1 Convexity, Convex Relaxations, and Global Optimization Algorithms[1]

1.1 Minima

Consider the nonlinear optimization problem in a Euclidean space:

$$\min_{\mathbf{x} \in S \subset \mathbb{R}^n} f(\mathbf{x}),$$

where $S \equiv$ the feasible region.

Definition (global minimum). A point $\mathbf{x}^* \in S$ is a global minimum of f if

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) \quad \forall\ x \in S.$$

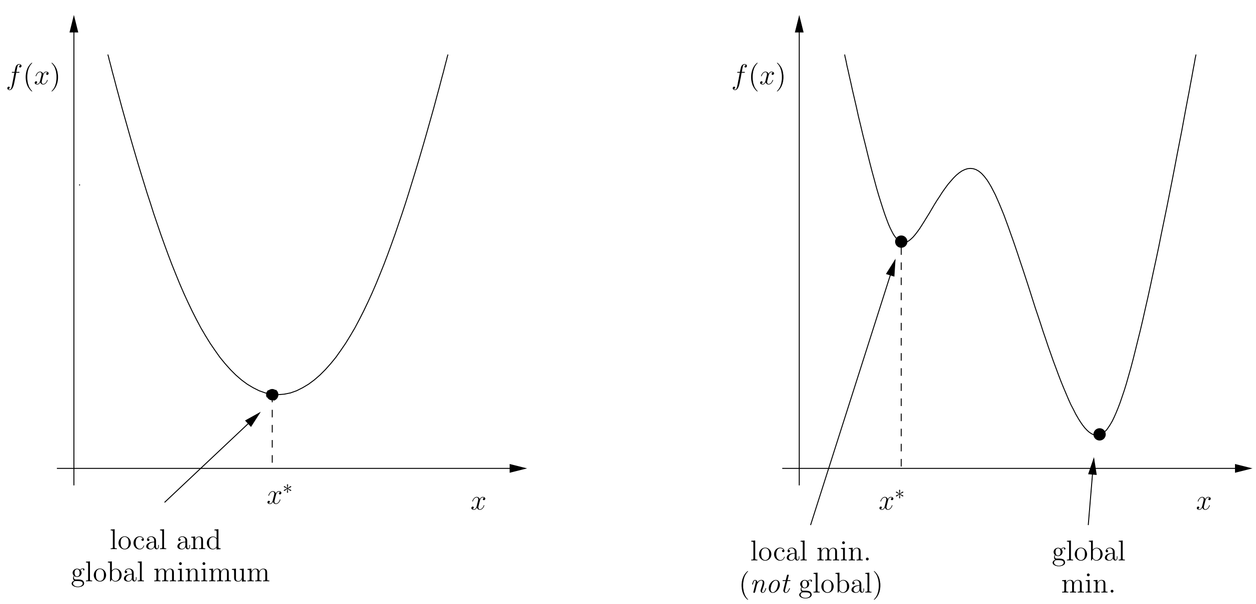
Synonyms: optimal solution, global optimal solution, or solution (to the optimization problem).

Definition (local minimum). A point $\mathbf{x}^* \in S$ is a local minimum if there exists an $\varepsilon > 0$ such that

$$f(\mathbf{x}) \geq f(\mathbf{x}^*)$$

$\forall\ \mathbf{x} \in S$ such that $||\mathbf{x} - \mathbf{x}^*|| < \varepsilon$.

Remark. Trivially, global minimum $\Rightarrow$ local minimum.



What is so special about the function on the left that the local minimum is also the global minimum? **CONVEXITY!** The property of convexity is so important because its definition implies that every local minimum is also a global minimum.

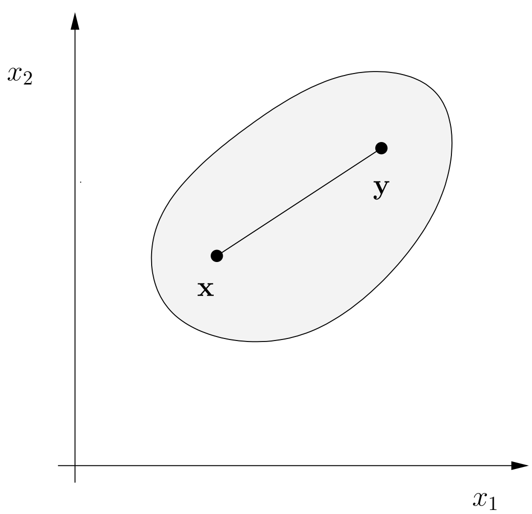
[1] ©PIB 1/2/02

1.2 Convex sets

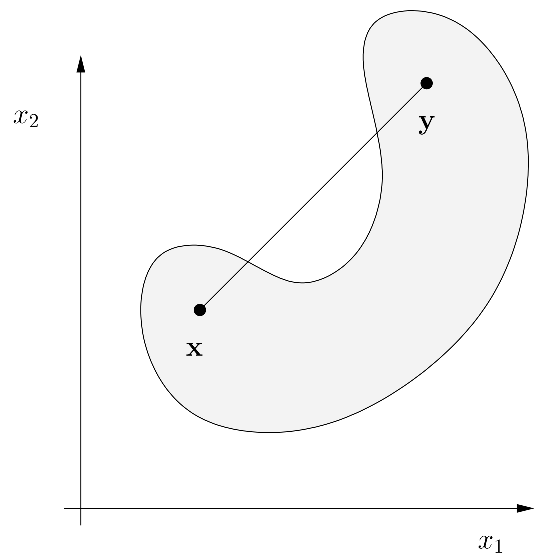
Definition (convex set). A set $S \subset \mathbb{R}^n$ is convex if for every $\mathbf{x},\ \mathbf{y} \in S$,

$$\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S \quad \forall\ \lambda \in [0,1].$$

Remark. In words, every point on the straight line connecting any pair of points in S must also be in S.



Convex Set



Nonconvex Set

N.B. Must hold for *all* pairs of points in the set.

The intersection of two convex sets is convex. By induction, the intersection of a finite number of convex sets is convex.

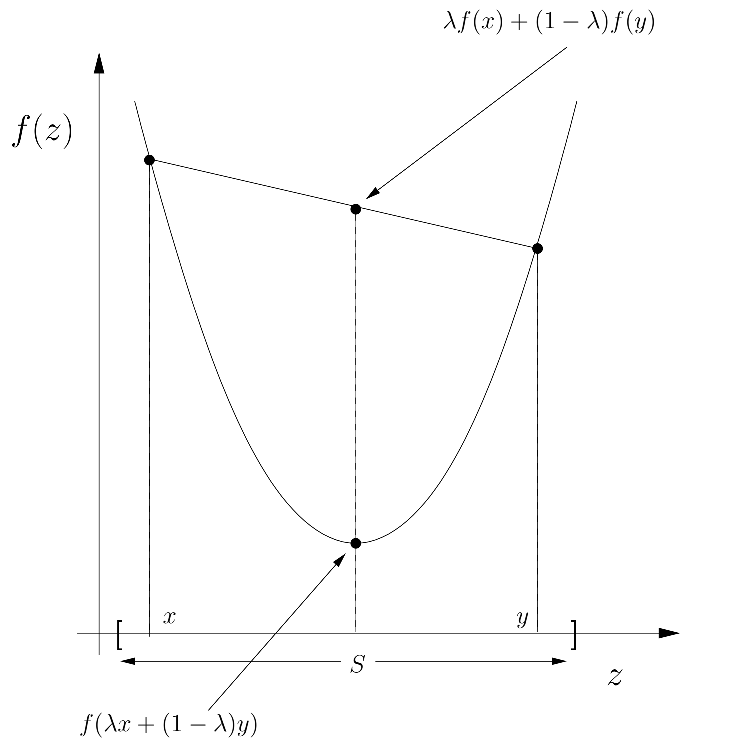
1.3 Convex functions

Definition (convex function). Let $f : S \to \mathbb{R}$, where S is a nonempty convex set in $\mathbb{R}^n$. The function is said to be convex on S if

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$$

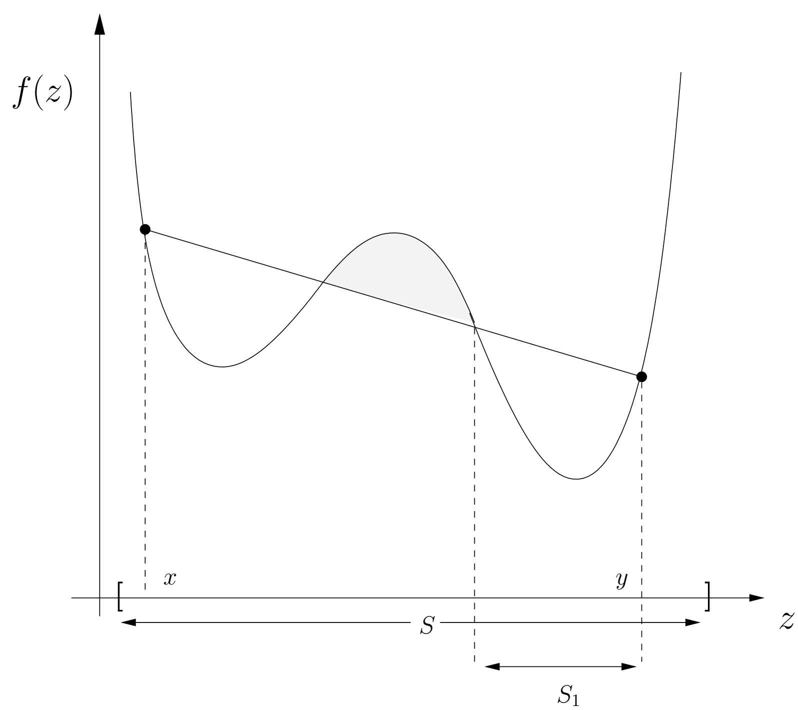
for each $\mathbf{x}, \mathbf{y} \in S$ and for each $\lambda \in (0,1)$.

Geometrically: Function always lies below straight line connecting any two points.



Convex Function

Must be satisfied for *all* pairs of points in S.



Nonconvex Function

Convexity depends on the set. In the above example, the function is convex on S_1 but nonconvex on S.

Lemma. Let S be a nonempty convex set in $\mathbb{R}^n$ and let $g : S \to \mathbb{R}$ be a convex function. Then, the set $\{\mathbf{x} \in S \,:\, g(\mathbf{x}) \leq \alpha,\ \alpha \in \mathbb{R}\}$ is convex.

1.4 Convex Optimization

Theorem. *Let S be a nonempty convex set in $\mathbb{R}^n$ and let $f : S \to \mathbb{R}$ be convex on S. Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Suppose $\mathbf{x}^* \in S$ is a local minimum. Then $\mathbf{x}^*$ is a global minimum.*

Proof. By hypothesis, $\mathbf{x}^*$ is a local minimum of f. Thus, by definition, there exists an $\varepsilon > 0$ such that

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) \quad \forall\ \mathbf{x} \in S \cap B_\varepsilon(\mathbf{x}^*)$$

where $B_\varepsilon(\mathbf{x}^*)$ is the open ball defined by $B_\varepsilon(\mathbf{x}^*) = \{x \,:\, ||\mathbf{x}-\mathbf{x}^*|| < \varepsilon\}$. Assume the contrary. That is, suppose that $\mathbf{x}^*$ is not a global minimum. This implies the existence of an $\hat{\mathbf{x}} \in S$ such that

$$f(\hat{\mathbf{x}}) < f(\mathbf{x}^*). \tag{1}$$

Choose $\lambda \in (0, 1)$ sufficiently small such that

$$\lambda\hat{\mathbf{x}} + (1-\lambda)\mathbf{x}^* \in S \cap B_\varepsilon(\mathbf{x}^*).$$

The existence of such a λ is clear by the Archimedean property of the real numbers. By the convexity of f and the Inequality (1), we have

$$\begin{aligned} f(\lambda\hat{\mathbf{x}} + (1-\lambda)\mathbf{x}^*) &\leq \lambda f(\hat{\mathbf{x}}) + (1-\lambda)f(\mathbf{x}^*) \\ &< \lambda f(\mathbf{x}^*) + (1-\lambda)f(\mathbf{x}^*) \\ &= f(\mathbf{x}^*). \end{aligned}$$

However, this contradicts that $\mathbf{x}^*$ is a local minimum of f.

□

In other words, for convex optimization (minimization of a convex function on a convex set):

local minimum ⇒ global minimum.

Why is this useful?

Example. Consider the unconstrained minimization:

$$\min_{\mathbf{x} \in S = \mathbb{R}^n} f(\mathbf{x}).$$

From elementary calculus, we know:

Theorem. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{x}^*$. If $\mathbf{x}^*$ is a local minimum, then*

$$\nabla f(\mathbf{x}^*) = \frac{\partial f(\mathbf{x}^*)}{\partial \mathbf{x}} = 0.$$

Remark. $\mathbf{x}$ such that $\nabla f(\mathbf{x}) = 0$ are known as stationary points.

Assume that we have an algorithm that can reliably find points such that

$$\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}^*) = 0.$$

Differentiable convex functions satisfy the following theorem.

Theorem. *Let S be a nonempty open convex set in $\mathbb{R}^n$, and let $f : S \to \mathbb{R}$ be differentiable on S. Then f is convex on S if and only if for any $\hat{\mathbf{x}} \in S$ we have*

$$f(\mathbf{x}) \geq f(\hat{\mathbf{x}}) + \nabla f(\hat{\mathbf{x}})^T(\mathbf{x} - \hat{\mathbf{x}}) \quad \forall\ \mathbf{x} \in S.$$

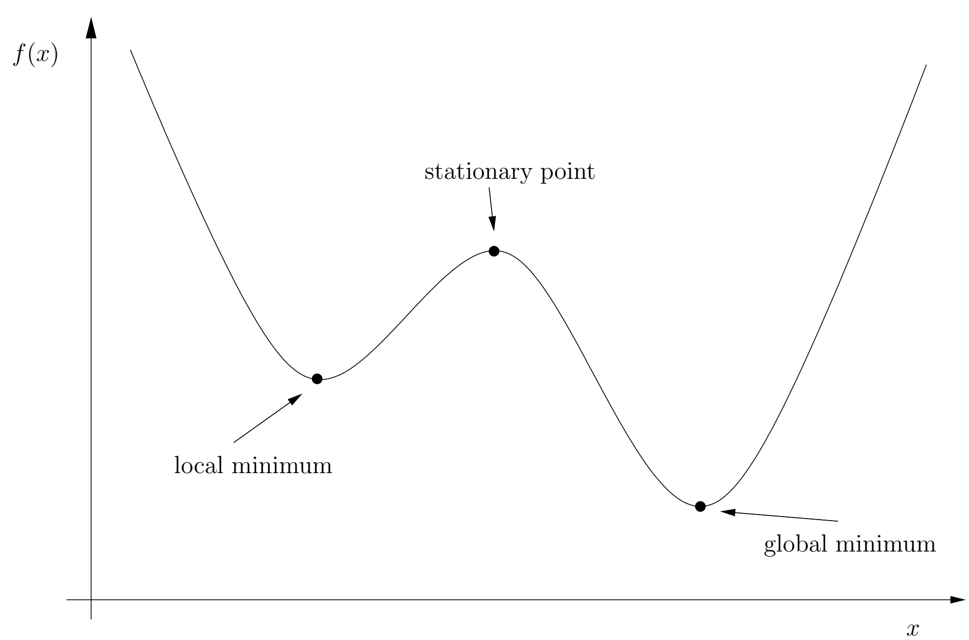
Now suppose $f(\mathbf{x})$ is differentiable and convex on $S = \mathbb{R}^n$. If $\nabla f(\mathbf{x}^*) = 0$ then

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) \quad \forall\ \mathbf{x} \in S.$$

Or, $\mathbf{x}^*$ is a global minimum on S.
Check: $\mathbb{R}^n$ is open and convex, so S satisfies the hypotheses.

This illustrates the general strategy: For convex optimization problems, if we have a reliable procedure for locating local minima, then we always get the global minimum. In fact, an infallible procedure to locate local minima of convex optimization problems appears very difficult to implement in practice, even if algorithms exist that promise this property theoretically. *N.B.* Might search for local minima directly, not stationary points.

1.5 Nonconvex Optimization



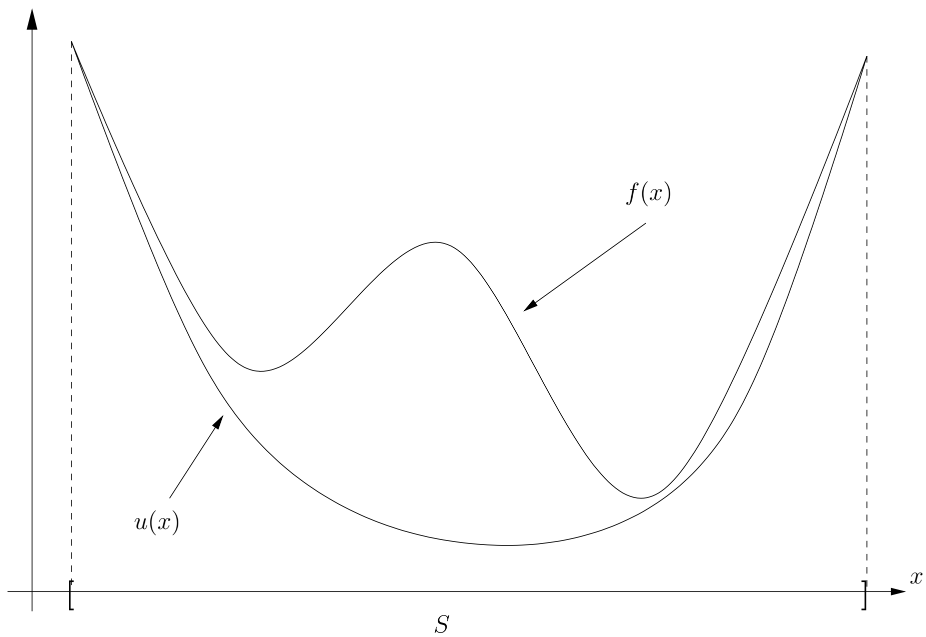
For nonconvex optimization problems, the strategy of finding local minima or stationary points fails. In the figure,

- All three points indicated are stationary points, one of which is actually a local maximum.
- Of the two local minima, only one is a global minimum.
- A simple descent strategy started to the left of the maximum will most likely locate the suboptimal local minimum.

A deterministic global optimization algorithm is designed to guarantee locating the global minimum objective function value within some ε tolerance with a finite number of iterations. There are many approaches to the design of deterministic global optimization algorithms but here we will only discuss the branch-and-bound approach (B&B). B&B relies on the notion of a convex relaxation of a nonconvex function.

Definition (convex relaxation). Let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is a nonempty convex set. Then a convex function $u : S \to \mathbb{R}$ is a convex relaxation of f if

$$u(\mathbf{x}) \leq f(\mathbf{x}) \quad \forall\, \mathbf{x} \in S.$$

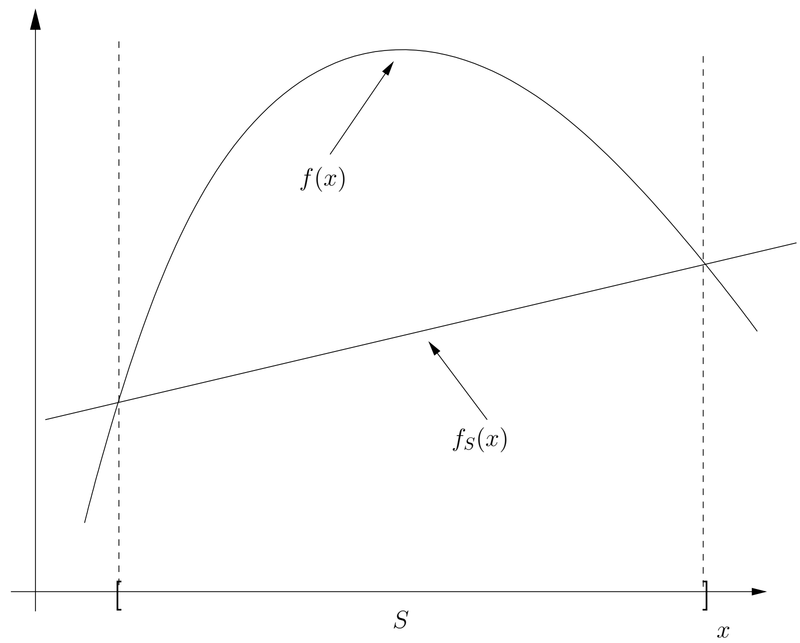


Definition (convex envelope). Let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is a nonempty convex set. The convex envelope of f over S (denoted f_S) is a convex relaxation such that for any other convex relaxation u of f on S, we have

$$f_S(\mathbf{x}) \geq u(\mathbf{x}) \quad \forall\, \mathbf{x} \in S.$$

Remark. The convex envelope is the tightest possible convex relaxation of a nonconvex function.

For a univariate concave function, the convex envelope is the secant joining the end points of the set S, as shown in the figure.



The convex envelopes of many functions are known. However, in general, finding the convex envelope of an arbitrary function is as hard as finding the global minimum. On the other hand, a number of polynomial algorithms exist for constructing convex relaxations of quite general classes of functions.

Suppose we have the nonconvex optimization problem

$$\min_{\mathbf{x}\in S} f(\mathbf{x})$$

such that

$$g(\mathbf{x}) \leq 0$$

where $S \subset \mathbb{R}^n$ is a nonempty convex compact set, $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}^m$ are continuous and potentially nonconvex. We can construct a convex optimization problem that is a relaxation of this problem via convex relaxations u and h of f and g respectively on S:

$$\min_{\mathbf{x}\in S} u(\mathbf{x})$$

such that

$$h(\mathbf{x}) \leq 0.$$

N.B. The set $S \cup \{\mathbf{x} : h(\mathbf{x}) \leq 0\}$ is convex, so every local minimum will also be a global minimum.

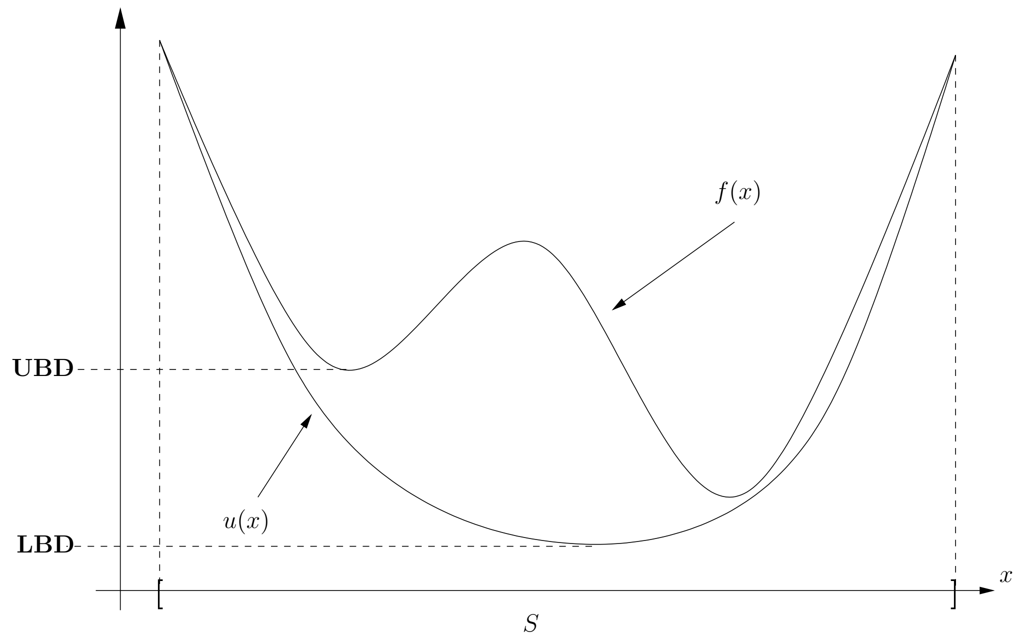
Consider the minimum of the nonconvex and convex problems, $\mathbf{x}^*$ and $\hat{\mathbf{x}}$, respectively. Both $\mathbf{x}^*$ and $\hat{\mathbf{x}}$ will be feasible for the convex problem because

$$h(\mathbf{x}^*) \leq g(\mathbf{x}^*) \leq 0.$$

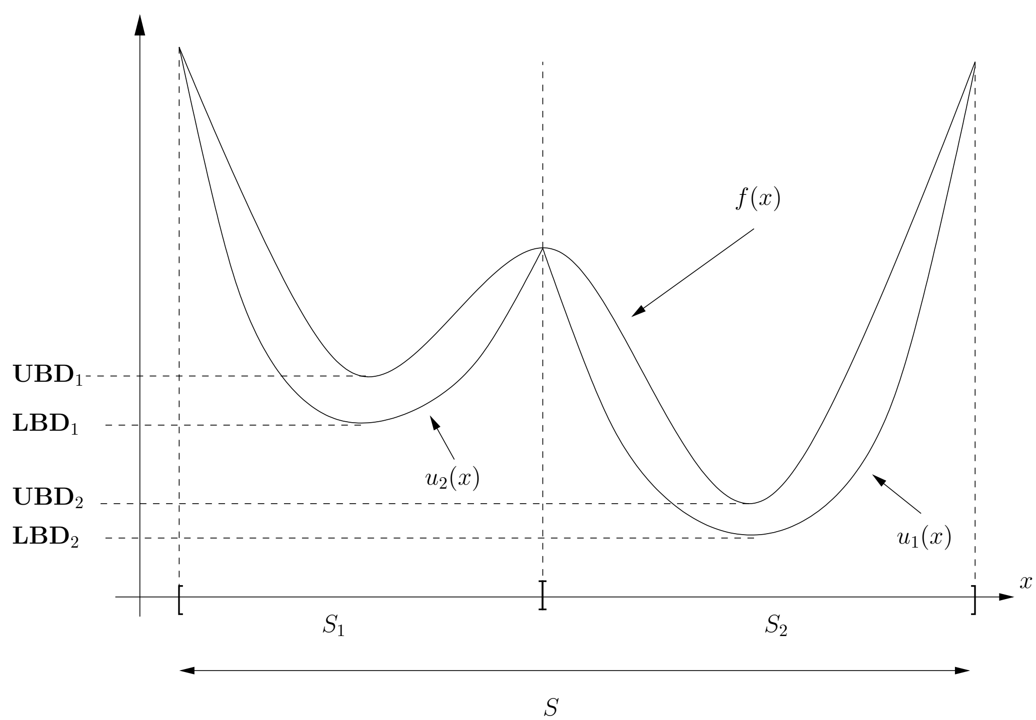
Moreover,

$$u(\hat{\mathbf{x}}) \leq u(\mathbf{x}^*) \leq f(\mathbf{x}^*).$$

i.e., the minimum for the convex problem is a lower bound on the minimum of the nonconvex problem. Hence, the term convex underestimating problem is frequently used. These preliminaries enable us to describe the B&B procedure for nonconvex optimization.

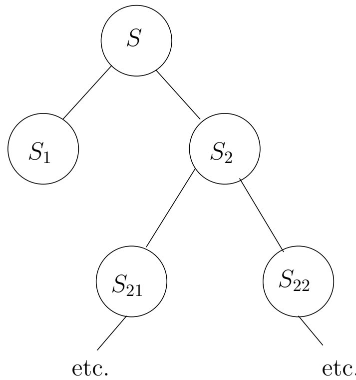


Partition the space S:



Now, because $\text{LBD}_1 \geq \text{UBD}_2$, the minimum cannot be attained in the set S_1. So, S_1 is excluded from further considerations, or *fathomed*. S_2 is further partitioned, fathoming as necessary until LBD on all sets comes within ε of the UBD.

The partitioning procedure can be illustrated by the following branch-and-bound tree:



The labels denote the partition over which the LBD and UBD are generated.

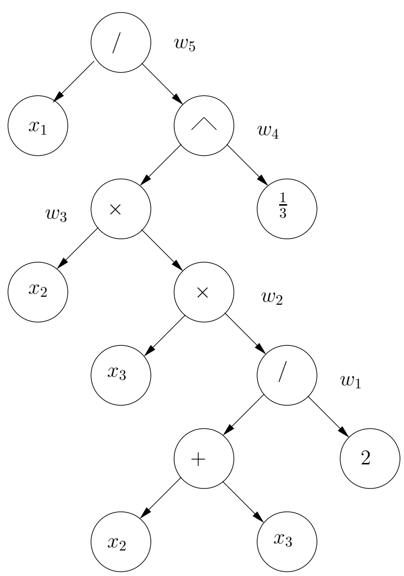
In order to have finite termination for ε tolerance, it is also necessary to have convergence of the convex relaxation to the nonconvex function as the size of the partition tends toward zero.

1.6 Constructing convex relaxations using AD

Example. Consider the following nonconvex function:

$$f(x) = \frac{x_1}{\left[x_2 x_3 \left(\frac{x_2+x_3}{2}\right)\right]^{1/3}}$$

which can be represented by the following binary tree:



or the elementary operation list:

$$\begin{aligned} w_1 &= (x_2 + x_3)/2 \\ w_2 &= x_3 w_1 \\ w_3 &= x_2 w_2 \\ w_4 &= w_3^{1/3} \\ w_5 &= x_1 / w_4 \end{aligned}$$

These representations indicate that the optimization problem

$$\min_{\mathbf{x} \in X} f(x)$$

can be reformulated as

$$\min_{\mathbf{x} \in X, \mathbf{w} \in W} w_5$$

such that

$$\begin{aligned} w_1 &= (x_2 + x_3)/2 \\ w_2 &= x_3 w_1 \\ w_3 &= x_2 w_2 \\ w_4 &= w_3^{1/3} \\ w_5 &= x_1 / w_4 \end{aligned}$$

A convex relaxation of this problem may be constructed via the known convex and concave envelopes of the RHS of the equality constraints. For example:

- linear constraints such as

  $$w_1 = (x_2 + x_3)/2$$

  always define a convex set.

- The convex and concave envelopes for any bilinear term such as $x_3 w_1$ are known, leading to the following relaxation:

  $$\begin{aligned} w_2 &\geq x_3^L w_1 + x_3 w_1^L - x_3^L w_1^L \\ w_2 &\geq x_3^U w_1 + x_3 w_1^U - x_3^U w_1^U \\ w_2 &\leq x_3^L w_1 + x_3 w_1^U - x_3^L w_1^U \\ w_2 &\leq x_3^U w_1 + x_3 w_1^L - x_3^U w_1^L \end{aligned}$$

  where x_3^L, x_3^U are lower and upper bounds on x_3 in the partition of interest, and w_1^L and w_1^U can be inferred by interval analysis from the elementary functions and the bounds on the x_i and other w_i