Householder Transformation

Let $v \in \mathbb{R}^n$, then $v^\perp$ is an $(n-1)$-dimensional subspace with normal $v$. For any given $x \in \mathbb{R}^n$, we can find another $y \in \mathbb{R}^n$ such that $y$ is symmetric to $x$ with respect to $v^\perp$:

Let $w := \text{proj}_v x = \frac{\langle v, x \rangle}{\langle v, v \rangle} v = \frac{v^t x}{v^t v} v$, the projection of $x$ onto $v$, then $x = \underbrace{w}_{\in \text{span}\{v\}} + (x - w)$ \in v^\perp

and so,

$$y = -w + (x - w) = x - 2w$$

$$= x - \frac{2v^t x}{v^t v} v$$ (v^t x \text{ is a number})

$$= x - \frac{2}{v^t v} v v^t x$$

$$= \left( I - \frac{2}{v^t v} v v^t \right) x.$$ \text{def} P_v

i.e. we found that there is a linear transformation $P_v$ such that $x \xrightarrow{P_v} y$. Such $P_v$ is the so-called Householder matrix/transformation, which transfers any $x \in \mathbb{R}^n$ into the one symmetric to $x$ w.r.t. $v^\perp$, i.e. let $y = P_v x$, then $x$ symmetric to $y$ w.r.t. $v^\perp$.

Clearly, we only need a vector $v$ to define a Householder matrix $P_v$, and it posses several good things: Intuitively, $P_w = P_v$ for any $0 \neq w \parallel v$, and since $P_v$ "flips" vectors, $P_v$ is an involutory, i.e. $P_v P_v = I$, and $P_v$ is symmetric, i.e. $P_v^t = P_v$, $\implies P_v$ is orthogonal.

Application – QR decomposition

If we want to “eliminate” all entries of $x$ after 1st one, i.e. $x = (x_1, x_2, \cdots, x_n)^t \xrightarrow{P_v} y = (*, 0, \cdots, 0)^t$, we can find a vector $v$ associated with Householder matrix $P_v$ to make this happen:

$$P_v x = k \hat{e}_1$$

for some multiplier $k$,

i.e. the role $P_v$ plays here is to transfer $x$ onto the axis that $\hat{e}_1$ lies. Apparently, $k$ must be either $-\|x\|$ or $\|x\|$ because $P_v$ is norm-preserving (orthogonal $\Rightarrow$ isometric). So our problem now is: How to determine $v$?

Since we want $P_v x = x - 2v^t x v = x - *v = k \hat{e}_1$, then $v \in \text{span}\{x, \hat{e}_1\}$, WLOG, we can assume that $v = x + \alpha \hat{e}_1$:

$$P_v x = x - \frac{2v^t x}{v^t v} v$$

$$= x - \frac{2(x^t x + \alpha x_1)}{x^t x + 2\alpha x_1 + \alpha^2} (x + \alpha \hat{e}_1)$$

$$= \left(1 - \frac{2(x^t x + \alpha x_1)}{x^t x + 2\alpha x_1 + \alpha^2}\right) x + *\hat{e}_1,$$

and we want $1 - \frac{2(x^t x + \alpha x_1)}{x^t x + 2\alpha x_1 + \alpha^2} = 0 \implies -x^t x + \alpha^2 = 0 \iff \alpha = \pm\|x\|, \implies k = \mp\|x\|$. Hence, if $v := x \pm \|x\| \hat{e}_1$, then $P_v x = k \hat{e}_1 = \mp\|x\| \hat{e}_1$. But, which one to choose? $+$ or $-$? If you write a program to handle this, but a given $x = x_1 \hat{e}_1$, then $v = x + \|x\| \hat{e}_1 = 0$ if $x_1 < 0$, $v = x - \|x\| \hat{e}_1$ may $= 0$ if $x_1 > 0$. To resolve this situation, letting $v := x + \text{sgn}(x_1)\|x\| \hat{e}_1$ is the best choice, where "sgn" agrees with sign and you can define your sgn(0) $= 1$ (or $-1$) consistently. i.e.

we take $v = x + \text{sgn}(x_1)\|x\| \hat{e}_1$,

then $P_v x = -\text{sgn}(x_1)\|x\| \hat{e}_1$.