What Is the Adjoint of a Linear System?

Omran Kouba and Dennis S. Bernstein
POC: D. S. Bernstein (dsbaero@umich.edu)

For About This Issue: In their tutorial article "What Is the Adjoint of a Linear System," Omran Kouba and Dennis Bernstein view a linear system as an infinite-dimensional operator and use this framework to derive the corresponding adjoint operator. Their goal is to show that the form of the adjoint operator explains the duality between estimation and control as well as the fact that, in classical optimal control, time runs backwards.

Summary: Duality is a bedrock principle of control and estimation, but where does this duality come from? The short answer is: adjoints. In fact, adjoints play a key role in numerous branches of engineering and computation, but what exactly is an adjoint? This article describes how adjoints are used to compute sensitivities and then derives the adjoint of a linear time-varying system as an operator between Hilbert spaces. A connection is made with the costate equation of classical optimal control theory, which is shown to be a partial adjoint. This tutorial article is intended for all students of systems and control theory who want to understand why duality shows up in unexpected ways.

Although controllability and observability are distinct properties, one of the fundamental—and most attractive—results of our field is the fact that (A, B) is controllable if and only if $(A^{\rm T}, B^{\rm T})$ is observable. As mentioned in "Summary," this duality provides a deep linkage between the linear-quadratic regulator (LQR), which seeks a feedback gain K such that $A+BK$ is asymptotically stable, and the linear-quadratic estimator (LQE), which seeks an output-error-injection gain F such that $A+FC$ is asymptotically stable. In the case of LQR, the controllability of (A, B) implies that there exists a feedback gain K that arbitrarily places the eigenvalues of $A + BK$, thus facilitating closed-loop asymptotic stability. In the dual case of LQE, the observability of (A, C) implies that there exists an error-injection gain F that arbitrarily places the eigenvalues of $A+FC$, thus facilitating closed-loop asymptotic stability of the error dynamics. A key distinction worth noting is that $A + BK$ is the dynamics matrix of a physical feedback loop, whereas $A + FC$ is the dynamics matrix of a nonphysical error system.

As mention in the Summary, the duality between estimation and control suggests that something deeper is going on. As shown by the elegant approach of [1, p. 198], these problems are related by the adjoint operator. In the case of an $n \times m$ real matrix A, the adjoint of A is simply its transpose A^{T}, which is an $m \times n$ matrix. Within the context of a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$, which can be thought of as an infinite-dimensional Euclidean space, the adjoint of a bounded linear operator $\mathcal{A}\colon \mathcal{H} \to \mathcal{H}$ is the linear operator $\mathcal{A}^*$ that satisfies $\langle \mathcal{A}v, w \rangle = \langle v, \mathcal{A}^* w \rangle$ for all $v, w \in \mathcal{H}$. Bounded and unbounded linear operators are extensively studied in mathematics [2], [3]. Although unbounded linear operators are not considered in this article, these operators represent partial differential operators, which define infinite-dimensional systems [4].

Adjoints are ubiquitous in computational mathematics and optimization theory [5, pp. 150–154, 159]. Data assimilation and optimization methods use adjoint models to compute the sensitivity of a cost function to parameter changes [6]–[9]. Within the context of optimal control theory, the adjoint system defines the costate, which appears in the statement of the minimum principle based on strong control variations and first-order variational necessary conditions based on weak control variations [10, pp. 188, 233], [11, p. 258]. The adjoint operator is used in [1, pp. 198, 199] to derive the optimal estimator from the optimal regulator; this approach thus reveals the origin of the deep duality between estimation and control mentioned above.

Within the context of differential equations, a linear system can be viewed as a linear operator that maps a vector-valued, square-integrable input u to a vector-valued, square-integrable output y. The linear dynamical system thus defines a bounded linear operator that maps one Hilbert space to another Hilbert space. The adjoint of this linear operator corresponds to a linear system that is different from the original linear system. The goal of this paper is to derive the dynamics of the adjoint system.

A bounded linear operator that maps one Hilbert space to another Hilbert space can be associated with its *adjoint operator*. For an operator defined by a linear dynamical system, the adjoint operator can be expressed in terms of another linear dynamical system; this *adjoint dynamical system* is closely related to the original linear dynamical system. The purpose of this article is to derive the adjoint dynamical system and show that it does, in fact, represent the adjoint of the bounded linear operator corresponding to the original dynamical system. Summaries of this result are given in [1, pp. 85, 86] and [12, pp. 68–70].

This note also presents the adjoint differential equation for the costate, which, as mentioned above, arises in classical optimal control theory. We call this the *partial adjoint* since it does not represent the adjoint operator per se but rather provides a computationally convenient representation of the state transition matrix. Details are provided in “The Costate Equation of Optimal Control as a Partial Adjoint.”

Using Adjoints to Determine Sensitivity

One of the most fundamental problems in mathematics and engineering is to compute the solution x of the matrix-vector equation

$$Ax = b, \tag{1}$$

where A is a matrix, b is a column vector, and x is a column vector. As discussed in "Existence and Uniqueness of Solutions to $Ax = b$ and $A^{\mathrm{T}}y = c$," (1) may or may not have a solution, and, if (1) does have a solution, then either it is the unique solution or (1) has infinitely many solutions. The problem of interest is to determine how a function of the form $f(x, c) = c^{\mathrm{T}}x$, where c is a given vector, changes as b changes. This dependence can be determined numerically by solving (1) for many different values of b; however, this is inconvenient in the case where the dimensions of x and b are large. Adjoints provide a computationally efficient approach to determining this sensitivity.

Let $A \in \mathbb{R}^{n\times m}$. For all $b \in \mathbb{R}^n$, define

$$\mathcal{X}(b) \triangleq \{x \in \mathbb{R}^m \colon Ax = b\}, \tag{2}$$

and, for all $x \in \mathcal{X}(b)$ and $c \in \mathbb{R}^m$, define the *primal cost* $f\colon \mathcal{X}(b) \times \mathbb{R}^m \to \mathbb{R}$ by

$$f(x, c) \triangleq c^{\mathrm{T}}x. \tag{3}$$

Furthermore, for all $c \in \mathbb{R}^m$, define

$$\mathcal{Y}(c) \triangleq \{y \in \mathbb{R}^n \colon A^{\mathrm{T}}y = c\}, \tag{4}$$

and, for all $y \in \mathcal{Y}(c)$ and $b \in \mathbb{R}^n$, define the *dual cost* $g\colon \mathcal{Y}(c) \times \mathbb{R}^n \to \mathbb{R}$ by

$$g(y, b) \triangleq b^{\mathrm{T}}y. \tag{5}$$

For $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^m$, the sets $\mathcal{X}(b)$ and $\mathcal{Y}(c)$ may be empty. Assuming that these sets are not empty, the following result shows that the primal and dual costs are equal. This result holds whether or not $\mathcal{X}(b)$ and $\mathcal{Y}(c)$ have unique elements.

Proposition 1. Let $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^m$, and assume that $\mathcal{X}(b)$ and $\mathcal{Y}(c)$ are not empty. Then, for all $x \in \mathcal{X}(b)$ and all $y \in \mathcal{Y}(c)$,

$$f(x, c) = g(y, b). \tag{6}$$

Proof. Note that

$$f(x, c) = x^{\mathrm{T}}c = x^{\mathrm{T}}A^{\mathrm{T}}y = (Ax)^{\mathrm{T}}y = b^{\mathrm{T}}y = g(y, b). \qquad \square$$

Note that $f(x,c)$ depends on b due to the dependence of x on b; in many applications, it is useful to determine this sensitivity. To do this, assume that A is either a nonsingular square matrix or a wide matrix with full row rank, which implies that, for all $b \in \mathbb{R}^n$, $Ax = b$ has at least one solution and, for all $c \in \mathbb{R}^m$, $A^{\mathrm{T}}y = c$ has at most one solution. Now, let $c \in \mathcal{R}(A^{\mathrm{T}})$, $b \in \mathbb{R}^n$, and $x \in \mathcal{X}(b)$. Furthermore, let $\delta b \in \mathbb{R}^n$ and let $\delta x \in \mathbb{R}^m$ satisfy $x + \delta x \in \mathcal{X}(b + \delta b)$; hence, $A\delta x = \delta b$. Letting $y \stackrel{\triangle}{=} (AA^{\mathrm{T}})^{-1}Ac$ denote the unique solution of $A^{\mathrm{T}}y = c$, it follows that the sensitivity $\delta f(x,c)$ of f due to δb is given by

$$
\begin{aligned}
\delta f(x,c) &\stackrel{\triangle}{=} f(x+\delta x, c) - f(x,c) \\
&= f(\delta x, c) \\
&= f(\delta x, A^{\mathrm{T}}y) \\
&= (A^{\mathrm{T}}y)^{\mathrm{T}}\delta x \\
&= y^{\mathrm{T}}A\delta x \\
&= y^{\mathrm{T}}\delta b \\
&= (\delta b)^{\mathrm{T}}y \\
&= g(y, \delta b). \qquad (7)
\end{aligned}
$$

It thus follows that, for all $b \in \mathbb{R}^n$, $x \in \mathcal{X}(b)$, $\delta b \in \mathbb{R}^n$, and $c \in \mathcal{R}(A^{\mathrm{T}})$,

$$\delta f(x,c) = g(y, \delta b), \qquad (8)$$

where $y \stackrel{\triangle}{=} (AA^{\mathrm{T}})^{-1}Ac$ is the unique solution of $A^{\mathrm{T}}y = c$. In the case where A is wide and thus not square, x is an arbitrary solution of $Ax = b$, which has infinitely many solutions.

The Adjoint of a Bounded Linear Operator

The $n \times m$ real matrix A defines the linear function f that maps $\mathbb{R}^m$ to $\mathbb{R}^n$ defined by $f(x) = Ax$. We equip $\mathbb{R}^k$ with the inner product

$$\langle x, y\rangle_{\ell_2^k} \stackrel{\triangle}{=} y^{\mathrm{T}}x, \qquad (9)$$

where the subscript ℓ_2^k is analogous with the notation used below for the inner product on the function space L_2. For all $x \in \mathbb{R}^k$, the corresponding norm on $\mathbb{R}^k$ is the Euclidean norm

$$\|x\|_{\ell_2^k} \stackrel{\triangle}{=} \langle x, x\rangle_{\ell_2^k}^{1/2} = \sqrt{x^{\mathrm{T}}x}. \qquad (10)$$

The adjoint of A is the unique $m \times n$ matrix A' that satisfies $\langle Ax, y\rangle_{\ell_2^n} = \langle x, A'y\rangle_{\ell_2^m}$ for all $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. Since

$$\begin{aligned}\langle Ax, y\rangle_{\ell_2^n} &= y^{\mathrm{T}} Ax \\ &= (Ax)^{\mathrm{T}} y \\ &= x^{\mathrm{T}} A^{\mathrm{T}} y \\ &= \langle x, A^{\mathrm{T}} y\rangle_{\ell_2^m},\end{aligned} \tag{11}$$

it follows that A' is the transpose $A^{\mathrm{T}} \in \mathbb{R}^{m\times n}$ of A.

The notion of the adjoint of a matrix can be extended to bounded linear operators between Hilbert spaces. A *Hilbert space* $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is a vector space over the field of real or complex numbers with an inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and with the property that the *normed vector space* $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ arising from $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is *complete* in the sense that every Cauchy sequence in $\mathcal{H}$ converges with respect to the norm $\|\cdot\|_{\mathcal{H}}$ induced by the inner product to an element of $\mathcal{H}$. More details can be found in [3, p. 112]. This article is restricted to the case of a real Hilbert space, where the field is $\mathbb{R}$. In particular, for all $x \in \mathcal{H}$,

$$\|x\|_{\mathcal{H}} \triangleq \sqrt{\langle x, x\rangle_{\mathcal{H}}}. \tag{12}$$

Real Hilbert spaces have several useful features. Perhaps the most useful feature is the notion of orthogonality, where the elements x and y of $\mathcal{H}$ are *orthogonal* if $\langle x, y\rangle_{\mathcal{H}} = 0$. Consequently, a Hilbert space can be viewed as a natural generalization of Euclidean space. Second, if $f\colon \mathcal{H} \to \mathbb{R}$ is a bounded linear functional, then there exists a unique element $y_f \in \mathcal{H}$ such that $f(x) = \langle x, y_f\rangle_{\mathcal{H}}$; this is the Riesz representation theorem [3, p. 345]. As discussed below, this theorem is the key to proving the existence of adjoints of bounded linear operators between Hilbert spaces. First, however, we explain the meaning of bounded linear operators and bounded linear functionals.

A linear functional $f\colon \mathcal{H} \to \mathbb{R}$ is *bounded* if there exists $M \geq 0$ such that, for all $x \in \mathcal{H}$,

$$|f(x)| \leq M\|x\|_{\mathcal{H}}. \tag{13}$$

This property is equivalent to saying that f is bounded on bounded subsets. It is also equivalent to continuity: If the sequence $(x_i)_{i=1}^{\infty}$ converges to 0 in $\mathcal{H}$, then $(f(x_i))_{i=1}^{\infty}$ converges to 0. Similarly, the linear operator $\mathcal{A}\colon \mathcal{H}_1 \to \mathcal{H}_2$ between Hilbert spaces $(\mathcal{H}_1, \langle\cdot,\cdot\rangle_{\mathcal{H}_1})$ and $(\mathcal{H}_2, \langle\cdot,\cdot\rangle_{\mathcal{H}_2})$ is *bounded* if there exists $M \geq 0$ such that, for all $x \in \mathcal{H}_1$,

$$\|\mathcal{A}x\|_{\mathcal{H}_2} \leq M\|x\|_{\mathcal{H}_1}. \tag{14}$$

This property is equivalent to saying that $\mathcal{A}$ is bounded on bounded subsets. It is also equivalent to continuity: If the sequence $(x_i)_{i=1}^{\infty}$ converges to 0 in $\mathcal{H}_1$, then $(\mathcal{A}x_i)_{i=1}^{\infty}$ converges to 0 in $\mathcal{H}_2$.

Let $\mathcal{A}\colon \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator. For $y \in \mathcal{H}_2$, the linear functional $f_y\colon x \mapsto \langle \mathcal{A}x, y\rangle_{\mathcal{H}_2}$ is bounded because $\mathcal{A}$ is bounded, and, by the Riesz representation theorem, there exists a unique vector $\mathcal{A}^*y \in \mathcal{H}_1$ such that, for all $x \in \mathcal{H}_1$, $f_y(x) = \langle x, \mathcal{A}^*y\rangle_{\mathcal{H}_1}$. Furthermore, the linearity of $\mathcal{A}$ implies the linearity of the mapping $y \mapsto \mathcal{A}^*y$, and the continuity (boundedness) of $\mathcal{A}$ implies the continuity (boundedness) of $\mathcal{A}^*$. Therefore, for every bounded linear operator $\mathcal{A}\colon \mathcal{H}_1 \to \mathcal{H}_2$ there exists a unique bounded linear operator $\mathcal{A}^*\colon \mathcal{H}_2 \to \mathcal{H}_1$ such that, for all $(x, y) \in \mathcal{H}_1 \times \mathcal{H}_2$,

$$\langle \mathcal{A}x, y\rangle_{\mathcal{H}_2} = \langle x, \mathcal{A}^*y\rangle_{\mathcal{H}_1}. \tag{15}$$

The bounded linear operator $\mathcal{A}\colon \mathcal{H} \to \mathcal{H}$, where $\mathcal{H}$ is a Hilbert space, is *self adjoint* if $\mathcal{A} = \mathcal{A}^*$. As discussed in [2], [3], self-adjoint operators, which are analogous to symmetric matrices, have many convenient properties.

An extremely useful Hilbert space is the vector space $\mathrm{L}_2^m[t_0, t_\mathrm{f}]$ of Lebesgue-measurable, square-integrable functions mapping $[t_0, t_\mathrm{f}]$ to $\mathbb{R}^m$, where $t_0 < t_\mathrm{f}$, equipped with the inner product

$$\langle f, g\rangle_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]} \triangleq \int_{t_0}^{t_\mathrm{f}} g(t)^\mathrm{T} f(t)\,\mathrm{d}t. \tag{16}$$

We write $\mathrm{L}_2[t_0, t_\mathrm{f}]$ in the case where $m = 1$. In fact, for finite t_0 and t_f, a Fourier series can be viewed as a basis-function expansion of an element of the Hilbert space $\mathrm{L}_2[t_0, t_\mathrm{f}]$. As in the case of $\mathbb{R}^m$, the norm induced by the inner product $\langle\cdot,\cdot\rangle_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]}$ is defined by

$$\|f\|_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]} \triangleq \langle f, f\rangle^{1/2}_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]}. \tag{17}$$

Note that

$$\|f\|_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]} = \left(\int_{t_0}^{t_\mathrm{f}} f(t)^\mathrm{T} f(t)\,\mathrm{d}t\right)^{1/2} = \left(\int_{t_0}^{t_\mathrm{f}} \|f(t)\|^2_{\ell_2^m}\,\mathrm{d}t\right)^{1/2}. \tag{18}$$

For example, let $K\colon [t_0, t_\mathrm{f}] \times [t_0, t_\mathrm{f}] \to \mathbb{R}$ be a Lebesgue-measurable function that satisfies

$$\int_{t_0}^{t_\mathrm{f}}\int_{t_0}^{t_\mathrm{f}} |K(t,\tau)|^2\,\mathrm{d}t\,\mathrm{d}\tau < \infty. \tag{19}$$

Now, let $u \in \mathrm{L}_2[t_0, t_\mathrm{f}]$, and define the function $\mathcal{A}u\colon [t_0, t_\mathrm{f}] \to \mathbb{R}$ by

$$(\mathcal{A}u)(t) \triangleq \int_{t_0}^{t_\mathrm{f}} K(t,\tau)u(\tau)\,\mathrm{d}\tau. \tag{20}$$

Then [13, Corollary 3.4.6] implies that $\mathcal{A}u$ is a Lebesgue-measurable function on $[t_0, t_\mathrm{f}]$. Furthermore, for all $u, v \in \mathrm{L}_2[t_0, t_\mathrm{f}]$,

$$\begin{aligned}
\langle \mathcal{A}u, v\rangle_{\mathrm{L}_2[t_0,t_\mathrm{f}]} &= \int_{t_0}^{t_\mathrm{f}} v(t) \int_{t_0}^{t_\mathrm{f}} K(t,\tau)u(\tau)\,\mathrm{d}\tau\mathrm{d}t \\
&= \int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t_\mathrm{f}} v(t)K(t,\tau)u(\tau)\,\mathrm{d}\tau\mathrm{d}t \\
&= \int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t_\mathrm{f}} u(\tau)K(t,\tau)v(t)\,\mathrm{d}\tau\mathrm{d}t \\
&= \int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t_\mathrm{f}} u(\tau)K(t,\tau)v(t)\,\mathrm{d}t\mathrm{d}\tau \\
&= \int_{t_0}^{t_\mathrm{f}} u(\tau) \int_{t_0}^{t_\mathrm{f}} K(t,\tau)v(t)\,\mathrm{d}t\mathrm{d}\tau \\
&= \langle u, \mathcal{A}^*v\rangle_{\mathrm{L}_2[t_0,t_\mathrm{f}]},
\end{aligned} \tag{21}$$

where the adjoint $\mathcal{A}^*$ of $\mathcal{A}$ is the linear operator given by

$$(\mathcal{A}^*v)(\tau) = \int_{t_0}^{t_\mathrm{f}} K(t,\tau)v(t)\,\mathrm{d}t. \tag{22}$$

It can be seen that $\mathcal{A}$ is self adjoint if and only if, for almost all $t, \tau \in [t_0, t_\mathrm{f}]$, $K(t,\tau) = K(\tau, t)$. The fact that $\mathcal{A}\colon \mathrm{L}_2[t_0, t_\mathrm{f}] \to \mathrm{L}_2[t_0, t_\mathrm{f}]$ is a bounded linear operator follows from the more general case considered below for vector-valued functions.

More generally, let $\sigma_{\max}$ denote the maximum singular value, and let $K\colon [t_0, t_\mathrm{f}] \times [t_0, t_\mathrm{f}] \to \mathbb{R}^{l\times m}$. Assume that K is Lebesgue measurable and that

$$M \triangleq \left(\int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t_\mathrm{f}} \sigma_{\max}^2(K(t,\tau))\,\mathrm{d}t\,\mathrm{d}\tau\right)^{1/2} < \infty, \tag{23}$$

Since $\mathbb{R}^{l\times m}$ is finite dimensional, all norms on this space are equivalent, and thus (23) is equivalent to

$$\left(\int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t_\mathrm{f}} \|K(t,\tau)\|_\mathrm{F}^2\,\mathrm{d}t\,\mathrm{d}\tau\right)^{1/2} < \infty, \tag{24}$$

where $\|\cdot\|_\mathrm{F}$ is the Frobenius norm. Therefore, (23) is equivalent to the condition that every entry of K is square integrable on $[t_0, t_\mathrm{f}] \times [t_0, t_\mathrm{f}]$.

Next, for all $u \in \mathrm{L}_2^m[t_0, t_\mathrm{f}]$, define the function $\mathcal{A}u\colon [t_0, t_\mathrm{f}] \to \mathbb{R}^l$ by

$$(\mathcal{A}u)(t) = \int_{t_0}^{t_\mathrm{f}} K(t,\tau)u(\tau)\,\mathrm{d}\tau. \tag{25}$$

Using the fact that, for all $A \in \mathbb{R}^{l \times m}$ and all $x \in \mathbb{R}^m$, $\|Ax\|_{\ell_2^l} \leq \sigma_{\max}(A)\|x\|_{\ell_2^m}$, it follows that, for all $t \in [t_0, t_\mathrm{f}]$,

$$\begin{aligned}
\|(\mathcal{A}u)(t)\|_{\ell_2^l} &= \left\| \int_{t_0}^{t_\mathrm{f}} K(t,\tau)u(\tau)\,\mathrm{d}\tau \right\|_{\ell_2^l} \\
&\leq \int_{t_0}^{t_\mathrm{f}} \|K(t,\tau)u(\tau)\|_{\ell_2^l}\,\mathrm{d}\tau \\
&\leq \int_{t_0}^{t_\mathrm{f}} \sigma_{\max}(K(t,\tau))\,\|u(\tau)\|_{\ell_2^m}\,\mathrm{d}\tau \\
&\leq \left(\int_{t_0}^{t_\mathrm{f}} \sigma_{\max}^2(K(t,\tau))\,\mathrm{d}\tau \right)^{1/2} \|u\|_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]}, \qquad (26)
\end{aligned}$$

where the third inequality follows from the Cauchy-Schwarz inequality in $\mathrm{L}_2[t_0, t_\mathrm{f}]$ [2, pp. 9,10]. Now, squaring and integrating yields

$$\begin{aligned}
\int_{t_0}^{t_\mathrm{f}} \|(\mathcal{A}u)(t)\|_{\ell_2^l}^2\,\mathrm{d}t &\leq \left(\int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t_\mathrm{f}} \sigma_{\max}^2(K(t,\tau)\,\mathrm{d}\tau\,dt \right) \|u\|_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]}^2 \\
&\leq M^2 \|u\|_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]}^2. \qquad (27)
\end{aligned}$$

Therefore, for all $u \in \mathrm{L}_2^m[t_0, t_\mathrm{f}]$, it follows that $\mathcal{A}u \in \mathrm{L}_2^l[t_0, t_\mathrm{f}]$ and $\|\mathcal{A}u\|_{\mathrm{L}_2^l[t_0,t_\mathrm{f}]} \leq M\|u\|_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]}$. Hence, $\mathcal{A}\colon \mathrm{L}_2^m[t_0, t_\mathrm{f}] \to \mathrm{L}_2^l[t_0, t_\mathrm{f}]$ is a bounded linear operator.

To determine the adjoint $\mathcal{A}^*$ of $\mathcal{A}$, let $u \in \mathrm{L}_2^m[t_0, t_\mathrm{f}]$ and $v \in \mathrm{L}_2^l[t_0, t_\mathrm{f}]$. It thus follows that

$$\begin{aligned}
\langle \mathcal{A}u, v \rangle_{\mathrm{L}_2^l[t_0,t_\mathrm{f}]} &= \int_{t_0}^{t_\mathrm{f}} v(t)^\mathrm{T} (\mathcal{A}u)(t)\,\mathrm{d}t \\
&= \int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t_\mathrm{f}} v(t)^\mathrm{T} K(t,\tau) u(\tau)\,\mathrm{d}\tau\,\mathrm{d}t \\
&= \int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t_\mathrm{f}} [K^\mathrm{T}(t,\tau) v(t)]^\mathrm{T} u(\tau)\,\mathrm{d}\tau\,\mathrm{d}t.
\end{aligned}$$

Therefore, $\mathcal{A}^* v$ is given by

$$(\mathcal{A}^* v)(\tau) = \int_{t_0}^{t_\mathrm{f}} K^\mathrm{T}(t,\tau) v(t)\,\mathrm{d}t. \qquad (28)$$

The Adjoint of a Linear Time-Invariant System

Consider the linear time-invariant system

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad (29)$$

$$x(t_0) = x_0, \qquad (30)$$

$$y(t) = Cx(t) + Du(t), \qquad (31)$$

where, for all $t \in [t_0, t_\mathrm{f}]$, $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input, $y(t) \in \mathbb{R}^l$ is the output, and A, B, C, D are real matrices of appropriate size. It thus follows that, for all $t \in [t_0, t_\mathrm{f}]$,

$$y(t) = Ce^{(t-t_0)A}x(t_0) + \int_{t_0}^{t} Ce^{(t-\tau)A}Bu(\tau)\,\mathrm{d}\tau + Du(t). \tag{32}$$

In the case where $x_0 = 0$ and $D = 0$, it follows that

$$y(t) = \int_{t_0}^{t} Ce^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau, \tag{33}$$

which can be written as

$$y(t) = \int_{t_0}^{t_\mathrm{f}} K(t,\tau)u(\tau)\,\mathrm{d}\tau, \tag{34}$$

where

$$K(t,\tau) \stackrel{\triangle}{=} \mathbb{I}_{[t_0,t]}(\tau)Ce^{(t-\tau)A}B, \tag{35}$$

and where $\mathbb{I}_X(x) = 1$ if $x \in X$ and $\mathbb{I}_X(x) = 0$ otherwise.

For all $u \in \mathrm{L}_2^m[t_0, t_\mathrm{f}]$, (34) can be written as

$$y = \mathcal{A}u, \tag{36}$$

where $\mathcal{A}\colon \mathrm{L}_2^m[t_0, t_\mathrm{f}] \to \mathrm{L}_2^l[t_0, t_\mathrm{f}]$ is a bounded linear operator. Using (28) it thus follows that, for all $v \in \mathrm{L}_2^l[t_0, t_\mathrm{f}]$,

$$(\mathcal{A}^*v)(\tau) = \int_{t_0}^{t_\mathrm{f}} \mathbb{I}_{[\tau,t_\mathrm{f}]}(t)B^\mathrm{T}e^{(t-\tau)A^\mathrm{T}}C^\mathrm{T}v(t)\,\mathrm{d}t \tag{37}$$

$$= \int_{\tau}^{t_\mathrm{f}} B^\mathrm{T}e^{(t-\tau)A^\mathrm{T}}C^\mathrm{T}v(t)\,\mathrm{d}t. \tag{38}$$

Finally, interchanging the roles of t and τ yields the adjoint operator

$$(\mathcal{A}^*v)(t) = \int_{t}^{t_\mathrm{f}} B^\mathrm{T}e^{(\tau-t)A^\mathrm{T}}C^\mathrm{T}v(\tau)\,\mathrm{d}\tau. \tag{39}$$

It can be seen that (39) differs from (33) in two key ways. First, the integration in (33) is from t_0 to t, whereas, in (39), the integration is from t to t_f. Furthermore, the matrices A, B, C in (33) are replaced, respectively, by $A^\mathrm{T}, C^\mathrm{T}, B^\mathrm{T}$ in (39). The first distinction shows that the adjoint operator operates in reverse time, while the second distinction shows that properties of A, B are replaced by properties of $A^\mathrm{T}, C^\mathrm{T}$.

If l and m are different, then $K(t,\tau)$ is rectangular, and thus $\mathcal{A}$ cannot be self adjoint. In the case where $l = m$, it follows from the fact that $K(t,\tau) = 0$ for all $\tau > t$ that $\mathcal{A}$ is not self adjoint.

The case where x_0 and D are not necessarily zero is considered in the next section, where the linear time-invariant dynamics (29), (31) are replaced by linear time-varying dynamics.

The Adjoint of a Linear Time-Varying System

For $t \in [t_0, t_\mathrm{f}]$, consider the linear, time-varying system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \tag{40}$$

$$x(t_0) = x_0, \tag{41}$$

$$y(t) = C(t)x(t) + D(t)u(t), \tag{42}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^l$, $A(t) \in \mathbb{R}^{n\times n}$, $B(t) \in \mathbb{R}^{n\times m}$, $C(t) \in \mathbb{R}^{l\times n}$, $D(t) \in \mathbb{R}^{l\times m}$, and $x_0 \in \mathbb{R}^n$. We assume for simplicity that A, B, C, D, and u are continuous functions of time. The solution to these equations can be expressed in terms of the state transition matrix $\Phi(t,\tau)$, which generalizes the exponential function $\exp((t-\tau)A)$ to the case where A is a function of t. Details on the state transition matrix are given in "Some Facts on the State Transition Matrix." For the following result, see [1, p. 82] and [14, pp. 46–49].

Theorem 1. Equations (40), (41) have a unique continuously differentiable solution on $[t_0, t_\mathrm{f}]$. In particular, for all $t \in [t_0, t_\mathrm{f}]$,

$$x(t) = \Phi(t,t_0)x_0 + \int_{t_0}^{t} \Phi(t,\tau)B(\tau)u(\tau)\,\mathrm{d}\tau \tag{43}$$

and

$$y(t) = C(t)\Phi(t,t_0)x_0 + \int_{t_0}^{t} C(t)\Phi(t,\tau)B(\tau)u(\tau)\,\mathrm{d}\tau + D(t)u(t). \tag{44}$$

Since each pair $(x_0, u) \in \mathbb{R}^n \times \mathrm{L}_2^m[t_0, t_\mathrm{f}]$ gives rise to an output $y \in \mathrm{L}_2^l[t_0, t_\mathrm{f}]$, we define the Hilbert space $\widetilde{\mathcal{H}} \triangleq \mathbb{R}^n \times \mathrm{L}_2^m[t_0, t_\mathrm{f}]$ with the inner product

$$\langle (a,u),(b,v)\rangle_{\widetilde{\mathcal{H}}} \triangleq b^\mathrm{T}a + \int_{t_0}^{t_\mathrm{f}} v(t)^\mathrm{T}u(t)\,\mathrm{d}t. \tag{45}$$

Completeness of $\widetilde{\mathcal{H}}$ follows from the fact that Cauchy sequences in $\mathcal{H}$ are couples of Cauchy sequences in $\mathbb{R}^n$ and $\mathrm{L}_2^m[t_0, t_\mathrm{f}]$, which converge in their respective spaces. Therefore, (44) can be represented by the linear operator $\mathcal{A}\colon \widetilde{\mathcal{H}} \to \mathrm{L}_2^l[t_0, t_\mathrm{f}]$ defined by

$$(\mathcal{A}(x_0,u))(t) \triangleq C(t)\Phi(t,t_0)x_0 + \int_{t_0}^{t} C(t)\Phi(t,\tau)B(\tau)u(\tau)\,\mathrm{d}\tau + D(t)u(t). \tag{46}$$

It can be seen that $\mathcal{A}$ has two parts, where one is defined by an integral and the other is the multiplication operator $(x_0,u) \mapsto C\Phi_{t_0}x_0 + Du$, where, for all $t \in [t_0, t_\mathrm{f}]$,

$$\Phi_\tau(t) \triangleq \Phi(t,\tau). \tag{47}$$

As shown in earlier, the continuity of B, C, D, and Φ implies that $\mathcal{A}\colon \widetilde{\mathcal{H}} \to \mathrm{L}_2^l[t_0, t_\mathrm{f}]$ is bounded.

Next, to determine $\mathcal{A}^*\colon \mathrm{L}_2^l[t_0,t_\mathrm{f}] \to \widetilde{\mathcal{H}}$, define, for all $t \in [t_0,t_\mathrm{f}]$,

$$\Gamma(t) \triangleq \int_{t_0}^{t} C(t)\Phi(t,\tau)B(\tau)u(\tau)\,\mathrm{d}\tau, \tag{48}$$

It thus follows that, for all $(x_0,u) \in \widetilde{\mathcal{H}}$ and $v \in \mathrm{L}_2^l[t_0,t_\mathrm{f}]$,

$$\begin{aligned}
\langle \mathcal{A}(x_0,u), v\rangle_{\mathrm{L}_2^l[t_0,t_\mathrm{f}]} &= \langle C\Phi_{t_0}x_0, v\rangle_{\mathrm{L}_2^l[t_0,t_\mathrm{f}]} + \langle Du, v\rangle_{\mathrm{L}_2^l[t_0,t_\mathrm{f}]} + \langle \Gamma, v\rangle_{\mathrm{L}_2^l[t_0,t_\mathrm{f}]} \\
&= \langle C\Phi_{t_0}x_0, v\rangle_{\mathrm{L}_2^l[t_0,t_\mathrm{f}]} + \langle Du, v\rangle_{\mathrm{L}_2^l[t_0,t_\mathrm{f}]} \\
&\quad + \int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t} \langle C(t)\Phi(t,\tau)B(\tau)u(\tau), v(t)\rangle_{\ell_2^l}\,\mathrm{d}\tau\mathrm{d}t \\
&= \langle x_0, \Phi_{t_0}^\mathrm{T} C^\mathrm{T} v\rangle_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]} + \langle u, D^\mathrm{T} v\rangle_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]} \\
&\quad + \int_{t_0}^{t_\mathrm{f}} \int_{t_0}^{t} \langle u(\tau), B(\tau)^\mathrm{T}\Phi(t,\tau)^\mathrm{T} C(t)^\mathrm{T} v(t)\rangle_{\ell_2^m}\,\mathrm{d}\tau\mathrm{d}t \\
&= \langle x_0, \Phi_{t_0}^\mathrm{T} C^\mathrm{T} v\rangle_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]} + \langle u, D^\mathrm{T} v\rangle_{\mathrm{L}_2^m[t_0,t_\mathrm{f}]} \\
&\quad + \int_{t_0}^{t_\mathrm{f}} \left\langle u(\tau), \int_{\tau}^{t_\mathrm{f}} B(\tau)^\mathrm{T}\Phi(t,\tau)^\mathrm{T} C(t)^\mathrm{T} v(t)\mathrm{d}t \right\rangle_{\ell_2^m} \mathrm{d}\tau \\
&= \langle x_0, a(v)\rangle_{\ell_2^m} + \int_{t_0}^{t_\mathrm{f}} \left\langle u(\tau), (\widetilde{\mathcal{A}}v)(\tau)\right\rangle_{\ell_2^m} \mathrm{d}\tau,
\end{aligned} \tag{49}$$

where $a\colon \mathrm{L}_2^l[t_0,t_\mathrm{f}] \to \mathbb{R}^m$ is defined by

$$a(v) \triangleq \int_{t_0}^{t_\mathrm{f}} \Phi(\tau,t_0)^\mathrm{T} C(\tau)^\mathrm{T} v(\tau)\,\mathrm{d}\tau \tag{50}$$

and $\widetilde{\mathcal{A}}\colon \mathrm{L}_2^l[t_0,t_\mathrm{f}] \to \mathrm{L}_2^m[t_0,t_\mathrm{f}]$ is defined by

$$(\widetilde{\mathcal{A}}v)(t) \triangleq D(t)^\mathrm{T} v(t) + \int_{t}^{t_\mathrm{f}} B(t)^\mathrm{T}\Phi(\tau,t)^\mathrm{T} C(\tau)^\mathrm{T} v(\tau)\,\mathrm{d}\tau. \tag{51}$$

Hence $\mathcal{A}^* v = (a(v), \widetilde{\mathcal{A}}v) \in \widetilde{\mathcal{H}}$.

Next, note that (51) can be written as

$$\begin{aligned}
(\widetilde{\mathcal{A}}v)(t) &= D(t)^\mathrm{T} v(t) + \int_{t_0}^{t_\mathrm{f}} B(t)^\mathrm{T}\Phi(\tau,t)^\mathrm{T} C(\tau)^\mathrm{T} v(\tau)\,\mathrm{d}\tau \\
&\quad - \int_{t_0}^{t} B(t)^\mathrm{T}\Phi(\tau,t)^\mathrm{T} C(\tau)^\mathrm{T} v(\tau)\,\mathrm{d}\tau.
\end{aligned} \tag{52}$$

Using the fact that $\Phi(\tau,t)^\mathrm{T} = \Phi(t_0,t)^\mathrm{T}\Phi(\tau,t_0)^\mathrm{T}$, it follows from (52) that

$$\begin{aligned}
\int_{t_0}^{t_\mathrm{f}} B(t)^\mathrm{T}\Phi(\tau,t)^\mathrm{T} C(\tau)^\mathrm{T} v(\tau)\,\mathrm{d}\tau &= B(t)^\mathrm{T}\Phi(t_0,t)^\mathrm{T} \int_{t_0}^{t_\mathrm{f}} \Phi(\tau,t_0)^\mathrm{T} C(\tau)^\mathrm{T} v(\tau)\,\mathrm{d}\tau \\
&= B(t)^\mathrm{T}\Phi(t_0,t)^\mathrm{T} a(v).
\end{aligned} \tag{53}$$

Hence

$$(\widetilde{\mathcal{A}}v)(t) = B(t)^{\mathrm{T}}\Phi(t_0,t)^{\mathrm{T}}a(v) + \int_{t_0}^{t} B(t)^{\mathrm{T}}\Phi(\tau,t)^{\mathrm{T}}(-C(\tau)^{\mathrm{T}})v(\tau)\,\mathrm{d}\tau + D(t)^{\mathrm{T}}v(t). \tag{54}$$

Letting Ψ denote the state transition matrix of $\dot{x} = -A(t)^{\mathrm{T}}x$, it follows from Corollary S1 in "Some Facts on the State Transition Matrix" that

$$\begin{aligned}(\widetilde{\mathcal{A}}v)(t) = {}& B(t)^{\mathrm{T}}\Psi(t,t_0)^{\mathrm{T}}a(v)\\ &+ \int_{t_0}^{t} B(t)^{\mathrm{T}}\Psi(t,\tau)(-C(\tau)^{\mathrm{T}})v(\tau)\,\mathrm{d}\tau + D(t)^{\mathrm{T}}v(t).\end{aligned} \tag{55}$$

Comparing (55) with (44), it follows that $z(t) \triangleq (\widetilde{\mathcal{A}}v)(t)$ is the output of the linear time-varying system

$$\dot{p}(t) = -A(t)^{\mathrm{T}}p(t) - C(t)^{\mathrm{T}}v(t), \tag{56}$$

$$p(t_0) = a(v), \tag{57}$$

$$z(t) = B(t)^{\mathrm{T}}p(t) + D(t)^{\mathrm{T}}v(t), \tag{58}$$

where $a(v)$ is defined by (50). Using Corollary S1, it follows that

$$p(t) = \int_{t}^{t_{\mathrm{f}}} \Phi(\tau,t)^{\mathrm{T}}C(\tau)^{\mathrm{T}}v(\tau)\,\mathrm{d}\tau, \tag{59}$$

which satisfies the initial condition (57) as well as the final condition $p(t_{\mathrm{f}}) = 0$. Consequently, (57) can be replaced by the $p(t_{\mathrm{f}}) = 0$. It thus follows that $\mathcal{A}^*v = (a(v), \widetilde{\mathcal{A}}v) = (p(t_0), z)$, where p and z are, respectively, the state and output of the linear time-varying system

$$\dot{p}(t) = -A(t)^{\mathrm{T}}p(t) - C(t)^{\mathrm{T}}v(t), \tag{60}$$

$$p(t_{\mathrm{f}}) = 0, \tag{61}$$

$$z(t) = B(t)^{\mathrm{T}}p(t) + D(t)^{\mathrm{T}}v(t). \tag{62}$$

We thus have the following duality result, which provides a restatement of the adjoint identity

$$\langle \mathcal{A}(x_0,u), v\rangle_{\mathrm{L}_2^l[t_0,t_{\mathrm{f}}]} = \langle (x_0,u), \mathcal{A}^*v\rangle_{\widetilde{H}}. \tag{63}$$

Theorem 2. Consider the linear time-varying system (40)–(42), where $u \in \mathrm{L}_2^m[t_0,t_{\mathrm{f}}]$, and the corresponding adjoint system (60)–(62), where $v \in \mathrm{L}_2^l[t_0,t_{\mathrm{f}}]$. Then

$$\langle y, v\rangle_{\widetilde{\mathcal{H}}} = \langle x(t_0), p(t_0)\rangle_{\ell_2^n} + \langle u, z\rangle_{\mathrm{L}_2^m[t_0,t_{\mathrm{f}}]}. \tag{64}$$

Time Reversal for the Adjoint System

The dynamics of the adjoint system (60)-(62) involve a final condition rather than an initial condition. In addition, if A constant and asymptotically stable, then $-A^{\mathrm{T}}$ is unstable. Both of these concerns can be addressed by reversing the direction of time. In particular, for the continuous function $t \mapsto W(t)$ defined on $[t_0, t_{\mathrm{f}}]$, let $\overleftarrow{W}$ denote the function $t \mapsto W(t_0 + t_{\mathrm{f}} - t)$. With this notation, let s and r denote, respectively, the state and output of the linear time-varying system

$$\dot{s}(t) = \overleftarrow{A}(t)^{\mathrm{T}} s(t) + \overleftarrow{C}(t)^{\mathrm{T}} \overleftarrow{v}(t), \tag{65}$$

$$s(t_0) = 0, \tag{66}$$

$$r(t) = \overleftarrow{B}(t)^{\mathrm{T}} s(t) + \overleftarrow{D}(t)^{\mathrm{T}} \overleftarrow{v}(t). \tag{67}$$

Then $p = \overleftarrow{s}$ and $z = \overleftarrow{r}$, and the duality expressed in (64) takes the form

$$\langle y, v \rangle = \langle x(t_0), s(t_{\mathrm{f}}) \rangle_{\ell_2^n} + \langle u, \overleftarrow{r} \rangle_{\mathrm{L}_2^m[t_0, t_{\mathrm{f}}]}. \tag{68}$$

Conclusions and Extensions

The goal of this article was to derive the adjoint of a linear time-varying system as the adjoint of a linear operator between Hilbert spaces. This objective was motivated by the fact that the adjoint of a linear system plays a fundamental role in control and estimation theory as demonstrated in [1]. In particular, the adjoint of a linear system explains the duality between controllability and observability, which is a bedrock principle of systems theory.

It was shown that the linear operator defined by a linear time-varying system is bounded—and thus continuous—but not self adjoint, even in the case of a square linear time-invariant system. An interesting research topic would be to investigate the implications of the lack of self adjointness in terms of the spectral properties of the operator.

The development in the present paper was confined to linear time-varying systems on a finite interval. In practice, however, the infinite-horizon case is important. This case is addressed in [1, pp. 188–194] through a limiting argument. An open problem is to directly address the infinite-horizon case by determining the adjoint of a linear system on an unbounded interval. This extension is challenging since it entails an unbounded linear operator whose domain is a dense, proper subspace of a Hilbert space.

Acknowledgments

The authors thank Dimitra Panagou for suggesting this topic and the reviewers for numerous helpful suggestions.

Sidebar 1: The Costate Equation of Optimal Control as a Partial Adjoint

An adjoint system appears in classical optimal control theory within the context of the minimum principle, where it is called the costate equation. The costate $p(t)$ appears in the Hamiltonian H defined by

$$H(x,u,p,t) \triangleq L(x,u) + p^{\mathrm{T}} f(x,u,t) \tag{S1}$$

for an optimal control problem with dynamics $\dot{x}(t) = f(x(t),u(t),t)$ and cost function $J(u) \triangleq \int_{t_0}^{t_\mathrm{f}} L(x(t),u(t),t)\,\mathrm{d}t$. The minimum principle states that $H(x(t),u(t),p(t),t)$ is minimized pointwise in time along the optimal state trajectory and control, where the costate p satisfies the partial adjoint equation $\dot{p}(t) = -A(t)^{\mathrm{T}}p(t)$, with the endpoint condition $p(t_\mathrm{f}) = p_\mathrm{f}$ in the case where the state trajectory of $\dot{x}(t) = f(x(t),u(t),t)$ has a specified initial condition. In the folklore of optimal control, p is sometimes viewed as a dynamic Lagrange multiplier, where the dynamics $\dot{x}(t) = f(x(t),u(t),t)$ are viewed as the constraint $\dot{x}(t) - f(x(t),u(t),t) = 0$ [10, p. 187].

Letting $\delta x(t)$ denote the solution to the linearized homogeneous dynamics $\delta\dot{x}(t) = A(t)\delta x(t)$, it follows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}[p(t)^{\mathrm{T}}\delta x(t)] &= \dot{p}(t)^{\mathrm{T}}x(t) + p(t)^{\mathrm{T}}\dot{x}(t)\\ &= [-A(t)^{\mathrm{T}}p(t)]^{\mathrm{T}}x(t) + p(t)^{\mathrm{T}}A(t)x(t)\\ &= -p(t)^{\mathrm{T}}A(t)x(t) + p(t)^{\mathrm{T}}A(t)x(t)\\ &= 0. \end{aligned}\tag{S2}$$

Hence, $p(t)^{\mathrm{T}}x(t)$ is constant, and thus $p(t) - p(0)$ and $x(t)$ are orthogonal along trajectories. The significance of this condition is explained in [15, p. 247] and [16, pp. 326, 327]. As shown below, the costate p provides a convenient representation of the state transition matrix.

Strictly speaking, the costate equation is not the adjoint of a linear system with an input and an output. We thus refer to the costate equation as a *partial adjoint*. The following result is stated for $l \geq 1$. In the case $l = 1$, the matrix P of the partial adjoint is the costate p.

Proposition S4. Let y be given by (40)–(42), and let $P\colon [t_0, t_\mathrm{f}] \to \mathbb{R}^{n\times l}$ satisfy

$$\dot{P}(t) = -A(t)^{\mathrm{T}}P(t), \tag{S3}$$

$$P(t_\mathrm{f}) = C^{\mathrm{T}}(t_\mathrm{f}). \tag{S4}$$

Then,

$$y(t_\mathrm{f}) = P(t_0)^{\mathrm{T}}x_0 + \int_{t_0}^{t_\mathrm{f}} P(\tau)^{\mathrm{T}}B(\tau)u(\tau)\,\mathrm{d}\tau + D(t_\mathrm{f})u(t_\mathrm{f}). \tag{S5}$$

Proof. It follows from (S3) and (40) that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}[P(t)^{\mathrm{T}}x(t)] &= \dot{P}(t)^{\mathrm{T}}x(t) + P(t)^{\mathrm{T}}\dot{x}(t)\\ &= -P(t)^{\mathrm{T}}A(t)x(t) + P(t)^{\mathrm{T}}[A(t)x(t) + B(t)u(t)]\\ &= P(t)^{\mathrm{T}}B(t)u(t).\end{aligned} \tag{S6}$$

Next, integrating (S6) from t_0 to t_{f} and using (S4), (41), and (42) with $t = t_{\mathrm{f}}$ yields

$$\begin{aligned}\int_{t_0}^{t_{\mathrm{f}}} P(t)^{\mathrm{T}}B(t)u(t)\,\mathrm{d}t &= P(t_{\mathrm{f}})^{\mathrm{T}}x(t_{\mathrm{f}}) - P(t_0)^{\mathrm{T}}x(t_0)\\ &= C(t_{\mathrm{f}})x(t_{\mathrm{f}}) - P(t_0)^{\mathrm{T}}x_0\\ &= y(t_{\mathrm{f}}) - D(t_{\mathrm{f}})u(t_{\mathrm{f}}) - P(t_0)^{\mathrm{T}}x_0,\end{aligned}$$

which is equivalent to (S5). □

It can be seen that (S5) is an alternative expression for $y(t_{\mathrm{f}})$ given by (44), that is,

$$y(t_{\mathrm{f}}) = C(t_{\mathrm{f}})\Phi(t_{\mathrm{f}}, t_0)x_0 + \int_{t_0}^{t_{\mathrm{f}}} C(t_{\mathrm{f}})\Phi(t_{\mathrm{f}}, \tau)B(\tau)u(\tau)\,\mathrm{d}\tau + D(t_{\mathrm{f}})u(t_{\mathrm{f}}). \tag{S7}$$

In order to obtain $\Phi(t_{\mathrm{f}}, \tau)$ to evaluate (S7), equation (S20) must be integrated from τ to t_{f} for each $\tau \in [t_0, t_{\mathrm{f}}]$. In contrast, (S5) requires only a single backward integration of P. To show the equivalence of (S5) and (S7), Corollary S1 in "Some Facts on the State Transition Matrix" implies that

$$P(t) = \Psi(t, t_{\mathrm{f}})C(t_{\mathrm{f}})^{\mathrm{T}} = \Phi(t_{\mathrm{f}}, t)^{\mathrm{T}}C(t_{\mathrm{f}})^{\mathrm{T}}, \tag{S8}$$

that is,

$$C(t_{\mathrm{f}})\Phi(t_{\mathrm{f}}, t) = C(t_{\mathrm{f}})\Psi(t, t_{\mathrm{f}})^{\mathrm{T}} = P(t)^{\mathrm{T}}. \tag{S9}$$

Replacing $C(t_{\mathrm{f}})\Phi(t_{\mathrm{f}}, t_0)$ and $C(t_{\mathrm{f}})\Phi(t_{\mathrm{f}}, \tau)$ in (S7) by $P(t_0)^{\mathrm{T}}$ and $P(\tau)^{\mathrm{T}}$, respectively, yields (S5).

In fact, P can be computed forward. To show this, define $Q(t) \triangleq \overleftarrow{P}(t) = P(t_0 + t_{\mathrm{f}} - t)$. Since $\dot{Q}(t) = -\dot{P}(t_0 + t_{\mathrm{f}} - t)$, it follows from (S3) that, for all $t \in [t_0, t_{\mathrm{f}}]$,

$$\dot{Q}(t) = A(t_0 + t_{\mathrm{f}} - t)^{\mathrm{T}}Q(t), \tag{S10}$$

$$Q(t_0) = C^{\mathrm{T}}(t_{\mathrm{f}}). \tag{S11}$$

Note that (S10), (S11) can be integrated forward in time. However, for online implementation, future values of $A(t_0 + t_{\mathrm{f}} - t)$ must be known.

Finally, denoting (40), (42) by the time-varying matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{S12}$$

it follows that the adjoint system can be represented by

$$\begin{bmatrix} -A^T & -C^T \\ B^T & D^T \end{bmatrix}. \tag{S13}$$

Then (S13) can be written as

$$\overleftarrow{\begin{bmatrix} -A^T & -C^T \\ B^T & D^T \end{bmatrix}}, \tag{S14}$$

whose upper left corner corresponds to the costate equation $\dot{p}(t) = -A(t)^{\mathrm{T}} p(t)$, thus justifying the name partial adjoint.

Sidebar 2: Existence and Uniqueness of Solutions to $Ax = b$ and $A^{\mathrm{T}}y = c$

For $A \in \mathbb{R}^{n\times m}$, let $A^{+} \in \mathbb{R}^{m\times n}$ denote the Moore-Penrose generalized inverse of A. Let "$\mathcal{R}$" denote range, and let "$\mathcal{N}$" denote null space.

Proposition S1. Let $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^{n}$, and $c \in \mathbb{R}^{m}$. Then the following statements hold:

- $i)$ If A is wide, then $Ax = b$ has either zero or infinitely many solutions. Both cases can occur.
- $ii)$ If A is square or tall, then $Ax = b$ has either zero, exactly one, or infinitely many solutions. All three cases can occur.
- $iii)$ The following statements are equivalent:
  - $a)$ $Ax = b$ has at least one solution.
  - $b)$ $b \in \mathcal{R}(A)$.
  - $c)$ $b = AA^{+}b$.
- $iv)$ The following statements are equivalent:
  - $a)$ $A^{\mathrm{T}}y = c$ has at least one solution.
  - $b)$ $c \in \mathcal{R}(A^{\mathrm{T}})$.
  - $c)$ $c = A^{+}Ac$.
- $v)$ $Ax = b$ has at most one solution if and only if $\mathcal{N}(A) = \{0\}$.
- $vi)$ $A^{\mathrm{T}}y = c$ has at most one solution if and only if $\mathcal{N}(A^{\mathrm{T}}) = \{0\}$.
- $vii)$ If $Ax = b$ has at least one solution, then all solutions are given by $A^{+}b + \mathcal{N}(A)$.
- $viii)$ If $A^{\mathrm{T}}y = c$ has at least one solution, then all solutions are given by $A^{+\mathrm{T}}b + \mathcal{N}(A^{\mathrm{T}})$.

As an alternative proof of Proposition 1, assume that $\mathcal{X}(b)$ and $\mathcal{Y}(c)$ are not empty. Since $b \in \mathcal{R}(A)$ and $c \in \mathcal{R}(A^{\mathrm{T}})$, it follows from $b = AA^{+}b$, $c = A^{+}Ac = A^{\mathrm{T}}A^{+\mathrm{T}}c$ given by [S1, Proposition 8.1.7] that $A^{+}b \in \mathcal{X}(b)$ and $A^{+\mathrm{T}}c \in \mathcal{Y}(c)$. Therefore, [S1, Proposition 3.7.5] implies that $\mathcal{X}(b)$ and $\mathcal{Y}(c)$ can be written as

$$\mathcal{X}(b) = A^{+}b + \mathcal{N}(A), \tag{S15}$$

$$\mathcal{Y}(c) = A^{+\mathrm{T}}c + \mathcal{N}(A^{\mathrm{T}}). \tag{S16}$$

Now, let $x \in \mathcal{X}(b)$ and $y \in \mathcal{Y}(c)$. Since, by [S1, Proposition 8.1.7], $\mathcal{N}(A) = \mathcal{R}(I - A^{+}A)$ and $\mathcal{N}(A^{\mathrm{T}}) = \mathcal{R}(I - AA^{+})$, it follows from (S15) and (S16) that there exist $\overline{x} \in \mathbb{R}^m$ and $\overline{y} \in \mathbb{R}^n$ such that $x = A^{+}b + (I - A^{+}A)\overline{x}$ and $y = A^{+\mathrm{T}}c + (I - AA^{+})\overline{y}$. Combining these facts yields

$$\begin{aligned}
f(x, c) &= c^{\mathrm{T}}x \\
&= c^{\mathrm{T}}[A^{+}b + (I - A^{+}A)\overline{x}] \\
&= c^{\mathrm{T}}A^{+}b + c^{\mathrm{T}}(I - A^{+}A)\overline{x} \\
&= c^{\mathrm{T}}A^{+}b + \overline{x}^{\mathrm{T}}(I - A^{+}A)c \\
&= c^{\mathrm{T}}A^{+}b \\
&= c^{\mathrm{T}}A^{+}b + \overline{y}^{\mathrm{T}}(I - AA^{+})b \\
&= b^{\mathrm{T}}A^{+\mathrm{T}}c + b^{\mathrm{T}}(I - AA^{+})\overline{y} \\
&= b^{\mathrm{T}}[A^{+\mathrm{T}}c + (I - AA^{+})\overline{y}] \\
&= b^{\mathrm{T}}y \\
&= g(y, b).
\end{aligned}$$

[S1] D. S. Bernstein, *Scalar, Vector, and Matrix Mathematics,* revised and expanded edition, Princeton University Press, 2018.

Sidebar 3: Some Facts on the State Transition Matrix

For $t \in [t_0, t_\mathrm{f}]$, consider the linear time-varying system

$$\dot{x}(t) = A(t)x(t), \tag{S17}$$

$$x(t_0) = x_0, \tag{S18}$$

where $A\colon [t_0, t_\mathrm{f}] \to \mathbb{R}^{n\times n}$ is continuous. The solution of (S17), (S18) is given by

$$x(t) = \Phi(t, t_0)x_0, \tag{S19}$$

where Φ is the state transition matrix given by the following result.

Proposition S2. Let $A\colon [t_0, t_\mathrm{f}] \to \mathbb{R}^{n\times n}$ be continuous. Then there exists a unique continuously differentiable function $\Phi\colon [t_0, t_\mathrm{f}] \times [t_0, t_\mathrm{f}] \to \mathbb{R}^{n\times n}$ such that, for all $t, \tau \in [t_0, t_\mathrm{f}]$, the following conditions are satisfied:

$$\frac{\partial}{\partial t}\Phi(t, \tau) = A(t)\Phi(t, \tau), \tag{S20}$$

$$\Phi(\tau, \tau) = I_n. \tag{S21}$$

Furthermore, for all $t, s, \tau \in [t_0, t_\mathrm{f}]$, the following conditions are satisfied:

$$\det \Phi(t, \tau) \neq 0, \tag{S22}$$

$$\Phi(t, \tau)^{-1} = \Phi(\tau, t), \tag{S23}$$

$$\Phi(t, s)\Phi(s, \tau) = \Phi(t, \tau). \tag{S24}$$

The following result gives conditions under which the state transition matrix can be written as a matrix exponential despite the fact that A is not constant. A proof is given in [14].

Proposition S3. Let $A\colon [t_0, t_\mathrm{f}] \to \mathbb{R}^{n\times n}$ be continuous, let Φ be given by Proposition S2, and consider the following conditions:

i) For all $t, \tau \in [t_0, t_\mathrm{f}]$, $A(t) = A(\tau)$.

ii) For all $t, \tau \in [t_0, t_\mathrm{f}]$, $A(t)A(\tau) = A(\tau)A(t)$.

iii) For all $t, \tau \in [t_0, t_\mathrm{f}]$, $A(t)\int_{t_0}^{\tau} A(s)\,\mathrm{d}s = \int_{t_0}^{\tau} A(s)\,\mathrm{d}s A(t)$.

iv) For all $t, \tau \in [t_0, t_\mathrm{f}]$, $\Phi(t, t_0) = e^{\int_{t_0}^{t} A(\tau)\,\mathrm{d}\tau}$.

Then, *i)* $\Longrightarrow$ *ii)* $\Longrightarrow$ *iii)* $\Longrightarrow$ *iv)*.

Development of the adjoint operator requires the following corollary of Proposition S2.

Corollary S1. Let $A(t)$ and $\Phi(t,\tau)$ be given by Proposition S2, and, for all $t,\tau \in [t_0, t_{\rm f}]$, define $\Psi(t,\tau) \stackrel{\triangle}{=} \Phi(\tau,t)^{\rm T}$. Then, for all $t,\tau \in [t_0, t_{\rm f}]$, $\Psi(t,\tau)$ is the state transition matrix of $\dot{x}(t) = -A(t)^{\rm T}x(t)$.

Proof. It follows from (S23) that $\Phi(t,\tau)\Phi(\tau,t) = I_n$. Differentiating this equation with respect to t yields

$$\left[\frac{\partial}{\partial t}\Phi(t,\tau)\right]\Phi(\tau,t) = -\Phi(t,\tau)\,\frac{\partial}{\partial t}\Phi(\tau,t).$$

Using (S20) yields

$$A(t)\Phi(t,\tau)\,\Phi(\tau,t) = -\Phi(t,\tau)\,\frac{\partial}{\partial t}\Phi(\tau,t),$$

and thus (S23) and (S24) imply that

$$\begin{aligned}\frac{\partial}{\partial t}\Phi(\tau,t) &= -\Phi(t,\tau)^{-1}A(t)\\ &= -\Phi(\tau,t)A(t).\end{aligned}$$

Therefore,

$$\frac{\partial}{\partial t}\Psi(t,\tau) = -A(t)^{\rm T}\Psi(t,\tau).$$

Finally, for all $\tau \in [t_0, t_{\rm f}]$, $\Psi(\tau,\tau) = \Phi(\tau,\tau)^{\rm T} = I_n$. Hence, for all $t,\tau \in [t_0, t_{\rm f}]$, the state transition matrix of $\dot{x}(t) = -A(t)^{\rm T}x(t)$ is given by $\Psi(t,\tau)$. $\square$

References

[1] M. Green and D. J. N. Limebeer, *Linear Robust Control*. Prentice Hall, 1994.

[2] G. Helmberg, *Introduction to Spectral Theory in Hilbert Space*. North-Holland, 1969.

[3] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. Springer, 1982.

[4] R. F. Curtain and H. J. Zwart, *An Introduction to Infinite-Dimensional Linear Systems Theory*. Springer, 1995.

[5] D. G. Luenberger, *Optimization by Vector Space Methods*. Wiley, 1969.

[6] R. L. Raffard, K. Amonlirdviman, J. D. Axelrod, and C. J. Tomlin, "An Adjoint-Based Parameter Identification Algorithm Applied to Planar Cell Polarity Signaling," *IEEE Trans. Autom. Contr.*, vol. 53, p. 109–121, 2008.

[7] M. B. Giles and N. A. Pierce, "An Introduction to the Adjoint Approach to Design," *Flow, Turbulence and Combustion*, vol. 65, pp. 393–415, 2000.

[8] R. M. Errico, "What Is an Adjoint Model?" *Bull. Amer. Meteor. Soc.*, vol. 78, no. 11, pp. 2577–2592, 1997.

[9] G. I. Marchuk, *Adjoint Equations and Analysis of Complex Systems*. Springer, 1994.

[10] D. E. Kirk, *Optimal Control Theory: An Introduction*. Prentice-Hall, 1970.

[11] J. Warga, *Optimal Control of Differential and Functional Equations*. Academic Press, 1972.

[12] H. D'Angelo, *Linear Time-Varying Systems: Analysis and Synthesis*. Allyn and Bacon, 1970.

[13] V. I. Bogachev, *Measure Theory, Volume I*. Springer-Verlag Berlin Heidelberg, 2007.

[14] W. J. Rugh, *Linear System Theory*, 2nd ed. Prentice Hall, 1996.

[15] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*. Wiley, 1967.

[16] T. L. Vincent and W. J. Grantham, *Nonlinear and Optimal Control Systems*. Wiley, 1997.

Author Biographies

Omran Kouba received the Sc.B. degree in Pure Mathematics from the University of Paris XI and the Ph.D. degree in Functional Analysis from Pierre and Marie Curie University in Paris, France. Currently he is a professor in the Department of Mathematics in the Higher Institute of Applied Sciences and Technology, Damascus (Syria). His interests are in real and complex analysis, inequalities, and problem solving.

Dennis S. Bernstein received the Sc.B. degree from Brown University and the Ph.D. degree from the University of Michigan, where he is currently a faculty member in the Aerospace Engineering Department. His interests are in identification, estimation, and control for aerospace applications. He is the author of *Scalar, Vector, and Matrix Mathematics*, published by Princeton University Press.