Sample covariance filtering

Benjamin Ménétrier

Last update: January 31, 2024

Copyright ©2015-2024: UCAR, CERFACS, METEO-FRANCE, IRIT and Meteorologisk Institutt

Contents

Part I.

Sampling covariances

1. Sampling theory

In this section, indices i, j, k and l refer to vector elements and indices p, q, r and s to ensemble members.

1.1. Sampling error

Let $\mathbf{x}^b \in \mathbb{R}^n$ be a random vector and $\boldsymbol{\varepsilon}^b$ its unbiased counterpart:

$$\boldsymbol{\varepsilon}^b = \mathbf{x}^b - \mathbb{E}\left[\mathbf{x}^b\right] \tag{1}$$

where $\mathbb{E}[\cdot]$ denotes the statistical expectation. The covariance of the distribution of $\mathbf{x}^b$ is denoted $\mathbf{B} \in \mathbb{R}^{n\times n}$ and its fourth order centered moment is denoted $\boldsymbol{\Xi} \in \mathbb{R}^{n\times n\times n\times n}$. They are respectively given by:

$$B_{ij} = \mathbb{E}\left[\varepsilon_i^b \varepsilon_j^b\right] \tag{2a}$$

$$\Xi_{ijkl} = \mathbb{E}\left[\varepsilon_i^b \varepsilon_j^b \varepsilon_k^b \varepsilon_l^b\right] \tag{2b}$$

Let $\left\{\widetilde{\mathbf{x}}_1^b, \dots, \widetilde{\mathbf{x}}_N^b\right\}$ be a set of N vectors that samples the distribution of $\mathbf{x}^b$ and $\left\{\delta\widetilde{\mathbf{x}}_1^b, \dots, \delta\widetilde{\mathbf{x}}_N^b\right\}$ their centered counterparts:

$$\delta\widetilde{\mathbf{x}}_p^b = \widetilde{\mathbf{x}}_p^b - \langle\widetilde{\mathbf{x}}^b\rangle \tag{3}$$

where $\langle\cdot\rangle$ denotes the ensemble mean:

$$\langle\widetilde{\mathbf{x}}^b\rangle = \frac{1}{N}\sum_{p=1}^{N} \widetilde{\mathbf{x}}_p^b \tag{4}$$

Centered moments (2a) et (2b) can be estimated from these perturbations by:

$$\widetilde{B}_{ij} = \frac{1}{N-1}\sum_{p=1}^{N} \delta\widetilde{x}_{i,p}^b \delta\widetilde{x}_{j,p}^b \tag{5a}$$

$$\widetilde{\Xi}_{ijkl} = \frac{1}{N}\sum_{p=1}^{N} \delta\widetilde{x}_{i,p}^b \delta\widetilde{x}_{j,p}^b \delta\widetilde{x}_{k,p}^b \delta\widetilde{x}_{l,p}^b \tag{5b}$$

If the ensemble size N goes to infinity, asymptotic $\widetilde{\mathbf{B}}$ and $\widetilde{\boldsymbol{\Xi}}$ converge to $\mathbf{B}$ and $\boldsymbol{\Xi}$ respectively:

$$\lim_{N\to\infty} \widetilde{\mathbf{B}} = \mathbf{B} \tag{6a}$$

$$\lim_{N\to\infty} \widetilde{\boldsymbol{\Xi}} = \boldsymbol{\Xi} \tag{6b}$$

The sampling errors on $\widetilde{\mathbf{B}}$ and $\widetilde{\boldsymbol{\Xi}}$ are respectively defined by:

$$\widetilde{\mathbf{B}}^e = \widetilde{\mathbf{B}} - \mathbf{B} \tag{7a}$$

$$\widetilde{\boldsymbol{\Xi}}^e = \widetilde{\boldsymbol{\Xi}} - \boldsymbol{\Xi} \tag{7b}$$

1.2. Expected sample mean

Since $\mathbb{E}\left[\widetilde{\mathbf{x}}^b_p\right]$ is the same for any member index p, it is hereafter denoted $\mathbb{E}\left[\widetilde{\mathbf{x}}^b\right]$. Thus, the sample mean expectation is given by:

$$\begin{aligned}
\mathbb{E}\left[\langle\widetilde{\mathbf{x}}^b\rangle\right] &= \mathbb{E}\left[\frac{1}{N}\sum_{p=1}^{N}\widetilde{\mathbf{x}}^b_p\right] \\
&= \frac{1}{N}\sum_{p=1}^{N}\mathbb{E}\left[\widetilde{\mathbf{x}}^b\right] \\
&= \mathbb{E}\left[\widetilde{\mathbf{x}}^b\right]
\end{aligned} \tag{8}$$

1.3. Covariance of sample means

Since ensemble members are mutually independent, their covariance is:

$$\mathrm{Cov}\left(\widetilde{x}^b_{i,p}, \widetilde{x}^b_{j,q}\right) = \begin{cases} B_{ij} & \text{if } p = q \\ 0 & \text{if } p \neq q \end{cases} \tag{9}$$

Thus, the expected product of sample means is given by:

$$\begin{aligned}
\mathbb{E}\left[\langle\widetilde{x}^b_i\rangle\langle\widetilde{x}^b_j\rangle\right] =& \mathbb{E}\left[\left(\frac{1}{N}\sum_{p=1}^{N}\widetilde{x}^b_{i,p}\right)\left(\frac{1}{N}\sum_{q=1}^{N}\widetilde{x}^b_{j,q}\right)\right] \\
&= \frac{1}{N^2}\sum_{1\leq p,q\leq N}\mathbb{E}\left[\widetilde{x}^b_{i,p}\widetilde{x}^b_{j,q}\right] \\
&= \frac{1}{N^2}\sum_{1\leq p,q\leq N}\left(\mathrm{Cov}\left(\widetilde{x}^b_{i,p}, \widetilde{x}^b_{j,q}\right) + \mathbb{E}\left[\widetilde{x}^b_i\right]\mathbb{E}\left[\widetilde{x}^b_j\right]\right) \\
&= \frac{1}{N^2}\left(\sum_{p=1}^{N}\mathrm{Cov}\left(\widetilde{x}^b_{i,p}, \widetilde{x}^b_{j,p}\right) + \sum_{\substack{1\leq p,q\leq N \\ p\neq q}}\mathrm{Cov}\left(\widetilde{x}^b_{i,p}, \widetilde{x}^b_{j,q}\right)\right) + \mathbb{E}\left[\widetilde{x}^b_i\right]\mathbb{E}\left[\widetilde{x}^b_j\right] \\
&= \frac{1}{N}B_{ij} + \mathbb{E}\left[\widetilde{x}^b_i\right]\mathbb{E}\left[\widetilde{x}^b_j\right]
\end{aligned} \tag{10}$$

so that the covariance of sample means is given by:

$$\begin{aligned}
\mathrm{Cov}\left(\langle\widetilde{x}^b_i\rangle, \langle\widetilde{x}^b_j\rangle\right) &= \mathbb{E}\left[\langle\widetilde{x}^b_i\rangle\langle\widetilde{x}^b_j\rangle\right] - \mathbb{E}\left[\widetilde{x}^b_i\right]\mathbb{E}\left[\widetilde{x}^b_j\right] \\
&= \frac{1}{N}B_{ij}
\end{aligned} \tag{11}$$

This equation is known as the "standard error" of the sample mean.

1.4. Equivalent estimator of sample covariance

Ensemble perturbations $\delta \mathbf{x}_p^b$ can be expanded in:

$$\begin{aligned}
\delta \widetilde{\mathbf{x}}_p^b &= \widetilde{\mathbf{x}}_p^b - \left\langle \widetilde{\mathbf{x}}^b \right\rangle \\
&= \widetilde{\mathbf{x}}_p^b - \frac{1}{N} \sum_{q=1}^{N} \widetilde{\mathbf{x}}_q^b \\
&= \frac{1}{N} \sum_{q=1}^{N} \left(\widetilde{\mathbf{x}}_p^b - \widetilde{\mathbf{x}}_q^b \right)
\end{aligned} \tag{12}$$

Thus, $\widetilde{\mathbf{B}}$ can be expressed as:

$$\begin{aligned}
\widetilde{B}_{ij} &= \frac{1}{N^2(N-1)} \sum_{1 \leq p,q,r \leq N} \left(\widetilde{x}_{i,p}^b - \widetilde{x}_{i,q}^b \right) \left(\widetilde{x}_{j,p}^b - \widetilde{x}_{j,r}^b \right) \\
&= \frac{1}{N^2(N-1)} \sum_{\substack{1 \leq p < q \leq N \\ 1 \leq r \leq N}} \left(\left(\widetilde{x}_{i,p}^b - \widetilde{x}_{i,q}^b \right) \left(\widetilde{x}_{j,p}^b - \widetilde{x}_{j,r}^b \right) + \left(\widetilde{x}_{i,q}^b - \widetilde{x}_{i,p}^b \right) \left(\widetilde{x}_{j,q}^b - \widetilde{x}_{j,r}^b \right) \right) \\
&= \frac{1}{N^2(N-1)} \sum_{\substack{1 \leq p < q \leq N \\ 1 \leq r \leq N}} \left(\widetilde{x}_{i,p}^b - \widetilde{x}_{i,q}^b \right) \left(\widetilde{x}_{j,p}^b - \widetilde{x}_{j,q}^b \right) \\
&= \frac{1}{N(N-1)} \sum_{1 \leq p < q \leq N} \left(\widetilde{x}_{i,p}^b - \widetilde{x}_{i,q}^b \right) \left(\widetilde{x}_{j,p}^b - \widetilde{x}_{j,q}^b \right)
\end{aligned} \tag{13}$$

This new covariance estimator is exactly equivalent to equation (5a).

1.5. Unbiased perturbations

Since we are interested in centered moments, it is possible to simplify the notations by using the unbiased counterparts of ensemble members:

$$\widetilde{\boldsymbol{\varepsilon}}_p^b = \widetilde{\mathbf{x}}_p^b - \mathbb{E}\left[\widetilde{\mathbf{x}}^b \right] \tag{14}$$

By definition, the expectation of the unbiased perturbation $\widetilde{\boldsymbol{\varepsilon}}_p^b$ vanishes, and its centered moments are equal to those of $\boldsymbol{\varepsilon}^b$. Unbiased perturbations $\left\{ \widetilde{\boldsymbol{\varepsilon}}_1^b, \ldots, \widetilde{\boldsymbol{\varepsilon}}_N^b \right\}$ are also mutually independent. As a summary, the following properties hold:

$$\mathbb{E}\left[\widetilde{\varepsilon}_{i,p}^b \right] = 0 \tag{15a}$$

$$\mathbb{E}\left[\widetilde{\varepsilon}_{i,p}^b \widetilde{\varepsilon}_{j,p}^b \right] = B_{ij} \tag{15b}$$

$$\mathbb{E}\left[\widetilde{\varepsilon}_{i,p}^b \widetilde{\varepsilon}_{j,p}^b \widetilde{\varepsilon}_{k,p}^b \widetilde{\varepsilon}_{l,p}^b \right] = \Xi_{ijkl} \tag{15c}$$

$$\mathbb{E}\left[\prod_p \prod_{i_p} \widetilde{\varepsilon}_{i_p,p}^b \right] = \prod_p \mathbb{E}\left[\prod_{i_p} \widetilde{\varepsilon}_{i_p,p}^b \right] \tag{15d}$$

1.6. Expected sample covariance

The covariance estimator (13) can be simply rewritten using unbiased perturbations:

$$\widetilde{B}_{ij} = \frac{1}{N(N-1)} \sum_{1\le p<q\le N} \left(\widetilde{\varepsilon}^b_{i,p} - \widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p} - \widetilde{\varepsilon}^b_{j,q}\right) \tag{16}$$

The expectation of $\widetilde{B}_{ij}$ given by the new estimator (13) is:

$$\mathbb{E}\left[\widetilde{B}_{ij}\right] = \frac{1}{N(N-1)} \sum_{1\le p<q\le N} \left(\mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,q}\right] - \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,q}\right] - \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,p}\right]\right) \tag{17}$$

Using properties (15b) and (15d):

$$\mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\right] = \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,q}\right] = B_{ij} \tag{18a}$$

$$\mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,q}\right] = \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,p}\right] = 0 \tag{18b}$$

Since the cardinality of the set $\{1 \le p < q \le N\}$ is $N(N-1)/2$, we get:

$$\mathbb{E}\left[\widetilde{B}_{ij}\right] = B_{ij} \tag{19}$$

Thus, covariance estimators (5a) and (13) are unbiased.

1.7. Expected product of sample covariances

The expected product of two sample covariances $\widetilde{B}_{ij}$ and $\widetilde{B}_{kl}$ is given by:

$$\begin{aligned}
\mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] &= \mathbb{E}\left[\left(\frac{1}{N(N-1)} \sum_{1\le p<q\le N} \left(\widetilde{\varepsilon}^b_{i,p} - \widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p} - \widetilde{\varepsilon}^b_{j,q}\right)\right)\right.\\
&\quad \left. \times \left(\frac{1}{N(N-1)} \sum_{1\le r<s\le N} \left(\widetilde{\varepsilon}^b_{k,r} - \widetilde{\varepsilon}^b_{k,s}\right)\left(\widetilde{\varepsilon}^b_{l,r} - \widetilde{\varepsilon}^b_{l,s}\right)\right)\right]\\
&= \frac{1}{N^2(N-1)^2} \sum_{\substack{1\le p<q\le N\\ 1\le r<s\le N}} \mathbb{E}\left[\left(\widetilde{\varepsilon}^b_{i,p} - \widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p} - \widetilde{\varepsilon}^b_{j,q}\right)\right.\\
&\quad \left. \times \left(\widetilde{\varepsilon}^b_{k,r} - \widetilde{\varepsilon}^b_{k,s}\right)\left(\widetilde{\varepsilon}^b_{l,r} - \widetilde{\varepsilon}^b_{l,s}\right)\right]
\end{aligned} \tag{20}$$

It should be emphasized that:

- expectation of unbiased perturbation $\widetilde{\boldsymbol{\varepsilon}}^b_p$ vanishes (property (15a)).
- unbiased perturbations are mutually independent (property (15d)),

As a consequence, the expectation of a product of several unbiased perturbations vanishes if one of them (i.e. one ensemble member index) is present only once in the product. For sake of clarity, such vanishing combinations are not explicitly mentioned hereafter. Besides, we can get the asymptotic sample covariance and asymptotic fourth-order centered moment from properties (15b) and (15c).

Indices p and q can be drawn first: there are $\dfrac{N(N-1)}{2}$ such draws for $1 \leq p < q \leq N$. Indices r and s are drawn in a second step, splitting the possibilities in three subsets depending on the cardinality of the intersection $\{p,q\} \cap \{r,s\}$, which can only takes values of 0, 1 and 2.

1. First case: $|\{p,q\} \cap \{r,s\}| = 0$.
   Indices r and s have to verify the condition $r,s \neq p$ and $r,s \neq q$. In the following diagram, $\varnothing$ denotes the forbidden values:

$$
\begin{array}{ccccccccc}
1 & & & p & & & q & & N \\
[& r,s &] & \varnothing & [& r,s &] \; \varnothing & [& r,s \;]
\end{array}
$$

   There are $\dfrac{(N-2)(N-3)}{2}$ draws verifying $1 \leq r < s \leq N$. Globally, there are $\dfrac{N(N-1)(N-2)(N-3)}{4}$ such terms in the sum, each one similar to:

$$
\begin{aligned}
& \mathbb{E}\left[\left(\widetilde{\varepsilon}^b_{i,p} - \widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p} - \widetilde{\varepsilon}^b_{j,q}\right)\left(\widetilde{\varepsilon}^b_{k,r} - \widetilde{\varepsilon}^b_{k,s}\right)\left(\widetilde{\varepsilon}^b_{l,r} - \widetilde{\varepsilon}^b_{l,s}\right)\right] \\
&= \mathbb{E}\left[\left(\widetilde{\varepsilon}^b_{i,p} - \widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p} - \widetilde{\varepsilon}^b_{j,q}\right)\right]\mathbb{E}\left[\left(\widetilde{\varepsilon}^b_{k,r} - \widetilde{\varepsilon}^b_{k,s}\right)\left(\widetilde{\varepsilon}^b_{l,r} - \widetilde{\varepsilon}^b_{l,s}\right)\right] \\
&= \left(\mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,q}\right]\right)\left(\mathbb{E}\left[\widetilde{\varepsilon}^b_{k,r}\widetilde{\varepsilon}^b_{l,r}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{k,s}\widetilde{\varepsilon}^b_{l,s}\right]\right) \\
&= \left(B_{ij} + B_{ij}\right)\left(B_{kl} + B_{kl}\right) \\
&= 4 B_{ij} B_{kl}
\end{aligned}
\tag{21}
$$

2. Second case: $|\{p,q\} \cap \{r,s\}| = 1$.
   Four sub-cases are possible to draw indices r and s verifying $1 \leq r < s \leq N$:

$$
\begin{array}{r|ccccccccccc}
 & 1 & & & p & & & & q & & & N \\
s = p: & [& r &] & s & [& & & \varnothing & & &] \\
s = q \text{ and } r \neq p: & [& r &] & \varnothing & [& r &] & s & [& \varnothing &] \\
s = q \text{ and } r \neq p: & [& \varnothing &] & r & [& s &] & \varnothing & [& s &] \\
p = q: & [& & & \varnothing & & &] & r & [& s &]
\end{array}
$$

   There are $2(N-2)$ possible draws. Globally, there are $N(N-1)(N-2)$ such terms in the sum, each one similar to:

$$\begin{aligned}
&\mathbb{E}\left[\left(\widetilde{\varepsilon}^b_{i,p}-\widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p}-\widetilde{\varepsilon}^b_{j,q}\right)\left(\widetilde{\varepsilon}^b_{k,p}-\widetilde{\varepsilon}^b_{k,s}\right)\left(\widetilde{\varepsilon}^b_{l,p}-\widetilde{\varepsilon}^b_{l,s}\right)\right] \\
&= \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\widetilde{\varepsilon}^b_{k,p}\widetilde{\varepsilon}^b_{l,p}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\right]\mathbb{E}\left[\widetilde{\varepsilon}^b_{k,s}\widetilde{\varepsilon}^b_{l,s}\right] \\
&\quad + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,q}\right]\left(\mathbb{E}\left[\widetilde{\varepsilon}^b_{k,p}\widetilde{\varepsilon}^b_{l,p}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{k,s}\widetilde{\varepsilon}^b_{l,s}\right]\right) \\
&= \Xi_{ijkl} + B_{ij}B_{kl} + B_{ij}\left(B_{kl}+B_{kl}\right) \\
&= \Xi_{ijkl} + 3B_{ij}B_{kl}
\end{aligned} \tag{22}$$

3. Third case: $|\{p,q\}\cap\{r,s\}| = 2$.
   Indices r and s have to verify $r = p$ and $s = q$:

$$\begin{array}{ccccccccccccc}
1 & & & p & & & & q & & & & N \\
\left[\right. & \varnothing & \left.\right] & r & \left[\right. & \varnothing & \left.\right] & s & \left[\right. & \varnothing & \left.\right]
\end{array}$$

   Globally, there are $\dfrac{N(N-1)}{2}$ such terms in the sum, each one similar to:

$$\begin{aligned}
&\mathbb{E}\left[\left(\widetilde{\varepsilon}^b_{i,p}-\widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p}-\widetilde{\varepsilon}^b_{j,q}\right)\left(\widetilde{\varepsilon}^b_{k,p}-\widetilde{\varepsilon}^b_{k,q}\right)\left(\widetilde{\varepsilon}^b_{l,p}-\widetilde{\varepsilon}^b_{l,q}\right)\right] \\
&= \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\widetilde{\varepsilon}^b_{k,p}\widetilde{\varepsilon}^b_{l,p}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,q}\widetilde{\varepsilon}^b_{k,q}\widetilde{\varepsilon}^b_{l,p}\right] \\
&\quad + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\right]\mathbb{E}\left[\widetilde{\varepsilon}^b_{k,q}\widetilde{\varepsilon}^b_{l,q}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{k,p}\right]\mathbb{E}\left[\widetilde{\varepsilon}^b_{j,q}\widetilde{\varepsilon}^b_{l,q}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{l,p}\right]\mathbb{E}\left[\widetilde{\varepsilon}^b_{j,q}\widetilde{\varepsilon}^b_{k,q}\right] \\
&\quad + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,q}\right]\mathbb{E}\left[\widetilde{\varepsilon}^b_{k,p}\widetilde{\varepsilon}^b_{l,p}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{k,q}\right]\mathbb{E}\left[\widetilde{\varepsilon}^b_{j,p}\widetilde{\varepsilon}^b_{l,p}\right] + \mathbb{E}\left[\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{l,q}\right]\mathbb{E}\left[\widetilde{\varepsilon}^b_{j,p}\widetilde{\varepsilon}^b_{k,p}\right] \\
&= 2\left(\Xi_{ijkl} + B_{ij}B_{kl} + B_{ik}B_{jl} + B_{il}B_{jk}\right)
\end{aligned} \tag{23}$$

Gathering all the terms, we get:

$$\begin{aligned}
\mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] &= \frac{1}{N^2(N-1)^2}\Big(N(N-1)(N-2)(N-3)B_{ij}B_{kl} \\
&\qquad + N(N-1)(N-2)\left(\Xi_{ijkl}+3B_{ij}B_{kl}\right) \\
&\qquad + N(N-1)\left(\Xi_{ijkl}+B_{ij}B_{kl}+B_{ik}B_{jl}+B_{il}B_{jk}\right)\Big) \\
&= P_1(N)\ \Xi_{ijkl} + P_2(N)\ B_{ij}B_{kl} + P_3(N)\left(B_{ik}B_{jl}+B_{il}B_{jk}\right)
\end{aligned} \tag{24}$$

with:

$$P_1(N) = \frac{N(N-1)(N-2)+N(N-1)}{N^2(N-1)^2} = \frac{1}{N} \tag{25a}$$

$$\begin{aligned}
P_2(N) &= \frac{N(N-1)(N-2)(N-3)+3N(N-1)(N-2)+N(N-1)}{N^2(N-1)^2} \\
&= \frac{N-1}{N}
\end{aligned} \tag{25b}$$

$$P_3(N) = \frac{N(N-1)}{N^2(N-1)^2} = \frac{1}{N(N-1)} \tag{25c}$$

In the case of a Gaussian distributed ensemble, the Wick-Isserlis theorem provides an explicit expression of higher order moments from covariances (Isserlis, 1916, 1918). Thus, we can get the fourth-order centered moment:

$$\Xi_{ijkl} = B_{ij}B_{kl} + B_{ik}B_{jl} + B_{il}B_{jk} \tag{26}$$

The expected product of sample covariances (24) is transformed into:

$$\mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] = B_{ij}B_{kl} + P_4(N)\left(B_{ik}B_{jl} + B_{il}B_{jk}\right) \tag{27}$$

with:

$$P_4(N) = P_1(N) + P_3(N) = \frac{1}{N-1} \tag{28}$$

This result can also be obtained from the Wishart theory (Wishart, 1928; Muirhead, 2005).

1.8. Expected sample fourth-order centered moment

With the development of equation (12), the sample fourth-order centered moment of equation (5b) becomes:

$$\widetilde{\Xi}_{ijkl} = \frac{1}{N^5} \sum_{1\leq p,q,r,s,t\leq N} \left(\widetilde{x}^b_{i,p} - \widetilde{x}^b_{i,q}\right)\left(\widetilde{x}^b_{j,p} - \widetilde{x}^b_{j,r}\right)\left(\widetilde{x}^b_{k,p} - \widetilde{x}^b_{k,s}\right)\left(\widetilde{x}^b_{l,p} - \widetilde{x}^b_{l,t}\right) \tag{29}$$

Since $\mathbb{E}\left[\widetilde{\mathbf{x}}^b_p\right]$ is the same for any member index p, the sample fourth-order centered moment can be rewritten using unbiased perturbations:

$$\widetilde{\Xi}_{ijkl} = \frac{1}{N^5} \sum_{1\leq p,q,r,s,t\leq N} \left(\widetilde{\varepsilon}^b_{i,p} - \widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p} - \widetilde{\varepsilon}^b_{j,r}\right)\left(\widetilde{\varepsilon}^b_{k,p} - \widetilde{\varepsilon}^b_{k,s}\right)\left(\widetilde{\varepsilon}^b_{l,p} - \widetilde{\varepsilon}^b_{l,t}\right) \tag{30}$$

The expectation of this sum is given by:

$$\mathbb{E}\left[\widetilde{\Xi}_{ijkl}\right] = \frac{1}{N^5} \sum_{1\leq p,q,r,s,t\leq N} \mathbb{E}\left[\left(\widetilde{\varepsilon}^b_{i,p} - \widetilde{\varepsilon}^b_{i,q}\right)\left(\widetilde{\varepsilon}^b_{j,p} - \widetilde{\varepsilon}^b_{j,r}\right)\left(\widetilde{\varepsilon}^b_{k,p} - \widetilde{\varepsilon}^b_{k,s}\right)\left(\widetilde{\varepsilon}^b_{l,p} - \widetilde{\varepsilon}^b_{l,t}\right)\right] \tag{31}$$

Index p can be drawn first: there are N such draws. Indices q, r, s and t are drawn in a second step, all being different from p since these cases do not bring any contribution to the sum. With the same remark as in the previous subsection, most of the terms vanish.

1. For the term Ξ_{ijkl}:
   - From the product $\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\widetilde{\varepsilon}^b_{k,p}\widetilde{\varepsilon}^b_{l,p}$:

     $$\begin{array}{ccccc} 1 & & p & & N \\ \left[\quad q,r,s,t \quad\right] & & \varnothing & & \left[\quad q,r,s,t \quad\right] \end{array}$$

     There are $N(N-1)^4$ such draws.

- From the product $\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,r}\widetilde{\varepsilon}^b_{k,s}\widetilde{\varepsilon}^b_{l,t}$:

$$\begin{array}{ccc} 1 & p & N \\ [\quad q=r=s=t \quad] & \varnothing & [\quad q=r=s=t \quad] \end{array}$$

  There are $N(N-1)$ such draws.

Globally, there are $N^2(N-1)(N^2-3N+3)$ such terms.

2. For the term $B_{ij}B_{kl}$:
   - From the product $\widetilde{\varepsilon}^b_{i,p}\widetilde{\varepsilon}^b_{j,p}\widetilde{\varepsilon}^b_{k,s}\widetilde{\varepsilon}^b_{l,t}$:

   $$\begin{array}{ccc} 1 & p & N \\ [\quad q,r,s=t \quad] & \varnothing & [\quad q,r,s=t \quad] \end{array}$$

     There are $N(N-1)^3$ such draws.
   - From the product $\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,r}\widetilde{\varepsilon}^b_{k,p}\widetilde{\varepsilon}^b_{l,p}$:

   $$\begin{array}{ccc} 1 & p & N \\ [\quad q=r,s,t \quad] & \varnothing & [\quad q=r,s,t \quad] \end{array}$$

     There are $N(N-1)^3$ such draws.
   - From the product $\widetilde{\varepsilon}^b_{i,q}\widetilde{\varepsilon}^b_{j,r}\widetilde{\varepsilon}^b_{k,s}\widetilde{\varepsilon}^b_{l,t}$:

   $$\begin{array}{ccc} 1 & p & N \\ [\quad q=r,s=t \quad] & \varnothing & [\quad q=r,s=t \quad] \end{array}$$

     There are $N(N-1)(N-2)$ such draws.

   Globally, there are $N^2(N-1)(2N-3)$ such terms. Indices permutations $B_{ik}B_{jl}$ and $B_{il}B_{jk}$ are also present in the same proportion.

Thus, the expectation of $\widetilde{\Xi}_{ijkl}$ is given by:

$$\mathbb{E}\left[\widetilde{\Xi}_{ijkl}\right] = P_5(N)\ \Xi_{ijkl} + P_6(N)\left(B_{ij}B_{kl} + B_{ik}B_{jl} + B_{il}B_{jk}\right) \tag{32}$$

with:

$$P_5(N) = \frac{(N-1)(N^2-3N+3)}{N^3} \tag{33a}$$

$$P_6(N) = \frac{(N-1)(2N-3)}{N^3} \tag{33b}$$

2. Multiple sub-ensembles with distinct distributions

2.1. Sub-ensembles characteristics

Instead of having an ensemble of size N that samples the distribution of a random vector $\mathbf{x}^b$, we assume in this section that we have R ensembles of respective sizes $N_1, \ldots, N_R$, with $\sum_{r=1}^{R} N_r = N$, which sample distinct random vectors $\mathbf{x}^b_1, \ldots, \mathbf{x}^b_R$. Each random vector $\mathbf{x}^b_r$ has a distribution characterized by its covariance $\mathbf{B}_r \in \mathbb{R}^{n\times n}$ and its fourth order centered moment $\mathbf{\Xi}_r \in \mathbb{R}^{n\times n\times n\times n}$. For each sub-ensemble $\left\{\widetilde{\mathbf{x}}^b_{r,1}, \ldots, \widetilde{\mathbf{x}}^b_{r,N_r}\right\}$, we define the centered counterparts:

$$\delta\widetilde{\mathbf{x}}^b_{r,p} = \widetilde{\mathbf{x}}^b_{r,p} - \langle\widetilde{\mathbf{x}}^b_r\rangle \tag{34}$$

where $\langle\cdot\rangle$ denotes the sub-ensemble mean:

$$\langle\widetilde{\mathbf{x}}^b_r\rangle = \frac{1}{N_r}\sum_{p=1}^{N_r}\widetilde{\mathbf{x}}^b_{r,p} \tag{35}$$

For each sub-ensemble, centered moments can be estimated from these perturbations by:

$$\widetilde{B}_{ij,r} = \frac{1}{N_r - 1}\sum_{p=1}^{N_r}\delta\widetilde{x}^b_{i,r,p}\delta\widetilde{x}^b_{j,r,p} \tag{36a}$$

$$\widetilde{\Xi}_{ijkl,r} = \frac{1}{N_r}\sum_{p=1}^{N_r}\delta\widetilde{x}^b_{i,r,p}\delta\widetilde{x}^b_{j,r,p}\delta\widetilde{x}^b_{k,r,p}\delta\widetilde{x}^b_{l,r,p} \tag{36b}$$

If the sub-ensemble size N_r goes to infinity, asymptotic $\widetilde{\mathbf{B}}_r$ and $\widetilde{\mathbf{\Xi}}_r$ converge to $\mathbf{B}_r$ and $\mathbf{\Xi}_r$ respectively:

$$\lim_{N_r\to\infty}\widetilde{\mathbf{B}}_r = \mathbf{B}_r \tag{37a}$$

$$\lim_{N_r\to\infty}\widetilde{\mathbf{\Xi}}_r = \mathbf{\Xi}_r \tag{37b}$$

2.2. Full ensemble sample covariance issues

If we consider the full ensemble $\left\{\widetilde{\mathbf{x}}_{1,1}, \ldots, \widetilde{\mathbf{x}}_{1,N_1}, \ldots, \widetilde{\mathbf{x}}_{R,1}, \ldots, \widetilde{\mathbf{x}}_{R,N_R}\right\}$, its ensemble mean is:

$$\langle\widetilde{\mathbf{x}}\rangle = \frac{1}{N}\sum_{r=1}^{R}\sum_{p_r=1}^{N_r}\widetilde{\mathbf{x}}_{r,p_r} \tag{38}$$

which can be expressed as:

$$\begin{aligned}
\langle\widetilde{\mathbf{x}}\rangle &= \frac{1}{N}\sum_{s=1}^{R}N_s\langle\widetilde{\mathbf{x}}_s\rangle \\
&= \langle\widetilde{\mathbf{x}}_r\rangle + \sum_{s=1}^{R}\frac{N_s}{N}\Big(\langle\widetilde{\mathbf{x}}_s\rangle - \langle\widetilde{\mathbf{x}}_r\rangle\Big)
\end{aligned} \tag{39}$$

The sample covariance matrix of the full ensemble is given by:

$$\widetilde{B}^F_{ij} = \frac{1}{N-1} \sum_{r=1}^{R} \sum_{p_r=1}^{N_r} \left(\widetilde{x}_{i,r,p_r} - \langle \widetilde{x}_i \rangle \right) \left(\widetilde{x}_{j,r,p_r} - \langle \widetilde{x}_j \rangle \right) \tag{40}$$

Using the expansion of the full ensemble mean, we get for each sub-ensemble r:

$$\begin{aligned}
&\sum_{p_r=1}^{N_r} \left(\widetilde{x}_{i,r,p_r} - \langle \widetilde{x}_i \rangle \right) \left(\widetilde{x}_{j,r,p_r} - \langle \widetilde{x}_j \rangle \right) \\
&= \sum_{p_r=1}^{N_r} \left(\widetilde{x}_{i,r,p_r} - \langle \widetilde{x}_{i,r} \rangle - \sum_{s=1}^{R} \frac{N_s}{N} \left(\langle \widetilde{x}_{i,s} \rangle - \langle \widetilde{x}_{i,r} \rangle \right) \right) \left(\widetilde{x}_{j,r,p_r} - \langle \widetilde{x}_{j,r} \rangle - \sum_{t=1}^{R} \frac{N_t}{N} \left(\langle \widetilde{x}_{j,t} \rangle - \langle \widetilde{x}_{j,r} \rangle \right) \right) \\
&= \sum_{p_r=1}^{N_r} \left(\widetilde{x}_{i,r,p_r} - \langle \widetilde{x}_{i,r} \rangle \right) \left(\widetilde{x}_{j,r,p_r} - \langle \widetilde{x}_{j,r} \rangle \right) + \sum_{p_r=1}^{N_r} \sum_{1 \le s,t \le R} \frac{N_s N_t}{N^2} \left(\langle \widetilde{x}_{i,s} \rangle - \langle \widetilde{x}_{i,r} \rangle \right) \left(\langle \widetilde{x}_{j,t} \rangle - \langle \widetilde{x}_{j,r} \rangle \right) \\
&= \left(N_r - 1 \right) \widetilde{B}_{ij,r} + \sum_{1 \le s,t \le R} \frac{N_r N_s N_t}{N^2} \left(\langle \widetilde{x}_{i,s} \rangle - \langle \widetilde{x}_{i,r} \rangle \right) \left(\langle \widetilde{x}_{j,t} \rangle - \langle \widetilde{x}_{j,r} \rangle \right)
\end{aligned} \tag{41}$$

Using the same process as for equation (13), the last term summed over all the sub-ensembles becomes:

$$\begin{aligned}
&\sum_{1 \le r,s,t \le R} \frac{N_r N_s N_t}{N^2} \left(\langle \widetilde{x}_{i,s} \rangle - \langle \widetilde{x}_{i,r} \rangle \right) \left(\langle \widetilde{x}_{j,t} \rangle - \langle \widetilde{x}_{j,r} \rangle \right) \\
&= \sum_{1 \le r < s \le R} \frac{N_r N_s}{R N^2} \left(\langle \widetilde{x}_{i,s} \rangle - \langle \widetilde{x}_{i,r} \rangle \right) \left(\langle \widetilde{x}_{j,s} \rangle - \langle \widetilde{x}_{j,r} \rangle \right)
\end{aligned} \tag{42}$$

Thus, the sample covariance matrix of equation for the full ensemble can be expressed as:

$$\widetilde{B}^F_{ij} = \frac{1}{N-1} \left(\sum_{r=1}^{R} \left(N_r - 1 \right) \widetilde{B}_{ij,r} + \sum_{1 \le r < s \le R} \frac{N_r N_s}{R N^2} \left(\langle \widetilde{x}_{i,s} \rangle - \langle \widetilde{x}_{i,r} \rangle \right) \left(\langle \widetilde{x}_{j,s} \rangle - \langle \widetilde{x}_{j,r} \rangle \right) \right) \tag{43}$$

It appears that the full ensemble sample covariance $\widetilde{\mathbf{B}}^F$ is not a linear combination of the sub-ensemble sample covariances. Indeed, there is also a term involving products of sub-ensemble means if these sample means differ between sub-ensembles.

The expectation of a sub-ensemble mean is simply the distribution mean:

$$\mathbb{E} \left[\langle \widetilde{x}_{i,r} \rangle \right] = \mathbb{E} \left[x_{i,r} \right] \tag{44}$$

Moreover, the covariance of two elements i and j of a same sub-ensemble mean is given equation (11):

$$\mathrm{Cov} \left[\langle \widetilde{x}_{i,r} \rangle, \langle \widetilde{x}_{j,r} \rangle \right] = \frac{1}{N_r} B_{ij,r} \tag{45}$$

whereas for distinct sub-ensembles $r \neq s$:

$$\mathrm{Cov}\left[\langle\widetilde{x}_{i,r}\rangle, \langle\widetilde{x}_{j,s}\rangle\right] = 0 \tag{46}$$

Using the classical formula for covariances, we get:

$$\begin{aligned}\mathbb{E}\left[\langle\widetilde{x}_{i,r}\rangle\langle\widetilde{x}_{j,r}\rangle\right] &= \mathrm{Cov}\left[\langle\widetilde{x}_{i,r}\rangle, \langle\widetilde{x}_{j,r}\rangle\right] + \mathbb{E}\left[\langle\widetilde{x}_{i,r}\rangle\right]\mathbb{E}\left[\langle\widetilde{x}_{j,r}\rangle\right] \\ &= \frac{1}{N_r} B_{ij,r} + \mathbb{E}\left[x_{i,r}\right]\mathbb{E}\left[x_{j,r}\right]\end{aligned} \tag{47}$$

and for distinct sub-ensembles $r \neq s$:

$$\mathbb{E}\left[\langle\widetilde{x}_{i,r}\rangle\langle\widetilde{x}_{j,s}\rangle\right] = \mathbb{E}\left[x_{i,r}\right]\mathbb{E}\left[x_{j,s}\right] \tag{48}$$

With these results, we can compute the expectation of the full ensemble sample covariance matrix:

$$\begin{aligned}\mathbb{E}\left[\widetilde{B}_{ij}\right] &= \frac{1}{N-1}\left(\sum_{r=1}^{R}\left(N_r - 1\right)B_{ij,r} + \sum_{1\leq r<s\leq R}\frac{N_rN_s}{RN^2}\left(\frac{1}{N_r}B_{ij,r} + \frac{1}{N_s}B_{ij,s}\right.\right. \\ &\quad \left.\left. +\mathbb{E}\left[x_{i,r}\right]\mathbb{E}\left[x_{j,r}\right] + \mathbb{E}\left[x_{i,s}\right]\mathbb{E}\left[x_{j,s}\right] - \mathbb{E}\left[x_{i,r}\right]\mathbb{E}\left[x_{j,s}\right] - \mathbb{E}\left[x_{i,s}\right]\mathbb{E}\left[x_{j,r}\right]\right)\right) \\ &= \frac{1}{N-1}\left(\sum_{r=1}^{R}\left(N_r - 1\right)B_{ij,r} + \frac{1}{RN^2}\sum_{1\leq r<s\leq R}\left(N_sB_{ij,r} + N_rB_{ij,s}\right)\right. \\ &\quad \left. + \sum_{1\leq r<s\leq R}\frac{N_rN_s}{RN^2}\left(\mathbb{E}\left[x_{i,s}\right] - \mathbb{E}\left[x_{i,r}\right]\right)\left(\mathbb{E}\left[x_{j,s}\right] - \mathbb{E}\left[x_{j,r}\right]\right)\right)\end{aligned} \tag{49}$$

This equation is a confirmation of the systematic impact of distributions means differences on the full ensemble expected sample covariance $\widetilde{\mathbf{B}}^F$, which is not a desired property.

2.3. Corrected sample moments

As a solution, we can remove sub-ensemble means for each sub-ensemble in the calculation of the full ensemble sample covariance. We define the sum of perturbations products $\widetilde{\mathbf{S}}^c$:

$$\begin{aligned}\widetilde{S}^c_{ij} &= \sum_{r=1}^{R}\sum_{p_r=1}^{N_r}\left(\widetilde{x}_{i,r,p_r} - \langle\widetilde{x}_{i,r}\rangle\right)\left(\widetilde{x}_{j,r,p_r} - \langle\widetilde{x}_{j,r}\rangle\right) \\ &= \sum_{r=1}^{R}\left(N_r - 1\right)\widetilde{B}_{ij,r}\end{aligned} \tag{50}$$

Its expectation is given by:

$$\mathbb{E}\left[\widetilde{S}^c_{ij}\right] = \sum_{r=1}^{R}\left(N_r - 1\right)B_{ij,r} \tag{51}$$

and we look for a corrected sample covariance matrix $\widetilde{\mathbf{B}}^c$ that would be proportional to $\widetilde{\mathbf{S}}^c$, with a proportionality coefficient α such that if $\mathbf{B}_1 = \cdots = \mathbf{B}_R = \mathbf{B}$, then $\mathbb{E}\left[\widetilde{\mathbf{B}}^c\right] = \mathbf{B}$. Since:

$$\begin{aligned}
\mathbb{E}\left[\widetilde{B}^c_{ij}\right] &= \alpha \mathbb{E}\left[\widetilde{S}^c_{ij}\right] \\
&= \alpha \sum_{r=1}^{R} \left(N_r - 1\right) B_{ij} \\
&= \alpha (N - R) B_{ij}
\end{aligned} \tag{52}$$

then $\alpha = (N-R)^{-1}$ and:

$$\widetilde{B}^c_{ij} = \frac{1}{N-R} \widetilde{S}^c_{ij} \tag{53}$$

This is consistent with the fact that the ensemble of N perturbations where R sub-ensemble means have been removed have only $N-R$ degrees of freedom left.

Similarly, for the sample fourth-order centered moment, we define the sum of perturbations products $\widetilde{\mathbf{\Psi}}^c$:

$$\begin{aligned}
\widetilde{\Psi}^c_{ijkl} &= \sum_{r=1}^{R} \sum_{p_r=1}^{N_r} \left(\widetilde{x}_{i,r,p_r} - \langle \widetilde{x}_{i,r} \rangle\right) \left(\widetilde{x}_{j,r,p_r} - \langle \widetilde{x}_{j,r} \rangle\right) \left(\widetilde{x}_{k,r,p_r} - \langle \widetilde{x}_{k,r} \rangle\right) \left(\widetilde{x}_{l,r,p_r} - \langle \widetilde{x}_{l,r} \rangle\right) \\
&= \sum_{r=1}^{R} N_r \widetilde{\Xi}_{ijkl,r}
\end{aligned} \tag{54}$$

Since the sample fourth-order centered moment estimator is biased, the previous method cannot be applied to find an appropriate normalization. Thus, we choose an arbitrary (but consistent) one for the corrected sample fourth-order centered moment:

$$\widetilde{\Xi}^c_{ijkl} = \frac{1}{N} \widetilde{\Psi}^c_{ijkl} \tag{55}$$

2.4. Asymptotic behavior

The corrected sample covariance $\widetilde{\mathbf{B}}^c$ is a linear combination of sub-ensemble sample covariances, the weights being related to the relative sizes of sub-ensembles. If we define $\boldsymbol{\gamma} \in [0,1]^R$ such that $N_r = \gamma_r N$, then we get:

$$\widetilde{B}^c_{ij} = \sum_{r=1}^{R} \frac{\gamma_r N - 1}{N-R} \widetilde{B}_{ij,r} \tag{56}$$

We denote $\mathbf{B}^c$ the asymptotic limit of $\widetilde{\mathbf{B}}^c$ while $\boldsymbol{\gamma}$ remains constant:

$$B^c_{ij} = \lim_{N \to \infty} \widetilde{B}^c_{ij} = \sum_{r=1}^{R} \gamma_r B_{ij,r} \tag{57}$$

The expected corrected sample covariance is:

$$\mathbb{E}\left[\widetilde{B}^c_{ij}\right] = \sum_{r=1}^{R} \frac{\gamma_r N - 1}{N - R} B_{ij,r} \tag{58}$$

so if $\gamma_r \neq (\gamma_r N - 1)/(N - R)$, asymptotic and expected corrected sample covariances are not equal:

$$B^c_{ij} \neq \mathbb{E}\left[\widetilde{B}^c_{ij}\right] \tag{59}$$

The corrected sample fourth-order centered moment $\widetilde{\boldsymbol{\Xi}}^c$ is also a linear combination of sub-ensembles sample fourth-order centered moments, with weigths $\gamma_r = N_r/N$:

$$\widetilde{\Xi}^c_{ijkl} = \sum_{r=1}^{R} \gamma_r \widetilde{\Xi}_{ijkl,r} \tag{60}$$

We denote $\boldsymbol{\Xi}^c$ the asymptotic limit of $\widetilde{\boldsymbol{\Xi}}^c$ while $\boldsymbol{\gamma}$ remains constant:

$$\Xi^c_{ijkl} = \lim_{N\to\infty} \widetilde{\Xi}^c_{ijkl} = \sum_{r=1}^{R} \gamma_r \Xi_{ijkl,r} \tag{61}$$

The expected corrected sample fourth-order centered moment is:

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\Xi}^c_{ijkl}\right] &= \sum_{r=1}^{R} \gamma_r \mathbb{E}\left[\widetilde{\Xi}_{ijkl,r}\right] \\
&= \sum_{r=1}^{R} \gamma_r \left(P_5(N_r)\, \Xi_{ijkl,r} + P_6(N_r) \left(B_{ij,r}B_{kl,r} + B_{ik,r}B_{jl,r} + B_{il,r}B_{jk,r}\right)\right)
\end{aligned} \tag{62}$$

2.5. Case of equal sub-ensemble sizes

In the case where all sub-ensemble sizes are equal ($N_1 = \cdots = N_R = N/R$), then $\gamma_r = 1/R$ and:

$$\frac{\gamma_r N - 1}{N - R} = \frac{N/R - 1}{N - R} = \frac{1}{R} \tag{63}$$

so that the corrected sample covariance $\widetilde{\mathbf{B}}^c$ becomes the regular average of sub-ensemble sample covariances:

$$\widetilde{B}^c_{ij} = \frac{1}{R} \sum_{r=1}^{R} \widetilde{B}_{ij,r} \tag{64}$$

As a consequence, the asymptotic and expected corrected sample covariances are equal:

$$B^c_{ij} = \frac{1}{R} \sum_{r=1}^{R} B_{ij,r} = \mathbb{E}\left[\widetilde{B}^c_{ij}\right] \tag{65}$$

The corrected sample fourth-order centered moment $\widetilde{\boldsymbol{\Xi}}^c$ becomes also the regular average of sub-ensemble sample fourth-order centered moment:

$$\widetilde{\Xi}^c_{ijkl} = \frac{1}{R}\sum_{r=1}^{R}\widetilde{\Xi}_{ijkl,r} \tag{66}$$

3. Iterative estimation of centered moments

3.1. General case

The mean of an ensemble of N members $\left\{\widetilde{\mathbf{x}}_1^b, \ldots, \widetilde{\mathbf{x}}_N^b\right\}$ is computed by:

$$\left\langle \widetilde{x}_i^b \right\rangle_N = \frac{1}{N} \sum_{p=1}^{N} \widetilde{x}_{i,p}^b \tag{67}$$

where the index N indicates the size of the ensemble over which the mean is computed. This equation can be transformed into:

$$\begin{aligned}
\left\langle \widetilde{x}_i^b \right\rangle_N &= \frac{1}{N} \left(\widetilde{x}_{i,N}^b + \sum_{p=1}^{N-1} \widetilde{x}_{i,p}^b \right) \\
&= \frac{1}{N} \left(\widetilde{x}_{i,N}^b + (N-1) \left\langle \widetilde{x}_i^b \right\rangle_{N-1} \right) \\
&= \left\langle \widetilde{x}_i^b \right\rangle_{N-1} + \frac{1}{N} \left(\widetilde{x}_{i,N}^b - \left\langle \widetilde{x}_i^b \right\rangle_{N-1} \right)
\end{aligned} \tag{68}$$

From this formula, $\left\langle \widetilde{x}_i^b \right\rangle_N$ can be updated knowing $\left\langle \widetilde{x}_i^b \right\rangle_{N-1}$ and a new member $\widetilde{x}_{i,N}^b$. Similarly, the sum $s_{\mathbf{k},N}$ where $\mathbf{k} \in \mathbb{N}^n$ is defined as:

$$s_{\mathbf{k},N} = \sum_{p=1}^{N} \prod_{i=1}^{n} \left(\widetilde{x}_{i,p}^b - \left\langle \widetilde{x}_i^b \right\rangle_N \right)^{k_i} \tag{69}$$

This sum can be split in two terms:

$$s_{\mathbf{k},N} = \sum_{p=1}^{N-1} \prod_{i=1}^{n} \left(\widetilde{x}_{i,p}^b - \left\langle \widetilde{x}_i^b \right\rangle_N \right)^{k_i} + \prod_{i=1}^{n} \left(\widetilde{x}_{i,N}^b - \left\langle \widetilde{x}_i^b \right\rangle_N \right)^{k_i} \tag{70}$$

Using equation (68) for the recursive computation of the mean and the binomial theorem, the first term of equation (70) can be expanded in:

$$\begin{aligned}
&\sum_{p=1}^{N-1} \prod_{i=1}^{n} \left(\widetilde{x}_{i,p}^b - \left\langle \widetilde{x}_i^b \right\rangle_N \right)^{k_i} \\
&= \sum_{p=1}^{N-1} \prod_{i=1}^{n} \left(\widetilde{x}_{i,p}^b - \left\langle \widetilde{x}_i^b \right\rangle_{N-1} - \frac{1}{N} \left(\widetilde{x}_{i,N}^b - \left\langle \widetilde{x}_i^b \right\rangle_{N-1} \right) \right)^{k_i} \\
&= \sum_{p=1}^{N-1} \prod_{i=1}^{n} \sum_{m_i=0}^{k_i} \binom{k_i}{m_i} \left(\widetilde{x}_{i,p}^b - \left\langle \widetilde{x}_i^b \right\rangle_{N-1} \right)^{m_i} \left(-\frac{1}{N} \left(\widetilde{x}_{i,N}^b - \left\langle \widetilde{x}_i^b \right\rangle_{N-1} \right) \right)^{k_i - m_i}
\end{aligned} \tag{71}$$

where $\mathbf{m} \in \mathbb{N}^n$. Rearranging the sums to extract p-dependent factors, we get:

$$\begin{aligned}
&\sum_{p=1}^{N-1} \prod_{i=1}^{n} \left(\widetilde{x}_{i,p}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N} \right)^{k_i} \\
&= \sum_{p=1}^{N-1} \sum_{m_1=0}^{k_1} \cdots \sum_{m_n=0}^{k_n} \prod_{i=1}^{n} \left(\binom{k_i}{m_i} \left(\frac{-1}{N} \right)^{k_i - m_i} \left(\widetilde{x}_{i,p}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right)^{m_i} \right. \\
&\quad \left. \times \left(\widetilde{x}_{i,N}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right)^{k_i - m_i} \right) \\
&= \sum_{m_1=0}^{k_1} \cdots \sum_{m_n=0}^{k_n} \left(\frac{-1}{N} \right)^{\sum_{i=1}^{n} (k_i - m_i)} \prod_{i=1}^{n} \left(\binom{k_i}{m_i} \left(\widetilde{x}_{i,N}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right)^{k_i - m_i} \right) \\
&\quad \times \sum_{p=1}^{N-1} \prod_{i=1}^{n} \left(\widetilde{x}_{i,p}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right)^{m_i} \\
&= \sum_{m_1=0}^{k_1} \cdots \sum_{m_n=0}^{k_n} \left(\frac{-1}{N} \right)^{\sum_{i=1}^{n} (k_i - m_i)} \prod_{i=1}^{n} \left(\binom{k_i}{m_i} \left(\widetilde{x}_{i,N}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right)^{k_i - m_i} \right) s_{\mathbf{m},N-1} \qquad (72)
\end{aligned}$$

The second term of (70) is simply modified in:

$$\begin{aligned}
\prod_{i=1}^{n} \left(\widetilde{x}_{i,N}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N} \right)^{k_i} &= \prod_{i=1}^{n} \left(\frac{N-1}{N} \left(\widetilde{x}_{i,N}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right) \right)^{k_i} \\
&= \left(\frac{N-1}{N} \right)^{\sum_{i=1}^{n} k_i} \prod_{i=1}^{n} \left(\widetilde{x}_{i,N}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right)^{k_i} \qquad (73)
\end{aligned}$$

Thus, bringing the two terms together:

$$\begin{aligned}
s_{\mathbf{k},N} = &\sum_{m_1=0}^{k_1} \cdots \sum_{m_n=0}^{k_n} \left(\frac{-1}{N} \right)^{\sum_{i=1}^{n} (k_i - m_i)} \prod_{i=1}^{n} \left(\binom{k_i}{m_i} \left(\widetilde{x}_{i,N}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right)^{k_i - m_i} \right) s_{\mathbf{m},N-1} \\
&+ \left(\frac{N-1}{N} \right)^{\sum_{i=1}^{n} k_i} \prod_{i=1}^{n} \left(\widetilde{x}_{i,N}^{b} - \left\langle \widetilde{x}_{i}^{b} \right\rangle_{N-1} \right)^{k_i} \qquad (74)
\end{aligned}$$

3.2. Covariance and fourth-order centered moment

For sake of clarity, we define an operator $/\ldots/$ taking as many integers as necessary as inputs, such as:

$$/ijkl\ldots/ = \begin{cases} 1 & \text{if indices } i,j,k,l,\ldots \text{ are all different} \\ 0 & \text{else} \end{cases} \qquad (75)$$

To estimate the fourth-order centered moment, at most four elements of the random vector are involved. To simplify the notations, we can assume that the size of the random vector is only $n = 4$, other elements being discarded. The algorithm to compute $\widetilde{\mathbf{B}}$ and $\widetilde{\Xi}$ iteratively is:

1. Load the first member $\widetilde{\mathbf{x}}^b_1$.
2. Initialize all the variables:
   - $\boldsymbol{\mu} = \widetilde{\mathbf{x}}^b_1$,
   - all the sums $s_{\mathbf{k},1}$ are set to zero.
3. For p between 2 and N :
   - Load p^{th} member $\widetilde{\mathbf{x}}^b_p$.
   - Update the fourth-order sum, $s_{(1,1,1,1),p}$:

$$
\begin{aligned}
s_{(1,1,1,1),p} = s_{(1,1,1,1),p-1} &- \frac{1}{p}\sum_{i=1}^{4}\left(\widetilde{x}^b_{i,p} - \langle \widetilde{x}^b_i \rangle_{p-1}\right) s_{(/1i/,/2i/,/3i/,/4i/),p-1} \\
&+ \frac{1}{p^2}\sum_{1\leq i,j\leq 4}\left(\widetilde{x}^b_{i,p} - \langle \widetilde{x}^b_i \rangle_{p-1}\right)\left(\widetilde{x}^b_{j,p} - \langle \widetilde{x}^b_j \rangle_{p-1}\right) \\
&\times s_{(/1ij/,/2ij/,/3ij/,/4ij/),p-1} \\
&+ \frac{(p-1)(p^2-3p+3)}{p^3}\prod_{i=1}^{4}\left(\widetilde{x}^b_{i,p} - \langle \widetilde{x}^b_i \rangle_{p-1}\right)
\end{aligned} \tag{76}
$$

   - Update third-order sums, of form $s_{\mathbf{k},p}$ where $\mathbf{k}$ is a permutation of the vector $(1,1,1,0)$. For instance with $s_{(1,1,1,0),p}$:

$$
\begin{aligned}
s_{(1,1,1,0),p} = s_{(1,1,1,0),p-1} &- \frac{1}{p}\sum_{i=1}^{3}\left(\widetilde{x}^b_{i,p} - \langle \widetilde{x}^b_i \rangle_{p-1}\right) s_{(/1i/,/2i/,/3i/,0),p-1} \\
&+ \frac{(p-1)(p-2)}{p^2}\prod_{i=1}^{3}\left(\widetilde{x}^b_{i,p} - \langle \widetilde{x}^b_i \rangle_{p-1}\right)
\end{aligned} \tag{77}
$$

   - Update second-order sums, of form $s_{\mathbf{k},p}$ where $\mathbf{k}$ is a permutation of the vector $(1,1,0,0)$. For instance with $s_{(1,1,0,0),p}$:

$$
s_{(1,1,0,0),p} = s_{(1,1,0,0),p-1} + \frac{p-1}{p}\left(\widetilde{x}^b_{1,p} - \langle \widetilde{x}^b_1 \rangle_{p-1}\right)\left(\widetilde{x}^b_{2,p} - \langle \widetilde{x}^b_2 \rangle_{p-1}\right) \tag{78}
$$

   - Update means:

$$
\boldsymbol{\mu} = \boldsymbol{\mu} + \frac{1}{p}\left(\widetilde{\mathbf{x}}^b_p - \boldsymbol{\mu}\right) \tag{79}
$$

4. Normalize sums to get centered moments:
   - Covariances:

$$
\widetilde{B}_{12} = \frac{1}{N-1} s_{(1,1,0,0),N} \tag{80}
$$

   - Fourth-order centered moments:

$$
\widetilde{\Xi}_{1234} = \frac{1}{N} s_{(1,1,1,1),N} \tag{81}
$$

Part II.

Filtering sample covariances

4. Use of the sampling theory

4.1. Random processes modeling

In linear filtering theory, the "truth" $\mathbf{B}$ is considered as a random variable, from which we have a noisy estimation $\widetilde{\mathbf{B}}$ only (Wiener, 1949). It should be noted that in the previous section, the truth was considered as a fixed and known value, which is not the case anymore.

We model the generation of sample centered moments $\widetilde{\mathbf{B}}$ and $\widetilde{\Xi}$ as a two-step random process:

1. First, a random process $\mathcal{R}$ generates the asymptotic centered moments $\mathbf{B}$ and Ξ.
2. Second, a random process $\widetilde{\mathcal{R}}$ generates the sample $\left\{\widetilde{\mathbf{x}}_1^b, \ldots, \widetilde{\mathbf{x}}_N^b\right\}$, whose statistical properties are consistent with the realization of $\mathbf{B}$ and Ξ obtained after the first random process $\mathcal{R}$. The sample centered moments $\widetilde{\mathbf{B}}$ and $\widetilde{\Xi}$ are then estimated from this sample.

Since they have very different natures, it seems reasonable to assume that these random processes are independent:

$$\mathcal{R} \perp\!\!\!\perp \widetilde{\mathcal{R}} \tag{82}$$

It is possible to take both processes into account in the sample centered moments statistics. Hereafter, the expectation symbol $\mathbb{E}\left[\cdot\right]$ thus applies to both processes together. In particular, equations (19), (24) and (32) becomes respectively:

$$\mathbb{E}\left[\widetilde{B}_{ij}\right] = \mathbb{E}\left[B_{ij}\right] \tag{83}$$

$$\begin{aligned}\mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] = P_1(N)\,\mathbb{E}\left[\Xi_{ijkl}\right] + P_2(N)\,\mathbb{E}\left[B_{ij}B_{kl}\right] \\ + P_3(N)\left(\mathbb{E}\left[B_{ik}B_{jl}\right] + \mathbb{E}\left[B_{il}B_{jk}\right]\right)\end{aligned} \tag{84}$$

and

$$\mathbb{E}\left[\widetilde{\Xi}_{ijkl}\right] = P_5(N)\,\mathbb{E}\left[\Xi_{ijkl}\right] + P_6(N)\left(\mathbb{E}\left[B_{ij}B_{kl}\right] + \mathbb{E}\left[B_{ik}B_{jl}\right] + \mathbb{E}\left[B_{il}B_{jk}\right]\right) \tag{85}$$

In the case of a Gaussian distributed ensemble, equation (27) becomes:

$$\mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] = \mathbb{E}\left[B_{ij}B_{kl}\right] + P_4(N)\left(\mathbb{E}\left[B_{ik}B_{jl}\right] + \mathbb{E}\left[B_{il}B_{jk}\right]\right) \tag{86}$$

As another consequence of the independence (82), the expected product of an asymptotic sample covariance and a covariance sampling error vanishes:

$$\mathbb{E}\left[\widetilde{B}^e_{ij}B_{kl}\right] = 0 \tag{87}$$

$$\Leftrightarrow \mathbb{E}\left[\widetilde{B}_{ij}B_{kl}\right] = \mathbb{E}\left[B_{ij}B_{kl}\right] \tag{88}$$

4.2. Linear relations between expected raw and asymptotic quantities

Noticing that the indices permutations of the sample fourth-order centered moment are all equal: $\widetilde{\widetilde{\Xi}}_{ijkl} = \widetilde{\widetilde{\Xi}}_{ikjl} = \widetilde{\widetilde{\Xi}}_{iljk}$, indices permutations of equation (84) and equation (85) lead to a linear system $\mathbf{A}_N \mathbf{x} = \widetilde{\mathbf{x}}$ with:

$$\mathbf{A}_N = \begin{pmatrix} P_2(N) & P_3(N) & P_3(N) & P_1(N) \\ P_3(N) & P_2(N) & P_3(N) & P_1(N) \\ P_3(N) & P_3(N) & P_2(N) & P_1(N) \\ P_6(N) & P_6(N) & P_6(N) & P_5(N) \end{pmatrix}, \ \mathbf{x} = \begin{pmatrix} \mathbb{E}\left[B_{ij}B_{kl}\right] \\ \mathbb{E}\left[B_{ik}B_{jl}\right] \\ \mathbb{E}\left[B_{il}B_{jk}\right] \\ \mathbb{E}\left[\Xi_{ijkl}\right] \end{pmatrix} \text{ and } \widetilde{\mathbf{x}} = \begin{pmatrix} \mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] \\ \mathbb{E}\left[\widetilde{B}_{ik}\widetilde{B}_{jl}\right] \\ \mathbb{E}\left[\widetilde{B}_{il}\widetilde{B}_{jk}\right] \\ \mathbb{E}\left[\widetilde{\widetilde{\Xi}}_{ijkl}\right] \end{pmatrix}$$

The matrix $\mathbf{A}_N$ can be inverted:

$$\mathbf{A}_N^{-1} = \begin{pmatrix} P_7(N) & P_8(N) & P_8(N) & P_9(N) \\ P_8(N) & P_7(N) & P_8(N) & P_9(N) \\ P_8(N) & P_8(N) & P_7(N) & P_9(N) \\ P_{10}(N) & P_{10}(N) & P_{10}(N) & P_{11}(N) \end{pmatrix} \tag{89}$$

with:

$$P_7(N) = \frac{2P_1(N)P_6(N) - P_5(N)\left(P_2(N) + P_3(N)\right)}{\left(P_2(N) - P_3(N)\right)\left(3P_1(N)P_6(N) - P_5(N)\left(P_2(N) + 2P_3(N)\right)\right)}$$

$$= \frac{(N-1)(N^2 - 3N + 1)}{N(N-2)(N-3)} \tag{90a}$$

$$P_8(N) = \frac{P_3(N)P_5(N) - P_1(N)P_6(N)}{\left(P_2(N) - P_3(N)\right)\left(3P_1(N)P_6(N) - P_5(N)\left(P_2(N) + 2P_3(N)\right)\right)}$$

$$= \frac{N-1}{N(N-2)(N-3)} \tag{90b}$$

$$P_9(N) = \frac{P_1(N)}{\left(3P_1(N)P_6(N) - P_5(N)\left(P_2(N) + 2P_3(N)\right)\right)}$$

$$= -\frac{N}{(N-2)(N-3)} \tag{90c}$$

$$P_{10}(N) = \frac{P_6(N)}{\left(3P_1(N)P_6(N) - P_5(N)\left(P_2(N) + 2P_3(N)\right)\right)}$$

$$= -\frac{(N-1)(2N-3)}{N(N-2)(N-3)} \tag{90d}$$

$$P_{11}(N) = -\frac{P_2(N) + 2P_3(N)}{\left(3P_1(N)P_6(N) - P_5(N)\left(P_2(N) + 2P_3(N)\right)\right)}$$

$$= \frac{N(N^2 - 2N + 3)}{(N-1)(N-2)(N-3)} \tag{90e}$$

to get $\mathbf{x} = \mathbf{A}_N^{-1}\widetilde{\mathbf{x}}$, which is equivalent to:

$$\mathbb{E}\left[B_{ij}B_{kl}\right] = P_7(N)\,\mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] + P_8(N)\left(\mathbb{E}\left[\widetilde{B}_{ik}\widetilde{B}_{jl}\right] + \mathbb{E}\left[\widetilde{B}_{il}\widetilde{B}_{jk}\right]\right) + P_9(N)\,\mathbb{E}\left[\widetilde{\Xi}_{ijkl}\right] \tag{91a}$$

$$\mathbb{E}\left[\Xi_{ijkl}\right] = P_{10}(N)\,\left(\mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] + \mathbb{E}\left[\widetilde{B}_{ik}\widetilde{B}_{jl}\right] + \mathbb{E}\left[\widetilde{B}_{il}\widetilde{B}_{jk}\right]\right) + P_{11}(N)\,\mathbb{E}\left[\widetilde{\Xi}_{ijkl}\right] \tag{91b}$$

In the case of a Gaussian distributed ensemble, the expected product of covariances (86) and its indices permutations lead to a linear system $\mathbf{A}_N\mathbf{x} = \widetilde{\mathbf{x}}$ with:

$$\mathbf{A}_N = \begin{pmatrix} 1 & P_4(N) & P_4(N) \\ P_4(N) & 1 & P_4(N) \\ P_4(N) & P_4(N) & 1 \end{pmatrix},\ \mathbf{x} = \begin{pmatrix} \mathbb{E}\left[B_{ij}B_{kl}\right] \\ \mathbb{E}\left[B_{ik}B_{jl}\right] \\ \mathbb{E}\left[B_{il}B_{jk}\right] \end{pmatrix} \text{ and } \widetilde{\mathbf{x}} = \begin{pmatrix} \mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] \\ \mathbb{E}\left[\widetilde{B}_{ik}\widetilde{B}_{jl}\right] \\ \mathbb{E}\left[\widetilde{B}_{il}\widetilde{B}_{jk}\right] \end{pmatrix}$$

The matrix $\mathbf{A}_N$ can be inverted:

$$\mathbf{A}_N^{-1} = \begin{pmatrix} P_{12}(N) & P_{13}(N) & P_{13}(N) \\ P_{13}(N) & P_{12}(N) & P_{13}(N) \\ P_{13}(N) & P_{13}(N) & P_{12}(N) \end{pmatrix} \tag{92}$$

with:

$$P_{12}(N) = \frac{1 + P_4(N)}{\Big(1 - P_4(N)\Big)\Big(1 + 2P_4(N)\Big)} = \frac{N(N-1)}{(N-2)(N+1)} \tag{93a}$$

$$P_{13}(N) = \frac{-P_4(N)}{\Big(1 - P_4(N)\Big)\Big(1 + 2P_4(N)\Big)} = -\frac{N-1}{(N-2)(N+1)} \tag{93b}$$

to get $\mathbf{x} = \mathbf{A}_N^{-1}\widetilde{\mathbf{x}}$, which is equivalent to:

$$\mathbb{E}\left[B_{ij}B_{kl}\right] = P_{12}(N)\ \mathbb{E}\left[\widetilde{B}_{ij}\widetilde{B}_{kl}\right] + P_{13}(N)\left(\mathbb{E}\left[\widetilde{B}_{ik}\widetilde{B}_{jl}\right] + \mathbb{E}\left[\widetilde{B}_{il}\widetilde{B}_{jk}\right]\right) \tag{94}$$

4.3. Particular case

For the rest of this document, we only deal with the subcase where $k = i$ and $l = j$.

Denoting $\widetilde{\boldsymbol{\xi}} \in \mathbb{R}^{n\times n}$ the matrix such that $\widetilde{\xi}_{ij} = \widetilde{\Xi}_{ijij}$, equations (84) and (85) become:

$$\mathbb{E}\left[\widetilde{B}_{ij}^2\right] = P_1(N)\mathbb{E}\left[\xi_{ij}\right] + P_{14}(N)\mathbb{E}\left[B_{ij}^2\right] + P_3(N)\mathbb{E}\left[B_{ii}B_{jj}\right] \tag{95a}$$

$$\mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right] = P_1(N)\mathbb{E}\left[\xi_{ij}\right] + 2P_3(N)\mathbb{E}\left[B_{ij}^2\right] + P_2(N)\mathbb{E}\left[B_{ii}B_{jj}\right] \tag{95b}$$

$$\mathbb{E}\left[\widetilde{\xi}_{ij}\right] = P_5(N)\mathbb{E}\left[\xi_{ij}\right] + P_6(N)\left(2\mathbb{E}\left[B_{ij}^2\right] + \mathbb{E}\left[B_{ii}B_{jj}\right]\right) \tag{95c}$$

with:

$$P_{14}(N) = P_2(N) + P_3(N) = \frac{N^2 - 2N + 2}{N(N-1)} \tag{96}$$

Reciprocally, equations (91a) and (91b) becomes:

$$\mathbb{E}\left[B_{ij}^2\right] = P_{15}(N)\mathbb{E}\left[\widetilde{B}_{ij}^2\right] + P_8(N)\mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right] + P_9(N)\mathbb{E}\left[\widetilde{\xi}_{ij}\right] \tag{97a}$$

$$\mathbb{E}\left[B_{ii}B_{jj}\right] = 2P_8(N)\mathbb{E}\left[\widetilde{B}_{ij}^2\right] + P_7(N)\mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right] + P_9(N)\mathbb{E}\left[\widetilde{\xi}_{ij}\right] \tag{97b}$$

$$\mathbb{E}\left[\xi_{ij}\right] = P_{10}(N)\left(2\mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right]\right) + P_{11}(N)\ \mathbb{E}\left[\widetilde{\xi}_{ij}\right] \tag{97c}$$

with:

$$P_{15}(N) = P_7(N) + P_8(N) = \frac{(N-1)^2}{N(N-3)} \tag{98}$$

In the case of a Gaussian distributed ensemble, equation (86) is used instead:

$$\mathbb{E}\left[\widetilde{B}_{ij}^2\right] = P_{16}(N)\mathbb{E}\left[B_{ij}^2\right] + P_4(N)\mathbb{E}\left[B_{ii}B_{jj}\right] \tag{99a}$$

$$\mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right] = 2P_4(N)\mathbb{E}\left[B_{ij}^2\right] + \mathbb{E}\left[B_{ii}B_{jj}\right] \tag{99b}$$

with:

$$P_{16}(N) = 1 + P_4(N) = \frac{N}{N-1} \tag{100}$$

Reciprocally, equation (94) becomes:

$$\mathbb{E}\left[B_{ij}^2\right] = P_{17}(N)\mathbb{E}\left[\widetilde{B}_{ij}^2\right] + P_{13}(N)\mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right] \tag{101a}$$

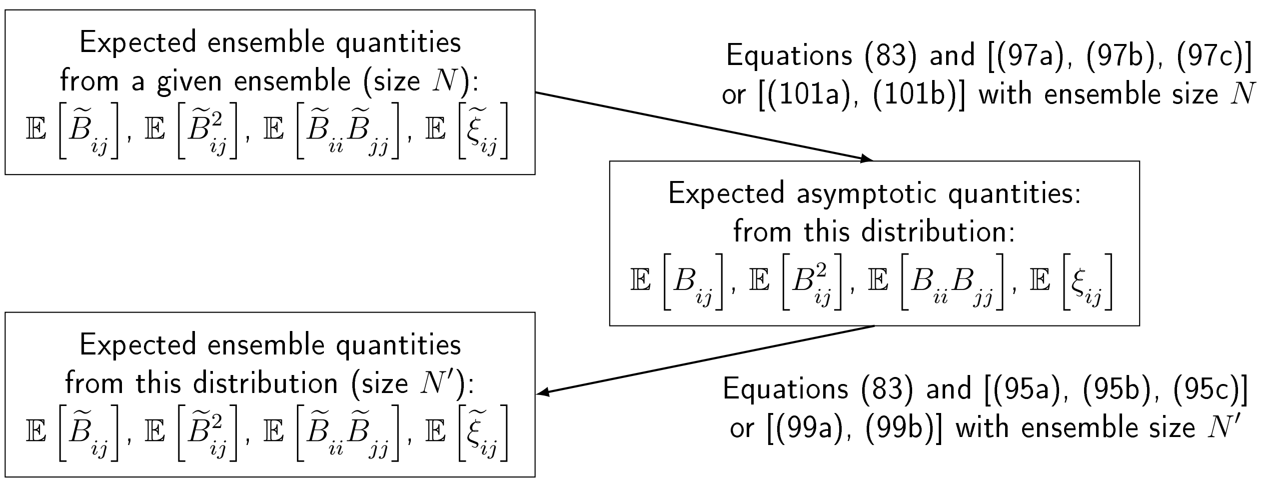
$$\mathbb{E}\left[B_{ii}B_{jj}\right] = 2P_{13}(N)\mathbb{E}\left[\widetilde{B}_{ij}^2\right] + P_{12}(N)\mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right] \tag{101b}$$

with:

$$P_{17}(N) = P_{12}(N) + P_{13}(N) = \frac{(N-1)^2}{(N-2)(N+1)} \tag{102}$$

4.4. Diagnostics for any ensemble size

Expected ensemble quantities can be estimated from an ensemble of size N, allowing the estimation of expected asymptotic quantities. Interestingly, these quantities can be used to estimate expected raw quantities for an hypothetic ensemble with the same distribution but of a different size N'.



4.5. Multiple sub-ensembles with distinct distributions

The computation of diagnostics for any ensemble size (previous section) need to be adapted in the framework of multiple sub-ensembles. For sake of simplicity, we consider the case of R sub-ensembles of equal sizes $N_r = N/R$ and we separate the procedure into two steps:

1. Starting from estimations of $\widetilde{\mathbf{B}}_r$ and $\widetilde{\boldsymbol{\xi}}_r$ for each sub-ensemble, we get from equations (64) and (66):

$$\mathbb{E}\left[\widetilde{B}^c_{ij}\right] = \frac{1}{R}\sum_{r=1}^{R}\mathbb{E}\left[\widetilde{B}_{ij,r}\right] \tag{103a}$$

$$\mathbb{E}\left[\widetilde{B}^{c2}_{ij}\right] = \frac{1}{R^2}\sum_{1\le r,s\le R}\mathbb{E}\left[\widetilde{B}_{ij,r}\widetilde{B}_{ij,s}\right] \tag{103b}$$

$$\mathbb{E}\left[\widetilde{B}^c_{ii}\widetilde{B}^c_{jj}\right] = \frac{1}{R^2}\sum_{1\le r,s\le R}\mathbb{E}\left[\widetilde{B}_{ii,r}\widetilde{B}_{jj,s}\right] \tag{103c}$$

$$\mathbb{E}\left[\widetilde{\xi}_{ij}\right] = \frac{1}{R}\sum_{r=1}^{R}\mathbb{E}\left[\widetilde{\xi}_{ij,r}\right] \tag{103d}$$

2. From this sampled quantities, we need to estimate asymptotic quantities. First equation is easy:

$$\mathbb{E}\left[B^c_{ij}\right] = \mathbb{E}\left[\widetilde{B}^c_{ij}\right] \tag{104}$$

   Using the independence (82), the expected product of aymptotic and sample covariances can be expressed as:

$$\begin{aligned}\mathbb{E}\left[B^c_{ij}\widetilde{B}^c_{kl}\right] &= \frac{1}{R^2}\sum_{1\le r,s\le R}\mathbb{E}\left[B_{ij,r}\widetilde{B}_{kl,s}\right]\\ &= \frac{1}{R^2}\left(\sum_{r=1}^{R}\mathbb{E}\left[B_{ij,r}B_{kl,r}\right] + \sum_{\substack{1\le r,s\le R\\ r\neq s}}\mathbb{E}\left[B_{ij,r}\widetilde{B}_{kl,s}\right]\right)\end{aligned} \tag{105}$$

   The first term can be computed using equation (91a) for each sub-ensemble. However, an additional assumption is required to estimate the second term. A possible one is to assume that:

   a) the sampling error $\widetilde{B}_{ij,r} - B_{ij,r}$ of sub-ensemble r is not correlated with the sampling error $\widetilde{B}_{kl,s} - B_{kl,s}$ of sub-ensemble s:

$$\mathbb{E}\left[\left(\widetilde{B}_{ij,r} - B_{ij,r}\right)\left(\widetilde{B}_{kl,s} - B_{kl,s}\right)\right] \approx \mathbb{E}\left[\widetilde{B}_{ij,r} - B_{ij,r}\right]\mathbb{E}\left[\widetilde{B}_{kl,s} - B_{kl,s}\right] = 0 \tag{106}$$

   b) the sampling error $\widetilde{B}_{ij,r} - B_{ij,r}$ of sub-ensemble r is not correlated with the asymptotic covariance matrix $B_{kl,s}$ of sub-ensemble s:

$$\mathbb{E}\left[\left(\widetilde{B}_{ij,r} - B_{ij,r}\right)B_{kl,s}\right] \approx \mathbb{E}\left[\widetilde{B}_{ij,r} - B_{ij,r}\right]\mathbb{E}\left[B_{kl,s}\right] = 0 \tag{107}$$

Thus, we obtain two interesting properties:

a) $\mathbb{E}\left[B_{ij,r}\widetilde{B}_{kl,s}\right] \approx \mathbb{E}\left[B_{ij,r}B_{kl,s}\right]$, so that $\mathbb{E}\left[B^c_{ij}\widetilde{B}^c_{kl}\right] \approx \mathbb{E}\left[B^c_{ij}B^c_{kl}\right]$, similar to equation (88),

b) $\mathbb{E}\left[B_{ij,r}\widetilde{B}_{kl,s}\right] \approx \mathbb{E}\left[\widetilde{B}_{ij,r}\widetilde{B}_{kl,s}\right]$, useful to estimate:

$$\mathbb{E}\left[B^c_{ij}\widetilde{B}^c_{kl}\right] = \frac{1}{R^2}\left(\sum_{r=1}^{R}\mathbb{E}\left[B_{ij,r}B_{kl,r}\right] + \sum_{\substack{1\leq r,s\leq R\\ r\neq s}}\mathbb{E}\left[\widetilde{B}_{ij,r}\widetilde{B}_{kl,s}\right]\right) \tag{108}$$

As a consequence, we get the second and third equations:

$$\mathbb{E}\left[B^{c2}_{ij}\right] = \frac{1}{R^2}\left(\sum_{r=1}^{R}\mathbb{E}\left[B^2_{ij,r}\right] + \sum_{\substack{1\leq r,s\leq R\\ r\neq s}}\mathbb{E}\left[\widetilde{B}_{ij,r}\widetilde{B}_{ij,s}\right]\right) \tag{109a}$$

$$\mathbb{E}\left[B^c_{ii}B^c_{jj}\right] = \frac{1}{R^2}\left(\sum_{r=1}^{R}\mathbb{E}\left[B_{ii,r}B_{jj,r}\right] + \sum_{\substack{1\leq r,s\leq R\\ r\neq s}}\mathbb{E}\left[\widetilde{B}_{ii,r}\widetilde{B}_{jj,s}\right]\right) \tag{109b}$$

where $\mathbb{E}\left[B^2_{ij,r}\right]$ and $\mathbb{E}\left[B_{ii,r}B_{jj,r}\right]$ are computed using equations [(97a), (97b)] or [(101a), (101b)] for each sub-ensemble. The last equation is simply:

$$\mathbb{E}\left[\xi^c_{ij}\right] = \frac{1}{R}\sum_{r=1}^{R}\mathbb{E}\left[\xi_{ij,r}\right] \tag{110}$$

where $\mathbb{E}\left[\xi_{ij,r}\right]$ is computed using equation (97c) for each sub-ensemble.

5. Filtering optimization

5.1. Optimization norms

We define $\widehat{\mathbf{B}}(\boldsymbol{\lambda})$ as the filtered estimate of $\widetilde{\mathbf{B}}$, where $\boldsymbol{\lambda} \in \mathbb{R}^M$ is a vector of parameters and $\widehat{\mathbf{B}}(\boldsymbol{\lambda})$ is a linear function of $\boldsymbol{\lambda}$.

To optimize $\boldsymbol{\lambda}$, we need to define a norm that should be minimized. A general choice is to use the expected squared Frobenius norm of the filtering error, projected into a subspace defined by the projection matrix $\mathbf{Q} \in \mathbb{R}^{n \times n'}$:

$$e(\boldsymbol{\lambda}) = \mathbb{E}\left[\left\|\mathbf{Q}^{\mathrm{T}}\left(\widehat{\mathbf{B}}(\boldsymbol{\lambda}) - \mathbf{B}\right)\mathbf{Q}\right\|_{\mathrm{F}}^2\right] \tag{111}$$

The asymptotic sample covariance $\mathbf{B}$ can be considered as the target of the filtering. The definition of the projection matrix $\mathbf{Q}$ depends on the considered system.

Hereafter, we will focus on a particular case where $\mathbf{Q}$ is square ($n' = n$) and diagonal, leading to:

$$\begin{aligned} e(\boldsymbol{\lambda}) &= \mathbb{E}\left[\left\|\mathbf{W} \circ \left(\widehat{\mathbf{B}}(\boldsymbol{\lambda}) - \mathbf{B}\right)\right\|_{\mathrm{F}}^2\right] \\ &= \mathbb{E}\left[\sum_{1 \le i,j \le n} \left(W_{ij}\left(\widehat{B}_{ij}(\boldsymbol{\lambda}) - B_{ij}\right)\right)^2\right] \\ &= \sum_{1 \le i,j \le n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}(\boldsymbol{\lambda}) - B_{ij}\right)^2\right] \end{aligned} \tag{112}$$

where $W_{ij} = Q_{ii} Q_{jj}$ defines a weight matrix applied via a Schur product (element-by-element) denoted $\circ$. The definition of the $\mathbf{W}$ is an open question, but $\mathbf{W}$ should not depend on the ensemble since it is extracted from the expectation $\mathbb{E}[\cdot]$.

5.2. Linear filtering optimization

The gradient of $e(\boldsymbol{\lambda})$ with respect to λ_m is given by:

$$\frac{\partial e}{\partial \lambda_m} = 2 \sum_{1 \le i,j \le n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}(\boldsymbol{\lambda}) - B_{ij}\right)\frac{\partial \widehat{B}_{ij}}{\partial \lambda_m}\right] \tag{113}$$

We can define the subset $\mathcal{S}^{\mathrm{opt}} \subset \mathbf{R}^M$ of the vectors $\boldsymbol{\lambda}^{\mathrm{opt}}$ canceling the gradient (113):

$$\forall m \in [1, M] \ , \ \left.\frac{\partial e}{\partial \lambda_m}\right|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\mathrm{opt}}} = 0 \tag{114}$$

Since $\widehat{\mathbf{B}}(\boldsymbol{\lambda})$ is a linear function of $\boldsymbol{\lambda}$, $e(\boldsymbol{\lambda})$ is a quadratic and nonnegative function of $\boldsymbol{\lambda}$. Thus any vector $\boldsymbol{\lambda}^{\mathrm{opt}}$ in $\mathcal{S}^{\mathrm{opt}}$ is minimizing $e(\boldsymbol{\lambda})$, and is an equivalently relevant solution for our problem.

5.3. Norm reduction

An interesting property of the linear filtering is the possibility to compute the optimal norm reduction. Since the filtering is linear:

$$\widehat{B}_{ij}\left(\boldsymbol{\lambda}\right) = \sum_{1\leq m\leq M} \frac{\partial \widehat{B}_{ij}}{\partial \lambda_m} \lambda_m \tag{115}$$

Equations (113) and (114) can be combined into:

$$\forall m \in [1, M]\ ,\ \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right) \left.\frac{\partial \widehat{B}_{ij}}{\partial \lambda_m}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{\mathrm{opt}}}\right] = 0 \tag{116}$$

The sum over $[1, M]$ of the previous equation, post-multiplied by λ_m^{opt}, yields:

$$\begin{aligned}
&\sum_{1\leq m\leq M} \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right) \left.\frac{\partial \widehat{B}_{ij}}{\partial \lambda_m}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{\mathrm{opt}}}\right] \lambda_m^{\mathrm{opt}} = 0 \\
\Leftrightarrow\ &\sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right) \sum_{1\leq m\leq M} \left.\frac{\partial \widehat{B}_{ij}}{\partial \lambda_m}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{\mathrm{opt}}} \lambda_m^{\mathrm{opt}}\right] = 0 \\
\Leftrightarrow\ &\sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right) \widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right)\right] = 0
\end{aligned} \tag{117}$$

We define $\boldsymbol{\lambda}^0$ the vector of parameters such as no filtering is applied: $\widehat{\mathbf{B}}\left(\boldsymbol{\lambda}^0\right) = \widetilde{\mathbf{B}}$. Using the previous property, we can derive:

$$\begin{aligned}
e\left(\boldsymbol{\lambda}^0\right) &= \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widetilde{B}_{ij} - B_{ij}\right)^2\right] \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right) - \left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - \widetilde{B}_{ij}\right)\right)^2\right] \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right)^2\right] + \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - \widetilde{B}_{ij}\right)^2\right] \\
&\quad - 2 \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right)\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - \widetilde{B}_{ij}\right)\right]
\end{aligned} \tag{118}$$

Reducing the last term with equation (117), we get:

$$e\left(\boldsymbol{\lambda}^0\right) - e\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) = \Delta e + 2 \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right) \widetilde{B}_{ij}\right] \tag{119}$$

where:

$$\Delta e = \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - \widetilde{B}_{ij}\right)^2\right] \tag{120}$$

The difference $e\left(\boldsymbol{\lambda}^0\right) - e\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right)$ is the norm reduction due to the filtering. If we can show that the last term of equation (119) vanishes, then Δe can be computed once $\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right)$ is known to estimate this norm reduction. Actually, an even more general condition given by $\mathbb{E}\left[\left(\widehat{B}_{ij}\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right) - B_{ij}\right)\widetilde{B}_{ij}\right] = 0$ will be verified in the following filtering methods.

Also:

$$\begin{aligned}
e\left(\boldsymbol{\lambda}^0\right) &= \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widetilde{B}_{ij} - B_{ij}\right)^2\right] \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2 \left(\mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \mathbb{E}\left[B_{ij}^2\right] - 2\mathbb{E}\left[\widetilde{B}_{ij} B_{ij}\right]\right) \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2 \left(\mathbb{E}\left[\widetilde{B}_{ij}^2\right] - \mathbb{E}\left[B_{ij}^2\right]\right) \qquad (121)
\end{aligned}$$

which can be used to compute a relative norm reduction $\dfrac{e\left(\boldsymbol{\lambda}^0\right) - e\left(\boldsymbol{\lambda}^{\mathrm{opt}}\right)}{e\left(\boldsymbol{\lambda}^0\right)}$.

6. Variance filtering

6.1. Filtering method

Variances are gathered in vectors $\mathbf{v} \in \mathbb{R}^n$:

$$v_i = B_{ii} \tag{122a}$$

$$\widetilde{v}_i = \widetilde{B}_{ii} \tag{122b}$$

$$\widehat{v}_i = \widehat{B}_{ii} \tag{122c}$$

To filter the sample variances $\widetilde{\mathbf{v}}$, we apply a linear filter (gain and offset). Thus $\widehat{\mathbf{v}}$ is defined as:

$$\widehat{\mathbf{v}} = \mathbf{F}\widetilde{\mathbf{v}} + \mathbf{f} \tag{123}$$

where $\mathbf{F} \in \mathbb{R}^{n\times n}$ is the filtering gain and $\mathbf{f} \in \mathbb{R}^n$ the filter offset.

6.2. Expected squared norm

We use the identity matrix as a weight: $\mathbf{W} = \mathbf{I}_n$. The expected squared norm of equation (112) adapted for the variance filtering becomes:

$$\begin{aligned}
e(\mathbf{F},\mathbf{f}) &= \sum_{1\leq i\leq n} \mathbb{E}\left[\left(\sum_{1\leq j\leq n} F_{ij}\widetilde{v}_j + f_i - v_i\right)^2\right] \\
&= \sum_{1\leq i\leq n} \left(\sum_{1\leq j,k\leq n} F_{ij}F_{ik}\mathbb{E}\left[\widetilde{v}_j\widetilde{v}_k\right] + f_i^2 + \mathbb{E}\left[v_i^2\right]\right. \\
&\qquad \left. +2f_i\sum_{1\leq j\leq n} F_{ij}\mathbb{E}\left[\widetilde{v}_j\right] - 2\sum_{1\leq j\leq n} F_{ij}\mathbb{E}\left[v_i\widetilde{v}_j\right] - 2f_i\mathbb{E}\left[v_i\right]\right)
\end{aligned} \tag{124}$$

Using the independence of random processes in (88), it can be simplified:

$$\begin{aligned}
e(\mathbf{F},\mathbf{f}) = \sum_{1\leq i\leq n} &\left(\sum_{1\leq j,k\leq n} F_{ij}F_{ik}\mathbb{E}\left[\widetilde{v}_j\widetilde{v}_k\right] + f_i^2 + \mathbb{E}\left[v_i^2\right]\right. \\
&\left. + 2f_i\left(\sum_{1\leq j\leq n} F_{ij}\mathbb{E}\left[\widetilde{v}_j\right] - \mathbb{E}\left[\widetilde{v}_i\right]\right) - 2\sum_{1\leq j\leq n} F_{ij}\mathbb{E}\left[v_iv_j\right]\right)
\end{aligned} \tag{125}$$

Its gradient is given by:

$$\frac{\partial e}{\partial F_{ij}} = 2\left(\sum_{1\leq k\leq n} F_{ik}\mathbb{E}\left[\widetilde{v}_j\widetilde{v}_k\right] + f_i\mathbb{E}\left[\widetilde{v}_j\right] - \mathbb{E}\left[v_iv_j\right]\right) \tag{126a}$$

$$\frac{\partial e}{\partial f_i} = 2\left(f_i + \sum_{1\leq j\leq n} F_{ij}\mathbb{E}\left[\widetilde{v}_j\right] - \mathbb{E}\left[\widetilde{v}_i\right]\right) \tag{126b}$$

6.3. Explicit optimality

Setting the gradient of $e(\mathbf{F},\mathbf{f})$ to zero, we get in a vectorial form:

$$\begin{aligned}
&2\left(\mathbf{F}^{\mathrm{opt}}\mathbb{E}\left[\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^{\mathrm{T}}\right]+\mathbf{f}^{\mathrm{opt}}\mathbb{E}\left[\widetilde{\mathbf{v}}^{\mathrm{T}}\right]-\mathbb{E}\left[\mathbf{v}\mathbf{v}^{\mathrm{T}}\right]\right)=0\\
\Leftrightarrow\ &\mathbf{F}^{\mathrm{opt}}\mathbb{E}\left[\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^{\mathrm{T}}\right]+\mathbf{f}^{\mathrm{opt}}\mathbb{E}\left[\widetilde{\mathbf{v}}^{\mathrm{T}}\right]=\mathbb{E}\left[\mathbf{v}\mathbf{v}^{\mathrm{T}}\right]
\end{aligned}\tag{127}$$

and:

$$\begin{aligned}
&2\left(\mathbf{F}^{\mathrm{opt}}\mathbb{E}\left[\widetilde{\mathbf{v}}\right]+\mathbf{f}^{\mathrm{opt}}-\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\right)=0\\
\Leftrightarrow\ &\mathbf{F}^{\mathrm{opt}}\mathbb{E}\left[\widetilde{\mathbf{v}}\right]+\mathbf{f}^{\mathrm{opt}}=\mathbb{E}\left[\widetilde{\mathbf{v}}\right]
\end{aligned}\tag{128}$$

From equation (128), we get the filter offset:

$$\mathbf{f}^{\mathrm{opt}}=\left(\mathbf{I}_n-\mathbf{F}^{\mathrm{opt}}\right)\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\tag{129}$$

so that equation (127) reads:

$$\begin{aligned}
&\mathbf{F}^{\mathrm{opt}}\mathbb{E}\left[\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^{\mathrm{T}}\right]+\left(\mathbf{I}_n-\mathbf{F}^{\mathrm{opt}}\right)\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\mathbb{E}\left[\widetilde{\mathbf{v}}^{\mathrm{T}}\right]=\mathbb{E}\left[\mathbf{v}\mathbf{v}^{\mathrm{T}}\right]\\
\Leftrightarrow\ &\mathbf{F}^{\mathrm{opt}}\left(\mathbb{E}\left[\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^{\mathrm{T}}\right]-\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\mathbb{E}\left[\widetilde{\mathbf{v}}^{\mathrm{T}}\right]\right)=\mathbb{E}\left[\mathbf{v}\mathbf{v}^{\mathrm{T}}\right]-\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\mathbb{E}\left[\widetilde{\mathbf{v}}^{\mathrm{T}}\right]\\
\Leftrightarrow\ &\mathbf{F}^{\mathrm{opt}}=\left(\mathbb{E}\left[\mathbf{v}\mathbf{v}^{\mathrm{T}}\right]-\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\mathbb{E}\left[\widetilde{\mathbf{v}}^{\mathrm{T}}\right]\right)\left(\mathbb{E}\left[\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^{\mathrm{T}}\right]-\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\mathbb{E}\left[\widetilde{\mathbf{v}}^{\mathrm{T}}\right]\right)^{-1}
\end{aligned}\tag{130}$$

Using the independence of random processes in (88), $\mathbb{E}\left[\widetilde{\mathbf{v}}\right]=\mathbb{E}\left[\mathbf{v}\right]$ so that:

$$\begin{aligned}
\mathbf{F}^{\mathrm{opt}}&=\left(\mathbb{E}\left[\mathbf{v}\mathbf{v}^{\mathrm{T}}\right]-\mathbb{E}\left[\mathbf{v}\right]\mathbb{E}\left[\mathbf{v}^{\mathrm{T}}\right]\right)\left(\mathbb{E}\left[\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^{\mathrm{T}}\right]-\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\mathbb{E}\left[\widetilde{\mathbf{v}}^{\mathrm{T}}\right]\right)^{-1}\\
&=\mathrm{Cov}\left(\mathbf{v}\right)\mathrm{Cov}\left(\widetilde{\mathbf{v}}\right)^{-1}
\end{aligned}\tag{131}$$

6.4. Properties

Equations (127) and (128) can be rewritten using the optimally filtered variances $\widehat{\mathbf{v}}^{\mathrm{opt}}=\mathbf{F}^{\mathrm{opt}}\widetilde{\mathbf{v}}+\mathbf{f}^{\mathrm{opt}}$:

$$\mathbb{E}\left[\widehat{\mathbf{v}}^{\mathrm{opt}}\widetilde{\mathbf{v}}^{\mathrm{T}}\right]=\mathbb{E}\left[\mathbf{v}\mathbf{v}^{\mathrm{T}}\right]\tag{132a}$$

$$\mathbb{E}\left[\widehat{\mathbf{v}}^{\mathrm{opt}}\right]=\mathbb{E}\left[\widetilde{\mathbf{v}}\right]\tag{132b}$$

6.5. Norm reduction

Using the independence of random processes in (88) and the optimality condition of equation (132a), we verify that:

$$\mathbb{E}\left[\left(\widehat{v}_i^{\mathrm{opt}}-v_i\right)\widetilde{v}_i\right]=\mathbb{E}\left[\widehat{v}_i^{\mathrm{opt}}\widetilde{v}_i\right]-\mathbb{E}\left[v_iv_i\right]=0\tag{133}$$

so that:

$$e\left(\mathbf{I}_n,\mathbf{0}\right)-e\left(\mathbf{F}^{\mathrm{opt}},\mathbf{f}^{\mathrm{opt}}\right)=\sum_{1\leq i\leq n}\mathbb{E}\left[\left(\widehat{v}_i^{\mathrm{opt}}-\widetilde{v}_i\right)^2\right]\tag{134}$$

7. Localization

7.1. Filtering method

To localize the sample covariances, we apply a localization matrix via a Schur product (element-by-element). Thus $\widehat{\mathbf{B}}$ is defined as:

$$\widehat{\mathbf{B}} = \mathbf{L} \circ \widetilde{\mathbf{B}} \tag{135}$$

where $\mathbf{L} \in \mathbb{R}^{n \times n}$ is the localization matrix.

7.2. Expected squared norm

The expected squared norm of equation (112) adapted for the localization becomes:

$$\begin{aligned} e(\mathbf{L}) &= \sum_{1 \leq i,j \leq n} W_{ij}^2 \mathbb{E}\left[\left(L_{ij}\widetilde{B}_{ij} - B_{ij}\right)^2\right] \\ &= \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(L_{ij}^2 \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \mathbb{E}\left[B_{ij}^2\right] - 2L_{ij}\mathbb{E}\left[\widetilde{B}_{ij}B_{ij}\right]\right) \end{aligned} \tag{136}$$

Using the independence of random processes in (88), it can be simplified:

$$e(\mathbf{L}) = \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(L_{ij}^2 \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \left(1 - 2L_{ij}\right)\mathbb{E}\left[B_{ij}^2\right]\right) \tag{137}$$

Its gradient is given by:

$$\frac{\partial e}{\partial L_{ij}} = 2W_{ij}^2 \left(L_{ij}\mathbb{E}\left[\widetilde{B}_{ij}^2\right] - \mathbb{E}\left[B_{ij}^2\right]\right) \tag{138}$$

7.3. Explicit optimality

Setting the gradient of $e(\mathbf{L})$ to zero, we get:

$$2W_{ij}^2 \left(L_{ij}^{\text{opt}}\mathbb{E}\left[\widetilde{B}_{ij}^2\right] - \mathbb{E}\left[B_{ij}^2\right]\right) = 0 \;\Leftrightarrow\; L_{ij}^{\text{opt}} = \frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} \tag{139}$$

We can notice that the weight matrix $\mathbf{W}$ does not affect $\mathbf{L}^{\text{opt}}$.

7.4. Properties

Using the independence of random processes in (88), we can notice that:

$$L_{ij}^{\text{opt}} = \frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\left(B_{ij} + \widetilde{B}_{ij}^e\right)^2\right]} = \frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[B_{ij}^2\right] + \mathbb{E}\left[\widetilde{B}_{ij}^{e2}\right]} \tag{140}$$

which shows three important properties:

- The optimal localization is bounded: $0 \leq L^{\text{opt}}_{ij} \leq 1$.
- For a finite ensemble size, sampling noise arises ($\mathbb{E}\left[\widetilde{B}^{e2}_{ii}\right] > 0$) so that L^{opt}_{ii} (at zero separation) is strictly lower than 1, contrary to what can be usually found in the literature.
- For an infinite ensemble size ($N \rightarrow \infty$), the sampling noise vanishes ($\mathbb{E}\left[\widetilde{B}^{e2}_{ij}\right] \rightarrow 0$) so that $L^{\text{opt}}_{ij} \rightarrow 1$: no localization is applied.

7.5. Norm reduction

Using the independence of random processes in (88) and the optimality condition of equation (139), we verify that:

$$\mathbb{E}\left[\left(\widehat{B}^{\text{opt}}_{ij} - B_{ij}\right)\widetilde{B}_{ij}\right] = L^{\text{opt}}_{ij}\mathbb{E}\left[\widetilde{B}^2_{ij}\right] - \mathbb{E}\left[B^2_{ij}\right] = 0 \tag{141}$$

so that:

$$e\left(\mathbf{1}\right) - e\left(\mathbf{L}^{\text{opt}}\right) = \sum_{1\leq i,j\leq n} W^2_{ij}\mathbb{E}\left[\left(\widehat{B}^{\text{opt}}_{ij} - \widetilde{B}_{ij}\right)^2\right] \tag{142}$$

8. Links between localization and correlation

8.1. Approximations of the optimal localization

The sample correlation matrix $\widetilde{\mathbf{C}}$ is defined as:

$$\widetilde{C}_{ij} = \frac{\widetilde{B}_{ij}}{\sqrt{\widetilde{B}_{ii}\widetilde{B}_{jj}}} \tag{143}$$

and the asymptotic limit of $\widetilde{\mathbf{C}}$ is denoted $\mathbf{C}$.

In the case of a Gaussian distributed ensemble, we can get two different approximations of the localization:

1. In a simplified modeling of the asymptotic sample covariance matrix $\mathbf{B}$, we can assume that the asymptotic sample correlation matrix $\mathbf{C}$ is exactly the same at every realization of its generating random process $\mathcal{R}$, with no variability. As a consequence, it can be extracted from expectations:

   $$\mathbb{E}\left[B_{ij}^2\right] = \mathbb{E}\left[B_{ii}B_{jj}C_{ij}^2\right] = \mathbb{E}\left[B_{ii}B_{jj}\right] C_{ij}^2 \tag{144}$$

   Equation (86) gives:

   $$\begin{aligned}\mathbb{E}\left[\widetilde{B}_{ij}^2\right] &= \Big(1 + P_4(N)\Big)\ \mathbb{E}\left[B_{ij}^2\right] + P_4(N)\ \mathbb{E}\left[B_{ii}B_{jj}\right] \\ &= \Big(\Big(1 + P_4(N)\Big) C_{ij}^2 + P_4(N)\Big)\ \mathbb{E}\left[B_{ii}B_{jj}\right]\end{aligned} \tag{145}$$

   so that the optimal localization of equation (139) is given by:

   $$\begin{aligned}L_{ij}^{\text{app1}} &= \frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} \\ &= \frac{\mathbb{E}\left[B_{ii}B_{jj}\right] C_{ij}^2}{\Big(\Big(1 + P_4(N)\Big) C_{ij}^2 + P_4(N)\Big)\ \mathbb{E}\left[B_{ii}B_{jj}\right]} \\ &= \frac{C_{ij}^2}{\Big(1 + P_4(N)\Big) C_{ij}^2 + P_4(N)}\end{aligned} \tag{146}$$

   if we assume (reasonably) that $\mathbb{E}\left[B_{ii}B_{jj}\right] \neq 0$.

2. A first order approximation of $\mathbb{E}\left[\widetilde{C}_{ij}^2\right]$ is given by:

   $$\mathbb{E}\left[\widetilde{C}_{ij}^2\right] = \mathbb{E}\left[\frac{\widetilde{B}_{ij}^2}{\widetilde{B}_{ii}\widetilde{B}_{jj}}\right] \approx \frac{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right]} \tag{147}$$

Plugging this result into equation (101a), we get:

$$\mathbb{E}\left[B_{ij}^2\right] \approx P_{17}(N)\ \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + P_{13}(N)\frac{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}{\mathbb{E}\left[\widetilde{C}_{ij}^2\right]} \tag{148}$$

Thus, the optimal localization $\mathbf{L}^{\text{opt}}$ of equation (139) can be approximated by:

$$L_{ij}^{\text{app2}} = P_{17}(N) + P_{13}(N)\frac{1}{\mathbb{E}\left[\widetilde{C}_{ij}^2\right]} \tag{149}$$

For a Gaussian distributed ensemble, Rady *et al.* (2005) proves that:

$$\begin{aligned}\mathbb{E}\left[\widetilde{C}_{ij}^2\right] = C_{ij}^2 & \\ & + \frac{2C_{ij}^4 - 3C_{ij}^2 + 1}{N-1} \\ & + \frac{8C_{ij}^6 - 14C_{ij}^4 + 6C_{ij}^2}{(N-1)^2} \\ & + \frac{48C_{ij}^8 - 104C_{ij}^6 + 68C_{ij}^4 - 12C_{ij}^2}{(N-1)^3} \\ & + O\left(\frac{1}{(N-1)^4}\right)\end{aligned} \tag{150}$$

This result can be used to compare $\mathbf{L}^{\text{app1}}$ and $\mathbf{L}^{\text{app2}}$. Figure 1 shows that the second approximation gives a more severe localization (for a given asymptotic sample correlation C_{ij}, $L_{ij}^{\text{app2}} \leq L_{ij}^{\text{app1}}$). Interetingly, both approximation seems to converge for $N \rightarrow \infty$.

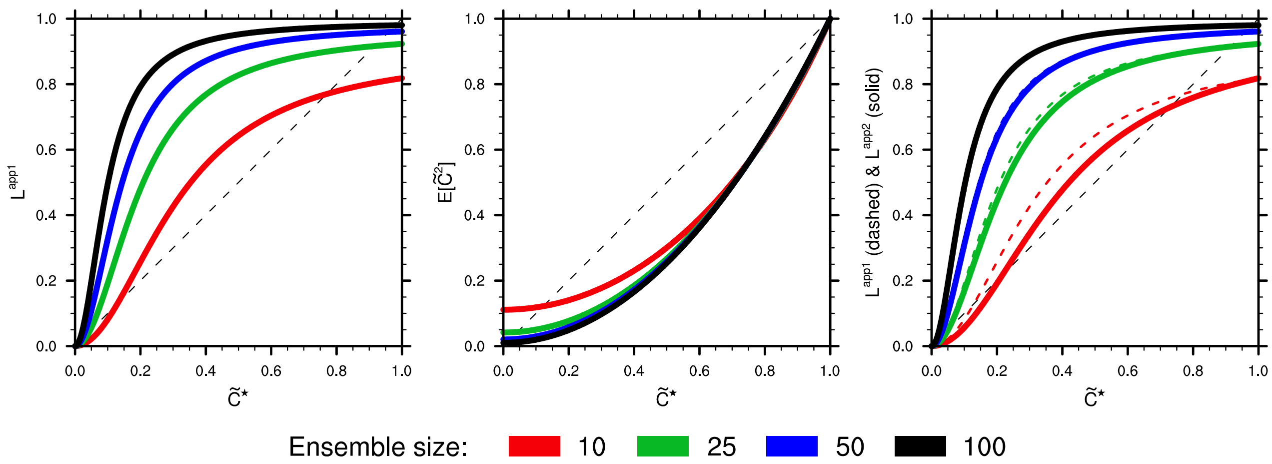


Figure 1

8.2. Local tensor definition

The local tensor is a matrix used to define the curvature of a function at its origin. Let $\mathbf{r} = \left(x_1, \ldots, x_d\right)^{\mathrm{T}} \in \mathbb{R}^d$ be a coordinate vector in a d-dimensional space. The local correlation tensor $\mathbf{H}^C \in \mathbb{R}^{d\times d}$ associated with the correlation function $C(\mathbf{r})$, assumed to be twice differentiable, is given by:

$$\mathbf{H}^C = -\nabla\nabla^{\mathrm{T}} C\big|_{\mathbf{r}=\mathbf{0}} \;\Leftrightarrow\; H^C_{\alpha,\beta} = -\left.\frac{\partial^2 C}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} \tag{151}$$

For a localization function $L(\mathbf{r})$ that is not normalized (i.e. $L(\mathbf{0}) \neq 1$), the local localization tensor definition has to be adapted:

$$\mathbf{H}^L = \frac{-\nabla\nabla^{\mathrm{T}} L\big|_{\mathbf{r}=\mathbf{0}}}{L(\mathbf{0})} \;\Leftrightarrow\; H^L_{\alpha,\beta} = \frac{-\left.\dfrac{\partial^2 L}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}}}{L(\mathbf{0})} \tag{152}$$

8.3. Using the asymptotic sample correlation tensor

The asymptotic sample correlation $\mathbf{C}$ and the localization $\mathbf{L}^{\mathrm{app1}}$ are transposed in continuous space and respectively denoted $C(\mathbf{r})$ and $L^{\mathrm{app1}}(\mathbf{r})$. Equation (149) becomes:

$$L^{\mathrm{app1}} = \frac{C^2}{\left(1+P_4(N)\right)C^2 + P_4(N)} \tag{153}$$

The gradient of L^{app1} is:

$$\frac{\partial L^{\mathrm{app1}}}{\partial x_\alpha} = \frac{2P_4(N)\;C\dfrac{\partial C}{\partial x_\alpha}}{\left(\left(1+P_4(N)\right)C^2 + P_4(N)\right)^2} \tag{154}$$

and its Hessian matrix is:

$$\begin{aligned}\frac{\partial^2 L^{\mathrm{app1}}}{\partial x_\alpha \partial x_\beta} = {} & \frac{2P_4(N)\;C\dfrac{\partial^2 C}{\partial x_\alpha \partial x_\beta}}{\left(\left(1+P_4(N)\right)C^2 + P_4(N)\right)^2} \\ & + \frac{2P_4(N)\left(P_4(N) - 3\left(1+P_4(N)\right)C^2\right)\dfrac{\partial C}{\partial x_\alpha}\dfrac{\partial C}{\partial x_\beta}}{\left(\left(1+P_4(N)\right)C^2 + P_4(N)\right)^3}\end{aligned} \tag{155}$$

However in $\mathbf{r} = \mathbf{0}$:

$$C(\mathbf{0}) = 1 \tag{156a}$$

$$\left.\frac{\partial C}{\partial x_\alpha}\right|_{\mathbf{r}=\mathbf{0}} = 0 \tag{156b}$$

so that:

$$\left.\frac{\partial^2 L^{\mathrm{app1}}}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} = \frac{2P_4(N)}{\Big(1+2P_4(N)\Big)^2} \left.\frac{\partial^2 C}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} \tag{157}$$

Thus, the local localization tensor associated with L^{app1} is given by:

$$\begin{aligned} H_{\alpha,\beta}^{L^{\mathrm{app1}}} &= \frac{-\left.\dfrac{\partial^2 L^{\mathrm{app1}}}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}}}{L^{\mathrm{app1}}(\mathbf{0})} \\ &= -\frac{2P_4(N)}{1+2P_4(N)} \left.\frac{\partial^2 C}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} \\ &= P_{18}(N)\ H_{\alpha,\beta}^{C} \end{aligned} \tag{158}$$

with:

$$P_{18}(N) = \frac{2P_4(N)}{1+2P_4(N)} = \frac{2}{N+1} \tag{159}$$

Since $P_{18}(N)$ is positive, the components of $\mathbf{H}^{L^{\mathrm{app1}}}$ and $\mathbf{H}^{C}$ have the same sign.

8.4. Using the squared sample correlation tensor

The sample correlation $\widetilde{\mathbf{C}}$ and the localization $\mathbf{L}^{\mathrm{app2}}$ are transposed in continuous space and respectively denoted $\widetilde{C}(\mathbf{r})$ and $L^{\mathrm{app2}}(\mathbf{r})$. Equation (149) becomes:

$$L^{\mathrm{app2}} = P_{17}(N) + P_{13}(N)\frac{1}{\mathbb{E}\left[\widetilde{C}^2\right]} \tag{160}$$

The gradient of L^{app2} is:

$$\frac{\partial L^{\mathrm{app2}}}{\partial x_\alpha} = -P_{13}(N)\ \frac{\dfrac{\partial \mathbb{E}\left[\widetilde{C}^2\right]}{\partial x_\alpha}}{\mathbb{E}\left[\widetilde{C}^2\right]^2} \tag{161}$$

and its Hessian matrix is:

$$\frac{\partial^2 L^{\mathrm{app2}}}{\partial x_\alpha \partial x_\beta} = -P_{13}(N)\ \frac{\dfrac{\partial^2 \mathbb{E}\left[\widetilde{C}^2\right]}{\partial x_\alpha \partial x_\beta}\mathbb{E}\left[\widetilde{C}^2\right] - 2\dfrac{\partial \mathbb{E}\left[\widetilde{C}^2\right]}{\partial x_\alpha}\dfrac{\partial \mathbb{E}\left[\widetilde{C}^2\right]}{\partial x_\beta}}{\mathbb{E}\left[\widetilde{C}^2\right]^3} \tag{162}$$

However in $\mathbf{r} = \mathbf{0}$:

$$\widetilde{C}(\mathbf{0}) = 1 \tag{163a}$$

$$\left.\frac{\partial \widetilde{C}}{\partial x_\alpha}\right|_{\mathbf{r}=\mathbf{0}} = 0 \tag{163b}$$

so that:

$$\left.\frac{\partial^2 L^{\mathrm{app2}}}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} = -P_{13}(N) \left.\frac{\partial^2 \mathbb{E}\left[\widetilde{C}^2\right]}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} \tag{164}$$

Thus, the local localization tensor associated with L^{app2} is given by:

$$\begin{aligned} H^{L^{\mathrm{app2}}}_{\alpha,\beta} &= \frac{-\left.\dfrac{\partial^2 L^{\mathrm{app2}}}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}}}{L^{\mathrm{app2}}(\mathbf{0})} \\ &= \frac{P_{13}(N)}{P_{17}(N) + P_{13}(N)} \left.\frac{\partial^2 \mathbb{E}\left[\widetilde{C}^2\right]}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} \\ &= P_{19}(N)\ H^{\mathbb{E}\left[\widetilde{C}^2\right]}_{\alpha,\beta} \end{aligned} \tag{165}$$

with:

$$P_{19}(N) = -\frac{P_{13}(N)}{P_{17}(N) + P_{13}(N)} = \frac{1}{N-2} \tag{166}$$

8.5. Using the sample correlation tensor

Since expectation and differentiation commute:

$$\frac{\partial^2 \mathbb{E}\left[\widetilde{C}^2\right]}{\partial x_\alpha \partial x_\beta} = \mathbb{E}\left[\frac{\partial^2 \widetilde{C}^2}{\partial x_\alpha \partial x_\beta}\right] \tag{167}$$

The gradient of $\widetilde{C}^2$ is:

$$\frac{\partial \widetilde{C}^2}{\partial x_\alpha} = 2\widetilde{C}\frac{\partial \widetilde{C}}{\partial x_\alpha} \tag{168}$$

and the Hessian matrix of $\widetilde{C}^2$ is:

$$\frac{\partial^2 \widetilde{C}^2}{\partial x_\alpha \partial x_\beta} = 2\left(\frac{\partial \widetilde{C}}{\partial x_\alpha}\frac{\partial \widetilde{C}}{\partial x_\beta} + \widetilde{C}\frac{\partial^2 \widetilde{C}}{\partial x_\alpha \partial x_\beta}\right) \tag{169}$$

so that:

$$\left.\frac{\partial^2 \widetilde{C}^2}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} = 2 \left.\frac{\partial^2 \widetilde{C}}{\partial x_\alpha \partial x_\beta}\right|_{\mathbf{r}=\mathbf{0}} \tag{170}$$

which gives, since expectation and differentiation commute:

$$H_{\alpha,\beta}^{\mathbb{E}\left[\widetilde{C}^2\right]} = 2H_{\alpha,\beta}^{\mathbb{E}\left[\widetilde{C}\right]} \tag{171}$$

Finally:

$$H_{\alpha,\beta}^{L^{\mathrm{app2}}} = 2P_{19}(N)\ H_{\alpha,\beta}^{\mathbb{E}\left[\widetilde{C}\right]} \tag{172}$$

Since $P_{19}(N)$ is positive, the components of $\mathbf{H}^{L^{\mathrm{app2}}}$ and $\mathbf{H}^{\mathbb{E}\left[\widetilde{C}\right]}$ have the same sign.

8.6. Disclaimer

The use of previous formulae linking the local localization tensor to local correlation tensors should be used *very carefully*. Indeed, local tensors provide information about the curvature of functions at their origin only, whereas the global shapes of localization and correlation functions are very different in general.

9. Static hybridization

9.1. Filtering method

Static hybridization is the linear combination of the localized sample covariance matrix with a static covariance matrix. Thus, $\widehat{\mathbf{B}}$ is defined as:

$$\widehat{\mathbf{B}} = \beta^{e2}\mathbf{L} \circ \widetilde{\mathbf{B}} + \beta^{c2}\overline{\mathbf{B}} \tag{173}$$

where β^e and β^c are the hybridization coefficients, repectively relative to the localized sample covariance matrix $\mathbf{L} \circ \widetilde{\mathbf{B}}$ and to the static covariance matrix $\overline{\mathbf{B}} \in \mathbb{R}^{n \times n}$.

9.2. Fixed localization

In this section, the localization matrix $\mathbf{L} = \mathbf{L}^f$ is considered as a fixed parameter. For sake of clarity, the coefficient β^{e2} is denoted β and the coefficient β^{c2} is denoted γ. Thus:

$$\widehat{\mathbf{B}} = \beta\mathbf{L}^f \circ \widetilde{\mathbf{B}} + \gamma\overline{\mathbf{B}} \tag{174}$$

9.2.1. Expected squared norm

The expected squared norm of equation (112) adapted for the hybridization becomes:

$$\begin{aligned}
e(\beta,\gamma) &= \sum_{1 \leq i,j \leq n} W_{ij}^2 \mathbb{E}\left[\left(\beta L_{ij}^f \widetilde{B}_{ij} + \gamma \overline{B}_{ij} - B_{ij}\right)^2\right] \\
&= \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(\beta^2 L_{ij}^{f2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \gamma^2 \overline{B}_{ij}^2 + \mathbb{E}\left[B_{ij}^2\right]\right. \\
&\quad \left. + 2\beta L_{ij}^f \gamma \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} - 2\beta L_{ij}^f \mathbb{E}\left[\widetilde{B}_{ij} B_{ij}\right] - 2\gamma \overline{B}_{ij} \mathbb{E}\left[B_{ij}\right]\right)
\end{aligned} \tag{175}$$

Using the independence of random processes in (88), it can be simplified:

$$\begin{aligned}
e(\beta,\gamma) = \sum_{1 \leq i,j \leq n} & W_{ij}^2 \left(\beta^2 L_{ij}^{f2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \gamma^2 \overline{B}_{ij}^2 + \left(1 - 2\beta L_{ij}^f\right) \mathbb{E}\left[B_{ij}^2\right]\right. \\
& \left. + 2\left(\beta L_{ij}^f - 1\right) \gamma \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij}\right)
\end{aligned} \tag{176}$$

Its gradient is given by:

$$\frac{\partial e}{\partial \beta} = 2 \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(\beta L_{ij}^{f2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] - L_{ij}^f \mathbb{E}\left[B_{ij}^2\right] + L_{ij}^f \gamma \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij}\right) \tag{177a}$$

$$\frac{\partial e}{\partial \gamma} = 2 \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(\gamma \overline{B}_{ij}^2 + \left(\beta L_{ij}^f - 1\right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij}\right) \tag{177b}$$

9.2.2. Explicit optimality

Setting the gradient of $e(\beta, \gamma)$ to zero, we get:

$$\beta \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \gamma \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f} \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} = \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f} \mathbb{E}\left[B_{ij}^2\right] \tag{178a}$$

$$\beta \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f} \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} + \gamma \sum_{1 \leq i,j \leq n} W_{ij}^2 \overline{B}_{ij}^2 = \sum_{1 \leq i,j \leq n} W_{ij}^2 \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \tag{178b}$$

The determinant of this system is:

$$d = \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] \sum_{1 \leq i,j \leq n} W_{ij}^2 \overline{B}_{ij}^2 - \left(\sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f} \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij}\right)^2 \tag{179}$$

We denote:

$$\begin{aligned} n_\beta = & \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f} \mathbb{E}\left[B_{ij}^2\right] \sum_{1 \leq i,j \leq n} W_{ij}^2 \overline{B}_{ij}^2 \\ & - \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f} \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \sum_{1 \leq i,j \leq n} W_{ij}^2 \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \end{aligned} \tag{180a}$$

$$\begin{aligned} n_\gamma = & \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] \sum_{1 \leq i,j \leq n} W_{ij}^2 \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \\ & - \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f} \mathbb{E}\left[B_{ij}^2\right] \sum_{1 \leq i,j \leq n} W_{ij}^2 L_{ij}^{f} \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \end{aligned} \tag{180b}$$

to write the solution of the system:

$$\beta = \frac{n_\beta}{d} \tag{181a}$$

$$\gamma = \frac{n_\gamma}{d} \tag{181b}$$

9.3. Optimized localization

In this section, the localization matrix $\mathbf{L}$ is optimized simultaneously with the hybridization weights. The coefficient β^{e2} can be included in the localization matrix: $\mathbf{L}^h = \beta^{e2}\mathbf{L}$, and the coefficient β^{c2} is denoted γ. Thus:

$$\widehat{\mathbf{B}} = \mathbf{L}^h \circ \widetilde{\mathbf{B}} + \gamma \overline{\mathbf{B}} \tag{182}$$

9.3.1. Expected squared norm

The expected squared norm of equation (112) adapted for the hybridization becomes:

$$\begin{aligned} e(\mathbf{L}^h, \gamma) = & \sum_{1 \leq i,j \leq n} W_{ij}^2 \mathbb{E}\left[\left(L_{ij}^h \widetilde{B}_{ij} + \gamma \overline{B}_{ij} - B_{ij}\right)^2\right] \\ = & \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(L_{ij}^{h2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \gamma^2 \overline{B}_{ij}^2 + \mathbb{E}\left[B_{ij}^2\right]\right. \\ & \left. + 2L_{ij}^h \gamma \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} - 2L_{ij}^h \mathbb{E}\left[\widetilde{B}_{ij} B_{ij}\right] - 2\gamma \overline{B}_{ij} \mathbb{E}\left[B_{ij}\right]\right) \end{aligned} \tag{183}$$

Using the independence of random processes in (88), it can be simplified:

$$\begin{aligned} e(\mathbf{L}^h, \gamma) = \sum_{1 \leq i,j \leq n} & W_{ij}^2 \left(L_{ij}^{h2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \gamma^2 \overline{B}_{ij}^2 + \left(1 - 2L_{ij}^h\right) \mathbb{E}\left[B_{ij}^2\right] \right. \\ & \left. +2\left(L_{ij}^h - 1\right) \gamma \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) \end{aligned} \tag{184}$$

Its gradient is given by:

$$\frac{\partial e}{\partial L_{ij}^h} = 2W_{ij}^2 \left(L_{ij}^h \mathbb{E}\left[\widetilde{B}_{ij}^2\right] - \mathbb{E}\left[B_{ij}^2\right] + \gamma \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) \tag{185a}$$

$$\frac{\partial e}{\partial \gamma} = 2 \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(\gamma \overline{B}_{ij}^2 + \left(L_{ij}^h - 1\right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) \tag{185b}$$

9.3.2. Explicit optimality

Setting the gradient of $e(\mathbf{L}^h, \gamma)$ to zero, we get:

$$\begin{aligned} & 2W_{ij}^2 \left(L_{ij}^{h,\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] - \mathbb{E}\left[B_{ij}^2\right] + \gamma^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) = 0 \\ \Leftrightarrow\ & L_{ij}^{h,\mathrm{opt}} = \frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} - \gamma^{\mathrm{opt}} \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij}}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} \end{aligned} \tag{186}$$

and:

$$\begin{aligned} & 2 \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(\gamma^{\mathrm{opt}} \overline{B}_{ij}^2 + \left(L_{ij}^{h,\mathrm{opt}} - 1\right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) = 0 \\ \Leftrightarrow\ & \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(\gamma^{\mathrm{opt}} \overline{B}_{ij}^2 + \left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} - \gamma^{\mathrm{opt}} \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij}}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} - 1 \right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) = 0 \\ \Leftrightarrow\ & \gamma^{\mathrm{opt}} = \frac{\displaystyle\sum_{1 \leq i,j \leq n} W_{ij}^2 \left(1 - \frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} \right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij}}{\displaystyle\sum_{1 \leq i,j \leq n} W_{ij}^2 \left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]^2}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} \right) \overline{B}_{ij}^2} \end{aligned} \tag{187}$$

9.4. Properties

Equation (187) shows an interesting property of the optimal hybridization coefficient γ^{opt}: it takes the amplitude of the specified static covariance matrix $\overline{\mathbf{B}}$ into account. For instance, if the specified static covariance matrix $\overline{\mathbf{B}}$ is multiplied by 2, then the optimal hybridization coefficient γ^{opt} is divided by 2.

For the case without hybridization, no localization is applied for an infinite ensemble size: $\mathbb{E}\left[\widetilde{B}_{ij}^2\right] \rightarrow$

$\mathbb{E}\left[B_{ij}^2\right]$. Thus, for the case with hybridization, the optimal hybridization coefficient vanishes ($\gamma^{\text{opt}} \rightarrow 0$) and no localization is applied ($L_{ij}^{h,\text{opt}} \rightarrow 1$).

9.5. Norm reduction

Using the independence of random processes in (88) and the optimality condition of equation (186), we verify that:

$$\mathbb{E}\left[\left(\widehat{B}_{ij}^{\text{opt}} - B_{ij}\right)\widetilde{B}_{ij}\right] = L_{ij}^{h,\text{opt}}\mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \gamma^{\text{opt}}\overline{B}_{ij}\mathbb{E}\left[\widetilde{B}_{ij}\right] - \mathbb{E}\left[B_{ij}^2\right] = 0 \tag{188}$$

so that:

$$e\left(\mathbf{1},0\right) - e\left(\mathbf{L}^{h,\text{opt}},\gamma^{\text{opt}}\right) = \sum_{1\leq i,j\leq n} W_{ij}^2\mathbb{E}\left[\left(\widehat{B}_{ij}^{\text{opt}} - \widetilde{B}_{ij}\right)^2\right] \tag{189}$$

9.6. Hybridization benefits

The impact of localization and hybridization versus localization alone can be computed via the difference between $e(\mathbf{L}^{h,\text{opt}},\gamma^{\text{opt}})$ and $e(\mathbf{L}^{\text{opt}})$:

$$\begin{aligned}
e(\mathbf{L}^{h,\text{opt}},\gamma^{\text{opt}}) &= \mathbb{E}\left[\sum_{1\leq i,j\leq n} W_{ij}^2\left(L_{ij}^{h,\text{opt}}\widetilde{B}_{ij} + \gamma^{\text{opt}}\overline{B}_{ij} - B_{ij}\right)^2\right] \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2\mathbb{E}\left[\left(\left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} - \gamma^{\text{opt}}\frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij}}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\right)\widetilde{B}_{ij} + \gamma^{\text{opt}}\overline{B}_{ij} - B_{ij}\right)^2\right] \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2\mathbb{E}\left[\left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij} - B_{ij} + \left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij}\right)\gamma^{\text{opt}}\overline{B}_{ij}\right)^2\right] \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2\mathbb{E}\left[\left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij} - B_{ij}\right)^2\right] \\
&\quad + 2\sum_{1\leq i,j\leq n} W_{ij}^2\mathbb{E}\left[\left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij} - B_{ij}\right)\left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij}\right)\right]\gamma^{\text{opt}}\overline{B}_{ij} \\
&\quad + \sum_{1\leq i,j\leq n} W_{ij}^2\mathbb{E}\left[\left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij}\right)^2\right]\gamma^{\text{opt}2}\overline{B}_{ij}^2
\end{aligned} \tag{190}$$

In the previous equation, the first term can be expressed with equation (139) as:

$$\sum_{1\leq i,j\leq n} W_{ij}^2\mathbb{E}\left[\left(L_{ij}^{\text{opt}}\widetilde{B}_{ij} - B_{ij}\right)^2\right] = e(\mathbf{L}^{\text{opt}}) \tag{191}$$

The second term can be simplified in:

$$
\begin{aligned}
& 2 \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij} - B_{ij}\right)\left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij}\right)\right]\gamma^{\mathrm{opt}}\overline{B}_{ij} \\
&= 2 \sum_{1\leq i,j\leq n} W_{ij}^2 \left(\left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} - 1\right)\mathbb{E}\left[\widetilde{B}_{ij}\right] - \frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\mathbb{E}\left[\widetilde{B}_{ij}\right] + \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\mathbb{E}\left[B_{ij}^2\right]\right)\gamma^{\mathrm{opt}}\overline{B}_{ij} \\
&= 2 \sum_{1\leq i,j\leq n} W_{ij}^2 \left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} - 1\right)\gamma^{\mathrm{opt}}\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij} \qquad (192)
\end{aligned}
$$

The third term is equal to:

$$
\begin{aligned}
& \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\widetilde{B}_{ij}\right)^2\right]\gamma^{\mathrm{opt}2}\overline{B}_{ij}^2 \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2 \left(1 - 2\frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]^2}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} + \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]^2}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\right)\gamma^{\mathrm{opt}2}\overline{B}_{ij}^2 \\
&= \sum_{1\leq i,j\leq n} W_{ij}^2 \left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]^2}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\right)\gamma^{\mathrm{opt}2}\overline{B}_{ij}^2 \qquad (193)
\end{aligned}
$$

Interestingly, the intermediate step of equation (187) leads to:

$$
\sum_{1\leq i,j\leq n} W_{ij}^2 \left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} - 1\right)\gamma^{\mathrm{opt}}\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij} = \sum_{1\leq i,j\leq n} W_{ij}^2 \left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]^2}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\right)\gamma^{\mathrm{opt}2}\overline{B}_{ij}^2 \qquad (194)
$$

Thus, we finally get:

$$
e(\mathbf{L}^{h,\mathrm{opt}}, \gamma^{\mathrm{opt}}) - e(\mathbf{L}^{\mathrm{opt}}) = - \sum_{1\leq i,j\leq n} W_{ij}^2 \left(1 - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]^2}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\right)\gamma^{\mathrm{opt}2}\overline{B}_{ij}^2 \leq 0 \qquad (195)
$$

since $\mathbb{E}\left[\widetilde{B}_{ij}^2\right] \geq \mathbb{E}\left[\widetilde{B}_{ij}\right]^2$. This result is very important since it shows the superiority of the hybrid formalism. Indeed, the optimally localized-hybridized covariance matrix is *always* more accurate than its optimally localized-only counterpart, *whatever the static covariance matrix $\overline{\mathbf{B}}$ specified for the hybridization*.

9.7. Optimization of the full static covariance matrix: a failed attempt

Instead of considering the static covariance matrix as an input data and optimizing the hybridization coefficient γ, we can try to optimize the full static covariance matrix $\overline{\mathbf{B}}^{\mathrm{full}} = \gamma\overline{\mathbf{B}}$.

In this case, the expected squared norm is:

$$e(\mathbf{L}^h, \overline{\mathbf{B}}^{\text{full}}) = \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(L_{ij}^{h2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + \overline{B}_{ij}^{\text{full}2} + \left(1 - 2L_{ij}^h\right) \mathbb{E}\left[B_{ij}^2\right] \right.$$
$$\left. +2\left(L_{ij}^h - 1\right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij}^{\text{full}} \right) \tag{196}$$

Its gradient with respect to $\overline{B}_{ij}^{\text{full}}$ is given by:

$$\frac{\partial e}{\partial \overline{B}_{ij}^{\text{full}}} = 2W_{ij}^2 \left(\overline{B}_{ij}^{\text{full}} + \left(L_{ij}^h - 1\right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \right) \tag{197}$$

Setting this gradient to zero leads to:

$$2W_{ij}^2 \left(\overline{B}_{ij}^{\text{full}} + \left(L_{ij}^h - 1\right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \right) = 0 \ \Leftrightarrow \ \overline{B}_{ij}^{\text{full}} = \left(1 - L_{ij}^h\right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \tag{198}$$

In this equation, it is very unlikely that $1 - L_{ij}^h$ will be the element of a positive semi-definite matrix, so that $\overline{\mathbf{B}}^{\text{full}}$ will not have any square-root. Thus, this attempt is useless in practice.

9.8. Hybrid target

Within the hybrid formalism, it is possible to define a hybrid target for the filtering: $\mathbf{B}$ can be replaced by $\mathbf{L}^\star \circ \mathbf{B} + \gamma^\star \overline{\mathbf{B}}$, where $\mathbf{L}^\star$ is a localization matrix and $\gamma^\star$ a hybridization weight, both specified by the user. This hybrid target could be closer to the "true" covariance matrix than $\mathbf{B}$, because of some issues in the ensemble generation.

The expected squared norm of equation (112) in this case becomes:

$$\begin{aligned} e(\mathbf{L}^h, \gamma) &= \sum_{1 \leq i,j \leq n} W_{ij}^2 \mathbb{E}\left[\left(L_{ij}^h \widetilde{B}_{ij} + \gamma \overline{B}_{ij} - \left(L_{ij}^\star B_{ij} + \gamma^\star \overline{B}_{ij} \right) \right)^2 \right] \\ &= \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(L_{ij}^{h2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + (\gamma - \gamma^\star)^2 \overline{B}_{ij}^2 + L_{ij}^{\star 2} \mathbb{E}\left[B_{ij}^2\right] \right. \\ &\left. +2L_{ij}^h (\gamma - \gamma^\star) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} - 2L_{ij}^h L_{ij}^\star \mathbb{E}\left[\widetilde{B}_{ij} B_{ij}\right] - 2(\gamma - \gamma^\star) \overline{B}_{ij} L_{ij}^\star \mathbb{E}\left[B_{ij}\right] \right) \end{aligned} \tag{199}$$

Using the independence of random processes in (88), it can be simplified:

$$e(\mathbf{L}^h, \gamma) = \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(L_{ij}^{h2} \mathbb{E}\left[\widetilde{B}_{ij}^2\right] + (\gamma - \gamma^\star)^2 \overline{B}_{ij}^2 + \left(L_{ij}^{\star 2} - 2L_{ij}^h L_{ij}^\star\right) \mathbb{E}\left[B_{ij}^2\right] \right.$$
$$\left. +2\left(L_{ij}^h - L_{ij}^\star\right) (\gamma - \gamma^\star) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) \tag{200}$$

Its gradient is given by:

$$\frac{\partial e}{\partial L_{ij}^h} = 2W_{ij}^2 \left(L_{ij}^h \mathbb{E}\left[\widetilde{B}_{ij}^2\right] - L_{ij}^\star \mathbb{E}\left[B_{ij}^2\right] + (\gamma - \gamma^\star)^2 \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) \tag{201a}$$

$$\frac{\partial e}{\partial \gamma} = 2 \sum_{1 \leq i,j \leq n} W_{ij}^2 \left((\gamma - \gamma^\star) \overline{B}_{ij}^2 + \left(L_{ij}^h - L_{ij}^\star\right) \mathbb{E}\left[\widetilde{B}_{ij}\right] \overline{B}_{ij} \right) \tag{201b}$$

Setting the gradient of $e(\mathbf{L}^h,\gamma)$ to zero, we get:

$$
\begin{aligned}
& 2W_{ij}^2\left(L_{ij}^{h,\mathrm{opt}}\mathbb{E}\left[\widetilde{B}_{ij}^2\right]-L_{ij}^\star\mathbb{E}\left[B_{ij}^2\right]+\left(\gamma^{\mathrm{opt}}-\gamma^\star\right)\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij}\right)=0\\
\Leftrightarrow\ & L_{ij}^{h,\mathrm{opt}}=L_{ij}^\star\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}-\left(\gamma^{\mathrm{opt}}-\gamma^\star\right)\frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij}}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}
\end{aligned}
\tag{202}
$$

and:

$$
\begin{aligned}
& 2\sum_{1\le i,j\le n}W_{ij}^2\left(\left(\gamma^{\mathrm{opt}}-\gamma^\star\right)\overline{B}_{ij}^2+\left(L_{ij}^{h,\mathrm{opt}}-L_{ij}^\star\right)\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij}\right)=0\\
\Leftrightarrow & \sum_{1\le i,j\le n}W_{ij}^2\left(\left(\gamma^{\mathrm{opt}}-\gamma^\star\right)\overline{B}_{ij}^2+\left(L_{ij}^\star\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}-\left(\gamma^{\mathrm{opt}}-\gamma^\star\right)\frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij}}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}-L_{ij}^\star\right)\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij}\right)=0\\
\Leftrightarrow\ & \gamma^{\mathrm{opt}}=\gamma^\star+\frac{\displaystyle\sum_{1\le i,j\le n}W_{ij}^2L_{ij}^\star\left(1-\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\right)\mathbb{E}\left[\widetilde{B}_{ij}\right]\overline{B}_{ij}}{\displaystyle\sum_{1\le i,j\le n}W_{ij}^2\left(1-\frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]^2}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\right)\overline{B}_{ij}^2}
\end{aligned}
\tag{203}
$$

We denote $\mathbf{L}^{h,\mathrm{old}}$ and γ^{old} the localization and hybridization parameter obtained using $\mathbf{B}$ as a target (equations (186) and (187)). Interestingly, if $L_{ij}^\star=L^\star$ (hybrid target without localization of $\mathbf{B}$), then:

$$
L_{ij}^{h,\mathrm{opt}}=L^\star L_{ij}^{h,\mathrm{old}}
\tag{204a}
$$

$$
\gamma^{\mathrm{opt}}=\gamma^\star+L^\star\gamma^{\mathrm{old}}
\tag{204b}
$$

10. Dual-ensemble hybridization

10.1. Filtering method

Instead of combining a localized sample covariance matrix with a static covariance matrix, we can perform a linear combination of two localized sample covariance matrices estimated from two distinct ensembles. Let $\widetilde{\mathbf{B}}_1$ be the main sample covariance matrix estimated from an ensemble of size N_1, and $\widetilde{\mathbf{B}}_2$ the secondary sample covariance matrix estimated from a distinct ensemble of size N_2. Thus, $\widehat{\mathbf{B}}$ is defined as:

$$\widehat{\mathbf{B}} = \mathbf{L}_1 \circ \widetilde{\mathbf{B}}_1 + \mathbf{L}_2 \circ \widetilde{\mathbf{B}}_2 \tag{205}$$

where $\mathbf{L}_1$ and $\mathbf{L}_2$ are the effective localization matrices applied to $\widetilde{\mathbf{B}}_1$ and $\widetilde{\mathbf{B}}_2$, which includes hybridization coefficients implicitly.

10.2. Expected squared norm

We assume that the asymptotic sample covariance matrix of the first ensemble $\mathbf{B}_1$ can be used as the target matrix. The expected squared norm of equation (112) adapted for the dual-ensemble hybridization becomes:

$$\begin{aligned}
e(\mathbf{L}_1, \mathbf{L}_2) &= \sum_{1 \leq i,j \leq n} W_{ij}^2 \mathbb{E}\left[\left(L_{ij,1}\widetilde{B}_{ij,1} + L_{ij,2}\widetilde{B}_{ij,2} - B_{ij,1}\right)^2\right] \\
&= \sum_{1 \leq i,j \leq n} W_{ij}^2 \left(L_{ij,1}^2 \mathbb{E}\left[\widetilde{B}_{ij,1}^2\right] + L_{ij,2}^2 \mathbb{E}\left[\widetilde{B}_{ij,2}^2\right] + \mathbb{E}\left[B_{ij,1}^2\right]\right. \\
&\quad + 2L_{ij,1}L_{ij,2}\mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right] - 2L_{ij,1}\mathbb{E}\left[\widetilde{B}_{ij,1}B_{ij,1}\right] \\
&\quad \left. -2L_{ij,2}\mathbb{E}\left[\widetilde{B}_{ij,2}B_{ij,1}\right]\right)
\end{aligned} \tag{206}$$

Using the independence of random processes in (88), it can be simplified:

$$\begin{aligned}
e(\mathbf{L}_1, \mathbf{L}_2) = \sum_{1 \leq i,j \leq n} W_{ij}^2 &\left(L_{ij,1}^2 \mathbb{E}\left[\widetilde{B}_{ij,1}^2\right] + L_{ij,2}^2 \mathbb{E}\left[\widetilde{B}_{ij,2}^2\right] + \left(1 - 2L_{ij,1}\right)\mathbb{E}\left[B_{ij,1}^2\right]\right. \\
&\left. +2L_{ij,1}L_{ij,2}\mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right] - 2L_{ij,2}\mathbb{E}\left[\widetilde{B}_{ij,2}B_{ij,1}\right]\right)
\end{aligned} \tag{207}$$

Its gradient is given by:

$$\frac{\partial e}{\partial L_{ij,1}} = 2W_{ij}^2 \left(L_{ij,1}\mathbb{E}\left[\widetilde{B}_{ij,1}^2\right] - \mathbb{E}\left[B_{ij,1}^2\right] + L_{ij,2}\mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right]\right) \tag{208a}$$

$$\frac{\partial e}{\partial L_{ij,2}} = 2W_{ij}^2 \left(L_{ij,2}\mathbb{E}\left[\widetilde{B}_{ij,2}^2\right] + L_{ij,1}\mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right] - \mathbb{E}\left[\widetilde{B}_{ij,2}B_{ij,1}\right]\right) \tag{208b}$$

10.3. Explicit optimality

Setting the gradient of $e(\mathbf{L}_1, \mathbf{L}_2)$ to zero, we get:

$$\begin{aligned} & 2W_{ij}^2 \left(L_{ij,1}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,1}^2\right] - \mathbb{E}\left[B_{ij,1}^2\right] + L_{ij,2}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right] \right) = 0 \\ \Leftrightarrow\ & L_{ij,1}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,1}^2\right] + L_{ij,2}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right] = \mathbb{E}\left[B_{ij,1}^2\right] \end{aligned} \tag{209}$$

and:

$$\begin{aligned} & 2W_{ij}^2 \left(L_{ij,2}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,2}^2\right] + L_{ij,1}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right] - \mathbb{E}\left[\widetilde{B}_{ij,2}B_{ij,1}\right] \right) = 0 \\ \Leftrightarrow\ & L_{ij,2}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,2}^2\right] + L_{ij,1}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right] = \mathbb{E}\left[\widetilde{B}_{ij,2}B_{ij,1}\right] \end{aligned} \tag{210}$$

Combining these equations, we get:

$$L_{ij,1}^{\mathrm{opt}} = \frac{\mathbb{E}\left[B_{ij,1}^2\right]\mathbb{E}\left[\widetilde{B}_{ij,2}^2\right] - \mathbb{E}\left[\widetilde{B}_{ij,2}B_{ij,1}\right]\mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right]}{\mathbb{E}\left[\widetilde{B}_{ij,1}^2\right]\mathbb{E}\left[\widetilde{B}_{ij,2}^2\right] - \mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right]^2} \tag{211a}$$

$$L_{ij,2}^{\mathrm{opt}} = \frac{\mathbb{E}\left[\widetilde{B}_{ij,2}B_{ij,1}\right]\mathbb{E}\left[\widetilde{B}_{ij,1}^2\right] - \mathbb{E}\left[B_{ij,1}^2\right]\mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right]}{\mathbb{E}\left[\widetilde{B}_{ij,1}^2\right]\mathbb{E}\left[\widetilde{B}_{ij,2}^2\right] - \mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right]^2} \tag{211b}$$

10.4. Norm reduction

Using the independence of random processes in (88) and the optimality condition of equation (209), we verify that:

$$\mathbb{E}\left[\left(\widehat{B}_{ij}^{\mathrm{opt}} - B_{ij,1}\right)\widetilde{B}_{ij,1}\right] = L_{ij,1}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,1}^2\right] + L_{ij,2}^{\mathrm{opt}} \mathbb{E}\left[\widetilde{B}_{ij,2}\widetilde{B}_{ij,1}\right] - \mathbb{E}\left[B_{ij,1}^2\right] = 0 \tag{212}$$

so that:

$$e(\mathbf{1},\mathbf{0}) - e\left(\mathbf{L}_1^{\mathrm{opt}}, \mathbf{L}_2^{\mathrm{opt}}\right) = \sum_{1\leq i,j\leq n} W_{ij}^2 \mathbb{E}\left[\left(\widehat{B}_{ij}^{\mathrm{opt}} - \widetilde{B}_{ij}\right)^2\right] \tag{213}$$

10.5. Additional assumption

The expected product $\mathbb{E}\left[\widetilde{B}_{ij,2}B_{ij,1}\right]$ cannot be estimated using the sampling theory: an extra assumption is required. If the sampling error $\widetilde{B}_{ij,1} - B_{ij,1}$ of the main ensemble was not correlated with the sample covariance $\widetilde{B}_{ij,2}$ of the secondary ensemble, then we would get:

$$\mathbb{E}\left[B_{ij,1}\widetilde{B}_{ij,2}\right] = \mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right] \tag{214}$$

The Cauchy-Schwarz inequality ensures that the denominator or $L^{\mathrm{opt}}_{ij,2}$ and $L^{\mathrm{opt}}_{ij,2}$ is nonnegative. However, if we want to keep $L^{\mathrm{opt}}_{ij,2}$ and $L^{\mathrm{opt}}_{ij,2}$ nonnegative, then we have to set respectively:

$$\frac{\mathbb{E}\left[B^2_{ij,1}\right]\mathbb{E}\left[\widetilde{B}^2_{ij,2}\right]}{\mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right]} \geq \mathbb{E}\left[B_{ij,1}\widetilde{B}_{ij,2}\right] \tag{215a}$$

$$\mathbb{E}\left[B_{ij,1}\widetilde{B}_{ij,2}\right] \geq \frac{\mathbb{E}\left[B^2_{ij,1}\right]\mathbb{E}\left[\widetilde{B}_{ij,1}\widetilde{B}_{ij,2}\right]}{\mathbb{E}\left[\widetilde{B}^2_{ij,1}\right]} \tag{215b}$$

In the case where the system (215) has no solution in practice, the localizations $L^{\mathrm{opt}}_{ij,1}$ and $L^{\mathrm{opt}}_{ij,2}$ should be set to zero.

11. Multi-block covariance, common filtering

11.1. Multi-block formalism

If the random vector $\mathbf{x}^b$ is built as a concatenation of P blocks of size n/P:

$$\mathbf{x}^b = \begin{pmatrix} \mathbf{x}^b_1 \\ \vdots \\ \mathbf{x}^b_P \end{pmatrix} \tag{216}$$

then the corresponding covariance matrices and weight matrices will have the structure of:

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \dots & \mathbf{B}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{P1} & \dots & \mathbf{B}_{PP} \end{pmatrix} \tag{217}$$

where $\mathbf{B}_{pq} \in \mathbb{R}^{n/P \times n/P}$. For sake of clarity, we introduce a new notations for the block-weighted average:

$$\overbrace{\underbrace{B_{ij,pq}}} = \sum_{1 \leq p,q \leq P} W^2_{ij,pq} B_{ij,pq} \tag{218}$$

Since the different blocks of covariance can have very different ranges of values, the optimization might be dominated by some blocks with high values and not take other blocks with small values into account. A simple solution would be to use the inverse climatologic standard deviation $\bar{\boldsymbol{\sigma}}$ as weights:

$$W_{ij} = \bar{\sigma}_i^{-1} \bar{\sigma}_j^{-1} \tag{219}$$

11.2. Common localization

A usual practice is to apply a common covariance matrix $\mathbf{L} \in \mathbb{R}^{n/P \times n/P}$ to each block $\widetilde{\mathbf{B}}_{pq}$:

$$\widehat{\mathbf{B}}_{pq} = \mathbf{L} \circ \widetilde{\mathbf{B}}_{pq} \tag{220}$$

The expected squared norm of equation (112) adapted for the common localization becomes:

$$\begin{aligned} e(\mathbf{L}) &= \sum_{1 \leq i,j \leq (n/P)} \sum_{1 \leq p,q \leq P} W^2_{ij,pq} \mathbb{E}\left[\left(L_{ij} \widetilde{B}_{ij,pq} - B_{ij,pq} \right)^2 \right] \\ &= \sum_{1 \leq i,j \leq (n/P)} \left(L^2_{ij} \overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}^2_{ij,pq} \right]}} + \overbrace{\underbrace{\mathbb{E}\left[B^2_{ij,pq} \right]}} - 2 L_{ij} \overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}_{ij,pq} B_{ij,pq} \right]}} \right) \end{aligned} \tag{221}$$

Using the independence of random processes in (88), it can be simplified:

$$e(\mathbf{L}) = \sum_{1 \leq i,j \leq (n/P)} \left(L^2_{ij} \overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}^2_{ij,pq} \right]}} + \left(1 - 2 L_{ij} \right) \overbrace{\underbrace{\mathbb{E}\left[B^2_{ij,pq} \right]}} \right) \tag{222}$$

Its gradient is given by:

$$\frac{\partial e}{\partial L_{ij}} = 2\left(L_{ij}\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}^2_{ij,pq}\right]}} - \overbrace{\underbrace{\mathbb{E}\left[B^2_{ij,pq}\right]}}\right) \tag{223}$$

Setting the gradient of $e(\mathbf{L})$ to zero, we get:

$$\begin{aligned}
& 2\left(L^{\text{opt}}_{ij}\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}^2_{ij,pq}\right]}} - \overbrace{\underbrace{\mathbb{E}\left[B^2_{ij,pq}\right]}}\right) = 0 \\
\Leftrightarrow\ & L^{\text{opt}}_{ij} = \frac{\overbrace{\underbrace{\mathbb{E}\left[B^2_{ij,pq}\right]}}}{\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}^2_{ij,pq}\right]}}}
\end{aligned} \tag{224}$$

11.3. Common static hybridization

If a common localization $\mathbf{L}^h$ and a common hybrid coefficient γ are used for the static hybridization, each filtered block is given by:

$$\widehat{\mathbf{B}}_{pq} = \mathbf{L}^h \circ \widetilde{\mathbf{B}}_{pq} + \gamma\overline{\mathbf{B}}_{pq} \tag{225}$$

The expected squared norm of equation (112) adapted for the common hybridization becomes:

$$\begin{aligned}
e(\mathbf{L}^h,\gamma) &= \sum_{1\le i,j\le (n/P)}\ \sum_{1\le p,q\le P} W^2_{ij,pq}\,\mathbb{E}\left[\left(L^h_{ij}\widetilde{B}_{ij,pq} + \gamma\overline{B}_{ij,pq} - B_{ij,pq}\right)^2\right] \\
&= \sum_{1\le i,j\le (n/P)} \left(L^{h2}_{ij}\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}^2_{ij,pq}\right]}} + \gamma^2\,\overbrace{\underbrace{\overline{B}^2_{ij,pq}}} + \overbrace{\underbrace{\mathbb{E}\left[B^2_{ij,pq}\right]}}\right. \\
&\qquad \left. + 2L^h_{ij}\gamma\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij}}} - 2L^h_{ij}\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}_{ij,pq}B_{ij,pq}\right]}} - 2\gamma\,\overbrace{\underbrace{\overline{B}_{ij,pq}\mathbb{E}\left[B_{ij,pq}\right]}}\right)
\end{aligned} \tag{226}$$

Using the independence of random processes in (88), it can be simplified:

$$\begin{aligned}
e(\mathbf{L}^h,\gamma) = \sum_{1\le i,j\le (n/P)} & \left(L^{h2}_{ij}\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}^2_{ij,pq}\right]}} + \gamma^2\,\overbrace{\underbrace{\overline{B}^2_{ij,pq}}}\right. \\
& \left. + \left(1 - 2L^h_{ij}\right)\overbrace{\underbrace{\mathbb{E}\left[B^2_{ij,pq}\right]}} + 2\left(L^h_{ij} - 1\right)\gamma\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}}}\right)
\end{aligned} \tag{227}$$

Its gradient is given by:

$$\frac{\partial e}{\partial L^h_{ij}} = 2\left(L^h_{ij}\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}^2_{ij,pq}\right]}} - \overbrace{\underbrace{\mathbb{E}\left[B^2_{ij,pq}\right]}} + \gamma\,\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}}}\right) \tag{228a}$$

$$\frac{\partial e}{\partial \gamma} = 2\sum_{1\le i,j\le (n/P)} \left(\gamma\,\overbrace{\underbrace{\overline{B}^2_{ij,pq}}} + \left(L^h_{ij} - 1\right)\overbrace{\underbrace{\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}}}\right) \tag{228b}$$

Setting the gradient of $e(\mathbf{L}^h,\gamma)$ to zero, we get:

$$
\begin{aligned}
&2\left(L_{ij}^{h,\mathrm{opt}}\,\mathbb{E}\left[\widetilde{B}_{ij,pq}^2\right]-\mathbb{E}\left[B_{ij,pq}^2\right]+\gamma^{\mathrm{opt}}\,\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}\right)=0\\
\Leftrightarrow\ &L_{ij}^{h,\mathrm{opt}}=\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}-\gamma^{\mathrm{opt}}\frac{\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}}{\mathbb{E}\left[\widetilde{B}_{ij,pq}^2\right]}
\end{aligned}
\tag{229}
$$

and:

$$
\begin{aligned}
&2\sum_{1\leq i,j\leq (n/P)}\left(\gamma^{\mathrm{opt}}\,\overline{B}_{ij,pq}^2+\left(L_{ij}^{h,\mathrm{opt}}-1\right)\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}\right)=0\\
\Leftrightarrow\ &\sum_{1\leq i,j\leq (n/P)}\left(\gamma^{\mathrm{opt}}\,\overline{B}_{ij,pq}^2+\left(\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}-\gamma^{\mathrm{opt}}\frac{\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}}{\mathbb{E}\left[\widetilde{B}_{ij,pq}^2\right]}-1\right)\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}\right)=0\\
\Leftrightarrow\ &\gamma^{\mathrm{opt}}=\frac{\displaystyle\sum_{1\leq i,j\leq (n/P)}\left(1-\frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}\right)\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}}{\displaystyle\sum_{1\leq i,j\leq (n/P)}\left(\overline{B}_{ij,pq}^2-\frac{\left(\mathbb{E}\left[\widetilde{B}_{ij,pq}\right]\overline{B}_{ij,pq}\right)^2}{\mathbb{E}\left[\widetilde{B}_{ij,pq}^2\right]}\right)}
\end{aligned}
\tag{230}
$$

11.4. Common dual-ensemble hybridization

For a dual-ensemble with common localization matrices $\mathbf{L}_1$ and $\mathbf{L}_2$, each filtered block is given by:

$$
\widehat{\mathbf{B}}_{pq}=\mathbf{L}_1\circ\widetilde{\mathbf{B}}_{pq,1}+\mathbf{L}_2\circ\widetilde{\mathbf{B}}_{pq,2}
\tag{231}
$$

The expected squared norm of equation (112) adapted for the common dual-ensemble hybridization becomes:

$$
\begin{aligned}
e(\mathbf{L}_1, \mathbf{L}_2) &= \sum_{1\le i,j\le (n/P)} \sum_{1\le p,q\le P} W_{ij,pq}^2 \mathbb{E}\left[\left(L_{ij,1}\widetilde{B}_{ij,pq,1} + L_{ij,2}\widetilde{B}_{ij,pq,2} - B_{ij,pq,1}\right)^2\right] \\
&= \sum_{1\le i,j\le (n/P)} \Bigg(L_{ij,1}^2\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}^2\right] + L_{ij,2}^2\, \mathbb{E}\left[\widetilde{B}_{ij,pq,2}^2\right] + \mathbb{E}\left[B_{ij,pq,1}^2\right] \\
&\quad + 2L_{ij,1}L_{ij,2}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] - 2L_{ij,1}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}B_{ij,pq,1}\right] \\
&\quad - 2L_{ij,2}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,2}B_{ij,pq,1}\right] \Bigg)
\end{aligned}
\tag{232}
$$

Using the independence of random processes in (88), it can be simplified:

$$
\begin{aligned}
e(\mathbf{L}_1, \mathbf{L}_2) = \sum_{1\le i,j\le (n/P)} &\Bigg(L_{ij,1}^2\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}^2\right] + L_{ij,2}^2\, \mathbb{E}\left[\widetilde{B}_{ij,pq,2}^2\right] + \left(1 - 2L_{ij,1}\right) \mathbb{E}\left[B_{ij,pq,1}^2\right] \\
&+ 2L_{ij,1}L_{ij,2}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] - 2L_{ij,2}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,2}B_{ij,pq,1}\right] \Bigg)
\end{aligned}
\tag{233}
$$

Its gradient is given by:

$$
\frac{\partial e}{\partial L_{ij,1}} = 2\left(L_{ij,1}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}^2\right] - \mathbb{E}\left[B_{ij,pq,1}^2\right] + L_{ij,2}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] \right) \tag{234a}
$$

$$
\frac{\partial e}{\partial L_{ij,2}} = 2\left(L_{ij,2}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,2}^2\right] + L_{ij,1}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] - \mathbb{E}\left[\widetilde{B}_{ij,pq,2}B_{ij,pq,1}\right] \right) \tag{234b}
$$

Setting the gradient of $e(\mathbf{L}_1, \mathbf{L}_2)$ to zero, we get:

$$
\begin{aligned}
&2\left(L_{ij,1}^{\mathrm{opt}}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}^2\right] - \mathbb{E}\left[B_{ij,pq,1}^2\right] + L_{ij,2}^{\mathrm{opt}}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] \right) = 0 \\
\Leftrightarrow\ & L_{ij,1}^{\mathrm{opt}}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}^2\right] + L_{ij,2}^{\mathrm{opt}}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] = \mathbb{E}\left[B_{ij,pq,1}^2\right]
\end{aligned}
\tag{235}
$$

and:

$$
\begin{aligned}
&2\left(L_{ij,2}^{\mathrm{opt}}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,2}^2\right] + L_{ij,1}^{\mathrm{opt}}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] - \mathbb{E}\left[\widetilde{B}_{ij,pq,2}B_{ij,pq,1}\right] \right) = 0 \\
\Leftrightarrow\ & L_{ij,2}^{\mathrm{opt}}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,2}^2\right] + L_{ij,1}^{\mathrm{opt}}\, \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] = \mathbb{E}\left[\widetilde{B}_{ij,pq,2}B_{ij,pq,1}\right]
\end{aligned}
\tag{236}
$$

Combining these equations, we get:

$$L^{\text{opt}}_{ij,1} = \frac{N_{ij,1}}{D_{ij}} \tag{237a}$$

$$L^{\text{opt}}_{ij,2} = \frac{N_{ij,2}}{D_{ij}} \tag{237b}$$

with:

$$N_{ij,1} = \mathbb{E}\left[B^2_{ij,pq,1}\right]\mathbb{E}\left[\widetilde{B}^2_{ij,pq,2}\right] - \mathbb{E}\left[\widetilde{B}_{ij,pq,2}B_{ij,pq,1}\right]\mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] \tag{238a}$$

$$N_{ij,2} = \mathbb{E}\left[\widetilde{B}_{ij,pq,2}B_{ij,pq,1}\right]\mathbb{E}\left[\widetilde{B}^2_{ij,pq,1}\right] - \mathbb{E}\left[B^2_{ij,pq,1}\right]\mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right] \tag{238b}$$

$$D_{ij} = \mathbb{E}\left[\widetilde{B}^2_{ij,pq,1}\right]\mathbb{E}\left[\widetilde{B}^2_{ij,pq,2}\right] - \mathbb{E}\left[\widetilde{B}_{ij,pq,1}\widetilde{B}_{ij,pq,2}\right]^2 \tag{238c}$$

Part III.

Homogeneous and isotropic filtering

12. Ergodicity assumption

12.1. Principle

The previous filtering equations are fully general, since no assumption has been made on the practical way of computing the expectations $\mathbb{E}[\cdot]$. However, it is then necessary to make an assumption about the estimation of these expectations from available data (often a single realization), in order to get usable formulations. Various ergodicity assumptions could be made, based on the location, the scale, or the coordinate of any base that can decompose the covariances.

Importantly, the filtering parameters must be consistent with the ergodicity assumption.

12.2. Averaging operator

To estimate the statistical expectations $\mathbb{E}[\cdot]$ in previous formulae, we choose the simplest method: a spatial and angular ergodicity assumption, leading to a homogeneous and isotropic filtering. The spatial and angular averages are combined into a single discrete operator $\lceil \cdot \rfloor$ that can be applied either to a vector or to a matrix:

- for a vector $\mathbf{a}$, $\lceil \mathbf{a} \rfloor$ is a scalar giving the spatial average of $\mathbf{a}$,
- for a matrix $\mathbf{A}$, $\lceil \mathbf{A} \rfloor$ is a vector whose component $\lceil \mathbf{A} \rfloor_m$ gives the spatial and angular average of the matrix $\mathbf{A}$ for the separation class m.

Using only an angular ergodicity assumption (not spatial) to get locally varying diagnostics is also possible. The diagnostic is performed for a selection of points and values are interpolated to all points with a nearest neighbor method.

12.3. Random sampling

The practical computation of the spatial and angular average can be achieved via a random sampling.

For each separation class m, a set of N_{sam} couples of random points is defined. For each couple, the separation d is such as $d_m \leq d < d_{m+1}$, where d_m and d_{m+1} are the distance bounds of separation class m. Centered moments are estimated for each couple, and moments are averaged over the N_{sam} couples for each class.

The number of couples N_{sam} is directly linked to the quality and the smoothness of the estimated spatial and angular average. It has been found experimentally that a number of couples $N_{\text{sam}} = 500$ at least is required to get relevant localization results, $N_{\text{sam}} = 5000$ giving sufficiently precise results.

12.4. Homogeneous sampling

To ensure an homogeneous sampling for a grid with varying cell areas, it is necessary to inverse the cell areas CDF denoted ϕ:

- let $\mathbf{c}$ be the vector of cell areas: c_i is the cell area of the grid point of index i,
- the PDF $\mathbf{n}$ contains the relative cell areas: $n_i = \dfrac{c_i}{\sum_j c_j}$,
- the CDF $\boldsymbol{\phi}$ is defined as:

$$\phi_1 = n_1 \tag{239a}$$
$$\text{for } i > 1\ ,\ \phi_i = \phi_{i-1} + n_i \tag{239b}$$

- draw a random number r from a uniform distribution in $[0, 1]$,
- the index i of the selected grid point is such as $|\phi_i - r|$ is minimized, with $\phi_i > r$ (a dichotomy method is very efficient since $\phi_{i+1} \geq \phi_i$).

13. Variance filtering

13.1. Issues with the explicit optimal filter

Using the explicit expression of the optimal filter (131), as in Raynaud *et al.* (2009), might lead to severe issues:

- in general, the estimation of $\mathrm{Cov}\left(\widetilde{\mathbf{v}}\right)$ is noisy, which makes its inversion unstable,
- the optimal filter $\mathbf{F}^{\mathrm{opt}}$ can have negative lobes, which might lead to negative filtered variances, especially if the sample variance signal shows strong gradients.

As a consequence, another approach is used here: a positive parametric filter $\mathbf{F}$ is choosed and its parameters are optimized to partially satisfy the optimality conditions. This method ensures the positivity of filtered variances and provide a robust - although suboptimal - estimation of the filter.

13.2. Simplified optimization

We define the filter $\mathbf{F}$ as a homogeneous and isotropic kernel parametrized with two parameters only: its amplitude a and its length-scale $\mathcal{L}^f$.

At optimality, equation (132b) is transformed into:

$$\left\lceil \widehat{\mathbf{v}}^{\mathrm{opt}} \right\rfloor = \left\lceil \widetilde{\mathbf{v}} \right\rfloor \tag{240}$$

which means that the filter kernel $\mathbf{F}$ must preserve the spatial mean. Thus, its amplitude a is directely derived from its length-scale $\mathcal{L}^f$, which is the only parameter to optimize. This length-scale has to be:

- large enough to significantly remove the sampling noise,
- small enough to keep the signal of interest (i.e. the asymptotic sample variances).

Equation (132a) is transformed into:

$$\left\lceil \widehat{\mathbf{v}}^{\mathrm{opt}} \widetilde{\mathbf{v}}^{\mathrm{T}} \right\rfloor = \left\lceil \mathbf{v}\mathbf{v}^{\mathrm{T}} \right\rfloor \tag{241}$$

which provides a series of conditions to verify in the optimization of $\mathcal{L}^f$ (one for each separation class). The problem is overdetermined (M constraints for one parameter), so we choose to optimize $\mathcal{L}^f$ by only verifying the condition for the zero-separation class m_0, which is equivalent to the spatial average of a vector:

$$\begin{aligned} & \left\lceil \widehat{\mathbf{v}}^{\mathrm{opt}} \widetilde{\mathbf{v}}^{\mathrm{T}} \right\rfloor_{m_0} = \left\lceil \mathbf{v}\mathbf{v}^{\mathrm{T}} \right\rfloor_{m_0} \\ \Leftrightarrow \ & \left\lceil \widehat{\mathbf{v}}^{\mathrm{opt}} \circ \widetilde{\mathbf{v}} \right\rfloor = \left\lceil \mathbf{v} \circ \mathbf{v} \right\rfloor \end{aligned} \tag{242}$$

The expected product of asymptotic sample variances can be derived from equation (97b):

$$\left\lceil \mathbf{v} \circ \mathbf{v} \right\rfloor = P_{20}(N) \left\lceil \widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}} \right\rfloor + P_9(N) \left\lceil \widetilde{\varsigma} \right\rfloor \tag{243}$$

where $\widetilde{\boldsymbol{\varsigma}} \in \mathbb{R}^n$ is a vector whose components are given by $\widetilde{\varsigma}_i = \widetilde{\xi}_{ii} = \widetilde{\Xi}_{iiii}$, and with:

$$P_{20}(N) = P_7(N) + 2P_8(N) = \frac{(N-1)(N^2-3N+3)}{N(N-2)(N-3)} \tag{244}$$

In the case of a Gaussian distributed ensemble, equation (101b) is used instead:

$$\lceil \mathbf{v} \circ \mathbf{v} \rfloor = P_{21}(N) \lceil \widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor \tag{245}$$

with:

$$P_{21}(N) = P_{12}(N) + 2P_{13}(N) = \frac{N-1}{N+1} \tag{246}$$

13.3. Solution unicity

In this subsection, we consider the filtered variances $\widehat{\mathbf{v}}$ as a function of the filtering length-scale $\mathcal{L}^f$ and try to find the approriate $\mathcal{L}^f$ for which equation (242) is verified.

Sample and filtered variances $\widetilde{\mathbf{v}}$ and $\widehat{\mathbf{v}}$ counterparts in spectral space are respectively denoted $\widetilde{\mathrm{s}}$ and $\widehat{\mathrm{s}}$. From Berre (2000), we know that a homogeneous and isotropic filtering in grid-point space is equivalent to a diagonal filter in spectral space:

$$\widehat{s}_k = \rho_k \widetilde{s}_k \tag{247}$$

where ρ_k are real coefficients of the spectral filter, lying between 0 and 1, with $\rho_0 = 1$ to preserve the spatial average as required by equation (240). The Plancherel formula gives the spatial average of a Schur product of two grid-point vectors as a spectral average of a Schur product of their spectral counterparts:

$$\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor = \sum_{k=0}^{K} \widehat{s}_k \widetilde{s}_k^* = \sum_{k=0}^{K} \rho_k \left| \widetilde{s}_k \right|^2 \tag{248}$$

where K is the maximum wavenumber index and * denotes the conjugate complex number. Thus, its derivative with respect to ρ_k is non-negative:

$$\frac{\partial}{\partial \rho_k} \left(\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor \right) = \left| \widetilde{s}_k \right|^2 \geq 0 \tag{249}$$

With usual filtering kernels, we can check that for $k > 0$:

$$\frac{\mathrm{d}\rho_k}{\mathrm{d}\mathcal{L}^f} < 0 \tag{250}$$

since for increasing filtering intensity (increasing $\mathcal{L}^f$), the ρ_k coefficients go to 0. For instance, in a continuous case, the Fourier transform of a Gaussian filter of length-scale $\mathcal{L}^f$ is given by:

$$\rho_k^G \left(\mathcal{L}^f \right) = \exp \left(-\frac{\left(\mathcal{L}^f k \right)^2}{2} \right) \tag{251}$$

so that:

$$\frac{\mathrm{d}\rho^G_k}{\mathrm{d}\mathcal{L}^f} = -\mathcal{L}^f k^2 \exp\left(-\frac{(\mathcal{L}^f k)^2}{2}\right) < 0 \tag{252}$$

As a consequence, the derivative of $\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor$ with respect to $\mathcal{L}^f$ is negative:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\mathcal{L}^f}\left(\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor\right) &= \sum_{k=1}^{K} \frac{\partial}{\partial \rho_k}\left(\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor\right) \frac{\mathrm{d}\rho_k}{\mathrm{d}\mathcal{L}^f} \\ &= \sum_{k=1}^{K} \left|\widetilde{s}_k\right|^2 \frac{\mathrm{d}\rho_k}{\mathrm{d}\mathcal{L}^f} < 0\end{aligned} \tag{253}$$

which proves that $\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor$ is a strictly decreasing function of $\mathcal{L}^f$.

From equations (5a) and (5b) used to compute $\widetilde{\mathbf{v}}$ and $\widetilde{\boldsymbol{\varsigma}}$ respectively, we can get:

$$(N-1)\widetilde{v}_i = \sum_{p=1}^{N} \left(\widetilde{x}^b_{i,p} - \left\langle \widetilde{x}^b_i \right\rangle\right)^2 \tag{254}$$

$$N\widetilde{\varsigma}_i = \sum_{p=1}^{N} \left(\widetilde{x}^b_{i,p} - \left\langle \widetilde{x}^b_i \right\rangle\right)^4 \tag{255}$$

and from the convexity property:

$$\sum_{p=1}^{N} \left(\widetilde{x}^b_{i,p} - \left\langle \widetilde{x}^b_i \right\rangle\right)^4 \geq \left(\sum_{p=1}^{N} \left(\widetilde{x}^b_{i,p} - \left\langle \widetilde{x}^b_i \right\rangle\right)^2\right)^2 \tag{256}$$

hence:

$$\widetilde{\varsigma}_i \geq \frac{(N-1)^2}{N} \widetilde{v}_i^2 \tag{257}$$

so that:

$$\lceil \widetilde{\boldsymbol{\varsigma}} \rfloor \geq \frac{(N-1)^2}{N} \lceil \widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor \tag{258}$$

Since $P_9(N) < 0$, equation (243) gives:

$$\lceil \mathbf{v} \circ \mathbf{v} \rfloor \leq P_{22}(N) \lceil \widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor \tag{259}$$

with:

$$P_{22}(N) = P_{20}(N) + P_9(N)\frac{(N-1)^2}{N} = -\frac{(N-1)(2N-3)}{N(N-2)(N-3)} \tag{260}$$

If there is no filtering ($\mathcal{L}^f = 0$), then $\widehat{\mathbf{v}} = \widetilde{\mathbf{v}}$. Since $P_{22}(N) < 1$, we have in this case:

$$\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor = \lceil \widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor > P_{22}(N) \lceil \widetilde{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor \geq \lceil \mathbf{v} \circ \mathbf{v} \rfloor \tag{261}$$

Thus, for a zero filtering lenght-scale, $\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor$ is larger than the right-hand side of equation (242).

As a conclusion, the right-hand side of equation (242) is obtained for a unique value of the filtering length-scale $\mathcal{L}^f$. Obviously, this result is also valid for the case of a Gaussian distributed ensemble.

13.4. Iterative method

We have shown that $\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor$ is a strictly decreasing function of the filtering length-scale $\mathcal{L}^f$, whose value for $\mathcal{L}^f = 0$ is larger than the right-hand side of equation (242). Thus, the optimal lenght-scale for which equation (242) is true can be found iteratively, for instance with a dichotomy method as in the following pseudo-code:

1. Compute the right-hand side of equation (242)
2. Choose a low-pass filter $\mathbf{F}$ preserving the average $\lceil \cdot \rfloor$, and a large enough initial filtering length-scale $\mathcal{L}^f > \mathcal{L}^{f,\mathrm{opt}}$. Initialize the step $\mathrm{d}\mathcal{L}^f$ at $\mathcal{L}^f/2$.
3. For a certain number of iterations:
   - Filter the sample variances $\widehat{\mathbf{v}}$ with $\mathbf{F}\left(\mathcal{L}^f\right)$ to get $\widehat{\mathbf{v}}$.
   - Compute $\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor$.
   - If $\lceil \widehat{\mathbf{v}} \circ \widetilde{\mathbf{v}} \rfloor$ is larger (resp. small) than the right-hand side of equation (242), then increase (resp. decrease) the length-scale $\mathcal{L}^f$ by the step $\mathrm{d}\mathcal{L}^f$.
   - Update the step: $\mathrm{d}\mathcal{L}^f = \mathrm{d}\mathcal{L}^f/2$

In practice, about 10 iterations are enough to get a satisfactory precision on $\mathcal{L}^f$. For instance, with this basic dichotomy algorithm, if the maximum filtering length-scale is estimated at 500 km, then the starting length-scale is set to 250 km, and after 10 iterations the obtained accuracy is $\mathrm{d}\mathcal{L}^f = 250/2^{10} \simeq 0.24$ km. This is likely to be satisfactory in practice, even if more iterations could be done until the machine precision is reached.

14. Static covariance matrix specification

To compute hybridization coefficients and a localization adapted to the hybridization, it is necessary to specify the spatial and angular average of the static covariance matrix $\overline{\mathbf{B}}$, i.e. $\lceil\overline{\mathbf{B}}\rfloor$. For some heterogeneous and anisotropic models of $\overline{\mathbf{B}}$, this task might be very complicated. However, there is a simple two-step solution via a randomization of $\overline{\mathbf{B}}$:

1. Generate a pseudo-ensemble $\left\{\bar{\mathbf{x}}^b_1, \ldots, \bar{\mathbf{x}}^b_{\overline{N}}\right\}$ of size $\overline{N}$ from the square-root of $\overline{\mathbf{B}}$ denoted $\overline{\mathbf{U}}$, with $\overline{\mathbf{U}}\overline{\mathbf{U}}^{\mathrm{T}} = \overline{\mathbf{B}}$:

$$\bar{\mathbf{x}}^b_p = \overline{\mathbf{U}}\boldsymbol{\eta}_p \tag{262}$$

   where $\left\{\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_{\overline{N}}\right\}$ is a set of $\overline{N}$ unbiased and uncorrelated random vectors (distributed with a zero mean and a identity covariance matrix). For instance, a Gaussian distribution can be used:

$$\boldsymbol{\eta}_p \sim \mathcal{N}\left(0, \mathbf{I}\right) \tag{263}$$

2. Estimate $\lceil\overline{\mathbf{B}}\rfloor$ from the pseudo-ensemble $\left\{\bar{\mathbf{x}}^b_1, \ldots, \bar{\mathbf{x}}^b_{\overline{N}}\right\}$.

Computing the spatial and angular average by using the same random sampling as for the "real" ensemble $\left\{\widetilde{\mathbf{x}}^b_1, \ldots, \widetilde{\mathbf{x}}^b_N\right\}$ ensures a better consistency of both estimations, at the lowest cost.

Acknowledgements

Many thanks to the careful readers that spotted some typos:

- Mayeul Destouches (CNRM-Météo France, Toulouse, France)
- Zofia Stanley (CIRES, Boulder, Colorado)

References

Berre L. 2000. Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Monthly Weather Review* **128**(3): 644–667.

Isserlis L. 1916. On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression. *Biometrika* **11**(3): 185–190.

Isserlis L. 1918. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12**(1/2): 134–139.

Muirhead RI. 2005. *Aspects of multivariate statistical theory*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Rady EHA, Fergany HA, Edress AM. 2005. An easy way to calculate the moments of sample correlation coefficients. *Communications Series A1 Mathematics & Statistics* **54**(1): 21–28.

Raynaud L, Berre L, Desroziers G. 2009. Objective filtering of ensemble-based background-error variances. *Quarterly Journal of the Royal Meteorological Society* **135**(642): 1177–1199.

Wiener N. 1949. *Extrapolation, interpolation, and smoothing of stationary time series, with engineering applications.* Technology Press of the Massachusetts Institute of Technology: Cambridge.

Wishart J. 1928. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A**(1/2): 32–52.