the method is determined by the largest and the smallest eigenvalues of A. Strang discusses this issue in the case of a 2×2 matrix. Convergence is significantly accelerated.

Another method is known as *Nesterov acceleration*. In this method,

$$u_{k+1} = u_k + \beta(u_k - u_{k-1}) - \rho \nabla J_{u_k + \gamma(u_k - u_{k-1})},$$

where β, ρ, γ are parameters. For details, see Strang [171] (Chapter VI, Section 4).

Lax also discusses other methods in which the step ρ_k is chosen using roots of Chebyshev polynomials; see Lax [113], Chapter 17, Sections 2–4.

A variant of Newton's method described in Section 41.2 can be used to find the minimum of a function belonging to a certain class of strictly convex functions. This method is the special case of the case where the norm is induced by a symmetric positive definite matrix P, namely $P = \nabla^2 J(x)$, the Hessian of J at x.

49.9 Newton's Method For Finding a Minimum

If $J\colon \Omega \to \mathbb{R}$ is a convex function defined on some open subset Ω of $\mathbb{R}^n$ which is twice differentiable and if its Hessian $\nabla^2 J(x)$ is symmetric positive definite for all $x \in \Omega$, then by Proposition 40.12(2), the function J is strictly convex. In this case, for any $x \in \Omega$, we have the quadratic norm induced by $P = \nabla^2 J(x)$ as defined in the previous section, given by

$$\|u\|_{\nabla^2 J(x)} = (u^\top \nabla^2 J(x)\, u)^{1/2}.$$

The steepest descent direction for this quadratic norm is given by

$$d_{\mathrm{nt}} = -(\nabla^2 J(x))^{-1} \nabla J_x.$$

The norm of d_{nt} for the the quadratic norm defined by $\nabla^2 J(x)$ is given by

$$\begin{aligned}
(d_{\mathrm{nt}}^\top \nabla^2 J(x)\, d_{\mathrm{nt}})^{1/2} &= \left(-(\nabla J_x)^\top (\nabla^2 J(x))^{-1} \nabla^2 J(x) (-(\nabla^2 J(x))^{-1} \nabla J_x)\right)^{1/2} \\
&= \left((\nabla J_x)^\top (\nabla^2 J(x))^{-1} \nabla J_x\right)^{1/2}.
\end{aligned}$$

Definition 49.9. Given a function $J\colon \Omega \to \mathbb{R}$ as above, for any $x \in \Omega$, the *Newton step* d_{nt} is defined by

$$d_{\mathrm{nt}} = -(\nabla^2 J(x))^{-1} \nabla J_x,$$

and the *Newton decrement* $\lambda(x)$ is defined by

$$\lambda(x) = \left((\nabla J_x)^\top (\nabla^2 J(x))^{-1} \nabla J_x\right)^{1/2}.$$

Observe that

$$\langle \nabla J_x, d_{\mathrm{nt}} \rangle = (\nabla J_x)^\top (-(\nabla^2 J(x))^{-1} \nabla J_x) = -\lambda(x)^2.$$