Integr. equ. oper. theory 99 (9999), 1–14
0378-620X/99000-0, DOI 10.1007/s00020-003-0000

Integral Equations and Operator Theory

Bi-Inner Dilations and Bi-Stable Passive Scattering Realizations of Schur Class Operator-Valued Functions

Damir Z. Arov and Olof J. Staffans

Abstract. Let $S(U;Y)$ be the class of all Schur functions (analytic contractive functions) whose values are bounded linear operators mapping one separable Hilbert space U into another separable Hilbert space Y, and which are defined on a domain $\Omega \subset \mathbb{C}$, which is either the open unit disk $\mathbb{D}$ or the open right half-plane $\mathbb{C}^+$. In the development of the Darlington method for passive linear time-invariant input/state/output systems (by Arov, Dewilde, Douglas and Helton) the following question arose: do there exist simple necessary and sufficient conditions under which a function $\theta \in S(U;Y)$ has a bi-inner dilation $\Theta = \begin{bmatrix} \theta_{11} & \theta \\ \theta_{21} & \theta_{22} \end{bmatrix}$ mapping $U_1 \oplus U$ into $Y \oplus Y_1$; here U_1 and Y_1 are two more separable Hilbert spaces, and the requirement that Θ is bi-inner means that Θ is analytic and contractive on Ω and has unitary nontangential limits a.e. on $\partial\Omega$. There is an obvious well-known necessary condition: there must exist two functions $\psi_r \in S(U;Y_1)$ and $\psi_l \in S(U_1;Y)$ (namely $\psi_r = \theta_{22}$ and $\psi_l = \theta_{11}$) satisfying $\psi_r^*(z)\psi_r(z) = I - \theta^*(z)\theta(z)$ and $\psi_l(z)\psi_l^*(z) = I - \theta(z)\theta^*(z)$ for almost all $z \in \partial\Omega$. We prove that this necessary condition is also sufficient. Our proof is based on the following facts. 1) A solution ψ_r of the first factorization problem mentioned above exists if and only if the minimal optimal passive realization of θ is strongly stable. 2) A solution ψ_l of the second factorization problem exists if and only if the minimal $*$-optimal passive realization of θ is strongly co-stable (the adjoint is strongly stable). 3) The full problem has a solution if and only if the balanced minimal passive realization of θ is strongly bi-stable (both strongly stable and strongly co-stable). This result seems to be new even in the case where θ is scalar-valued.

Mathematics Subject Classification (2000). Primary 47A48, 93B15; Secondary 94C05.

Damir Z. Arov thanks Åbo Akademi for its hospitality and the Academy of Finland for its financial support during his visit to Åbo in August–September 2003. He also gratefully acknowledges the partial financial support by the joint grant UM1-2567 OD 03 from the U.S. Civilian Research and Development Foundation (CRDF) and the Ukrainian Government.

Keywords. Darlington method, optimal passive realization, $*$-optimal passive realization, balanced passive realization.

1. Introduction

The well-known Darlington method in passive circuit theory can be used to synthesize a lossy finite passive circuit network with a given frequency characteristic in an impedance, (input) scattering, or transmission (chain scattering) setting. Typically the given frequency characteristic function is a rational matrix function of the appropriate class (positive real, contractive, or J-contractive in the open right half-plane $\mathbb{C}^+$), and the Darlington synthesis is carried out by first constructing a lossless network and then loading it with a number of resistors. This number is required to be as small as possible. Darlington [16] introduced this method in the context of a one-pole with a given scalar rational positive real characteristic function, which was realized as the impedance of a network. This approach was extended to multi-poles in a scattering setting by Belevitch [14, Chapter 12].

Let $S^{p\times q}$ be the class of all contractive holomorphic matrix-valued function of size $p \times q$ defined on $\mathbb{C}^+$. In the scattering setting the Darlington method is related to the representation of a rational (usually real) $\theta \in S^{p\times q}$ as one block of a rational bi-inner matrix-valued function

$$\Theta = \begin{bmatrix} \theta_{11} & \theta \\ \theta_{21} & \theta_{22} \end{bmatrix} \tag{1}$$

of size $m \times m$, where m is required to be as small as possible. If θ is real, the Θ should also be real. If $p = q$ (which is usually the case), then $r = m - p$ is the minimal number of resistors in a circuit whose input scattering matrix is θ. Thus, we obtain a lossy network with input scattering matrix θ by dropping a total of r exterior branches in a lossless network with scattering matrix Θ.

Later the Darlington method was extend from finite to infinite networks (with both lumped and distributed parameters). In this setting the given frequency characteristic is no longer rational (see [5, 9]). In order to use the same method to realize a given holomorphic contractive non-rational matrix-valued function θ defined on $\mathbb{C}^+$ as the scattering matrix (transfer function) of a conservative or passive linear time-invariant input/state/output system one again needs to solve an extension problem, where θ is embedded as a block of a bi-inner non-rational matrix-valued function Θ of dimension $m \times m$ (see [2, 17]). As in [8, 9] we again require m to be as small as possible.

Arov [2] and Dewilde [17] discovered that $\theta \in S^{p\times q}$ has a representation (1) with some inner matrix function Θ of size $m \times m$ if and only if θ has a meromorphic pseudo-continuation θ_- into the left half-plane $\mathbb{C}^-$ with bounded Nevanlinna characteristic in $\mathbb{C}^-$. Recall that a meromorphic function θ_- is said to have a *bounded Nevanlinna characteristic* (or that it is of *bounded type*) in $\mathbb{C}^-$ if θ_- can be represented as a quotient $\theta_-(z) = a(z)^{-1}b(z)$, where a and b are bounded holomorphic functions on $\mathbb{C}^-$, a being scalar-valued and b matrix-valued. The statement that θ_- is a *pseudo-continuation* of θ means that for almost all $y \in \mathbb{R}$ the limits $\theta(iy) := \lim_{x\downarrow 0} \theta(x+iy)$ and $\theta_-(iy) := \lim_{x\uparrow 0} \theta_-(x+iy)$ are equal. See, e.g., [18, Section 1] or [23, Sections 4.2 and 6.3] for more detailed discussions of these two notions.

At the next stage of generality we replace the matrix-valued function $\theta \in S^{p\times q}$ by an operator-valued function. Let U (the input space) and Y (the output space) be separable Hilbert spaces, and let $S(U;Y)$ be the class of all Schur functions (analytic contractive functions) whose values are bounded linear operators mapping U into Y, defined on $\mathbb{C}^+$. Given a function $\theta \in S(U;Y)$ we look for a bi-inner dilation $\Theta \in S(\widetilde{U};\widetilde{Y})$ of θ of the following type. The spaces $\widetilde{U}$ and $\widetilde{Y}$ are dilations of U and Y, i.e., $\widetilde{U} = U_1 \oplus U$ and $\widetilde{Y} = Y \oplus Y_1$ where U_1 and Y_1 are two auxiliary separable Hilbert spaces, and Θ has the block matrix form (1), where this time $\theta_{11} \in S(U_1;Y)$, $\theta_{21} \in S(U_1;Y_1)$, and $\theta_{22} \in S(U;Y_1)$, and

$$\theta(z) = P_Y \Theta(z)_{|U}, \tag{2}$$

where P_Y is the orthogonal projection in $\widetilde{Y}$ onto Y. We recall that a function $\Theta \in S(\widetilde{U};\widetilde{Y})$ is called inner, co-inner, or bi-inner if its almost everywhere defined boundary values $\Theta(iy) := \lim_{x\downarrow 0} \Theta(x+iy)$ are isometric, co-isometric, or unitary, respectively, a.e. on $i\mathbb{R}$. A representation (2) of θ with a bi-inner function Θ is called a $\widetilde{D}$-representation of θ, and the function Θ is called a bi-inner dilation of θ; see [8] and [18].

As we saw above, a function $\theta \in S^{p\times q}$ has a matrix-valued bi-inner dilation if and only if θ has a meromorphic pseudo-continuation θ_- into the left half-plane $\mathbb{C}^-$ with bounded Nevanlinna characteristic. In the case where $\theta \in S(U;Y)$ (with possibly infinite-dimensional U and Y) the property of θ of having a meromorphic pseudo-continuation θ_- into the left half-plane $\mathbb{C}^-$ with bounded Nevanlinna characteristic is still a sufficient condition for the existence of a bi-inner dilation; see [2, 4, 8], and [18] (we define pseudo-continuation and bounded Nevanlinna characteristic in the same way as above; in particular, we require the denominator a to be a scalar function). However, this condition is no longer necessary, as we will prove below. Neither is it necessary in the case where $\theta \in S^{p\times q}$, but we allow the dilated spaces $\widetilde{U}$ and $\widetilde{Y}$ to be infinite-dimensional.

If U or Y is infinite-dimensional, then the additional condition that the dilated spaces $\widetilde{U}$ and $\widetilde{Y}$ should have a minimal dimension is no longer meaningful. The appropriate notion is instead that the dilated function should have *minimal losses*, a notion which was introduced by Arov in [4, 8]. A bi-inner dilation Θ of θ has minimal losses if the multiplication operator by θ_{11} is injective on $L^2(i\mathbb{R};U)$, or equivalently, if the multiplication operator by θ_{22}, acting on $L^2(i\mathbb{R};U_1)$, has a dense range. A $\widetilde{D}$-representation with minimal losses is called a D-representation, and in the matrix-valued case this is a standard Darlington representation with minimal dimension described earlier. For a more detailed description of how this condition on minimal losses should be interpreted we refer the reader to [8]. Arov [8] also obtains a necessary and sufficient condition for the existence of a D-representation.

An obvious (and well-known) necessary condition for the existence of a bi-inner dilation of $\theta \in S(U;U)$ is that there must exist two functions $\psi_r \in S(U;Y_1)$ and $\psi_l \in S(U_1;Y)$ (namely $\psi_r = \theta_{22}$ and $\psi_l = \theta_{11}$) satisfying

$$\psi_r^*(z)\psi_r(z) = I - \theta^*(z)\theta(z), \;\; \psi_l(z)\psi_l^*(z) = I - \theta(z)\theta^*(z), \;\; \text{a.e. on } \partial\Omega. \tag{3}$$

The main result of this paper is that the converse is also true: the solvability of the two factorization problems (3) in the Schur class of functions is not only necessary, but also sufficient for the existence of a bi-inner dilation. This answers a question posed more than 30 years ago by Douglas and Helton [18, p. 66].

Above we have concentrated on the case where the function θ is defined on the open half-plane $\mathbb{C}^+$. A similar theory is valid when θ is defined on the open unit disc $\mathbb{D}$ instead. The latter case is technically slightly simpler, and we shall therefore in the sequel mainly concentrate on the case where θ is defined on $\mathbb{D}$, and only at the very end return to the case where θ is defined on $\mathbb{C}^+$. In particular, below we let $S^{p\times q}$ and $S(U;Y)$ stand for the class of Schur functions defined on $\mathbb{D}$ (instead of being defined on $\mathbb{C}^+$).

Darlington synthesis for time-dependent systems has been studied by Pick [22].

2. Preliminaries

Let U, X, and Y be separable Hilbert spaces, and let $\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ be a quadruple of bounded linear operators mapping $X \oplus U$ into $X \oplus Y$. With these operators we associate the following discrete-time system:

$$\Sigma\colon \quad \begin{aligned} x(n+1) &= Ax(n) + Bu(n), \\ y(n) &= Cx(n) + Du(n). \end{aligned} \tag{4}$$

We call A the *main operator*, B the *control operator*, C the *observation operator*, and D the *feedthrough operator* of $\Sigma = [\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]; (Y,X,U)]$. The transfer function of Σ is given by

$$\theta_\Sigma(z) = zC(I - zA)^{-1}B + D, \quad 1/z \in \rho(A)$$

(where $\rho(A)$ is the resolvent set of A). The system Σ is (scattering) *passive* if $\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ is a contraction from $X \oplus U$ to $X \oplus Y$, and *conservative* if $\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ is unitary. In these cases the transfer function is defined on all of $\mathbb{D}$, and it belongs to $S(U;Y)$ (it is a Schur function on $\mathbb{D}$). Conversely, given a function $\theta \in S(U;Y)$, we call the system Σ in (4) a *passive or conservative realization* of θ if Σ is passive or conservative, and the transfer function of Σ is θ in the sense that

$$\theta(z) = \theta_\Sigma(z) = zC(I - zA)^{-1}B + D, \quad z \in \mathbb{D}.$$

The system Σ (and its main operator A) is *strongly stable* if $A^k x \to 0$ as $k \to \infty$ for all $x \in X$, and it is *strongly co-stable* if $(A^*)^k x \to 0$ as $k \to \infty$ for all $x \in X$ (where both limits are taken in the strong sense).

Let $\mathbb{Z}^+ = 0, 1, 2, \dots$ and $\mathbb{Z}^- = -1, -2, -3, \dots$. We define the *reachability map* of Σ to be the operator $\mathfrak{B}$ which maps a sequence $\{u(-k)\}_{k=1}^\infty$ with only finitely many nonzero elements into

$$\mathfrak{B}u = \sum_{k=1}^{\infty} A^{k-1} B u(-k).$$

The *observability map* of Σ is the operator which maps $x \in X$ into the sequence

$$\mathfrak{C}x = \{CA^k x\}_{k=0}^\infty.$$

(In the case of a (scattering) passive system the operator $\mathfrak{B}$ can be extended to a bounded linear operator on $\ell^2(\mathbb{Z}^-;U)$, and $\mathfrak{C}$ is bounded from X to $\ell^2(\mathbb{Z}^+;Y)$.) The *reachable subspace* $\mathfrak{R}$ is the closure of the range of $\mathfrak{B}$, and the *unobservable subspace* $\mathfrak{U}$ is the kernel of $\mathfrak{C}$. We call Σ *controllable* if $\mathfrak{R} = X$, it is *observable* if $\mathfrak{U} = \{0\}$, it is *minimal* if it is both controllable and observable, and it is *simple* if $\mathfrak{U} \cap \mathfrak{R}^\perp = \{0\}$.

As first shown by Sz.-Nagy and Foiaş [25] and Brodskiĭ [15], every $\mathcal{B}(U;Y)$-valued Schur function θ has a simple conservative realization, which is unique up to unitary similarity. It also has a minimal passive realization (as was noticed by Arov [3]). The latter realization is not unique. Any two minimal passive realizations $\Sigma = [[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}]; (Y, X, U)]$ and $\widetilde{\Sigma} = \left[\left[\begin{smallmatrix} \widetilde{A} & \widetilde{B} \\ \widetilde{C} & \widetilde{D} \end{smallmatrix}\right]; (Y, \widetilde{X}, U)\right]$ of θ are pseudo-similar to each other in the sense that there is a closed (possibly unbounded) injective linear operator Q with dense domain $\mathcal{D}(Q) \subset X$ and dense range $\mathcal{R}(Q) \subset \widetilde{X}$ such that

$$\begin{bmatrix} \widetilde{A} & \widetilde{B} \\ \widetilde{C} & \widetilde{D} \end{bmatrix} \begin{bmatrix} Qx \\ u \end{bmatrix} = \begin{bmatrix} QA & QB \\ C & D \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}, \qquad x \in \mathcal{D}(Q), \qquad u \in U.$$

In particular, $\mathcal{R}(B) \subset \mathcal{D}(Q)$, and A maps $\mathcal{D}(Q)$ into itself. We refer the reader to [12] or [24, Section 9.2] for details.[1]

Among all minimal passive realizations of θ there is one whose norm is the weakest possible one (in the sense that the pseudo-similarity which maps the state space of any other minimal passive realization into the state space of this particular realization is a contraction). It is clearly unique up to unitary similarity. We call this realization a *minimal optimal passive* realization, and denote it by $\Sigma_\circ = \left[\left[\begin{smallmatrix} A_\circ & B_\circ \\ C_\circ & D_\circ \end{smallmatrix}\right]; (Y, X_\circ, U)\right]$. There is also another minimal passive realization of θ whose norm is the strongest possible one. We call this realization a *minimal $*$-optimal* one and denote it by $\Sigma_\bullet = \left[\left[\begin{smallmatrix} A_\bullet & B_\bullet \\ C_\bullet & D_\bullet \end{smallmatrix}\right]; (Y, X_\bullet, U)\right]$. See [7, 8, 11] for more details on these two realizations (and also for non-minimal versions[2] of these two types of realizations, as well as passive realizations in general).

[1]The domain of such a pseudo-similarity Q is not unique in general. However, if Q is bounded (or Q^{-1} is bounded), then the domain (or the range) of Q is the whole space, and in this case it is unique.

[2]Every (non-minimal) optimal system is a passive dilation of a minimal optimal system, and every $*$-optimal system is a passive dilation of a minimal $*$-optimal system.

3. The Balanced Passive Realization

From the minimal optimal and $*$-optimal realizations of $\theta \in S(U;Y)$ we can construct still another one, the *balanced passive realization*, by using interpolation. In the control literature this realization is often called the *bounded real balanced realization* (see, e.g., [20, Section 5]), and it was originally introduced by Opdenacker and Jonckheere [21] in a continuous time impedance setting. The most common among all balanced realization is the *Hankel balanced realization*, whose infinite-dimensional version was first developed by Young [26].

Let $\Sigma_\circ = \left[\left[\begin{smallmatrix} A_\circ & B_\circ \\ C_\circ & D_\circ \end{smallmatrix}\right]; (Y, X_\circ, U)\right]$ be a minimal optimal realization of θ, and let $\Sigma'_\bullet = \left[\left[\begin{smallmatrix} A'_\bullet & B'_\bullet \\ C'_\bullet & D'_\bullet \end{smallmatrix}\right]; (Y, X'_\bullet, U)\right]$ be a minimal $*$-optimal realization of θ. These two systems are pseudo-similar. Let Q' be the pseudo-similarity mapping $X'_\bullet \supset \mathcal{D}(Q') \to \mathcal{R}(Q') \subset X_\circ$. Since $\Sigma_\circ$ has the weakest possible norm and $\Sigma'_\bullet$ has the strongest possible norm among all minimal passive realizations of θ, Q' is a contraction. In particular, as Q' is closed, $\mathcal{D}(Q') = X'_\bullet$. Let $X_\bullet = \mathcal{R}(Q') \subset X_\circ$. Then $X_\bullet$ is dense in $X_\circ$. We make $X_\bullet$ into a Hilbert space by defining $\|x\|_{X_\bullet} = \|(Q')^{-1}x\|_{X'_\bullet}$. Let Q be the operator that we get from Q' by interpreting Q as an operator $X'_\bullet \to X_\bullet$ (i.e., it is otherwise the same operator as Q, but its range space is $X_\bullet$ instead of $X_\circ$). Clearly Q is unitary $X'_\bullet \to X_\bullet$. Let $\Sigma_\bullet = \left[\left[\begin{smallmatrix} A_\bullet & B_\bullet \\ C_\bullet & D_\bullet \end{smallmatrix}\right]; (Y, X_\bullet, U)\right]$ be the system defined by

$$\begin{bmatrix} A_\bullet & B_\bullet \\ C_\bullet & D_\bullet \end{bmatrix} = \begin{bmatrix} QA'_\bullet Q^{-1} & QB'_\bullet \\ C'_\bullet Q^{-1} & D'_\bullet \end{bmatrix}.$$

Then $\Sigma_\bullet$ and $\Sigma'_\bullet$ are unitarily similar. In particular $\Sigma_\bullet$ is minimal and $*$-optimal. This system can be interpreted as a *restriction* of $\Sigma_\circ$ in the sense that

$$\begin{bmatrix} A_\bullet & B_\bullet \\ C_\bullet & D_\bullet \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} A_\circ & B_\circ \\ C_\circ & D_\circ \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}, \qquad \begin{bmatrix} x \\ u \end{bmatrix} \in X_\bullet \oplus U.$$

The pseudo-similarity $Q_\bullet \colon X_\bullet = \mathcal{D}(Q_\bullet) \to \mathcal{R}(Q_\bullet) \subset X_\circ$ between these two systems is simply the contractive embedding operator $X_\bullet \to X_\circ$, i.e., for all $x \in X_\bullet$ we have $Q_\bullet x = x \in X_\circ$ and $\|Q_\bullet x\|_{X_\circ} = \|x\|_{X_\circ} \le \|x\|_{X_\bullet}$.

Let $E_\bullet \in \mathcal{B}(X_\bullet)$ be the Gram operator corresponding to the embedding $X_\bullet \subset X_\circ$, i.e., $E_\bullet$ is the positive self-adjoint injective contraction on $X_\bullet$ which is determined by the fact that $\langle x, y\rangle_{X_\circ} = \langle x, E_\bullet y\rangle_{X_\bullet}$ for all x, $y \in X_\bullet$. This operator can be extended to an operator $E_\circ \in \mathcal{B}(X_\circ)$ in the following way. Since $E_\bullet$ is a self-adjoint contraction in $X_\bullet$, we have $E_\bullet^3 \le E_\bullet$, and therefore, for all $x \in X_\bullet$,

$$\begin{aligned} \|E_\bullet x\|^2_{X_\circ} &= \langle E_\bullet x, E_\bullet x\rangle_{X_\circ} = \langle E_\bullet x, E_\bullet^2 x\rangle_{X_\bullet} = \langle x, E_\bullet^3 x\rangle_{X_\bullet} \\ &\le \langle x, E_\bullet x\rangle_{X_\bullet} = \|x\|^2_{X_\circ}. \end{aligned}$$

Thus, we may interpret $E_\bullet$ as a densely defined contraction $X_\circ \to X_\circ$. By continuity, it can be extended to a contraction $E_\circ \in \mathcal{B}(X_\circ)$. To see that $E_\circ$ is self-adjoint in $X_\circ$ we argue as follows. For all x, $y \in X_\bullet$ we have

$$\begin{aligned} \langle y, E_\circ x\rangle_{X_\circ} &= \langle y, E_\bullet x\rangle_{X_\circ} = \langle y, E_\bullet^2 x\rangle_{X_\bullet} = \langle E_\bullet y, E_\bullet x\rangle_{X_\bullet} \\ &= \langle E_\bullet y, x\rangle_{X_\circ} = \langle E_\circ y, x\rangle_{X_\circ}. \end{aligned}$$

By continuity, the identity $\langle y, E_\circ x\rangle_{X_\circ} = \langle E_\circ y, x\rangle_{X_\circ}$ must hold for all $x,\ y \in X_\circ$. Thus, $E_\circ$ is self-adjoint. We remark that $X_\bullet = \mathcal{R}\big(E_\circ^{1/2}\big)$ (where $E_\circ^{1/2}$ is the positive square root of $E_\circ$ in $X_\circ$), and that

$$\langle x, y\rangle_\bullet = \langle E_\circ^{-1/2} x, E_\circ^{-1/2} y\rangle_{X_\circ}, \qquad x,\ y \in X_\bullet.$$

Let $X_\odot = \mathcal{R}\big(E_\circ^{1/4}\big)$ (here $E_\circ^{1/4}$ is the positive 4^{th} root of $E_\circ$ in $X_\circ$). This subspace of $X_\circ$ becomes a Hilbert space if we equip it with the inner product

$$\langle x, y\rangle_\odot = \langle E_\circ^{-1/4} x, E_\circ^{-1/4} y\rangle_{X_\circ} = \langle E_\circ^{1/4} x, E_\circ^{1/4} y\rangle_{X_\bullet}, \qquad x,\ y \in X_\odot.$$

With this inner product, we have $X_\bullet \subset X_\odot \subset X_\circ$, with contractive and dense embeddings. (The space $X_\odot$ is known as the *Riesz interpolation space* between $X_\circ$ and $X_\bullet$ with exponent $\frac{1}{2}$. A full scale of Hilbert spaces parameterized by $\alpha \in (0,1)$ is constructed in [19, p. 142].)

Theorem 3.1. *Let $X_\bullet \subset X_\odot \subset X_\circ$ be the spaces defined above. Define the operators $A_\odot$, $B_\odot$, $C_\odot$, and $D_\odot$ by*

$$\begin{bmatrix} A_\odot & B_\odot \\ C_\odot & D_\odot \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} A_\circ & B_\circ \\ C_\circ & D_\circ \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}, \qquad \begin{bmatrix} x \\ u \end{bmatrix} \in X_\odot \oplus U. \tag{5}$$

Then $\left[\begin{smallmatrix} A_\odot & B_\odot \\ C_\odot & D_\odot \end{smallmatrix}\right] \in \mathcal{B}(X_\odot \oplus U; X_\odot \oplus U)$, and $\big[\left[\begin{smallmatrix} A_\odot & B_\odot \\ C_\odot & D_\odot \end{smallmatrix}\right]; (Y, X_\odot, U)\big]$ is a minimal passive realization of θ.

We call $\Sigma_\odot$ (and any other system which is unitarily similar to $\Sigma_\odot$) a *balanced passive realization* of θ.

The proof of this theorem is based on the following result on Riesz interpolation.

Lemma 3.2. *Let $X_\bullet$, $X_\circ$, $Y_\bullet$, and $Y_\circ$ be four Hilbert spaces, with $X_\bullet \subset X_\circ$ and $Y_\bullet \subset Y_\circ$ (with continuous embeddings). Let $X_\odot$ and $Y_\odot$ be the Riesz interpolation spaces with exponent $\frac{1}{2}$ between $X_\bullet$ and $X_\circ$ respectively $Y_\bullet$ and $Y_\circ$ (constructed as explained above). If $A_\circ$ is a contraction from $X_\circ$ into $Y_\circ$ with the property that $A_\bullet := A_{\circ|X_\bullet}$ is a contraction from $X_\bullet$ into $Y_\bullet$, then $A_\odot := A_{\circ|X_\odot}$ is a contraction from $X_\odot$ into $Y_\odot$ (in particular, the range of $A_\odot$ lies in $Y_\odot$).*

This lemma is contained in [19, Theorem 9.1, p. 144] (take the exponent to be $\frac{1}{2}$ in that theorem). The case where $X_\circ = Y_\circ$ and $X_\bullet = Y_\bullet$ is also found in [24, Lemma 9.5.8]. An even more general version is given in [1, Theorem C.4, p. 283]. There also exponents $\alpha \in (0,1)$ different from $\frac{1}{2}$ are allowed, and our requirements $X_\bullet \subset X_\circ$ and $Y_\bullet \subset Y_\circ$ have been relaxed to the requirement that these subspaces should be compatible in the sense of interpolation theory. If $A_\circ$ and $A_\bullet$ are just bounded operators rather than contractions, then the conclusion is that $A_\odot$ is a bounded operator with $\|A_\odot\| \le \|A_\bullet\|^\alpha \|A_\circ\|^{1-\alpha}$.

Proof of Theorem 3.1. We begin by showing that that the ranges of $A_\odot$ and $B_\odot$ (with domains $X_\odot$ respective U) are contained in $X_\odot$. The latter inclusion follows trivially from the fact that $\mathcal{R}(B_\odot) = \mathcal{R}(B_\bullet) \subset X_\bullet \subset X_\odot$. The former inclusion is a consequence of Lemma 3.2 with $Y_\circ = X_\circ$ and $Y_\bullet = X_\bullet$. (In addition, we find that $A_\odot$ is a contraction on $X_\odot$.)

We next show that $\Sigma_\odot$ is passive. To do this we use Lemma 3.2 with the following substitutions:

$$X_\circ \to X_\circ \oplus U,\ X_\bullet \to X_\bullet \oplus U,\ Y_\circ \to X_\circ \oplus Y,\ Y_\bullet \to X_\bullet \oplus Y,\ A_\circ \to \left[\begin{smallmatrix} A_\circ & B_\circ \\ C_\circ & D_\circ \end{smallmatrix}\right].$$

According to Lemma 3.2, $\left[\begin{smallmatrix} A_\odot & B_\odot \\ C_\odot & D_\odot \end{smallmatrix}\right]$ is contractive, hence $\Sigma_\odot$ is passive.

To see that $\Sigma_\odot$ is controllable we argue as follows. By construction, the reachability maps $\mathfrak{B}_\bullet$ and $\mathfrak{B}_\odot$ of $\Sigma_\bullet$ respectively $\Sigma_\odot$ have the same range. By assumption, $\Sigma_\bullet$ is controllable, i.e., $\mathcal{R}(\mathfrak{B}_\bullet)$ is dense in $X_\bullet$. Since $X_\bullet$ is dense in $X_\odot$, also $\mathcal{R}(\mathfrak{B}_\odot) = \mathcal{R}(\mathfrak{B}_\bullet)$ is dense in $X_\odot$. Thus $\Sigma_\odot$ is controllable. That $\Sigma_\odot$ is observable follows from the fact that the null space of the observability map $\mathfrak{C}_\odot$ of $\Sigma_\odot$ is a subset of the null space of the observability map $\mathfrak{C}_\circ$ of $\Sigma_\circ$, and the latter is trivial since $\Sigma_\circ$ is observable.

That the transfer function of $\Sigma_\odot$ is θ follows from the fact that for all z with $|z| > 0$ and all $u \in U$ (recall that $\mathcal{R}(B_\odot) = \mathcal{R}(B_\bullet) \subset X_\bullet$)

$$[C_\odot(z - A_\odot)^{-1}B_\odot + D_\odot]u = [C_\bullet(z - A_\bullet)^{-1}B_\bullet u + D_\bullet]u = \theta(z)u. \quad \square$$

4. Inner Dilations

A function $\theta \in S(U;Y)$ (on the unit disk $\mathbb{D}$) has almost everywhere defined limits $\theta(e^{i\varphi}) := \lim_{r\uparrow 1} \theta(re^{i\varphi})$ on the unit circle $\mathbb{T}$ in the strong sense. By an *inner* $\mathcal{B}(U;Y)$-valued function on $\mathbb{D}$ we mean a function $\theta \in S(U;Y)$ satisfying $\theta(z)^*\theta(z) = I$ for almost all z with $|z| = 1$. The function $\theta \in S(U;Y)$ is *co-inner* is $\theta(\overline{z})^*$ is inner, and it is *bi-inner* if it is both inner and co-inner.

Definition 4.1. Let $\theta \in S(U;Y)$.

1) By an *inner dilation* of θ we mean an inner function Θ of the form $\Theta = \left[\begin{smallmatrix} \theta \\ \theta_r \end{smallmatrix}\right]$, where $\theta_r \in S(U;Y_1)$ for some Hilbert space Y_1.
2) By a *co-inner dilation* of θ we mean a co-inner function Θ of the form $\Theta = \begin{bmatrix} \theta_l & \theta \end{bmatrix}$, where $\theta_l \in S(U_1;Y)$ for some Hilbert space U_1.
3) By a *bi-inner dilation* of θ we mean a bi-inner function $\Theta = \left[\begin{smallmatrix} \theta_{11} & \theta \\ \theta_{21} & \theta_{22} \end{smallmatrix}\right] \in S(U_1 \oplus U; Y \oplus Y_1)$ for some Hilbert spaces U_1 and Y_1.

Not every $\theta \in S(U;Y)$ has an inner, or co-inner, or bi-inner dilation. But obviously, if θ has a bi-inner dilation, then it has both an inner dilation and a co-inner dilation. Our main theorem, stated below, says that the converse statement is also true.

Theorem 4.2. *Let $\theta \in S(U;Y)$. Then θ has a bi-inner dilation if and only if it has both an inner dilation and a co-inner dilation.*

This is a part of Corollary 4.6 below, which in turn follows from Lemmas 4.3–4.5 below.

Lemma 4.3 (see [8, Proposition 3]**).**

1) *θ has an inner dilation if and only if θ has a minimal passive strongly stable realization.*
2) *θ has a co-inner dilation if and only if θ has a minimal passive strongly co-stable realization.*
3) *θ has a bi-inner dilation if and only if θ has a minimal passive realization with is both strongly stable and strongly co-stable.*

Proof. If we ignore the word "minimal", then this is [8, Proposition 3]. However, from any non-minimal realization we can always get a minimal one by first replacing the state space by the reachable subspace, restricting A and C to this subspace, and then projecting the state space onto the orthogonal complement of the unobservable subspace. (See [8, 11] for details.) □

Lemma 4.4.

1) *θ has an inner dilation if and only if the minimal optimal passive realization of θ is strongly stable.*
2) *θ has a co-inner dilation if and only if the minimal $*$-optimal passive realization of θ is strongly co-stable.*

This lemma could be derived from [6, Theorem 3]. For the convenience of the reader we include a proof.

Proof. We prove only 1) and leave the analogous proof of 2) to the reader.

Suppose that θ has an inner dilation. Then it has a minimal strongly stable realization Σ. The norm of the minimal optimal passive realization $\Sigma_\circ$ is the weakest one among all passive realizations, and therefore the pseudo-similarity Q which maps the state space X of Σ into the state space $X_\circ$ of $\Sigma_\circ$ is a contraction. The image of X under Q is dense in $X_\circ$, and each (autonomous) trajectory which starts in this set is the image of a trajectory of Σ, hence it tends to zero. Since all trajectories in $\Sigma_\circ$ are bounded, this implies that all trajectories of $\Sigma_\circ$ tend to zero. Thus, $\Sigma_\circ$ is strongly stable.

The converse statement is trivial. □

Lemma 4.5. *The balanced passive realization of θ is strongly stable if and only if the minimal optimal realization of θ is strongly stable. The balanced passive realization of θ is strongly co-stable if and only if the minimal $*$-optimal realization of θ is strongly co-stable.*

Proof. It follows from Lemmas 4.3 and 4.4 that the minimal optimal realization $\Sigma_\circ$ is strongly stable whenever the balanced realization $\Sigma_\odot$ is strongly stable, and that the minimal $*$-optimal realization $\Sigma_\bullet$ is strongly co-stable whenever $\Sigma_\odot$ is strongly co-stable.

Suppose that $\Sigma_\circ$ is strongly stable. Let $x \in X_\bullet$. Then for all $n = 0, 1, 2, \ldots$, by the relationships between the different inner products and the Schwartz inequality,

$$\begin{aligned}\|A^n x\|^2_{X_\odot} &= \|E^{1/4} A^n x\|^2_{X_\bullet} = \langle E^{1/4} A^n x, E^{1/4} A^n x\rangle_{X_\bullet} \\ &= \langle A^n x, E^{1/2} A^n x\rangle_{X_\bullet} \le \|A^n x\|_{X_\bullet} \|E^{1/2} A^n x\|_{X_\bullet} \\ &= \|A^n x\|_{X_\bullet} \|A^n x\|_{X_\circ}.\end{aligned}$$

Let $n \to \infty$. Then $\|A^n x\|_{X_\circ} \to 0$ whereas $\|A^n x\|_{X_\bullet}$ stays bounded. Thus $\|A^n x\|_{X_\odot} \to 0$. This being true on a dense subset (and since $A_{X_\odot}$ is a contraction), the same statement must be true for all $x \in X_\odot$. Thus, $\Sigma_\odot$ is strongly stable, as claimed.

That $\Sigma_\odot$ is strongly co-stable whenever $\Sigma_\bullet$ is strongly co-stable is proved in a similar way. □

Corollary 4.6. *Let θ be a $\mathcal{B}(U; Y)$-valued Schur function on the open unit disk $\mathbb{D}$. Then the following conditions are equivalent.*

1) *θ has a bi-inner dilation.*
2) *θ has both an inner dilation and a co-inner dilation.*
3) *The balanced passive realization of θ is both strongly stable and strongly co-stable.*
4) *The minimal optimal passive realization of θ is strongly stable and the minimal $*$-optimal passive realization of θ is strongly co-stable.*
5) *θ has a minimal passive realization with is both strongly stable and strongly co-stable.*

This follows from Lemmas 4.3, 4.4, and 4.5 (and from [8, Proposition 3] for the non-minimal version of 5)).

By Definition 4.1, $\theta \in S(U; Y)$ has an inner dilation if and only if the factorization problem

$$\psi_r^*(z)\psi_r(z) = I - \theta^*(z)\theta(z), \text{ a.e. on } \mathbb{T}, \tag{6}$$

has a solution $\theta_r \in S(U; Y_1)$ for some Hilbert space Y_1, and $\theta \in S(U; Y)$ has a co-inner dilation if and only if the factorization problem

$$\psi_l(z)\psi_l^*(z) = I - \theta(z)\theta^*(z), \text{ a.e. on } \mathbb{T}, \tag{7}$$

has a solution $\theta_l \in S(U_1; Y)$ for some Hilbert space U_1. Both of these problems are special cases of the *right or left spectral factorization* that we get by replacing the right-hand sides of (6) or (7) by an operator-valued function f which is integrable and strictly positive a.e. on $\mathbb{T}$. In the case of a nonnegative scalar function f the Szegö theorem gives necessary and sufficient conditions for the existence of H^2-solutions of these spectral factorization problems. Analogous conditions for matrix-valued functions f satisfying $f(z) > 0$ for almost all $z \in \mathbb{T}$ are given by the Zasuhin–Krein theorem. Finally, for operator-valued functions f the factorization theorem of Devinatz gives sufficient conditions. All of these conditions have the property that if they hold when we replace f by $I - \theta^*\theta$, then they also hold if we replace f by $I-\theta\theta^*$. By combining these criteria with Corollary 4.6 we obtain the following result.

Corollary 4.7.

1) *A scalar-valued Schur function θ, which is not inner, has a bi-inner dilation if and only if*
$$\int_{\mathbb{T}} -\ln\bigl(1-|\theta(z)|\bigr)d|z|<\infty.$$
2) *An $n\times n$ matrix-valued Schur function θ which satisfies $\theta^*(z)\theta(z)<I$ for almost all $z\in\mathbb{T}$ has a bi-inner dilation if and only if*
$$\int_{\mathbb{T}} -\ln\det\bigl(I-\theta^*(z)\theta(z)\bigr)d|z|<\infty.$$
3) *A function $\theta\in S(U;Y)$ (where U and Y are allowed to be infinite-dimensional) has a bi-inner dilation if*
$$\int_{\mathbb{T}} -\ln\bigl(1-\|\theta(z)\|^2\bigr)d|z|<\infty.$$

Proof. Assertion 1) follows from Corollary 4.6 together with Szegö's theorem (see [23, p. 110]), which says the following: given a nonnegative function $f\in L^1(\mathbb{T})$, the factorization problem
$$|\psi(z)|^2=f(z) \text{ a.e. on } \mathbb{T},$$
has a solution ψ in the Hardy class $H^2(\mathbb{D})$ if and only if $\int_{\mathbb{T}}-\ln f(z)d|z|<\infty$. We take $f(z)=1-|\theta(z)|^2$, and notice that $\ln\bigl(1-|\theta(z)|\bigr)^2=\ln(1-|\theta(z)|)+\ln(1+|\theta(z)|)$, where the latter function is essentially bounded, hence $\int_{\mathbb{T}}-\ln\bigl(1-|\theta(z)|\bigr)d|z|<\infty$ if and only if $\int_{\mathbb{T}}-\ln f(z)d|z|<\infty$. Cauchy's formula and the boundedness of ψ on $\mathbb{T}$ imply that ψ must actually be a Schur function.

Assertion 2) is proved in a similar way. We replace Szegö's theorem by the Zasuhin–Krein theorem (see [25, part c) of Proposition 7.1, p. 227] and also the discussion on [25, p. 236]). According to this theorem, if $\theta^*(z)\theta(z)<I$ for almost all $z\in\mathbb{T}$, or equivalently, if $\theta(z)\theta^*(z)<I$ for almost all $z\in\mathbb{T}$, then the factorization problem (6) has a H^2-solution if and only if $\int_{\mathbb{T}}-\ln\det\bigl(I-\theta^*(z)\theta(z)\bigr)d|z|<\infty$, whereas (7) has a H^2-solution if and only if $\int_{\mathbb{T}}-\ln\det\bigl(I-\theta(z)\theta^*(z)\bigr)d|z|<\infty$. However, $\det\bigl(I-\theta^*(z)\theta(z)\bigr)=\det\bigl(I-\theta(z)\theta^*(z)\bigr)$, so if one of the two problems has a solution, then so has the other. Again Cauchy's formula implies that the factors ψ_r and ψ_l in (3) will be Schur functions.

The proof of assertion 3) is similar to the proof of assertion 2), but this time we use the Devinatz factorization theorem (see, [25, part b) of Proposition 7.1, p. 227]). □

Example 4.8. Let $\theta(z)=(3+z)^{-1/2}$, where we take the branch of the square root which is analytic on $\mathbb{C}\setminus(-\infty,-3]$). This function is analytic on $\mathbb{D}$. It is easy to see that it is contractive, with an absolute value which is bounded away from both zero and 1. Therefore, by part 1) of Corollary 4.7, this function has an operator-valued bi-inner dilation (with infinite-dimensional input and output spaces). However, according to Arov [2] and Dewilde [17], it cannot have a matrix-valued bi-inner dilation since it is does not have a pseudo-continuation to $\mathbb{C} \setminus \overline{\mathbb{D}}$ with bounded Nevanlinna characteristic (it has a branch point at -3).

The above example also sheds some additional light on the theory about the existence of a *D-realization*[3] of a given Schur function θ and how this property is related to the existence of other passive realizations of θ with some special properties. According to Arov [8], θ has a D-representation if an only if θ has a realization *with minimal losses* which is both strongly stable and strongly co-stable. On the other hand, by [2], [17], a scalar-valued or matrix-valued Schur function θ has a D-representation if and only if θ has a pseudo-continuation to $\mathbb{C}\setminus\overline{\mathbb{D}}$ with bounded Nevanlinna characteristic. Thus, the balanced realization of the Schur function in Example 4.8 is both strongly stable and strongly co-stable, but it *does not have minimal losses.* The function θ in Example 4.8 is a typical example of a scalar *Schur function which has a bi-inner dilation but no D-representation.*

5. Continuous Time Bi-Inner Dilations

Results analogous to those presented in the last two sections are valid also in the case where the given Schur function θ is defined on the open right-half plane $\mathbb{C}^+$ (instead of on the open unit disk $\mathbb{D}$). In this case we use the L^2-well-posed continuous time realizations described in, e.g., [13] and [24]. All the proofs remain essentially the same, except for the fact that references to known discrete time results are replaced by references to the analogous continuous time results, all of which are found in [24] (and many of them also in [13]). In particular, the minimal optimal, $*$-optimal and balanced passive realizations are described in [24, Section 11.8]. Alternatively, the continuous time case can be reduced to the discrete time case by means of the Cayley transform.

References

[1] Jim Agler and John E. McCarthy, *Pick interpolation and Hilbert function spaces*, Graduate Studies in Mathematics, vol. 44, American Mathematical Society, Providence, RI, 2002.

[2] Damir Z. Arov, *Darlington's method for dissipative systems*, Dokl. Akad. Nauk SSSR. **201** (1971), 559–562, Translated in Soviet Physics Dokl. 16 (1971), 954–956.

[3] ———, *Scattering theory with dissipation of energy*, Dokl. Akad. Nauk SSSR. **216** (1974), 713–716, Translated in Soviet Math. Dokl. 15 (1974), 848–854.

[4] ———, *Unitary couplings with losses (a theory of scattering with losses)*, Funkcional. Anal. i Priložen. **8** (1974), 5–22, Translated in Funct. Anal. Appl. 8 (1974), 280–294.

[3] The notions of a D-representation and a passive system with minimal losses was explained briefly in the introduction.

[5] ———, *The realization of a canonical system with dissipative boundary conditions at one end of a segment in terms of the coefficient of dynamic flexibility*, Sibir. Mat. Zh. **16** (1975), 440–463, Translated in Sib. Math. J. 16 (1975), 335-352 (1976).

[6] ———, *Optimal and stable passive systems*, Dokl. Akad. Nauk SSSR. **247** (1979), 265–268, Translated in Soviet Math. Dokl. 20 (1979), 676–680.

[7] ———, *Passive linear stationary dynamic systems*, Sibir. Mat. Zh. **20** (1979), 211–228, Translated in Sib. Math. J. 20 (1979), 149-162.

[8] ———, *Stable dissipative linear stationary dynamical scattering systems*, J. Operator Theory **1** (1979), 95–126, Translated in [10].

[9] ———, *Scattering matrix and impedance of a canonical differential system with a dissipative boundary condition whose coefficient is a rational matrix-valued function of the spectral parameter (in Russian)*, Algebra Anal. **13** (2001), 26–53, Translated in St. Petersburg Math. J. 13, No.4, 527–547 (2002).

[10] ———, *Stable dissipative linear stationary dynamical scattering systems*, Interpolation Theory, Systems Theory, and Related Topics. The Harry Dym Anniversary Volume (Basel Boston Berlin), Operator Theory: Advances and Applications, vol. 134, Birkhäuser-Verlag, 2002, English translation of the article in J. Operator Theory 1 (1979), 95–126, pp. 99–136.

[11] Damir Z. Arov, Marinus A. Kaashoek, and Derk R. Pik, *Minimal and optimal linear discrete time-invariant dissipative scattering systems*, Integral Equations Operator Theory **29** (1997), 127–154.

[12] ———, *The Kalman–Yakubovich–Popov inequality and infinite dimensional discrete time dissipative systems*, J. Operator Theory (2005), to appear.

[13] Damir Z. Arov and Michael A. Nudelman, *Passive linear stationary dynamical scattering systems with continuous time*, Integral Equations Operator Theory **24** (1996), 1–45.

[14] Vitold Belevitch, *Classical network theory*, Holden-Day, San Francisco, Calif.-Cambridge-Amsterdam, 1968.

[15] V. M. Brodskiĭ, *On operator colligations and their characteristic functions*, Soviet Math. Dokl. **12** (1971), 696–700.

[16] Sidney Darlington, *Synthesis of reactance 4-poles which produce prescribed insertion loss characteristics including special applications to filter design*, Journal of Mathematics and Physics **18** (1939), 257–353.

[17] Patrick Dewilde, *Input-output description of roomy systems*, SIAM J. Control Optim. **14** (1971), 712–736.

[18] Ronald G. Douglas and J. William Helton, *Inner diations of analytic matrix functions and Darlington synthesis*, Acta Sci. Math (Szeged) **34** (1973), 61–67.

[19] Selim G. Kreĭn and Ju. I. Petunin, *Scales of Banach spaces*, Uspehi Mathem. Nauk SSSR **21** (1966), 89–168, Translated in Russian Math. Surveys 21 (1966), 85–159.

[20] Raimund Ober, *Balanced parameterization of classes of linear operators*, SIAM J. Control Optim. **29** (1991), 1251–1287.

[21] Philippe C. Opdenacker and Edmond A. Jonckheere, *A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds*, IEEE Trans. Circuits Systems **35** (1988), 184–189.

[22] Derk R. Pik, *Time-variant Darlington synthesis and induced realizations*, Internat. J. Appl. Math. Comput. Sci. **6** (2001), 1331–1360.

[23] Marvin Rosenblum and James Rovnyak, *Hardy classes and operator theory*, Oxford University Press, New York, Oxford, 1985.

[24] Olof J. Staffans, *Well-posed linear systems*, Cambridge University Press, Cambridge and New York, 2005.

[25] Béla Sz.-Nagy and Ciprian Foiaş, *Harmonic analysis of operators on Hilbert space*, North-Holland, Amsterdam London, 1970.

[26] Nicolas J. Young, *Balanced realizations in infinite dimensions*, Operator Theory and Systems (Basel Boston Berlin), Operator Theory: Advances and Applications, vol. 19, Birkhäuser-Verlag, 1986, pp. 449–471.

Damir Z. Arov
Department of Mathematics
South-Ukrainian Pedagogical University
270020 Odessa, Ukraine

Olof J. Staffans
Åbo Akademi University
Department of Mathematics
FIN-20500 Åbo, Finland
e-mail: `http://www.abo.fi/~staffans/`