1. QR faktorizacija

1.1. Produkt Householderovih reflektora i WY reprezentacija

Prisjetimo se, Householderov reflektor zadan je vektorom $u \in \mathbb{R}^m$, $u \neq 0$ koji izgleda ovako

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

Primijetite da vektor u možemo normirati i tako da vrijedi $u^T u = 2$, pa tada možemo pisati Householderove reflektore u jednostavnijoj formi. Budući da slovo u podsjeća na 'unit' (jedinični), onda je uobičajeno za takve u pisati v. Jedna takva Householderova matrica je

$$H = I - vv^T.$$

Sjetite se da smo pri dobivanju QR faktorizacije trebali primjenjivati produkte Householderovih matrica. Neugoda je što produkti Householderovih matrica nisu Householderove matrice. Ipak, postoji način kako efikasno implementirati produkt Householderovih matrica (Bischof i Van Loan, 1987.). To je tzv. WY reprezentacija.

Pretpostavimo da moramo izračunati produkt točno r Householderovih matrica

$$H_i = I - v_i v_i^T,$$

tako da dobijemo

$$Q_r = H_r H_{r-1} \cdots H_1.$$

Tvrdimo da se Q_r može napisati u obliku

$$Q_r = I + W_r Y_r^T, \quad W_r, Y_r \in \mathbb{R}^{m \times r},$$

gdje je Y_r donja trapezoidna matrica. Definiramo rekurziju

$$W_i = [W_{i-1}, -v_i], \quad Y_i = [Y_{i-1}, Q_{i-1}^T v_i],$$

pri čemu je

$$W_1 = -v_1, \quad Y_1 = v_1.$$

Dokaz činjenice je indukcijom po r. Ako je $r = 1$, onda je

$$Q_1 = I + W_1 Y_1^T = I - v_1 v_1^T = H_1.$$

Korak indukcije. Pretpostavimo da je Q_r može napisati kao

$$Q_r = I + W_r Y_r^T.$$

Tada vrijedi

$$\begin{aligned} Q_{r+1} &= H_{r+1}Q_r = (I - v_{r+1}v_{r+1}^T)Q_r = Q_r - v_{r+1}v_{r+1}^TQ_r \\ &= I + W_rY_r^T - v_{r+1}v_{r+1}^TQ_r. \end{aligned} \tag{1.1}$$

S druge strane, iz definicije

$$W_{r+1} = [W_r, -v_{r+1}], \quad Y_{r+1} = [Y_r, Q_r^Tv_{r+1}],$$

slijedi

$$W_{r+1}Y_{r+1}^T = [W_r, -v_{r+1}] \cdot \begin{bmatrix} Y_r^T \\ v_{r+1}^TQ_r \end{bmatrix} = W_rY_r^T - v_{r+1}v_{r+1}^TQ_r. \tag{1.2}$$

Uvrstimo li (1.2) u (1.1), izlazi da je

$$Q_{r+1} = I + W_{r+1}Y_{r+1}^T,$$

što smo i htjeli pokazati.

1.2. QR faktorizacija reflektorima

Neka je

$$A = [a_1, a_2, \ldots, a_n], \quad a_k \in \mathbb{R}^m,$$

tj. neka su a_k stupci matrice A.

Standardna ideja: napredovanje po stupcima kako množimo reflektorima. Neka je $A_0 = A$, onda imamo

$$A_k = H_kA_{k-1}, \quad k = 1, \ldots, n.$$

Algoritam bi tad izgledao ovako (takav je, recimo bio u LINPACK-u)

for $k = 1, n$
  izračunaj v_k tako da a_k postane c_ke_k,
  (e_k k-ti vektor kanonske baze)
  transformiraj a_k
  for $j = k + 1, n$
    $a_j = a_j - 2(v_k^Ta_j)v_k$
  end
end

Međutim, mana ovog algoritma je što k-ti stupac točno k puta dovlači u memoriju, što je sporo.

Algoritam se može i bolje organizirati, na sljedeći način.

for $k = 1, n$
  /* **izračunaj** $a_k = H_{k-1} \cdots H_1a_k$*/
  for $j = 1, k - 1$
    $a_k = a_k - 2(v_j^Ta_k)v_j$
  end
  izračunaj v_k tako da a_k postane c_ke_k,

(e_k k-ti vektor kanonske baze)
end

Primijetite da u ovom algoritmu unutarnja petlja svaki stupac a_k dovlači u memoriju točno jednom i nad njim obavlja sve prethodno izračunate transformacije, tj. unutarnja pretlja, zapravo računa $Q_{k-1}a_k$.

Ovako zapisan algoritam ima puno BLAS-1 operacija, jer se u unutarnjoj petlji operacija `xAXPY`, tj. računaju se produkti oblika

$$y := \alpha x + y.$$

1.3. Kompaktna WY reprezentacija

Ni forma s WY reprezentacijom nije optimalna, jer za spremanje Y i W nam treba dva polja dimenzija $m \times r$. U kompaktnoj verziji to se može smanjiti na samo jedno polje veličine $m \times r$.

Ideja je matricu Q umjesto u obliku $Q = I + WY^T$, gdje su W i Y dimenzija $m \times r$ napisati u obliku

$$Q = I + YTY^T, \tag{1.3}$$

gdje je $Y \in \mathbb{R}^{m\times r}$ donja trapezoidna, a $T \in \mathbb{R}^{r\times r}$ gornjetrokutasta, pa zajedno stanu u jedno polje. Relaciju (1.3) zovemo kompaktna WY reprezentacija.

Da bismo napravili efikasnu QR faktorizaciju bogatu matričnim operacijama, pretpostavimo da je zadana matrica $A \in \mathbb{R}^{m\times n}$, gdje je $n = rN$. Matricu ćemo dijeliti u blokove veličine r, pa ćemo dobiti blok verziju QR faktorizacije.

/* **blok QR faktorizacija** */
for $k = 1, N$
  /* **izračunaj indeks početka bloka** */
  $s = (k-1)r + 1$
  odredi blok-reflektor Q tako da je
  $Q^T A(s:m, s:s+r-1) = C$, pri čemu $C(r+1:m,:) = 0$
  primijeni blok reflektor na preostale retke i stupce u A
  $A(s:m, S:n) = Q^T A(s:m, s:n)$
end

Primijetimo da je ažuriranje matrice $B := A(s:m, s:n)$ bogato matričnim množenjima, ako je Q dan ili kao WY^T ili kao YTY^T. Posljednje ažuriranje onda glasi

$$B := (I + YW^T)B = B + Y(W^T B),$$

što je bogato BLAS-3 operacijama `xGEMM`.

Dakle, u slučaju 'obične' WY reprezentacije, u k-tom koraku W i Y generiraju se iz Householderovih matrica $H_j = I - v_j v_j^T$, za $k = 1, \ldots r$ na sljedeći način

/* **generiranje WY reprezentacije** */
for $j = 1, r$
  if $j = 1$
    $w = [-2v_1], \quad Y = [v_1]$
  else

$z = -2(I + WY^T)v_j, \quad W = [W \ z], \quad Y = [Y \ v_j]$
end if
end

Teorem 1.1 (Kompaktno WY ažuriranje)**.** *Neka je* $Q = I + YTY^T \in \mathbb{R}^{m\times m}$ *ortogonalna matrica i neka je* $Y \in \mathbb{R}^{m\times j}$, $m > j$ *i* $T \in \mathbb{R}^{j\times j}$ *gornjetrokutasta. Neka je* $H = I - 2vv^T$ *Householderova matrica, s tim da je* $v \in \mathbb{R}^m$ *i* $\|v\|_2 = 1$*. Ako označimo*

$$Q_+ = QH,$$

tada je

$$Q_+ = I + Y_+T_+Y_+^T,$$

gdje je

$$Y_+ = [Y \ v] \in \mathbb{R}^{m\times(n+1)}$$

i

$$T_+ = \begin{bmatrix} T & z \\ 0 & \rho \end{bmatrix},$$

gdje je $\rho = -2$ *i* $z = -2TY^Tv$*.*

Dokaz. Izračunajmo tražene veličine

$$I + Y_+T_+Y_+^T = I + \begin{bmatrix} Y & v \end{bmatrix} \begin{bmatrix} T & z \\ 0 & \rho \end{bmatrix} \begin{bmatrix} Y^T \\ v^T \end{bmatrix} = I + YTY^T + Yzv^T + \rho vv^T.$$

To je jednako

$$Q_+ = QH = (I + YTY^T)(I - 2vv^T) = I + YTY^T - 2YTY^Tvv^T - 2vv^T$$

kad je $\rho = -2$ i $z = -2TY^Tv$. □

Konačno, napišimo algoritam generacije matrica Y i T
/* **generiranje YT***/
for $j = 1, r$
if $j = 1$
$Y = [v_1], \quad T = [-2]$
else
$z = -2TY^Tv_j, \quad Y = [Y \ v_j],$

$$T = \begin{bmatrix} T & z \\ 0 & -2 \end{bmatrix}$$

end if
end
Uočimo da se onda matrica $B = Q^TB$ može izračunati kao

$$B = (I + YT^TY^T)B = B + (Y \cdot T^T)(Y^T \cdot B) = B + Y(T^T(Y^T \cdot B)).$$

Ispravni izbor formule ovisi o dimenzijama matrica.