iterations. The term $\log_2 \log_2(\epsilon_0/\epsilon)$ grows *extremely slowly* as ϵ goes to zero, and for practical purposes it can be considered constant, say five or six (six iterations gives an accuracy of about $\epsilon \approx 5 \cdot 10^{-20}\epsilon_0$).

In summary, the number of Newton iterations required to find a minimum of J is approximately bounded by

$$\frac{J(u_0) - p^*}{\gamma} + 6.$$

Examples of the application of Newton's method and further discussion of its efficiency are given in Boyd and Vandenberghe [29] (Section 9.5.4). Basically, Newton's method has a faster convergence rate than gradient or steepest descent. Its main disadvantage is the cost for forming and storing the Hessian, and of computing the Newton step, which requires solving a linear system.

There are two major shortcomings of the convergence analysis of Newton's method as sketched above. The first is a pracical one. The complexity estimates involve the constants m, M, and L, which are almost never known in practice. As a result, the bound on the number of steps required is almost never known specifically.

The second shortcoming is that although Newton's method itself is affine invariant, the analysis of convergence is very much dependent on the choice of coordinate system. If the coordinate system is changed, the constants m, M, L also change. This can be viewed as an aesthetic problem, but it would be nice if an analysis of convergence independent of an affine change of coordinates could be given.

Nesterov and Nemirovski discovered a condition on functions that allows an affine-invariant convergence analysis. This property, called *self-concordance*, is unfortunately not very intuitive.

Definition 49.10. A (partial) convex function f defined on $\mathbb{R}$ is *self-concordant* if

$$|f'''(x)| \leq 2(f''(x))^{3/2} \quad \text{for all } x \in \mathbb{R}.$$

A (partial) convex function f defined on $\mathbb{R}^n$ is *self-concordant* if for every nonzero $v \in \mathbb{R}^n$ and all $x \in \mathbb{R}^n$, the function $t \mapsto J(x + tv)$ is self-concordant.

Affine and convex quadratic functions are obviously self-concordant, since $f''' = 0$. There are many more interesting self-concordant functions, for example, the function $X \mapsto -\log\det(X)$, where X is a symmetric positive definite matrix.

Self-concordance is discussed extensively in Boyd and Vandenberghe [29] (Section 9.6). The main point of self-concordance is that a coordinate system-invariant proof of convergence can be given for a certain class of strictly convex self-concordant functions. This proof is given in Boyd and Vandenberghe [29] (Section 9.6.4). Given a starting value u_0, we assume that the sublevel set $\{x \in \mathbb{R}^n \mid J(x) \leq J(u_0)\}$ is closed and that J is bounded below. Then there are two parameters η and γ as before, but *depending only on the parameters α, β involved in the line search*, such that: