

Solutions to Exercises

Terrell L. Hodge¹, Blair R. Szymczyna² and Todd J. Barkman³

¹Department of Mathematics, Western Michigan University, Kalamazoo, MI, USA, ²Department of Chemistry, Western Michigan University, Kalamazoo, MI, USA, ³Department of Biological Sciences, Western Michigan University, Kalamazoo, MI, USA

Exercise 11.1. [MetaCyc] For a pictorial view of this pathway in standard biochemical reaction diagram form (in mathematical format, a *hypergraph*), go to <http://metacyc.org> [1]. In the search box in the site's upper right-hand corner, type "caffeine biosynthesis." From the search results, choose "caffeine biosynthesis 1." At the time of writing, this will produce the interactive picture shown in Figure S11.1.

In the style of the biochemical diagram shown above, each single reaction, represented notationally by a single arrow " \longrightarrow " in equations like Eq. (11.1), is now graphically represented by a multi-pronged arrow that, in mathematical language, is an example of a *hyperedge*.

1. Mouse over each hyperedge to see the standard-form equation for the corresponding reaction. What do the abbreviations SAH and SAM in Eq. (11.1) represent? Are they abbreviations that appear in MetaCyc?

Solution. Given the simplicity of the reactions, one may infer that SAM is "*S*-adenosyl-L-methionine" and SAH is "*S*-adenosyl-L-homocysteine." These abbreviations appear on this page of MetaCyc when mousing over the compound names and are commonly used among biologists.

2. Does the set of reactions appearing on this MetaCyc page correspond to the system in (11.2–11.5) labeled above?

Solution. They should, unless the database has changed to reflect additional information that may have become available.

3. To see a representation in terms of the molecular structure of each principle compound, select the pull-down box "More Detail" next to "Enzyme View" at the top middle of this MetaCyc page. What is the impact of the reaction on the molecule of xanthosine in the first step (i.e., first reaction)? That is, what does "methylation" seem to correspond to, here, at a molecular level? *Hints:* It may help to recall that three of the four reactions in (11.2–11.5) are methylation steps, so one may try comparing across all the reaction steps. It may also help to look up the relevant molecules in a different database. For example, for the first reaction step, go to <http://www.genome.jp/kegg/> [2]. Enter "xanthosine" in the search box, scroll down to "KEGG COMPOUND" and click on the link for xanthosine ("C01762"); among other information, a picture of the molecule is displayed. Likewise, use KEGG to find a picture of the molecular structure of 7-methylxanthosine (also available from the compound list for xanthosine). Compare these diagrams with those in MetaCyc to inform your deductions.

Solution. The meaning of methylation is shown through the molecular diagrams and other information on the sites. It is the addition of a CH₃ group or, alternatively, the substitution for another atom or group by a CH₃ group. Methylation on the MetaCyc diagrams is represented only by the addition of a single bond line; in KEGG, the bond line with the CH₃ attached is visible, so is more informative. In each of the three steps that are methylation

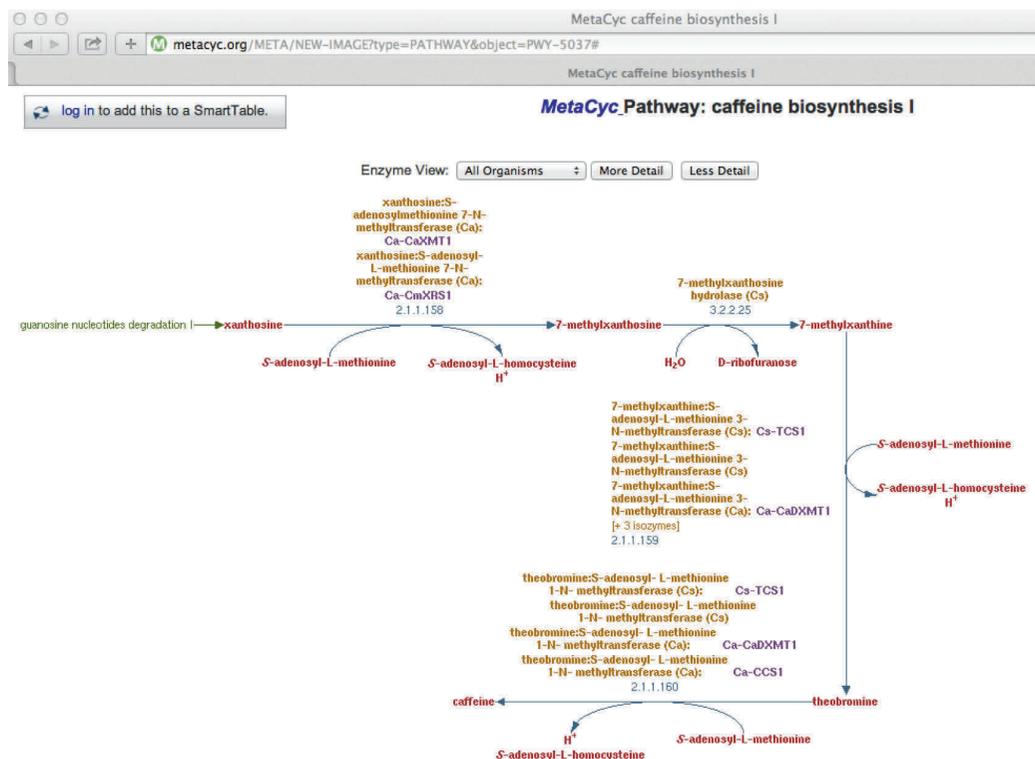


FIGURE S11.1 A typical ^1H NMR spectrum. Coupling between hydrogen nuclei that are close together in a molecule leads to peak multiplicity. The number of coupled hydrogen atoms is one less than the number of associated peaks. First.

steps, SAM donates a methyl group, with SAH (along with H^+) as the remainders. *Additional Note:* One can see this further in KEGG by clicking on the link to “caffeine metabolism” from, for example, the page found earlier for the compound xanthosine, and then clicking on the rectangular box labeling the reaction on the resulting graph that links the nodes for xanthosine and 7-methylxanthosine. The resulting page includes a complete picture of the molecular diagrams for all metabolites involved in the first reaction step. Overall, the picture for “caffeine metabolism” in KEGG is broader and more involved, compared to that for “caffeine biosynthesis 1” in MetaCyc.

- Continuing with the previous MetaCyc page for caffeine biosynthesis 1, select “Less Detail” to return to the previous view. Above each hyperedge, a list of enzymes is provided. Mouse over an enzyme to see a description of the related standard taxonomic species in which this enzyme has been demonstrated to catalyze the reaction (including the nature of evidence for the entry) and one or more genes associated with these enzymes. For which of the reactions listed in (11.2–11.5) is *Coffea arabica* experimentally indicated?

Solution. As above, mousing over the enzymes above shows that *C. arabica* is indicated in the first, third, and fourth reactions, but not the second.

- Compare the results of your previous answer with the outcome of selecting *C. arabica* from the pull-down menu for the button currently labeled “All Organisms” next to “Enzyme View.” Then, change to “*Camellia sinensis*.” Does the set of reactions (i.e., the *pathway*) change? And what is the experimental evidence for this pathway?

Solution. In this new view, one see that second hyperedge for *C. arabica* is unmarked with an enzyme. For *Camellia synthesis*, the same set of four reactions is listed, but this time, experimental evidence is cited for the second, third, and fourth reactions, but not the first.

6. Instead of going straight to the “caffeine biosynthesis 1” link from the search page, one can select the “temporary Smart Table” option.¹ When this opens, try using the pull-down menu on the resulting table page to add various columns to the table that consolidate the information listed above, including lists of the relevant compounds, enzymes, reactants, and products associated with each system of reactions, taken in its totality. Note that the reactions are labeled by their enzymes; this is way to record and include the enzyme data, although they are omitted from the complexes.

Solution. Follow the instructions here. This set-up is useful if one wants to have data about a reaction or reaction system in a concise form ready to input into other programs. □

Exercise 11.2.

1. Write out the two reaction vectors (10-tuples) $\mathbf{y}'_3 - \mathbf{y}_3$, $\mathbf{y}'_4 - \mathbf{y}_4$ that describe the remaining two reactions in the caffeine biosynthesis pathway from the equation system (11.2–11.5).

Solution. Here, carrying out the component-wise subtraction gives $\mathbf{y}'_3 - \mathbf{y}_3 = (0, -1, 0, 1, 1, 0, -1, 0, 1, 0)$, and $\mathbf{y}'_4 - \mathbf{y}_4 = (0, -1, 0, 1, 1, 0, 0, 0, -1, 1)$.

2. Take the sum of the four reaction vectors $(\mathbf{y}'_1 - \mathbf{y}_1) + (\mathbf{y}'_2 - \mathbf{y}_2) + (\mathbf{y}'_3 - \mathbf{y}_3) + (\mathbf{y}'_4 - \mathbf{y}_4)$ and interpret the outcome in English.

Solution. Here, the sum $(\mathbf{y}'_1 - \mathbf{y}_1) + (\mathbf{y}'_2 - \mathbf{y}_2) + (\mathbf{y}'_3 - \mathbf{y}_3) + (\mathbf{y}'_4 - \mathbf{y}_4) = (-1, -3, 0, 3, 3, -1, 0, 1, 0, 1)$. This calculation represents the outcome of taking the metabolic pathway of caffeine biosynthesis, given by the chain of successive reactions in the equation system (11.2–11.5); alternatively, we are treating these four reactions as four steps in one single reaction $XR + 3SAM + H_2O \rightarrow 3SAH + 3H^+ + Drib + Cf$. Ultimately, producing caffeine by this pathway involves the input of xanthosine, with accompanying amounts of water and *S*-adenosyl-L-methionine, and with additional byproducts of *S*-adenosyl-L-homocysteine, H^+ , and D-ribofuronase, all in the amounts specified by the vector $(-1, -3, 0, 3, 3, -1, 0, 1, 0, 1)$. □

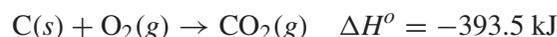
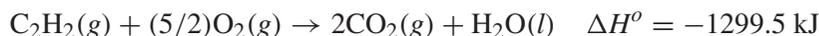
Exercise 11.3 (Hess’s Law and Spanning Sets). The arithmetic of tuples, as in Exercise 11.2 and the material preceding it, appears in many useful contexts. For example, Hess’s Law (in physical chemistry) ensures that enthalpy of a given chemical reaction is constant, regardless of the number of steps taken in the reaction. In practice, that lets one solve for the enthalpy of a reaction for which one may not have the experimental data by relating it to other reactions in ways that we can now state in terms of linear combinations of complexes and/or reaction vectors.²

1. As an illustration, consider the following typical sample problem, and an accompanying solution, for calculating enthalpy using Hess’s Law. These are posed as Problem 11.1 at <http://www.chemteam.info/Thermochem/HessLawIntro1a.html>. They are also available with the Supplementary Materials for this chapter. Note that “???” denotes the unknown enthalpy value ΔH° sought for the given reaction.

Problem 11.1. Calculate the enthalpy for this reaction:



given the following thermochemical equations:

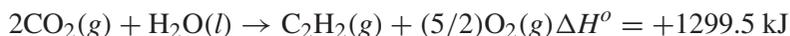


1. *Warning:* In current practice, this is a bit buggy after the initial request.

2. *Note:* Mathematically inclined students can work out the exercises below without even knowing what enthalpy is, only that it is preserved as shown therein. However, you may wish to start this exercise by doing a web search on Hess’s Law, for additional background, if it is unfamiliar.

Solution.

- (1) Determine what we must do to the three given equations to get our target equation:
- (a) first equation: flip it so as to put C_2H_2 on the product side
 - (b) second equation: multiply it by two to get $2C$
 - (c) third equation: do nothing. We need one H_2 on the reactant side and that is what we have.
- (2) Rewrite all three equations with changes applied:



Notice that the ΔH values changed as well.

- (3) Examine what cancels: $2CO_2$ from first and second equations H_2O from first and third equations $(5/2)O_2$ from first equation and sum of second and third equations
- (4) Add up ΔH values for our answer:

$$+1299.5 \text{ kJ} + (-787 \text{ kJ}) + (-285.8 \text{ kJ}) = +226.7 \text{ kJ}$$

In the language of this chapter, the metabolites are C, H_2 , C_2H_2 , O_2 , CO_2 , H_2O ; with this order, the target reaction $2C(s) + H_2(g) \rightarrow C_2H_2(g)$ is represented by $\mathbf{d} = (-2, -1, 1, 0, 0, 0)$. The three given equations can be represented as $\mathbf{a} = (0, 0, -1, -\frac{5}{2}, 2, 1)$, $\mathbf{b} = (-1, 0, 0, -1, 1, 0)$, and $\mathbf{c} = (0, -1, 0, -\frac{1}{2}, 0, 1)$. The effect of the operations on the equations suggested there is to produce $-\mathbf{a} = (0, 0, 1, \frac{5}{2}, -2, -1)$, $2\mathbf{b} = (-2, 0, 0, -2, 2, 0)$, and $1\mathbf{c} = (0, -1, 0, -\frac{1}{2}, 0, 1)$ as before. Then $\mathbf{d} = -\mathbf{a} + 2\mathbf{b} + 1\mathbf{c}$, so according to Hess's Law, the corresponding enthalpies share the same relationship: $\Delta H_{\mathbf{d}} = -\Delta H_{\mathbf{a}} + 2\Delta H_{\mathbf{b}} + 1\Delta H_{\mathbf{c}}$. This enables the calculation of $\Delta H_{\mathbf{d}}$ from the other reactions.

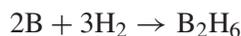
In a similar format as just given for Problem 11.1, see Problem 11.2 below and try to rework the solution to it. For additional practice, consider likewise (re)working one or more other problems at <http://www.chemteam.info/Thermochem/HessLawIntro1a.html>; see also the Supplementary Materials for this chapter.

Problem 11.2. The standard molar enthalpy of formation, $\Delta_f H^\circ$, of diborane cannot be determined directly because the compound cannot be prepared by reaction of boron and hydrogen. However, the value can be calculated. Calculate the standard enthalpy of formation of gaseous diborane (B_2H_6) using the following thermochemical information:

- (a) $4B(s) + 3O_2(g) \rightarrow 2B_2O_3(s) \quad \Delta H^\circ = -2509.1 \text{ kJ}$
- (b) $2H_2(g) + O_2(g) \rightarrow 2H_2O(l) \quad \Delta H^\circ = -571.7 \text{ kJ}$
- (c) $B_2H_6(g) + 3O_2(g) \rightarrow B_2O_3(s) + 3H_2O(l) \quad \Delta H^\circ = -2147.5 \text{ kJ}$

Solution.

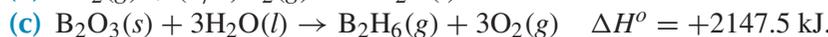
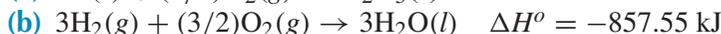
- (1) An important key is to know what equation we are aiming for. The answer is in the word *formation*:



Remember that formation means forming 1 mole of the target substance. This means that a one MUST be in front of the B_2H_6 .

- (2) In order to get to the formation reaction, the following must happen to equations (a), (b), and (c):
- equation (a)—divide through by 2
 - equation (b)—multiply through by $3/2$
 - equations (c)—flip Why?
 - equation (a)—this gives us $2B$ (from $4B$) for the final equation
 - equation (b)—this gives us $3H_2$ for the final equation
 - equations (c)—this puts B_2H_6 on the right-hand side of the final equation.

- (3) The above manipulations have consequences for the coefficients AND the ΔH° values. Rewrite equations (a), (b), and (c):



Add the three equations and the ΔH° values: $2\text{B} + 3\text{H}_2 \rightarrow \text{B}_2\text{H}_6 \quad \Delta H^\circ = +35.4 \text{ kJ}$.

Solution. Solutions vary depending upon choice of problems, but for Problem 11.2, the metabolites are B, O₂, B₂O₃, H₂, H₂O, B₂H₆; with this order, the target reaction $2\text{B} + 3\text{H}_2 \rightarrow \text{B}_2\text{H}_6$ is represented by $\mathbf{d} = (-2, 0, 0, -3, 0, 1)$. The three given equations can be represented as $\mathbf{a} = (-4, -3, 2, 0, 0, 0)$, $\mathbf{b} = (0, -1, 0, -2, 2, 0)$, and $\mathbf{c} = (0, -3, 1, 0, 3, -1)$. The effect of the operations on the equations suggested there is to produce $\frac{1}{2}\mathbf{a} = (-2, \frac{-3}{2}, 1, 0, 0, 0)$, $\frac{3}{2}\mathbf{b} = (0, \frac{-3}{2}, 0, 3, 0, \frac{-3}{2})$, and $(-1)\mathbf{c} = (0, 3, -1, 0, -3, 1)$ as before. Then $\mathbf{d} = \frac{1}{2}\mathbf{a} + \frac{3}{2}\mathbf{b} - \mathbf{c}$, so according to Hess's Law, the corresponding enthalpies share the same relationship: $\Delta H_{\mathbf{d}} = \frac{1}{2}\Delta H_{\mathbf{a}} + \frac{3}{2}\Delta H_{\mathbf{b}} - \Delta H_{\mathbf{c}}$. This enables the calculation of $\Delta H_{\mathbf{d}}$ from the other reactions.

2. Return again to part (1) (Problem 11.1). In the formal language of linear algebra, the *span* $\text{Span}(\mathbf{a}, \mathbf{b}, \mathbf{c})$ consists of all \mathbb{R} -linear combinations $\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{b} + \alpha_3\mathbf{c}$ of the reaction vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$. We showed the reaction vector $\mathbf{d} \in \text{Span}(\mathbf{a}, \mathbf{b}, \mathbf{c})$. Reframe Hess's Law in this light.

Solution. Hess's Law says that, if a reaction has a reaction vector d with $\mathbf{d} \in \text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_t)$ for some set of reaction vectors $\mathbf{a}_1, \dots, \mathbf{a}_t$, then the enthalpy of \mathbf{d} satisfies the same linear relationship determined by \mathbf{d} 's membership in $\text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_t)$. That is, if $\mathbf{d} = \sum_{i=1}^t \alpha_i \mathbf{a}_i$, then $\Delta H_{\mathbf{d}} = \sum_{i=1}^t \alpha_i \Delta H_{\mathbf{a}_i}$.

3. Suppose we take instead $\mathbf{u} = -\mathbf{a} + 3\mathbf{b}$, and $\mathbf{v} = \mathbf{1b} + (-1)\mathbf{c} = \mathbf{b} - \mathbf{c}$. Then $\mathbf{d} \in \text{Span}(\mathbf{u}, \mathbf{v})$, with $\mathbf{d} = \mathbf{1u} - \mathbf{1v}$. In terms of the original metabolites, to what biochemical reactions do \mathbf{u}, \mathbf{v} correspond?

Solution. Using the metabolites with the ordering C, H₂, C₂H₂, O₂, CO₂, H₂O as earlier for Problem 11.1, $\mathbf{u} = (-3, 0, 1, \frac{3}{2}, 1, -1)$, with the associated formal reaction $3\text{C} + \text{H}_2\text{O} \rightarrow \text{C}_2\text{H}_2 + \frac{3}{2}\text{O}_2 + \text{CO}_2$, and $\mathbf{v} = (-1, 1, 0, \frac{-1}{2}, 1, -1)$, with the associated formal reaction $\text{C} + \frac{1}{2}\text{O}_2 + \text{H}_2\text{O} \rightarrow \text{H}_2 + \text{CO}_2$. We note that reactions are often instead expressed in terms of integer coefficients (whole numbers of moles/molecules), but the linear algebraic formalism here captures all the inherent relative ratios of substrates and products needed for these reactions to occur. \square

Exercise 11.4 (Determining Spanning and Redundancy (i.e., Linear Dependence) Algorithmically). For Problem 11.1 as in Exercise 11.3 part (1),

$$[A|\mathbf{d}] = \left[\begin{array}{ccc|c} 0 & -1 & 0 & -2 \\ 0 & 0 & -1 & -1 \\ -1 & 0 & 0 & 1 \\ \frac{-5}{2} & -1 & \frac{-1}{2} & 0 \\ \frac{2}{2} & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{array} \right]; \text{RREF}([A|\mathbf{d}]) = \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]; \text{and } \text{RREF}(A) = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right];$$

implying $\text{rank}(A) = 3 = \text{rank}([A|\mathbf{d}])$. Thus all columns of A are basic columns, and, from the last column of $\text{RREF}([A|\mathbf{d}])$, $\mathbf{d} = -\mathbf{1a} + \mathbf{2b} + \mathbf{1c}$, as we saw before. The bars separating elements of A from \mathbf{d} in $[A|\mathbf{d}]$ help us recall the special role of the last column but do not otherwise affect the matrix calculations of the RREFs, and could be omitted. Calculations of the RREFs may be done using many calculators or MATLAB.³

3. If a matrix B has been entered into MATLAB, the needed command is `rref(B)` for a matrix B . For small matrices, the needed Gaussian elimination can also be carried out by hand; see, for example, [3, Section 1.2], [4, Ch. 1], or any basic reference on Gaussian elimination for linear systems. Another technology option is to use Octave (<http://www.gnu.org/software/octave/>), a freely available software that imitates MATLAB.

1. Relative to Exercise 11.3 part (2), following the discussion above, use the appropriate corresponding $RREF([A|\mathbf{d}])$ and $RREF(A)$ to compute the desired reaction vector \mathbf{d} as a linear combination of the reaction vectors for the given equations of those problems you had selected from <http://www.chemteam.info/Thermochem/HessLawIntro1a.html>. Use this matrix formulation to carry out another one of the problems at that site. Better yet, go to the related link <http://www.chemteam.info/Thermochem/HessLawIntro1b.html>, or the Supplementary Materials for this chapter, for calculations using four reactions, and see how these matrix calculations algorithmically reduce the *ad hoc* work needed for Problem #5 there.

Solution. Direct computation using $RREF([A|\mathbf{d}])$, $RREF(A)$ just as above; answers will vary depending upon choice of problems. For example, for Problem 11.2 worked out above, using the MATLAB code

```
A = [0 -1 0 -2; 0 0 -1 -1; -1 0 0 1; -5/2 -1 -1/2 0; 2 1 0 0; 1 0 1 0]; rref(A)
```

that enters the matrix A (but does not print it; to see it printed, use `print(A)` or enter A as above, but without the semicolon) and prints $RREF(A)$. One can similarly define a matrix X representing $[A|\mathbf{d}]$ by

```
X = [0 -1 0 -2; 0 0 -1 -1; -1 0 0 1; -5/2 -1 -1/2 0; 2 1 0 0; 1 0 1 0; -2 -1 1 0 0 0];
```

or by

```
d = [-2; -1; 1; 0; 0; 0];
```

```
X = [A d];
```

and compute `rref(X)` as above. (See also the solution to part 2 of this exercise below.) The results are:

$$[A|\mathbf{d}] = \left[\begin{array}{cccc|c} -4 & 0 & 0 & -2 \\ -3 & -1 & -3 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & -2 & 0 & -3 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & -1 & 1 \end{array} \right]; RREF([A|\mathbf{d}]) = \left[\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]; \text{ and } RREF(A) = \left[\begin{array}{cccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right];$$

implying $rank(A) = 3 = rank([A|\mathbf{d}])$, all columns of A are basic columns, and, from the last column of $RREF([A|\mathbf{d}])$, $\mathbf{d} = \frac{1}{2}\mathbf{a} + \frac{3}{2}\mathbf{b} + (-1)\mathbf{c}$, as we saw before. \square

2. Returning to the setting of Problem 11.1 as in Exercise 11.3 part (1), keep \mathbf{d} as before, but now consider $Span(\mathbf{a}, \mathbf{u}, \mathbf{b}, \mathbf{c}, \mathbf{v})$, for \mathbf{u}, \mathbf{v} as in part (3) of Exercise 11.3. We have already seen redundancy in this system, because by their definition, already $\mathbf{u}, \mathbf{v} \in Span(\mathbf{a}, \mathbf{b}, \mathbf{c})$. Set up a matrix A with columns in the order $\mathbf{a}, \mathbf{u}, \mathbf{b}, \mathbf{c}, \mathbf{v}$ to represent $Span(\mathbf{a}, \mathbf{u}, \mathbf{b}, \mathbf{c}, \mathbf{v})$, and compute $RREF(A)$. Which columns of A are basic, and which are nonbasic? For those that are nonbasic, what are their representations in terms of the other columns?

Solution. Columns 1, 2, and 4 are basic; Column 3 equals $\frac{1}{3}$ (Column 1) + $\frac{1}{3}$ (Column 2), while Column 5 = $\frac{1}{3}$ (Column 1) + $\frac{1}{3}$ (Column 2) - 1 (Column 4), in keeping with the same linear relationships for \mathbf{b} in terms of \mathbf{a} , and \mathbf{u} and then \mathbf{v} in terms of \mathbf{a}, \mathbf{u} , and \mathbf{c} . One set of MATLAB code lines to implement the calculations to producing $RREF(A)$ appears below:

```
a = [0; 0; -1; -5/2; 2; 1]
```

```
b = [-1; 0; 0; -1; 1; 0]
```

```
c = [0; -1; 0; -1/2; 0; 1]
```

```
u = -a + 3*b
```

```
v = b-c
```

```
A=[a u b c v]
```

```
rref(B)
```

3. Continuing Part (2): Furthermore, in Exercise 11.3, we have seen two distinct linear combinations of \mathbf{d} as an element of $\text{Span}(\mathbf{a}, \mathbf{u}, \mathbf{b}, \mathbf{c}, \mathbf{v})$, one in terms of $\mathbf{a}, \mathbf{b}, \mathbf{c}$, and the other in terms of \mathbf{u}, \mathbf{v} . Is this situation represented by the calculation of $RREF([A|\mathbf{d}])$, or is a different solution produced? What transpires if A is replaced by B , with columns ordered as $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}, \mathbf{v}$?

Solution. For the columns of A ordered as instructed, the linear combination is different from before, in the sense that it is presented as $\mathbf{d} = \frac{-1}{3}\mathbf{a} + \frac{2}{3}\mathbf{u} + 1\mathbf{c}$. If one uses B , one obtains $\mathbf{d} = -\mathbf{a} + 2\mathbf{b} + \mathbf{c}$.

4. Continuing Part (3): We know $\mathbf{d} \in \text{Span}(A)$ and $\mathbf{d} \in \text{Span}(\mathbf{u}, \mathbf{v})$. Is $\text{Span}(A) = \text{Span}(\mathbf{u}, \mathbf{v})$?

Solution. No; $\text{rank}(A) = 3$, but $\text{rank}([\mathbf{u}, \mathbf{v}]) = 2$. □

Exercise 11.5. Write out the stoichiometry matrix for the caffeine biosynthesis pathway (11.2–11.5).

Solution.

$$S = \begin{bmatrix} -1 & 0 & 0 & 0 \\ -1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

□

Exercise 11.6. Replicate Eq. (11.12), but now for the full caffeine network (11.2–11.5).

Solution.

$$\begin{bmatrix} \frac{d[XR]}{dt} \\ \frac{d[SAM]}{dt} \\ \frac{d[SevenMXR]}{dt} \\ \frac{d[SAH]}{dt} \\ \frac{d[Hplus]}{dt} \\ \frac{d[H2O]}{dt} \\ \frac{d[SevenMX]}{dt} \\ \frac{d[Drib]}{dt} \\ \frac{d[Tb]}{dt} \\ \frac{d[Cf]}{dt} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ -1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \\ v_4(t) \end{bmatrix} = \begin{bmatrix} -v_1 \\ -v_2 - v_3 - v_4 \\ v_1 - v_2 \\ v_1 + v_3 + v_4 \\ v_1 + v_3 + v_4 \\ -v_2 \\ v_2 - v_3 \\ v_2 \\ v_3 - v_4 \\ v_4 \end{bmatrix}.$$

□

Exercise 11.7 (Re: Stoichiometric Matrix—Exercise and Project).

1. (Exercise) Research in progress suggests that the picture of caffeine biosynthesis, among the whole collection of caffeine-producing plant families, is more complex than current literature, that is, based on (11.2–11.5) and explored in Exercise 11.1, suggests. Determine the dimensions of and write out the stoichiometric matrix S for the sample extended caffeine reaction network in the accompanying picture (caffeine biosynthesis.pdf) posted in the Supplemental Materials for this chapter.

Solution. From the file (caffeine biosynthesis.pdf), there are 15 distinct metabolites (the last abbreviation on the file is for [SAH and Hplus], representing one of SAH and one of H^+ , so yields no news species) and 17 reactions. Thus, the stoichiometric matrix S has dimensions 15×17 . Construction of the matrix S proceeds as in previous examples and exercises in this chapter, with the reaction vectors $\mathbf{y}'_i - \mathbf{y}_i$, $1 \leq i \leq 17$, corresponding to reactions $R_1 - R_{17}$, represented as column vectors, and is left to the reader. Actual matrices will vary, depending upon the ordering of the metabolites used. \square

2. (Project) (Optional; Continuation of Exercise 11.7 Part (1), and Referencing [5]) Extend the stoichiometric matrix to include the external input (xanthosine) and output (caffeine) to the system. One can calculate a spanning (generating) set and basis (generating set of minimal size) for the nullspace $N(S) = \{\mathbf{x} | S\mathbf{x} = \mathbf{0}\}$ from $RREF(S)$, (see [5] or [3, Section 4.2]). Is the outcome a “biologically good” basis, as described in [5]?

Can you find a set of extreme paths for this system? What, if anything, do you learn?

Solution Sketch. To create the new matrix S for this project, the original matrix (reabeled as \tilde{S}), from part (1) of this exercise, should be augmented by two additional columns, corresponding to the two “external” (i.e., exchange) fluxes, the first corresponding to the intake of xanthosene and the second to the output of caffeine. As per conventions described in [5, p. 247], the old \tilde{S} will correspond to the “internal” fluxes v_1, \dots, v_{17} and the two new columns of the new S to the “external” fluxes $v_{18} := b_1, v_{19} := b_2$. If the i th column \mathbf{s}_{*i} of S corresponds to the i th reaction R_i , $1 \leq i \leq 17$, then $\mathbf{s}_{*i} = \tilde{\mathbf{s}}_{i*}$, while there will be two new “reactions,” consisting of R_{18} given by \mathbf{e}_j^T , where the j th ordered metabolite is XR , and R_{19} given by $-\mathbf{e}_k^T$, where the k th ordered metabolite is Cf . (In terms of possible conventions, setting all external fluxes to heading “outward” from the involved metabolite would also be fine, that is, using $\mathbf{s}_{*j} = -\mathbf{e}_j^T$.) Note that, because the number of rows (15) of S is less than the number of columns (19), there will be at least $19 - 15 = 4$ nonbasic columns, and hence at least four free variables, whence the nullspace $N(S)$ has a dimension of at least four and is nontrivial. Solving $RREF(S)\mathbf{v} = \mathbf{0}$ for the basic variables in terms of the free variables (as in [3, p. 175]) produces a spanning set (that is a basis) for $N(S)$, from $RREF(S)$. The resulting vectors \mathbf{v} can be matched with the pathways on the original diagram that they indicate and checked to see if they are “biologically valid,” meaning the coefficients on the internal flux entries are positive. If so, these produce the extreme paths. If not, check whether linear combinations produce a set that still spans $N(S) \cap \mathbb{R}_+^{17}$ can be observed; technology for creating “biologically good” bases may also be employed, if accessible. See [5] for details. Experimenting with adding additional exchange fluxes would extend this project. \square

Exercise 11.8. Go to paper [6]⁴ and examine the small reaction network (futile cycle) there of Example 2.2. Then try to write out the equations for the mass-action kinetics of this reaction system, as per Eq. (11.14). You can check your answer by turning to [6, Example 2.3] (accounting for notational differences: [6] uses \mathbf{c} in place of $[\mathbf{X}]$).

Solution. Solution appears in [6, Example 2.3]. \square

Exercise 11.9.

1. Write out the vectors that will generate the stoichiometric subspace for the reaction system in Exercise 11.8.

Solution. Using the species (metabolites) ordered as S_1, \dots, S_6 , and the six given reactions $r_1 - r_6$, upon letting \mathbf{e}_i be the column vector with a 1 in the i th component and 0 elsewhere, S has columns \mathbf{s}_{*j} , corresponding to $r_j : \mathbf{y}'_j - \mathbf{y}_j$, $1 \leq j \leq 6$; specifically, $\mathbf{s}_{*1} = -\mathbf{e}_1 - \mathbf{e}_3 + \mathbf{e}_5$; $\mathbf{s}_{*2} = -\mathbf{s}_{*1}$; $\mathbf{s}_{*3} = \mathbf{e}_1 + \mathbf{e}_4 - \mathbf{e}_5$; $\mathbf{s}_{*4} = -\mathbf{e}_2 - \mathbf{e}_4 + \mathbf{e}_6$; $\mathbf{s}_{*5} = -\mathbf{s}_{*4}$, and $\mathbf{s}_{*6} = \mathbf{e}_2 + \mathbf{e}_3 - \mathbf{e}_6$.

2. What is the *dimension* of this space, that is, the smallest number of linearly independent vectors that span the stoichiometric subspace? Can you find a minimal set of generators for this stoichiometric subspace? (*Hint:* See Exercise 11.4 and the discussion preceding it.)

4. Included in the Supplemental Materials for this chapter; preprint also available from the Math ARXIV.

Solution. We seek the rank of S and a minimal set of generators (linearly independent set of generators) for $\text{Span}(S)$. We claim $\text{rank}(S) = 3$, and provide a minimal spanning set below. For a computational approach, one can calculate $RREF(S)$ and find $\text{rank}(S)$ as the number of pivots of $RREF(S)$; then the associated basic columns of S form a minimal spanning set. Alternatively, using the definitions of a linear combination and a spanning set, $\text{Span}(S) \subseteq \text{Span}(\mathbf{s}_{*1}, \mathbf{s}_{*3}, \mathbf{s}_{*4}, \mathbf{s}_{*6})$, because $\mathbf{s}_{*2} = -\mathbf{s}_{*1}$ and $\mathbf{s}_{*5} = \mathbf{s}_{*4}$. So, the dimension of the stoichiometric subspace is no larger than 4. It suffices to check whether $\mathbf{s}_{*1}, \mathbf{s}_{*3}, \mathbf{s}_{*4}, \mathbf{s}_{*6}$ are linearly independent. One can now check whether $RREF([\mathbf{s}_{*1}, \mathbf{s}_{*3}, \mathbf{s}_{*4}, \mathbf{s}_{*6}])$ has three pivots, with \mathbf{s}_{*6} nonbasic. Alternatively, prior to doing the computation, one can perhaps see through direct observation that $\mathbf{s}_{*6} = -\mathbf{s}_{*1} - \mathbf{s}_{*3} - \mathbf{s}_{*4}$, but that $\mathbf{s}_{*1}, \mathbf{s}_{*3}, \mathbf{s}_{*4}$ are independent. For example, in this list of three, \mathbf{e}_3 is a summand only of \mathbf{s}_{*1} , \mathbf{e}_5 is a summand of only \mathbf{s}_{*3} , and \mathbf{e}_6 is a summand only of \mathbf{s}_{*4} , so none of these three columns of S can be combinations of the other two.

3. Check your answers via [6, Example 3.2].

Solution. Relative to the solution to part (1) of this exercise, [6, Example 3.2] uses $\mathbf{s}_{*2}, \mathbf{s}_{*3}, \mathbf{s}_{*5}$ as the minimal spanning set (expressed in terms of the reactions r_2, r_3, r_5). Given the linear dependencies identified in the solution to part (1) of this problem, it should be clear that these also give $\text{Span}(S)$.

4. For the reaction system in Exercise 11.8, check that $\frac{d[X_1]}{dt} + \frac{d[X_5]}{dt} = 0$, so the change in $[X_1] + [X_5]$ must remain constant over time, and, hence, it equals the constant $[X_1](0) + [X_5](0)$, the sum of initial concentrations for the first and fifth metabolites. The equation $[X_1] + [X_5] = [X_1](0) + [X_5](0)$ is an example of a conservation law.

Solution. $\frac{d[X_1]}{dt} + \frac{d[X_5]}{dt} = 0$ follows immediately from the expressions found for each summand in Exercise 11.8.

5. Conservation laws correspond to vectors \mathbf{x} in \mathbb{R}^m satisfying $S^T \mathbf{x} = \mathbf{0}$, also called the left-nullspace $N(S^T)$ of S , for S^T the transpose of S . (Recall that, for any $s \times t$ matrix $B = [b_{ij}] = [\mathbf{b}_{*1} \dots \mathbf{b}_{*t}]$, its *transpose* B^T

is the $t \times s$ matrix $B^T = [b_{j,i}] = \begin{bmatrix} \mathbf{b}_{*1} \\ \vdots \\ \mathbf{b}_{*t} \end{bmatrix}$ obtained by taking i th row of B^T to be the i th column of B .) The

number of linearly independent conservation laws is $m - \text{rank}(S)$. For the reaction system in Exercise 11.8, how many such independent conservation laws will there be? Can you find any others than those in part (3) of this exercise above?

Solution. By part (2), $\text{rank}(S) = 3$, so $m - \text{rank}(S) = 6 - 3 = 3$. Thus, one expects two more conservation laws in addition to the one from part (4) of this Exercise. By observation, these are given by $\frac{d[X_2]}{dt} + \frac{d[X_6]}{dt} = 0$, and $\frac{d[X_3]}{dt} + \frac{d[X_4]}{dt} + \frac{d[X_5]}{dt} + \frac{d[X_6]}{dt} = 0$. In [6], these three conservation laws are related to certain elements (“total amounts”) of the orthogonal space to $\text{Span}(S)$. (Note: For those familiar with fundamental results in linear algebra, $m - \text{rank}(S)$ is the dimension of $N(S^T)$, and this problem can be approached using standard techniques for finding bases for nullspaces.) \square

Exercise 11.10 (Optional Project; Presumes Familiarity with Linear Algebra). Following a large body of research on the topic of determining the steady states (or positive steady states) of (bio)chemical reaction networks, the authors of [6] use further analysis of this setting and the polynomial map $f_k([\mathbf{X}])$ to provide a condition that prevents the existence of multiple positive steady states in terms of the injectivity of $f_k([\mathbf{X}])$ via a condition on the determinant of the associated linear Jacobian map (matrix) of $f_k([\mathbf{X}])$. This condition can be assessed effectively by using symbolic computation software like Mathematica, Maple, or MATLAB; a version of the steps/pseudocode appears in [6, Section 11]. Use, for example, MATLAB to implement this procedure outlined in [6, Section 11] to determine the potential for a biochemical reaction network of your choosing to have multiple positive steady-state solutions.

Solution. Solutions will vary. One can try out and/or test one’s implementation by using it to check the outcomes described for the networks, as in [6, Examples 9.2, 9.4, 9.5]. \square

Exercise 11.11 (Magnetic Resonance and \mathbb{R}^3).

1. For an overview of the underlying principles of NMR illustrated in the 2D case of a compass needle oscillating back and forth in the plane, please see the short video⁵: <http://www.youtube.com/watch?v=1OrPCNVSA4o>.
2. Now, to pass from the compass needle to the 3D situation discussed in this chapter, let's get oriented⁶ with 3D space. To view and work with a standard orientation (one that obeys the "right-hand rule") of 3D space \mathbb{R}^3 given by X -, Y -, and Z -axes, please access the online interactive demonstration "Sets of Linear Combinations and Their Images Under Linear Transformation"⁷ at <http://demonstrations.wolfram.com/SetsOfLinearCombinationsAndTheirImagesUnderLinearTransformation/>. Set the coordinates of vector \mathbf{a} there to be $a_1 = 10$, $a_2 = 0 = a_3$; those for vector \mathbf{b} to be $b_1 = 0 = b_3$, $b_2 = 10$; and for vector \mathbf{c} , $c_1 = c_2 = 0$, $c_3 = 10$. Set the coordinates of point \mathbf{p} to be $p_1 = 0 = p_2 = p_3$, and remove the box checks for "predetermined vector \mathbf{u} " and "predetermined vector \mathbf{v} " (so that these do not display). Set the sliders for r_1, r_2, r_3 at 0. The entries for matrix A may be set to any values at present. Then, in the display window, the blue vector lies along the X -axis, the purple vector lies along the Y -axis, and the green vector lies along the Z -axis. Each vector has length⁸ 10 (chosen so the visualization will be clear).
 - (a) Alter the lengths of vectors \mathbf{a} , \mathbf{b} , and \mathbf{c} by changing the entries a_1, b_2 , and c_3 , respectively, to see the impact of multiplication by a scalar on these vectors. Changing a_1 to be 20, for instance, scales the previous blue vector by a factor of 2.

Solution. Answers vary, as readers play with the settings. Only positive multiples are allowed by the settings, however.

- (b) Return again to the settings given initially in part (a) above (e.g., $a_1 = 10, a_2 = 0 = a_3$, etc.). Repeat similar operations, but by using the sliders for r_1, r_2 , and r_3 . As per the caption on the demonstration, you should see green dots representing the corresponding integer multiples of each of vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$. (Notice that $r_1 = 1$ yields two green dots, one for the blue vector \mathbf{a} , and one for its negative $-\mathbf{a}$. Notice also that the green dots reflect the alternate representation of vectors as arrows emanating from the origin and as points corresponding to the locations of the arrowheads.)

Solution. Answers vary, as readers play around with the settings. Readers will see green dots representing the negatives of the given scalar multiples, as well as additional lattice points allowed up to the window settings.

- (c) Return again to the settings given in part (a) above ($a_1 = 10, a_2 = 0 = a_3$, etc.). Set $r_1 = 1$. Now, implement the stretching (scalar multiplication) of the blue vector \mathbf{a} by a factor of two, this time by applying the linear transformation associated with taking the matrix product $A\mathbf{x}$ of matrix A with vector \mathbf{x} , here for $\mathbf{x} = \mathbf{a}$, by setting $A_{1,1} = 2$, and leaving all other $A_{i,j} = 0$. As per the caption on the demonstration, you should see black dots, representing the images of the green dots, appear. Check that this coincides with the effect in part (a) of changing a_1 to equal 20. What vector does the additional black dot represent?

Solution. The vector $(-20, 0, 0)$.

3. Return again to the settings given initially in part (a) above ($a_1 = 10, a_2 = 0 = a_3$, etc.). In the initial description above of magnetic resonance, the vectors \mathbf{M} and \mathbf{B}_o will lie along the Z -axis, while radiation will be applied along the X -axis. The vector \mathbf{M} will be some length that differs from that of \mathbf{B}_o . Without regard to what might be actual "reasonable" lengths, represent this situation by setting the blue vector and

5. The corresponding software illustrated there is also freely available.

6. Pun intended.

7. If you have not previously used Wolfram demonstrations, you may need to download a free viewer from Wolfram; this is quick and easy. Follow their instructions as prompted.

8. More generally, we take the *length* $\|\mathbf{x}\|$ of a vector \mathbf{x} with components $x = (x_1, \dots, x_n)$ to be $\|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_n^2}$, generalizing the familiar notion of the distance between two points in a plane when one of the points is the origin.

purple vector both to have length 5, with the blue vector lying on the X -axis and setting the purple vector (representing \mathbf{M}) to lie on the Z -axis. Set the green vector, representing \mathbf{B}_0 , to have length 10, and keep it on the Z -axis. Set $r_2 = 1$. The (eventual) “rotation” of \mathbf{M} into the Y -axis is accomplished by a linear transformation.⁹ Implement this by setting $A_{1,1} = 1 = A_{2,2} = A_{3,2}$, and all other $A_{i,j} = 0$. Observe the resulting change from green dots to black dots.

Solution. Simply follow directions here and observe the rotation via the black dots lying on the Y -axis.

4. The columns of A (as in part (b) above) determine the images, under the transformation given by matrix multiplication of the original unit (i.e., length 1) vectors in each coordinate axis direction (respectively, $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$), and the transformation respects the scaling of each of these unit vectors that produces the vectors \mathbf{a} , \mathbf{b} , and \mathbf{c} (i.e., blue, purple, and green, respectively). In essence, the first two columns of A say to leave the unit vectors in the X - and Y -directions as is, but to move the unit vector in the Z -direction into the Y -direction. Because we used the setting $r_2 = 1$, we see the effect of transforming the purple vector \mathbf{M} to the Y -direction.

- (a) Implement settings that will demonstrate the impact of the same linear transformation (matrix A as in part (3)) on the green vector.

Solution. Set $r_1 = r_2 = 0$, and $r_3 = 1$.

- (b) Implement settings that would return the blue, purple, and green vectors to the respective X -, Y -, and Z -axis directions, but then “rotate” the blue vector to the Y -axis direction.

Solution. Return again to the settings given initially in part (a) above (e.g., $a_1 = 10, a_2 = 0 = a_3$, etc.). Then set $r_1 = 1$, keep $r_2 = 0 = r_3$, and set $A_{1,2} = 1 = A_{2,2} = A_{3,3}$, with all other $A_{i,j} = 0$.

- (c) Repeat the previous part, but rotate the blue vector to lie on the line $Y = X$ in the XY -plane.

Solution. Return again to the settings given initially in part (a) above. Then set $r_1 = 1$, keep $r_2 = 0 = r_3$, and set $A_{1,2} = 1 = A_{2,1} = A_{2,2} = A_{3,3}$, with all other $A_{i,j} = 0$.

5. While the effect in the situations above is to apparent rotation of a selected vector into another, the full transformations performed in parts (3) and (4) are in fact examples of *projections*: linear transformations that collapse a higher-dimensional space onto a lower-dimensional one inside it (here collapsing \mathbb{R}^3 onto a copy of a plane \mathbb{R}^2 in \mathbb{R}^3). For instance, in part (3), matrix multiplication by A sends $\text{Span}((1, 0, 0), (0, 1, 0), (0, 0, 1))$ to $\text{Span}((1, 0, 0), (0, 1, 0))$. Linear transformations that are *rotations* preserve the length of any vector, a property not shared by projections.

For an illustration and the definition of true rotation matrices that perform general rotations around X -, Y -, and Z -axes, go to <http://demonstrations.wolfram.com/VectorRotationsIn3D/>. Two points (blue and red) on a sphere define a vector that can then be rotated around each of the axis by choosing the axis and moving the rotation slider from 0° to 360° . The remaining buttons (“vector start point,” “vector end point”) and sliders (“ ϕ start,” “ θ start”) provide the means of changing to a new vector.¹⁰ To check your understanding of the matrix mechanisms in play, using the rotation matrices provided with the demo:

- (a) To what component of the demo does the γ correspond?

Solution. The notation γ is the entry for the angle of rotation in the three canonical rotation matrices given in the demo’s explanation. It is controlled by the rotation slider.

- (b) Write out the matrix that would perform a rotation of 90° around the Y -axis.

Solution. Use the Y -axis rotation matrix shown in the demonstration, and take $\gamma = \frac{\pi}{2}$.

9. As are all true rotations, but see part (5) of Exercise 11.11.

10. These move the blue and red points around using *spherical coordinates* that are defined using ϕ and θ , but one does not have to know the theory of spherical coordinates to play with this demo.

- (c) Write out a matrix that would perform the rotation that takes the unit vector $\mathbf{e}_3 = (0, 0, 1)$ in the Z -direction onto the positive Y -axis. Under the corresponding linear transformation, where would the unit vector $\mathbf{e}_2 = (0, 1, 0)$ be carried?

Solution. Use the X -axis rotation matrix with $\gamma = -\frac{\pi}{2}$. In this case, the Y -axis is carried to the negative Z -axis direction.

- (d) Compare the true rotation matrix you just created in part (5(c)) to that of the matrix employed in the setting of part (3) for magnetic resonance, wherein radiation is applied along the X -axis, the magnetic field \mathbf{B}_o lies along the Z -axis, and the vector \mathbf{M} rotates from the \mathbf{B}_o direction to the Y -axis. How do these differ in their effects on the standard unit vectors $\mathbf{e}_1 = (1, 0, 0)$, \mathbf{e}_2 , and \mathbf{e}_3 , and on the X -axis, \mathbf{M} , and \mathbf{B}_o ? Is either a desirable representation of this physical situation?

Solution. As may be observed through viewing the columns, or by carrying out matrix multiplication, for example, $\mathbf{A}\mathbf{x}$, for the relevant matrices A and vectors \mathbf{x} (written as columns), the rotation matrix fixes the X direction and sends $(0, 1, 0)$ to $(0, 0, -1)$ and $(0, 0, 1)$ to $(0, 1, 0)$. In particular, it sends both \mathbf{M} and \mathbf{B}_o to the Y -axis. The other matrix fixes the X - and Y -directions, but carries the Z to the Y -axis, hence moves both \mathbf{M} and \mathbf{B}_o to align with the Y -axis. This is the consequence of linearity! because the \mathbf{B}_o vector would not be moving, it may seem neither transformation is desirable. On the other hand, viewed as a transformation from a copy of \mathbb{R}^3 in which \mathbf{B}_o and X label a coordinate system for which the applied incident radiation is orthogonal to the magnetic field vector \mathbf{B}_o , and \mathbf{M} is parallel to \mathbf{B}_o , the desired rotation of \mathbf{M} is indeed implemented by either transformation (more minimally by the first).

6. Having played with the 3D setting of \mathbb{R}^3 , now view a more dynamic simulation of the 3D phenomenon of magnetic resonance at <http://www.mathworks.com/matlabcentral/fileexchange/41295-nuclear-magnetic-resonance-simulation>.

Solution. Access and observe the simulation as instructed. \square

The simulation can be downloaded and accessed through MATLAB, but a video of its implementation is also available at the site.

Exercise 11.12 (Optional; for Those Not Faint of Heart). To better understand the process and complexity of accounting for J -coupling and to get a taste of some software used to interpret and process NMR spectra more generally, consider downloading a copy of the Chenomx NMR Suite (<http://www.chenomx.com>). Follow the easy instructions for downloading and selecting the evaluation option.¹¹ Open the accompanying tutorial PDF to Chapter 4 (Spin Simulator), and read (and, as inclined, follow) the detailed instructions for simulating the spectrum for valine using the sample files that accompany the download. This “simple” example runs over three pages and takes 25 steps. The subsequent tutorial, for a more “complex” example (proline) requires six pages of instructions!

Solution. Left to the reader; following the instructions through the Chenomx download. \square

Exercise 11.13.

1. Access the Biological Magnetic Resonance Data Bank (BMRB) metabolomics database (<http://www.bmrb.wisc.edu/metabolomics/>) to easily obtain the NMR spectrum of a molecule, such as glutaconic acid, by entering the name into the “enter search: field” and clicking on the “data” tab.
2. It is important to note the conditions under which the spectrum was acquired, because the conditions of your sample you may be different and these variables may affect the spectrum that results. According to the data from part (1) of this exercise, what was the pH at which the entry for glutaconic acid was acquired, and at what temperature?

Solution. Glutaconic acid was acquired at a pH of 7.4 and a temperature of 298 K.

For comparison purposes, the experimenter should collect the urine sample using the same conditions.

11. With this free option, however, work cannot be saved.

3. The use of the same reference molecule is also important, because commonly used reference molecules can have slightly different chemical shifts. According to the entry in the BMRB, what is the reference molecule for the listed acquisition of glutamic acid?

Solution. The reference molecule is DSS. □

Exercise 11.14 (Biomarkers for Glue Sniffers, I).

1. To identify the molecules that are associated with the set of peaks (at 7.83, 7.64, 7.55, and 3.96 ppm) where the signals increased, log into the metabolomics section of the BMRB <http://www.bmrb.wisc.edu/metabolomics/>, and “Search 1D Peak Lists.” Then enter the individual ppm values into the search box individually, with only spaces separating the individual numbers: 7.83 7.64 7.55 3.96. After the numbers have been entered, press the “Peak Search” button. How many entries are in the resulting table?

Solution. Results varied between 257 and 261, at the time of writing.

2. Verify that only three of the molecules have an optimal \hat{O} Peak Match \acute{o} score of 1.00: guanine, hippuric acid, and glycolate, while all the other hits have a score of 0.67 or lower. To determine which of these three molecules is associated with the peaks, look up the individual spectra of the molecules. Clicking on the name of the molecule links to a page that provides a wealth of information on the molecule.

Solution. A table appears listing the molecules by their matching scores, the closest matches first, and the table results coincide with the claims made in the exercise.

3. To access the NMR chemical shift data for guanine, for example, click on the “bmse000090” tab near the top of the web page. Guanine is a very common chemical species in biological systems because it is a precursor of guanosine, one of the five nitrogenous bases used to make DNA and RNA molecules. The molecule has the chemical formula $C_5H_5N_5O$, but only one of the hydrogen atoms is attached to a carbon atom. The other four hydrogen atoms are attached to nitrogen atoms (primary and secondary amines) and are in exchange with water. Consequently, the 1H NMR spectrum of guanine consists of a single signal.¹² Explain why the chemical shift of the carbon bound proton is at 7.6 ppm. Because this is similar to (<0.05 ppm away from) our signal at 7.64 ppm, we can, hypothesize that guanine levels in urine may be affected by a component in glue.

Solution. Clicking on the tab as instructed provides a wealth of information, including a picture of the spectrum for guanine; the chemical shift is one parameter shown.

4. Another possibility is that glycolate levels in urine are affected as well. Glycolate is a small molecule with a chemical formula of $C_2H_4O_3$, which is an intermediate in the photorespiration pathway of plants but can also be produced by humans, is commonly used in skin care products, and is found in urine. In glycolate, two of the protons are bound to oxygen atoms and not observable by 1H NMR, but the other two protons are bound to the same carbon and form a methylene group ($-CH_2-$). Because the two hydrogen atoms are virtually identical, the chemical shift of both is 3.93 ppm, and has a similar chemical shift to our signal (<0.05 ppm away from) at 3.96 ppm.
5. Explain why, thus far, neither guanine or glycolate explain all the signals that increase in the urine of the glue-sniffing individuals.

Solution. Using, for example, the spectra of guanine and glycolate, the sum of the spectra does not yet produce the observed spectrum in [8]; not all peaks nor peak heights are accounted for.

6. Hippuric acid remains another possibility. Hippuric acid is an acyl glycine molecule that is formed when benzoic acid and the amino acid glycine are linked together via an amide bond. A diet that includes a large amount of foods rich in phenolic compounds, such as tea, wine, and fruit juices, leads to increased amounts of hippuric acid in urine. Hippuric acid has a chemical formula of $C_9H_9NO_3$ and has four chemical shifts:

12. The 1D 1H , 2D 1H - 1H TOCSY and 2D 1H - ^{13}C heteronuclear single quantum spectroscopy (HSQC) spectra on the page also confirm this; see the appendix [7] for additional information about these alternative spectra.

7.808, 7.618, 7.53, and 3.944 ppm. Are these values similar to those observed for the signals that increased in the glue-sniffing individuals?

Solution. All four values are very similar to those observed for the signals that increased in the glue-sniffing individuals. Consequently, the best hypothesis we have is that hippuric acid levels in urine increase when a person sniffs glue. It is important to note, however, that we cannot exclude the hypotheses that guanine and glycolate levels increase as well.

7. Remarks: Those who have downloaded the evaluation copy of Chenomx can also go to the library manager (see the tutorial PDF in Exercise 11.12 of this chapter) and upload the 500-MHz compound set (the NMR setting used in [8]), and then check the library for spectra for each of guanine, glycolate, and hippuric acid. Currently, one of these is not in the database exactly as is, but you can use BMRB or Human Metabolome Database (HMDB) to find it.¹³ The remaining two are under slightly different names (try clicking on the “alternate names” option for the pages of your best guesses).

8. Use MetaCyc, KEGG, or another similar database to investigate links between toluene and hippuric acid via metabolic pathways. Does this provide further evidence for hippuric acid as a biomarker for glue sniffing?

Solution. Results will vary. For example, a search on MetaCyc for toluene shows it is converted into hippuric acid via metabolic pathways and benzoic acid, so there is, *a priori*, an established link, and one can expect hippuric acid to appear “downstream” (as a degradation product) of toluene.

9. Of the three possibilities (guanine, glycolate, and hippuric acid), for which one is the hypothesis that it increases among glue-sniffing individuals best supported? Can the hypothesis that the levels of the other two increase be ruled out?

Solution. The best fit in terms of peaks and chemical shifts is the hippuric acid spectrum; the other two cannot be completely ruled out, because they may be appearing as part of the signal (linear combination). □

Exercise 11.15 (Biomarkers for Glue-Sniffers, II). Which molecule has decreased levels in glue sniffing individuals? Recall that signals at 4.06 and 3.05 ppm decreased in value in people that identified as glue sniffers.

Solution. Using the BMRB and following steps as in the previous exercise, one can determine that citrate is reduced in the glue sniffers. □

Exercise 11.16.

1. Check that the matrices (from part (5) of Exercise 11.11) that implement rotations around the X -, Y -, and Z -axes in \mathbb{R}^3 are orthogonal. (*Hint:* Use one or more basic trigonometric identities.)

Solution Sketch. An easy way is to use the definition of an orthogonal matrix directly, plus the fact that $\cos^2(\gamma) + \sin^2(\gamma) = 1$, to show that the columns/rows of each matrix are orthonormal, by direct calculation. For example, the columns of the rotation matrix about the Z -axis are $(\cos(\gamma), \sin(\gamma), 0)^T$, $(-\sin(\gamma), \cos(\gamma), 0)^T$, $(0, 0, 1)^T$. From the definition of the matrix product (vector dot product), the dot product of the third column with itself is 1, while the dot product of the third column with the other two is 0. The dot product of the first column with itself is $\cos^2(\gamma) + \sin^2(\gamma) = 1$, as is true for the second column. Finally, the dot product of the first two columns is $\cos(\gamma)(-\sin(\gamma)) + \sin(\gamma)\cos(\gamma) + 0 = 0$; therefore, by definition, the given rotation matrix is orthogonal.

2. For later use, it is helpful to note¹⁴ that, for any $r \times s$ matrix B and any $s \times t$ matrix C , $(BC)^T = C^T B^T$. Presuming this relationship, if there is a third matrix D of size $t \times u$, find an expression for the $u \times r$ matrix $(BCD)^T$ in terms of B^T , C^T , and D^T .

Solution. $(BCD)^T = ((BC)D)^T = D^T(BC)^T = D^T C^T B^T$.

13. A direct link to the latter and other external databases will appear on the Chenomx page for your chosen metabolite.

14. For a proof, one can compare the j th column of each side.

3. Show that if A is orthogonal, so is A^T .

Solution. One approach: Check that $(B^T)^T = B$, for any matrix B , so $(A^T)^T A^T = A A^T = I$, because A is orthogonal, while likewise $A^T (A^T)^T = A^T A = I$. Thus A^T is also orthogonal.

4. Use $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}^T}$ for any row vector \mathbf{x} , or, equivalently, $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \cdot \mathbf{x}}$ for any column vector \mathbf{x} , to show Theorem 11.1(3).

Solution. Because A is orthogonal, $A^T A = I$. Then $\|A\mathbf{x}\|^2 = (A\mathbf{x})^T (A\mathbf{x}) = (\mathbf{x}^T A^T) (A\mathbf{x}) = \mathbf{x}^T (A^T A) \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|^2$, and the result follows. \square

Exercise 11.17 (SVD in MATLAB). Calculation of the singular value decomposition (SVD) of a matrix A entered in MATLAB is given simply by the command `[U,S,V]=svd(A)`; the command `svd(A)` alone produces the list of singular values. In MATLAB, enter `A = [1.4015 -1.0480; -0.4009 1.0133]`; `eigshow(A)` to obtain an interactive demonstration of the geometric content of the SVD for the 2×2 matrix A in terms of its singular values and right and left singular vectors. In the resulting pop-up, follow the help button instructions (for `svd` mode) to complete the illustration of a figure illustrating the SVD for A in terms of distortion of the unit circle in \mathbb{R}^2 . Then calculate the singular values for A and the actual matrix factorization (SVD) of A . Compare your results with the illustration in the summary article [9].

Solution. Entering `A` and `eigshow(A)` as instructed produces a pop-up in which one obtains a new coordinate frame, with the effect of A represented by the stretching of the unit circle by the singular values computed below. Entering the commands as instructed produces the singular values and the SVD as follows:

```
svd(A)
ans=
2.0000
0.5000
[U,S,V]=svd(A)
U=
-0.8660 0.5000
0.5000 0.8660
S=
2.0000 0
0 0.5000
V=
-0.7071 0.7071
0.7071 0.7071
```

The resulting picture from `eigshow(A)` is the same as that shown in [9], another helpful survey article on the SVD and its importance well worth reading. Check out the additional references to applications given there. \square

Exercise 11.18 (PCA Example with ^1H NMR Data). Refer to [10, Section 2.3].¹⁵ Three types of samples (plasma, urine, and saliva) were drawn once from 150 subjects.

1. For each type of sample, the relevant data matrix was either 150×409 or 150×493 . What formed the rows and columns of each data matrix X , what was the size of the data matrix, and how was the data entry x_{ij} defined?

Solution. Each X had 150 rows corresponding to the 150 subjects. After the exclusion of certain regions on the chemical shift axis of the spectra, the remaining regions were defined as certain buckets of size 0.02 ppm or some variable region sizes in which large variations in chemical shift were expected; such regions were determined through an in-house MATLAB routine of the authors. For the plasma and saliva data, this resulted in matrices of size 150×493 . The bucketed regions gave 493 measurements (variable values) for each subject

15. Available free via Googling or from this chapter's Supplementary Materials.

(spectrum). The columns represented these different bucket regions (discretized chemical shift axis). Each entry x_{ij} represented the concentration of the i th spectrum over the j th bucket/variable along the relevant chemical shift axis. The same is true for the urine data, but here X was size 150×409 . *Note:* These matrices X are the transposes of types described earlier, that is, with subjects/spectra corresponding to columns. \square

2. These data were further separated into a testing set of 50 subjects and a remaining training set of 100. After some rescaling, principal component analysis (PCA) was applied to the training data for the remaining three types of sample data, so to X of size 100×493 or 100×409 , as relevant. We do not learn the value of $r = \text{rank}(X)$, but it is no larger than $100 = \min(100, 493)$ ($\min(100, 409)$), so there are at most $r = 100$ principal components; likely many fewer. The first two principal components, PC 1 and PC 2, were examined. We do not know what the vector representations of these as elements of \mathbb{R}^{100} would be, but by SVD PC 1 and PC 2 are orthogonal vectors and it makes sense to represent them as two new coordinate axes vectors (e.g., the x -axis for PC 1 and y -axis for PC 2). Refer to [10, Figure 1]. As shown there, the *scores plots* relative to PC 1 and PC 2 are given as follows: Each point on the scores plot corresponds to a subject/spectrum, an element of \mathbb{R}^{493} or \mathbb{R}^{409} . PCA projects each spectrum onto the span $\text{Span}(U_1)$ for U_1 as in Theorem 11.2. As per previous remarks and Exercise 11.20 below, in fact $\text{Span}(Y) = \text{Span}(PC\ 1, PC\ 2, \mathbf{y}_3, \dots, \mathbf{y}_r) = \text{Span}(U_1)$, an r -dimensional space, for $Y = XV$, where V arises from Theorem 11.2 by setting $A = X$. In the new principal coordinates, each spectrum is represented uniquely as a linear combination $\sum_{j=1}^r \alpha_j \mathbf{y}_{*j}$ of the columns of Y (or U_1 , replacing α_j by α_j/σ_j , using the singular values σ_j , $1 \leq j \leq r$). If we take only the first two multiples α_1 and α_2 , that is, the coordinates for the principal component axes PC 1 and PC 2, then from a spectrum one gets an ordered pair (α_1, α_2) . These pairs form the score plots. What information relative to the participants' gender do you glean from these three plots?

Solution. The plasma PCA plot shows clear clustering of the data by gender, indicating that gender can be distinguished by the plasma spectrum data. The other two are inconclusive and suggest that, modulo the total variation explained, gender is not distinguished by the first two principal components of urine or saliva data (see below).

3. As per [10, Figure 1], for each sample type, what percent of the variation was explained by the first principal component PC 1? By PC 2? Do these results suggest to you that reducing to $\text{Span}(PC\ 1, PC\ 2)$ was sufficient to capture the behavior of the data?

Solution. For plasma, it is approximately 54% for PC 1 and 15% for PC 2, for a total of approximately 70%. For urine they are 25% and 11%, respectively (26% total); for saliva, 30% and 19% (49% total). Answers on the sufficiency of capturing the data will necessarily vary with experience and context, but at the very least we can say the following: the clear clustering by gender and the relatively high total variance for the plasma data versus the relatively low total variance of the urine data suggest further testing of the impact of a third or higher principal component may be a good idea. However, because no remaining principal component can explain more than 11% of the total variation, we may not see much additional information. \square

Exercise 11.19 (Application of PCA to Biomarkers for Glue Sniffers). Among the species identified by prior analysis of the urine spectra in [8], PCA identified three of greatest variability. What do the PCA score plots for two of the principal components (see [8, Figure 4a]) and three of the principal components (see [8, Figure 4b]) imply? If PCA ordered metabolites by the greatest variability, why were PC 1 and PC 3 chosen for Figure 4a instead of PC 1 and PC 2? Are the results consistent with Exercises 11.14 and 11.15?

Solution. As per [8, Figures 3 and 4; p. 123], PC 1 and PC 3 are hippuric acid and citrate; creatinine is ruled out due to too much variability between individuals, but the 3D PC plot still shows clustering with these three used. The score plots demonstrate groupings along the species by controls and glue sniffers, and further grouping of the 2W glue sniffer and the rest. This is consistent with the results of Exercises 11.14 and 11.15. \square

Exercise 11.20 (PCA by SVD). Let X be an $m \times n$ data matrix X of m measurement types (types of metabolites) with n trials (spectra), and let $X = U\Sigma V^T$ be the SVD of X as in Theorem 11.2. Set $Y = XV$. Recall that by the assumptions of Theorem 11.2, V is orthogonal.

1. Explain why, for each j with $1 \leq j \leq r$, the variances $\text{var}(\mathbf{y}_{*j})$ are successively maximized (with the largest variance associated to \mathbf{y}_{*1} , the next largest to \mathbf{y}_{*2} , and so on. (*Hint:* $\mathbf{y}_{*j} = X\mathbf{v}_{*j}$. Now compare with Theorem 11.2(3).)

Solution Sketch. For each $1 \leq j \leq r$, $\text{var}(\mathbf{y}_{*j}) = \frac{1}{n-1} \|\mathbf{y}_{*j}\|^2 = \frac{1}{n-1} \|X\mathbf{v}_{*j}\|^2$, a nonzero scalar multiple of $\|X\mathbf{v}_{*j}\|^2$. because V is orthogonal, $\|\mathbf{v}_{*j}\| = 1$. Now, for $j = 1$, Theorem 11.2(3) implies \mathbf{v}_{*1} maximizes $\|X\mathbf{z}\|^2$ over all \mathbf{z} with $\|\mathbf{z}\| = 1$. Thus, in particular, \mathbf{v}_{*1} maximizes $\|X\mathbf{v}_{*j}\|^2$ over all $1 \leq j \leq r$; therefore, \mathbf{v}_{*1} maximizes the values $\text{var}(\mathbf{y}_{*j})$, for all $1 \leq j \leq r$. Moreover, it follows from Theorem 11.2(4) and the orthogonality of U that $\|X\mathbf{v}_{*j}\| = \|\sigma_j \mathbf{u}_{*j}\| = \sigma_j \|\mathbf{u}_{*j}\| = \sigma_j$, for each $1 \leq j \leq r$. Thus, σ_1 must be the largest among all the values σ_j , $1 \leq j \leq r$.

Now, V is orthogonal, so $\mathbf{v}_{*1}^T \mathbf{v}_{*2} = 0$. Thus, for any real number λ_1 , $(\lambda_1 \mathbf{v}_{*1})^T \mathbf{v}_{*2} = 0$, that is, \mathbf{v}_{*2} is orthogonal to $\text{Span}(\mathbf{v}_{*1})$. Repeating the argument above with \mathbf{v}_{*2} in place of \mathbf{v}_{*1} , and restricting the index j to $2 \leq j \leq r$, one can apply Theorem 11.2(3) and (4) to show the variances $\text{var}(\mathbf{y}_{*j})$, $2 \leq j \leq r$, are maximized, with the largest variance associated with \mathbf{y}_{*2} . The corresponding singular value σ_2 satisfies $\sigma_2 \leq \sigma_1$, but is otherwise the largest among the σ_j , $2 \leq j \leq r$.

For $j = 3$, \mathbf{v}_{*3} , the orthogonality of V implies \mathbf{v}_{*3} is orthogonal to both \mathbf{v}_{*1} and \mathbf{v}_{*2} ; then $(\lambda_1 \mathbf{v}_{*1} + \lambda_2 \mathbf{v}_{*2})^T \mathbf{v}_{*3} = (\lambda_1 \mathbf{v}_{*1}^T + \lambda_2 \mathbf{v}_{*2}^T) \mathbf{v}_{*3} = (\lambda_1 \mathbf{v}_{*1}^T \mathbf{v}_{*3}) + (\lambda_2 \mathbf{v}_{*2}^T \mathbf{v}_{*3}) = 0 + 0 = 0$, so \mathbf{v}_{*3} is orthogonal to $\text{Span}(\mathbf{v}_{*1}, \mathbf{v}_{*2})$. One can thus proceed again as above, applying Theorem 11.2(3) and (4). Proceeding informally (or more formally with induction) in this way completes the proof. \square

2. Explain why the columns of Y are orthogonal. What does this imply about the covariances $\text{cov}(\mathbf{y}_{*i}, \mathbf{y}_{*j})$?

Solution. By Theorem 11.2(4), the columns of Y are $\mathbf{y}_{*j} = \sigma_j \mathbf{u}_{*j}$. because U is orthogonal, for $i \neq j$, $\mathbf{u}_{*i}^T \mathbf{u}_{*j} = 0$, so $\text{cov}(\mathbf{y}_{*i}, \mathbf{y}_{*j}) = \sigma_i \sigma_j \mathbf{u}_{*i}^T \mathbf{u}_{*j} = 0$.

3. Explain why $\text{Span}(Y) = \text{Span}(X)$. (*Hint:* Show $\text{Span}(Y) \subseteq \text{Span}(X)$. Explain why $\dim(\text{Span}(Y)) = r$, and also $\dim(\text{Span}(X)) = r$, so then $\text{Span}(Y) = \text{Span}(X)$ is forced.)

Solution. Each $\mathbf{y}_{*j} = X\mathbf{v}_{*j}$ is in $\text{Span}(X)$, because by Eq. (11.9) $X\mathbf{v}_{*j} = \sum_{k=1}^n \mathbf{x}_k v_{kj}$. Now, the nonzero columns of XV correspond to XV_1 , because $Xv_{*j} = 0$ for $r+1 \leq j \leq n$, by Theorem 11.2(4). Thus, $\text{Span}(Y)$ is the set of all linear combinations $\sum_{j=1}^r \alpha_j \mathbf{y}_{*j}$, and, since each $\mathbf{y}_{*j} \in \text{Span}(X)$, it is also true that $\sum_{j=1}^r \alpha_j \mathbf{y}_{*j}$ is contained in $\text{Span}(X)$. But, the r column vectors of Y are linearly independent (covariances are all 0), so the dimension of $\text{Span}(Y)$ is at least r , whereas $\dim(\text{Span}(X)) = \text{rank}(X) = r$, so $\text{Span}(Y) = \text{Span}(X)$.

4. Parts (1)-(4) above verify that $Y = XV$ satisfies the conditions for PCA, hence the columns \mathbf{y}_{*j} , $1 \leq j \leq r$, are, by definition, the principal components of X . Note that for the purpose of determining the principal components of X , it suffices to consider the matrix product $Y = XV_1$. Explain why the columns of U_1 are also often called the *normalized principal components of X* , and give an orthonormal set of vectors lying in the directions of greatest variation for the data X , given by $\sigma_1 \geq \dots \geq \sigma_r > 0$. Thus, the left singular vectors of X and the right singular vectors of X , both associated to the nonzero singular values $\sigma_1, \dots, \sigma_r$, provide the mechanisms for PCA.

Solution. The first observation regarding XV_1 is immediate by Theorem 11.2(4). Also by Theorem 11.2(4), for $1 \leq j \leq r$, $\mathbf{y}_{*j} = \sigma_j \mathbf{u}_{*j}$, so $\|\mathbf{y}_{*j}\| = \|\sigma_j \mathbf{u}_{*j}\| = \sigma_j \|\mathbf{u}_{*j}\| = \sigma_j \cdot 1 = \sigma_j$, because U is orthogonal. Thus, the “normalized” principal component vectors are $\frac{1}{\sigma_j} \mathbf{y}_{*j} = \mathbf{u}_{*j}$, giving unit vectors in the direction of the greatest variance of X , by the calculation used in part (2) above. \square

REFERENCES

- [1] Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, et al., The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 2014;42:D459-71. Available from: <http://metacyc.org>.
- [2] Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database: <http://www.genome.jp/kegg/pathway.html>, Kanehisa Laboratories.
- [3] Meyer C. *Matrix analysis and applied linear algebra*. Philadelphia, PA: SIAM; 2000.
- [4] Leon SJ. *Linear algebra with applications*. Upper Saddle River, NJ: Pearson/Prentice Hall; 2006.

- [5] Hodge TL. Metabolic pathway analysis: a linear algebraic approach. In: Robeva R, Hodge T, editors. *Mathematical concepts and methods in modern biology*. London, UK: Academic Press/Elsevier; 2013.
- [6] Feliu E, Wulf C. Preclusion of switch behavior in reaction networks. *Appl Math Comput* 2012;219:1449-67. Available from: <http://dx.doi.org/10.1016/j.amc.2012.07.048>.
- [7] Szymczyna BR. Appendix to this chapter: two- dimensional NMR.
- [8] Kwon B, Kim S, Lee DK, Park YJ, Kim MD, Lee JS. ¹H NMR spectroscopic identification of a glue sniffing biomarker. *Forensic Sci Int* 2011;209(1-3):120-5.
- [9] Moler C. Professor SVD. 2006-91425v00. Available from: <http://www.mathworks.com/company/newsletters/articles/professor-svd.html>.
- [10] Ramadan Z, Jacobs D, Grigorov M, Kochhar S. Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. *Talanta* 2006;68:1683-91. <http://dx.doi.org/10.1016/j.talanta.2005.08.042>.