June 26, 2022

Proposition 1. *Solve linear equation:*

$$Ax = b, \tag{1}$$

where A and b are both large matrix.

Solution. Method of Steepest Descent [Wang(2008)]
We can rewrite this problem by minimizing the quadratic function:

$$J(x) = \frac{1}{2} min_x \|Ax - b\|_2. \tag{2}$$

The method of steepest descent tries to find an update in the direction of the steepest descent of the quadratic function (2):

$$J'(x^{(i)}) = Ax^{(i)} - b = -r^{(i)}, \tag{3}$$

where x^i is the solution for i-th iteration, $r^{(i)}$ is the gradient for i-th iteration.

Assume that we have the step α_i for i-th iteration to minimize the loss function J (2),yield

$$\min_{\alpha_i > 0} J\left(x^{(i)} + \alpha_i r^{(i)}\right). \tag{4}$$

The α_i should be chosen such that:

$$\begin{aligned} J'\left(x^{(i)} + \alpha_i r^{(i)}\right) &= 0 \\ \Leftrightarrow \left(b - A\left(x^{(i)} + \alpha_i r^{(i)}\right)\right)^T r^{(i)} &= 0 \\ \Leftrightarrow \alpha_i &= \frac{r^{(i)} r^{(i)}}{r^{(i)T} A r^{(i)}} \end{aligned} \tag{5}$$

Then we get the following steepest descent method for $i = 1, 2, \ldots$:

$$\begin{array}{ccc} r^{(i)} & := & b - Ax^{(i)}, \\ \alpha_i & := & \frac{r^{(i)T} r^{(i)}}{r^{(i)T} A r^{(i)}}, \\ x^{(i+1)} & = x^{(i)} & +\alpha_i r^{(i)}, \end{array} \tag{6}$$

and the termination condition is $\|r^{(i)}\| \leq \epsilon$, where ϵ is a positive constant. □

Solution. Conjugate Gradient Method [Nazareth(2009)]
If A is symmetric and positive-definite, then we can use conjugate gradient method to solve linear equations, which is more effective than steepest descent.

Assume we have initial states x_0. The solution x^* is unknown. r_k is the residual at k-th step:

$$\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k \tag{7}$$

r_k is the negative gradient at k-th step. But in CG, the direction p_k is conjugate to each other, given by:

$$\mathbf{p}_k = \mathbf{r}_k - \sum_{i<k} \frac{\mathbf{p}_i^\top \mathbf{A}\mathbf{r}_k}{\mathbf{p}_i^\top \mathbf{A}\mathbf{p}_i} \mathbf{p}_i \tag{8}$$

Then the next optimal states are given by:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k \tag{9}$$

with $\alpha_k = \frac{\mathbf{p}_k^\top(\mathbf{b} - \mathbf{A}\mathbf{x}_k)}{\mathbf{p}_k^\top \mathbf{A}\mathbf{p}_k} = \frac{\mathbf{p}_k^\top \mathbf{r}_k}{\mathbf{p}_k^\top \mathbf{A}\mathbf{p}_k}$. Only when r_k is sufficiently small, we return the x_{k+1} as the result. □

Solution. Biconjugate gradient stabilized method [Van der Vorst(1992)]
Unlike the conjugate gradient method, this algorithm does not require the matrix A to be symmetric and positive-definite. Firstly, we define the initial states x_0, initial negative gradient $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$, $\rho_0 = \alpha = \omega_0 = 1$, $\boldsymbol{v}_0 = \boldsymbol{p}_0 = \mathbf{0}$. Then for $i = 1, 2, ..., n$:

$$\begin{aligned}
&\rho_i = (\hat{\boldsymbol{r}}_0, \boldsymbol{r}_{i-1}) \\
&\beta = (\rho_i/\rho_{i-1})\,(\alpha/\omega_{i-1}) \\
&\boldsymbol{p}_i = \boldsymbol{r}_{i-1} + \beta\left(\boldsymbol{p}_{i-1} - \omega_{i-1}\boldsymbol{v}_{i-1}\right) \\
&\boldsymbol{v}_i = \boldsymbol{A}_i \\
&\alpha = \rho_i/\left(\hat{\boldsymbol{r}}_0, \boldsymbol{v}_i\right) \\
&\boldsymbol{s} = \boldsymbol{r}_{i-1} - \alpha\boldsymbol{v}_i \\
&\boldsymbol{t} = \boldsymbol{A}\boldsymbol{s} \\
&\omega_i = (\boldsymbol{t}, \boldsymbol{s})/(\boldsymbol{t}, \boldsymbol{t}) \\
&\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + \alpha\boldsymbol{p}_i + \omega_i\boldsymbol{s} \\
&\boldsymbol{r}_i = \boldsymbol{s} - \omega_i\boldsymbol{t},
\end{aligned} \tag{10}$$

if r_i is sufficiently small, we return the x_i as the result. □

Solution. Fast PseudoInverse [Jung and Sael(2020)]
Code implementation with notes and comments can be found in
`https://github.com/zhoujy53/Literature-Notes/blob/main/report/fastpi_with_notes.ipynb` □

Proposition 2. *Using CauchySchwarz inequality to prove Pythagorean theorem.*
Cauchy inequality:

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\| \tag{11}$$

where $\langle\cdot,\cdot\rangle$ is the inner product.
Pythagorean theorem: *As shown in Fig. 1, the area of the square whose side is the hypotenuse (the side opposite the right angle) is equal to the sum of the areas of the squares on the other two sides.*

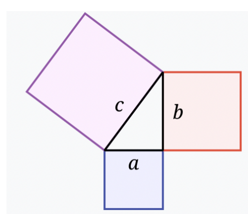

Figure 1: Right triangle

$$a^2 + b^2 = c^2 \tag{12}$$

Solution.

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle \\
&= \|\mathbf{u}\|^2 + \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle + \|\mathbf{v}\|^2 \quad \text{where } \langle \mathbf{v}, \mathbf{u}\rangle = \overline{\langle \mathbf{u}, \mathbf{v}\rangle} \\
&= \|\mathbf{u}\|^2 + 2\,\mathrm{Re}\langle \mathbf{u}, \mathbf{v}\rangle + \|\mathbf{v}\|^2 \\
&\leq \|\mathbf{u}\|^2 + 2|\langle \mathbf{u}, \mathbf{v}\rangle| + \|\mathbf{v}\|^2
\end{aligned} \tag{13}$$

Then using Cauchy inequality (11) yields:

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\end{aligned} \tag{14}$$

From (14), we get the triangle inequality:

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \tag{15}$$

The Pythagorean theorem is a special case of triangle inequality, only when the induced angle of $\mathbf{u}$ and $\mathbf{v}$ is right angle. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle = 0$ implies that:

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \tag{16}$$

$\square$

References

[Jung and Sael(2020)] Jinhong Jung and Lee Sael. Fast and accurate pseudoinverse with sparse matrix reordering and incremental approach. *Machine Learning*, 109(12):2333–2347, 2020.

[Nazareth(2009)] John L Nazareth. Conjugate gradient method. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(3):348–353, 2009.

[Van der Vorst(1992)] Henk A Van der Vorst. Bi-cgstab: A fast and smoothly converging variant of bi-cg for the solution of nonsymmetric linear systems. *SIAM Journal on scientific and Statistical Computing*, 13(2):631–644, 1992.

[Wang(2008)] Xu Wang. Method of steepest descent and its applications. *IEEE Microwave and Wireless Components Letters*, 12:24–26, 2008.