1 March 16th, 2021

1.1 CG as a Direct Method

As proved before, GC will get the exact solution after t most n steps. In addition, the complexity per step is:

$$1 \text{ matrix-vector product } + \text{opeartions of } O(n)$$

Note that one matrix vector product is $O(m+n)$ where m is the number of nonzero entries in A. This means that the total computational cost is $O(mn+n^2)$ in the worse case.

- If A is the 1D Discrete Laplacian matrix, this is no better than Cholesky decomposition, which is $O(n)$.
- However if A is the 2D Discrete Laplacian, both are $O(n^2)$.

1.2 GC as an Iterative Method

CG can give a very accurate solution even if $k \ll n$.

Theorem 1.1

Assume A is SPD. Then $\{x_k\}$ generated by CG satisfies:

1. If A has only s distinct eigenvalues, then:
$$x_k = x_* \text{for all } k \geq s.$$
2. For a genera A: Let $\gamma = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$ be the condition number, then we have:
$$\|x_k - x_*\|_A \leq 2\left(\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1}\right)^k \|x_0 - x_*\|_A.$$
3. If eigenvalues of A satisfies:
$$0 < \lambda_1 \leq \ldots \leq \lambda_s \leq \alpha \leq \lambda_{s+1} \leq \ldots \leq \lambda_{n-t} \leq \beta \leq \lambda_{n-t+1} \leq \ldots \leq \lambda_n$$
Where α is close to β, (i.e. most eigenvalues are close together barring s small and t large outlying eigenvalues), then:
$$\|x_k - x_*\|_A \leq 2\left(\frac{\sqrt{\beta/\alpha}-1}{\sqrt{\beta/\alpha}+1}\right)^{k-s-t} \left(\max_{\lambda\in[\alpha,\beta]} \prod_{\ell\in\{1,\ldots,t\}\cup\{n-t+1,\ldots,n\}} \left|\frac{\lambda-\lambda_\ell}{\lambda_\ell}\right|\right)$$
Note that the right factor is a constant.

Corollary 1.2

From Theorem 1.1 (2), we have that the convergence speed depends on $O(\sqrt{\gamma})$, where as for steepest descent, it is $O(\gamma)$, meaning that the CG is much faster than steepest descent.

Example 1.3

If $A = (I + vv^T)$, then there are only two distinct eigenvalues, meaning that CG will converge in only two steps.

Proof. By the optimality of CG, we have:

$$\begin{aligned}
\|x_k - x_*\|_A &= \min_{x \in x_0 + K_k} \|x_* - x\|_A \\
&= \min_{c \in \mathbb{R}^k} \left\| x_* - (x_0 + \sum_{j=0}^{k-1} c_j A^j r_0) \right\|_A \\
&= \min_{c \in \mathbb{R}^k} \left\| (x_* - x_0) + \sum_{j=0}^{k-1} c_j A^{j+1} (x_* - x_0) \right\|_A \\
&= \min_{c \in \mathbb{R}^k} \left\| \left(I + \sum_{j=1}^{k} c_{j-1} A^j \right) (x_* - x_0) \right\|_A \\
&= \min_{p \in \mathbb{P}_k, p(0)=1} \|p(A)(x_* - x_0)\|_A \\
&\leq \left(\min_{p \in \mathbb{P}_k, p(0)=1} \|p(A)\|_A \right) \|(x_* - x_0)\|_A \\
&= \left(\min_{p \in \mathbb{P}_k, p(0)=1} \|p(A)\|_2 \right) \|(x_* - x_0)\|_A .
\end{aligned}$$

Where $\mathbb{P}_k$ is the set of polynomial of degree k.

Since A is symmetric, $p(A)$ is also symmetric. Thus, we have:

$$\begin{aligned}
\|x_k - x_*\|_A &\leq \left(\min_{p \in \mathbb{P}_k, p(0)=1} \|p(A)\|_2 \right) \|(x_* - x_0)\|_A \\
&= \left(\min_{p \in \mathbb{P}_k, p(0)=1} \max_{i \in \{1,\ldots,n\}} |p(\lambda_i)| \right) \|(x_* - x_0)\|_A .
\end{aligned}$$

1. If A has only s distinct eigenvalues, say $\lambda_1, \ldots, \lambda_s$, we have:

$$\min_{p \in \mathbb{P}_k, p(0)=1} \max_{i \in \{1,\ldots,n\}} |p(\lambda_i)| \leq \max_{i \in \{1,\ldots,n\}} |q(\lambda_i)| \quad \forall q \begin{cases} q \in \mathbb{P}_k \\ q(0) = 1 \end{cases}$$

   Let us choose q by:

$$q(\lambda) = \prod_{i=1}^{s} \left(\frac{\lambda_i - \lambda}{\lambda_i} \right)$$

We have check that $q \in \mathbb{P}_s \subset \mathbb{P}_k$ and that $q(0) = 1$. With this, we have:

$$\begin{aligned}\min_{p\in\mathbb{P}_k, p(0)=1} \max_{i\in\{1,\ldots,n\}} |p(\lambda_i)| &\leq \max_{i\in\{1,\ldots,n\}} |q(\lambda_i)| \\ &= \max_{i\in\{i,\ldots,s\}} |q(\lambda_i)| = 0.\end{aligned}$$

2. We relax the estimation by:

$$\begin{aligned}\|x_k - x_*\|_A &\leq \left(\min_{p\in\mathbb{P}_k, p(0)=1} \max_{i\in\{1,\ldots,n\}} |p(\lambda_i)|\right) \|(x_* - x_0)\|_A \\ &\leq \left(\min_{p\in\mathbb{P}_k, p(0)=1} \max_{\lambda\in[\lambda_{\min},\lambda_{\max}]} |p(\lambda)|\right) \|(x_* - x_0)\|_A .\end{aligned}$$

Now we use a change of variable to estimate $\min\max|p(\lambda)|$. Define:

$$\mu = 2\frac{\lambda - \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}} - 1.$$

i.e. $\lambda = \lambda_{\min} \implies \mu = -1$, $\lambda = \lambda_{\max} \implies \mu = 1$. Thus, we estimate:

$$\min_{p\in\mathbb{P}_k, p(\mu_0)=1} \max_{\mu\in[-1,1]} |p(\mu)|$$

where $\mu_0 = 2\frac{-\lambda_{\min}}{\lambda_{\max}-\lambda_{\min}} - 1 = -\frac{\lambda_{\max}+\lambda_{\min}}{\lambda_{max}-\lambda_{\min}}$.

The solution of the minimax is given by the **Chebychev polynomial**.

Lemma 1.4

If $\mu_0 \neq [-1, 1]$, then:

$$\frac{C_k(\mu)}{C_k(\mu_0)} = \arg\min_{p\in\mathbb{P}_k, p(\mu_0)=1} \max_{\mu\in[-1,1]} |p(\mu)|$$

where:

$$C_k(\mu) = \begin{cases} \cos(k \cdot \arccos(\mu)) & \mu \in [-1, 1] \\ \cosh(k \cdot \operatorname{arccosh}(\mu)) & \mu \geq 1 \\ (-1)^k \cosh(k \cdot \operatorname{arccosh}(-\mu)) & \mu \leq 1 \end{cases}$$

Proof. First we check that $C_k \in \mathbb{P}_k$. Indeed

$$\begin{aligned}C_0(\mu) &= 1 \in \mathbb{P}_0 \\ C_1(\mu) &= \mu \in \mathbb{P}_1.\end{aligned}$$

Also, by:

$$\begin{cases} \cos((k+1)\theta) + \cos((k-1)\theta) = 2\cos\theta\cos(k\theta) \\ \cosh((k+1)\theta) + \cosh((k-1)\theta) = 2\cosh\theta\cosh(k\theta) \end{cases}$$

Choosing $\theta = \arccos\mu$ if $|\mu| \leq 1$ or $\arccos h|\mu|$ if $|\mu| \geq 1$ and $k = k+1$, we have:

$$\begin{aligned}C_k(\mu) + C_{k-2}(\mu) &= 2\mu C_{k-1}(\mu) \\ \implies C_k(\mu) &= 2\mu C_{k-1}(\mu) - C_{k-2}(\mu) \in \mathbb{P}_k.\end{aligned}$$

$\square$

$\square$