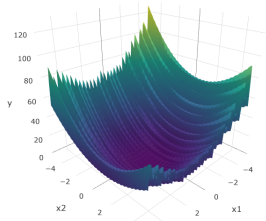
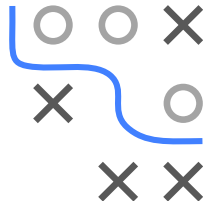


Optimization in Machine Learning

Multi-Start Optimization



Learning goals

- Multimodal functions
- Basins of Attractions
- Simple multi-start procedure

MOTIVATION

- So far: derivative-free methods for *unimodal* objective function (exception: simulated annealing)
- With multimodal objective functions, methods converge to **local minima**.
- Optimum found may differ for different starting values $\mathbf{x}^{[0]}$



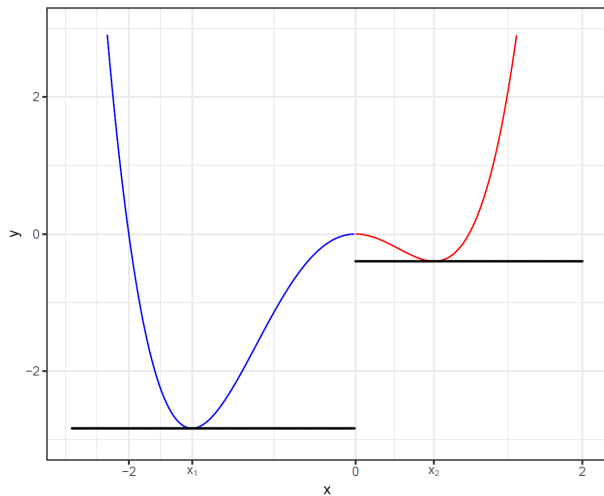
Attraction areas:

- Let f_1^*, \dots, f_k^* be local minimum values of f with $f_i^* \neq f_j^* \quad \forall i \neq j$.
- Notation: $A(\mathbf{x}^{[0]})$ denotes result of algorithm A started at $\mathbf{x}^{[0]}$
- Then: Set

$$\mathcal{A}(f_i^*, A) = \{\mathbf{x} : A(\mathbf{x}) = f_i^*\}$$

is called *attraction area/basin of attraction* of f_i^* for algorithm A

ATTRACTION AREAS

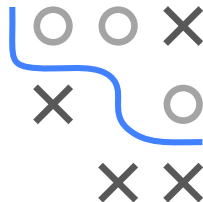


basin of attraction paths

— x_1

— x_2

● optimum found by Nelder-Mead

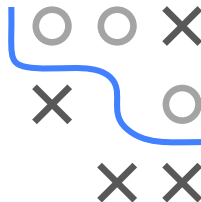
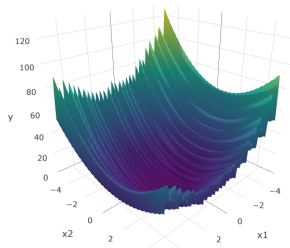


MULTI-STARTS

Levy function:

$$f(\mathbf{x}) = \sin^2(3\pi x_1) + (x_1 - 1)^2[1 + \sin^2(3\pi x_2)] + (x_2 - 1)^2[1 + \sin^2(2\pi x_2)]$$

- Global minimum: $f(\mathbf{x}^*) = 0$ at $\mathbf{x}^* = (1, 1)^\top$
- Optimize f by BFGS method with random starting point in $[-2, 2]^2$ and collect result
- Repeat 100 times



Distribution of results (y values):

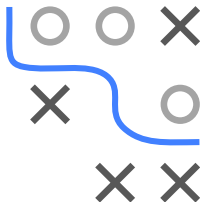
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0000	0.1099	0.5356	2.4351	1.9809	18.3663

MULTI-STARTS / 2

Idea: use multiple starting points $\mathbf{x}^{[1]}, \dots, \mathbf{x}^{[k]}$ for algorithm A

Algorithm Multistart optimization

- 1: Given: optimization algorithm $A(\cdot)$, $f : \mathcal{S} \mapsto \mathbb{R}$, $\mathbf{x} \mapsto f(\mathbf{x})$
 - 2: $k = 0$
 - 3: **repeat**
 - 4: Draw starting point $\mathbf{x}^{[k]}$ from \mathcal{S} (e.g. uniform if \mathcal{S} is of finite volume)
 - 5: **if** $k = 0$ **then** $\hat{\mathbf{x}} = \mathbf{x}^{[0]}$
 - 6: **end if**
 - 7: Initialize algorithm with start value $\mathbf{x}^{[k]} \Rightarrow \tilde{\mathbf{x}} = A(\mathbf{x}^{[k]})$
 - 8: **if** $f(\tilde{\mathbf{x}}) < f(\hat{\mathbf{x}})$ **then** $\hat{\mathbf{x}} = \tilde{\mathbf{x}}$
 - 9: **end if**
 - 10: $k = k + 1$
 - 11: **until** Stop criterion fulfilled
 - 12: **return** $\hat{\mathbf{x}}$
-



MULTI-STARTS / 3

BFGS with Multistart gives us the true minimum of the Levy function:

```
iters = 20 # number of starts
xbest = c(runif(1, -2, 2), runif(1, -2, 2))

for (i in 1:iters) {
  x1 = runif(1, -2, 2)
  x2 = runif(1, -2, 2)
  res = optim(par = c(x1, x2), fn = f, method = "BFGS")
}

if (res$value < f(xbest)) {
  xbest = res$par
}

xbest
## [1] 1 1
```

