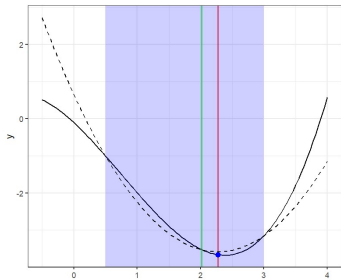


# Optimization in Machine Learning

## Univariate optimization

### Brent's method



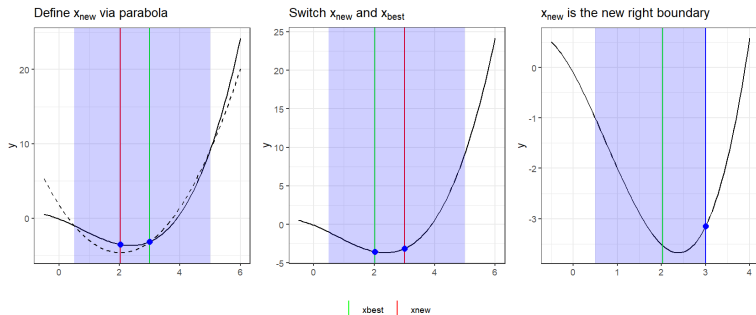
#### Learning goals

- Quadratic interpolation
- Brent's procedure

# QUADRATIC INTERPOLATION

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

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



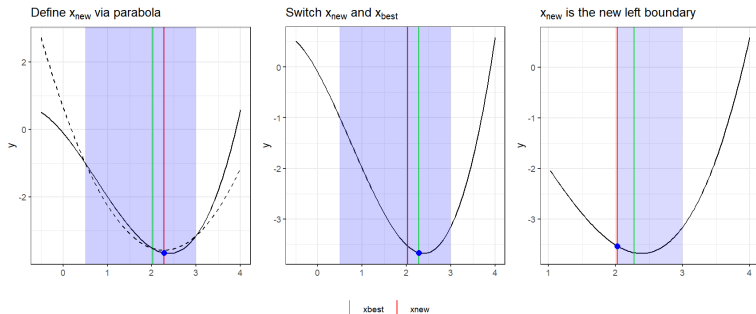
Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with  $x^{\text{best}}$ . Right: New interval.



# QUADRATIC INTERPOLATION

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

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



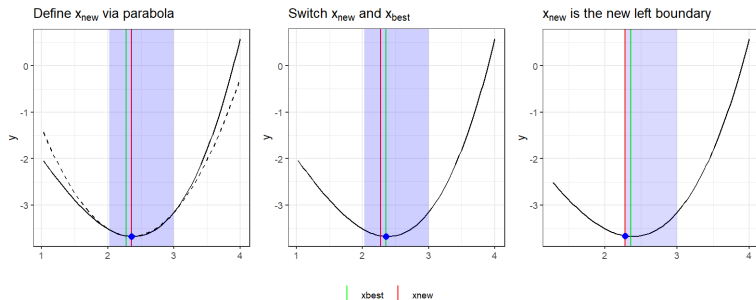
Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with  $x^{\text{best}}$ . Right: New interval.



# QUADRATIC INTERPOLATION

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

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



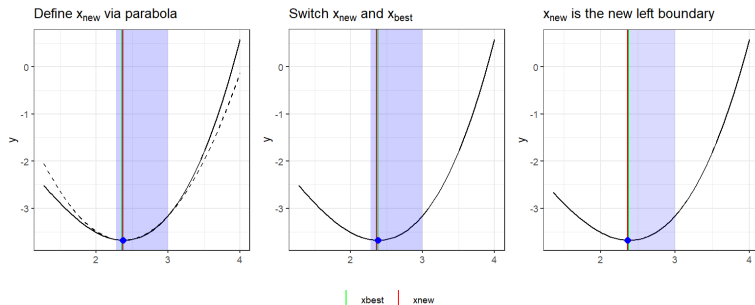
Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with  $x^{\text{best}}$ . Right: New interval.



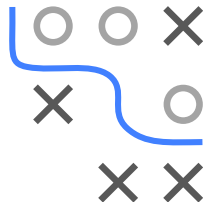
# QUADRATIC INTERPOLATION

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

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



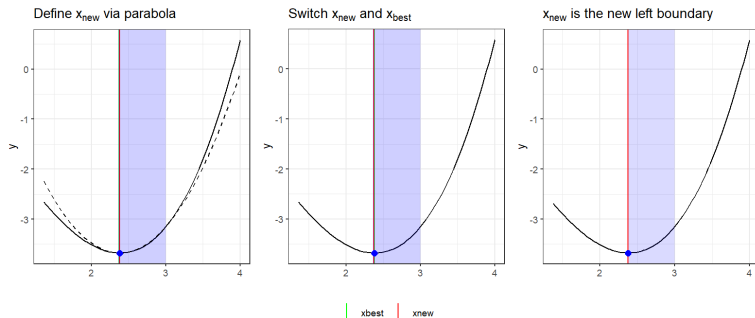
Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with  $x^{\text{best}}$ . Right: New interval.



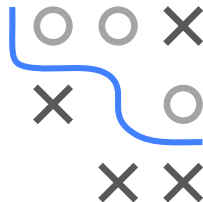
# QUADRATIC INTERPOLATION

Similar to golden ratio procedure but select  $x^{\text{new}}$  differently:  $x^{\text{new}}$  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  $x^{\text{best}}$ . Right: New interval.



# QUADRATIC INTERPOLATION COMMENTS

- Quadratic interpolation **not robust**. The following may happen:
  - Algorithm suggests the same  $x^{\text{new}}$  in each step,
  - $x^{\text{new}}$  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)  $x^{\text{new}}$  falls within  $[x^{\text{left}}, x^{\text{right}}]$  (ii)  $x^{\text{new}}$  sufficiently far away from  $x^{\text{best}}$   
(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()`

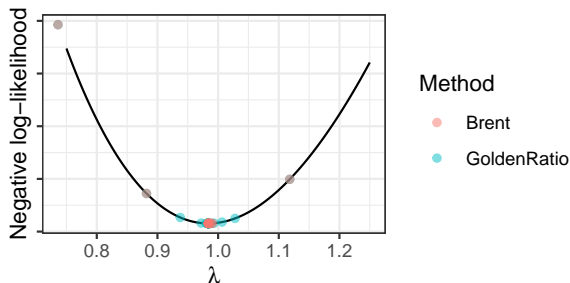




# EXAMPLE: MLE POISSON

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

$$-\ell(\lambda, \mathcal{D}) = -\log \prod_{i=1}^n f(x^{(i)} | \lambda) = -\sum_{i=1}^n \log f(x^{(i)} | \lambda)$$



GR and Brent converge to minimum at  $x^* \approx 1$ .

**But:** GR needs  $\approx 45$  it., Brent only needs  $\approx 15$  it. for same tolerance.

