

Notes on Unconstrained Optimization — Formulations, Existence and Convergence Proofs

Compiled for an Undergraduate Statistics Student

September 16, 2025

1 Introduction and sources

We study unconstrained optimization problems

$$\min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is sufficiently smooth. These notes follow the lecture slides and textbook material; see the course slides for background and notation.

2 Preliminaries

Definition 2.1: Gradient and Hessian

For $f \in C^2$,

$$g(x) = \nabla f(x) \in \mathbb{R}^n, \quad G(x) = \nabla^2 f(x) \in \mathbb{R}^{n \times n}.$$

We will use the Taylor expansion (second order): for d small,

$$f(x + d) = f(x) + g(x)^T d + \frac{1}{2} d^T G(x) d + o(\|d\|^2).$$

3 Optimality conditions (brief)

Theorem 3.1: First- and second-order conditions

If x^* is a local minimizer and $f \in C^1$ then $g(x^*) = 0$. If moreover $f \in C^2$ and $g(x^*) = 0$, then $G(x^*) \succeq 0$. If $G(x^*) \succ 0$ then x^* is a strict local minimizer.

Proof. Standard: see Taylor expansion and directional derivative argument (short proofs omitted here as they are elementary calculus; see slides). \square

4 Line-search methods: existence and convergence

4.1 Problem set-up

Given x_k and a descent direction d_k (i.e. $g_k^T d_k < 0$), consider the one-dimensional function

$$\varphi(\alpha) = f(x_k + \alpha d_k), \quad \alpha > 0.$$

A line-search chooses a step length α_k satisfying some criterion (exact or inexact). See lecture notes for the algorithmic structure.

4.2 Existence of inexact step lengths (Wolfe / Goldstein)

We state and prove the standard existence result.

Theorem 4.1: Existence of α satisfying Wolfe / Goldstein

Let $f \in C^1$ and assume $\varphi(\alpha)$ is bounded below on $\alpha > 0$. If $g_k^T d_k < 0$ then there exists $\alpha > 0$ satisfying the Wolfe conditions (and hence Goldstein for suitable parameters).

Proof (concise). Because $g_k^T d_k < 0$ we have $\varphi'(0) = g_k^T d_k < 0$. By continuity of φ' , there exists $\epsilon > 0$ such that $\varphi'(\alpha) < 0$ for $\alpha \in [0, \epsilon)$. If φ attains its global minimum at some $\alpha^* > 0$ then $\varphi'(\alpha^*) = 0$ and the Wolfe conditions can be satisfied with small ρ and σ . If not, proceed by the following constructive bracket-and-shrink argument: start with α_0 small and increase until $\varphi(\alpha)$ stops decreasing (this finds an interval $[0, \bar{\alpha}]$ containing a minimizer under mild regularity). Then apply the intermediate value property to $\psi(\alpha) := \varphi'(\alpha) - \sigma\varphi'(0)$. Initially $\psi(0) = \varphi'(0) - \sigma\varphi'(0) = (1 - \sigma)\varphi'(0) < 0$. At a sufficiently large α where $\varphi'(\alpha)$ becomes nonnegative, $\psi(\alpha) \geq -\sigma\varphi'(0) > 0$. By continuity there is α with $\psi(\alpha) = 0$, which gives the curvature Wolfe condition. The Armijo (sufficient decrease) condition can be enforced by choosing a point not too large in the bracket. This standard argument (see textbooks) yields existence. A similar argument gives Goldstein existence because Goldstein is a two-sided Armijo condition which holds on a nonempty interval. \square

4.3 Global convergence of line-search methods (Zoutendijk-type result)

We give a concise proof that under standard assumptions, a class of line-search methods with Wolfe conditions yields $\lim_{k \rightarrow \infty} \|g_k\| = 0$.

Assumptions.

1. f is bounded below on the level set $\{x : f(x) \leq f(x_0)\}$.
2. $f \in C^1$ and g is Lipschitz continuous on the level set: $\|g(x) - g(y)\| \leq L\|x - y\|$.
3. Directions d_k are descent directions and the angle between d_k and $-g_k$ is uniformly bounded away from $\pi/2$: there exists $\mu > 0$ with

$$-\frac{g_k^T d_k}{\|g_k\| \|d_k\|} \geq \mu > 0 \quad \forall k.$$

4. α_k satisfy the Wolfe conditions (with fixed $0 < \rho < \sigma < 1$).

Theorem 4.2: Global convergence (gradient vanishes)

Under the above assumptions either the algorithm terminates at a stationary point or $\lim_{k \rightarrow \infty} \|g_k\| = 0$.

Sketch. From the Armijo (sufficient decrease) Wolfe condition,

$$f(x_k) - f(x_{k+1}) \geq -\rho \alpha_k g_k^T d_k = \rho \alpha_k \|g_k\| \|d_k\| \cos \theta_k,$$

where $\cos \theta_k = -g_k^T d_k / (\|g_k\| \|d_k\|) \geq \mu$. The curvature Wolfe condition and Lipschitz continuity of g imply a lower bound on α_k :

$$g(x_{k+1})^T d_k \geq \sigma g_k^T d_k \implies \|g(x_{k+1}) - g_k\| \|d_k\| \geq (1 - \sigma) |g_k^T d_k|.$$

Using $\|g(x_{k+1}) - g_k\| \leq L \alpha_k \|d_k\|$ gives

$$\alpha_k \geq \frac{(1 - \sigma) |g_k^T d_k|}{L \|d_k\|^2} = \frac{(1 - \sigma) \|g_k\| \cos \theta_k}{L \|d_k\|}.$$

Combine with the decrease bound:

$$f(x_k) - f(x_{k+1}) \geq \rho \frac{(1 - \sigma) \|g_k\|^2 \cos^2 \theta_k}{L}.$$

By the uniform angle bound $\cos \theta_k \geq \mu$, we obtain

$$f(x_k) - f(x_{k+1}) \geq C \|g_k\|^2, \quad C := \rho(1 - \sigma)\mu^2/L > 0.$$

Summing from $k = 0$ to $N - 1$ gives

$$f(x_0) - f(x_N) \geq C \sum_{k=0}^{N-1} \|g_k\|^2.$$

Because f is bounded below, the left-hand side is bounded, so $\sum_k \|g_k\|^2 < \infty$. Hence $\|g_k\| \rightarrow 0$. This is the usual Zoutendijk-type argument; full details are in standard textbooks. □

5 Common algorithms and local convergence

5.1 Steepest descent (negative gradient) — brief comment

Steepest descent with exact line search is globally convergent but can be slow. With inexact line search under the previous assumptions it satisfies the global convergence result above. For quadratic functions it converges linearly; details omitted.

5.2 Newton's method: quadratic convergence

Newton iteration:

$$x_{k+1} = x_k - G(x_k)^{-1}g(x_k),$$

assuming $G(x_k)$ is invertible.

Theorem 5.1: Local quadratic convergence of Newton's method

Let $f \in C^2$ and x^* be a point with $g(x^*) = 0$ and $G(x^*)$ nonsingular. If G is Lipschitz continuous in a neighborhood of x^* , then for x_0 sufficiently close to x^* the Newton iterates are well-defined and converge to x^* quadratically:

$$\|x_{k+1} - x^*\| \leq K \|x_k - x^*\|^2$$

for some $K > 0$.

Sketch. Write the Newton step $s_k = -G(x_k)^{-1}g(x_k)$. Using Taylor expansion of g at x^* :

$$g(x_k) = G(x^*)(x_k - x^*) + r_k, \quad \|r_k\| \leq \frac{1}{2}L\|x_k - x^*\|^2,$$

with L the Lipschitz constant of G . Then

$$x_{k+1} - x^* = x_k - x^* - G(x_k)^{-1}g(x_k).$$

Replace $g(x_k)$ with the expansion and use continuity of G^{-1} near x^* to get

$$\|x_{k+1} - x^*\| \leq C \|x_k - x^*\|^2$$

for x_k close, which yields quadratic convergence. See standard derivations. □

5.3 Quasi-Newton (BFGS) — superlinear convergence

Quasi-Newton methods build an approximation B_k to the Hessian (or its inverse) updating it each step (BFGS is the common choice). Under standard assumptions (local smoothness, B_k bounded and positive definite, line search satisfying strong Wolfe conditions), BFGS achieves superlinear convergence to a local minimizer when started sufficiently close.

Theorem 5.2: Local superlinear convergence of BFGS (sketch)

Assume $f \in C^2$, $G(x^*) \succ 0$, and BFGS updates are used with line searches satisfying the strong Wolfe conditions. If initial x_0 sufficiently close to x^* and B_0 positive definite, then $x_k \rightarrow x^*$ and the rate is superlinear.

Sketch. The proof combines (i) the Dennis–More condition which shows that the quasi-Newton update approximates the true Hessian sufficiently fast, and (ii) the fact that strong Wolfe line searches control the step lengths so that the curvature condition holds, ensuring stability of updates. From these, one shows $\|B_k - G(x^*)\| \rightarrow 0$ and the step s_k behaves like the Newton step asymptotically, which yields superlinear convergence. See textbook proofs for the full technical steps. □

6 Trust-region methods

6.1 Subproblem and existence of solution

At iterate x_k build quadratic model

$$q_k(d) = f(x_k) + g_k^T d + \frac{1}{2} d^T G_k d,$$

and solve the trust-region subproblem

$$\min_{d \in \mathbb{R}^n} q_k(d) \quad \text{s.t. } \|d\| \leq \Delta_k.$$

Proposition 6.1. *Existence of global minimizer for subproblem For any symmetric G_k and any $\Delta_k > 0$, the trust-region subproblem has at least one global minimizer.*

Proof. The feasible set $\{d : \|d\| \leq \Delta_k\}$ is compact and q_k is continuous, hence it attains a minimum on that set by the extreme value theorem. □

Uniqueness. Uniqueness holds if q_k is strictly convex on the ball, e.g., when $G_k \succ 0$ (positive definite) so the unconstrained minimizer $d_N = -G_k^{-1} g_k$ is unique and either lies inside the ball (then it is the unique solution) or the constrained solution on the boundary is unique under mild conditions.

6.2 Basic convergence idea

After solving the subproblem we compute the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + d_k)}{q_k(0) - q_k(d_k)}$$

and accept the step if ρ_k is sufficiently large, updating Δ_k by rules such as those in the lecture slides. Under standard assumptions (same Lipschitz and boundedness assumptions as in line-search), trust-region methods are globally convergent in the sense that either a stationary point is found or $\|g_k\| \rightarrow 0$. The proof uses the fact that the model reduction approximates the actual reduction, and poor model quality (small ρ_k) leads to shrinking Δ_k , whereas good ρ_k increases Δ_k , guaranteeing eventual progress. □

6.3 Dogleg method (brief)

Dogleg is a practical strategy when G_k is positive definite: combine the Cauchy (steepest-descent scaled) step d_{SD} and Newton step d_N along a piecewise linear path to find a point on the trust-region boundary with acceptable reduction. The Dogleg algorithm produces a candidate d_k satisfying model decrease; global convergence of the trust-region method using Dogleg follows from the general trust-region convergence theory because Dogleg delivers a (near-)optimal subproblem solution and the standard ρ_k -based radius updates apply. □