4.3.3 Appendix: Chemical Concentrations

4.3.3.1 Geometry of the Manifold

There are different definitions of the *concentration* in chemistry. For instance, when one consider the mass concentration of a product i in a mixing of n products, one defines

$$c^i = \frac{\text{mass of } i}{\text{total mass}} \quad , \tag{4.43}$$

and one has the constraint

$$\sum_{i=1}^{n} c^i = 1 \quad , \tag{4.44}$$

the range of variation of the concentration being

$$0 \leq c^i \leq 1 \quad . \tag{4.45}$$

To have a Jeffreys quantity (that should have a range of variation between zero and infinity) we can introduce the *eigenconcentration*

$$K^i = \frac{\text{mass of } i}{\text{mass of not } i} \quad . \tag{4.46}$$

Then,

$$0 \leq K^i \leq \infty \quad . \tag{4.47}$$

The inverse parameter $1/K^i$ having an obvious meaning, we clearly now face a Jeffreys quantity. The relations between concentration and eigenconcentration are easy to obtain:

$$K^i = \frac{c^i}{1 - c^i} \quad ; \qquad c^i = \frac{K^i}{1 + K^i} \quad . \tag{4.48}$$

The constraint in equation 4.44 now becomes

$$\sum_{i=1}^{n} \frac{K^i}{1 + K^i} = 1 \quad . \tag{4.49}$$

From the Jeffreys quantities K^i we can introduce the *logarithmic eigenconcentrations*

$$k^i = \log K^i \quad , \tag{4.50}$$

that are Cartesian quantities, with the range of variation

$$-\infty \leq k^i \leq +\infty \quad , \tag{4.51}$$

subjected to the constraint

$$\sum_{i=1}^{n} \frac{e^{k^i}}{1 + e^{k^i}} = 1 \quad . \tag{4.52}$$

Should we not have the constraint expressed by the equations 4.44, 4.49 and 4.52, we would face an n-dimensional manifold, with different choices of coordinates, the coordinates $\{c^i\}$, the coordinates $\{K^i\}$, or the coordinates $\{k^i\}$. As the quantities k^i, logarithm of the Jeffreys quantities K^i, play the role of Cartesian coordinates, the distance between a point k_a^i and a point k_b^i is

$$D = \sqrt{\sum_{i=1}^{n} (k_b^i - k_a^i)^2} \quad . \tag{4.53}$$

Replacing here the different definition of the different quantities, we can express the distance by any of the three expressions

$$D_n = \sqrt{\sum_{i=1}^{n} \left(\log \frac{c_b^i\,(1-c_a^i)}{c_a^i\,(1-c_b^i)} \right)^2} = \sqrt{\sum_{i=1}^{n} \left(\log \frac{K_b^i}{K_a^i} \right)^2} = \sqrt{\sum_{i=1}^{n} (k_b^i - k_a^i)^2} \ . \tag{4.54}$$

The associated distance elements are easy to obtain (by direct differentiation):

$$ds_n^2 = \sum_{i=1}^{n} \left(\frac{dc^i}{c^i\,(1-c^i)} \right)^2 = \sum_{i=1}^{n} \left(\frac{dK^i}{K^i} \right)^2 = \sum_{i=1}^{n} (dk^i)^2 \quad . \tag{4.55}$$

To express the volume element of the manifold in these different coordinates we just need to evaluate the metric determinant $\sqrt{\mathbf{g}}$, to easily obtain

$$dv_n = \frac{dc^1}{c^1\,(1-c^1)}\,\frac{dc^2}{c^2\,(1-c^2)}\,\cdots = \frac{dK^1}{K^1}\,\frac{dK^2}{K^2}\,\cdots = dk^1\,dk^2\,\cdots \quad . \tag{4.56}$$

In reality, we do not work in this n-dimensional manifold. As we have n quantities and one constraint (that expressed by the equations 4.44, 4.49 and 4.52), we face a manifold with dimension $n-1$. While the n-dimensional manifold can se seen as an Euclidean manifold (that accepts the Cartesian coordinates $\{k^i\}$), this $(n-1)$-dimensional manifold is not Euclidean, as the constraint 4.52 is not a linear constraint in the Cartesian coordinates. Of course, under the form 4.44 the constraint is formally linear, but the coordinates $\{c^i\}$ are not Cartesian.

The metric over the $(n-1)$-dimensional manifold is that induced by the metric over the n-dimensional manifold. It is easy to evaluate this induced metric, and we are going to use now of the possible methods.

Because the simplicity of the metric may be obscured when addressing the general case, let us make the derivation when we have only three chemical elements, i.e., when $n = 3$. From this special case, the general formulas for the n-dimensional will be easy to write. Also, in what follows, let us consider only the quantities c^i (the ordinary concentrations), leaving as an exercise for the reader to obtain equivalent results for the eigenconcentrations K^i or the eigenconcentrations k^i.

When we have only three chemical elements, the constraint in equation 4.44, becomes, explicitly,

$$c^1 + c^2 + c^3 \;=\; 1 \quad , \tag{4.57}$$

and the distance element (equation 4.55) becomes

$$ds_3^2 \;=\; \left(\frac{dc^1}{c^1\,(1-c^1)}\right)^2 + \left(\frac{dc^2}{c^2\,(1-c^2)}\right)^2 + \left(\frac{dc^3}{c^3\,(1-c^3)}\right)^2 \quad . \tag{4.58}$$

As coordinates over the two-dimensional manifold defined by the constraint, let us arbitrarily choose the first two coordinates $\{c^1, c^2\}$, dropping c^3. Differentiating the constraint 4.57 gives $dc^3 = -dc^1 - dc^2$, expression that we can insert in 4.58, to obtain the following expressions for the distance element over the two-dimensional manifold:

$$ds_2^2 \;=\; \left(\frac{1}{Q^1} + \frac{1}{Q^3}\right)(dc^1)^2 + \left(\frac{1}{Q^2} + \frac{1}{Q^3}\right)(dc^2)^2 + \frac{2\,dc^1\,dc^2}{Q^3} \quad , \tag{4.59}$$

where

$$\begin{aligned} Q^1 &= (c^1)^2\,(1-c^1)^2 \\ Q^2 &= (c^2)^2\,(1-c^2)^2 \\ Q^3 &= (c^3)^2\,(1-c^3)^2 \quad , \end{aligned} \tag{4.60}$$

and where $c^3 = 1 - c^1 - c^2$. From this expression we evaluate the metric determinant $\sqrt{\mathbf{g}}$, to obtain the volume element (here, in fact, surface element):

$$dv_2 \;=\; \frac{\sqrt{1 + (Q^1 + Q^2)/Q^3}}{\sqrt{Q^1\,Q^2}}\, dc^1\, dc^2 \quad . \tag{4.61}$$

This volume density (in fact, surface density) is represented in figure 4.10.

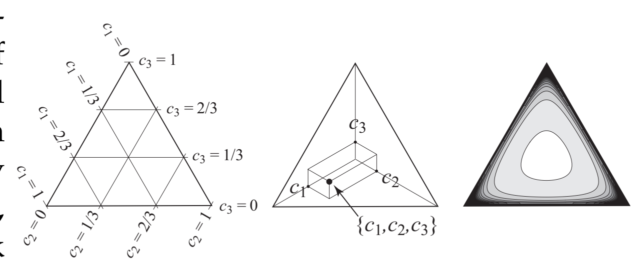

Figure 4.10: Top left, when one has three quantities $\{c^2, c^2, c^3\}$ related by the constraint $c^1 + c^2 + c^3 = 1$ one may use any of the two equivalent representations the usual one (left) or a 'cube corner' representation (middle). At the right, the volume density $\sqrt{\mathbf{g}}$, as expressed by equation 4.61 (here, in fact, we have a 'surface density'). Dark grays correspond to large values of the volume density.

4.3.3.2 A Simple Volumetric Probability

In every kind of problem it is convenient to have some simple probability distributions, that can be used as an approximation for experimental histograms. Let us obtain the simplest probability model for concentrations.

We can start working with the logarithmic eigenconcentrations. As these are Cartesian quantities, we can consider a Gaussian distribution,

$$f_3 \;=\; \frac{1}{(2\pi)^{3/2}\,\sqrt{\sigma^1\sigma^2\sigma^3}}\;\exp\Big[-\frac{1}{2}\sum_{i=1}^{3}\frac{(k^i-k_0^i)^2}{(\sigma^i)^2}\Big] \quad . \tag{4.62}$$

It is centered at point $\{k_0^1, k_0^2, k_0^3\}$, and the standard deviations are[5] σ^1, σ^2, and σ^3.

As a volumetric probability is an invariant quantity[6], to express the volumetric probability in the variables $\{K^1, K^2, K^3\}$ we only need to replace the variables. This gives

$$f_3 \;=\; \frac{1}{(2\pi)^{3/2}\,\sqrt{\sigma^1\sigma^2\sigma^3}}\;\exp\Big[-\frac{1}{2}\sum_{i=1}^{3}\Big(\frac{1}{\sigma^i}\log\frac{K^i}{K_0^i}\Big)^2\Big] \quad . \tag{4.63}$$

This is clearly a three-dimensional lognormal volumetric probability. It is centered at point $\{K_0^1, K_0^2, K_0^3\}$, and the parameters σ^1, σ^2, and σ^3 keep their interpretation of standard deviations of the logarithmic variables.

In terms of the ordinary concentrations c^i the volumetric probability is

$$f_3(c^1,c^2,c^3) \;=\; \frac{1}{(2\pi)^{3/2}\,\sqrt{\sigma^1\sigma^2\sigma^3}}\;\exp\Big[-\frac{1}{2}\sum_{i=1}^{3}\Big(\frac{1}{\sigma^i}\log\frac{c^i\,(1-c_0^i)}{c_0^i\,(1-c^i)}\Big)^2\Big] \quad . \tag{4.64}$$

It is centered at point $\{c_0^1, c_0^2, c_0^3\}$, and the parameters σ^1, σ^2, and σ^3 keep their original interpretation.

So far, the volumetric probability f is defined over the three-dimensional manifold, and the constraint $c^1 + c^2 + c^3 = 1$ has not been introduced. Introducing it is clearly a matter of conditional volumetric probability. Formula 2.66 (page 40) tells us that the conditional volumetric probability is obtained, excepted for a normalizing factor, by just inputing the constraint in the original distribution. This gives

$$f_2(c^1,c^2) \;=\; \text{const.}\;\exp\Big[-\frac{1}{2}\sum_{i=1}^{3}\Big(\frac{1}{\sigma^1}\log\frac{c^1\,(1-c_0^1)}{c_0^1\,(1-c^1)}\Big)^2\Big] \quad , \tag{4.65}$$

where $c^3 = 1 - c^1 - c^2$. Typical shapes of the volumetric probability are displayed in figure 4.11.

---

[5] We could, of course, consider a general Gaussian, with nonvanishing covariances.

[6] This would not be the case if we were using a probability density. See chapter 2 for details.

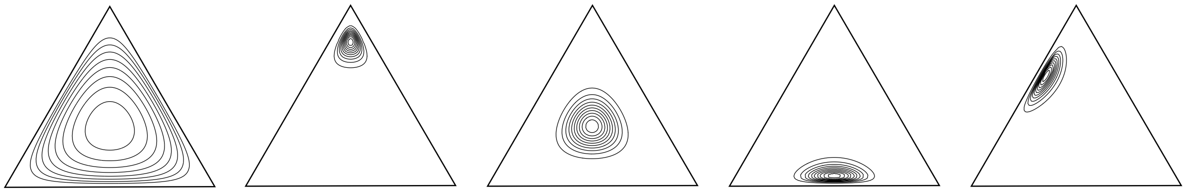

Figure 4.11: Different instances of the proposed probability model, derived from a Gaussian (the maximum values of the volumetric probability are in the middle of the 'circles'). At the left, the probability distribution is centered at point $c^1 = c^2 = c^3 = 1/3$, and the standard deviations are large ($\sigma^1 = \sigma^2 = \sigma^3 = 2$). In the other four cases, the standard deviations have always the values $\sigma^1 = \sigma^2 = \sigma^3 = 0.5$, and are respectively centered at the points $\{0.1, 0.1, 0.8\}$, $\{1/3, 1/3, /1/3\}$, $\{0.47, 0.47, 0.16\}$, $\{0.333, 0.050, 0.627\}$. In order for the five volumetric probabilities displayed here to be normed to one, the constant appearing in equation 4.65 has been evaluated by numerical integration, and has the values 0.12, 0.11, 0.60, 2.80, and 0.41 respectively.

To compute the finite probability of a domain $\mathcal{D}$, of the 'triangle' $\{c^1, c^2\}$ we must, of course, use the volume element obtained above (equation 4.61). Explicitly,

$$P(\mathcal{D}) \;=\; \int_{\mathcal{D}} dc^1\, dc^2\, \frac{\sqrt{1 + (Q^1 + Q^2)/Q^3}}{\sqrt{Q^1\, Q^2}}\, f_2(c^1, c^2) \quad , \tag{4.66}$$

where the intermediary variables Q^i were introduced in equations 4.60.

As a final comment, should one work with the more common probability densities (instead of volumetric probabilities), the proposed model would correspond to the probability density

$$\overline{f}_2(c^1, c^2) \;=\; \frac{\sqrt{1 + (Q^1 + Q^2)/Q^3}}{\sqrt{Q^1\, Q^2}}\, f_2(c^1, c^2) \quad , \tag{4.67}$$

and the finite probability of a domain $\mathcal{D}$ would be computed as

$$P(\mathcal{D}) \;=\; \int_{\mathcal{D}} dc^1\, dc^2\, \overline{f}_2(c^1, c^2) \quad . \tag{4.68}$$

4.3.3.3 From 3 to n Dimensions

Given n variables c^i such that $\sum_{i=1}^n c^i = 1$, given n constants c_0^i such that $0 \leq c_0^i \leq 1$, and given n positive constants σ^i, one selects $n-1$ of the n variables c^i, to define the $(n-1)$-dimensional volumetric probability (obvious generalization of expression 4.65)

$$f(c^1, \dots, c^{n-1}) \;=\; \text{const.}\ \exp\Big[-\frac{1}{2} \sum_{i=1}^{n} \Big(\frac{1}{\sigma^i} \log \frac{c^i\,(1 - c_0^i)}{c_0^i\,(1 - c^i)} \Big)^2 \Big] \quad . \tag{4.69}$$

where the c^n appearing in the last term of the sum is not one of the variables of the volumetric probability, but is the value

$$c^n = 1 - \sum_{i=1}^{n-1} c^i \quad . \tag{4.70}$$

The constant factor in equation 4.69 can always be chosen[7] such that the total probability is normed to one[8]:

$$\underbrace{\int_{\mathcal{D}} dc^1 \, \dots \, dc^{n-1}}_{\sum_{i=1}^{n-1} c^i \leq 1} \; \frac{\sqrt{1 + \sum_{i=1}^{n-1} Q^i / Q^n}}{\sqrt{\prod_{i=1}^{n-1} Q^i}} \; f(c^1, \dots, c^{n-1}) \quad , \tag{4.71}$$

To compute the probability of a domain $\mathcal{D}$, inside the simplex defined by the conditions $\sum_{i=1}^{n-1} c^i \leq 1$, we use the expression

$$P(\mathcal{D}) \;=\; \int_{\mathcal{D}} dc^1 \, \dots \, dc^{n-1} \; \frac{\sqrt{1 + \sum_{i=1}^{n-1} Q^i / Q^n}}{\sqrt{\prod_{i=1}^{n-1} Q^i}} \; f(c^1, \dots, c^{n-1}) \quad , \tag{4.72}$$

where the intermediary variables Q^i are (see equations 4.60)

$$Q^i \;=\; (c^i)^2 \, (1 - c^i)^2 \quad , \tag{4.73}$$

the quantity Q^n being computed using the value of c^n given in equation 4.70.

The interpretation of the constants c^i_0 and σ^i is as follows. One may pass from the quantities c^i to the quantities

$$k^i \;=\; \log \frac{c^i}{1 - c^1} \quad . \tag{4.74}$$

In the n-dimensional (flat) manifold where the k^i are Cartesian quantities, one has an n-dimensional Gaussian distribution, centered at the point whose coordinates are

$$k^i \;=\; \log \frac{c^i_0}{1 - c^1_0} \quad , \tag{4.75}$$

and with standard deviations that are the σ^i. By construction, the $(n-1)$-dimensional volumetric probability in equation 4.69 is the 'intersection' of this Gaussian by the (non flat) hypersurface defined by the constraint $\sum_{i=1}^{n} c^i = 1$.

[7] I have not been able to analytically evaluate the expression of the constant that would ensure that the volumetric probability is automatically normed to one. This does not prevent practical use, as the constant can always be easily evaluated by numerical integration (using standard mathematical software). For instance, all the volumetric probabilities appearing in figure 4.11 were normed so the integrals equal one (although the plotting software has a renormalization of its own).

[8] The integral can, for instance, be written $\int_0^1 dc^1 \int_0^{1-c^1} dc^2 \dots \int_0^{1-c^1-c^2-\dots-c^{n-2}} dc^{n-1}$.

4.3.3.4 Dirichlet Distribution

One often finds in the literature (e.g., DeGroot, 1970; Wilks, 1962) a probability distribution, also defined for a set of variables that sum to one, whose 'shape' is quite similar to that of the probability distribution f we have just obtained (i.e., the shape that has been represented, in the 3D case, in figure 4.11).

As in usual texts the notion of volume of the manifold is not considered, they can only consider probability densities, not volumetric probabilities. Then, given n variables c^i such that $\sum_{i=1}^n c^i = 1$, one selects $n-1$ of these variables, and given n positive constants α^i , to define the *Dirichlet probability density*

$$D(c^1, \dots, c^{n-1}) \;=\; \frac{\Gamma(\sum_{i=1}^n \alpha^i)}{\prod_{i-1}^n \Gamma(\alpha^i)} \prod_{i=1}^n (c^i)^{\alpha^i - 1} \quad , \tag{4.76}$$

where the c^n appearing in the last term of the product is not one of the variables of the probability density, but is the value

$$c^n = 1 - \sum_{i=1}^{n-1} c^i \quad . \tag{4.77}$$

This probability density is normalized by the condition (see footnote 8)

$$\underbrace{\int dc^1 \dots dc^{n-1}}_{\sum_{i=1}^{n-1} c^i \leq 1} D(c^1, \dots, c^{n-1}) \;=\; 1 \quad . \tag{4.78}$$

In this text, we refrain from defining mean and standard deviation unless the manifold is metric. But let us do, for the Dirichlet probability density, what normal texts do. The mean of the variable c^i is defined as (and one finds the result)

$$\underbrace{\int dc^1 \dots dc^{n-1}}_{\sum_{i=1}^{n-1} c^i \leq 1} c^i\, D(c^1, \dots, c^{n-1}) \;=\; \frac{\alpha^i}{A} \quad , \tag{4.79}$$

where $A = \sum_{i=1}^n \alpha^i$. The variance of the variable c^i is defined as (and one finds the result)

$$\underbrace{\int dc^1 \dots dc^{n-1}}_{\sum_{i=1}^{n-1} c^i \leq 1} \left(c^i - \frac{\alpha^i}{A} \right)^2 D(c^1, \dots, c^{n-1}) \;=\; \frac{\alpha^i\,(A - \alpha^i)}{A^2\,(A+1)} \quad . \tag{4.80}$$

In these two equations, we must understand that, when applying them for $i = n$, we must replace c^n by $1 - \sum_{i=1}^{n-1} c^i$ in the integrand. Note that we do not have the freedom to independently fix the mean values and the variances of the variables.