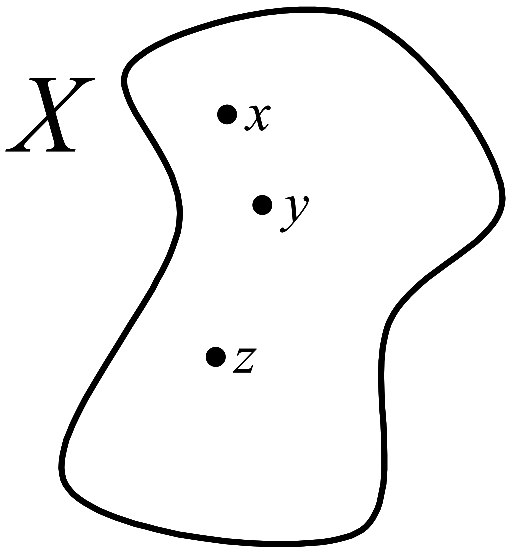
# 21 Manove's Mathematics for Micro

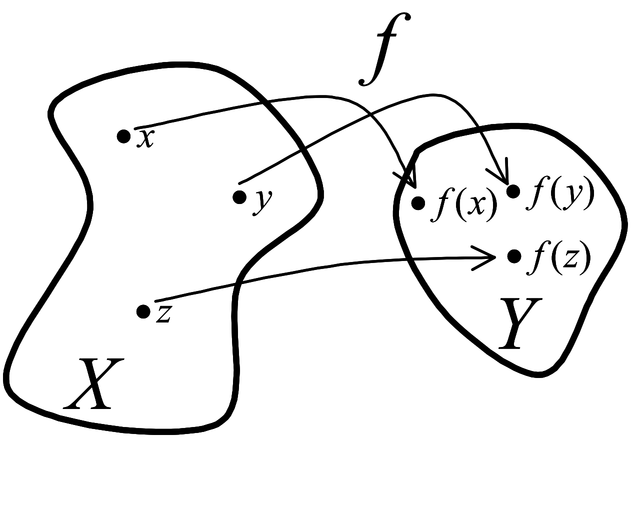
**We will review these notes from in class from time to time. Students will not be directly examined on this material, but some of it will be important for understanding other parts of the course.**

## 21.1 Mathematical Spaces

- Space: say $X$, a set [its members are called points]. Ideas like closeness, distance, size, other operations need not be defined.



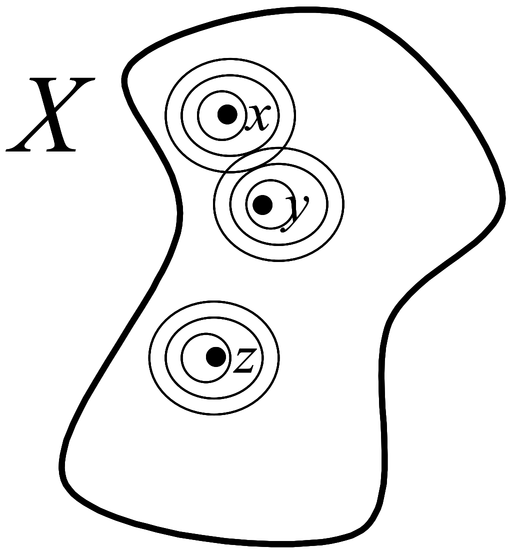

- Functions: A function $f$ maps $X$ into $Y$, written $f : X \to Y$, if $f$ assigns a point in $Y$ to each point in $X$.



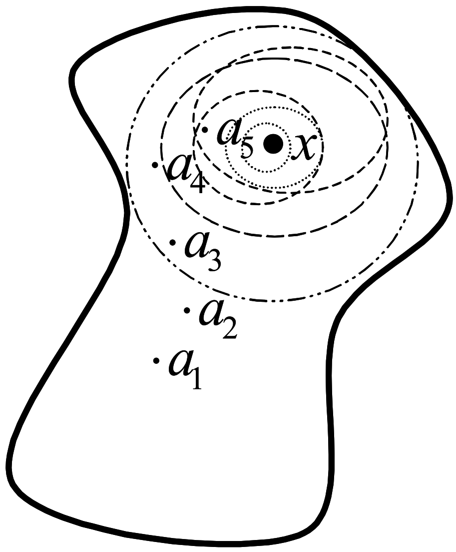

-
  - Functions can do things to spaces (transform them). Examples: shrinking, stretching, flattening, etc.
  - If $f : X \to Y$ and $S \subset X$, then the **image** of $S$ (written $f(S)$) is the set of all points in $Y$ that come from $S$.

- If $f : X \to Y$ and $R \subset Y$, then the **inverse image** of $R$ (written $f^{-1}(R)$) is the set of all points in $X$ that go into $R$.

- Topological Space: a space, say $X$, in which the idea of closeness (continuity, convergence) is defined. To define closeness we use structure called a topology. The topology contains various subsets of $X$ that are called **open sets.**
  - The identity of the open sets is all you need to know to talk about closeness.
  - Usually, we think of open sets as the insides of spheres (without their surfaces) or the unions of spheres (still without surface points).
  - Open sets must follow certain axioms:
    1. The union of any number of open sets is an open set.
    2. The intersection of a finite number of open sets is an open set
    3. The entire space $X$ and the empty set $\emptyset$ are open sets.
  - If a point is contained in an open set, the open set is called a neighborhood of that point.




  - Think of it this way: a point $s$ is closer to $x$ than other points are if it is in more (and smaller) neighborhoods of $x$.
  - A sequence is an ordered list of points, usually an infinite number of them: for example,
$$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots$$
is an example of an infinite sequence.
  - A sequence is **eventually** in a set $U$ if it starts either inside or outside of $U$ gets inside and then doesn't come out again.
  - A sequence **converges** to a point $x$ if the sequence is eventually in every neighborhood of $x$.

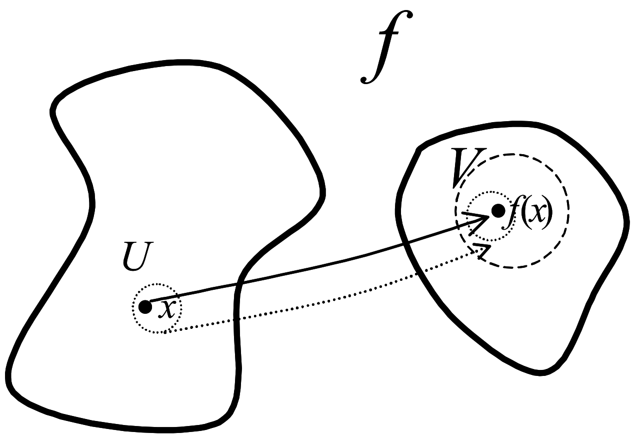



–

- Sets form a **basis** for a topology if all open sets can be constructed from unions of those sets.

**Example 21.1** *Suppose open intervals of the form* $(a, b) \equiv \{x \mid a < x < b\}$ *are a basis for the topology of the real line* $\mathbb{R}$*. Prove that* $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ *converges to* $0$*. Prove that* $\frac{1}{2}, 1, \frac{1}{3}, 1, \frac{1}{4}, 1, \ldots$ *does not converge to* $0$*.*

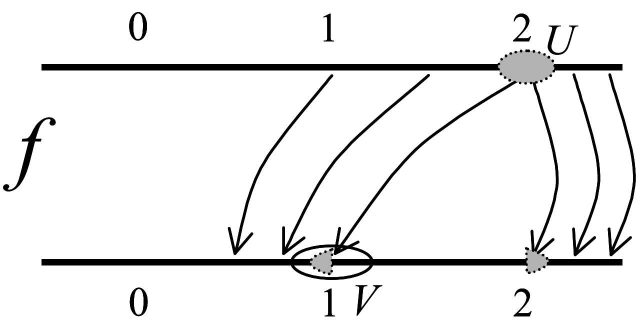
- A function $f : X \to Y$ is continuous at a point $x$ if for any neighborhood $V$ of $f(x)$ there is a neighborhood $U$ of $x$ such that $f(U) \subset V$. [Points near $x$ end up near $f(x)$]




**Example 21.2** $f : \mathbb{R} \to \mathbb{R}$, *where*

$$f(x) = \begin{cases} x & \text{for } \; x > 2 \\ .5x & \text{for } \; x \leq 2 \end{cases}$$

*Continuous at every point except* $x = 2$.




- – A function $f : X \to Y$ is continuous (everywhere) if it is continuous at every point.

**Proposition 21.1** *A function $f : X \to Y$ is continuous if and only if the inverse image of every open set is an open set.*

**Proof.** Suppose that the inverse image of every open set is open. We show that $f$ is continuous at every point $x$. Choose $x \in X$, and let $V$ be an open set around $f(x)$. Then $x \in f^{-1}(V)$, and since $f^{-1}(V)$ is open, it is a neighborhood of $x$. Furthermore $f(f^{-1}(V)) \subset V$ because $f(f^{-1}(V)) = V$. Thus $f$ is continuous at every $x$.

Suppose that $f$ is continuous (at every $x \in X$). We show that the inverse image of very open set is open. Let $V \subset Y$ be open, and let $U = f^{-1}(V)$. We must show $U$ is open. Choose $x \in U$. From continuity at $x$, there is a neighborhood $U_x$ of $x$ such that $f(U_x) \subset V$, so that $U_x \subset U$. Thus $U$ is the union of the $U_x$s, and the union of open sets is an open set. ■

**Problem 21.***MM.1. Suppose $X$ is a topological space and every subset of $X$ is open. How does this affect convergence? Give an example of a sequence that converges to* 0.

**Problem 21.***MM.2. For each of the following series, prove either that it converges or that it does not converge:*

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + ...$$

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + ...$$

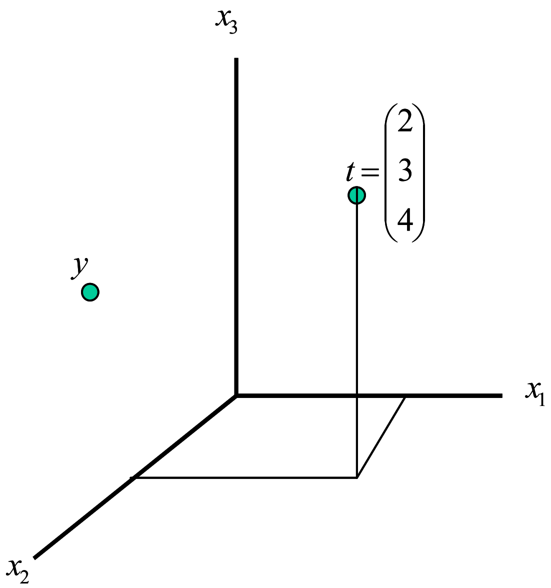
**Problem 21.***MM.3. Is the function*

$$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ \\ x & \text{for} \quad x \geq 0 \end{cases}$$

*continuous? Explain.*

## 21.2 Vector Spaces

- Normed vector space [normed linear space]: a space, say $X$, with a lot of structure defined: we want to know exactly where points are, distances, directions, how to go from one point to another.
- For normed vector spaces, we must have
  - vector addition
  - scalar multiplication
  - a norm (distance from the origin).
  - A topology is implied by the norm.




- Normed vector spaces can be either finite-dimensional or infinite dimensional, real or complex. We use Euclidean spaces of the form $\mathbb{R}^n$, which denotes an $n$-dimensional real vector space with the Euclidean norm $\|x\| \equiv \sqrt{\sum_i x_i^2}$ (conventional notion of distance from the origin).
- Notation: Usually use vector notation:
  - 
    - $X = n$-dimensional vector space; if $x \in X$, then $x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$
    - $Y = m$-dimensional vector space; $f : X \to Y$ is a function.

$$y = f(x)$$

means

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_n \end{bmatrix}\right)$$

or in scalar notation:

$$\begin{aligned} y_1 &= f_1(x_1, ...x_n) \\ y_2 &= f_2(x_1, ...x_n) \\ &\cdots \\ y_m &= f_m(x_1, ...x_n) \end{aligned}$$

**Definition 21.1** *A set of points (vectors)* $\{e_1, e_2, ..., e_n\}$ *is called a* ***basis*** *for a vector space, if any point* $y$ *in the spaced can be expressed in the form*

$$y = \alpha_1 e_1 + \alpha_2 e_2 + ... + \alpha_n e_n$$

*but none of the* $e_i$ *can be expressed similarly in terms of the others.*

**Definition 21.2** *The number of elements in any basis is called the* ***dimension*** *of the space.*

- The standard basis for a Euclidean space is the set of vectors with one 1 and the rest 0's:

$$\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

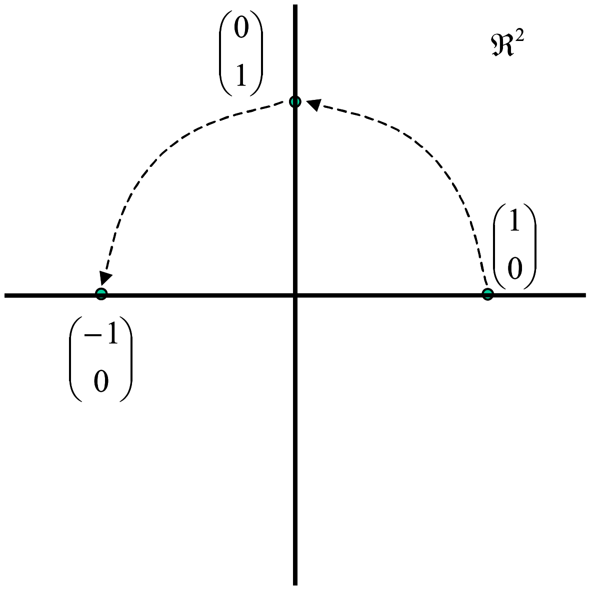
**Definition 21.3** *A function* $f : X \to Y$ *from one vector space to another is called a linear transformation if*

- $f(x + y) = f(x) + f(y)$
- *and* $f(\alpha x) = \alpha f(x)$

**Example 21.3** $f : \mathbb{R}^2 \to \mathbb{R}^2$

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$

*Is this linear? How does it transform the space?*




*Basis vectors of* $\mathbb{R}^2$ *are* $\begin{bmatrix}1\\0\end{bmatrix}$ *and* $\begin{bmatrix}0\\1\end{bmatrix}$

$$f\left(\begin{bmatrix}1\\0\end{bmatrix}\right)=\begin{bmatrix}0\\1\end{bmatrix},\quad f\left(\begin{bmatrix}0\\1\end{bmatrix}\right)=\begin{bmatrix}-1\\0\end{bmatrix}$$

*The matrix is*

$$f\equiv\begin{bmatrix}0&-1\\1&0\end{bmatrix}$$

*Because of the properties of linear transformations, the matrix can be used in place of the function:*

$$\begin{bmatrix}2\\3\end{bmatrix}=2\begin{bmatrix}1\\0\end{bmatrix}+3\begin{bmatrix}0\\1\end{bmatrix}$$

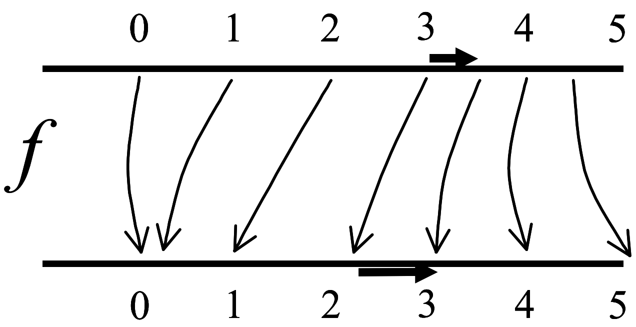
$$f\left(\begin{bmatrix}2\\3\end{bmatrix}\right)=2f\left(\begin{bmatrix}1\\0\end{bmatrix}\right)+3f\left(\begin{bmatrix}0\\1\end{bmatrix}\right)=2\begin{bmatrix}0\\1\end{bmatrix}+3\begin{bmatrix}-1\\0\end{bmatrix}$$

$$=\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}=\begin{bmatrix}-3\\2\end{bmatrix}$$

## 21.3 Derivatives

- Suppose $f:\mathbb{R}\to\mathbb{R}$ is a nice (differentiable) function. The **derivative** $f'$ of $f$ is the rate of change of $f(x)$ as $x$ changes.
  - If $f(x)$ changes twice as fast as $x$, then $f'(x)=2$.
  - If $\Delta x$ is a small change in $x$,then $\Delta f(x)=$ (approx) $f'(x)\Delta x$.

**Example 21.4** $f(x)=\frac{1}{4}x^2$




$$f(3) = \frac{9}{4}, \quad f(3.5) = \frac{49}{16}, \quad \Delta x = \frac{1}{2}, \quad \Delta f(x) = \frac{13}{16}$$
$$f'(x) = \frac{1}{2}x, \quad f'(3) = \frac{3}{2}, \quad \Delta f(x) = \text{(approx) } f'(x)\Delta x = \frac{3}{4}$$

*Why are the two results different?*

- Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is a nice (differentiable) function. The **Jacobian matrix** describes how $f(x)$ changes when $x$ changes.
  - If $\Delta x$ (a vector!) is a small change in $x$,then
  $$\Delta f(x) = \text{(approx)} \frac{\partial f(x)}{\partial x} \Delta x.$$
  - If $f$ is differentiable, the Jacobian matrix of all the partial derivatives is denoted by
  $$\frac{\partial f(x)}{\partial x} \equiv \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_n(x)}{\partial x_n} \end{bmatrix}.$$
    * Hint: Treat vectors like scalars.
    * I try to keep my notation the same as M-C's, but in this case his notation $D_x f(x)$ for $\frac{\partial f(x)}{\partial x}$ seems confusing.

**Example 21.5** *Linear transformation:*

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$
$$\frac{\partial f(x)}{\partial x} \equiv \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

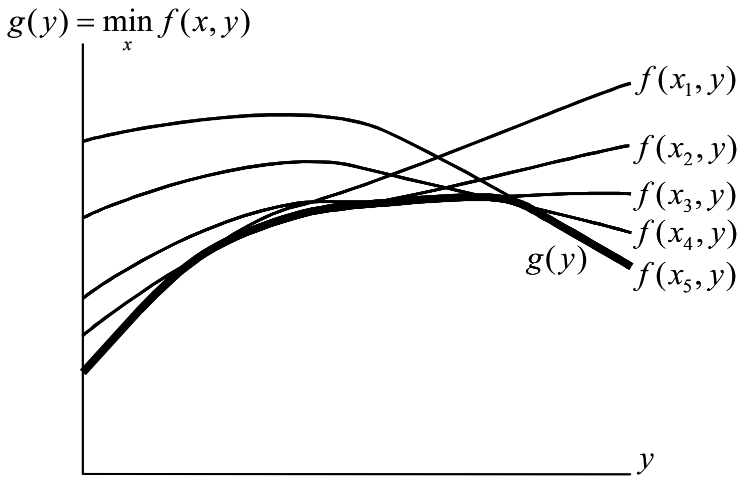
*Suppose $x$ goes from* $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ *to* $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$, *so that* $\Delta x = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

*$f(x)$ goes from* $\begin{bmatrix} -2 \\ 3 \end{bmatrix}$ *to* $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$, *so that* $\Delta f(x) = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$

$$\textit{Does } \Delta f(x) = \textit{(approx) } \frac{\partial f(x)}{\partial x} \Delta x \textit{ ?}$$

*Yes, the result is exact, why?*

- The Envelope Theorem is a theorem about creating a new function out of a family of functions by always choosing the minimum (or maximum) value from every function in the family. In the figure below, the family $f(x, y)$ and the new function $g$ are plotted as functions of $y$. The function $g(y)$ is called the lower **envelope** of $f(x, y)$.




- The theorem says that the slope of the envelope at any point is the same as the slope of the member of the family that it touches.
- M-C has a more general version of the theorem: don't worry about it, because it is quite messy.

**Proposition 21.2** *(Envelope Theorem) Let* $g(y) = \min_x f(x, y)$*, where* $f(x, y)$ *is differentiable. Then*

$$g'(y) = \left.\frac{\partial f(x, y)}{\partial y}\right]_{x=\hat{x}(y)}$$

*where* $\hat{x}(y)$ *is the value of* $x$ *that minimizes* $f(x, y)$*.*

- The intuition: as $y$ changes $x$ must change because $x$ must always minimizes $f(x, y)$. Therefore if $y$ changes by $\Delta y$, $\Delta g(y)$ comes from two sources, directly from $\Delta y$ and from $\Delta x$ caused by $\Delta y$. Then envelope theorem says that if $\Delta y$ is small, the part of $\Delta g$ that comes from $\Delta x$ (labeled $\Delta g_x$ on the graph) is close to 0.
- The reason is that the curves are flat at $\hat{x}(y)$, because $\hat{x}(y)$ minimizes $f(x, y)$. So even large changes in $\hat{x}$ produces small changes in $f(\hat{x}, y)$.

**Proof.** Let $\hat{x}(y)$ be the solution of $\min_x f(x, y)$. The f.o.c for $\hat{x}(y)$ is

$$\left.\frac{\partial f}{dx}\right]_{x=\hat{x}(y)} = 0.$$

We can now write:

$$g(y) = f(\hat{x}(y), y)),$$

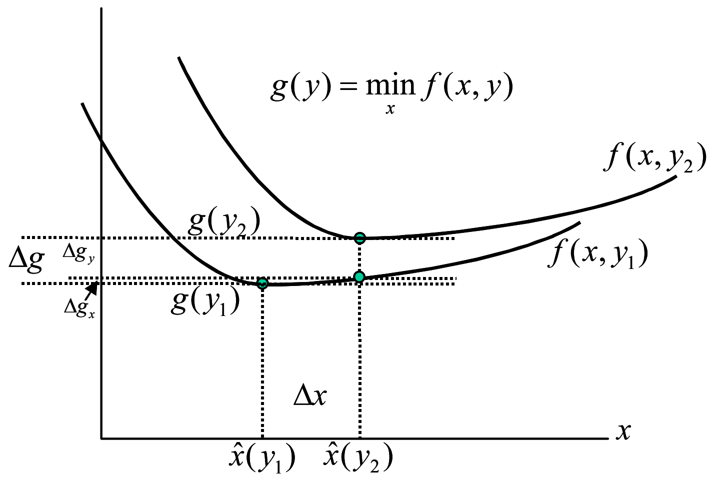


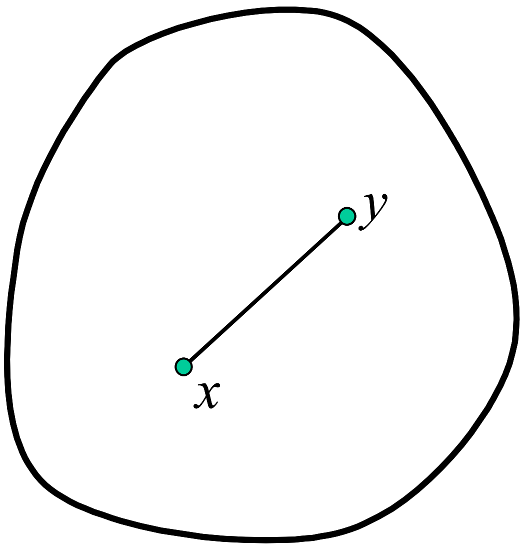

Figure 1:

so, by the chain rule,

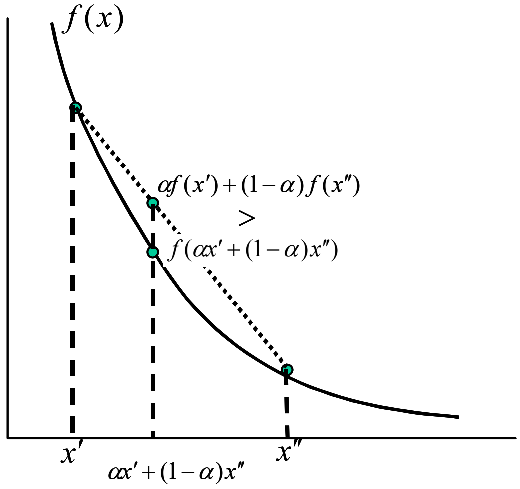
$$g'(y) = \frac{\partial f}{dx}\bigg]_{x=\hat{x}(y)} \hat{x}'(y) + \frac{\partial f}{\partial y}.$$

The first term is 0. ■

- Convexity:




**Definition 21.4** *A set $X$ in a vector space is **convex** if for any $x, y \in X$, the line $\overline{xy} \subset X$, that is if $\alpha x + (1-\alpha)y \in X$ whenever $0 \leq \alpha \leq 1$.*



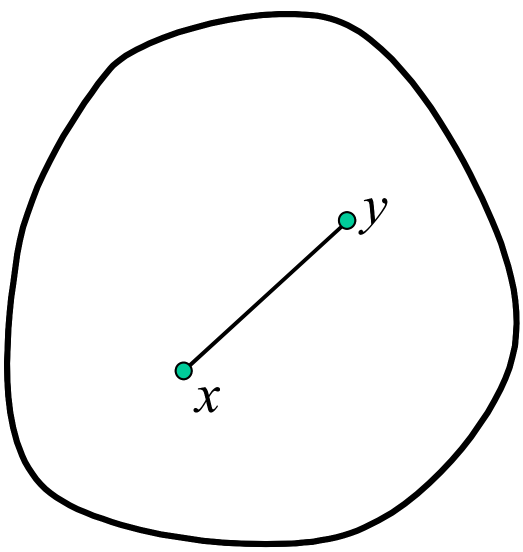

**Definition 21.5** *A function $f : X \to \mathbb{R}$ is **convex** if for all $x, y \in X$, and $\alpha \in [0, 1]$, $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - a)f(y)$; that is, the line between any two points on the graph of the function, lies on or above the graph. [This is true if and only if the set $F = \{(x, z) \,|\, x \in X, z \geq f(x)\}$ is convex.]*

**Definition 21.6** *A function $g : X \to \mathbb{R}$ is concave if and only if $-g$ is a convex function.*

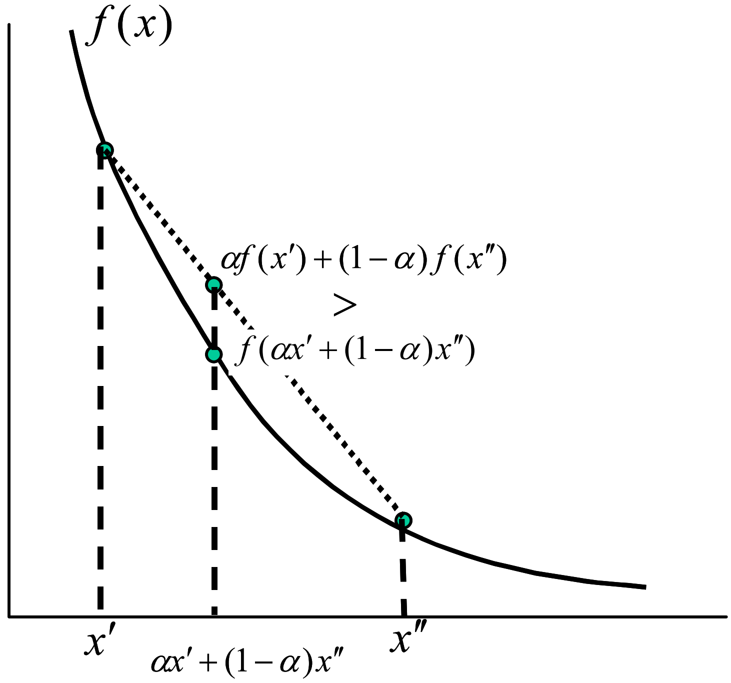
**Problem 21.***MM.4. Find a 2-dimensional matrix that stretches the horizontal axis by a factor of 2 and that shrinks the vertical access by a factor of 1/2.*

- *What happens to* $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *?*
- *Does any vector have its direction unchanged?*

## 21.4 Convexity and Concavity



**Definition 21.7** *A set $X$ in a vector space is **convex** if for any $x, y \in X$, the line $\overline{xy} \subset X$, that is if $\alpha x + (1-\alpha)y \in X$ whenever $0 \leq \alpha \leq 1$.*




**Definition 21.8** *A function $f : X \to \mathbb{R}$ is **convex** if*

- *for all $x, y \in X$, and $\alpha \in [0,1]$, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-a)f(y)$*
- *that is, the line between any two points on the graph of the function, lies on or above the graph.*
- *This is true if and only if the set $F = \{(x, z) \,|\, x \in X, z \geq f(x)\}$ is convex.*

**Definition 21.9** *A function $g : X \to \mathbb{R}$ is concave if and only if $-g$ is a convex function.*

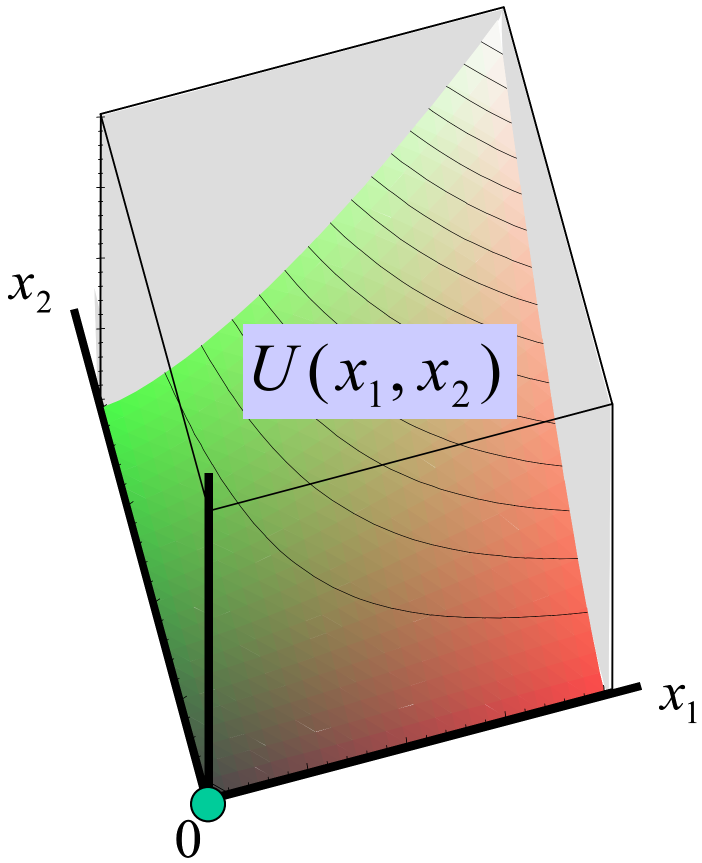
**Definition 21.10** *A function $U$ is quasiconcave if for all $x$ the sets $G(x) = \{y \,|\, U(y) \geq U(x)\}$ are convex. A function $V$ is quasiconvex if for all $p$ the sets $B(p) = \{q \,|\, V(q) \leq V(p)\}$ are convex.*

**Proposition 21.3** *All convex functions are quasiconvex, and all concave functions are quasiconcave (but not the other way round).*

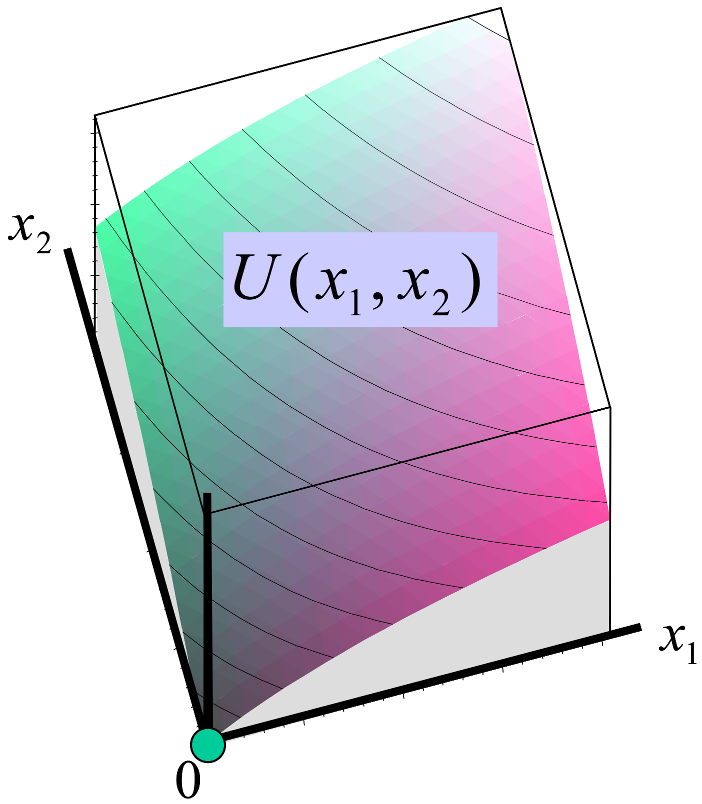
**Proof.** Try it as an exercise. ■

- The following are 3-dimensional graphs of two utility functions
- The graph below represents $U(x_1, x_2) \equiv (x_1 x_2)^{1.5}$.
  - The function is quasiconcave, because the $G$ sets are convex sets.

- However, the function is not concave, because the set under the surface is not convex.
- The reason is that the surface curves up as you move away from the origin along a ray.




- The graph below represents $U(x_1, x_2) \equiv 3\ln(x + 1)(y + 1)$.
  - This utility function is concave, so we can be sure that the function is also quasiconcave.
  - The surface curves downwards not only from side-to-side as above, but also along a ray from the origin.



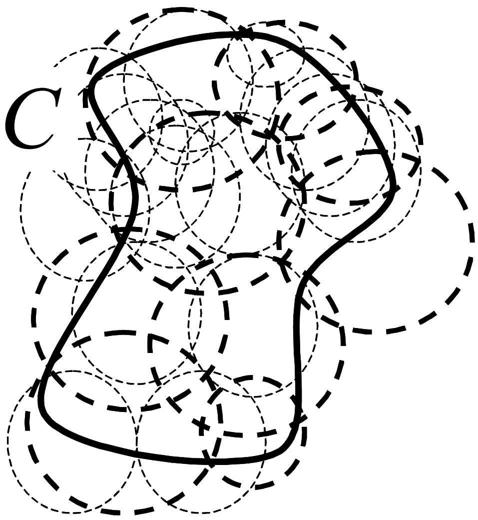

- Notice that both utility functions represent very similar preferences.
- Quasiconcave functions can be made concave by monotonic transformations, so the preferences that they represent need not change.

## 21.5 Compact Sets

**Definition 21.11** *A family of open sets* $\{U_i\}$ *forms an* open cover *for a set* $V$ *if* $V \subset \cup_i\{U_i\}$.

**Definition 21.12** *A set* $C$ *in a topological space is* compact *if EVERY open cover of* $C$ *contains a finite subcover.*

- If an infinite number of open sets covers a compact set, then only a finite number of those sets are in fact necessary.
- Note: if you can find *one* open cover of $C$ that lacks a finite subcover, then $C$ is not compact. It does not matter if other open covers of $C$ have finite subcovers. If there is *one* open cover with a finite subcover, the the set fails the test for being compact.




**Example 21.6** *The open interval* $(0,1)$ *is not compact! This is because* $(0,1)$ *is covered by the sets* $\{(\frac{1}{2},1),(\frac{1}{4},1),(\frac{1}{8},1),(\frac{1}{16},1),...\}$*, but it is not possible to select a finite number of these sets that also cover* $(0,1)$*. [Note: there are many other open covers of* $(0,1)$ *that have finite subcovers, but this is not relevant!]*

**Proposition 21.4** *If* $C$ *is compact and* $f$ *is a continuous function, then* $f(C)$ *is compact.*

**Proof.** Choose any open cover $\{U_i\}$ of $f(C)$, and take the inverse images $\{f^{-1}(U_i)\}$. These are an open cover for $C$. Because $C$ is compact you can chose a finite subcover $\{f^{-1}(U_1),...,f^{-1}(U_n)\}$. But then $\{U_1,...,U_n\}$ must be a finite subcover of $f(C)$. Therefore $f(C)$ is compact. ■

**Proposition 21.5** *Suppose* $C$ *is compact and suppose* $F$ *is closed and* $F\subset C$*. Then* $F$ *is compact.*

**Proof.** Let $\{U_i\}$ be an open cover of $F$, and define $\hat{U}=F_C$ (the complement of $F$). Then, $\hat{U}$ is an open set and $\hat{U}$ and $\{U_i\}$ form a cover of $C$. Because $C$ is compact, we can select the finite subcover $\{U_{1,...},U_n,\hat{U}\}$ of $C$ and thus also of $F$. [$\hat{U}$ may or may not have to be in the finite subcover of $C$, but it can't hurt.] But $\hat{U}$ does NOT help cover $F$, so $\{U_{1,...},U_n\}$ must be a finite cover of $F$. ■

**Definition 21.13** *A set* $X$ *is* bounded *if it is contained in a sphere (of finite radius).*

**Proposition 21.6** *In Euclidean vector spaces (including the real line), any compact set is closed and bounded.*

**Proof.** Let $X$ be a compact subset of $\mathbb{R}^n$.

1. We show that $X$ is bounded. Consider spheres of the form $S_n=\{x_n \mid \|x_n\|<n\}$ for all integers $n$. These spheres eventually include every point in $\mathbb{R}^n$Chose a finite number subcover of $X$, and let $S_{\hat{n}}$ be the largest set in the subcover. Then $X\subset S_{\hat{n}}$ and is thus bounded.

2. If $X$ is not closed then $X$ is not compact: If $X$ is not closed, we can choose a boundary point $b$ of $X$ such that $b \notin X$. Now consider the closed spheres of the form $\bar{S}_i = \{x \mid \|x - b\| \leq \frac{1}{2^i}\}$. Each $\bar{S}_i$ contains points in $X$. The intersection of all these sets is exactly the point $b$. Now take the complement of each set, $\bar{S}_{iC}$. These sets form a cover for $X$ because everything but $b$ (which is not in $X$) is in one of the $\bar{S}_{iC}$. Suppose there is a finite subcover, and suppose $\bar{S}_{kC}$ is the largest set in the subcover. Then the subcover does not include any points in $\bar{S}_k$, but $\bar{S}_k$ includes points of $X$, a contradiction.

■

- We need to show that every closed an bounded set in Euclidean space is compact, but this theorem relies on a fundamental property of real numbers.
- We will accept the following proposition without proof: it comes from the way real numbers are constructed as limits of sequences of rational numbers.

**Proposition 21.7** *The real line $\mathbb{R}$ and Euclidean vector spaces $\mathbb{R}^n$ are* complete; *that is, every Cauchy sequence converges.*

**Proposition 21.8** *Suppose $I_i = [a_i, b_i]$ are closed intervals in $\mathbb{R}$ with $I_{i+1} \subset I_i$. Then there is a number $x$ such that $x \in \cap\{I_i\}$.*

**Proof.** $\{a_i\}$ is increasing and bounded from above by $b_1$. It must be a Cauchy sequence, so it has a limit. Let $x = \lim_{i \to \infty} a_i$. Then for all $i$, $a_i \leq x \leq b_i$ [why], so $x \in \cap\{I_i\}$. ■

**Proposition 21.9 (Heine-Borel)** *The closed real interval $[0, 1]$ is compact.*

**Proof in Words.** Suppose otherwise. Then there is some open cover $\{U_i\}$ of $[0, 1]$ that does not contain a finite subcover of $[0, 1]$. But this means that $\{U_i\}$ either does not contain a finite subcover for $[0, \frac{1}{2}]$ or does not contain a finite subcover for $[\frac{1}{2}, 1]$. If we select the half-interval without a finite subcover, we can divide that in half and find a smaller interval without a finite subcover in $\{U_i\}$. But now the intersection of the entire sequence of intervals without finite subcovers must contain a point $x$. And $x \in U_i$ for some $i$ (because $\{U_i\}$ is a cover). But that $U_i$ is an open set with a finite width, so it must also contain a one of the small intervals around $x$. A contradiction, because that one open set would constitute a finite subcover of the interval. ■

- The same proof works for closed higher-dimensional "cubes" in $\mathbb{R}^n$, and since any closed and bound set must be contained in some such cube, and since we have shown that any closed subset of a compact set is compact, we know:

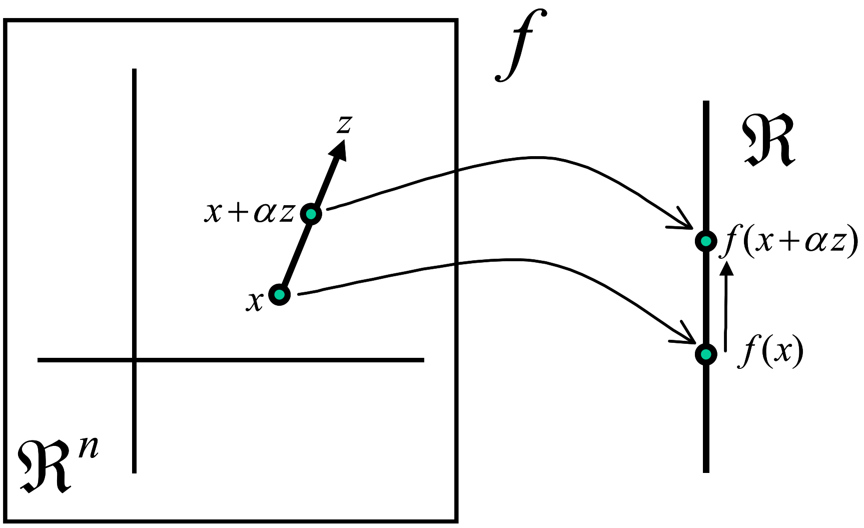
**Proposition 21.10** *Any closed and bounded set in $\mathbb{R}^n$ is compact.*

**Proposition 21.11 (Maximum value theorem)** *Suppose $C \subset \mathbb{R}^n$ is compact and suppose $f : C \to \mathbb{R}$ is continuous. Then $f$ takes a maximum value on $C$.*

**Proof.** $f(C)$ is compact, and therefore it is closed and bounded. Because $f(C)$ is bounded it must have a least upper bound ($\mathbb{R}$ is complete), and because $f(C)$ is closed, it contains its least upper bound, which then must be the maximum point in $f(C)$. ■

**Definition 21.14** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then $\frac{\partial f(x)}{\partial x} \equiv \nabla f(x) \equiv [\frac{\partial f(x)}{\partial x_1}, ..., \frac{\partial f(x)}{\partial x_n}]$ is called the gradient of $f$ at $x$.*

**Definition 21.15** *Let $z$ be a vector of unit length ($\|z\| = 1$). Define $h(\alpha) = f(x + \alpha z)$. The directional derivative of $f$ at $x$ in the direction $z$, is given by $h'(0)$.*




**Proposition 21.12** *If $z$ is any vector of unit length, then $\nabla f(x) \cdot z$ is the directional derivative of $f$ in the direction $z$, so that*

$$f(x + \alpha z) = f(x) + (\nabla f(x) \cdot z)\alpha + \text{ (second-order-small error)}$$

*The direction of $\nabla f(x)$ is that of the maximum rate of change of $f$ at $x$, and its magnitude is the maximum directional derivative for all directions.*

**Proof.** By the chain rule,

$$h'(\alpha) = \sum_i \frac{\partial f(x)}{\partial x_i} z_i \equiv \nabla f(x + \alpha z) \cdot z,$$

so that $h'(0) \equiv \nabla f(x) \cdot z$. We know that $h(\alpha)$ =(approx) $h(0) + h'(0)\alpha$, and the approximation for $f$ follows. From trigonometry we know that $u \cdot v = \|u\| \, \|v\| \cos \theta$, where $\theta$ is the angle between $u$ and $v$. Therefore, for $\bar{z} = \frac{1}{\|\nabla f\|} \nabla f$, we have $\nabla f \cdot \bar{z} = \max\{\nabla f \cdot z \mid \|z\| = 1\}$, which means that the gradient $\nabla f$ is the direction of maximum increase of $f$. ■

**Proposition 21.13 (Kuhn-Tucker Necessary Conditions)** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable and let $B \subset \mathbb{R}^n$ with $x^* \in B$. Then $f(x^*)$ =max$\{f(x) \mid x \in B\}$ only if the directional derivative of $f$ at $x^*$ is 0 or negative in every direction $z$ within $B$ (that is, in every direction $z$ such that for all positive and sufficiently small $\alpha$, we have $x^* + \alpha z \in B$).*

**Proof.** Suppose $x^* + \alpha z \in B$ for all $\alpha$ sufficiently small. Then if $\nabla f(x^*) \cdot z > 0$, $f(x^* + \alpha z) > f(x^*)$ for small $\alpha$, so that $f(x^*)$ is not a maximum. ■

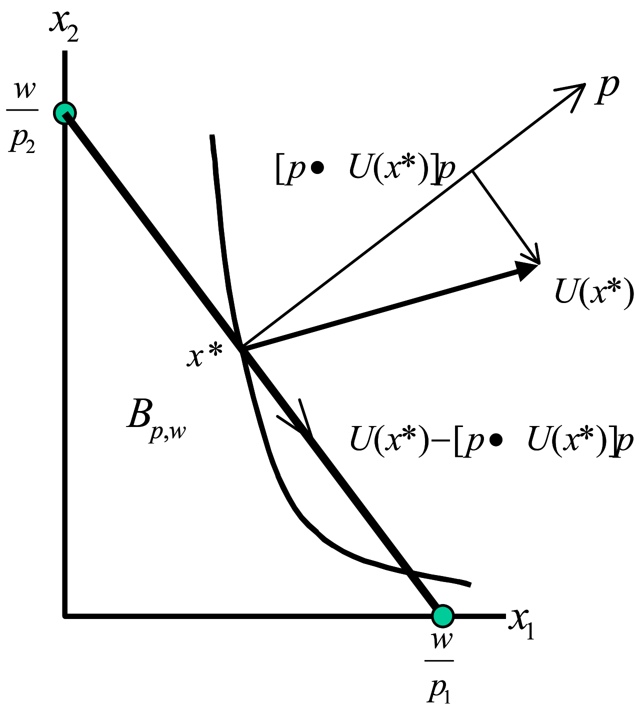



Figure 2:

## 21.6 Separating Hyperplane Theorem

**Definition 21.16** *The set*

$$H_{p,x_0} \equiv \{x \,|\, x \text{ and } x_0 \in \mathbb{R}^n,\, p(x-x_0)=0\}$$

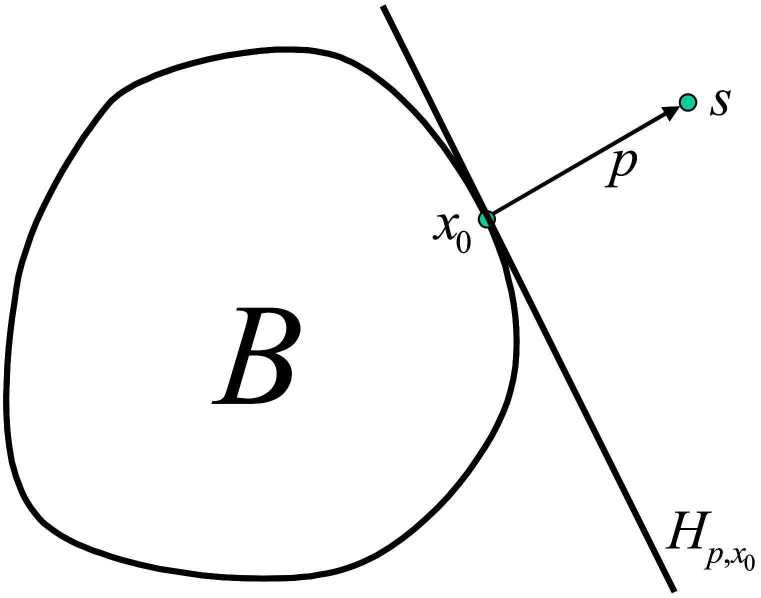
*is a* **`hyperplane`** *of dimension $n-1$ orthogonal to $p$ through the point $x_0$.*

**Proposition 21.14** *(Separating Hyperplane Theorem) If $B \subset \mathbb{R}^n$ is convex and closed, and if $s \notin B$, then*

- *there is a hyperplane $H$ of dimension $n-1$ such that $B$ lies on one side of $H$ and $s$ lies on the other;*
- *more precisely, there is some $x_0$ and $p \neq 0$, such that*

$$pb \leq px_0 < ps$$

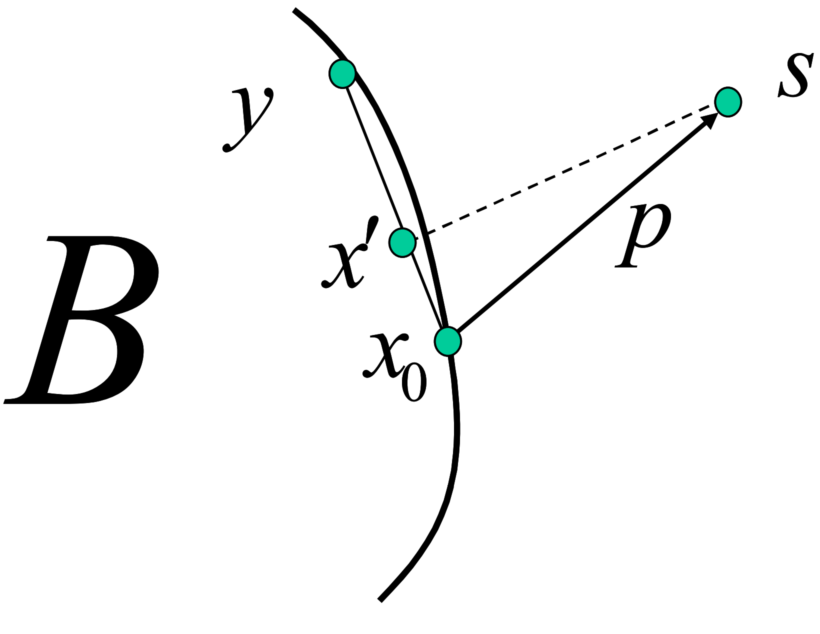
*for all $b \in B$.*




**Proof.** Let $x_0$ be the closest point in $B$ to $s$. Then

- set $p \equiv s - x_0$
- so that $0 < p \cdot p = p(s - x_0)$.
- Therefore $px_0 < ps$.
- Suppose for some $y \in B$, $py > px_0$.

■




**Proof.**

- Then $p(y - x_0) > 0$.
- Set $x' = \alpha y + (1 - \alpha)x_0$ where $\alpha \in (0, 1)$.
- By convexity, $x' \in B$.
- We show that for small enough $\alpha$, $x'$ is closer to $s$ than $x_0$ is, a contradiction.
  - We have $s - x' = (s - x_0) - \alpha(y - x_0) = p - \alpha(y - x_0)$.
  - For $\alpha$ small we have,

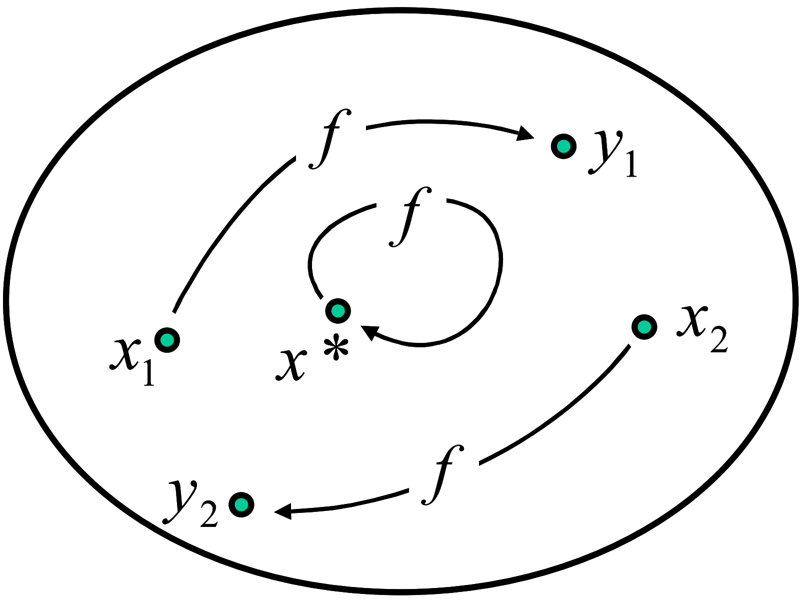
$$\begin{aligned} \|s - x'\|^2 &= (s - x')(s - x') \\ &= (s - x_0)(s - x_0) \\ &\quad + \alpha^2(y - x_0)(y - x_0) \\ &\quad - 2\alpha p(y - x_0) \\ &< \|s - x_0\| \ \text{[Why?]} \end{aligned}$$

■

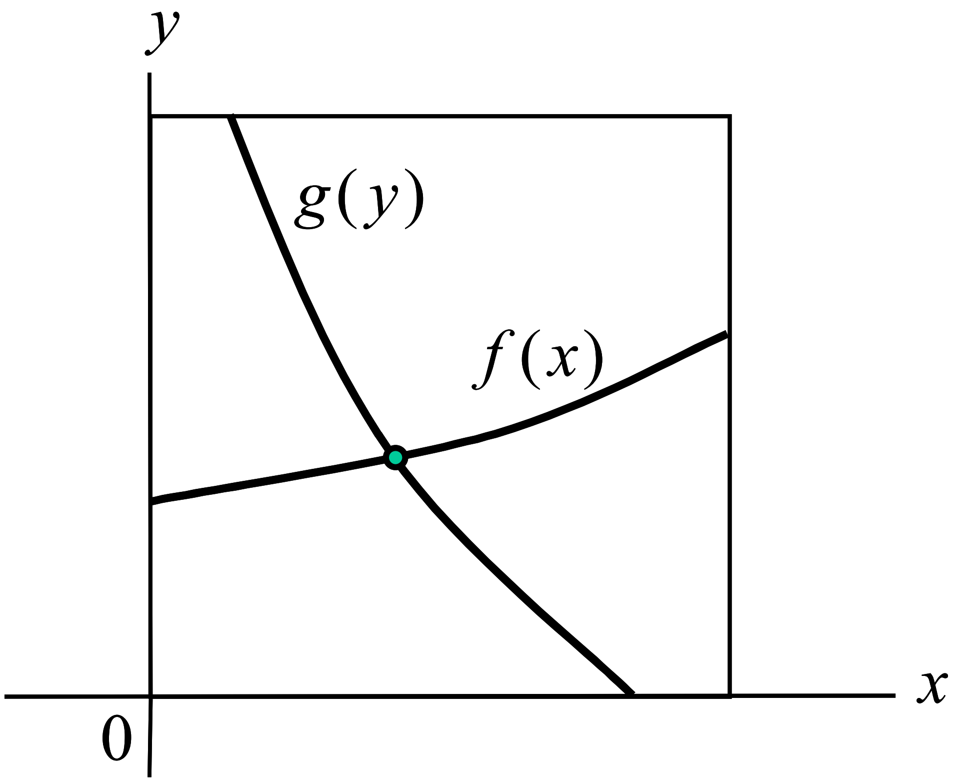
## 21.7 Fixed-Point Theorems

- Fixed point theorems can be viewed in two ways:
- If $f : X \to X$ has the right properties, then there is an $x^* \in X$ such that $f(x^*) = x^*$.
  - Example: $f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \equiv \begin{bmatrix} y^2 \\ x^2 \end{bmatrix}$

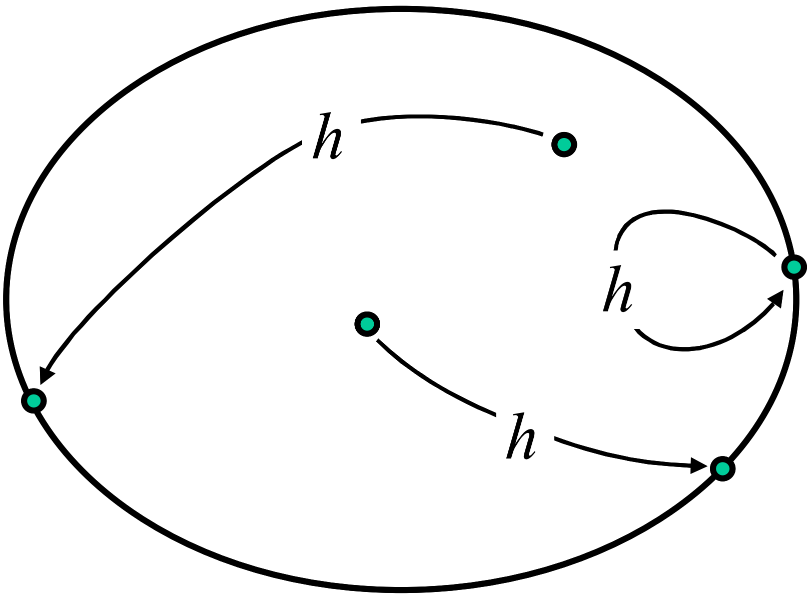
    has two fixed points: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
  - Example:



- Consider the unit square, and suppose $y = f(x)$ and $x = g(y)$ are both continuous functions. Then the fixed point theorems say that their graphs must intersect.




- This theorem is applies to higher dimensional spaces as well.



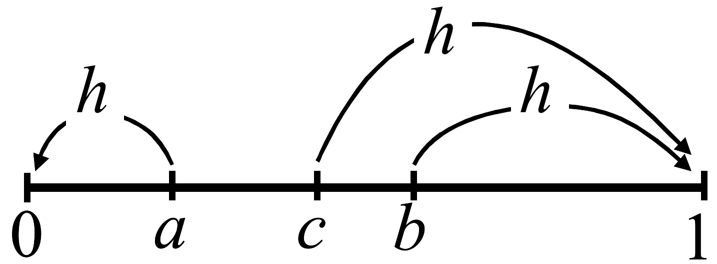

**Definition 21.17** *Let $X$ be a closed set with boundary $\bar{X}$.*

- *A function is called a* `retract` *on $X$ if it maps $X$ onto its boundary, and leaves the boundary fixed.*

- *That is, a function $h : X \to \bar{X}$ is a retract if $h(\bar{x}) = \bar{x}$ for all $\bar{x} \in \bar{X}$*

**Proposition 21.15** *(Sperner's Lemma). Suppose $X \subset \mathbb{R}^n$ is compact and convex. Then every retract defined on $X$ must be* ***discontinuous****.*

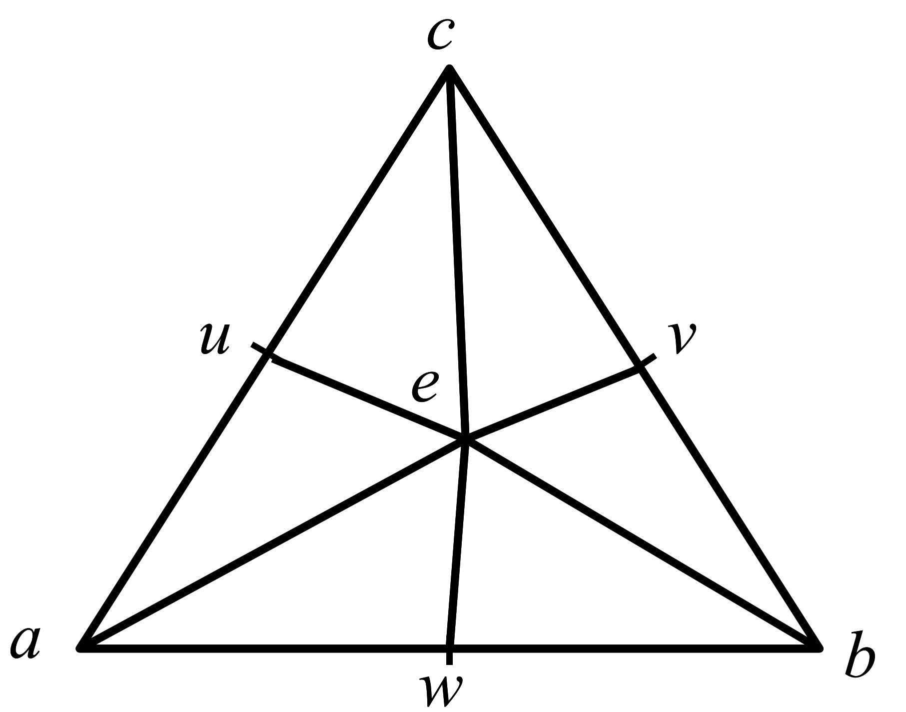
- The proof of this theorem has a lot of details, but the idea isn't too difficult.
- First we will do it for $X = [0, 1]$, the real unit interval.




  - We will call a line segment "split" if $h$ maps one endpoint to 0 and the other to 1.
  - In the drawing, the line segment $\overline{ab}$ is split.
  - Claim: if we divide any split line segment into two subsegments, then one of the subsegments is also split.
    * Proof: suppose $\overline{ab}$ is a split segment so that $h(a) = 0$ and $h(b) = 1$ (or vice versa).
    * Suppose $c$ divides $\overline{ab}$ into $\overline{ac}$ and $\overline{cb}$ (as above).
    * If $h(c) = 0$,then $\overline{cb}$ is split.
    * If $h(c) = 1$,then $\overline{ac}$ is split.
  - Now we start with the whole segment $[0, 1]$ which, by the definition of a retract, is split.
  - Divide that segment into two parts, and select the split subsegment.
  - Continue this process ad infinitum to obtain a sequence of nested split segments:
$$[0, 1] \supset [a_1, b_1] \supset [a_2, b_2] \supset, ...$$
  - By compactness, the intersection of these nested sets must contain a point, say $c^*$.
  - Claim: $h$ is not continuous at $c^*$.
    * Proof: Choose a small neighborhood $V$ of $h(c^*)$
    * choose any neighborhood $U$ of $c^*$.
    * For $n$ sufficiently large the split segment $[a_n, b_n] \subset U$.
    * But $h(a_n)$ or $h(b_n)$ are far apart so one of them cannot be in $V$.
  - QED

- The proof for a retract defined on a simplex $S \subset \mathbb{R}^n$ is similar.

- Suppose $h : S \to \bar{S}$.
- We define another function $\hat{h}$, where $\hat{h}(x)$ is the `vertex` of $S$ closest to $h(x)$,which is on the boundary of $S$.
- Uses induction on the dimension of the simplex.
- We've already proved the theorem for the one dimensional simplex.
- For a two dimensional simplex (an ordinary triangle),
- Note that $\hat{h}$ maps the three original vertices to themselves.
- Divide each face of the simplex (side of the triangle) into subsimplexes.
- By induction, from each face has a split subsimplex.
- Say, for example $\overline{uc}$, $\overline{vc}$ and $\overline{bw}$.
- Place a point in (near) the center of the simplex (point $e$) and use it to divide the simplex into subsimplices.
- Consider only the subsimplices created out of split faces: $\overline{euc}$, $\overline{evc}$ and $\overline{ebw}$.
- Claim: one of them must also be split
  - If $\hat{h}(e) = a$, then $\overline{evc}$ must be split, etc.

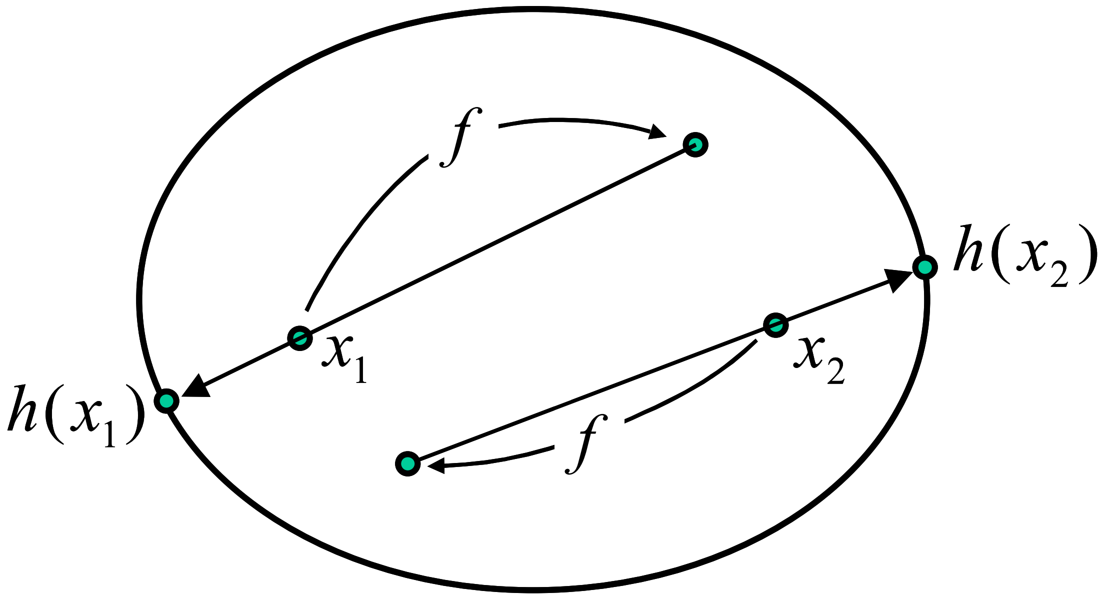



- Now repeat the process with the split subsimplex.
- We get a nested sequence of subsimplices, getting smaller and smaller.

- Intersection of all of them must contain a point $c^*$.
- As above, $h$ is discontinuous at $c^*$.
- QED

**Proposition 21.16** *Suppose $A \subset \mathbb{R}^n$ is nonempty, compact and convex, and $f : A \to A$ is continuous, then $f$ has a fixed point. That is, for some $x^* \in A$, $f(x^*) = x^*$.*

**Proof.** Suppose $f$ has no fixed point. Then construct a retract $h : A \to \bar{A}$ as follows:




- For any point $x$ find $f(x)$.
- Project a straight line from $f(x)$ through $x$ to the boundary of $A$.
- Define $h(x)$ to be the boundary point determined by that line.
- Note: $h(x)$ is a retract, because it leaves all boundary points fixed.
- Because $f$ and the projection process are continuous, $h$ must be continuous.
- This contradicts Sperner's Lemma.
- Therefore, $f$ must have a fixed point (in which case $h$ would not be defined at the fixed point).

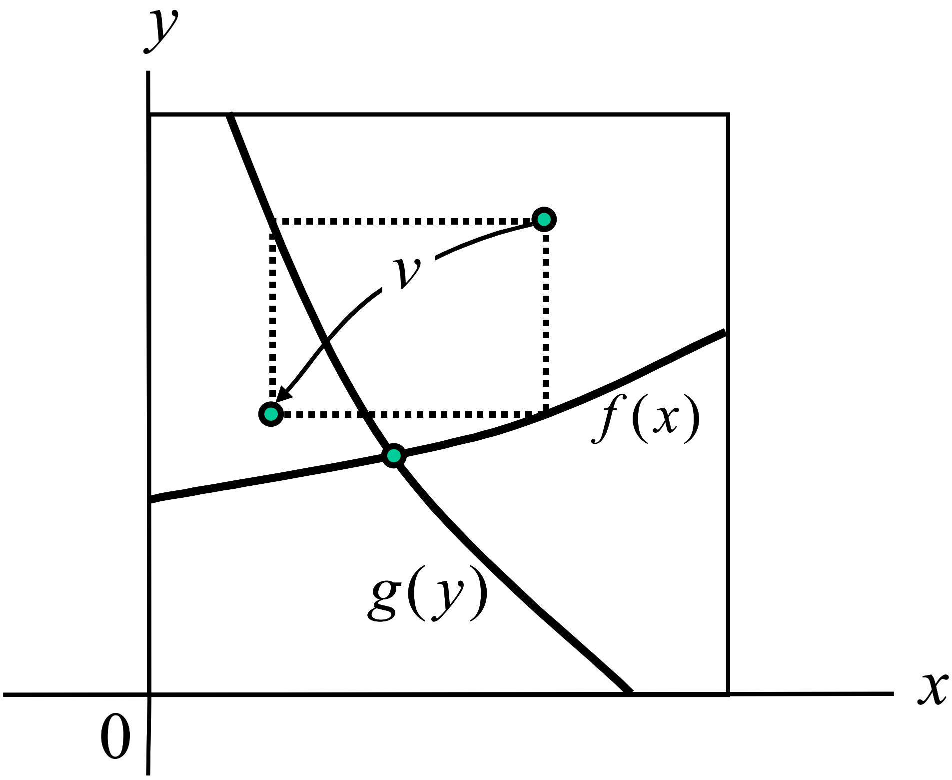
■

- We can now prove that under appropriate conditions, curves intersect.
- We demonstrate the result for real-valued functions of one variable.

**Proposition 21.17** *Consider the unit square, and suppose $y = f(x)$ and $x = g(y)$ are both continuous functions. Then their graphs must intersect.*

**Proof.** Define the function $v : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$v\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} g(y) \\ f(x) \end{bmatrix}.$$




- Because $f$ and $g$ are continuous, $v$ must be continuous.
- $v$ has a fixed point $\begin{bmatrix} x^* \\ y^* \end{bmatrix}$.
- The curves must intersect there:
  - $y^* = f(x^*)$, $x^* = g(y^*)$

■