Notes on Householder transformations

Herman Jaramillo

May 10, 2016

Abstract

Just for my own understanding.

1 The Householder transformation

The QR factorization of a matrix could be done by the use of the Cholesky factorization, Fast Givens, Grand–Schmidt or Householder transformation methods (among others).

Here I want to explain about the Householder transformation method.

Given two vectors with equal norm, $\boldsymbol{x}$ and $\boldsymbol{y}$ we want to find a matrix that transforms $\boldsymbol{x}$ into $\boldsymbol{y}$. If the vectors are the same, the problem is trivial and the matrix is the identity I, then we assume $\boldsymbol{x} \neq \boldsymbol{y}$.

We can build the unit vector that sends us from the starting point $\boldsymbol{x}$ to the end point $\boldsymbol{y}$, that is

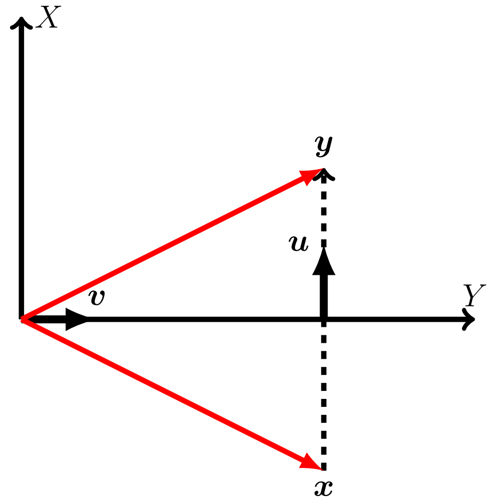
$$\boldsymbol{u} = \frac{\boldsymbol{x} - \boldsymbol{y}}{\|\boldsymbol{x} - \boldsymbol{y}\|}.$$

So we want a transformation that reflect the vector $\boldsymbol{x}$ to its mirror image $\boldsymbol{y}$ with respect to the axis $\boldsymbol{v}$, which would be the bisector of the two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. That is

$$\boldsymbol{v} = \frac{\boldsymbol{x} + \boldsymbol{y}}{\|\boldsymbol{x} + \boldsymbol{y}\|}.$$

Since the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are in a plane (spanned by both vectors) we can assume that we are living in this plane with some Cartesian coordinates X and Y, as shown in figure 1. Let us assume that the vector $\boldsymbol{x}$ is below

Figure 1: An illustration of the construction of the Householder transformation



the X It is clear that to go from $\boldsymbol{x}$ to $\boldsymbol{y}$ we have to add the dashed vector that joins $\boldsymbol{x}$ to $\boldsymbol{y}$. But this vector is 2 times the the projection of $\boldsymbol{x}$ along the direction $\boldsymbol{u}$. That is

$$\boldsymbol{x} - \boldsymbol{y} = 2\langle \boldsymbol{x}, \boldsymbol{u} \rangle \boldsymbol{u},$$

from which

$$\boldsymbol{y} = \boldsymbol{x} - 2\langle \boldsymbol{x}, \boldsymbol{u} \rangle \boldsymbol{u} = \boldsymbol{x} - 2\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{x} = (I - 2\boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x} \tag{1.1}$$

I claim that it was something like this that motivated Householder to define his general matrix

$$H = I - 2\boldsymbol{u}\boldsymbol{u}^T.$$

in $\mathbb{R}^{n\times n}$. In words: The Householder transformation H, reflects the vector $\boldsymbol{x}$ with respect to the plane orthogonal to $\boldsymbol{u}$, into $\boldsymbol{y}$.

The properties and use of this matrix are interesting. Let us list a few properties.

1.1 properties

- Besides $H\boldsymbol{x} = \boldsymbol{y}$ we have the particular mappings

$$\begin{aligned} H\boldsymbol{v} &= \boldsymbol{v} \quad \text{and} \\ H\boldsymbol{u} &= -\boldsymbol{u} \end{aligned}$$

  That is, both $\boldsymbol{v}$ and $\boldsymbol{u}$ are eigenvectors of H, and the $\lambda_u = 1$ is an eigenvalue of multiplicity $n-1$, (since there are $n-1$ dimensions in the hyperplane normal to $\boldsymbol{v}$), and λ_v is an eigenvalue of multiplicity 1.

  The proof for this is by simple substitution:

$$\begin{aligned} H\boldsymbol{v} &= (I - 2\boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{v} \\ &= \boldsymbol{v} - 2\boldsymbol{u}(\cancelto{0}{\boldsymbol{u}^T\boldsymbol{v}}) \\ &= \boldsymbol{v} \end{aligned}$$

$$\begin{aligned} H\boldsymbol{u} &= (I - 2\boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{u} \\ &= \boldsymbol{u} - 2(\boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{u} \\ &= \boldsymbol{u} - 2\boldsymbol{u}(\cancelto{1}{\boldsymbol{u}^T\boldsymbol{u}}) \\ &= -\boldsymbol{u} \end{aligned}$$

- H is symmetric. That is

$$H^T = (I - 2\boldsymbol{u}\boldsymbol{u}^T)^T = I - 2\boldsymbol{u}\boldsymbol{u}^T,$$

- H is orthogonal. That is, $HH^T = H^T H = I$. Let us proof one of these equalities.

  Since H is symmetric

$$\begin{aligned} HH^T &= (I - 2\boldsymbol{u}\boldsymbol{u}^T)(I - 2\boldsymbol{u}\boldsymbol{u}^T) \\ &= (I - 4\boldsymbol{u}\boldsymbol{u}^T + 4(\boldsymbol{u}\boldsymbol{u}^T)(\boldsymbol{u}\boldsymbol{u}^T)) \\ &= I - 4\boldsymbol{u}\boldsymbol{u}^T + 4\boldsymbol{u}(\cancelto{1}{\boldsymbol{u}^T\boldsymbol{u}})\boldsymbol{u}^T \\ &= I. \end{aligned}$$

- H is idempotent. That is $H^2 = I$. This is immediate from the symmetric and orthogonality conditions above.

1.2 The appealing use of geometric algebra

.

1.2.1 Notes on Clifford Algebras

I use Eric Chisolm notes on geometric algebra for this section. I found interesting the simplification of the Householder transformation under the geometric algebra (also known as Clifford algebra).

Up to this point I used bold face fonts to represent vectors. Vectors of a geometric algebra space come in all dimensions and there are no distinction except for the context, so I will drop the bold face fonts and use regular non–boldface fonts for all vector objects.

The motivation that Eric presents is that of generalizing the concept of vector based on its properties and in a coordinate free space.

By taking the intrinsic properties of a vector

1. An *attitude:* exactly which subspace is represented.

2. A *weight* a measure, length, area, volume, etc.

3. An *orientation:* positive/negative, forward/backward, clockwise/counterclockwise. No matter the dimension a space has always two dimensions.

By extracting this intrinsic properties we factor out all tensor that we know.

- Tensors of order zero:
  1. Have the attitude of scalars.
  2. Have an absolute value (weight).
  3. Could be positive or negative.
- Tensors of order one
  1. Have the attitude of traditional vectors.
  2. Have a norm (length).
  3. Could be positive or negative.
- Tensors of order two

1. Have the attitude of matrices.
2. Have a norm. However here we want to think about the determinant instead of the norm, because the determinant measures the volume corresponding to the matrix.
3. Could be positive or negative (left or right handed).

In general a tensor of any order can have those characteristics.

1. The attitude is the rank of the tensor.
2. The weight is?
3. The orientation is?

It is not clear what the weight and orientation could be for tensors of rank 3 and higher. For tensors of order 2 (matrices) we know that the determinant plays the role of area and volume and it has a sign but, what is the generalization of determinant for higher order tensors?

In geometric algebra the dot and wedge products are defined as well as a new definition of product called "geometric product", which allows to represent objects of any dimension and all as part of the same space (which is not possible in traditional linear algebra where each space can be represented by objects of only one dimension n)

In traditional linear algebra we project a vector v along a vector u with the operation

$$P_u(v) = \frac{u \cdot v}{|u|^2} u, \tag{1.2}$$

where $u \cdot v$ is the inner product and $|u|^2 = u \cdot u$ is the square of the length of u. the cross product represents the oriented plane defined by u and v; which point along the normal, and the direction indicates the orientation. Beyond 3D the definition of cross product does not work Differential geometry provides a definition called the wedge "$\wedge$" product which we introduce here in a different way and that works as a good extension for the three dimensional cross–product.

I now introduce the motivation and definition of the new product in Clifford algebras. We want a product uv that follows the basic properties of algebra, that is, it has a module, it is associative, it has closure, and element has an inverse (these are the properties of a group, so the algebra would be a group under this operations). If this is the case, then we have that

$$uv = \frac{1}{2}(uv + vu) + \frac{1}{2}(uv - vu). \tag{1.3}$$

The first term looks like an inner product and has the properties of the inner product, so we define

$$u \cdot v \equiv \frac{1}{2}(uv + vu).$$

See with this that $u^2 = u.u = |u|^2$, so the square of any vector is just its squared length. Then we can define

$$u^{-1} \equiv \frac{u}{u^2},$$

(assuming $u \neq 0$) is the multiplicative inverse of u, since $uu^{-1} = \frac{u^2}{u^2} = 1$. So in this sense we can divide vectors. Also, the projection formula 1.2 turns out to be

$$P_u(v) = (v.u)u^{-1}.$$

Here $u \neq 0$, if $u = 0$ we do not need any projection to 0, do we?

The second term of the product uv in 1.3 is defined as the *wedge* product. That is

$$u \wedge v = \frac{1}{2}(uv - vu).$$

First we see here clear the similarity with the determinant formula. The wedge product is anti–symmetric. Let us investigate more of its properties. 1.3 we see

$$\begin{aligned} uv &= u \cdot v + u \wedge v \\ vu &= u \cdot v - u \wedge v, \end{aligned}$$

we multiply these equations together to get

$$uvvu = (u \cdot v)^2 - (u \wedge v)^2.$$

Now $vv = |v|^2, uu = |u|^2$,so

$$\begin{aligned}(u \wedge v)^2 &= uvvu - (u \cdot v)^2 \\ &= |u|^2|v|^2 \cos^2\theta - |u|^2|v|^2 \\ &= -|u|^2|v|^2 \sin^2\theta.\end{aligned}$$

So the square of the wedge product is a negative scalar and has the magnitude of the square of the cross product.

These two properties are interesting. The wedge product between to vectors represents the area formed by their construction parallelogram (as the cross product does) and it is a complex number representing a 90 degrees phase shift ($\mathrm{i} = \mathrm{e}^{\pi\mathrm{i}/2}$). The product $u \wedge v$ is called a *simple bisector* or *2–blade* , so 2-blades represent planes with an area and orientation (interchange u and v and you change the sign of $u \wedge v$). There are blades of all orders, for example

$$u \wedge v \wedge w,$$

which can be seen as $u \wedge (\wedge v \wedge u)$. So by doing binary operations we can extend the definition of any order (here 3, or 3-blades).

Let us now use the learned elements to show the Householder transformation from the point of view of Clifford algebra.

Referring to figure 1 again,

Let x be the vector we want to reflect and v be a vector along the reflection axis. Then

$$\begin{aligned}x &= x(vv^{-1}) \\ &= (xv)v^{-1} \\ &= (x.v)v^{-1} + (x \wedge v)v{-}1.\end{aligned}$$

The first term is the projection of x along v. So the other term is the component of x perpendicular to v also called the *orthogonal rejection* of x from v. Now let y be the reflected vector. Its component perpendicular to v has the same sign, while its component along v should have opposite, so the reflected vector is given by

$$\begin{aligned}y &= -(x \cdot v)v^{-1} + (x \wedge v)v^{-1} \\ &= [-(v \cdot x) + (v \wedge x)]v^{-1} \\ &= -(vx)v^{-1}\end{aligned}$$

This is a great simplification over the Householder reflection 1.1. We further note the following symmetries

- $(v \cdot x)v^{-1}$ is the projection from x to v.
- $-(v \wedge x)v^{-1}$ is the orthogonal rejection of from x from v.
- $-(vx)v^{-1}$ is the reflection of x with respect to v.

1.3 Use in the QR factorization

We ask if it is possible to find

$$H\boldsymbol{x} = \|\boldsymbol{x}\|\boldsymbol{e_1},$$

That is, a matrix H such that takes the vector $\boldsymbol{x}$, into a vector with first component $\|\boldsymbol{x}\|$ and the rest of the components 0.

Figure 2: An illustration of the construction of the Householder transformation, that suppress all components but the first on a vector.

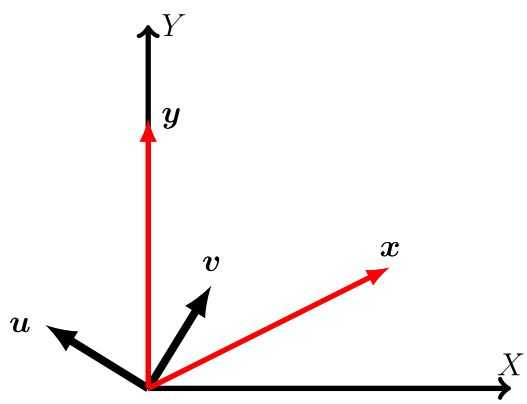


Figure 2 illustrates this situation. A given vector $\boldsymbol{x}$ and its x_1 component suppress by a reflection with respect to the vector $\boldsymbol{v}$. The vector

$$\boldsymbol{u} = \frac{\boldsymbol{x} - \boldsymbol{y}}{\|\boldsymbol{x} - \boldsymbol{y}\|}$$

should do the work of reference vector for the Householder transformation.

Let us consider an example. Given $\boldsymbol{x} = (0, -3, 4)^T$, construct a Householder matrix H such that $H\boldsymbol{x} = c\,\boldsymbol{e_1}$. We find

$$c = \|\boldsymbol{x}\| = 5,$$

so since $(0, -3, 4)^T - 5(1, 0, 0)^T = (-5, -3, 4)^T$, and

$$\|(-5, -3, 4)^T\| = \sqrt{25 + 9 + 15} = \sqrt{50}$$

then

$$\boldsymbol{u} = \frac{1}{\sqrt{50}}(-5, -3, 4)^T$$

and

$$\begin{aligned} H = I - 2\boldsymbol{u}\boldsymbol{u}^T &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{2}{50}\begin{pmatrix} 25 & 15 & -20 \\ 15 & 9 & -12 \\ -20 & -12 & 16 \end{pmatrix} \\ &= \frac{1}{25}\begin{pmatrix} 0 & -15 & 20 \\ -15 & 16 & 12 \\ 20 & 12 & 9 \end{pmatrix} \end{aligned}$$

We check that

$$\frac{1}{25}\begin{pmatrix} 0 & -15 & 20 \\ -15 & 16 & 12 \\ 20 & 12 & 9 \end{pmatrix}\begin{pmatrix} 0 \\ -3 \\ 4 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix}$$

The QR factorization continue inductively for lower dimensions by reducing the size of the matrix each time by 1, from the upper most left entry down to the lower most right entry.