Scientific Computing HW 8

Ryan Chen

October 23, 2024

1. To show that the algorithms are equivalent, we rewrite $\alpha_k, r_{k+1}, \beta_{k+1}$. Rewrite α_k as

$$\begin{aligned}
\alpha_k &= -\frac{r_k^T p_k}{p_k^T A p_k} \\
&= -\frac{r_k^T(-r_k + \beta_k p_{k-1})}{p_k^T A p_k} && p_k = -r_k + \beta_k p_{k-1} \\
&= \frac{r_k^T r_k}{p_k^T A p_k} && r_k^T p_{k-1} = 0 \text{ by Theorem 5.2 in [NW]}
\end{aligned}$$

Rewrite r_{k+1} as

$$\begin{aligned}
r_{k+1} &= Ax_{k+1} - b \\
&= A(x_k + \alpha_k p_k) - b \\
&= Ax_k - b + \alpha_k A p_k \\
&= r_k + \alpha_k A p_k
\end{aligned}$$

The expressions for α_k, r_{k+1} give

$$Ap_k = \frac{r_{k+1} - r_k}{\alpha_k} = \frac{p_k^T A p_k (r_{k+1} - r_k)}{r_k^T r_k}$$

Use this to rewrite β_{k+1} as

$$\begin{aligned}
\beta_{k+1} &= \frac{r_{k+1}^T A p_k}{p_k^T A p_k} \\
&= \frac{r_{k+1}^T (r_{k+1} - r_k)}{r_k^T r_k} \\
&= \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} && r_{k+1}^T r_k = 0 \text{ by Theorem 5.3 in [NW]}
\end{aligned}$$

2. As a preliminary, define the Krylov subspaces (the first definition is a natural extension to $k = 0$)

$$\mathcal{K}_0 := \{0\}, \quad \mathcal{K}_k = \mathcal{K}(A, r, k) := \operatorname{span}\left\{A^i r : 0 \leq i \leq k-1\right\}, \ k \geq 1$$

We will divide the proof into the following claims:

- Claim 1: We can define the greatest integer k' such that $\dim \mathcal{K}_{k'} = k'$.
Observe that $\dim \mathcal{K}_0 = \dim\{0\} = 0$. Also observe that A has n rows, so that for all k' we have that $\mathcal{K}_{k'}$ is a subspace of $\mathbb{R}^n$ hence $\dim \mathcal{K}_{k'} \leq n$. Thus there is a nonzero finite number of integers k' satisfying $\dim \mathcal{K}_{k'} = k'$, so that we can pick k' to be the greatest such integer.
- Claim 2: $\mathcal{K}_p = \mathcal{K}_{k'}$ for all $p \geq k'+1$. We will prove this by induction.
Base case $p = k'+1$: We have $\dim \mathcal{K}_{k'+1} \leq k'+1$, and by Claim 1 we have $\dim \mathcal{K}_{k'+1} \neq k'+1$, hence
$$\dim \mathcal{K}_{k'+1} < k'+1 \implies \dim \mathcal{K}_{k'+1} \leq k' \implies \dim \mathcal{K}_{k'+1} \leq \dim \mathcal{K}_{k'}$$
Since $\mathcal{K}_{k'}$ is a subspace of $\mathcal{K}_{k'+1}$, we have $\dim \mathcal{K}_{k'} \leq \dim \mathcal{K}_{k'+1}$. Thus $\dim \mathcal{K}_{k'} = \dim \mathcal{K}_{k'+1}$, and using again the fact $\mathcal{K}_{k'}$ is a subspace of $\mathcal{K}_{k'+1}$, we have $\mathcal{K}_{k'} = \mathcal{K}_{k'+1}$.
Inductive step: Fix $p \geq k'+1$ and assume $\mathcal{K}_p = \mathcal{K}_{k'}$. This gives
$$A^i r \in \mathcal{K}_{k'}, \ 0 \leq i \leq p-1 \quad (2.1)$$
It remains to show $A^p r \in \mathcal{K}_{k'}$ so that we can conclude $\mathcal{K}_{p+1} = \mathcal{K}_{k'}$. Setting $i = p-1$ in (2.1) gives $A^{p-1} r \in \mathcal{K}_{k'}$, so $A^{p-1} r = \sum_{i=0}^{k'-1} c_i A^i r$ for some scalars c_i. Then
$$A^p r = \sum_{i=0}^{k'-1} c_i A^{i+1} r = \sum_{i=1}^{k'} c_{i-1} A^i r = \sum_{i=1}^{k'-1} c_{i-1} A^i r + c_{k'-1} A^{k'} r$$
The RHS sum over i is in $\mathcal{K}_{k'}$ by definition of $\mathcal{K}_{k'}$, and the last RHS term is in $\mathcal{K}_{k'}$ by setting $i = k'$ in (2.1), thus $A^p r \in \mathcal{K}_{k'}$.
- Claim 3: If $r = By$ for some $y \in \mathbb{R}^k$ then $k = k'$. We conclude that $\mathcal{K}_p = \mathcal{K}_k$ for all $p \geq k+1$.
We aim to show that $k' \leq k$ and $k' \geq k$. To show that $k' \leq k$, consider that for all $i \geq 0$,
$$A^i r = A^i B y = B C^i y \in \operatorname{col} B$$
Thus $\mathcal{K}_{k'}$ is a subspace of $\operatorname{col} B$. This along with $\operatorname{rank} B = k$ and Claim 1 gives $k' \leq k$.
To show that $k' \geq k$, suppose for the sake of contradiction that $k' < k$. This along with Claim 2 gives $\mathcal{K}_k = \mathcal{K}_{k'}$, hence $\dim \mathcal{K}_k = k'$. This means that $A^i r, \ 0 \leq i \leq k-1$ are linearly dependent, so there exist scalars α_i, not all zero, such that $\sum_{i=0}^{k-1} \alpha_i A^i r = 0$. Let B have columns $b_j, \ 1 \leq j \leq k$ and rewrite the LHS,
$$\sum_{i=0}^{k-1} \alpha_i A^i r = \sum_{i=0}^{k-1} B \underbrace{\alpha_i C^i y}_{=: z_i} = \sum_{i=0}^{k-1} B z_i = \sum_{i=0}^{k-1} \sum_{j=1}^{k} z_{ij} b_j = \sum_{j=1}^{k} \left[\sum_{i=0}^{k-1} z_{ij}\right] b_j$$
Since the b_j's are linearly independent,
$$\sum_{i=0}^{k-1} z_{ij} = 0, \ 1 \leq j \leq k \quad (2.2)$$
The LHS of (2.2) is a multivariate polynomial in terms of the entries of C that is nonzero (the α_i's are not all zero) and has degree at most $k-1$. But (2.2) implies that it has at least k distinct roots, so its degree is at least k, a contradiction. We conclude that $k' \geq k$.

3. (a) Define $Q \in \mathcal{P}_{k+1}$ by

$$Q(\lambda) := C\left[\frac{1}{2}(\lambda_1 + \lambda_{n-k}) - \lambda\right] \prod_{i=n-k+1}^{n} (\lambda_i - \lambda)$$

To find C we impose $Q(0) = 1$.

$$1 = Q(0) = \frac{C}{2}(\lambda_1 + \lambda_{n-k}) \prod_{i=n-k+1}^{n} \lambda_i \implies C = \frac{1}{\frac{1}{2}(\lambda_1 + \lambda_{n-k})} \prod_{i=n-k+1}^{n} \frac{1}{\lambda_i}$$

Hence

$$Q(\lambda) = \frac{\frac{1}{2}(\lambda_1 + \lambda_{n-k}) - \lambda}{\frac{1}{2}(\lambda_1 + \lambda_{n-k})} \prod_{i=n-k+1}^{n} \left(1 - \frac{\lambda}{\lambda_i}\right)$$

(b) Factor theorem: Given $R \in \mathcal{P}_k$, we have $R(\lambda) = (\lambda - \lambda_0)P(\lambda)$ for some $P \in \mathcal{P}_{k-1}$ iff $R(\lambda_0) = 0$. From $Q(0) - 1 = 0$ and the theorem, $Q(\lambda) - 1 = \lambda P(\lambda)$ for some $P \in \mathcal{P}_k$, hence $P(\lambda) = \frac{Q(\lambda)-1}{\lambda}$.

(c) Using $P \in \mathcal{P}_k$ from part (b),

$$\min_{P_k \in \mathcal{P}_k} \max_{1 \le i \le n} [1 + \lambda_i P_k(\lambda_i)]^2 \le \max_{1 \le i \le n} [1 + \lambda_i P(\lambda_i)]^2 = \max_{1 \le i \le n} Q^2(\lambda_i)$$

Plug this into the ansatz to get

$$\|x_{k+1} - x^*\|_A^2 \le \max_{1 \le i \le n} Q^2(\lambda_i) \|x_0 - x^*\|_A^2$$

(d) From part (a),

$$Q^2(\lambda) = \left(\frac{\lambda - \frac{1}{2}(\lambda_1 + \lambda_{n-k})}{\frac{1}{2}(\lambda_1 + \lambda_{n-k})}\right)^2 \prod_{i=n-k+1}^{n} \left(1 - \frac{\lambda}{\lambda_i}\right)^2$$

We have $Q^2(\lambda_i) = 0$ for $n - k + 1 \le i \le n$, and the assumption $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ gives

$$Q^2(\lambda_i) \le \left(\frac{\lambda - \frac{1}{2}(\lambda_1 + \lambda_{n-k})}{\frac{1}{2}(\lambda_1 + \lambda_{n-k})}\right)^2, \ 1 \le i \le n - k$$

Together, these computations give

$$\max_{1 \le i \le n} Q^2(\lambda_i) \le \max_{1 \le i \le n-k} \left(\frac{\lambda_i - \frac{1}{2}(\lambda_1 + \lambda_{n-k})}{\frac{1}{2}(\lambda_1 + \lambda_{n-k})}\right)^2 \le \max_{\lambda \in [\lambda_1, \lambda_{n-k}]} \left(\frac{\lambda - \frac{1}{2}(\lambda_1 + \lambda_{n-k})}{\frac{1}{2}(\lambda_1 + \lambda_{n-k})}\right)^2$$

This along with the result of part (c) gives

$$\|x_{k+1} - x^*\|_A^2 \le \max_{\lambda \in [\lambda_1, \lambda_{n-k}]} \left(\frac{\lambda - \frac{1}{2}(\lambda_1 + \lambda_{n-k})}{\frac{1}{2}(\lambda_1 + \lambda_{n-k})}\right)^2 \|x_0 - x^*\|_A^2$$

(e) The function

$$f(\lambda) := \left(\frac{\lambda - a}{a}\right)^2, \quad a := \frac{1}{2}(\lambda_1 + \lambda_{n-k})$$

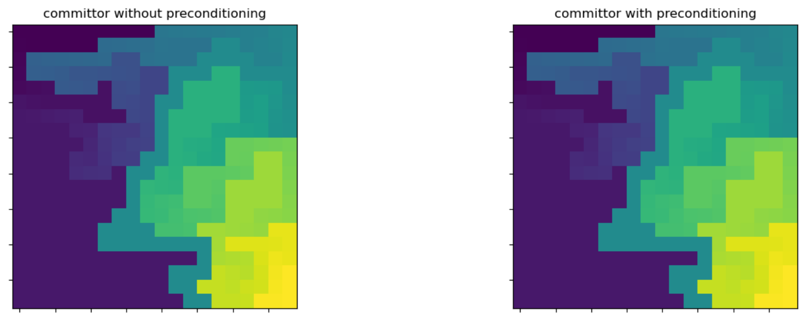
is a concave up parabola which is even wrt $\lambda = a$. With a being the midpoint between λ_1 and λ_{n-k}, we see that f attains its maximum on the interval $[\lambda_1, \lambda_{n-k}]$ at the endpoints λ_1 and λ_{n-k}.

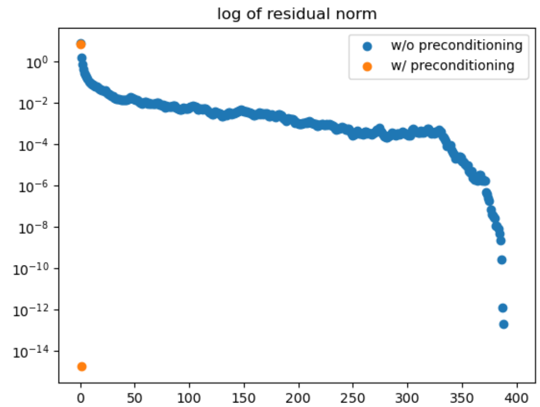
$$\max_{\lambda \in [\lambda_1, \lambda_{n-k}]} f(\lambda) = f(\lambda_1) = \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1}\right)^2$$

(f) From parts (d) and (e),

$$\|x_{k+1} - x^*\|_A^2 \le \left(\max_{\lambda \in [\lambda_1, \lambda_{n-k}]} f(\lambda)\right) \|x_0 - x^*\|_A^2 = \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1}\right)^2 \|x_0 - x^*\|_A^2$$

4. Code: `https://github.com/RokettoJanpu/scientific-computing-1-redux/blob/main/hw8/hw8.ipynb`





CG with conditioning only took one iteration, resulting in only two plotted points.