Math 381 - Fall 2022

Jay Newby

University of Alberta

Week 12

Last Week

1. Low rank matrix approximations
2. Eigenvalue problem

This Week

New topic: Optimization

1. Examples of optimization problems
2. Continuous optimization problems in 1D
3. Continuous optimization problems in higher dimensions
4. Constrained optimization

Warmup Example: motivating some basic differential geometry concepts

x = position
y = velocity ($y = x'$)

$k=1, m=1$

Let $\cancel{y = m\dot{x}}$

$x''(t) = -x(t)$

$y' = -x, \; x' = y$

$$E(x, y) = \frac{1}{2}kx^2 + \frac{1}{2m}y^2 = \frac{1}{2}x^2 + \frac{1}{2}y^2$$

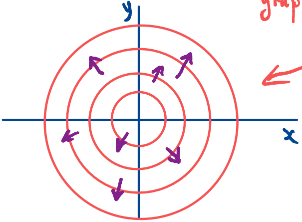
Energy

Solutions

$x(t) = A\cos(t) + B\sin(t)$

$$\nabla E = \begin{bmatrix} x \\ y \end{bmatrix}$$

graph trajectories $(x(t), y(t))$



Level curves of $E(x,y)$

$$\frac{d}{dt} E(x(t), y(t)) = \nabla E^T \begin{pmatrix} x' \\ y' \end{pmatrix}$$

$$= xx' + yy' = xy + y(-x) = 0$$

Friction:

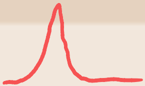
$y' = -x - \gamma y$

Example: dynamical system (simple mass-spring)

Example: maximum likelihood

Normal distribution

$$p(x|\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$



Given iid data x_j we want to find the value of σ that maximizes the likelihood function, define as

$$L(\sigma) = \prod_{j=1}^{n} p(x_j \mid \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\sum_{j=1}^{n} \frac{x_j^2}{2\sigma^2}\right].$$

$$\mathcal{L}(\sigma) = \log(L(\sigma)) = -\frac{1}{\sigma^2}\left[\sum_{j=1}^{n} \frac{x_j^2}{2}\right] - \frac{n}{2}\log(2\pi\sigma^2)$$

Example: maximum likelihood

$$\mathcal{L}(\sigma) = -\frac{1}{\sigma^2}\left[\sum_{j=1}^{n} \frac{x_j^2}{2}\right] - n\log(\sigma) + C$$

Let $q = \sum_{j=1}^{n} \frac{x_j^2}{2}$

$$\mathcal{L}'(\sigma) = \frac{2}{\sigma^3} q - \frac{n}{\sigma} = 0$$

$$\Rightarrow q - n\sigma^2 = 0$$

$$\Rightarrow \sigma^2 = \frac{1}{n} q = \frac{1}{2n}\sum_{j=1}^{n} x_j^2$$

Example: machine learning: neural networks

MNIST

Example: machine learning: neural networks

Loss function for a simple image classifier

Suppose that for a given training image, the ground truth classifications (e.g., dog, cat, boat, etc) is given by the probability distribution $p \in \mathbb{R}^m$ with $p \geq 0$ and $\sum_{j-1}^{m} p_j = 1$. Suppose we have a neural network with parameters $w \in \mathbb{R}^n$. For a given input image, the neural network generates an estimated distribution $q \in \mathbb{R}^m$, which depend on the parameters w. The cross entropy loss function is defined as

$$L(w) = -\sum_{j=1}^{m} p_j \log(q_j(w))$$

Continuous optimization problems

Continuous optimization problem

$$\min_x f(x) \quad \text{subject to} \quad g(x) = 0, \quad h(x) \leq 0.$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$.

Local and global minima

Global optimization is well posed in some special cases

- convex problems
- finite sets
- closed and bounded sets

Existence and uniqueness

Existence on closed and bounded sets

If f continuous on a closed and bounded set S then there exists a global minimum of f on S. If S is unbounded or not closed then there might not be local or global minimum (e.g., $f(x) = x$ on (a, b))

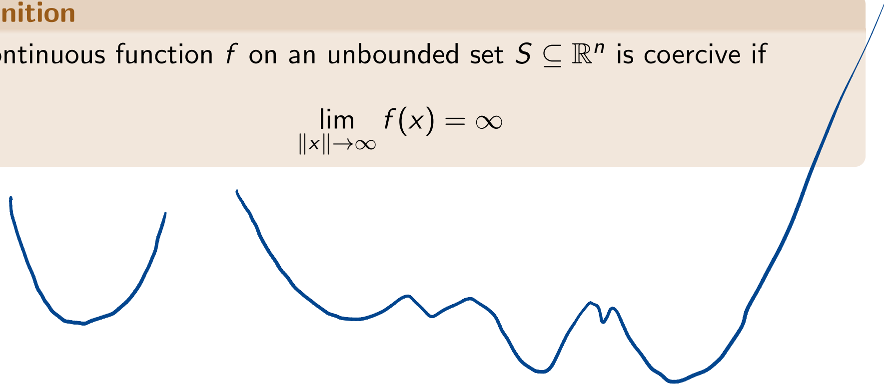
Existence for closed and unbounded sets

If f is coercive on a closed and unbounded set S then f has a global minimum on S.

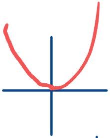
Definition

A continuous function f on an unbounded set $S \subseteq \mathbb{R}^n$ is coercive if

$$\lim_{\|x\| \to \infty} f(x) = \infty$$

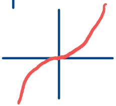


Example: $f(x) = x^2$



coercive

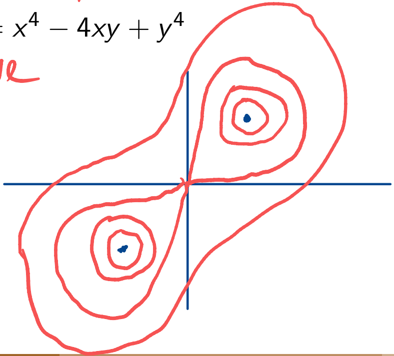
Example: $f(x) = x^3$



not coercive

Example: $f(x, y) = x^4 - 4xy + y^4$

coercive



Level sets

Existence from sublevel sets

If f is continuous on a set $S \subseteq \mathbb{R}^n$ and has nonempty sublevel set that is closed and bounded, then f has a global minimum on S.

Level sets

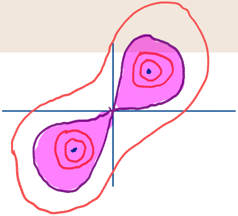
$f\colon S \subseteq \mathbb{R}^n \to \mathbb{R}$ Level set is the set of points in S for which f is equal to a constant

$$S_\gamma = \{x \in S : f(x) = \gamma\}$$

Definition: Sublevel set

Given a constant γ

$$L_\gamma = \{x \in S : f(x) \leq \gamma\}$$



We can say something about uniqueness for convex problems

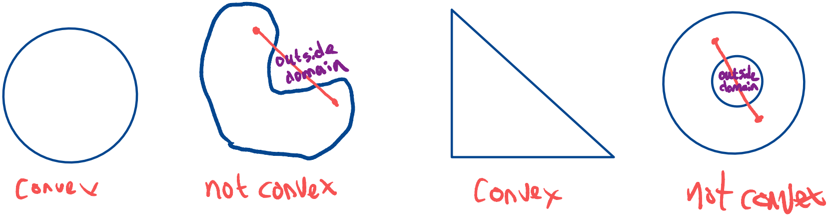
Definition: convex set

The set $S \subseteq \mathbb{R}^n$ is convex if

$$\{\alpha x + (1-\alpha)y : 0 \leq \alpha \leq 1\} \subseteq S$$

for all $x, y \in S$.

Examples:



We can say something about uniqueness for convex problems

Wednesday...

Definition: Convex function

A function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$, with S a convex set, is a convex function if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y),$$

for all $\alpha \in [0,1]$ and all $x, y \in S$.

Definition: Strictly convex function

A convex function is strictly convex if

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y),$$

for all $\alpha \in (0,1)$ and all $x, y \in S$.

Examples:

Uniqueness of the global minimum for strictly convex function on convex sets

- Sublevel sets of a convex function are convex
- Any local minimum of a convex function f on a convex set S is a global minimum on S
- Any local minimum of a strictly convex function f on a convex set S is a unique global minimum on S