Gauss-Newton Methods

- Problem Formulation
- Gauss-Newton Methods
- Local Convergence Analysis

Contents

Problem Formulation

Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$ we are searching for solutions of the nonlinear least-squares problem

$$\min_x \frac{1}{2} \|f(x)\|_2^2 \ .$$

Examples:

- For $f(x) = Ax - b$ this amounts to solving a standard least-squares problem,
- If we want to estimate the parameters x of a nonlinear model $h : \mathbb{R}^n \to \mathbb{R}^m$ from given measurements, we often set

$$f(x) = h(x) - \eta \ .$$

Problem Formulation

Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$ we are searching for solutions of the nonlinear least-squares problem

$$\min_x \frac{1}{2} \|f(x)\|_2^2 \; .$$

Examples:

- For $f(x) = Ax - b$ this amounts to solving a standard least-squares problem,
- If we want to estimate the parameters x of a nonlinear model $h : \mathbb{R}^n \to \mathbb{R}^m$ from given measurements, we often set

$$f(x) = h(x) - \eta \; .$$

Problem Formulation

Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$ we are searching for solutions of the nonlinear least-squares problem

$$\min_x \frac{1}{2} \|f(x)\|_2^2 \ .$$

Examples:

- For $f(x) = Ax - b$ this amounts to solving a standard least-squares problem,
- If we want to estimate the parameters x of a nonlinear model $h : \mathbb{R}^n \to \mathbb{R}^m$ from given measurements, we often set

$$f(x) = h(x) - \eta \ .$$

Contents

Gauss-Newton method

In order to solve the nonlinear least-squares optimization problem $f(x)$, we start with an initial guess x_0 and solve the standard least-squares problem

$$\min_{\Delta x_k} \frac{1}{2} \, \|f(x_k) + f'(x_k)\Delta x_k\|_2^2 \ ,$$

and update $x_{k+1} = x_k + \alpha_k \Delta x_k$ for $k \in \{0, 1, 2, \ldots\}$, where $\alpha_k \in (0, 1]$ is a line search parameter.

If the Jacobian matrix $f'(x_k)$ has full-rank, the step direction can alternatively be written in the form

$$\Delta x_k = -\left(f'(x_k)^T f'(x_k)\right)^{-1} f'(x_k)^T f(x_k) \ .$$

Gauss-Newton method

In order to solve the nonlinear least-squares optimization problem $f(x)$, we start with an initial guess x_0 and solve the standard least-squares problem

$$\min_{\Delta x_k} \frac{1}{2} \left\| f(x_k) + f'(x_k)\Delta x_k \right\|_2^2 \ ,$$

and update $x_{k+1} = x_k + \alpha_k \Delta x_k$ for $k \in \{0, 1, 2, \ldots\}$, where $\alpha_k \in (0, 1]$ is a line search parameter.

If the Jacobian matrix $f'(x_k)$ has full-rank, the step direction can alternatively be written in the form

$$\Delta x_k = -\left(f'(x_k)^T f'(x_k) \right)^{-1} f'(x_k)^T f(x_k) \ .$$

Interpretation as Newton-type methods

If we apply a Newton-type method to solve the minimization problem

$$\min_x F(x) \qquad \text{with} \qquad F(x) = \frac{1}{2} \|f(x)\|_2^2$$

we obtain a step direction of the form

$$\Delta x_k = -M(x_k)^{-1} F'(x_k)^T = -M(x_k)^{-1} f'(x_k)^T f(x_k) \; .$$

Thus, Gauss-Newton methods are special class of Newton type methods which employ the Hessian approximation

$$F''(x_k) \approx M(x_k) = f'(x_k)^T f'(x_k) \; .$$

Properties of Gauss-Newton Hessian approximations

The exact Hessian matrix is given by

$$F''(x_k) = f''(x_k)^T f(x_k) + f'(x_k)^T f'(x_k) = f''(x_k)^T f(x_k) + M(x_k) \ .$$

- The approximation error $f''(x_k)^T f(x_k)$ is small if either the residuum $f(x_k)$ or the second derivative $f''(x_k)$ is small.
- The Gauss-Newton Hessian approximation $M(x_k) = f'(x_k)^T f'(x_k)$ is always positive semi-define.
- Similar to exact Newton methods, Gauss-Newton methods are invariant under scaling.
- $M(x_k)$ depends on first order derivatives only.

Properties of Gauss-Newton Hessian approximations

The exact Hessian matrix is given by

$$F''(x_k) = f''(x_k)^T f(x_k) + f'(x_k)^T f'(x_k) = f''(x_k)^T f(x_k) + M(x_k) .$$

- The approximation error $f''(x_k)^T f(x_k)$ is small if either the residuum $f(x_k)$ or the second derivative $f''(x_k)$ is small.
- The Gauss-Newton Hessian approximation $M(x_k) = f'(x_k)^T f'(x_k)$ is always positive semi-define.
- Similar to exact Newton methods, Gauss-Newton methods are invariant under scaling.
- $M(x_k)$ depends on first order derivatives only.

Properties of Gauss-Newton Hessian approximations

The exact Hessian matrix is given by

$$F''(x_k) = f''(x_k)^T f(x_k) + f'(x_k)^T f'(x_k) = f''(x_k)^T f(x_k) + M(x_k)\ .$$

- The approximation error $f''(x_k)^T f(x_k)$ is small if either the residuum $f(x_k)$ or the second derivative $f''(x_k)$ is small.
- The Gauss-Newton Hessian approximation $M(x_k) = f'(x_k)^T f'(x_k)$ is always positive semi-define.
- Similar to exact Newton methods, Gauss-Newton methods are invariant under scaling.
- $M(x_k)$ depends on first order derivatives only.

Properties of Gauss-Newton Hessian approximations

The exact Hessian matrix is given by

$$F''(x_k) = f''(x_k)^T f(x_k) + f'(x_k)^T f'(x_k) = f''(x_k)^T f(x_k) + M(x_k) \; .$$

- The approximation error $f''(x_k)^T f(x_k)$ is small if either the residuum $f(x_k)$ or the second derivative $f''(x_k)$ is small.
- The Gauss-Newton Hessian approximation $M(x_k) = f'(x_k)^T f'(x_k)$ is always positive semi-define.
- Similar to exact Newton methods, Gauss-Newton methods are invariant under scaling.
- $M(x_k)$ depends on first order derivatives only.

Properties of Gauss-Newton Hessian approximations

The exact Hessian matrix is given by

$$F''(x_k) = f''(x_k)^T f(x_k) + f'(x_k)^T f'(x_k) = f''(x_k)^T f(x_k) + M(x_k)\ .$$

- The approximation error $f''(x_k)^T f(x_k)$ is small if either the residuum $f(x_k)$ or the second derivative $f''(x_k)$ is small.
- The Gauss-Newton Hessian approximation $M(x_k) = f'(x_k)^T f'(x_k)$ is always positive semi-define.
- Similar to exact Newton methods, Gauss-Newton methods are invariant under scaling.
- $M(x_k)$ depends on first order derivatives only.

Contents

Local Convergence Analysis

As Gauss-Newton methods are special Newton-type methods the local convergence bound has the form

$$\|x_{k+1} - x^*\| \leq \kappa \|x_k - x^*\| + \frac{\omega}{2} \|x_k - x^*\|_2^2 \ ,$$

where

$$\left\|I - M(x_k)^{-1} F''(x_k)\right\| \leq \left\|M(x_k)^{-1}\right\| \|f(x_k)\| \|f''(x_k)\| \leq \kappa$$

- If we have $f''(x) = 0$, the method converges in one step.
- If we have $f(x^*) = 0$ at the limit point x^*, we have

$$\kappa \leq \mathbf{O}(\|x_k - x^*\|)$$

and the method converges with locally quadratic rate.

Local Convergence Analysis

As Gauss-Newton methods are special Newton-type methods the local convergence bound has the form

$$\|x_{k+1} - x^*\| \leq \kappa \|x_k - x^*\| + \frac{\omega}{2} \|x_k - x^*\|_2^2 \ ,$$

where

$$\left\| I - M(x_k)^{-1} F''(x_k) \right\| \leq \left\| M(x_k)^{-1} \right\| \|f(x_k)\| \, \|f''(x_k)\| \leq \kappa$$

- If we have $f''(x) = 0$, the method converges in one step.
- If we have $f(x^*) = 0$ at the limit point x^*, we have

$$\kappa \leq \mathbf{O}(\|x_k - x^*\|)$$

  and the method converges with locally quadratic rate.

Local Convergence Analysis

As Gauss-Newton methods are special Newton-type methods the local convergence bound has the form

$$\|x_{k+1} - x^*\| \leq \kappa \|x_k - x^*\| + \frac{\omega}{2} \|x_k - x^*\|_2^2 \ ,$$

where

$$\left\| I - M(x_k)^{-1} F''(x_k) \right\| \leq \left\| M(x_k)^{-1} \right\| \|f(x_k)\| \, \|f''(x_k)\| \leq \kappa$$

- If we have $f''(x) = 0$, the method converges in one step.
- If we have $f(x^*) = 0$ at the limit point x^*, we have

$$\kappa \leq \mathbf{O}(\|x_k - x^*\|)$$

  and the method converges with locally quadratic rate.