1 March 18th, 2021

1.1 Preconditioned with Projection Methods

For one dimensional projection, we have steepest descent + preconditioning. Here we search for d_k on $\{d : \|d\|_A = \beta\}$. We obtain d_k by:

$$\min_{d_k} \|x_* - (x_k + \alpha d_k)\|_A^2 \quad \text{s.t } \|d_k\|_A = \beta$$

How ever, this is unrealistic because d_k would depend on x_*. Instead, we approximate it by searching d_k on $\{d : \|d\|_2 = \beta\}$. So, d_k is obtained by:

- Fixing $\alpha > 0$ and small
- Solving approximately (since we are consider the ellipsoid under $\|\cdot\|_2$):

$$\min_{d_k} \|x_* - (x_k + \alpha d_k)\|_2^2 \quad \text{s.t } \|d_k\|_2 = \beta$$

When the condition number of A is big, the set $\{d : \|d\|_2 = \beta\}$ is very flat in A-inner product space. To improve it, we search for d on $\S_P = \{d : \|d\|_P = \beta\}$, where P is such that S_P is rounder than S_2. Where

$$\|x\|_P = x^T P x$$

With this, we have:

$$d_k = \arg\min_{d_k \in S_p} \|(x_k + \alpha d_k) - x_*\|_A^2$$

for small $\alpha > 0$.

Remark 1.1 — The main idea is we change the metric from 2-norm to P norm, and we take a rough approximation by restricting it to S_P .

We have:

$$\begin{aligned} \|(x_k + \alpha d_k) - x_*\|_A^2 &= \|x_k - x_*\|_A^2 + 2\alpha \langle d_k, x_k - x_* \rangle_A + \alpha^2 \|d_k\|_A^2 \\ &\approx \|x_k - x_*\|_A^2 + 2\alpha \langle d_k, x_k - x_* \rangle_A . \end{aligned}$$

Thus:

$$\min_{d_k \in S_P} \|(x_k + \alpha d_k) + x_*\|_A^2 \iff \min_{d_k \in S_P} \langle d_k, x_k - x_* \rangle_A$$

Note that this is because $\alpha \gtrsim 0$ is close to zero.

We have:

$$\begin{aligned} \langle d_k, x_k - x_* \rangle_A &= \langle d_k, Ax_k - b \rangle \\ &= \langle d_k, -r_k \rangle \\ &= \left\langle d_k, -P^{-1} r_k \right\rangle_P \\ &\geq \|d_k\|_P \|P^{-1} r_k\|_P \quad \text{(by Cauchy-Schwarz)} \\ &= \beta \|P^{-1} r_k\|_P . \end{aligned}$$

The lower bound is attained when d_k is in the opposite direction of $-P^{-1}r_k$. Thus, we choose:

$$d_k = P^{-1}r_k$$

since only the direction of d_k matters. By calculation, we have:

$$\begin{aligned}\alpha_k = \arg\min_{\alpha\in\mathbb{R}} \|(x_k + \alpha d_k) - x_*\|_A^2 \\ &= \frac{\langle r_k, d_k\rangle}{\langle Ad_k, d_k\rangle}.\end{aligned}$$

This gives us Algorithm 1.

Algorithm 1 Preconditioned Steepest Descent

1: **for** $k = 0, 1, \dots$ **do**
2: $r_k = b - Ax_k$
3: Solve d_k from $Pd_k = r_k$
4: $\alpha_k = \frac{\langle r_k, d_k\rangle}{\langle Ad_k, d_k\rangle}$
5: $x_{k+1} = x_k + \alpha_k d_k$
6: **end for**

Remark 1.2 — If we choose $P = I$ then we get the non-preconditioned steepest descent.

For each iteration, we need to perform:

- 2 mat-vec products of A
- solve 1 system of linear equation of P
- operations of $O(n)$

We can improve this to using only 1 mat-vec product of A by introducing a new variable $p_k = Ad_k$, since:

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

This can be seen in Algorithm 2.

Algorithm 2 Improved Preconditioned Steepest Descent

1: $r_0 = b - Ax_0$
2: **for** $k = 0, 1, \dots$ **do**
3: Solve d_k from $Pd_k = r_k$
4: $p_k = Ad_k$
5: $\alpha_k = \frac{\langle r_k, d_k\rangle}{\langle Ad_k, d_k\rangle}$
6: $x_{k+1} = x_k + \alpha_k d_k$
7: $r_{k+1} = r_k - \alpha_k p_k$
8: **end for**