{R}^n$ be a local maximizer of f on $\mathcal{D}$ or local minimizer of f on $\mathcal{D}$. Then $Df(x) = 0$.*

First Order Conditions.

- This theorem only provides a necessary condition for an unconstrained local optimum. It is *not* sufficient.
- The theorem does *not* say that if $Df(x) = 0$ for some x in the interior of the feasible set, then x must be an unconstrained local optimizer. The following example illustrates this fact.

Example 6. Let $\mathcal{D} = \mathbb{R}$, and let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$ for all $x \in \mathbb{R}$.

- Then, we have $f'(0) = 0$ but 0 is neither a local maximizer nor a local minimizer of f on $\mathbb{R}$.

- Any open ball about 0 must contain a point $x > 0$ and a point $z < 0$. At $x > 0$, we have $f(x) = x^3 > 0 = f(0)$ while at $z > 0$ we have $f(z) = z^3 < 0 = f(0)$. ♦

Definition. We call a point $x \in \mathcal{D}$ that satisfies the first order condition $Df(x) = 0$ a *critical point* of f on $\mathcal{D}$. ▲

- So, the previous theorem says that every unconstrained local optimizer must be a critical point.

- But as we saw in our example, there may exist critical points that are neither local maximizers nor local minimizers.

- The first order conditions for unconstrained local optima do not distinguish between maxima and minima.

- To do so we need to look at the behaviour of the second derivative, i.e. D^2f (the Hessian of f).

Second Order Conditions.

- Just as we look at a linear approximation when we want to say something about the slope of a function, we look at a quadratic approximation when we want to say something about the curvature of a function.

- A (very) sketch(y) proof of the second order conditions we will see later is as follows.

- Let x be a critical point of f. We can represent the value of a C^2 function, at a point $x + h$ close to x, as the sum of the Taylor polynomial of order two about x and the remainder term $R(h)$:

  $$f(x + h) = f(x) + Df(x)h + \frac{1}{2}h^T D^2 f(x)h + R(h).$$

  As h goes to zero, the remainder term goes to 0 very quickly, so we will ignore this negligibly small term.

- Now, since x is a critical point, the derivative of f at x i.e. $Df(x)$ is zero. Thus

  $$f(x + h) - f(x) \approx \frac{1}{2}h^T D^2 f(x)h.$$

  for small h.

- Thus if $h^T D^2 f(x)h < 0$, it follows that for all small $h \neq 0$, we have

  $$f(x + h) < f(x).$$

  Hence, x is a (strict) local maximizer of f.

- We say that the $n \times n$ Hessian matrix $D^2 f(x)$ is negative definite if $y^T D^2 f(x) y < 0$ for all $y \in \mathbb{R}^n$ with $y \neq 0$.
- Thus, the definiteness, or otherwise, of $D^2 f(x)$ will provide the conditions we are after.
- So before we present the second order conditions, we need to look at the definiteness of general quadratic forms.

Definiteness of Quadratic Forms.

- We now take a detour to look at the definiteness of quadratic forms.

Definition. Let A be an $n \times n$ symmetric matrix. Then A is said to be

- *positive definite* if $x^T Ax > 0$ for all x in $\mathbb{R}^n$, $x \neq 0$;
- *positive semidefinite* if $x^T Ax \geq 0$ for all x in $\mathbb{R}^n$;
- *negative definite* if $x^T Ax < 0$ for all x in $\mathbb{R}^n$, $x \neq 0$;
- *negative semidefinite* if $x^T Ax \leq 0$ for all x in $\mathbb{R}^n$;
- *indefinite* if $x^T Ax > 0$ and $y^T Ay < 0$ for some x and y in $\mathbb{R}^n$. ▲

- Note that a matrix that is positive definite (negative definite) is always positive semidefinite (negative semidefinite).
- Also, notice that A is positive (semi)definite iff the matrix $-A$ is negative (semi)definite.

Example 7.

1. The quadratic form A defined by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is positive definite, since for any $(x_1, x_2) \in \mathbb{R}^2$, we have $x^T Ax = x_1^2 + x_2^2$, and this is positive whenever $x \neq 0$.

2. Consider the quadratic form A defined by

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

For any $(x_1, x_2) \in \mathbb{R}^2$, we have $x^T Ax = x_1^2$, which can be zero even if $x \neq 0$. Hence A is not positive definite. However, it is true that we always have $x^T Ax \geq 0$, so A is positive semidefinite.

3. As an example of an indefinite quadratic form, consider

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

For any $(x_1, x_2) \in \mathbb{R}^2$, we have $x^T Ax = 2x_1x_2$. When $x = (1, 1)$, $x^T Ax = 2 > 0$, so A is not negative semidefinite. But when $x = (-1, 1)$, $x^T Ax = -2 < 0$, so that A is not positive semidefinite either. ♦

- But, in practice, how do we determine the definiteness of a matrix?
- To answer this question, we need some more terminology.

Definition. Let A be an $n \times n$ matrix.

- The *kth order leading principal submatrix* of A, denoted by A_k is the submatrix obtained by retaining only the first k rows and columns.
- The *kth order leading principal minor* of A, denoted by $|A_k|$ is the determinant of the kth order leading principal submatrix of A. ▲

Theorem 5. *Let A be an $n \times n$ symmetric matrix. Then*

1. *A is positive definite iff $|A_k| > 0$ for all $k \in \{1, \ldots, n\}$;*
2. *A is negative definite iff $(-1)^k |A_k| > 0$ for all $k \in \{1, \ldots, n\}$.*

Furthermore, a positive (negative) semidefinite matrix A is positive (negative) definite iff $|A| \neq 0$.

- So a symmetric matrix is positive definite iff all its leading principal minors are positive.
- A symmetric matrix is negative definite iff its leading principal minors alternate in sign, starting with negative, i.e. $|A_1| < 0$, $|A_2| > 0$, $|A_3| < 0$, etc.

Definition. Let A be an $n \times n$ matrix.

- A *kth order principal submatrix* of A is a submatrix obtained by retaining the same k rows and columns.
- A *kth order principal minor* of A is a determinant of a kth order principal submatrix of A. ▲

Example 8. Consider the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

The principal submatrices of order 1 are (1), (5), and (9). The first of these is the the leading principal submatrix A_1. The principal submatrices of order two are

$$\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix},$$

with the first being the leading second order principal submatrix A_2. The third order (leading) principal submatrix is $A_3 = A$ itself. ♦

Theorem 6. *Let A be an $n \times n$ symmetric matrix. Then*

1. *A is positive semidefinite iff every principal minor of A is nonnegative;*
2. *A is negative semidefinite iff every principal minor of odd order is nonpositive and every principal minor of even order is nonnegative.*

- In our example above, there were $1+3+3 = 7$ principal submatrices. In general there will be

  $$\sum_{k=1}^{n} \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n-1} + \binom{n}{n} = 2^n - 1$$

  principal submatrices of an $n \times n$ matrix.

Second Order Conditions.

- Before we state the second order conditions, another definition.

Definition. A point $x \in \mathcal{D}$ is a *strict local maximizer* of f on $\mathcal{D}$ if there is some $r > 0$ such that $f(x) > f(y)$ for all $y \in \mathcal{D} \cap B_r(x)$, $y \neq x$. ▲

Theorem 7. *Let f be a C^2 function on $\mathcal{D} \subseteq \mathbb{R}^n$ and let $x \in \operatorname{int} D$.*

1. *If f has a local maximum at x, then $D^2 f(x)$ is negative semidefinite.*
2. *If f has a local minimum at x, then $D^2 f(x)$ is positive semidefinite.*
3. *If $Df(x) = 0$ and $D^2 f(x)$ is negative definite at x, then x is a strict local maximizer of f on $\mathcal{D}$.*
4. *If $Df(x) = 0$ and $D^2 f(x)$ is positive definite at x, then x is a strict local minimizer of f on $\mathcal{D}$.*

- This first two parts of the theorem give necessary conditions which the second derivative $D^2 f$ must satisfy at interior local maxima and minima.
  - Parts (1) and (2) of the theorem therefore allow us to rule out from contention as optima any interior critical points at which the Hessian is indefinite.
  - Such points, which are maximizers of f in some directions and minimizers of f in other directions are called saddle points.

- The second two parts give *sufficient* conditions on the second derivative that identify specific points as being local optimizers.
- However the sufficient conditions, requiring definiteness at the optimal point, are different to the necessary conditions which require semidefiniteness.
- This can be problematic when we come to apply the theorem.
- Suppose that x is a critical point at which $D^2 f(x)$ is positive semidefinite (but not positive definite).
  - Part (3) of the theorem does not allow us to conclude that x must be a local minimizer.
  - Neither can we use part (2) to rule out the possibility of x being a local minimizer, because we only need $D^2 f(x)$ to be positive semidefinite at a local maximizer.
- The next example shows that we cannot strengthen parts (1) and (2) of the theorem to say
  - $D^2 f$ must be negative (positive) definite at a local maximizer (minimizer) x.

Example 9. Let $\mathcal{D} = \mathbb{R}$, and let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be defined, respectively, by $f(x) = x^4$ and $g(x) = -x^4$ for all $x \in \mathbb{R}$.

- Clearly, since $x^4 \geq 0$ everywhere and $f(0) = g(0) = 0$, 0 is a global minimum of f on $\mathcal{D}$ and a global maximum of g on $\mathcal{D}$.
- However, $f''(0) = g''(0) = 0$, so $f''(0)$ is positive semidefinite but not positive definite, while $g''(0)$ is negative semidefinite, but not negative definite. ♦
- Now we show by example that we cannot strengthen parts (3) and (4) of the theorem to say
  - If $Df(x) = 0$ and $D^2 f(x)$ is negative (positive) semidefinite, then x is a local maximizer (minimizer) of f on $\mathcal{D}$.

Example 10. Let $\mathcal{D} = \mathbb{R}$, and let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$ for all $x \in \mathbb{R}$.

- Then $f'(x) = 3x^2$ and $f''(x) = 6x$, so $f'(0) = f''(0) = 0$ and so $f''(0)$ is both positive semidefinite and negative semidefinite (but not positive or negative definite).
- However 0 is neither a local maximizer nor a local minimizer of f on $\mathcal{D}$. ♦

Using the First and Second Order Conditions.

- The first order condition on its own is of limited use in finding solutions to optimization problems for two reasons.
  - The condition only applies to cases where the optimum occurs at an interior point, but in most applications some or all constraints will bind.
  - The condition is only a *necessary* one.
- Even when we combine the first order condition with the second order conditions, we do not gain a lot.
  - First, parts (1) and (2)are only *necessary* conditions.
  - Second, although parts (3) and (4) give *sufficient* conditions, they are sufficient for all unconstrained local optima, and not just global optima.
- At best, the second order conditions help determine if a point is a *local* optimum, they are of no use in determining if a local optimum is a global optimum or not.
- Global optima may fail to exist even if several critical points, including local optima exist.

Example 11. Let $\mathcal{D} = \mathbb{R}$ and let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 2x^3 - 3x^2$.

- First, f is a C^2 function on $\mathbb{R}$.
- There are exactly two critical points at $x = 0$ and $x = 1$.
- We find that $f''(0) = -6 < 0$ while $f''(1) = 6 > 0$.
- Thus the point 0 is a strict local maximizer of f on $\mathcal{D}$, while 1 is a strict local minimizer of f on $\mathcal{D}$.
- However, the first and second order conditions do not help determine whether these are global optima. Clearly they are not – there are no global optima in this problem, since $\lim_{x\to+\infty} f(x) = +\infty$ and $\lim_{x\to-\infty} f(x) = -\infty$. ⧫

- On the other hand, in problems in which we know beforehand that a global optimum exists, we can sometimes use the first order conditions to compute the solution.
- For example, say we know there is an unconstrained solution.
  - In this case the set of all points that meet the first order conditions must contain the optimum point, and the problem is reduced to finding the point that optimizes the objective function over this set.
- More generally, if we do not know that the optimum is in the interior of the set, we can modify the above procedure as long as the set of boundary points of the constraint set is small.

- The optimum must either occur on the boundary or in the interior of the constraint set. In the latter case the first order conditions must be satisfied.
- Thus, to identify the optimum, simply compare the optimum value of the objective function on the boundary with the value of the function at those points in the interior that satisfy the first order conditions.

Example 12. Consider the problem of maximizing $f(x) = 4x^3 - 5x^2 + 2x$ over the unit interval $[0, 1]$.

- Since $[0, 1]$ is compact, and f is continuous on this interval, the Weierstrass theorem tells us that f has a (global) maximum on this interval.
- Either the maximum occurs at one of the boundary points 0 or 1, or it is an unconstrained maximum.
- If it is an unconstrained maximum, it must meet the first order condition $f'(x) = 12x^2 - 10x + 2 = 0$. The only points satisfying this condition are $x = 1/2$ and $x = 1/3$.
- Evaluating f at the four points $0, 1/3, 1/2$ and 1 shows that $x = 1$ is the point where f attains its maximum on $[0, 1]$. Simlarly we can show that $x = 0$ is the point where f is minimized on $[0, 1]$. ⧫

- In most economic applications, the set of boundary points is large and so carrying out the comparisons, as in the previous example, is a nontrivial task.
- For example, in the utility maximization problem, the entire line segment

$$\{x \in \mathbb{R}^n \mid p \cdot x = I, x \geq 0\},$$

  is part of the boundary of the budget set $\mathcal{B}(p, I)$.
- Still, the procedure used in the example indicates how *global* optima may be identified by combining knowledge of the existence of a solution with first order necessary conditions for *local* optima.
- As we saw, if a solution is known to exist, it can be identified by simply comparing the value of the objective function at points on the boundary and at points satisfying the first order condtions.
- The only role second order conditions can play is to cut down on the number of points at which the comparison has to be carried out.
- If, say, we are seeking a maximum of f then all critical points that also satisfy the second order conditions for a strict local minimum can be ruled out.
- The most that second order conditions can do is help identify *local* optima.