Geometry Optimization

Frequent problem class

Energy

- Find most stable molecular geometry: Compare conformers
- Find transition state geometries: Identify reaction pathways

Residuals

- Fitting experimental data: Model observations
- Potential fitting: Simplify calculations
- Machine learning: Surrogate models

Definitions

Solution coefficients x

- Molecular geometries
- Fitting coefficients
- Model coefficients

Scalar objective function f

- Energy
- Residual norm
- Here: smooth, i.e. differentiable function

$$f(x_1, x_2, \ldots, x_n) = f(\mathsf{x}) = y \tag{18}$$

Domain X

- Valid parameter range
- Any solution within accepted

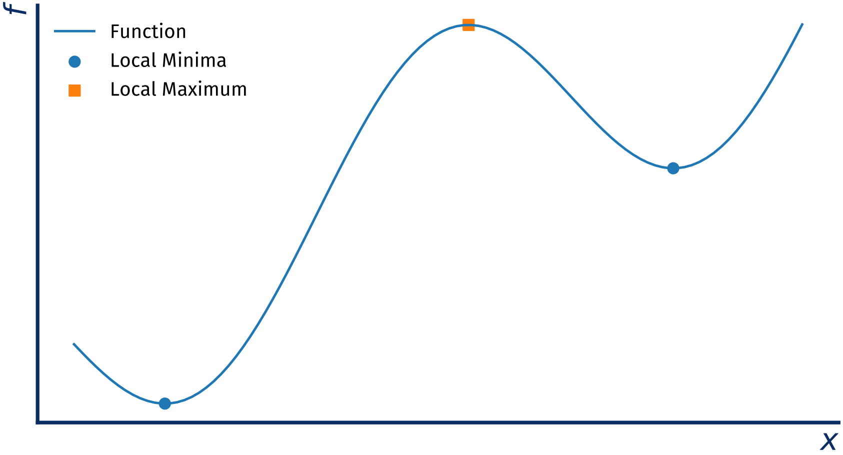
Target x_0

- Maximise or minimise y (over domain)

$$\begin{aligned} \mathsf{x}_0 &\equiv \arg\min_{\mathsf{x} \in X} f(\mathsf{x}) \\ &= \{\mathsf{x} \mid \mathsf{x}, \mathsf{y} \in X : f(\mathsf{x}) \leq f(\mathsf{y})\} \end{aligned} \tag{19}$$

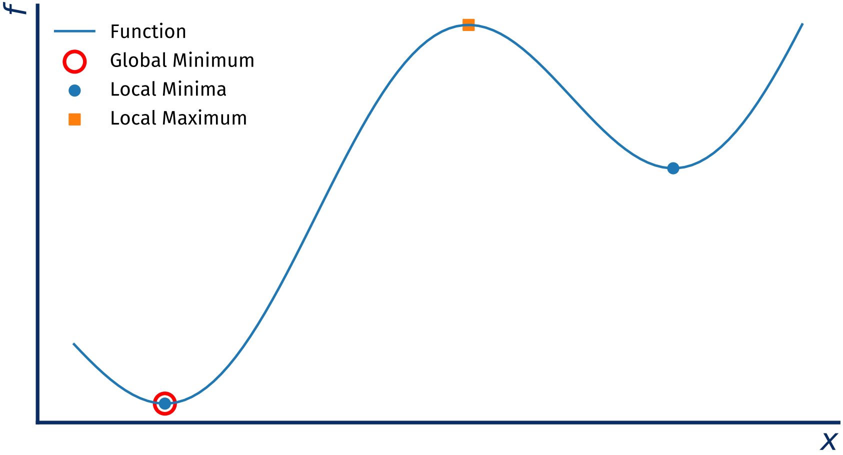
Local Minimum

$$\exists \epsilon > 0 \; : \; \forall y \in [x_0 - \epsilon,\, x_0 + \epsilon] \; : \; f(x_0) \leq f(y) \tag{20}$$



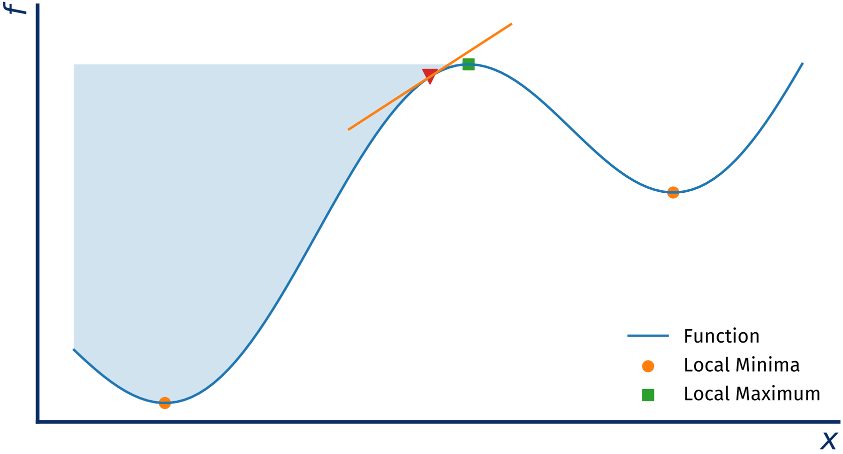
Global Minimum

$$\forall y \in X : \ f(x_0) \leq f(y) \tag{21}$$



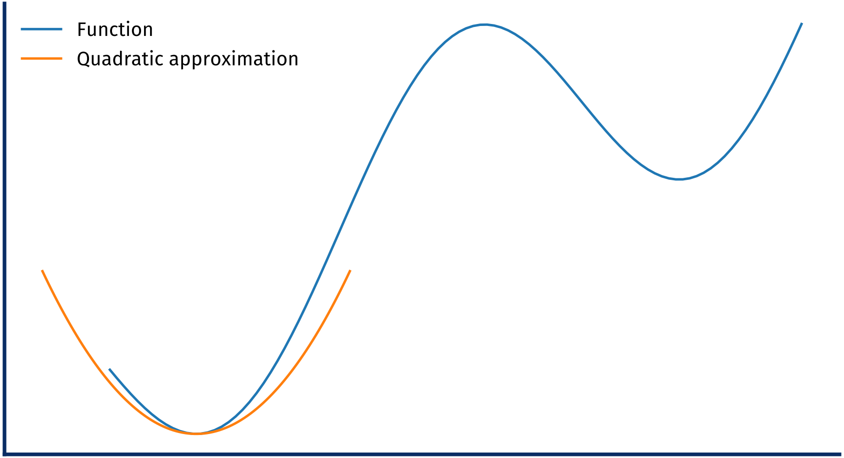
Attractive Basin

All values that if the gradient is followed reach a given minimum.

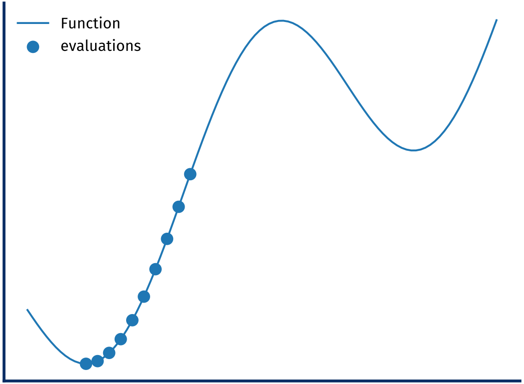


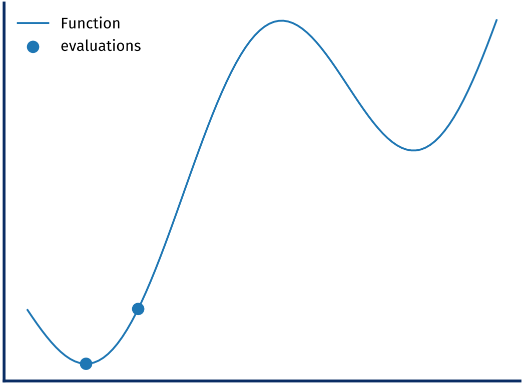
Quadratic Region

All values where a Taylor series expansion up to second order around the minimum is a *good* approximation.



Iterative vs Direct





Iterative

- Edging closer to the minimum
- Continue until close enough

Direct

- One-step optimization
- Analytical expression, no gradients

Optimisation Strategies

Pure strategies

- Follow gradient and/or Hessian: (Quasi-)Newton methods
  - Transition states: eigenmode following
- Reduce dimensionality: Subspace methods
- (Quasi-)randomly pick points: Stochastic optimisation
- Regularly pick points: Grid refinement

Hybrid

- Problem specific
- Typically global optimization
- E.g. stochastic first, then Newton

Series notation for iterative approaches

$$\{a_n\} \quad ; \quad \lim_{n\to\infty} = x_0 \tag{22}$$

Coordinates Matter

Cartesian coordinates inefficient

- Convergence rate slow
- Physical intuition: collective motions
- Tiny steps
- Numerical instability

Issues

- Separability of internal motions
- Condition number of Hessian matrix (ratio of smallest and largest eigenvalue)
- Handling of constraints and symmetries (typical physical constraints)

Problems with Cartesian Coordinates

Spurious degrees of freedom

- 3 translational modes: $\mathbf{R} = \frac{1}{M}\sum_I m_I \mathbf{r}_I$
- 3 rotational modes

Poor conditioning

- Large eigenvalue spread in Hessian matrix
- Coupling between bond stretching and molecular rotation

Example: Dimer

- Embedding dimensions: $2 \cdot 3$
- Actual dimensionality: $2 \cdot 3 - 5$ (linear molecule)
- Internal coordinates: 1

Hessian eigenvalues:

$$\lambda_{\text{translation}} = \lambda_{\text{rotation}} = 0$$

$$\lambda_{\text{vibration}} \gg 0$$

Internal Coordinates

☰ Natural molecular descriptors

- Bond lengths: $d_{IJ} = |\mathbf{r}_I - \mathbf{r}_J|$
- Bond angles: $\theta_{IJK} = \arccos \frac{(\mathbf{r}_I - \mathbf{r}_J)\cdot(\mathbf{r}_K - \mathbf{r}_J)}{d_{IJ} d_{JK}}$
- Dihedral angles: ϕ_{IJKL} (torsion around JK bond)

✓ Advantages for optimization

- Direct chemical interpretation
- Natural constraint handling
- Better conditioned Hessian matrices
- Efficient sampling of conformational space

Internal coordinates:

$$\mathbf{q} = \{d_{IJ}, \ldots, \theta_{IJK}, \ldots, \phi_{IJKL} \ldots\}$$

Redundant Internal Coordinates

Overcompleteness

- Minimal coordinate set can have singularities
- Avoids linear dependencies near planar/linear configurations
- More robust, easier to automate

Complete vs. minimal sets

- Minimal: exactly $3N - 6$ coordinates
- Complete: all possible bonds, angles, dihedrals
- Redundant: $M > 3N - 6$ coordinates

Gradient Projection

Coordinate transformation problem

- Energy gradient available in Cartesian coordinates: $\mathbf{g}_\mathbf{x} = \nabla_\mathbf{x} E$
- Need gradient in internal coordinates: $\mathbf{g}_\mathbf{q} = \nabla_\mathbf{q} E$
- Chain rule application required

$$\frac{\partial E}{\partial q_i} = \sum_{j=1}^{3N} \frac{\partial E}{\partial x_j} \frac{\partial x_j}{\partial q_i} \tag{23}$$

$$\mathbf{g}_\mathbf{q} = \underbrace{\mathbf{B}^+}_{\text{Pseudo-inverse}} \mathbf{g}_\mathbf{x} \tag{24}$$

Wilson B-matrix properties

- Dimensions: $M \times 3N$ where M = number of internal coordinates
- For redundant coordinates: $M > 3N - 6$
- Rank deficient: $\text{rank}(\mathbf{B}) = 3N - 6$

$$\mathbf{B}_{ij} = \frac{\partial q_i}{\partial x_j} \tag{25}$$

Handling Overcompleteness

Singular value decomposition approach

$$\mathsf{B} = \mathsf{U}\boldsymbol{\Sigma}\mathsf{V}^T$$

where $\boldsymbol{\Sigma}$ contains singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$

$$\underbrace{\mathsf{B}^+}_{\text{pseudo-inverse}} = \mathsf{V}\boldsymbol{\Sigma}^+\mathsf{U}^T \tag{26}$$

$$\boldsymbol{\Sigma}^+_{ii} = \begin{cases} \sigma_i^{-1} & \text{if } \sigma_i > \epsilon \\ 0 & \text{if } \sigma_i \leq \underbrace{\epsilon}_{\text{threshold}} \end{cases} \tag{27}$$

Optimization in projected space

- Eliminates redundant degrees of freedom
- Ensures progress along physically meaningful directions
- Maintains numerical stability

Advantages of SVD approach

- Automatic rank determination
- Numerical robustness
- Handles near-linear dependencies gracefully
- Preserves maximum information content

Summary

Concepts

- Relevant points are extrema on the potential energy surface
- Internal coordinates more accessible
- Requires gradient evaluation in most cases: will be expensive