Convex Optimization

A Gentle Introduction

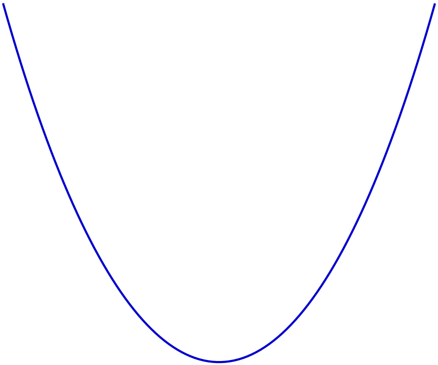
S.V. N. (vishy) Vishwanathan

Purdue University
vishy@purdue.edu

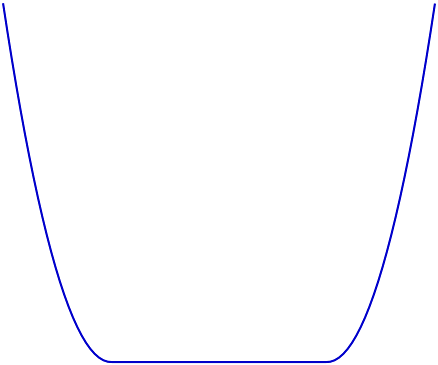
April 15, 2013

Outline

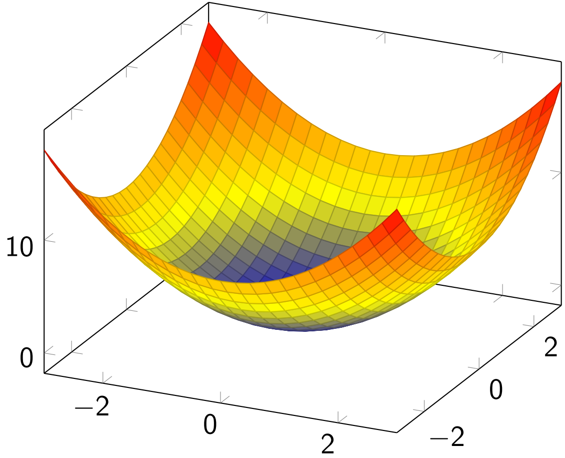
Convex Functions



Convex Functions



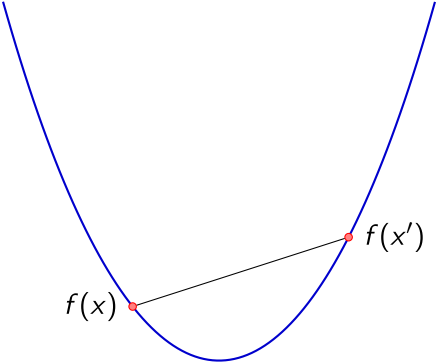
Convex Functions



Disclaimer

- My focus is on intuition
- Not mathematically rigorous

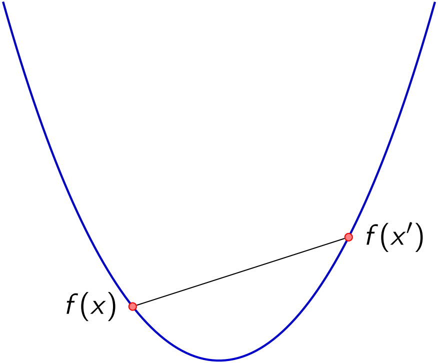
Convex Function



A function f is convex if, and only if, for all x, x' and $\lambda \in (0, 1)$

$$f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x')$$

Convex Function



A function f is strictly convex if, and only if, for all x, x' and $\lambda \in (0, 1)$

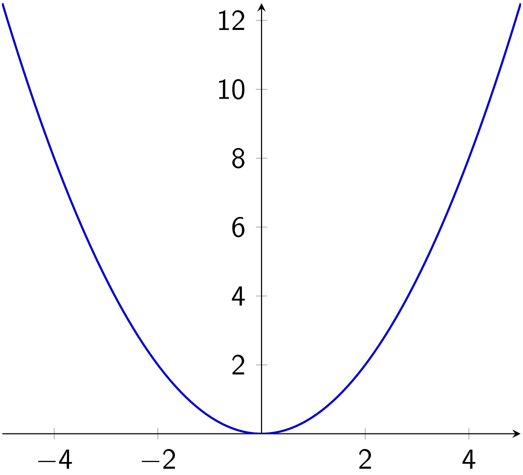
$$f(\lambda x + (1 - \lambda)x') < \lambda f(x) + (1 - \lambda)f(x')$$

Exercise: Jensen's Inequality

- Extend the definition of convexity to show that if f is convex, then for all $\lambda_i \geq 0$ such that $\sum_i \lambda_i = 1$ we have

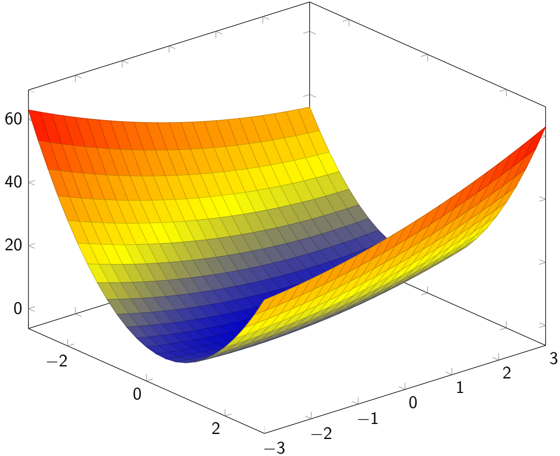
$$f\left(\sum_i \lambda_i x_i\right) \leq \sum_i \lambda_i f(x_i)$$

Some Familiar Examples



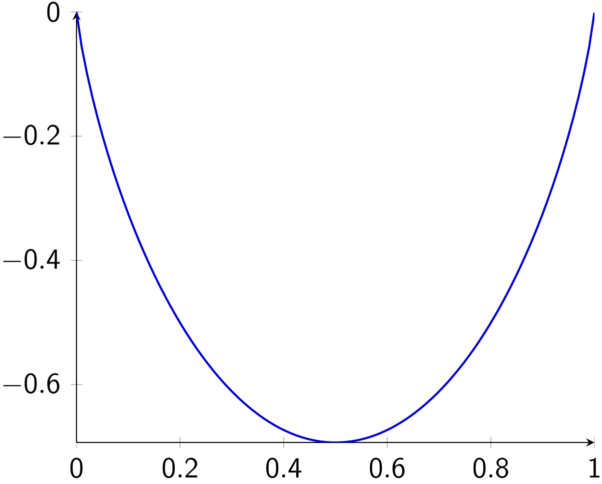
$$f(x) = \frac{1}{2}x^2 \text{ (Square norm)}$$

Some Familiar Examples



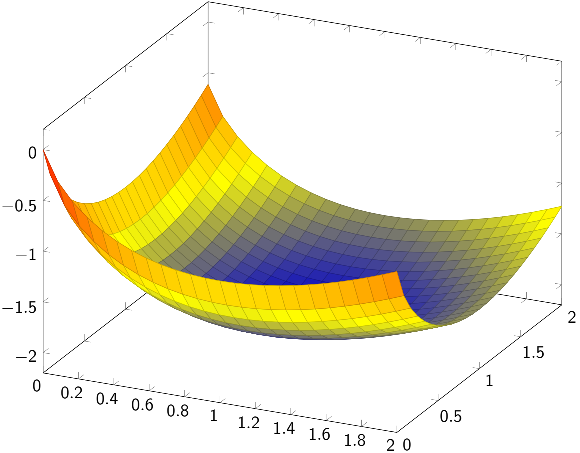
$$f(x, y) = \frac{1}{2} \begin{bmatrix} x, y \end{bmatrix} \begin{bmatrix} 10, 1 \\ 2, 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Some Familiar Examples



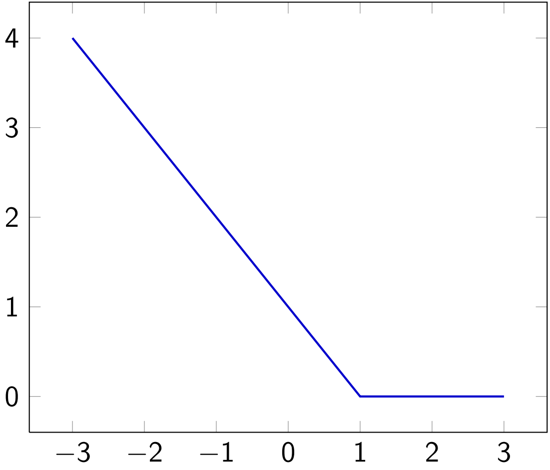
$$f(x) = x \log x + (1 - x) \log(1 - x) \text{ (Negative entropy)}$$

Some Familiar Examples



$$f(x, y) = x \log x + y \log y - x - y \text{ (Un-normalized negative entropy)}$$

Some Familiar Examples

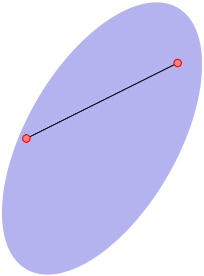


$$f(x) = \max(0, 1 - x) \text{ (Hinge Loss)}$$

Some Other Important Examples

- Linear functions: $f(x) = ax + b$
- Softmax: $f(x) = \log \sum_i \exp(x_i)$
- Norms: For example the 2-norm $f(x) = \sqrt{\sum_i x_i^2}$

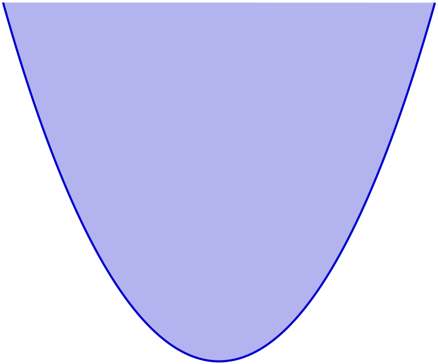
Convex Sets



A set C is convex if, and only if, for all $x, x' \in C$ and $\lambda \in (0,1)$ we have

$$\lambda x + (1-\lambda)x' \in C$$

Convex Sets and Convex Functions



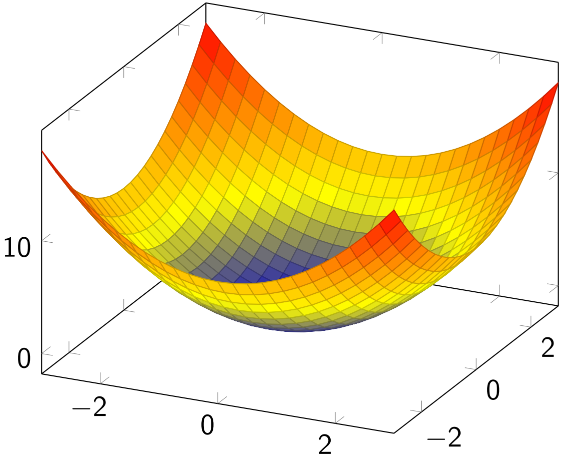
A function f is convex if, and only if, its epigraph is a convex set

Convex Sets and Convex Functions

- Indicator functions of convex sets are convex

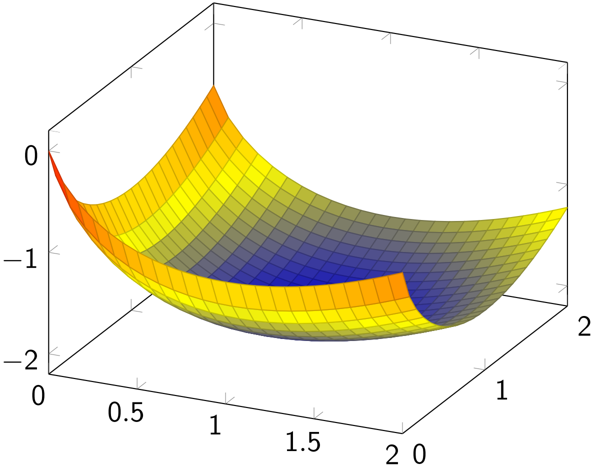
$$I_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise.} \end{cases}$$

Below sets of Convex Functions



$$f(x,y) = x^2 + y^2$$

Below sets of Convex Functions

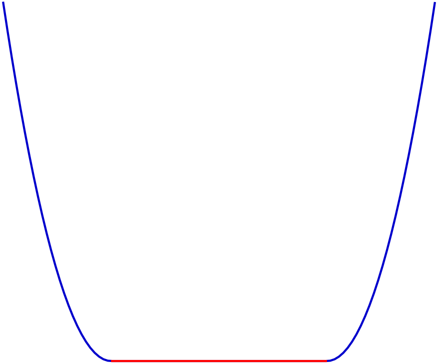


$$f(x, y) = x \log x + y \log y - x - y$$

Below sets of Convex Functions

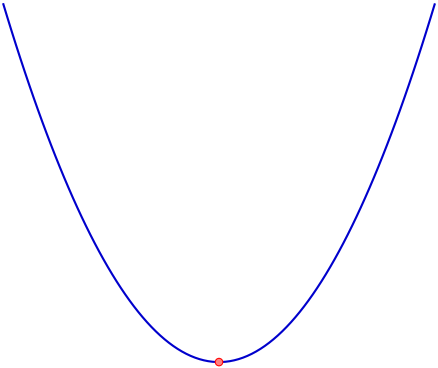
- If f is convex, then all its level sets are convex
- Is the converse true? (Exercise: construct a counter-example)

Minima on Convex Sets



- Set of minima of a convex function is a convex set
- Proof: Consider the set $\{x : f(x) \leq f^*\}$

Minima on Convex Sets



- Set of minima of a strictly convex function is a singleton
- Proof: try this at home!

Outline

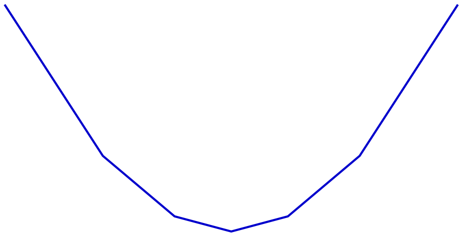
Set Operations

- Intersection of convex sets is convex
- Image of a convex set under a linear transformation is convex
- Inverse image of a convex set under a linear transformation is convex

Function Operations

- Linear Combination with non-negative weights: $f(x) = \sum_i w_i f_i(x)$ s.t. $w_i \geq 0$
- Pointwise maximum: $f(x) = \max_i f_i(x)$
- Composition with affine function: $f(x) = g(Ax + b)$
- Projection along a direction: $f(\eta) = g(x_0 + \eta d)$
- Restricting the domain on a convex set: $f(x)$s.t. $x \in \mathcal{C}$

One Quick Example

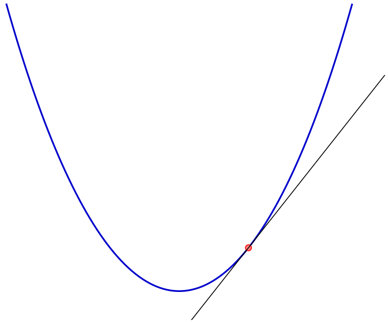


The piecewise linear function $f(x) := \max_i \langle u_i, x \rangle$ is convex

Outline

First Order Taylor Expansion

The First Order Taylor approximation globally lower bounds the function



For any x and x' we have

$$f(x) \geq f(x') + \left\langle x - x', \nabla f(x') \right\rangle$$

Identifying the Minimum

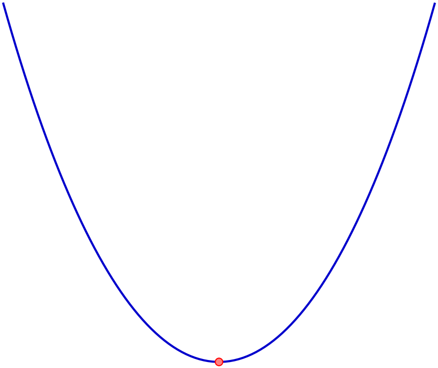
- Let $f : X \to \mathbb{R}$ be a differentiable convex function. Then x is a minimizer of f, if, and only if,

$$\left\langle x' - x, \nabla f(x) \right\rangle \geq 0 \text{ for all } x'.$$

- One way to ensure this is to set $\nabla f(x) = 0$
- Minimizing a smooth convex function is the same as finding an x such that $\nabla f(x) = 0$

Outline

Problem Statement



- Given a black-box which can compute $J : \mathbb{R} \to \mathbb{R}$ and $J' : \mathbb{R} \to \mathbb{R}$ find the minimum value of J

Increasing Gradients

- From the first order conditions

$$J(w) \geq J(w') + (w - w') \cdot J'(w')$$

and

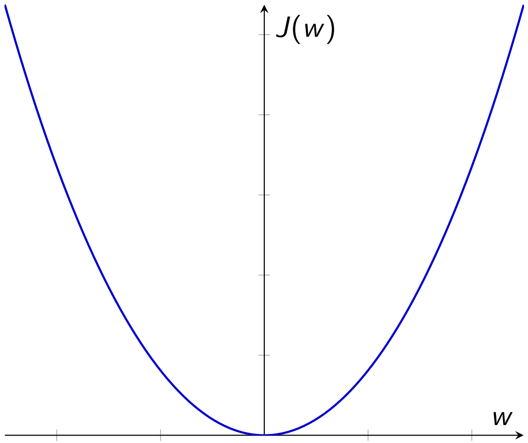
$$J(w') \geq J(w) + (w' - w) \cdot J'(w)$$

- Add the two

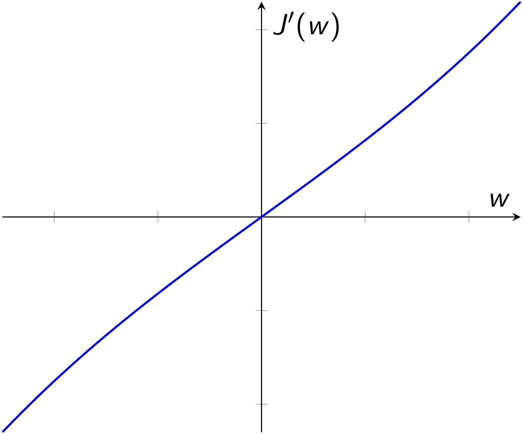
$$(w - w') \cdot (J'(w) - J'(w')) \geq 0$$

$w \geq w'$ implies that $J'(w) \geq J'(w')$

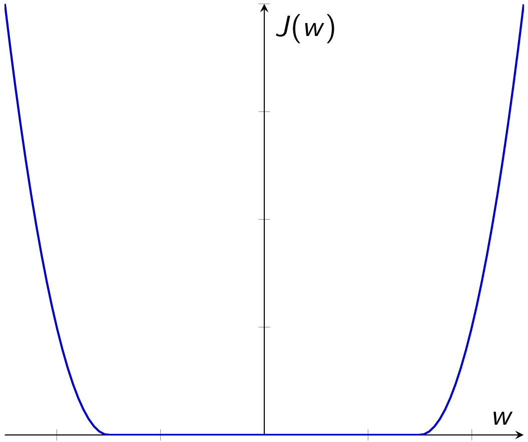
Increasing Gradients



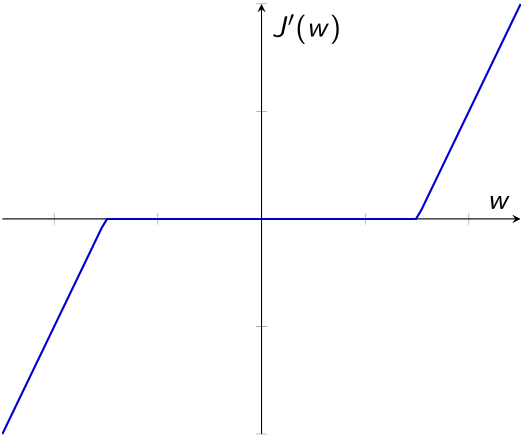
Increasing Gradients



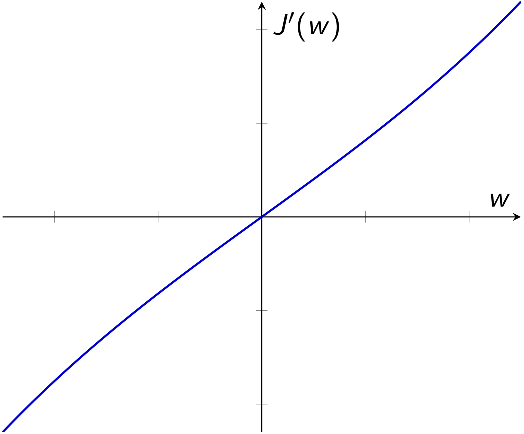
Increasing Gradients



Increasing Gradients

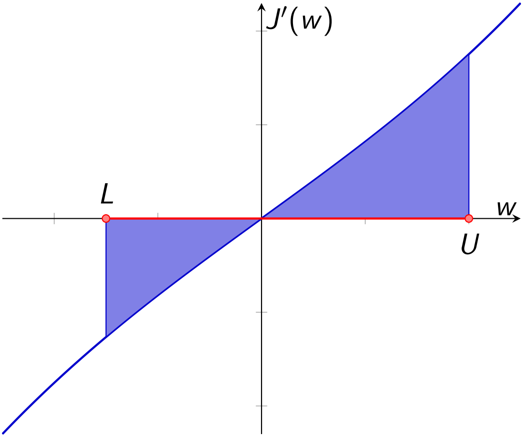


Problem Restatement

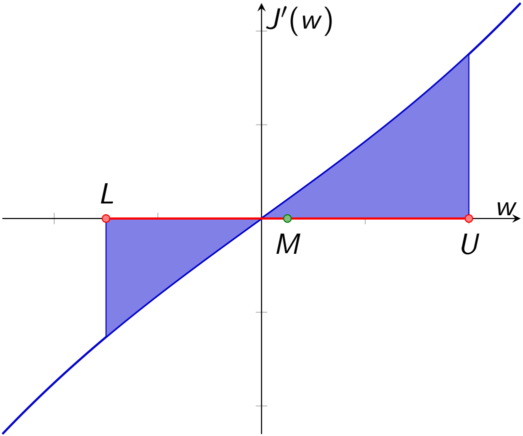


- Identify the point where the increasing function J' crosses zero

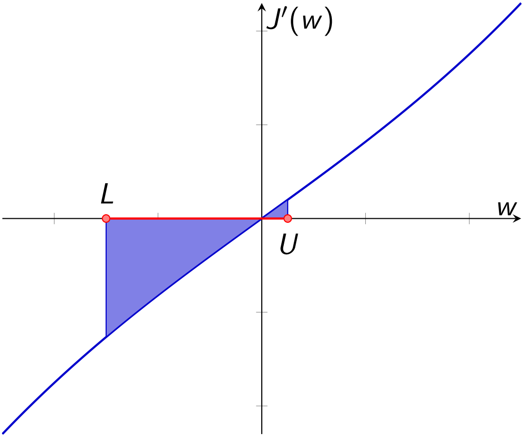
Bisection Algorithm



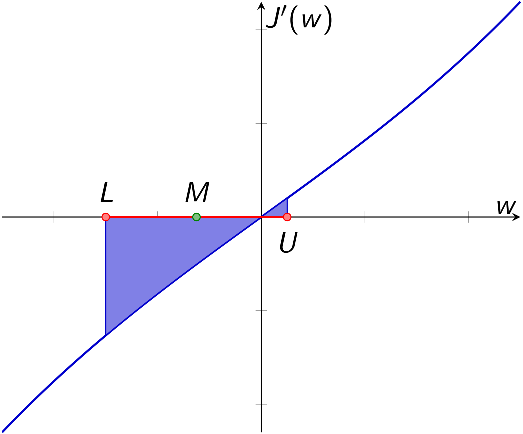
Bisection Algorithm



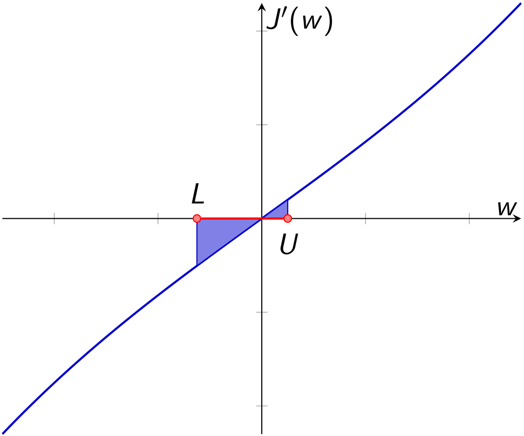
Bisection Algorithm



Bisection Algorithm



Bisection Algorithm



Interval Bisection

Require: L, U, ϵ

$\quad maxgrad \leftarrow J'(U)$

$\quad$ **while** $(U - L) \cdot maxgrad > \epsilon$ **do**

$\qquad M \leftarrow \frac{U+L}{2}$

$\qquad$ **if** $J'(M) > 0$ **then**

$\qquad\quad U \leftarrow M$

$\qquad$ **else**

$\qquad\quad L \leftarrow M$

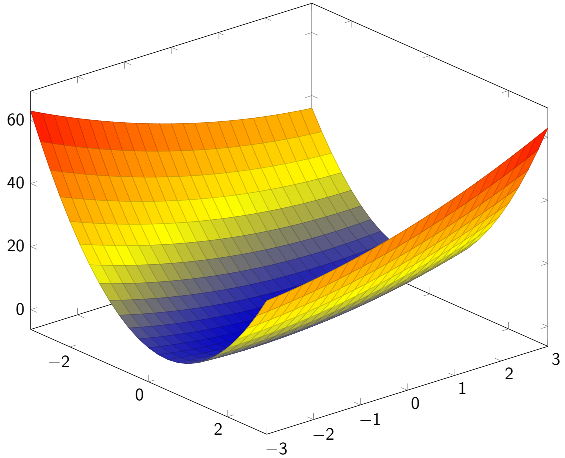
$\qquad$ **end if**

$\quad$ **end while**

$\quad$ **return** $\frac{U+L}{2}$

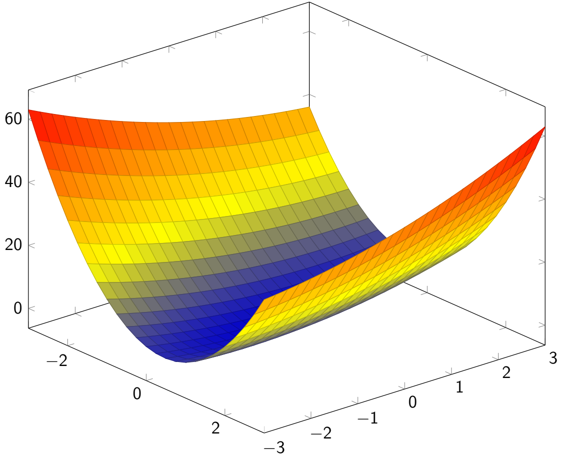
Outline

Problem Statement



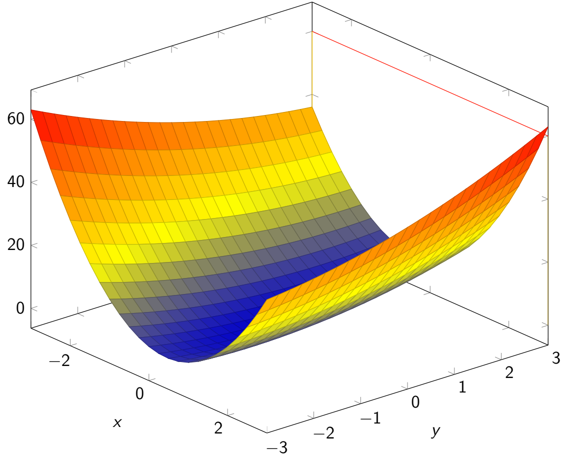
- Given a black-box which can compute $J : \mathbb{R}^n \to \mathbb{R}$ and $\nabla J : \mathbb{R}^n \to \mathbb{R}^n$ find the minimum value of J

Concrete Example



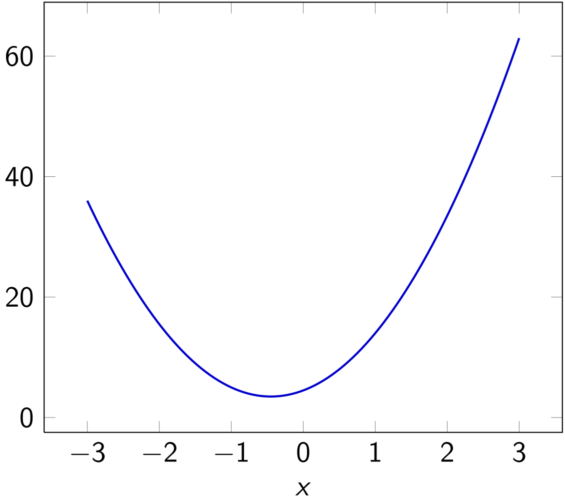
$$f(x, y) = \frac{1}{2} \left[\; x, y \; \right] \left[\begin{array}{c} 10, 1 \\ 2, 1 \end{array} \right] \left[\begin{array}{c} x \\ y \end{array} \right]$$

Concrete Example



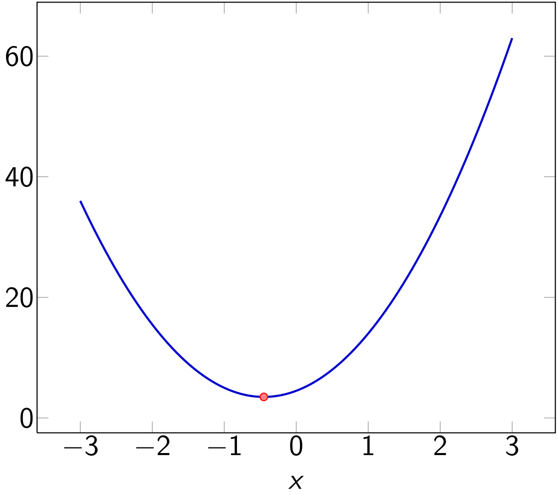
$$f(x,3) = \frac{1}{2}\left[\ x, 3\ \right]\left[\begin{array}{c} 10, 1 \\ 2, 1 \end{array}\right]\left[\begin{array}{c} x \\ 3 \end{array}\right]$$

Concrete Example



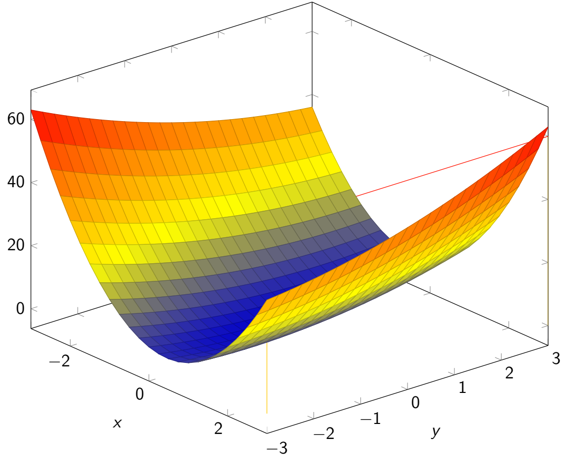
$$f(x, 3) = 5x^2 + \frac{9}{2}x + \frac{9}{2}$$

Concrete Example

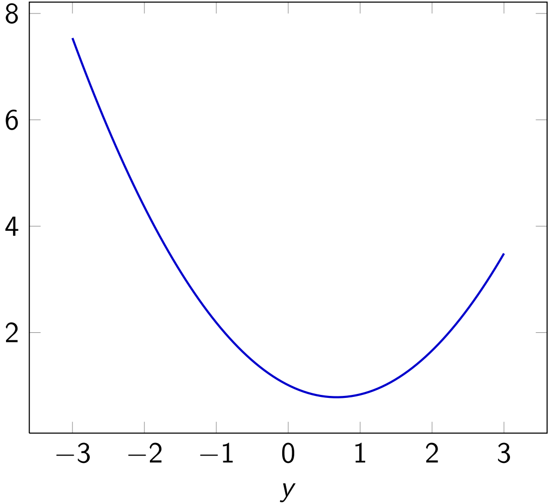


$$f(x, 3) = 5x^2 + \frac{9}{2}x + \frac{9}{2} \qquad \text{Minima: } x = -\frac{9}{20}$$

Concrete Example

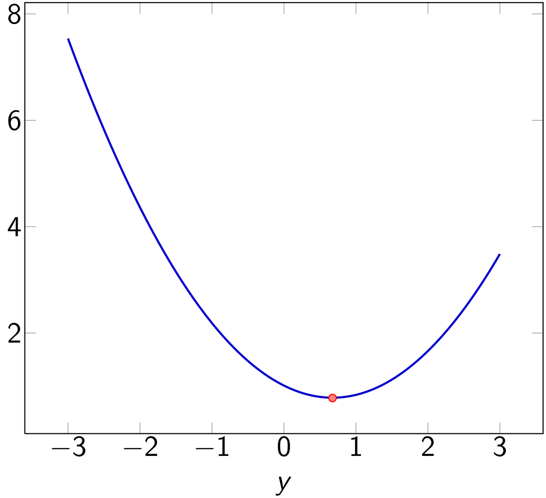


Concrete Example



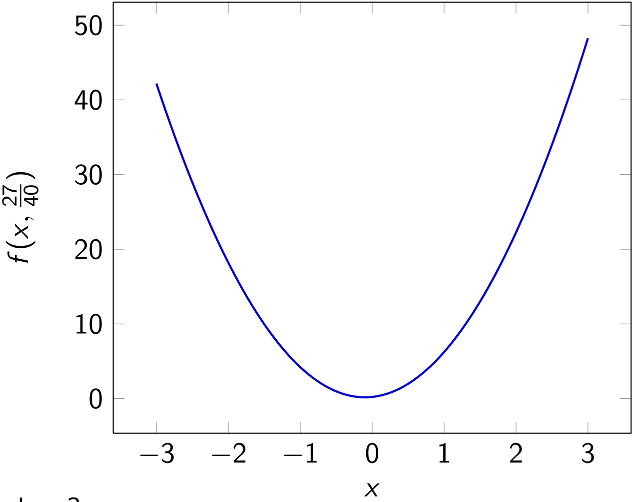
$$f(-\frac{9}{20}, y) = \frac{1}{2}y^2 - \frac{27}{40}y + \frac{81}{80}$$

Concrete Example



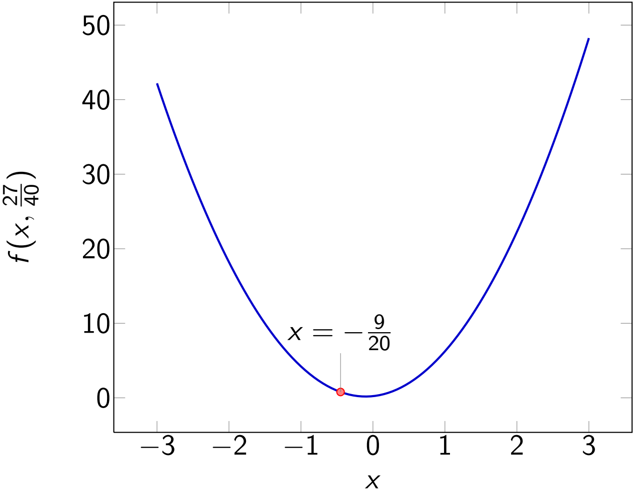
$$f(-\frac{9}{20}, y) = \frac{1}{2}y^2 - \frac{27}{40}y + \frac{81}{80} \qquad \text{Minima: } y = \frac{27}{40}$$

Concrete Example



- Are we done?

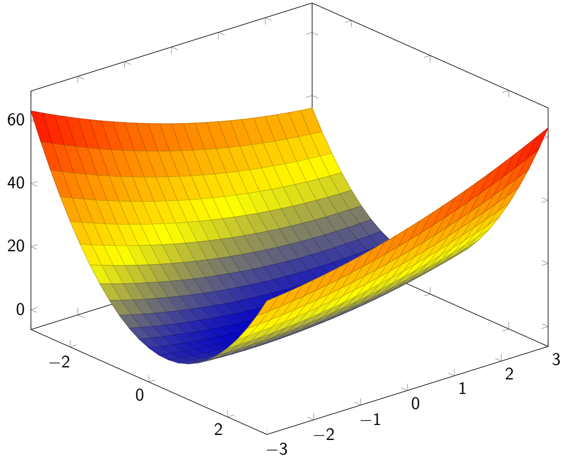
Concrete Example



- Are we done?

Outline

Problem Statement



- Given a black-box which can compute $J : \mathbb{R}^n \to \mathbb{R}$ and $\nabla J : \mathbb{R}^n \to \mathbb{R}^n$ find the minimum value of J

Basic Idea

- Given a location w_t at iteration t update

$$w_{t+1} = w_t - \eta_t \nabla J(w_t),$$

- η_t is a scalar stepsize

Gradient Descent Algorithm

```
Input: Initial point w_0, gradient norm tolerance ε
Set t = 0
while ‖∇J(w_t)‖ ≥ ε do
    w_{t+1} = w_t − η_t ∇J(w_t)
    t = t + 1
end while
Return: w_t
```

Line Search Strategies - I

- **Exact:** $J(w_t - \eta \nabla J(w_t))$ is a one dimensional convex function in η.
- **Inexact:** Armijio-Goldstein (Wolfe) conditions

$$\begin{aligned} J(w_{t+1}) &\leq J(w_t) + c_1 \eta_t \left\langle \nabla J(w_t), w_{t+1} - w_t \right\rangle \text{(sufficient decrease)} \\ \left\langle \nabla J(w_{t+1}), w_{t+1} - w_t \right\rangle &\geq c_2 \left\langle \nabla J(w_t), w_{t+1} - w_t \right\rangle \text{(curvature)} \end{aligned}$$

with $0 < c_1 < c_2 < 1$.

Line Search Strategies - II

- **Decaying Stepsize:** Use a stepsize which decays according to a fixed schedule, for example, $\eta_t = 1/\sqrt{t}$
- **Fixed Stepsize:** Suppose J has a Lipschitz continuous gradient with modulus L. Set $\eta_t = \frac{1}{L}$.