1 March 23rd, 2021

1.1 Convergence of Preconditioned Steepest Descent

We have:

$$\begin{aligned}
\|x_{k+1}-x_*\|_A &= \min_{x\in x_k+\operatorname{span}\{P^{-1}r_k\}} \|x-x_*\|_A \\
&= \min_{\alpha\in\mathbb{R}} \|x_k+\alpha P^{-1}(b-Ax_k)-x_*\|_A \\
&= \min_{\alpha\in\mathbb{R}} \|x_k-\alpha P^{-1}A(x_k-x_*)-x_*\|_A \\
&= \min_{\alpha\in\mathbb{R}} \|(I-\alpha P^{-1}A)(x_k-x_*)\|_A \\
&= \min_{p\in\mathbb{P}_1, p(0)=1} \|p(P^{-1}A)(x_k-x_*)\|_A \\
&\le \min_{p\in\mathbb{P}_1, p(0)=1} \|p(P^{-1}A)\|_A\cdot\|x_k-x_*\|_A.
\end{aligned}$$

As we have proven before, we have:

$$\begin{aligned}
\|p(P^{-1}A)\|_A &= \|A^{-\frac{1}{2}}p(P^{-1}A)A^{-\frac{1}{2}}\|_2 \\
&= \|p(A^{\frac{1}{2}}P^{-1}A^{\frac{1}{2}})\|_2 \\
&= \max_{i=1,\dots,n}\left|p(\lambda_i(A^{\frac{1}{2}}P^{-1}A^{\frac{1}{2}}))\right|.
\end{aligned}$$

Since:

$$\{\lambda_i(AB): i\in 1,\dots,n\}\cup\{0\} = \{\lambda_i(BA): i\in 1,\dots,n\}\cup\{0\}$$

i.e. the non-zero eigenvalues of AB are the same as BA, we have:

$$\begin{aligned}
\|p(P^{-1}A)\|_A &= \max_{i=1,\dots,n}\left|\lambda_i(A^{\frac{1}{2}}P^{-1}A^{\frac{1}{2}})\right| \\
&= \max_{i=1,\dots,n}\left|p(\lambda_i(P^{-1}A))\right|.
\end{aligned}$$

As such, we have:

$$\begin{aligned}
\|x_{k+1}-x_*\|_A &\le \min_{p\in\mathbb{P}_1, p(0)=1} \|p(P^{-1}A)\|_A\cdot\|x_k-x_*\|_A \\
&\le \min_{p\in\mathbb{P}_1, p(0)=1}\max_{i=1,\dots,n}\left|p(\lambda_i(P^{-1}A))\right|\cdot\|x_k-x_*\|_A \\
&\le \left(\frac{\gamma(P^{-1}A)-1}{\gamma(P^{-1}A)+1}\right)\|x_k-x_*\|_A.
\end{aligned}$$

Where $\gamma(P^{-1}A)=\frac{\lambda_{\max}(P^{-1}A)}{\lambda_{\min}(P^{-1}A)}$

Remark 1.1 — The analysis is similar to the steepest descent, but just with $kP^{-1}A$ instead of A.

By a change of variable:

$$\tilde{r_k}=P^{-\frac{1}{2}}r_k,\quad \tilde{x_k}=P^{\frac{1}{2}}x_k,\quad \tilde{A}=P^{-\frac{1}{2}}AP^{-\frac{1}{2}}, \tilde{b}=P^{-\frac{1}{2}}b$$

if we apply steepest descent to $\tilde{A}\tilde{x} = \tilde{b}$, we have:

$$\begin{aligned}\tilde{r_k} = \tilde{b} - \tilde{A}\tilde{r_k} &= P^{-\frac{1}{2}}b - P^{-\frac{1}{2}}AP^{-\frac{1}{2}}P^{-\frac{1}{2}}x_k \\ &= P^{-\frac{1}{2}}(b - AP^{-1}x_k) \\ &= P^{-\frac{1}{2}}r_k.\end{aligned}$$

$$\begin{aligned}\tilde{\alpha_k} &= \frac{\langle \tilde{r_k}, \tilde{r_k}\rangle}{\langle A\tilde{r_k}, \tilde{r_k}\rangle} \\ &= \frac{\left\langle P^{-\frac{1}{2}}r_k, P^{-\frac{1}{2}}r_k\right\rangle}{\left\langle P^{-\frac{1}{2}}AP^{-\frac{1}{2}}P^{-\frac{1}{2}}r_k, P^{-\frac{1}{2}}r_k\right\rangle} \\ &= \frac{\langle P^{-1}r_k, r_k\rangle}{\langle AP^{-1}r_k, P^{-1}r_k\rangle} \\ &= \frac{\langle r_k, d_k\rangle}{\langle Ad_k, d_k\rangle} = \alpha_k.\end{aligned}$$

Thus, we have:

$$\begin{aligned}\tilde{x_{k+1}} &= \tilde{x_k} + \alpha_k\tilde{r_k} \\ P^{\frac{1}{2}}x_{k+1} &= P^{\frac{1}{2}}x_k + \alpha_k P^{-\frac{1}{2}}r_k \\ &= P^{\frac{1}{2}}(x_k + \alpha_k P^{-1}r_k) \\ \iff x_{k+1} &= x_k + \alpha_k P^{-1}r_k = x_k + \alpha_k d_k.\end{aligned}$$

Remark 1.2 — This means that solving the steepest descent for $\tilde{A}\tilde{x} = \tilde{b}$, we have the preconditioned steepest descent for $Ax = b$ with preconditioner P.

Thus the convergence of Preconditioned Steepest Descent depends on:

$$\gamma(\tilde{A}) = \gamma(P^{-1}A)$$

1.2 Finding a Good Preconditioner P

A good preconditioner should satisfy:

- P should be SPD (otherwise the norm constraint will not be satisfied)
- The solution of $Pd = r$ should be easy to compute.
- $P^{-1}A$ should have a small condition number (or, roughly, $P^{-1}A \approx I$ since I has the smallest condition number). This gives us $P \approx A$.

  Remark 1.3 — If $P = A$, then $Pd = r$ is not easy to solve, thus we need to strike a balance.

Example 1.4

We can choose $P = D$, the diagonal of A.

1.3 Preconditioned CG (PCG)

PCG is the most popular algorithm for SPD matrix. In standard CG, we choose:

$$K = \text{span}\{r_k, d_{k-1}\}$$

similarly, in PCG, we choose:

$$K = \text{span}\{P^{-1}r_k, d_{k-1}\}.$$

Remark 1.5 — d_{k-1} is the momentum component $(x_k - x_{k-1})$, thus we still keep the same.

Thus, we have:

$$\begin{cases} d_k = P^{-1}r_k + \beta_k d_{k-1} & \text{where } \beta_k = -\frac{\langle P^{-1}r_k, Ad_{k-1}\rangle}{\langle Ad_{k-1}, d_{k-1}\rangle} \\ x_{k+1} = x_k + \alpha_k d_k & \text{where } \alpha_k = \frac{\langle d_k, r_k\rangle}{\langle Ad_k, d_k\rangle} \end{cases}$$

Remark 1.6 — The analysis is the same, except r_k is replace by $P^{-1}r_k$.

This gives us Algorithm 1.

Algorithm 1 Preconditioned Conjugate Gradient (PCG)

```
1: for k = 0, 1, ... do
2:     r_k = b - A x_k
3:     β_k = -<P^{-1} r_k, A d_{k-1}> / <A d_{k-1}, d_{k-1}>
4:     d_k = P^{-1} r_k + β_k d_{k-1}
5:     α_k = <d_k, r_k> / <A d_k, d_k>
6:     x_{k+1} = x_k + α_k d_k
7: end for
```

We can once again introduce intermediate variable to improve the computational cost, giving us Algorithm 2 by letting $p_k = Ad_k$, $s_k = P^{-1}r_k$. Thus, we have:

$$r_{k+1} = b - Ax_{k+1} = b - A(x_k + \alpha_k d_k) = r_k - \alpha_k p_k$$

This has computational cost:

- 1 mat-vec product of A
- Solve 1 linear equation of P
- Operations of $O(n)$

Algorithm 2 Improved Preconditioned Conjugate Gradient (PCG)

$r_0 = b - Ax_0$
$p_{-1} = 0$
$d_{-1} = 0$
for $k = 0, 1, \ldots$ **do**
  Solve s_k from $Ps_k = r_k$
  $\beta_k = -\frac{\langle s_k, p_{k-1}\rangle}{\langle p_{k-1}, d_{k-1}\rangle}$
  $d_k = s_k + \beta_k d_{k-1}$
  $p_k = Ad_k$
  $\alpha_k = \frac{\langle d_k, r_k\rangle}{\langle p_k, d_k\rangle}$
  $x_{k+1} = x_k + \alpha_k d_k$
  $r_{k+1} = r_k - \alpha_k p_k$
end for

1.4 Convergence of PCG

By change of variable:

$$\tilde{r_k} = P^{-\frac{1}{2}} r_k,\ \tilde{b} = P^{-\frac{1}{2}} b,\ \tilde{x_k} = P^{\frac{1}{2}x_k},\ \tilde{A} = P^{-\frac{1}{2}} A P^{-\frac{1}{2}},\ \tilde{d_k} = P^{\frac{1}{2}} d_k,\ \tilde{x} = P^{\frac{1}{2}} x$$

we will see that:

$$\text{CG for solving } \tilde{A}\tilde{x} = \tilde{b} \iff \text{PCG for solving } Ax = b \text{ with preconditioner } P$$

Moreover:

$$\begin{aligned} \tilde{x_{k+1}} &= \operatorname*{arg\,min}_{\tilde{x} \in \tilde{x_0} + K_k} \|\tilde{x} - \tilde{x_*}\|_{\tilde{A}} \quad \text{where } K_k = \operatorname{span}\{\tilde{r_0}, P^{-\frac{1}{2}} A P^{-\frac{1}{2}} \tilde{r_0}, \ldots\} \\ \iff x_{k+1} &= \operatorname*{arg\,min}_{x \in x_0 + K_{P,k}} \|x - x_*\|_A \quad \text{where } K_{P,k} = \operatorname{span}\{r_0, P^{-1} A r_0, \ldots\}. \end{aligned}$$

Thus the convergence is:

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

Therefore, all of the convergence results for standard CG can be applied:

- $x_k = x_*$ for all $k \ge s$, where s is the number of distinct eigenvalues of $P^{-1}A$.
- $\|x_k - x_*\|_A \le 2\left(\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1}\right)^k \|x_0 - x_*\|_A$, where $\gamma = \frac{\lambda_{\max}(P^{-1}A)}{\lambda_{\min}(P^{-1}A)}$
- If λ_i are eigenvalues of $P^{-1}A$ satisfying:

  $$0 \le \lambda_1 \le \ldots \le \lambda_s \le \alpha \le \lambda_{s+1} \le \ldots \le \lambda_{n-t} \le \beta \le \lambda_{n-t+1} \le \ldots \le \lambda_n$$

  then:

  $$\|x_k - x_*\|_A \le 2\left(\frac{\sqrt{\beta/\alpha}-1}{\sqrt{\beta/\alpha}+1}\right)^{k-s-t} \cdot \max_{\lambda \in [\alpha,\beta]} \left(\prod_{i \in \{1,\ldots,n\} \cup \{n-t+1,\ldots,n\}} \frac{|\lambda - \lambda_i|}{|\lambda_i|}\right) \cdot \|x_0 - x_*\|_A$$

Corollary 1.7

If all eigenvalues of $P^{-1}A$ satisfies:

$$0 \leq \lambda_1 \leq \ldots \leq \lambda_s \leq 1-\delta \leq \lambda_{s+1} \leq \ldots \leq \lambda_{n-t} \leq 1+\lambda \leq \ldots \leq \lambda_n$$

then:

$$\|x_k - x_*\|_A \leq 2\delta^{k-s-t}\left(\frac{1+\delta}{\lambda_{\min}}\right)^s$$

Proof. For any $\lambda_j > 1+\delta$, we have:

$$\max_{\lambda\in[1-\delta,1+\delta]} \frac{|\lambda-\lambda_j|}{|\lambda_j|} \leq 1$$

For any $0 < \lambda_j \leq 1-\delta$, we have:

$$\max_{\lambda\in[1-\delta,1+\delta]} \frac{|\lambda-\lambda_j|}{|\lambda_j|} = \frac{|(1+\delta)-\lambda_j|}{|\lambda_j|} \leq \frac{1+\delta}{\lambda_{\min}}$$

In addition, assuming $\delta > 0$, we have:

$$\frac{\sqrt{\frac{1+\delta}{1-\delta}}-1}{\sqrt{\frac{1+\delta}{1-\delta}}+1} = \frac{\sqrt{\frac{1+\delta}{1-\delta}}-1}{\sqrt{\frac{1+\delta}{1-\delta}}+1}\frac{\sqrt{\frac{1+\delta}{1-\delta}}-1}{\sqrt{\frac{1+\delta}{1-\delta}}+1} = \frac{\left(\sqrt{\frac{1+\delta}{1-\delta}}-1\right)^2}{\frac{1+\delta}{1-\delta}+1} = \frac{\frac{2}{1-\delta}-2\sqrt{\frac{1+\delta}{1-\delta}}}{\frac{2\delta}{1-\delta}} = \frac{1-\sqrt{(1+\delta)(1-\delta)}}{\delta} \leq \delta$$

Remark 1.8 — The last inequality uses taylor expansion of $\sqrt{(1+\delta)(1-\delta)}$.

With this, we can directly plug into the equation before. □

If the assumption in the corollary is satisfied, then:

$$2\delta^{k-s-t}\left(\frac{1+\delta}{\lambda_{\min}}\right)^s \leq \epsilon$$

is sufficient for an ϵ-solution.