1 Background

Basis abundances:

$$\boldsymbol{w} = (w_1, \ldots, w_D) \in \mathscr{R}_+^D \tag{1}$$

Composition:

$$\boldsymbol{x} = (x_1, \ldots, x_D) = \mathscr{C}(\boldsymbol{w}) = \boldsymbol{w}/(w_1 + \cdots + w_D) \in \mathscr{S}^d \tag{2}$$

Centered log-ratios:

$$\begin{aligned}
\boldsymbol{z} = \operatorname{clr}(\boldsymbol{x}) = \log(\boldsymbol{x}/g(\boldsymbol{x})) = \log(\boldsymbol{w}/g(\boldsymbol{w})),\ g(\boldsymbol{x}) &= \left(\prod_i x_i\right)^{1/D} && (3) \\
&= \left(\log x_1 - \frac{1}{D}\sum_i \log x_i, \ldots, \log x_D - \frac{1}{D}\sum_i \log x_i\right) && (4) \\
&= \left(\log w_1 - \frac{1}{D}\sum_i \log w_i, \ldots, \log w_D - \frac{1}{D}\sum_i \log w_i\right) && (5)
\end{aligned}$$

Log-ratio variances:

$$\tau_{ij} = \operatorname{var}\log(x_i/x_j) \tag{6}$$

Basis covariances:

$$\operatorname{var}(\log \boldsymbol{w}) = \boldsymbol{\Omega} \tag{7}$$

Centered log-ratio covariances:

$$\operatorname{var}(\operatorname{clr}\boldsymbol{x}) = \operatorname{var}(\operatorname{clr}\boldsymbol{w}) = \boldsymbol{\Gamma} \tag{8}$$

Variation matrix:

$$[\tau_{ij}] = [\operatorname{var}(\log(x_i/x_j))] = \boldsymbol{T} \tag{9}$$

$\boldsymbol{\Omega} \to \boldsymbol{\Gamma}$:

$$\begin{aligned}
\gamma_{ij} &= \omega_{ij} - \omega_{i\cdot} - \omega_{j\cdot} + \omega_{\cdot\cdot} && (10) \\
\boldsymbol{\Gamma} &= \boldsymbol{G}\boldsymbol{\Omega}\boldsymbol{G},\ \boldsymbol{G} = \boldsymbol{I} - D^{-1}\boldsymbol{J},\ \boldsymbol{J} = [1]_{D\times D} && (11)
\end{aligned}$$

$\mathbf{\Omega} \to \boldsymbol{T}$:

$$\tau_{ij} = \omega_{ii} + \omega_{jj} - 2\omega_{ij} \tag{12}$$
$$\boldsymbol{T} = \boldsymbol{J}\,\mathrm{diag}(\mathbf{\Omega}) + \mathrm{diag}(\mathbf{\Omega})\boldsymbol{J} - 2\mathbf{\Omega} \tag{13}$$

$\mathbf{\Gamma} \leftrightarrow \boldsymbol{T}$:

$$\tau_{ij} = \gamma_{ii} + \gamma_{jj} - 2\gamma_{ij} \tag{14}$$
$$\boldsymbol{T} = \boldsymbol{J}\,\mathrm{diag}(\mathbf{\Gamma}) + \mathrm{diag}(\mathbf{\Gamma})\boldsymbol{J} - 2\mathbf{\Gamma} \tag{15}$$
$$\gamma_{ij} = \frac{1}{2}(\tau_{i\cdot} + \tau_{j\cdot} - \tau_{ij} - \tau_{\cdot\cdot}) \tag{16}$$
$$\mathbf{\Gamma} = -\frac{1}{2}\boldsymbol{GTG} \tag{17}$$

2 Findings

$\mathbf{\Gamma} \to \mathbf{\Omega}$ parameterized by $\omega_{11}, \ldots, \omega_{DD}$:

$$\omega_{ij} = \frac{1}{2}\left(\omega_{ii} + \omega_{jj} - \tau_{ij}\right) \tag{18}$$
$$= \frac{1}{2}\left(\omega_{ii} + \omega_{jj} - \gamma_{ii} - \gamma_{jj} + 2\gamma_{ij}\right) \tag{19}$$
$$= \gamma_{ij} + \frac{1}{2}\left(\omega_{ii} - \gamma_{ii} + \omega_{jj} - \gamma_{jj}\right) \tag{20}$$
$$\mathbf{\Omega} = \frac{1}{2}\left(\boldsymbol{J}\,\mathrm{diag}(\mathbf{\Omega}) + \mathrm{diag}(\mathbf{\Omega})\boldsymbol{J} - \boldsymbol{T}\right) \tag{21}$$
$$= \frac{1}{2}\left(\boldsymbol{J}\,\mathrm{diag}(\mathbf{\Omega}) + \mathrm{diag}(\mathbf{\Omega})\boldsymbol{J} - \boldsymbol{J}\,\mathrm{diag}(\mathbf{\Gamma}) - \mathrm{diag}(\mathbf{\Gamma})\boldsymbol{J} + 2\mathbf{\Gamma}\right) \tag{22}$$
$$= \mathbf{\Gamma} + \frac{1}{2}\left[\boldsymbol{J}\left(\mathrm{diag}(\mathbf{\Omega}) - \mathrm{diag}(\mathbf{\Gamma})\right) + \left(\mathrm{diag}(\mathbf{\Omega}) - \mathrm{diag}(\mathbf{\Gamma})\right)\boldsymbol{J}\right] \tag{23}$$

Notation for set of potential basis covariances associated with a given clr covariance matrix:

$$\mathscr{B}(\mathbf{\Gamma}) = \{\mathbf{\Omega} : \mathbf{\Omega} \text{ is symmetric positive semi-definite and } \boldsymbol{G}\mathbf{\Omega}\boldsymbol{G} = \mathbf{\Gamma}\} \tag{24}$$

$\mathbf{\Gamma}$ has minimal total variance among all $\mathbf{\Omega}$s which have corresponding clr covariances $\mathbf{\Gamma}$:

1. Given a clr covariance matrix $\mathbf{\Gamma}$, suppose there exists $\mathbf{\Omega} \in \mathscr{B}(\mathbf{\Gamma})$ such that $\mathrm{tr}(\mathbf{\Omega}) < \mathrm{tr}(\mathbf{\Gamma})$.

2. For $j = (1, \ldots, 1)^\intercal$,

$$
\begin{aligned}
j^\intercal \mathbf{\Omega} j &= \sum_i \sum_j \omega_{ij} && (25)\\
&= \sum_i \sum_j (\gamma_{ij} + \frac{1}{2}[\omega_{ii} - \gamma_{ii} + \omega_{jj} - \gamma_{jj}]) && (26)\\
&= \sum_i \left\{ 0 + \frac{1}{2} \left[p\omega_{ii} - p\gamma_{ii} + \mathrm{tr}(\mathbf{\Omega}) - \mathrm{tr}(\mathbf{\Gamma}) \right] \right\} && (27)\\
&= \frac{1}{2} \left\{ p\,\mathrm{tr}(\mathbf{\Omega}) - p\,\mathrm{tr}(\mathbf{\Gamma}) + p\,\mathrm{tr}(\mathbf{\Omega}) - p\,\mathrm{tr}(\mathbf{\Gamma}) \right\} && (28)\\
&= \mathrm{tr}(\mathbf{\Omega}) - \mathrm{tr}(\mathbf{\Gamma}) && (29)\\
&< 0 && (30)
\end{aligned}
$$

3. Thus, $\mathrm{tr}(\mathbf{\Omega}) < \mathrm{tr}(\mathbf{\Gamma})$ implies that $\mathbf{\Omega}$ is not positive semi-definite, which contradicts the premise that $\mathbf{\Omega} \in \mathscr{B}(\mathbf{\Gamma})$.

4. Therefore, it must be that for all $\mathbf{\Omega} \in \mathscr{B}(\mathbf{\Gamma})$, $\mathrm{tr}(\mathbf{\Omega}) \geq \mathrm{tr}(\mathbf{\Gamma})$.

For $D = 2$, it is possible to describe $\mathscr{B}(\mathbf{\Gamma})$ exactly. We can see which basis correlations and partial correlations are possible for a given $\mathbf{\Gamma}$, and visualize the structure of $\mathscr{B}(\mathbf{\Gamma})$.

1. For $D = 2$, $\mathbf{\Gamma}$ is completely determined by one entry because $\gamma_{11} = -\gamma_{12} = -\gamma_{21} = \gamma_{22}$. And $\gamma_{11} = \omega_{11} - \omega_{i\cdot} - \omega_{j\cdot} + \omega_{\cdot\cdot} = -\omega_{12} + \frac{1}{4}(\omega_{11} + 2\omega_{12} + \omega_{22}) = \frac{1}{4}(\omega_{11} - 2\omega_{12} + \omega_{22}) = \mathrm{var}[\frac{1}{2}\log(w_1/w_2)] = \frac{1}{4}\tau_{12}$.

2. Positive semi-definiteness of $\mathbf{\Omega}$ requires that for any $\boldsymbol{x} \in \mathscr{R}^2$, $\boldsymbol{x} \neq \mathbf{0}$, $\omega_{11}x_1^2 + 2\omega_{12}x_1x_2 + \omega_{22}x_2^2 \geq 0$. First, if $x_1 = 0$ then $\omega_{22} \geq 0$, and if $x_2 = 0$ then $\omega_{11} \geq 0$. Let $x_2 \neq 0$ be fixed. Then we need the quadratic expression $\omega_{11}x_1^2 + 2\omega_{12}x_2x_1 + \omega_{22}x_2^2 \geq 0$ for all x_1. That requires that the determinant $4\omega_{12}^2x_2^2 - 4\omega_{11}\omega_{22}x_2^2 = 4x_2^2(\omega_{12}^2 - \omega_{11}\omega_{22}) \leq 0$ which requires that $\omega_{12}^2 - \omega_{11}\omega_{22} \leq 0$.

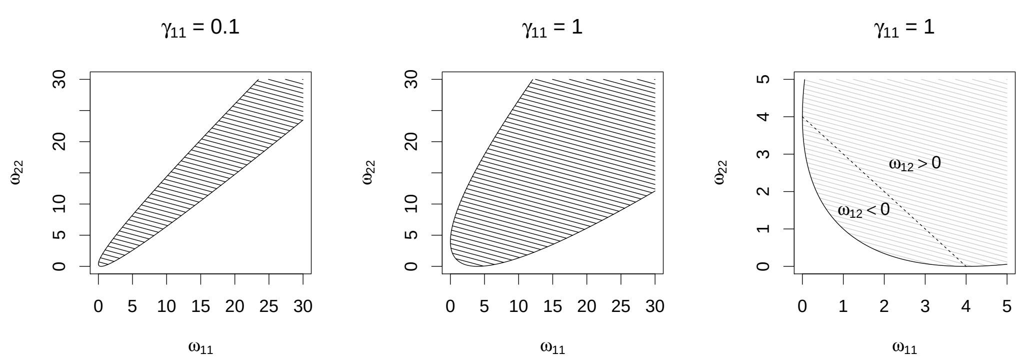
3. $\omega_{12} = \gamma_{12} + \frac{1}{2}(\omega_{11} - \gamma_{11} + \omega_{22} - \gamma_{22}) = -\gamma_{11} + \frac{1}{2}\omega_{11} - \frac{1}{2}\gamma_{11} + \frac{1}{2}\omega_{22} - \frac{1}{2}\gamma_{11} = \frac{1}{2}(\omega_{11} + \omega_{22}) - 2\gamma_{11}$.

4. $\omega_{12}^2 - \omega_{11}\omega_{22} = \frac{1}{4}\omega_{22}^2 - (\frac{1}{2}\omega_{11} + 2\gamma_{11})\omega_{22} + (\frac{1}{4}\omega_{11}^2 - 2\gamma_{11}\omega_{11} + 4\gamma_{11}^2) = 0$ when $\omega_{22} = \left[\frac{1}{2}\omega_{11} + 2\gamma_{11} \pm 2\sqrt{\gamma_{11}\omega_{11}}\right]/\frac{1}{2} = (\sqrt{\omega_{11}} \pm 2\sqrt{\gamma_{11}})^2$.

5. $\omega_{12}^2 - \omega_{11}\omega_{22} \leq 0$ when $(\sqrt{\omega_{11}} - 2\sqrt{\gamma_{11}})^2 \leq \omega_{22} \leq (\sqrt{\omega_{11}} + 2\sqrt{\gamma_{11}})^2$. In combination with $\omega_{12} = \frac{1}{2}(\omega_{11} + \omega_{22}) - 2\gamma_{11}$, this defines $\mathscr{B}(\mathbf{\Gamma})$ for $D = 2$.

6. Within the set $\mathscr{B}(\mathbf{\Gamma})$, what basis correlations are possible? $\omega_{12} = 0$ when $\omega_{22} = 4\gamma_{11} - \omega_{11}$, $\omega_{12} < 0$ when $\omega_{22} < 4\gamma_{11} - \omega_{11}$, and $\omega_{12} > 0$ when $\omega_{22} > 4\gamma_{11} - \omega_{11}$. In particular, $\omega_{12}^2 = \omega_{11}\omega_{22}$ (i.e. $\rho_{12} = \pm 1$) along the line bounding the set. Here the partial correlations are the same as the marginal correlations.

7. $\mathbf{\Gamma}$ is at the "vertex" of the set $\mathscr{B}(\mathbf{\Gamma})$.



For any $D \geq 2$, given $\mathbf{\Gamma}$, there exists $\mathbf{\Omega} \in \mathscr{B}(\mathbf{\Gamma})$ in which $\omega_{ij} > 0$ and there exists $\mathbf{\Omega} \in \mathscr{B}(\mathbf{\Gamma})$ in which $\omega_{ij} < 0$, provided $\tau_{ij} > 0$. (A particular basis correlation always could be either negative or positive if there are no restrictions on $\mathbf{\Omega}$ and $\mathrm{var}[\log(x_i/x_j)] > 0$.)

1. First, ω_{ij} can be > 0: For clr data $\boldsymbol{z} = (\mathrm{clr}\, x_1, \ldots, \mathrm{clr}\, x_D)$ with covariance matrix $\mathbf{\Gamma}$, if $\gamma_{ij} > 0$, then $\mathbf{\Gamma}$ is such a $\mathbf{\Omega}$. If $\gamma_{ij} \leq 0$, let $\boldsymbol{y} = \boldsymbol{z} + e$ where e is any random variable uncorrelated with every z_i and with $\mathrm{var}(e) > \gamma_{ij}$. Then $\mathrm{var}(\boldsymbol{y}) \in \mathscr{B}(\mathbf{\Gamma})$ because $\mathrm{clr}(\boldsymbol{y}) = \boldsymbol{z}$, and $\mathrm{cov}(y_i, y_j) = \gamma_{ij} + \mathrm{var}(e) > 0$.

2. Now, ω_{ij} can be < 0 provided $\tau_{ij} > 0$: For clr data $\boldsymbol{z} = (\operatorname{clr} x_1, \ldots, \operatorname{clr} x_D)$ with covariance matrix $\boldsymbol{\Gamma}$, if $\gamma_{ij} < 0$, then $\boldsymbol{\Gamma}$ is such a $\boldsymbol{\Omega}$. If $\gamma_{ij} \geq 0$, let $\boldsymbol{y} = \boldsymbol{z} - \frac{1}{2}(z_i + z_j)$. Then $\operatorname{cov}(y_i, y_j) = \operatorname{cov}(\frac{1}{2} \operatorname{clr} x_i - \frac{1}{2} \operatorname{clr} x_j, -\frac{1}{2} \operatorname{clr} x_i + \frac{1}{2} \operatorname{clr} x_j) = -\frac{1}{4}(\gamma_{ii} - 2\gamma_{ij} + \gamma_{jj}) = \frac{1}{4}\tau_{ij} < 0$ and in fact $\operatorname{var}(y_i) = \operatorname{var}(y_j) = \frac{1}{4}\tau_{ij}$, so $\operatorname{cor}(y_i, y_j) = -1$.

For $D > 2$, based on empirical investigation, any value is possible for a given partial correlation if there are no restrictions on $\boldsymbol{\Omega}$.

For a given $\boldsymbol{\Gamma}$, potential $\boldsymbol{\Omega}$s can be randomly generated within the set constrained by $\operatorname{tr}(\boldsymbol{\Omega}) \leq V_{max}$.

1. Generate a random unit-length vector $\boldsymbol{a}$ using a re-scaled D-dimensional normal(0, 1) vector.
2. Let $\boldsymbol{z} = (z_1, \ldots, z_D)^\intercal$ be the clr vector. Calculate maximum r such that $\sum_{i=1}^{D} \operatorname{var}(z_i + r\boldsymbol{a}^\intercal \boldsymbol{z}) \leq V_{max}$. Call that r_{max}.
3. Generate a random r from the uniform(0, r_{max}) distribution.
4. Calculate $v_{max} = V_{max} - \sum_{i=1}^{D} \operatorname{var}(z_i + r\boldsymbol{a}^\intercal \boldsymbol{z})$.
5. Generate a random v from the uniform(0, v_{max}) distribution.
6. Let the log abundances be $y_i = z_i + r\boldsymbol{a}^\intercal \boldsymbol{z} + e$ where e is an independent random variable with $\operatorname{var}(e) = v$.
7. Then $\boldsymbol{\Omega} = \boldsymbol{\Gamma} + r\boldsymbol{\Gamma}\boldsymbol{A} + r\boldsymbol{A}^\intercal\boldsymbol{\Gamma} + r^2\boldsymbol{A}^\intercal\boldsymbol{\Gamma}\boldsymbol{A} + v\boldsymbol{J}$.