Intuitive chemometrics: Linear analysis of absorption spectra

Determining concentration of light-absorbing substances in a solution by the amount of light the solution absorbs when a beam of light passes through it.

1. Introduction.

Chemometrics is the science of extracting information from chemical systems by data-driven means[WP].

One important branch of chemometric analysis involves the use of UV-Visible absorption spectroscopy to analyze the composition of a solution of multiple substances, either without or with first adding reagents that react with one or more substance to form colored derivatives whose concentrations are directly related to the amount of the original substance(s). We use the general term "substances" here, but in general we will only be considering substances in the mixture that absorb light in the wavelength range of the spectrum being analyzed. In some cases we will distinguish between "analytes" (substances being analyzed, substances whose concentration will be estimated by the procedure used) and interferents (substance that will not be determined but whose light absorption must be taken into account in determination of the analyte(s).

First a disclaimer- we are not presenting any new methods. The methods presented here have been worked out and are well understood. Instead we are presenting a way of looking at these methods which provides a graphical/geometrical interpretation and makes them much more intuitive than the abstract formulas that have been derived. This approach depends heavily on two concepts: the realization that digital spectra are vectors, and the concept of a "probe" spectrum, which will be described below. We also emphasize the importance of a "goodness of fit" criterion in analyzing samples, so that if a sample is not represented by your model, you have some indication that the results are not reliable.

We will be applying this approach to several different classes/methods of analysis-

1. Conentrations from Spectra - The first methods involve the case where a known number of substances may be present in the sample, no other substance in the sample absorbs at the wavelengths used for the analysis, accurate spectra of the pure substances at unit concentration are available, and you have measured the spectrum of the unknown mixture.

We will break this down into three cases based on the number of substances and the number of wavelengths used for analyzing them: first the trivial case of a single substance being analyzed at a single wavelength, then multiple substances being analyzed at a number of wavelengths equal to the number of substances, then multiple substances being measured at a number of wavelengths greater than the number of substances, usually a whole spectrum

measured at intervals along the wavelength range where the substances absorb light.

2. Spectra from concentrations- Given a number of "training" spectra of mixtures of the substances, in which the substances are present in different, known, proportions, determine the spectra of the pure components. This is the reverse of the first procedure, and can be broken down into three categories based on number of substances and the number of training spectra. The pure spectra would then be used to analyze spectra of unknown samples. This can be useful in cases where the substances cannot be prepared in pure form without altering their absorbance properties, for example cytochromes whose spectra depend on their environment in a protein complex and are irreversibly altered when the complex is split. Or in some cases it may be more economical to analyze a number of samples by more exhaustive means to determine the amount of each substance, and then use these samples to make training spectra for analyzing further samples by taking a simple spectrum.

3. Concentrations from concentrations- Is combining procedures 1 and 2 to determine concentrations in the unknown based on spectra of the training solutions and knowledge of the concentration of each substance in each training solution, without explicitly determining the standard spectra of the pure substances. The motivation would be as in procedure 2, but additionally it has been claimed that this gives better results than method 1 even when the spectra of the pure substances are available, in cases where Beer's law (see below) is not strictly applicable, because the training spectra incorporate the deviations in some way to provide a better estimate.

(This is the one where textbook says the number of wavelengths must be less than the number of components?,but we disagree)

4. Concentrations from concentrations- Method 4 involves the case where you are only interested in determining one substance, but there are other substances ("interferents") that absorb light at the same wavelengths and would interfere in determination of the desired substance. You have a set of training spectra which are representative of all the samples you are likely to analyze, but the identities, spectra, and concentrations of the interferents are not known. You only have the concentration of the substance of interest. For example you are analyzing samples of ore from different parts of a mine to see which are richest in the metal your company is producing, but the ore contains other substances that affect the spectrum in the assay.

Ideally, for greatest accuracy, you would want to find out what the different interferents are, determine their spectra, and use method 1 to analyze the samples. However if the training samples truly are representative, this method allows you to get a good estimate of the substance of interest without knowing what the interferents are or their spectra.

5. Reducing the dimension, segregating signal - Principal components analysis/Singular Value Decomposition. This works together with the previous methods. At the least, it provides a way to represent all the information in a set of spectra with a smaller number of data points. More importantly, it provides

segregation of random, one-time noise from spectral features recurring throughout the dataset. Other advantages can be explained later with linear algebra.

Note on the English language: absorb is a verb meaning to soak up, into internal structure, as a sponge absorbs water. (*adsorb* is to accumulate a substance on the surface- charcoal adsorbs aniline, aniline "adsorbs to" charcoal). In the corresponding *processes*, b becomes p: absorption, adsorption. The capacity of a sponge to absorb liquid is its "absorbency". Physicists/chemists likened the extinction of light passing through a substance (or that part of the extinction not due to reflection or light scattering) as "absorption" of the light by the substance (the energy of the light is "absorbed" into electronic transitions and perhaps eventually to heat energy). In order to quantitate light absorption there needed to be a precise term. The term most used today is "absorbance" (note the a instead of e as in absorbency), which will be defined in the next section.

2. The Beer-Lambert law-

TLDR summary: Attenuation of light on passing through a solution can be measured as absorbance of the solution. The absorbance of the solution is then equal to the sum of the extinction coefficients of each substance in the solution multiplied by the concentration of that substance.

The Beer-Lambert law, or Beer's law for short, encapsulates three observations about the fraction of monochromatic light passing through a solution containing substance(s) that absorb the light: the dependence on path length through the solution, on the concentration of the absorbing substance(s), and on the presence of multiple absorbing substances. As a beam passes through a homogeneous solution, its power decays exponentially. Say the first 1 mm absorbs 10% of the light passing through, the next mm absorbs 10% of the remaining, and so on. This is the formula for exponential decay. Likewise as concentration of the absorbing molecules increase, the additional molecules can only absorb what was not already being absorbed. So if a solution containing 1 μM of a substance absorbs 10% of the light passing through, adding another 1 μM will absorb 10% of the remaining , and so on. So the percent of light transmitted through a substance (percent transmittance or %T) is exponentially related to the path length (L) and the concentration (c), and in fact to the product of path length and concentration: The higher the concentration, the more each mm absorbs, so the faster the intensity falls off with distance into the solution.

$\%T \equiv I/I_o$	Intensity transmitted relative to incident intensity
$\%T = 10^{-k \times L \times c}$	k is a constants determining the rate of fall-off
$-\log(\%T) = k \times L \times c$	with distance or concentration

So - log(%T) is proportional to both L and c. As path length or concentration increase, less light gets through, %T is smaller, -log(%T) is larger. -log(%T) goes to infinity when no light passes through, and zero when 100% passes through (log(1)=0).

- log(%T), the negative log to the base 10 of %T, is called the absorbance (A) of the solution, and the constant of proportionality is called the extinction coefficient, usually represented as "ε". This gives:

$$A = c \times L \times \varepsilon \qquad [1]$$

This ε is a property of the absorbing substance, being larger if the substance absorbs a lot, zero if it doesn't absorb at all. It is also dependent on wavelength in a way that is characteristic for the absorbing substance.

We will assume the path length L is 1 (cm), the most common width for spectrophotometer cuvets, in which case $A = c\varepsilon$. If c is also one, then the absorbance A is equal to ε, so we can also define the extinction coefficient ε as the absorbance of the pure substance at unit concentration (in whatever units you choose to use, in a cuvet with path length 1, assuming Beer's law holds at unit concentration.)

Finally, if there are two substances absorbing, at a concentration such that either one by itself would absorb 20% of the light (resulting in %T of 80%, each can only absorb 20% of the 80% not absorbed by the other substance, so %T of the mixture is 80% of 80% or 64%, rather than 60%. Thus the effect of multiple substances on transmittance (%T) is multiplicative, and the affect on absorbance (-Log(%T)) is additive. So the absorbance of a mixture is the sum of the absorbances due to each substance:

$$A = c_1\varepsilon_1 + c_2\varepsilon_2 + c_3\varepsilon_3 + c_4\varepsilon_4 = \sum C_j\varepsilon_j \qquad [2]$$

where c_j is the concentration of the j^{th} substance and ε_j is its extinction coefficient.

3. Digital Spectra-

TLDR summary: Digital spectra consist of a series of numbers representing the absorbance of the sample at different wavelengths.

The value of ε, and hence the amount of light absorbed by a substance, depends on wavelength in a way that is characteristic of the substance. This is due to the need to match the energy of the photon to the energy difference of the electronic transition being excited in the molecule. A plot of absorption vs wavelength is called the spectrum of the substance or mixture. For reasons beyond the scope of this text, UV-visible absorption spectra tend to have broad, overlapping peaks rather than sharp lines at specific wavelengths.

In principle, spectra are continuous plots of absorbance as a function of wavelength, acquired in olden times by recording the spectrophotometer signal on a strip-chart or X-Y recorder as the monochromater is continuously varied from the starting wavelength to the final wavelength. In order to do non-analog calculations on the spectra, they need to be digitized, and this is done in the same way that audio waveforms are digitized, by sampling the signal at discrete wavelengths.

Thus a digital spectrum is a series of values (numbers) representing the absorbance of the sample at different wavelengths. In order to do math on the spectra, they have to be recorded at the same wavelengths in the same order so that for example subtracting the background from the sample spectrum, it only makes sense to subtract absorbances at the same wavelengths. We often refer

to absorbance at a particular wavelength λ as A_λ, for example A_{260}, A_{280}, A_{420} etc. In the case of a fully digitized spectrum, it makes more sense to refer to the values by index: A_1, A_2, A_3, . . . where A_1 is the absorbance at λ_1, A_2 is the absorbance at λ_2, and in general A_i is the absorbance at λ_i.

4. Spectra as vectors-

TLDR summary: Digital spectra can be considered as vectors. Many of the standard operations carried out on spectra, such as addition, subtraction, scalar multiplication, and formation of linear combinations of different spectra, can be seen as standard vector operations

A vector in linear algebra is an "ordered n-tuple" of numbers, where n is the number of elements. These N elements of the vector can also be called "coordinates", or "components" of the vector, in reference to a graphical representation in which the elements locate the tip of the vector in n-dimensional space, or represent the vector when it is broken down into its "components" along the x, y, z, etc principal axes.

In the graphical representation, each number represents displacement of the tip of the vector in n-dimensional space along each of a set of orthogonal axes: (x, y, z axes in 3-space) or multipliers for unit-length basis vectors along the axes: e_1, e_2, e_3, . . . that add up to the vector.

Digital spectra fit the definition of ordered n-tuple, where n is the number of points at which the spectrum is sampled. They are vectors in an n-dimensional vector space. They are implicitly treated as vectors in the math behind the methods listed in the introduction. In this text we are attempting to exploit the graphical representation of vectors to make that math more intuitive, and give the user a better understanding of what goes on in the equations and "black box" procedures.

Several operations commonly applied to vectors are also relevant for spectra. Vectors can be added or subtracted (subtraction of one vector from the other is simply addition of the negative of the vector), and they can be multiplied by a scalar. Two vectors can be multiplied, in at least three different ways- the one that we will be concerned with is the "inner" or "dot" product, represented by (A•B). The product A•B is a scalar number equal to the sum of the products of corresponding elements of the two vectors: $A \bullet B = a_1b_1 + a_2b_2 + a_3b_3 + \text{etc.}$ We will make use of the dot product a great deal in this text. We refer to the operation of forming the dot product of A and B as "dotting" A "into" B.

Vectors can also be multiplied by a matrix, resulting in another vector (of the same or different dimension), and this operation will also be important in the following sections.

As mentioned above when subtracting spectra to remove background or to see the spectral change, we subtract point-wise, subtracting the absorbance values of the two spectra at each wavelength. This is the definition for addition or subtraction of vectors: $A - B = a_1-b_1, a_2-b_2, a_3-b_3$).

We can simulate the spectrum of a substance at arbitrary concentration c by applying Beer's law for a single substance: at each of the wavelengths , we

multiply the extinction coefficient at that wavelength by c. This is making a scalar multiple of the vector of extinction coefficients, which vector we call E:

$E = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \ . \ . \ ., \varepsilon_i, \ . \ . \ .)$

for a solution containing the substance at concentration c,

$A = (A_1, A_2, A_3 \ . \ .) = (c\varepsilon_1, c\varepsilon_2, c\varepsilon_3, \ . \ . \ ., c\varepsilon_i, \ . \ . \ .) = c\mathbf{E}$,

the scalar multiple of vector E by c.

Since the extinction coefficient ε_i is the absorbance of a solution of the pure substance at unit concentration at wavelength λ_i, the vector E consisting of all the ε is just the spectrum of the pure substance at unit concentration. We call this the "standard spectrum" of the substance

In Beer's law for a mixture of substances, (equation 3, $A = \sum c_j \boldsymbol{\varepsilon}_j$), the right hand side can be seen as a dot product of a vector of concentrations of each substance, with the vector of extinction coefficients of the corresponding substances, all at the same wavelength. This does not involve spectra as vectors, but is important in the operations that follow.

Finally if we apply Beers law to the spectrum of a mixture of compounds, by applying equation 2 at every wavelength i, we get the spectrum A whose elements A_i are given by:

$$A_i = \sum c_j \boldsymbol{\varepsilon}_{ij} \qquad \textbf{[3]}$$

Where A_i is the measured absorbance at the i^{th} wavelength, c_j is the concentration of the j^{th} substance, and $\boldsymbol{\varepsilon}_{ij}$ is the extinction coefficient of the j^{th} substance, at the i^{th} wavelength.

Equation 3 can also be expressed by saying the spectrum **A** is a linear combination of the standard spectra, weighted by the concentration of each:

$$\mathbf{A} = \sum c_j \mathbf{E}_j \qquad [3a]$$

or as a matrix multiplication:

$$\mathbf{A} = \mathbf{[E]C} \qquad [3b]$$

where A, the observed absorbance spectrum, is a column vector, [E] is a matrix whose columns are the standard spectra $\mathbf{E}_j$, and **C** is a column vector of the concentrations of the J different substances. This is Beers law in matrix form, and will be discussed in more detail in a later section.

To minimize confusion, we try to use consistent conventions for subscripts. To the extent it is practical, we will use the subscript i to refer to different sampling points in the spectrum, that is the different wavelengths used. j will refer to the different substances, and in the case that more than one spectrum A is being analyzed, k will refer to the different spectra.

The set of real numbers can be considered a one-dimensional vector space, with each number being a "1-tuple". The graphical representation of this one-dimensional apace is the number line, with arrows going left or right from the origin at zero to the value of the number. Applying the definition of the dot product gives $a \bullet b = a_1 \times b_1 = ab$, the conventional product of the two numbers.

In line with this, we can consider an absorbance measurement at a single wavelength to be a 1-dimensional "spectrum". Obviously if there are two substances at unknown concentration absorbing at that wavelength, we cannot tell from the absorbance how much is due to each. We can only determine the

concentration if we know that only one substance absorbing at that wavelength is present. In general, the number of wavelengths measured must be greater than or equal to the number of substances being analyzed- otherwise multiple solutions exist- the problem is "under-determined".

5. Some Linear Algebra:

TLDR summary: Length of a vector, relation between dot product of two vectors and the angle between, expressing one vector as the some of a component *along* and a component *perpendicular* to another vector.

When a vector is expressed as an ordered n-tuple, the Pythagorean theorem shows that the length, or "norm", of a vector A, written |A|. is the square root of the sum of squares of the elements, or equivalently the square root of the dot product of the vector with itself:

$$|A| = \text{sqrt}(a_1^2 + a_2^2 + a_3^2 + \ldots) = \text{sqrt}(A \bullet A) . \qquad 4.$$

This is a special case of another rule, that the dot product of vector A with vector B is equal to the product of the lengths of A and B, multiplied by the cosine of the angle between:

$$A \bullet B = |A| \times |B| \times \text{Cos}\ \theta, \qquad 5.$$

because if A is dotted into itself, the angle θ is zero, and Cos(0) is 1, so:
$A \bullet A = |A| \times |A| \times 1 = |A|^2$, and $\text{sqrt}(A \bullet A) = |A|$.

At the opposite extreme, if the angle between A and B is 90° (they are perpendicular, or orthogonal), Cos θ is zero, and so the dot product of A and B is zero. While we normally refer to 90° angles when in 2- or 3-dimensional space, the concept of orthogonality, as defined by a dot product of zero, generalizes to any number of dimensions. In particular, all of the principal axes are orthogonal to each other, as for example you can see that the dot product of (0,0,0,1,0,0, . . .) and (0,0,1,0,0,0, . . .) is zero. Any two vectors span a plane in multidimensional space, and the angle between the two vectors in that plane is still determined by equation 4.

Note that if we take the real numbers as a 1-dimensional vector space, the norm $|a| = \text{sqrt}(a \bullet a)$ is just $\text{sqrt}(a^2)$, or the magnitude or absolute value of a. The norm of a vector is sometimes denoted with double bars ||A|| to distinguish it from the absolute value of a scalar, denoted |a|, but in a sense they are the same thing. For simplicity we use single bars, and refer to the quantity as the norm, length, or magnitude of the vector as seems appropriate.

A vector **A** can be expressed as the sum of two vectors, one in the same direction as another vector **B** (a scalar multiple of **B**) and a second one perpendicular to **B**. These are referred to as the component of A *along* B, and the component of A *perpendicular*,(or *orthogonal*) to B. The component of A along B can also be called the orthogonal projection of A onto B. Note that this is the projection along a line orthogonal to B, not to A.

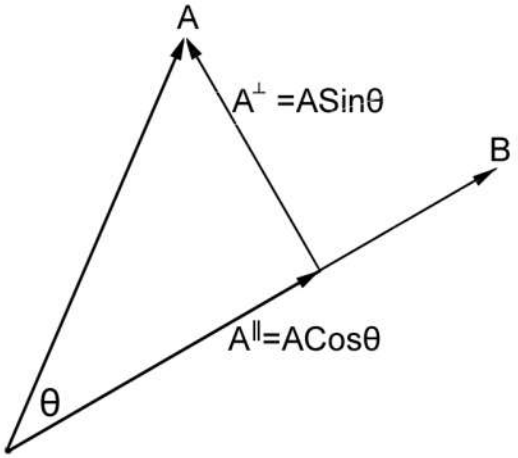

Figure 1. Decomposition of vector A into a component along another vector B ($A^{\parallel}$) and a component perpendicular to B ($A^{\perp}$).

Being perpendicular, these two component vectors form a right triangle with the original vector **A** as the hypotenuse (Figure 1). Trigonometry tells us that, if the angle between A and B is θ, then the length of the component of A along **B, $|\mathbf{A}^{\parallel}|$,** is equal to the length of **A** multiplied by Cos θ; while the length of the perpendicular component $|A^{\perp}|$ is |A|×Sin θ.

$$|A^{\parallel}| = |A|\text{Cos}\theta = A.B/|B| \quad (6.)$$

$$|A^{\perp}| = |A|\text{Sin}\theta \quad (7.)$$

Thus the dot product of two vectors A and B is a measure of the length of the component of A along B, ranging from zero when the vectors are orthogonal to |A|×|B| when they are colinear. In particular, if B is a unit length vector, the dot product ranges from 0 to |A| and is numerically equal to the length of the component of A along B. This is because A•B = |A||B|Cosθ, so if |B| =1 it is equal to |A|Cosθ which by inspection of the figure is the component of A along B. When B is not unit-length, the length of the component of A along B is A•B/|B| = |A| Cosθ.

It is important to distinguish between "the *length* of the component of A along B" (a real number) and "the component of A along B", a vector. To make the latter, we take a unit-length vector along **B**, which is 1/|B|×**B**, and multiply by the length of the component of A along B, A•B/|B|. This gives $\mathbf{A}^{\parallel}$ = (A•B/$|B|^2$)×**B**, or (A•B/B•B)×**B**. Given $\mathbf{A}^{\parallel}$, we can subtract that vectorially from **A** to get $\mathbf{A}^{\perp}$, the orthogonal component.

$$A^{\parallel} = (A\bullet B/B\bullet B) \times \mathbf{B} \quad \mathbf{(8.)}$$

$$A^{\perp} = \mathbf{A} - (A\bullet B/B.B) \times \mathbf{B} \quad (9.)$$

Given that the spectra generally consist of measurements at many more than 2 wavelength points, we need to generalize to higher dimension. Consider a plane in three dimensions, spanned by the vectors B and C, A vector is perpendicular to that plane if it is perpendicular to every vector in the plane and, in particular, perpendicular to both B and C. An arbitrary vector A can be divided

into a component perpendicular to the plane and a component that lies within the plane. Again these two components make a right triangle with A as the hypotenuse, and the above relations hold.

To further generalize, in N dimensions, a "subspace" can be defined as the span of a number of vectors less than N, just as the plane in the example above was defined by the span of vectors B and C. An arbitrary vector can be divided into a component orthogonal to that subspace (i.e. orthogonal to each of its defining vectors) and a component within the subspace. This generalization includes the above examples of vector A having a component within, and a component orthogonal to, the plane (a 2-dimensional subspace spanned by the vectors B and C), or the original example of vector A having a component along, and a component orthogonal to, the line (a one-dimensional subspace spanned by the vector B).

And regardless of the number of dimensions, the vector and its components within, and orthogonal to, the subspace define a plane and a right triangle in that plane. So we can draw the same 2-dimensional picture as in Figure 1, and the same relations derived above (equations 4 - 9) still hold. However now instead of showing the x-y plane it shows a plane through the origin at a particular angle in multi-dimensional space.

This is relevant for the next section, where we say the probe spectrum for substance j is orthogonal to the spectra of all the other substances present in the sample. The direction (but not the magnitude) of the probe spectrum is defined by the component of the standard spectrum, E, orthogonal to the subspace spanned by the spectra of all the other components. In addition, when we solve the least squares problem to get the "best fit" linear combination of the standard spectra to an experimental spectrum A, That linear combination is the component of A within the span of the standard spectra, while the difference between that and the observed spectrum A (the "residuals" is the component of A orthogonal to the span of the standard spectra.

The definition for the length, or norm, of a vector also provides a definition for distance in Euclidean space: the distance between two points is the length of a vector connecting them, so square root of $((x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2)$

6. "Probe" spectra:

For every substance being determined by one of the procedures covered here, we define its probe spectrum as a spectrum that, when dotted into the spectrum of a mixture, gives an estimate of the concentration of that substance in the mixture. The probe spectrum for a substance depends not only on the spectrum of that substance but also on the spectra of other absorbing substances that may be present. In problems of the first type, where we know the standard spectra E for each potential substance, we can calculate a probe spectrum P for each substance.

The probe spectra (P's) and the standard spectra (E's) for the different substances share a special relationship, co-orthonormality. This means that the probe spectrum for the jth substance ($\mathbf{P}_j$) must satisfy, for every j and every k,

$P_j \bullet E_k = 0, \; j \neq k$ (i.e. the vectors P_j and E_k, dotted along wavelength i) [4]
$= 1, \; j = k$

The first requirement is equivalent to saying that the probe spectrum is orthogonal to all of the standard spectra except the one which it measures. The probe spectrum's dot product with that standard spectrum gives the concentration represented by that spectrum, namely 1.

This relationship allows the probe spectrum, when dotted into the spectrum of a mixture, to pick out the concentration of its corresponding substance and ignore the others. If the spectrum A is a linear combination of the standard spectra of the substances weighted by concentration c (equation 2), then:

$$\mathbf{A} = c_1 \times \mathbf{E}_1 + c_2 \times \mathbf{E}_2 + c_3 \times \mathbf{E}_3 + c_4 \times \mathbf{E}_4 + \; . \; . \; .$$

If we dot the probe spectrum for the 2nd substance into A:

$$\mathbf{P_2 \bullet A} = c_1 \times \mathbf{P_2 \bullet E_1} + c_2 \times \mathbf{P_2 \bullet E_2} + c_3 \times \mathbf{P_2 \bullet E_3} + c_4 \times \mathbf{P_2 \bullet E_4} + \; . \; . \; .$$

$$\mathbf{P_2 \bullet A} = c_1 \times 0 + c_2 \times 1 + c_3 \times 0 + c_4 \times 0 + \; . \; . \; . \quad = \; c_2$$

(This is like the Fourier transform, where all the component frequencies are orthogonal, so you can integrate a given frequency against the composite waveform to get the amplitude of that frequency in the waveform)

Because the dot product of the probe spectrum with the corresponding standard spectrum is 1, the probe spectrum can be considered an "inverse" of the standard spectrum. The inverse of a number is its reciprocal, and it is unique. In vector space of dimension greater than 1, there can be any number of vectors that, when dotted into a particular vector, give 1, so the inverse is not unique. The probe spectrum is a particular inverse.

7. Graphical interpretation-

TLDR: Although we cannot visualize spaces with greater than three dimension, the concepts and procedures described can be illustrated in two ways- first by choosing examples in two or three dimensions, and secondly by illustrating a 2D plane in multidimensional space, or the projection of multidimensional space onto that plane.

We want to use the graphical representation of vectors to make the analytical procedures more intuitive The problem arises that we can't really visualize spaces with more than three dimensions, and in general the spectra will have measurements at much more than three, often thousands, of wavelengths. There are two ways to visualize what is going on in the procedures. One is to use an example where the number of wavelengths is limited to two or three. For example, the Warburg and Christian method[1] of analyzing protein and nucleic acid (NA) by absorbance at 260 nm and 280 nm, or Arnon's measurement[2] of chlorophyll A and chlorophyll B at 645 and 663 nm.

We can take a sheet of graph paper and label the axes as A_{260} and A_{280}. Then plot the 2-wavelength "absorption spectra" of the two substances, as in Figure 2C. Protein absorbs more at 280 than at 260, while the opposite is true for nucleic acid, so the two vectors point in different directions in the first quadrant. The points labeled Protein and Nucleic Acid are the "standard" spectra at concentration = 1, i.e. the absorbance that would be measured at 260 and 280 nm for solutions of concentration 1 (in whatever units). For a pure solution of either substance at different concentrations, both A_{260} and A_{280} would be scaled in proportion, so they would lie along the same lines as the corresponding standard spectra. In Figure 2C We have plotted the points for a solution of protein at concentration 0.4 and a solution of NA at concentration 0.3 units. In terms of linear algebra, all solutions of pure protein lie in a one-dimensional subspace (red line) "spanned" by the standard spectrum of Protein, and likewise for NA solutions (green line). And the same is true in any number of dimensions.

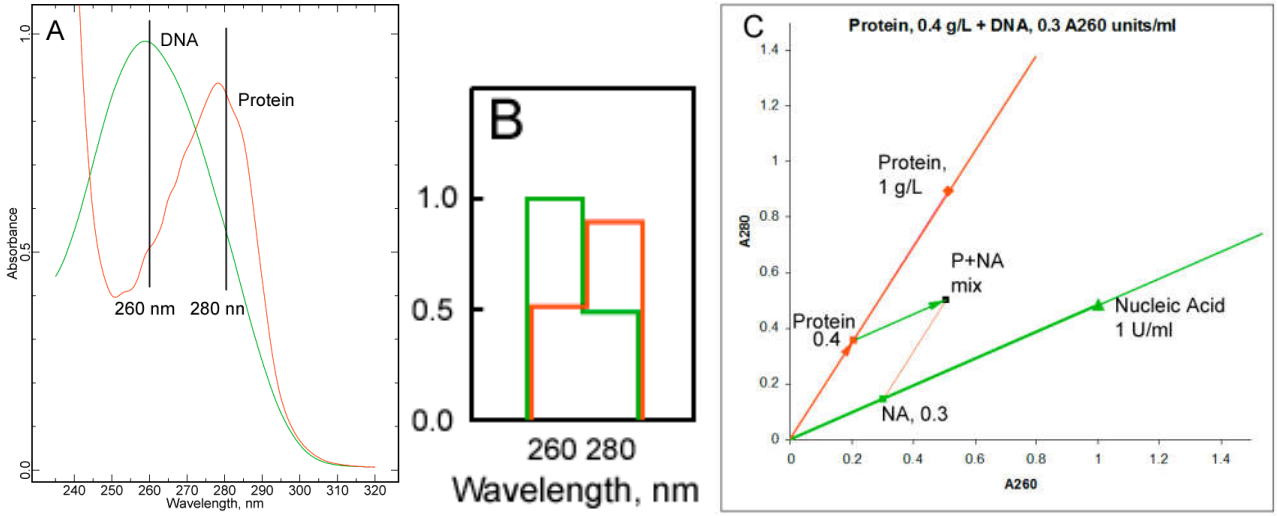

Figure 2. Absorbance spectra of Nucleic Acid and Protein. A, B: spectral view with 901 or 2 wavelength points, respectively. C: Vector view of the two 2-wavelength spectra and a hypothetical mixture.

The spectrum of any mixture will be a linear combination of these vectors. In Figure 2C we have plotted the coordinates of a mixture containing Protein at 0.4 and NA at 0.3 units concentration (black dot). Any mixture of protein and NA will lie in the 2-dimensional subspace spanned by the two standard spectra. That isn't saying much in the present case, because they span the entirety of our 2D vector space. (footnote: If you consider that concentrations are always positive, all solutions lie in between the positive standard vectors; but in dealing with difference spectra we often have negative changes in absorbance).

But suppose we use three wavelengths, adding a point at 270 nm as the third dimension Z. Both substances absorb at 270, so their standard spectra will no longer lie in the 260-280 (x-y) plane but will slant upward in the 270 (z) direction. Again any solution will be a linear combination of the two spectra, but they span a plane in three dimensions, in which all solutions of mixtures must fall. And the same is true in any higher dimension.

Now we can get a feeling for how "probe" spectra work. The probe spectrum for protein must satisfy two conditions- its dot product with the standard protein spectrum must be 1, and it's product with the NA spectrum must be zero. Remember that the dot product of a vector A with a unit-length vector in any direction, gives the length of the component of vector A along that vector. We will consider later the (rather trivial) problem of scaling the probe spectrum so its product with the protein spectrum is one. For it's product with the NA spectrum to be zero, it must be orthogonal (at right angles to) that spectrum. In 2D there is only one direction which is orthogonal to a given line. In Figure 3 we have drawn a blue line through the origin perpendicular to the NA spectrum. The "probe" spectrum for protein lies along this line. Notice that the direction of the protein probe is determined not by the protein spectrum, but the NA spectrum which it must be orthogonal to. This is generally the case when the number of substances to be determined is equal to the dimension of the spectra, i.e. the number of wavelengths. In general the blue line would be a line or subspace that is orthogonal to all the spectra other than protein.

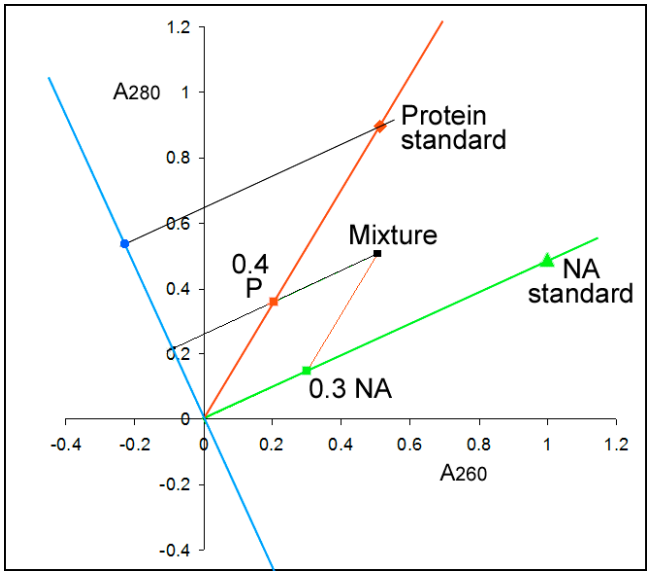

Figure 3. The probe spectrum for protein (blue) is orthogonal to the spectrum for nucleic acid (green). The length of the projection of any point onto the probe spectrum is a measure of the amount of protein, in the presence or absence of nucleic acid. And this is what is measured by taking the dot product of the spectrum with the probe spectrum.

Because the spectrum of NA at any concentration would lie along the green line, its component along the blue line is zero (its projection onto the blue line is zero). furthermore adding a portion of NA to any solution moves the point in a direction perpendicular to the blue line, so does not change the component of that spectrum on the blue line: Notice that the red dot "0.4 P" and the black dot labeled "mixture", which have the same concentration of protein, have the same projection on the blue line. and that projection is 0.4 of the way to the point where the protein standard spectrum projects.

And since the dot product of a vector A with a unit-length vector in any direction, gives the length of the projection of vector A on to that vector, it is clear that the dot product of a spectrum with the probe spectrum for a substance is a measure of the amount of that substance in the spectrum. This will be discussed more in a later section.

A second way to visualize processes in multi-dimensional space is to look at the projection of vectors onto a plane, a 2-dimensional subspace of the n-dimensional space.

This is analogous to using a camera to make a 2-dimensional picture of a three-dimensional scene- what we actually have is the projection of the scene in x,y,z space onto an x-y plane, perhaps with some perspective cues to let our brains regenerate the 3-D image. With the spectra, while we can't hope to visualize the n-dimensional scene, we can use linear algebra operations, corresponding to the graphical operations we did in the 2-D projection, to generate probe spectra that are orthogonal to the other components in n dimensions and give 1 when dotted into the corresponding component.

The two-wavelength method we used above to illustrate the protein/nucleic acid problem is actually one example of this, as two wavelengths span a plane in n-dimensional space, just as the x-y plane in our picture of the 3D world is spanned by the x and y vectors. All the other dimensions are perpendicular to these two and don't show up in the projection, so the figures we drew were actually the projections of the n-dimensional spectra onto the 260-280 plane.

A more general approach is to depict the projection, not onto two principle axes like (x, y) or (260, 280), but onto a plane spanned by two substances of interest. Any linear combination of the two will lie in the same plane. In this case the plane is not aligned with any of the principal axes, so we can't label the axes, except in terms of Euclidean distance in absorbance units. But we can get the angle between the vectors in N-dimensional space from their dot product, and their length from the norms of the vectors. One use for this is where a substance has two spectral forms; so two substances, and the sum of the concentrations of the two is constant as one gets converted into the other.

Figure 4 shows an example where cytochrome c is converted stepwise from the oxidized form to the reduced form. The gray spectrum, the one with the lowest absorbance at 550 nm, is the fully oxidized form. The spectra with successively increasing A_{550} represent stepwise conversion of some of the oxidized form into the reduced form, as the sample is titrated with small portions of reducing agent.

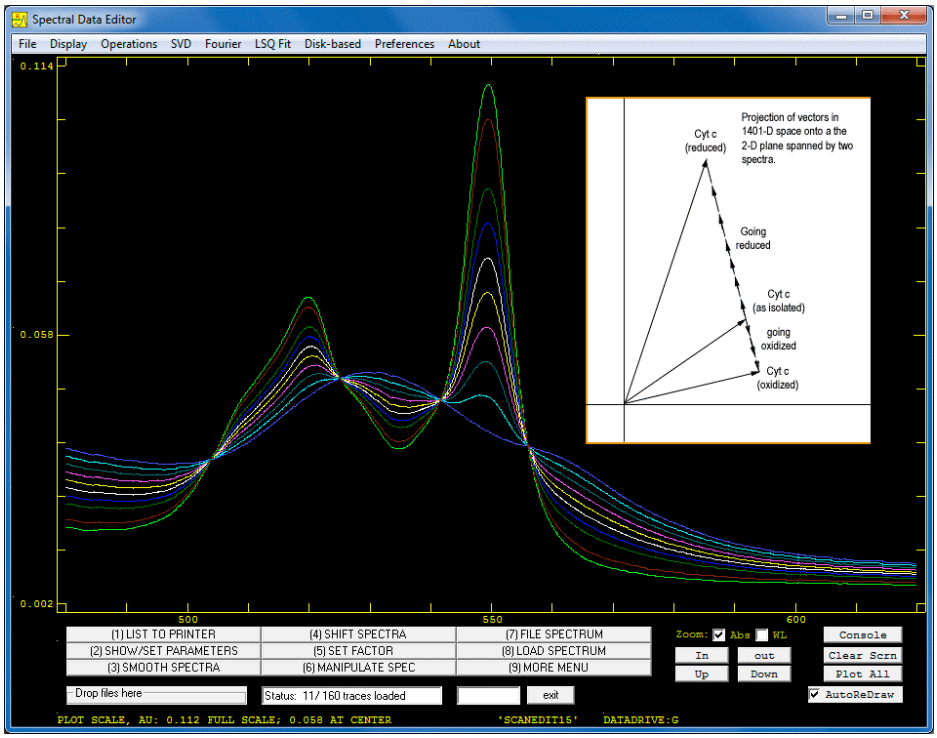

Figure 4. Successive spectra as oxidized cytochrome c is converted into the reduced form by small additions of reductant. The presence of isobestic points, where the spectrum doesn't change due to both forms having the same absorbance, indicates that it is a one-dimensional change. The inset shows vectors for the oxidized and reduced forms and one intermediate, as well as how the incremental portions of the diffrence spectra add to the oxidized form to eventually make the reduced form.

The total concentration of cyt c (oxidized and reduced) remains the same. Since the increase in concentration in one is equal to the decrease in concentration of the other, the only spectral changes occurring are multiples of the (reduced minus oxidized) difference spectrum. At wavelengths where the oxidized and reduced have the same extinction coefficient, the absorbance does not change, so all the lines intersect at these "isosbestic" points. Isosbestic points are an indication that a single chemical process is happening, since it is unlikely that two different processes would each make no change at the same precise wavelength points.

That is the spectral view. The inset in Figure 4 shows the vectorial view. As mentioned, we are looking in the plane spanned by the two forms of cytochrome c. The difference spectrum, being a linear combination of the two, also lies in this plane, as would any mixture of oxidized and reduced form in various proportions and total amount. (The X and Y axes are misleading since none of the principal axes of the N-dimensional space would lie in the plane.)

Since the only change occurring is the conversion of oxidized to reduced form, all the changes lie in the same direction, that of the reduced minus oxidized difference spectrum. This is illustrated by appending short vectors to the oxidized spectrum, along the line between oxidized and reduced. While all combinations of oxidized and reduced forms would span the plane, if we are given the restriction that the total cytochrome is constant and the only change is interconversion, all possible spectra lie along a line between the fully oxidized and reduced spectra at that concentration. Sort of a 1-dimensional space, but offset from the origin by the absolute (as opposed to difference) spectrum.

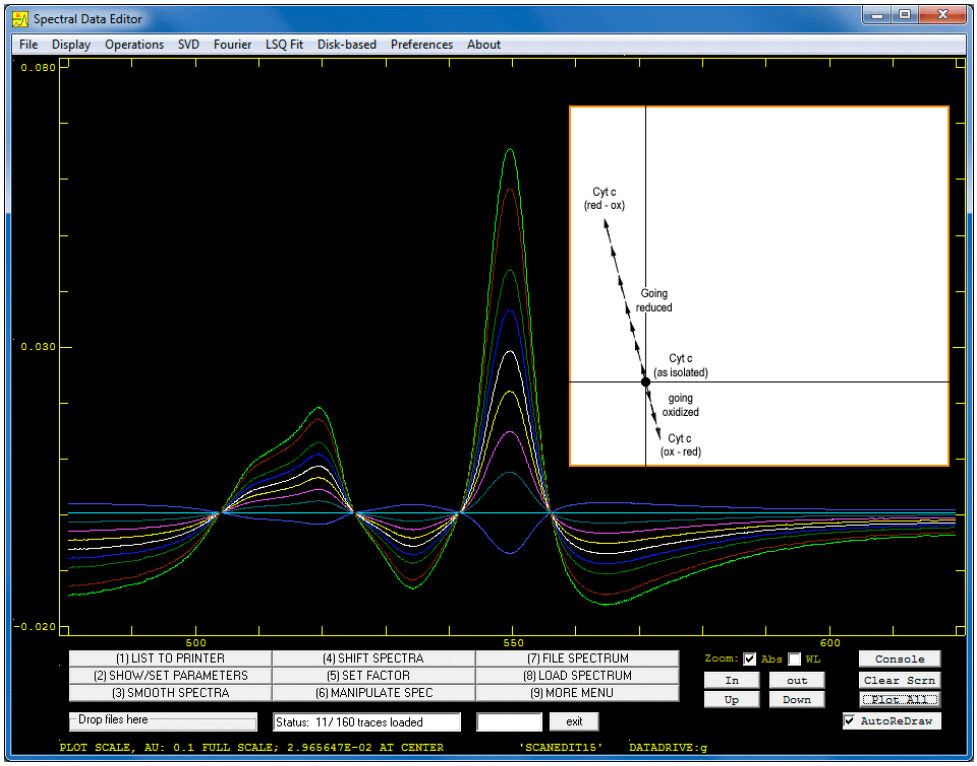

Figure 5. In the simulated experiment of Figure 4, the spectrum of one of the intermediate forms has been subtracted from all. Now it is clear from the spectral view that all spectra are multiples of the same thing, the diffrence spectrum. The vector view shows that the intermediate that was subtracted has moved to the origin, and the series of points makes a line through the origin.

In Figure 5 we have subtracted one of the spectra in Figure 4 (specifically the light blue one) from all the spectra to show the difference spectra. Clearly all the spectra are multiples of each other, and the isobestic points are at zero, because there is no change at those points. In vector view, subtracting one spectrum moves that spectrum to the origin, and moves all the other spectra by the same amount, resulting in the line of points passing through the origin. Now this is truly a one-dimensional space, the subspace spanned by the difference spectrum.

8. **More on the Probe spectra**

If the number of substances is equal to the number of wavelengths sampled, i.e. the dimension of the space, then the requirement that the probe spectrum be orthogonal to all other substances makes the probe spectrum unique- in n dimensions, there can be only one direction that is orthogonal to n -1 linear independent vectors. (In two dimensions, this means there is only one direction that is perpendicular to a line; in three dimensions there is only one direction which is perpendicular to two vectors or the plane that they span.)

However when the number of wavelengths is greater than the number of substances, which is definitely the recommended approach, multiple vectors (a whole subspace) would satisfy the co-orthonormality requirement. For two substances in 3 dimensions, the probe spectrum for the second has to be orthogonal to the standard spectrum for the first (and not orthogonal to its own standard spectrum). The subspace perpendicular to the first spectrum is a plane, and every vector in that plane (except one, which is also orthogonal to the second spectrum) is a potential probe spectrum for the second, after scaling to give 1 when dotted into the standard spectrum. The procedure for determining

the probe spectra should give the ones that will give the most accurate results in the face of ever-present measurement error/noise. It can be shown that the best choice is still the component of E that is orthogonal to the subspace spanned by the spectra of all the other components.

9. Error analysis, and the Standard Deviation Multiplier.

Measurements are always associated with some degree of error, and it is important to know how that affects any results we calculate from the measurements. This requires applying the rules for error propagation to the equation by which concentration is calculated from the absorbance measurements. We assume that the extinction coefficients, the standard spectra, are perfectly accurate and that the wavelength positions are accurate- that may not be the case, but that is a problem to be dealt with elsewhere. Then the only significant source of error is the error in measuring absorbance of the sample at each wavelength. For simplicity, assume each absorbance measurement includes a random error which is normally distributed with an average of zero and a standard deviation of σ.

For measurement at a single wavelength, the concentration is calculated by dividing the absorbance by the extinction coefficient at that wavelength, or, to make the analogy with the probe spectrum, by multiplying the absorbance by the reciprocal of the extinction coefficient (the absorbance-measurement is a one-wavelength spectrum, and i/ε is its one-wavelength probe spectrum):

$c = 1/\varepsilon \times A$

(since the extinction coefficient is by definition the absorbance of a solution at unit concentration, if the concentration of our test solution is 1, A = ε and c = 1. For higher or lower concentrations, the absorbance will be proportionately higher or lower, and c will be the correct concentration.) Now if A includes an error represented by σ that error will also get multiplied by 1/ε:

$$c = 1/\varepsilon \times (A \pm \sigma)$$
$$= (1/\varepsilon \times A) \pm (1/\varepsilon \times \sigma)$$
$$= c^\circ \pm (1/\varepsilon \times \sigma)$$

where c° is the actual concentration we would get from a perfectly accurate absorbance measurement. Or, denoting the standard deviations in absorbance and concentration as σ_A, σ_c

$\sigma_c = 1/\varepsilon \times \sigma_A$

so the standard deviation of the calculated concentration is the standard deviation of the absorbance measurement multiplied by the reciprocal of the extinction coefficient. For a given error in the absorbance measurement, the larger the extinction coefficient the smaller the resulting error in the concentration. This is why for this method one measures the absorbance at a peak in the spectrum, where the extinction is highest. As a special case of the principle described below, we can also call the quantity 1/ε, the Standard Deviation Multiplier.

In the case of measurement at multiple wavelengths, such as every point in the whole spectrum, and with the possibility of multiple absorbing species, the

concentration of the jth species is calculated as the dot product of the probe spectrum for that species with the absorbance spectrum.

$C_j = \mathbf{P_j \bullet A} = \sum P_{ij} A_i$ (sum over wavelength i)

where P_j is the probe spectrum obtained by whatever procedure, and A is the measured spectrum of the solution.

The rule for error propagation on addition is that the variance (σ^2) adds, so the variance of the concentration calculated from the absorbance measurements is:

$\sigma^2(C_j) = \sum \sigma^2(P_{ij} A_i)$ (where σ(x) denotes the standard deviation of x)

As in the single wavelength case, the rule for multiplication by a constant is that σ is also multiplied by the constant, so

$\sigma_i(P_{ij} A_i) = P_{ij} \times \sigma_i(A_i)$

and

$\sigma_i^2(P_{ij} A_i) = P_{ij}^2 \times \sigma_i^2(A_i)$

Putting that into the previous,

$\sigma^2(c_j) = \sum (P_{ij}^2 \times \sigma_i^2(A_i))$

for simplicity we will assume σ_i is the same at each wavelength i, which is a good approximation when the absorbance is not too high and the wavelength range not too long (The error will increase due to shot noise when very little light is getting through or the wavelength is at the limit of the range of the light source or detector). Thus:

$\sigma_i(A_i) = \sigma_A$, a constant

So we replace $\sigma_i(A_i)$ with constant σ_A and factor it out of the sum:

$\sigma^2(c_j) = \sum (P_{ij}^2 \times \sigma_i^2(A_i))$

$\sigma^2(c_j) = \sum (P_{ij}^2 \times \sigma_A^2)$

$= \sigma_A^2 \times \sum (P_{ij})^2$

and taking the square root,

$\sigma(c_j) = \sigma_A \times \mathrm{sqrt}(\sum (P_{ij})^2)$

So the standard deviation in measurement of the jth substance is just the standard deviation of absorbance measurement multiplied by the square root of the sum of squares of the elements of the probe spectrum, that is, the length, or norm, (|P|) of its probe spectrum. We call this quantity the "standard deviation multiplier" (SDM) for the jth substance. Given a choice between potential probe spectra, we should choose the one with the smallest SDM to minimize sensitivity to measurement error. In cases where the number of wavelengths is greater than the number of substances, the choice of the probe spectrum is not fixed by the requirement that it be perpendicular to the spectra of all other substances, so we define in a way that minimizes its magnitude and thus the effect of errors on the calculated concentration. (This is achieved by taking its direction as the component of the standard spectrum E_j that is orthogonal to all the other standard spectra E_k, k≠j. later?)

10. The magnitude of the probe spectrum-

So far we have concentrated on the direction of the probe spectrum. The probe spectrum for a substance must be orthogonal to all the other spectra so that its dot product with the spectrum of a mixture is insensitive to the presence of the other substances, while giving a result proportional to the amount of the substance of interest. Now we consider what the magnitude of the probe spectrum should be in order that its dot product with the standard spectrum will give a value of 1, required by the orthonormality condition.

As described in section 5, the dot product of two vectors is equal to the product of their magnitudes and the cosine of the angle between them. This ensures that the dot product of the spectrum of an unknown with the probe spectrum is proportional to the amount of the standard spectrum in the unknown (i.e. concentration of this substance in the solution). It is also proportional to the magnitude of the probe spectrum so, by adjusting that, we can arrange that the dot product with the standard spectrum is 1, thus calibrating it so that the dot product with an unknown solution is the concentration of the substance in the mixture.

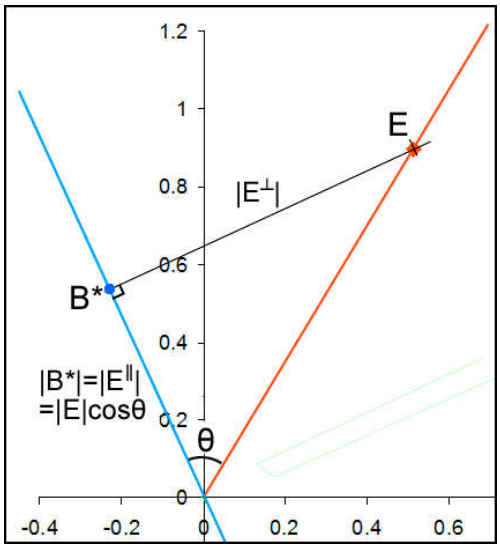

Figure 6. Decomposition of the standard spectrum E (red) into components along and perpendicular to the direction of the probe spectrum (blue).

Let B be a vector in the direction of the probe spectrum, along the blue line in Figure 6. Now as in section 5 divide the standard spectrum E into the sum of two vectors, one orthogonal to B and one along B (Figure 6). We will call the component along B, B*. The length of B* is $|B^*| = |E|\text{Cos}\theta$.

The dot product of E with B is $|E||B|\cos\theta$. Let the length of B be 1, that is B is a unit-length vector in a direction orthogonal to the spectra of the other components. (If it is not unit length, this can be achieved by replacing B with B/|B|). Now the dot product of E with unit-length B is $|E|\times1\times\cos\theta$, $= |E|\cos\theta = |B^*|$. For the probe spectrum, we want the dot product with E to be 1, so we divide the unit-length **B** by |B*| to make $\mathbf{P} = 1/|B^*|\times\mathbf{B}$. Since B is unit-length, the length of the probe is $|P| = 1/|B^*|$, the reciprocal of the length of the component of E along B. Or, in terms of the component B* of E along B, we could make the unit vector as $B = 1/|B^*|\times\mathbf{B}^*$, in which case the probe is $P = 1/|B^*|^2 \times \mathbf{B}^*$, again with $|P|=1/|B^*|$.

In the previous section it was concluded that minimizing |P|, the standard deviation multiplier, minimizes sensitivity to error in the absorbance measurements. Since |P| is the reciprocal of |B*| this is equivalent to maximizing |B*|, the length of the component of E along the probe. Since |B*| = |E|cosθ, this is achieved for large E and small θ (Cos(0)=1). This is sensible because with large E, large changes in absorbance represent small changes in concentration, and if θ is close to zero we are not losing too much of that E in the process of making it orthogonal to the other spectra.

Note that the magnitude of B* will always be less than or equal that of the standard spectrum E. The ratio |B*|/|E|, which is less than or equal to1, can be called the "linear independence" of E from the other spectra. It is the fraction of E that remains after the component that is linearly dependent on the other spectra is subtracted out, leaving the orthogonal, linearly independent component of **E**.

In section 6 it was said that any spectrum whose dot product with a given spectrum is 1 could be considered an "inverse" of that spectrum, and that if the inverse is colinear with the given spectrum, the length of the inverse must be the reciprocal of the length of the given spectrum. The probe spectrum is an inverse of this sort for B*. It is also an inverse for E which is not colinear. In general the length of the inverse spectrum must be the inverse of the length of the component (B*) of the given spectrum (E in this case) along the inverse spectrum.

Now lets apply this to the protein/nucleic acid case, in which there was only one other substance (NA); and bring the spectrum of that substance into the picture (green line in Figure 7).

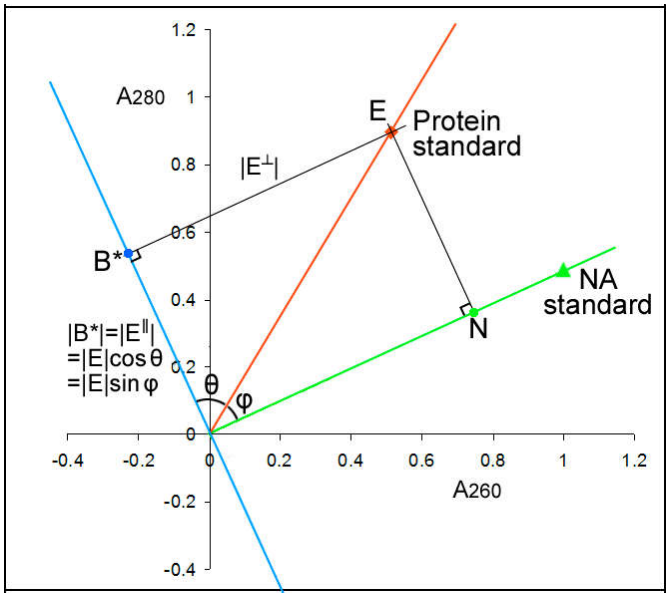

Figure 7. The component of E along the probe spectrum for one substance (protein) is the component of E perpendicular to the standard spectrum of the other substance (NA).
In general, it is the component orthogonal to to the standard spectra of all the other substances.

Remember this is analyzing protein and nucleic acid based on absorbance at two wavelengths, 260 and 280 nm. The direction of the probe spectrum for protein is defined by the blue line perpendicular to the green nucleic acid standard. The standard spectrum of protein E has been divided into components

along and perpendicular to this blue line. The component of E along the blue line is the vector B*.

The components of E along and perpendicular to the nucleic acid spectrum are also drawn in Figure 7; from the origin to N (green) and from N to E (black). These segments complete a rectangle, the diagonal of which is E. This illustrates the fairly obvious fact that B*, the component of E *along* the blue line orthogonal to NA, is equal to the component of E *orthogonal to* the NA spectrum (black line from N to E). If B is orthogonal to NA, then the component of E *along* B is the component of E *orthogonal to* NA, and vice versa.

This allows B* to be calculated directly from E and NA: Take the component of E along NA ($E \bullet NA/|NA|^2 \times \mathbf{NA}$) and subtract vectorially from E to get the orthogonal component, $\mathbf{B^*} = \mathbf{E} - (E \bullet NA/|NA|^2 \times \mathbf{NA})$. In terms of the angle between E and NA, labelled in Figure 7, |B*| is |E| times the sin φ. Note that φ, is the complement of the angle θ, and the sine of (90° - θ) is equal to the cos θ, so we don't have a contradiction with the previous $|B^*| = |E| \cos \theta$.

~~This allows a more concise description of B* and the probe spectrum. B* is the component of E orthogonal to NA (or in general, orthogonal to all the other spectra). Take B* and multiply by 1/|B*| to make a unit-length vector in the direction of B*. Now multiply by 1/|B*| again to make the probe spectrum: $\mathbf{P} = (1/|B^*|^2) \times \mathbf{B^*}$. Notice that |P| calculated this way is still equal to 1/|B*|. To make the probe spectrum P, take the component of E orthogonal to all the other spectra and multiply it by the reciprocal of its length squared.~~

Note that the magnitude of B* will always be less than or equal that of the standard spectrum E. The ratio |B*|/|E| can be called the "linear independence" of E from the other spectra. It is the fraction of |E| that remains after the component that is linearly dependent on the other spectra is subtracted out, leaving the orthogonal, linearly independent portion. If two of the spectra being analyzed are very similar, their orthogonal components (B*) will be small, so their linear independence is small and their SDM's will be large.

This was done assuming the number of wavelengths (the dimension of the vector space) is equal to the number of substances being analyzed, so that the direction of B* (the blue line) is fixed by the need to be orthogonal to n-1 other components. In cases where the number of wavelengths is greater than the number of substances, there are multiple choices for the direction of B* that would be orthogonal to the other components. For any of these choices, the component of E along that direction would define a B*, and unless the direction is orthogonal to E, a probe spectrum could be calculated. However it can be shown (appendix x) that choosing B* as the component of E orthogonal to all the other spectra gives the largest |B*|, hence the smallest |P| and the least sensitivity to error in the measurements.

<stop here>

Appendix 3 (have a better). B* is the orthogonal projection of E onto the orthogonal space (because E-B* is colinear with NA which defines the orthogonal space), therefore E-B* is the shortest distance from E to the orthogonal space. If

B* is replaced by the orthogonal projection of E onto any other direction in the orthogonal space, E-B* prime would be longer, and since E-B* and B* are two legs of a right triangle whose hypotenuse is E, B* would be shorter.

example- take Figure 7 as a 3D problem, imagine the blue line (orthogonal space) is a disk or plane perpendicular to NA, viewed edge on. Now rotate so that we are looking down the green NA vector. Now we are looking flat onto the orthogonal plane, with the origin in the center. The NA vector becomes a green dot at the center. The protein (E) vector is coming up above the plane. Its projection onto the plane, B* is on the plane directly under E. Now make vectors in the plane, radiating out form the origin in all directions. The orthogonal projection of E onto each of these vectors will still appear orthogonal in this view (the vector is in the plane of the paper so its orthogonal space makes a plane perpendicular to the paper which will appear as a perpendicular line, any vector orthogonal to it will be within that line). So we can draw lines from E to the vectors, intersecting the vectors at tight angles, and clearly the projection of E onto these vectors gets shorter as they move away from the projection onto the plane.

Wikipedia refers to subspaces U and V of vectors space W. P is projection from W onto U, and Q is projection from W onto P. Any vector in W can be divided into a component within U, P(x), and the remainder x-P(x). The spaces U and V are complementary, as are the operators: Q = $\mathbf{I}$ - P, Q = $\mathbf{I}$ - P.
https://en.wikipedia.org/wiki/Projection_(linear_algebra)#Complementarity_of_image_and_kernel

The projection operators can be expressed as matrices. For projection onto (like x axis, or xy plane) they are just identity matrices with some diagonal elements zero.

To generate a matrix that takes the component within a space, multiply x on the left by a matrix whose rows are an orthonormal basis for the subspace. This gives a column vector whose elements are the coefficients to express P(x) as an ZLC of these vectors. Then multiply that, on the left, by a matrix whose columns are the basis, to make that LC. Looks like M^TM, except M has to be orthonormalized first. Then MTM will be an identity matrix? except it is not square. and they are multiplied wrong way around, to give a large, singular matrix.

Image and Kernel: The image of a general function is the set of all values it could produce operating on all possible inputs. For orthogonal projection onto a subspace, the subspace is the image. The kernel of a linear map is the set of all input values that are mapped to the zero vector. The projection is W => W, but to a subset U of W. U is the image of P. Anything orthogonal to U (to all vectors in U) maps to the zero vector. S other orthogonal space V is the kernel of P. for projection Q, which is the projection onto V, the roles reverse.

Now it is to be shown that~~, when the direction of B is not determined by the orthogonality requirement, that is when the number of wavelengths is greater~~

~~than the number of substances,~~ The choice for direction of B that maximizes |B*| is along the component of E orthogonal to the span of the other spectra.

In the case that the dimension of the vector space (the number of wavelengths, (N) is equal to the number of substances being analyzed (M), that is the only choice. But when the dimension is greater than the number of substances, the vector space can be divided into two subspaces: the subspace spanned by the M-1 other spectra, and a subspace spanned by all the vectors orthogonal to the other spectra. This orthogonal subspace has dimension (N - (M-1)). In the case that N = M, it has dimension 1, there is only one direction that is orthogonal to n-1 linearly independent vectors, as in the two-dimensional Warburg-Christian plot of Figure ?.

In the discussion of section ?, it was mentioned that the decomposition of a vector into components along and perpendicular to another vector could be generalized to decomposition of a vector into a component within a subspace plus a component orthogonal to that subspace. By "within" a subspace, we mean a vector that can be expressed as a linear combination of the vectors whose span defines the subspace, in this case the other spectra. By "orthogonal to" a subspace, we mean orthogonal to each of the vectors within that subspace.

Now remember from section x that the length of the probe spectrum must be the reciprocal of

Given this result, we can explain why the best choice for the direction of the probe spectrum is always the component of E orthogonal to the span of all the other spectra. Remember from section ? that an arbitrary vector, in this case Ej, can be expressed as the sum of a component within a subspace (n this case the span of the other E vectors), and a component orthogonal to that subspace. These two components, together with the vector E as hypotenuse, form a right triangle. It can be proved, but it should be obvious in case the subspace is xxxx

(this was originally in component along . . ., but parts needed to be moved later. I think they are all there now, this can delete?)

The first application we will make of this is when vector B is the component of E_j orthogonal to the spectra of all other substances, defining the direction (but not the magnitude) of the probe spectrum P_j. Given the conclusion based on error analysis, that the length of the probe should be as small as possible, you might think that the component of E orthogonal to all the others should be as small as possible. But we will see in a later section that after "normalization" to make the dot product with E equal to one, the length of the probe spectrum is the inverse of the length of its component orthogonal to the other spectra, as fitting since we said the probe spectrum is a sort of inverse of standard spectrum, Thus this length should be maximized, which means the angle between E and P should be minimized.

Beer's law in matrix form

Chapter 3. description of the procedures

Procedure 1: matrix inversion and classical least squares fitting.

We said this can be divided into three cases- (1) single wavelength, single substance; (2) multiple substances with a number of wavelengths equal to the number of substances; or (3) with a number of wavelengths greater than the number of substances to be analyzed.

Actually the first can be considered a special case of the second. The field of scalars can be considered a vector space of dimension 1, where the vectors have a single element (x) instead of (x,y) or (x,y,z). Defining the dot product of two vectors as the sum of products of corresponding elements of each vector, the dot product in one dimension reduces to the product of the two numbers. Now the standard spectrum of a substance is just the extinction coefficient ε of the substance at the single wavelength chosen. The probe spectrum must be orthogonal to all other standard spectra. But there can be no other substances, since we can't analyze more substances than the number of wavelengths. And it must give 1 when dotted into the standard spectrum. This will happen only if its single element is the reciprocal of the single element of the standard spectrum, and our probe spectrum is just (1/ε). So we measure the absorbance of our sample at the specified wavelength, and dot into (multiply by) (1/ε) to get the concentration. But this is what we have always been doing, except we express it as dividing by the extinction coefficient rather than multiplying by its reciprocal.

Noise analysis- The absorption measurement is associated with a certain amount of random error, which we will assume has a more or less gaussian distribution, characterized by a standard deviation σ. When we multiply the measurement by 1/ε we multiply the error by the same, so our calculated concentration has a standard deviation of σ/ε. Thus the most accurate measurements are made when 1/ε is small, i.e. ε is large. Assuming roughly constant σ as a function of wavelength, this is done by taking the measurement at an absorbance peak. This result is also obtained using the standard deviation multiplier defined for probe spectra, that being the norm or root of sum of squares of the elements of the probe spectrum. Since the probe spectrum has only one element, 1/ε, this reduces to just the value of that one element, 1/ε.

For the cases where there are more than one substance to analyze and an equal or greater number of wavelength points in the spectra, we need procedures to generate a probe spectrum for each substance, orthogonal to the other spectra of the other spectra, and giving 1 as inner product with the standard spectrum of that substance. We can't just subtract the components along each other spectrum, because they are not orthogonal. Say we subtract the component of spectrum one along spectrum 2 from spectrum 1; Now 1 and 2 are orthogonal. But spectrum 3 contains a component along 2, so when we subtract the component along 3 from 1 we re adding back some of component 2, and 1 is n longer orthogonal to

12. How can we generate standard spectra (in problems of type 1)?

Fortunately, there is another very common application that requires developing a co-orthonormal set of vectors, that you are probably familiar with: matrix inversion.

A concise way of expressing the co-orthonormality relation is in matrix form: Let $\mathbf{P^T}$ be a matrix whose rows are the probe spectra P_j of the substances, and **E** be a matrix whose columns are the standard spectra Ej of the different substances. Consider the matrix product $\mathbf{P^T E}$:

$$\begin{matrix} ---P_1--- \\ ---P_2--- \\ ---P_3--- \end{matrix} \quad \begin{matrix} | & | & | \\ \times & E_1 & E_2 E_3 \\ | & | & | \end{matrix} = \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}$$

, the identity matrix.

This means that $\mathbf{P^T}$ is equal to the inverse of the matrix E; the probe spectra are the rows of $[E]^{-1}$.. The inverse of a square matrix is unique, so the probe spectra are unique. Whatever means we use to generte the probe spectra, n the case that the number of wavelengths is equal to the number of substances considered, the result will be the same as the rows of E^{-1}.

For the cases where there are more than one substance to analyze and an equal or greater number of wavelength points in the spectra, we return to equation 3, $A_i = \sum c_j \boldsymbol{\varepsilon}_{ij}$.

11. Beer's law in matrix form

This will come up over and over again, it is the fundamental equation of this study, so it is important to have a good intuitive grasp of what it means.

Now we have all we need to express the spectrum of a mixture as a linear combination of the spectra of the substances. Applying beers law for a mixture (2) to each (I^{th}) wavelength of the spectrum gives:

[3]

Which for four substances and four wavelengths expands to:

$$A_1 = E_{11}C_1 + E_{12}C_2 + E_{13}C_3 + E_{14}C_4$$
$$A_2 = E_{21}C_1 + E_{22}C_2 + E_{23}C_3 + E_{24}C_4$$
$$A_3 = E_{31}C_1 + E_{32}C_2 + E_{33}C_3 + E_{34}C_4$$
$$A_4 = E_{41}C_1 + E_{42}C_2 + E_{43}C_3 + E_{44}C_4$$

Here we recognize the columns of the right-hand-side as being the standard spectra E of each substance multiplied by the concentration of that substance.

So if we measure at 4 wavelengths, we have four simultaneous equations, and if the extinction coefficients are sufficiently different (linearly independent) we can solve for the C, given the A, by successive elimination.

We can also notice that [3] ($A_i=\sum E_{ij}C_j$) looks like matrix multiplication, where A_i is a column vector of the measured absorbances, E_{ij} is the matrix of extinction coefficients (with the jth column being the standard spectrum of the jth substance), and C_j is a column vector of the concentrations of the different substances, like:

$$\begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} = \begin{bmatrix} E_{11} & E_{12} & E_{13} & E_{14} \\ E_{21} & E_{22} & E_{23} & E_{24} \\ E_{31} & E_{32} & E_{33} & E_{34} \\ E_{41} & E_{42} & E_{43} & E_{44} \end{bmatrix} \times \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{matrix}$$

A quick reminder of how matrix multiplication works: when a matrix pre-multiplies a vector (on its right), the result is a linear combination of the columns of the matrix, each column being multiplied by the corresponding element of the vector: A commonly taught mnemonic for matrix multiplication is to take the vector on the right and lay it out across the top of the matrix, and for each row calculate Ai by multiplying each Eij in the row by the Cj at the top and summing them (dot product of the C vector and the i^{th} row of the matrix)

$$\begin{matrix} & C_1 & C_2 & C_3 & C_4 \\ & \times & \times & \times & \times \\ A_1 & E_{11} & E_{12} & E_{13} & E_{14} \\ A_2 = & E_{21} & E_{22} & E_{23} & E_{24} \\ A_3 & E_{31} & E_{32} & E_{33} & E_{34} \\ A_4 & E_{41} & E_{42} & E_{43} & E_{44} \end{matrix}$$

Notice each (jth) column of E is the 4-wavelength "spectrum" of the corresponding (jth) substance. So the matrix multiplication is making a linear combination of the standard spectra of the individual substances, the amount of each being determined by the concentration of that substance. This means that, although in general spectra taken at I wavelengths exist in an I-dimensional vector space, the absorbance spectrum A exists in a J-dimensional vector space spanned by the spectra of the J different substances. If I is greater than J (more wavelengths than absorbing substances), all spectra of mixtures of the substances live in an J-dimensional subspace of the I-dimensional space of the spectra (in a perfect world with no noise). In this context we also refer to the

standard spectra (columns of E) as the "basis" spectra, because they provide a basis for the space in which all spectra of mixtures must reside. They are what you need for analyzing the spectrum of a mixture by the first procedure.

next put a figure with actual spectra drawn in for A and E.

11b. dealing with more than one spectrum.
training spectra, time course, titration

1. Matrix inversion: If the number of wavelengths I is equal to the number of substances J then E is square and we can solve [3] for the C_j by matrix inversion,

$$\mathbf{C} = [\mathbf{E}]^{-1}\,\mathbf{A}$$

Now to approach it from a different viewpoint- the concept of "probe spectra". The "probe spectrum" for a given substance (in a mixture containing some or all of a defined set of other absorbing substances) Is a spectrum which when dotted into the absorbance spectrum of the mixture, gives the concentration of that substance. This requires that the probe spectrum Pj of the jth substance is orthogonal to the spectra of all the other substances in the solution, and that its dot product with the spectrum of the jth substance is one (assuming the spectra all represent unit concentration).

$P_j \bullet E_k = 0,\ j \neq k$ (i.e. the vectors P_j and E_k, dotted along i) [4]
$= 1,\ j = k$

Then if the spectrum A is a linear combination of the spectra of the substances:

$$A = C_1 \times E_1 + C_2 \times E_2 + C_3 \times E_3 + C_4 \times E_4 + \ldots$$

say we take the probe spectrum for the 2nd component and dot into A:

$$P_2 \bullet A = C_1 \times P_2 \bullet E_1 + C_2 \times P_2 \bullet E_2 + C_3 \times P_2 \bullet E_3 + C_4 \times P_2 \bullet E_4 + \ldots$$

$$P_2 \bullet A = C_1 \times 0 + C_2 \times 1 + C_3 \times 0 + C_4 \times 0 + \ldots = C_2$$

(This is like the Fourier transform, where all the component frequencies are orthogonal, so you can integrate a given frequency against the composite waveform to get the amplitude of that frequency in the waveform)

So how can we make a probe spectrum? If you think about it,

If we make the matrix [P] with the probe spectra as rows, and use it to multiply the matrix [E] which has the standard spectra as columns, then [4] (co-orthonormality of the P and E spectra):

$P_j \bullet E_k = 0,\ j \neq k$ (i.e. the vectors P_j and E_k, dotted along i)
$= 1,\ j = k$

implies [P][E] = [I],

[I] being the identity matrix with 1 along the diagonal and 0 elsewhere.
because the j,kth element of the product matrix is the ith row of P dotted into the jth column of E. If I=J, the matrix is square and (in general) invertible, and the inverse matrix satisfies $[E]^{-1}[E] = [I]$. So the rows of the inverse matrix are the probe spectra for the standard spectra in E, and what we are doing when we solve simultaneous equations by matrix inversion:

$$\mathbf{C} = [\mathbf{E}]^{-1}\,\mathbf{A}$$

is dotting the A vector successively into each row of the $[E]^{-1}$ matrix, i.e. each probe spectrum, into each of the probe spectra, to get concentration C_j of the corresponding substance.

And the probe spectra are unique, as the inverse of a square matrix is unique. Each probe spectrum has to be orthogonal to n-1 other spectra in n-space, and there is only one more dimension it can go along. Like if you are drawing lines on a paper (2- space) , once you draw one line there is only one direction perpendicular to it, or in 3 dimensions, given two lines they span a plane in 3 space, and there is only one direction perpendicular to the plane. And once the direction is determined, the magnitude is fixed by the requirement that it give 1 when dotted into the standard spectrum.

But matrix inversion, using the same number of wavelengths as substances to be determined (substances), is a lousy method of analysis- first there is no "goodness of fit" test. Given 4 substances at 4 wavelengths, whatever 4 numbers you plug in for the absorbance, you will get a set of four concentrations. And if you back-calculate what the spectrum should be given those concentrations, it will match your spectrum exactly (to within the precision of the arithmetic). Matrix inversion works, and $\mathbf{C}=[\mathbf{E}]^{-1}\mathbf{A}$ implies $\mathbf{A}=[\mathbf{E}]\mathbf{C}$. So if there is an additional absorbing substance, or the molecules of the known substances interact in a way that changes their spectra, you will have no warning (unless the concentration comes out significantly negative).
Secondly, there is the question of choosing the n wavelengths that best differentiate the spectra. For well-separated absorption peaks, you would just choose the highest absorption peak for each component. But for well-separated peaks, you don't need matrix inversion. For overlapping peaks, it is more complicated to determine which wavelengths give the best differentiation of the substances; which wavelengths maximize the orthogonality of the spectra (although this could be handled).

Much better to take more wavelengths than the number of substances, ideally a whole spectrum across the range where they absorb. Now if there is an additional spectral component, the spectrum will be outside of the span of the

basis spectra and no linear combination of them will give a reasonable fit. (Whereas with I = J, the J spectra span the entire I-dimensional space, and any spectrum can be fit by them.)

But how can we use the extra information from say, 1001 points instead of 3 or 4? We can't use conventional matrix inversion, because the E matrix is not square, so not invertible. The problem is over-determined, we have too many data to determine a small number of parameters.

One way would be to select a number of wavelengths equal to the number of components and do matrix inversion as above. In an ideal world with perfect accuracy and zero "noise", any choice of J wavelengths should give the same answer. In the real world different sets of wavelengths will give different answers, and some choices will give greatly improved accuracy and noise immunity than others.

2. Classical least squares method

But in any case, you would not be making the best use of the available data. Why not let every wavelength point add whatever it can to the result? A way to do this is called least-squares fitting, because it minimizes the sum of the discrepancies between the predicted spectrum and the observed spectrum by its choice of the concentrations. Because of the Gaussian nature of error probability distribution this roughly translates into selecting the set of concentrations that would have the maximum likelihood of resulting in the observed spectrum (see appendix 1).

The equations for finding the least squares solution are usually derived using calculus, the point where the derivative of the sum of squares of "residuals" is zero so the sum of squares must be at a minimum wrt each concentration. For me it is much more intuitive to do it with geometry and linear algebra, as follows:

We want to find the set of concentrations that would predict a spectrum (within the span of the basis spectra) which would have the greatest likelihood of being measured as the actual measured spectrum, which is slightly outside of the span due to error/noise. Seeing the spectra as points in I-dimensional space (I being the number of wavelengths), it would intuitively seem that the point within the span that is "closest" to the actual measured point outside, would be the most likely. And because distance in Euclidean spaces corresponds to sqrt of sum of the squares along each dimension (Pythagorean theorem), the point that is closest is the point that minimizes sum of squares and (if std deviation of the measurement at each wavelength is the same) maximizes the likelihood of having resulted in the observed spectrum.

So how do we find the closest point? There is a theorem that the distance from a point within a subspace, to a point outside, is minimum when the vector between them is orthogonal to the subspace. This is obvious in two or three dimensions- if your subspace is a line, the point in the line is closest to some arbitrary point when the distance between the points is perpendicular to the line. If you draw that line, and then move up or down the line from that point and draw a line from there to the point outside, you are drawing the hypotenuse of a right triangle of which the shortest, perpendicular, line is one leg. Or if you have a two-

dimensional subspace, the subspace is a plane in the higher dimensional space, and the shortest distance from a point outside the plane to a point inside the plane is a line from the outside point "straight" (i.e. perpendicular) to the plane. (This can also be described as taking the projection of a vector outside onto the subspace, or finding the component of the vector within the subspace.)

So we want to find a set of concentrations that predicts a spectrum (inside the span) such that the distance (vector difference) between that and the observed spectrum is minimized, and the theorem tells us that this will occur when that difference is orthogonal to the subspace, i.e. has no component "along" any of the basis spectra. We call this difference R for the vector of "residuals", the leftover stuff that our equation can't fit. The real spectrum is the sum of the ideal spectrum and the residuals:

A = predicted spectrum + R

R = A - predicted spectrum

R = A - EC (predicting A by [3])

To solve, we set the dot product of R with each of the basis spectra to zero, giving I equations in I unknowns.

The (relatively simple) derivation is in appendix 2. The solution for best-fit concentration is $C_{bf} = E^{\perp}A$

Where $E^{\perp} = [E^TE]^{-1}[E^T]$, called the generalized inverse, or least-squares inverse, of E. $[E^TE]$ (and its inverse) are I by I square matrices. E^T has the basis spectra as rows, and multiplying by $[E^TE]^{-1}$ makes linear combinations of these rows so that the rows of the product are co-orthonormal with the original spectra, subtracting out the component of each other spectrum from the spectrum in question to make the probe spectrum. So while with the matrix inversion method, the probe spectra are the rows of the inverse matrix E^{-1}, with least squares the probe spectra are the rows of the generalized inverse matrix $E^{\perp}$.

Note that this is linear least squares. The result is a unique solution, obtained by a single matrix operation on the spectrum, unlike non-linear least squares which usually involves iteration, questions of convergence, dependence on starting point, and multiple local minima; which may have given least-squares fitting an unfavorable reputation.

should make a distinction between:
A is sum of the ideal spectrum and the errors
A is the sum of the "best-fit" spectrum and the residuals.
The error vector will not be precisely orthogonal to the span so the best-fit spectrum will be slightly different from the ideal spectrum.

3. Find the pure spectra from spectra of mixtures (training set) of which the composition is known.

Hybrid LLSQ/NLSQ methods for reaction time course, titration, elution profile

4."inverse calibration" Find the probe spectra given a training set -

- use 3 then 2

- combine those in a single step to make the probe spectra without determining the pure spectra.

5. Find the probe spectrum for component of interest from a training set with other (unknown but representative) components, but concentrations in the training sample are only available for the one of interest.
6. Principle components analysis.

2.5? Single analyte, multi-wavelength.

This is very instructive but of little practiacal importance. If there was only one absorbing compound, and bseline drift is negligible, you would probably measure absorbance at a peak, averaging ofer time to get an accurate value.

It is instructive to consider the case where there is only one absorbing substance, and we take it's spectrum in a region where it absorbs. We could calculate the concentratin from any single wavelength measurement where we have the extinction coefficient, preferably at a peak in the spectrum.

But suppose the spectrum is rather noisy, perhaps because the solution was very dilute. And examining the spectrum (Figure ?) we see that the wavelength we want use, say 280 nm, is significantly lower than the surrounding points at 129.9 and 280.1. If we average the values at 279.9 through 280.1 and use that as A280, would that give a mor accurate value? Probably yes.

But then, a more ambitious idea. We have a very accurate spectrum of this substance, so in principle we know the extinction coefficient at every wavelength and can estimate the concentration from every wavelength point. Can we do this in a way such that it further increases the accuracy, and does not dilute out the result from the pea wavelength by adding more noise than signal?

Yes- A standard statistical approach, weighted averaging, will give us the best answer,and we will go through that below. But first lets consider this in terms of the probe spectrum and standard deviation multiplier (which as we will see gives the same result). From section ??, to make the probe spectrum we take the component of the standard spectrum orthogonal to the spectra of all other substances being analyzed, and we multiply that by the inverse square of its norm. Since there are no other absorbing substances, the orthogonal component is just the standard spectrum itself, B* = E. The probe spectrum is then:

$$P = 1/|E|2 \times \mathbf{E} = 1/(E \bullet E) \times \mathbf{E}$$

Now to get the concentration C, we take the dot product of this with our noisy sample spectrum A:

$$C = P \bullet A = (E \bullet A)/(E \bullet E) = \sum E_i A_i / \sum E_i^2$$ (summation over the different wavelength points)

We can also recognize this as the length of the component of A along E, divided by |E| so that it gives 1 when that length is just the norm of P. Remember that theoretically, any solution of this one substance should be a multiple of the

standard vector E, and thus colinear with E. In terms of linear algebra, E spans the 1-dimensional subspace in whih all solution of this subsance should lie. But due to measurement error, the actual measurement A will be off the line. By expressing A as the sum of its component along E and its component othogonal to E, We are saying that the component along A represents the true spectrum of the solution, and the component of A orthogonal to E (the difference between A and its component along E) is due to error.

Because this difference is the component of A orthogonal to E, the component of A along E is the point along E that minimizes the difference, i.e. the assumed error, i.e the pont along (the span of) E that is closest to A. Intuitively it seems this should be the point along E that has the highest likelihood of resulting in the observed spectrum A, and we will see later that this is the case.

Now consider the precision, or reproducibility, of determining the concentration this way, that is, the sensitivity to random errors in the measurements. Can we show that this estimate is more accurate than just using absorbance at the peak? In section x we saw that the standard deviation of the calculated conentration is the standard deviation of the absorbance measurement multiplied by the norm of the probe spectrum which we call the standard deviation multiplier. In equation ? above, we said that the probe spectrum for a single substance is just the the standard spectrum for that component () multiplied by the inverse square of its norm ($1/|E|^2$). The norm of the probe spectrum is thus:

SDM = $|P| = 1/|E|^2) * |E| = 1/|E|$

If we just use the peak absorbance, say A280, then concenration is clculated from absorbance and extinction coeeficient at that wavelength by Beer's law:

$C = A_{280}/\varepsilon_{280}$, $\sigma(C) = \sigma(A_{280})/|\varepsilon_{280}|$,
SDM = $1/|\varepsilon_{280}|$

In other words, the error in measurement is multiplied by $1/E_{280}$. Clearly if |E| is greater than ε_{280}, which it is if the E includes the peak measurement at 280 nm as well as some other wavelength points, using more wavelengths give a more precise measurement.

Here $|\varepsilon_{280}|$ indicates absolute value. Normally this would be irrelevant as absolute extinction coeeficients are positive, but whndealing with difference extinction coefficients they cn be negative, and whether you mulltiply by a positive or negative number, the variance gets multiplied by the magnitude of that number. This way we never end up with negative standard deviations.

Since there is only one absorbing species present,its spectrum E spans the (one-dimensional) space of all possible spectra. By taking the dot product of the unknown spectrum A with E and dividing bty |E|, we are dividing the spectrum A into a component in that space, and a component orthogonal to that space.

Now it is clear that the larger |E| is, the smaller the SDM is, and the more accurate the result will be for a given level of error () in A. And it is clear that adding any additional wavelength points wher E is not zero, will increase |E| and make the result more accurate. Adding points where E is precisely zero will have no effect on the result or its error.

We cannot yet say this for the case where multiple substances are being analyzed. The size of the orthogonal component of E (B*) depends on the size of E and its linear independance. Suppose we have two substances whose absorption peaks are well separated but showe spectra overlap in between the peaks. If we take only two points, at the two peaks, where tha absorbance is mainly due to one or the other. The E value will be largely independent. Now if we include a point in the middle, where both have about the same absorbance, this would make |E| bigger for both substances, but make them less independant, that increase in |E| will not all show up in |B*|.

Now lets attack the same problem using weighted averaging. We will calculate the concentration from each measurement (from each wavelength oint) and make a properly weighted average of the results. ~~This involves seeing how the error in each sample is propogated through the averaging process, and weighting accordingly (??).~~

~~(Study bevington and "propagation of uncertainty" on wikipedia. Assume error at adjacent points is independent, even if absorbance is highly correlated.~~

~~This is actually the same as what we did to show that norm is the SDM.?~~

To make a weighted average, each of the values being averaged is multiplied by a weighting factor w before summing, and the sum, instead of dividing by N, is divided by the sum of the weights:

$C_{wa} = \sum W_i C_i / \sum W_i$

Clearly if each weighting factor is 1, the sum of the weights is N and the weighted average is the same as the conventional average. Likewise for any value of w, as long as they are all the same.

Statistics tells us that the weighting factor for each term should be the reciprocal of the variance of the estimate, or reciprocal square of the standard deviation σ. The values with larger stadard deviation are less reliable, and get less weight in the sum. This gives:

$$W_i = 1/\sigma^2(C_i)$$

where C_i is the concentration calculated from the ith point.

at each point the concentration estimate is calculated from Beer's law at a single wavelength:

$C_i = A_i/E_i$,

and as we saw in section ?, the standard deviation of the resul is that of the absorbance measurement times $1/E_i$.

$\sigma(C_i) = (1/E_i)\ \sigma(Ai)$

so

$W_i = 1/\sigma^2(C_i) = E_i^2/\sigma^2(A_i)$

and $W_iC_i = E_i^2/\sigma^2(Ai) \times Ai/Ei$

$= EiAi/\sigma^2(Ai)$

This looks peculiar because instead if dividing absorbance by the extinction coefficint, we are multiplyin by it. But remember the sum will be divided by the sum of weighting factors, which involve E_i^2:

$$C_{wa} = \frac{\sum W_iC_i}{\sum W_i}$$

$$= \frac{\sum E_iA_i/\sigma^2(A_i)}{\sum E_i^2/\sigma^2(A_i)}$$

Now if again we assume the standard deviation is the same at the different wavelengths, say σ_A, we can factor it out of the nemerator and denominater and it cancels, leaving:

$$C_{wa} = \frac{\sum E_iA_i}{\sum E_i^2} = \frac{E \bullet A}{E \bullet E} = (1/|E|2 \times E) \bullet A$$

Now lets go back to the idea mentioned above, that we are dividing A into a component along E, which we ake as the real spectrum, and a component orthogonal to E, which we assign to error (Fig ?). This is probably not right, as the random error is unlikely to be exactly orthogonl to E. But by choosing the error orthogonal to E, we are choosing the point along E which is closest to the observed spectrum. For anything like a Gaussian distribution, the arger the error the least likely it is for that error t occur, So we are choosing the point along E which has the gratest liklihood of generating the obsered spectrum.

But is our intuition that the shorter the error vector is, the more likely that such an error could occur; correct? Absolutely. Remember that the length, or norm, of a vector is given by the square root of sum of the squares of the components. In our two-dimensinal case, the error vector $A^{\dagger}$ is the vecor sum of the error in A280 and the error in A260 (Figure), and its length is the square root of the sum of the squares of those two errors. Thus by minimizing the length of , we are minimizing the squares of the errors in the two measurements.

For a gaussian distribution, the relative probability of an error of magnitude r is proportional to $e^{-r/\sigma}$. This would apply to both absorption measurements, and, if the measurements are independent, the probability of have error r280 on the A280 measurement and R260 on the A260 measurement is the product of their probabilitys or:

$e^{-r280\text{^}2/\sigma} \times e^{-r260\text{^}2/\sigma} = e^{-(r280\text{^}2/\sigma + r260\text{^}2/\sigma)}$

which is maximized by minimizing the negative exponent ($r280^2/\sigma$ + $r260^2/\sigma$). Thus the probablility is inversely related to the exponent of the sum of squrs of the errors, i.e. the length of the error vector.

While Figure x was drawn fo the two-dimensional case, this is directly generalizable to N dimensins. We could say that Figure x reprsents the projection of N-dimensional space onto the vectors E and A. Again we divide A into components along and perpendicular to E, by dotting A into E and dividing by $|E|^2$ and multiplying E by that to get the component along E, then subtracting that vectorially from A to get $A^{\dagger}$. The length of the error vector $A^{\dagger}$ is the sum of squares of the errors at each of the wavelengths, and by making it orthogonal to E we minimize the length and maximize the likelihood that an ideal spectrum $A^{\parallel}$ could result in the observed spectrum A.

But once we know this, we don't need to think about $A^{\dagger}$. Just get the multiplier $(A \bullet E)/|E|^2$. Since E represents unit concentration and this multiplier times E gives the corrected spectrum, this multiplier is the concentration of the substance in A.

The SDM has units of concentration, whatever the units standard spectra were expresed in. A large SDM for uM spectra may not be a problem if the samples contain the component at mM concentrations. more concerning is the linear independance.

Including the baseline as one component greatly increases the SDM? For one real componentn this decreases the max value at any wavelength from the true maximim to something like +/- 1/2 of that? Only if the mean value is ~1/2 the max. If there is a lot of baseline in the standard, the average subtracted out will be a smaller fraction of the max. While the max of absolute of is smaller, a large number of points that were near zero are now significantly negative.May not change SDM much. And if there are multiple components, they will each have their own DC components, which will have to be subtracted to make them orthogonal to each other?
Have to rely on the shape of the spectrum, not its DC position.
Dual wavelength ext coeff suffer the same, if the low point is well above zero- say measuring cyt c not at 550 but 50-535, the extinction is smaller, random error will be more significant.

Selecting a few wavelengths for analysis- should be extrema of the different probe spectra?

Procedure 1- finish up by extending to multiple unknown spectra, introduce "time course"
The rows of A (TC) are linear combinations of the rows of C (TC of each analyte)
The cols of A (spectra) are LC of the columns of E (spectra of each component)
Procedure 2- finding the standard spectra given training spectra with known composition
(In chapter? we discuss where the composition is unknown, but a parametrized model will give the concentrations,and by alternately optimizing the parameters given the current spectra and optimizing the spectra based on concentrations from the model with those parameters, a best fit can be obtained)

Go back to Beer's law in matrix form, A = EC, which we solved by multiplying both sides on the left by E^{-1}. defined such that $E^{-1}E = I$. E^{-1} here is called the "left inverse" of E, and it exists if the rank of E is not less than the number of columns of E, as explined earlier.

Earlier:
Remember: Given a **matrix M_{IJ}**,
The **elements** of the matrix are organized into **row**s and **column**s
I is the number of rows and the length of the columns
J is the number of columns, and the length of the rows.
However we also use $\mathbf{M_{ij}}$ as the i, jth element of M_{IJ}; that is the element in the ith row and jth column, in which case i amd j are are indices, or pointers, into the matrix::
i is the row index of M_{ij}
j is the column index of M_{ij}

Beer's law in matrix form: $A_{ik} = E_{ij}C_{jk}$ can refer to the elements, in which case this is the definition of matrix multiplication, or A=EC

The columns of E, are one for each analyte, so number of columns is the number of unknowns being solved for. The number of rows is the number of wavelengths, which are the data we will use. It is generally reckognized that the number of unknowns cannot be greater than the number of data, so the number of rows (length of columns) must be greater than or equal to the number of spectra (columns).

This can be be explained in terms of vectors spanning a (sub)space:

The columns of **E** are the spectra, the number of columns (if linerly independent) is the dimension of their span. The rows are the wavelengths, so the number of rows (= number of wavelegths, length of columns) is the dimension of the space in which the spectra exist. If the number of rows is less than the number of columns, the dimension of the space is smaller, and so the columns (spectra) cannot be linearly independent. There will be multiple (infinite) exact solutions, only one of which will be correct in the real world.

If the number of rows is equal to the number of columns (square matrix), then the number of spectra is equal to the dimension of their space, and (if linearly independent) they span that space. Any spectrum in that space (that number of wavelengths) can be expressed exactly as a linear combination of the columns, and that solution is unique. That unique solution is given by $C = E^{-1}A$, where E^{-1} is the standard inverse of the square matrix E.

If the number of rows (wavelengths) is greater than the number of columns (spectra), then the dimension of the vector space is greater than the number of spectra, so those spectra cannot span the space, and there will be vectors in the space that cannot be expressed as a linear combination of the basis spectra. However theoretically, since we assume that the analytes represented by the spectra are the only absorbing species in the solution, the spectrum of the solution should be a linear combination of the spectra, and so be within their span. Any deviation is due to noise, and a solution that minimiizes this deviation that we have to attribute to noise is given by $C = E^{-1}A$, where E^{-1} is the left generalized inverse of E, given by $=[E^TE]^{-1}E^T$.

This is a lot to remember, so going forward we can just summarize it as:
In order matrix M to have a left inverse, the columns of M must be independent, and the columns must be long (at least as long as the rows). (If the rows are longer than the columns, the columns cannot be independant, and $[E^TE\}$ will be singular and so not invertible) (/Earlier)

Aside- thinking about problem 2:
Suppose instead we know the concentrations (the spectra are "training spectra"), and we want to determine the standard spectra of the analytes. for a number of wavelengths equal to thenumber of analytes, the matrix C is square. Square matices the inverse C^{-1} functions as both a left inverse and a right inverse, so the solution is given by using the inverse on the right::
$A = EC$
$AC^{-1} = ECC^{-1} = E$
$E = AC^{-1}$.
We can do the same thing with more spectra different from the number of components, if a right generalized inverse of the non-square matrix C exists

The right generalized inverse exists, it is a matrix whose columns are orthonormal to the rows of C. Remeber the rows of C are the "time courses" at the different wavelengths. In case of training spectra, they are lists of the absorbance of each training spectrum at the given wavelength. If the training time coures span the vector space of the

in problem 1 the number of waewlengths in E could not be less than the number of compounds N, otherwise there could not be N independent spectra, and decomposition would be ambiguous. Each standard spectrum could be fitted as itself or as an LC of the others. so any LC of the standard spectra could be the same as an infinite number of other LCs.

so that they span the space,and any LC of the (unknown) standard spectra is within that span. We can find a LC of the spectra that fits any spectrum in the space. If the number of wavelegthswas less than the number of components, the spectra would be linearly dependant.N-1 could span the space and so fit the nth. each spectrum would be a linear combination of the other N-1 spectra. But the rank would not be zero, it would in general be N-1.

In problem 2 (We'll do it for just one wavelength first) we are trying to find a LC of the training time courses (at one wavelength) that matches the time course of each component. This means we need N independent time courses for the N different cmponents, and this can only exist if the number of time points (dimension of the space) is not less than the number of components

rank: if one spectrum can be expressed as a linear comb of all the others, the rank is zero? and any spectrum is a LC of the others. No you have to say each spectrum is LC of the others.
If two spectra are lin dep on each other but the rest are independent, the rand is one less than the number of spectra.

continue:
if we take the transpose of both sides we get $A^T = C^T E^T$
which can be solved as before with $C^{T-1}A^T = E^T$ providing C^T has an inverse marix C^{T-1} such that $C^{T-1}\ C^T = I$. As before this requires the rank of C' to be ~~greater than or~~ equal to the number of columns, in turn requiring the number of rows of C^T >= the nmber of columns.The rows of C^T (columns of C) correspond to the different time points, with the elements of each row giving the concentration at that time point. The columns correspond to the time courses at each wavelength, with the elements of each column giving the concentration of that analyte at each time point. Figure ? (prev chapter) illustrates this. The requirement thus amounts to the number of different samples (time points) being greater than the number of analytes contributing to the spectra.
(We would previously have discussed rank and square matrix inverse and generalized inverse of a non-square matrix; beers law in which the dataset A has columns (spectra) that are LC of the columns of E and rows (time courses) are LC of the rows of C)

multiplying a matrix by another on the right makes LC of the rows of the matrix. Thus multiplying C by E on the right as in beers law first form, each row of A (time course at that wavelength) is a linear combination of the rows of C (time course of the individual analytes.
(just as in the first problem multiplying E by C on the left was making linear combinations of the spectra (columns of E) that make up the spectra, and multiplying A by E^-1 on the right

While the number of time points needs to be greater or equal to the number of analytes, there is no such restriction on the number of wavelengths: Just as the first problem can be solved when A contains a single spectrum (column), the 2nd can be solved when A contains a single row (the timecourse at a single wavelength). Also no upper limit on the number of wavelengths, the problem can be solved at each wavelength

We can solve for the spectra directly from beers law without taking the transpose, if we introduce the concept of a left inverse matrix C^{-1} such that $C^{-1}C = I$. If such an inverse exists, then taking the transpose of the above gives $C^T [C^{-1}]^T = I' = I$

The transpose of the right inverse is the right inverse of the transpose of the matrix. So one way of finding the right inverse is to transpose the matrix, take its inverse, and transpose that. from the above considerations a right inverse will exist if the matrix has more columns than rows, i.e. the rows are longer than the columns

In problem 2 we are given a number of training spectra, and given the concentration of each of N analytes that are present in each of the solutions from which the spectra were obtained. From this we want to determine the standard spectra of each of the N components. Agin we will set this up using beer-lambert law inmatrix form:

A = EC

Now A is the matrix whose columns are the training spectra, E is the matrix whose columns are the (to be determined) standard spectra of the analytes, and the columns of C are the concentrations (given) of each analyte in the coresponding spectrum. Thus the number of columns of A and C are the number of training spectra. The number of rows in A is the number of wavelengths, and the number of rows in C is the number of analytes.

To simplify, we will consider just one wavelength at first, determining the absorbane of each analyte at that wavelength given the absorbtion at that wavelength for each of the solutions and the concentrations of each analyte in the solution. This can be repeated at every wavelength to obtain the spectra)

This means A is a row vector, the elements being the absorbance of each training solution at the given wavelength, E is a row vector, the elements being the (unknown) absorbance of each analyte at that wavelength. C is a full MxN matrix (M rows and N columns), N being the number of analytes and M being the number of spectra:

```
                                        [          |          ]
(.----------Ak----) = (-------Ej--------)  [---- C(j,k) ------  ]
                                        [          |          ]
```

where k goes from 1 to m, the number oftraining spectra, and J goes from 1 to N, the number of analytes.

Multiplying C by E on the right makes a linear combination of the rows of C. Since the columns of C are the list of concntrations of the different analytes in thee corresponding spectrum, the rows are the time course of the corresponding analyte in the series of spectra. So we are saying A, the time course of absorption at a wavelength, is a linear combination of the time-courses of the different analytes, being weighted by the extinctin coefficient of each analyte at that wavelengh. This is very siilar to problem 1 with a single spectrum which we want to express as a linear cmbination of the spectra of the individual analytes, weighted by the concentration of each analyte in the solution from which the spectrum was obtained.

In terms of vector spaces, A is a vector in M-dimensional time-course space, M being the number o f training spectra, and it is to be expressed as a LC of N idependent time courses of the N different analytes. (if they are not independant, one can be expressed as an LC of the others providing an infinite number of soulutions).

In order for th N time courses of the N analytes to be independent, the dimension M of the space must be no less than the number N of analytes. The number M of unknowns (the A_k) must be no less than the number N of parameters to be determined (the E_j). (It also depends on whoever was making the training solutions not making the them linearly dependant, for example one solution contains 1, 2, 3 units of the three components and another contains 2, 4, 6 units)

If M = N, C is a square matrix and the problemis immediately solved by multiplying both sides of the equation by C^{-1} on the right:

A = EC

$AC^{-1} = ECC^{-1} = E$

$E = AC^{-1}$

. This is almost exactly what we did in problem1 when E was square (number of wavelengths = nuber of analytes. The only difference is we were multiplying both sides by E^{-1} on the *left*:

$E^{-1}A = E^{-1}EC = C$

So can we ue a generalized inverse to solve the second problem when the number of training spectra is greater than the number of analytes? Yes, but . . .

When a matrix is square, its left inverse is the same as its right inverse:

$M^{-1}M = MM^{-1} = I$

Non-square matrices can be multiplied together only if the number of columns in the left matrix is equal to the number of rows in the right matrix. So given a non-square left generalized inverse M-1, such that $M^{-1}M = I$, that cannot be equal to MM^{-1}, because in general MM^{-1} doesn't even make sense. We can generate a right inverse in the same manner we generated the left inverse, by requirng the vector **R** of residuals in

$\mathbf{E} = \mathbf{AC}^{-1} + \mathbf{R}$

to be minimal, meaning that R must be orthogonal to the time course of every component (columns of E). But if we rephrae the problem by taking the transpose of each side, we can solve it with a left transpose, and if we transpose the solution we can get an equation for the right transpose if desired:

$A = EC$

$A^T = C^TE^T$. (trnspose of a product is product of transposes)

$A' = C'E'$. (where $A' = A^T$ etc) exactly like prob 1 with E,C reversed

then $C'^{-1}A' = C'^{-1}C'E' = E'$ where is a laft generalized inverse of C'

$E' = C'^{-1}A'$

$E'^T = A'^TC'^{-1T}$

$E^{TT} = A^{TT}C^{T-1T}$

$E = AC^{T-1T}$

Since we are solving the same problem when we solve the transpose of the problem, this shows that the right generalized inverse of a matrix is the transpose of the left generalized invere of the transpose of the matrix. for a matrix to have a left inverse, the number of rows must be no less than the number of columns (columns long). To have a right inverse, the number of columns must be no less than the number of rows (rows long). If the number of rows is the same as the number of columns (square matrix) a matrix can have both left and right inverses and they are the same.

We saw that the generalized left inverse is given by $[E^TE]^{-1}E^T$.

Substituting C^T for E and noting the transpose of the transpose is the original,

the left inverse of the transpose of C is

$[C^{TT}C^T]^{-1}C^{TT}$

$[CC^T]^{-1}C$

transposing that gives the right inverse of C:

$C^T[CC^T]^{-1T}$

CCT is diagonal, but is its inverse?

$M = M^T$

$M^{-1}M = I$

$M^TM^{-1T} = I^T = I$

$MM^{-1T} = I^T = I$

$M^{-1} = M^{-1T}$ (for square matrices where the inverse is defined)

MM^T is square, with dimension = first dimension (number of rows) of M

$[MM^T]$ to be invertible, its dimenion should be the smaller of the dimensions of M

so for MM^T the rows must be few, clumns short (rows long)

M^TM has dimension the second dimension (number of clumns) of M

so for M^TM the columns must be few, rows short, columns long.

The procedure is highly analogous to the procedure 1. Instead of finding a linear combination of the standard spectra that fits the unknown we find a linear combination of the "time course" of known concen

trations that fits the time course at each wavelength

$A_{ik} = E_{ij}C_{jk}$

(mult on right by C makes LC of the columns of matrix E. Col of A_{ik} are LC of colums of E_{ij})
(mult on left by E makes LC of the rows of matrix C. Col of A_{ik} are LC of colums of E_{ij})
(Note this could be done one time point (spectrum k) at a time, iwc A and C are col vectors
Or it could be done one wavelength (i) at a time, iwc A and E are row vectors

$[E^{-1}]_{ji}A_{ik} = C_{jk}$ This is our standard proced for fitting spectra, j>=k, k can be 1

$A_{ik} = E_{ij}C_{jk}$

mult on the left makes LC of the rows of matrix C, the timecourses of the individuasl analytes.
(If only one of the analytes absorbs at that wl,
suppose there is a matrix $[C^{-1}]_{k,j}$ that satifies $C_{j,k} *[C^{-1}]_{kj} = I_{j,j}$. Mult by it on left:

$A_{i,k}[C^{-1}]_{kj} = E_{i,j}[I]_{j,j} = E_{i,j}$

(Mult $A_{i,k}$ by $[C^{-1}]_{kj}$ on the left makes LC of the rows of A that are equal to the rows of E, TCs)
(note this could be done one WL at a time if A has only one row (one wavelength i, one TC)
(LC of the rows of
or, to make it even more like the first proc (and to allow use of the same softare), take the transpose of the first equation.

$A_{ik} = E_{ij} C_{jk}$

$A^T_{ki} = C^T_{kj} E^T_{ji}$ Mult on right makes LC of the cols of E^T (rows of E, TCs)

$[C^{T-1}]_{kj} A^T_{ki} = E^T_{ji}$ J>=i, i can be one

[C-1]kj satisfies: [C-1]kj * Cj,k = Ik,k
This means the every row of C-1 is orthogonal to every columns if Cj,k except the coresponding, and gives 1 with it
If J=K the C matrix is square and can be inverted by standard procedures to give C-1. If the number of spectra are greater than the number of analytes, (they cannot be fewer) we can use the same least-squares inversion procedur to find the basis spectra that est fit the test spectra given the known concentrations in terms of minimizing sum of squares of residuals.

For SVD, i and k are the same as above, j are replaced by h, the principal components.

$A_{i,k} = U_{i,h} [\sum V^T]_{h,k}$

The number of rows of V^t has to equal the number of columns of U, = h) (In order to dot columns of V into rows of U)
The number of columns of V has to equal the number of columns of U, = h
The rows of Vt (columns of V) are singTC, columns of U are eigenspectra
The columns of VT are the amount of each singular component at that time point

$A_{i,k} = U_{i,h} [\sum V^T]_{h,k}$

$E_{i,j} = A_{i,k}[C^{-1}]_{kj} = U_{i,h} [\sum V^T]_{h,k}[C^{-1}]_{kj}$

The first mult on right is finding LC of SingTCs that make each real component,
then multiply those coefficients by the spectra to make spectra coresponding to each real component
working with the transpose:

$A_{k,i} = [\sum V^T]_{kh} U_{hi}$

$E^T_{ji} = [C^{T-1}]_{kj} A^T_{ki} = [C^{T-1}]_{kj}[\sum V^T]_{kh} U_{hi}$

Apply C^{T-1} to the sing timecourses $\sum V^T$, make the corresponding LC of the sing spectra U_{hi}

Why do we fit model TC to LC of (eigenTC × eigenvalue) rather than just eigenTC? (This is a little weak because we don't actually measure eigentc by fitting spectra. but the values would be exactly the same if we did)

When fitting the TC to eigenTC, we use the TC weighted by singular value. Note this is singular, not eigen, value. If we use the unweighted eigenvectors, the last, which is very noisy, gets the same weight as the first.

The standard deviation of the determined amount of the eigenvector due to random error in the measurement depends on SDM. Since the eigenvectors are orthogonl, the probe spectrum for each is the eigenvector divided by eigenval, as in the single analyte case.

$|P| = |(\mathbf{B}^*/|B^*|^2)| = |B^*|/|B^*|^2 = 1/|B^*|$;where B is std spectrum and B* orth comp.
B* is the orthogonal component of the std spectrum, or TC The amount of each component in each spectrum (each point) could be determined by dotting the experimental time course into each col of V *or* dotting the correp experimental spectrum into each column of U: coeff = AU or V^TA
(1) Since the singular TC are orthogonal, B* = E . . .?
start over-
(2)Precision of the TC values depends on the SDM of the sing spectra, not TC.*:*
If the eigenTC were determined by fitting the spectra to a LC of the eigenspectra, the amount of each eigenTC at a given point (spectrum) would have an error determined by the SD of measurement and the SDM of the probe spectrum.
To regenerate the exptl spectra, we sum the eigenspectra times the value of the correponding TC at that point, times the singular value. Thus the standard spectrum for the eigenTC are the singular spectra times the correponding eigenvalues, |B*|=|E| = SingVal, SDM=1/SV. This makes the SDM smaller for the eigentc with larger SV. However if we are fitting the eigenTC multiplied by the corresp eval, the probe spectra are the eigen spectra themselves, which are normalized so all have norm 1, |P| = 1/|B| = 1. So if the std error of measurement is independent of wavelength and absorbance value, the std dev of the amount of each weighted TC is equal (in magnitude not percent) to that std deviation of absorbance measurement. Same units because the singular time courses are LC of the time courses at each wavelength, i.e absorbance measurements.

About left, right multiplication:
The columns of the left get dottted into the rows of the right matrix.
multiplication of a matrix by a vector on the right makes a LC of the columns
multiplication of a matrix by a vector on the left makes a LC of the rows
multiplication of a matrix by another on the right makes LCs of the columns
multiplication of a matrix by another on the left makes LCs of the rows
mult by a diagonal matrix on the tight multiplies each colum by the corresponding diagonal element

About inverses:
The columns of the left get dotted into the rows of the right matrix.
$A^{-1}A = I$ implies the rows of A^{-1} are co-orthonormal with cols of A
$AA^{-1} = I$ implies the cols of A^{-1} are co-orthonormal with rows of A
$[A^{-1}A]T = A^TA^{-1T} = I^T = I$
This implies that the inverse of the transpose is the transpose of the inverse. for square A?
Rows of a left inverse are co-orthonormal with the columns of the matrix
Columns of a right inverse are co-orthonormal with the rows of the matrix
$Y = AX$, $X = A^{-1}Y$, $A = XY^{-1}$
Left inverse A^{-1} is finding a linear comb of the columns of A that fit column(s) of Y, and putting the coefficients in X

Right inverse Y^{-1} is finding a linear comb of the rows of Y that fit X, and putting the coefficients in A

About transpose:

A^TA diagonal implies the columns of A make an orthogonal set

AA^T diagonal implies the rows of A make an orthogonal

The dimension of I cannot be greater than rank of A, because you couldnt have more vectors orthogonal to every vector in A other than corresponding.

So a nonsquare matrix interacts with its inverse along the long dimension.

$A_{mn}{}^{-1}A_{nm} = I$ implies $n \geq m$

$A_{nm}A^{-1}{}_{mn} = I$ implies $m \geq n$

If the columns are long (more rows than columns, n>m), it has a left inverse

If the rows are long (more col than rows, m>n) it has a right inverse.

In standard problem 1 E_{ij} has more wavelengths than analytes, i>j, so has a left inverse from the same equation, to solve problem 2, C_{jk} must have more spectra k than analytes j (why?), so has a right inverse.

our prog is set up to invert E with columns long, so we are calculating a left inverse

$X_{bf} = \{[M^TM]^{-1}M^T\}Y$

For n<m, $X_{bf} = Y\{M[MM^T]^{-1}\}$

Probe spectra form an inverse matrix to the standard spectra,

$P_j \bullet E_k = 0, \; j \neq k$ (i.e. the vectors P_j and E_k, dotted along wavelength i) [4]

$= 1, \; j = k$

This can be expressed in matrix form: let P be the matrix whose Jth row is the probe spectrum for the Jth analyte, and let E be the matrix whose Jth column is the Spectrum of the Jth analyte, and multiply P×E. By definition of matrix multiplication, the j,k^{th} element of the product is the dot product of the j^{th} row of P with the k^{th} column of E, so the roduct matrix is zere except where k = J, the diagonal elements, which will be 1. Such a matrix is called the identity matrix. In the case where E and P are square this defines P as the inverse of E, which is unique. So the rows of the inverse of E are the probe spectra for the correspondong standard spectra in E.

If E is not square, the product PE is still square, and P is called a left inverse of E (because it gives the identity matrix when multiplying E on the left). This inverse is not unique, as there are multiple spectra that would give give I when multiplying E on the right, But there is one out of these taht will result in the smallest SDMs when taken as probe spectra.

One of the suboptimal left inverses would be to truncate the spectra after the first M points, invert that, and fill out to N with zeros. The dot products will only involve the first M points, where the spectra are co-orthonormal. But the truncation limits the probe spectra to an M-dimensional subspace, and the portion of the standard spectra outside of this subspace would have contributed to increasing the norm2 divisor and decreasing SDM.

Appendix 1. Probabilistic implications of least-square fitting.

Assume measurement of absorbance at a single wavelength has an error distribution such that A = |A|+er
where |A| is the "population mean" or the actual value of A and er has a normal distribution about zero, so the measurement has a normal distribution about |A|.

The probability of measuring A' when the real absorbance is |A| is given by:
$P(A') \sim e^{-((A'-|A|)/\sigma)2}$

If we measure A at I different wavelengths, assuming the errors are independent, the probability of measuring a particular set of absorbance A'_1, A'_2, A'_3 . . . is just the product of the probabilities of each individual measurement:

$$P(A'_1, A'_2, A'_3 \ldots) \text{ given } |A_1|, |A_2|, |A_3| \ldots$$
$$\sim e^{-((A'-|A|)/\sigma)2} \times e^{-((A'-|A|)/\sigma)2} \times e^{-((A'-|A|)/\sigma)2} \times \ldots$$

multiplying exponentials corresponds to adding the exponents, so:

$$\sim e^{-((A'-|A|)/\sigma)2 +((A'-|A|)/\sigma)2 +((A'-|A|)/\sigma)2 + \ldots)}$$

This will be largest when the negative exponent is smallest, so to maximize probability we minimize the exponent:

$$(((A'-|A|)/\sigma)^2 +((A'-|A|)/\sigma)^2 +((A'-|A|)/\sigma)^2 + \ldots)$$

if σ is the same at each wavelength, it factors out and we minimize the sum of squared errors (least squares fitting):

$$1/\sigma \times ((A'-|A|)^2 +(A'-|A|)^2 +(A'-|A|)^2 + \ldots)$$

(if the σ are different, we can minimize the sum of squared (A'-|A|/σ), which is called weighted least squares fitting. In terms of fitting spectra, we just divide the absorbance and the extinction coefficients at each wavelength by an estimate of σ at that wavelength, which could be estimated by scanning the spectrum 5 times and calculating the standard deviation at each point. Transforming our I-dimensional absorbance space, into an I-dimensional weighted absorbance space)

Our model (lin comb of standard spectra) is saying what the real value of A is, which is |A|. So we adjust the parameters of our model (the concentrations) to minimize the sum of squares between caculated and observed values, thus finding the model that has the highest probability of predicting the observed result.

Appendix 2. Derivation of the formula for LSInverse matrix

The real observed spectrum is the sum of the ideal spectrum and the residuals:

A = predicted spectrum + R

R = A - predicted spectrum

R = A - EC (predicting A by [3])

where R is the spectrum of residuals and the C (concentrations) are the variables to vary in order to minimize |R|; ∑squares. And we want to make R orthogonal to each of the basis spectra, columns of E.

To solve, we set the dot product of R and each of the basis spectra to zero, giving I equations in I unknowns. A and EC are column vectors, so R is also. To make R orthogonal, we set its dot product with every basis spectrum (every column in E) to zero. Since there are I basis spectra, this gives us I equations in I unknowns (the so-called "normal equations") which we can in solve. Again, it is convenient to express it as a matrix equation. If we take the transpose of E the rows will be the standard spectra, and if we premultiply R this by we are taking the dot products of R with each row, with each spectrum. Set this to zero and solve for C:

$0 = [E^T] \times R$ where 0 is a column vector of all zeros

$0 = [E^T](A - EC)$

$0 = [E^T]A - [E^T E]C$

$[E^T]A = [E^T E]C$ solve by taking the inverse of $[E^T E]$

$[E^T E]^{-1} [E^T]A = C$

$\{[E^T E]^{-1}[E^T]\} A = C$

$E^{\perp} A = C$

where $E^{\perp}$ is the quantity in { }, $[E^T E]^{-1}[E^T]$, a J × I matrix called the generalized inverse of E. When it multiplies A to give C, each row gets dotted into A to give the C, the concentration of the corresponding substance. Thus each row is a probe spectrum for the corresponding substance, just as in matrix inversion. $E^{\perp}$ satisfies 4, with

$E^{\perp} E = [I]$

$\{[E^T E]^{-1}[E^T]\} E = [I]$

$[E^T E]^{-1}[E^T E] = [I]$ any matrix, times its inverse, = [I]

To derive a generalized inverse, not necc LS
want something like $E^{-1} E = I$
but E is not invertible becuse not square
multiply E by E^T to make it square before inverting:

$[E^T E]^{-1}[E^T E] = [I]$

$\{[E^T E]^{-1}[E^T]\} E = [I]$

$E^{\perp} E = [I]$

Rows of $[E^T]$ span the space of the M spectra
multiplying by $[E^T E]^{-1}$ on left gives a matrix that still spans

so we haven't lost any dimension
so the rows of $E^{\perp}$ must be the components of E orthogonal to all others, and not some other vector in the left null space?
Suppose instead of mutiplying by E^T before inverted, we multiplied by some arbitray matrix M with the dimension of E^T
$[ME]^{-1}[ME] = [I]$ provided ME is invertible
$\{[ME]^{-1}[M]\}E = [I]$

How do we know E^TE is invertible? Does it have the same rank as E?
The columns of the product are LC of the columns of E^T, and the rows are LC of the rows of E.

Appendix 3. Argument that adding more wavelengths will never degrade the result (in classic least squares or any method that can be described by "probe" spectra?)

In any chemometric process that works by multiplying the spectrum by a matrix (e.g. generalized inverse) the matrix can be seen as a collection of "probe" spectra, each of which is orthogonal to all but one of the spectra and gives 1 when dotted into that spectrum. This can be seen in the case of the LSI matrix from the fact that the LSI × matrix is diagonal unity. (So we can reverse the logic and use this as the definition of the probe spectra, and note that matrix iversion (simple or generalized) is a convenient way to generate probe spectra.)

To determine concentration of the j^{th} component in a solution, we take a spectrum and dot it into the probe spectrum. It is required to give a result that is proportional the the amount of component J and insensitive to the amount of any other substance considered. This implies that the probe spectrum is orthogonal to all components except component J. This sets specific requirements on the direction of the probe spectrum.

Since this direction was required to be orthogonal to the spectra of any contaminants, the result is proportional only to the concentration of this particular compound (A). And since dotting the standard, unit-concetration spectrum into the probe gives 1, the result with an arbitray spectrum is equal to the concentration of component A in the solution.

When the number of wavelengths are greater than the number of components, the probe spectrum is not unique- there is a subspace of the spectral space that is orthogonal to M-1 components, it has dimension of N - (M-1). For N=M it is 1-dimensional and the direction is uniquely determined. For n>M it is 2- or more-dimensioned, so multiple probe spectra could be found that satisfy the orthogonality condition.

Once the direction of the probe has been chosen, it must be standardized so that it reports 1 when dotted into the standard, unit-concentration spectrum E_j. First the vector is normalized to give a unit vector in the direction, then dotted into the standard spectrum E_j to give the length of the component of the standard spectrum in this direction, E'. In order for the probe to give 1 instead of this length when dotted into the standard spectrum, we divide the normalized spectrum by this length $\|E_j'\|$. It's length is now the reciprocal of the length of the component of the standard spectrum in this direction.

Or to avoid the ambiguity of the original spectrum's length, Lets start with the component of the standard spectrum along the direction of the proposed probe spectrum, call it E_j' as above. To get a unit length vector in this direction, divide by $\|E_j'\|$. The norm is now 1, so dotting Ej into it gives again the length of the component of E_j in this direction. To standardize it, divide again by $\|E_j'\|$. The norm is now $1/\|E_j'\|$, and dotting E_j into it gives 1. So the probe spectrum is given by $E_j'/\|E_j'\|^2$, where Ej' is the component of Ej along any vector orthogonal to all the other spectra.

What is the error? The concentation is calculated as the dot product of the probe spectrum with the experimental spectrum being analyzed, that is $C_j = \sum p_{ji} a_i$ where a_i are the absorbance at each wavelength point, and p_{ji} are the elements of the probe spectrum for compound J at the same wavelengths.

The expected values of the result is the correct answer, (assuming the standard spectra are correct) for any probe that satisfies the orthogonality conditions, since multiplying the standard spectra by the probe gives 1 for compound J and zero for the rest, and so multiplying any linear combination of the standard spectra will give the amount of J's spectrum in the linear combination. However the precision of the measurement, the sensitivity of the measurement to random error in the absorption measurements, depends on the choice of the probe, when the orthogonal sub space has more than one dimension. We can examine this by applying the rules of error propagation to the above equation:

Multiplying a_i by p_{ji} multiplies its standard deviation by the same factor p_{ji}, and adding all the different terms adds their variance, so the variance of measured Cj is $\sigma^2 = \sum (p_{ji}\sigma_i)^2$. where σ_i is the standard deviation of absorption measurement at the i^{th} wavelength. In the absence of other information, suppose σ is the same at every wavelegth. Then it can be factored out, and after taking the square root, σ for the calculated concetration is σ for the absorbance meaurement multiplied by $\text{sqrt}(\sum p_{ji})$, the norm of the probe spectrum for compound j. (As described above,) this has to be the inverse of the norm of thecomponent of the standard spectrum along the probe spectrum. the error will be inversely proportional to the norm of that component, E'. Thus of all the spectra that satisfy the orthogonality requirement, the one that will give the more precise result is the one along which the standard spectrum has the longest component.

All the potential probe spectra occupy a subspace of the output space which is orthogonal to the span of the other compound's spectra. The one that maximizes the component of the standard spectrum along it is the orthogonal component of the standard spectrum (express the standard spectrum as the sum of its component within the span of the other spectra and its comonent orthogonal to them, take the latter. Any other vector in the orthogonal space would be related to it as one leg of a right triangle of which it is the hypoteneue. (needs more justification)

Now suppose we have a matrix for analyzing the spectrum at a limited number of wavelengths. And the probe spectrum for each is the optimum, i.e. the orthogonal component of the standard spectrum divided by the norm^2 of that orthogonal component. i.e. a unit vector in the direction of the orthogonal component of A, divided by the norm of the component. (if n-m we have no choice). So the error in measurement will be the std deviation of absorbance divided by the norm of that component.

Then we are considering extending the wavelength range used to analyze. Now we need a probe spectrum that includes the additinal points. One way to make this wuld be to take the original probe spectrum and extend it with zeros at each of the new wavelengths. This will give exactly the same result as using the shorter range, with the same error (extending the spectrum with zeros doesn't affect the norm). This will still be orthogonal to all the other spectra in the new range.

But from what we said above the optimal probe spectrum will be the component of A orthogonal to B, C, D, etc; over the whole new wavelength range. Any other vector in the orthogonal space in the new range is shorter, being related as a leg to the hypoteneuse of the optimal spectrum. Therefore the optimal probe spectrum in the new range has norm greater than that in the shorter range, and the additional wavelengths improve the acuracy.

But note taking the optimal probe in the wider range and setting the extra wavelengths to zero will not give the same thing as taking the optimal probe in the sparse spectrum and extendign with zero. We can't compare the norms by saying one has zeros where the other is nonzero. They will be different in the region where both are nonzero. So we are just saying it is different from the best, so cannot be as good.

Taking the projection of a vector onto a subspace always makes it smaller (pythagoras) can we say both the colinear and orthogonal projections have shortened? But what i orthogonal has also changed. two lines that are perpendicular in 3 are not necessarily perpendicular in projection onto a plane.

Our experiment demonstrating the benefit of adding wavelengths is in ??

The above depends on the claim that the LSinverseprobe is the best. see also:
https://en.wikipedia.org/wiki/Gauss%E2%80%93Markov_theorem
Gauss-Morkov theorem says LLSQ solution hs the lowest sampling variance:
In statistics, the Gauss–Markov theorem states that the ordinary least squares (OLS) estimator has the lowest sampling variance within the class of linear unbiased estimators, if the errors in the linear regression model are uncorrelated, have equal variances and expectation value of zero.[1] The errors do not need to be normal, nor do they need to be independent and identically distributed (only uncorrelated with mean zero and homoscedastic with finite variance).

This also has a proof that the eigenvalues of the correlation matrix are positive.

comparison ofcalibration methods (simulated)
https://terpconnect.umd.edu/~toh/models/index.html
https://terpconnect.umd.edu/~toh/models/CLS.html

Lost in space(s):
transform from V to W, domain to co-domain
space in V that goes to 0 in W is the null space, or kernel.
Space in V that does not go to 0 in W is the row space. If the input vector is in the span of the rows, it will give a nonzero result when dotted into at least one of them during multiplication.
Space in W that can be reached by the transform is the column space, image, or range.
dimension of the range is the rank of the matrix.
Space that cannot be reached is the left null space

When we talk about the space orthogonal to all "other" vetors, w mean the space orthogonal to the column space of E after removing the column corresponding to the analyte in question, the left null space of that matrix.
E is a matrix operator that operates on vector C to give the spectrum of the mixture.
The column space is the subspace that can be reached by varying concentrations of all.
The column space of the decremented matrix is all spectra that could be obtined without that analyte
The left nullspace of the decremented matrix is the set of all spectra that could not be reached without it ???
The left null space of the decremented matrix within the column space of the full matrix is he set of all spectra that could serve as probe spectra.

If there are more rows than columns, the rank of the matrix is the number of columns (dimension of the domain) unless the spectra are linearly dependent.
If one spectrum is a multiple of another (or more generally a LC of the other spectra, then in every row the two elements will be in the same proportion. When one is eliminated by row reduction the other will be also.

In the absence of lin dependence, the null space is zero so the dimension of the column space is the dimensionof the domain.

What should we call the "unstripping" procedure?
A variant of a "reverse" procedure?

It seems surprising that you only need the concentration profile for the one analyte. You have all the variation in the training spectra being divided up between the concentration profiles of the different substances. These will not be orthogonal. So some of what you are attributing to the analyte should go to another component? Apparently not.

Multivariate curve resolution (wikipedia "chemometrics")

In chemometric parlance, multivariate curve resolution seeks to deconstruct data sets with limited or absent reference information and system knowledge. Some of the earliest work on these techniques was done by Lawton and Sylvestre in the early 1970s.[17][18] These approaches are also called self-modeling mixture analysis, blind source/signal separation, and spectral unmixing. For example, from a data set comprising fluorescence spectra from a series of samples each containing multiple fluorophores, multivariate curve resolution methods can be used to extract the fluorescence spectra of the individual fluorophores, along with their relative concentrations in each of the samples, essentially unmixing the total fluorescence spectrum into the contributions from the individual components. The problem is usually ill-determined due to rotational ambiguity (many possible solutions can equivalently represent the measured data), so the application of additional constraints is common, such as non-negativity, unimodality, or known interrelationships between the individual components (e.g., kinetic or mass-balance constraints).[19][20]

a

Product/reactant ratio from change in absorbance:

When ony one reaction is occurring,

$P/R = (A_0-A)/(A-A_{100})$,

where Ao, A, A_{100} coiuld be absorbance at a single wvelength, or the norm of the difference spectrum.

P + R = T = 100 (better to call it T)

$A = \varepsilon_R[R] + \varepsilon_P[P]$

$A_{100} = \varepsilon_P [T]$

$A_0 = \varepsilon_R [T]$

$A-A_0 = ([R]-T)\varepsilon_R + \varepsilon_P[P]$ $\quad [P] = T - [R]$

$A-A_0 = -[P]\varepsilon_R + \varepsilon_P[P]$

$A-A_0 = (\varepsilon_P - \varepsilon_R)[P]$ $\quad$ (($\varepsilon_P - \varepsilon_R$) is the difference extinction coefficient)

$A_{100} - A = -\varepsilon_R[R] + \varepsilon_P(T - [P])[R] = [T] - [P]$,

$A_{100} - A = -\varepsilon_R[R] + \varepsilon_P([R])$

$A_{100} - A = (\varepsilon_P-\varepsilon_R)[R]$

$(A-A_0)/(A_{100}-A) = (\varepsilon_P - \varepsilon_R)[P]/(\varepsilon_P-\varepsilon_R)[R] =[P]/[R]$

This holds whether A increases or decreases during the reaction. If it decreases, both $(A - A_0)$ and $(A_{100} - A)$ will be negative, the authors reversed the order to make them positive, but the ratio is the same.

or

$R = T - P$

$P/R = P/(T-P) = 1/(T/P - 1)$

$P/R = (T-R)/R = T/R - 1$

multivalued function of multple variables; vector-valued function of a vector.

Common curve-fitting programsfit a single-valued function of a single variable. but with multiple parameters.

Fitting spectra as a function of concentrations is a vector-vector problem, but it is linear so easily handled.

Nernst curve is a scalar function of a single variable (Eh) with multiple parameters,

Spectra as a function of EH is a vector-valued function of a single variable with multiple parameters.

[1] O. Warburg and W. Christian, Isolierung und Kristallisation des Garungsferments Enolase, Biochem. Zeitschrift 310 (1942) 384-421.

[2] D.I. Arnon, Copper Enzymes in Isolated Chloroplasts. Polyphenoloxidase in Beta Vulgaris, Plant Physiol 24 (1949) 1-15.

Do we need to do QR decomp to improve the "condition" of the solution?
What is the difference between condition number and SDM?
see wkipedia "QR decomposition" and "condition number"

pdf/specralanalysis/chemometrics.pdf has references showing "reverse" methods are superior to "classical" methods. Better check them out.

Symbols: