Householder Reflectors

The Householder reflector is arguably the most important tool in (dense) numerical linear algebra. Let $u \in \mathbb{R}^{n \times 1}$. Then the Householder reflector defined by u is given by

$$H = H(u) = I - \beta u u^T, \quad \text{where } \beta = 2/(u^T u).$$

Algebraically: $H = H^{-1}$ is a symmetric (Hermitian) rank-1 perturbation of I. Analytically: H is an orthogonal (unitary) matrix. Geometrically: Hv is the reflection of v about the hyperplane orthogonal to u (as a function: $u \to H(u)$ has domain $\mathbb{R}P^{n-1}$, and as an operator: $H : v \to Hv$ is an orthogonal reflector on $\mathbb{R}^n$).

Typically, H is used in matrix factorizations to introduce zeros into some other matrix. To see how it works, suppose we would like an arbitrary vector x to be sent to a multiple of some vector y under the action of H, i.e. find u such that $Hx = \alpha y$. Since H is orthogonal, $\|x\|_2 = \|Hx\|_2 = |\alpha|\|y\|_2$, giving $|\alpha| = \|x\|_2/\|y\|_2$. If $(I - \beta u u^T)x = \alpha y$, then $\eta u = x - \alpha y$, where $\eta = \beta(u^T x) \in \mathbb{R}$. Since $H(u) = H(\gamma u)$, we may take u to be any (nonzero) multiple of $x \pm \alpha y$.

Introducing zeros into a matrix is usually cast as introducing zeros below a given element, so we will take y above to be e_1 (zeros below the first element). In that case u will be a multiple of $x \pm \alpha e_1$. Now put on your error analysis hat and show that we should take u to be a multiple of $x + \operatorname{sign}(x_1)\|x\|_2\, e_1$ (hint: what happens if $x \approx e_1$?). Such a u is called a *Householder vector* for x.

Notice that except for the first entry, u is x. The only computational task, therefore, is to find $\|x\|_2$, and the only challenge there is to avoid underflow or overflow (which should be incorporated into any 2–norm code anyway (scale)).

So for any x we can easily compute a Householder vector u such that $Hx = \pm\|x\|_2 e_1$. In order to zero entries $k{+}1{:}n$ of a vector y, we simply compute a Householder vector $\tilde{u}$ for $y(k{:}n)$. Then embedding $\tilde{u}$ in u: $u^T = (0, \tilde{u}^T)$ gives an embedding of $\tilde{H} \equiv I - \beta \tilde{u}\tilde{u}^T$ in $H = I - \beta u u^T$:

$$u = \begin{pmatrix} 0 \\ \tilde{u} \end{pmatrix} \quad \Longrightarrow \quad H = \begin{bmatrix} I & 0 \\ 0 & \tilde{H} \end{bmatrix}.$$

A discussion of Householder reflectors wouldn't be complete without looking at *how* we compute HB for some matrix $B \in \mathbb{R}^{n \times p}$. H is $n \times n$, but is completely defined by $u \in \mathbb{R}^n$, and as such we should expect that we can take advantage of the structure. Firstly, we don't explicitly form H. It would be wasteful of both memory and computation. Instead, we just remember (store) u. We don't need H:

$$HB = (I - \beta u u^T)B = B - (\beta u)(u^T B).$$

Some think that we should save memory by scaling u so that $u(1) = 1$ (and since it is known implicitly, it doesn't need to be stored), others suggest scaling u so that $\beta = 1$ ($\|u\|_2 = \sqrt{2}$, and thus (βu) doesn't require any computation), and I prefer a base-2 scaling that avoids a bit of rounding error and can be fast. These ideas are all fine, but rather inconsequential if n is very large.