SIAM REVIEW
Vol. 64, No. 4, pp. 1031–1061

A Generalized Dual Transform: Linear Algebra and Geometry of (Pseudo)Inverting a Matrix*

L. P. Withers, Jr.†

Abstract. A new theory of a generalized dual transform of a list of vectors is presented, emphasizing its geometry as well as its linear algebra. We review the theory for independent vectors and find a new butterfly identity. This leads us to a parallel process to compute the dual vectors. Next we take the general case of arbitrary vectors, which may be linearly dependent. We start with two very symmetric axioms for the general dual transform. We show how the dual transform is closely related to the pseudoinverse of a matrix, defined by the Moore and Penrose axioms. We find that each dual vector is a sort of contrast of its corresponding original vector against the background of the rest of the vectors. To get the dual vector, we operate on the original vector by either an orthogonal projector or a protractor defined in terms of the rest of the vectors. Which operator is used depends on whether or not the original vector is linearly independent of the rest. When we update the protractor to include a new vector, a stereographic projection appears unexpectedly. Basic identities for dual vectors are proved, which result in closed-form algebraic processes to compute the dual list. The processes transform any list of vectors into its dual list, regardless of dependencies among the primal vectors. Linear-algebraic recipes are developed for a complete, reversible Gram–Schmidt process, a vector version of the Greville process, and a simple parallel butterfly process to get the dual list.

Key words. dual transform, generalized biorthogonality, pseudoinverse, orthogonal projector, protractor, stereographic projection, complete Gram–Schmidt process, Greville process, parallel butterfly process, Levinson process

MSC codes. 15A03, 15A09, 15A06

DOI. 10.1137/19M1270410

Contents

*Received by the editors June 24, 2019; accepted for publication (in revised form) January 25, 2022; published electronically November 3, 2022.
https://doi.org/10.1137/19M1270410
†School of Physics and Astronomy, George Mason University, Fairfax, VA 22030 USA (langwithers@gmail.com).

1. Introduction. Linear algebra contains many geometric ideas that carry over from Euclidean space $\mathbb{R}^3$ to any complete linear space $\mathbf{V}$ with an inner product $\langle\cdot,\cdot\rangle$. One such idea is orthogonality. We say that two vectors a and b in $\mathbf{V}$ are *orthogonal* if $\langle a, b\rangle = 0$, and we write this as $a \perp b$ to suggest intuitively that the vectors are perpendicular to each other. The norm $\|a\| = \sqrt{\langle a, a\rangle}$ measures the length of a vector a. Then the Pythagorean theorem is true in this linear space: if $a \perp b$, then $\|a + b\|^2 = \|a\|^2 + \|b\|^2$. (See the books [1], [2], [3] for linear space axioms.)

Orthogonality makes a list of several vectors easier to work with. It disentangles them by making the inner product zero for every pair of distinct vectors. Consider a list of linearly independent vectors $a_1, a_2, \ldots, a_n$ that span a subspace $\mathbf{A}$ in $\mathbf{V}$. (Such a list is called a basis of $\mathbf{A}$.) To orthogonalize them, we use the classic Gram–Schmidt process (we cover this in detail in the next section). It transforms the original list into an orthogonal list, $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n$, such that $\bar{a}_j \perp \bar{a}_k$ for every pair with $j \neq k$, and $\langle \bar{a}_k, \bar{a}_k\rangle = \|\bar{a}_k\|^2 = 1$.

The new list spans the same subspace $\mathbf{A}$. This means that any vector b in $\mathbf{A}$ is equal to a linear combination of the vectors, that is, $b = \sum_{k=1}^{n} \alpha_k \bar{a}_k$ for some scalar combining coefficients α_k. Now, using orthogonality, we can take the inner product of each $\bar{a}_j$ with both sides of the equality and easily find that $\alpha_j = \langle \bar{a}_j, b\rangle$. Orthogonality gives us the simple formula

$$b = \sum_{k=1}^{n} \langle \bar{a}_k, b\rangle \bar{a}_k. \tag{1.1}$$

How could we express any vector b in $\mathbf{A}$ as a linear combination of the original, unorthogonalized vectors? It might sound like a messy job, but it is again surprisingly easy, this time using orthogonality in a new way. We transform the original list into a *biorthogonal* list, $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n$, such that $\hat{a}_j \perp a_k$ for every pair with $j \neq k$, and $\langle \hat{a}_k, a_k\rangle = 1$. The new list still spans the subspace $\mathbf{A}$. We call it the *dual* list. It complements the original (primal) list, rather than replacing it.

Since b is any vector in $\mathbf{A}$, we can expand it in terms of the original basis as $b = \sum_{k=1}^{n} \alpha_k a_k$ for some scalar combining coefficients α_k. What are the coefficients? Using biorthogonality, we can take the inner product of each dual vector $\hat{a}_j$ with both sides of the equality and easily find that $\alpha_j = \langle \hat{a}_j, b\rangle$. Biorthogonality gives us a second simple formula,

$$b = \sum_{k=1}^{n} \langle \hat{a}_k, b\rangle a_k. \tag{1.2}$$

This paper is about how to compute the dual list for any given list of vectors. It turns out that the Gram–Schmidt method produces orthogonal vectors that belong to successive dual (biorthogonal) lists. To handle the general case of independent or dependent vectors, we build the theory of the dual on two axioms. General purpose versions of the dual transform methods result. Worked examples and figures are given to illustrate the theory and methods.

Because the dual list is an abstract linear-algebraic entity, it has a wide spectrum of applications and it enables algebraic computing methods. Applications include (pseudo)inverting a matrix, solving linear least squares problems, (bi)orthogonalizing polynomials [1], [2], [3], and operating on frames [4]. The dual is also used in crystallography [5], beamforming for antenna arrays [6], [7], and general relativity [8], [9].

Here is the outline of the rest of this paper. We begin with the important special case of independent vectors in section 2 and find a new butterfly identity and parallel linear-space algorithm for the dual transform. In section 3, we give linear-space axioms for the general dual and show their equivalence to the axioms of Moore and Penrose for the pseudoinverse of a matrix. Two kinds of linear operators, orthogonal projectors P and protractors Q, are our tools for developing the theory of the dual list in section 4. A highlight here is finding a hidden stereographic projection when we update the generalized biorthogonality coefficients. This geometric insight yields new facts about their analysis. Linear-space recipes to compute the generalized dual are provided in section 5, with two appendices, including a general (and reversible) Gram–Schmidt process, a vector version of the Greville process, and a general parallel butterfly process. As an example, we show that the Greville process implies a complete Levinson recursion to solve a linear prediction problem.

2. The Dual Transform of Independent Vectors.

Let's begin with the dual transform of a list or sequence of independent vectors, $a_1, a_2, \ldots, a_n$ in $\mathbf{V}$, an inner product space with a scalar field consisting of the complex numbers $\mathbb{C}$ [1], [3]. An example is the space $\mathbb{C}^m$, with the complex-valued inner product of any two vectors $c, d \in \mathbb{C}^m$ given by $\langle c, d\rangle = \sum_{k=1}^{m} c_k^* d_k$. Here z^* denotes the complex conjugate of $z \in \mathbb{C}$. Everything in this paper reduces in an obvious manner for scalars in the field of real numbers $\mathbb{R}$.

Let $\mathbf{A} = \operatorname{span}\{a_1, \ldots, a_n\}$ be the subspace of $\mathbf{V}$ spanned by all the linear combinations of our list of independent vectors. We define the *dual transform* of the given list as the complementary list $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n \in \mathbf{A}$ that has the *biorthogonal* relations

$$\langle \hat{a}_j, a_k\rangle = \delta_{jk} \tag{2.1}$$

for $j, k = 1, \ldots, n$, where δ_{jk} is 1 when $j = k$, and 0 otherwise [1], [3].

There is an interesting geometric way to construct the dual list, using orthogonal complements and orthogonal projections. Let's review these concepts. The *orthogonal complement* of a set $\mathbf{A}$ in $\mathbf{V}$ is defined as the set $\mathbf{A}^\perp = \{v \in \mathbf{V} : v \perp a \text{ for all } a \in \mathbf{A}\}$. If $\mathbf{A}$ is a linear subspace of $\mathbf{V}$, then $\mathbf{A}^\perp$ is a subspace too, and we have $\mathbf{A} \cap \mathbf{A}^\perp = \{0\}$. An example we can visualize is in $\mathbb{R}^3$. Choose a vector $v \neq 0$, and let $\mathbf{A}$ be the one-dimensional subspace $\operatorname{span}\{v\}$. Then $\mathbf{A}^\perp$ is the plane of vectors orthogonal to its normal vector v and passing through the origin (0), a two-dimensional subspace.

Given any vector v in a linear space $\mathbf{V}$ of dimension 3 or higher, there are lots of right triangles with one leg a in $\mathbf{A}$ and the other leg w perpendicular to a, and with hypotenuse (sum of the legs) equal to v. That is, we can decompose v as $v = a + w$, $w \perp a$, in many ways. But we have not required $w \perp \mathbf{A}$, that is, $w \in \mathbf{A}^\perp$. A remarkable fact is that we can decompose every vector v as $v = a + w$ for unique $a \in \mathbf{A}, w \in \mathbf{A}^\perp$. Only one right triangle with one leg in $\mathbf{A}$ and one leg in $\mathbf{A}^\perp$ can be built on a given hypotenuse v. To prove this, we use the orthogonal basis $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n$ of the subspace $\mathbf{A}$ from formula (1.1) and verify that the unique $a \in \mathbf{A}$ and $w \in \mathbf{A}^\perp$

that sum to v are actually given by (see [1])

$$a = \sum_{k=1}^{n} \langle \bar{a}_k, v \rangle \bar{a}_k, \quad w = v - \sum_{k=1}^{n} \langle \bar{a}_k, v \rangle \bar{a}_k. \tag{2.2}$$

Now we are ready to define the *orthogonal projector* $P_{\mathbf{A}}$ by (see [3])

$$P_{\mathbf{A}}(a + w) = a \quad \text{for all } a \in \mathbf{A}, \text{ for all } w \in \mathbf{A}^{\perp}. \tag{2.3}$$

$P_{\mathbf{A}}$ is a linear operator that projects vectors orthogonally onto a subspace $\mathbf{A}$ of $\mathbf{V}$. Clearly it comes as a pair with the projector onto the orthogonal complement, $P_{\mathbf{A}^{\perp}} = I - P_{\mathbf{A}}$, where I is the identity operator. For we have $P_{\mathbf{A}^{\perp}}(a+w) = w = (a+w) - a = (I - P_{\mathbf{A}})(a + w)$.

An orthogonal projector is characterized by two properties. First, it is disposable or idempotent in that after using it once, you can throw it away; using it again has no effect: $P_{\mathbf{A}}^2 = P_{\mathbf{A}}$. This is true because $P_{\mathbf{A}}(P_{\mathbf{A}}(a + w)) = P_{\mathbf{A}}(a) = a = P_{\mathbf{A}}(a + w)$.

The second property is that it is self-adjoint or Hermitian: $P_{\mathbf{A}}^* = P_{\mathbf{A}}$. The adjoint O^* of a linear operator O is defined in terms of exchanging their roles inside inner products, such that $\langle O^* x, y\rangle = \langle x, Oy\rangle$ for all $x, y \in \mathbf{V}$. A projector $P_{\mathbf{A}}$ is its own adjoint, because for any $v = a + w$, $x = a' + w'$, we have $\langle P_{\mathbf{A}} x, v\rangle = \langle a', a + w\rangle = \langle a', a\rangle = \langle a' + w', a\rangle = \langle x, P_{\mathbf{A}} v\rangle$. For proofs of the converse, that the two properties imply that P is an orthogonal projector, see [1], [2], [3].

Now back to the dual transform. We get a little technical and consider subspaces of $\mathbf{A}$ spanned by all but one of the independent vectors $a_1, a_2, \ldots, a_n$. We leave out the jth vector and define $\mathbf{A}_{(j)} = \operatorname{span}\{a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_n\}$ for $j = 1, \ldots, n$. Then, as promised, here is the geometric formula for the dual transform $\{\hat{a}_j\}_{j=1}^n$ of the vectors $\{a_j\}_{j=1}^n$ (see [3]):

$$\boxed{\hat{a}_j = \frac{P_{\mathbf{A}_{(j)}^{\perp}} a_j}{\|P_{\mathbf{A}_{(j)}^{\perp}} a_j\|^2}.} \tag{2.4}$$

Geometrically, each dual vector is just the component of the original vector a_j normal to the subspace spanned by the rest of the vectors, with a scale factor. Figure 1 depicts this for three vectors in $\mathbb{R}^3$. The dual vector $\hat{a}_1$ is the projection $p_1 = P_{\operatorname{span}\{a_2,a_3\}^{\perp}} a_1$, scaled by dividing by $\|p_1\|^2$. Scaling the projections this way, we can demonstrate that they fulfill the biorthogonal relations (2.1),

$$\langle \hat{a}_j, a_k\rangle = \langle P_{\mathbf{A}_{(j)}^{\perp}} a_j / \|P_{\mathbf{A}_{(j)}^{\perp}} a_j\|^2, a_k\rangle = \langle a_j, P_{\mathbf{A}_{(j)}^{\perp}} a_k\rangle / \, \|P_{\mathbf{A}_{(j)}^{\perp}} a_j\|^2 = \delta_{jk},$$

since $P_{\mathbf{A}_{(j)}^{\perp}} a_k = 0$ for $j \neq k$, and otherwise, for $j = k$, $\langle a_j, P_{\mathbf{A}_{(j)}^{\perp}} a_k\rangle = \langle a_j, P^2_{\mathbf{A}_{(j)}^{\perp}} a_j\rangle = \langle P_{\mathbf{A}_{(j)}^{\perp}} a_j, P_{\mathbf{A}_{(j)}^{\perp}} a_j\rangle = \|P_{\mathbf{A}_{(j)}^{\perp}} a_j\|^2$.

First we will show that the forward and inverse transforms are the same. We use the fact that biorthogonality is a unique symmetric relation.

LEMMA 1. *Let $a_1, \ldots, a_n$ be linearly independent vectors that span a subspace $\mathbf{A} \subseteq \mathbf{V}$. If, for $j, k = 1, \ldots, n$, b_j and c_k are in the subspace $\mathbf{A}$ and there are biorthogonal relations $\langle a_k, b_j\rangle = \delta_{kj}$, $\langle b_j, c_k\rangle = \delta_{jk}$, then $c_k = a_k$.*

Proof. Since $\{a_i\}_{i=1}^n$ is a basis of $\mathbf{A}$, for each k we have $c_k = \sum_{i=1}^n \alpha_{ik} a_i$ for some scalars α_{ik}. Then $\delta_{jk} = \langle b_j, c_k\rangle = \sum_{i=1}^n \alpha_{ik}\langle b_j, a_i\rangle = \alpha_{jk}$. Therefore $c_k = a_k$. ∎

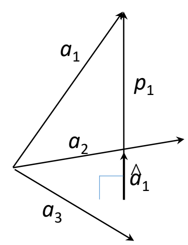


Fig. 1 *Geometry of the dual transform. For $n = 3$ independent vectors a_1, a_2, a_3 in $\mathbb{R}^3$, one dual vector $\hat{a}_1$ is shown. It is the component p_1 of a_1 orthogonal to the subspace span$\{a_2, a_3\}$, divided by $\|p_1\|^2$. The other two dual vectors, $\hat{a}_2$ and $\hat{a}_3$, are constructed similarly.*

This implies that the dual transform is reflexive.

Theorem 2. *For $k = 1, \ldots, n$, $a_k = \hat{\hat{a}}_k$.*

Proof. By definition the dual vectors $\hat{a}_k \in \mathbf{A} = \text{span}\{a_k\}_{k=1}^n$ and satisfy the biorthogonal relations (2.1). Similarly, $\hat{\hat{a}}_k \in \text{span}\{\hat{a}_k\}_{k=1}^n$, so $\hat{\hat{a}}_k \in \text{span}\{a_k\}_{k=1}^n$ and $\langle \hat{a}_j, \hat{\hat{a}}_k \rangle = \delta_{jk}$. The conclusion follows by Lemma 1. ∎

We also have $\text{span}\{\hat{a}_k\}_{k=1}^n = \text{span}\{a_k\}_{k=1}^n$, since $\text{span}\{a_k\} = \text{span}\{\hat{\hat{a}}_k\} \subseteq \text{span}\{\hat{a}_k\}$. Thus the $\hat{a}_k$ are independent and may be called the dual or reciprocal basis of the subspace $\mathbf{A}$ with basis $\{a_k\}_{k=1}^n$.

We have defined a concrete dual basis, one that occupies the same subspace $\mathbf{A}$ as the original basis. A more abstract dual basis lives in the dual space $\hat{\mathbf{A}}$ of all linear mappings α of vectors in $\mathbf{A}$ into scalars. Then the dual basis for $\hat{\mathbf{A}}$ is defined by the mappings $\hat{\alpha}_k(a_j) = \delta_{jk}$ (biorthogonal relations) and $\hat{\hat{\mathbf{A}}}$ is isomorphic to $\mathbf{A}$ [1].

For independent vectors $a_1, \ldots, a_n \in \mathbb{C}^m, m \geq n$, the dual transform can be thought of as a matrix transform instead of a vector transform. It maps the column matrix $A = [a_1 \vdots a_2 \vdots \cdots \vdots a_n] \mapsto \hat{A} = [\hat{a}_1 \vdots \hat{a}_2 \vdots \cdots \vdots \hat{a}_n]$. The conjugated dual vectors turn out to be the rows of the Moore–Penrose pseudoinverse A^+; that is, $A^+ = (\hat{A})^*$. We will cover the relationship of the dual transform to matrix inversion in general in section 3.

We now turn to the relationship of the dual transform to Gram–Schmidt orthogonalization.

Gram–Schmidt Orthogonality. The basic principle of orthogonality in the Gram–Schmidt process is expressed by an eponymous identity for orthogonal projectors [10]. From this identity, we can derive the Gram–Schmidt process and show that the orthogonal vectors it generates as its list of given vectors expands are actually dual vectors. Other interesting facts about dual vectors of a set of independent vectors also come from the Gram–Schmidt identity, leading to a parallel butterfly process for computing the whole dual transform.

Theorem 3 (Gram–Schmidt identity). *For the orthogonal projectors as defined above,*

$$P_{\mathbf{A}^\perp} = P_{\mathbf{A}_{(j)}^\perp} - P_{P_{\mathbf{A}_{(j)}^\perp} a_j}. \tag{2.5}$$

Proof. For any subspace $\mathbf{H}$ of $\mathbf{V}$ decomposed into orthogonal complements $\mathbf{M}$ and $\mathbf{N}$, $P_{\mathbf{H}} = P_{\mathbf{M}} + P_{\mathbf{N}}$ [3]. A special case of this is an identity to inflate the subspace $\mathbf{H}$ of a projector by a vector g,

$$P_{\text{span}\{\mathbf{H},g\}} = P_{\mathbf{H}+\text{span}\{P_{\mathbf{H}^\perp} g\}} = P_{\mathbf{H}} + P_{P_{\mathbf{H}^\perp} g}.$$

Subtracting each side from the identity operator I gives another version,

$$P_{\text{span}\{\mathbf{H},g\}^\perp} = P_{\mathbf{H}^\perp} - P_{P_{\mathbf{H}^\perp} g}.$$

Now the projector on the left-hand side and the projector subtracted on the right-hand side project onto orthogonal subspaces, since the identity implies $P_{\text{span}\{\mathbf{H},g\}^\perp}(P_{\mathbf{H}^\perp} g) = 0$. The identity (2.5) follows when we put $\mathbf{H} = \mathbf{A}_{(j)}$ and $g = a_j$. ∎

The right-hand term of the Gram–Schmidt identity (2.5) appears a bit convoluted, but it is only a projector onto the span of a single vector (namely, the projection $p_j = P_{\mathbf{A}_{(j)}^\perp} a_j$). What makes the Gram–Schmidt identity so valuable is that this is a projector we know how to compute! The formula for the orthogonal projection of a vector a onto the span of one vector $v \neq 0$ is

$$P_v a = \frac{\langle v, a\rangle}{\|v\|^2} v. \tag{2.6}$$

This formula is easy to derive, for we have $P_v a = \alpha v$ for some scalar α and $a - \alpha v \perp v$.

How do we orthogonalize a list of several independent vectors $a_1, a_2, \ldots, a_n$? An immediate answer in the language of orthogonal projectors is to write down the list

$$a_1,\ P_{\text{span}\{a_1\}^\perp} a_2,\ P_{\text{span}\{a_1,a_2\}^\perp} a_3,\ \ldots,\ P_{\text{span}\{a_1,\ldots,a_{n-1}\}^\perp} a_n. \tag{2.7}$$

Note that this is an orthogonal list, because each successive projection is orthogonal to all of its predecessors. Let us label the projections as $p_k = P_{\mathbf{A}_{k-1}^\perp} a_k$, where the subspaces are denoted by $\mathbf{A}_k = \text{span}\{a_1, \ldots, a_k\}$, $k = 1, \ldots, n$. Now we can rescale this list (2.7) by dividing each vector p_k by its squared norm $\|p_k\|^2$. Then, by our earlier geometric formula (2.4), $p_k/\|p_k\|^2 = \hat{a}_k^k$, the last dual vector in the dual list of $a_1, \ldots, a_k$. The rescaled orthogonal list (2.7) is a sequence of dual vectors,

$$\hat{a}_1^1,\ \hat{a}_2^2,\ \ldots,\ \hat{a}_n^n. \tag{2.8}$$

The dual vectors $\hat{a}_k^k$ are an orthogonal basis for the subspace $\mathbf{A}$ spanned by the a_k's. An easy induction shows that $\mathbf{A}_k = \text{span}\{a_1, \ldots, a_k\} = \text{span}\{\hat{a}_1^1, \ldots, \hat{a}_k^k\}$. Initially (for $k = 1$) we have $\hat{a}_1^1 = a_1/\|a_1\|^2$, whose span equals that of a_1. By hypothesis, for any $k-1$, $\mathbf{A}_{k-1}$ can be expressed as either kind of span. Because the projection $p_k = P_{\mathbf{A}_{k-1}^\perp} a_k = (I - P_{\mathbf{A}_{k-1}}) a_k = a_k - P_{\mathbf{A}_{k-1}} a_k$, we evidently have $\hat{a}_k^k \propto p_k \in \text{span}\{a_1, \ldots, a_k\}$ and $a_k \in \text{span}\{\hat{a}_1^1, \ldots, \hat{a}_k^k\}$. So the two spans contain each other, and must be equal.

How can we compute the orthogonal list (2.8)? This is where the Gram–Schmidt identity (2.5) goes to work. Before we rescale them as dual vectors $\hat{a}_k^k$, the projections p_k in this list (2.8) are $p_k = P_{\mathbf{A}_{k-1}^\perp} a_k = a_k - P_{\mathbf{A}_{k-1}} a_k$. The projector $P_{\mathbf{A}_{k-1}}$ can be expressed by the formula

$$P_{\mathbf{A}_{k-1}} = \sum_{j=1}^{k-1} \langle \hat{a}_j^j, \cdot\rangle p_j. \tag{2.9}$$

We can prove (2.9) by induction. For $k = 1$, we have $P_\emptyset = 0$. Assume (2.9) is true for $k-1$. Then we must show it is true for the case k. Subtracting both sides of the Gram–Schmidt identity (2.5) from the identity operator I tells us that we can replace the projector $P_{\mathbf{A}_{k-1}}$ by $P_{\mathbf{A}_{k-2}} + P_{p_{k-1}}$. Also, from our formula (2.6) to project onto a single vector, we have $P_{p_{k-1}} = \langle p_{k-1}, \cdot\rangle p_{k-1}/\|p_{k-1}\|^2 = \langle \hat{a}_{k-1}^{k-1}, \cdot\rangle p_{k-1}$. This completes the induction.

Now we substitute (2.9) into $p_k = a_k - P_{\mathbf{A}_{k-1}} a_k$, to get the formula

$$p_k = a_k - \sum_{j=1}^{k-1} \langle \hat{a}_j^j, a_k\rangle p_j. \tag{2.10}$$

This gives us a way to compute the orthogonal list (2.8) from scratch: the Gram–Schmidt orthogonalization process. At each step, it projects a new vector to make it orthogonal to its predecessors [1], [2]. In fact, it computes one dual vector at each step, using the original linearly independent vectors $a_1, a_2, \ldots, a_n$:

Gram–Schmidt process

$\hat{a}_1^1 = a_1/\|a_1\|^2$
for $k = 2, \ldots, n$
    $p_k = a_k - \sum_{j=1}^{k-1} \tilde{\alpha}_{jk} p_j$, where $\tilde{\alpha}_{jk} = \langle \hat{a}_j^j, a_k\rangle$
    $\hat{a}_k^k = p_k/\|p_k\|^2$
end

In this version of the process, the projected components p_k are not scaled as usual by their inverse norms to make them unit vectors. Instead, they are scaled by their inverse squares as in formula (2.4), so that each $\hat{a}_k^k$ is the last dual vector in the dual transform of the first k vectors $a_1, \ldots, a_k$. This scaling convention has the virtue that it preserves information about the vector lengths, rather than erasing it as normalizing the vectors does. The Gram–Schmidt process becomes reversible. To recover the original vectors, one runs the process again for the dual vectors. (Taking the dual of the dual returns the original vector a_k at each step.)

Example 1. The Gram–Schmidt process can orthogonalize the list of polynomials $x^0, x^1, x^2, \ldots, x^n$, considered as functions of x defined over the closed interval $[-1, 1]$. We choose the inner product $\langle f, g\rangle = \int_{-1}^1 f(x)g(x)\,dx$ for integrable functions f, g. Then the Gram–Schmidt process produces the orthogonal polynomials $P_0(x), P_1(x), P_2(x), \ldots, P_n(x)$, named for Legendre. (Together with their derivatives, they are the main ingredients for constructing the spherical harmonics used in physics [11].) These are the first few Legendre polynomials, scaled as dual polynomials:

$$\begin{aligned}
\hat{a}_1^1 &= \tfrac{1}{2}\cdot 1 &&= \tfrac{1}{2}\cdot P_0(x),\\
\hat{a}_2^2 &= \tfrac{3}{2}\cdot x &&= \tfrac{3}{2}\cdot P_1(x),\\
\hat{a}_3^3 &= \tfrac{15}{4}\cdot \tfrac{1}{2}(3x^2-1) &&= \tfrac{15}{4}\cdot P_2(x),\\
\hat{a}_4^4 &= \tfrac{35}{4}\cdot \tfrac{1}{2}x(5x^2-3) &&= \tfrac{35}{4}\cdot P_3(x),\\
\hat{a}_5^5 &= \tfrac{315}{16}\cdot \tfrac{1}{8}(35x^4-30x^2+3) &&= \tfrac{315}{16}\cdot P_4(x).
\end{aligned}$$

Next we show that biorthogonal relations are fulfilled uniquely by the dual transform given by formula (2.4).

THEOREM 4. *Suppose* $\{a_j\}_{j=1}^n$ *span a subspace* $\mathbf{A}$ *of* $\mathbf{V}$. *If, for a vector* $b_k \in \mathbf{A}$, $\langle a_j, b_k\rangle = \delta_{jk}$ *for* $j = 1, \ldots, n$, *then* $b_k = \hat{a}_k$, *and* a_k *is independent of the other vectors* $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n$.

Proof. From the biorthogonal relations, $b_k \neq 0$ (otherwise it would nullify them) and $P_{\mathbf{A}_{(k)}} b_k = 0$. Since $b_k \in \mathbf{A}$, $P_{\mathbf{A}} b_k = b_k$. Substituting these two expressions in the Gram–Schmidt identity, we find $b_k = P_{P_{\mathbf{A}_{(k)}^\perp} a_k} b_k = \big(\frac{\langle p_k, b_k\rangle}{\langle p_k, p_k\rangle}\big) p_k$, where $p_k = P_{\mathbf{A}_{(k)}^\perp} a_k \neq 0$. This implies that a_k is independent of the other vectors, i.e., that it does not belong to $\mathbf{A}_{(k)}$. Since an orthogonal projector is self-adjoint, $\langle p_k, b_k\rangle = \langle P_{\mathbf{A}_{(k)}^\perp} a_k, b_k\rangle = \langle a_k, P_{\mathbf{A}_{(k)}^\perp} b_k\rangle = \langle a_k, b_k\rangle = 1$, so that $b_k = \big(\frac{1}{\langle p_k, p_k\rangle}\big) p_k = \hat{a}_k$. ∎

Another characterization of the dual vectors now follows.

COROLLARY 5. *For each k, the vector* b_k *with minimum norm* $\|b_k\|$, *which satisfies the biorthogonal constraints* $\langle b_k, a_j\rangle = \delta_{kj}$ *for* $j = 1, \ldots, n$, *is given uniquely by* $b_k = \hat{a}_k$.

Proof. The minimum is attained for some vector, since the norm is continuous over the closed subspace of vectors that satisfy the constraints. We will show that a vector of minimum norm has to be in the subspace $\mathbf{A}$. The conclusion will then follow by Theorem 4.

Let b be any vector that satisfies the biorthogonal constraints. We may write it as a sum $b = b^{\|} + b^{\perp}$, where $b^{\|} \in \mathbf{A}$ and $b^\perp \perp \mathbf{A}$. Then $\delta_{kj} = \langle b, a_j\rangle = \langle b^{\|}, a_j\rangle$, so $b^{\|}$ also satisfies the constraints. However (by the Pythagorean theorem), its norm is less if $b^\perp \neq 0$: $\|b\|^2 = \|b^{\|}\|^2 + \|b^\perp\|^2 > \|b^{\|}\|^2$. Therefore a vector with smallest norm must be in $\mathbf{A}$. ∎

To construct the dual list for a given list of several vectors, we must include new vectors one at a time. When we introduce a new vector a_k, this identity updates the dual list. We let $\hat{a}_{j(k)}$ denote the vector dual to a_j in the dual transform of $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n$.

THEOREM 6 (butterfly identity). *Let* a_j, a_k *each be independent of* $\{a_i\}_{i=1, i\neq j,k}^n$. *For* $j, k = 1, \ldots, n$, $j \neq k$,

$$\begin{aligned} \gamma \hat{a}_k &= \hat{a}_{k(j)} - \alpha_{kj}^* \hat{a}_{j(k)}, \\ \gamma \hat{a}_j &= \hat{a}_{j(k)} - \alpha_{jk}^* \hat{a}_{k(j)}, \end{aligned} \tag{2.11}$$

where $\alpha_{jk} = \langle \hat{a}_{j(k)}, a_k\rangle$, $\alpha_{kj} = \langle \hat{a}_{k(j)}, a_j\rangle$, *and* $\gamma = 1 - \alpha_{jk}^* \alpha_{kj}^* \in [0.1]$.

Proof. Take the Gram–Schmidt identity, with a_k left out for now. It separates the projection onto the component of a_j normal to the remaining vectors:

$$P_{\operatorname{span}\{a_i\ ; i\neq k\}^\perp} = P_{\operatorname{span}\{a_i\ ; i\neq j,k\}^\perp} - P_{p_{j(k)}},$$

where $p_{j(k)} = P_{\operatorname{span}\{a_i\ ; i\neq j,k\}^\perp} a_j$. Now we use this Gram–Schmidt identity to project a_k, to find

$$\begin{aligned} p_k &= p_{k(j)} - \alpha_{jk} p_{j(k)}, \\ p_j &= p_{j(k)} - \alpha_{kj} p_{k(j)}, \end{aligned} \tag{2.12}$$

where the projection $p_k = P_{\operatorname{span}\{a_i\ ; i\neq k\}^\perp} a_k$ and the inner products are $\alpha_{jk} = \langle \frac{p_{j(k)}}{\|p_{j(k)}\|^2}, a_k\rangle = \langle \hat{a}_{j(k)}, a_k\rangle$ and $\alpha_{kj} = \langle \frac{p_{k(j)}}{\|p_{k(j)}\|^2}, a_j\rangle = \langle \hat{a}_{k(j)}, a_j\rangle$. The similar identity below (2.12) follows by exchanging the roles of a_j and a_k. Note that p_k is orthogonal to $p_{j(k)}$. Taking squared norms of both sides of (2.12), using the Pythagorean theorem, we find

$$\|p_k\|^2 = \left\|p_{k(j)}\right\|^2 - |\alpha_{jk}|^2 \cdot \left\|p_{j(k)}\right\|^2. \tag{2.13}$$

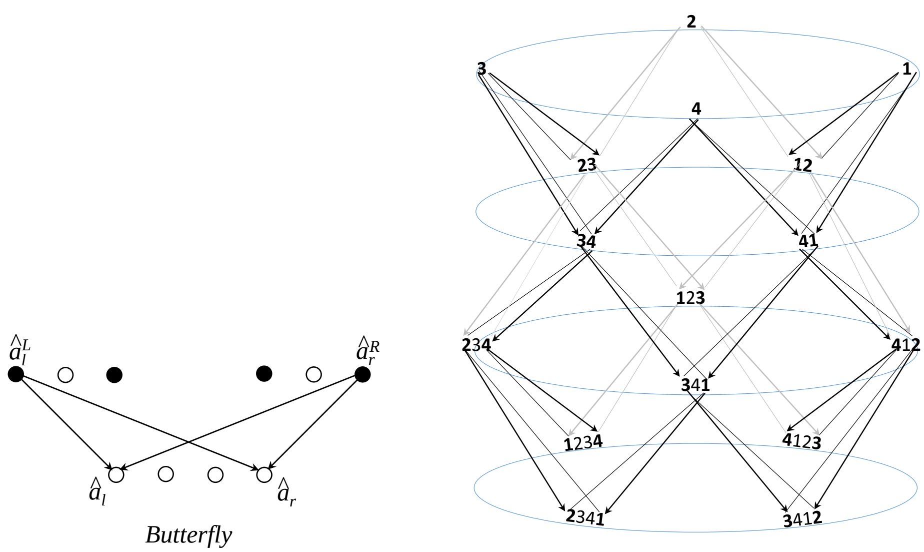
The same projection, normal to the span of all the vectors except a_j and a_k, operates inside both inner products α_{jk} and α_{kj}. Using self-adjointness of the projector, it is easy to check that

$$\alpha_{kj}^* \left\|p_{k(j)}\right\|^2 = \alpha_{jk} \left\|p_{j(k)}\right\|^2. \tag{2.14}$$

Substituting the left side of (2.14) directly into (2.13) gives

$$\|p_k\|^2 = \left\|p_{k(j)}\right\|^2 \left(1 - \alpha_{kj}^* \alpha_{jk}^*\right). \tag{2.15}$$

(This implies $\gamma = 1 - \alpha_{kj}^* \alpha_{jk}^* \geq 0$, which for $n = 2$ vectors is the Cauchy–Schwarz inequality.) Dividing both sides of (2.12) by corresponding sides of (2.15), multiplying both sides of the result by γ, and substituting (2.14) again, we have the butterfly identities (2.11). The lower identity of the pair comes from exchanging j and k. We will prove the bounds on γ in the general case in section 5; it is nonzero when all the vectors are independent. ∎



(a) The vector butterfly updates at one processing node from its left (L) and right (R) parent nodes.

(b) Parallel processor configuration for the butterfly dual transform.

Fig. 2 *Butterfly dual process.*

A parallel process for the dual transform can be built using butterfly identities on a cylinder of n rings of n processors each, shown for $n = 4$ in Figure 2(b). Each processing node is labeled by the indices of the list of vectors whose dual vectors it computes. For example, the node labeled **1234** computes the left and right dual vectors of its dual list, $\hat{a}_1^4$ and $\hat{a}_4^4$, but not the rest of the dual list; i.e., in this case, $\hat{a}_2^4$ and $\hat{a}_3^4$ are left out. Only the left and right dual vectors of the node's own dual list are actually known and updated in each node. We begin with just one vector in each of the n nodes at the top level. (It serves initially as both the left and right vector of each node.) As we go from one level down to the next, the lists for the left and right parent nodes merge to form a new list. Each node on the next level then has a list with one more vector than either of its parent nodes.

In each node, the left and right outer dual vectors $\hat{a}_\ell$ and $\hat{a}_r$ are computed from those in the left and right (L and R) parent nodes in a butterfly pattern (shown in Figure 2(a)):

$$\begin{aligned} \gamma\hat{a}_\ell &= \hat{a}_\ell^L - \alpha_\ell^{L*}\hat{a}_r^R, \\ \gamma\hat{a}_r &= \hat{a}_r^R - \alpha_r^{R*}\hat{a}_\ell^L, \end{aligned} \tag{2.16}$$

where $\gamma = 1-\alpha_\ell^{L*}\alpha_r^{R*}$, and the inner products $\alpha_\ell^L = \langle \hat{a}_\ell^L, a_r\rangle$ and $\alpha_r^R = \langle \hat{a}_r^R, a_\ell\rangle$. This butterfly is the pair of vector differences from Theorem 6 for $j, k = \ell, r$. Here is the process for the case that $a_1, \dots, a_n$ are linearly independent:

<u>Butterfly dual process</u>

```
for node j = 1,...,n                          ! initialize the first level
    â_j = a_j/‖a_j‖²
end
for level k = 2,...,n
    for each node
        α_ℓ^L = ⟨â_ℓ^L, a_r⟩
        α_r^R = ⟨â_r^R, a_l⟩
        γ = 1 − α_ℓ^{L*} α_r^{R*}
        â_ℓ = γ^{-1} (â_ℓ^L − α_ℓ^{L*} â_r^R)   ! butterfly for left,right dual vectors
        â_r = γ^{-1} (â_r^R − α_r^{R*} â_ℓ^L)
    end
end
```

At each level, each node inherits its left index ℓ from its left parent node, and its right index r from its right parent node. The vectors (primal and dual) with both of these indices are also inherited, but no other vectors are needed. The butterfly identity updates the dual vectors $\hat{a}_\ell, \hat{a}_r$ in every node. These two vectors belong to the dual transform of the primal list $a_\ell, \dots, a_r$. For each node in level k, the list has length k, though only the two outer dual vectors are ever used or known inside the node. The butterfly net illustrated in Figure 2(b) acts as a kind of distributed, collective memory. No node has full information about its own dual list, but as a group, the nodes are able to compute a complete dual list.

At the bottom level, the end result is two copies of every vector of the dual transform of the original set of vectors. Thus, in particular, a matrix of full column rank can be inverted merely by linearly combining pairs of vectors! The next example will show how this works.

Example 2. We take a 3×3 Hilbert matrix [2],

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}.$$

To invert this matrix, using the butterfly recipe, we begin with its column vectors, labeled a_1, a_2, a_3, so that we can write $A = [a_1\, a_2\, a_3]$. We can get the dual list of vectors in three steps or levels ($n = 3$). Since the given matrix A has all rational entries, we can compute using rational arithmetic. (Here the MATLAB Symbolic Math Toolbox was used.) Then our answers are exact at every step. No approximations or convergence are necessary.

Level 1 *(initial level)*:

$\|a_1\|^2 = 49/36, \qquad \|a_2\|^2 = 61/144, \qquad \|a_3\|^2 = 769/3600$

$$\hat{a}_1 = \begin{bmatrix} 36/49 \\ 18/49 \\ 12/49 \end{bmatrix}, \qquad \hat{a}_2 = \begin{bmatrix} 72/61 \\ 48/61 \\ 36/61 \end{bmatrix}, \qquad \hat{a}_3 = \begin{bmatrix} 1200/769 \\ 900/769 \\ 720/769 \end{bmatrix}.$$

Level 2:

$\alpha_1, \alpha_2 = 27/49, 108/61, \qquad \alpha_2, \alpha_3 = 216/305, 1080/769, \qquad \alpha_3, \alpha_1 = 1890/769, 27/70$

$\gamma_{12} = 73/2989, \qquad \gamma_{23} = 253/46909, \qquad \gamma_{31} = 40/769$

$\hat{a}_1, \hat{a}_2 =$ $\begin{bmatrix} \frac{252}{73} \\ \frac{-198}{73} \\ \frac{-240}{73} \end{bmatrix}, \begin{bmatrix} \frac{-360}{73} \\ \frac{408}{73} \\ \frac{468}{73} \end{bmatrix}$, $\qquad \hat{a}_2, \hat{a}_3 =$ $\begin{bmatrix} \frac{3528}{253} \\ \frac{-1968}{253} \\ \frac{-3420}{253} \end{bmatrix}, \begin{bmatrix} \frac{-4560}{253} \\ \frac{3060}{253} \\ \frac{5040}{253} \end{bmatrix}$, $\qquad \hat{a}_3, \hat{a}_1 =$ $\begin{bmatrix} \frac{-33}{7} \\ \frac{36}{7} \\ \frac{45}{7} \end{bmatrix}, \begin{bmatrix} \frac{1251}{490} \\ \frac{-396}{245} \\ \frac{-219}{98} \end{bmatrix}$.

Level 3:

$\alpha_1, \alpha_3 = -27/146, -1350/253, \qquad \alpha_2, \alpha_1 = 1404/253, 351/1960, \qquad \alpha_3, \alpha_2 = 27/28, 378/365$

$\gamma_{13} = 244/18469, \qquad \gamma_{21} = 769/123970, \qquad \gamma_{32} = 1/730$

$\hat{a}_1, \hat{a}_3 =$ $\begin{bmatrix} 9 \\ -36 \\ 30 \end{bmatrix}, \begin{bmatrix} 30 \\ -180 \\ 180 \end{bmatrix}$, $\qquad \hat{a}_2, \hat{a}_1 =$ $\begin{bmatrix} -36 \\ 192 \\ -180 \end{bmatrix}, \begin{bmatrix} 9 \\ -36 \\ 30 \end{bmatrix}$, $\qquad \hat{a}_3, \hat{a}_2 =$ $\begin{bmatrix} 30 \\ -180 \\ 180 \end{bmatrix}, \begin{bmatrix} -36 \\ 192 \\ -180 \end{bmatrix}$.

Here we end up with two copies of each dual vector, as expected. To form the inverse matrix, all that remains is to transpose the three dual vectors and stack them as the rows of a new matrix:

$$A^{-1} = \begin{bmatrix} \hat{a}_1^t \\ \hat{a}_2^t \\ \hat{a}_3^t \end{bmatrix} = \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}.$$

It is pleasant that, after all the fractions, the final results in this example are integers.

3. Axioms for the Dual Transform. To generalize the dual transform to any list of vectors, we will adopt two linear-space axioms. First we express them (in Theorem 7) in a matrix-operator form, to show their equivalence with the axioms of Moore and Penrose for the pseudoinverse A^+ of any $m \times n$ matrix A with real or complex entries, whose columns may be linearly dependent [13], [14], [15], [16].

The dual matrix is the adjoint (conjugate transpose) of the pseudoinverse; that is, $\hat{A} = (A^+)^*$.

It is well known that the pseudoinverse solves the linear least squares problem. Given a vector b of observed data and a matrix A to model the data, this problem is to find the closest point to b in the subspace $\mathbf{A} = \text{Im } A \subset \mathbb{C}^n$, the image or range of A, the span of its columns. The squared distance or error $\|b - a\|^2$ is minimum at the point $a = A\alpha$ for $\alpha = A^+b$. At this point a, we also have $b - a \perp \mathbf{A}$ [2], [3]. Moreover, α is the unique solution with minimum 2-norm. In section 4, we solve this problem in the setting of a linear vector space.

A robust numerical recipe to compute the matrix A^+ in floating point arithmetic is based on the singular value decomposition (SVD) of A [2], [12]. Numerical experiments indicate that the recipes in this paper are as accurate as the SVD-based method for pseudoinverses of random complex matrices and of matrices whose column vectors are points on an m-dimensional torus (the set of points of the form $(e^{i\theta_1}, \ldots, e^{i\theta_m})^t$, with $\theta_1, \ldots, \theta_m \in [0, 2\pi)$). (For example, the phased response at one time instant of an array of m isotropic radio antennas to a unit-amplitude plane wave arriving from a given direction lies on this torus [7].)

Two conditions that characterize A^+ in general are $A^+A = P_{\text{Im } A^+}$ and $AA^+ = P_{\text{Im } A}$. These two Moore conditions are easily seen to be equivalent to the four Penrose conditions $A^+AA^+ = A^+$, $AA^+A = A$, $(A^+A)^* = A^+A$, and $(AA^+)^* = AA^+$ [19].

From the Penrose conditions, we also have $A^+A = P_{\text{ker } A^\perp}$, since $A^+Ak = 0$ when and only when $Ak = AA^+Ak = 0$; that is, k is in the kernel or nullspace of A, ker A. Note that for A of full column rank, its kernel is trivial (A only maps 0 to 0), so we have $A^+A = I$. Thus A^+ is a left inverse of A. In terms of the dual matrix, this becomes $\hat{A}^*A = I$, which is the same as biorthogonality (2.1).

To characterize the dual transform in an inner product space, we adopt two new axioms. We give them in matrix form first, to relate them to the axioms for the pseudoinverse. A and its dual transform $\hat{A}$ have the same kernels and images, respectively, mapped to each other in a strongly symmetric way.

THEOREM 7. *The Moore–Penrose conditions are equivalent to the following two conditions:*

1. $A\alpha = 0$ *if and only if* $\hat{A}\alpha = 0$.
2. $A\alpha_1 = b$ *and* $\hat{A}\alpha_2 = b$ *either both have solutions, given by* $\alpha_1 = \hat{A}^*b$ *and* $\alpha_2 = A^*b$*, or else neither has a solution.*

It follows from these conditions that when the solutions in condition 2 *exist, they are the unique solutions orthogonal to ker A. They are related as* $\alpha_1 = \hat{A}^*\hat{A}\alpha_2$ and $\alpha_2 = A^*A\alpha_1$.

Proof. The Penrose conditions imply conditions 1 and 2. To show condition 1, if $\hat{A}^*(A\alpha) = A^*(\hat{A}\alpha) = 0$, then $A\hat{A}^*(A\alpha) = A\alpha = 0$ and $\hat{A}A^*(\hat{A}\alpha) = \hat{A}\alpha = 0$. So either dependency, $A\alpha = 0$ or $\hat{A}\alpha = 0$, implies both.

2. For any solution of one of the equations, there exists a solution of the other: If $A\alpha_1 = b$, then $A\hat{A}^*A\alpha_1 = \hat{A}(A^*A\alpha_1) = b$, so $\alpha_2 = A^*A\alpha_1 = A^*b$ satisfies $\hat{A}\alpha_2 = b$. Similarly, if $\hat{A}\alpha_2 = b$, then $\alpha_1 = \hat{A}^*\hat{A}\alpha_2 = \hat{A}^*b$ satisfies $A\alpha_1 = b$.

If $\alpha_1 = \hat{A}^*b$ solves $A\alpha_1 = b$, then $\alpha = \alpha_1 + \alpha_0$ is another solution for any $\alpha_0 \in$ ker A. But $\alpha_1 = \hat{A}^*b$ is the unique solution orthogonal to ker A, since, in general, Im $\hat{A}^* = (\text{ker } \hat{A})^\perp$ [2], [3], and by condition 1, $(\text{ker } \hat{A})^\perp = (\text{ker } A)^\perp$. To show it is unique, suppose α is a solution of $\hat{A}\alpha = b$ orthogonal to ker A. Then clearly $\alpha - \alpha_1$ is both orthogonal to and in ker A, so $\alpha - \alpha_1 = 0$.

Conditions 1 and 2 imply the Moore conditions: For every $b \in$ Im A, condition 2 implies that $A\hat{A}^*b = b = \hat{A}A^*b$. Condition 2 also implies that $(\text{Im } A)^\perp = (\text{Im } \hat{A})^\perp$. We also have the fundamental relations of linear algebra, $(\text{Im } A)^\perp = \ker A^*$ and $(\text{Im } \hat{A})^\perp = \ker \hat{A}^*$. Thus, for every $c \in$ the orthogonal complement $(\text{Im } A)^\perp$, we have $A\hat{A}^*q = 0 = \hat{A}A^*q$. Therefore $A\hat{A}^* = \hat{A}A^* = P_{\text{Im}A}$.

Again, condition 1 implies that the kernels of A and $\hat{A}$ are equal, so their orthogonal complements are equal. Thus we have $\text{Im } \hat{A}^* = (\ker \hat{A})^\perp = (\ker A)^\perp = \text{Im } A^*$. Now we must show that for every vector $\alpha \in \text{Im } \hat{A}^*$, there is a vector b_1 such that $b_1 = A\alpha$. For some b_1, $\alpha = \hat{A}^*b_1$. We may assume that $b_1 \in \text{Im}A = (\ker A^*)^\perp = (\ker \hat{A}^*)^\perp$, since any component of $b_1 \in \ker \hat{A}^*$ vanishes, leaving α unchanged if we remove that component. By condition 2, the equation $A\alpha_1 = b_1$ has unique solution $\alpha_1 = \hat{A}^*b_1$ for $b_1 \in \text{Im } A$. Thus $\alpha = \alpha_1$ and we have $b_1 = A\alpha$. (To show that b_1 is the unique solution for a given α, suppose b is any solution of $\hat{A}^*b = \alpha$ orthogonal to $\ker A^*$. Then clearly $b - b_1$ is both orthogonal to and in $\ker A^*$, so $b - b_1 = 0$.) Similarly, we can show that, given the same vector α, $\alpha = A^*b_2$ for the specific solution $b_2 = \hat{A}\alpha$. It now follows that $\hat{A}^*A\alpha = \alpha = A^*\hat{A}\alpha$. Also, for every $\beta \in \ker A = \ker \hat{A}$, $\hat{A}^*A\beta = 0 = A^*\hat{A}\beta$. Therefore $\hat{A}^*A = A^*\hat{A} = P_{\text{Im}\hat{A}^*} = P_{\text{Im}A^+}$. ∎

Having related the axioms for the dual transform and the pseudoinverse, we now promote the dual transform axioms to a linear-space setting. We do so by replacing the columns of the matrix A by vectors a_k in a space $\mathbf{V}$ for $k = 1, \ldots, n$, keeping the components of the solution vectors α as combining coefficients of the vectors in $\mathbf{V}$.

Then the linear map $A : \mathbb{C}^n \to \mathbf{V}$ is given by $\alpha \mapsto \sum_k \alpha_k a_k$. The image Im A of this map is the span of the vectors, $\text{span}\{a_k\}_{k=1}^n$. The adjoint map $A^* : \mathbf{V} \to \mathbb{C}^n$ maps $b \mapsto \alpha = \{\langle a_k, b\rangle\}_{k=1}^n$. To check that this is the adjoint, take any $\alpha \in \mathbb{C}^n, b \in \mathbf{V}$. We have $\langle b, A\alpha\rangle = \langle b, \sum_k \alpha_k a_k\rangle = \sum_k \alpha_k \langle b, a_k\rangle = \sum_k \langle a_k, b\rangle^* \alpha_k = \langle \{\langle a_k, b\rangle\}_{k=1}^n, \{\alpha_k\}_{k=1}^n\rangle = \langle A^*b, \alpha\rangle$.

The two axioms in Theorem 7 remain as stated, but now for linear maps $A, \hat{A}$ and their adjoint maps $A^*, \hat{A}^*$, rather than matrices. It will be convenient to restate the two axioms directly as follows.

DEFINITION 8. *Consider a list of vectors $a_1, a_2, \ldots, a_n$ in a linear space $\mathbf{V}$ with real or complex inner product. Its* dual *list $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n$ obeys the following two axioms:*

1. Mirrored linear dependencies. *For combining coefficients $\alpha_1^\circ, \ldots, \alpha_n^\circ$, we have*

$$\sum_k \alpha_k^\circ a_k = 0 \quad \textit{if and only if} \quad \sum_k \alpha_k^\circ \hat{a}_k = 0.$$

2. Mirrored linear span. *A vector $b \in \mathbf{V}$ is expressible as a linear combination*

$$\sum_k \breve{\alpha}_k a_k = b \quad \textit{if and only if} \quad \sum_k \alpha_k \hat{a}_k = b.$$

When they exist, the combining coefficients are given by

$$\breve{\alpha}_k = \langle \hat{a}_k, b\rangle \quad \textit{and} \quad \alpha_k = \langle a_k, b\rangle,\ k = 1, \ldots, n.$$

From these two linear-space axioms, we will develop the theory and linear-algebraic algorithms for the dual transform. From now on, when we refer to axiom 1 or 2, we always mean the respective axiom 1 or 2 of Definition 8.

4. Dual Theory of Projectors and Protractors. The expression given in axiom 2 is the orthogonal projector onto $\mathbf{A}$, the span of $a_1, \ldots, a_n$. By axiom 2, $\mathbf{A}$ equals the span of $\hat{a}_1, \ldots, \hat{a}_n$ too.

THEOREM 9. *The linear operator* $P_{\mathbf{A}} = \sum_k \langle \hat{a}_k, \cdot \rangle a_k$ *is the orthogonal projector onto* $\mathbf{A}$*. It is also given by* $P_{\mathbf{A}} = \sum_k \langle a_k, \cdot \rangle \hat{a}_k$*.*

Proof. For any vector $b \in \mathbf{A}$, $P_{\mathbf{A}}(b) = b$ by axiom 2 for either form of the projector. For any vector $c \in \mathbf{A}^{\perp}$, clearly $P_{\mathbf{A}}(c) = 0$. ∎

We can easily demonstrate that $P_{\mathbf{A}}$ is self-adjoint. Again, for any vector $c \in \mathbf{A}^{\perp}$, $P_{\mathbf{A}}(c) = 0$. So by linearity of $P_{\mathbf{A}}$, it is sufficient to consider any pair of vectors $a, b \in \mathbf{A}$. We have $\langle a, P_{\mathbf{A}}(b) \rangle = \langle a, \sum_k \langle \hat{a}_k, b \rangle a_k \rangle = \sum_k \langle \hat{a}_k, b \rangle \langle a, a_k \rangle = \sum_k \langle a_k, a \rangle^* \langle \hat{a}_k, b \rangle = \langle \sum_k \langle a_k, a \rangle \hat{a}_k, b \rangle = \langle P_{\mathbf{A}}(a), b \rangle$, using the mirrored form of $P_{\mathbf{A}}$ given by axiom 2.

By symmetry of the axioms, it is clear that the dual transform is reflexive, that is, $\hat{\hat{a}}_k = a_k$, for $k = 1, \ldots, n$. If $a_k = 0$ for some k, let its coefficient alone be nonzero in axiom 1. Then we always have $\hat{0} = 0$. For simplicity in what follows, unless otherwise noted, we will assume that the vectors $a_k \neq 0$. For only one nonzero vector, since $\langle \hat{a}, a \rangle = 1$, with $b = a$ in axiom 2, it follows that $\hat{a} = a/\|a\|^2$. If the $\{a_k\}_{k=1}^n$ are an orthonormal set, then by axiom 2, $\hat{a}_k = a_k$.

Given any vector b in a linear space $\mathbf{V}$, what is the closest vector a in the subspace $\mathbf{A}$ to b? That is, for what $a \in \mathbf{A}$ is the squared distance or error $\|b - a\|^2$ minimum? This is the linear least squares problem, posed for a linear space. With the (generalized) dual list, we can immediately solve it for any primal list $a_1, \ldots, a_n$ that spans the subspace $\mathbf{A}$.

THEOREM 10. *For any* $b \in \mathbf{V}$*, a set of scalar coefficients* α_j *that combine the vectors* a_j *with least squared error*

$$\left\| b - \sum_{j=1}^{n} \alpha_j a_j \right\|^2 \tag{4.1}$$

is given by

$$\alpha_j = \langle \hat{a}_j, b \rangle \quad \text{for } j = 1, 2, \ldots, n. \tag{4.2}$$

Proof. First, project b orthogonally onto the subspace $\mathbf{A}$; i.e., put $a = P_{\mathbf{A}} b = \sum_k \langle \hat{a}_k, b \rangle a_k$, by Theorem 9. This means that $b - a \perp \mathbf{A}$. Consider any other point a' in $\mathbf{A}$. We can write its distance vector to b as $b - a' = (b - a) + (a - a')$, where $a - a' \in \mathbf{A}$ is nonzero. By the Pythagorean theorem, $\|b - a'\|^2 = \|(b - a)\|^2 + \|(a - a')\|^2$. Thus $\|b - a\|^2$ is minimum for this a. ∎

Orthogonal projectors can't do everything. In a linearly dependent situation, we need another kind of linear operator, called a protractor, which rearranges the vectors inside its own subspace $\mathbf{A}$ and ignores anything outside it. In particular, this linear operator simply maps each primal vector to its dual vector.

DEFINITION 11. *Given a list of vectors* $a_1, a_2, \ldots, a_n$ *in* $\mathbf{V}$ *and its dual list* $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n$*, the* dualizer *or* protractor *is the linear operator* $Q : \mathbf{V} \to \mathbf{A}$ *such that* $Qb = 0$ *for every* $b \in \mathbf{A}^{\perp}$ *and*

$$Qa_j = \hat{a}_j \text{ for every } j = 1, \ldots, n.$$

THEOREM 12. $Q = \sum_k \langle \hat{a}_k, \cdot \rangle \hat{a}_k$*. Restricted to* $\mathbf{A}$*,* $Q^+ = \sum_k \langle a_k, \cdot \rangle a_k$ *is the inverse operator; moreover,* $QQ^+ = Q^+ Q = P_{\mathbf{A}}$*.*

Proof. For $j = 1, \ldots, n$, we have $a_j = P_{\mathbf{A}} a_j = \sum_k \langle \hat{a}_k, a_j \rangle a_k$. By axiom 1, $\hat{a}_j = \sum_k \langle \hat{a}_k, a_j \rangle \hat{a}_k = Qa_j$. The rest of the proof is easy. ∎

THEOREM 13. *Q is self-adjoint. Therefore, we have*

$$\langle \hat{a}_j, a_k\rangle = \langle \hat{a}_k, a_j\rangle^* = \langle a_j, \hat{a}_k\rangle \text{ for } j, k = 1, \ldots, n.$$

For $j = k$, we have $\langle \hat{a}_k, a_k\rangle = \langle \hat{a}_k, a_k\rangle^*$, so that this is always real valued.

Proof. By Theorem 12, for any vector $c \in \mathbf{A}^\perp$, $Qc = 0$. So by linearity of Q, it is sufficient to consider any pair of vectors $a, b \in \mathbf{A}$. We have $\langle a, Qb\rangle = \langle a, \sum_k \langle \hat{a}_k, b\rangle \hat{a}_k\rangle = \sum_k \langle \hat{a}_k, b\rangle\langle a, \hat{a}_k\rangle = \sum_k \langle \hat{a}_k, a\rangle^* \langle \hat{a}_k, b\rangle = \langle \sum_k \langle \hat{a}_k, a\rangle \hat{a}_k, b\rangle = \langle Qa, b\rangle$. In particular, we can choose $a = Qa_j = \hat{a}_j$ and $b = a_k$. □

THEOREM 14 (biorthogonality generalized). *The orthogonal projector onto the adjoint operator's image space $\mathbf{A}^* = \mathrm{Im}\, A^*$ is given by the matrix $P_{\mathbf{A}^*} = [\langle a_\ell, \hat{a}_k\rangle]_{\ell,k=1}^n$.*

Proof. First we must show that for any vector $\alpha \in \mathrm{Im}\, A^*$, $P_{\mathbf{A}^*}(\alpha) = \alpha$. For some $b \in \mathbf{V}$, we have $\alpha = \{\alpha_k\}_{k=1}^n = \{\langle a_k, b\rangle\}_{k=1}^n$. Then

$$\begin{aligned} P_{\mathbf{A}^*}\alpha &= [\langle a_\ell, \hat{a}_j\rangle]_{\ell,j=1}^n [\langle a_j, b\rangle]_{j=1}^n \\ &= \left[\left\langle \sum_{j=1}^n \langle a_\ell, \hat{a}_j\rangle^* a_j, b\right\rangle\right]_{\ell=1}^n = \left[\left\langle \sum_{j=1}^n \langle \hat{a}_j, a_\ell\rangle a_j, b\right\rangle\right]_{\ell=1}^n = [\langle a_\ell, b\rangle]_{\ell=1}^n = \alpha, \end{aligned}$$

by axiom 2.

Second, note that the matrix $P_{\mathbf{A}^*}$ is conjugate symmetric, since by Theorem 13, $\langle a_\ell, \hat{a}_j\rangle = \langle a_j, \hat{a}_\ell\rangle^*$. Consider a complex vector $\beta \perp \mathrm{Im}\, A^*$. For any $b \in \mathbf{V}$, this means that $\sum_j \langle a_j, b\rangle^* \cdot \beta_j = 0$. Taking $b = \hat{a}_k$, this is $\sum_j \langle a_j, \hat{a}_k\rangle^* \cdot \beta_j = \sum_j \langle a_k, \hat{a}_j\rangle \cdot \beta_j = 0$. This last sum expresses the product of the kth row of $P_{\mathbf{A}^*}$ and β. Thus, we find that $P_{\mathbf{A}^*}\beta = 0$. □

We shall need the following observation about the adjoint image space projector acting on its left.

LEMMA 15. *The orthogonal projector matrix $P_{\mathbf{A}^*} = [\langle a_\ell, \hat{a}_k\rangle]_{\ell,k=1}^n$ operates to its left to project a row of vectors in $\mathbf{V}$ as follows:*

$$\begin{aligned} (\hat{a}_1, \ldots, \hat{a}_n) &= (\hat{a}_1, \ldots, \hat{a}_n) \cdot P_{\mathbf{A}^*}, \\ (a_1, \ldots, a_n) &= (a_1, \ldots, a_n) \cdot P_{\mathbf{A}^*}. \end{aligned}$$

Proof. By axiom 2, we can write

$$\begin{aligned} \hat{a}_k &= \sum_{j=1}^n \langle a_j, \hat{a}_k\rangle \hat{a}_j, \quad k = 1, \ldots, n, \\ (\hat{a}_1, \ldots, \hat{a}_n) &= (\hat{a}_1, \ldots, \hat{a}_n) \cdot [\langle a_j, \hat{a}_k\rangle]_{j,k=1}^n = (\hat{a}_1, \ldots, \hat{a}_n) \cdot P_{\mathbf{A}^*}, \end{aligned}$$

and similarly, now also using $\langle \hat{a}_j, a_k\rangle = \langle a_j, \hat{a}_k\rangle$,

$$\begin{aligned} a_k &= \sum_{j=1}^n \langle \hat{a}_j, a_k\rangle a_j, \quad k = 1, \ldots, n, \\ (a_1, \ldots, a_n) &= (a_1, \ldots, a_n) \cdot [\langle a_j, \hat{a}_k\rangle]_{j,k=1}^n = (a_1, \ldots, a_n) \cdot P_{\mathbf{A}^*}. \end{aligned}$$ □

Now we come to two key identities. They relate the dual transform of a sequence of n vectors to the dual transforms of subsequences of $n - 1$ vectors. For cases where the list has all but one vector, we now need a somewhat flexible notation for the dual vectors. Let $\hat{a}^n_{j(k)}$ denote the vector dual to a_j in the dual transform of $\{a_j\}_{j=1,\ldots,n,j\neq k}$. When there is no ambiguity, we can omit some of the labels. If the order or length n of the complete list is fixed and an arbitrary vector a_k is left out of the list, we can suppress the label n and simply write $\hat{a}_{j(k)}$. In the case in which $k = n$ and a_n is always left off the end of the list, it is sometimes convenient to write $\hat{a}^{n-1}_j$ instead of $\hat{a}^n_{j(n)}$. Similarly, when the context is clear, for dual vectors of the complete list, we write $\hat{a}_j$ for $\hat{a}^n_j$ to denote the vector dual to a_j in the dual transform of $\{a_j\}_{j=1,\ldots,n}$.

These two identities are almost equivalent, taken in pairs. We proved the butterfly identity for a case of independence in Theorem 6.

THEOREM 16. *For* $j, k = 1, \ldots, n,\ j \neq k$,
1. (Greville identity) $\quad \hat{a}_j = \hat{a}_{j(k)} - \alpha^*_{jk}\hat{a}_k$,
2. (Butterfly identity) $\quad \gamma\hat{a}_j = \hat{a}_{j(k)} - \alpha^*_{jk}\hat{a}_{k(j)}$,

where $\alpha_{jk} = \langle \hat{a}_{j(k)}, a_k\rangle$, $\alpha_{kj} = \langle \hat{a}_{k(j)}, a_j\rangle$, *and* $\gamma = 1 - \alpha^*_{jk}\alpha^*_{kj}$.

Proof. 1. We give a vector-space version of Greville's matrix proof [17, equation (8)]. Without loss of generality, we prove this identity for $k = n$. Consider the projection $P_{\mathbf{A}_n}$ of the vectors $\hat{a}_{1(n)}, \ldots, \hat{a}_{n-1(n)} \in \mathbf{A}_n = \text{Im } A_n = \text{span}\{a_j\}^n_{j=1}$, which leaves them unchanged. For $k = 1, \ldots, n-1$,

$$\hat{a}_{k(n)} = \sum_{j=1}^{n}\langle a_j, \hat{a}_{k(n)}\rangle\hat{a}_j = \sum_{j=1}^{n-1}\langle a_j, \hat{a}_{k(n)}\rangle\hat{a}_j + \langle \hat{a}_{k(n)}, a_n\rangle^*\hat{a}_n$$
$$= (\hat{a}_1, \ldots, \hat{a}_{n-1})[\langle a_j, \hat{a}_{k(n)}\rangle]^{n-1}_{j=1} + \langle \hat{a}_{k(n)}, a_n\rangle^*\hat{a}_n.$$

By Theorem 14, the orthogonal projector $P_{\text{Im } A^*_{n-1}} = [\langle a_j, \hat{a}_{k(n)}\rangle]^{n-1}_{j,k=1}$. Let $\hat{A}^n_{n-1}$ map $\{\alpha_k\}^{n-1}_{k=1} \mapsto \sum^{n-1}_{k=1}\alpha_k\hat{a}^n_k$. Then $P_{\text{Im } A^*_{n-1}} = P_{\text{Im } \hat{A}^{n*}_{n-1}}$, because the adjoint spaces Im A^*_{n-1} and Im $\hat{A}^{n*}_{n-1}$ are equal. This follows because they are orthogonal complements of the corresponding kernels ker A_{n-1} and ker $\hat{A}^n_{n-1}$, which are equal by axiom 1 with combining coefficient $\alpha^\circ_n = 0$. Thus, by Lemma 15, $(\hat{a}_1, \ldots, \hat{a}_{n-1})\,[\langle \hat{a}_j, a_k\rangle]^{n-1}_{j=1} = \hat{a}_k$ for $k = 1, \ldots, n-1$.

2. We may write two cases of the Greville identity as a 2×2 linear system of vector equations,

$$\begin{pmatrix} \hat{a}_{j(k)} \\ \hat{a}_{k(j)} \end{pmatrix} = \begin{bmatrix} 1 & \alpha^*_{jk} \\ \alpha^*_{kj} & 1 \end{bmatrix} \begin{pmatrix} \hat{a}_j \\ \hat{a}_k \end{pmatrix}. \tag{4.3}$$

Then, if the determinant $\gamma = 1 - \alpha^*_{jk}\alpha^*_{kj}$ is not zero, we have

$$\gamma \begin{pmatrix} \hat{a}_j \\ \hat{a}_k \end{pmatrix} = \begin{bmatrix} 1 & -\alpha^*_{jk} \\ -\alpha^*_{kj} & 1 \end{bmatrix} \begin{pmatrix} \hat{a}_{j(k)} \\ \hat{a}_{k(j)} \end{pmatrix}. \tag{4.4}$$

This is identity 2. If $\gamma = 0$, that is, $\alpha^*_{jk}\alpha^*_{kj} = 1$, then we premultiply both sides of the upper equation of (4.3) by α^*_{kj} and compare the result with the lower equation, to find that $\hat{a}_{j(k)} = \alpha^{*-1}_{kj}\hat{a}_{k(j)}$. Similarly, $\hat{a}_{k(j)} = \alpha^{*-1}_{jk}\hat{a}_{j(k)}$. This verifies that (4.4) is still true when $\gamma = 0$. So the pair of Greville identities implies the corresponding pair of butterfly identities. ∎

Stereographic Projection. We saw in section 2 that orthogonal projection is used to update the dual transform when we add a new vector that is independent of the other vectors. (We say that the vector $a_k \neq 0$ is linearly *independent of the rest*,

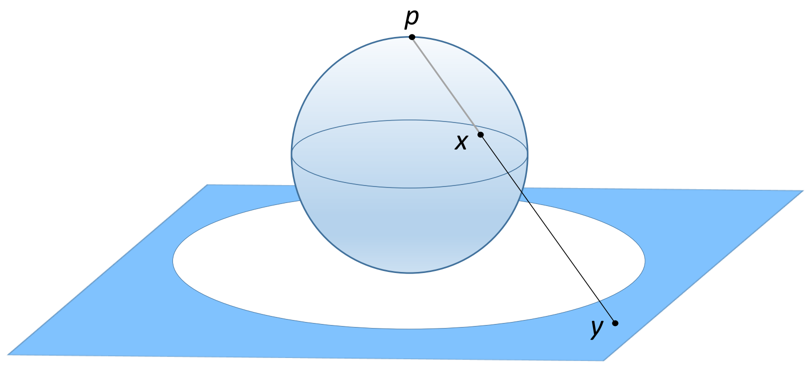


Fig. 3 *Stereographic projection in $n = 3$ dimensions for a sphere of radius $1/2$, tangent to the plane at its south pole $(0, 0, 0)$. Its north pole is $p = (0, 0, 1)$. This projection occurs naturally for arbitrary n in the course of updating the combining coefficients $\{\alpha_{jn} = \langle \hat{a}_j, a_n \rangle\}_{j=1}^n$. In this example, we update $y = (\alpha_{12}, \alpha_{22})$ to $x = (\alpha_{13}, \alpha_{23}, \alpha_{33})$ when a_3 is dependent on a_1, a_2. The case of a_3 independent of a_1, a_2 maps $y = \infty$ to the north pole $x = p = \{\delta_{j3}\}_{j=1}^3$. The coefficient vectors x on the sphere represent generalized biorthogonal relations between the primal and dual vectors.*

$a_1, a_2, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n$, if any linear combination of all n vectors that equals 0 must have a zero coefficient of a_k.) We now come to a surprising fact: stereographic projection occurs naturally, without human artifice, in the process of updating the dual when the new vector is dependent on the others.

To make this connection clear, first we recall the definition of this kind of projection [20], [21]. It maps all but one point of a sphere S^{n-1} in $\mathbb{C}^n$ onto the subspace (hyperplane) $\mathbb{C}^{n-1}$ that it is tangent to. Let the sphere have radius $1/2$, its south pole at the origin, and its north pole at the point $p = (0, \ldots, 0, 1)$. Its center is $p/2 = (0, \ldots, 0, 1/2)$. Let $x = (x_1, \ldots, x_n)$ be any point on the sphere other than p. We project the ray from p through x until it intersects the hyperplane at point y (see Figure 3). Considering similar right triangles normal to the hyperplane, with hypotenuses px and py along the ray, we easily get the continuous map $f : S^{n-1} - p \to \mathbb{C}^{n-1}$ given by

$$y = f(x_1, \ldots, x_n) = \frac{1}{p_n - x_n}(x_1, \ldots, x_{n-1}) \,. \tag{4.5}$$

(We have fixed the north pole p to be a real vector. For the complex sphere, there is a degree of freedom in the choice of this pole. If we multiply the points p and x on the sphere by a phasor $e^{i\theta}, \theta \in [0, 2\pi)$, the projected point y in the hyperplane stays the same.) The inverse map g is given by

$$x = g(y_1, \ldots, y_{n-1}) = (ty_1, \ldots, ty_{n-1}, 1 - t) \,, \tag{4.6}$$

where $t = 1/(1 + |y_1|^2 + \cdots + |y_{n-1}|^2)$, based on the relation $p_n - x_n = 1 - (1 - t) = t$. We can confirm that, for this value of t, $x = g(y)$ is not merely a point on the ray, but also a point on the sphere, by means of the identity

$$|y_1|^2 + \cdots + |y_{n-1}|^2 + \left(|y_1|^2 + \cdots + |y_{n-1}|^2 - 1\right)^2/4 = \tfrac{1}{4}(1 + |y_1|^2 + \cdots + |y_{n-1}|^2)^2. \tag{4.7}$$

THEOREM 17 (forward stereographic projection). *If a_n is independent of the rest, we have biorthogonal relations*

$$\langle \hat{a}_n^n, a_k\rangle = \delta_{kn}, \quad k = 1, \dots, n. \tag{4.8}$$

If $a_n \neq 0$ is dependent on the rest, we have $\langle \hat{a}_n^n, a_n\rangle \in (0, 1)$ and

$$\langle \hat{a}_k^{n-1}, a_n\rangle = \langle \hat{a}_k^n, a_n\rangle/(1 - \langle \hat{a}_n^n, a_n\rangle), \quad k = 1, \dots, n-1. \tag{4.9}$$

The relationship (4.9) *for the dependent case has the form* (4.5) *of a forward stereographic projection taking $x = \{\langle \hat{a}_k^n, a_n\rangle\}_{k=1}^n$ to $y = \{\langle \hat{a}_k^{n-1}, a_n\rangle\}_{k=1}^{n-1}$. The biorthogonal relations* (4.8) *for the independent case represent the north pole $x = p$, which maps to the point at infinity $y = \infty$.*

Proof. Using axiom 2, take the projection of a_n onto the full span of $a_1, \dots, a_n$:

$$a_n = \sum_{k=1}^{n} \langle \hat{a}_k^n, a_n\rangle a_k,$$

$$(1 - \langle \hat{a}_n^n, a_n\rangle)a_n = \sum_{k=1}^{n-1} \langle \hat{a}_k^n, a_n\rangle a_k = \sum_{k=1}^{n-1} \langle a_k, \hat{a}_n^n\rangle a_k. \tag{4.10}$$

Equation (4.10) represents a linear combination equal to 0. If a_n is independent of the rest, its coefficient must be 0. Thus $\langle \hat{a}_n^n, a_n\rangle = 1$. Then, taking the inner product of both sides of (4.10) with $\hat{a}_n^n$, we have $0 = \sum_{k=1}^{n-1} |\langle a_k, \hat{a}_n^n\rangle|^2$. Therefore, $\hat{a}_n^n \perp a_k$, for $k = 1, \dots, n-1$.

If a_n depends on the rest, taking the inner product of both sides of (4.10) with $\hat{a}_n^n$, we get $(1 - \langle \hat{a}_n^n, a_n\rangle)\langle \hat{a}_n^n, a_n\rangle = \sum_{k=1}^{n-1} |\langle a_k, \hat{a}_n^n\rangle|^2 \geq 0$. This requires that $\langle \hat{a}_n^n, a_n\rangle \in [0, 1]$. In fact, $\langle \hat{a}_n^n, a_n\rangle \neq 1, 0$. Suppose $\langle \hat{a}_n^n, a_n\rangle = 1$. As we saw in the previous paragraph, this would imply biorthogonal relations $\langle \hat{a}_n^n, a_k\rangle = \delta_{kn}$. But by Theorem 4, this implies that a_n is independent of the rest, contradicting our assumption. Next, suppose $\langle \hat{a}_n^n, a_n\rangle = 0$. Then, similarly, $\sum_{k=1}^{n-1} |\langle \hat{a}_k^n, a_n\rangle|^2 = 0$. Thus $\langle \hat{a}_k^n, a_n\rangle = 0$ for $k = 1, \dots, n$, or $a_n \perp \mathbf{A}_n$. But $a_n \in \mathbf{A}_n$, by axiom 2. Therefore, $a_n = 0$. Thus, in general, we have the definiteness property that $\langle \hat{a}_n^n, a_n\rangle = 0$ if and only if $a_n = 0$.

So, when a_n depends on the rest, we may divide all members of (4.10) by $1 - \langle \hat{a}_n^n, a_n\rangle \neq 0$ to obtain

$$a_n = \sum_{k=1}^{n-1} \langle \hat{a}_k^n, a_n\rangle a_k/(1 - \langle \hat{a}_n^n, a_n\rangle). \tag{4.11}$$

By axiom 2, a_n also projects onto the rest directly:

$$a_n = \sum_{k=1}^{n-1} \langle \hat{a}_k^{n-1}, a_n\rangle a_k. \tag{4.12}$$

Since the combining coefficients for a_n are unique and orthogonal to the kernel of the rest, they are equal in both expansion (4.11) and expansion (4.12). ∎

THEOREM 18 (Greville's alternative update [17]).

1. *If $a_n \notin \text{span}\{a_1, \dots, a_{n-1}\}$, then $\hat{a}_n^n = p_n/\|p_n\|^2$,*
where $p_n = P_{\mathbf{A}_{n-1}^\perp} a_n = a_n - \sum_{j=1}^{n-1} \alpha_{jn} a_j$ and $\alpha_{jn} = \langle \hat{a}_j^{n-1}, a_n\rangle$.

2. *If $a_n \in \text{span}\{a_1, \dots, a_{n-1}\}$, then $\hat{a}_n^n = q_n/\beta_{nn}$,*
where $q_n = Q_{\mathbf{A}_{n-1}} a_n = \sum_{j=1}^{n-1} \alpha_{jn} \hat{a}_j^{n-1}$ and $\beta_{nn} = 1 + \sum_{j=1}^{n-1} \alpha_{jn}^* \alpha_{jn}$.

The test for independence of a_n from the rest of the vectors is that $\|p_n\|^2 > 0$.

Proof. 1. By Theorem 17, if a_n is independent of the rest, we have biorthogonal relations $\langle \hat{a}_n^n, a_k\rangle = \delta_{kn}$ for $\hat{a}_n^n$. Then Theorem 4 implies that $\hat{a}_n^n = p_n/\|p_n\|^2$, where $p_n = P_{\mathbf{A}_{n-1}^\perp} a_n$. By Theorem 9, $p_n = a_n - \sum_{j=1}^{n-1} \alpha_{jn} a_j$.

2. Since a_n is in the span of $a_1, \ldots, a_{n-1}$, by axiom 2, $a_n = \sum_{k=1}^{n-1} \langle \hat{a}_k^{n-1}, a_n\rangle a_k$. By axiom 1, the dual dependency also holds: $\hat{a}_n^n = \sum_{k=1}^{n-1} \langle \hat{a}_k^{n-1}, a_n\rangle \hat{a}_k^n$. Then, by the Greville identity, Theorem 16(1),

$$\begin{aligned} \hat{a}_n^n &= \sum_{k=1}^{n-1} \langle \hat{a}_k^{n-1}, a_n\rangle \hat{a}_k^{n-1} - \sum_{k=1}^{n-1} |\langle \hat{a}_k^{n-1}, a_n\rangle|^2 \hat{a}_n^n \\ &= \sum_{k=1}^{n-1} \langle a_k, \beta_{nn}^{-1} q_n\rangle \hat{a}_k^{n-1} = \beta_{nn}^{-1} q_n, \end{aligned} \tag{4.13}$$

by axiom 2, since q_n is in the span of $\hat{a}_1^{n-1}, \ldots, \hat{a}_{n-1}^{n-1}$ by its definition as $Q_{\mathbf{A}_{n-1}} a_n$. We also used the self-adjoint property of $Q_{\mathbf{A}_{n-1}}$. ∎

LEMMA 19. *For $a_n \neq 0$ dependent on the rest, let the positive scalar $\beta^{-1} = 1 - \langle \hat{a}_n^n, a_n\rangle$. Then $\beta = \beta_{nn} = 1 + \sum_{k=1}^{n-1} |\langle \hat{a}_k^{n-1}, a_n\rangle|^2$ and*

$$\begin{aligned} \beta^2 \sum_{k=1}^{n-1} |\langle \hat{a}_k^n, a_n\rangle|^2 &= \sum_{k=1}^{n-1} |\langle \hat{a}_k^{n-1}, a_n\rangle|^2 = \beta - 1, \\ 0 < \langle \hat{a}_n^n, a_n\rangle = \sum_{k=1}^{n} |\langle \hat{a}_k^n, a_n\rangle|^2 &= \beta^{-1} \sum_{k=1}^{n-1} |\langle \hat{a}_k^{n-1}, a_n\rangle|^2 \equiv 1 - 1/\beta < 1. \end{aligned}$$

Proof. Taking inner products of both sides of (4.13) with a_n and substituting the original coefficients from equations (4.9) in Theorem 17, we find

$$\begin{aligned} \langle \hat{a}_n^n, a_n\rangle = 1 - \beta^{-1} &= \sum_{k=1}^{n-1} \langle a_k, \hat{a}_n^n\rangle^* \langle \hat{a}_k^{n-1}, a_n\rangle = \beta \sum_{k=1}^{n-1} \langle \hat{a}_k^n, a_n\rangle^* \langle \hat{a}_k^n, a_n\rangle \\ &= \sum_{k=1}^{n-1} \langle \hat{a}_k^n, a_n\rangle^* \langle \hat{a}_k^{n-1}, a_n\rangle = \beta^{-1} \sum_{k=1}^{n-1} \langle \hat{a}_k^{n-1}, a_n\rangle^* \langle \hat{a}_k^{n-1}, a_n\rangle. \end{aligned}$$

Then $\sum_{k=1}^{n-1} |\langle \hat{a}_k^n, a_n\rangle|^2 + |\langle \hat{a}_n^n, a_n\rangle|^2 = \beta^{-1}(1-\beta^{-1}) + (1-\beta^{-1})^2 = 1 - \beta^{-1} = \langle \hat{a}_n^n, a_n\rangle$. Note that this last identity is also true when $a_n \neq 0$ is independent of the rest, because then it reduces to $1^2 = 1$. ∎

THEOREM 20 (reverse stereographic projection). *For a_n dependent on the rest,*

$$\begin{aligned} &\langle \hat{a}_k^n, a_n\rangle = \langle \hat{a}_k^{n-1}, a_n\rangle/\beta_{nn}, \quad k = 1, \ldots, n-1, \\ &\langle \hat{a}_n^n, a_n\rangle = 1 - 1/\beta_{nn}, \end{aligned}$$

where $\beta_{nn} = 1 + \sum_{k=1}^{n-1} |\langle \hat{a}_k^{n-1}, a_n\rangle|^2$. These are coordinates of a point on a sphere $S^{n-1} \subset \mathbb{C}^n$ of radius $1/2$ and center $(0, \ldots, 0, 1/2)$, since

$$\sum_{k=1}^{n-1} |\langle \hat{a}_k^{n-1}, a_n\rangle/\beta_{nn}|^2 + (1/2 - 1/\beta_{nn})^2 = 1/4.$$

Proof. We reverse Theorem 17, now replacing $\beta = 1/(1 - \langle \hat{a}_n^n, a_n \rangle)$ with the formula $\beta_{nn} = 1 + \sum_{k=1}^{n-1} |\langle \hat{a}_k^{n-1}, a_n \rangle|^2$ by Lemma 19. The equation of the sphere also follows by Lemma 19, since $(1/\beta)(1 - 1/\beta) = 1/4 - (1/2 - 1/\beta_{nn})^2$. ∎

Thus a forward update (from $n-1$ to n vectors) of the combining coefficients, given by the inner products in $\mathbf{A}^* = \text{Im } A^* \subset \mathbb{C}^n$, is a reverse stereographic projection. In particular, column k and row k of the adjoint-space projector $P_{\mathbf{A}^*}$, representing generalized biorthogonality in Theorem 14, are conjugate points on the sphere of radius $1/2$ and center $\frac{1}{2}\{\delta_{jk}\}_{j=1}^n$. If a_k is independent of the rest, the points are both at the kth north pole $\{\delta_{jk}\}_{j=1}^n$.

We can now show an analogue of the Gram–Schmidt identity (Theorem 3) for protractors instead of orthogonal projectors.

COROLLARY 21. *If a_j depends on $a_1, \dots, a_{j-1}, a_{j+1}, \dots, a_n$, then*

$$Q_{\mathbf{A}} = Q_{\mathbf{A}_{(j)}} - \langle q_j, \cdot \rangle \hat{a}_j.$$

Proof. Since the operators are linear, it is sufficient to verify the identity for each primal vector a_k. Operating with both sides on a_j gives

$$\begin{aligned} Q_{\mathbf{A}} a_j &= Q_{\mathbf{A}_{(j)}} a_j - \langle q_j, a_j \rangle \hat{a}_j, \\ (1 + \langle q_j, a_j \rangle) \hat{a}_j &= q_j, \end{aligned}$$

which is true by Theorem 18(2). Operating on a_k, $k \neq j$, we have

$$\begin{aligned} Q_{\mathbf{A}} a_k &= Q_{\mathbf{A}_{(j)}} a_k - \langle q_j, a_k \rangle \hat{a}_j, \\ \hat{a}_k &= \hat{a}_{k(j)} - \langle Q_{\mathbf{A}_{(j)}} a_j, a_k \rangle \hat{a}_j, \\ \hat{a}_k &= \hat{a}_{k(j)} - \langle a_j, \hat{a}_{k(j)} \rangle \hat{a}_j, \end{aligned}$$

by Theorem 18(2) and Theorem 16(1), and by self-adjointness of Q. It is clear that both sides of the identity are 0 when they operate on any vector orthogonal to $\text{span}\{a_k\}_{k=1}^n$. ∎

5. Processes to Generate the General Dual Transform. Assembling the identities proved in the previous section, we can now construct three completely general processes for the dual transform.

Greville Process. Taken together, the recursive identities of Greville from Theorems 16(1) and 18 comprise a vector version of his method to compute the complete dual transform of n vectors [17], [18]:

<u>Greville process</u>

$\hat{a}_1^1 = a_1/|a_1|^2$
for $k = 2, \dots, n$
    $p_k = a_k - \sum_{j=1}^{k-1} \alpha_{jk} a_j$, where $\alpha_{jk} = \langle \hat{a}_j^{k-1}, a_k \rangle$
    if $\|p_k\|^2 > 0$
        $\hat{a}_k^k = p_k/\|p_k\|^2$
    else
        $\beta_{kk} = 1 + \sum_{j=1}^{k-1} |\alpha_{jk}|^2$
        $\hat{a}_k^k = q_k/\beta_{kk}$, where $q_k = \sum_{j=1}^{k-1} \alpha_{jk} \hat{a}_j^{k-1}$
    end

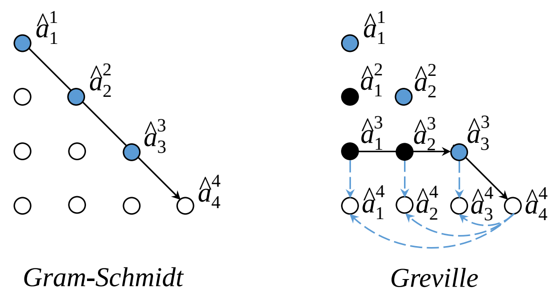


Fig. 4 *Comparison of update flows: Gram–Schmidt process and Greville process.*

  for $j = 1, \dots, k-1$
    $\hat{a}_j^k = \hat{a}_j^{k-1} - \alpha_{jk}^* \hat{a}_k^k$
  end
end

In this process, we try to get $\hat{a}_k^k$ from p_k, the projection of a_k off of (orthogonal to) the span of the preceding a_j's given by Theorem 18(1). But if a_k is dependent on some of the preceding a_j's (the projection p_k is zero), our alternative is to compute the protraction q_k to get $\hat{a}_k^k$ from Theorem 18(2). Here we can see the stereographic projection $\{\alpha_{jk}\}_{j=1}^{k-1}/\beta_{kk}$ (without its kth component $1 - 1/\beta_{kk}$ on the sphere's axis) of the coefficients $\{\alpha_{jk}\}_{j=1}^{k-1}$. Then the remaining dual transform vectors can be filled in from $\hat{a}_k^k$, using the Greville identity (Theorem 16(1)).

The update flows in the Gram–Schmidt and Greville processes are compared in Figure 4. To obtain the next $\hat{a}_k^k$, the Gram–Schmidt process combines the dual vectors on the diagonal, while the Greville process combines the original or dual vectors on the last row (solid arrows). The Greville process then updates the rest of the dual transform (dashed arrows).

The Gram–Schmidt process is often used to show that the orthogonal basis of a vector space exists. The vectors that the Greville process constructs can be shown to fulfil axioms 1 and 2 for the dual vectors. Thus, the dual of any list of vectors always exists.

Example 1 (*continued*). We saw earlier that the Gram–Schmidt process generates the Legendre polynomials from the list of polynomials $x^0, x^1, x^2, \dots, x^n$ defined over the interval $[-1, 1]$, using the inner product of the form $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,dx$. Let us apply the Greville process to get the complete set of dual polynomials of this kind:

k	$\hat{a}_1^k$	$\hat{a}_2^k$	$\hat{a}_3^k$	$\hat{a}_4^k$	$\cdots$
1	$\frac{1}{2} \cdot [\quad 1 \ ]$				
2	$\frac{3}{2} \cdot [\ \frac{1}{3}(1)$	$x\ ]$			
3	$\frac{15}{4} \cdot [\ \frac{1}{10}(3 - 5x^2)$	$\frac{2}{5}x$	$\frac{1}{2}(3x^2 - 1)\ ]$		
4	$\frac{35}{4} \cdot [\ \frac{3}{70}(3 - 5x^2)$	$-\frac{3}{14}x(7x^2 - 5)$	$\frac{3}{14}(3x^2 - 1)$	$\frac{1}{2}x(5x^2 - 3)\ ]$	
$\vdots$					$\ddots$

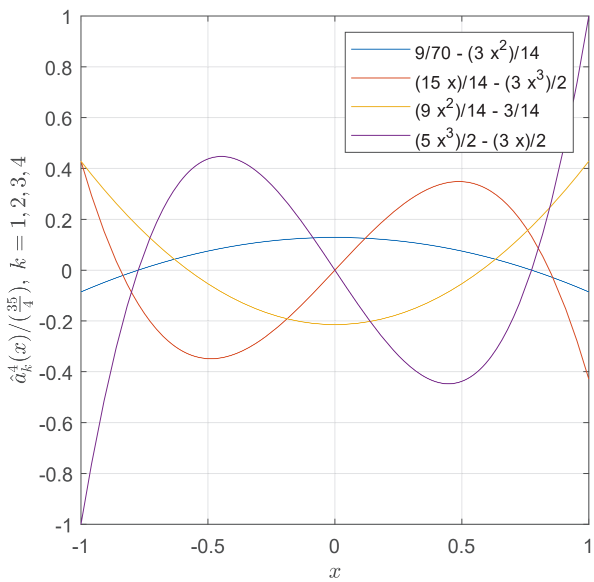


Fig. 5 *Graphs of the dual polynomials of* $1, x, x^2, x^3$.

The last dual polynomial $\hat{a}_k^k$ in every row list is the Legendre polynomial $P_{k-1}(x)$, apart from the common factor on the left side. (We kept this scale factor aside to make sure the Legendre polynomial with its usual coefficients can be plainly seen at the right end of every row.) We can keep going indefinitely, as the dots above signify. There is an interesting even-odd pattern as we go from one dual row to the next. When we add an even power of x to the list, an even function over $[-1, 1]$, only the even dual polynomials need to be updated; the odd ones stay the same. Similarly, when we add an odd power of x to the list, only the odd dual polynomials need to be updated; the even ones stay the same. The graphs of the dual polynomials in the fourth row are displayed in Figure 5. They are of degrees 2 and 3.

Example 3. There is a handy formula for the inverse of a 2×2 matrix of full rank, to wit, $\left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right]^{-1} = \left[\begin{smallmatrix} d & -c \\ -b & a \end{smallmatrix}\right]/(ad-bc)$. It is easy to verify. Is there a formula for the pseudoinverse of any 2×2 matrix A of rank 1? We can write it in general as

$$A = \begin{bmatrix} ca & da \\ cb & db \end{bmatrix} \tag{5.1}$$

for a or $b \neq 0$ and c or $d \neq 0$.

To apply the Greville process, we begin with the columns of A, $a_1 = c\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $a_2 = d\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$, as our primal list of vectors. (If $c = 0, d \neq 0$, reverse the roles of a_1, a_2.) Then our initial step is $\hat{a}_1^1 = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] / \left(c^*(|a|^2 + |b|^2)\right)$. One more step ($k = 2$) will complete the job:

$\alpha_{12} = \left(d(|a|^2 + |b|^2)\right) / \left(c(|a|^2 + |b|^2)\right) = d/c$,
$\hat{p}_2 = d\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] - (d/c) \cdot c\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] = 0$, a dependent case.
So we take the alternative:
$\beta_{22} = 1 + |\alpha_{12}|^2 = 1 + |d|^2/|c|^2$, and
$q_2 = \alpha_{12}\hat{a}_1^1 = \left(d/|c|^2\right)\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] / \left(|a|^2 + |b|^2\right)$,
$\hat{a}_2^2 = q_2/\beta_{22} = d\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] / \left(\left(|a|^2 + |b|^2\right)\left(|c|^2 + |d|^2\right)\right)$.

Finally, by the Greville identity,

$$\begin{aligned}\hat{a}_1^2 &= \hat{a}_1^1 - \alpha_{12}^* \hat{a}_2^2 \\ &= \begin{bmatrix} a \\ b \end{bmatrix} / \left(c^*(|a|^2 + |b|^2)\right) \cdot \left(1 - d^* d / \left(|c|^2 + |d|^2\right)\right) \\ &= c \begin{bmatrix} a \\ b \end{bmatrix} / \left(\left(|a|^2 + |b|^2\right)\left(|c|^2 + |d|^2\right)\right).\end{aligned}$$

Now the 2×2 pseudoinverse of a rank-1 2×2 matrix A is

$$A^+ = \left[\hat{a}_1^2 \ \hat{a}_2^2\right]^* = A^* / \left(\left(|a|^2 + |b|^2\right)\left(|c|^2 + |d|^2\right)\right). \tag{5.2}$$

For instance, let $A = \left[\begin{smallmatrix} 2 & 6 \\ 4 & 12 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1\cdot 2 & 3\cdot 2 \\ 1\cdot 4 & 3\cdot 4 \end{smallmatrix}\right]$. Then $A^+ = A^t / \left(\left(|2|^2 + |4|^2\right)\left(|1|^2 + |3|^2\right)\right) = \left[\begin{smallmatrix} 2 & 4 \\ 6 & 12 \end{smallmatrix}\right] /200$.

Example 4. The Levinson–Durbin process for linear prediction is a thinly disguised special application of the Greville process. To show this, we first introduce the linear prediction problem [22], [23], [24].

Suppose you are given a stream of data samples $\ldots, x_n, x_{n-1}, \ldots, x_3, x_2, x_1$, numbered backwards in time. How would you predict the next sample x_0 from a linear combination of the past n samples? The samples can be, for example, from a 10ms speech segment inside your cell phone, or sunspot count data with its 11 year cycle.

A good first example of how this works is the case of a pure tone, a sine wave of radian frequency ω being sampled at regular time intervals Δt. The samples have the values $x_2 = \sin\omega(k-1)\Delta t$, $x_1 = \sin\omega k\Delta t$, $x_0 = \sin\omega(k+1)\Delta t$. Applying the addition theorem for sines to $\sin\omega(k \pm 1)\Delta t$ and adding the two results, we get

$$x_0 = c_1 x_1 + c_2 x_2$$

for constant coefficients $c_1 = 2\cos\omega\Delta t$, $c_2 = -1$. If we know the frequency ω, this is an exact model that always predicts the next sample value of a sine wave from the preceding two samples.

More generally, we consider the samples to be generated by underlying random variables $\ldots, X_n, \ldots, X_2, X_1, X_0$. They belong to a linear space with an inner product of any two random variables X and Y given by their correlation,

$$\langle X, Y\rangle = \mathsf{E}(X^* Y),$$

where $\mathsf{E}(Z)$ denotes the expected value of a random variable Z. We assume the random variables are complex valued, as signal data often are. If they are real valued, just ignore the asterisks for complex conjugates.

We also assume that the data stream is stationary, in the sense that the correlations depend only on how far apart the pair of random variables are spaced in time. That is, we assume $\langle X_i, X_j\rangle = \mathsf{E}(X_i^* X_j) = r_{i-j}$. They stay the same under time translations. We can estimate them from a fair amount of sample data as $r_k = \frac{1}{M-k}\sum_{m=k}^{M}\left(x_{m-k}^* x_m\right)$ for every spacing $k = i - j$. When we correlate the random samples in reverse time order, we must take the complex conjugate of the correlation; that is, $\mathsf{E}(X_j^* X_i) = r_{j-i} = r_{i-j}^*$. For example, $r_{-3} = r_3^*$.

The inner product gives us the mean square as the norm squared, $\|X\|^2 = \langle X, X\rangle = \mathsf{E}(X^* X)$. Now we can ask for the coefficients c_j to combine the random variables X_j to predict $Y = X_0$ with least squared error

$$\left\| Y - \sum_{j=1}^{n} c_j X_j \right\|^2 . \tag{5.3}$$

From Theorem 10, we know the solution is

$$c_j = \langle \hat{X}_j^n, Y \rangle \text{ for } j = 1, 2, \ldots, n. \tag{5.4}$$

For any other data window of n consecutive samples $X_{m+n}, X_{m+n-1}, \ldots, X_{m+1}$ and $Y = X_m$, the least squared error (5.3) is the same, because when we expand the squared norm, each term is of the form $c_i^* c_j r_{i-j}$, regardless of the time translation m. Therefore, the solution (5.4) with least squared error gives us a model of order n for the entire stationary data stream. This autoregressive model can be used to estimate its frequency spectrum and for many other signal processing purposes [22], [23].

We still need to use the Greville recipe to get the dual random variables $\hat{X}_j^n$ to put into our solution (5.4). This will give us the optimum coefficients $c_1, \ldots, c_n$. We begin by working out steps $k = 1, 2, 3$ of the recipe:

Step $k = 1$:
$\hat{X}_1^1 = X_1 / \|X_1\|^2 = X_1 / r_0$,
$c_1^1 = \langle \hat{X}_1^1, Y \rangle = \langle X_1, Y \rangle / r_0 \equiv \langle X_1, X_0 \rangle / r_0 = r_1 / r_0$.

Step $k = 2$:

By symmetry for equal time gaps, for any k, $\alpha_{jk} \equiv \langle \hat{X}_j^{k-1}, X_k \rangle = \langle X_0, \hat{X}_{k-j}^{k-1} \rangle^* = \langle \hat{X}_{k-j}^{k-1}, X_0 \rangle \equiv \langle \hat{X}_{k-j}^{k-1}, Y \rangle$. We will use this equality often in these steps. Thus,
$p_2 = X_2 - \langle \hat{X}_1^1, X_2 \rangle X_1 = X_2 - \langle \hat{X}_1^1, Y \rangle X_1 = X_2 - c_1^1 X_1$,
$\|p_2\|^2 = r_0 + \frac{|r_1|^2}{r_0} - 2\frac{|r_1|^2}{r_0} = r_0 - \frac{|r_1|^2}{r_0} = r_0 - r_1 c_1^1 \geq 0$,
if $\|p_2\|^2 > 0$, $c_2^2 = \langle \hat{X}_2^2, Y \rangle = \langle p_2, Y \rangle / \|p_2\|^2 = \left(r_2 - r_1 c_1^1\right) / \left(r_0 - r_1 c_1^1\right)$.

Otherwise, $\|p_2\|^2 = 0$ if and only if $p_2 = 0$, by the definiteness property of an inner product [1]. Thus $X_2 - c_1^1 X_1 = 0$, a linear-algebraic dependency. This means that X_2 and X_1 are perfectly correlated random variables, with squared correlation coefficient $\rho^2 \equiv |\langle X_2, X_1 \rangle|^2 / \left(\|X_1\|^2 \|X_2\|^2\right) = 1 = |r_1|^2 / r_0^2$. This is equivalent to $\det R_2 = 0$ for $R_2 = [r_{i-j}]_{i,j=1}^2 = \left[\begin{smallmatrix} r_0 & r_1^* \\ r_1 & r_0 \end{smallmatrix}\right]$, a Toeplitz matrix. We can work out the Greville alternative:
$\beta_{22} = 1 + |\langle \hat{X}_1^1, X_2 \rangle|^2 = 1 + |\langle \hat{X}_1^1, Y \rangle|^2 = 1 + |c_1^1|^2$,
$\hat{X}_2^2 = q_2 / \beta_{22}$, where $q_2 = \langle \hat{X}_1^1, X_2 \rangle \hat{X}_1^1 = \langle \hat{X}_1^1, Y \rangle \hat{X}_1^1 = c_1^1 \hat{X}_1^1$,
$c_2^2 = \langle \hat{X}_2^2, Y \rangle = \langle q_2, Y \rangle / \beta_{22} = |c_1^1|^2 / \left(1 + |c_1^1|^2\right)$.

Either way, all that remains is to fill out the missing dual element by means of the Greville identity:
$\hat{X}_1^2 = \hat{X}_1^1 - \langle \hat{X}_1^1, X_2 \rangle^* \hat{X}_2^2$,
$c_1^2 = \langle \hat{X}_1^2, Y \rangle = \langle \hat{X}_1^1, Y \rangle - \langle \hat{X}_1^1, X_2 \rangle \langle \hat{X}_2^2, Y \rangle = \langle \hat{X}_1^1, Y \rangle - \langle \hat{X}_1^1, Y \rangle \langle \hat{X}_2^2, Y \rangle$,
$c_1^2 = c_1^1 - c_1^1 c_2^2$.

Step $k = 3$:

$p_3 = X_3 - \langle \hat{X}_1^2, X_3 \rangle X_1 - \langle \hat{X}_2^2, X_3 \rangle X_2,$
$p_3 = X_3 - \langle \hat{X}_2^2, Y \rangle X_1 - \langle \hat{X}_1^2, Y \rangle X_2 = X_3 - c_2^2 X_1 - c_1^2 X_2.$

Recalling that $p_3 \perp \left(\langle \hat{X}_1^2, X_3 \rangle X_1 + \langle \hat{X}_2^2, X_3 \rangle X_2 \right)$, we have
$\|p_3\|^2 = \langle X_3, p_3 \rangle = r_0 - r_2 c_1^2 - r_1 c_2^2.$
If $\|p_3\|^2 > 0$,
$c_3^3 = \langle \hat{X}_3^3, Y \rangle = \langle p_3, Y \rangle / \|p_3\|^2 = \left(r_3 - r_2 c_1^2 - r_1 c_2^2 \right) / \left(r_0 - r_2 c_1^2 - r_1 c_2^2 \right).$

Otherwise, $p_3 = 0 = X_3 - c_2^2 X_1 - c_1^2 X_2$.
Then $0 = \langle X_3 - c_1^2 X_2 - c_2^2 X_1,\ X_j \rangle,\ j = 3, 2, 1$:
$0 = r_0 - c_1^2 r_1^* - c_2^2 r_2^*,$
$0 = r_1 - c_1^2 r_0 - c_2^2 r_1^*,$
$0 = r_2 - c_1^2 r_1 - c_2^2 r_0.$
This is a linear dependency among the columns of the Toeplitz matrix

$$R_3 = [r_{i-j}]_{i,j=1}^3 = \begin{bmatrix} r_0 & r_1^* & r_2^* \\ r_1 & r_0 & r_1^* \\ r_2 & r_1 & r_0 \end{bmatrix},$$

so that $\det R_3 = 0$.

Following the Greville alternative formula,
$\beta_{33} = 1 + |\langle \hat{X}_1^2, X_3 \rangle|^2 + |\langle \hat{X}_2^2, X_3 \rangle|^2 = 1 + |\langle \hat{X}_1^2, Y \rangle|^2 + |\langle \hat{X}_2^2, Y \rangle|^2,$
$\hat{X}_3^3 = q_3 / \beta_{33}$, where $q_3 = \langle \hat{X}_1^2, Y \rangle \hat{X}_1^2 + \langle \hat{X}_2^2, Y \rangle \hat{X}_2^2 = c_1^2 \hat{X}_1^2 + c_2^2 \hat{X}_2^2,$
$c_3^3 = \langle \hat{X}_3^3, Y \rangle = \langle q_3, Y \rangle / \beta_{33} = \left(|c_1^2|^2 + |c_2^2|^2 \right) / \left(1 + |c_1^2|^2 + |c_2^2|^2 \right).$

Again we fill in the rest of the dual elements by means of the Greville identity:
$\hat{X}_j^3 = \hat{X}_j^2 - \langle \hat{X}_j^2, X_j \rangle^* \hat{X}_3^3,\ j = 1, 2,$
$c_j^3 = \langle \hat{X}_j^3, Y \rangle = \langle \hat{X}_j^2, Y \rangle - \langle \hat{X}_j^2, X_j \rangle \langle \hat{X}_3^3, Y \rangle = \langle \hat{X}_j^2, Y \rangle - \langle \hat{X}_{3-j}^2, Y \rangle \langle \hat{X}_3^3, Y \rangle,$
$c_j^3 = c_j^2 - c_{3-j}^2 c_3^3,\ j = 1, 2.$

Our reasoning for Step 3 extends directly to any Step k. This allows us to rewrite the Greville process to solve for the optimum coefficients of the linear prediction model (5.3) for a stationary time series of random variables, as follows:

Levinson process (general version)

$c_1^1 = r_1 / r_0$
for $k = 2, \dots, n$
  $\|p_k\|^2 = r_0 - r_{k-1} c_1^{k-1} - \cdots - r_1 c_{k-1}^{k-1}$
  if $\|p_k\|^2 > 0$
    $c_k^k = \left(r_k - \sum_{j=1}^{k-1} r_{k-j} c_j^{k-1} \right) / \left(r_0 - \sum_{j=1}^{k-1} r_{k-j} c_j^{k-1} \right)$
  else
    $c_k^k = \left(|c_1^{k-1}|^2 + \cdots + |c_{k-1}^{k-1}|^2 \right) / \left(1 + |c_1^{k-1}|^2 + \cdots + |c_{k-1}^{k-1}|^2 \right)$
  end
  for $j = 1, \dots, k-1$
    $c_j^k = c_j^{k-1} - c_{k-j}^{k-1} c_k^k$
  end
end

The Greville process directly yields the complete Levinson process. The alternative formula for the "else" case above patches a hole in the process and appears to be something new.

Generalized Gram–Schmidt Process. We saw that, at each step, the Gram–Schmidt process computes the last vector of the dual transform of the current list of independent vectors. Interpreting it this way, it is natural to ask if it can be extended to cases with dependent vectors. Here is a complete Gram–Schmidt process.

Gram–Schmidt process (general version)

$\hat{a}_1^1 = a_1/\|a_1\|^2$
$\beta_{11} = 1$
for $k = 2, \ldots, n$
  $p_k = a_k - \sum_{j=1}^{k-1} \tilde{\alpha}_{jk} \|p_j\|^2 \hat{a}_j^j$, where $\tilde{\alpha}_{jk} = \langle \hat{a}_j^j, a_k \rangle$

  for $j = 1, \ldots, k$
    $\beta_{jk} = \delta_{jk} - \sum_{i=1}^{k-1} \beta_{ij}^* \tilde{\alpha}_{ik}$
    $\beta_{kj} = \beta_{jk}^*$
  end

  if $\|p_k\|^2 > 0$
    $\hat{a}_k^k = p_k/\|p_k\|^2$
  else
    $\hat{a}_k^k = -\sum_{j=1}^{k-1} \beta_{jk} \hat{a}_j^j / \beta_{kk}$
  endif
end

This process has its original expression for p_k and an alternative for dependent vectors that is derived in Appendix A. The expression for the projection p_k (related to the dual vector by Theorem 18(1)) follows, as for the original Gram–Schmidt process, by induction from the Gram–Schmidt identity (Theorem 3): $P_{\mathbf{A}_{k-1}^\perp} = P_{\mathbf{A}_{k-2}^\perp} - P_{p_{k-1}}$. If $p_{k-1} = 0$, then $P_{\mathbf{A}_{k-1}^\perp} = P_{\mathbf{A}_{k-2}^\perp}$, and we can omit the contribution of $\hat{a}_{k-1}^{k-1}$ to the projector.

This process is also reversible. We recover the primal vectors by running the process again on the dual vectors. This general process still produces an orthogonal basis of a sequence of vectors. The basis vectors are just the dual vectors computed without the alternative, that is, those whose "signature" $\|p_k\|^2$ is positive. Each alternative dual vector is not orthogonal to the previous basis vectors, but is a linear combination of them. Counting the basis vectors reveals the dimension of the subspace they span.

Parallel Butterfly Dual Process. The butterfly dual process at the end of section 2 can be extended to a general version that will handle dependencies among the a_k's. When $\gamma = 0$, the alternative is a weighted sum of the middle dual vectors in one of the parent nodes. The alternative vector itself can be updated recursively, without computing the middle vectors. This is shown in Appendix B. Each processing node now contains a left and a right alternative vector q, each with a supporting scalar, as well as the original (primal) and dual vectors on the left and right ends. We keep the alternative updated at every node, so that it is ready to use whenever a new dependency forms. The general butterfly dual process, for any given list of vectors, now follows:

Butterfly dual process (general version)

```
for node j = 1, ..., n                              ! initialize the first level
    β_ℓ = β_r = 1
    q_ℓ = q_r = 0
    â_ℓ = â_r = a_j/||a_j||^2
end
for level k = 2, ..., n
    for node j = 1, ..., n
        α_ℓ^L = ⟨â_ℓ^L, a_r⟩
        α_r^R = ⟨â_r^R, a_l⟩
        η = ⟨q_ℓ^L, a_r⟩ = ⟨q_r^R, a_ℓ⟩*
        β_ℓ = β_ℓ^L + |α_r^R|^2 β_r^R − 2Re(α_r^R η)
        β_r = β_r^R + |α_ℓ^L|^2 β_ℓ^L − 2Re(α_ℓ^L η*)
        q_ℓ = q_ℓ^L − α_r^R q_r^R + (α_r^R β_r^R − η*) â_r^R
        q_r = q_r^R − α_ℓ^L q_ℓ^L + (α_ℓ^L β_ℓ^L − η) â_ℓ^L
        γ = 1 − α_ℓ^{L*} α_r^{R*}
        if γ > 0
            â_ℓ = γ^{-1} (â_ℓ^L − α_ℓ^{L*} â_r^R)     ! butterfly for left,right dual vectors
            â_r = γ^{-1} (â_r^R − α_r^{R*} â_ℓ^L)
        else
            â_ℓ = q_ℓ/β_ℓ                          ! alternative
            â_r = q_r/β_r
        endif
    end
end
```

This process was briefly reported, without supporting theory, in [7]. Since the lines of communication from parent to child processing nodes remain the same from level to level, the algebraically exact butterfly dual process can also be implemented recursively using one ring of n parallel processors n times over.

THEOREM 22. *γ lies in the unit interval $[0,1]$, and $\gamma = 0$ if and only if a new dependency occurs when left and right vectors a_ℓ and a_r are both added to the middle list at a node $\{a_{\ell+1}, \ldots, a_{r-1}\}$.*

Proof. The butterfly update coefficient in equations (2.16) is $\gamma = 1 - \alpha_\ell^{L*}\alpha_r^{R*}$ for inner products $\alpha_\ell^L = \langle \hat{a}_\ell^L, a_r\rangle$ and $\alpha_r^R = \langle \hat{a}_r^R, a_\ell\rangle$. The parent nodes each contain a common set of vectors, their middle $\{a_{\ell+1}, \ldots, a_{r-1}\}$, and differ only by one outer vector (the left vector a_ℓ of the left parent node and the right vector a_r of the right parent node). We will distinguish four cases in terms of dependencies between the outer vectors and the middle set. Let $\mathbf{A}_M = \operatorname{span}\{a_{\ell+1}, \ldots, a_{r-1}\}$.

The first case is when each parent outer vector is independent of the middle: $(a_\ell \notin \mathbf{A}_M,\ \mathbf{A}_M \not\ni a_r)$. Here $0 \le \gamma \le 1$ or, equivalently, $0 \le \alpha_\ell^L \alpha_r^R \le 1$. This follows by the Cauchy–Schwarz inequality, since by Theorem 18(1), the left and right end dual vectors are $\hat{a}_\ell^L = P_{\mathbf{A}_M^\perp} a_\ell / \|P_{\mathbf{A}_M^\perp} a_\ell\|^2$ and $\hat{a}_r^R = P_{\mathbf{A}_M^\perp} a_r / \|P_{\mathbf{A}_M^\perp} a_r\|^2$. In particular,

$\gamma = 0$ if and only if $P_{\mathbf{A}_M^\perp} a_\ell \propto P_{\mathbf{A}_M^\perp} a_r$, which is true if and only if ($a_\ell \notin \mathbf{A}_M$, $\mathbf{A}_M \not\ni a_r$, but $\operatorname{span}\{a_\ell, \mathbf{A}_M\} \ni a_r$). In other words, $\gamma = 0$ if and only if a new dependency occurs.

Next consider the case ($a_\ell \notin \mathbf{A}_M$, $\mathbf{A}_M \ni a_r$). Here $\hat{a}_\ell^L \perp \{a_{\ell+1}^L, \ldots, a_{r-1}^L\}$ and so $\hat{a}_\ell^L \perp a_r$, making $\gamma = 1$. By symmetry, $\gamma = 1$ for the case when ($a_\ell \in \mathbf{A}_M$, $\mathbf{A}_M \not\ni a_r$), too.

Last, take the case ($a_\ell \in \mathbf{A}_M$, $\mathbf{A}_M \ni a_r$). Because a_r depends on the middle set, the end Greville update for the right parent node (from the middle grandparent node, which has just the middle vectors $\mathbf{A}_M$) degenerates to a dependency $a_r = \sum_{j=\ell+1}^{r-1} \alpha_j^{(M,r)} a_j$, with alternative $\hat{a}_r^R = \frac{1}{\beta^{(M,r)}} \sum_{j=\ell+1}^{r-1} \alpha_j^{(M,r)} \hat{a}_j^M$, with $\alpha_j^{(M,r)} = \langle \hat{a}_j^M, a_r \rangle$. We also have the similar formulas for the left parent node, also updated from the middle grandparent node. So $\alpha_r^R = \langle \hat{a}_r^R, a_\ell \rangle = \frac{1}{\beta^{(M,r)}} \sum_{j=\ell+1}^{r-1} \alpha_j^{(M,r)*} \alpha_j^{(\ell,M)}$, and then

$$\begin{aligned} \alpha_\ell^L \alpha_r^R &= \frac{1}{\beta^{(\ell,M)}} \frac{1}{\beta^{(M,r)}} \left| \sum_{j=\ell+1}^{r-1} \alpha_j^{(M,r)*} \alpha_j^{(\ell,M)} \right|^2 \\ &< \frac{1}{\beta^{(\ell,M)} - 1} \frac{1}{\beta^{(M,r)} - 1} \left| \sum_{j=\ell+1}^{r-1} \alpha_j^{(M,r)*} \alpha_j^{(\ell,M)} \right|^2 \le 1, \end{aligned}$$

from the definition of the β's and the Cauchy–Schwarz inequality. Therefore $0 < \gamma \le 1$. When the left and right outer vectors depend on disjoint subsets of the middle vectors, then for each j, either $\alpha_j^{(M,r)}$ or $\alpha_j^{(\ell,M)}$ is zero (or both are); so in this subcase, $\gamma = 1$. ∎

Appendix A. Gram–Schmidt Alternative Updates. This appendix gives an alternative for the Gram–Schmidt process in the case of dependent vectors. It is equivalent to the alternative taken in the Greville process, but only uses the end dual vectors $\hat{a}_k^k$ at each level k of Figure 4. The identities below provide the alternative updates.

We introduce coefficients

$$\beta_{jk} = \sum_{i=1}^{\min(j,k)} \alpha_{ij}^* \alpha_{ik}$$

for all $j, k = 1, \ldots, n$, where $\alpha_{jk} = \langle \hat{a}_j^{k-1}, a_k \rangle$ for $j < k$, $\alpha_{kk} = -1$, and $\alpha_{jk} = 0$ for $j > k$. Note that from this definition, $\beta_{11} = 1$ and $\beta_{jk}^* = \beta_{kj}$. (In matrix terms, we have $B = [\beta_{jk}] = U^* U$ for the upper triangular matrix $U = [\alpha_{jk}]$.) Then we have these vector identities.

LEMMA 23. *For $j, k \ge 1$,*

$$\sum_{i=1}^{k} \beta_{ij} \hat{a}_i^i = - \sum_{i=1}^{\min(j,k)} \alpha_{ij} \hat{a}_i^k.$$

Proof by induction. The identities for $k = 1$ are easily checked. Assuming the identities are true for $k - 1$, we show they must be true for k. The left-hand side is $\sum_{i=1}^{k} \beta_{ij} \hat{a}_i^i = \sum_{i=1}^{k-1} \beta_{ij} \hat{a}_i^i + \beta_{kj} \hat{a}_k^k = -\sum_{i=1}^{\min(j,k-1)} \alpha_{ij} \hat{a}_i^{k-1} + \sum_{i=1}^{\min(j,k)} \alpha_{ik}^* \alpha_{ij} \hat{a}_k^k$ (by the induction hypothesis and the definition of β_{kj}) $= -\sum_{i=1}^{\min(j,k)} \alpha_{ij} \hat{a}_i^k$ (by the Greville identity). This completes the induction step. ∎

Taking the inner product of a_k with both sides of the identities for $k-1$ in Lemma 23 immediately yields a scalar identity.

LEMMA 24. *For* $j, k \geq 1$,

$$\sum_{i=1}^{k} \beta_{ij}^* \tilde{\alpha}_{ik} = -\alpha_{jk}^*,$$

where $\tilde{\alpha}_{ik} = \langle \hat{a}_i^i, a_k \rangle$ *for* $i < k$, $\tilde{\alpha}_{kk} = 1$, *and* $\tilde{\alpha}_{ik} = 0$ *for* $i > k$. *A fine point to note here is that* $\tilde{\alpha}_{kk} = 1$ *is an update coefficient used before we know* $\hat{a}_k^k$. *From the two cases of Theorem* 18, *we have* $\langle \hat{a}_k^k, a_k \rangle = 1$ *for the first case, but* $\langle \hat{a}_k^k, a_k \rangle = 1 - \frac{1}{\beta_{kk}}$ *for the alternative case.*

From their definitions, we have $\alpha_{k-1,k} = \tilde{\alpha}_{k-1,k}$ for every $k \geq 2$. The next identity relates the α's and $\tilde{\alpha}$'s in general.

LEMMA 25. *For* $j, k \geq 1$,

$$-\sum_{i=j}^{k} \alpha_{ji} \tilde{\alpha}_{ik} = \delta_{jk}.$$

Proof by induction. The cases for $j \geq k$ and $j = k-1$ are easy to check by definition. These include the cases for $k = 1, 2$. For $k = 3$, we have $\alpha_{13} = \langle \hat{a}_1^2, a_3 \rangle$. In this inner product, substitute for $\hat{a}_1^2$ with the Greville identity $\hat{a}_1^2 = \hat{a}_1^1 - \alpha_{12}^* \hat{a}_2^2$. This gives $\alpha_{13} = \tilde{\alpha}_{13} - \alpha_{12}\tilde{\alpha}_{23}$. Similarly, for the $j = k-2$ case, we have in general that $\alpha_{k-2,k} = \tilde{\alpha}_{k-2,k} - \alpha_{k-2,k-1}\tilde{\alpha}_{k-1,k}$.

Next we suppose the lemma is true for $k-1$. Then for the $k-1$ vectors $a_1, a_2, \ldots, a_{k-2}, a_k$, we have, for $j = 1, \ldots, k-3$,

$$\langle \hat{a}_j^{k-2}, a_k \rangle = \langle \hat{a}_j^j, a_k \rangle - \sum_{i=j+1}^{k-2} \langle \hat{a}_j^{i-1}, a_i \rangle \langle \hat{a}_i^i, a_k \rangle.$$

Using the Greville identity, $\hat{a}_j^{k-1} = \hat{a}_j^{k-2} - \langle \hat{a}_j^{k-2}, a_{k-1} \rangle^* \hat{a}_{k-1}^{k-1}$, in the left-hand inner product gives

$$\langle \hat{a}_j^{k-1}, a_k \rangle = \langle \hat{a}_j^j, a_k \rangle - \sum_{i=j+1}^{k-2} \langle \hat{a}_j^{i-1}, a_i \rangle \langle \hat{a}_i^i, a_k \rangle - \langle \hat{a}_j^{k-2}, a_{k-1} \rangle \langle \hat{a}_{k-1}^{k-1}, a_k \rangle.$$

This is the lemma for k and $j = 1, \ldots, k-3$. We observed earlier that the cases for $j = k-2$, $k-1$, and k are true, so the induction step is complete. ∎

Lemma 25 is a biorthogonal relation, and Lemma 24 is a "half-biorthogonal" relation, since its right side equals δ_{jk} for $j \leq k$. To put these lemmata in matrix terms, define another upper triangular matrix $\tilde{U} = [\tilde{\alpha}_{jk}]$. Then Lemma 25 is $U\tilde{U} = -I$. Since $B = U^*U$, we confirm Lemma 24 again: $B\tilde{U} = -U^*$.

Appendix B. Alternative Butterfly Updates. We saw in Theorem 22 that $\gamma = 0$ is one of the cases where a_r depends linearly on the rest of the vectors, $\{a_\ell, \ldots, a_{r-1}\}$. The Greville alternative (Theorem 18(2)) applies to these cases. This alternative (without its scalar $1/\beta_r$) is $q_r = \sum_{j=\ell}^{r-1} \alpha_j^L \hat{a}_j^L$. To avoid computing this growing sum at junctures when $\gamma = 0$, as well as to avoid storing and updating the middle dual vectors, we will now derive a vector butterfly identity to recursively update the alternative sums. We will also derive a scalar butterfly identity to update the scalars $\beta_r = 1 + \sum_{j=\ell}^{r-1} |\alpha_j^L|^2$. We will also need updates for the left sum $q_\ell = \sum_{j=\ell+1}^{r} \alpha_j^R \hat{a}_j$ and its scalar β_ℓ.

First, we use the Greville identity to relate the coefficients in these two sums back to the coefficients from the middle grandparent node (M). For the left parent node (L), we have $\alpha_j^L = \langle \hat{a}_j^L, a_r \rangle = \langle a_j^M - \alpha_j^{(\ell,M)*} \hat{a}_\ell^L, a_r \rangle$, so that for $j = (\ell+1), \ldots, (r-1)$,

$$\alpha_j^L = \alpha_j^{(M,r)} - \alpha_j^{(\ell,M)} \alpha_\ell^L. \tag{B.1}$$

Similarly, for the right node (R) and $j = (\ell+1), \ldots, (r-1)$,

$$\alpha_j^R = \alpha_j^{(\ell,M)} - \alpha_j^{(M,r)} \alpha_r^R. \tag{B.2}$$

Using the left coefficient relations (B.1), we find

$$\begin{aligned} q_r &= \alpha_\ell^L \hat{a}_\ell^L + \sum_{j=\ell+1}^{r-1} \alpha_j^L \hat{a}_j \\ &= \sum \alpha_j^{(M,r)} \hat{a}_j^M - \alpha_\ell^L \sum \alpha_j^{(\ell,M)} \hat{a}_j^M \\ &\quad + \left(\alpha_\ell^L \left[1 + \sum \alpha_j^{(\ell,M)*} \alpha_j^{(\ell,M)} \right] - \sum \alpha_j^{(\ell,M)*} \alpha_j^{(M,r)} \right) \hat{a}_\ell^L, \end{aligned} \tag{B.3}$$

or

$$q_r = q_r^R - \alpha_\ell^L q_\ell^L + \left(\alpha_\ell^L \beta_\ell^L - \eta \right) \hat{a}_\ell^L, \tag{B.4}$$

where η is the cross-term $\eta = \sum_{j=\ell+1}^{r-1} \alpha_j^{(\ell,M)*} \alpha_j^{(M,r)}$. Likewise, from (B.2), there is also a left sum update,

$$q_\ell = q_\ell^L - \alpha_r^R q_r^R + \left(\alpha_r^R \beta_r^R - \eta^* \right) \hat{a}_r^R. \tag{B.5}$$

To update the scalars β_ℓ and β_r, first we use (B.1) once more:

$$\begin{aligned} \beta_r &= 1 + |\alpha_\ell^L|^2 + \sum_{j=\ell+1}^{r-1} |\alpha_j^L|^2 \\ &= 1 + |\alpha_\ell^L|^2 + \sum |\alpha_j^{(M,r)} - \alpha_j^{(\ell,M)} \alpha_\ell^L|^2 \\ &= \left(1 + \sum |\alpha_j^{(M,r)}|^2 \right) + |\alpha_\ell^L|^2 \left(1 + \sum |\alpha_j^{(\ell,M)}|^2 \right) - 2\mathrm{Re} \left(\alpha_\ell^{L*} \sum \alpha_j^{(\ell,M)*} \alpha_j^{(M,r)} \right) \\ &= \beta_r^R + |\alpha_\ell^L|^2 \beta_\ell^L - 2\mathrm{Re} \left(\alpha_\ell^{L*} \eta \right). \end{aligned} \tag{B.6}$$

Similarly, there is

$$\beta_\ell = \beta_\ell^L + |\alpha_r^R|^2 \beta_r^R - 2\mathrm{Re} \left(\alpha_r^R \eta \right). \tag{B.7}$$

The only thing missing is a way to compute the cross-term η:

$$\begin{aligned} \eta &= \sum_{j=\ell+1}^{r-1} \alpha_j^{(\ell,M)*} \alpha_j^{(M,r)} = \sum \langle \hat{a}_j^M, a_\ell \rangle \langle \hat{a}_j^M, a_r \rangle \\ &= \left(\sum \langle \hat{a}_j^M, a_\ell \rangle \hat{a}_j^M \right)^* \cdot \left(\sum \langle \hat{a}_j^{(M,r)}, a_r \rangle a_j \right) \\ &= \left(\sum \langle \hat{a}_j^M, a_\ell \rangle \hat{a}_j^M \right)^* \cdot P_{\mathbf{A}_M} a_r \\ &= \langle q_\ell^L, a_r \rangle, \end{aligned} \tag{B.8}$$

where we have applied the fact that $P_{\mathbf{A}_M}\hat{a}_j^M = \hat{a}_j^M$. Similarly, $\eta^* = \langle q_r^R, a_\ell \rangle$. These identities ((B.4), (B.5), (B.6), (B.7), (B.8)) provide a complete update for the q's and β's.

Acknowledgments. My thanks go to Paul Swett for introducing me to formula (2.4) and to John Whelchel and Robert Schreiber for their early encouragement of the parallel process. I also thank the reviewer for comments that made the paper better and more reader-friendly.

REFERENCES

[1] S. Axler, *Linear Algebra Done Right*, 3rd ed., Springer, Berlin, 2015. (Cited on pp. 1032, 1033, 1034, 1035, 1037, 1054)

[2] G. Strang, *Introduction to Linear Algebra*, 5th ed., Wellesley-Cambridge Press, Cambridge, MA, 2016. (Cited on pp. 1032, 1033, 1034, 1037, 1040, 1042)

[3] I. Gohberg, S. Goldberg, and M. A. Kaashoek, *Basic Classes of Linear Operators*, Birkhäuser, Boston, 2003. (Cited on pp. 1032, 1033, 1034, 1036, 1042)

[4] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Boston, 2003. (Cited on p. 1033)

[5] G. D. Mahan, *Condensed Matter in a Nutshell*, Princeton University Press, 2011. (Cited on p. 1033)

[6] S. G. Glisic, *Advanced Wireless Communications: 4G Technologies*, Wiley, New York, 2009. (Cited on p. 1033)

[7] L. P. Withers, Jr., *A parallel algorithm for generalized inverses of matrices, with applications to optimum beamforming*, in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing I, 1993, pp. 369–372. (Cited on pp. 1033, 1042, 1057)

[8] J. B. Hartle, *Gravity: An Introduction to Einstein's Theory of Relativity*, Addison-Wesley, San Francisco, 2003. (Cited on p. 1033)

[9] K. S. Thorne and R. D. Blandford, *Modern Classical Physics*, Princeton University Press, 2017. (Cited on p. 1033)

[10] I. Ziskind and M. Wax, *Maximum likelihood localization of multiple sources by alternating projection*, IEEE Trans. Acoust. Speech Signal Process., 36 (1988), pp. 1553–1560. (Cited on p. 1035)

[11] D. J. Griffiths and D. F. Schroeter, *Introduction to Quantum Mechanics*, 3rd ed., Cambridge University Press, 2018. (Cited on p. 1037)

[12] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996, https://doi.org/10.1137/1.9781611971484. (Cited on p. 1042)

[13] E. H. Moore, *On the reciprocal of the general algebraic matrix*, Bull. Amer. Math. Soc., 26 (1920), pp. 394–395. (Cited on p. 1041)

[14] E. H. Moore, *General analysis*, Part 1, Mem. Amer. Phil. Soc., 1 (1935), pp. 197–209. (Cited on p. 1041)

[15] A. Bjerhammar, *Rectangular reciprocal matrices with special reference to geodetic calculations*, Bull. Géodésique, 52 (1951), pp. 188–220. (Cited on p. 1041)

[16] R. Penrose, *A generalized inverse for matrices*, Proc. Cambridge Phil. Soc., 51 (1955), pp. 406–413. (Cited on p. 1041)

[17] T. N. E. Greville, *Some applications of the pseudoinverse of a matrix*, SIAM Rev., 2 (1960), pp. 15–22, https://doi.org/10.1137/1002004. (Cited on pp. 1046, 1048, 1050)

[18] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, Springer, Berlin, 2003. (Cited on p. 1050)

[19] S. L. Campbell and C. D. Meyer, *Generalized Inverses of Linear Transformations*, SIAM, Philadelphia, 2009, https://doi.org/10.1137/1.9780898719048. (Cited on p. 1042)

[20] J. Munkres, *Topology*, 2nd ed. (Classic Version), Pearson, London, 2017. (Cited on p. 1047)

[21] K. Knopp, *Elements of the Theory of Functions*, Dover, New York, 2016. (Cited on p. 1047)

[22] S. M. Kay, *Modern Spectral Estimation*, Prentice-Hall, Englewood Cliffs, NJ, 1988. (Cited on pp. 1053, 1054)

[23] S. M. Kay, *Intuitive Probability and Random Processes Using MATLAB*, Springer, New York, 2006. (Cited on pp. 1053, 1054)

[24] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, 2007 (Cited on p. 1053)