Another rule for choosing the stepsize is *Armijo's rule*.

These methods, and others, are discussed in detail in Berstekas [17].

Boyd and Vandenberghe discuss steepest descent methods for various types of norms besides the Euclidean norm; see Boyd and Vandenberghe [29] (Section 9.4). Here is brief summary.

49.8 Steepest Descent for an Arbitrary Norm

The idea is to make $\langle \nabla J_{u_k}, d_k \rangle$ as negative as possible. To make the question sensible, we have to limit the size of d_k or normalize by the length of d_k.

Let $\| \ \|$ be any norm on $\mathbb{R}^n$. Recall from Section 14.7 that the *dual norm* is defined by

$$\|y\|^D = \sup_{\substack{x \in \mathbb{R}^n \\ \|x\|=1}} |\langle x, y \rangle|.$$

Definition 49.8. A *normalized steepest descent direction* (with respect to the norm $\| \ \|$) is any unit vector $d_{\mathrm{nsd},k}$ which achieves the minimum of the set of reals

$$\{\langle \nabla J_{u_k}, d \rangle \mid \|d\| = 1\}.$$

By definition, $\|d_{\mathrm{nsd},k}\| = 1$.

A *unnormalized steepest descent direction* $d_{\mathrm{sd},k}$ is defined as

$$d_{\mathrm{sd},k} = \|\nabla J_{u_k}\|^D \, d_{\mathrm{nsd},k}.$$

It can be shown that

$$\langle \nabla J_{u_k}, d_{\mathrm{sd},k} \rangle = -(\|\nabla J_{u_k}\|^D)^2;$$

see Boyd and Vandenberghe [29] (Section 9.4).

The *steepest descent method (with respect to the norm $\| \ \|$)* consists of the following steps: Given a starting point $u_0 \in \mathrm{dom}(J)$ do:

repeat

(1) Compute the steepest descent direction $d_{\mathrm{sd},k}$.

(2) Line search. Perform an exact or backtracking line search to find ρ_k.

(3) Update. $u_{k+1} = u_k + \rho_k d_{\mathrm{sd},k}$.