Kaczmarz's Algorithm For Solving Large Linear Systems*

Ziv Yaniv

Version 1.0
January 30 2012

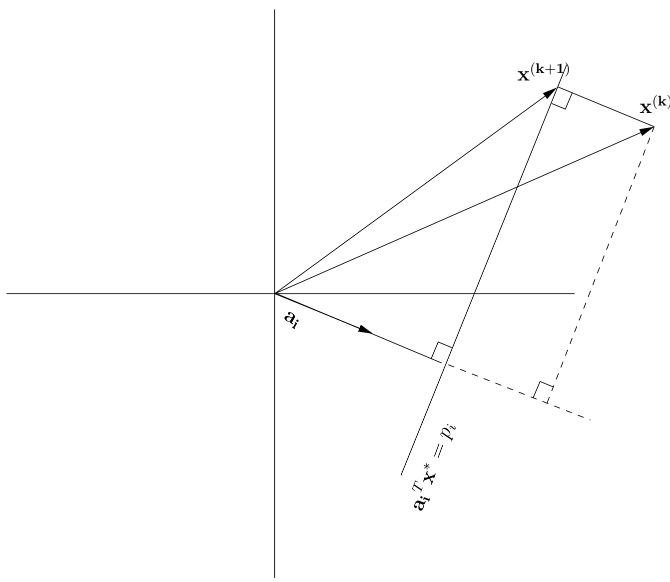

Figure 1: Hyper-plane representing one linear equation, with two consecutive approximate solutions. The plane's normal is given by $\frac{\mathbf{a_i}}{\sqrt{\mathbf{a_i^T a_i}}}$, and the distance from the origin is given by $\frac{p_i}{\sqrt{\mathbf{a_i^T a_i}}}$.

Given a large set of linear equations, $A\mathbf{x} = \mathbf{p}$, we seek a solution which does not require inversion of the matrix A. This summary describes Kaczmarz's [1] projection based approach for estimating this solution.

*Notation:
Matrices, A.
Column vectors,$\mathbf{a}$.
Scalars, a.

In expanded form we have:

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots a_{1,N}x_N = p_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots a_{2,N}x_N = p_2 \\ \vdots \\ a_{M,1}x_1 + a_{M,2}x_2 + \cdots a_{M,N}x_N = p_M \end{aligned}$$

Each of the equations defines a hyper-plane. If there exists a solution to the equation system, it is the intersection point of all planes. Figure 1 illustrates the projection based approach of Kaczmarz's algorithm.

Given an approximate solution $\mathbf{x^{(k)}}$ we project it onto a hyper-plane to obtain our next approximation $\mathbf{x^{(k+1)}}$. Without prior knowledge, the initial approximation can be arbitrarily set to $\mathbf{x^{(0)}} = \mathbf{0}$.

Note that the initialization allows explicit incorporation of prior knowledge, entries, x_i, that have known values.[1]

We thus have:

$$\begin{aligned} \mathbf{x^{(k+1)}} &= \mathbf{x^{(k)}} - \frac{\mathbf{a_i^T x^{(k)}} - p_i}{\sqrt{\mathbf{a_i^T a_i}}} \frac{\mathbf{a_i}}{\sqrt{\mathbf{a_i^T a_i}}} \\ &= \mathbf{x^{(k)}} - \frac{\mathbf{a_i^T x^{(k)}} - p_i}{\mathbf{a_i^T a_i}} \mathbf{a_i} \end{aligned} \tag{1}$$

We now show that this projection based method converges, and that it converges to the solution $\mathbf{x}^*$. This is done by analyzing the relationship between the residuals of successive approximations $\mathbf{x^{(k)}}, \mathbf{x^{(k+1)}}$.

$$\|\mathbf{x^{(k+1)}} - \mathbf{x}^*\|^2 = \|\mathbf{x^{(k)}} - \mathbf{x}^* - \frac{\mathbf{a_i^T x^{(k)}} - p_i}{\mathbf{a_i^T a_i}} \mathbf{a_i}\|^2$$

Using the identity $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x}^T\mathbf{y} + \|\mathbf{y}\|^2$ we have:

$$\begin{aligned} \|\mathbf{x^{(k+1)}} - \mathbf{x}^*\|^2 &= \|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2 - 2\frac{\mathbf{a_i^T x^{(k)}} - p_i}{\mathbf{a_i^T a_i}} \mathbf{a_i^T}(\mathbf{x^{(k)}} - \mathbf{x}^*) + \frac{(\mathbf{a_i^T x^{(k)}} - p_i)^2}{\mathbf{a_i^T a_i}} \\ &= \|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2 - \frac{\mathbf{a_i^T x^{(k)}} - p_i}{\mathbf{a_i^T a_i}}(2\mathbf{a_i^T x^{(k)}} - 2\mathbf{a_i^T x^*} - \mathbf{a_i^T x^{(k)}} + p_i) \end{aligned}$$

Substituting $\mathbf{a_i^T x^*} = p_i$ into the equation above we get:

$$\begin{aligned} \|\mathbf{x^{(k+1)}} - \mathbf{x}^*\|^2 &= \|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2 - \frac{\mathbf{a_i^T x^{(k)}} - p_i}{\mathbf{a_i^T a_i}}(2\mathbf{a_i^T x^{(k)}} - 2p_i - \mathbf{a_i^T x^{(k)}} + p_i) \\ &= \|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2 - \frac{(\mathbf{a_i^T x^{(k)}} - p_i)^2}{\mathbf{a_i^T a_i}} \end{aligned}$$

[1] In the case of computed tomography we may know which areas are empty space (air) and thus have a value of $x_i = -1000$HU.

The second term of the equation above is non-negative. We thus have:

$$\|\mathbf{x^{(k+1)}} - \mathbf{x}^*\|^2 \leq \|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2$$

Therefore,$\{\|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2\}$ is a non-increasing sequence, bounded from below, $\|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2 \geq 0$. Hence it converges.

We now observe that:

$$\|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2 = \|\mathbf{x^{(0)}} - \mathbf{x}^*\|^2 - \sum_{i=0}^{k-1} \frac{(\mathbf{a^T_{r(i)}} \mathbf{x^{(i)}} - p_{r(i)})^2}{\mathbf{a^T_{r(i)}} \mathbf{a_{r(i)}}}$$

where $r(i)$ is the index of the row used for projection of $\mathbf{x^{(i)}}$.

as the sequence $\{\|\mathbf{x^{(k)}} - \mathbf{x}^*\|^2\}$ converges we see that

$$\sum_{i=0}^{\infty} \frac{(\mathbf{a^T_{r(i)}} \mathbf{x^{(i)}} - p_{r(i)})^2}{\mathbf{a^T_{r(i)}} \mathbf{a_{r(i)}}} < \infty$$

consequentially (after setting $p_i = \mathbf{a^T_{r(i)}} \mathbf{x}^*$)

$$\lim_{k \to \infty} \mathbf{a^T_{r(i)}} (\mathbf{x^{(k)}} - \mathbf{x}^*) = 0$$

Giving us

$$\lim_{k \to \infty} \mathbf{x^{(k)}} = \mathbf{x}^*$$

References

[1] S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen*, Bull. Acad. Polon. Sci. ct Lettres A, vol. 35, pp. 355–357, 1937.