Newton's Method

- Problem Formulation
- Newton's method
- Local Convergence Analysis
- Unconstrained Optimization
- Globalization Techniques

Contents

Problem Formulation

Given a function $f : \mathbb{R}^n \to \mathbb{R}^n$ we are searching for solutions of the nonlinear equation

$$f(x) = 0 \ .$$

Examples:

- For $f(x) = Ax - b$ this amounts to solving a linear equation system,
- For $f(x) = x^2 + 1$: no solution can be found,
- For $f(x) = x^3 - x$: three solutions exist.

Problem Formulation

Given a function $f : \mathbb{R}^n \to \mathbb{R}^n$ we are searching for solutions of the nonlinear equation

$$f(x) = 0 \ .$$

Examples:

- For $f(x) = Ax - b$ this amounts to solving a linear equation system,
- For $f(x) = x^2 + 1$: no solution can be found,
- For $f(x) = x^3 - x$: three solutions exist.

Problem Formulation

Given a function $f : \mathbb{R}^n \to \mathbb{R}^n$ we are searching for solutions of the nonlinear equation

$$f(x) = 0 \; .$$

Examples:

- For $f(x) = Ax - b$ this amounts to solving a linear equation system,
- For $f(x) = x^2 + 1$: no solution can be found,
- For $f(x) = x^3 - x$: three solutions exist.

Problem Formulation

Given a function $f : \mathbb{R}^n \to \mathbb{R}^n$ we are searching for solutions of the nonlinear equation

$$f(x) = 0 \ .$$

Examples:

- For $f(x) = Ax - b$ this amounts to solving a linear equation system,
- For $f(x) = x^2 + 1$: no solution can be found,
- For $f(x) = x^3 - x$: three solutions exist.

Contents

Main Idea

In order to solve the nonlinear equation $f(x)$, we start with an initial guess x_0 and solve the linear equation systems

$$f(x_k) + M(x_k)(x_{k+1} - x_k) = 0 \ ,$$

for $k \in \{0, 1, 2, \ldots\}$. Here, the matrix $M(x_k) \in \mathbb{R}^{n \times n}$ is chosen in such a way that

$$f(x_k) + M(x_k)(x - x_k) \approx f(x) \ ,$$

is an approximation of the function f. For example, if f is differentiable, we might choose $M(x_k) = f'(x_k)$, which corresponds to the so called Newton method.

Main Idea

In order to solve the nonlinear equation $f(x)$, we start with an initial guess x_0 and solve the linear equation systems

$$f(x_k) + M(x_k)(x_{k+1} - x_k) = 0 \ ,$$

for $k \in \{0, 1, 2, \ldots\}$. Here, the matrix $M(x_k) \in \mathbb{R}^{n \times n}$ is chosen in such a way that

$$f(x_k) + M(x_k)(x - x_k) \approx f(x) \ ,$$

is an approximation of the function f. For example, if f is differentiable, we might choose $M(x_k) = f'(x_k)$, which corresponds to the so called Newton method.

Main Idea

In order to solve the nonlinear equation $f(x)$, we start with an initial guess x_0 and solve the linear equation systems

$$f(x_k) + M(x_k)(x_{k+1} - x_k) = 0 \ ,$$

for $k \in \{0, 1, 2, \ldots\}$. Here, the matrix $M(x_k) \in \mathbb{R}^{n \times n}$ is chosen in such a way that

$$f(x_k) + M(x_k)(x - x_k) \approx f(x) \ ,$$

is an approximation of the function f. For example, if f is differentiable, we might choose $M(x_k) = f'(x_k)$, which corresponds to the so called Newton method.

Newton (or Newton-type) method

If the matrix $M(x_k)$ is invertible, the method can also be written in the form

$$x_{k+1} = x_k - M(x_k)^{-1} f(x_k) \ ,$$

for $k \in \{0, 1, 2, \ldots\}$.

- In practice, we usually work with approximations $M(x_k) \approx f'(x_k)$.
- If $M(x_k)$ is independent of x_k, we only need to decompose M once (e.g., using LR or QR decomposition)
- Some methods try to update M at every step withough re-computing the Jacobian.

Newton (or Newton-type) method

If the matrix $M(x_k)$ is invertible, the method can also be written in the form

$$x_{k+1} = x_k - M(x_k)^{-1} f(x_k) \ ,$$

for $k \in \{0, 1, 2, \ldots\}$.

- In practice, we usually work with approximations $M(x_k) \approx f'(x_k)$.
- If $M(x_k)$ is independent of x_k, we only need to decompose M once (e.g., using LR or QR decomposition)
- Some methods try to update M at every step withough re-computing the Jacobian.

Newton (or Newton-type) method

If the matrix $M(x_k)$ is invertible, the method can also be written in the form

$$x_{k+1} = x_k - M(x_k)^{-1} f(x_k) \ ,$$

for $k \in \{0, 1, 2, \ldots\}$.

- In practice, we usually work with approximations $M(x_k) \approx f'(x_k)$.
- If $M(x_k)$ is independent of x_k, we only need to decompose M once (e.g., using LR or QR decomposition)
- Some methods try to update M at every step withough re-computing the Jacobian.

Newton (or Newton-type) method

If the matrix $M(x_k)$ is invertible, the method can also be written in the form

$$x_{k+1} = x_k - M(x_k)^{-1} f(x_k) \ ,$$

for $k \in \{0, 1, 2, \ldots\}$.

- In practice, we usually work with approximations $M(x_k) \approx f'(x_k)$.
- If $M(x_k)$ is independent of x_k, we only need to decompose M once (e.g., using LR or QR decomposition)
- Some methods try to update M at every step withough re-computing the Jacobian.

Scaling Properties

If x^* satisfies $f(x^*) = 0$ it also satisfies $S \cdot f(x^*) = 0$, where $S \in \mathbb{R}^{n \times n}$ can be any (invertible) scaling matrix. If we apply the above recursion to the scaled equation

$$\widetilde{f}(x) = S \cdot f(x) = 0$$

we obtain the iterates $x_{k+1} = x_k - M(x_k)^{-1} S \cdot f(x_k)$, which do in general not coincide with the iterates that are obtained without scaling f. However, if we use exact Jacobians, we have

$$M(x_k) = \widetilde{f}'(x) = S \cdot f'(x) \qquad \Longrightarrow \qquad x_{k+1} = x_k - f'(x_k)^{-1} f(x_k) \; .$$

This implies that Newton's methods with exact Jacobians is invariant under scaling.

Scaling Properties

If x^* satisfies $f(x^*) = 0$ it also satisfies $S \cdot f(x^*) = 0$, where $S \in \mathbb{R}^{n \times n}$ can be any (invertible) scaling matrix. If we apply the above recursion to the scaled equation

$$\tilde{f}(x) = S \cdot f(x) = 0$$

we obtain the iterates $x_{k+1} = x_k - M(x_k)^{-1} S \cdot f(x_k)$, which do in general not coincide with the iterates that are obtained without scaling f. However, if we use exact Jacobians, we have

$$M(x_k) = \tilde{f}'(x) = S \cdot f'(x) \qquad \Longrightarrow \qquad x_{k+1} = x_k - f'(x_k)^{-1} f(x_k) \, .$$

This implies that Newton's methods with exact Jacobians is invariant under scaling.

Scaling Properties

If x^* satisfies $f(x^*) = 0$ it also satisfies $S \cdot f(x^*) = 0$, where $S \in \mathbb{R}^{n \times n}$ can be any (invertible) scaling matrix. If we apply the above recursion to the scaled equation

$$\widetilde{f}(x) = S \cdot f(x) = 0$$

we obtain the iterates $x_{k+1} = x_k - M(x_k)^{-1} S \cdot f(x_k)$, which do in general not coincide with the iterates that are obtained without scaling f. However, if we use exact Jacobians, we have

$$M(x_k) = \widetilde{f}'(x) = S \cdot f'(x) \qquad \Longrightarrow \qquad x_{k+1} = x_k - f'(x_k)^{-1} f(x_k)\ .$$

This implies that Newton's methods with exact Jacobians is invariant under scaling.

Contents

Local Convergence Analysis

Assumptions:

- There exists a point x^* with $f(x^*) = 0$.
- The point x_0 is already in a small neighborhood of x^*.
- The scaled Jacobian matrix $M(x_k)^{-1} f'(x)$ is Lipschitz continuous w.r.t. x in a neighborhood of x^* with Lipschitz constant $\omega \geq 0$.

The basic idea is to estimate the distance of the iterates to x^*:

$$
\begin{aligned}
& \|x_{k+1} - x^*\| \\
= \; & \left\| x_k - x^* - M(x_k)^{-1} f(x_k) \right\| \\
= \; & \left\| x_k - x^* - M(x_k)^{-1} \int_0^1 J(x^* + s(x_k - x^*))(x_k - x^*) \mathrm{d}s \right\| \\
\leq \; & \left\| x_k - x^* - M(x_k)^{-1} J(x_k)(x_k - x^*) \right\| + \tfrac{\omega}{2} \left\| x_k - x^* \right\|_2^2 \; .
\end{aligned}
$$

Local Convergence Analysis

Assumptions:

- There exists a point x^* with $f(x^*) = 0$.
- The point x_0 is already in a small neighborhood of x^*.
- The scaled Jacobian matrix $M(x_k)^{-1} f'(x)$ is Lipschitz continuous w.r.t. x in a neighborhood of x^* with Lipschitz constant $\omega \geq 0$.

The basic idea is to estimate the distance of the iterates to x^*:

$$
\begin{aligned}
& \|x_{k+1} - x^*\| \\
= \; & \left\| x_k - x^* - M(x_k)^{-1} f(x_k) \right\| \\
= \; & \left\| x_k - x^* - M(x_k)^{-1} \int_0^1 J(x^* + s(x_k - x^*))(x_k - x^*) \mathrm{d}s \right\| \\
\leq \; & \left\| x_k - x^* - M(x_k)^{-1} J(x_k)(x_k - x^*) \right\| + \tfrac{\omega}{2} \left\| x_k - x^* \right\|_2^2 \; .
\end{aligned}
$$

Local Convergence Analysis

In summary, we find the estimate

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

as long as $\left\| I - M(x_k)^{-1} J(x_k) \right\| \leq \kappa$. Here, κ can be interpreted as a bound on the accuracy of the Jacobian approximation M. If we have $\kappa < 1$ and $\|x_0 - x^*\| < \frac{2}{\omega}(1-\kappa)$ the iterates contract and we have

$$\lim_{k \to \infty} x_k \ \to \ x^* \ .$$

Local Convergence Analysis

In summary, we find the estimate

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

as long as $\left\| I - M(x_k)^{-1} J(x_k) \right\| \le \kappa$. Here, κ can be interpreted as a bound on the accuracy of the Jacobian approximation M. If we have $\kappa < 1$ and $\|x_0 - x^*\| < \frac{2}{\omega}(1-\kappa)$ the iterates contract and we have

$$\lim_{k\to\infty} x_k \ \to \ x^* \ .$$

Local Convergence Analysis

In summary, we find the estimate

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

as long as $\left\|I - M(x_k)^{-1} J(x_k)\right\| \leq \kappa$. Here, κ can be interpreted as a bound on the accuracy of the Jacobian approximation M. If we have $\kappa < 1$ and $\|x_0 - x^*\| < \frac{2}{\omega}(1 - \kappa)$ the iterates contract and we have

$$\lim_{k \to \infty} x_k \ \to \ x^* \ .$$

Convergence Rate

The convergence rate estimate

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

implies that

- if we have $\kappa \neq 0$ the convergence rate is in general linear.
- if we choose $M(x_k) = J(x_k)$ (Newton's method), we have $\kappa = 0$ and

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

  In this case, the convergence rate is called quadratic. (the number of correct internal decimal places roughly doubles in every step).

Convergence Rate

The convergence rate estimate

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

implies that

- if we have $\kappa \neq 0$ the convergence rate is in general linear.
- if we choose $M(x_k) = J(x_k)$ (Newton's method), we have $\kappa = 0$ and

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

  In this case, the convergence rate is called quadratic. (the number of correct internal decimal places roughly doubles in every step).

Convergence Rate

The convergence rate estimate

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

implies that

- if we have $\kappa \neq 0$ the convergence rate is in general linear.
- if we choose $M(x_k) = J(x_k)$ (Newton's method), we have $\kappa = 0$ and

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

  In this case, the convergence rate is called quadratic. (the number of correct internal decimal places roughly doubles in every step).

Degeneracy Handling

If the exact Jacobian $J(x^*)$ is singular (has eigenvalues that are equal to zero), Newton's method is not applicable, since the matrices $J(x_k)$ converge to a singular matrix that cannot be inverted.

If the matrix $M(x_k)$ is chosen in such a way that the convergence condition

$$\left\| I - M(x_k)^{-1} J(x_k) \right\| < 1$$

is maintained. This is possible even if J is singular, although special care has to be taken, if M is ill-conditioned. If we choose M such that

$$\left\| I - M(x_k)^{-1} J(x_k) \right\| \leq \mathbf{O}(\|x_k - x^*\|) \ ,$$

a (locally) quadratic convergence rate can be recovered.

Degeneracy Handling

If the exact Jacobian $J(x^*)$ is singular (has eigenvalues that are equal to zero), Newton's method is not applicable, since the matrices $J(x_k)$ converge to a singular matrix that cannot be inverted.

If the matrix $M(x_k)$ is chosen in such a way that the convergence condition

$$\left\| I - M(x_k)^{-1} J(x_k) \right\| < 1$$

is maintained. This is possible even if J is singular, although special care has to be taken, if M is ill-conditioned. If we choose M such that

$$\left\| I - M(x_k)^{-1} J(x_k) \right\| \leq \mathbf{O}(\|x_k - x^*\|) \; ,$$

a (locally) quadratic convergence rate can be recovered.

Degeneracy Handling

If the exact Jacobian $J(x^*)$ is singular (has eigenvalues that are equal to zero), Newton's method is not applicable, since the matrices $J(x_k)$ converge to a singular matrix that cannot be inverted.

If the matrix $M(x_k)$ is chosen in such a way that the convergence condition

$$\left\| I - M(x_k)^{-1} J(x_k) \right\| < 1$$

is maintained. This is possible even if J is singular, although special care has to be taken, if M is ill-conditioned. If we choose M such that

$$\left\| I - M(x_k)^{-1} J(x_k) \right\| \le \mathbf{O}(\|x_k - x^*\|) \ ,$$

a (locally) quadratic convergence rate can be recovered.

Contents

Nonlinear Least-Squares Problems

The unconstrained nonlinear least-squares problem is given by

$$\min_x \|f(x)\|_2^2 \ .$$

- If we can find a $x^* \in \mathbb{R}^n$ with $f(x^*) = 0$, then x^* is minimizer of the above problem.
- For $f(x) = Ax - b$ this problem is a least-squares problem in standard form.
- Makes sense for any function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m \neq n$ in general.

Nonlinear Least-Squares Problems

The unconstrained nonlinear least-squares problem is given by

$$\min_x \|f(x)\|_2^2 \ .$$

- If we can find a $x^* \in \mathbb{R}^n$ with $f(x^*) = 0$, then x^* is minimizer of the above problem.
- For $f(x) = Ax - b$ this problem is a least-squares problem in standard form.
- Makes sense for any function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m \neq n$ in general.

Nonlinear Least-Squares Problems

The unconstrained nonlinear least-squares problem is given by

$$\min_x \|f(x)\|_2^2 \ .$$

- If we can find a $x^* \in \mathbb{R}^n$ with $f(x^*) = 0$, then x^* is minimizer of the above problem.
- For $f(x) = Ax - b$ this problem is a least-squares problem in standard form.
- Makes sense for any function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m \neq n$ in general.

Nonlinear Least-Squares Problems

The unconstrained nonlinear least-squares problem is given by

$$\min_x \|f(x)\|_2^2 \ .$$

- If we can find a $x^* \in \mathbb{R}^n$ with $f(x^*) = 0$, then x^* is minimizer of the above problem.
- For $f(x) = Ax - b$ this problem is a least-squares problem in standard form.
- Makes sense for any function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m \neq n$ in general.

Unconstrained Optimization Problems

An even more general class of problems are the unconstrained optimization problems

$$\min_x \; F(x) \; .$$

This contains the nonlinear least-squares problems as a special case, since we can choose $F(x) = \|f(x)\|_2^2$.

If F is twice Lipschitz-continuously differentiable, a minimizer can be found by applying Newton's method to

$$F'(x) = 0$$

If a solution x^* satisfies $F''(x) \succ 0$, it must be a local minimizer.

Unconstrained Optimization Problems

An even more general class of problems are the unconstrained optimization problems

$$\min_x \; F(x) \; .$$

This contains the nonlinear least-squares problems as a special case, since we can choose $F(x) = \|f(x)\|_2^2$.

If F is twice Lipschitz-continuously differentiable, a minimizer can be found by applying Newton's method to

$$F'(x) = 0$$

If a solution x^* satisfies $F''(x) \succ 0$, it must be a local minimizer.

Unconstrained Optimization Problems

An even more general class of problems are the unconstrained optimization problems

$$\min_x \; F(x) \; .$$

This contains the nonlinear least-squares problems as a special case, since we can choose $F(x) = \|f(x)\|_2^2$.

If F is twice Lipschitz-continuously differentiable, a minimizer can be found by applying Newton's method to

$$F'(x) = 0$$

If a solution x^* satisfies $F''(x) \succ 0$, it must be a local minimizer.

Newton-Type Methods for Optimization

In detail, Newton-type methods for unconstrained optimization problems can be written in the form

$$x_{k+1} = x_k - M(x_k)^{-1} F'(x_k)^T \ ,$$

where $M(x_k) \approx F''(x_k)$ is a suitable Hessian approximation.

- In practice, we often choose a symmetric Hessian approximation M, since F'' is symmetric.
- If $M(x_k)$ is symmetric and positive definite, the iterate x_{k+1} is the minimizer of the quadratic function

$$\min_{x_{k+1}} \ F(x_k) + F'(x_k)(x_{k+1} - x_k) + \frac{1}{2}\left(x_{k+1} - x_k\right)^T M(x_k)\left(x_{k+1} - x_k\right) \ ,$$

which can be interpreted as a quadratic model of F.

Newton-Type Methods for Optimization

In detail, Newton-type methods for unconstrained optimization problems can be written in the form

$$x_{k+1} = x_k - M(x_k)^{-1} F'(x_k)^T \ ,$$

where $M(x_k) \approx F''(x_k)$ is a suitable Hessian approximation.

- In practice, we often choose a symmetric Hessian approximation M, since F'' is symmetric.
- If $M(x_k)$ is symmetric and positive definite, the iterate x_{k+1} is the minimizer of the quadratic function

$$\min_{x_{k+1}} \ F(x_k) + F'(x_k)(x_{k+1} - x_k) + \frac{1}{2} (x_{k+1} - x_k)^T M(x_k) (x_{k+1} - x_k) \ ,$$

which can be interpreted as a quadratic model of F.

Newton-Type Methods for Optimization

In detail, Newton-type methods for unconstrained optimization problems can be written in the form

$$x_{k+1} = x_k - M(x_k)^{-1} F'(x_k)^T \ ,$$

where $M(x_k) \approx F''(x_k)$ is a suitable Hessian approximation.

- In practice, we often choose a symmetric Hessian approximation M, since F'' is symmetric.
- If $M(x_k)$ is symmetric and positive definite, the iterate x_{k+1} is the minimizer of the quadratic function

  $$\min_{x_{k+1}} \ F(x_k) + F'(x_k)(x_{k+1} - x_k) + \frac{1}{2} \left(x_{k+1} - x_k\right)^T M(x_k) \left(x_{k+1} - x_k\right) \ ,$$

  which can be interpreted as a quadratic model of F.

Contents

Line Search Methods

So far, we have only analyzed the local convergence properties of Newton-type methods. If we start far from a local solution, Newton type methods are often take "too big" steps and are divergent.

One way to fix this problem is to first compute a step-direction by solving

$$\min_{x_{k+1}} \; F(x_k) + F'(x_k)\Delta x_k + \frac{1}{2}\Delta x_k^T M(x_k)\Delta x_k \; ,$$

and update the iterate as

$$x_{k+1} = x_k + \alpha_k \Delta x_k \; .$$

Here, $\alpha_k \in (0,1]$ is a so-called line-search parameter, which is found by (approximately) solving the scalar optimization problem

$$\min_{\alpha_k \in [0,1]} \; F(x_k + \alpha_k \Delta x_k) \; .$$

Line Search Methods

So far, we have only analyzed the local convergence properties of Newton-type methods. If we start far from a local solution, Newton type methods are often take "too big" steps and are divergent.

One way to fix this problem is to first compute a step-direction by solving

$$\min_{x_{k+1}} \; F(x_k) + F'(x_k)\Delta x_k + \frac{1}{2}\Delta x_k^T M(x_k)\Delta x_k \; ,$$

and update the iterate as

$$x_{k+1} = x_k + \alpha_k \Delta x_k \; .$$

Here, $\alpha_k \in (0, 1]$ is a so-called line-search parameter, which is found by (approximately) solving the scalar optimization problem

$$\min_{\alpha_k \in [0,1]} \; F(x_k + \alpha_k \Delta x_k) \; .$$

Line Search Methods

So far, we have only analyzed the local convergence properties of Newton-type methods. If we start far from a local solution, Newton type methods are often take "too big" steps and are divergent.

One way to fix this problem is to first compute a step-direction by solving

$$\min_{x_{k+1}} \; F(x_k) + F'(x_k)\Delta x_k + \frac{1}{2}\Delta x_k^T M(x_k)\Delta x_k \; ,$$

and update the iterate as

$$x_{k+1} = x_k + \alpha_k \Delta x_k \; .$$

Here, $\alpha_k \in (0,1]$ is a so-called line-search parameter, which is found by (approximately) solving the scalar optimization problem

$$\min_{\alpha_k \in [0,1]} \; F(x_k + \alpha_k \Delta x_k) \; .$$

Armijo Line Search Conditions

In practice the line search optimization problem

$$\min_{\alpha_k \in [0,1]} F(x_k + \alpha_k \Delta x_k) \ .$$

is not solved exactly (too expensive), but only approximately.

One way to implement this is by using back-tracking until the Armijo condition

$$F(x_k + \alpha_k \Delta x_k) \leq F(x_k) + c\alpha_k F'(x_k) \Delta x_k$$

for a constant $c \ll 1$ is satisfied. This condition ensures that the line search parameter is not excessively large, although it is not sufficient to prove convergence in general.

Armijo Line Search Conditions

In practice the line search optimization problem

$$\min_{\alpha_k \in [0,1]} F(x_k + \alpha_k \Delta x_k) \ .$$

is not solved exactly (too expensive), but only approximately.

One way to implement this is by using back-tracking until the Armijo condition

$$F(x_k + \alpha_k \Delta x_k) \leq F(x_k) + c \alpha_k F'(x_k) \Delta x_k$$

for a constant $c \ll 1$ is satisfied. This condition ensures that the line search parameter is not excessively large, although it is not sufficient to prove convergence in general.

Armijo Line Search Conditions

If we substitute $\Delta x_k = -M(x_k)^{-1}F'(x_k)^T$ the Armijo line search condition can alternatively be written in the form

$$F(x_k + \alpha_k \Delta x_k) \leq F(x_k) - c\alpha_k F'(x_k) M(x_k)^{-1} F'(x_k)^T \; .$$

Thus, if M is positive definite, the Armijo condition ensures that we get a strict descent of the objective function whenever we apply a (damped) Newton step.

Positive definite approximations M in combination with Armijo line search work extremely well in practice, but other variants exist.

Armijo Line Search Conditions

If we substitute $\Delta x_k = -M(x_k)^{-1}F'(x_k)^T$ the Armijo line search condition can alternatively be written in the form

$$F(x_k + \alpha_k \Delta x_k) \leq F(x_k) - c\alpha_k F'(x_k) M(x_k)^{-1} F'(x_k)^T \ .$$

Thus, if M is positive definite, the Armijo condition ensures that we get a strict descent of the objective function whenever we apply a (damped) Newton step.
Positive definite approximations M in combination with Armijo line search work extremely well in practice, but other variants exist.