ECON2125/4021/8013

Lecture 25

John Stachurski

Semester 1, 2015

Announcements

1. This week's lectures will be revision
   - Today's lecture is a review of optimization and linear algebra
   - Tomorrow will review probability, analysis and dynamics
2. Final practice question set is up on GitHub (set 3)

Optimization Review

Consider a maximization problem such as

$$\max_{\mathbf{x} \in D} f(\mathbf{x}) \quad \text{where} \quad f \colon D \to \mathbb{R}$$

A maximizer is a point $\mathbf{x}^* \in D$ such that

$$f(\mathbf{x}^*) \geq f(\mathbf{x}) \quad \forall\, \mathbf{x} \in D$$

In general,

- there may be one, zero, or many maximizers
- maximizers can be interior or on boundaries
- similar story for minimizers

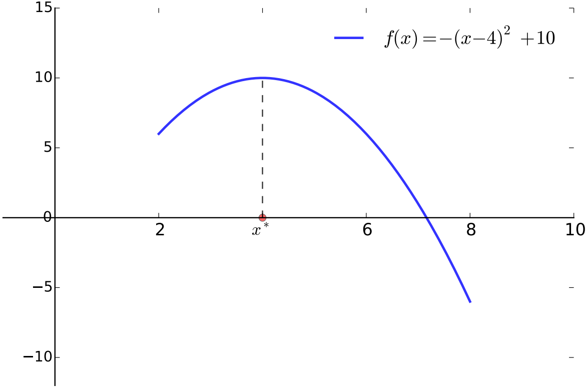


Figure : f has a unique maximizer on $D = [2, 8]$

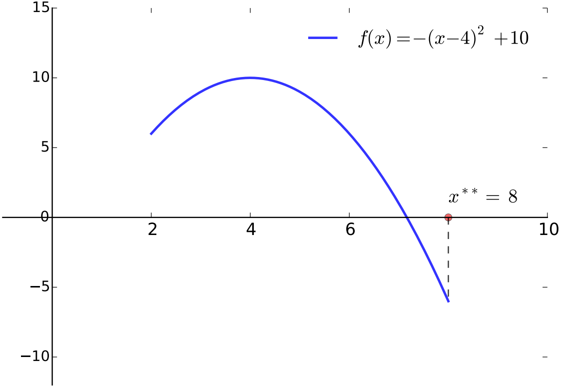


Figure : f has a unique minimizer on $D = [2, 8]$

In these pictures, the maximizer x^* is interior

It is also stationary, meaning

$$f'(x^*) = 0$$

For multivariate f, stationarity requires

$$\frac{\partial}{\partial x_i} f(\mathbf{x}) = 0 \quad \text{for all } i$$

Intuitively, the function is "flat" at such an $\mathbf{x}$

- zero slope in all directions

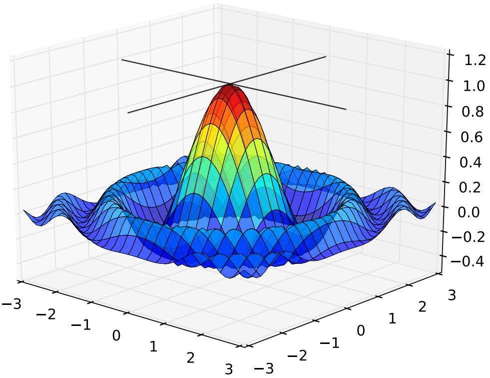


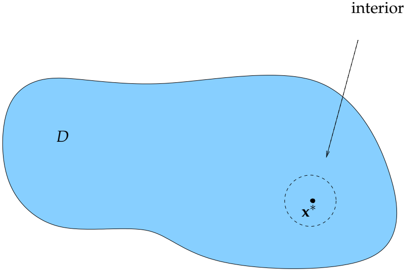
Figure : $(0, 0)$ is a stationary point of this f

Key Idea. For differentiable functions, any interior maximizer or minimizer must be stationary

Intuition: Suppose that $\mathbf{x}^*$ is an interior maximizer

Since $\mathbf{x}^*$ is interior, $\exists$ an ϵ-ball around $\mathbf{x}^*$ that lies inside D

Thus, we can move a little way in every direction without leaving D



If this is true and $\mathbf{x}^*$ is a maximizer, then f must be stationary at this point

For suppose this isn't true

Then

1. we can find an uphill direction on the graph of f
2. we can move a little way in that direction without leaving D

This contradicts $\mathbf{x}^*$ being a maximizer over all $\mathbf{x} \in D$

Similar story for minimizers

Example. Let

$$D := B_4(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| < 4\}$$

and

$$f(\mathbf{x}) = f(x_1, x_2) = x_1^2 - x_1 x_2 + 4x_2^2$$

Claim The point $\mathbf{1} := (1, 1)$ is not a maximizer of f on D

Proof: It suffices to show that $\mathbf{1}$ is interior and non-stationary

Clearly $\mathbf{1} \in D$ because $\|\mathbf{1}\| = \sqrt{1^2 + 1^2} = \sqrt{2} < 4$

Moreover $\mathbf{1}$ is interior to D because ϵ-balls are open (and so?)

Finally $\mathbf{1}$ is not stationary because $f_1'(x_1, x_2) = 2x_1 - x_2$ and hence

$$f_1'(\mathbf{1}) = f_1'(1, 1) = 2 - 1 = 1$$

Necessary Conditions

In the setting of smooth functions + interior points, stationarity is a necessary condition for maxima

- maximizer $\implies$ stationary
- not stationary $\implies$ not maximizer

When searching for maximizers, this helps us narrow down candidates

Any maximizer must be either

1. a stationary point, or
2. non-interior (i.e., on the boundary)

Example. Consider the problem $\max_{x \in D} f(x)$ where

$$f(x) = x^4 - 3x^3 - 4x^2 - x + 1, \qquad D = [-2, 4]$$

Stationary points are solutions to

$$4x^3 - 9x^2 - 8x - 1 = 0$$

One can solve this cubic (you don't need to) to find zeros at

$$x_1 = -0.153, \quad x_2 = -0.552, \quad x_3 = 2.96$$

The only possibilities for maxima are these points and $-2, 4$

Evaluating one at a time shows that $f(-2)$ is the largest

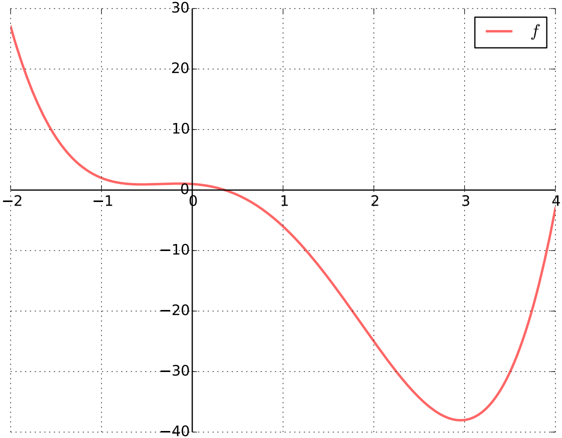


Figure : The function $f(x) = x^4 - 3x^3 - 4x^2 - x + 1$

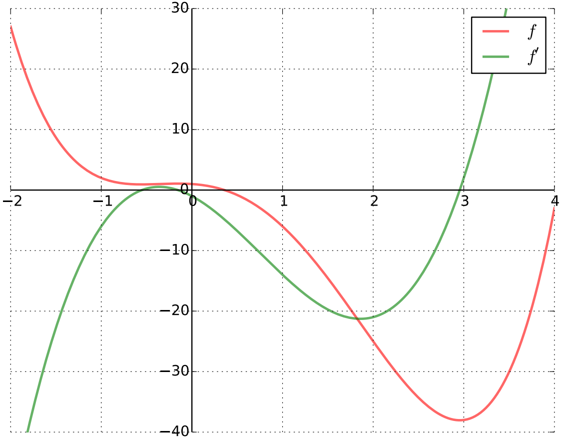


Figure : The function f and its derivative f'

Constrained Optimization Review

In a way, all optimization problems are in some sense constrained

- $\max_{\mathbf{x} \in D} f(\mathbf{x})$ constrains us to search within D

But for economists, "constrained" usually means that

1. there's some additional constraint
2. that constraint is typically binding

Examples.

- a consumer maximizing utility over their budget set
- a firm that produces at minimal cost

When constraints bind, maxima and minima are not usually stationary

If we're constrained,

- we can't move freely in every direction
- hence we can't always exploit a non-zero slope

Hence stationarity is not a necessary condition

We have to look for another one

This leads us to tangency conditions

Key Idea. When f and g are both differentiable functions on D, every solution to

$$\max_{x_1,x_2} f(x_1,x_2)$$
$$\text{s.t. } g(x_1,x_2)=0$$

in the interior of D must satisfy

$$\frac{f_1(x_1,x_2)}{f_2(x_1,x_2)}=\frac{g_1(x_1,x_2)}{g_2(x_1,x_2)}$$

For if not we can shift along the constraint to a better point

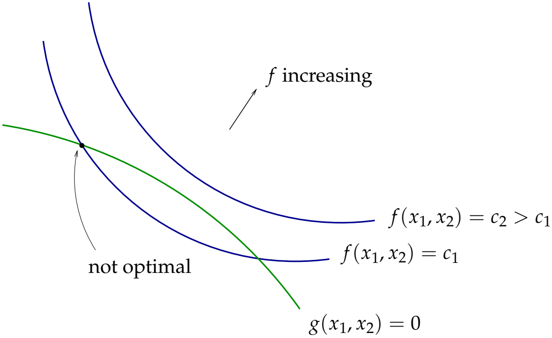


Figure : Tangency necessary for optimality

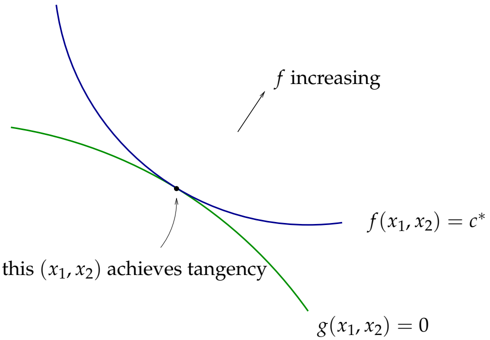


Figure : Tangency necessary for optimality

Example. Consider the problem

$$\max_{x_1,x_2} f(x_1,x_2) = x_1^{1/2} + x_2^{1/2} \quad \text{s.t.} \quad x_1^2 + x_2^2 = 1$$

and $x_i > 0$ for $i = 1,2$

Setting $g(x_1,x_2) = x_1^2 + x_2^2 - 1$, the tangency condition becomes

$$\frac{x_1^{-1/2}}{x_2^{-1/2}} = \frac{x_1}{x_2} \quad \Longleftrightarrow \quad \frac{x_1^{-3/2}}{x_2^{-3/2}} = 1 \quad \Longleftrightarrow \quad x_1 = x_2$$

Plugging this back into the constraint $x_1^2 + x_2^2 = 1$ gives

$$x_1^* = \sqrt{1/2}, \qquad x_2^* = \sqrt{1/2}$$

This is the only solution and the only candidate for maximizer

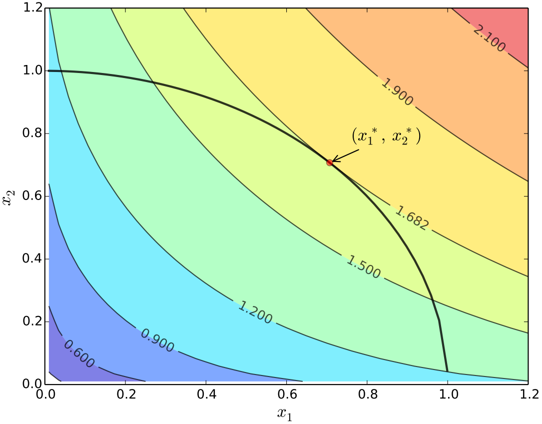


Figure : Maximizer at the tangent

Existence of Optima Review

Not every function has a maximizer / minimizer

Example. Let $\mathbf{A}$ be $N \times N$ and indefinite

If $Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$, then Q has neither a max nor min on $\mathbb{R}^N$

To see that no maximizer exists, observe that

$$\exists\, \mathbf{z} \in \mathbb{R}^N \text{ s.t. } Q(\mathbf{z}) = \mathbf{z}'\mathbf{A}\mathbf{z} > 0$$

(Otherwise $\mathbf{A}$ would be nonpositive definite)

No $\mathbf{x} \in \mathbb{R}^N$ can maximize Q because it is dominated, for sufficiently large n, by

$$Q(n\mathbf{z}) = n^2 \mathbf{z}'\mathbf{A}\mathbf{z} \to \infty$$

Even functions on bounded domains can fail to have max / min

Example. Consider maximizing $f(x) = 1/x$ on $D := (0, 1)$

No maximizer of f exists in D

Indeed, suppose to the contrary that $z \in D$ is a maximizer

Then $f(z) \geq f(x)$ for all $x \in (0, 1)$

Since $0 < z < 1$, we have $0 < z/2 < 1$, and hence $z/2 \in D$

But

$$f(z/2) = \frac{2}{z} > \frac{1}{z} = f(z)$$

Contradiction

Key Idea. Continuous functions on closed bounded sets have both maximizers and minimizers

Consider the problem

$$\max \sum_{t=1}^{T} \left(\frac{1}{2}\right)^{t} \sqrt{x_t}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} x_t \leq 1 \quad \text{and} \quad 0 \leq x_t, \quad t = 1, \ldots, T$$

This is a planning problem (similar to the one from lecture 21)

Let's show that a maximizer exists

Step 1: Let's write the constraint set as

$$D := \left\{ \mathbf{x} \in \mathbb{R}^T \ : \ \mathbf{1}'\mathbf{x} \leq 1, \ \mathbf{x} \geq \mathbf{0} \right\}$$

Claim D is closed

Let $\{\mathbf{x}_n\}$ be a sequence in D converging to some $\mathbf{x} \in \mathbb{R}^T$

We claim that $\mathbf{x} \in D$

Note first that $\mathbf{1}'\mathbf{x}_n \to \mathbf{1}'\mathbf{x}$

- because $\mathbf{x}_n \to \mathbf{x} \implies \mathbf{a}'\mathbf{x}_n \to \mathbf{a}'\mathbf{x}$ for any $\mathbf{a} \in \mathbb{R}^T$

Since $\mathbf{1}'\mathbf{x}_n \leq 1$ for all n, the same is true for $\mathbf{1}'\mathbf{x}$

- weak inequalities are preserved under limits (see lecture 16)

It remains to show that $\mathbf{x} \geq \mathbf{0}$

This also follows from preservation of weak inequalities under limits

Since $\mathbf{x}_n \in D$ for all n, we have $\mathbf{x}_n \geq \mathbf{0}$ for all n

Since $\mathbf{x} = \lim_{n \to \infty} \mathbf{x}_n$, the same is true for $\mathbf{x}$

In summary, $\mathbf{1}'\mathbf{x} \leq 1$ and $\mathbf{x} \geq \mathbf{0}$

Hence $\mathbf{x} \in D$

We conclude that the limit of any sequence in D also lies in D

Hence D is closed as claimed

Claim D is bounded

Proof: Recall that $D = \left\{\mathbf{x} \in \mathbb{R}^T \ : \ \mathbf{1}'\mathbf{x} \leq 1, \ \mathbf{x} \geq \mathbf{0}\right\}$

We need to show that

$$\exists\, M \in \mathbb{R} \quad \text{s.t.} \quad \|\mathbf{x}\| \leq M, \quad \forall\, \mathbf{x} \in D$$

This holds with $M := \sqrt{T}$ because

$$\mathbf{x} \in D \quad \Longrightarrow \quad 0 \leq x_t \leq 1, \ \forall\, t$$

and hence

$$\|\mathbf{x}\| = \sqrt{\sum_{t=1}^{T} x_t^2} \leq \sqrt{\sum_{t=1}^{T} 1} = \sqrt{T}$$

To complete the proof of existence, we need to show that

$$f(\mathbf{x}) = f(x_1, \ldots, x_T) = \sum_{t=1}^{T} \left(\frac{1}{2}\right)^t \sqrt{x_t}$$

is continuous on D

We know (lecture 17) that

- $\sqrt{\cdot}$ is a continuous function
- continuous function $\times$ scalar $=$ continuous function
- continuous $+$ continuous $=$ continuous

Hence f is a continuous function... and has a maximizer on D

Aside on Open / Closed Sets

As a rule of thumb,

- if you see strict inequalities, think "open set"
- if you see weak inequalities, think "closed set"
- if you see a mix, think "neither"

Examples.

- $(a, b) = \{x \in \mathbb{R} : a < x < b\}$ is open
- $B_\epsilon(\mathbf{a}) = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x} - \mathbf{a}\| < \epsilon\}$ is open
- $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ is closed
- $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$ is neither

Uniqueness of Optima Review

Key Idea. For functions defined on a convex set,

- a strictly concave function has at most one maximizer
- a strictly convex function has at most one minimizer

Most of the time, strict concavity / convexity are checked using derivative conditions

The most important ones are

1. positive definite Hessian $\implies f$ strictly convex
2. negative definite Hessian $\implies f$ strictly concave

Example. Above we showed existence of a maximizer in the problem

$$\max f(\mathbf{x}) = \sum_{t=1}^{T} \left(\frac{1}{2}\right)^t \sqrt{x_t}$$

$$\text{over } D := \left\{ \mathbf{x} \in \mathbb{R}^T \ : \ \mathbf{1}'\mathbf{x} \leq 1, \ \mathbf{x} \geq \mathbf{0} \right\}$$

Now let's prove uniqueness

This will be established if we can show that

- D is a convex subset of $\mathbb{R}^T$
- $f(\mathbf{x}) = \sum_{t=1}^{T} \left(\frac{1}{2}\right)^t \sqrt{x_t}$ is strictly concave on D

Regarding convexity of D, we have already shown (lecture 19) that

- $P := \{\mathbf{x} \in \mathbb{R}^T : \mathbf{x} \geq \mathbf{0}\}$ is convex
- Intersections of convex sets are convex

Moreover, $D = C \cap P$ where

$$C := \{\mathbf{x} \in \mathbb{R}^T : \mathbf{1}'\mathbf{x} \leq 1\}$$

Hence it suffices to show that C is convex, or

$$\mathbf{x}, \mathbf{y} \in C \text{ and } \lambda \in [0,1] \quad \Longrightarrow \quad \mathbf{z} := \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in C$$

This follows from $\mathbf{1}'\mathbf{x} \leq 1$ and $\mathbf{1}'\mathbf{y} \leq 1$, which gives

$$\mathbf{1}'\mathbf{z} = \lambda\mathbf{1}'\mathbf{x} + (1-\lambda)\mathbf{1}'\mathbf{y} \leq \lambda + (1-\lambda) = 1$$

It remains to show that

$$f(\mathbf{x}) = \sum_{t=1}^{T} \left(\frac{1}{2}\right)^{t} \sqrt{x_t}$$

is a strictly concave function on D

To see this, note that

$$f_{ij} := \frac{\partial}{\partial x_i \partial x_j} f(\mathbf{x}) = \begin{cases} -\left(\frac{1}{2}\right)^{i+2} x_i^{-3/2} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Let

$$\gamma_i := -\left(\frac{1}{2}\right)^{i+2} x_i^{-3/2}$$

The Hessian matrix of f at $\mathbf{x}$ is then

$$H(\mathbf{x}) := \begin{pmatrix} f_{11}(\mathbf{x}) & \cdots & f_{1T}(\mathbf{x}) \\ & \vdots & \\ f_{T1}(\mathbf{x}) & \cdots & f_{TT}(\mathbf{x}) \end{pmatrix} = \mathrm{diag}(\gamma_1, \ldots, \gamma_T)$$

Hence, for $\mathbf{z} = (z_1, \ldots, z_T) \neq \mathbf{0}$ we have

$$\mathbf{z}' H(\mathbf{x}) \mathbf{z} = \sum_{t=1}^{T} \gamma_t z_t^2 < 0$$

Hence $H(\mathbf{x})$ is negative definite

Hence f is strictly concave... and the maximizer is unique

Linear Algebra / Vector Space Review

We spent a lot of time working with vector space concepts

- span
- independence
- bases

But when we do applications it's almost always with matrices

Why do we need to think about vector spaces?

Answer: Because the concepts are clearer when we strip away matrix structure, reducing linear operations to their simplest form

Linear Combinations

$\mathbb{R}^N :=$ the set of N-tuples $\mathbf{x} = (x_1, \ldots, x_N)$ with $x_n \in \mathbb{R}$

We have two fundamental linear operations that act on vectors

1. scalar multiplication
2. vector addition

Consider a collection of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$ in $\mathbb{R}^N$

We can combine these with operations 1 & 2 to produce new vectors, such as

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_K \mathbf{x}_K$$

- $\mathbf{y}$ is a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_K$

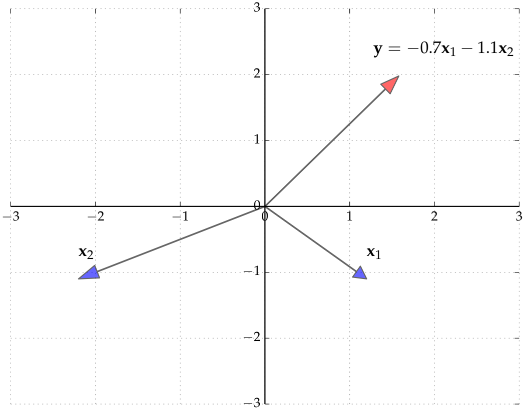


Figure : $\mathbf{y}$ is a linear combination of $\mathbf{x}_1, \mathbf{x}_2$

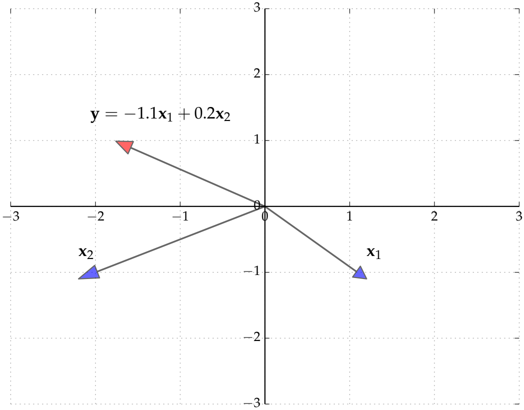


Figure : $\mathbf{y}$ is a linear combination of $\mathbf{x}_1, \mathbf{x}_2$

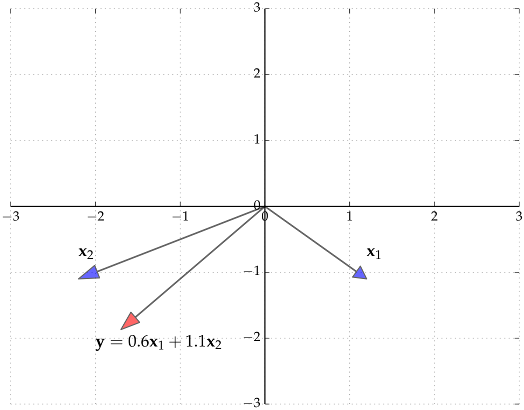


Figure : $\mathbf{y}$ is a linear combination of $\mathbf{x}_1, \mathbf{x}_2$

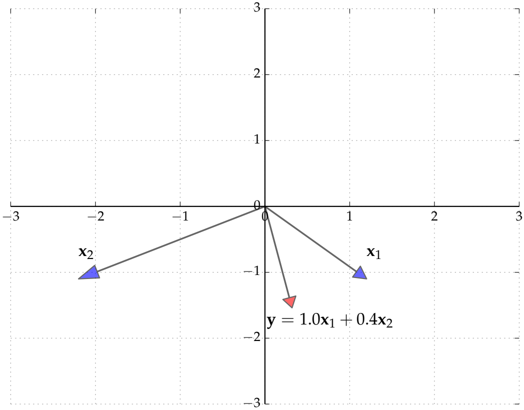


Figure : $\mathbf{y}$ is a linear combination of $\mathbf{x}_1, \mathbf{x}_2$

The **span** of $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ is the set of linear combinations we can form using these vectors

That is, $\operatorname{span}(X)$ is all vectors $\mathbf{y}$ we can create by varying the scalars in

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_K \mathbf{x}_K$$

Key Idea. You cannot span $\mathbb{R}^N$ with less than N vectors

For example, consider the case of $\mathbb{R}^3$

- The span of one vector is just a one dimensional line
- The span of two vectors is at most a two dimensional plane

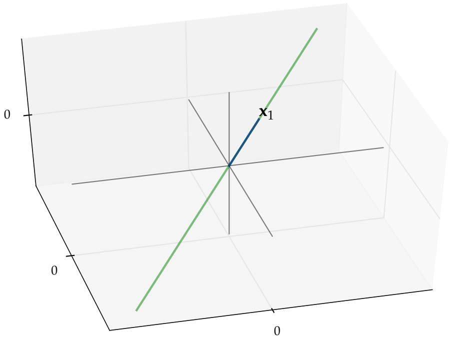


Figure : The span of $\{\mathbf{x}_1\}$ alone is a line

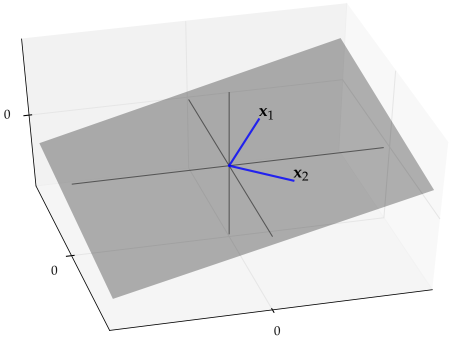


Figure : The span of $\{\mathbf{x}_1, \mathbf{x}_2\}$ is a plane

Hence we need at least three vectors to span $\mathbb{R}^3$

However, even 3 vectors won't span $\mathbb{R}^3$ if some don't contribute

For example, suppose

- we already have $\{\mathbf{x}_1, \mathbf{x}_2\}$
- we now add another vector $\mathbf{x}_3$...
- but $\mathbf{x}_3$ lies in the span of $\{\mathbf{x}_1, \mathbf{x}_2\}$

Then no overall contribution will be made

Hence we fail to span $\mathbb{R}^3$

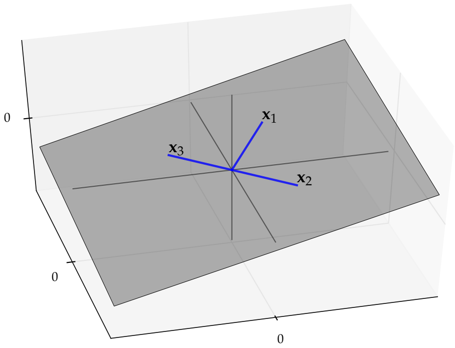


Figure : Linear dependence – the new vector $\mathbf{x}_3$ doesn't contribute

Key Idea. A set of vectors is linearly independent when they all contribute to their span

In particular,

Key Idea. For N vectors to span $\mathbb{R}^N$ they need to be linearly independent

That is, for N vectors in $\mathbb{R}^N$

$$\text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_N\} = \mathbb{R}^N \quad \Longleftrightarrow$$

$$\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \text{ linearly independent}$$

Any N linearly independent vectors in $\mathbb{R}^N$ is called a **basis** of $\mathbb{R}^N$

Key Idea. Every $\mathbf{y}$ in $\mathbb{R}^N$ has exactly one representation as a linear combination of basis vectors

That is, for any basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$,

1. Every $\mathbf{y}$ in $\mathbb{R}^N$ can be written as a linear combination

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_N \mathbf{x}_N$$

2. The representation is unique

Application: Finding Linear Combinations

Consider the following two vectors in $\mathbb{R}^2$

$$\mathbf{x}_1 = \begin{pmatrix} 1.2 \\ -1.1 \end{pmatrix}, \qquad \mathbf{x}_2 = \begin{pmatrix} -2.2 \\ -1.1 \end{pmatrix}$$

Given arbitrary $\mathbf{y}$ in $\mathbb{R}^2$, can we always find scalars α_1, α_2 such that

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2$$

If so, how can I compute them?

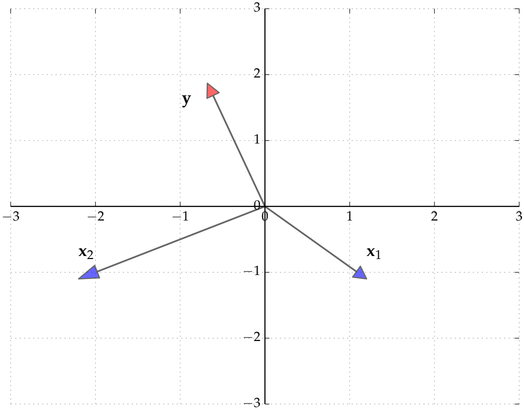


Figure : Can any $\mathbf{y} \in \mathbb{R}^2$ be realized as a linear combination of $\mathbf{x}_1, \mathbf{x}_2$?

By the preceding discussion, if $\{\mathbf{x}_1, \mathbf{x}_2\}$ is linearly independent, then yes

In particular,

$$\{\mathbf{x}_1, \mathbf{x}_2\} \text{ is linearly independent} \iff \{\mathbf{x}_1, \mathbf{x}_2\} \text{ is a basis of } \mathbb{R}^2$$

In this case,

$$\forall\, \mathbf{y} \in \mathbb{R}^2, \;\; \exists \;\text{ unique pair } \alpha_1, \alpha_2 \;\text{ s.t. }\; \mathbf{y} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2$$

How can we check whether $\{\mathbf{x}_1, \mathbf{x}_2\}$ is linearly independent?

Recall: This will be true iff

$$\alpha_1 \begin{pmatrix} 1.2 \\ -1.1 \end{pmatrix} + \alpha_2 \begin{pmatrix} -2.2 \\ -1.1 \end{pmatrix} = \mathbf{0} \quad \Longrightarrow \quad \alpha_1 = \alpha_2 = 0$$

That is,

$$\begin{aligned} 1.2\alpha_1 &= 2.2\alpha_2 \\ -1.1\alpha_1 &= 1.1\alpha_2 \end{aligned} \quad \Longrightarrow \quad \alpha_1 = \alpha_2 = 0$$

This is true: If both equations on the left hold then

$$\alpha_1 = -\alpha_2 \quad \text{and} \quad \alpha_1 = (2.2/1.2)\alpha_2$$

The only possibility is that $\alpha_1 = \alpha_2 = 0$

Hence $\{\mathbf{x}_1, \mathbf{x}_2\}$ is a basis of $\mathbb{R}^2$

In particular, for any given $\mathbf{y} \in \mathbb{R}^2$, there is a unique pair of scalars α_1, α_2 such that

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2$$

Remaining question: how to compute α_1, α_2?

Make a matrix with $\mathbf{x}_1$ and $\mathbf{x}_2$ as its columns

$$\mathbf{X} := \begin{pmatrix} 1.2 & -2.2 \\ -1.1 & -1.1 \end{pmatrix}$$

Given $\mathbf{y} \in \mathbb{R}^2$ we seek α_1, α_2 such that $\mathbf{y} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2$

Equivalently, we see α_1, α_2 such that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1.2 & -2.2 \\ -1.1 & -1.1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

How to solve for (α_1, α_2)?

Since $\mathbf{X}$ is nonsingular (why?), the solution is

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1.2 & -2.2 \\ -1.1 & -1.1 \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$= \frac{1}{-1.32 - 2.42} \begin{pmatrix} -1.1 & 2.2 \\ 1.1 & 1.2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

The general problem: Solve a system of linear equations

Given square matrix $\mathbf{X}$ and vector $\mathbf{y}$, can we find $\boldsymbol{\alpha}$ such that

$$\mathbf{X}\boldsymbol{\alpha} = \mathbf{y}$$

This is the same problem as finding scalars α_i such that

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_N \mathbf{x}_N, \qquad \mathbf{x}_i = i\text{-th column of } \mathbf{X}$$

If $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ linearly independent, they form a basis of $\mathbb{R}^N$, and

1. we can always find such scalars (existence)
2. we only find one such set of scalars (uniqueness)
3. they are equal to $\mathbf{X}^{-1}\mathbf{y}$