Unconstrained Optimization

An unconstrained non-linear optimization problem has the form:

$$\min_{x \in \mathbb{R}^n} \; f(x)$$

where x is the vector of *decision variables*, $f(x) : \mathbb{R}^n \to \mathbb{R}$ is the *objective function* that is to be minimized. There are no constraints on the decision variables (except that they are to be real numbers).

What about maximization problems?

Observe that

$$\max f(x) = \min -f(x)$$

So to maximize you can minimize the negative of the original objective function.

Goal: Find the Optimal Solution

Definition

A vector x^* is a *global optimal solution* for an unconstrained problem if $f(x) \geq f(x^*)$ for all $x \in \mathbb{R}^n$.

A vector x^* is a *strict global optimal solution* for an unconstrained problem if $f(x) > f(x^*)$ for all $x \in \mathbb{R}^n$. and x and x* are different

Example

Consider the unconstrained problem:

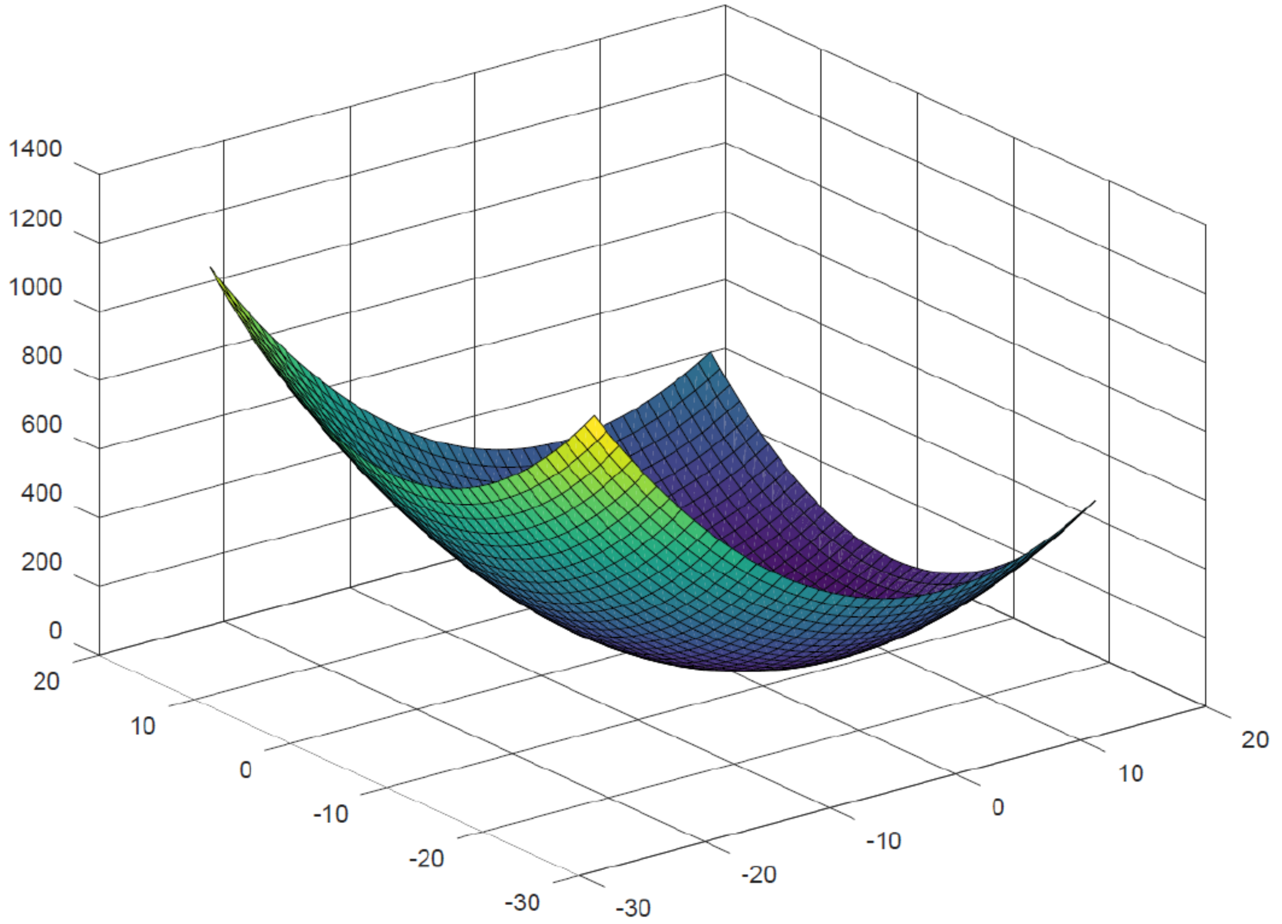
$$\min_{x \in \mathbb{R}^2} f(x) = (x_1 - 3)^2 + (x_2 + 4)^2$$

Note: $x \in \mathbb{R}^2$ means the vector can be written as $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

What is the optimal solution?

Graph of the function

$$\min_{x} \; f(x) = (x_1 - 3)^2 + (x_2 + 4)^2$$



Optimal Solution

What is the optimal solution(s) for

$$\min_{x \in \mathbb{R}^2} f(x) = (x_1 - 3)^2 + (x_2 + 4)^2$$

Try the vector $x^* = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$ Why is this the strict global optimal solution?

Key Insight

Observe that
$f(x) = (x_1 - 3)^2 + (x_2 + 4)^2 \geq 0$ for all $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$

and at $x^* = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$ we get $f(x^*) = 0$ so this establishes that x^* is a global minimum

(optimal) solution i.e.

$$f(x) = (x_1 - 3)^2 + (x_2 + 4)^2 \geq 0 = f(x^*)$$

But we are not quite done, why?

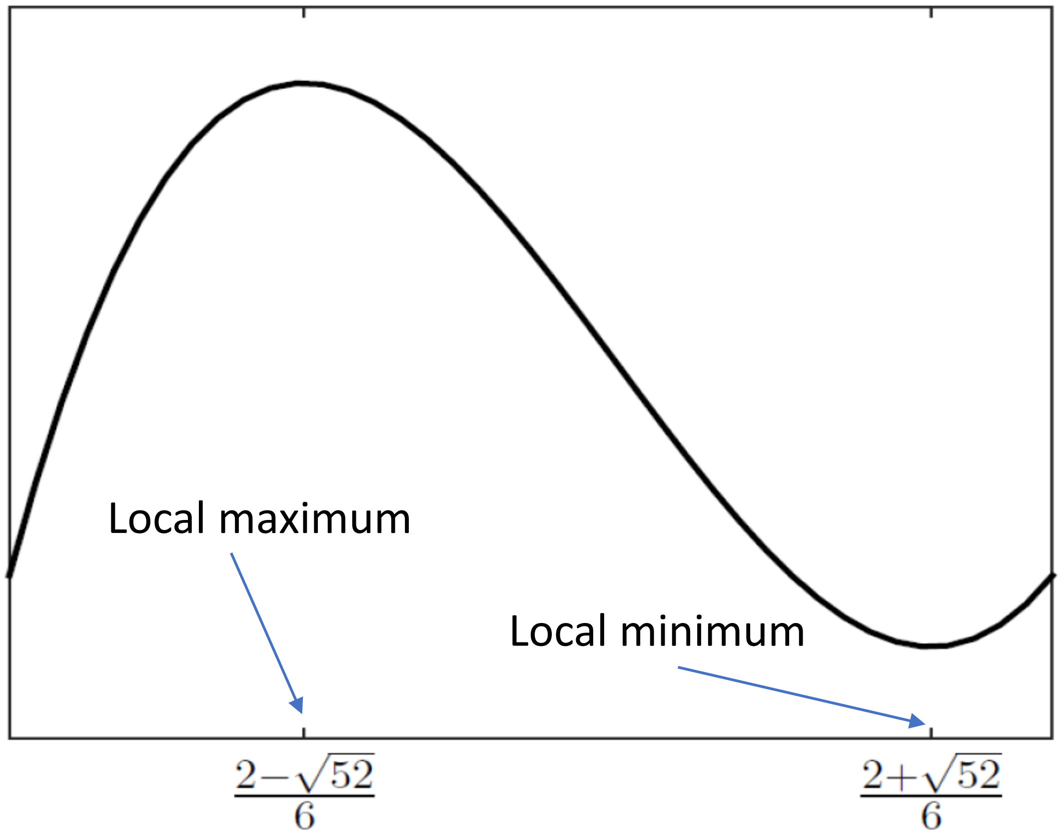
Functions with no global optimal solution

Consider the unconstrained problem:

$$\min_{x} \; f(x) = x^3 - x^2 - 4x - 6.$$

Observe that $f(x) \to -\infty$ as $x \to -\infty$ so there is no global minimum.

Graph



$$f(x) = x^3 - x^2 - 4x - 6$$

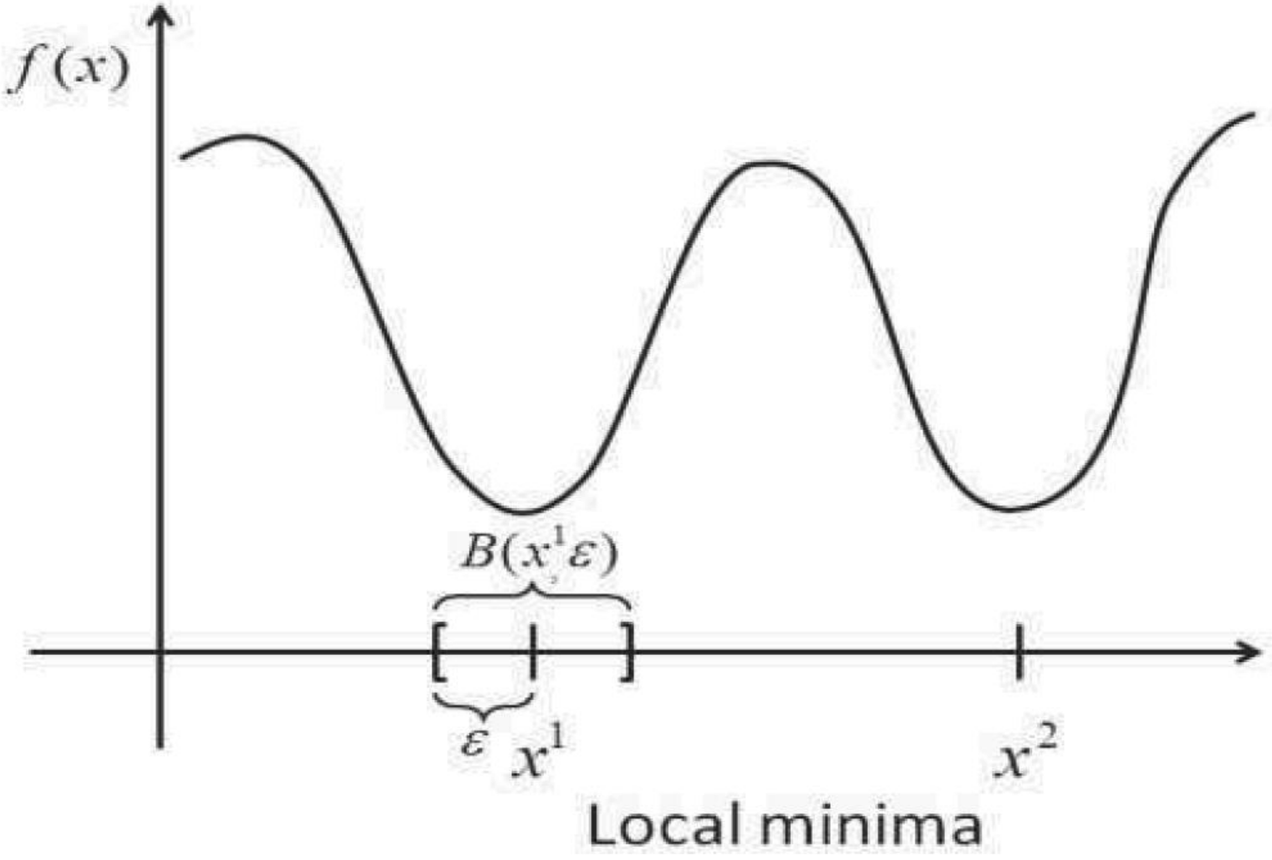
Local optima

Intuitively, a local minimum x^* of $f(x)$ is a vector x^*such that $f(x^*)$ is the lowest value in some *neighborhood* of x^*. That is, for any vector $\bar{x}$ in the neighborhood $f(x^*) \leq f(\bar{x})$ where the neighborhood is usually some strict subset of $\mathbb{R}^n$.

Formalizing the idea of neighborhood

Let $B(x^*, \varepsilon) = \{x \in R^n | \, \|x - x^*\|_2 \leq \varepsilon\}$, which is a ball in R^n with radius ε. In R^1, $B(x^*, \varepsilon)$ is the line segment $[x^* - \varepsilon, x^* + \varepsilon]$, in R^2 it is the disc with center $\bar{x}$ and radius ε, in R^3 is a sphere with center x^* and radius ε, etc.

Illustration



Optimality Conditions (FONC)

First-Order Necessary Condition (FONC): Consider an unconstrained nonlinear optimization problem of the form:

$$\min_{x \in \mathbb{R}^n} \; f(x).$$

If x^* is a local minimum, then $\nabla f(x^*) = 0$.

Review: Gradient of a function

If $f(x) : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then the function ∇f defined by

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

is called the gradient of f. The gradient is a function from $\mathbb{R}^n \to \mathbb{R}^n$.

Example Revisited

$$\min_{x \in \mathbb{R}^2} f(x) = (x_1 - 3)^2 + (x_2 + 4)^2$$

To find stationary points, we set the gradient of $f(x)$ equal to zero, which gives:

$$\nabla f(x) = \begin{pmatrix} 2(x_1 - 3) \\ 2(x_2 + 4) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

or:

$$x^* = \begin{pmatrix} 3 \\ -4 \end{pmatrix}.$$

FONC

FONC says

"Any vector $\bar{x}$ for which $\nabla f(\bar{x}) \neq 0$ can't be a local minimum."

On the other hand if a vector $\bar{x}$ satisfies $\nabla f(\bar{x}) = 0$ it is not necessarily true that $\bar{x}$ is a local minimum.

Any vector $\bar{x}$ that satisifes the condition $\nabla f(\bar{x}) = 0$ is called a *stationary point.*

Limitations of FONC for identifying local minima

Although this condition eliminates non-stationary points that cannot be local minima, it does not distinguish between local minima and local maxima.

Example re-visited

Consider the unconstrained problem:

$$\min_{x} \; f(x) = x^3 - x^2 - 4x - 6.$$

Example continued

To find stationary points, we set the gradient of $f(x)$ equal to zero, which gives:

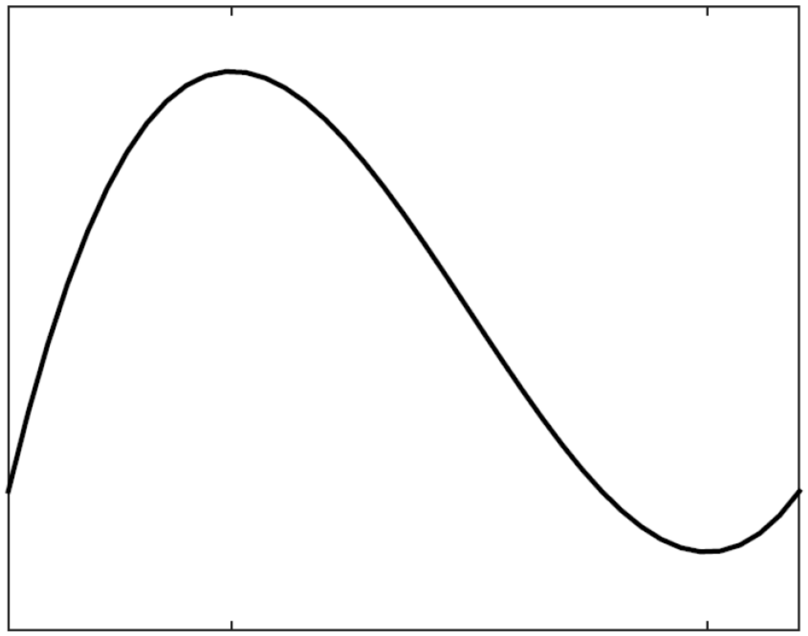
$$\nabla f(x) = 3x^2 - 2x - 4 = 0,$$

or:

$$x^* \in \left\{ \frac{2 - \sqrt{52}}{6}, \frac{2 + \sqrt{52}}{6} \right\}.$$

Both of these are stationary points and thus candidate local minima, based on the FONC.

Example continued



It is clear that only one of these two stationary points, $x^* = \frac{2+\sqrt{52}}{6}$, is a local minimum, whereas $\frac{2-\sqrt{52}}{6}$ is a local maximum.

Example of problem where no vector satisfies FONC

Consider the unconstrained problem:

$$\min_x \; f(x) = x_1 - 2x_2.$$

Example continued

The gradient of this objective function is:

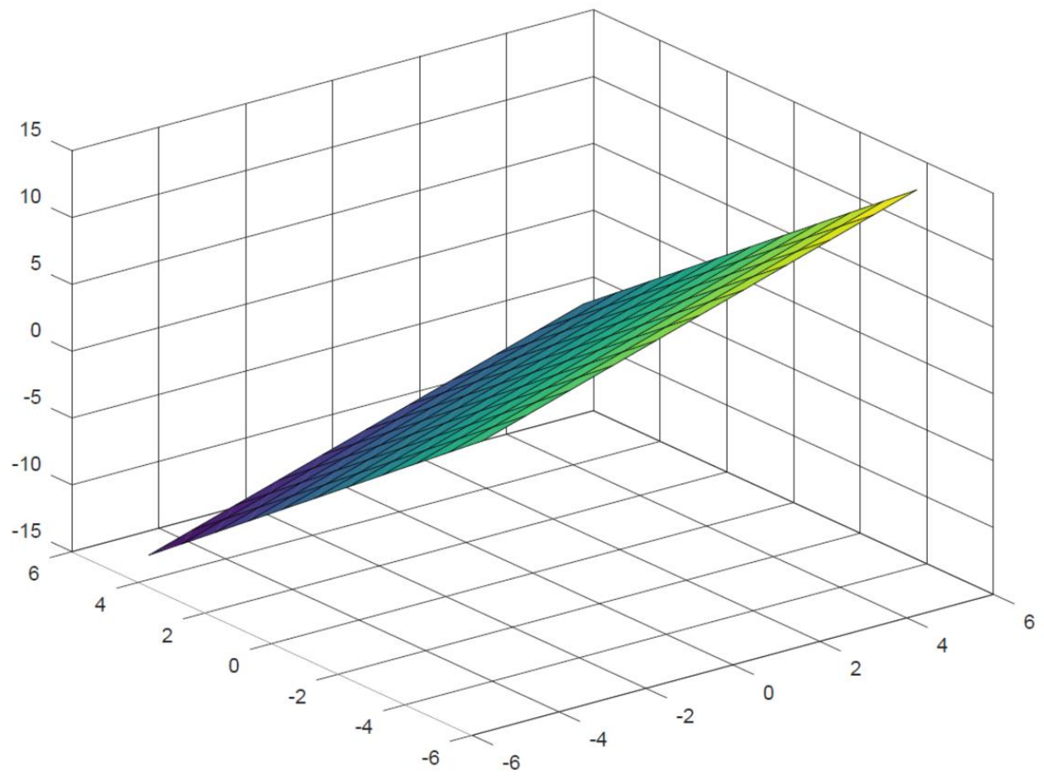
$$\nabla f(x) = \begin{pmatrix} 1 \\ -2 \end{pmatrix},$$

which cannot be made to equal zero. Because this problem does not have any stationary points, it cannot have any local minima. In fact, this problem is unbounded. To see this, note that:

$$\lim_{x_1 \to -\infty, x_2 \to +\infty} f(x) = -\infty.$$

Graph of Example

$$\min_{x} \; f(x) = x_1 - 2x_2$$



Limitations of FONC: Analyzing stationary points

- Assume that we can compute stationary points for an unconstrained minimization problem.
- How can we ascertain whether stationary points are a local min or a local max when a graph is not possible to generate (due to higher dimensions i.e. larger number of decision variables) when the FONC is unable to differentiate.
- Does one have to use ad hoc mathematical reasoning to determine whether stationary points are a local min or max? This can be very challenging or not even possible.

Some good news: Convex Optimization

There is a special class of unconstrained (and constrained problems) for which a stationary point (assuming $f(x)$ is differentiable) will not only be a local minimum but a global minimum!

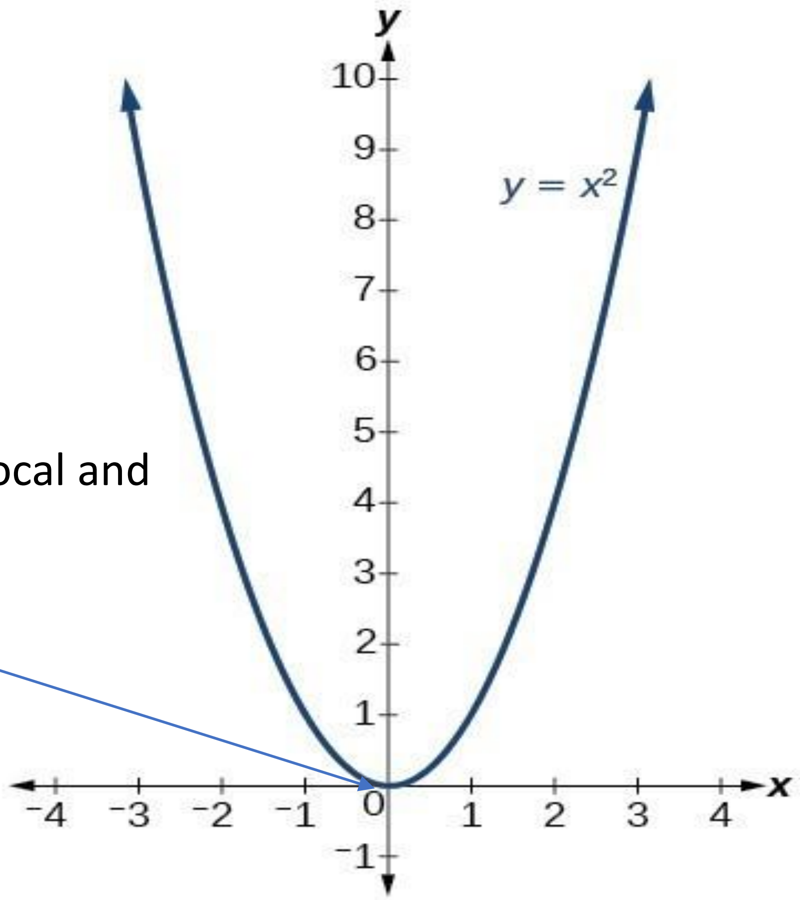
Consider the problem

$$\min_{x \in \mathbb{R}} x^2$$

Now $\nabla f(x) = 2x$ and so $x = 0$ satisfies $\nabla f(x) = 0$ and is the unique stationary point.

Graph of $y = f(x) = x^2$

x = 0 is a stationary point that is both a strict local and global minimum



Convex optimization

- What is special about the previous example? Did we just get lucky or are there some essential properties that can be found in many other instances of unconstrained minimization?
- Key properties
- [1] The domain of f(x) (in this case the real number line) is a convex set S.
- [2] The objective function f(x) is a convex function over the its domain.