Optimization: Homework 4

Ben Larson

September 29, 2016

1 Question 1, 5.1

Implementing the conjugate gradient method as in Algorithm 5.2. We used it to solve the Hilbert matrix defined as:

$$A_{i,j} = \frac{1}{i+j-1} \quad \texttt{and} \quad b = (1,1,....1)^T \,|\, with|x_0 = 0$$

for various dimensions. This algorithm works by trying to minimize the residual in the function:

$$\phi(x) = x^T Ax - b^T x$$

For each step:

$$stepsize = \alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$$

Then we update the new x with the usual way we have. By a step length above and direction given:

$$p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$$

This algorithm will converge in n steps. This means that each step/direction is in a conjugate direction. I believe this means that it makes a maximum move in the respective dimension that minimizes the projection of that vector in the subspace of overall function. So each step is a reduction on that dimension.

2 Question 2, 5.2

Show that non-zero vectors $p_0, p_1, ...p_n$ satisfy:

$$p_i^T A p_j = 0 \quad \texttt{for all} \quad i \neq j$$

where A is symmetric positive definite, then these vectors are linearly independent.
We know if the dot product of two vectors, $p_i^T p_j = 0$ we have orthogonal vectors. If we say that a set of vectors $p_1, p_2...p_n$ can be written linearly independent if $\sum_{i=1}^n \alpha_i x_i = 0$ implies that α_k=0. This is saying the only way to represent 0 is by setting α =0. In a similar saying, we can say that the only way for any set of vectors to have 0 inner product is if they are orthogonal vectors. $\sum \alpha_i p_i^T A p_j = 0$ only if α_i=0.

3 Question 3, 5.11

When applied to a quadratic function, both PR and HS reduce down to FR.

$$FR = B_{k+1} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$$

$$PR = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{||\nabla f_k||^2}$$

$$HS = \frac{\nabla_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{(\nabla f_{k+1} - \nabla f_k)^T p_k}$$

1. Starting with Polak-Riviere is the easiest to show. We know that for the conjugate gradient methods we get orthogonal vectors with each step. So any vector in v_{k+1} will be orthogonal to v_k. This indicates that $v^T v = 0$.

$$PR = \frac{\nabla f_{k+1}^T(\nabla f_{k+1} - \nabla f_k)}{||\nabla f_k||^2} = \frac{\nabla f_{k+1}^T \nabla f_{k+1} - \nabla f_k \nabla f_{k+1}^T)}{\nabla f_k^T \nabla f_k} = \frac{\nabla f_{k+1}^T \nabla f_{k+1} - 0}{\nabla f_k^T \nabla f_k}$$

And we end up with:

$$\frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$$

2. This one I'm not completely sure how to reduce down to the Feltcher Reeves. Starting with:

$$HS = \frac{\nabla_{k+1}^T(\nabla f_{k+1} - \nabla f_k)}{(\nabla f_{k+1} - \nabla f_k)^T p_k} = \frac{\nabla f_{k+1}^T \nabla f_{k+1} - 0}{(\nabla f_{k+1} - \nabla f_k)^T p_k} = ??$$

The consecutive search directions are conjugate with respect to the average hessian on the interval $[x_k, x_{k+1}]$.

$$G_k = \int \nabla^2 f(x_k + \tau \alpha_k p_k) d\tau$$

$$\nabla f_{k+1} = \nabla f_k + \alpha_k G_k p_k$$

$$p_{k+1} = -\nabla f_{k+1} + \beta_{k+1} p_k$$

The condition $p_{k+1}^T G_k p_k = 0$ requires that β_k be given by above. Substitute in ∇f_{k+1} to the denominator.

$$\frac{\nabla f_{k+1}^T \nabla f_{k+1} - 0}{(\nabla f_{k+1} - \nabla f_k)^T p_k} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{(\nabla f_k + \alpha_k G_k p_k - \nabla f_k)^T p_k} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{(\alpha_k G_k p_k)^T p_k}$$

I haven't yet thought of a way to further reduce this.

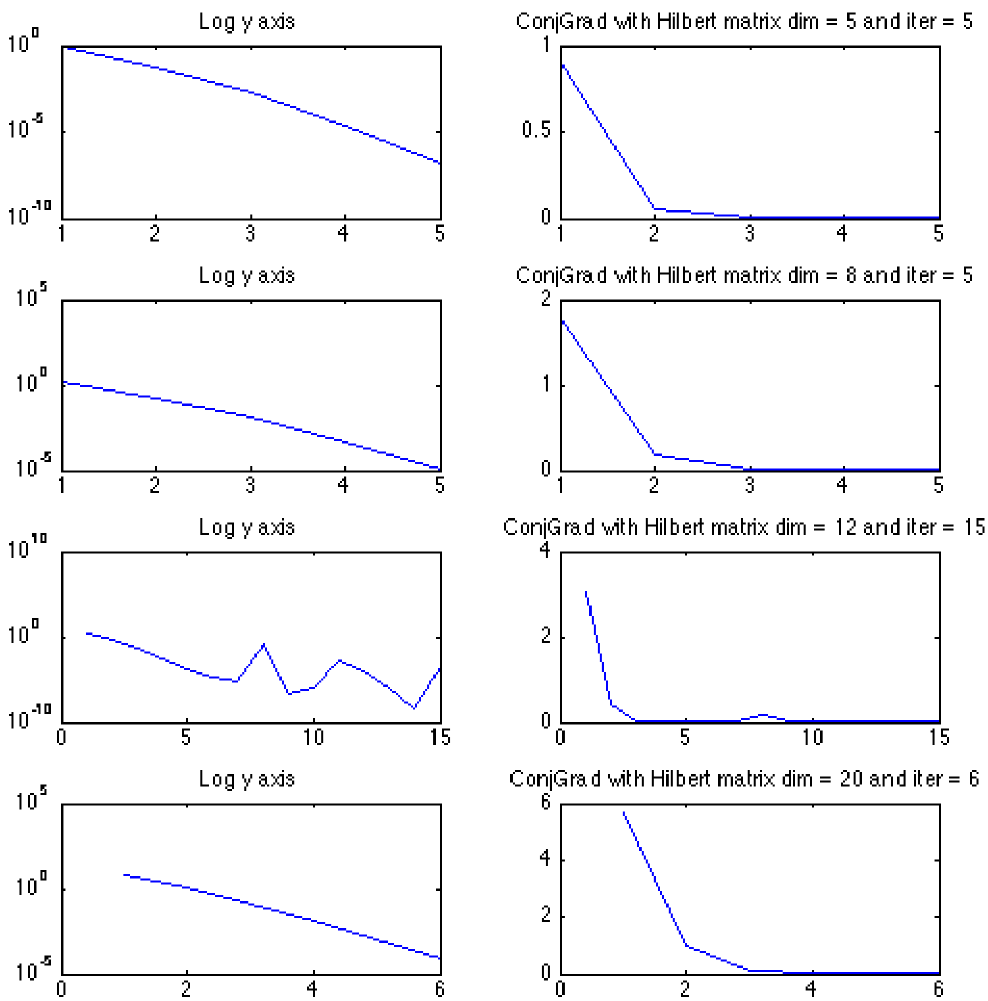

Figure 1: Plots of the conjugate gradient minimization on the Hilbert matrix. I tried with dimensions 5,8,12, and 20. Dim = [5, 8] we have the same rate of convergence. Convergence was defined as the residual error less that 10^{-6}.