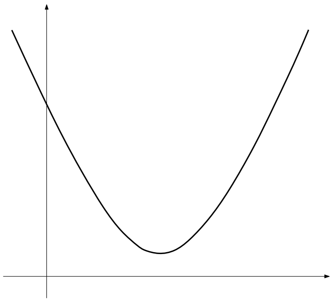
Convex Analysis at Infinity

An Introduction to Astral Space

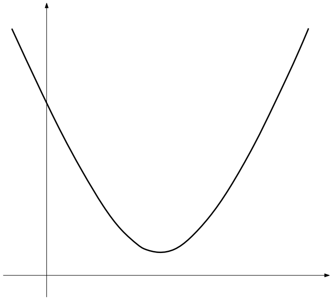
Miro Dudík
Rob Schapire
Matus Telgarsky

further reading at: aka.ms/astral

Convex functions

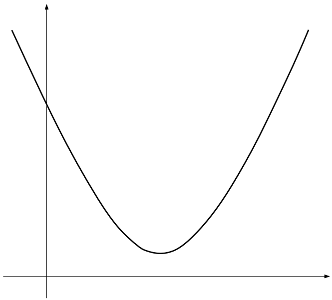


Convex functions



- minimizing convex functions is basis for many methods in machine learning and statistics (and other fields)
  - e.g.: maximum likelihood, maximum entropy, linear regression, logistic regression, boosting, SVM's, ...

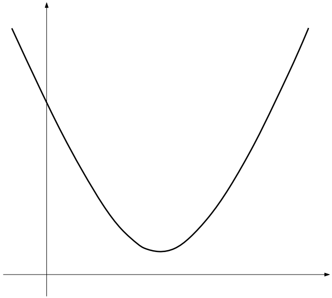
Convex functions



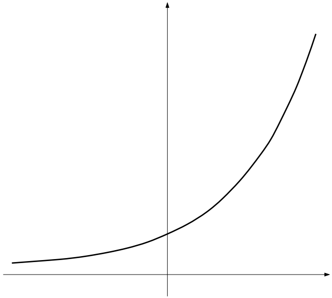
- minimizing convex functions is basis for many methods in machine learning and statistics (and other fields)
  - e.g.: maximum likelihood, maximum entropy, linear regression, logistic regression, boosting, SVM's, ...
- convex functions are really nice!
  - local minimum must be global minimum
  - if gradient $= 0$ then must be global minimum
  - usually easier to find and analyze minimization algorithms
  - beautiful properties

Finite minimizers

- common to assume function actually attains minimum at a finite point

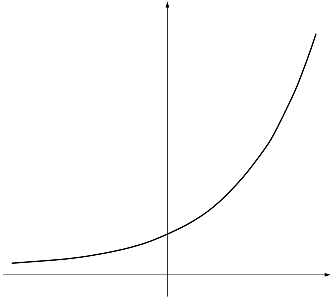


Finite minimizers



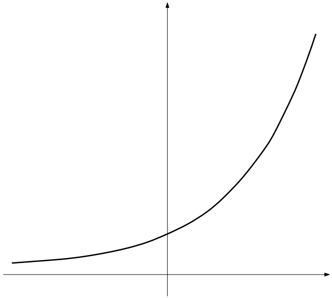
- common to assume function actually attains minimum at a finite point
- problem: some convex functions have no finite minimizer
  - function then must be minimized by sequence heading "to infinity"

Finite minimizers



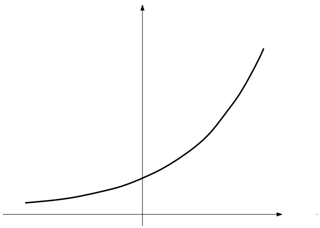
- common to assume function actually attains minimum at a finite point
- problem: some convex functions have no finite minimizer
  - function then must be minimized by sequence heading "to infinity"
  - certainly includes cases of practical interest
  - analyzing convergence often requires carefully tailored techniques

Finite minimizers



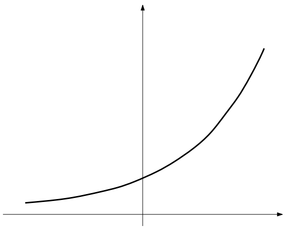
- common to assume function actually attains minimum at a finite point
- problem: some convex functions have no finite minimizer
  - function then must be minimized by sequence heading "to infinity"
  - certainly includes cases of practical interest
  - analyzing convergence often requires carefully tailored techniques
- this talk: develop theory for studying such minimizers at infinity

Example: exponential function



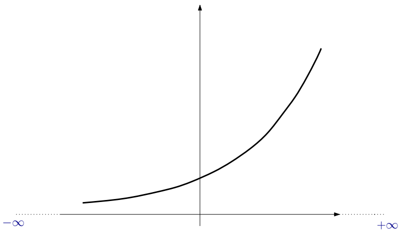
- e.g.: $f(x) = e^x$

Example: exponential function



- e.g.: $f(x) = e^x$
- no finite point $x \in \mathbb{R}$ where minimum attained
- instead, minimized by any sequence (x_t) with $x_t \to -\infty$
- wish could say: "minimized at $-\infty$"

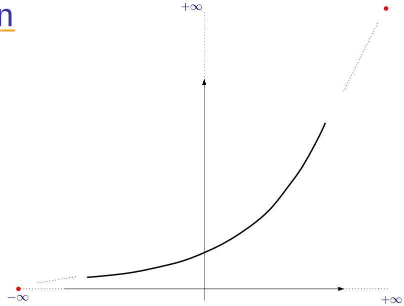
Example: exponential function



- e.g.: $f(x) = e^x$
- no finite point $x \in \mathbb{R}$ where minimum attained
- instead, minimized by any sequence (x_t) with $x_t \to -\infty$
- wish could say: "minimized at $-\infty$"
- can do by:
  - extending $\mathbb{R}$ to include $\pm\infty$:

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} = [-\infty, +\infty]$$

Example: exponential function

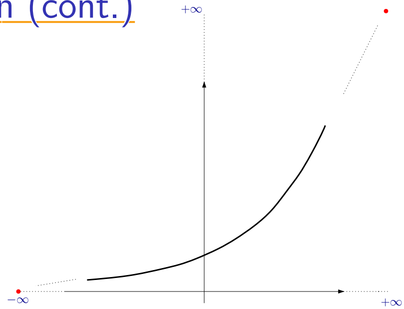


- e.g.: $f(x) = e^x$
- no finite point $x \in \mathbb{R}$ where minimum attained
- instead, minimized by any sequence (x_t) with $x_t \to -\infty$
- wish could say: "minimized at $-\infty$"
- can do by:
  - extending $\mathbb{R}$ to include $\pm\infty$:

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} = [-\infty, +\infty]$$

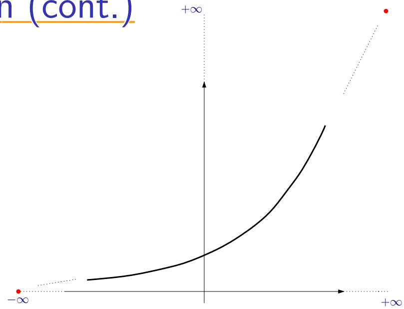
  - extending f to $\overline{\mathbb{R}}$ by setting:
    $f(-\infty) = 0$ and $f(+\infty) = +\infty$

Example: exponential function (cont.)



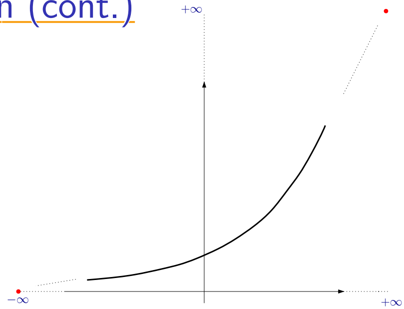
- extended f:
  - is continuous over $\overline{\mathbb{R}}$
  - attains minimum at $-\infty$
  - "feels" convex
  - maybe can extend derivatives so that $f'(-\infty) = 0$

Example: exponential function (cont.)



- extended f:
  - is continuous over $\overline{\mathbb{R}}$
  - attains minimum at $-\infty$
  - "feels" convex
  - maybe can extend derivatives so that $f'(-\infty) = 0$
- in $n = 1$ dimensions, seems clear how to
  - add "points at infinity"
  - extend functions to enlarged space, capturing minimizers at infinity

Example: exponential function (cont.)



- extended f:
  - is continuous over $\overline{\mathbb{R}}$
  - attains minimum at $-\infty$
  - "feels" convex
  - maybe can extend derivatives so that $f'(-\infty) = 0$
- in $n = 1$ dimensions, seems clear how to
  - add "points at infinity"
  - extend functions to enlarged space, capturing minimizers at infinity
- what about in $n \geq 2$ dimensions?

Goals

- extend $\mathbb{R}^n$ to include points at infinity

Goals

- extend $\mathbb{R}^n$ to include points at infinity
- aim to reveal structure of how convex functions behave at infinity

Goals

- extend $\mathbb{R}^n$ to include points at infinity
- aim to reveal structure of how convex functions behave at infinity
- trying to build up foundations so relatively easy, for instance, to prove convergence of algorithms

Goals

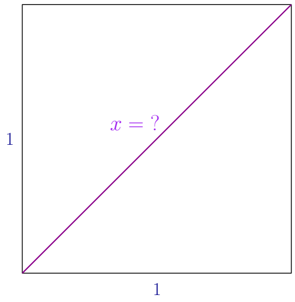
- extend $\mathbb{R}^n$ to include points at infinity
- aim to reveal structure of how convex functions behave at infinity
- trying to build up foundations so relatively easy, for instance, to prove convergence of algorithms
- hope to make study of convex functions more "complete" and "regular"
  - e.g., so every convex function, when extended to new space, has a minimizer

Goals

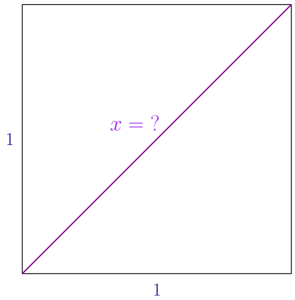
- extend $\mathbb{R}^n$ to include points at infinity
- aim to reveal structure of how convex functions behave at infinity
- trying to build up foundations so relatively easy, for instance, to prove convergence of algorithms
- hope to make study of convex functions more "complete" and "regular"
  - e.g., so every convex function, when extended to new space, has a minimizer
- want compatible with key notions of convex analysis

Analogy

- if only working in $\mathbb{Q} =$ rationals, then no number equals length of diagonal of a unit square



Analogy

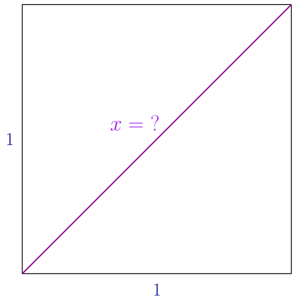


- if only working in $\mathbb{Q} =$ rationals, then no number equals length of diagonal of a unit square
- can approach value with sequences in $\mathbb{Q}$:

  $\frac{1}{1},\ \frac{3}{2},\ \frac{7}{5},\ \frac{17}{12},\ \frac{41}{29},\ \frac{99}{70},\ \frac{239}{169},\ \ldots \to \sqrt{2}$

  $1,\ 1.4,\ 1.41,\ 1.414,\ 1.4142,\ 1.41421,\ \ldots \to \sqrt{2}$

Analogy

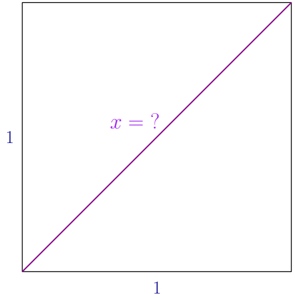


- if only working in $\mathbb{Q}$ = rationals, then no number equals length of diagonal of a unit square
- can approach value with sequences in $\mathbb{Q}$:

  $\frac{1}{1}, \frac{3}{2}, \frac{7}{5}, \frac{17}{12}, \frac{41}{29}, \frac{99}{70}, \frac{239}{169}, \ldots \to \sqrt{2}$

  $1, 1.4, 1.41, 1.414, 1.4142, 1.41421, \ldots \to \sqrt{2}$
- so don't "need" reals
  (can just work over sequences of rationals)

Analogy



- if only working in $\mathbb{Q}$ = rationals, then no number equals length of diagonal of a unit square
- can approach value with sequences in $\mathbb{Q}$:

  $\frac{1}{1}, \frac{3}{2}, \frac{7}{5}, \frac{17}{12}, \frac{41}{29}, \frac{99}{70}, \frac{239}{169}, \ldots \to \sqrt{2}$

  $1, 1.4, 1.41, 1.414, 1.4142, 1.41421, \ldots \to \sqrt{2}$
- so don't "need" reals (can just work over sequences of rationals)
- far preferable to extend $\mathbb{Q}$ to $\mathbb{R}$
  - much more complete, regular, well-structured

Analogy (cont.)

- in same way, can continue to use sequences to study minimizers of convex functions
  - might be much nicer to study minimizers at infinity as mathematical objects in their own right
  - can hope larger space would be more complete, regular, and revealing of structure

This work

- introduce astral space, extension of $\mathbb{R}^n$ with points at infinity
- extend functions on $\mathbb{R}^n$ to astral space
- study key properties and topics extended to astral space, especially from convex analysis

Outline

- what can minimizers at infinity look like?
- constructing astral space
- what are astral points like?
- extending functions to astral space
- convergence of iterative algorithms

Notation

- $n =$ dimension
- scalars (in $\mathbb{R}$): $x, y, \ldots$
- vectors (in $\mathbb{R}^n$): $\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}, \ldots$
  - as tuple: $\boldsymbol{x} = (x_1, \ldots, x_n)$

Notation

- $n =$ dimension
- scalars (in $\mathbb{R}$): $x, y, \ldots$
- vectors (in $\mathbb{R}^n$): $\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}, \ldots$
  - as tuple: $\boldsymbol{x} = (x_1, \ldots, x_n)$
- all sequences indexed by $t = 1, 2, \ldots$
- limits and convergence always as $t \to +\infty$
- $(\boldsymbol{x}_t)$ is sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$

Outline

- what can minimizers at infinity look like?
- constructing astral space
- what are astral points like?
- extending functions to astral space
- convergence of iterative algorithms

Minimizers at infinity

- given convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- if no finite minimizer, can only be minimized by sequence $(\mathbf{x}_t)$ to infinity
- what can such "minimizers at infinity" look like?

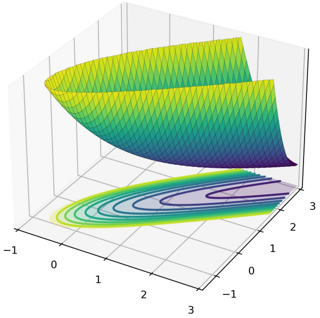
Minimizers at infinity

- given convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- if no finite minimizer, can only be minimized by sequence $(\mathbf{x}_t)$ to infinity
- what can such "minimizers at infinity" look like?
- in $n = 1$ dimensions, can only converge to $\pm\infty$

Minimizers at infinity

- given convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- if no finite minimizer, can only be minimized by sequence $(\mathbf{x}_t)$ to infinity
- what can such "minimizers at infinity" look like?
- in $n = 1$ dimensions, can only converge to $\pm\infty$
- in $n \geq 2$ dimensions, many possibilities
  - for example...

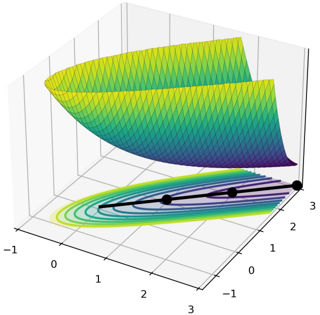
Example: Diagonal valley



- in $\mathbb{R}^2$, say

$$f(\boldsymbol{x}) = f(x_1, x_2) = \underbrace{e^{-x_1}} + \underbrace{(x_2 - x_1)^2}$$

Example: Diagonal valley

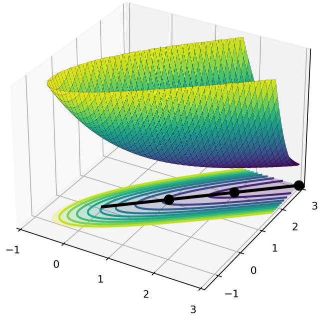


- in $\mathbb{R}^2$, say

$$f(\boldsymbol{x}) = f(x_1, x_2) = \underbrace{e^{-x_1}}_{\substack{\Downarrow \\ x_1 \to +\infty}} + \underbrace{(x_2 - x_1)^2}_{\substack{\Downarrow \\ x_2 - x_1 \to 0}}$$

- to minimize, must follow "diagonal valley"
  - e.g., set $x_1 = x_2 = t$ and let $t \to +\infty$
  - i.e., $\boldsymbol{x}_t = (t, t)$: $f(\boldsymbol{x}_t) = f(t, t) = e^{-t} \to 0$

Example: Diagonal valley

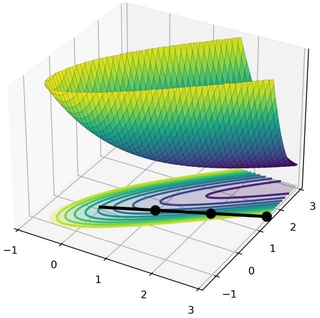


- in $\mathbb{R}^2$, say

$$f(\boldsymbol{x}) = f(x_1, x_2) = \underbrace{e^{-x_1}}_{\substack{\Downarrow \\ x_1 \to +\infty}} + \underbrace{(x_2 - x_1)^2}_{\substack{\Downarrow \\ x_2 - x_1 \to 0}}$$

- to minimize, must follow "diagonal valley"
  - e.g., set $x_1 = x_2 = t$ and let $t \to +\infty$
  - i.e., $\boldsymbol{x}_t = (t, t)$: $f(\boldsymbol{x}_t) = f(t, t) = e^{-t} \to 0$
- matters how sequence goes to infinity!

Example: Diagonal valley

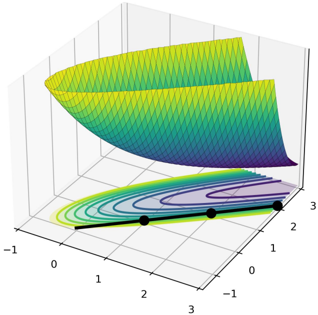


- in $\mathbb{R}^2$, say

$$f(\boldsymbol{x}) = f(x_1, x_2) = \underbrace{e^{-x_1}}_{\substack{\Downarrow \\ x_1 \to +\infty}} + \underbrace{(x_2 - x_1)^2}_{\substack{\Downarrow \\ x_2 - x_1 \to 0}}$$

- to minimize, must follow "diagonal valley"
  - e.g., set $x_1 = x_2 = t$ and let $t \to +\infty$
  - i.e., $\boldsymbol{x}_t = (t, t)$: $f(\boldsymbol{x}_t) = f(t, t) = e^{-t} \to 0$
- matters how sequence goes to infinity!
  - direction matters

Example: Diagonal valley

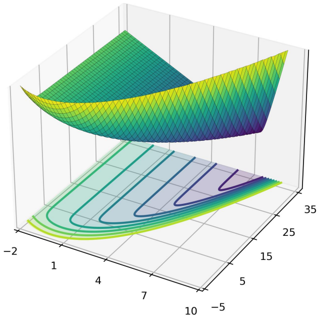


- in $\mathbb{R}^2$, say

$$f(\boldsymbol{x}) = f(x_1, x_2) = \underbrace{e^{-x_1}}_{\substack{\Downarrow \\ x_1 \to +\infty}} + \underbrace{(x_2 - x_1)^2}_{\substack{\Downarrow \\ x_2 - x_1 \to 0}}$$

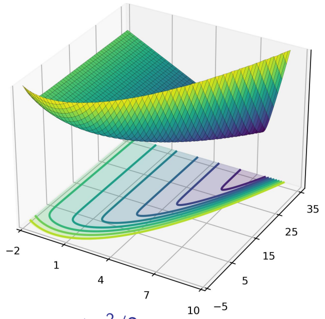
- to minimize, must follow "diagonal valley"
  - e.g., set $x_1 = x_2 = t$ and let $t \to +\infty$
  - i.e., $\boldsymbol{x}_t = (t, t)$: $f(\boldsymbol{x}_t) = f(t, t) = e^{-t} \to 0$
- matters how sequence goes to infinity!
  - direction matters
  - offset also matters

Example: Two-speed exponential



- can every convex function be minimized along a ray?

Example: Two-speed exponential

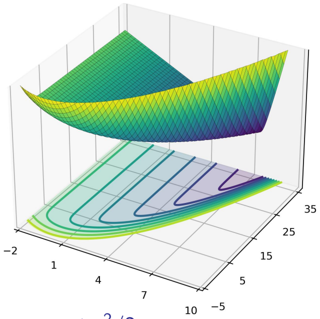


- can every convex function be minimized along a ray? no!
- e.g., in $\mathbb{R}^2$, let

$$f(\boldsymbol{x}) = f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

- to minimize, need:
  - $x_1 \to +\infty$

Example: Two-speed exponential

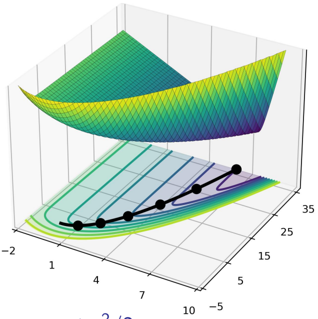


- can every convex function be minimized along a ray? no!
- e.g., in $\mathbb{R}^2$, let

$$f(\boldsymbol{x}) = f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

- to minimize, need:
  - $x_1 \to +\infty$
  - $x_2 \to +\infty$ much faster than x_1 so that $-x_2 + x_1^2/2 \to -\infty$

Example: Two-speed exponential

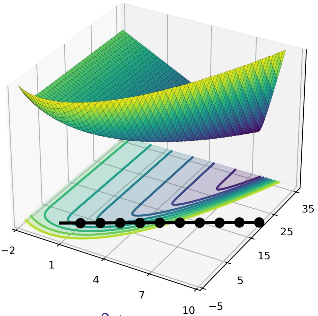


- can every convex function be minimized along a ray? no!
- e.g., in $\mathbb{R}^2$, let

$$f(\boldsymbol{x}) = f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

- to minimize, need:
  - $x_1 \to +\infty$
  - $x_2 \to +\infty$ much faster than x_1 so that $-x_2 + x_1^2/2 \to -\infty$
  - e.g. $\boldsymbol{x}_t = (t, t^2)$

Example: Two-speed exponential

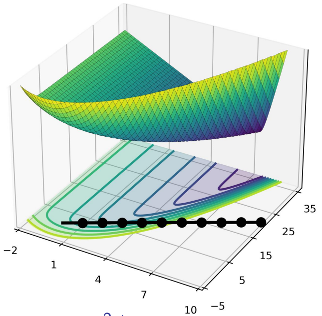


- can every convex function be minimized along a ray? no!
- e.g., in $\mathbb{R}^2$, let

$$f(\boldsymbol{x}) = f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

- to minimize, need:
  - $x_1 \to +\infty$
  - $x_2 \to +\infty$ much faster than x_1 so that $-x_2 + x_1^2/2 \to -\infty$
  - e.g. $\boldsymbol{x}_t = (t, t^2)$
- no minimizing sequence along straight ray

Example: Two-speed exponential



- can every convex function be minimized along a ray? no!
- e.g., in $\mathbb{R}^2$, let

$$f(\boldsymbol{x}) = f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

- to minimize, need:
  - $x_1 \to +\infty$
  - $x_2 \to +\infty$ much faster than x_1 so that $-x_2 + x_1^2/2 \to -\infty$
  - e.g. $\boldsymbol{x}_t = (t, t^2)$
- no minimizing sequence along straight ray
- how to construct space capturing such minimizers at infinity?

Outline

- what can minimizers at infinity look like?
- constructing astral space
- what are astral points like?
- extending functions to astral space
- convergence of iterative algorithms

Basic idea

- sequences to infinity don't converge because
  nothing to converge to

Basic idea

- sequences to infinity don't converge because nothing to converge to
- idea: add "new" points to $\mathbb{R}^n$ that can be limits of such sequences

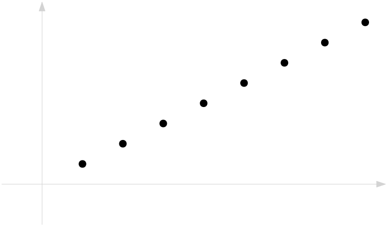
Basic idea

- sequences to infinity don't converge because nothing to converge to
- idea: add "new" points to $\mathbb{R}^n$ that can be limits of such sequences
- key questions:
  - which sequences should have limits?
  - when should two sequences have same limit?

Basic idea

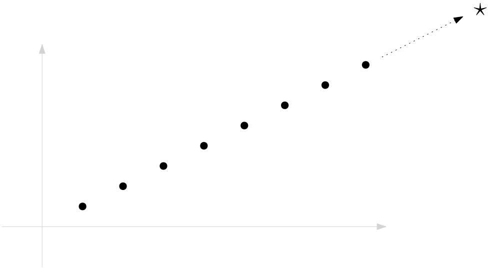
- sequences to infinity don't converge because nothing to converge to
- idea: add "new" points to $\mathbb{R}^n$ that can be limits of such sequences
- key questions:
  - which sequences should have limits?
  - when should two sequences have same limit?
- once answered, can construct space:
  - add "new" points to be limits of each group of sequences that should all have same limit

Examples and intuitions



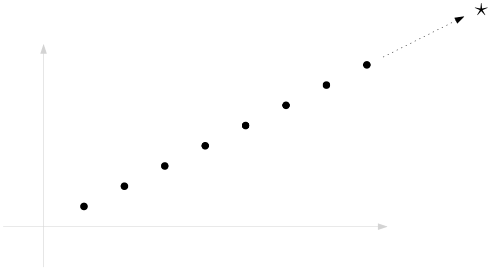
- any sequence following ray to infinity should have a limit
  e.g., $\mathbf{x}_t = (2t, t) = t\mathbf{v}$ where $\mathbf{v} = (2, 1)$

Examples and intuitions



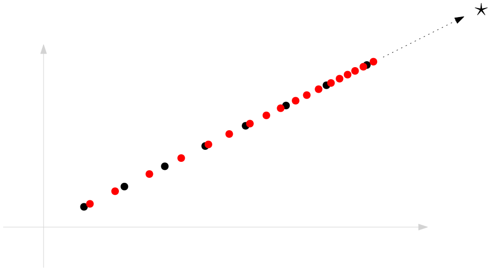
- any sequence following ray to infinity should have a limit
  e.g., $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$

Examples and intuitions



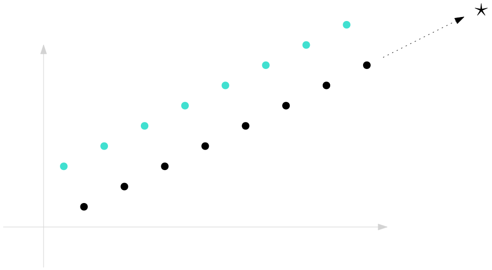
- any sequence following ray to infinity should have a limit
  e.g., $\mathbf{x}_t = (2t, t) = t\mathbf{v}$ where $\mathbf{v} = (2, 1)$
- what sequences should have same limit?

Examples and intuitions



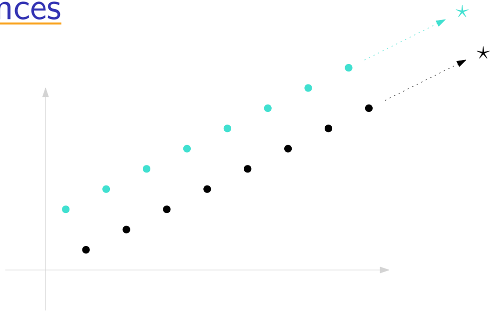
- any sequence following ray to infinity should have a limit
  e.g., $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
- what sequences should have same limit?
  - e.g., if change rate converging to infinity

Example: parallel sequences



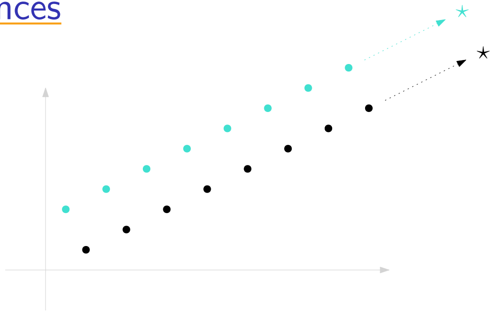
- say shift sequence by fixed offset
- $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
  $\boldsymbol{x}'_t = (2t - 1, t + 2) = t\boldsymbol{v} + \boldsymbol{w}$ where $\boldsymbol{w} = (-1, 2)$

Example: parallel sequences



- say shift sequence by fixed offset
- $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
  $\boldsymbol{x}'_t = (2t - 1, t + 2) = t\boldsymbol{v} + \boldsymbol{w}$ where $\boldsymbol{w} = (-1, 2)$
- should two sequences have same limit?

Example: parallel sequences



- say shift sequence by fixed offset
- $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
  $\boldsymbol{x}'_t = (2t-1, t+2) = t\boldsymbol{v} + \boldsymbol{w}$ where $\boldsymbol{w} = (-1, 2)$
- should two sequences have same limit?
- we believe no because:
  - offset matters for minimization
  - in applications, often care about such offsets, not just overall direction of minimization

A basic principle

- how to capture these intuitions?

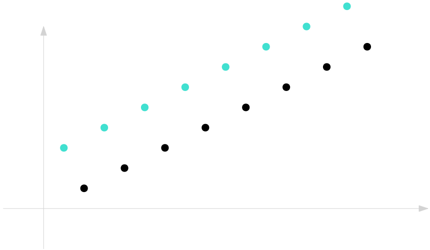
A basic principle

- how to capture these intuitions?
- basic principle: focus on limits in every direction $\boldsymbol{u} \in \mathbb{R}^n$
  i.e., along one-dimensional projections of the sequence

A basic principle

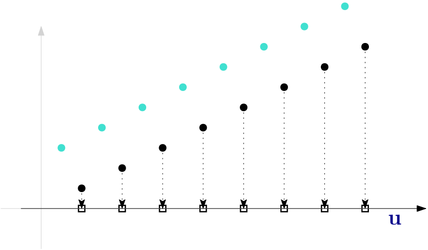
- how to capture these intuitions?
- basic principle: focus on limits in every direction $\boldsymbol{u} \in \mathbb{R}^n$
  i.e., along one-dimensional projections of the sequence
  - for example...

Example: parallel sequences



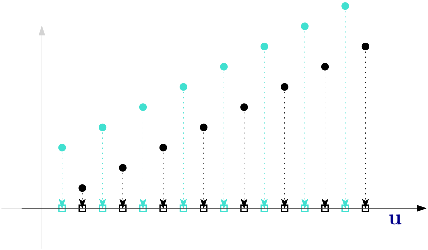
- as before:
  $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
  $\boldsymbol{x}'_t = (2t - 1, t + 2) = t\boldsymbol{v} + \boldsymbol{w}$ where $\boldsymbol{w} = (-1, 2)$

Example: parallel sequences



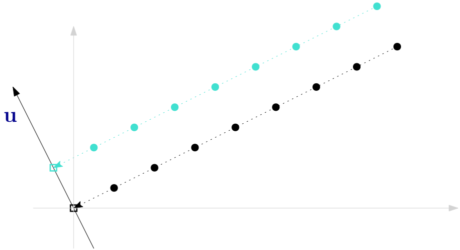
- as before:
  $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
  $\boldsymbol{x}'_t = (2t - 1, t + 2) = t\boldsymbol{v} + \boldsymbol{w}$ where $\boldsymbol{w} = (-1, 2)$
- e.g. $\boldsymbol{u} = (1, 0)$
  $\boldsymbol{x}_t \cdot \boldsymbol{u} = 2t \to +\infty$

Example: parallel sequences



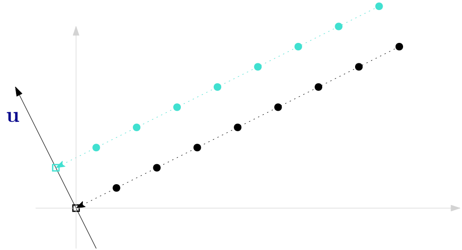
- as before:
  $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
  $\boldsymbol{x}'_t = (2t-1, t+2) = t\boldsymbol{v} + \boldsymbol{w}$ where $\boldsymbol{w} = (-1, 2)$
- e.g. $\boldsymbol{u} = (1, 0)$
  $\boldsymbol{x}_t \cdot \boldsymbol{u} = 2t \to +\infty$
  $\boldsymbol{x}'_t \cdot \boldsymbol{u} = 2t - 1 \to +\infty$

Example: parallel sequences



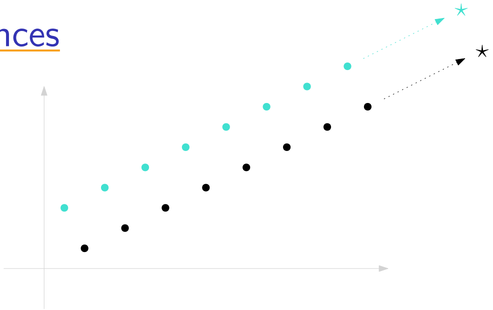
- as before:
  $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
  $\boldsymbol{x}'_t = (2t - 1, t + 2) = t\boldsymbol{v} + \boldsymbol{w}$ where $\boldsymbol{w} = (-1, 2)$
- e.g. $\boldsymbol{u} = (1, 0)$
  $\boldsymbol{x}_t \cdot \boldsymbol{u} = 2t \to +\infty$
  $\boldsymbol{x}'_t \cdot \boldsymbol{u} = 2t - 1 \to +\infty$
- e.g. $\boldsymbol{u} = (-1, 2)$
  $\boldsymbol{x}_t \cdot \boldsymbol{u} = 0 \to 0$
  $\boldsymbol{x}'_t \cdot \boldsymbol{u} = 5 \to 5$

Example: parallel sequences



- as before:
  $\boldsymbol{x}_t = (2t, t) = t\boldsymbol{v}$ where $\boldsymbol{v} = (2, 1)$
  $\boldsymbol{x}'_t = (2t - 1, t + 2) = t\boldsymbol{v} + \boldsymbol{w}$ where $\boldsymbol{w} = (-1, 2)$
- e.g. $\boldsymbol{u} = (1, 0)$
  $\boldsymbol{x}_t \cdot \boldsymbol{u} = 2t \to +\infty$
  $\boldsymbol{x}'_t \cdot \boldsymbol{u} = 2t - 1 \to +\infty$
- e.g. $\boldsymbol{u} = (-1, 2)$
  $\boldsymbol{x}_t \cdot \boldsymbol{u} = 0 \to 0$
  $\boldsymbol{x}'_t \cdot \boldsymbol{u} = 5 \to 5$
- so: in some direction, sequences have different limits

Example: parallel sequences



- as before:
  $x_t = (2t, t) = t v$ where $v = (2, 1)$
  $x'_t = (2t - 1, t + 2) = t v + w$ where $w = (-1, 2)$
- e.g. $u = (1, 0)$
  $x_t \cdot u = 2t \to +\infty$
  $x'_t \cdot u = 2t - 1 \to +\infty$
- e.g. $u = (-1, 2)$
  $x_t \cdot u = 0 \to 0$
  $x'_t \cdot u = 5 \to 5$
- so: in some direction, sequences have different limits
- therefore: require (x_t) and (x'_t) to have different limits

Our approach

- which sequences $(\boldsymbol{x}_t)$ should have limits?

Our approach

- which sequences $(\mathbf{x}_t)$ should have limits?
  - exactly those that converge in all directions
  - meaning: $\lim(\mathbf{x}_t \cdot \mathbf{u})$ exists for all $\mathbf{u} \in \mathbb{R}^n$

Our approach

- which sequences $(\boldsymbol{x}_t)$ should have limits?
  - exactly those that converge in all directions
  - meaning: $\lim(\boldsymbol{x}_t \cdot \boldsymbol{u})$ exists for all $\boldsymbol{u} \in \mathbb{R}^n$
- when should two sequences $(\boldsymbol{x}_t)$ and $(\boldsymbol{x}'_t)$ have same limit?

Our approach

- which sequences $(\boldsymbol{x}_t)$ should have limits?
  - exactly those that converge in all directions
  - meaning: $\lim(\boldsymbol{x}_t \cdot \boldsymbol{u})$ exists for all $\boldsymbol{u} \in \mathbb{R}^n$
- when should two sequences $(\boldsymbol{x}_t)$ and $(\boldsymbol{x}'_t)$ have same limit?
  - exactly when they are all-directions equivalent, i.e., have same limit in every direction
  - meaning: $\lim(\boldsymbol{x}_t \cdot \boldsymbol{u}) = \lim(\boldsymbol{x}'_t \cdot \boldsymbol{u})$ for all $\boldsymbol{u} \in \mathbb{R}^n$

Our approach

- which sequences $(\boldsymbol{x}_t)$ should have limits?
  - exactly those that converge in all directions
  - meaning: $\lim(\boldsymbol{x}_t \cdot \boldsymbol{u})$ exists for all $\boldsymbol{u} \in \mathbb{R}^n$
- when should two sequences $(\boldsymbol{x}_t)$ and $(\boldsymbol{x}'_t)$ have same limit?
  - exactly when they are all-directions equivalent, i.e., have same limit in every direction
  - meaning: $\lim(\boldsymbol{x}_t \cdot \boldsymbol{u}) = \lim(\boldsymbol{x}'_t \cdot \boldsymbol{u})$ for all $\boldsymbol{u} \in \mathbb{R}^n$
- note that limits can be in $\overline{\mathbb{R}}$

Astral space

- when expand $\mathbb{R}^n$ according to these criteria, get

$$\overline{\mathbb{R}^n} = \text{astral space}$$

Astral space

- when expand $\mathbb{R}^n$ according to these criteria, get

$$\overline{\mathbb{R}^n} = \text{astral space}$$

- in $n = 1$ dimensions, only add $\pm\infty$ so $\overline{\mathbb{R}^1}$ same as $\overline{\mathbb{R}}$

Some topological properties

- astral space is compact
  - very powerful property
    (e.g. to prove convergence of iterative algorithms)

Some topological properties

- astral space is compact
  - very powerful property
    (e.g. to prove convergence of iterative algorithms)
- every linear function can be extended continuously

Some topological properties

- astral space is compact
  - very powerful property
    (e.g. to prove convergence of iterative algorithms)
- every linear function can be extended continuously
  - implies "anything linear" likely to behave "nicely" in astral space
    - especially relevant to convex analysis
    - means can generalize: linear maps, hyperplanes, halfspaces, convex sets and functions, conjugates, differential theory

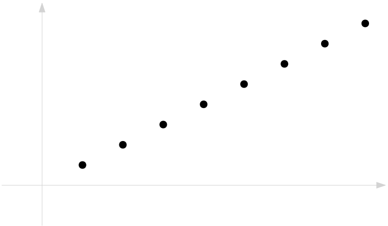
Some topological properties

- astral space is compact
  - very powerful property
    (e.g. to prove convergence of iterative algorithms)
- every linear function can be extended continuously
  - implies "anything linear" likely to behave "nicely" in astral space
    - especially relevant to convex analysis
    - means can generalize: linear maps, hyperplanes, halfspaces, convex sets and functions, conjugates, differential theory
- astral space is not a vector space, nor a metric space

Outline

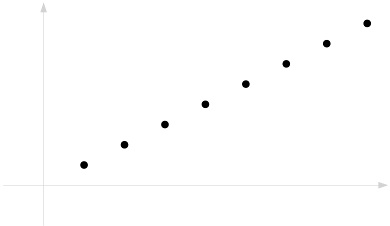
- what can minimizers at infinity look like?
- constructing astral space
- what are astral points like?
- extending functions to astral space
- convergence of iterative algorithms

Astrons



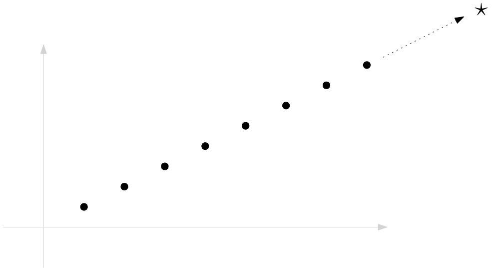
- simplest sequence to infinity is along ray from origin: $\mathbf{x}_t = t\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^n$

Astrons



- simplest sequence to infinity is along ray from origin: $\mathbf{x}_t = t\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^n$
- converges in all directions

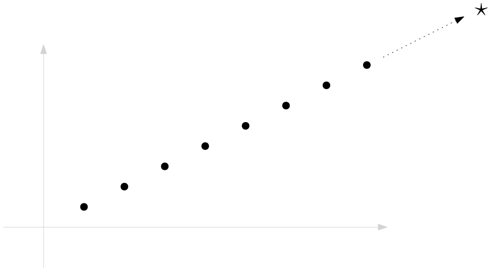
Astrons



- simplest sequence to infinity is along ray from origin: $\mathbf{x}_t = t\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^n$
- converges in all directions
- therefore, has limit in $\overline{\mathbb{R}^n}$ called astron $\omega\mathbf{v}$:

$$\omega\mathbf{v} = \lim t\mathbf{v}$$

Astrons

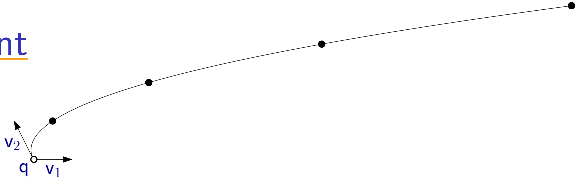


- simplest sequence to infinity is along ray from origin: $\mathbf{x}_t = t\mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^n$
- converges in all directions
- therefore, has limit in $\overline{\mathbb{R}^n}$ called astron $\omega\mathbf{v}$:

$$\omega\mathbf{v} = \lim t\mathbf{v}$$

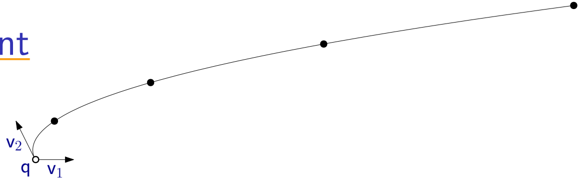
- turn out to be building blocks for all astral points

Example astral point



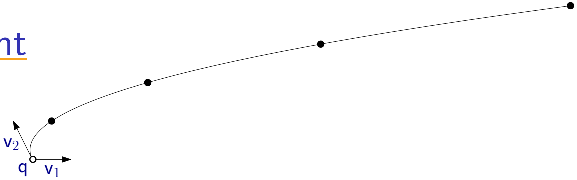
- say $\boldsymbol{x}_t = t^2 \boldsymbol{v}_1 + t\boldsymbol{v}_2 + \boldsymbol{q}$ for some $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{q} \in \mathbb{R}^n$

Example astral point



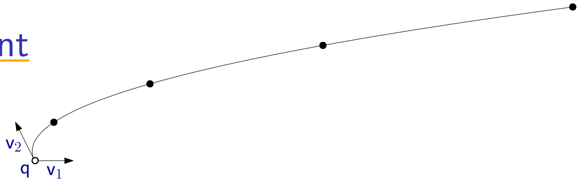
- say $\mathbf{x}_t = t^2 \mathbf{v}_1 + t\mathbf{v}_2 + \mathbf{q}$ for some $\mathbf{v}_1, \mathbf{v}_2, \mathbf{q} \in \mathbb{R}^n$
- intuitively:
  - converges to infinity most strongly in direction of $\mathbf{v}_1$
  - secondary convergence to infinity in direction of $\mathbf{v}_2$
  - finite shift or offset by $\mathbf{q}$

Example astral point



- say $\boldsymbol{x}_t = t^2 \boldsymbol{v}_1 + t\boldsymbol{v}_2 + \boldsymbol{q}$ for some $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{q} \in \mathbb{R}^n$
- intuitively:
  - converges to infinity most strongly in direction of $\boldsymbol{v}_1$
  - secondary convergence to infinity in direction of $\boldsymbol{v}_2$
  - finite shift or offset by $\boldsymbol{q}$
- converges in all directions, so has astral limit $\overline{\boldsymbol{x}}$

Example astral point



- say $\mathbf{x}_t = t^2\mathbf{v}_1 + t\mathbf{v}_2 + \mathbf{q}$ for some $\mathbf{v}_1, \mathbf{v}_2, \mathbf{q} \in \mathbb{R}^n$
- intuitively:
  - converges to infinity most strongly in direction of $\mathbf{v}_1$
  - secondary convergence to infinity in direction of $\mathbf{v}_2$
  - finite shift or offset by $\mathbf{q}$
- converges in all directions, so has astral limit $\overline{\mathbf{x}}$
- turns out, can write in form:

$$\overline{\mathbf{x}} = \underbrace{\omega\mathbf{v}_1 \mathbin{\dot{+}} \omega\mathbf{v}_2}_{\text{astrons}} \mathbin{\dot{+}} \mathbf{q}$$

- operation $\dot{+}$ is leftward addition:
  - similar to vector addition but not commutative
  - gives kind of "dominance" to term on left

Representing astral points

- in general: every astral point $\overline{\boldsymbol{x}}$ can be written in form

$$\overline{\boldsymbol{x}} = \underbrace{\omega \boldsymbol{v}_1 \mathbin{\dot{+}} \cdots \mathbin{\dot{+}} \omega \boldsymbol{v}_k}_{\text{astrons}} \mathbin{\dot{+}} \underbrace{\boldsymbol{q}}_{\substack{\text{finite}\\ \text{part}}}$$

  for some orthonormal $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$
  and some $\boldsymbol{q} \in \mathbb{R}^n$ orthogonal to the $\boldsymbol{v}_i$'s

Representing astral points

- in general: every astral point $\overline{\mathbf{x}}$ can be written in form

$$\overline{\mathbf{x}} = \underbrace{\omega \mathbf{v}_1 \mathbin{+\!\!\!\!\cdot} \cdots \mathbin{+\!\!\!\!\cdot} \omega \mathbf{v}_k}_{\text{astrons}} \mathbin{+\!\!\!\!\cdot} \underbrace{\mathbf{q}}_{\substack{\text{finite}\\\text{part}}}$$

  for some orthonormal $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$
  and some $\mathbf{q} \in \mathbb{R}^n$ orthogonal to the $\mathbf{v}_i$'s
- astral rank $= k$ (number of astrons in $\overline{\mathbf{x}}$'s representation)
  - astral rank $= 0 \Rightarrow \overline{\mathbf{x}} \in \mathbb{R}^n$
  - astral rank $= 1 \Rightarrow \overline{\mathbf{x}}$ is limit of sequence along ray

Outline

- what can minimizers at infinity look like?
- constructing astral space
- what are astral points like?
- extending functions to astral space
- convergence of iterative algorithms

Extending a function to astral space

- given convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- want to extend to astral space:

$$\bar{f} : \overline{\mathbb{R}^n} \to \overline{\mathbb{R}}$$

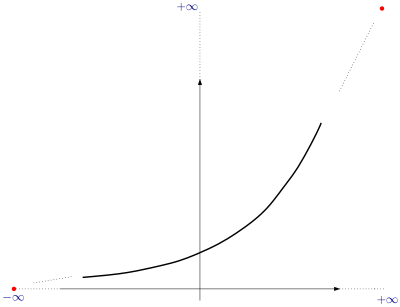
Extending a function to astral space

- given convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- want to extend to astral space:

$$\bar{f} : \overline{\mathbb{R}^n} \to \overline{\mathbb{R}}$$

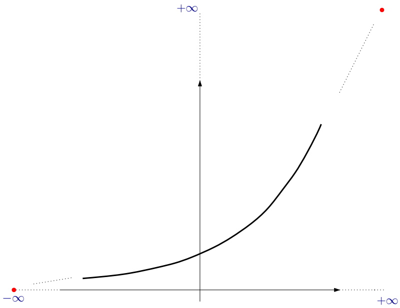
- how to define?

Example: exponential function



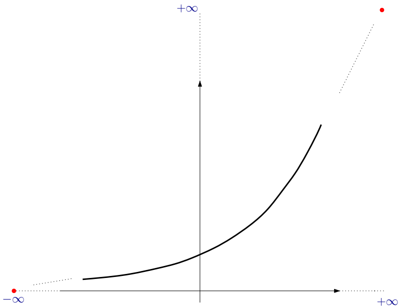
- say $f(x) = e^x$ for $x \in \mathbb{R}$

Example: exponential function



- say $f(x) = e^x$ for $x \in \mathbb{R}$
- want to define $\bar{f}$ = extension of f to $\overline{\mathbb{R}} = \overline{\mathbb{R}^1}$:
  - $\bar{f}(-\infty) = 0$ because: if $x_t \to -\infty$ then $f(x_t) \to 0$
  - $\bar{f}(+\infty) = +\infty$ because: if $x_t \to +\infty$ then $f(x_t) \to +\infty$

Example: exponential function



- say $f(x) = e^x$ for $x \in \mathbb{R}$
- want to define $\bar{f}$ = extension of f to $\overline{\mathbb{R}} = \overline{\mathbb{R}^1}$:
  - $\bar{f}(-\infty) = 0$ because: if $x_t \to -\infty$ then $f(x_t) \to 0$
  - $\bar{f}(+\infty) = +\infty$ because: if $x_t \to +\infty$ then $f(x_t) \to +\infty$
- only way to extend to $\overline{\mathbb{R}}$ continuously

Continuous extensions

- in general, given convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- want to define its extension $\bar{f} : \overline{\mathbb{R}^n} \to \overline{\mathbb{R}}$
- given $\overline{\mathbf{x}}$, how to define $\bar{f}(\overline{\mathbf{x}})$?

Continuous extensions

- in general, given convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- want to define its extension $\bar{f} : \overline{\mathbb{R}^n} \to \overline{\mathbb{R}}$
- given $\overline{\mathbf{x}}$, how to define $\bar{f}(\overline{\mathbf{x}})$?
- say $\lim f(\mathbf{x}_t)$ exists and is the same for every sequence $\mathbf{x}_t \to \overline{\mathbf{x}}$

Continuous extensions

- in general, given convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- want to define its extension $\bar{f} : \overline{\mathbb{R}^n} \to \overline{\mathbb{R}}$
- given $\overline{\mathbf{x}}$, how to define $\bar{f}(\overline{\mathbf{x}})$?
- say $\lim f(\mathbf{x}_t)$ exists and is the same for every sequence $\mathbf{x}_t \to \overline{\mathbf{x}}$
  - then only reasonable to set: $\bar{f}(\overline{\mathbf{x}}) = \lim f(\mathbf{x}_t)$ for any (and every) sequence $\mathbf{x}_t \to \overline{\mathbf{x}}$

Continuous extensions

- in general, given convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$
- want to define its extension $\bar{f} : \overline{\mathbb{R}^n} \to \overline{\mathbb{R}}$
- given $\overline{\mathbf{x}}$, how to define $\bar{f}(\overline{\mathbf{x}})$?
- say $\lim f(\mathbf{x}_t)$ exists and is the same
  for every sequence $\mathbf{x}_t \to \overline{\mathbf{x}}$
  - then only reasonable to set:
    $\bar{f}(\overline{\mathbf{x}}) = \lim f(\mathbf{x}_t)$ for any (and every) sequence $\mathbf{x}_t \to \overline{\mathbf{x}}$
- if holds for all $\overline{\mathbf{x}} \in \overline{\mathbb{R}^n}$ then
  $\bar{f}$ is (unique) continuous extension of f to $\overline{\mathbb{R}^n}$

Example: Linear functions

- say $f(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{u}$ for some $\boldsymbol{u} \in \mathbb{R}^n$ [e.g., $f(x_1, x_2) = 2x_1 - x_2$]

Example: Linear functions

- say $f(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{u}$ for some $\boldsymbol{u} \in \mathbb{R}^n$ [e.g., $f(x_1, x_2) = 2x_1 - x_2$]
- if $\boldsymbol{x}_t \to \overline{\boldsymbol{x}}$ then

$$\lim f(\boldsymbol{x}_t) = \lim(\boldsymbol{x}_t \cdot \boldsymbol{u})$$

exists and same for every sequence converging to $\overline{\boldsymbol{x}}$

[by construction!]

Example: Linear functions

- say $f(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{u}$ for some $\boldsymbol{u} \in \mathbb{R}^n$ [e.g., $f(x_1, x_2) = 2x_1 - x_2$]
- if $\boldsymbol{x}_t \to \overline{\boldsymbol{x}}$ then

$$\lim f(\boldsymbol{x}_t) = \lim(\boldsymbol{x}_t \cdot \boldsymbol{u})$$

  exists and same for every sequence converging to $\overline{\boldsymbol{x}}$

  [by construction!]

- therefore, f has continuous extension $\bar{f}$

Example: Linear functions

- say $f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{u}$ for some $\mathbf{u} \in \mathbb{R}^n$ [e.g., $f(x_1, x_2) = 2x_1 - x_2$]
- if $\mathbf{x}_t \to \overline{\mathbf{x}}$ then

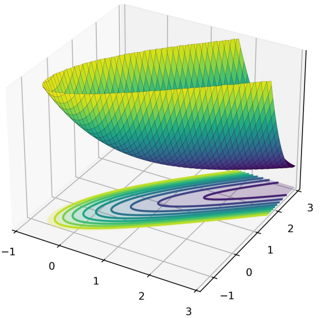
$$\lim f(\mathbf{x}_t) = \lim(\mathbf{x}_t \cdot \mathbf{u})$$

  exists and same for every sequence converging to $\overline{\mathbf{x}}$

  [by construction!]

- therefore, f has continuous extension $\bar{f}$
- so: every linear function can be continuously extended to $\overline{\mathbb{R}^n}$

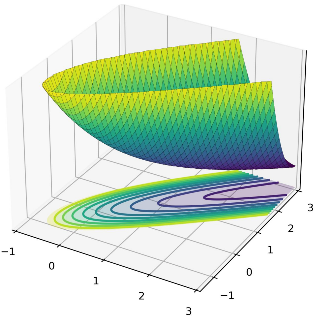
Example: Diagonal valley



- in $\mathbb{R}^2$, recall

$$f(x_1, x_2) = e^{-x_1} + (x_2 - x_1)^2$$

Example: Diagonal valley

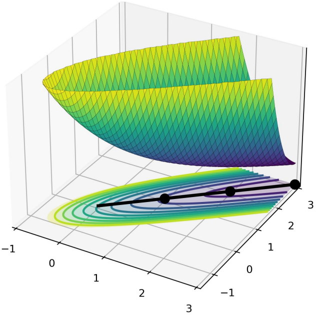


- in $\mathbb{R}^2$, recall

$$f(x_1, x_2) = e^{-x_1} + (x_2 - x_1)^2$$

- can show extends continuously to $\overline{\mathbb{R}^2}$

Example: Diagonal valley

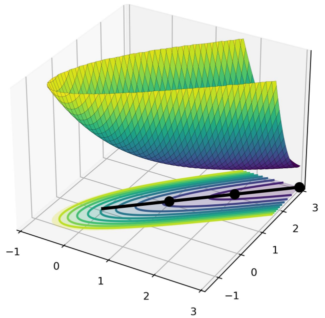


- in $\mathbb{R}^2$, recall

$$f(x_1, x_2) = e^{-x_1} + (x_2 - x_1)^2$$

- can show extends continuously to $\overline{\mathbb{R}^2}$
- saw f minimized by $\mathbf{x}_t = t\mathbf{v}$ where $\mathbf{v} = (1, 1)$

Example: Diagonal valley

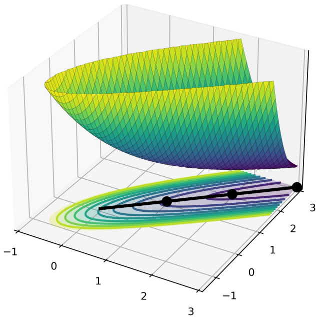


- in $\mathbb{R}^2$, recall

$$f(x_1, x_2) = e^{-x_1} + (x_2 - x_1)^2$$

- can show extends continuously to $\overline{\mathbb{R}^2}$
- saw f minimized by $\boldsymbol{x}_t = t\boldsymbol{v}$ where $\boldsymbol{v} = (1, 1)$
- converges to astron $\omega\boldsymbol{v}$
- $\bar{f}$ minimized at $\overline{\boldsymbol{x}} = \omega\boldsymbol{v}$

Example: Diagonal valley



- in $\mathbb{R}^2$, recall

$$f(x_1, x_2) = e^{-x_1} + (x_2 - x_1)^2$$

- can show extends continuously to $\overline{\mathbb{R}^2}$
- saw f minimized by $\mathbf{x}_t = t\mathbf{v}$ where $\mathbf{v} = (1, 1)$
- converges to astron $\omega\mathbf{v}$
- $\bar{f}$ minimized at $\overline{\mathbf{x}} = \omega\mathbf{v}$
- since continuous, f also minimized by any sequence $\mathbf{x}'_t \to \omega\mathbf{v}$

Extending functions in general

- f might not have continuous extension

Extending functions in general

- f might not have continuous extension
- continuous case:
  $\bar{f}(\overline{\mathbf{x}}) = \lim f(\mathbf{x}_t)$ over all sequences $\mathbf{x}_t \to \overline{\mathbf{x}}$

Extending functions in general

- f might not have continuous extension
- general case:
  $\bar{f}(\overline{\mathbf{x}}) =$ minimum of $\lim f(\mathbf{x}_t)$ over all sequences $\mathbf{x}_t \to \overline{\mathbf{x}}$

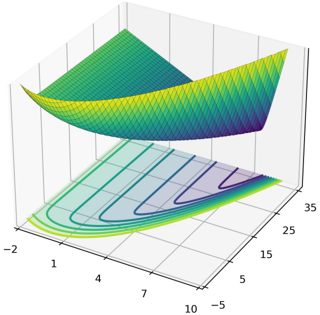
Extending functions in general

- f might not have continuous extension
- general case:
  $\bar{f}(\overline{\mathbf{x}}) =$ minimum of $\lim f(\mathbf{x}_t)$ over all sequences $\mathbf{x}_t \to \overline{\mathbf{x}}$
- properties:
  - $\bar{f}$ must attain minimizer in $\overline{\mathbb{R}^n}$ [since compact]

Extending functions in general

- f might not have continuous extension
- general case:
  $\bar{f}(\overline{\mathbf{x}}) =$ minimum of $\lim f(\mathbf{x}_t)$ over all sequences $\mathbf{x}_t \to \overline{\mathbf{x}}$
- properties:
  - $\bar{f}$ must attain minimizer in $\overline{\mathbb{R}^n}$ [since compact]
  - $\overline{\mathbf{x}}$ minimizes $\bar{f}$ iff
    there exists sequence $\mathbf{x}_t \to \overline{\mathbf{x}}$ minimizing f

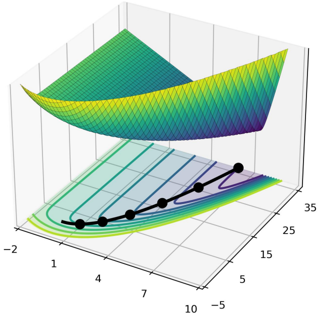
Example: Two-speed exponential



- in $\mathbb{R}^2$, recall:

  $$f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

Example: Two-speed exponential



- in $\mathbb{R}^2$, recall:

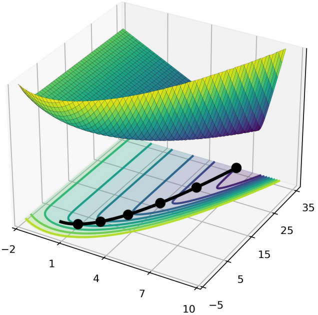
  $$f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

- minimized by sequence:

  $$\boldsymbol{x}_t = (t, t^2) = t^2 \boldsymbol{e}_2 + t\boldsymbol{e}_1$$

  where $\boldsymbol{e}_1 = (1, 0)$ and $\boldsymbol{e}_2 = (0, 1)$

Example: Two-speed exponential



- in $\mathbb{R}^2$, recall:

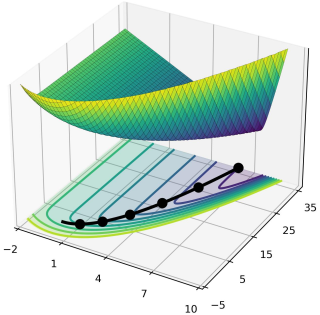
  $$f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

- minimized by sequence:

  $$\boldsymbol{x}_t = (t, t^2) = t^2\boldsymbol{e}_2 + t\boldsymbol{e}_1 \;\rightarrow\; \overline{\boldsymbol{x}} = \omega\boldsymbol{e}_2 \dotplus \omega\boldsymbol{e}_1$$

  where $\boldsymbol{e}_1 = (1, 0)$ and $\boldsymbol{e}_2 = (0, 1)$

Example: Two-speed exponential



- in $\mathbb{R}^2$, recall:

  $$f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

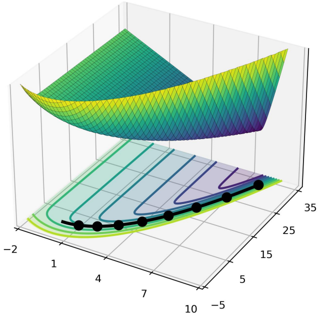
- minimized by sequence:

  $$\mathbf{x}_t = (t, t^2) = t^2 \mathbf{e}_2 + t\mathbf{e}_1 \;\rightarrow\; \overline{\mathbf{x}} = \omega \mathbf{e}_2 \mathbin{\cdot\!\!+} \omega \mathbf{e}_1$$

  where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$

- $\bar{f}$ minimized by $\overline{\mathbf{x}}$
  - $\overline{\mathbf{x}}$ has astral rank 2
  - can show no other minimizers

Example: Two-speed exponential



- in $\mathbb{R}^2$, recall:

  $$f(x_1, x_2) = e^{-x_1} + e^{-x_2 + x_1^2/2}$$

- minimized by sequence:

  $$\mathbf{x}_t = (t, t^2) = t^2 \mathbf{e}_2 + t\mathbf{e}_1 \;\rightarrow\; \overline{\mathbf{x}} = \omega \mathbf{e}_2 \mathbin{+\!\!\!\!\cdot}\; \omega \mathbf{e}_1$$

  where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$

- $\bar{f}$ minimized by $\overline{\mathbf{x}}$
  - $\overline{\mathbf{x}}$ has astral rank 2
  - can show no other minimizers

- $\bar{f}$ not continuous at $\overline{\mathbf{x}}$: e.g.:

  $$\mathbf{x}'_t = (t, \tfrac{1}{2}t^2) = \tfrac{1}{2}t^2 \mathbf{e}_2 + t\mathbf{e}_1 \;\rightarrow\; \overline{\mathbf{x}}$$

  but $f(\mathbf{x}'_t) \rightarrow 1 \neq \bar{f}(\overline{\mathbf{x}})$

Outline

- what can minimizers at infinity look like?
- constructing astral space
- what are astral points like?
- extending functions to astral space
- convergence of iterative algorithms

Convergence of gradient-based methods

- in standard convex analysis, if $\nabla f(\boldsymbol{x}) = 0$ then $\boldsymbol{x}$ minimizes f

Convergence of gradient-based methods

- in standard convex analysis, if $\nabla f(\boldsymbol{x}) = 0$ then $\boldsymbol{x}$ minimizes f
- say $(\boldsymbol{x}_t)$ sequence with $\nabla f(\boldsymbol{x}_t) \to 0$
  - must $f(\boldsymbol{x}_t) \to \inf f$?

Convergence of gradient-based methods

- in standard convex analysis, if $\nabla f(\boldsymbol{x}) = 0$ then $\boldsymbol{x}$ minimizes f
- say $(\boldsymbol{x}_t)$ sequence with $\nabla f(\boldsymbol{x}_t) \to 0$
  - must $f(\boldsymbol{x}_t) \to \inf f$? no!

Convergence of gradient-based methods

- in standard convex analysis, if $\nabla f(\boldsymbol{x}) = 0$ then $\boldsymbol{x}$ minimizes f
- say $(\boldsymbol{x}_t)$ sequence with $\nabla f(\boldsymbol{x}_t) \to 0$
  - must $f(\boldsymbol{x}_t) \to \inf f$? no!
- e.g., in $\mathbb{R}^2$, let

$$f(x_1, x_2) = \begin{cases} x_1^2/x_2 & \text{if } x_2 > |x_1| \\ 2|x_1| - x_2 & \text{else} \end{cases}$$

  - f is convex, continuous, finite, nonnegative, continuously differentiable everywhere of interest
  - let $\boldsymbol{x}_t = (t^2, t^3)$

Convergence of gradient-based methods

- in standard convex analysis, if $\nabla f(\boldsymbol{x}) = 0$ then $\boldsymbol{x}$ minimizes f
- say $(\boldsymbol{x}_t)$ sequence with $\nabla f(\boldsymbol{x}_t) \to 0$
  - must $f(\boldsymbol{x}_t) \to \inf f$? no!
- e.g., in $\mathbb{R}^2$, let

$$f(x_1, x_2) = \begin{cases} x_1^2/x_2 & \text{if } x_2 > |x_1| \\ 2|x_1| - x_2 & \text{else} \end{cases}$$

  - f is convex, continuous, finite, nonnegative, continuously differentiable everywhere of interest
  - let $\boldsymbol{x}_t = (t^2, t^3)$
  - then $\nabla f(\boldsymbol{x}_t) = (\frac{2}{t}, -\frac{1}{t^2}) \to 0$
  - however, $f(\boldsymbol{x}_t) = t \to +\infty$

Convergence and astral continuity

- in this case:
  - $\boldsymbol{x}_t = t^3\boldsymbol{e}_2 + t^2\boldsymbol{e}_1 \;\rightarrow\; \overline{\boldsymbol{x}} = \omega\boldsymbol{e}_2 \mathbin{\cdot\!\!+} \omega\boldsymbol{e}_1$
  - $\bar{f}$ not continuous at $\overline{\boldsymbol{x}}$

Convergence and astral continuity

- in this case:
  - $\mathbf{x}_t = t^3\mathbf{e}_2 + t^2\mathbf{e}_1 \;\rightarrow\; \overline{\mathbf{x}} = \omega\mathbf{e}_2 \mathbin{+\!\!\!\!\cdot}\, \omega\mathbf{e}_1$
  - $\bar{f}$ not continuous at $\overline{\mathbf{x}}$
- not coincidence!
- when assume continuity, must get convergence to minimum

Convergence and astral continuity

- in this case:
  - $\mathbf{x}_t = t^3\mathbf{e}_2 + t^2\mathbf{e}_1 \;\rightarrow\; \overline{\mathbf{x}} = \omega\mathbf{e}_2 \mathbin{\dot{+}} \omega\mathbf{e}_1$
  - $\bar{f}$ not continuous at $\overline{\mathbf{x}}$
- not coincidence!
- when assume continuity, must get convergence to minimum
- general theorem: if:
  - $\nabla f(\mathbf{x}_t) \rightarrow 0$
  - $\mathbf{x}_t \rightarrow \overline{\mathbf{x}}$ and $\bar{f}$ continuous at $\overline{\mathbf{x}}$ (with $\bar{f}(\overline{\mathbf{x}}) < +\infty$)

  then: $f(\mathbf{x}_t) \rightarrow \inf f$

Convergence and astral continuity

- in this case:
  - $\mathbf{x}_t = t^3\mathbf{e}_2 + t^2\mathbf{e}_1 \;\rightarrow\; \overline{\mathbf{x}} = \omega\mathbf{e}_2 \mathbin{\dot{+}} \omega\mathbf{e}_1$
  - $\bar{f}$ not continuous at $\overline{\mathbf{x}}$
- not coincidence!
- when assume continuity, must get convergence to minimum
- general theorem: if:
  - $\nabla f(\mathbf{x}_t) \rightarrow 0$
  - $\mathbf{x}_t \rightarrow \overline{\mathbf{x}}$ and $\bar{f}$ continuous at $\overline{\mathbf{x}}$ (with $\bar{f}(\overline{\mathbf{x}}) < +\infty$)

  then: $f(\mathbf{x}_t) \rightarrow \inf f$
- reveals structure and regularity not otherwise apparent

Convergence and astral continuity (cont.)

- can use to prove convergence of standard iterative methods applied to various ML/statistical settings
  - e.g.: gradient descent, coordinate descent, steepest descent
  - e.g.: logistic regression, boosting, maximum likelihood (which all have continuous extensions)

Convergence and astral continuity (cont.)

- can use to prove convergence of standard iterative methods applied to various ML/statistical settings
  - e.g.: gradient descent, coordinate descent, steepest descent
  - e.g.: logistic regression, boosting, maximum likelihood (which all have continuous extensions)
- don't require finite minimizer
- algorithms operate in $\mathbb{R}^n$, but use astral methods in proofs
  - rely on astral continuity properties
    (without which results do not hold, in general)

How to connect to generalization??

How to connect to generalization??

- can only speculate!

How to connect to generalization??

- can only speculate!
- in ML, usually minimize empirical risk function based on random examples
- sometimes, maybe, some (astral?) property of minimizer can imply generalization

How to connect to generalization??

- can only speculate!
- in ML, usually minimize empirical risk function based on random examples
- sometimes, maybe, some (astral?) property of minimizer can imply generalization
- e.g. AdaBoost minimizes exponential loss
  - finds solution with large-margin property, implying generalization
  - really an astral property of minimizer at infinity (namely, of first astron in representation)

Summary

- tried to give a taste of astral space:
    - its construction
    - structure of astral points
    - how to extend convex functions

Summary

- tried to give a taste of astral space:
  - its construction
  - structure of astral points
  - how to extend convex functions
- aim: expand foundations of convex analysis to encompass points at infinity
  - e.g. to enable easier, more general proofs of convergence

Summary

- tried to give a taste of astral space:
  - its construction
  - structure of astral points
  - how to extend convex functions
- aim: expand foundations of convex analysis to encompass points at infinity
  - e.g. to enable easier, more general proofs of convergence
- far more not covered
  - details at: aka.ms/astral [or arxiv.org/abs/2205.03260] (will eventually be published as a book)