A basic note on iterative matrix inversion

Alen Alexanderian*

Abstract

We provide a short proof of the iteration $\mathbf{X}_{n+1} = \mathbf{X}_n(2I - A\mathbf{X}_n)$ for computing the inverse of a matrix.

1 Introduction and preliminaries

This note concerns the well known Newton iteration[1] for computing the inverse of a matrix [1, 3, 2, 4]:

$$\mathbf{X}_{n+1} = \mathbf{X}_n(2\mathbf{I} - \mathbf{A}\mathbf{X}_n). \tag{1.1}$$

For an arbitrary matrix $\mathbf{A}$ the above iteration allows for computing its generalized inverse (pseudo inverse). The motivation for writing this basic note was to provide a short proof of convergence for the case $\mathbf{A}$ is a real non-singular matrix.

Let us recall the following basic result which will be used later in this note.

Lemma 1.1. *Suppose $\mathbf{A}$ is an $n \times n$ complex matrix, with spectral radius $\rho(\mathbf{A})$. Then, $\lim_{n\to\infty} \mathbf{A}^n = 0$ if and only if $\rho(\mathbf{A}) < 1$.*

In what follows, we denote the eigenvalues of an $n \times n$ matrix $\mathbf{S}$ by $\lambda_i(\mathbf{S})$, $i = 1, \ldots, n$. In case $\mathbf{S}$ is symmetric positive definite matrix $\mathbf{S} \in \mathbb{R}^{n\times n}$, we can order its eigenvalues,

$$\lambda_1(\mathbf{S}) \geq \lambda_2(\mathbf{S}) \geq \cdots \geq \lambda_n(\mathbf{S}) > 0.$$

2 Proof of convergence

Here we state and prove a theorem concerning the convergence of (1.1). As mentioned before, the result considered here is a special case of a more general and well known result which concerns the iterative computation of pseudo-inverses.

Theorem 2.1. *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a non-singular matrix and define the sequence $\{\mathbf{X}_n\}_{n\geq 0}$ of matrices as follows:*

$$\begin{cases} \mathbf{X}_0 = \alpha \mathbf{A}^T, & \text{with } \alpha \in \left(0, \dfrac{2}{\lambda_1(\mathbf{A}\mathbf{A}^T)}\right), \\ \mathbf{X}_{n+1} = \mathbf{X}_n(2\mathbf{I} - \mathbf{A}\mathbf{X}_n). \end{cases} \tag{2.1}$$

Then, $\mathbf{X}_n \to \mathbf{A}^{-1}$ as $n \to \infty$.

---

*The University of Texas at Austin, USA. E-mail: alen@ices.utexas.edu

Last revised: July 12, 2013

[1]The reference to Newton is due to the observation that the iteration considered here can be obtained by applying the Newton's method for solving the equation $f(\mathbf{X}) = \mathbf{0}$ with $f(\mathbf{X}) = \mathbf{A} - \mathbf{X}^{-1}$.

Proof. Let $\mathbf{R}_n = \mathbf{I} - \mathbf{A}\mathbf{X}_n$. Then, we note that $\mathbf{X}_{n+1} = \mathbf{X}_n(\mathbf{I} + \mathbf{R}_n)$. We first show that $\rho(\mathbf{R}_0) = \rho(\mathbf{I} - \alpha\mathbf{A}\mathbf{A}^T) < 1$. We note that the eigenvalues $\lambda_i(\mathbf{R}_0)$ are given by,

$$\lambda_i(\mathbf{R}_0) = 1 - \alpha\lambda_i(\mathbf{A}\mathbf{A}^T).$$

Thus, by the choice of α in (2.1) it follows that $|\lambda_i(\mathbf{R}_0)| < 1$, and thus $\rho(\mathbf{R}_0) < 1$. Therefore, by Lemma 1.1 we have

$$\lim_{n\to\infty} (\mathbf{R}_0)^n = 0. \tag{2.2}$$

Next, we note that,

$$\begin{aligned}\mathbf{R}_n = \mathbf{I} - \mathbf{A}\mathbf{X}_n &= \mathbf{I} - \mathbf{A}\mathbf{X}_{n-1}(\mathbf{I} + \mathbf{R}_{n-1}) \\ &= \mathbf{I} - \mathbf{A}\mathbf{X}_{n-1} - \mathbf{A}\mathbf{X}_{n-1}\mathbf{R}_{n-1} \\ &= \mathbf{R}_{n-1} - \mathbf{A}\mathbf{X}_{n-1}\mathbf{R}_{n-1} = (\mathbf{I} - \mathbf{A}\mathbf{X}_{n-1})\mathbf{R}_{n-1} = (\mathbf{R}_{n-1})^2.\end{aligned}$$

Therefore, inductively, we have $\mathbf{R}_n = (\mathbf{R}_0)^{2^n}$. Hence, $\lim_{n\to\infty} \mathbf{R}_n = \lim_{n\to\infty} (\mathbf{R}_0)^{2^n} = \mathbf{0}$, where the last equality follows from (2.2). Finally, from the definition of $\mathbf{R}_n$ we note that $\mathbf{X}_n = \mathbf{A}^{-1}(\mathbf{I} - \mathbf{R}_n)$. Therefore,

$$\lim_{n\to\infty} \mathbf{X}_n = \lim_{n\to\infty} \mathbf{A}^{-1}(\mathbf{I} - \mathbf{R}_n) = \mathbf{A}^{-1}. \qquad \square$$

Remark 2.2. *The statement of the above theorem suggests that to determine an appropriate α one needs to approximate the largest eigenvalue of $\mathbf{A}\mathbf{A}^T$. However, as noted in [2] this is not necessary. Denoting the entries of $\mathbf{A}\mathbf{A}^T$ by $(b_{ij})_{i,j=1}^n$ and using Gerschgorin's Theorem, we know that, $\lambda_1(\mathbf{A}\mathbf{A}^T)$ is bounded by the maximum absolute value row sum of $\mathbf{A}\mathbf{A}^T$, that is, $\lambda_1(\mathbf{A}\mathbf{A}^T) \le \max_{i\in\{1,\ldots,n\}} \sum_{j=1}^n |b_{ij}|$. Thus, letting $R = \max_i \sum_j |b_{ij}|$, we can select α according to,*

$$\alpha \in \left(0, \frac{2}{R}\right).$$

References

[1] Adi Ben-Israel. An iterative method for computing the generalized inverse of an arbitrary matrix. *Math. Comp.*, 19:452–455, 1965.

[2] Adi Ben-Israel. A note on an iterative method for generalized inversion of matrices. *Mathematics of Computation*, 20(95):439–440, 1966.

[3] Adi Ben-Israel and Dan Cohen. On iterative computation of generalized inverses and associated projections. *SIAM J. Numer. Anal.*, 3:410–419, 1966.

[4] Jr. H. P. Decell and S. W. Kuhng. An iterative method for computing the generalized inverse of a matrix. Report D-3464, NASA TN, 1966.