On the tripling algorithm for large-scale nonlinear matrix equations with low rank structure

Ning Dong [a,b], Bo Yu [b,*]

[a] *School of Mathematics and Computational Science, Xiangtan University, Xiangtan, 411105, PR China*
[b] *School of Science, Hunan University of Technology, Zhuzhou, 412000, PR China*

ARTICLE INFO

ABSTRACT

For the large-scale nonlinear matrix equations with low rank structure, the well-developed doubling algorithm in low rank form (DA-LR) is known as an efficient method to compute the stabilizing solution. By further analyzing the global efficiency index constructed in this paper, we propose a tripling algorithm in low rank form (TA-LR) from two points of view, the cyclic reduction and the symplectic structure preservation. The new presented algorithm shares the same pre-processing complexity with that of DA-LR, but can attain the prescribed normalized residual level within less iterations by only consuming some negligible iteration time as an offset. Under the solvability condition, the proposed algorithm is demonstrated to inherit a cubic convergence and is capable of delivering errors from the current iteration to the next with the same order. Numerical experiments including some from nano research show that the TA-LR is highly efficient to compute the stabilizing solution of large-scale nonlinear matrix equations with low rank structure.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The nonlinear matrix equation (NME)

$$X + BX^{-1}A = Q \tag{1.1}$$

and its dual form

$$Y + AY^{-1}B = Q \tag{1.2}$$

with A, B, Q, X and $Y \in \mathbb{C}^{n \times n}$ arise from many real-life applications such as in surface acoustic simulations [1–3], quadratic eigenvalue problems in fast train [4–6] or computation of Green's function in nano research [7–9]. The coefficient matrices in NME (1.1), obtained by discretizing the Hamiltonian system from concrete problems, are usually provided with the structure of sparsity and low rank. Here in this paper we always suppose that the matrix Q is Hermitian and nonsingular with $O(n)$ complexity for solving the linear system $Qx = a$ with $x, a \in \mathbb{C}^n$ and, matrices A and B respectively own ranks $r_a, r_b \ll n$ such that they admit some low rank decompositions. It is widely approved that one favorite doubling algorithm, also called the SDA-2, works very well for solving the NME (1.1). It is also argued by Chu et al. [10] that the doubling algorithm of type 1 (SDA-1), when solving another class of discrete-time algebraic Riccati equations (DAREs) [11,12], is better than a third-order structure-preserving tripling iterative strategy (STA-1) in defiance of the special low rank structure. Conceivably,

---

* Corresponding author.
  E-mail address: boyu_hut@126.com (B. Yu).

similar conclusions may be drawn for SDA-2 when solving NME (1.1). Here in this paper we are going to give a theoretical interpretation for this by using the efficiency index (EI). Particularly, we are more interested in the global efficiency of the whole algorithm for large-scale NMEs with low rank structure, which may bring to light the motivation for developing tripling algorithm in low rank form (TA-LR).

A reasonable index, introduced by Traub [13], to evaluate the efficiency of the iterative method for solving the scalar equation $f(x) = 0$ is

$$\mathrm{EI} = p^{1/q},$$

where p is the order of the method and q represents the number of pieces of information for one iteration. For example, the classic Newton's method according to the EI above, has the efficiency $\sqrt{2} = 1.4142$ since it converges quadratically with the requirement of two pieces of information (one function evaluation and one derivative evaluation [13,14]). Unfortunately, such an index seems too rough to be applied to iterative methods for solving matrix equations with dimension $n \geq 1$. Here we turn to Guo's efficiency index [15], measured by flops counts, to re-interpret the efficiency of the doubling algorithm.

It is well-known that SDA-2 has a second order convergence in the non-critical case and the computational complexity is about $\frac{40}{3}n^3$ for one iteration [16,9]. So one has $\mathrm{EI}_{\mathrm{sda}} = 1.4142$ by regarding $\frac{20}{3}n^3$ flops as one piece of information. For the structure-preserving tripling algorithm of type 2 (STA-2, see next section), additional $\frac{52}{3}n^3$ flops are to be seen required for establishing the third order convergence, which means $\mathrm{EI}_{\mathrm{sta}} = 3^{\frac{20}{3}n^3/(\frac{40}{3}n^3 + \frac{52}{3}n^3)} = 1.2697$. Therefore, one may argue similarly to [10] that the SDA-2 is still better than STA-2 for solving NME (1.1) in generic cases.

However for large-scale NMEs with $r_a, r_b \ll n$, the calculation of the above EI at each iteration seems not so important since that the entire computations of SDA-2 in low rank form are mainly dominated by the $O(n)$ pre-processing procedure [17]. So it is more reasonable to consider the EI of the whole algorithm rather than that of each iteration. To be specifical, suppose that l_d and l_t iterations are required respectively for the doubling algorithm and the tripling algorithm before the termination and $r_a = r_b := r$. The global efficiency index (GEI) for SDA-2 and STA-2 could be defined as

$$\mathrm{GEI}_{\mathrm{sda}} = 2^{\frac{c}{l_d \cdot c_d \cdot r^3}} \quad \text{and} \quad \mathrm{GEI}_{\mathrm{sta}} = 3^{\frac{c}{l_t \cdot c_t \cdot r^3}}, \tag{1.3}$$

respectively, where c is some proper positive constant, c_d and c_t stand for the complexity constants in the doubling algorithm and the tripling algorithm respectively. Although complexity analysis in Section 5 indicates that c_t is greater than c_d in (1.3), it is still possible that the $\mathrm{GEI}_{\mathrm{sta}}$ might be sufficiently approximative to the $\mathrm{GEI}_{\mathrm{sda}}$ provided that l_t is less than d_t to some degree. Such a hedging trade-off will be validated practically and explained detailedly by some numerical experiments in the last section. On the other but more vital hand, the whole iteration complexities of two algorithms are usually far less than the dimension n for large-scale NMEs with low rank structure. Then their discrepancy on iteration time could be largely compressed so as to be ignorable. These facts directly contribute our original motivation to reconsider the tripling algorithm in low rank form (TA-LR). In fact, one will see in numerical experiments that for large-scale problems, the TA-LR is capable of attaining the prescribed residual level within less iterations (compared with the doubling algorithm in low rank form (DA-LR)), only sacrificing some negligible iteration time as an offset.

Throughout this paper, we always assume that the following solvability condition in [18,8] holds true. So that both the DA-LR and the TA-LR can obtain the stabilizing solution. Note that NME (1.1) is also closely related with the quadratic polynomial matrix equation $BZ^2 - QZ + A = 0$ with $Z = X^{-1}A$ and its solvability conditions can be found in [19–24].

Theorem 1.1. *Let $A = C + iD$, $B = C^* + iD^*$ be $n \times n$ complex matrices with $C = (A + B^*)/2$, $D = (A - B^*)/(2i)$ and the superscript "$*$" representing the conjugate transpose. Let $Q = \mathrm{Re}(Q) + i\mathrm{Im}(Q)$. If the matrix $\psi(z) = zD^* + \mathrm{Im}(Q) + z^{-1}D$ is positive definite for each z on the unit circle in the complex plane. Then the NME (1.1) (DNME (1.2)) has a stabilizing solution $X_s = \mathrm{Re}(X_s) + i\,\mathrm{Im}(X_s)$ $(\widehat{X}_s = \mathrm{Re}(\widehat{X}_s) + i\,\mathrm{Im}(\widehat{X}_s))$ with the spectral radius of $X_s^{-1}A$ $(\widehat{X}_s^{-1}B)$ less than one.*

The rest of the paper is organized as follows. Section 2 develops the tripling algorithm for NME (1.1) from viewpoints of the cyclic reduction and the structure preservation. A detailed convergence analysis is established in Section 3 and the corresponding low rank algorithm TA-LR is constructed in Section 4, respectively. Section 5 is devoted to the complexity and errors analysis. Numerical experiments are reported in the last section to show the efficiency of the TA-LR for computing the stabilizing solution of large-scale NMEs.

Notation. Symbols $\mathbb{N}^+$, $\mathbb{R}^{n \times n}$ and $\mathbb{C}^{n \times n}$ in this paper stand for sets of positive integers, $n \times n$ real matrices and $n \times n$ complex matrices, respectively. $\mathbb{T}$ represents the unit circle in the complex plane. I_n (or simply I if its dimension is clear from the context) is the $n \times n$ identity matrix. For a matrix $A \in \mathbb{C}^{n \times n}$, $\sigma(A)$ and $\rho(A)$ denote respectively the spectrum and spectral radius of A, and $\mathrm{Re}(A)$ and $\mathrm{Im}(A)$ stand respectively for the real part and the image part of A. For Hermitian matrices A and $B \in \mathbb{C}^{n \times n}$, we say $A > B$ $(A \geq B)$ if $A - B$ is a positive definitive (positive semi-definite) matrix.

2. Tripling algorithm

In order to describe the tripling algorithm explicitly and conveniently, this section is divided into two parts to depict the construction procedure from two points of view: (a) the cyclic reduction; (b) the symplectic structure preservation.