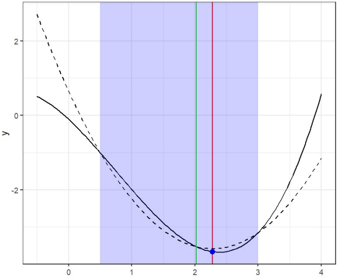
Optimization in Machine Learning

Univariate optimization: Brent's method



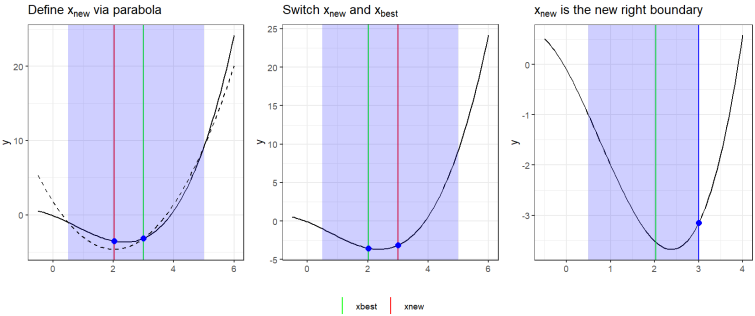
Learning goals

- Quadratic interpolation
- Brent's procedure

QUADRATIC INTERPOLATION

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

$$(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$$

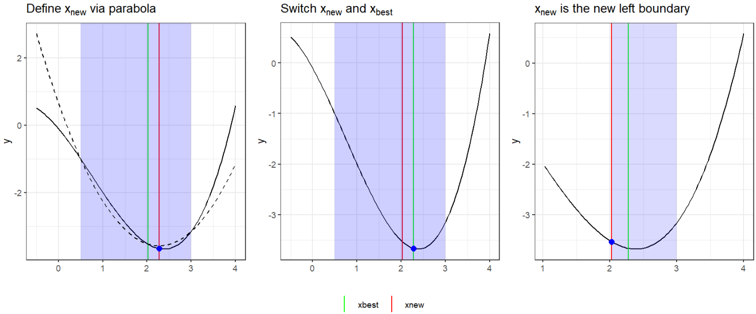


Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x^{best}. Right: New interval.

QUADRATIC INTERPOLATION

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

$$(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$$

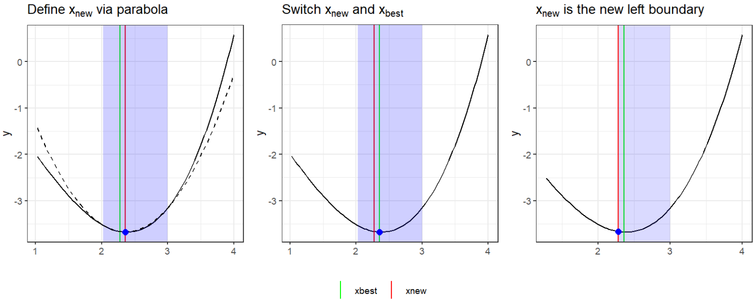


Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x^{best}. Right: New interval.

QUADRATIC INTERPOLATION

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

$$(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$$

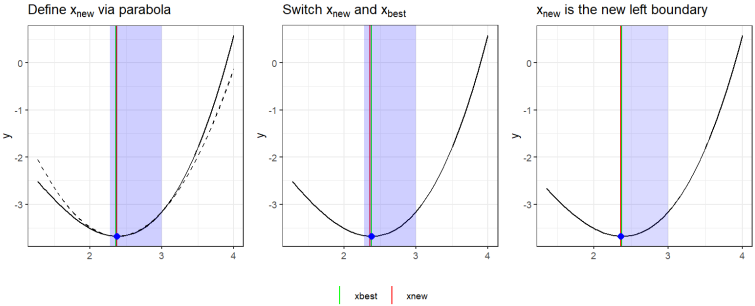


Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x^{best}. Right: New interval.

QUADRATIC INTERPOLATION

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

$$(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$$

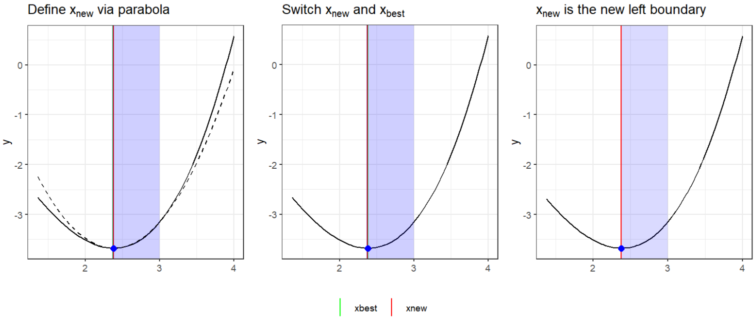


Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x^{best}. Right: New interval.

QUADRATIC INTERPOLATION

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

$$(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$$



Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x^{best}. Right: New interval.

QUADRATIC INTERPOLATION COMMENTS

- Quadratic interpolation **not robust**. The following may happen:
  - Algorithm suggests the same x^{new} in each step,
  - x^{new} outside of search interval,
  - Parabola degenerates to line and no real minimum exists
- Algorithm must then abort, finding a global minimum is not guaranteed.

BRENT'S METHOD

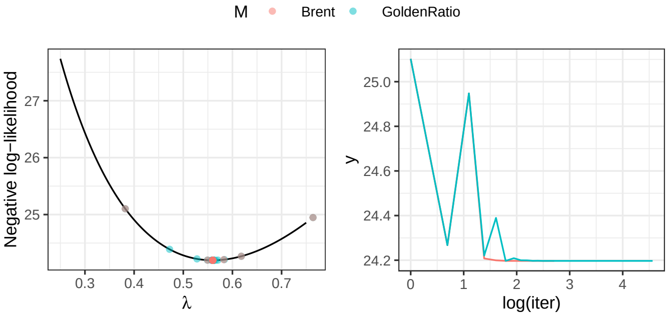
- Brent proposed an algorithm (1973) that alternates between golden ratio search and quadratic interpolation as follows:
  - Quadratic interpolation step acceptable if: (i) x^{new} falls within $[x^{\text{left}}, x^{\text{right}}]$ (ii) x^{new} sufficiently far away from x^{best}

    (Heuristic: Less than half of movement of step before last)
  - Otherwise: Proposal via golden ratio
- Benefit: Fast convergence (quadratic interpolation), unstable steps (e.g. parabola degenerated) stabilized by golden ratio search
- Convergence guaranteed if the function f has a local minimum
- Used in R-function `optimize()`

EXAMPLE: ML CAUCHY

- Density Poisson: $f(k \mid \lambda) := \mathbb{P}(x = k) = \frac{\lambda^k \cdot \exp(-\lambda)}{k!}$
- Negative log-Likelihood for n observations:

$$-\ell(\lambda, \mathcal{D}) \quad = \quad -\log L(\lambda, \mathcal{D}) = -\log \prod_{i=1}^{n} f\left(x^{(i)} \mid \lambda\right) = -\sum_{i=1}^{n} \frac{\lambda^{x^{(i)}} \cdot \exp(-\lambda)}{x^{(i)}!}$$



GR and Brent converge to global opt at $x^* = 0.75$.
But GR needs $\approx$ 93 iters, Brent needs $\approx$ 18 for same level of precision.