Derivative Free Optimization

Computational Methods in Optimization

CS 520, Purdue

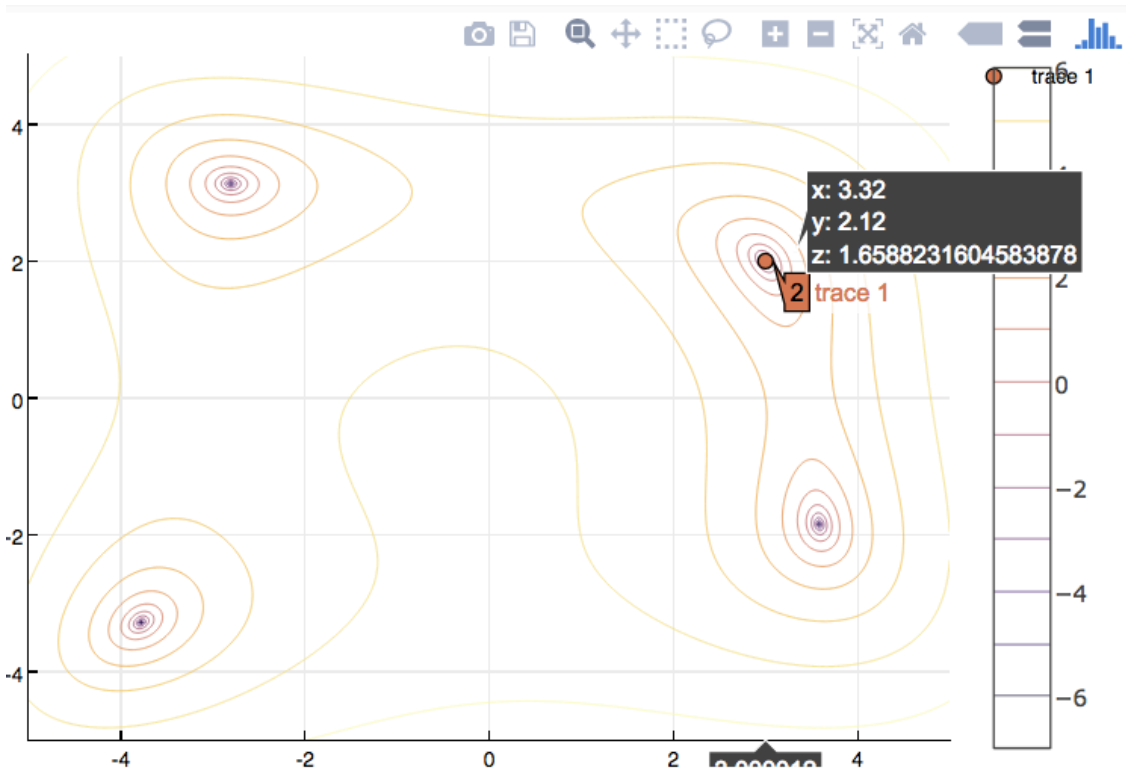
David F. Gleich

Purdue University

2017-04-20

A demo

```
f = x -> (x[1].^2 + x[2] - 11).^2 +
      (x[1] + x[2].^2 - 7).^2
sol = optimize(f, [0.0;0.0], NelderMead())
xs = sol.minimizer
```



Derivative-free optimization (DFO)

Chapter 9

Question

How would you do optimization without derivatives?

Solution 1

Use finite differences

$$f'(x) \approx \frac{1}{\gamma}(f(x+\gamma) - f(x))$$

Solution 1

Use finite differences

$$f'(x) \approx \frac{1}{\gamma}(f(x+\gamma) - f(x))$$

How to pick gamma?
How much work?

Solution 2

Interpolate and update a quadratic model (Section 9.2)

$$m_k(\mathbf{x}_k + \mathbf{p}) = c + \mathbf{q}^T\mathbf{p} + \tfrac{1}{2}\mathbf{p}^T\mathbf{G}\mathbf{p}$$

$$m_k(\mathbf{y}_\ell) = f(\mathbf{y}_\ell) \quad 1 \leq \ell \leq \text{total parameters}$$

Then use a trust-region method.

Solution 2

- How to find *c*, *q*, and *G*?

  $O(n^2)$ parameters

  How to choose the point set y_l?

- How to update *c*, *q*, and *G*?

  Details of interpolation methods. See the book, or references.

Solution 2

- How to find *c*, *q*, and *G*? $O(n^2)$ parameters

  Use interpolation condition to form an n^2 by n^2 linear system

  $O(n^6)$ to solve

  $O(n^4)$ to update

  How to choose the point set y_l?

Solution 3

Fix a sequence of search directions that span R^n, and cycle among them

$$\text{“p = ”}\quad \mathbf{e}_1, -\mathbf{e}_1, \ldots, \mathbf{e}_n, -\mathbf{e}_n, \mathbf{e}_1, -\mathbf{e}_1 \ldots, \mathbf{e}_n, \ldots$$
$$\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_n, \pm\mathbf{e}_{n-1}, \ldots, \pm\mathbf{e}_1, \ldots$$

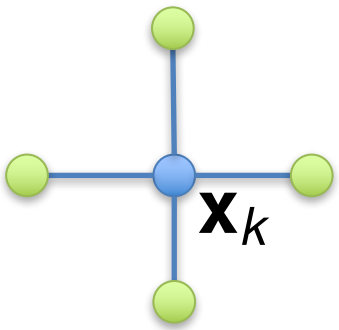
brutally slow in general

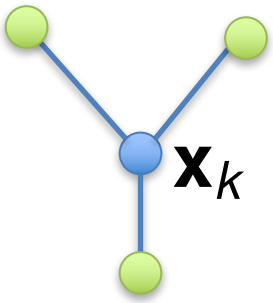
wickedly fast when applicable

(like a scalpel)

Solution 4

Pick a stencil around the current point





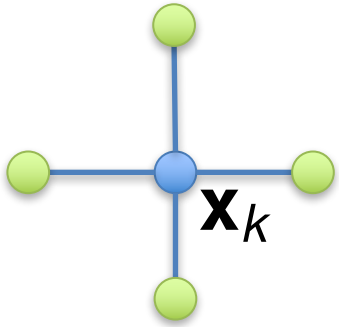
Example stencils in R^2

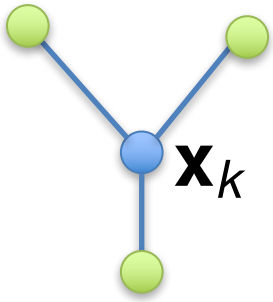
$$= \mathbf{x}_k + \gamma_k \mathbf{p}_k, \mathbf{p}_k \in \mathcal{D}_k$$

Move to the best point if "good enough" (sufficient decrease)

Otherwise, reduce gamma and revaluate

Solution 4





Example stencils in R^2

$$= \mathbf{x}_k + \gamma_k \mathbf{p}_k, \mathbf{p}_k \in \mathcal{D}_k$$

We need $\beta_{\text{min}} \leq \mathbf{p} \leq \beta_{\text{max}}$

$$\min_{\mathbf{v} \in \mathbb{R}^n} \max_{\mathbf{p} \in \mathcal{D}_k} \frac{\mathbf{v}^T \mathbf{p}}{\|\mathbf{p}\| \, \|\mathbf{v}\|} \geq \delta$$

This ensures we can always get at least a delta projection on any gradient.

for $\mathbf{p} \in \mathcal{D}_k$ to satisfy Zoutendijk

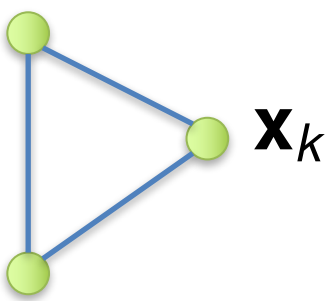
Solution 5 – Nelder-Mead

Consider a simplex of points

A simplex consists of n+1 non-colinear points

$$f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \ldots \leq f(\mathbf{x}_{n+1})$$

We order the vertices by decreasing function value.

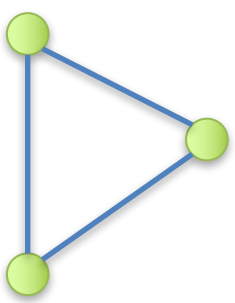


Such a simplex gives us a local "linear" model of our function!

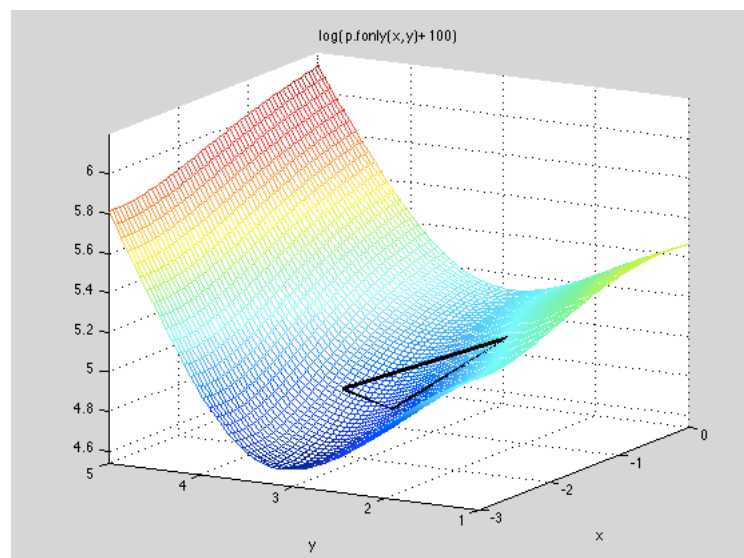
Solution 5 – Nelder-Mead

Use the "slope" of the simplex to find a good direction

$$f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \ldots \leq f(\mathbf{x}_{n+1})$$



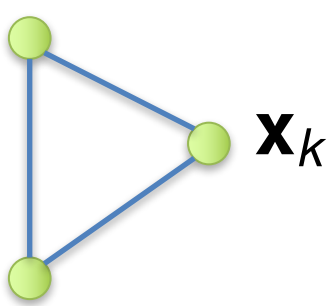
$\mathbf{x}_k$



Solution 5 – Nelder-Mead

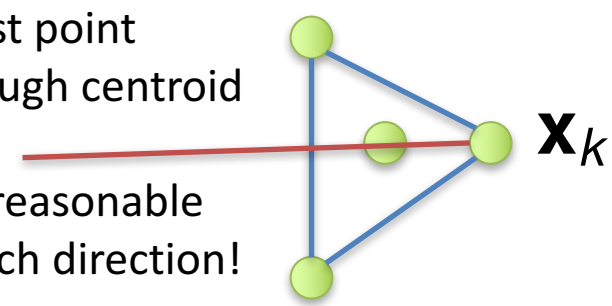
Use the "slope" of the simplex to find a good direction

$$f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \ldots \leq f(\mathbf{x}_{n+1})$$



The line from the worst point through centroid

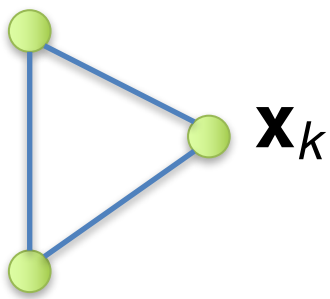
is a reasonable search direction!



Solution 5 – Nelder-Mead

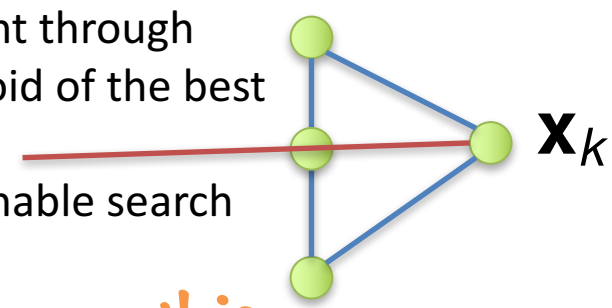
Use the "slope" of the simplex to find a good direction

$$f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \ldots \leq f(\mathbf{x}_{n+1})$$



The line from the worst point through the centroid of the best

is a reasonable search direction!



Why can't we just use this direction and be done?

Nelder-Mead

Because we need a simplex at the next step too!

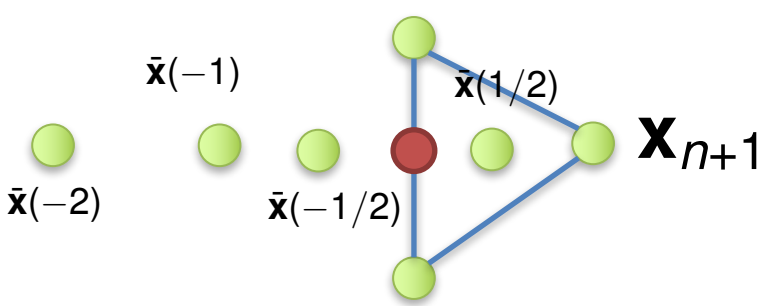
Can’t be too big.

Can’t be too small.

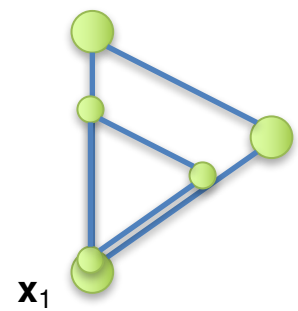
Solution 5 – Nelder-Mead

Use the slope of the simplex to find a good direction.

Reflecting the *worst point* in the simplex around the centroid of what's left to find a better point



Or shrink to the best point

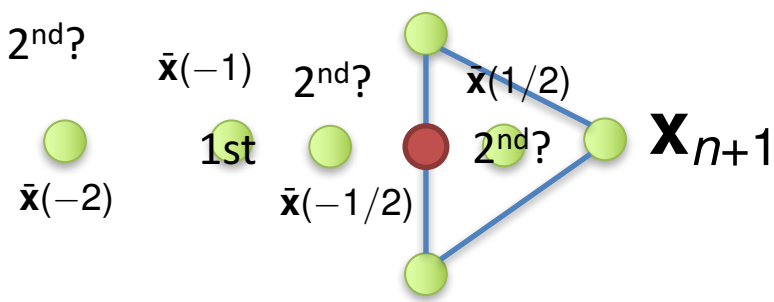


$$\bar{\mathbf{x}}(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i + t\left(\mathbf{x}_{n+1} - \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\right)$$

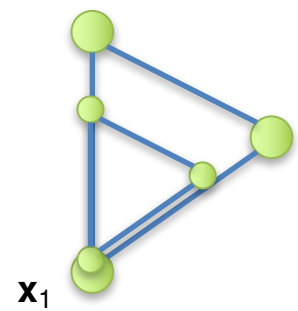
Solution 5 – Nelder-Mead

Use the slope of the simplex to find a good direction.

Reflecting the *worst point* in the simplex around the centroid of what's left to find a better point



Or shrink to the best point if no point is better



$$\bar{\mathbf{x}}(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i + t\left(\mathbf{x}_{n+1} - \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\right)$$

Quiz

Why is this better than pattern-search?