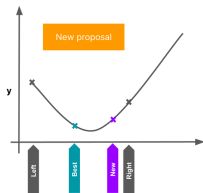


Optimization in Machine Learning

Univariate optimization

Golden ratio



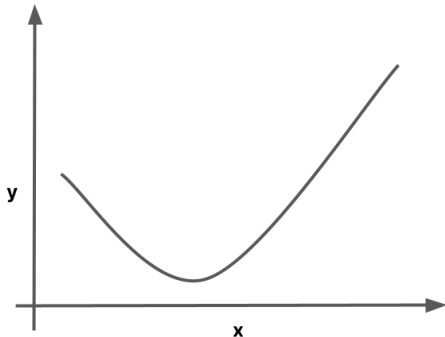
Learning goals

- Simple nesting procedure
- Golden ratio

UNIVARIATE OPTIMIZATION

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

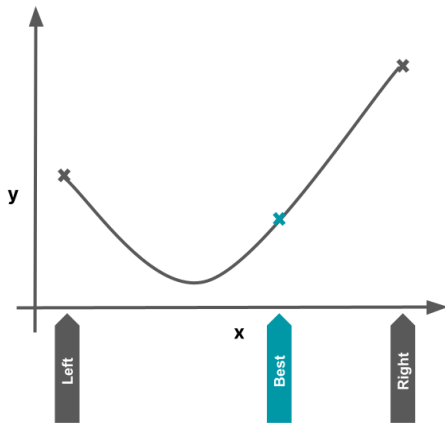
Goal: Iteratively improve eval points. Assume function is unimodal. Will not rely on gradients, so this also works for black-box problems.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

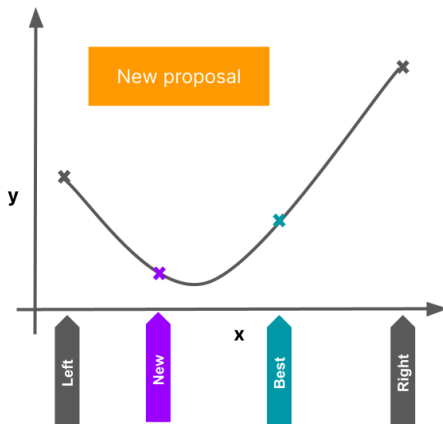
Always maintain three points: left, right, and current best.



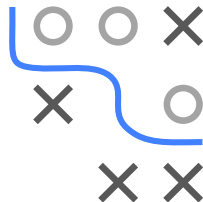
SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

Propose random point in interval.



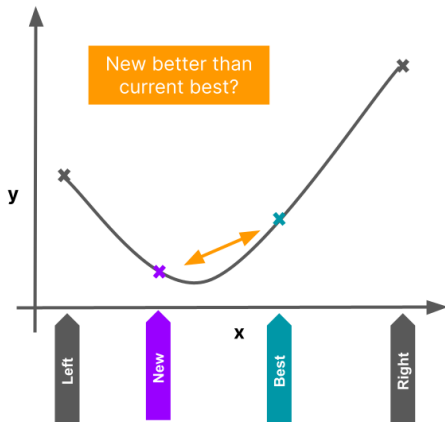
NB: Later we will define the optimal choice for a new proposal.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

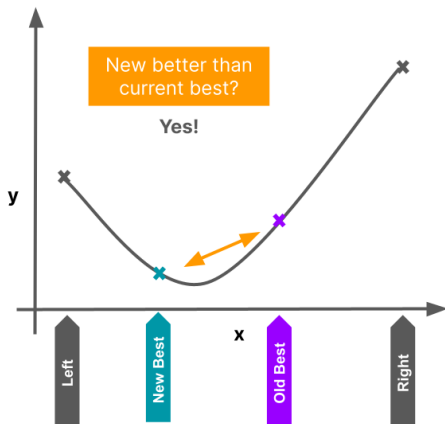
Compare proposal against current best.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

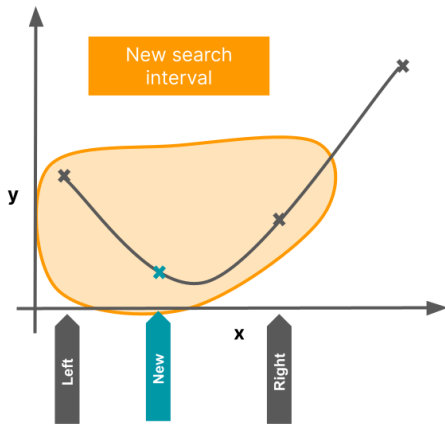
If it is better: proposal becomes current best.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

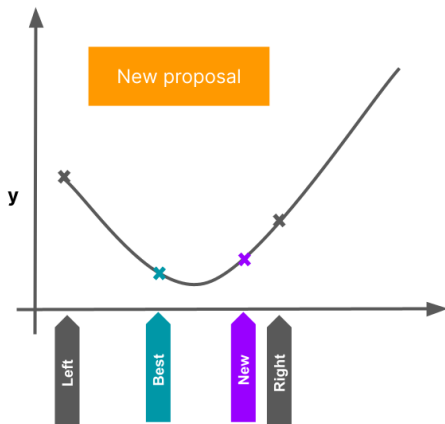
New search interval: around current best.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

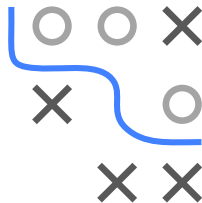
Propose a random point.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

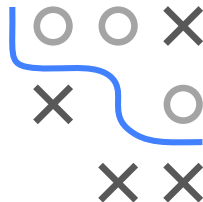
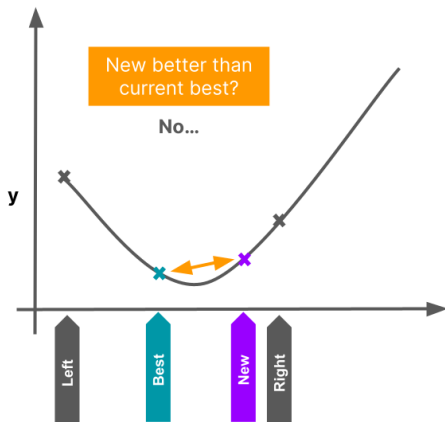
Compare proposal against current best.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

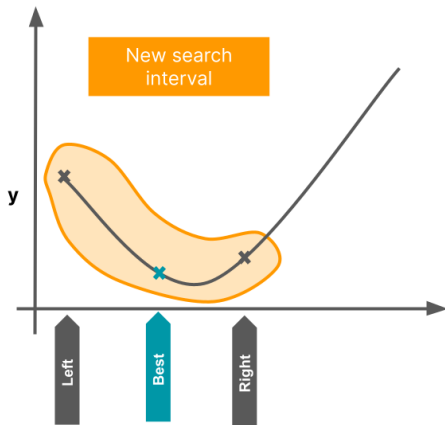
If it is better: proposal becomes current best.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

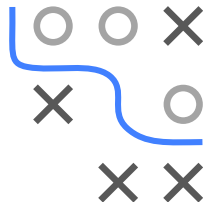
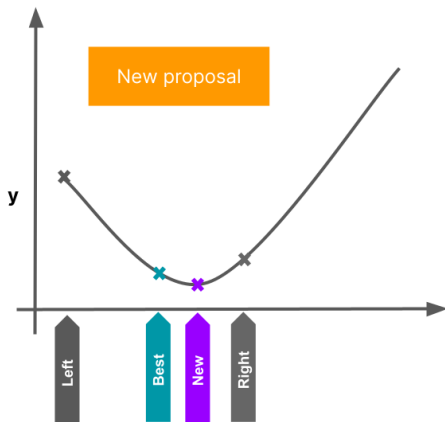
New search interval: around current best.



SIMPLE NESTING PROCEDURE

Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

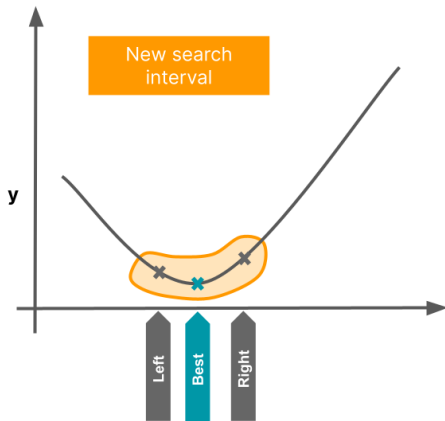
Propose a random point.



SIMPLE NESTING PROCEDURE

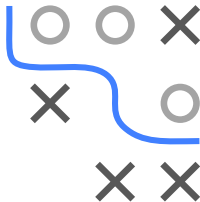
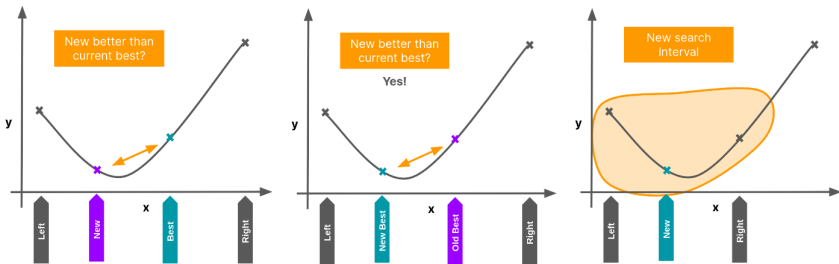
Let $f : \mathbb{R} \rightarrow \mathbb{R}$.

New search interval: around current best.



SIMPLE NESTING PROCEDURE

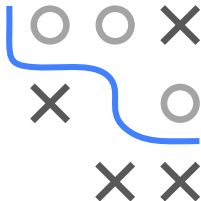
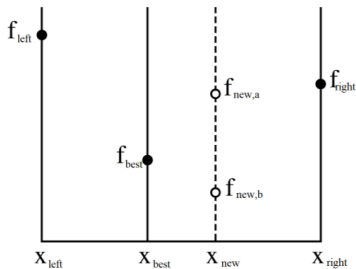
- **Initialization:** Search interval $(x^{\text{left}}, x^{\text{right}})$, $x^{\text{left}} < x^{\text{right}}$
- Choose x^{best} randomly.
- For $t = 0, 1, 2, \dots$
 - Choose x^{new} randomly in $[x^{\text{left}}, x^{\text{right}}]$
 - If $f(x^{\text{new}}) < f(x^{\text{best}})$:
 - $x^{\text{best}} \leftarrow x^{\text{new}}$
 - New interval: Points around x^{best}



GOLDEN RATIO

Key question: How can x^{new} be chosen better than randomly?

- **Insight 1:** Always in bigger subinterval to maximize reduction.
- **Insight 2:** x^{new} symmetrically to x^{best} for uniform reduction.

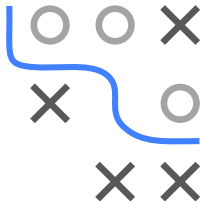
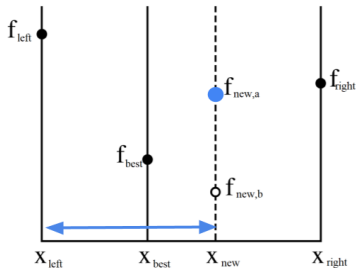


Consider two hypothetical outcomes x^{new} : $f_{\text{new},a}$ and $f_{\text{new},b}$.

GOLDEN RATIO / 2

If $f_{new,a}$ is the outcome, x_{best} stays best and we search around x_{best} :

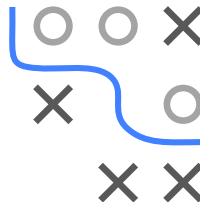
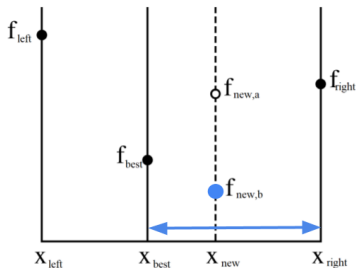
$[X_{left}, X_{new}]$



GOLDEN RATIO / 3

If $f_{new,b}$ is outcome, x_{new} becomes best point and search around x_{new} :

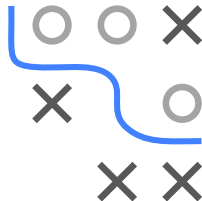
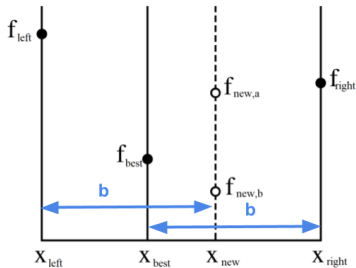
$$[x_{best}, x_{right}]$$



GOLDEN RATIO / 4

For uniform reduction, require the two potential intervals equal sized:

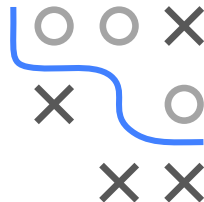
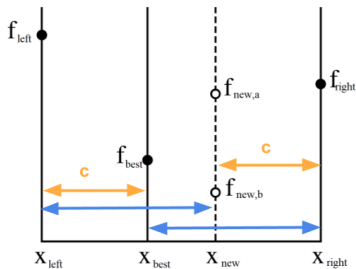
$$b := x_{\text{right}} - x_{\text{best}} = x_{\text{new}} - x_{\text{left}}$$



GOLDEN RATIO / 5

One iteration ahead: require again the intervals to be of same size.

$$C := x_{\text{best}} - x_{\text{left}} = x_{\text{right}} - x_{\text{new}}$$



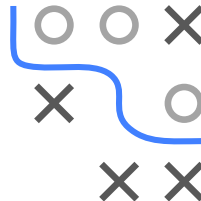
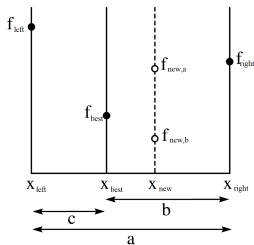
GOLDEN RATIO / 6

To summarize, we require:

$$a = x^{right} - x^{left},$$

$$b = x_{right} - x_{best} = x_{new} - x_{left}$$

$$c = x_{best} - x_{left} = x_{right} - x_{new}$$



GOLDEN RATIO / 7

- We require the same percentage improvement in each iteration
- For φ reduction factor of interval sizes (a to b , and b to c)

$$\varphi := \frac{b}{a} = \frac{c}{b}$$

$$\varphi^2 = \frac{b}{a} \cdot \frac{c}{b} = \frac{c}{a}$$

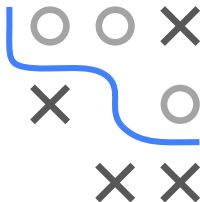
- Divide $a = b + c$ by a :

$$\frac{a}{a} = \frac{b}{a} + \frac{c}{a}$$

$$1 = \varphi + \varphi^2$$

$$0 = \varphi^2 + \varphi - 1$$

- Unique positive solution is $\varphi = \frac{\sqrt{5}-1}{2} \approx 0.618$.



GOLDEN RATIO / 8

- With x^{new} we always go φ percentage points into the interval.
- Given x^{left} and x^{right} it follows

$$\begin{aligned}x^{\text{best}} &= x^{\text{right}} - \varphi(x^{\text{right}} - x^{\text{left}}) \\ &= x^{\text{left}} + (1 - \varphi)(x^{\text{right}} - x^{\text{left}})\end{aligned}$$

and due to symmetry

$$\begin{aligned}x^{\text{new}} &= x^{\text{left}} + \varphi(x^{\text{right}} - x^{\text{left}}) \\ &= x^{\text{right}} - (1 - \varphi)(x^{\text{right}} - x^{\text{left}}).\end{aligned}$$



GOLDEN RATIO / 9

Termination criterion:

- A reasonable choice is the absolute error, i.e. the width of the last interval:

$$|x^{best} - x^{new}| < \tau$$

- In practice, more complicated termination criteria are usually applied, for example in *Numerical Recipes in C, 2017*

$$|x^{right} - x^{left}| \leq \tau(|x^{best}| + |x^{new}|)$$

is proposed as a termination criterion.

