Continuing the MATH 4260 Story

Here are some ways to do it. Since you have the Meckeses' book at hand, you might enjoy reading some parts that we did not cover before moving on, which are detailed below. For complete and comprehensive coverage of *finite*-dimensional linear algebra, see *Linear Algebra* by Kenneth Hoffman and Ray Kunze. Equally rigorous, but slightly gentler, treatments are (as recommended in the syllabus) *Linear Algebra Done Right* by Sheldon Axler (now `available as a free pdf online`); *The Less Is More Linear Algebra of Vector Spaces and Matrices* by Daniela Calvetti and Erikki Somersalo (written by colleagues of the Meckeses at Case Western); *Linear Algebra* by Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence; and *Advanced Linear Algebra* by Hugo J. Woerdeman. Again, these books treat the *finite*-dimensional theory very well (and certainly present many results that hold without dimension counting). For immense coverage of *matrices* in particular (as opposed to more abstract finite-dimensional matters) with some analysis prerequisites, consult *Matrix Analysis* by Roger A. Horn and Charles R. Johnson.

But we can't stop with finite dimensions, and that is where we need analysis and topology. I recommend *Introductory Functional Analysis with Applications* by Erwin Kreyszig and *An Introduction to Hilbert Space* by Nicholas Young. Both presume some familiarity with real analysis—more so limits, continuity, and basic topology than derivatives or integrals. A unique introduction to real analysis is *Real Analysis for the Undergraduate: With an Invitation to Functional Analysis* by Matthew A. Pons. This does exactly what it claims: all of the material for the real analysis sequence, with topics in *linear analysis* at the end of each chapter. Something of a bridge between linear algebra in all dimensions and abstract algebra is *Advanced Linear Algebra* by Steven Roman, which will present more connections with modern algebra than analysis. Finally, linear algebraic theory is wholly incomplete without computations, both theoretical and implemented—"Be wise, linearize" $\mapsto$ "Be wise, discretize." For that, see the stunningly beautiful *Numerical Linear Algebra* by Lloyd N. Trefethen and David Bau III.

1. There are three major, related topics in the book that we did not discuss: diagonalization, the singular value decomposition (SVD), and the spectral theorem. (In my opinion, the first two are better topics for our *first* course in linear algebra, because they depend so heavily on finite-dimensionality, and because they can be phrased entirely in the language of matrices. Increasingly, I think of Linear I as the progression of stories "Solve $A\mathbf{x} = \mathbf{b} \mapsto$ Understand $A\mathbf{x} = \mathbf{b} \mapsto$ Understand $A \mapsto$ Approximate A," and the SVD is one of the keys to approximating A.)

First, let $\mathcal{V}$ be a finite-dimensional vector space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. **Diagonalization** answers the question of when $\mathcal{V}$ has a basis that interacts "best" with $\mathcal{T}$: when there is a basis $(v_1, \ldots, v_n)$ and scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ such that

$$\mathcal{T}\left(\sum_{k=1}^{n} \alpha_k v_k\right) = \sum_{k=1}^{n} \alpha_k \lambda_k v_k.$$

That is, $(v_1, \ldots, v_n)$ is a basis consisting of *eigenvectors* of $\mathcal{T}$. When such a basis exists, $\mathcal{T}$ is **diagonalizable**. Properties of diagonalizable operators are discussed in Sections 3.5 and 3.6.

2. Second, let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional inner product spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. Since we are not requiring $\mathcal{V} = \mathcal{W}$, it does not make sense to think about eigenvalues for $\mathcal{T}$. But it does make sense to ask for a substitute for eigenvalues—and to ask for bases of $\mathcal{V}$ and $\mathcal{W}$ with which $\mathcal{T}$ interacts particularly nicely. We might as well ask for *orthonormal* bases along the way.

Suppose $\mathcal{T} \neq 0$. The **SINGULAR VALUE DECOMPOSITION** asserts that $\mathcal{V}$ and $\mathcal{W}$ have orthonormal bases $(u_1, \ldots, u_n)$ and $(\widetilde{u}_1, \ldots, \widetilde{u}_m)$ and that there exist $\sigma_1, \ldots, \sigma_r \in \mathbb{R}$ with $1 \le r \le \min\{m, n\}$ and $0 < \sigma_r \le \sigma_{r-1} \le \cdots \le \sigma_1$ such that

$$\mathcal{T} u_k = \begin{cases} \sigma_k \widetilde{u}_k, \ 1 \le k \le r \\ 0, \ r+1 \le k \le n. \end{cases}$$

This is discussed and proved at length in Sections 5.1 and 5.2. See Problem 5.1.13 for an analogous "decomposition" of $\mathcal{T}$ as a linear combination of very special operators.

An important corollary of the SVD is that we can construct a best approximation to $\mathcal{T}$ in the operator norm out of all of the "rank k" operators. That is, if we want to find the operator $\mathcal{T}_k$ that minimizes $\|\mathcal{T} - \mathcal{T}_k\|_{\mathcal{V} \to \mathcal{W}}$ with $\dim[\mathcal{T}_k(\mathcal{V})] = k$, then the SVD tells us how to do that. At the level of matrix multiplication operators, this gives us another notion of approximating a given problem. Previously we thought about how to "solve" $A\mathbf{x} = \mathbf{b}$ by replacing $\mathbf{b} \notin \operatorname{col}(A)$ with some $\widehat{\mathbf{b}} \in \operatorname{col}(A)$ and then taking $\widehat{\mathbf{x}}$, which solves $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$ exactly, as an approximation to our "impossible" problem $A\mathbf{x} = \mathbf{b}$. What if instead we solved $A_k\mathbf{x} = \mathbf{b}$, with A_k as the best rank-k approximation to A? Or what if we just viewed A as a collection of data, possibly a large collection, and we wanted to approximate that collection for "storage" purposes? Again, the SVD tells us how to do it.

3. Third, let $\mathcal{V}$ be a finite-dimensional inner product space and $\mathcal{T} \in \mathbf{L}(\mathcal{V})$. Diagonalizability says that $\mathcal{V}$ has a basis consisting of eigenvectors of $\mathcal{T}$. When does $\mathcal{V}$ have an *orthonormal* basis of eigenvectors of $\mathcal{T}$? (Gram–Schmidt will not give us this directly from diagonalizability—why not?) The answer here is the **SPECTRAL THEOREM**: $\mathcal{V}$ has an orthonormal basis of eigenvectors of $\mathcal{T}$ if and only if $\mathcal{T}$ and $\mathcal{T}^*$ commute. (If $\mathcal{T}\mathcal{T}^* = \mathcal{T}^*\mathcal{T}$, then $\mathcal{T}$ is **NORMAL**.) This is proved in Section 5.4 using the SVD; there are other routes.

4. Now here is where we left off. We could have said much more about norms. First, many norms are really measuring the same thing, even if they look different. Two norms $\|\cdot\|$ and $|||\cdot|||$ on a space $\mathcal{V}$ are **EQUIVALENT** if there are constants C_1, $C_2 > 0$ such that

$$C_1 \|v\| \le |||v||| \le C_2 \|v\|$$

for all $v \in \mathcal{V}$. The idea is that knowing how v behaves with respect to one norm yields its behavior, up to a multiplicative constant, with respect to the other norm. In the special case of the p-norms on $\mathbb{F}^n$, it is possible to determine "optimal" constants for these inequalities at which equalities are actually attained. As another example, if $\mathcal{V}$ is an inner product space with $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$ for some subspace $\mathcal{U}$, putting $|||v||| := \|\mathcal{P}_{\mathcal{U}} v\| + \|v - \mathcal{P}_{\mathcal{U}} v\|$ gives an equivalent norm on $\mathcal{V}$.

It turns out that when $\mathcal{V}$ is finite-dimensional, all norms on $\mathcal{V}$ are equivalent—big surprise, as everything is nicer in finite dimensions. It is possible to construct examples of norms on specific infinite-dimensional spaces that are not equivalent.

5. We could also say much more about the operator norm. First, we could give an example of an operator for which the maxima/minima of the sets in Problem 45.10 and equation (45.2) do not exist, and so we need the notion of supremum/infimum from analysis for operator norms on infinite-dimensional spaces. Second, we could think about a matrix $A \in \mathbb{F}^{m \times n}$ both as an operator in $\mathbf{L}(\mathbb{F}^n, \mathbb{F}^m)$ and (up to isomorphism) as a vector in $\mathbb{F}^{mn}$. We could put different p-norms on $\mathbb{F}^n$ and $\mathbb{F}^m$ and look at the operator norm of A with respect to those underlying norms and how that operator norm might compare to another p-norm on $\mathbb{F}^{mn}$.

Third, we said a lot about the orthogonal projection onto a subspace of an inner product space, but we never said that it was unique. If $\mathcal{V}$ is an inner product space and $\mathcal{U}$ is a subspace, there are many operators $\mathcal{P} \colon \mathcal{V} \to \mathcal{U}$ such that $\mathcal{P}(\mathcal{V}) = \mathcal{U}$, $\mathcal{P}^2 = \mathcal{P}$, and $\mathcal{P}^* = \mathcal{P}$. Adding the condition $\|\mathcal{P}\|_{\mathcal{V}\to\mathcal{V}} \le 1$ ensures that $\mathcal{P} = \mathcal{P}_{\mathcal{U}}$.

While we can say a lot about the algebraic dual space $\mathcal{V}'$, more interesting, typically, is the plain **DUAL SPACE** $\mathcal{V}^* := \mathbf{B}(\mathcal{V}, \mathbb{F})$, which is a subspace of $\mathcal{V}'$.

6. We could also break norms. Say that $\mathcal{C}_{\mathsf{pw}}([0,1])$ is the space of piecewise-continuous functions on $[0,1]$. For example,

$$f(x) := \begin{cases} 1, & x = 0 \\ 0, & 0 < x \le 1 \end{cases}$$

is piecewise continuous on $[0,1]$. Using the notion of integral extended to piecewise-continuous functions as we do in calculus, we have $\|f\|_1 = 0$ but $f \neq 0$. Then $\|\cdot\|_1$ is no longer a norm but a **SEMINORM**: a map that satisfies all of the properties of a norm except definiteness.

If $p \colon \mathcal{V} \to \mathbb{R}$ is a seminorm on $\mathcal{V}$, it turns out that $\mathcal{V}_0 := \{v \in \mathcal{V} \mid p(v) = 0\}$ is a subspace of $\mathcal{V}$. By introducing the notion of **QUOTIENT SPACE** $\mathcal{V}/\mathcal{V}_0$, we can construct a norm on $\mathcal{V}$ out of p. That is, we put

$$v + \mathcal{V}_0 := \{v + w \mid w \in \mathcal{V}_0\} \quad \text{and} \quad \mathcal{V}/\mathcal{V}_0 := \{v + \mathcal{V}_0 \mid v \in \mathcal{V}\},$$

so $\mathcal{V}/\mathcal{V}_0$ is a set of sets. We define arithmetic in $\mathcal{V}/\mathcal{V}_0$ by

$$(v_1 + \mathcal{V}_0) + (v_2 + \mathcal{V}_0) := (v_1 + v_2) + \mathcal{V}_0 \quad \text{and} \quad \alpha(v + \mathcal{V}_0) := (\alpha v) + \mathcal{V}_0,$$

and this turns out to be well-defined. In particular, the zero vector in $\mathcal{V}/\mathcal{V}_0$ is $\mathcal{V}_0$. Then putting $\|v + \mathcal{V}_0\| := p(v)$ gives a well-defined norm on $\mathcal{V}_0$. In short, "quotienting out by $\mathcal{V}_0$" allows us to treat every $v \in \mathcal{V}_0$ with $p(v) = 0$ as the zero vector in $\mathcal{V}/\mathcal{V}_0$.

The quotient space might be the first vector space in our course that was not at some level a function space! Quotient spaces give another proof of rank–nullity, too.

7. There is much more to say about linear functionals, too. A sneaky little result that we never needed, but that is surprisingly useful, is that if $\varphi_1, \ldots, \varphi_n, \varphi \in \mathcal{V}'$ with $\cap_{k=1}^n \ker(\varphi_k) \subseteq \ker(\varphi)$, then $\varphi \in \operatorname{span}(\varphi_1, \ldots, \varphi_n)$. This is very helpful to know.

More excitingly, it turns out that the "evaluation" functional

$$\varphi \colon \mathcal{C}([0,1]) \to \mathbb{R} \colon f \mapsto f(0)$$

cannot be written in the form $\varphi(f) = \int_0^1 f(x)\rho(x)\, dx$ for some $\rho \in \mathcal{C}([0,1])$, and so the Riesz representation theorem can break.

We studied linear functionals on inner product spaces extensively, but we can also work on more general vector spaces. Recall that for an inner product space $\mathcal{V}$ and $\mathcal{U} \subseteq \mathcal{V}$, the **ORTHOGONAL COMPLEMENT** of $\mathcal{U}$ is $\mathcal{U}^\perp = \{v \in \mathcal{V} \mid \langle v, u \rangle = 0 \text{ for all } u \in \mathcal{U}\}$. For a vector space $\mathcal{V}$ and $\mathcal{U} \subseteq \mathcal{V}$, the **ANNIHILATOR** of $\mathcal{U}$ is

$$\mathcal{U}^\circ := \{\varphi \in \mathcal{V}' \mid \varphi(u) = 0 \text{ for all } u \in \mathcal{U}\} = \{\varphi \in \mathcal{V}' \mid \mathcal{U} \subseteq \ker(\varphi)\}.$$

How are $\mathcal{U}^\circ$ and $\mathcal{U}^\perp$ related in an inner product space? For $\mathcal{U} \subseteq \mathcal{V}'$, we put

$$^\circ\mathcal{U} := \{v \in \mathcal{V} \mid \varphi(v) = 0 \text{ for all } \varphi \in \mathcal{U}\} = \cap_{\varphi \in \mathcal{U}} \ker(\varphi).$$

One might ask about the relationship of $^\circ(\mathcal{U}^\circ)$ to $\mathcal{U}$, just as we did for $(\mathcal{U}^\perp)^\perp$. It is possible to develop "range equals kernel circ" relationships with annihilating sets.

We also have a more general notion of adjoint, which we might inelegantly call the "algebraic adjoint." For vector spaces $\mathcal{V}$ and $\mathcal{W}$ and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$, define $\mathcal{T}' \in \mathbf{L}(\mathcal{W}', \mathcal{V}')$ by

$$(\mathcal{T}'\psi)(v) = \psi(\mathcal{T}v).$$

Parse through this equality to understand why $\mathcal{T}'\psi$ is a linear functional on $\mathcal{V}$ for each linear functional ψ on $\mathcal{W}'$. It turns out that if $A \in \mathbb{F}^{m \times n}$, then $[\mathcal{T}_A'] = A^\mathsf{T}$. That is, the adjoint $\mathcal{T}'$ is represented by the *ordinary* transpose of $\mathcal{T}$, not the conjugate transpose. If $\mathcal{V}$ and $\mathcal{W}$ are inner product spaces and $\mathcal{T}$ has an adjoint, we might explore how $\mathcal{T}'$ and $\mathcal{T}^*$ interact.

And if we have one dual space, why not two? We can show that if $\mathcal{V}$ is finite-dimensional, then $\mathcal{V}$ is isomorphic to $\mathcal{V}'' := (\mathcal{V}')'$. In particular, there is a "natural" isomorphism: define $\mathcal{J} \colon \mathcal{V} \to \mathcal{V}''$ by $(\mathcal{J}v)\varphi := \varphi(v)$ for $\varphi \in \mathcal{V}'$. Operators defined by "evaluate at" are quite common! A normed space $\mathcal{V}$ is **REFLEXIVE** if $\mathcal{J}$ is an isomorphism from $\mathcal{V}$ to $\mathcal{V}^*$, and reflexive spaces have many nice properties (here it is important that we use $\mathcal{J}$, not any old isomorphism, and $\mathcal{V}^*$, not $\mathcal{V}'$).

8. There is also an algebraic structure that we have neglected: an **ALGEBRA**. This is a vector space $\mathcal{V}$ on which there is defined a "multiplication of vectors" operation that interacts with vector addition and scalar multiplication as one would expect, although the multiplication does not have to be commutative. The prototypical example is operator composition in $\mathbf{L}(\mathcal{V})$. Operator composition interacts well with the operator norm on $\mathbf{B}(\mathcal{V})$ when $\mathcal{V}$ is a normed space: $\|\mathcal{T}_1\mathcal{T}_2\|_{\mathcal{V}\to\mathcal{V}} \le \|\mathcal{T}_1\|_{\mathcal{V}\to\mathcal{V}} \|\mathcal{T}_2\|_{\mathcal{V}\to\mathcal{V}}$. Operator composition also interacts with the adjoint when $\mathcal{V}$ is an inner product space (or not) in that $(\mathcal{T}_1\mathcal{T}_2)^* = \mathcal{T}_2^*\mathcal{T}_1^*$. Otherwise, taking the adjoint resembles algebraic properties of conjugating complex numbers, e.g., $(\mathcal{T}^*)^* = \mathcal{T}$.

More generally, if $\mathcal{V}$ is an algebra with a norm, we want that norm to be **SUBMULTIPLICATIVE**: $\|vw\| \le \|v\| \|w\|$. And we are interested in the existence of an **INVOLUTION**: a map $\mathcal{V} \to \mathcal{V} \colon v \mapsto v^*$ that mimics the essential properties of the adjoint, e.g., $(vw)^* = w^*v^*$.

9. The majority of this course has involved *equalities*. Three major exceptions are the Cauchy–Schwarz inequality, the triangle inequality for norms (whether induced by an inner product or not), and the best approximation result from projections (also that approximation result from the SVD, which we did not discuss). Having a norm allows us to compare sizes: a vector v is "small" if $\|v\|$ is a "small" real number. If $\|v - w\|$ is "small," then v and w are "close"—and once we have a notion of measuring closeness, we can do analysis: limits, continuity, derivatives, integrals. We could redevelop much of introductory real analysis by replacing x and y with v and w and $|x-y|$ with $\|v-w\|$. This allows us to develop a notion of limits and continuity for functions between normed spaces—not necessarily linear operators, and not functions that are vectors within a space. For example, we could study the continuity of $\mathcal{F}\colon \mathcal{C}([0,1]) \to \mathcal{C}([0,1])\colon f \mapsto g \circ f$ with $g \in \mathcal{C}(\mathbb{R})$ fixed, where we norm $\mathcal{C}([0,1])$ by $\|\cdot\|_\infty$.

What about derivatives? When we are dealing with vectors, not real numbers, slopes and rates of change are less meaningful. But remember that the derivative is a *local linear approximation*: $f(x+h) \approx f(x) + f'(x)h$. We can ultimately encode this for a function $\mathcal{F}\colon \mathcal{V} \to \mathcal{W}$ by saying that f is differentiable at $v \in \mathcal{V}$ if there is an operator $\mathcal{T} \in \mathbf{B}(\mathcal{V}, \mathcal{W})$ such that

$$\lim_{h \to 0} \frac{\|\mathcal{F}(v+h) - \mathcal{F}(v) - \mathcal{T}h\|_{\mathcal{W}}}{\|h\|_{\mathcal{V}}} = 0,$$

whatever "$\lim_{h\to 0}$" means here. We call $\mathcal{T}$ the **DERIVATIVE** of $\mathcal{F}$ at v and write $D\mathcal{F}(v) := \mathcal{T}$. The power of the derivative is, as in real-variable calculus, *local linear approximation*: $\mathcal{F}(v+h) \approx \mathcal{F}(v) + D\mathcal{F}(v)h$. What is easier to study: some arbitrary map $\mathcal{F}$ or a linear operator $D\mathcal{F}(v)$? We do know a lot about linear operators.

10. Inequalities are what make analysis work, and analysis is what helps us answer so many lingering questions.

(i) Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and $\mathcal{T} \in \mathbf{L}(\mathcal{V}, \mathcal{W})$. What guarantees that $\mathcal{T}$ has an adjoint?

(ii) Let $\mathcal{V}$ be an inner product space and let $\mathcal{U}$ be a subspace of $\mathcal{V}$. What guarantees that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^\perp$ and $\mathcal{U} = (\mathcal{U}^\perp)^\perp$?

(iii) Let $\mathcal{V}$ be an inner product space and $\varphi \in \mathcal{V}'$. What guarantees the existence of $\rho \in \mathcal{V}$ such that $\varphi(v) = \langle v, \rho \rangle$ for all $v \in \mathcal{V}$?

(iv) Let $\mathcal{V}$ be an inner product space. What guarantees the existence of an orthonormal set $\{\phi_k\}_{k=1}^\infty$ such that

$$\lim_{n \to \infty} \left\| v - \sum_{k=1}^n \langle v, \phi_k \rangle \phi_k \right\| = 0$$

for all $v \in \mathcal{V}$?

(v) Let $\|\cdot\|$ and $|||\cdot|||$ be norms on a vector space $\mathcal{V}$. What guarantees that $\|\cdot\|$ and $|||\cdot|||$ are equivalent?

The first step in answering all of these questions is to involve **COMPLETENESS**: every **CAUCHY SEQUENCE** in the space(s) must converge. This is exactly the same definition of completeness as in analysis. Starting with completeness, adding some other topological machinery, and doing a *lot* of estimates will provide guarantees (sometimes sufficient, sometimes necessary) as requested above. This is the power and the glory of **FUNCTIONAL ANALYSIS**.