

## RECURSIVE ALGORITHMS FOR SOLVING A CLASS OF NONLINEAR MATRIX EQUATIONS WITH APPLICATIONS TO CERTAIN SENSITIVITY OPTIMIZATION PROBLEMS\*

WEI-YONG YAN<sup>†</sup>, JOHN B. MOORE<sup>‡</sup>, AND UWE HELMKE<sup>§</sup>

**Abstract.** This paper is concerned with solving a class of nonlinear algebraic matrix equations. Two recursive algorithms are proposed in terms of matrix *difference* equations and are studied. A set of initial values is characterized, from which the convergence of the algorithms can be guaranteed.

Based on the general results, several effective algorithms are presented to compute  $L^2$ -sensitivity optimal realizations, as well as Euclidean norm balancing realizations, of a given linear system. A locally exponential convergence property is proved for one of them. As is shown by simulation in this paper, these algorithms prove to be far more practical for digital computer implementation than the gradient flows previously proposed.

**Key words.** matrix equations, difference equations, linear systems, sensitivity, minimal realizations

**AMS subject classifications.** 93B40, 15A24, 39A10

**1. Introduction.** Consider the algebraic matrix equation of the form

$$(1.1) \quad \mathcal{F}(X) - X\mathcal{G}(X)X = 0, \quad X \in \mathcal{P}(n),$$

with  $\mathcal{F}(\cdot)$  and  $\mathcal{G}(\cdot)$  continuous operators from  $\mathcal{P}(n)$  to itself, where  $\mathcal{P}(n)$  denotes the set of all positive definite symmetric  $n \times n$  matrices. In this paper we are interested in finding the solution to (1.1) under the following basic assumptions:

*Assumption 1.*  $\mathcal{F}(\cdot)$  and  $\mathcal{G}(\cdot)^{-1}$  are nondecreasing, that is, for any  $X_1, X_2 \in \mathcal{P}(n)$  with  $X_2 \geq X_1$ , there hold  $\mathcal{F}(X_2) \geq \mathcal{F}(X_1)$  and  $\mathcal{G}(X_2) \leq \mathcal{G}(X_1)$ , where the notation  $X_2 \geq X_1$  ( $X_2 > X_1$ ) means that  $X_2 - X_1$  is positive semidefinite (definite).

*Assumption 2.* Equation (1.1) has a unique solution  $\bar{X}$  in  $\mathcal{P}(n)$ .

This type of matrix equation often arises in systems and control. For example, it has been recently found [1], [3] that solving the problems of  $L^2$ -sensitivity minimization and Euclidean norm balancing can be reduced to solving certain highly nonlinear equations of the form (1.1). Unfortunately, there is no explicit formula for their unique solution to these algebraic equations. The only computation method available to date is to solve certain related nonlinear *differential* equations. For high-order problems, this method may be impractical or inefficient using conventional digital computers. Therefore, it is desirable to develop suitable iterative algorithms in terms of *difference* equations whose solution can converge to the required solution from appropriate initial conditions.

Of course, it is not always possible to relate (1.1) to some optimization problem or differential equation so that certain numerical methods such as Euler approximation and Newton–Raphson can be applied. Moreover, even if possible, the existing numerical methods may not always work well and may be inefficient because their success heavily depends on intricate step size adjustment, which is sometimes time consuming. In contrast, the method to be proposed in this paper for solving (1.1) does not require any step size adjustment, as will be seen soon.

\* Received by the editors March 2, 1992; accepted for publication (in revised form) June 23, 1993. This work was partially supported by the Boeing Commercial Aircraft Corporation (BCAC).

<sup>†</sup> Department of Mathematics, University of Western Australia, Nedlands, Western Australia 6009, Australia.

<sup>‡</sup> Department of Systems Engineering, Research School of Physical Sciences and Engineering, Australian National University, GPO Box 4, Canberra, Australian Capital Territory 2601, Australia.

<sup>§</sup> Department of Mathematics, University of Regensburg, 8400 Regensburg, Germany.

To suggest a workable algorithm for (1.1), let us consider the trivial case where  $\mathcal{F}(P) = F \in \mathcal{P}(n)$  and  $\mathcal{G}(P) = G \in \mathcal{P}(n)$ , that is, the operators  $\mathcal{F}$  and  $\mathcal{G}$  are constant. Quite obviously, Assumptions 1 and 2 are fulfilled in this case. Equation (1.1) then reduces to

$$(1.2) \quad F - XGX = 0,$$

which is apparently a very special form of the algebraic Riccati equation in continuous time. Thus, by the bilinear transformation given in [2], (1.2) is equivalent to the following algebraic Riccati equation in discrete time:

$$(1.3) \quad X = X - 2(X + F)(2X + F + G^{-1})^{-1}(X + F) + 2F.$$

It is known from [2] that the solution of the Riccati difference equation

$$(1.4) \quad X_{i+1} = X_i - 2(X_i + F)(2X_i + F + G^{-1})^{-1}(X_i + F) + 2F$$

converges to the solution of (1.2) or (1.3) from any initial condition  $X_0 \in \mathcal{P}(n)$ .

The above-mentioned fact inspires us to come up with the following difference equation for the general purpose:

$$(1.5) \quad X_{i+1} = X_i - 2[X_i + \mathcal{F}(X_i)][2X_i + \mathcal{F}(X_i) + \mathcal{G}(X_i)^{-1}]^{-1}[X_i + \mathcal{F}(X_i)] + 2\mathcal{F}(X_i),$$

which is obtained by respective substitution of the operators  $\mathcal{F}$  and  $\mathcal{G}$  for  $F$  and  $G$  into (1.4). A natural question arises as to whether the solution of (1.5) can still converge to the solution of (1.1) from any initial condition  $X_0 \in \mathcal{P}(n)$  in the general case. In particular, can (1.5) serve as an iterative algorithm in the practical cases of interest?

*Remark 1.1.* It is worth emphasizing that the operators  $\mathcal{F}$  and  $\mathcal{G}$  will not be required to be smooth in the development to follow. We hope that this would widen the potential applications of the algorithms to be developed in the paper.

*Remark 1.2.* Note that computing the value of the operators  $\mathcal{F}$  and  $\mathcal{G}$  is required at each iteration of the algorithm (1.5), which may be undesirable or difficult in the situation where  $\mathcal{F}$  and  $\mathcal{G}$  are complicated or there are even no explicit expressions for them. As will be seen in the sequel, this difficulty can be overcome by way of incorporating two additional difference equations with (1.5).

*Remark 1.3.* If the operators  $\mathcal{F}$  and  $\mathcal{G}$  satisfy differentiability conditions, homotopy methods might be used to find the solution of (1.1). However, this kind of method eventually rests in solving a differential equation [4]. Moreover, its success crucially depends on the construction of a homotopy map satisfying certain requirements; otherwise, there is no guarantee that the method is globally convergent (see, e.g., [4]).

In the next section we prove some auxiliary results. Section 3 is devoted to studying the convergence properties of two types of general nonlinear difference equations including (1.5). Section 4 discusses specific iterative schemes for solving (1.1) in the cases of  $L^2$ -sensitivity minimization and Euclidean norm balancing. Section 5 proves that the convergence of the proposed algorithm for the  $L^2$ -sensitivity minimization problem is locally exponential, and §6 presents some simulation results. Conclusions appear in §7.

**2. Preliminary results.** Define

$$(2.1) \quad \mathcal{R}(X) \triangleq X - 2[X + \mathcal{F}(X)][2X + \mathcal{F}(X) + \mathcal{G}(X)^{-1}]^{-1}[X + \mathcal{F}(X)] + 2\mathcal{F}(X)$$

for  $X \in \mathcal{P}(n)$ , where  $\mathcal{F}(\cdot), \mathcal{G}(\cdot) : \mathcal{P}(n) \mapsto \mathcal{P}(n)$ . The following fundamental properties of  $\mathcal{R}(\cdot)$  are instrumental in discussing the convergence of the algorithm  $X_{i+1} = \mathcal{R}(X_i)$  subsequently.

LEMMA 2.1. Suppose  $\mathcal{F}(X)$  and  $\mathcal{G}(X)^{-1}$  are nondecreasing with respect to  $X \in \mathcal{P}(n)$ . Then  $\mathcal{R}(\cdot)$  maps  $\mathcal{P}(n)$  to itself; moreover, for any  $X, Y \in \mathcal{P}(n)$  there hold

$$(2.2) \quad X \geq Y \implies \mathcal{R}(X) \geq \mathcal{R}(Y),$$

$$(2.3) \quad \mathcal{R}(X) \geq X \iff \mathcal{F}(X) \geq X\mathcal{G}(X)X,$$

$$(2.4) \quad \mathcal{R}(X) \leq X \iff \mathcal{F}(X) \leq X\mathcal{G}(X)X,$$

$$(2.5) \quad \mathcal{R}(X) = X \iff \mathcal{F}(X) = X\mathcal{G}(X)X.$$

*Proof.* Upon noting from the matrix inversion formula that

$$(2.6) \quad \mathcal{R}(X) = 2\{[X + \mathcal{F}(X)]^{-1} + [X + \mathcal{G}(X)^{-1}]^{-1}\}^{-1} - X,$$

it follows that  $X \in \mathcal{P}(n)$  implies  $\mathcal{R}(X) \in \mathcal{P}(n)$ . Now assume  $X \geq Y$ . Because  $\mathcal{F}(X)$  and  $\mathcal{G}(X)^{-1}$  are nondecreasing, we obtain

$$\begin{aligned} \mathcal{R}(X) &\geq 2\{[X + \mathcal{F}(Y)]^{-1} + [X + \mathcal{G}(Y)^{-1}]^{-1}\}^{-1} - X \\ &= X + 2\mathcal{F}(Y) - 2[X + \mathcal{F}(Y)][2X + \mathcal{F}(Y) + \mathcal{G}(Y)^{-1}]^{-1}[X + \mathcal{F}(Y)] \\ &= \mathcal{F}(Y) + [\mathcal{G}(Y)^{-1} - \mathcal{F}(Y)][2X + \mathcal{F}(Y) + \mathcal{G}(Y)^{-1}]^{-1}[X + \mathcal{F}(Y)] \\ &= \frac{1}{2}[\mathcal{F}(Y) + \mathcal{G}(Y)^{-1}] - \frac{1}{2}[\mathcal{G}(Y)^{-1} - \mathcal{F}(Y)] \\ &\quad \cdot [2X + \mathcal{F}(Y) + \mathcal{G}(Y)^{-1}]^{-1}[\mathcal{G}(Y)^{-1} - \mathcal{F}(Y)] \\ &\geq \frac{1}{2}[\mathcal{F}(Y) + \mathcal{G}(Y)^{-1}] - \frac{1}{2}[\mathcal{G}(Y)^{-1} - \mathcal{F}(Y)] \\ &\quad \cdot [2Y + \mathcal{F}(Y) + \mathcal{G}(Y)^{-1}]^{-1}[\mathcal{G}(Y)^{-1} - \mathcal{F}(Y)] \\ &= \mathcal{R}(Y), \end{aligned}$$

implying that (2.2) holds. Further, from the identity

$$(2.7)$$

$$\mathcal{R}(X) = X + 2\mathcal{F}(X) - 2\{[X + \mathcal{F}(X)]^{-1} + [X + \mathcal{F}(X)]^{-1}[X + \mathcal{G}(X)^{-1}][X + \mathcal{F}(X)]^{-1}\}^{-1}$$

it can be seen that

$$\begin{aligned} \mathcal{R}(X) &\geq X \\ &\iff \mathcal{F}(X)^{-1} \leq [X + \mathcal{F}(X)]^{-1} + [X + \mathcal{F}(X)]^{-1}[X + \mathcal{G}(X)^{-1}][X + \mathcal{F}(X)]^{-1} \\ &\iff [X + \mathcal{F}(X)]\mathcal{F}(X)^{-1}[X + \mathcal{F}(X)] \leq [X + \mathcal{F}(X)] + [X + \mathcal{G}(X)^{-1}] \\ &\iff X\mathcal{F}(X)^{-1}X \leq \mathcal{G}(X)^{-1} \\ &\iff \mathcal{F}(X) \geq X\mathcal{G}(X)X, \end{aligned}$$

that is, (2.3) holds. In the same spirit, (2.4) can be proved and (2.5) is obtained as a direct result of (2.3) and (2.4).  $\square$

COROLLARY 2.1. Let

$$(2.8) \quad f(X, Y, Z) = X - 2(X + Y)(2X + Y + Z)^{-1}(X + Y) + 2Y$$

for  $(X, Y, Z) \in \mathcal{P}(n) \times \mathcal{Q}(n) \times \mathcal{Q}(n)$ , where  $\mathcal{Q}(n)$  denotes the set of all  $n \times n$  nonnegative definite symmetric matrices. Then

$$(2.9) \quad 0 < X_1 \leq X_2, \quad 0 \leq Y_1 \leq Y_2, \quad 0 \leq Z_1 \leq Z_2$$

imply

$$(2.10) \quad f(X_1, Y_1, Z_1) \leq f(X_2, Y_2, Z_2).$$

In particular, if  $Y, Z \geq A$  for some  $A \in \mathcal{P}(n)$ , there holds

$$(2.11) \quad f(X, Y, Z) \geq A, \quad \forall X \in \mathcal{P}(n).$$

*Proof.* From (2.2) of Lemma 2.1 and (2.6) it follows that

$$\begin{aligned} f(X_1, Y_1, Z_1) &\leq f(X_2, Y_1, Z_1) \\ &= 2\{(X_2 + Y_1)^{-1} + (X_2 + Z_1)^{-1}\}^{-1} - X_2 \\ &\leq 2\{(X_2 + Y_2)^{-1} + (X_2 + Z_2)^{-1}\}^{-1} - X_2 \\ &= f(X_2, Y_2, Z_2), \end{aligned}$$

which gives (2.10).  $\square$

*Remark 2.1.* From (2.5), we can see that under Assumption 1 in §1 the equilibrium point of the difference equation  $\mathcal{R}(X_{i+1}) = X_i$  exactly equals the solution of (1.1).

The next two results are only used in developing an efficient iterative scheme for finding  $L^2$ -sensitivity optimal realizations.

**LEMMA 2.2.** *Given a minimal realization  $(A, b, c)$  with the eigenvalues of  $A$  being in the open unit disk, let  $U(\mu)$  be the solution of the following Lyapunov equation:*

$$(2.12) \quad Q = \begin{bmatrix} A' & c'b' \\ 0 & A' \end{bmatrix} Q \begin{bmatrix} A & 0 \\ bc & A \end{bmatrix} + \begin{bmatrix} c'c & 0 \\ 0 & \mu I \end{bmatrix}.$$

Then there holds

$$(2.13) \quad \lim_{\mu \rightarrow \infty} U(\mu)^{-1} = 0.$$

*Proof.* Let  $V(\mu)$  be the solution of the Lyapunov equation

$$(2.14) \quad Q = \begin{bmatrix} A' & c'b' \\ 0 & A' \end{bmatrix} Q \begin{bmatrix} A & 0 \\ bc & A \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \mu I \end{bmatrix}.$$

Then it is quite evident that  $U(\mu) \geq V(\mu) = \mu V(1)$ . Because the realization  $(A, b, c)$  is minimal, it follows that the pair

$$\left( \begin{bmatrix} A' & c'b' \\ 0 & A' \end{bmatrix}, \begin{bmatrix} 0 \\ I \end{bmatrix} \right)$$

is controllable. Thus,  $V(1)$  is positive definite, leading to  $\lim_{\mu \rightarrow \infty} U(\mu)^{-1} = 0$ .  $\square$

**LEMMA 2.3.** *Given a minimal realization  $(A, b, c)$  with the eigenvalues of  $A$  being in the open unit disk, consider the difference equation*

(2.15)

$$Q_{i+1} \triangleq \begin{bmatrix} Q_{i+1}^{11} & Q_{i+1}^{12} \\ Q_{i+1}^{21} & Q_{i+1}^{22} \end{bmatrix} = \begin{bmatrix} A' & c'b' \\ 0 & A' \end{bmatrix} \begin{bmatrix} Q_i^{11} & Q_i^{12} \\ Q_i^{21} & Q_i^{22} \end{bmatrix} \begin{bmatrix} A & 0 \\ bc & A \end{bmatrix} + \begin{bmatrix} c'c & 0 \\ 0 & P_i \end{bmatrix}$$

with an initial symmetric matrix  $Q_0$ .

(i) If  $P_i$  is nonnegative definite for all  $i = 0, 1, \dots$ , then there exists a constant  $\beta > 0$  and an integer  $k$  such that

$$(2.16) \quad Q_i^{11} > \beta I, \quad \forall i \geq k.$$

(ii) If there hold  $P_i \geq \mu I, i = 0, 1, \dots$ , for some constant  $\mu > 0$ , then there exists an integer  $k$  such that

$$(2.17) \quad Q_i > U(\mu/2), \quad \forall i \geq k,$$

where  $U(\cdot)$  is defined as in Lemma 2.2.

*Proof.* By close inspection, it can be seen that  $\lim_{i \rightarrow \infty} Q_i^{11} = \bar{Q}$ , where  $\bar{Q}$  is the unique positive definite solution to the Lyapunov equation

$$(2.18) \quad A'QA - c'c = Q.$$

From this, (i) immediately follows.

As for (ii), note that  $Q_i \geq \tilde{Q}_i$  for all  $i \geq 0$ , where  $\tilde{Q}_i$  is the solution to

$$\tilde{Q}_{i+1} = \begin{bmatrix} A' & c'b' \\ 0 & A' \end{bmatrix} \tilde{Q}_i \begin{bmatrix} A & 0 \\ bc & A \end{bmatrix} + \begin{bmatrix} c'c & 0 \\ 0 & \mu I \end{bmatrix},$$

with  $\tilde{Q}_0 = Q_0$ . Because

$$\lim_{i \rightarrow \infty} \tilde{Q}_i = U(\mu) > U(\mu/2),$$

(ii) is concluded.  $\square$

**3. General convergence analysis.** Recall that the main purpose of this paper is to solve the matrix equation

$$(3.1) \quad \mathcal{F}(X) - X\mathcal{G}(X)X = 0, \quad X \in \mathcal{P}(n),$$

with  $\mathcal{F}(\cdot), \mathcal{G}(\cdot) : \mathcal{P}(n) \mapsto \mathcal{P}(n)$  continuous operators. For this purpose, we propose the following difference equation:

$$(3.2) \quad \Phi_{i+1} = \mathcal{R}(\Phi_i),$$

with

$$(3.3) \quad \mathcal{R}(X) \triangleq X - 2[X + \mathcal{F}(X)][2X + \mathcal{F}(X) + \mathcal{G}(X)^{-1}]^{-1}[X + \mathcal{F}(X)] + 2\mathcal{F}(X).$$

Our first convergence result characterizes a set of initial values for which the solution of (3.2) will converge to the unique solution of (3.1).

**THEOREM 3.1.** *Adopt Assumptions 1 and 2 in §1. Suppose there exist  $X_1, X_2 \in \mathcal{P}(n)$  with  $X_1 \leq X_2$  such that*

$$(3.4) \quad \mathcal{F}(X_1) \geq X_1\mathcal{G}(X_1)X_1 \quad \text{and} \quad \mathcal{F}(X_2) \leq X_2\mathcal{G}(X_2)X_2.$$

Then the solution  $\Phi_i$  of (3.2) converges to the solution  $\bar{X}$  of (3.1) from any initial condition  $\Phi_0$  satisfying

$$(3.5) \quad X_1 \leq \Phi_0 \leq X_2.$$

*Proof.* Let  $\Phi_0^{(1)} = X_1$ ,  $\Phi_0^{(2)} = X_2$ , and  $\Phi_0$  satisfy (3.5). Then it follows from (2.3)–(2.4) in Lemma 2.1 that

$$\Phi_0^{(1)} \leq \mathcal{R}(\Phi_0^{(1)}) = \Phi_1^{(1)} \quad \text{and} \quad \Phi_0^{(2)} \geq \mathcal{R}(\Phi_0^{(2)}) = \Phi_1^{(2)}.$$

Thus, making use of (2.2) in Lemma 2.1 leads to

$$\Phi_0^{(1)} \leq \Phi_1^{(1)} \leq \Phi_1^{(2)} \leq \Phi_0^{(2)}.$$

By induction, the following relation is established

$$(3.6) \quad \Phi_0^{(1)} \leq \Phi_1^{(1)} \leq \dots \leq \Phi_k^{(1)} \leq \Phi_k^{(2)} \leq \dots \leq \Phi_1^{(2)} \leq \Phi_0^{(2)} \quad \forall k \geq 1.$$

Therefore,  $\lim_{k \rightarrow \infty} \Phi_k^{(1)}$  and  $\lim_{k \rightarrow \infty} \Phi_k^{(2)}$  exist and are in  $\mathcal{P}(n)$  because  $\{\Phi_k^{(1)}\}$  and  $\{\Phi_k^{(2)}\}$  are bounded above by  $\Phi_0^{(2)}$  and below by  $\Phi_0^{(1)}$ , respectively. Consequently, these two limits satisfy  $\mathcal{R}(X) = X$  and therefore are solutions of (3.1) because of (2.5) in Lemma 2.1. By Assumption 2, it turns out that

$$(3.7) \quad \lim_{k \rightarrow \infty} \Phi_k^{(1)} = \lim_{k \rightarrow \infty} \Phi_k^{(2)} = \bar{X}.$$

Further, note that  $\Phi_0^{(1)} \leq \Phi_0 \leq \Phi_0^{(2)}$ . Hence, again by induction and using (2.2) in Lemma 2.1, there results

$$\Phi_k^{(1)} \leq \Phi_k \leq \Phi_k^{(2)}, \quad \forall k \geq 0.$$

This together with (3.7) yields that  $\lim_{k \rightarrow \infty} \Phi_k = \bar{X}$ .  $\square$

*Remark 3.1.* It is worth mentioning that Theorem 3.1 is still valid if (3.5) is replaced with the existence of an integer  $N \geq 0$  such that

$$(3.8) \quad X_1 \leq \Phi_N \leq X_2.$$

*Remark 3.2.* Observe that (3.1) is equivalent to

$$(3.9) \quad \mathcal{F}_\alpha(X) - X\mathcal{G}_\alpha(X)X = 0, \quad X \in \mathcal{P}(n),$$

where  $\mathcal{F}_\alpha(X) \triangleq \mathcal{F}(X)/\alpha$  and  $\mathcal{G}_\alpha(X) \triangleq \mathcal{G}(X)/\alpha$  for any  $\alpha > 0$ . This suggests that an infinite number of algorithms can be generated for (3.1) by substituting  $\mathcal{F}_\alpha$  and  $\mathcal{G}_\alpha$  for  $\mathcal{F}$  and  $\mathcal{G}$  and setting different values to  $\alpha$ . Naturally, we might expect  $\alpha$  to play some role in achieving a satisfactory convergence rate. Later simulation results do demonstrate that a suitable choice of  $\alpha$  can significantly speed up the convergence of the algorithm.

Because the equation

$$(3.10) \quad \mathcal{G}(Y^{-1}) - Y\mathcal{F}(Y^{-1})Y = 0, \quad Y \in \mathcal{P}(n)$$

is equivalent to (3.1) with  $Y = X^{-1}$ , it follows from Theorem 3.1 that the solution of the difference equation

$$(3.11) \quad \begin{aligned} \Sigma_{i+1} &= \Sigma_i - 2[\Sigma_i + \mathcal{G}(\Sigma_i^{-1})] \\ &\quad \times [2\Sigma_i + \mathcal{G}(\Sigma_i^{-1}) + \mathcal{F}(\Sigma_i^{-1})^{-1}]^{-1} [\Sigma_i + \mathcal{G}(\Sigma_i^{-1})] + 2\mathcal{G}(\Sigma_i^{-1}) \end{aligned}$$

converges to  $\bar{X}$  from the initial condition  $\Sigma_0 = \Phi_0^{-1}$ , where  $\Phi_0$  satisfies (3.5). What is more interesting is the following result.

**THEOREM 3.2** *With the same hypotheses and notation as in Theorem 3.1, let  $\Sigma_i$  denote the solution of the difference equation (3.11) with the initial condition  $\Sigma_0 = \Phi_0^{-1}$ . Then there holds*

$$(3.12) \quad \Phi_i \Sigma_i = I, \quad i = 0, 1, \dots$$

*Proof.* Simple matrix manipulations yield

$$(3.13) \quad \begin{aligned} \Phi_{i+1} &= 2\{[\Phi_i + \mathcal{F}(\Phi_i)]^{-1} + [\Phi_i + \mathcal{G}(\Phi_i)^{-1}]^{-1}\}^{-1} - \Phi_i \\ &= \{2\{[I + \mathcal{F}(\Phi_i)\Phi_i^{-1}]^{-1} + [I + \mathcal{G}(\Phi_i)^{-1}\Phi_i^{-1}]^{-1}\}^{-1} - I\}\Phi_i \\ &= \{[I + \Phi_i\mathcal{F}(\Phi_i)^{-1}]^{-1} + [I + \Phi_i\mathcal{G}(\Phi_i)^{-1}]\} \\ (3.14) \quad &\times \{[I + \mathcal{F}(\Phi_i)\Phi_i^{-1}]^{-1} + [I + \mathcal{G}(\Phi_i)^{-1}\Phi_i^{-1}]^{-1}\}^{-1}\Phi_i. \end{aligned}$$

Similarly, we have

$$(3.15) \quad \begin{aligned} \Sigma_{i+1} &= \Sigma_i \{[I + \mathcal{G}(\Sigma_i^{-1})^{-1}\Sigma_i]^{-1} + [I + \mathcal{F}(\Sigma_i^{-1})\Sigma_i]^{-1}\} \\ &\times \{[I + \Sigma_i^{-1}\mathcal{G}(\Sigma_i^{-1})]^{-1} + [I + \Sigma_i^{-1}\mathcal{F}(\Sigma_i^{-1})^{-1}]^{-1}\}^{-1}. \end{aligned}$$

Quite obviously,  $\Phi_k \Sigma_k = I$  implies  $\Phi_{k+1} \Sigma_{k+1} = I$ . Thus, by induction, (3.12) is proved.  $\square$

Note that implementing (3.2) requires computation of the values of the operators  $\mathcal{F}$  and  $\mathcal{G}$  at each iteration. However, in some situations, the operators  $\mathcal{F}$  and  $\mathcal{G}$  may be so complicated that evaluating them is too time consuming if not impossible, which severely diminishes the efficiency of the algorithm. On the other hand, it is sometimes possible to approximate  $\mathcal{F}$  and  $\mathcal{G}$  by simpler operators somehow. In this case, it does not seem unreasonable to calculate the approximate values of  $\mathcal{F}$  and  $\mathcal{G}$  instead at each iteration. Let us now consider this strategy and address the related convergence issue in detail.

Assume that there exist two continuous operators

$$S : \mathcal{Q}(k) \times \mathcal{P}(n) \mapsto \mathcal{Q}(k) \quad \text{and} \quad T : \mathcal{Q}(l) \times \mathcal{P}(n) \mapsto \mathcal{Q}(l),$$

and two constant matrices  $K \in \mathbb{R}^{k \times n}$ ,  $L \in \mathbb{R}^{l \times n}$  such that we have the following assumptions.

*Assumption 3.*  $S$  is nondecreasing with respect to each of its arguments, and  $T$  is nondecreasing with respect to its first argument and nonincreasing with respect to its second argument.

*Assumption 4.* For any given  $X \in \mathcal{P}(n)$ , the equations

$$(3.16) \quad U = S(U, X) \quad \text{and} \quad V = T(V, X)$$

have unique solutions  $U(X) \in \mathcal{Q}(k)$  and  $V(X) \in \mathcal{Q}(l)$ , which satisfy

$$(3.17) \quad \mathcal{F}(X) = K'U(X)K \quad \text{and} \quad \mathcal{G}(X) = L'V(X)L.$$

With these assumptions, we suggest the following modified algorithm for solving (3.1):

$$(3.18) \quad \begin{aligned} \Psi_{i+1} &= \Psi_i - 2(\Psi_i + K'U_iK) \times [2\Psi_i + K'U_iK + (L'V_iL)^{-1}]^{-1} \\ &\quad \cdot (\Psi_i + K'U_iK) + 2K'U_iK, \end{aligned}$$

$$(3.19) \quad U_{i+1} = S(U_i, \Psi_i),$$

$$(3.20) \quad V_{i+1} = T(V_i, \Psi_i).$$

Its convergence property is stated below.

**THEOREM 3.3.** *Consider the system of difference equations (3.18)–(3.20) with the initial condition  $(P_0, U_0, V_0) \in \mathcal{P}(n) \times \mathcal{Q}(k) \times \mathcal{Q}(l)$ . Let Assumptions 2–4 be enforced. Suppose that there exist  $X_1, X_2 \in \mathcal{P}(n)$  with  $X_1 \leq X_2$  such that (3.4) is met. If there exists an integer  $N \geq 0$  such that*

$$(3.21) \quad X_1 \leq \Psi_N \leq X_2, \quad U(X_1) \leq U_N \leq U(X_2), \quad V(X_2) \leq V_N \leq V(X_1),$$

there holds

$$(3.22) \quad \lim_{i \rightarrow \infty} (\Psi_i, U_i, V_i) = (\bar{X}, U(\bar{X}), V(\bar{X})).$$

*Proof.* Without loss of generality we assume  $N = 0$ . Put

$$(\Psi_0^{(1)}, U_0^{(1)}, V_0^{(1)}) \triangleq (X_1, U(X_1), V(X_1)) \quad \text{and} \quad (\Psi_0^{(2)}, U_0^{(2)}, V_0^{(2)}) \triangleq (X_2, U(X_2), V(X_2)).$$

Then it is obvious that

$$U_0^{(2)} = \mathcal{S}(U_0^{(2)}, \Psi_0^{(2)}) = U_1^{(2)} \quad \text{and} \quad V_0^{(2)} = \mathcal{T}(V_0^{(2)}, \Psi_0^{(2)}) = V_1^{(2)}.$$

By (2.4) of Lemma 2.1, it follows from the second inequality of (3.4) that

$$\Psi_1^{(2)} = \mathcal{R}(\Psi_0^{(2)}) \leq \Psi_0^{(2)}.$$

Now assume that for some positive integer  $m$ ,

$$(3.23) \quad \Psi_m^{(2)} \leq \Psi_{m-1}^{(2)}, \quad U_m^{(2)} \leq U_{m-1}^{(2)}, \quad V_m^{(2)} \geq V_{m-1}^{(2)}.$$

Then by Corollary 2.1, we have  $\Psi_{m+1}^{(2)} \leq \Psi_m^{(2)}$ . Moreover,

$$\begin{aligned} U_{m+1}^{(2)} &= \mathcal{S}(U_m^{(2)}, \Psi_m^{(2)}) \leq \mathcal{S}(U_{m-1}^{(2)}, \Psi_{m-1}^{(2)}) = U_m^{(2)}, \\ V_{m+1}^{(2)} &= \mathcal{T}(V_m^{(2)}, \Psi_m^{(2)}) \geq \mathcal{T}(V_{m-1}^{(2)}, \Psi_{m-1}^{(2)}) \geq V_m^{(2)}. \end{aligned}$$

Therefore by induction, (3.23) is valid for any integer  $m \geq 0$ . In the same manner, it can be shown that

$$(3.24) \quad \Psi_m^{(1)} \geq \Psi_{m-1}^{(1)}, \quad U_m^{(1)} \geq U_{m-1}^{(1)}, \quad V_m^{(1)} \leq V_{m-1}^{(1)}, \quad m = 1, 2, \dots$$

Again from Corollary 2.1 together with (3.21), it can be inductively established that

$$(3.25) \quad \Psi_i^{(1)} \leq \Psi_i \leq \Psi_i^{(2)}, \quad U_1^{(1)} \leq U_i \leq U_i^{(2)}, \quad V_i^{(2)} \leq V_i \leq V_i^{(1)}, \quad i = 0, 1, \dots$$

As a consequence of (3.23)–(3.25), it follows that

$$(3.26) \quad \Psi_0^{(1)} \leq \Psi_1^{(1)} \leq \dots \leq \Psi_i^{(1)} \leq \Psi_i \leq \Psi_i^{(2)} \leq \dots \leq \Psi_1^{(2)} \leq \Psi_0^{(2)},$$

$$(3.27) \quad U_0^{(1)} \leq U_1^{(1)} \leq \dots \leq U_i^{(1)} \leq U_i \leq U_i^{(2)} \leq \dots \leq U_2^{(1)} \leq U_0^{(2)},$$

$$(3.28) \quad V_0^{(2)} \leq V_1^{(2)} \leq \dots \leq V_i^{(2)} \leq V_i \leq V_i^{(1)} \leq \dots \leq V_1^{(1)} \leq V_0^{(1)},$$

which imply that the limit

$$\lim_{i \rightarrow \infty} (\Psi_i^{(j)}, U_i^{(j)}, V_i^{(j)})$$

exists and is in  $\mathcal{P}(n) \times \mathcal{Q}(k) \times \mathcal{Q}(l)$  for  $j = 1, 2$ . Let  $(\Psi^{(j)}, U^{(j)}, V^{(j)})$  denote the limit. By continuity of the operators  $\mathcal{S}$  and  $\mathcal{T}$  and invertibility of  $L'V^{(j)}L$ ,  $(\Psi^{(j)}, U^{(j)}, V^{(j)})$  is a fixed point of the system of difference equations (3.18)–(3.20). In particular, we have

$$(3.29) \quad U^{(j)} = \mathcal{S}(U^{(j)}, \Psi^{(j)}) \quad \text{and} \quad V^{(j)} = \mathcal{T}(V^{(j)}, \Psi^{(j)})$$

which, by Assumption 4, leads to

$$\mathcal{F}(\Psi^{(j)}) = K'U^{(j)}K \quad \text{and} \quad \mathcal{G}(\Psi^{(j)}) = L'V^{(j)}L.$$

It turns out that  $\Psi^{(j)} = \mathcal{R}(\Psi^{(j)})$ , that is,  $\Psi^{(j)}$  satisfies (3.1). By Assumption 2, there results  $\Psi^{(j)} = \bar{X}$ . This together with (3.29) gives rise to

$$(\Psi^{(j)}, U^{(j)}, V^{(j)}) = (\bar{X}, U(\bar{X}), V(\bar{X})), \quad j = 1, 2.$$

Therefore, again from (3.26)–(3.28), (3.22) follows.  $\square$

**COROLLARY 3.1.** *With the same hypotheses as in Theorem 3.1, the solution of the second-order difference equation*

$$(3.30)$$

$$\Phi_{i+1} = \Phi_i - 2[\Phi_i + \mathcal{F}(\Phi_{i-1})][2\Phi_i + \mathcal{F}(\Phi_{i-1}) + \mathcal{G}(\Phi_{i-1})^{-1}]^{-1}[\Phi_i + \mathcal{F}(\Phi_{i-1})] + 2\mathcal{F}(\Phi_{i-1})$$

converges to the solution  $\bar{X}$  of (3.1) from any initial condition  $(\Phi_{-1}, \Phi_0) \in \mathcal{P}(n) \times \mathcal{P}(n)$  satisfying

$$(3.31) \quad X_1 \leq \Phi_{-1}, \quad \Phi_0 \leq X_2,$$

with  $X_1, X_2$  given as in Theorem 3.1.

*Proof.* Consider the system of difference equations (3.18)–(3.20) with

$$(3.32) \quad (\Psi_0, U_0, V_0) = (\Phi_0, \mathcal{F}(\Phi_{-1}), \mathcal{G}(\Phi_{-1}))$$

and

$$(3.33) \quad \mathcal{S}(U, X) \triangleq \mathcal{F}(X), \quad \mathcal{T}(V, X) \triangleq \mathcal{G}(X), \quad K = L = I.$$

Then it is straightforward to check that  $\Psi_i = \Phi_i$  for all  $i \geq 0$ , where  $\Psi_i$  denotes the first component of the solution of (3.18)–(3.20) and  $\Phi_i$  is the solution of (3.30). It is also trivial to see that the operators  $\mathcal{S}$  and  $\mathcal{T}$  fulfill Assumptions 3 and 4. Finally, note that

$$(3.34) \quad \mathcal{F}(X_1) \leq U_0 \leq \mathcal{F}(X_2) \quad \text{and} \quad \mathcal{G}(X_2) \leq V_0 \leq \mathcal{G}(X_1).$$

Thus, by Theorem 3.3, it is concluded that

$$\lim_{i \rightarrow \infty} \Phi_i = \lim_{i \rightarrow \infty} \Psi_i = \bar{X}. \quad \square$$

**Remark 3.3.** By using a similar argument, Theorem 3.1 can also be proved as a consequence of Theorem 3.3.

**4. Iterative computation of  $L^2$ -sensitivity optimal realizations and Euclidean norm-balancing realizations.** In this section, we apply the established general results to two specific problems in system realization theory. One problem is to find  $L^2$ -sensitivity optimal realizations of a given system and the other is to find Euclidean norm-balancing realizations. Several iterative algorithms are proposed and proved to possess the convergence property.

Now consider a discrete-time, single-input–single-output, stable system with a transfer function  $H(z)$  of order  $n$ . Assume that  $H(z)$  has an initial minimal realization as follows:

$$(4.1) \quad \left[ \begin{array}{c|c} A & b \\ \hline c & d \end{array} \right] \triangleq c(zI - A)^{-1}b + d.$$

The  $L^2$ -sensitivity index of the system  $H(z)$  with respect to the realization  $(A, b, c, d)$  is defined by

$$(4.2) \quad \Gamma_1(A, b, c) \triangleq \left\| \frac{\partial H}{\partial A} \right\|_2^2 + \left\| \frac{\partial H}{\partial b} \right\|_2^2 + \left\| \frac{\partial H}{\partial c} \right\|_2^2$$

$$(4.3) \quad = \frac{1}{2\pi i} \text{trace} \left\{ \oint [\mathcal{A}(z)\mathcal{A}(z)^* + \mathcal{B}(z)\mathcal{B}(z)^* + \mathcal{C}(z)^*\mathcal{C}(z)] \frac{dz}{z} \right\},$$

where

$$(4.4) \quad \mathcal{A}(z) = \left[ \begin{array}{cc|c} A & bc & 0 \\ 0 & A & I \\ \hline I & 0 & 0 \end{array} \right], \quad \mathcal{B}(z) = \left[ \begin{array}{c|c} A & b \\ \hline I & 0 \end{array} \right], \quad \mathcal{C}(z) = \left[ \begin{array}{c|c} A & I \\ \hline c & 0 \end{array} \right].$$

The  $L^2$ -sensitivity minimization problem is to find a similarity transformation  $T$  so that the  $L^2$ -sensitivity index  $\Gamma_1(TAT^{-1}, Tb, cT^{-1})$  of  $H(z)$  with respect to the transformed realization is minimized. Regarding this problem, we summarize the main known facts from [1] as follows.

*Fact 1.*  $\Gamma_1(TAT^{-1}, Tb, cT^{-1})$  achieves its minimum at  $T = T_0$  if and only if  $T_0'T_0$  is an equilibrium point of the differential equation

$$(4.5) \quad \begin{aligned} \dot{P}(t) = & \frac{1}{2\pi i} \oint \{ P(t)^{-1}[\mathcal{A}(z)^*P(t)\mathcal{A}(z) + \mathcal{C}(z)^*\mathcal{C}(z)]P(t)^{-1} \\ & - \mathcal{A}(z)P(t)^{-1}\mathcal{A}(z)^* - \mathcal{B}(z)\mathcal{B}(z)^* \} \frac{dz}{z}. \end{aligned}$$

*Fact 2.* Equation (4.5) has a unique equilibrium  $\bar{P}$  in  $\mathcal{P}(n)$ .

*Fact 3.* The solution  $P(t)$  of (4.5) exponentially converges to  $\bar{P}$  from any initial value  $P(0) \in \mathcal{P}(n)$ .

Although Fact 3 suggests that the equilibrium  $\bar{P}$  can be found by solving an initial value problem associated with (4.5), this method lacks computational efficiency and can be numerically ill conditioned, especially when the order  $n$  of the system is high.

A similar situation arises when we try to minimize the Euclidean norm defined by

$$(4.6) \quad \Gamma_2(A, b, c) \triangleq \text{trace}(AA' + bb' + c'c)$$

with respect to realizations of  $H(z)$ . There are three analogous facts [3], but here the relevant equation is

$$(4.7) \quad \dot{P}(t) = P(t)^{-1}[A'P(t)A + c'c]P(t)^{-1} - AP(t)^{-1}A' - bb'.$$

Although (4.7) looks much simpler than (4.5), likewise a computationally attractive method to find its unique equilibrium point has not been proposed.

We are in a position to present several iterative algorithms for finding the unique solution to the matrix equation

(4.8)

$$\frac{1}{2\pi i} \oint \{P^{-1}[A(z)^*PA(z) + C(z)^*C(z)]P^{-1} - A(z)P^{-1}A(z)^* - B(z)B(z)^*\} \frac{dz}{z} = 0.$$

PROPOSITION 4.1. *Given the initial minimal realization of  $H(z)$  as in (4.1) with  $\bar{P}$  denoting the solution of (4.8), define*

(4.9) 
$$W_c(P) \triangleq \frac{1}{2\pi i} \oint [A(z)P^{-1}A(z)^* + B(z)B(z)^*] \frac{dz}{z},$$

(4.10) 
$$W_o(P) \triangleq \frac{1}{2\pi i} \oint [A(z)^*PA(z) + C(z)^*C(z)] \frac{dz}{z}.$$

Let  $\Phi_i$  be the solution of the first-order difference equation

(4.11) 
$$\begin{aligned} \Phi_{i+1} &= \Phi_i - 2[\Phi_i + W_o(\Phi_i)/\alpha] \\ &\quad \times [2\Phi_i + W_o(\Phi_i)/\alpha + \alpha W_c(\Phi_i)^{-1}]^{-1} [\Phi_i + W_o(\Phi_i)/\alpha] + 2W_o(\Phi_i)/\alpha \end{aligned}$$

from an initial condition  $\Phi_0 \in \mathcal{P}(n)$  and  $\Pi_i$  the solution of the second-order difference equation

(4.12)

$$\begin{aligned} \Pi_{i+1} &= \Pi_i - 2[\Pi_i + W_o(\Pi_{i-1})/\alpha] \\ &\quad \times [2\Pi_i + W_o(\Pi_{i-1})/\alpha + \alpha W_c(\Pi_{i-1})^{-1}]^{-1} [\Pi_i + W_o(\Pi_{i-1})/\alpha] + 2W_o(\Pi_{i-1})/\alpha \end{aligned}$$

from  $(\Pi_{-1}, \Pi_0) \in \mathcal{P}(n) \times \mathcal{P}(n)$ , where  $\alpha$  is any fixed positive constant. Then there holds

(4.13) 
$$\lim_{i \rightarrow \infty} \Phi_i = \lim_{i \rightarrow \infty} \Pi_i = \bar{P}.$$

*Proof.* Letting

(4.14) 
$$\mathcal{F}(P) = W_o(P)/\alpha \quad \text{and} \quad \mathcal{G}(P) = W_c(P)/\alpha,$$

we can easily see that both  $\mathcal{F}(P)$  and  $\mathcal{G}(P)^{-1}$  are nondecreasing and continuous with respect to  $P \in \mathcal{P}(n)$  and that  $\bar{P}$  is the unique solution of  $\mathcal{F}(P) = P\mathcal{G}(P)P$  in  $\mathcal{P}(n)$ . Because for any fixed  $P \in \mathcal{P}(n)$  there hold

(4.15) 
$$\lim_{\mu \rightarrow 0^+} [\mathcal{F}(\mu P) - (\mu P)\mathcal{G}(\mu P)(\mu P)] = \frac{1}{2\pi\alpha i} \oint C(z)^*C(z) \frac{dz}{z} > 0,$$

(4.16) 
$$\lim_{\nu \rightarrow +\infty} [\mathcal{F}(\nu P) - (\nu P)\mathcal{G}(\nu P)(\nu P)]/\nu^2 = -\frac{1}{2\pi\alpha i} \oint PB(z)B(z)^*P \frac{dz}{z} < 0.$$

Thus, the theorem follows by directly applying Theorem 3.1 and Corollary 3.1. □

*Remark 4.1.* It is readily seen from Theorem 3.2 that the algorithm

(4.17)

$$\begin{aligned} \Sigma_{i+1} = & \Sigma_i - 2[\Sigma_i + W_c(\Sigma_i^{-1})/\alpha] \\ & \times [2\Sigma_i + W_c(\Sigma_i^{-1})/\alpha + \alpha W_o(\Sigma_i^{-1})^{-1}]^{-1} [\Sigma_i + W_c(\Sigma_i^{-1})/\alpha] + 2W_c(\Sigma_i^{-1})/\alpha \end{aligned}$$

with the initial value  $\Sigma_0 \in \mathcal{P}(n)$  also provides an alternative way to compute the equilibrium  $\bar{P}$ .

Note that the calculation of  $W_c(P)$  and  $W_o(P)$  inevitably involves intensive iterations given a  $P$ . To overcome this drawback, we propose the following modified algorithm, which only needs to evaluate much simpler expressions than  $W_c(P)$  and  $W_o(P)$  at each iteration.

**PROPOSITION 4.2.** *Adopt the same hypotheses and notation as in Proposition 4.1. Then for any given  $\alpha > 0$  and initial condition  $(\Psi_0, U_0, V_0) \in \mathcal{P}(n) \times \mathcal{P}(2n) \times \mathcal{P}(2n)$ , the solution  $(\Psi_i, U_i, V_i)$  of the system of difference equations*

$$(4.18) \quad \begin{aligned} \Psi_{i+1} = & \Psi_i - 2(\Psi_i + U_i^{11}/\alpha) \\ & \times [2\Psi_i + U_i^{11}/\alpha + \alpha(V_i^{11})^{-1}]^{-1} (\Psi_i + U_i^{11}/\alpha) + 2U_i^{11}/\alpha, \end{aligned}$$

(4.19)

$$U_{i+1} \triangleq \begin{bmatrix} U_{i+1}^{11} & U_{i+1}^{12} \\ U_{i+1}^{21} & U_{i+1}^{22} \end{bmatrix} = \begin{bmatrix} A' & c'b' \\ 0 & A' \end{bmatrix} \begin{bmatrix} U_i^{11} & U_i^{12} \\ U_i^{21} & U_i^{22} \end{bmatrix} \begin{bmatrix} A & 0 \\ bc & A \end{bmatrix} + \begin{bmatrix} c'c & 0 \\ 0 & \Psi_i \end{bmatrix},$$

(4.20)

$$V_{i+1} \triangleq \begin{bmatrix} V_{i+1}^{11} & V_{i+1}^{12} \\ V_{i+1}^{21} & V_{i+1}^{22} \end{bmatrix} = \begin{bmatrix} A & bc \\ 0 & A \end{bmatrix} \begin{bmatrix} V_i^{11} & V_i^{12} \\ V_i^{21} & V_i^{22} \end{bmatrix} \begin{bmatrix} A' & 0 \\ c'b' & A' \end{bmatrix} + \begin{bmatrix} bb' & 0 \\ 0 & \Psi_i^{-1} \end{bmatrix}$$

converges to its unique fixed point  $(\bar{P}, \bar{U}, \bar{V}) \in \mathcal{P}(n) \times \mathcal{P}(2n) \times \mathcal{P}(2n)$ .

*Proof.* Define the operators  $S$  and  $T$  on  $\mathcal{Q}(2n) \times \mathcal{P}(n)$  by

$$(4.21) \quad S(U, X) \triangleq \begin{bmatrix} A' & c'b' \\ 0 & A' \end{bmatrix} U \begin{bmatrix} A & 0 \\ bc & A \end{bmatrix} + \begin{bmatrix} c'c & 0 \\ 0 & X \end{bmatrix},$$

$$(4.22) \quad T(V, X) \triangleq \begin{bmatrix} A & bc \\ 0 & A \end{bmatrix} V \begin{bmatrix} A' & 0 \\ c'b' & A' \end{bmatrix} + \begin{bmatrix} bb' & 0 \\ 0 & X^{-1} \end{bmatrix},$$

and let  $L = K = \begin{bmatrix} I \\ 0 \end{bmatrix} \in \mathbb{R}^{2n \times n}$  with  $I$  an  $n \times n$  identity matrix. Because  $(A, b, c)$  is minimal, the operators  $S$  and  $T$  continuously map  $\mathcal{P}(2n) \times \mathcal{P}(n)$  into  $\mathcal{P}(2n)$ . Quite obviously,  $S$  and  $T$  meet Assumption 3 in the previous section. Moreover, it is clear that the Lyapunov equations

$$U = S(U, X) \quad \text{and} \quad V = S(V, X)$$

have unique solutions  $U(X) \in \mathcal{P}(2n)$  and  $V(X) \in \mathcal{P}(2n)$  for any  $X \in \mathcal{P}(n)$ . In addition to this, it is known from [5] that

$$\mathcal{F}(X) = K'U(X)K \quad \text{and} \quad \mathcal{G}(X) = L'V(X)L,$$

where  $\mathcal{F}(\cdot)$  and  $\mathcal{G}(\cdot)$  are defined as in (4.14). Thus, Assumption 4 is fulfilled.

Next, it is routine to check inductively that  $\Psi_i > 0$  for all  $i \geq 0$ . Hence, from (i) of Lemma 2.3, there exists an integer  $k_1$  such that

$$(4.23) \quad U_i^{11}, V_i^{11} > \beta I, \quad \forall i \geq k_1$$

for some constant  $\beta > 0$ . By Corollary 2.1, this implies that  $\Psi_i > (\beta/\alpha)I$  for all  $i > k_1$ . Making use of (ii) of Lemma 2.3 yields that there exists  $k_2 > k_1$  such that

$$(4.24) \quad U_i > \tilde{U} \quad \text{and} \quad V_i > \tilde{V}, \quad \forall i \geq k_2,$$

where  $\tilde{U}$  and  $\tilde{V}$  are the solutions to the Lyapunov equations

$$(4.25) \quad U = \begin{bmatrix} A' & c'b' \\ 0 & A' \end{bmatrix} U \begin{bmatrix} A & 0 \\ bc & A \end{bmatrix} + \begin{bmatrix} c'c & 0 \\ 0 & 0 \end{bmatrix},$$

$$(4.26) \quad V = \begin{bmatrix} A & bc \\ 0 & A \end{bmatrix} V \begin{bmatrix} A' & 0 \\ c'b' & A' \end{bmatrix} + \begin{bmatrix} bb' & 0 \\ 0 & 0 \end{bmatrix},$$

respectively. Now by Lemma 2.2, for sufficiently small  $\mu > 0$  and sufficiently large  $\nu > 0$  there hold

$$(4.27) \quad U_{k_2} < U(\nu I) \quad \text{and} \quad V_{k_2} < V(\mu I).$$

Because

$$(4.28) \quad \lim_{\mu \rightarrow 0} U(\mu I) = \tilde{U} \quad \text{and} \quad \lim_{\nu \rightarrow \infty} V(\nu I) = \tilde{V},$$

it is seen from (4.24) that for sufficiently small  $\mu > 0$  and sufficiently large  $\nu > 0$  there hold

$$(4.29) \quad U_{k_2} > U(\mu I) \quad \text{and} \quad V_{k_2} > V(\nu I).$$

Also, it is clear that

$$(4.30) \quad \mu I < \Psi_{k_2} < \nu I$$

for sufficiently small  $\mu > 0$  and sufficiently large  $\nu > 0$ . In view of (4.15)–(4.16), a direct application of Theorem 3.3 leads to

$$\lim_{i \rightarrow \infty} (\Psi_i, U_i, V_i) = (\bar{P}, U(\bar{P}), V(\bar{P})). \quad \square$$

Finally, regarding the Euclidean norm-balancing problem with an initial minimal realization  $(A, B, C)$ , we claim that with the new definitions

$$(4.31) \quad W_o(P) \triangleq A'PA + C'C \quad \text{and} \quad W_c(P) \triangleq AP^{-1}A' + BB',$$

(4.11) and (4.12) are convergent to the unique solution  $\mathbb{P}$  of

$$(4.32) \quad (A'PA + C'C) - P(AP^{-1}A' + BB')P = 0$$

in  $\mathcal{P}(n)$ . In fact, it suffices to verify by Theorem 3.1 that for any  $P \in \mathcal{P}(n)$  there exist  $P_1, P_2 \in \mathcal{P}(n)$  with  $P_1 \leq P \leq P_2$  such that

$$(4.33) \quad W_o(P_1) \geq P_1 W_c(P_1) P_1 \quad \text{and} \quad W_o(P_2) \leq P_2 W_c(P_2) P_2.$$

However, this is true upon noting that

$$\begin{aligned} W_o(\mu\mathbb{P}) - (\mu\mathbb{P})W_c(\mu\mathbb{P})(\mu\mathbb{P}) &= \mu A'\mathbb{P}A + C'C - \mu^2\mathbb{P}(\mu^{-1}A\mathbb{P}^{-1}A' + BB')\mathbb{P} \\ &= \mu(A'\mathbb{P}A - \mathbb{P}A\mathbb{P}^{-1}A'\mathbb{P}) + (C'C - \mu^2\mathbb{P}BB'\mathbb{P}) \\ &= \mu(\mathbb{P}BB'\mathbb{P} - C'C) + (C'C - \mu^2\mathbb{P}BB'\mathbb{P}) \\ &= (1 - \mu)(C'C + \mu\mathbb{P}BB'\mathbb{P}). \end{aligned}$$

Hence, the claim is true.

**5. A result on convergence rate.** In this section we prove that the convergence of (4.11) is locally exponential by showing that their unique equilibrium is sink. In other words, the linearization of (4.11) at the equilibrium has all its eigenvalues in the open unit disk.

PROPOSITION 5.1. *The linearization of (4.11) is asymptotically stable at the fixed point  $\bar{P}$ .*

*Proof.* We might as well assume  $\alpha = 1$ . Now with

(5.1)

$$S_1(X) \triangleq [X + W_o(X)]^{-1}, \quad S_2(X) = [X + W_c(X)^{-1}]^{-1}, \quad T(X) = [S_1(X) + S_2(X)]^{-1},$$

(4.11) can be rewritten as

$$(5.2) \quad P_{i+1} = \mathcal{R}(P_i) \triangleq 2T(P_i) - P_i.$$

Note that  $S_1(X), S_2(X), T(X)$  are defined for the realization  $(A, b, c)$ . Accordingly, we can define  $\mathbb{A}(z), \mathbb{W}_o(X)$ , and  $\mathbb{S}_1$ , and so on for the  $L^2$ -sensitivity optimal realization  $(\bar{P}^{\frac{1}{2}}A\bar{P}^{-\frac{1}{2}}, \bar{P}^{\frac{1}{2}}b, c\bar{P}^{-\frac{1}{2}})$ . Then it is routine to check the following relations:

$$(5.3) \quad \bar{P}^{\frac{1}{2}}S_1(\bar{P})\bar{P}^{\frac{1}{2}} = \mathbb{S}_1(I), \quad \bar{P}^{\frac{1}{2}}S_2(\bar{P})\bar{P}^{\frac{1}{2}} = \mathbb{S}_2(I), \quad \bar{P}^{-\frac{1}{2}}T(\bar{P})\bar{P}^{-\frac{1}{2}} = \mathbb{T}(I).$$

Because  $\mathcal{R}(X)$  is an operator from  $R^{n \times n}$  to  $R^{n \times n}$ , its Fréchet derivative at  $X = \bar{P}$  is a linear operator from  $R^{n \times n}$  to  $R^{n \times n}$  given by

$$(5.4) \quad \begin{aligned} \mathcal{R}'(\bar{P})X &= 2T(\bar{P})\{S_1(\bar{P})[X + W'_o(\bar{P})X]S_1(\bar{P}) \\ &\quad + S_2(\bar{P})[X - W'_c(\bar{P})^{-1}W'_c(\bar{P})XW_c(\bar{P})^{-1}]S_2(\bar{P})\}T(\bar{P}) - X, \end{aligned}$$

with

(5.5)

$$W'_o(\bar{P})X = \frac{1}{2\pi i} \oint \mathcal{A}(z)^* X \mathcal{A}(z) \frac{dz}{z} \quad \text{and} \quad W'_c(\bar{P})X = -\frac{1}{2\pi i} \oint \mathcal{A}(z) \bar{P}^{-1} X \bar{P}^{-1} \mathcal{A}(z)^* \frac{dz}{z}.$$

Using (5.3), we have

(5.6)

$$\begin{aligned} &\bar{P}^{-\frac{1}{2}}[\mathcal{R}'(\bar{P})X]\bar{P}^{-\frac{1}{2}} \\ &= 2\mathbb{T}(I)\{\mathbb{S}_1(I) \left[ \bar{P}^{-\frac{1}{2}}X\bar{P}^{-\frac{1}{2}} + \frac{1}{2\pi i} \oint \mathcal{A}(z)^*(\bar{P}^{-\frac{1}{2}}X\bar{P}^{-\frac{1}{2}})\mathcal{A}(z) \frac{dz}{z} \right] \mathbb{S}_1(I) \\ &\quad + \mathbb{S}_2(I) \left[ \bar{P}^{-\frac{1}{2}}X\bar{P}^{-\frac{1}{2}} + \mathbb{W}_c(I)^{-1} \frac{1}{2\pi i} \oint \mathcal{A}(z)\bar{P}^{-\frac{1}{2}}X\bar{P}^{-\frac{1}{2}}\mathcal{A}(z)^* \frac{dz}{z} \mathbb{W}_c(I)^{-1} \right] \\ &\quad \cdot \mathbb{S}_2(I)\}\mathbb{T}(I) - \bar{P}^{-\frac{1}{2}}X\bar{P}^{-\frac{1}{2}} \\ &= \mathbb{R}'(I)(\bar{P}^{-\frac{1}{2}}X\bar{P}^{-\frac{1}{2}}), \end{aligned}$$

from which it is not hard to see that the linear operator  $\mathcal{R}'(\bar{P})$  is asymptotically stable in the discrete time sense if and only if  $\mathbb{R}'(I)$  is also. Because  $\mathbb{W}_c(I) = \mathbb{W}_o(I)$ , it follows that

$$\mathbb{R}'(I)X$$

$$(5.7) \quad = 2\mathbb{T}\mathbb{S}_1 \left[ X + \mathbb{W}_o X \mathbb{W}_o + \frac{1}{2\pi i} \oint [\mathbb{A}(z)^* X \mathbb{A}(z) + \mathbb{A}(z) X \mathbb{A}(z)^*] \frac{dz}{z} \right] \mathbb{S}_1 \mathbb{T} - X$$

$$(5.8) \quad = 2\mathbb{S}_1 \left[ X + \mathbb{W}_o X \mathbb{W}_o + \frac{1}{2\pi i} \oint [\mathbb{A}(z)^* X \mathbb{A}(z) + \mathbb{A}(z) X \mathbb{A}(z)^*] \frac{dz}{z} \right] \mathbb{S}_1 - X,$$

where  $\mathbb{T}$  is understood to be  $\mathbb{T}(I)$  and so on. Thus, the matrix representation of  $\mathbb{R}'(I)$  is expressed by

$$(5.9) \quad 2(\mathbb{S}_1 \otimes \mathbb{S}_1) \left\{ I + \mathbb{W}_o \otimes \mathbb{W}_o + \frac{1}{2\pi i} \oint [\mathbb{A}(z)^* \otimes \mathbb{A}(z)^\tau + \mathbb{A}(z) \otimes \mathbb{A}(\bar{z})] \frac{dz}{z} \right\} - I.$$

It remains to show that this matrix has all its eigenvalues in the open unit disk, which is equivalent to saying that so is the matrix

$$\Gamma \triangleq 2(\mathbb{S}_1^{1/2} \otimes \mathbb{S}_1^{1/2}) \left\{ I + \mathbb{W}_o \otimes \mathbb{W}_o + \frac{1}{2\pi i} \oint [\mathbb{A}(z)^* \otimes \mathbb{A}(z)^\tau + \mathbb{A}(z) \otimes \mathbb{A}(\bar{z})] \frac{dz}{z} \right\} \cdot (\mathbb{S}_1^{1/2} \otimes \mathbb{S}_1^{1/2}) - I.$$

Since  $\Gamma$  is symmetric and  $\Gamma > -I$ , it suffices to prove that  $\Gamma < I$ . To do this, note that

$$\begin{aligned} \Gamma < I &\iff I + \mathbb{W}_o \otimes \mathbb{W}_o + \frac{1}{2\pi i} \oint [\mathbb{A}(z)^* \otimes \mathbb{A}(z)^\tau + \mathbb{A}(z) \otimes \mathbb{A}(\bar{z})] \frac{dz}{z} < \mathbb{S}_1^{-1} \otimes \mathbb{S}_1^{-1} \\ &\iff \frac{1}{2\pi i} \oint [\mathbb{A}(z)^* \otimes \mathbb{A}(z)^\tau + \mathbb{A}(z) \otimes \mathbb{A}(\bar{z})] \frac{dz}{z} < \mathbb{W}_o \otimes I + I \otimes \mathbb{W}_o \\ &\iff \frac{1}{2\pi i} \oint [\mathbb{A}(z)^* \otimes \mathbb{A}(z)^\tau + \mathbb{A}(z) \otimes \mathbb{A}(\bar{z})] \frac{dz}{z} \\ &< \frac{1}{2\pi i} \oint [\mathbb{A}(z)^* \mathbb{A}(z) \otimes I^\tau + I \otimes \mathbb{A}(\bar{z}) \mathbb{A}(z)^\tau + \mathbb{C}(z)^* \mathbb{C}(z) \otimes I + I \otimes \mathbb{B}(z) \mathbb{B}(z)^*] \frac{dz}{z} \\ &\iff \frac{1}{2\pi i} \oint (\mathbb{A}(z) \otimes I - I \otimes \mathbb{A}(z)^\tau)^* (\mathbb{A}(z) \otimes I - I \otimes \mathbb{A}(z)^\tau) \frac{dz}{z} + \Sigma_o \otimes I + I \otimes \Sigma_c \\ &> 0, \end{aligned}$$

where  $\Sigma_o$  and  $\Sigma_c$  denote the observability and controllability Gramians. Hence, it follows that  $\Gamma < I$ .  $\square$

**6. Simulation results.** The purpose of this section is to demonstrate the effectiveness of the algorithms proposed in the previous section by simulation on a SPARCstation. To do this, consider a specific minimal statespace realization  $(A, b, c)$  with

$$(6.1) \quad A = \begin{bmatrix} 0.5 & 0 & 1.0 \\ 0 & -0.25 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad c = [1 \quad 5 \quad 10].$$

Recall that there exists a unique positive definite matrix  $\bar{P} \in \mathbb{R}^{3 \times 3}$  such that the realization  $(TAT^{-1}, Tb, cT^{-1})$  is  $L^2$ -sensitivity optimal for any similarity transformation  $T$  with  $T'T =$

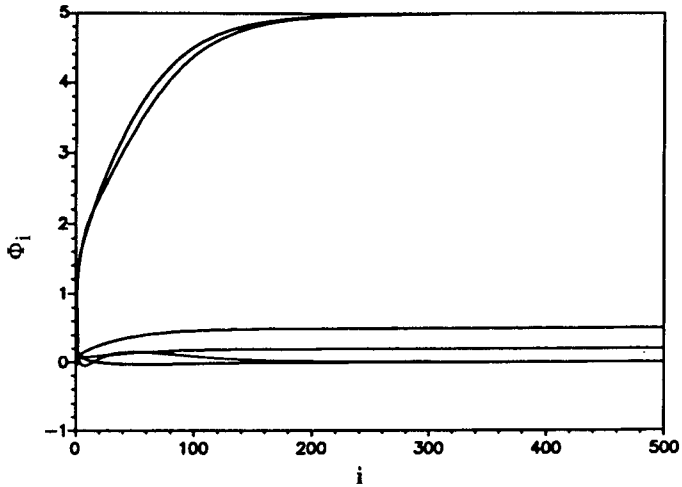


FIG. 6.1. The trajectory of  $\Phi_i$  of (4.11) with  $\alpha = 300$ .

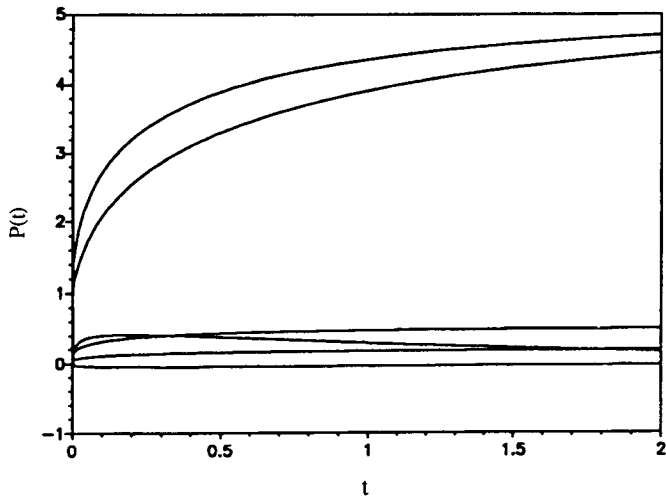


FIG. 6.2. The trajectory of  $P(t)$  of (4.5).

$\bar{P}$ . It turns out that  $\bar{P}$  is exactly given by

$$(6.2) \quad \bar{P} = \begin{bmatrix} 0.2 & 0 & 0.5 \\ 0 & 5.0 & 0 \\ 0.5 & 0 & 5.0 \end{bmatrix},$$

which indeed satisfies (4.8).

We first take (4.11) with  $\alpha = 300$  and implement it starting from the identity matrix using MATLAB. The resulting trajectory  $\Phi_i$  during the first 500 iterations is shown in Fig. 6.1 and is clearly seen to converge very fast to  $\bar{P}$ . The time taken for this implementation is less than three minutes. In contrast, if an ordinary differential equation (ODE) algorithm in MABLAB is used to solve (4.5) with the same initial condition, it is found that it takes about 45 minutes to compute the solution  $P(t)$  on the time interval  $[0, 2]$ , which is depicted in Fig. 6.2. In fact, more than 2,400 iterations are performed during that time interval. Even so, the solution does not appear to be close enough to  $\bar{P}$  though it very slowly tends to  $\bar{P}$ .

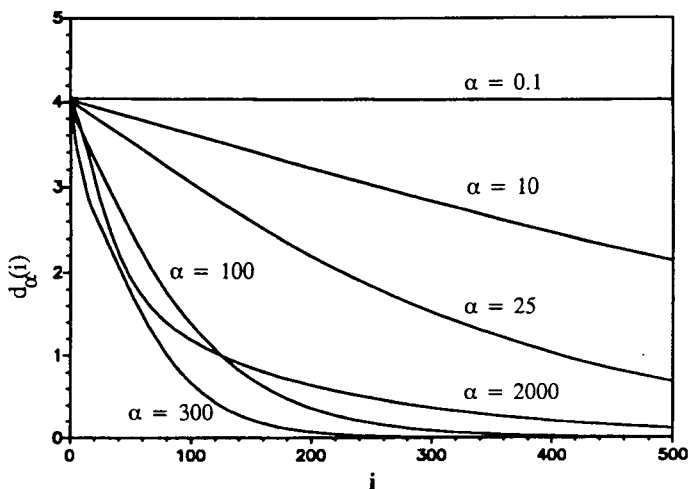


FIG. 6.3. Effect of different  $\alpha$  on the convergence rate of (4.11).

Next, we examine the effect of  $\alpha$  on the convergence rate (4.11). For this purpose, define the deviation between  $\Phi_i$  and the true solution  $\bar{P}$  in (6.2) as

$$(6.3) \quad d_\alpha(i) = \|\Phi_i - \bar{P}\|_2,$$

where  $\|\cdot\|_2$  denotes the spectral norm of a matrix. Implement (4.11) with

$$\alpha = 0.1, 10, 25, 100, 300, 2000,$$

respectively, and depict the evolution of the associated deviation  $d_\alpha(i)$  for each  $\alpha$  in Fig. 6.3. Then we can see that  $\alpha = 300$  is the best choice. In addition, as long as  $\alpha \leq 300$ , the larger  $\alpha$ , the faster the convergence of the algorithm. On the other hand, it should be observed that a larger  $\alpha$  is not always better than a smaller  $\alpha$  and that too small an  $\alpha$  can make the convergence extremely slow.

Finally, let us turn to two algorithms (4.12) and (4.18)–(4.20) with  $\alpha = 300$ . All the initial matrices required for the implementation are set to identity matrices of appropriate dimension. Define

$$f(i) = \|\Pi_i - \bar{P}\|_2 \quad \text{and} \quad g(i) = \|\Psi_i - \bar{P}\|_2$$

as the deviations from the true solution  $\bar{P}$  for the two algorithms, respectively. Their evolutions are depicted in Fig. 6.4 and manifestly exhibit the convergence of the algorithms. Indeed, the algorithm (4.18)–(4.20) is fastest in terms of the execution time, but with the same number of iterations it does not produce a solution as satisfactory as (4.11) or (4.12).

Some concluding remarks are in order.

*Remark 6.1.* Adding a scalar factor to (4.5) does not help much in reducing the CPU time required for solving it on a digital computer.

*Remark 6.2.* Because  $\alpha$  does play an important role in speeding up the algorithms, it is worthwhile to do further study to find some helpful guidelines for choosing a suitable  $\alpha$ .

*Remark 6.3.* It appears that the proposed algorithms are quite robust against nonsymmetric or indefinite disturbances. This is demonstrated by implementing (4.11) with  $\alpha = 300$  and the initial matrix

$$\Phi_0 = \begin{bmatrix} 1 & 10 & -10 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

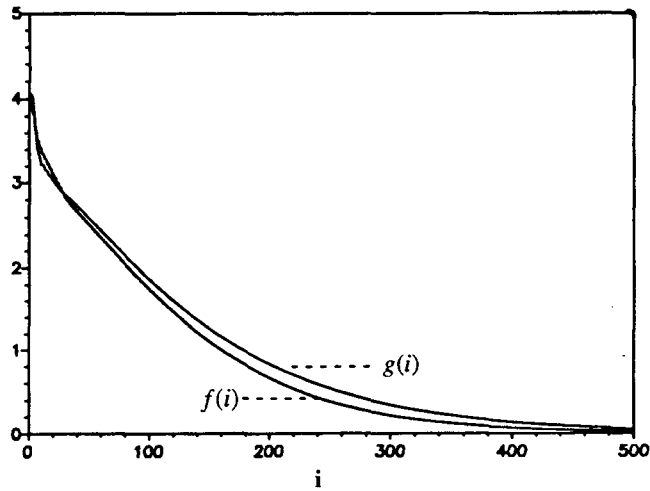


FIG. 6.4. Convergence of algorithms (4.12) and (4.18)–(4.20) with  $\alpha = 300$ .

which is obviously nonsymmetric and indefinite. It turns out that the effect of the nonsymmetry and indefiniteness almost completely vanishes after 30 iterations. Afterwards, the algorithm converges to  $\bar{P}$  with no oscillatory behavior.

**7. Conclusions.** Two types of difference equation have been proposed and studied with the aim of solving a class of nonlinear matrix equations. The main contribution of this paper is twofold. First, we characterize a set of initial conditions from which the solution of the proposed difference equations is guaranteed to converge to the solution of the matrix equation of concern. Second, the general results have been successfully employed to derive a number of efficient iterative algorithms for finding  $L^2$ -sensitivity optimal realizations and Euclidean norm-balancing realizations. These algorithms are simple to implement without any requirement of step-size adjustment. Another feature of them is that their convergence rate can be significantly improved by proper choice of a constant scalar in advance. In addition, the convergence of one algorithm is locally exponential. The effectiveness of the algorithms are demonstrated by simulation.

#### REFERENCES

- [1] U. HELMKE AND J. B. MOORE,  $L^2$  sensitivity minimization of linear system representations via gradient flows, *J. Math. Systems Control Theory*, to appear.
- [2] K. L. HITZ AND B. D. O. ANDERSON, *Iterative method of computing the limiting solution of the matrix Riccati differential equation*, *Proceedings of IEE*, 119 (1972), pp. 1402–1406.
- [3] J. E. PERKINS, U. HELMKE, AND J. B. MOORE, *Balanced realizations via gradient flows*, *Systems Control Lett.*, 14 (1990), pp. 369–380.
- [4] L. T. WATSON, *Engineering Applications of the Chow–Yorke Algorithm, Homotopy Methods and Global Convergence*, Plenum Press, New York, 1983, pp. 287–307.
- [5] W.-Y. YAN AND J. B. MOORE, *On  $L^2$ -sensitivity minimization of linear state-space systems*, *IEEE Trans. Circuits and Systems*, 39 (1992), pp. 641–648.