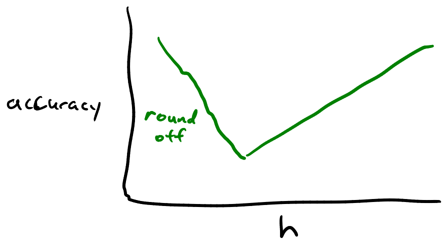
Derivative - Free - Optimization

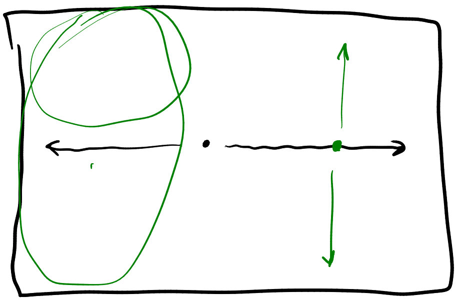
Zeroth-Order - f
First Order - $f, \vec{g}$
Second Order - $f, \vec{g}, H$



Options if you don't have $\vec{g}$

1. Auto-diff for $\vec{g}$ (and H)
2. Finite diff for $\vec{g}$ (usually not H) ← fmincon
   - Problems: 1. hard to choose step size h
     2. noise ←
     3. discontinuities
3. Derivative Free

Direct	Stochastic	Population
pattern	randomly sampling from distribution, improving distribution	Evolve a population of points



Line Search

Cyclic Coordinate Search

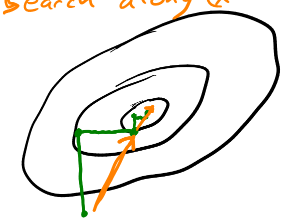
n-dimensions

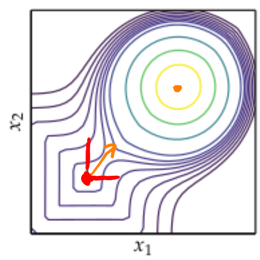
"full iteration" — full iteration produces a small change

initialize $\vec{x}^0$
loop until convergence
  for $i \in 1..n$
    $\alpha^* \leftarrow \underset{\alpha \in (-\infty,\infty)}{\text{argmax}} f(\vec{x}^k + \alpha \vec{e}^i)$ — approx line search; $\vec{e}^i$: cardinal unit vector
    $\vec{x}^{k+1} \leftarrow \vec{x}^k + \alpha^* \vec{e}^i$
    k++

$$\vec{e}^2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

"accelerated" — search along $(\vec{x}^k - \vec{x}^{k-n})$





Powell's method

initialize $\vec{x}^0$, $\vec{u}^1 \dots \vec{u}^n \leftarrow \vec{e}^1 \dots \vec{e}^n$

loop until convergence

  for $i \in 1..n$

    $\alpha^* \leftarrow \underset{\alpha \in (-\infty, \infty)}{\operatorname{argmin}} f(\vec{x}^k + \alpha \vec{u}^i)$

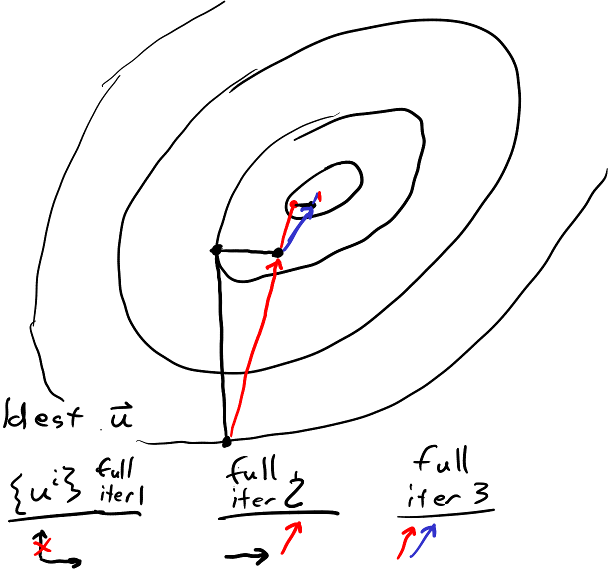
    $\vec{x}^{k+1} \leftarrow \vec{x}^k + \alpha^* \vec{u}^i$

    k++

  for $i \in 1..n-1$

    $\vec{u}^i \leftarrow \vec{u}^{i+1}$ // drop oldest $\vec{u}$

  $\vec{u}^n \leftarrow \vec{x}^k - \vec{x}^{k-n}$

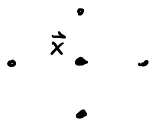


- After k full iterations, last k directions are mutually conjugate
- will minimize quadratic in $n(n+1)$ line searches

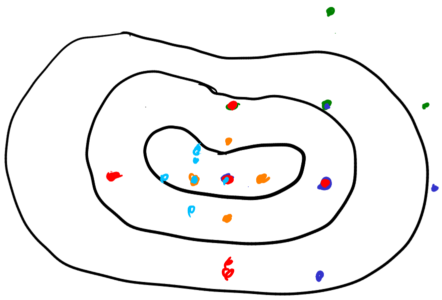
Pattern Search

Hooke-Jeeves

Pattern: $\vec{x} \pm \alpha \vec{e}^i$



If no improvement, decrease α



$2n$ evaluations at every step

Generalized Pattern Search

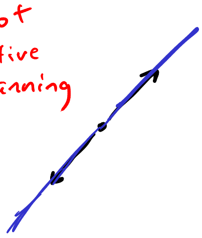
Pattern: $\vec{x} + \alpha \vec{d}^i$ $\{\vec{d}^i\}$ is a positive spanning set

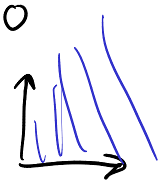
A set of vectors is a positive spanning set if any point in $\mathbb{R}^n$ can be constructed with a non-negative linear combination of the vectors.

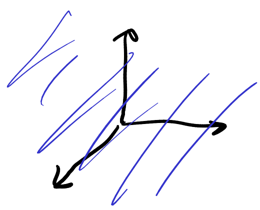
$$\vec{x} = \alpha_1 \vec{d}^1 + \alpha_2 \vec{d}^2 + \alpha_3 \vec{d}^3 \dots \qquad n+1$$

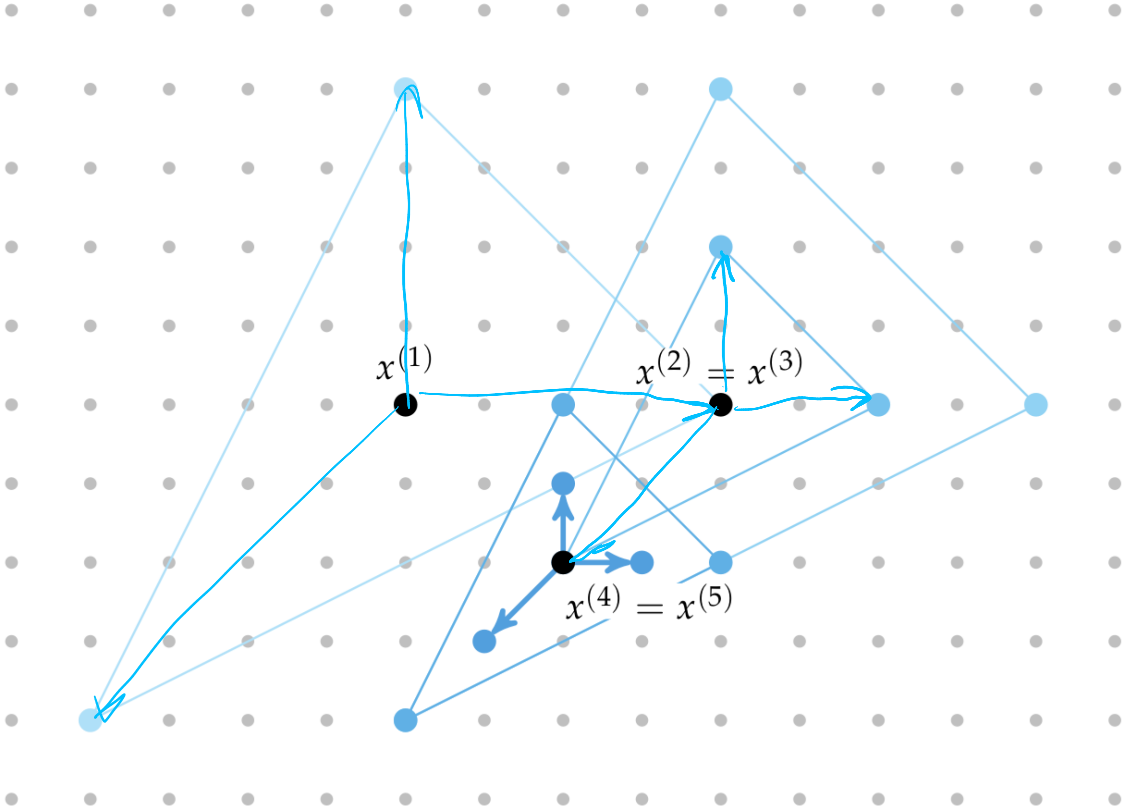
$$\alpha_i \geq 0$$

Not positive spanning

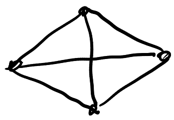








* Nelder – Mead (Simplex) (not related to Simplex for LP)

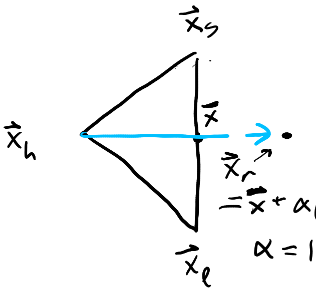
2D  3D  n+1 points

$\vec{x}_h$: highest (worst)
$\vec{x}_s$: second highest
$\vec{x}_l$: lowest

$\bar{x}$ average of all points except $\vec{x}^h$

Four operations

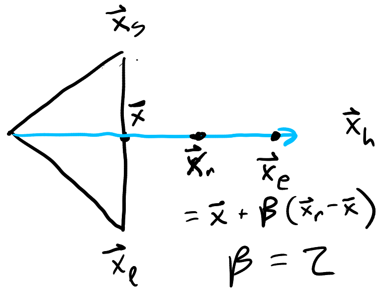
Reflection



$\vec{x}_r = \bar{x} + \alpha(\bar{x} - \vec{x}_h)$

$\alpha = 1$

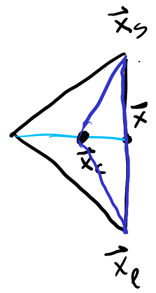
Expansion



$= \bar{x} + \beta(\vec{x}_r - \bar{x})$

$\beta = 2$

Contraction

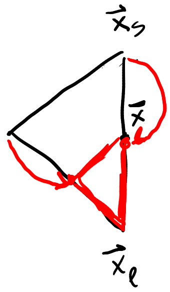


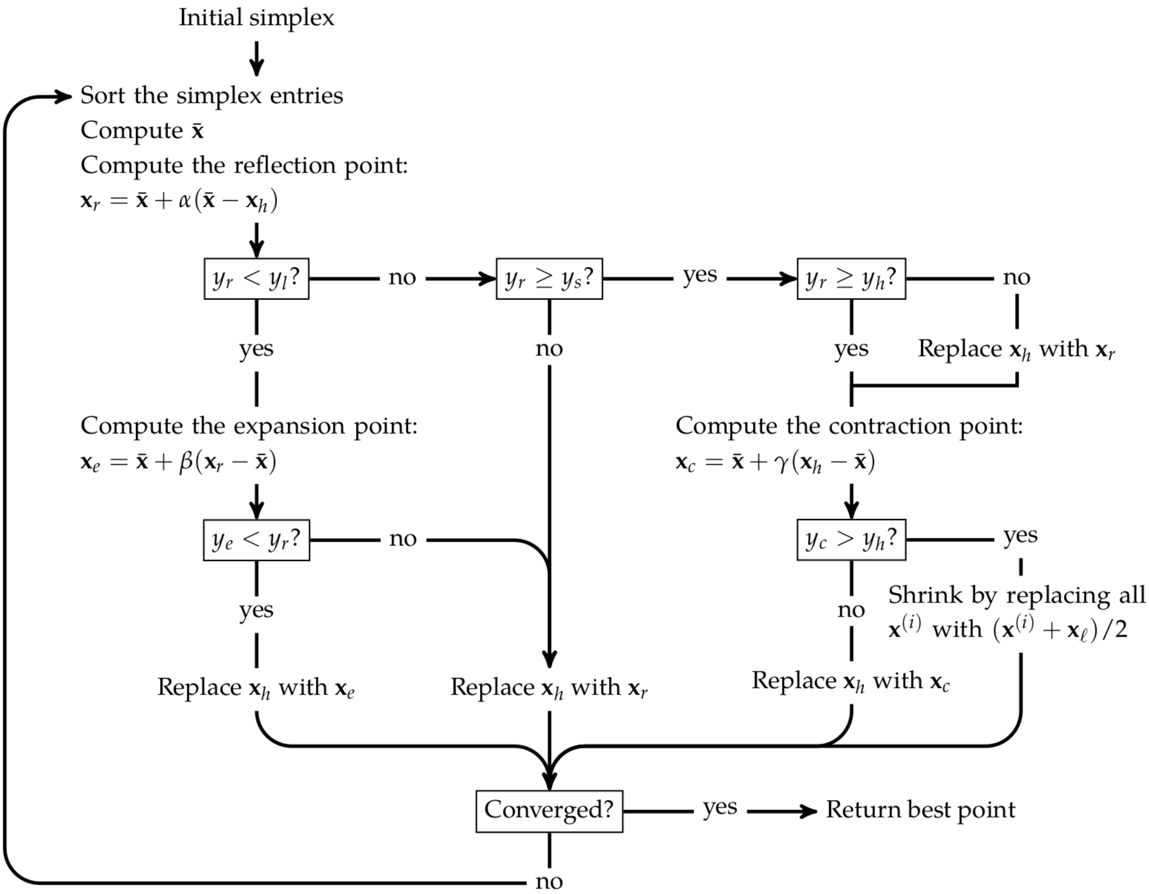
$\vec{x}_h$

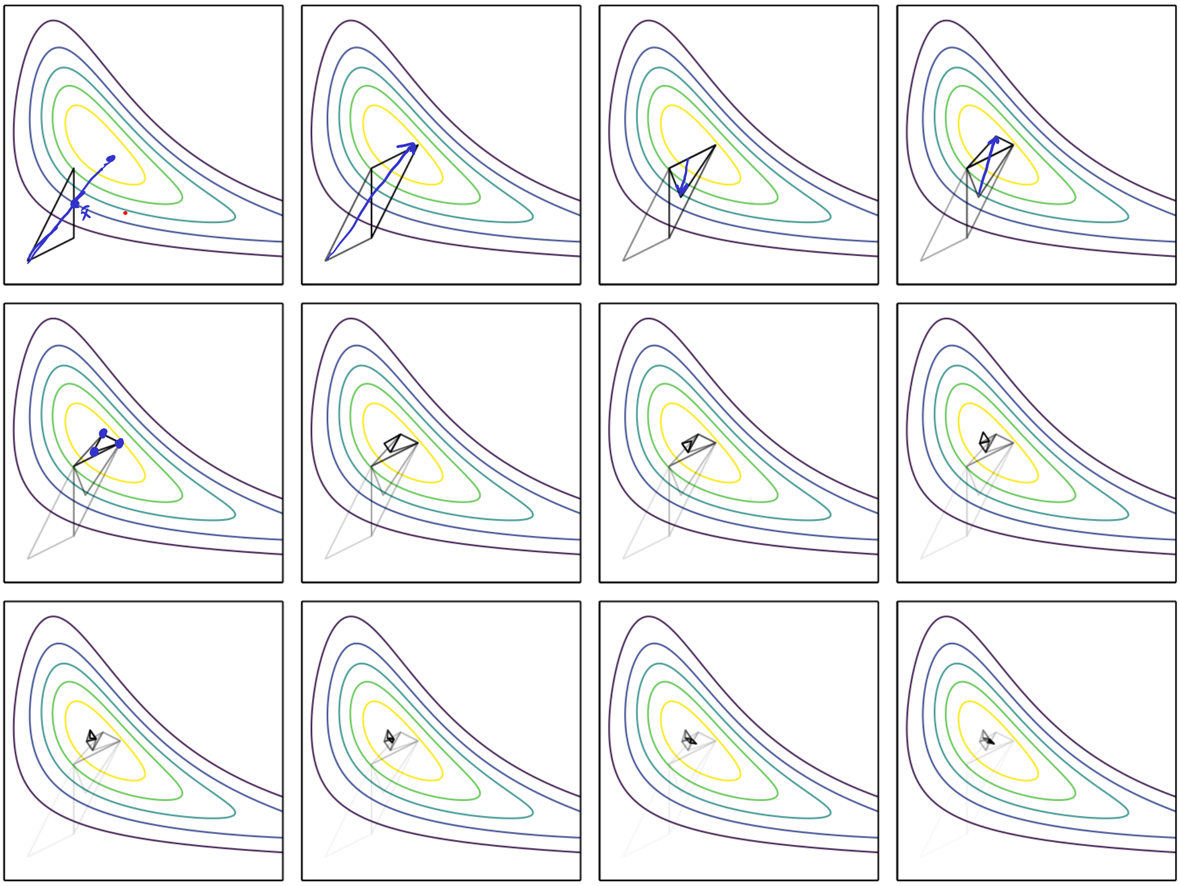
$\vec{x}_c = \bar{x} + \gamma(\vec{x}_h - \bar{x})$

$\gamma = 0.5$

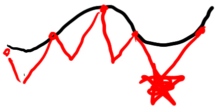
Shrinkage

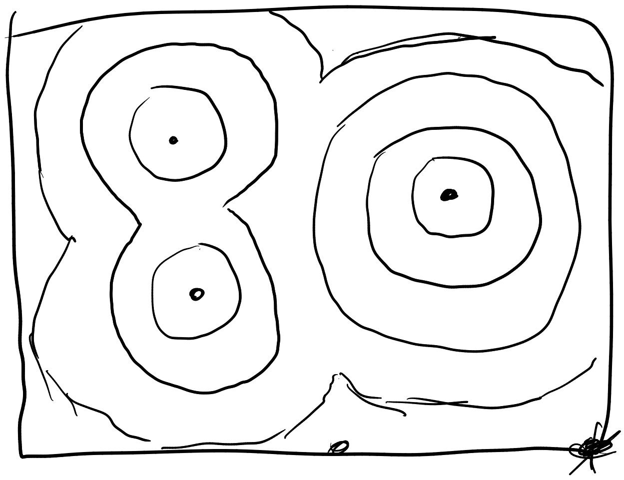



Initial simplex
Sort the simplex entries
Compute $\bar{\mathbf{x}}$
Compute the reflection point:
$\mathbf{x}_r = \bar{\mathbf{x}} + \alpha(\bar{\mathbf{x}} - \mathbf{x}_h)$
$y_r < y_l$?
no
$y_r \geq y_s$?
yes
$y_r \geq y_h$?
no
yes
no
yes
Replace $\mathbf{x}_h$ with $\mathbf{x}_r$
Compute the expansion point:
$\mathbf{x}_e = \bar{\mathbf{x}} + \beta(\mathbf{x}_r - \bar{\mathbf{x}})$
Compute the contraction point:
$\mathbf{x}_c = \bar{\mathbf{x}} + \gamma(\mathbf{x}_h - \bar{\mathbf{x}})$
$y_e < y_r$?
no
$y_c > y_h$?
yes
yes
Shrink by replacing all
$\mathbf{x}^{(i)}$ with $(\mathbf{x}^{(i)} + \mathbf{x}_\ell)/2$
no
Replace $\mathbf{x}_h$ with $\mathbf{x}_e$
Replace $\mathbf{x}_h$ with $\mathbf{x}_r$
Replace $\mathbf{x}_h$ with $\mathbf{x}_c$
Converged?
yes
Return best point
no



Schubert-Piyavskii for multiple dimensions

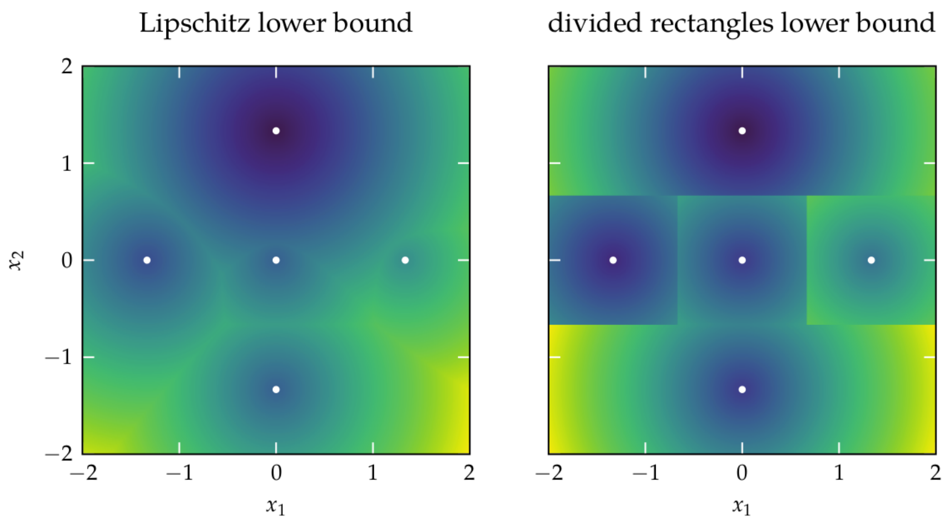




Lipschitz

ℓ

DIRECT (DIvided RECTangles)



(Does not assume a known ℓ)

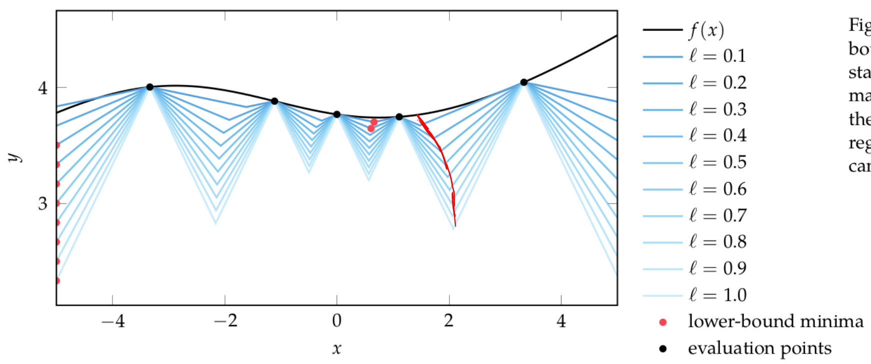


Figure 7.13. The Lipschitz lower bound for different Lipschitz constants ℓ. Not only does the estimated minimum change locally as the Lipschitz constant is varied, the region in which the minimum lies can vary as well.

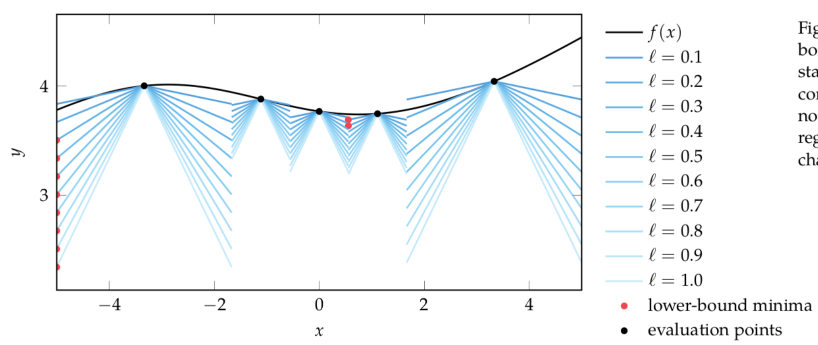
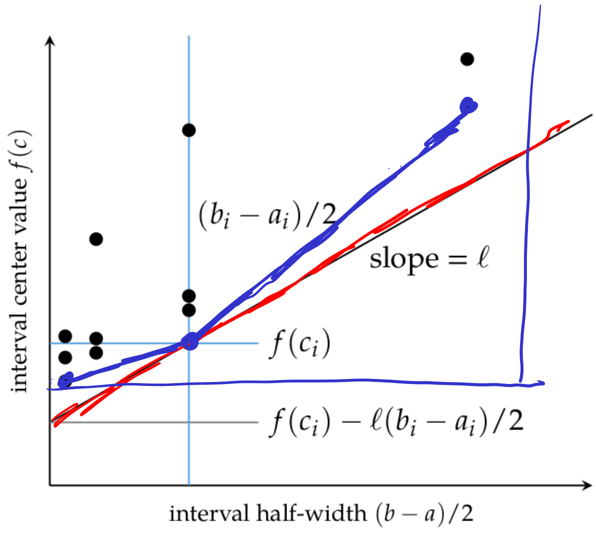
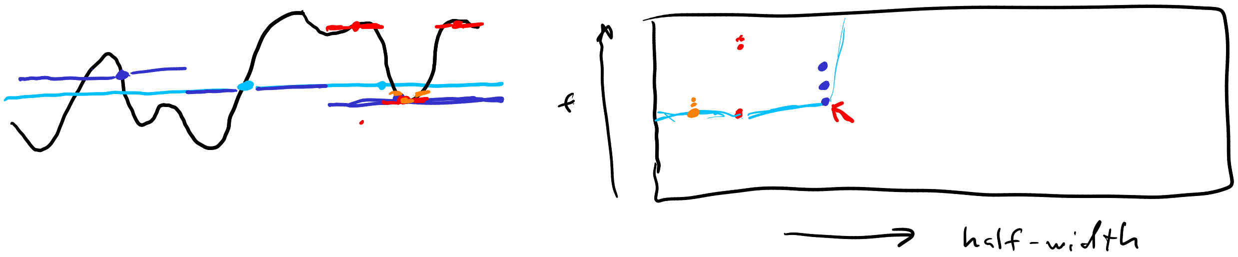
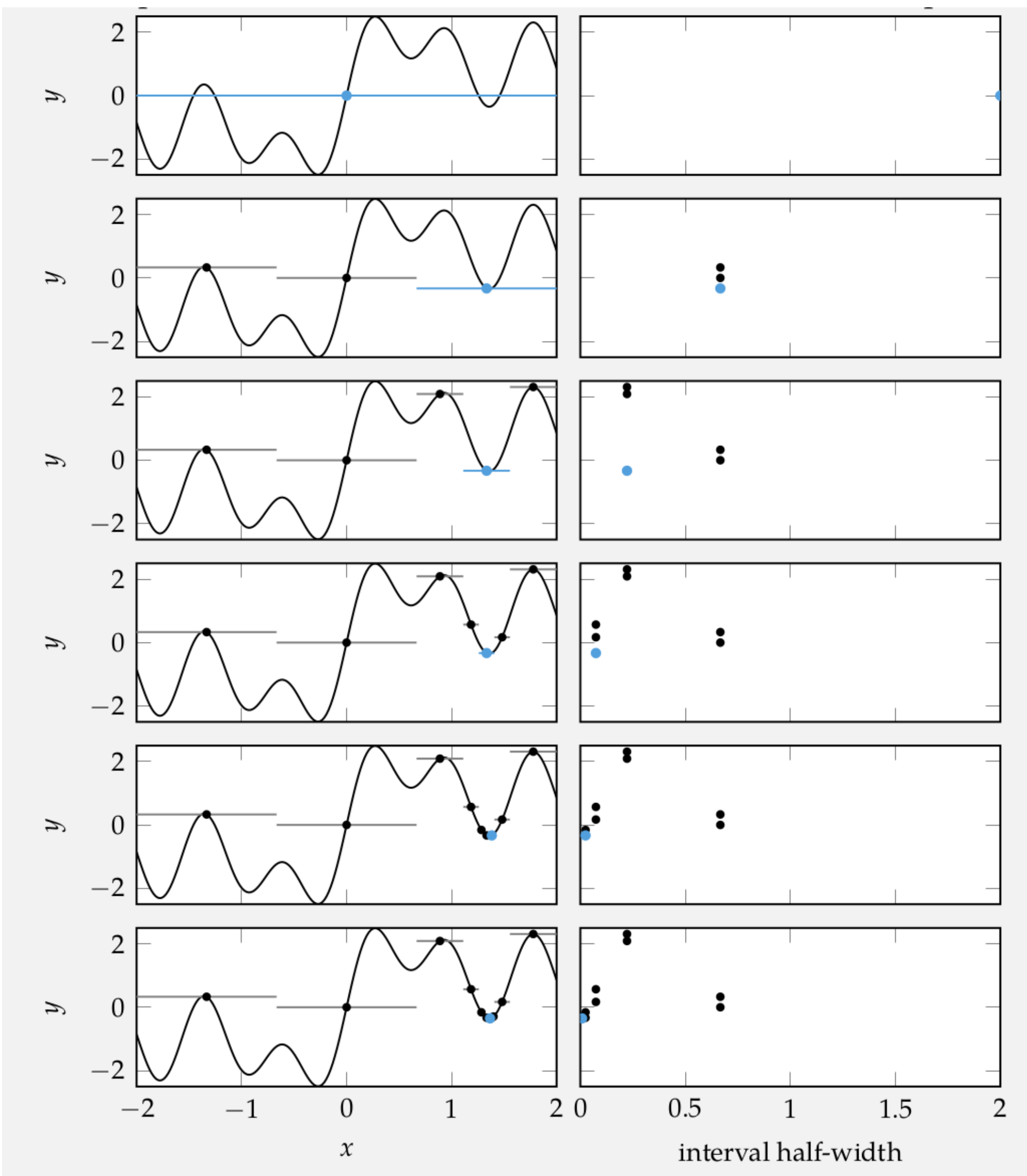
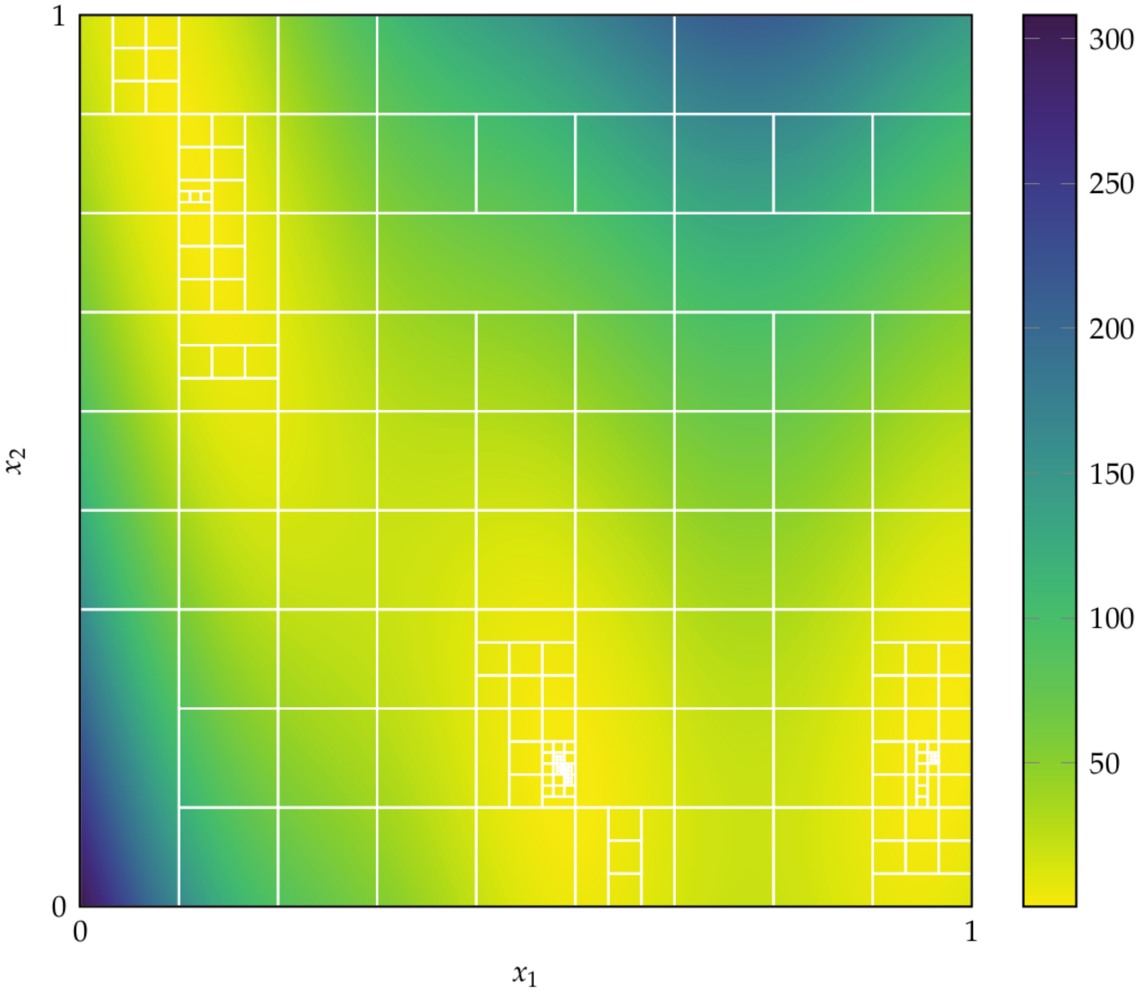


Figure 7.14. The DIRECT lower bound for different Lipschitz constants ℓ. The lower bound is not continuous. The minimum does not change locally but can change regionally as the Lipschitz constant changes.


interval center value $f(c)$
$(b_i - a_i)/2$
slope = ℓ
$f(c_i)$
$f(c_i) - \ell(b_i - a_i)/2$
interval half-width $(b - a)/2$


f
half-width


y
x
interval half-width


1
0
0
1
x_1
x_2
300
250
200
150
100
50