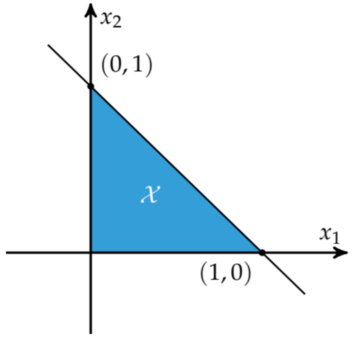
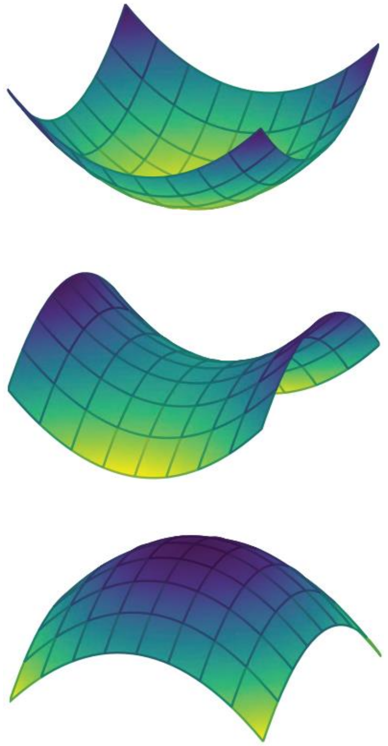
Tecnológico Nacional de México

Instituto Tecnológico de Orizaba/Celaya

Modern Computing Algorithms for Process Optimization with Julia Programming



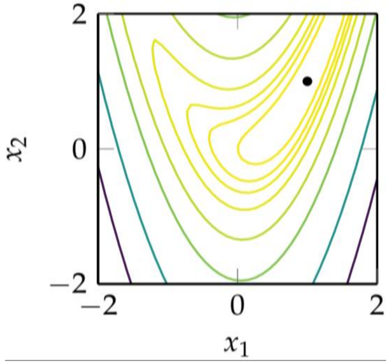
"6 – Second-Order Methods"



By:

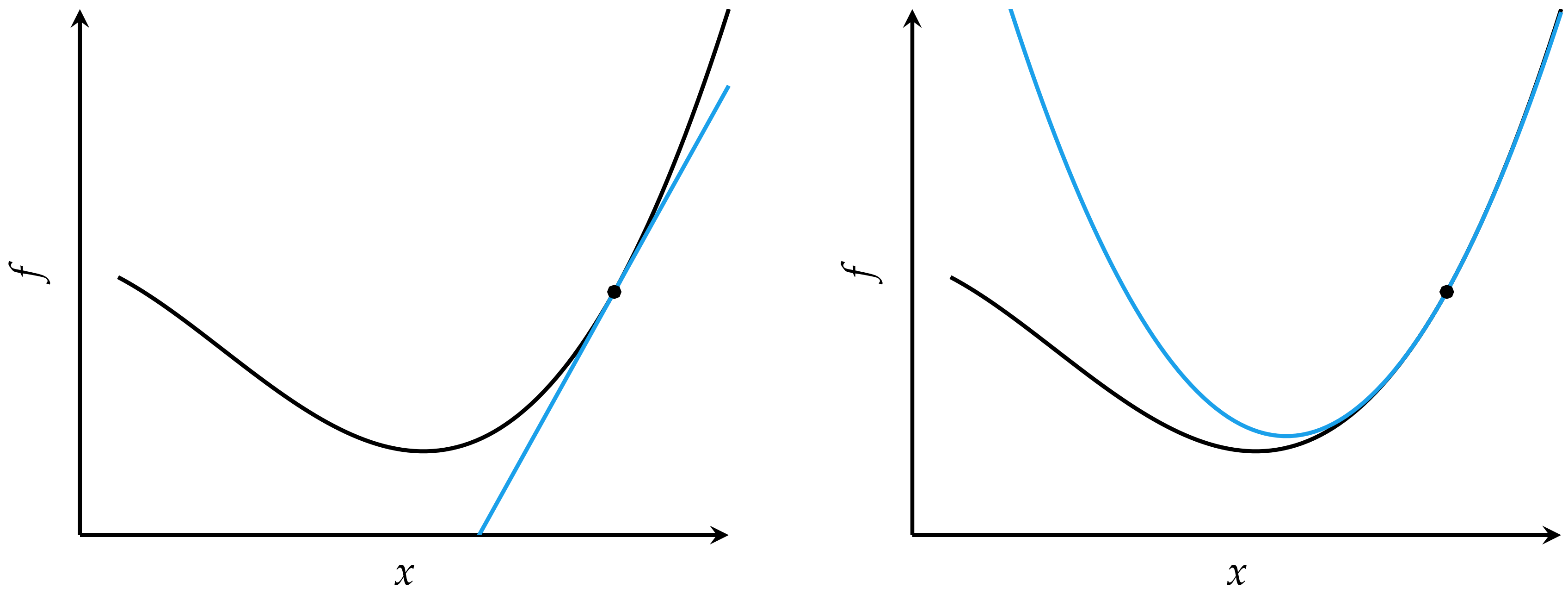
Dr. Kelvyn Baruc Sánchez Sánchez

Postdoctoral Researcher/I.T. Celaya



Second-Order Methods

- Locally approximate function as quadratic
- Comparison of first-order and second order approximations



Newton's Method

- Approximate a function using second-order Taylor series expansion
- Univariate function

$$q(x) = f(x^{(k)}) + (x - x^{(k)})f'(x^{(k)}) + \frac{(x - x^{(k)})^2}{2} f''(x^{(k)})$$

$$\frac{\partial}{\partial x} q(x) = f'(x^{(k)}) + (x - x^{(k)}) f''(x^{(k)}) = 0$$

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

Newton's Method

- Multivariate function

$$f(\mathbf{x}) \approx q(\mathbf{x}) = f(\mathbf{x}^{(k)}) + (\mathbf{g}^{(k)})^\top (\mathbf{x} - \mathbf{x}^{(k)}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^{(k)})^\top \mathbf{H}^{(k)} (\mathbf{x} - \mathbf{x}^{(k)})$$

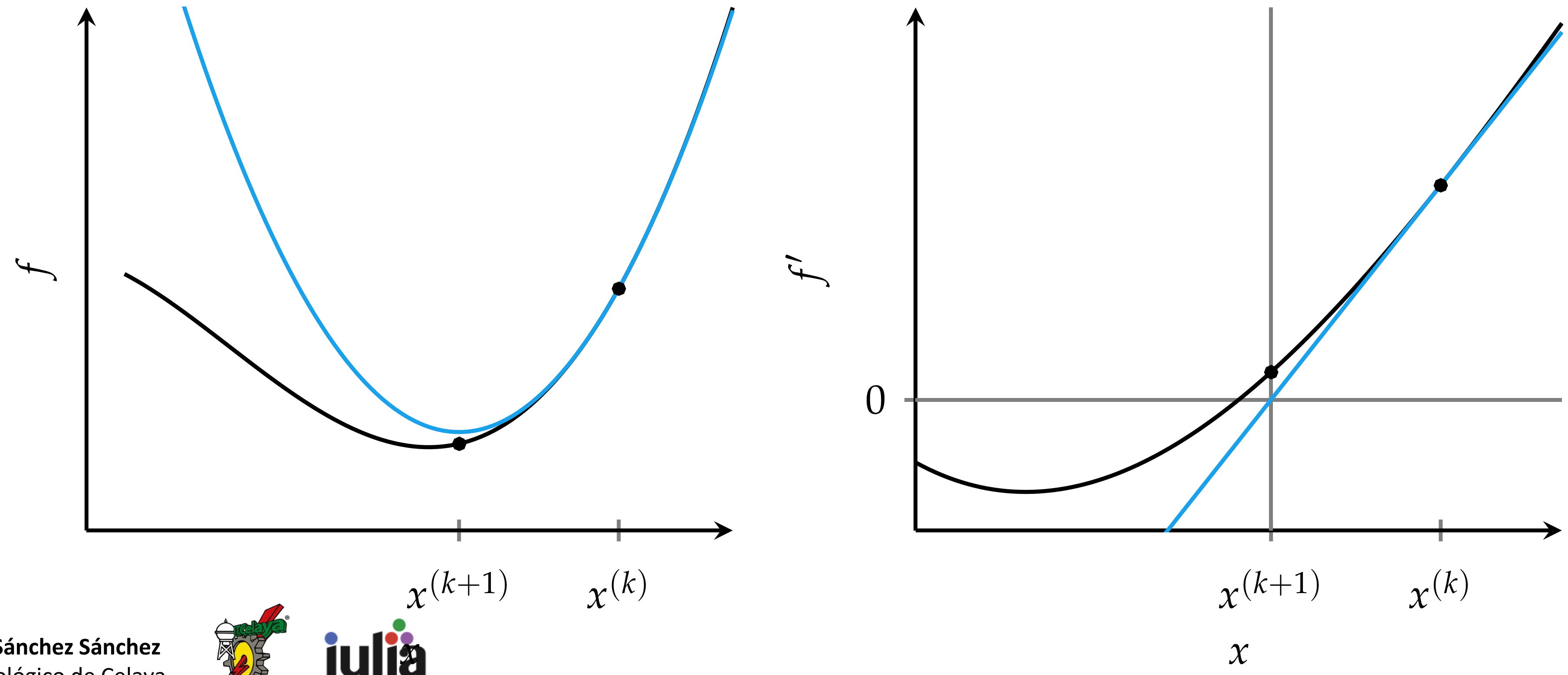
- Multivariate update rule

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{H}^{(k)})^{-1} \mathbf{g}^{(k)}$$

- **H** is the Hessian matrix

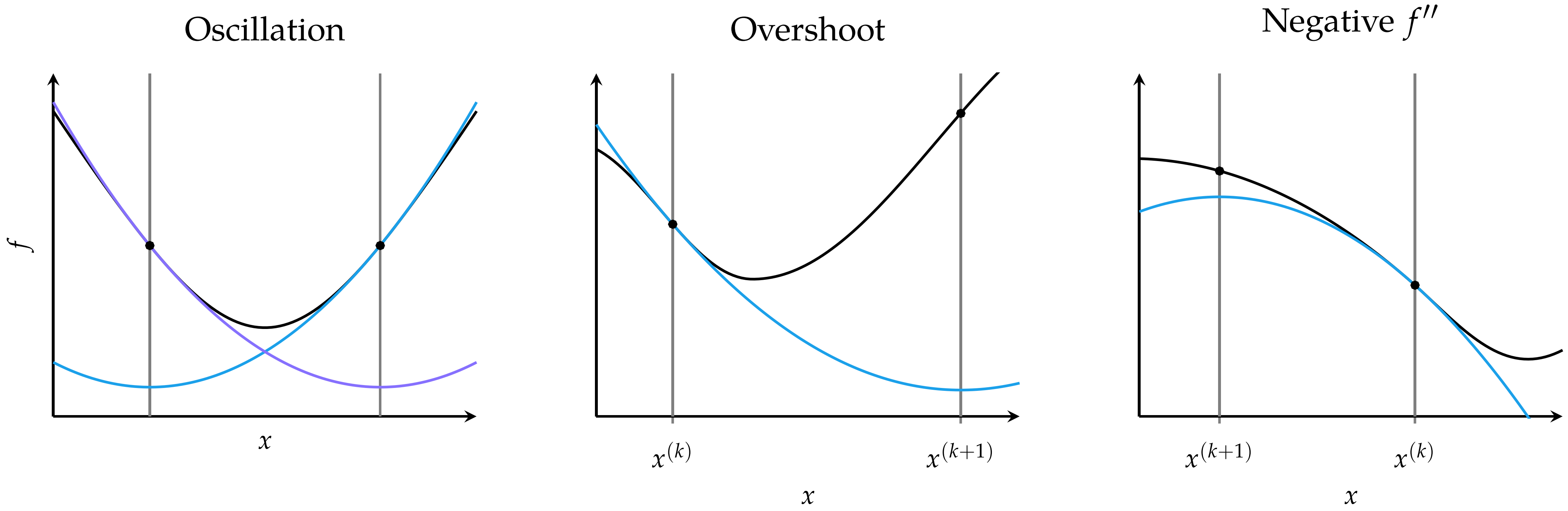
Newton's Method

- Newton's method for optimization is equivalent to finding the roots of the derivative function



Newton's Method

- Common causes of error in Newton's method

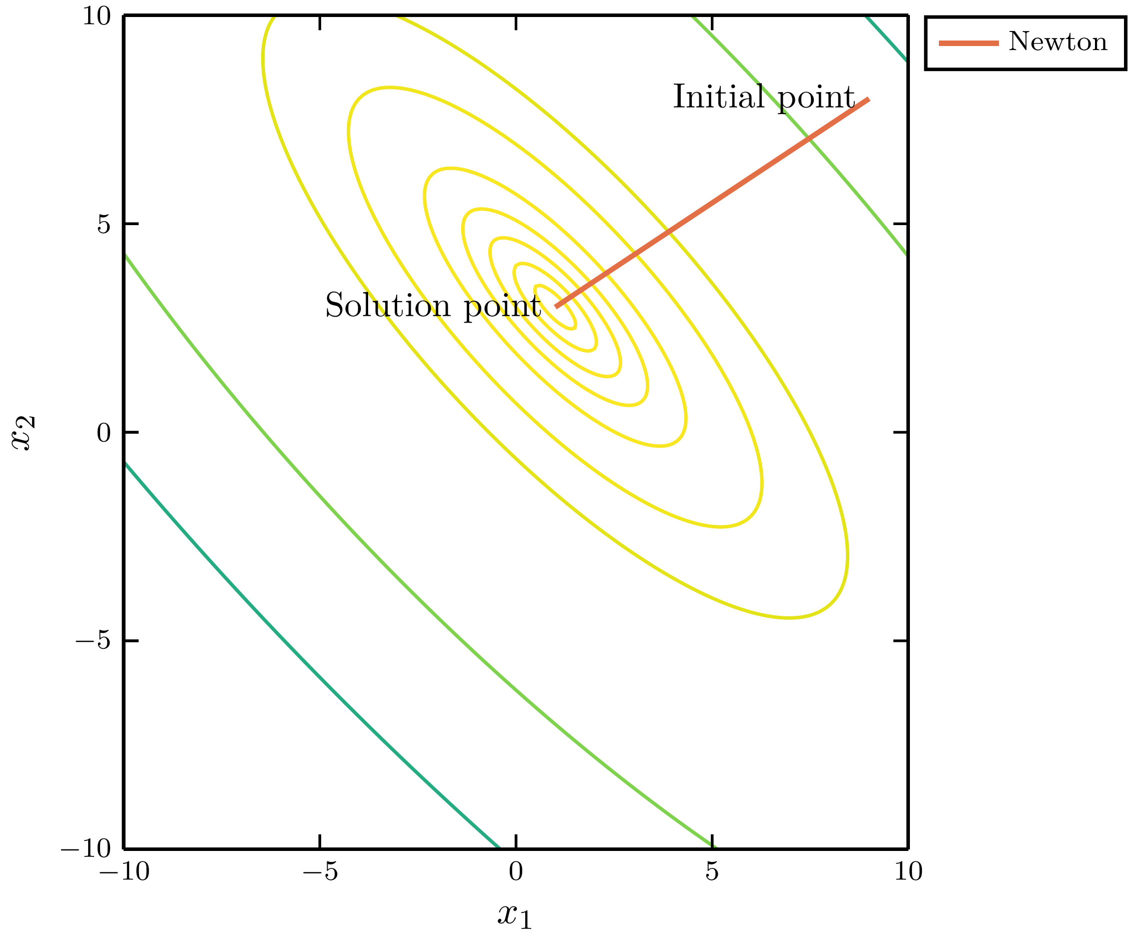


Dr. Kelvyn B. Sánchez Sánchez
Instituto Tecnológico de Celaya

Newton's Method

Booth's Function

$$f(\mathbf{x}) = (x_1 + 2x_2 - 7)^2 + (2x_1 + x_2 - 5)^2$$



Secant Method

- For univariate functions, if the second derivative is unknown, it can be approximated using the secant method

$$f''(x^{(k)}) \approx \frac{f'(x^{(k)}) - f'(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}$$

- Update equation

$$x^{(k+1)} \leftarrow x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f'(x^{(k)}) - f'(x^{(k-1)})} f'(x^{(k)})$$

Quasi-Newton Methods

- For multivariate functions, the inverse Hessian may not be available or infeasible to compute, so can be approximated using a variety of Quasi-Newton methods, each appropriate in different circumstances
- Davidon-Fletcher-Powell (DFP) method
- Broyden-Fletcher-Goldfarb-Shanno (BFGS) method
- Limited-memory BFGS (L-BFGS) method

Quasi-Newton Methods

- Quasi-Newton method updates have the form:

$$\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - \alpha^{(k)} \mathbf{Q}^{(k)} \mathbf{g}^{(k)}$$

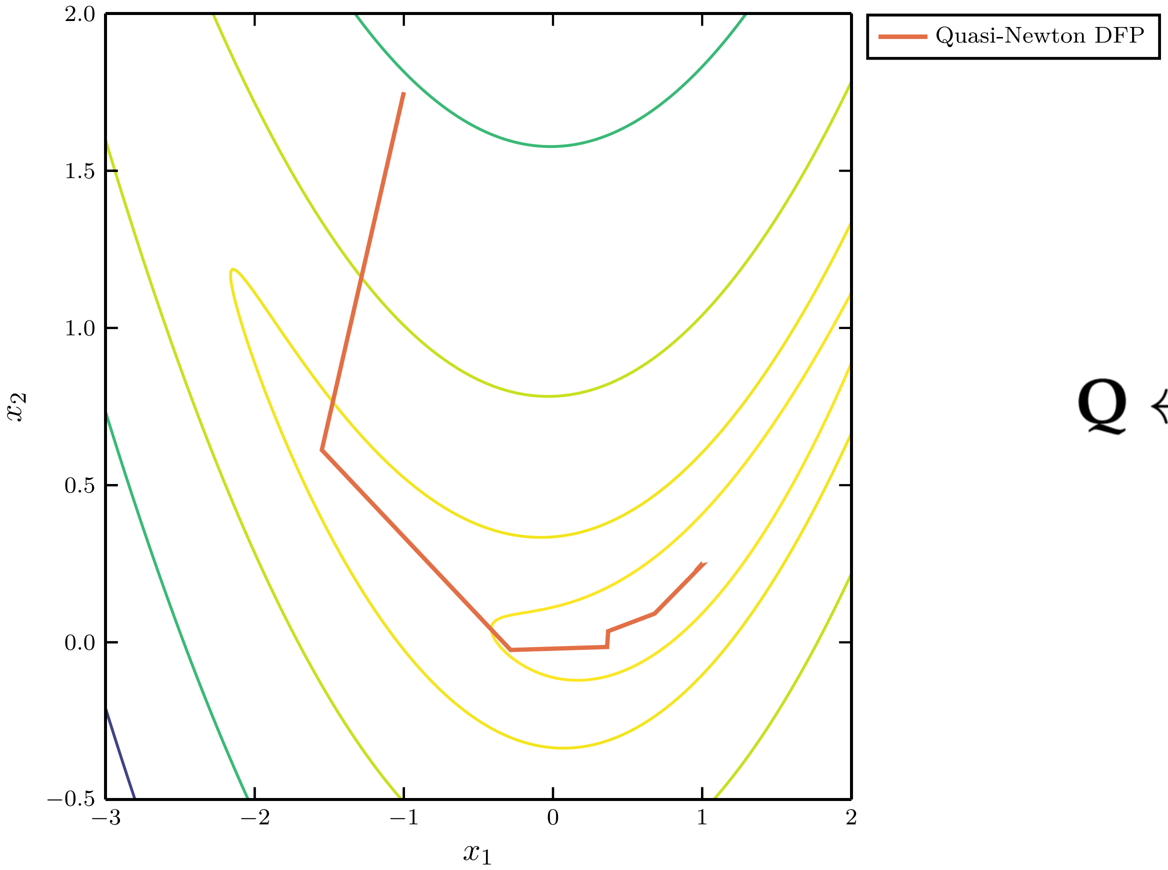
where $\alpha^{(k)}$ is a scalar step factor and $\mathbf{Q}^{(k)}$ approximates the inverse of the Hessian at $\mathbf{x}^{(k)}$.

These methods typically set $\mathbf{Q}^{(1)}$ to the identity matrix, and they then apply updates to reflect information learned with each iteration. To simplify the equations for the various quasi-Newton methods, we define the following:

$$\boldsymbol{\gamma}^{(k+1)} \equiv \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$$

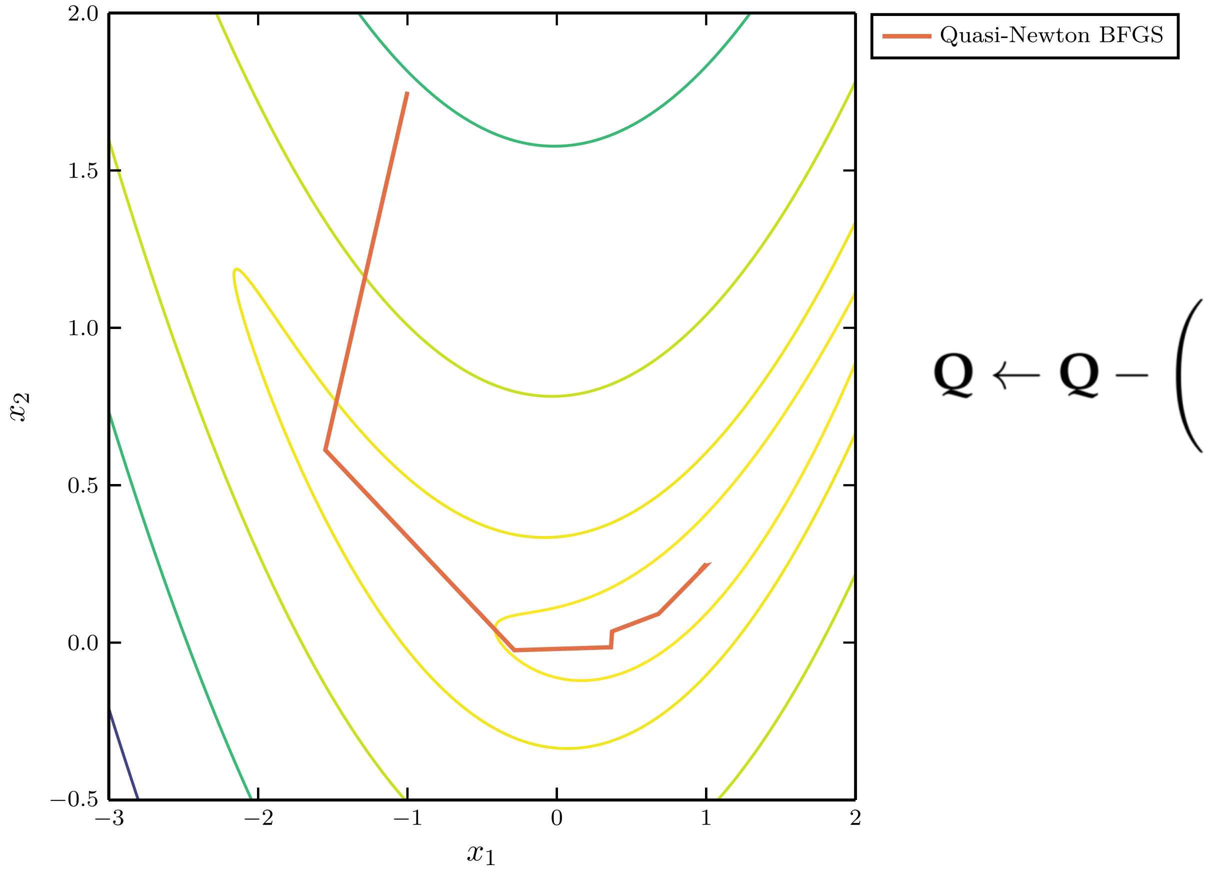
$$\boldsymbol{\delta}^{(k+1)} \equiv \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$$

Davidon-Fletcher-Powell (DFP) method



$$\mathbf{Q} \leftarrow \mathbf{Q} - \frac{\mathbf{Q}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top\mathbf{Q}}{\boldsymbol{\gamma}^\top\mathbf{Q}\boldsymbol{\gamma}} + \frac{\boldsymbol{\delta}\boldsymbol{\delta}^\top}{\boldsymbol{\delta}^\top\boldsymbol{\gamma}}$$

Broyden-Fletcher-Goldfarb-Shanno (BFGS) method



$$\mathbf{Q} \leftarrow \mathbf{Q} - \left(\frac{\boldsymbol{\delta}\boldsymbol{\gamma}^\top \mathbf{Q} + \mathbf{Q}\boldsymbol{\gamma}\boldsymbol{\delta}^\top}{\boldsymbol{\delta}^\top \boldsymbol{\gamma}}\right) + \left(1 + \frac{\boldsymbol{\gamma}^\top \mathbf{Q}\boldsymbol{\gamma}}{\boldsymbol{\delta}^\top \boldsymbol{\gamma}}\right)\frac{\boldsymbol{\delta}\boldsymbol{\delta}^\top}{\boldsymbol{\delta}^\top \boldsymbol{\gamma}}$$

Summary

- Incorporating second-order information in descent methods often speeds convergence.
- Newton’s method is a root-finding method that leverages second-order information to quickly descend to a local minimum.
- The secant method and quasi-Newton methods approximate Newton’s method when the second-order information is not directly available.