1 GMRES VS. FOM

See notes.

The major point is that FOM is like CG in that it solves a linear system based on the truncated Arnoldi factorization.

The large scale study by Peter Brown (http://dx.doi.org/10.1137/0912003) concluded: there is little difference, but liked the minimum residual property of GMRES.

2 EFFICIENT GMRES

Recall the prototype-GMRES method.

Given $A, b$ where we can only multiply by $A$.

for $i=1$ to maxiter
  Update the Arnoldi factorization $Q_k, H_{k+1}$.
  Solve for $z_k$ by minimizing $\| H_{k+1} z_k - [b] e_1 \|_2$.
  i.e. $z_k = \text{argmin} \| H_{k+1} z_k - [b] e_1 \|_2$
  Let $x_k = Q_k z_k$.
  Check $\parallel A x_k - b \parallel_2$
end

To implement this, we need to solve a least squares problem at each step. This takes $O(k^2)$ work because it's a Hessenberg matrix. Then we need to construct the solution and check the residual. These take $O(nk)$ and another matrix-vector product. We can do all of these steps more efficiently!

Here is the outline for the essential idea to optimize GMRES.

we only need to check the residual at each step, and do not need to compute $x_k$.

So the method we’ll look at optimizing is:

Given $A, b$ where we can only multiply by $A$.

for $i=1$ to maxiter
  Update the Arnoldi factorization $Q_k, H_{k+1}$.
  Compute $\| r_k \|$ where
  
  $r_k = A x_k - b$
  $x_k = Q_k z_k$
  $z_k = \text{argmin} \| H_{k+1} z_k - [b] e_1 \|_2$

  and stop once $\| r_k \|$ is sufficiently small, i.e.
  
  Update $\| r_k \| \rightarrow \| r_{k+1} \|$ and stop if it’s small enough.

Explicitly compute $z_k = \text{argmin} \| H_{k+1} z_k - [b] e_1 \|_2$

and return $x_k = Q_k z_k$ only at the end of the iteration

2.1 THE OPTIMIZATION IDEA

Let’s study the quantity we want to compute, let $\| b \| = \beta_0$, then

$\| r_k \| = \| b - A x_k \| = \| b - A Q_k y_k \| = \| H_k y_k - \beta_0 e_1 \|.$

After a four steps, this is:

$$
\begin{pmatrix}
\ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
y_4 \\
\vdots \\
\beta_0 1 \\
0 0 0
\end{pmatrix} = H_k
$$
We solve least squares problems via QR, so suppose that
\[ H_k = U_k R_k \]
is the QR factorization after \( k \)-steps. Then
\[ \| r_k \| = \| U_k R_k y_k - \beta_0 e_1 \| = \| R_k y_k - \beta_0 U_k^T e_1 \|. \]
Showing this after a few steps gives us the idea more clearly:
\[ \| r_k \| = \| \begin{bmatrix} 0 & \cdots & 0 & y_4 & \cdots & 0 \end{bmatrix} U_4 = \beta_0 y_5. \]
(Remember we solve for \( y_k \) such that this term is zero in the first four components. So we just need to figure out what \( y_5 \) is to get \( \| r_k \| \).

### 2.2 Taking It Deeper

We need to note two things here to continue our optimization:

1. We only need Givens rotations to get \( H_k \rightarrow R_k \).
2. We only need one rotation to update \( R_k \rightarrow R_{k+1} \). (Woah!)

Let’s review step 1 and see how that will help us with step 2.

Now, suppose we have \( U_4 \) and \( R_4 \), how do we get \( U_5, R_5 \)?

\[
H_5 = \begin{bmatrix} H_4 & h \\ 0 & h_{6,5} \end{bmatrix} = \begin{bmatrix} x & \cdots & x \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x \\ \vdots \\ x \\ 0 \end{bmatrix}
\]

If we rotate by \( U_4^T \), we get:

\[
\begin{bmatrix} U_4^T & 0 \\ 0 & 1 \end{bmatrix} H_5 = \begin{bmatrix} U_4^T & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} H_4 & h \\ 0 & h_{6,5} \end{bmatrix} = \begin{bmatrix} R_4 & U_4^T z_4 \\ 0 & h_{6,5} \end{bmatrix}
\]

So at this point, we just have the one Givens rotation: \( J_4 \) that we need to do to fix up the element \( h_{6,5} \) and so \( U_5^T = J_5 J_4 \cdots J_1 \), which is just one update.

### 2.3 Seeking Gamma.

Note that the elements of gamma are just the first columns of \( U_k^T \). Let \( g_k = U_k^T e_1 = [y_1 \ y_2 \ \cdots \ y_k]^T \). Then by our previous relationship:

\[ g_{k+1} = U_{k+1}^T e_1 = J_{k+1} U_k^T e_1 = J_{k+1} g_k. \]

But this is weird, because \( J \) is a \( k \times k \) matrix and \( g_k \) is a length \( k \) vector. So what we really mean is

\[ J_{k+1} \begin{bmatrix} g_k \\ 0 \end{bmatrix} \]

where we grew the vector by one element in order to make it work. Note that we don’t need to actually update \( g_k \) even though it should change.
2.4 THE WHOLE ALGORITHM

\[ g = \beta \text{new} \]

for \( k=1 \) to ... 

Update \( Q_k, H_k \)

Let \( \eta_{k+1} = H_{k+1,k} \)

Let \( z_k = H_{k,k,k} \)

Apply \( J_1 \ldots J_{k-1} \) to \( z_k \), and update \( H \)

Create \( J_k \) to eliminate \( \eta_{k+1} \).

Determine \( g_k \) from \( J_k g_{k-1} \) growing by zeros as needed.

If \( g_k(\text{end}) \) is small enough, then stop iterating.

At this point, \( H \) has the factor \( R \), and (if we do keep \( g \) accurate), then \( g \) is the right hand side, so we can just solve \( R_k y_k = g_k \) and then output \( x = Q_k y_k \).