Introduction to Spectroscopy and Data Science

How many dyes are in a collection of colorful samples?

Introduction

Absorbance spectroscopy – getting to know your spectrometer

Perhaps you've walked into a home improvement store or paint store and have seen the variety of available shades in color palettes and the mixing equipment that can precisely match an existing paint. There are many of colorful pigments in most paints, from white pigments like titanium dioxide to red ochre made of iron oxide. Or how about different colors of gelatin (like Jell-O) and chocolates (like M&Ms)? In most cases, a single dye does not get the desired color, so dyes are mixed and matched in industrial labs to make, for example, purple Jell-O from red and blue dyes. But just how many dyes are in these mixes? If you were a paint chemist or a food scientist with the mission to figure this out, how would you do this? Dyes are chemicals, and so chemical analysis, with a healthy dose of computing and data science, can reveal the secret.

This leads us to our next question: what techniques for chemical analysis do we have to help us solve this problem? What techniques do we have to try and identify atoms and molecules, and how do we probe how they behave or change over the course of a reaction or chemical process?

One of the most important techniques chemists use to answer these questions is spectroscopy. **Spectroscopy** is a technique that allows us to study a system by shining light on it. This technique takes advantage of the fact that atoms and molecules will absorb, emit, and scatter electromagnetic radiation (including light) in a manner that is based on their fundamental physical properties. Spectroscopy has been key in discovering and verifying many of the theories we have for how molecules function, and it is a commonly used technique for identifying specific molecules and how they change or behave during a chemical reaction or process. In fact, every molecule in the universe has a distinct spectroscopic 'signature' that one can use to identify it from light years away with a telescope– or right on your benchtop in lab with a **spectrometer**.

A spectrometer is a common chemical instrument that allows us to collect spectroscopic data. In the first half of your lab today (procedure part A–page 9) you will use your kits from Trimontana to build your own spectrometer with computer interface, learn about its components, and collect absorbance data on a set of dyes. After some practice runs with single dyes, the data will tell you which wavelengths of visible light are absorbed by a mix containing an unknown number of dyes. Computational data science will then allow you to count the dyes in the mix, so you can report your best-guess count. You'll be able to compare with other groups in the class who reproduce the same experiment and see if there is a consensus value. This is an important step in science because different people can get somewhat different answers even if they do everything right– there's always noise in the data!

While you are building your spectrometer, you and your class will discuss the components of the spectrometer and how it operates and work together to answer the following questions:

(1) What purpose does each of the following components in your spectrometer serve, and how do they work together to give you an absorbance measurement?
    a. The light source
    b. The slit and diffraction grating
    c. The swivel arm/detector

(2) How does our software determine an absorbance value from the information it receives from the spectrometer? How might scattering affect this measurement?

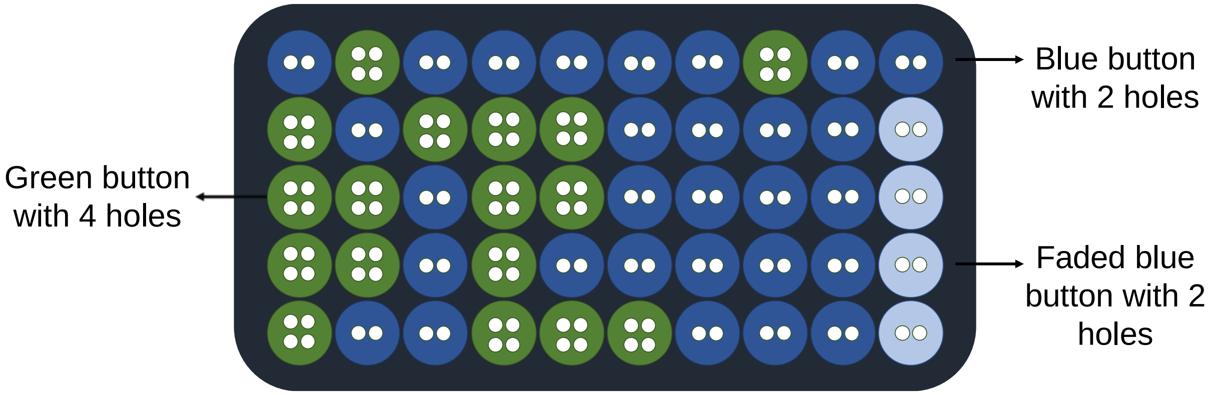
(3) List out 2-3 steps involved in collecting data (e.g. steps in the sample prep, operating the spectrometer, etc.). What aspects of these steps do you think are the most important for collecting consistent data? Where might you introduce any errors in your measurements?

Data analysis and singular value decomposition

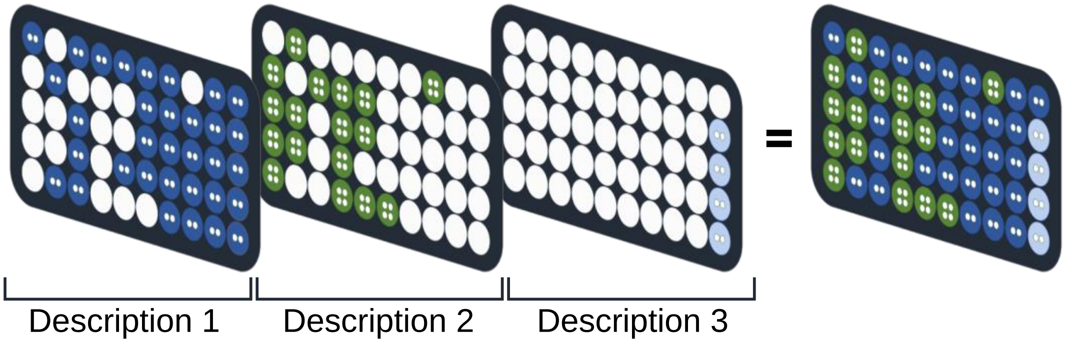
After we collect data in experiments, there can be a variety of techniques used to interpret what that data is telling us. Choosing an appropriate analysis technique is an essential step in trying to answer the specific questions you have about your system. If we want to know the concentration of a molecule that absorbs light at a known wavelength and molar absorptivity, we can use the Beer-Lambert law. If we want to know information about the rate constant for a reaction given a known reaction order and kinetics data, we can fit our data to a known rate law. Fundamental techniques like these are widely applicable and are commonly used in chemistry today, but what do chemists do when they do not know enough information about their system to use these techniques alone? What happens when our data set is too large to sort through manually? These questions become particularly relevant when studying new and complicated chemical systems and often require us to apply more advanced techniques to study properties of our system. This work often falls into the field of **Data Science**, and sometimes in chemistry, chemometrics.

Singular value decomposition (SVD) is an advanced data analysis technique that is useful for interpreting seemingly complicated data sets, like mixes with an unknown number of components. SVD effectively allows us to redescribe our data in a simpler format, helping highlight ‘how many pieces’ make up our data. It has broad applications ranging from image processing to recommender systems (back in 2009, it was a key component of an algorithm that won a prize for improving Netflix’s user recommendations for new movies). It remains a technique that is applied widely everywhere in science and engineering, including in the analysis of spectroscopic data in chemistry.

To start to understand how SVD is useful, let’s look at the following example. Say you have a collection of 50 buttons, like the one shown below:



If you were asked to describe your collection of 50 buttons, how would you do this? You probably would not list out every individual button in your collection, so what is the most concise way to describe your collection? For this collection, you might start by describing that you have 35 blue buttons with two holes (we will refer to this as description 1) and 15 green buttons with four holes (description 2). These two descriptions have already allowed us to accurately describe most of our collection. You could even go further and say that 4 blue buttons are faded (description 3), giving a full picture of your collection when you add up the three descriptions:



While this is simple for our 50-button collection, you can imagine that doing this on a collection of 1000 buttons with different features (e.g. different shapes, sizes, etc.) would be difficult. SVD is an algorithm that allows a computer to simplify large sets in the spirit of what we did above and can help you identify major features or ways to group your data (like our green and blue buttons), as well as highlighting noise (like our faded blue buttons). It keeps detecting groups when things get more complicated, say for example, if you have some green buttons with 2 holes and blue buttons with 4 holes as well, so your ‘colors’ and ‘holes’ get shuffled.

We will be using SVD today in the context of its application to data analysis in chemistry, so let’s now think about how this idea could be applied to analyzing spectroscopy data. In the context of physical chemistry, we will often collect spectroscopic data with features that are attributed to multiple chemical species or states. SVD is one technique that can help us distinguish how many unique features we have in our system and can help us distinguish how many unique features we have in our system and can be especially useful for interpreting data with a significant amount of noise.

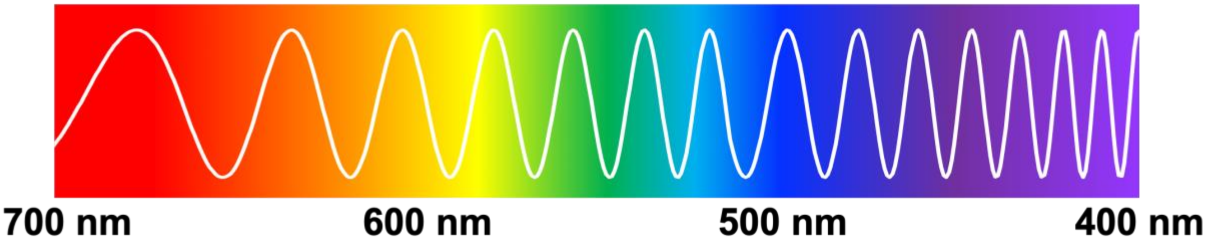


Figure 1. Visible light spectrum from longer wavelengths (red) to shorter wavelengths (violet)

Say you have a collection of 60 different spectra from various mixes of "X" pigments, where each spectrum is for a different mix. Even with just X=2 or 3 or 4 pigments, it's easy to make 60 different colors by mixing them in different ratios, so it's not obvious how many were used. If we are trying to reverse-engineer what "X" is, and our chemical knowledge tells us that each pigment has a unique signature of what wavelengths in the visible spectrum it absorbs (**Figure 1**), we can use SVD to sort through this data and determine how many unique absorbance features show up over the course of our reaction. This can be very difficult to do 'by hand' if we have X>2 or unwanted noise in our data. SVD can be applied to help us sort through our data using the following workflow, also outlined in **Figure 2** on the next page:

Data organization: Absorbance spectra are organized in a table or "matrix", with each column containing absorbance values for a single spectrum. For this example, as we move down through each row, we will be moving across the spectra, reading absorbance values at a particular wavelength. As we move across the columns, we will be moving from one mix to the next. Note that this figure only shows 5 columns, but you can imagine a table that extends out to 60 columns for 60 mixtures, or 200 rows if you have 200 wavelength points in each spectrum.

SVD: The singular value decomposition is computed, producing three unique outputs that work together to describe our original dataset.

SVD interpretation: All three outputs from our SVD are important, but the specific output we will focus on today are the singular values. Singular values help tell us how important each unique component found in our data is, and thus how many unique components we need to describe our data. For example, only two SVD singular values are large in the plot in Figure 2 at the bottom right, so there are likely only two major components in the mix.

When we compute the singular values for data collected from the real world (i.e. data that contains noise), we will often find a few singular values that are large, indicating those features are probably real components, and a collection of singular values that are close to zero, indicating they are more likely noise. When we exclude smaller singular values closer to zero, the remaining singular values help us determine how many unique components we need to describe our data. To make that decision, we have to pick a cut-off. One way to do this is by cutting off extra singular values after we've described a certain percentage of our data (e.g. 95%).

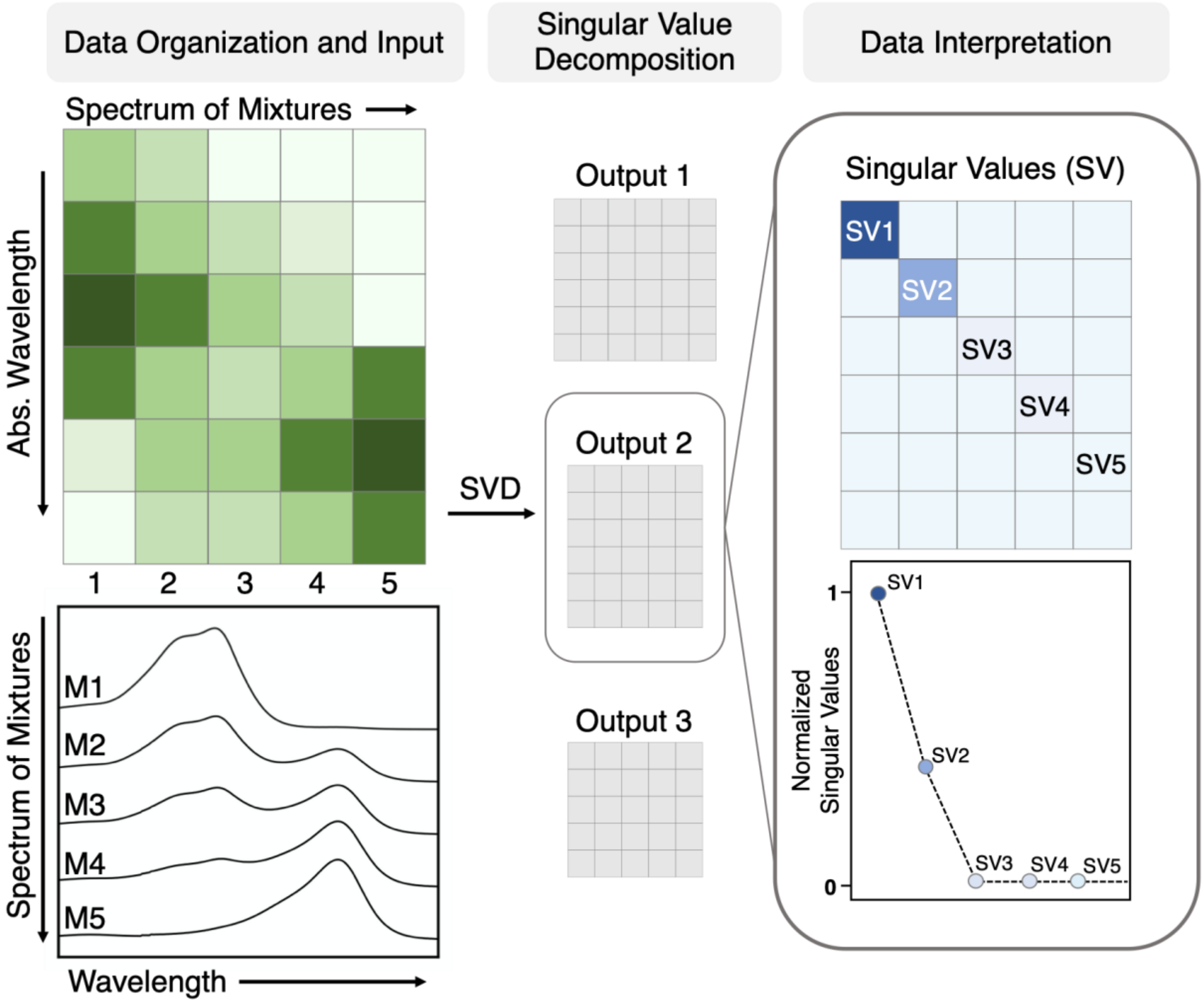


Figure 2. Illustration of singular value decomposition workflow for a collection of absorbance spectra. Data organization involves taking a data set of spectra (shown in black) and ordering absorbance values for each spectrum in columns, where rows correspond to the absorbance at a particular wavelength (our "matrix", shown in green). When SVD is performed on this collection of data, we get three outputs that will allow us to describe our data. The second output (shown in blue) will contain our singular values along the diagonal entries in the table, which will gradually decrease in magnitude as we move along the columns. Note that only five columns/spectra and singular values are displayed here, and that the actual input would include a 60-column table, with a singular value output table containing 60 columns as well.

A few important things to note. A more thorough understanding of how exactly SVD works requires some previous knowledge of linear algebra, which we will not require for the purpose of today's experiment. If you are interested in learning more about why we organize our data in the format outlined in the previous steps and the math behind how SVD works, you can visit the following series of videos on data-driven science and engineering here: Data Driven Science Science and Engineering Playlist – Brunton

Determining components in unknown mixtures of dyes

In the following examples and in the second half of your lab today, **we will apply SVD to determine how many unique components we have in mixtures of unknown dyes**.

For our first example, take the following absorbance data for 6 mixtures, M1 – M6 (**Figure 3**).

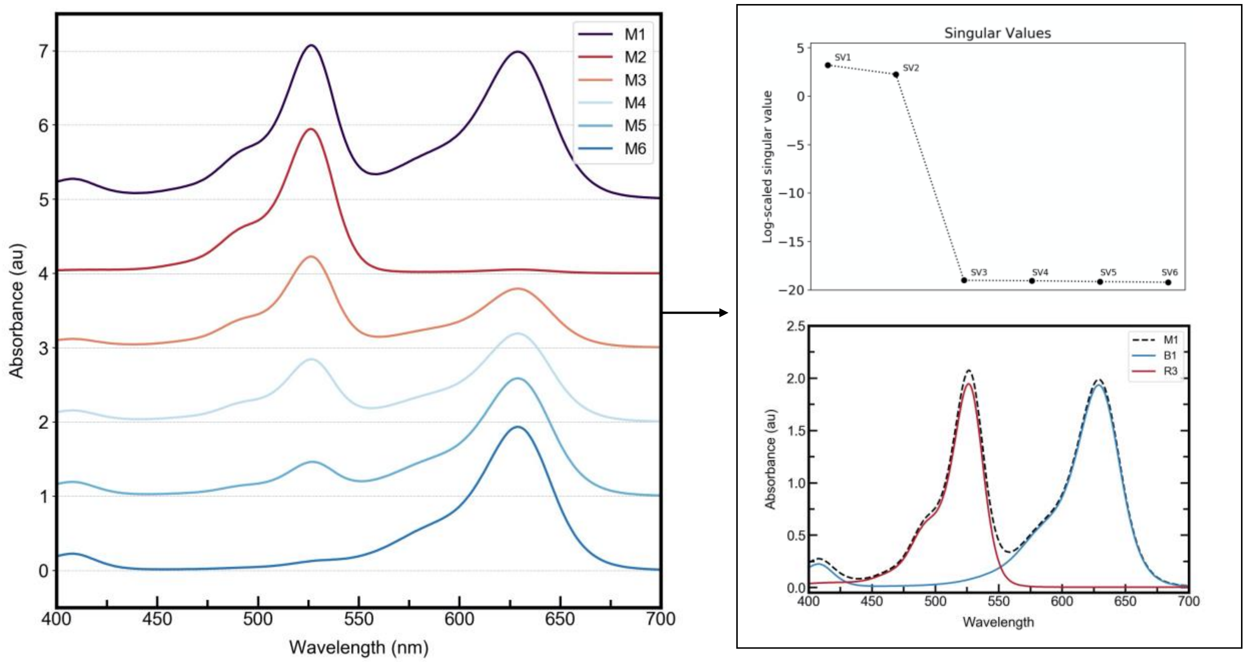


Figure 3. Absorbance data for six mixtures of dyes (left) and corresponding SVD results (top right). Singular values (SV1 – SV6) are plotted on a log scale to distinguish larger values (>1) more easily from smaller values (≈0) (note that this means singular values <1 are plotted as negative values). An overlay of the dyes used to produce these mixtures (Red 3=R3, Blue 1=B1) and mixture 1 (M1) is shown in the bottom right.

These mixtures contain different concentrations of food dyes Red 3 and Blue 1. Evaluating this data by eye, we can easily predict that there are two components in this solution, as there are two peaks that change in intensity independently. When we perform SVD on this data we find that we have two distinct larger values (SV1 and SV2), and the remaining four values are close to zero (SV3 - SV6). This confirms our visual guess that there are only two main components in our solution.

This analysis becomes particularly valuable when we have mixtures that are more difficult to interpret, as seen in **Figure 4**. These ten mixtures (M1-M10) contain varying ratios of five different food dyes with overlapping absorbance spectra, which make it increasingly difficult to identify distinct species. However, given a large enough body of data (here, a set of ten different mixtures was plenty), SVD can be used to help confirm that we have five unique components in our data.

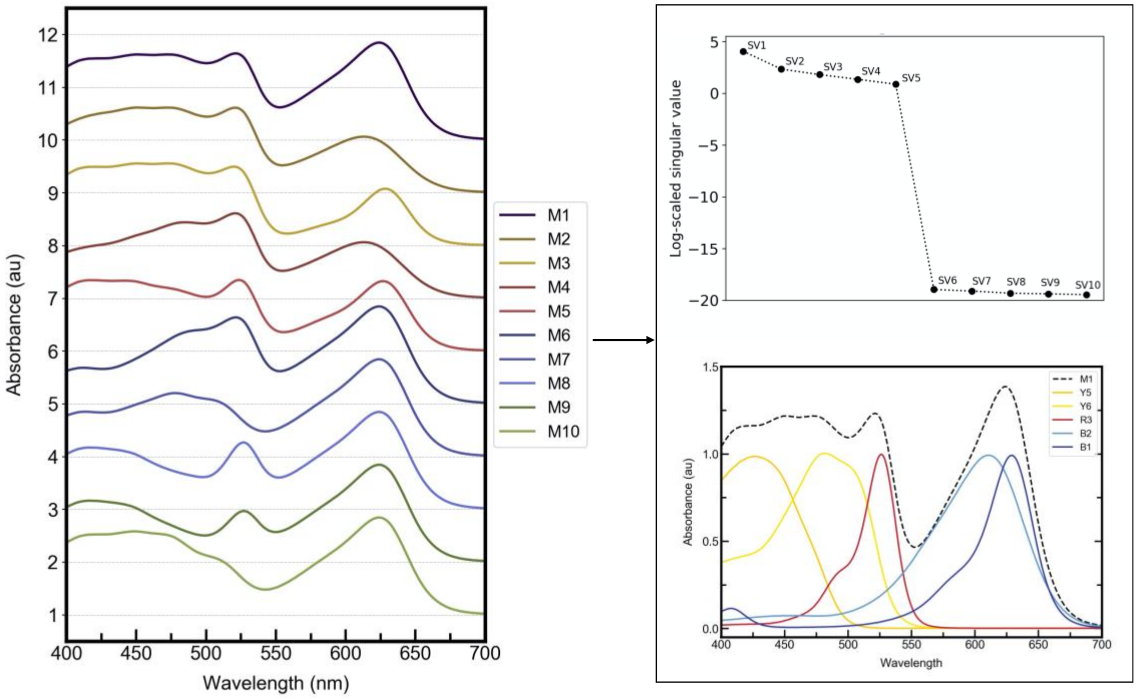


Figure 4. Absorbance data for ten mixtures of dyes (left) and corresponding SVD results (top right). An overlay of the dyes used to produce these mixtures (Yellow 5=Y5, Yellow 6=Y6, Red 3=R3, Blue 2=B2, Blue 1=B1) and mixture 1 (M1) is shown in the bottom right.

For your lab today, you'll be stepping into the role of a food scientist given five unknown mixtures of dyes, tasked with determining how many unique dyes are present in these mixtures overall.

You'll collect absorbance data on your Trimontana spectrometer like the examples shown with in Figures 2 and 3. However, your real data will be noisier than the artificial data in these figures! Using your data, you will then determine how many unique components you have. During your analysis, you will also compute the singular values for:

(1) "artificial" mixtures of the same dyes with noise removed and
(2) mixtures of the same dyes collected on a commercial UV-vis spectrometer

After performing SVD on all three data sets, you will compare your results to identify any differences between these three sources of data.

Learning objectives

Students will:

1. Understand how absorbance spectroscopy works and the basic components of a spectrometer (Pt 1 of lab)
2. Know how to collect absorbance data (Pts 1 and 2 of lab)
3. Know how to apply SVD and interpret the results (including the effects of data quality and noise on analysis) (Pts 2 and 3 of lab)
4. Use Excel and Python (through Jupyter) for data organization and analysis (Pts 2 and 3 of lab)

Procedure

Spectroscopy Lab I: Introduction to Spectroscopy and Data Science

**This version of the procedure includes options for both online and offline data collection.

A. Getting familiar with your spectrometer

(1) Build your Trimontana spectrometer following the demonstration from the Trimontana website:
https://docs.trimontana.tech/tutorialonline.html#spectroscope-assembly

The following instructional videos can also be very helpful for building your spectrometers and collecting data (Vimeo Password: Trimontana):
https://docs.trimontana.tech/tutorialonline.html#video-documentation

Ask your instructor whether or not you'll be using your spectrometers **offline** or **online** before proceeding to step 2.

(2) Plug your spectrometer in either (1) through your Raspberry/Orange Pi (**offline operation**) or (2) directly to your computer (**online/JupyterHub operation**), and paste the following addresses into three different windows:

   a. To access **instructions for data collection**:
      i. Offline operation through Raspberry/Orange Pi: 192.168.78.1
      ii. Online operation through JupyterHub: http://doc.trimontana.tech

   b. To access the **Jupyter notebooks for wavelength calibration and absorbance data collection**:
      i. Offline operation through Raspberry/Orange Pi: 192.168.78.1 (then scroll to sections for data collection and calibration for Jupyter notebooks)
      ii. Online operation through JupyterHub: http://hub.trimontana.tech/ (login information provided by your instructor)

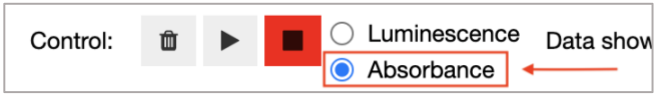
   c. To **download your data**:
      i. Offline operation through Raspberry Pi/Orange Pi: 192.168.78.1:8888/tree
      ii. Online operation through JupyterHub: http://hub.trimontana.tech/

Note: Jupyter notebooks are a handy way that data scientists in real-life commonly store, organize and manipulate their data.

(3) Calibrating your spectrometer: using the instructions outlined on the Trimontana website, run through the wavelength calibration.

(4) Collecting absorbance data: following your wavelength calibration, watch the instructional video for how to collect data on your Trimontana spectrometer, and follow the written instructions on the Trimontana website to collect absorbance data for two different samples. Make sure to save your absorbance data after each run.

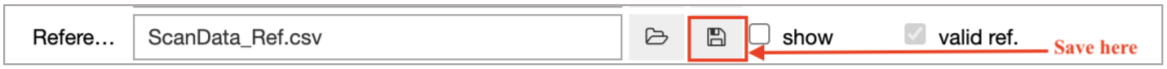
**Do not forget to select absorbance before starting data collection:



○ Luminescence
● Absorbance

a. Sample 1: 2 mL water (water blank, water absorbance spectrum)

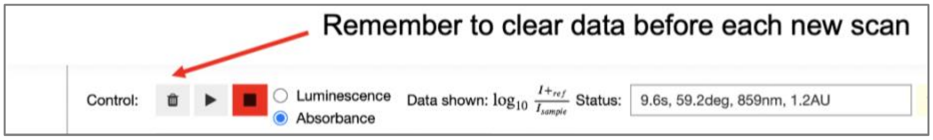
   i. Blank/background/reference: Use 2mL of water to collect your blank/background data. This scan will show up in red and be labeled “sample” in the Trimontana Jupyter notebook for data collection. Don’t forget to save your reference/background spectrum after you’ve collected this data:



Refere… ScanData_Ref.csv □ show ☑ valid ref.

Save here

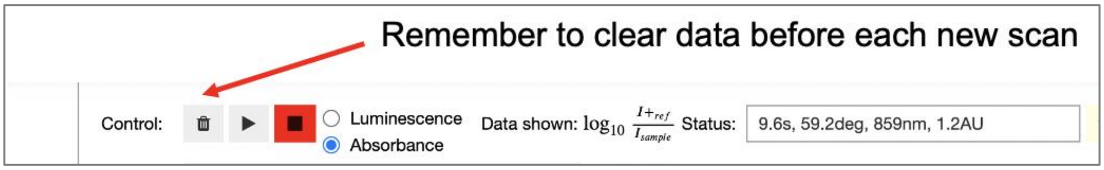
**Before moving onto your next absorbance scan, do not forget to press the trash can button to clear your current data memory. If you do not do this, you may write your new data onto a file with data from previous scans.



Remember to clear data before each new scan

Control: ○ Luminescence ● Absorbance Data shown: $\log_{10}\frac{I+_{ref}}{I_{sample}}$ Status: 9.6s, 59.2deg, 859nm, 1.2AU

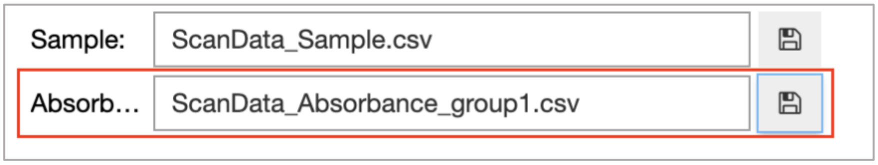
   ii. Absorbance data: Use 2 mL of water to collect an absorbance spectrum. This scan will show up in orange and be labeled “absorbance” in the Trimontana Jupyter notebook for data collection. This scan should not show any absorbance peaks if you collected your data correctly, apart from normal noise associated with the spectrometer (with a significant amount of noise from 400 - 500 nm).



Remember to clear data before each new scan

Control: ○ Luminescence ● Absorbance Data shown: $\log_{10}\frac{I+_{ref}}{I_{sample}}$ Status: 9.6s, 59.2deg, 859nm, 1.2AU

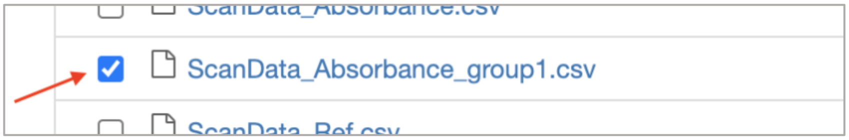
   iii. *Optional:* Save your absorbance data in the data collection window. If your instructor does not give you other instructions, pick <u>one</u> name for your group and keep this name for the entire lab (e.g. ScanData_Absorbance_group1.csv).



Sample: ScanData_Sample.csv

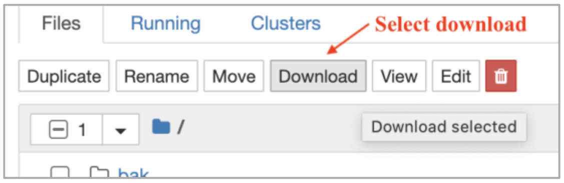
Absorb… ScanData_Absorbance_group1.csv

You will overwrite the data kept under this file name after you save a new scan, so be sure to download any important data and save it to your computer after each run. To do this, go to the address you opened in part 2(b). Select your group's absorbance data:



□ ScanData_Absorbance.csv
☑ ScanData_Absorbance_group1.csv
□ ScanData_Ref.csv

Then select download on the top of the window:



b. Sample 2: 2 mL of food dye (water blank/background, dye absorbance spectrum)
   i. Blank/background/reference: Use 2 mL of water to collect your blank/background data, just like you did in part 4(a) (page 10).
   ii. Absorbance data: Use 2 mL of a food dye solution to collect an absorbance spectrum. You should see a single peak in your absorbance spectrum. Examples of absorbance spectra for food dyes are shown below for both (1) data collected on a Trimontana spectrometer and (2) data collected on a commercial UV-Vis spectrometer.
   iii. *Optional:* Save and download your absorbance data.

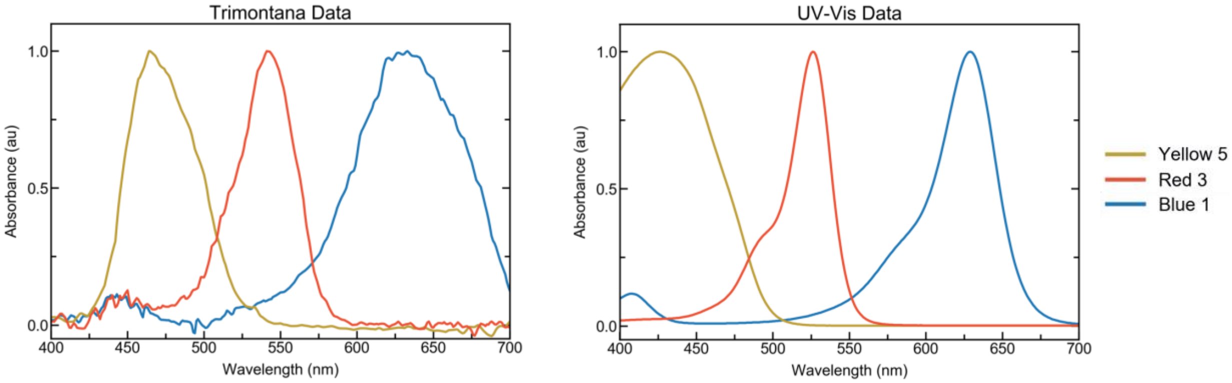


Figure 5. Examples of food dye absorbance spectra collected on two different types of spectrometers.

During this portion of the lab, discuss the questions from the lab introduction (at the top of page 2) with your lab partners.

B. Collecting data for unknown dye mixtures

*****Important note before starting:***

Do not dispose of any of your unknown solutions unless your instructor tells you otherwise.
Do not return your solutions to the stock solutions after you have taken your 2.5 mL fraction.
Do not use the same pipette for two different stock solutions.

(1) Transfer ~2.5 mL of each unknown sample from the 5 stock solutions to 5 different vials (one vial for each unique stock, labeled M1 – M5). Each of the 5 samples should look different but will contain mixtures of the same dyes. Your job today will be to determine how many individual dyes are present across all mixtures in a set.
(2) Make sure all your cuvettes are clean before starting and before measuring a new dye to ensure your unknown solutions are not contaminated with residual dye.
(3) Prepare a blank/background cuvette with ~2 mL water (just like you did for the known food dye samples) and a sample cuvette with ~2 mL of your first unknown solution. Use a plastic pipette to measure out each sample for this part of the lab, making sure to use a separate pipette for each unknown and your blank/background sample.
(4) Collect a background spectrum and save it as you did in the previous section of the lab. If you already have a 2 mL water background saved, you can skip this step. Remember to check your blank/background scan, which will appear in RED on the absorbance graph in the Trimontana Jupyter notebook. If there are any sharp spikes or peaks that deviate from your reference background spectra (included with your Trimontana spectrometers), you will need to re-collect this data. (Discussion Q: why do you think this might happen, and why is it important?)
(5) Collect two absorbance spectra for your first unknown solution, making sure to download and save each spectrum after it is collected (refer to instructions from the previous section for more details). Download and save your data as either a .xlsx (Excel) or .csv file after each scan.

**If time allows, plot both spectra in Excel to make sure your two spectra for the same dye look similar (example shown above)

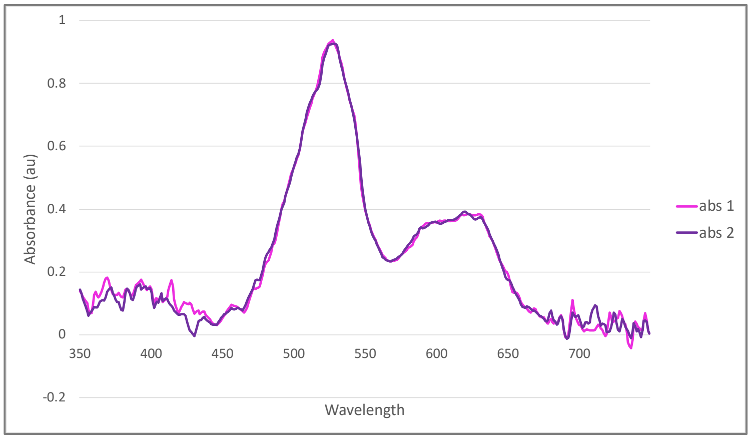


Figure 6. Example of two spectra (abs 1 and abs 2) for the same mixture. Notice that peaks are approximately the same for both scans.

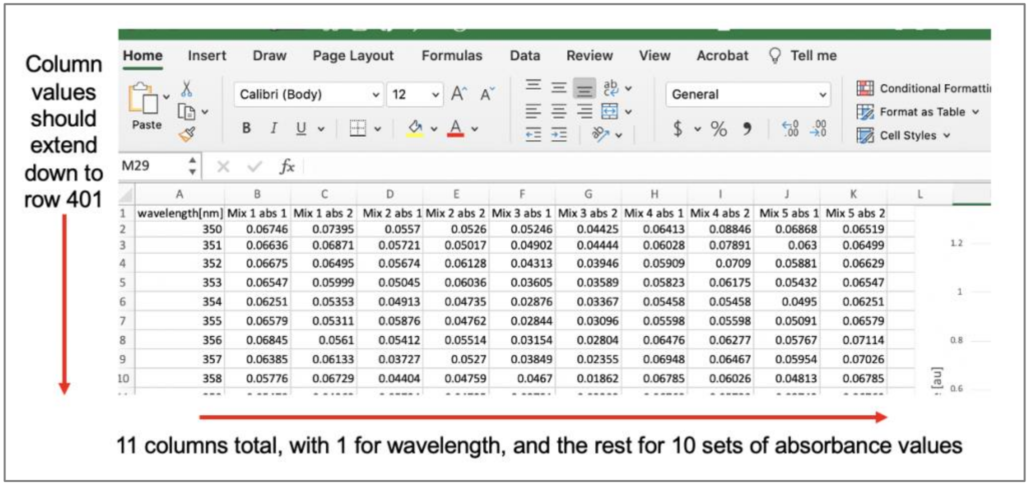
(6) Once you have collected two absorbance spectra of your first unknown, return the 2 mL of solution to your original sample vial (NOT the stock solutions provided to the class) using your designated pipette for that solution.

(7) Clean the cuvette you used for your first unknown solution thoroughly with water, making sure the cuvette is completely dry before moving on to your next mixture.

(8) Repeat steps 2 – 5 with the remaining four unknown dye mixtures, skipping re-collection of your blank spectrum with water. If you have not refreshed your browser with the Jupyter notebook for data collection, and you have not adjusted/bumped the diffraction grating for your spectrometer, you should not need to recollect and save your blank data after you have completed step 2 once.

C. Singular Value Decomposition Analysis

(1) Organize your data in Excel by copying over all of the absorbance values. An example of how to format your data is shown below:



Column values should extend down to row 401

	A	B	C	D	E	F	G	H	I	J	K
1	wavelength[nm]	Mix 1 abs 1	Mix 1 abs 2	Mix 2 abs 1	Mix 2 abs 2	Mix 3 abs 1	Mix 3 abs 2	Mix 4 abs 1	Mix 4 abs 2	Mix 5 abs 1	Mix 5 abs 2
2	350	0.06746	0.07395	0.0557	0.0526	0.05246	0.04425	0.06413	0.08846	0.06868	0.06519
3	351	0.06636	0.06871	0.05721	0.05017	0.04902	0.04444	0.06028	0.07891	0.063	0.06499
4	352	0.06675	0.06495	0.05674	0.06128	0.04313	0.03946	0.05909	0.0709	0.05881	0.06629
5	353	0.06547	0.05999	0.05045	0.06036	0.03605	0.03589	0.05823	0.06175	0.05432	0.06547
6	354	0.06251	0.05353	0.04913	0.04735	0.02876	0.03367	0.05458	0.05458	0.0495	0.06251
7	355	0.06579	0.05311	0.05876	0.04762	0.02844	0.03096	0.05598	0.05598	0.05091	0.06579
8	356	0.06845	0.0561	0.05412	0.05514	0.03154	0.02804	0.06476	0.06277	0.05767	0.07114
9	357	0.06385	0.06133	0.03727	0.0527	0.03849	0.02355	0.06948	0.06467	0.05954	0.07026
10	358	0.05776	0.06729	0.04404	0.04759	0.0467	0.01862	0.06785	0.06026	0.04813	0.06785

11 columns total, with 1 for wavelength, and the rest for 10 sets of absorbance values

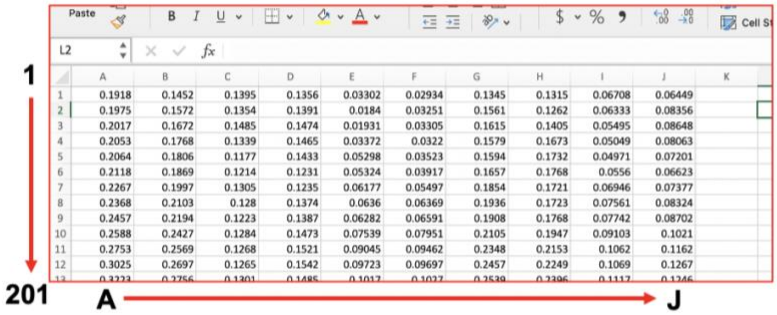
(2) Create an absorbance vs. wavelength graph for all five samples (example of plot for one sample shown above).
**See Excel tips slides if you need help plotting, copying, or organizing your data.

(3) Before proceeding to the singular value decomposition (SVD) analysis steps, hypothesize how many different dyes are in this sample based on absorbance features in the 500 - 700 nm range (as this spectrometer is less accurate in the 400 - 500 nm range). What features of your data led you to this conclusion?
Hint: All samples will have between 1 and 6 dyes

Estimated number of dyes = __________

(4) Create a separate .csv file with your absorbance data from 500 - 700 nm. This will serve as the data table that you'll input for SVD. Do this by copying and pasting the absorbance values from 500 - 700 nm for each unknown dye mixture spectrum. Do not include titles at the top of your columns. This final .csv file should look something like the following photo:

Final spreadsheet should have values in columns A – J and extend down to row 201



	A	B	C	D	E	F	G	H	I	J	K
1	0.1918	0.1452	0.1395	0.1356	0.03302	0.02934	0.1345	0.1315	0.06708	0.06449	
2	0.1975	0.1572	0.1354	0.1391	0.0184	0.03251	0.1561	0.1262	0.06333	0.08356	
3	0.2017	0.1672	0.1485	0.1474	0.01931	0.03305	0.1615	0.1405	0.05495	0.08648	
4	0.2053	0.1768	0.1339	0.1465	0.03372	0.0322	0.1579	0.1673	0.05049	0.08063	
5	0.2064	0.1806	0.1177	0.1433	0.05298	0.03523	0.1594	0.1732	0.04971	0.07201	
6	0.2118	0.1869	0.1214	0.1231	0.05324	0.03917	0.1657	0.1768	0.0556	0.06623	
7	0.2267	0.1997	0.1305	0.1235	0.06177	0.05497	0.1854	0.1721	0.06946	0.07377	
8	0.2368	0.2103	0.128	0.1374	0.0636	0.06369	0.1936	0.1723	0.07561	0.08324	
9	0.2457	0.2194	0.1223	0.1387	0.06282	0.06591	0.1908	0.1768	0.07742	0.08702	
10	0.2588	0.2427	0.1284	0.1473	0.07539	0.07951	0.2105	0.1947	0.09103	0.1021	
11	0.2753	0.2569	0.1268	0.1521	0.09045	0.09462	0.2348	0.2153	0.1062	0.1162	
12	0.3025	0.2697	0.1265	0.1542	0.09723	0.09697	0.2457	0.2249	0.1069	0.1267	
13	0.3223	0.2756	0.1301	0.1485	0.1017	0.1027	0.2539	0.2396	0.1117	0.1246	

(5) Log into the Trimontana JupyterHub for your course (http://hub.trimontana.tech), and upload your formatted data to your group folder. Name your data using the following format:

[DataType]_[CourseNumber]_[Group].csv

(6) Open the “Dye Lab SVD Analysis” notebook in your group folder and run each cell to:

a. Import your data from the .csv file (replace the file path with the file path for your .csv file). You’ll be doing this with the Python library Pandas, which we’ll import as “pd”.
b. Compute the singular values from your formatted data. You’ll be doing this with the Python library SciPy.
c. Print and plot your computed singular values.
d. Save your singular values to a .csv file.

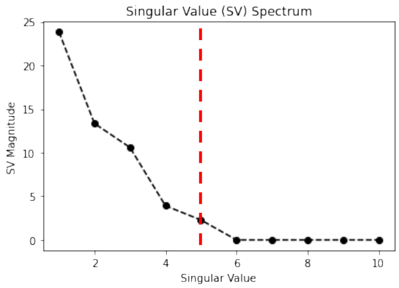
(7) If our data was perfect, we would be able to easily distinguish between larger singular values corresponding more clearly to how many components we need to describe our solution, and smaller insignificant singular values on the order of zero. However, given our measurements are not perfect (e.g. we have fluctuations in the data from noise/the accuracy of the spectrometer), we will need to approximate which singular values should be considered and which should be excluded. There are several ways to do this, but your class will use the following heuristics for today’s lab:

a. Plot your singular values on a (1) standard and (2) natural log scale (this allows us to better distinguish smaller singular values, compared to the first singular value, which will be large). You did this in step 6 (c) above, and will just need to copy or save these graphs.

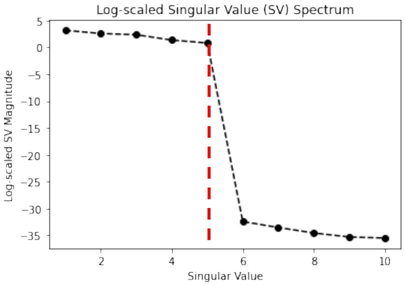
In your first plot (the raw singular value spectrum), try to find the last point in your plot where there is a drop in value. The remining values should be close to one another in value, and approximately 0.

In your second plot (the log-scaled singular values), find the last point above 0. All remaining values may vary in magnitude, but should be > 0.

All singular values at or above these points will be considered significant, and should give us a good approximation for how many dyes we have in our unknown mixtures. The number of significant values should be the same, however if you determine two different numbers from your plots, select the number of values determined from the log-scaled plot (as this plot allows us to more closely look at differences in smaller values). An example of this is shown below for a data set with 10 spectra analyzed via SVD, containing 5 unique components across all unknown mixtures (here, "w" = the singular value).



For your SV spectrum, find the point in your data where remaining singular values are close to 0.



For your log scaled SV spectrum, find the point where the remaining values are <0.

Answer for this data set:
5 unique dyes

b. To better understand our results in part (a), you'll next determine approximately what percentage of your data you're describing with these singular values. Find the sum of all the singular values, then find the sum of the significant/larger singular values you identified in part (a) above. Make sure to use the raw singular values and not the log scaled singular values.

**If you use Excel to calculate this, you can download these values directly from the .csv file you saved in step 6 (d).

Divide the sum of your larger singular values by the sum of all your singular values, and you should get a fraction above 0.9. This fraction indicates roughly what percentage of our data we can describe with the largest singular values (e.g. 0.95 = 95% of the data can be described using these singular values). Given the amount of noise associated with the Trimontana spectrometers, accepting the fewest number of singular values that will describe over 90% of your data should give you the approximate number of dyes in solution.

(8) Repeat steps 2 – 5 with the following data sets (provided by your instructor, found in the course materials folder):

a. Absorbance data for the same 5 mixtures collected on a commercial UV-vis Spectrometer (File name should start with "UVvis")

b. Artificial/simulated mixtures of the same dye(s) (File name should start with "Artificial")

Note for repeating step 7 (b) with UV-Vis and Artificial data:

**Given the low amount of noise associated with the UV-Vis data and the Artificial Data when compared to the Trimontana Spectrometer, accepting the fewest number of singular values that will describe over 95% of your data should give you the approximate number of dyes in solution.