Chapter III. Dual Spaces

III.1. Contours, Covariance, Contravariance, Dual Basis

Note. In this section X and Y denote finite dimensional real vector spaces. We take X to be n-dimensional and Y to be m-dimensional. Therefore, by the Fundamental Theorem of Finite Dimensional Spaces we have $X \cong \mathbb{R}^n$ and $Y \cong \mathbb{R}^m$ (see my online notes on "3.3. Coordinatization of Vectors" for sophomore Linear Algebra [MATH 2010]).

Note/Definition. The set of all linear maps from X to Y, denoted $L(X, Y)$, is itself a vector space. When $Y = \mathbb{R}$, the elements of $L(X, \mathbb{R})$ are called *linear functionals* on X (or *dual vectors* or *covariant vectors*). By contrast, the vectors in X are called *contravariant vectors*. Vector space $L(X, \mathbb{R})$ is the *dual space* of X, denoted X^*.

Note. Dodson and Poston sing the praises of geometrically interpreting the elements of X for which a functional is constant as a level contour (line, surface, etc.; see page 57). When a nonzero functional on an n-dimensional space is set equal to a constant, the subset of X satisfying the equation is an $n-1$ dimensional subspace (i.e., an "affine hyperplane").

Note/Definition. For $\mathbf{A} \in L(X, Y)$ and $\mathbf{f} \in Y^*$ we have $\mathbf{A} : X \to Y$ and $\mathbf{f} : Y \to \mathbb{R}$, so we can define $\mathbf{f} \circ \mathbf{A} : X \to \mathbb{R}$ where $\mathbf{f} \circ \mathbf{A} \in X^*$. So we define the *dual map* of $\mathbf{A} \in L(X, Y)$ as $\mathbf{A}^* \in L(Y^*, X^*)$ by $\mathbf{A}^*(\mathbf{f}) = \mathbf{f} \circ \mathbf{A}$ for $\mathbf{f} \in Y^*$.

Note. We know that X^* is itself a vector space. We are interested in finding a basis for X^* and finding a way to change bases of X^* (as they relate to changes in bases of X). First, we need to consider the dimension of X and X^*.

Lemma III.1.04. Let X be an n-dimensional real vector space with dual space X^*. Then $\dim(X^*) = \dim(X)$.

Definition. For n-dimensional vector space X with basis $\beta = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$, the basis $\beta^* = \{\mathbf{b}^1, \mathbf{b}^2, \ldots, \mathbf{b}^n\}$ (where $\mathbf{b}^j : X \to \mathbf{R}$ is defined by mapping $\mathbf{x} = a^1\mathbf{b}_1 + a^2\mathbf{b}_2 + \cdots + a^n\mathbf{b}_n \mapsto a^j$) is the *dual basis* of β.

Note.III.1.A. With $\mathbf{b}^i \in \beta^*$ and $\mathbf{b}_j \in \beta$ we have

$$\mathbf{b}^i\mathbf{b}_j = \mathbf{b}^i(0\mathbf{b}_1 + 0\mathbf{b}_2 + \cdots + 0\mathbf{b}_{j-1} + 1\mathbf{b}_j + 0\mathbf{b}_{j+1} + \cdots + 0\mathbf{b}_n) = \delta^i_j = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j. \end{cases}$$

Lemma III.1.A. Given a linear functional $\mathbf{f} \in X^*$ where $X = \mathbb{R}$, there is $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{f}(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle$ (the inner product on $\mathbb{R}^n$), and conversely for each $\mathbf{y} \in \mathbb{R}^n$ the mapping $\mathbf{x} \mapsto \langle \mathbf{x}, \mathbf{y} \rangle$ is a linear functional in X^*. That is, X^* is isomorphic to $\mathbb{R}^n$ when $X = \mathbb{R}^n$.

Note. Dodson and Poston say "It is tempting to identify X^* and X... However, this has great disadvantages because the isomorphism depends very much on the choice of bases." See page 59. As seen in Lemma III.1.A, for $X = \mathbb{R}^n$ we have $X^* = \mathbb{R}^n$. However, in the proof of Lemma III.1.A we used inner products to represent linear functionals. In this section we want to stay in the setting of vectors and matrices. In this case, we can take $X = \mathbb{R}^n$ where the elements of X are *column* vectors and then $X^* \cong \mathbb{R}^n$ where the elements of X^* are *row* vectors. then the matrix product $\mathbf{yx}$, where $\mathbf{y} \in X^*$ is $1 \times n$ and $\mathbf{x} \in X$ is $n \times 1$, yields the scalar $\langle \mathbf{y}, \mathbf{x}^t \rangle$ (or technically the 1×1 matrix containing $\langle \mathbf{y}, \mathbf{x}^t \rangle$). Here we use "$t$" the represent the transpose of a matrix, since later we will use "T" to represent that adjoint of a linear operator.

Note. In infinite dimensional vector spaces, it may not be the case that $X \cong A^*$. For example, if $X = L^p(E)$ where $1 \leq p < \infty$ (the classical Banach spaces on Lebesgue measurable set E: $L^p(E) = \{f$ measurable function on $E \mid \int_E |f|^p < \infty\}$) ten $X^* \cong L^q(E)$ where $1/P + 1/q = 1$ (and $p = 1$ implies $q = \infty$), But $L^p(E) \not\cong L^q(E)$ unless $p = q = 2$. In fact, X^* is not *equal* to $L^q(E)$ in this case but instead X^* is *isometrically isomorphic* to $L^q(E)$. See my Real Analysis (MATH 5210/5220) online notes on 8.1. The Riesz Representation for the Dual of L^p, $1 \leq p < \infty$ and 19. General L^p Spaces: Completeness, Duality, and Weak Convergence.

Note III.1.A. Let X have basis $\beta = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ and X^* have basis $\beta^* = \{\mathbf{b}^1, \mathbf{b}^2, \ldots, \mathbf{b}^n\}$ where β^* is the dual basis of β. Then for $\mathbf{x} \in X$, say $\mathbf{x} = x^1\mathbf{b}_1 + x^2\mathbf{b}_2 + \cdots + x^n\mathbf{b}_n = (x^1, x^2, \ldots, x^n)$, and $\mathbf{f} \in X^*$, say $\mathbf{f} = f_1\mathbf{b}^1 + f_2\mathbf{b}^2 + \cdots + f_n\mathbf{b}^n = (f_1, f_2, \ldots, f_n)$ (here we use coordinate vectors relative to the ordered bases β and β^*), then we have

$$\begin{aligned}
\mathbf{f}(\mathbf{x}) &= \mathbf{f}(x^1\mathbf{b}_1 + x^2\mathbf{b}_2 + \cdots + x^n\mathbf{b}_n) \text{ by the definition of } \mathbf{x} \\
&= \mathbf{f}(x^j\mathbf{b}_j) \text{ by the Einstein summation convention} \\
&= (f_1\mathbf{b}^1 + f_2\mathbf{b}^2 + \cdots + f_n\mathbf{b}^n)(x^j\mathbf{b}_j) \text{ by the definition of } \mathbf{f} \\
&= f_i x^j(\mathbf{b}^i\mathbf{b}_j) \text{ by commutivity of scalar multiplication} \\
&= f_i x^j \delta^i_j \text{ by Note III.1.A} \\
&= f_i x^i.
\end{aligned}$$

So similar to Lemma III.1.A we see that a linear function on $X = \mathbb{R}^n$ is represented by an inner product (here, the inner product of the coordinate vectors of $\mathbf{f}$ and $\mathbf{x}$ with respect to the ordered bases β^* and β, respectively).

Theorem III.1.A. Let $\beta = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for X and $\beta' = \{\mathbf{b}'_1, \mathbf{b}'_2, \ldots, \mathbf{b}'_m\}$ be a basis for Y such that the $m \times n$ matrix $A = [\mathbf{A}]^{\beta'}_{\beta}$ represents a linear transformation from X to Y with respect to ordered bases β and β'. Let β^* and β'^* be the dual bases of X^* and Y^*, respectively. Then the $n \times m$ matrix $A^* = [\mathbf{A}^*]^{\beta^*}_{\beta'^*} = \left([\mathbf{A}]^{\beta'}_{\beta}\right)^t$, where t represents the transpose operator on a matrix.

Note. As seen in the proof of Theorem III.1.A, in order to represent a linear transformation from X to Y, or a linear transformation from Y^* to X^*, as a matrix we need to represent elements of X or elements of Y^*, respectively, as column vectors (with respect to ordered bases β of X and β'^* of Y^*, respectively) when multiplying these vectors on the left by the matrices. Notice that Dodson and Poston have the matrices on the left but still use row vectors; see page 60.

Note. If β and β' are both ordered bases for X then there is an $n \times n$ matrix which converts coordinate vectors with respect to β to coordinate vectors with respect to β'. We denote this matrix as $[\mathbf{I}]_\beta^{\beta'}$, as in Section I.2.08. The columns of this matrix are the coordinate vectors of β written in terms of the basis β'. So if $\mathbf{x} \in X$, the (column) coordinate vector of $\mathbf{x}$ with respect to the ordered basis β in $[\mathbf{x}]^\beta$, and the (column) coordinate vector of $\mathbf{x}$ with respect to ordered basis β' is $[\mathbf{x}]^{\beta'}$ then we have $[\mathbf{I}]_\beta^{\beta'}[\mathbf{x}]^\beta = [\mathbf{x}]^{\beta'}$. Also, the inverse of $[\mathbf{I}]_\beta^{\beta'}$ is $[\mathbf{I}]_{\beta'}^{\beta}$ so that $[\mathbf{I}]_\beta^{\beta'}[\mathbf{I}]_{\beta'}^{\beta} = [\mathbf{I}]_\beta^\beta = [\delta_i^j]$.

Theorem III.1.B. Let $\beta = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ and $\beta' = \{\mathbf{b}'_1, \mathbf{b}'_2, \ldots, \mathbf{b}'_n\}$ be bases for X and let $\beta^* = \{\mathbf{b}^1, \mathbf{b}^2, \ldots, \mathbf{b}^n\}$ be the dual basis of β (so that β^* is a basis of X^*). With $\beta'^* = \{\mathbf{b}'^1, \mathbf{b}'^2, \ldots, \mathbf{b}'^n\}$ the dual basis of β', for $\mathbf{f} \in X^*$ where

$$\mathbf{f} = \sum_{i=1}^n f_i \mathbf{b}^i = f_i \mathbf{b}^i = \sum_{i=1}^n f'_i \mathbf{b}'^i = f'_i \mathbf{b}'^i$$

we have $f'_i = b_i^j f_j$ where the b_i^j are coordinates of $\mathbf{b}'_i \in \beta'$ with respect to ordered basis β (that is, b_i^j satisfies $\mathbf{b}'_i = \sum_{j=1}^n b_i^j \mathbf{b}_j = b_i^j \mathbf{b}_j$).

Note. Theorem III.1.B shows us that covariant vector $\mathbf{f}$ "transforms covariantly"; that is, to transform from one basis to another in X^* we have sums over superscripts ("upper indices") for the components relative to a basis, as in $f'_i = b^j_i f_j$ and $\mathbf{b}'_i = b^j_i \mathbf{b}_j$. Also, vectors (or "contravariant vectors") transform contravariantly. As observed in the previous Note and in the proof of Theorem III.1.B, $[\mathbf{I}]^{\beta'}_{\beta}[\mathbf{x}]^{\beta} = [\mathbf{x}]^{\beta'}$ and $[\mathbf{I}]^{\beta}_{\beta'}[\mathbf{x}]^{\beta'} = [\mathbf{x}]^{\beta}$, so that for $\mathbf{x} \in X$ we have

$$\begin{bmatrix} b^1_1 & b^1_2 & \cdots & b^1_n \\ b^2_1 & b^2_2 & \cdots & b^2_n \\ \vdots & \vdots & \ddots & \vdots \\ b^n_1 & b^n_2 & \cdots & b^n_n \end{bmatrix} \begin{bmatrix} x'^1 \\ x'^2 \\ \vdots \\ x'^n \end{bmatrix} = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix},$$

so that

$$\begin{bmatrix} x'^1 \\ x'^2 \\ \vdots \\ x'^n \end{bmatrix} = \begin{bmatrix} \tilde{b}^1_1 & \tilde{b}^1_2 & \cdots & \tilde{b}^1_n \\ \tilde{b}^2_1 & \tilde{b}^2_2 & \cdots & \tilde{b}^2_n \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{b}^n_1 & \tilde{b}^n_2 & \cdots & \tilde{b}^n_n \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}$$

where $[\tilde{b}^j_i] = [b^j_i]^{-1}$, or $\tilde{b}^i_k b^k_j = \delta^i_j$, and so $x'^i = \tilde{b}^i_j x^j$. That is, to change from one basis to another in X we have sums over subscripts ("lower indices") for the components relative to a basis. The pattern is: upper indices on objects (covariant vectors or components of covariant vectors) indicate covariance, and lower indices on objects (contravariant vectors or components of contravariant vectors) indicate contravariance.

Note III.1.B. We now consider $(X^*)^*$. since X and X^* are isomorphic real vector spaces (when X is finite dimensional), we should not be surprised to learn $(X^*)^*$ is also isomorphic to X. In Exercise III.1.1(a), for each $\mathbf{x} \in X$ we define ${}^*\mathbf{x}^* \in (X^*)^*$ (so ${}^*\mathbf{x}^* : X^* \to \mathbb{R}$) ad ${}^*\mathbf{x}^*(\mathbf{f}) = \mathbf{f}(\mathbf{x}) \in \mathbb{R}$ for all $\mathbf{f} \in X^*$. In Exercise III.1.1(b), we define $\theta : X \to (X^*)^*$ as $\theta(\mathbf{x}) = {}^*\mathbf{x}^*$ and show that θ is a linear transformation. Let $\beta = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for X, let $\beta^* = \{\mathbf{b}^1, \mathbf{b}^2, \ldots, \mathbf{b}^n\}$ be the dual basis of β, and let $(\beta^*)^* = \{\mathbf{b}^{1\,*}, \mathbf{b}^{2\,*}, \ldots, \mathbf{b}^{n\,*}\}$ be the deal basis to β^*. So β^* is a basis for X^* and $(\beta^*)^*$ is a basis for $(X^*)^*$. For $\mathbf{f} \in X^*$, say $\mathbf{f} = f_i\mathbf{b}^i$, we have

$$\begin{aligned}
\mathbf{b}^{i\,*}(\mathbf{f}) &= \mathbf{b}^{i\,*}(f_i\mathbf{b}^i) = \mathbf{b}^{i\,*}(f_1, f_2, \ldots, f_n) \\
&= f_i \text{ by the definition of dual basis of } \beta^* \\
&= (f_1\mathbf{b}^1 + f_2\mathbf{b}^2 + \ldots + f_n\mathbf{b}^n)(\mathbf{b}_i) \text{ since } \mathbf{b}^j(\mathbf{b}_i) = \delta_i^j \text{ by Note III.1.1} \\
&= \mathbf{f}(\mathbf{f}) \text{ by the definition of } {}^*\mathbf{x}^* \text{ when } \mathbf{x} = \mathbf{b}_i \in X \\
&= (\theta(\mathbf{b}_i))(\mathbf{f}) \text{ by definition of } \theta.
\end{aligned}$$

Therefore θ maps basis β to basis $(\beta^*)^*$ with $\mathbf{b}_i \mapsto \mathbf{b}^{i\,*}$ for each $i \in \{1, 2, \ldots, n\}$. Since θ maps β one to one and onto $(\beta^*)^*$ then θ maps X one to one and onto $(X^*)^*$. Therefore θ is a vector space isomorphism and gives a "natural" way to identify $(X^*)^*$ with X, and to identify $\mathbf{b}^{i\,*}$ with $\mathbf{b}_i$. Dodson and Poston use this observation to justify the statement: "We shall simply regard X and X^* as each other's duals, and forget about $(X^*)^*$; in fact that is why the word 'dual' is used here at all." See page 63.

Note. The previous argument only holds for finite dimensional vector space. For example, on a measurable set E, $(L^1)^* \cong L^\infty$ (by the Riesz Representation Theorem), but $(L^\infty)^* \ncong L^1$ (by the Kantorovitch Representation Theorem) and so $((L^1)^*)^* \ncong L^1$. See my online notes for Real Analysis (MATH 5210/5220) on The Riesz Representation Theorem and The Kantorovich Representation Theorem.

Revised: 5/19/2019