Banded Matrix Fraction Representation of Triangular Input Normal Pairs

Andrew P. Mullhaupt* and Kurt S. Riedel† *

August 6, 2001

Abstract

An input pair (A, B) is triangular input normal if and only if A is triangular and $AA^* + BB^* = \mathbb{I}_n$, where $\mathbb{I}_n$ is the identity matrix. Input normal pairs generate an orthonormal basis for the impulse response. Every input pair may be transformed to a triangular input normal pair. A new system representation is given: (A, B) is triangular normal and A is a matrix fraction, $A = M^{-1}N$, where M and N are triangular matrices of low bandwidth. For single input pairs, M and N are bidiagonal and an explicit parameterization is given in terms of the eigenvalues of A. This band fraction structure allows for fast updates of state space systems and fast system identification. When A has only real eigenvalues, one state advance requires $3n$ multiplications for the single input case.

EDICS Numbers: 2-SREP, 2-SYSM

Key words. System representations, state space, balanced systems, orthonormal representations, system identification

1 INTRODUCTION.

An increasingly popular technique in system identification is the use of orthonormal forms with fixed denominators [12, 19, 2, 13, 17, 15, 14]. In this approach, the

*S.A.C. Capital Management, LLC, 540 Madison Ave, New York, NY 10022, † Present Address: Millennium Partners, 666 Fifth Ave, New York 10103

unknown impulse response, $G(q;\{c_k\})$, is represented as a linear sum of orthonormal basis functions:

$$G(q,\{c_k\}) = \sum_{k=1}^{n} c_k \mathcal{B}_k(q) \ , \tag{1.1}$$

where q is the forward shift operator. Here $\mathcal{B}_k(q)$ are prescribed rational transfer functions with poles $\{\lambda_j, j = 1 \ldots n\}$. The unknown coefficients are then identified using least squares identification or a robust analog [6]. The utility of this approach is now widely documented [14, 15]. In system identification, the use of orthonormal pairs improves the condition of the estimate [14, 11]. We specialize to the case where $\mathcal{B}_k(q)$ corresponds to the kth row in the semi-infinite impulse response matrix, $\Omega(A,B)$:

$$\Omega(A,B) \equiv (B, AB, A^2B \ldots) \ . \tag{1.2}$$

Here A is the $n \times n$ state transition matrix and B is a $n \times d$ matrix with $n \geq d$. Let C be the $p \times n$ matrix whose kth column is c_k. The jth term or lead of the impulse response in (1.1) is $CA^{j-1}B$.

In this paper, we develop a computationally convenient representation of the state transition matrices that arise from orthonormal bases. We then show how these representations allow the rapid identification of orthonormal single input models with prescribed poles. Our system representation results are general and well-suited for filter design as well. In the single input case, our filters are a particular realization of the orthogonal filters of [2, 13].

In system representation theory, one seeks families of system triples, (A,B,C), that parameterize the set of all transfer functions subject to the similarity equivalence. (Two systems are equivalent if there exists an invertible matrix, T, such that $(A',B',C') = (TAT^{-1}, TB, CT^{-1})$). We examine representations of the input pair, (A,B). The input pair is in triangular input normal (TIN) form if and only if $AA^* + BB^* = \mathbb{I}_n$, with A *lower triangular* (LT) and $\mathbb{I}_n$ is the identity matrix. The input normal condition is equivalent to the row orthogonality of the semi-infinite impulse response matrix, Ω.

The state transition matrix A is represented as a matrix fraction: $A = M^{-1}N$. Our results are not related to the extensive literature on representations of the transfer function $G(q)$ as a matrix fraction: $G(q) = N(q)/D(q)$. We show that generic linear state space systems with forcing ϵ_t have an equivalent representation of the form:

$$Mz_{t+1} = Nz_t + \hat{B}\,\epsilon_t \ , \tag{1.3}$$

where M and N are lower triangular matrices with bandwidth d. In addition, M and N are chosen such that *the covariance matrix of the state vector z_t tends to the identity matrix* for white noise forcing.

We believe that our work is the first systematic study of system representations, where A is a matrix fraction of banded matrices, $A = M^{-1}N$. The band fraction, $A = M^{-1}N$, is never computed in practice. To advance the state vector, we first compute $r_t \equiv N z_t + \hat{B}\,\epsilon_t$ and then solve the banded matrix equation $M z_{t+1} = r_t$. Constructing r_t requires $n(d+1)$ multiplications and solving for z_{t+1} requires nd multiplications since $M_{i,i} = 1$. Thus the total multiplication count for a state vector update is $(2d+1)n$. For SISO systems with real eigenvalues, the state transition matrix is identical to the cascade systems proposed in [2]. However, the parameterization/realizations of the (A, B) pair differ.

In Section 2, the existence and identifiability of TIN pairs is examined. Generically, TIN pairs are unique up to reordering the eigenvalues and phase rotations of the coordinates: $B_{j,k} \leftarrow B_{j,k} \exp i(\theta_j)$, $A_{i,j} \leftarrow A_{i,j} \exp i(\theta_i - \theta_j)$.

In Sections 3-4, we prove that generic TIN pairs have a band fraction representation: $A = M^{-1}N$, $B = M^{-1}\tilde{B}$, where M and N are triangular matrices of bandwidth d. We derive the eigenvectors of A in the $d = 1$ case and prove the numerical stability of the representation when the eigenvalues of A are order in increasing magnitude. In Section 5, we discuss the numerical details of the band fraction representation and other realizations of IN filters. We relate our band fraction representation in the $d = 1$ case, to the orthonormal basis functions of [13]. In Section 6, we show the advantage of our representations in system identification.

Notation: In the remainder of the article, we suppress the word 'lower". ('Triangular" means lower triangular.) Here A is a $n \times n$ matrix with eigenvalues $\{\lambda_i\}$ By $A_{i:j,k:m}$, we denote the $(j-i+1) \times (m-k+1)$ subblock of A from row i to row j and from column k to column m. We abbreviate $A_{i:j,1:n}$ by $A_{i:j,:}$. The matrix A has upper bandwidth d if $A_{i,j} = 0$ when $j > i + d$. We denote by $(B|A)$ the $n \times (n+d)$ matrix formed by the concatenation of B and A. 'Stable" means $|\lambda| < 1$. The $n \times n$ identity matrix is $\mathbb{I}_n$ and e_k is the unit vector in the kth coordinate. We define the equivalence $(A, B) \approx (A_2, B_2)$ when $(A_2 \equiv EAE^{-1}, B_2 \equiv EB)$, where E_n is a diagonal unitary matrix: $E_{j,k} = \exp(i\theta_j)\delta_{j-k}$.

2 TRIANGULAR INPUT NORMAL PAIRS

In this section, we present the fundamental representation results for the reduction to triangular normal form.

Definition 2.1 *An input pair, (A, B), is input normal (IN) if and only if*

$$\mathbb{I}_n - AA^* = BB^* \quad . \tag{2.1}$$

The input pair is a TIN pair if it is input normal and A is lower triangular.

IN pairs are not required to be stable or controllable. (From (2.1), A must be at least marginally stable.) In [8], 'input balanced" has a more restrictive definition of (2.1) and the additional requirement that the observability Grammian be diagonal. We do not impose any such condition on the observability Grammian. We choose this language so that 'normal" denotes restrictions on only one Grammian while 'balanced" denotes simultaneous restrictions on both Grammians.

Thus (A, B) is an IN pair if and only if the concatenated $n \times (n+d)$ matrix $(A|B)$ is row orthogonal. When (A, B) is input normal, then the identity matrix solves Stein's equation (also known as the discrete Lyapunov equation):

$$P - APA^* = BB^* \quad . \tag{2.2}$$

For stable A, $P = \sum_{j=0}^{\infty} A^j BB^* A^{j*}$, So (A, B) is input normal when $\Omega(A, B)^*\Omega(A, B)$ $= \mathbb{I}_n$ or equivalently when the basis, $\{\mathcal{B}_k(q)\}$, constitutes an orthonormal set.

Lemma 2.2 *Let (A, B) and $(A', B') = (TAT^{-1}, TB)$ be equivalent IN pairs, with (A, B) stable and controllable. Then T is unitary.*

Proof: Both $\mathbb{I}_n$ and $T^{-1}T^{-*}$ solve the Stein equation. Since (A, B) is stable and controllable, the solution of Stein's equation is unique. ∎

Theorem 2.3 *Every stable, controllable input pair (A, B), is similar to a lower triangular input normal pair $(\tilde{A} \equiv TAT^{-1}, \tilde{B} \equiv TB)$ with $\|\tilde{B}\|^2 \leq 1$. The order of the eigenvalues of $\tilde{A}$ may be specified arbitrarily. If (A, B) is real and A has real eigenvalues, then $(\tilde{A}, \tilde{B})$ and T may be chosen to be real.*

Proof: Let L be the unique Cholesky lower triangular factor of P with positive diagonal entries: $P = LL^*$. Here L is invertible. We set $\hat{A} = L^{-1}AL$ and $\hat{B} = L^{-1}B$. Note $(\hat{A}, \hat{B})$ satisfies (2.1). By Schur's unitary triangularization theorem (Horn and Johnson 2.3.1), there exists a unitary matrix, Q such that $Q\hat{A}Q^*$ is lower triangular. The proof of the real Schur form is in Sec. 7.4.1 of [1] and Theorem 2.3.4 of [4]. The eigenvalues of $\tilde{A}$ may be placed in any order [4]. Clearly, $\tilde{A} = Q\hat{A}Q^*$ and $\tilde{B} = Q\hat{B}$ satisfy (2.1) and therefore $\|\tilde{B}\|^2 \leq 1$. ∎

We now study the number of equivalent TIN pairs:

Theorem 2.4 *Let A and $\hat{A}$ be $n \times n$ LT matrices with $A_{ii} = \hat{A}_{ii}$ and let A and $\hat{A}$ be unitarily similar: $\hat{A}U = UA$ with U unitary. Let m be the number of distinct eigenvalues. Partition A, $\hat{A}$ and U into m blocks corresponding to the repeated eigenvalue blocks. Let n_i be the multiplicity of the ith eigenvalue, $1 \leq i \leq m$. Then U has block diagonal form: $U = U_1 \oplus U_2 \oplus \ldots \oplus U_m$, where U_i is a unitary matrix of size $n_i \times n_i$.*

Proof: From $\hat{A}_{m,m}U_{m,1} = U_{m,1}A_{1,1}$. If $m > 1$, then $\hat{A}_{m,m}$ and $A_{1,1}$ have no common eigenvalues. By Lemma 7.1.5 of [1], $U_{m,1} \equiv 0$. Repeating this argument shows $U_{m-k,1} \equiv 0$ for $k = 0, 1 \ldots < m-1$. By orthogonality, $U_{1,j} = 0$ for $1 < j < m$. We continue this chain showing that $U_{i,2} = 0$ for $i \neq 2$, etc. Proof by finite induction. ■

Corollary 2.5 *Let A be an $n \times n$ matrix with distinct eigenvalues. Then A is unitarily similar to triangular matrix $\hat{A}$ with ordered eigenvalues and $\hat{A}$ is unique up to diagonal unitary similarities: $\hat{A} \leftarrow E\hat{A}E^*$, where $E_{i,j} = \exp(i\theta_j)\delta_{i-j}$.*

If A is invertible and (A, B) is TIN, then A is the Cholesky factor of $\mathbb{I}_n - BB^*$ times a diagonal unitary matrix:

Theorem 2.6 *Let (A, B) be a TIN pair with rank $(\mathbb{I}_n - BB^*) = n$, then*

$$A = \text{Cholesky}\,(\mathbb{I}_n - BB^*)\, E_n \ , \tag{2.3}$$

where E_n is a diagonal unitary matrix. In the real case, E_n is a signature matrix.

Proof: Let L be the Cholesky factor of $(\mathbb{I}_n - BB^*)$. Clearly $A = LQ$ some unitary Q with $L_{ii} \geq 0$. By the uniqueness of the LQ factorization, L and Q are unique. Note $Q = L^{-1}A$ is lower triangular and therefore diagonal. ■

3 BAND FRACTION REPRESENTATIONS OF SINGLE INPUT TIN PAIRS.

For the single input case ($d = 1$), TIN pairs have an explicit band fraction representation $A = M^{-1}N$, $B = \rho_1 M^{-1}e_1$, where M and N are bidiagonal. The representation is parameterized by the eigenvalues of A. For each eigenvalue, λ_k, of A, we define $\rho_k = \sqrt{1 - |\lambda_k|^2}$, $\mu_k = \rho_{k+1}/\rho_k$ and $\gamma_k = \lambda_k^*\mu_k$. We define the matrices:

$$M \equiv \text{bidiag}\begin{pmatrix} 1 & & 1 & & 1 & & \cdots & \\ & \gamma_1 & & \gamma_2 & & \gamma_3 & & \cdots \end{pmatrix} , \tag{3.1}$$

$$N \equiv \text{bidiag}\begin{pmatrix} \lambda_1 & & \lambda_2 & & \lambda_3 & & \cdots & \\ & \mu_1 & & \mu_2 & & \mu_3 & & \cdots \end{pmatrix} , \tag{3.2}$$

where the top row contains the diagonal elements of M or N and the bottom row contains the $(n-1)$ elements of first subdiagonal of the bidiagonal matrices, M and N.

Theorem 3.1 *Let $\hat{A} \equiv M^{-1}N$ and $\hat{B} = M^{-1}\rho_1 e_1$, where M and N are given by (3.1- 3.2) with $|\lambda_k| < 1$. Then $(\hat{A}, \hat{B})$ is TIN with eigenvalues $\{\lambda_k\}$.*

Proof: Explicit evaluation shows $MM^* - NN^* = \rho_1^2 e_1 e_1^*$. ■

The Hautus criterion states that if a stable matrix, A, is nonderogatory (There is only one Jordan block for each eigenvalue of A in the eigendecomposition of A.), then there is a vector B such that (A, B) is controllable. Thus the Hautus criterion together with Theorem 2.3 imply that

Corollary 3.2 *Let A be a stable, nonderogatory matrix. Then there exists a similarity transformation, T, such that $\hat{A} = TAT^{-1}$ has the bidiagonal fraction representation of Theorem 3.1.*

The bidiagonal fraction can be casted in a more aestetic form $A = M^{-1}N = M_0^{-1}N_0$, where M_0 and N_0 are defined as

$$M_0 \equiv \text{bidiag}\begin{pmatrix} c_1 & & c_2 & & c_3 & & \cdots & \\ & s_1^* & & s_2^* & & s_3^* & & \cdots \end{pmatrix}, \tag{3.3}$$

$$N_0 \equiv \text{bidiag}\begin{pmatrix} s_1 & & s_2 & & s_3 & & \cdots & \\ & c_1 & & c_2 & & c_3 & & \cdots \end{pmatrix}, \tag{3.4}$$

with $c_k = 1 \Big/ \sqrt{1 - |\lambda_k|^2}$ and $s_k \equiv \lambda_k / \sqrt{1 - |\lambda_k|^2}$.

In a series of excellent articles [19, 2, 13, 17, 15, 14], a number of researchers have constructed IN filters using cascade realizations. We now consider the case single input case where A has real eigenvalues. In this case, the cascade realization of the IN filter is triangular. By Corollary 2.5, there is a unique triangular state advance matrix with the given eigenvalue ordering (up to diagonal similarity transformations with ± 1 elements). Thus all state space constructions of IN filters are simply different realizations of the same (A, B) pair (up to sign flips of the coordinates). In Section 5, we discuss the numerics of the various realizations of IN filters.

We now give a more constructive proof of equivalence of the band fraction representation with the orthogonal basis functions of [13]. Given a set of decay rates/poles of the frequency responses, $\{\lambda_n\}$, Ninness and Gustafsson derive a set of orthonormal basis functions in the frequency domain:

$$\hat{z}_n(w) = \sum_{t=0}^{\infty} z_n(t)\, w^{-t} = w^{\alpha} \frac{\sqrt{1 - |\lambda_n|^2}}{w - \lambda_n} \prod_{k=1}^{n-1} \left(\frac{1 - \lambda_k^* w}{w - \lambda_k} \right), \tag{3.5}$$

where α is 0 or 1. From this, we derive the relation:

$$(w - \lambda_k)\hat{z}_k(w) \Big/ \sqrt{1 - |\lambda_k|^2} = (1 - \lambda_{k-1}^* w)\hat{z}_{k-1}(w) \Big/ \sqrt{1 - |\lambda_{k-1}|^2} . \tag{3.6}$$

for $k \geq 1$ with $\lambda_0 \equiv 0$ and $\hat{z}_0(w) \equiv z^\alpha$. Equating like powers of w yields

$$z_k\,(t+1) + \lambda_{k-1}^* \mu_{k-1} z_{k-1}\,(t+1) = \lambda_k z_k\,(t) + \mu_{k-1} z_{k-1}\,(t) \;\; , \tag{3.7}$$

with $z_0(t) \equiv \delta_{0,-\alpha}$. This shows that the bidiagonal matrix fraction representation generates the same basis functions as in [13]. The k-th component of z_t is to be identified with the t-th lead of the k-th orthonormal transfer function, $\mathcal{B}_k(q)$.

We show that if the eigenvalues are sorted in order of increasing magnitude, then M^{-1} does not have large elements. We now give two results concerning the bidiagonal matrix fraction representation ($d = 1$). The first result shows that if the eigenvalues are sorted in order of increasing magnitude, then the bidiagonal factorization is well-conditioned. We then give the eigenvectors of A.

A. Explicit M inversion.

We can interpret the bidiagonal matrix fraction representation as an LR factorization of the augmented system: $(B|A) = M^{-1}(e_1|N)$. We now explicitly invert M in (3.1) and show that the numerical conditioning of the matrix inversion is good without pivoting. We define the $x-$condition number of an invertible matrix, G, as $\kappa_x(G) \equiv \|G\|_x \|G^{-1}\|_x$, where $x = 1, 2$, or ∞ [1, 3]. Here $\|G\|_1$ is the maximum column sum norm and $\|G\|_\infty$ is the maximum row sum norm.

Theorem 3.3 *Let (A, B) be a stable TIN pair ($d = 1$). If the eigenvalue magnitudes are in ascending order, $|\lambda_{k+1}| \geq |\lambda_k|$, then $|(M^{-1})_{i,j}| < 1$ for $i > j$ and $\kappa_2(M^{-1}) \leq 2n$.*

Proof: The LR factorization is $(B|A) = (M^{-1})\,(e_1|N)$. Now M is unit lower triangular with $(M)_{k+1,k} = \lambda_k^*$. So

$$(M^{-1})_{i,j} = (-1)^{i-j} \prod_{j \leq k < i} \lambda_k^* \mu_k = (-1)^{i-j} \left(\frac{1 - |\lambda_i|^2}{1 - |\lambda_j|^2} \right)^{1/2} \prod_{j \leq k < i} \lambda_k^* \;, \tag{3.8}$$

for $i > j$. Note $\|M^{-1}\|_x \leq n$ and $\|M\|_x \leq 2$ for $x = 1$ and $x = \infty$. Thus $\kappa_2(M^{-1}) \leq \kappa_1(M^{-1})\kappa_\infty(M^{-1}) \leq 2n$. ■ This implies that the LR factorization of $(B|\ A)$ is numerically stable without pivoting.

B. Eigenvectors

We now evaluate the eigenvectors of the single input TIN system using the parameterization (3.1)-(3.2) for the case of distinct eigenvalues. It is well known that the eigenvectors of triangular matrices satisfy a recursion formula. Let V_{ij} be the i-th component of the j-th eigenvector with $V_{ij} = 0$ for $i < j$. Write $V = [V_{ij}]$ and $\Lambda \equiv \text{diag}\,(\lambda_1, \ldots \lambda_n)$. The eigenvector equation $AV = V\Lambda$ becomes $NV = MV\Lambda$ or element by element:

$$\lambda_k \left(\lambda_{j-1}^* \mu_{j-1} V_{j-1,k} + V_{jk} \right) = \left(\mu_{j-1} V_{j-1,k} + \lambda_j V_{jk} \right), \tag{3.9}$$

We set $V_{kk} = 1$, $V_{jk} = 0$ for $j < k$ and solve the recursion for V_{jk} when $j > k$:

$$V_{jk} = \frac{\left(1 - |\lambda_j|^2\right)^{1/2} \left(1 - |\lambda_k|^2\right)^{1/2}}{\lambda_k - \lambda_j} \left[\prod_{k<j'<j} \left(\frac{1 - \lambda_k \lambda_{j'}^*}{\lambda_k - \lambda_{j'}} \right) \right] , \tag{3.10}$$

where the bracketed term is equal to 1 for $j = k + 1$.

In the next section, we show that real single input IN pairs with complex conjugate eigenvalues have a bidiagonal fraction representation as well. However, we do not give an explicit parameterization of M and N in terms of the eigenvalues.

4 MIMO BAND FRACTION REPRESENTATIONS OF TIN PAIRS.

We now prove that generic TIN pairs have a band fraction representation: $A = M^{-1}N$, $B = M^{-1}\hat{B}$, where M and N are triangular matrices of lower bandwidth d and $\hat{B}$ is upper triangular. The banded matrix fraction structure allows state space updates, $z_{t+1} \leftarrow Az_t + B\epsilon_t$, in $(2d+1)n$ multiplications using (1.3). Theorem 4.3 and Theorem 4.4 show how to parameterize TIN input pairs using nd parameters.

Theorem 4.1 *Let (A, B) satisfy $D - ADA^* = BB^*$, where D is a diagonal, positive definite matrix and A is LT and B is an $n \times d$ matrix. Let $(B|A)$ have nonvanishing principal subminors, $(B|A)_{1:k,1:k}$ for $k < n$, then (B, A) has an unique band fraction representation: $(B, A) = M^{-1}(\hat{B}, N)$, where M and N are $n \times n$ LT matrices of bandwidth d and $M_{i,i} = 1$ and $\hat{B}_{j,k} = 0$ for $j > k$.*

Proof: By Theorem 3.2.1 of [1], $(B|A) = L^{-1}R$ has a unique $(B|A_{1:n,1:(n-d)}) = L^{-1}R$ has a unique representation where R is UT and L is LT with $L_{i,i} = 1$. Let $\tilde{R}$ be the submatrix of R containing columns $(d + 1)$ through $(n + d)$. Since A is LT, $\tilde{R} = LA$ is LT of bandwidth d. Note $R(\mathbb{I}_d \oplus D)R^* = LDL^*$. By Theorem 4.3.1 of [1], L has bandwidth d. We set $M = L$ and $N = \tilde{R}$ ■

For $d = 1$, every controllable TIN pair has nonvanishing principal minors. The condition that $(B|A)$ have nonvanishing principal minors for $k < n$ is generically true. If $(B|A)$ has an LR decomposition and a vanishing principal minor for $k < n$, then the induced TIN pair, (A, B), does not have a unique LR decomposition of $(B|A)$ with $M_{i,i} = 1$. From $MM^* = NN^* + \hat{B}\hat{B}^*$, the singular values of M are singular values of $(\hat{B}|N)$. When the condition that $M_{i,i} = 1$ is relaxed, other band fraction representations may be generated by $(\hat{B}, N) \leftarrow D'(\hat{B}, N)$, $M \leftarrow D'M$, where D' is a nonsingular diagonal matrix.

The band fraction structure implies that the $n \times (n+d)$ matrix $Y \equiv (\hat{B}\,|N)$ is upper triangular with upper bandwidth d. We now examine parameterizations of the band fraction representation. We define

Definition 4.2 *Let $\mathcal{Y}$ denote the set of all upper triangular $n \times (n+d)$ matrices with upper bandwidth d. Let $\mathcal{Y}_1$ be the set of $Y \in \mathcal{Y}$ such that $Y_{i,i} = 1$ for $i < n$ and $\|Y_{n,:}\| = 1$ with a positive first nonzero element in the nth row.*

Note $\mathcal{Y}$ has $n(d+1)$ nonzero coordinates while $\mathcal{Y}_1$ is an nd dimensional manifold. We now show that we can construct a TIN pair from any Y in $\mathcal{Y}$ using the LQ decomposition.

Theorem 4.3 *Let $Y \in \mathcal{Y}$ have rank(n) and define M and $\hat{Q}$ as the LQ decomposition of Y. Define $B = \hat{Q}_{\,:,1:d}$ and $A = \hat{Q}_{\,:,(d+1):(n+d)}$. Then M has lower bandwidth d and is invertible and (A, B) is a TIN pair.*

Proof: Note rank(M) =rank(Y) $= n$. Let $(\hat{B}\,|N) = Y$. Note $MM^* = YY^* = NN^* + \hat{B}\,\hat{B}^*$. By Theorem 4.3.1 of [1], M has bandwidth d. From $A = M^{-1}N$, A is triangular and $AA^* + BB^* = \mathbb{I}_n$ implies (A, B) is a TIN pair. ■

We now examine the map $f(Y) = (A = \hat{Q}_{\,:,(d+1):(n+d)}, B = \hat{Q}_{\,:,1:d})$, where $M_{ii}(Y) > 0$. From Theorem 9.1 of [3], $(B|A) = M^{-1}Y$ has nonzero principal minors for $k < n$ if and only if Y does. We can rescale Y and M by a nonsingular diagonal matrix D: $Y \leftarrow DY$ and $M \leftarrow DM$, and preserve the induced TIN pair, (A, B). This freedom and complex phase equivalence allows us to restrict our consideration to $Y \in \mathcal{Y}_1$:

Theorem 4.4 *For each TIN pair, (A, B), with nonvanishing principal minors in $(B|A)$ for $k < n$, there exists a unique $Y \in \mathcal{Y}_1$ and a unique diagonal unitary matrix E such that (EAE^*, EB) is generated by Y: $f(Y) = (EAE^*, EB)$.*

Proof: Let $(\tilde{B}, \tilde{N}, \tilde{M})$ be the unique band fraction decomposition of $(B|A)$ with $\tilde{M}_{ii} = 1$. For an arbitrary diagonal unitary matrix E, the set of band fraction representations of (EAE^*, EB) is $\{(DE\tilde{B}, DE\tilde{N}E^*, DE\tilde{M}E^*)|\}$, where D is nonsingular and diagonal. Suppose both Y_1 and $Y_2 \in \mathcal{Y}_1$ generate an equivalent version of (A, B), i.e. $f(Y_i) = (E_iAE_i^*, E_iB)$. Thus $Y_i = (D_iE_i\tilde{B}|D_iE_i\tilde{N}E_i^*)$ for some D_i and E_i where D_i is nonsingular and diagonal and E_i is diagonal and unitary. Since $M_i = D_iE_i\tilde{M}E_i^*$, the condition $M_{i,i} > 0$ shows that D_1 and D_2 are positive matrices. The requirement that $Y_i \in \mathcal{Y}_1$ forces $|D_1| = |D_2|$ and $E_1 = E_2$. ■

We can parameterize input normal systems using $Y \in \mathcal{Y}_1$ for $d > 1$. There are two difficulties with this parameterization. First, the representations $(M^{-1}N, M^{-1}\hat{B}\,)$ for

$(\hat{B}|N)$ in $\mathcal{Y}_1$ contain many equivalent TIN representations corresponding to different orderings of the eigenvalues of A. Second, in the $\mathcal{Y}_1$ parameterization, the eigenvalues of A are only known after one solves M by computing the LQ decomposition of Y.

5 NUMERICS

We now discuss the computational speed of various system representations.

The bidiagonal band fraction representation, $A = M^{-1}N$, requires only $3n$ multiplications for a state vector update in a single input system. First, we compute $r_t \equiv Nz_t + e_1\epsilon_t$ in $2n$ multiplications. We then solve the bidiagonal matrix equation $Mz_{t+1} = r_t$ in n multiplications since $M_{i,i} = 1$. multiplications. The matrix operations in the band fraction advance may be implemented using the *dtbmv* and *dtbtrs* routines from the optimized LAPACK software [5]. We caution that if A has nonreal eigenvalues, then M and N are complex. Thus the computational advantage of the bidiagonal matrix fraction representation is primarily limited to the case when the eigenvalues of A are real.

In [2], Heuberger et al. propose a cascade realization of IN filters. This cascade representation requires $4n$ multiplications for a state vector update in a single input system. A different cascade realization of SISO IN pairs is given in Figure 1 of [13]. The Ninness and Gustafson representation uses a lossless cascade followed by a set of AR(1) models . This realization requires at least $5n$ multiplications for a state update.

The direct form representations [12] have even faster state advances, but these representations are not IN and can have very ill-conditioned Grammians [11, 14]. We advise against using high order models that are not in input normal form. The tridiagonal form [7] is another fast representation that is not input normal.

In [12, 18], embedded lossless representations are constructed. For $d = 1$, these embedded lossless filters $8n$ multiplications per advance [12]. These embedded filters include the cost of evaluating $C \cdot z_t$ while the IN filter representation has an additional cost of n multiplications to compute $C \cdot z_t$, so we should credit the embedded filters with n multiplications, yielding $7n$ multiplications. Unfortunately, embedded lossless systems are not input normal.

In [10], we give a different representation of Hessenberg and triangular input pairs by treating (B, A) as a projection of a product of nd Givens rotations. The multiplication count for these Givens product representations is $4dn$, which for large d is twice as large as the band fraction representations. If Householder transformations are used, the multiplication count is again $2nd$ asymptotically. The Givens product representations are a MIMO generalization of the cascade architecture. This approach may be more natural for multivariate case. In the $d = 1$ case, the band fraction representation has the advantage that the eigenvalues of A are parameters of the system. This eigenvalue parameterization is convenient when the eigenvalues are prespecified or adaptively estimated.

The Givens product representations in [10] may be less sensitive to roundoff error since they are orthogonal matrices [12]. The computational advantage of the band fraction representation is that fast numerical algorithms for band matrix inversion and multiplication are readily available.

6 FAST SYSTEM IDENTIFICATION

The band fraction representations may be used for the rapid identification of impulse responses. We remain in the framework where a fixed basis of orthogonal basis functions are given as described in [12, 19, 13, 17, 15, 14] and summarized in the introduction. We then use the band fraction representation of Sections 3-4, so $(A, B) \equiv M^{-1}(N, \hat{B})$ is given. The state vector evolves according to (1.3). We effectively compute the second moment matrices, $\hat{P}_t^\delta \equiv \sum_{i=1}^t \delta^{t-i} z_i z_i^*$ and $\hat{d}_t^\delta = \sum_{i=i}^t \delta^{t-1} z_i y_i^*$, where δ is the forgetting factor. The unknown coefficients, $\hat{C}$, using recursive least squares (RLS). At each time step, we reestimate $\hat{C}$ by solving $\hat{P}_t^\delta \hat{C}_t = \hat{d}_t^\delta$. This is the normal equations for the least squares estimate of $\hat{C}_t$. The kth component of z_t represents the kth orthonormal transfer function, $\mathcal{B}_k(q)$, applied to the sequence $\{u_1, u_2, \ldots u_t\}$. The resulting estimate of the the jth lead of the impulse response is $\hat{C} A^{j-1} B$. To solve for $\hat{C}_t$., we recommend using a QR update of the square root of normal equations [16]. Even faster methods are available that use the displacement rank structure of $\hat{P}_t^\delta$ [9, 16]. The interested reader can consult [6, 14] for a comprehensive description of adaptive estimation.

The input normal filter representations are advantageous for many reasons: First, $\hat{P}_t^\delta \stackrel{t\to\infty}{\longrightarrow}$ constant $\times\ \mathbb{I}_n$, when the 'true" state space model is used and the forcing noise is white. Thus the regression for $\hat{C}_t$ is well-conditioned. Similarly, IN filter structures are resistant to roundoff error [12]. Thus IN representations will time asymptotically satisfy the ansatz need by least mean squares (LMS) identification algorithms . This leads to a second advantage of IN filters: Gradient algorithms such as the least mean squares (LMS) algorithm often perform well enough in certain applications to obviate the need for more complicated and computationally intensive RLS algorithms. Third, an $\mathcal{O}(n)$ update of $\hat{C}_t$ is possible [9]. Finally, when the advance matrix is triangular, the IN pairs form nested families: $(A_{1:k,1:k}, B_{1:k,:})$ is TIN for $k \leq n$ when (A, B) is TIN. This nesting may be used in adaptive order selection.

These are advantages of all TIN filter representations. For single input pairs, The advantages of the band fraction representation over the cascade representation are primarily numerical, the faster computational speed ($3n$ multiplications versus $4n$ or more) and the availability of fast linear algebra software for banded matrices.

7 SUMMARY

We have shown that TIN pairs generically have a matrix fraction representation, $A = M^{-1}N$, where M and N have lower bandwidth d. This structure allows fast state space updates, $Mz_{t+1} \leftarrow Nz_t + \hat{B}\epsilon_t$, and fast solution of matrix equations: $Ax = f$. The total operations count is $(2d+1)n$ multiplications. We believe that our work is the first study of system representations, where A is a matrix fraction of banded matrices, $A = M^{-1}N$.

For single input filter design, our representations are parameterized explicitly in terms of the eigenvalues, $\{\lambda_k\}$. When M is bidiagonal, solving $Mz_{t+1} = f_t$ requires only $\mathcal{O}(n)$ multiplications and may be implemented as a systolic array with order independent latency. As showed in Theorem 3.3, if the eigenvalues are ordered in ascending magnitude, then the matrix inversion is well-conditioned and suitable for fixed point operations. For system identification, the covariance of z_t tends to the identity matrix when the forcing noise is white. Thus least squares regression is both well-conditioned and computationally fast. Since A is triangular, one can compute the evolution of all embedded models of dimension $n' < n$ simply by projecting the model of dimension n onto the first n' coordinates. Thus the triangular structure simplifies the evaluation of this nested family of models for model order selection.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations*, third edition, Baltimore: John Hopkins University Press, 1996.

[2] P.S.C. Heuberger, P.M.J. Van den Hof, and O.H. Bosgra, 'A generalized orthonormal basis for linear dynamical systems," *IEEE Trans. Aut. Cont.*, vol. 40, pp. 451-465, 1995.

[3] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, Philadelphia: SIAM Press, 1996.

[4] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge: Cambridge University Press, 1985.

[5] E. Anderson, et al. *LAPACK User's Guide* Philadelphia: SIAM Press, 1999.

[6] L. Ljung and T. Söderström, *Theory and practice of recursive identification*, Cambridge, MA: The MIT Press, 1983.

[7] T. McKelvey and A. Helmersson, 'State-space parameterizations of multivariate linear systems using tridiagonal matrix forms," *Proc. 35th IEEE Conf. on Decision and Control*, pp. 1666-1671, IEEE Press, 1996.

[8] B. Moore, 'Principal components analysis in linear systems: controllability, observability, model reduction," *IEEE Trans. Aut. Cont.*, vol. 26, pp. 17-32.

[9] A. Mullhaupt and K. S. Riedel, 'Fast identification of innovations filters," *IEEE Trans. Signal Processing*, vol. 45, pp. 2616-2619, 1997.

[10] A. Mullhaupt and K. S. Riedel, 'Hessenberg and Schur output normal pair representations," submitted.

[11] A. Mullhaupt and K. S. Riedel, 'Bounds on the condition number of solutions of the Stein equation," submitted.

[12] R. A. Roberts and C. T. Mullis, *Digital Signal Processing*, Addison Wesley, Reading, MA, 1987.

[13] B. Ninness and F. Gustafsson, 'A unifying construction of orthonormal bases for system identification," *IEEE Trans. Aut. Cont.* vol. 42, pp. 515-521, 1997.

[14] B. Ninness, 'The utility of orthonormal bases in system identification," Technical Report 9802, Dept. of EECE, University of Newcastle, Australia, 1998.

[15] B. Ninness, H. Hjalmarsson and F. Gustafsson, 'The fundamental role of orthonormal bases in system identification," *IEEE Trans. Aut. Cont.* vol. 44, pp. 1384-1407, 1999.

[16] A. H. Sayed and T. Kailath, 'A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, vol. 11, pp. 18-60, 1994.

[17] P.M.J. Van den Hof, P.S.C. Heuberger, and J. Bokor, 'System identification with generalized orthonormal basis functions," *Automatica*, vol. 31, pp. 1821-1831, 1995.

[18] A.J. van der Veen and M. Viberg, 'Minimal continuous state-space parametrizations," In *Proc. Eusipco*, Trieste (Italy), pp.523-526, 1996.

[19] B. Wahlberg, 'System identification using Laguerre models," *IEEE Trans. Aut. Cont.*, vol. 36, pp. 551-562, 1991.