Supplementary Information

This Supplementary Information contains:

1. Mathematical description of the auxiliary model of the SSLM used in the second case study
2. Mathematical description of the auxiliary model used in the third case study on cancer cells
3. Partitioning of the joint likelihood
4. Derivation of the conditional variance
5. MCMC convergence analysis
6. Justification of equation (4)

1. Mathematical description of the auxiliary model of the SSLM used in the second case study

We derive the dynamics of the auxiliary model of the SSLM. Recall that the auxiliary model consists of agents of three types: surviving agents with mark O, new appearing agents with mark $+$ and dead agents with mark $-$. The dynamics of this model are described by the Markov operator

$$L = L_1 + L_2 + L_3$$

where

$$L_1 F(\gamma) = \sum_{x \in \gamma_O \cup \gamma_+} \int_{\mathbb{R}^d} a^+(x-y) \left[F(\gamma_O, \gamma_+ \cup y, \gamma_-) - F(\gamma)\right] dy,$$

describes that surviving and newly appearing agents produce and disperse offspring. Further,

$$L_2 F(\gamma) = \sum_{x \in \gamma_O} m \left[F(\gamma_O \setminus x, \gamma_+, \gamma_- \cup x) - F(\gamma)\right] + \sum_{x \in \gamma_+} m \left[F(\gamma_O, \gamma_+ \setminus x, \gamma_-) - F(\gamma)\right],$$

where the first summand accounts for surviving agents dying and hence becoming dead agents and the second summand accounts for new agents dying. Finally,

$$\begin{aligned} L_3 F(\gamma) = &\sum_{x \in \gamma_O} \sum_{y \in \gamma_O \setminus x \cup \gamma_+} a^-(x-y) \left[F(\gamma_O \setminus x, \gamma_+, \gamma_- \cup x) - F(\gamma)\right] \\ &+ \sum_{x \in \gamma_+} \sum_{y \in \gamma_+ \setminus x \cup \gamma_O} a^-(x-y) \left[F(\gamma_O, \gamma_+ \setminus x, \gamma_-) - F(\gamma)\right], \end{aligned}$$

where the first and second line refer to surviving and new agents dying due to competition with surviving and new agents. We use Mathematica code to derive the differential equations of the leading terms of the cumulants. For the mean-field density we get

$$\frac{d}{dt} q_{t,O} = -q_{t,O} d_q$$

$$\frac{d}{dt} q_{t,+} = b_q - q_{t,+} d_q$$

$$\frac{d}{dt} q_{t,-} = q_{t,O} d_q,$$

with the death rate $d_q = m + A^-(q_{t,O} + q_{t,+})$ and the birth rate $b_q = A^+(q_{t,O} + q_{t,+})$. Further,

$$\frac{d}{dt} p_{t,O} = -2\pi \int_0^\infty k \tilde{a}^-(k) \left[\tilde{g}_{t,O,O}(k) + \tilde{g}_{t,O,+}(k)\right] dk - \left(p_{t,O} d_q + q_{t,O} d_p\right),$$

$$\frac{d}{dt} p_{t,+} = -2\pi \int_0^\infty k \tilde{a}^-(k) \left[\tilde{g}_{t,O,+}(k) + \tilde{g}_{t,+,+}(k)\right] dk - \left(p_{t,+} d_q + q_{t,+} d_p\right) + b_p,$$

$$\frac{d}{dt} p_{t,-} = 2\pi \int_0^\infty k \tilde{a}^-(k) \left[\tilde{g}_{t,O,O}(k) + \tilde{g}_{t,O,+}(k)\right] dk + p_{t,O} d_q + q_{t,O} d_p,$$

with the correction to the death rate $d_p = A^-(p_{t,O} + p_{t,+})$ and the correction to the birth rate $b_p = A^+(p_{t,O} + p_{t,+})$. Finally,

$$\frac{d}{dt}\tilde{g}_{t,O,O}(k) = -2\big(\tilde{a}^-(k)q_{t,O}\big[\tilde{g}_{t,O,+}(k) + q_{t,O}\big] + \tilde{g}_{t,O,O}(k)\big[\tilde{a}^-(k)q_{t,O} + d_q\big]\big),$$

$$\begin{aligned}\frac{d}{dt}\tilde{g}_{t,O,+}(k) = {}& \tilde{a}^+(k)\big[\tilde{g}_{t,O,O}(k) + \tilde{g}_{t,O,+}(k) + q_{t,O}\big] \\ & - \tilde{a}^-(k)\big[\tilde{g}_{t,+,+}(k)q_{t,O} + \big(\tilde{g}_{t,O,O}(k) + 2q_{t,O}\big)q_{t,+}\big] - \tilde{g}_{t,O,+}(k)\big(\tilde{a}^-(k)\big[q_{t,O} + q_{t,+}\big]\big) \\ & + 2d_q,\end{aligned}$$

$$\frac{d}{dt}\tilde{g}_{t,O,-}(k) = \tilde{a}^-(k)q_{t,O}\big[\tilde{g}_{t,O,+}(k) - \tilde{g}_{t,+,-}(k) + q_{t,O}\big] + \big[\tilde{g}_{t,O,O}(k) - \tilde{g}_{t,O,-}(k)\big]\big[\tilde{a}^-(k)q_{t,O} + d_q\big],$$

$$\begin{aligned}\frac{d}{dt}\tilde{g}_{t,+,+}(k) = {}& -2\big(\tilde{a}^-(k)q_{t,O}\big[\tilde{g}_{t,O,+}(k) + q_{t,O}\big] - \tilde{a}^+(k)\big[\tilde{g}_{t,O,+}(k) + \tilde{g}_{t,+,+}(k) + q_{t,O}\big] \\ & + \tilde{g}_{t,+,+}(k)\big[\tilde{a}^-(k)q_{t,+} + d_q\big]\big),\end{aligned}$$

$$\begin{aligned}\frac{d}{dt}\tilde{g}_{t,+,-}(k) = {}& d_q\big[\tilde{g}_{t,O,+}(k) - \tilde{g}_{t,+,-}(k)\big] + \tilde{a}^+(k)\big[\tilde{g}_{t,O,-}(k) - \tilde{g}_{t,+,-}(k)\big] \\ & + \tilde{a}^-(k)\left[q_{t,O}\big(\tilde{g}_{t,O,+}(k) + \tilde{g}_{t,+,+}(k) + q_{t,+}\big) - q_{t,+}\left(\tilde{g}_{t,O,-}(k) + \tilde{g}_{t,+,-}(k)\right)\right],\end{aligned}$$

$$\frac{d}{dt}\tilde{g}_{t,-,-}(k) = 2\big(\tilde{a}^-(k)q_{t,O}\tilde{g}_{t,+,-}(k) + \tilde{g}_{t,O,-}(k)\big[\tilde{a}^-(k)q_{t,O} + d_q\big]\big).$$

2. Mathematical description of the auxiliary model used in the third case study on cancer cells

For the auxiliary model of the cancer cells, we denote the configurations $\gamma = (\gamma_S, \gamma_R) = (\gamma_{S,O}, \gamma_{S,+}, \gamma_{S,-}, \gamma_{R,O}, \gamma_{R,+}, \gamma_{R,-})$ where the first subscript refers to the cell type and the second subscript to the auxiliary type. The dynamics of the auxiliary model are described by the Markov operator

$$L = L_1 + L_2 + L_3,$$

where

$$L_1F(\gamma) = \sum_{\substack{M,N\in\{S,R\}\\ M\neq N}} \sum_{x\in\gamma_{M,O}\cup\gamma_{M,+}} \int_{\mathbb{R}^d} b_M(x-y)\left[F\big(\gamma_{M,O}, \gamma_{M,+}\cup x, \gamma_{M,-}, \gamma_N\big) - F(\gamma)\right]dy$$

describes birth events of both types of cells. Next,

$$\begin{aligned}L_2F(\gamma) = {}& \sum_{M\in\{S,R\}} \sum_{x\in\gamma_{M,O}} m\left[F\big(\gamma_{M,O}\setminus x, \gamma_{M,+}, \gamma_{M,-}\cup x\big) - F(\gamma)\right] \\ & + \sum_{x\in\gamma_{M,+}} m\left[F\big(\gamma_{M,O}, \gamma_{M,+}\setminus x, \gamma_{M,-}\big) - F(\gamma)\right],\end{aligned}$$

which describes the density-independent mortality. Finally,

$$L_3F(\gamma) = \sum_{\substack{M,N\in\{S,R\}\\ M\neq N}} \sum_{x\in\gamma_{M,O}} \left(\sum_{y\in\gamma_{M,O}\setminus x\cup\gamma_{M,+}} a_{MM}(x-y) + \sum_{z\in\gamma_{N,O}\cup\gamma_{N,+}} a_{MN}(x-z)\right)$$

$$[F(\gamma_{M,O} \setminus x, \gamma_{M,+}, \gamma_{M,-} \cup x, \gamma_N) - F(\gamma_M, \gamma_N)]$$
$$+ \sum_{x \in \gamma_{M,+}} \left(\sum_{y \in \gamma_{M,O} \setminus x \cup \gamma_{M,+}} a_{MM}(x-y) + \sum_{z \in \gamma_{N,O} \cup \gamma_{N,+}} a_{MN}(x-z) \right)$$
$$[F(\gamma_{M,O}, \gamma_{M,+} \setminus x, \gamma_{M,-}, \gamma_N) - F(\gamma_M, \gamma_N)],$$

which accounts for surviving agents dying (and becoming dead agents) and new agents dying (and disappearing) by intra- and interspecific interactions. We derive the differential equations for the leading terms of the cumulants for this model. For agents of type $M \in \{S, R\}$, the evolution of the mean-field densities is

$$\frac{d}{dt} q_{t,M,O} = -q_{t,M,O} d_{q,M}$$

$$\frac{d}{dt} q_{t,M,+} = B_M(q_{t,M,O} + q_{t,M,+}) - q_{t,M,+} d_{q,M},$$

$$\frac{d}{dt} q_{t,M,-} = q_{t,M,-} d_{q,M},$$

with the death term $d_{q,M} = m_M + A_{MM}(q_{t,M,O} + q_{t,M,+}) + A_{MN}(q_{t,N,O} + q_{t,N,+})$ consisting of density-independent death, and intra- & interspecific interaction. Next, the evolution of the first-order correction to the densities is given by

$$\frac{d}{dt} p_{t,M,O} = -2\pi \int_0^\infty k \tilde{a}_{MM}(k) [\tilde{g}_{t,M,O,M,O}(k) + \tilde{g}_{t,M,O,M,+}(k)]$$
$$- 2\pi \int_0^\infty k \tilde{a}_{MN}(k) [\tilde{g}_{t,M,O,N,O}(k) + \tilde{g}_{t,M,O,N,+}(k)] - d_{M,p} \ ,$$

with the correction to the death term

$$d_{M,p} = m_M p_{t,M,O} + A_{MM}[(p_{t,M,+} q_{t,M,O} + p_{t,M,O} q_{t,M,+}) + 2 p_{t,M,O} q_{t,M,O}]$$
$$+ A_{MN}[(p_{t,N,O} q_{t,M,O} + p_{t,M,O} q_{t,N,O}) + (p_{t,N,+} q_{t,M,O} + p_{t,M,O} q_{t,N,+})],$$

Finally, recall that $M \in \{S, R\}$ and let $N \in \{S, R\}$ with $N \neq M$, then the evolution of the leading term of the two-point cumulants is

$$\frac{d}{dt} \tilde{g}_{t,M,O,M,O}(k)$$
$$= -2 \Big(\tilde{g}_{t,M,O,M,O}(k) [d_{q,M} + \tilde{a}_{MM}(k) q_{t,M,O}]$$
$$+ q_{t,M,O} \Big[\tilde{a}_{MN}(k) [\tilde{g}_{t,M,O,N,O}(k) + \tilde{g}_{t,M,O,N,+}(k)]$$
$$+ \tilde{a}_{MM}(k) [\tilde{g}_{t,M,O,M,+}(k) + q_{t,M,O}] \Big] \Big),$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,O,M,+}(k) \\
&= -2d_{q,M}\tilde{g}_{t,M,O,M,+}(k) \\
&- q_{t,M,O}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,M,+}(k) + \tilde{g}_{t,M,+,M,+}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,M,+,N,O}(k) + \tilde{g}_{t,M,+,N,+}(k)\big]\big) \\
&+ \tilde{b}_M(k)\big[\tilde{g}_{t,M,O,M,O}(k) + \tilde{g}_{t,M,O,M,+}(k) + q_{t,M,O}\big] \\
&- q_{t,M,+}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,M,O}(k) + \tilde{g}_{t,M,O,M,+}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,M,O,N,O}(k) + \tilde{g}_{t,M,O,N,+}(k)\big]\big) - 2\tilde{a}_{MM}(k)q_{t,M,O}q_{t,M,+},
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,O,M,-}(k) \\
&= d_{q,M}\left(\tilde{g}_{t,M,O,M,O}(k) + \tilde{g}_{t,M,O,M,-}(k)\right) \\
&+ q_{t,M,O}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,M,O}(k) + \tilde{g}_{t,M,O,M,+}(k) - \tilde{g}_{t,M,O,M,-}(k) \\
&- \tilde{g}_{t,M,+,M,-}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,M,O,N,O}(k) + \tilde{g}_{t,M,O,N,+}(k) - \tilde{g}_{t,M,-,N,O}(k) - \tilde{g}_{t,M,-,N,+}(k)\big]\big) \\
&+ \tilde{a}_{MM}(k)q_{t,M,O}^2,
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,+,M,+}(k) \\
&= 2\tilde{b}_M(k)\big(\tilde{g}_{t,M,O,M,+}(k) + \tilde{g}_{t,M,+,M,+}(k) + q_{t,M,+}\big) \\
&- 2\tilde{g}_{t,M,+,M,+}(k)\big(d_{q,M} + \tilde{a}_{MM}(k)q_{t,M,+}\big) \\
&- 2q_{t,M,+}\big(\tilde{a}_{MN}(k)\big[\tilde{g}_{t,M,+,N,O}(k) + \tilde{g}_{t,M,+,N,+}(k)\big] \\
&+ \tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,M,+}(k) + q_{t,M,+}\big]\big),
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,+,M,-}(k) \\
&= \tilde{b}_M(k)\left(\tilde{g}_{t,M,O,M,-}(k) + \tilde{g}_{t,M,+,M,-}(k)\right) \\
&+ d_{q,M}\left(\tilde{g}_{t,M,O,M,+}(k) - \tilde{g}_{t,M,+,M,-}(k)\right) \\
&+ q_{t,M,O}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,M,+}(k) + \tilde{g}_{t,M,+,M,+}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,M,+,N,O}(k) + \tilde{g}_{t,M,+,N,+}(k)\big]\big) \\
&- q_{t,M,+}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,M,-}(k) + \tilde{g}_{t,M,+,M,-}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,M,-,N,O}(k) + \tilde{g}_{t,M,-,N,+}(k)\big]\big) + \tilde{a}_{MM}(k)q_{t,M,O}q_{t,M,+},
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,-,M,-}(k) \\
&= 2d_{q,M}\tilde{g}_{t,M,O,M,-}(k) \\
&+ 2q_{t,M,O}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,M,-}(k) + \tilde{g}_{t,M,+,M,-}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,M,-,N,O}(k) + \tilde{g}_{t,M,-,N,+}(k)\big]\big),
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,O,N,O}(k)
&= -q_{t,M,O}\big(\tilde{a}_{MM}(k)\tilde{g}_{t,M,+,N,O}(k) + \tilde{a}_{MN}(k)\big[\tilde{g}_{t,N,O,N,O}(k) + \tilde{g}_{t,N,O,N,+}(k)\big]\big) \\
&- q_{t,N,O}\big(\tilde{a}_{NN}(k)\tilde{g}_{t,M,O,N,+}(k) + \tilde{a}_{NM}(k)\big[\tilde{g}_{t,M,O,M,O}(k) + \tilde{g}_{t,M,O,M,+}(k)\big]\big) \\
&- q_{t,M,O}q_{t,N,O}\big(\tilde{a}_{MN}(k) + \tilde{a}_{NM}(k)\big) \\
&- \tilde{g}_{t,M,O,N,O}(k)\big[d_{q,M} + d_{q,N} + \tilde{a}_{MM}(k) + \tilde{a}_{NN}(k)\big],
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,O,N,+}(k)
&= \tilde{b}_N(k)\left(\tilde{g}_{t,M,O,N,O}(k) + \tilde{g}_{t,M,O,N,+}(k)\right) \\
&- q_{t,M,O}\big(\tilde{a}_{MM}(k)\tilde{g}_{t,M,+,N,+}(k) + \tilde{a}_{MN}(k)\big[\tilde{g}_{t,N,O,N,+}(k) + \tilde{g}_{t,N,+,N,+}(k)\big]\big) \\
&- q_{t,N,+}\big(\tilde{a}_{NN}(k)\tilde{g}_{t,M,O,N,+}(k) + \tilde{a}_{NM}(k)\big[\tilde{g}_{t,M,O,M,O}(k) + \tilde{g}_{t,M,O,M,+}(k)\big]\big) \\
&- q_{t,M,O}q_{t,N,+}\big(\tilde{a}_{MN}(k) + \tilde{a}_{NM}(k)\big) \\
&- \tilde{g}_{t,M,O,N,+}(k)\big(d_{q,M} + d_{q,N} + \tilde{a}_{MM}(k)q_{t,M,O} + \tilde{a}_{NN}(k)q_{t,N,+}\big),
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,O,N,-}(k)
&= d_{q,N}\tilde{g}_{t,M,O,N,O}(k) - d_{q,M}\tilde{g}_{t,M,O,N,-}(k) \\
&- q_{t,M,O}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,N,-}(k) + \tilde{g}_{t,M,+,N,-}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,N,O,N,-}(k) + \tilde{g}_{t,N,+,N,-}(k)\big]\big) \\
&+ q_{t,N,O}\big(\tilde{a}_{NN}(k)\big[\tilde{g}_{t,M,O,N,O}(k) + \tilde{g}_{t,M,O,N,+}(k)\big] \\
&+ \tilde{a}_{NM}(k)\big[\tilde{g}_{t,M,O,M,O}(k) + \tilde{g}_{t,M,O,M,+}(k)\big]\big) + q_{t,M,O}q_{t,N,O}\tilde{a}_{NM}(k),
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,+,N,+}(k)
&= \tilde{b}_N(k)\left(\tilde{g}_{t,M,+,N,O}(k) + \tilde{g}_{t,M,+,N,+}(k)\right) \\
&+ \tilde{b}_M(k)\left(\tilde{g}_{t,M,O,N,+}(k) + \tilde{g}_{t,M,+,N,+}(k)\right) - \left(d_{q,M} + d_{q,N}\right)\tilde{g}_{t,M,+,N,+}(k) \\
&- q_{t,M,+}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,N,+}(k) + \tilde{g}_{t,M,+,N,+}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,N,O,N,+}(k) + \tilde{g}_{t,N,+,N,+}(k)\big]\big) \\
&- q_{t,N,+}\big(\tilde{a}_{NN}(k)\big[\tilde{g}_{t,M,+,N,O}(k) + \tilde{g}_{t,M,+,N,+}(k)\big] \\
&+ \tilde{a}_{NM}(k)\big[\tilde{g}_{t,M,O,M,+}(k) + \tilde{g}_{t,M,+,M,+}(k)\big]\big) - q_{t,M,+}q_{t,N,+}\big(\tilde{a}_{MN}(k) + \tilde{a}_{NM}(k)\big),
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,+,N,-}(k)
&= d_{q,N}\tilde{g}_{t,M,+,N,O}(k) - d_{q,M}\tilde{g}_{t,M,+,N,-}(k) \\
&+ \tilde{b}_M(k)\left(\tilde{g}_{t,M,O,N,-}(k) + \tilde{g}_{t,M,+,N,-}(k)\right) \\
&- q_{t,M,+}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,N,-}(k) + \tilde{g}_{t,M,+,N,-}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,N,O,N,-}(k) + \tilde{g}_{t,N,+,N,-}(k)\big]\big) \\
&+ q_{t,N,O}\big(\tilde{a}_{NN}(k)\big[\tilde{g}_{t,M,+,N,O}(k) + \tilde{g}_{t,M,+,N,+}(k)\big] \\
&+ \tilde{a}_{NM}(k)\big[\tilde{g}_{t,M,O,M,+}(k) + \tilde{g}_{t,M,+,M,+}(k)\big]\big) + q_{t,M,+}q_{t,N,+}\tilde{a}_{NM}(k),
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\tilde{g}_{t,M,-,N,-}(k) \\
&= d_{q,M}\tilde{g}_{t,M,O,N,-}(k) + d_{q,N}\tilde{g}_{t,M,-,N,O}(k) \\
&+ q_{t,M,O}\big(\tilde{a}_{MM}(k)\big[\tilde{g}_{t,M,O,N,-}(k) + \tilde{g}_{t,M,+,N,-}(k)\big] \\
&+ \tilde{a}_{MN}(k)\big[\tilde{g}_{t,N,O,N,-}(k) + \tilde{g}_{t,N,+,N,-}(k)\big]\big) \\
&+ q_{t,N,O}\big(\tilde{a}_{NN}(k)\big[\tilde{g}_{t,M,-,N,O}(k) + \tilde{g}_{t,M,-,N,+}(k)\big] \\
&+ \tilde{a}_{NM}(k)\big[\tilde{g}_{t,M,O,M,-}(k) + \tilde{g}_{t,M,+,M,-}(k)\big]\big).
\end{aligned}
$$

3. Partitioning of the joint likelihood

The ordering of the grid cells may numerically effect the results of the likelihood calculations with Eq. 1-3. Here we tested and show in Fig.1 the log-likelihood profiles of Eq. 5-7 using a fixed ordering of the grid cells (black line) and with three random orderings (coloured). The numerical differences of the profiles are negligible.

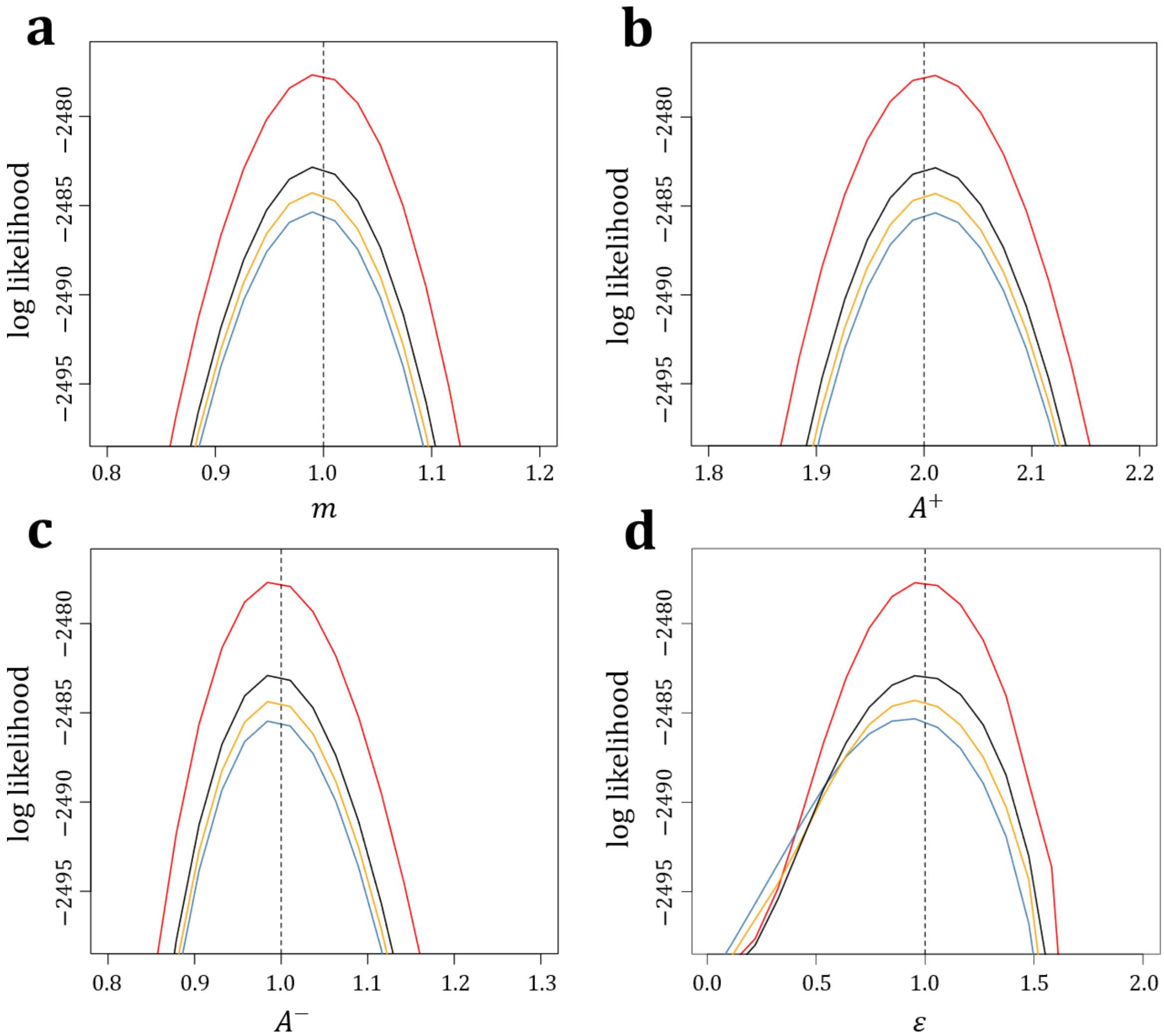

Figure 1: Log-likelihood profiles of the model parameters. *We use a fixed ordering of the grid cells in Eq. 5-7 (black) and three random orderings (coloured). Profiles are calculated varying one model parameter and using true parameter values for all other*

model parameters. ***a****, Profile for density-independent mortality* m*.* ***b****, Profile for reproduction and dispersal rate* A^+*.* ***c****, Profile for competition rate* A^-*.* ***d****, Profile for scaling of interaction kernels* ε*.*

4. Derivation of the conditional variance

Let $\mathbf{Z} = (Z_1, \dots, Z_n)^{\mathrm{T}}$ be a multivariate random variable with $E[\mathbf{Z}] = \boldsymbol{\mu}$ and the variance-covariance matrix $\boldsymbol{\Sigma}$. For any $k \times n$ dimensional matrix $\mathbf{U}$ of rank k and vector $\mathbf{b}$ of length k let $\mathbf{UZ} = \mathbf{b}$, $\mathbf{A} = \mathbf{I}_n - \mathbf{CU}$ such that $\boldsymbol{Z} = \mathbf{AZ} + \mathbf{CUZ}$ where $\mathbf{C} = \boldsymbol{\Sigma}\mathbf{U}^T(\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T)^{-1}$. Then,

$$\mathrm{Var}(\mathbf{Z}|\mathbf{UZ} = \mathbf{b}) = \mathrm{Var}(\mathbf{AZ} + \mathbf{CUZ}|\mathbf{UZ} = \mathbf{b})$$

$$= \mathrm{Var}(\mathbf{AZ}|\mathbf{UZ} = \mathbf{b}) + \mathrm{Var}(\mathbf{CUZ}|\mathbf{UZ} = \mathbf{b}) + \mathrm{Cov}(\mathbf{AZ}, \mathbf{CUZ}) + \mathrm{Cov}(\mathbf{CUZ}, \mathbf{AZ})$$

$$= \mathrm{Var}(\mathbf{AZ}|\mathbf{UZ} = \mathbf{b}) + \mathrm{Var}(\mathbf{CUZ}|\mathbf{UZ} = \mathbf{b}) + \mathbf{C}\mathrm{Cov}(\mathbf{AZ}, \mathbf{UZ}) + \mathrm{Cov}(\mathbf{UZ}, \mathbf{AZ})\mathbf{C}^T$$

Since from $\mathrm{Cov}(\mathbf{AZ}, \mathbf{UZ}) = 0$, if $\mathbf{AZ}$ and $\mathbf{UZ}$ would be independent (as explained in *Methods*) and $\mathrm{Var}(\mathbf{CUZ}|\mathbf{UZ} = \mathbf{b}) = \mathbf{C}\mathrm{Var}(\mathbf{UZ}|\mathbf{UZ} = \mathbf{b})\mathbf{C}^T = 0$, the expression simplifies to

$$\mathrm{Var}(\mathbf{Z}|\mathbf{UZ} = \mathbf{b}) = \mathrm{Var}(\mathbf{AZ}|\mathbf{UZ} = \mathbf{b}) = \mathbf{A}\mathrm{Var}(\mathbf{Z})\mathbf{A}^T,$$

where $\mathrm{Var}(\mathbf{z}) = \mathrm{diag}(\boldsymbol{\Sigma})$.

5. MCMC convergence analysis

Table 1: MCMC convergence analysis for case study 1. For each non-spatial and spatial model, 4 chains were run for 11,000 iterations, of which we discarded the first 1000 iterations during which adaptation was applied, and thinned the remaining by 10 to obtain 1000 samples per chain. The mean and upper confidence interval (in brackets) for potential scale reduction factor are reported.

	Density-independent death rate m	Fecundity rate A^+	Death by competition rate A^-	Length scale of interaction ε
Non-spatial	1.32 (1.78)	1.02 (1.07)	1.42 (1.97)	1.00 (1.00)
Spatial Normal	1.02 (1.05)	1.02 (1.05)	1.02 (1.04)	1.00 (1.01)
Spatial Poisson	1.01 (1.01)	1.01 (1.01)	1.02 (1.02)	1.01 (1.02)

Table 2: MCMC convergence analysis for case study 2. For each non-spatial and spatial model, 4 chains were run for 11,000 iterations, of which we discarded the first 1000 iterations during which adaptation was applied, and thinned the remaining by 10 to obtain 1000 samples per chain. The mean and upper confidence interval (in brackets) for potential scale reduction factor are reported.

a) Unmarked Agents

	Density-independent death rate m	Fecundity rate A^+	Death by competition rate A^-	Length scale of interaction ε
Non-spatial	1.00 (1.01)	1.01 (1.02)	1.04 (1.09)	1.01 (1.01)
Spatial Normal	1.04 (1.10)	1.00 (1.00)	1.03 (1.08)	1.03 (1.08)
Spatial Poisson	1.01 (1.02)	1.03 (1.06)	1.01 (1.03)	1.02 (1.06)
b) Marked Agents				
	Density-independent death rate m	Fecundity rate A^+	Death by competition rate A^-	Length scale of interaction ε
Non-spatial	1.06 (1.15)	1.00 (1.00)	1.12 (1.30)	1.02 (1.07)
Spatial Normal	1.03 (1.06)	1.00 (1.00)	1.04 (1.11)	1.02 (1.06)
Spatial Poisson	1.06 (1.17)	1.00 (1.01)	1.10 (1.27)	1.03 (1.08)

Table 3: MCMC convergence analysis for case study 3. For each of the 9 models, 4 chains were run for 11,000 iterations, of which we discarded the first 1000 iterations during which adaptation was applied, and thinned the remaining by 10 to obtain 1000 samples per chain. If visual inspection of the trace plots suggested that the chains were still in a transient phase, we continued the MCMC algorithm for additional 1000 adaptive iterations and additional 10,000 iterations which thinned by 10 obtained 1000 samples. The mean and upper confidence interval (in brackets) for potential scale reduction factor are reported. Asterisks indicate that one (*) or two (**) chains were discarded due to non-convergence.

	0.0	0.27*	0.8*	2.5	7.4	22.2	66.7**	200.0*	600.0
Fecundity of sensitive cells B_S	1.00 (1.00)	1.00 (1.01)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.99 (0.99)	0.99 (1.00)	1.00 (1.01)
Fecundity of resistant cells B_R	1.00 (1.00)	1.00 (1.01)	1.00 (1.02)	0.99 (1.00)	1.01 (1.04)	1.00 (1.01)	0.99 (1.00)	1.00 (1.00)	1.00 (1.01)
Death by competition rate A_{SS}	1.00 (1.01)	1.00 (1.00)	1.03 (1.06)	1.00 (1.00)	1.00 (1.01)	1.00 (1.01)	1.00 (1.02)	1.00 (1.01)	1.00 (1.00)
Death by competition rate A_{SR}	1.00 (1.00)	1.00 (1.01)	1.00 (1.00)	1.00 (1.00)	1.00 (1.01)	1.00 (1.00)	1.00 (1.01)	1.00 (1.00)	1.00 (1.00)
Death by competition rate A_{RS}	1.00 (1.00)	1.00 (1.00)	1.00 (1.01)	1.00 (1.01)	1.07 (1.17)	1.00 (1.01)	0.99 (0.99)	0.99 (1.00)	1.00 (1.01)
Death by competition rate A_{RR}	1.04 (1.10)	1.00 (1.01)	1.01 (1.03)	0.99 (1.00)	1.07 (1.17)	1.00 (1.01)	1.00 (1.01)	1.00 (1.00)	1.00 (1.00)
Density-independen	1.00 (1.00)	1.00 (1.01)	1.01 (1.03)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.01)	1.00 (1.00)

t mortality m_S									
Density-independent mortality m_R	1.00 (1.01)	1.00 (1.01)	1.00 (1.02)	0.99 (1.00)	1.03 (1.09)	1.00 (1.01)	1.00 (1.00)	1.00 (1.00)	1.00 (1.01)
Length scale of dispersal l_S	1.00 (1.01)	1.00 (1.00)	1.00 (1.01)	1.00 (1.00)	1.00 (1.00)	1.00 (1.01)	1.00 (1.00)	1.00 (1.00)	1.00 (1.01)
Length scale of dispersal l_R	1.04 (1.11)	1.00 (1.00)	1.01 (1.02)	1.00 (1.00)	1.05 (1.13)v	1.00 (1.01)	1.00 (1.01)	1.03 (1.04)	1.00 (1.01)
Length scale of interaction l_A	1.00 (1.00)	1.00 (1.00)	1.01 (1.01)	1.01 (1.02)	1.06 (1.16)	1.00 (1.01)	1.01 (1.04)	1.01 (1.03)	1.00 (1.00)

6 Justfication of formula (4)

6.1 Notations and main statement

We consider the Euclidean space $\mathbb{R}^n$, $n \geq 1$. We interpret elements from $\mathbb{R}^n$ as column-vectors: $X = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. Let $\mathcal{M}_{n,m}$ denote the space of $n \times m$ matrices.

For any two vector spaces $\mathtt{X}$ and $\mathtt{Y}$, we denote by $\mathcal{L}(\mathtt{X}, \mathtt{Y})$ the vector space of linear mappings from $\mathtt{X}$ to $\mathtt{Y}$. In particular, $\mathcal{M}_{n,m} = \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Consider the Hilbert space L^2 of random vectors $X \in \mathbb{R}^n$ with finite second moments: $\mathbb{E}[\|X\|^2] < \infty$ and the inner product given by $\mathbb{E}[X_1^T X_2]$ for $X_1, X_2 \in L^2$ (note that here X_1^T is a row-vector).

Let Y be a random vector in $\mathbb{R}^k$ with $1 \leq k \leq n$. Denote by L^2_Y the closed subspace of L^2 which consists of $\sigma(Y)$-measurable random vectors from L^2, i.e. $X \in L^2_Y$ iff $X \in L^2$ and there exists a Borel mapping $f : \mathbb{R}^k \to \mathbb{R}^n$ with $X = f(Y)$.

Let $Z \in L^2$. It is well-known that $\mathbb{E}[Z \mid Y]$ is the orthogonal projection of Z on L^2_Y, i.e.

$$\mathbb{E}\big[W^T\big(Z - \mathbb{E}[Z \mid Y]\big)\big] = 0 \tag{1}$$

for all $W \in L^2_Y$; moreover,

$$\mathbb{E}\big[\mathbb{E}[Z \mid Y]\big] = \mathbb{E}[Z]. \tag{2}$$

We denote also

$$\widetilde{Y} := Y - \mathbb{E}[Y], \qquad \Sigma_{ZY} := \mathrm{Cov}(Z, Y) \in \mathcal{M}_{n,k},$$
$$\widetilde{Z} := Z - \mathbb{E}[Z], \qquad \Sigma_{YY} := \mathrm{Cov}(Y, Y) \in \mathcal{M}_{k,k}.$$

Let us introduce the mapping $T \in \mathcal{L}(\mathbb{R}^k, \mathcal{M}_{k,k})$ by

$$(Tv)_{ij} = \sum_{m=1}^{k} \mathbb{E}[\widetilde{Y}_i \widetilde{Y}_j \widetilde{Y}_m] v_m, \qquad 1 \leq i, j \leq k. \tag{3}$$

Then, we define

$$S(Y) := \widetilde{Y}\widetilde{Y}^T - \Sigma_{YY} - T(\Sigma_{YY}^{-1}\widetilde{Y}) \in \mathcal{M}_{k,k}, \tag{4}$$

or, entry-wise, for $1 \leq i, j \leq k$,

$$S(Y)_{ij} = \widetilde{Y}_i \widetilde{Y}_j - (\Sigma_{YY})_{ij} - \sum_{m=1}^{k} \mathbb{E}[\widetilde{Y}_i \widetilde{Y}_j \widetilde{Y}_m] (\Sigma_{YY}^{-1}\widetilde{Y})_m. \tag{5}$$

Next, let $\mathrm{vec} : \mathcal{M}_{k,k} \to \mathbb{R}^{k^2}$ be the mapping which stacks the columns of a matrix from $\mathcal{M}_{k,k}$ into a vector from $\mathbb{R}^{k^2}$; namely, for $M \in \mathcal{M}_{k,k}$, one has

$$\mathrm{vec}(M)_{i+k(j-1)} = M_{ij}, \qquad 1 \leq i, j \leq k. \tag{6}$$

We define then

$$s(Y) := \mathrm{vec}(S(Y)) \in \mathbb{R}^{k^2}, \tag{7}$$

Now, we can introduce

$$\Sigma_{ZS} := \mathbb{E}\big[\widetilde{Z}\,(s(Y))^T\big] = \mathrm{Cov}\big(\widetilde{Z}, s(Y)\big) \in \mathcal{M}_{n,k^2} \tag{8}$$

$$\Sigma_{SS} := \mathbb{E}\big(s(Y)(s(Y))^T\big] = \mathrm{Cov}\big(s(Y), s(Y)\big) \in \mathcal{M}_{k^2,k^2}. \tag{9}$$

Theorem 1. *Assume that $\Sigma_{YY} \in \mathcal{M}_{k,k}$ and $\Sigma_{SS} \in \mathcal{M}_{k^2,k^2}$ are invertible matrices. Then*

$$\mathbb{E}[Z \mid Y] = \mathbb{E}[Z] + C(Y - \mathbb{E}[Y]) + Ds(Y) + R_3(Y), \tag{10}$$

where

$$C := \Sigma_{ZY}\big(\Sigma_{YY}\big)^{-1} \in \mathcal{M}_{n,k}, \tag{11}$$

$$D := \Sigma_{ZS}\big(\Sigma_{SS}\big)^{-1} \in \mathcal{M}_{n,k^2}, \tag{12}$$

and $R_3(Y) \in L^2_Y$ is orthogonal to all linear and quadratic polynomials of $\widetilde{Y}$ in L^2_Y. The entries of matrices Σ_{ZS} and Σ_{SS} can be found by formulas

$$(\Sigma_{ZS})_{l,i+j(k-1)} = \mathbb{E}[\widetilde{Z}_l\widetilde{Y}_i\widetilde{Y}_j] - \sum_{r=1}^{k} \mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r]C_{lr}, \tag{13}$$

and

$$\begin{aligned}(\Sigma_{SS})_{i+k(j-1),p+k(q-1)} &= \mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_p\widetilde{Y}_q] - (\Sigma_{YY})_{ij}(\Sigma_{YY})_{pq} \\ &\quad - \sum_{r=1}^{k}\sum_{s=1}^{k} \mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_s]\,\mathbb{E}[\widetilde{Y}_p\widetilde{Y}_q\widetilde{Y}_r]\,(\Sigma_{YY}^{-1})_{rs}.\end{aligned} \tag{14}$$

6.2 Proof of Theorem 1

We consider the orthogonal expansion of the centred (i.e. with zero mean) vector $\mathbb{E}[Z \mid Y] - \mathbb{E}[Z]$ through "powers" of $\widetilde{Y}$:

$$\mathbb{E}[Z \mid Y] - \mathbb{E}[Z] = \sum_{m\geq 1} r_m(Y), \tag{15}$$

where $r_m(Y) \in L^2_Y$ is a "polynomial" of order m of $\widetilde{Y}$ such that

$$\mathbb{E}[r_m(Y)] = \mathbb{E}[r_m(Y)^T r_k(Y)] = 0, \qquad 1 \leq k < m. \tag{16}$$

It will correspond to the orthogonal decomposition of L^2_Y:

$$L_2^Y = \bigoplus_{m\geq 0} H_Y^m = H_Y^0 \oplus H_Y^1 \oplus H_Y^2 \oplus, \ldots \tag{17}$$

where $H_Y^0 := \mathbb{R}^n$ (the space of "constant random vectors") and H_Y^m is a subspace of the space of all polynomial expressions of order m of $\widetilde{Y}$ which orthogonal to all lower order polynomials.

Namely, let $H_Y^1 := \operatorname{span}(Y - \mathbb{E}[Y])$, i.e. H_Y^1 is the subspace of L_Y^2 consisting of all (centred,) random vectors of the form $A(Y - \mathbb{E}[Y])$ for some (deterministic) matrix $A \in \mathcal{M}_{n,k}$.

Lemma 2. *Let $X \in \mathbb{R}^n$ be such that $X \perp A\widetilde{Y}$ for any $A \in \mathcal{M}_{n,k}$. Then*

$$\mathbb{E}[X\widetilde{Y}^T] = 0 \in \mathcal{M}_{n,k}. \tag{18}$$

Proof. We have that, for any $A \in \mathcal{M}_{n,k}$,

$$0 = \mathbb{E}[X^T(A\widetilde{Y})] = \mathbb{E}\big[\operatorname{tr}((A\widetilde{Y})X^T)\big] = \operatorname{tr}\big(A\mathbb{E}[\widetilde{Y}X^T]\big).$$

It is well-known that if, for a matrix $M \in \mathcal{M}_{k,n}$, one has that $\operatorname{tr}(AM) = 0$ for any $A \in \mathcal{M}_{n,k}$, then $M = 0$. Therefore,

$$\mathbb{E}[\widetilde{Y}X^T] = 0,$$

and we can transpose both parts that implies (18). □

Let $r_1(Y)$ be the orthogonal projection of the centred random vector $\mathbb{E}[Z \mid Y] - \mathbb{E}[Z] \in L_Y^2$ onto H_Y^1, i.e.

$$\mathbb{E}\big[W^T\big(\mathbb{E}[Z \mid Y] - \mathbb{E}[Z] - r_1(Y)\big)\big] = 0, \tag{19}$$

for all $W \in H_Y^1$.

Proposition 3. *Suppose that Σ_{YY} is invertible. Then*

$$r_1(Y) = C(Y - \mathbb{E}[Y]), \tag{20}$$

with C given by (11).

Proof. Since $r_1(Y) \in H_Y^1$, we have that (20) holds for some matrix $C \in \mathcal{M}_{n,k}$. Then, each $W \in H_Y^1$ satisfies $W = A\widetilde{Y}$ for some matrix $A \in \mathcal{M}_{n,k}$. Therefore, (19) reads

$$\mathbb{E}\big[(A\widetilde{Y})^T\big(\mathbb{E}[Z \mid Y] - \mathbb{E}[Z] - C\widetilde{Y}\big)\big] = 0 \tag{21}$$

for all $A \in \mathcal{M}_{n,k}$. By Lemma 2,

$$\mathbb{E}\big[\big(\mathbb{E}[Z \mid Y] - \mathbb{E}[Z] - C\widetilde{Y}\big)\widetilde{Y}^T\big] = 0, \tag{22}$$

$$\mathbb{E}\big[\big(\mathbb{E}[Z \mid Y] - \mathbb{E}[Z]\big)\widetilde{Y}^T\big] = C\mathbb{E}[\widetilde{Y}\widetilde{Y}^T]. \tag{23}$$

Next,

$$\mathbb{E}[\widetilde{Y}\widetilde{Y}^T] = \mathbb{E}\big[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^T\big] = \operatorname{Cov}(Y, Y) = \Sigma_{YY}, \tag{24}$$

and

$$\begin{aligned}
\mathbb{E}\big[\big(\mathbb{E}[Z \mid Y] - \mathbb{E}[Z]\big)\widetilde{Y}^T\big] &= \mathbb{E}\big[\big(\mathbb{E}[Z \mid Y] - \mathbb{E}[Z]\big)(Y - \mathbb{E}[Y])^T\big] \\
&= \mathbb{E}\big[\mathbb{E}[Z \mid Y]Y^T\big] - \mathbb{E}\big[\mathbb{E}[Z \mid Y]\big]\mathbb{E}\big[Y^T\big] \\
&= \mathbb{E}\big[ZY^T\big] - \mathbb{E}\big[Z\big]\mathbb{E}\big[Y^T\big] \\
&= \operatorname{Cov}(Z, Y) = \Sigma_{ZY}.
\end{aligned} \tag{25}$$

Therefore, (23) reads

$$\Sigma_{ZY} = C\Sigma_{YY}. \tag{26}$$

Since Σ_{YY} is assumed to be invertible, (11) holds. □

Since $\mathbb{E}[A(Y-\mathbb{E}[Y])]=0$, we have that $H^0_Y \perp H^1_Y$ in L^2_Y. We have, hence, the orthogonal decomposition

$$L^2_Y = H^0_Y \oplus H^1_Y \oplus H^{\geqslant 2}_Y, \tag{27}$$

where $H^{\geqslant 2}_Y := (H^0_Y \oplus H^1_Y)^\perp$ is the orthogonal complement to $H^0_Y \oplus H^1_Y$.

Corollary 4. *For any $Z \in L^2$, the following expansion of $\mathbb{E}[Z \mid Y] \in L^2_Y$ holds*

$$\mathbb{E}[Z \mid Y] = \mathbb{E}[Z] + C(Y - \mathbb{E}[Y]) + R_2(Y), \tag{28}$$

where C is given by (11) *and* $R_2(Y) \perp (H^0_Y \oplus H^1_Y)$.

We can continue the orthogonal expansion. Note that that H^1_Y can be described as the subspace of elements from

$$P^1_Y := \operatorname{span}(H^0_Y, \widetilde{Y}) = \{c + A\widetilde{Y} : c \in \mathbb{R}^n, A \in \mathcal{M}_{n,k}\} \tag{29}$$

orthogonal to H^0_Y. Similarly, since $\widetilde{Y}\widetilde{Y}^T \in \mathcal{M}_{k,k}$, we consider the space

$$\begin{aligned} P^2_Y :&= \operatorname{span}(H^0_Y, \widetilde{Y}, \widetilde{Y}\widetilde{Y}^T) \\ &= \{c + A\widetilde{Y} + B(\widetilde{Y}\widetilde{Y}^T) : c \in \mathbb{R}^n, A \in \mathcal{M}_{n,k}, B \in \mathcal{L}(\mathcal{M}_{k,k}, \mathbb{R}^n)\}. \end{aligned} \tag{30}$$

Next, we define H^2_Y as the set of vectors from P^2_Y orthogonal to $H^0_Y \oplus H^1_Y$.

Therefore, for an $X \in H^2_Y$, we have that $X = c + A\widetilde{Y} + B(\widetilde{Y}\widetilde{Y}^T)$ for some $c \in \mathbb{R}^n$, $A \in \mathcal{M}_{n,k}$, $B \in \mathcal{L}(\mathcal{M}_{k,k}, \mathbb{R}^n)$ and also $X \perp H^0_Y$ and $X \perp H^1_Y$. The first orthogonality means that

$$\begin{aligned} 0 &= \mathbb{E}[X] = c + B\Sigma_{YY}, \\ c &= -B\Sigma_{YY}. \end{aligned}$$

Next, $X \perp H^1_Y$ means that $X \perp A\widetilde{Y}$ for each $A \in \mathcal{M}_{n,k}$. By Lemma 2,

$$\begin{aligned} 0 = \mathbb{E}[X\widetilde{Y}^T] &= \mathbb{E}[(c + A\widetilde{Y} + B(\widetilde{Y}\widetilde{Y}^T))\widetilde{Y}^T] \\ &= A\Sigma_{YY} + \mathbb{E}[B(\widetilde{Y}\widetilde{Y}^T)\widetilde{Y}^T] \end{aligned} \tag{31}$$

Consider the action of B: it's a tensor $(B_{lij})_{1\le l\le n, 1\le i,j\le k}$ such that, for any $M = (M_{ij})_{1\le i,j\le k}$, we have

$$\big(B(M)\big)_l = \sum_{i,j=1}^k B_{lij} M_{ij}, \qquad 1 \le l \le n. \tag{32}$$

Therefore, since $(\widetilde{Y}\widetilde{Y}^T)_{ij} = \widetilde{Y}_i\widetilde{Y}_j$, we have that:

$$\big(B(\widetilde{Y}\widetilde{Y}^T)\widetilde{Y}^T\big)_{lm} = \sum_{i,j=1}^k B_{lij}\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_m, \qquad 1 \le l \le n, 1 \le m \le k;$$

hence, from (31), we obtain:

$$A\Sigma_{YY} = -B\kappa_{YYY}, \tag{33}$$

where

$$(B\kappa_{YYY})_{l,m} = \sum_{i,j=1}^{k} B_{lij}\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_m}], \qquad 1 \le l \le n, 1 \le m \le k. \tag{34}$$

Therefore,

$$A = -(B\kappa_{YYY})\Sigma_{YY}^{-1}, \tag{35}$$

and hence, $X \in H_Y^2$ iff

$$X = -B\Sigma_{YY} - (B\kappa_{YYY})\Sigma_{YY}^{-1}\widetilde{Y} + B\widetilde{Y}\widetilde{Y}^T. \tag{36}$$

Recall that $T \in \mathcal{L}(\mathbb{R}^k, \mathcal{M}_{k,k})$ is given by (3). Then, since $B\kappa_{YYY} \in \mathcal{M}_{n,k}$, we have for $v \in \mathbb{R}^k$,

$$\begin{aligned}\big((B\kappa_{YYY})v\big)_l &= \sum_{m=1}^{k}(B\kappa_{YYY})_{l,m}v_m = \sum_{m=1}^{k}\sum_{i,j=1}^{k} B_{lij}\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_m}]v_m \\ &= \sum_{i,j=1}^{k} B_{lij}(Tv)_{ij} = \big(B(Tv)\big)_l,\end{aligned}$$

for all $1 \le l \le n$, i.e.

$$(B\kappa_{YYY})v = B(Tv), \qquad v \in \mathbb{R}^k. \tag{37}$$

As a result, $X \in H_Y^2$ iff

$$X = B\big(S(Y)\big), \tag{38}$$

where $S(Y)$ is given by (4).

By Corollary 4,

$$R_2(Y) = \mathbb{E}[Z \mid Y] - \mathbb{E}[Z] - C(Y - \mathbb{E}[Y]) \perp H_Y^0 \oplus H_Y^1. \tag{39}$$

Let $r_2(Y)$ be the next term of the orthogonal expansion of $\mathbb{E}[Z \mid Y]$, i.e. let $r_2(Y)$ be the projection of $R_2(Y)$ on H_Y^2.

Let $D = D(B) \in \mathcal{M}_{n,k^2}$ be the unique matrix such that

$$B(M) = D\ \mathrm{vec}(M), \qquad M \in \mathcal{M}_{k,k}.$$

Namely, all coordinates of and we can set, for $v \in \mathbb{R}^{k^2}$ and $1 \le l \le n$,

$$(Dv)_l := \sum_{i,j=1}^{k} B_{lij}v_{i+k(j-1)}.$$

As a result,

$$B(S(Y)) = Ds(Y). \tag{40}$$

Now, we can find the conditions on D to have that

$$r_2(Y) = D\,s(Y) \tag{41}$$

would be the orthogonal projection of $R_2(Y)$, given by (39), on H_Y^2; it can be done similarly to how it was obtained for $r_1(Y)$.

Namely, since $r_2(Y) \in H_Y^2$, (41) holds with some $D \in \mathcal{M}_{n,k^2}$. Next, $D\,s(Y)$ is the orthogonal projection of $R_2(Y)$ on H_Y^2 iff

$$R_2(Y) - D\,s(Y) \perp W, \qquad W \in H_Y^2.$$

Each $W \in H_Y^2$ can be represented as $W = As(Y)$, for some $A \in \mathcal{M}_{n,k^2}$. Therefore,

$$R_2(Y) - D\,s(Y) \perp As(Y)$$

for each $A \in \mathcal{M}_{n,k^2}$. Applying Lemma 2 with k replaced by k^2 and $\widetilde{Y}$ replaced by $s(Y)$, we have that

$$\mathbb{E}\big[\big(R_2(Y) - D\,s(Y)\big)s(Y)^T\big] = 0 \in \mathcal{M}_{n,k^2}. \tag{42}$$

Next, by (4),

$$\mathbb{E}[S(Y)] = \mathbb{E}[\widetilde{Y}\widetilde{Y}^T] - \Sigma_{YY} - T(\Sigma_{YY}^{-1}\mathbb{E}[\widetilde{Y}]) = 0 \in \mathcal{M}_{k,k},$$

therefore, by (7),

$$\mathbb{E}[s(Y)] = 0 \in \mathbb{R}^{k^2}. \tag{43}$$

Furthermore, $s(Y)\widetilde{Y}^T \in \mathcal{M}_{k^2,k}$, and

$$\mathbb{E}\big[s(Y)\widetilde{Y}^T\big]_{i+k(j-1),m} = \mathbb{E}\big[s(Y)_{i+k(j-1)}\widetilde{Y}_m\big] = \mathbb{E}\big[S(Y)_{ij}\widetilde{Y}_m\big], \tag{44}$$

where we used (6). By (5),

$$S(Y)_{ij} = \widetilde{Y}_i\widetilde{Y}_j - (\Sigma_{YY})_{ij} - \sum_{r=1}^{k}\sum_{s=1}^{k}\mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r](\Sigma_{YY}^{-1})_{rs}\widetilde{Y}_s. \tag{45}$$

Hence,

$$\mathbb{E}\big[S(Y)_{ij}\widetilde{Y}_m\big] = \mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_m] - \sum_{r=1}^{k}\sum_{s=1}^{k}\mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r](\Sigma_{YY}^{-1})_{rs}\mathbb{E}[\widetilde{Y}_s\widetilde{Y}_m] = 0,$$

since

$$\begin{aligned}
\sum_{r=1}^{k}\sum_{s=1}^{k}\mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r](\Sigma_{YY}^{-1})_{rs}\mathbb{E}[\widetilde{Y}_s\widetilde{Y}_m] &= \sum_{r=1}^{k}\sum_{s=1}^{k}\mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r](\Sigma_{YY}^{-1})_{rs}(\Sigma_{YY})_{sm} \\
&= \sum_{r=1}^{k}\mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r](\Sigma_{YY}^{-1}\Sigma_{YY})_{rm} \\
&= \sum_{r=1}^{k}\mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r]\delta_{rm} = \mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_m].
\end{aligned}$$

Therefore, by (44),

$$\mathbb{E}[s(Y)\widetilde{Y}^T] = 0 \in \mathcal{M}_{k^2,k},$$

and taking transposition, we obtain that

$$\mathbb{E}[\widetilde{Y}s(Y)^T] = 0 \in \mathcal{M}_{k,k^2},$$

and therefore,

$$\mathbb{E}[(C\widetilde{Y})s(Y)^T] = 0 \in \mathcal{M}_{n,k^2}, \tag{46}$$

where $C \in \mathcal{M}_{n,k}$ is given by (11).

Then, from (42), (44), (46), and (39), we have a simplified condition on D:

$$\mathbb{E}\big[\big(\mathbb{E}[Z \mid Y] - Ds(Y)\big)s(Y)^T\big] = 0 \in \mathcal{M}_{n,k^2},$$

that gives the normal equation:

$$\mathbb{E}\big[\mathbb{E}[Z \mid Y]s(Y)^T\big] = D\mathbb{E}[s(Y)s(Y)^T];$$

where we used that $(Ds(Y))s(Y)^T = D(s(Y)s(Y)^T) \in \mathcal{M}_{n,k^2}$. Since $s(Y)$ is $\sigma(Y)$-measurable,

$$\mathbb{E}\big[\mathbb{E}[Z \mid Y]s(Y)^T\big] = \mathbb{E}\big[\mathbb{E}[Zs(Y)^T \mid Y]\big] = \mathbb{E}[Zs(Y)^T],$$

and hence we can rewrite the normal equation:

$$\mathbb{E}\big[Zs(Y)^T\big] = D\mathbb{E}[s(Y)s(Y)^T].$$

By (44), we can replace here Z by $\widetilde{Z} := Z - \mathbb{E}[Z]$. and therefore, we will get by (8) that

$$\Sigma_{ZS} = D\Sigma_{SS}.$$

Since we assume that Z_{SS} is invertible, we will get then (12).

As a result,

$$r_2(Y) = \Sigma_{ZS}(\Sigma_{SS})^{-1} \operatorname{vec} S(Y) \tag{47}$$

We can compute also the entries of $\Sigma_{ZS} \in \mathcal{M}_{n,k^2}$. Namely, for $1 \le l \le n$ and $1 \le i,j \le k$, we have by (8) and (45) that

$$\begin{aligned}(\Sigma_{ZS})_{l,i+j(k-1)} &= \mathbb{E}[\widetilde{Z}_l(S(Y))_{ij}] \\ &= \mathbb{E}[\widetilde{Z}_l\widetilde{Y}_i\widetilde{Y}_j] - \sum_{r=1}^{k}\sum_{s=1}^{k}\mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r](\Sigma_{YY}^{-1})_{rs}\mathbb{E}[\widetilde{Z}_l\widetilde{Y}_s].\end{aligned}$$

Since Σ_{YY} is a symmetric matrix, $(\Sigma_{YY}^{-1})_{rs} = (\Sigma_{YY}^{-1})_{sr}$, and then

$$\sum_{s=1}^{k}(\Sigma_{YY}^{-1})_{sr}\mathbb{E}[\widetilde{Z}_l\widetilde{Y}_s] = (\Sigma_{ZY}\Sigma_{YY}^{-1})_{lr} = C_{lr},$$

where C is given by (11). Therefore, (13) holds.

Next, we compute the entries of $\Sigma_{SS} \in \mathcal{M}_{k^2,k^2}$. We have, for $1 \le i,j,p,q \le k$,

$$(\Sigma_{SS})_{i+k(j-1),p+k(q-1)} = \mathbb{E}[s(Y)_{i+k(j-1)}s(Y)_{p+k(q-1)}] = \mathbb{E}[S(Y)_{ij}S(Y)_{pq}].$$

Denote

$$\begin{aligned}\widetilde{\Sigma}_{ij}(Y) :&= \widetilde{Y}_i\widetilde{Y}_j - (\Sigma_{YY})_{ij}, \\ H_{ij}(Y) :&= \sum_{r=1}^{k}\sum_{s=1}^{k}\mathbb{E}[\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_r](\Sigma_{YY}^{-1})_{rs}\widetilde{Y}_s,\end{aligned}$$

so that, by (45),

$$S(Y)_{ij} = \widetilde{\Sigma}_{ij}(Y) - H_{ij}(Y).$$

Then

$$\begin{aligned}
\mathbb{E}[S(Y)_{ij}S(Y)_{pq}] &= \mathbb{E}[(\widetilde{\Sigma}_{ij}(Y) - H_{ij}(Y))(\widetilde{\Sigma}_{pq}(Y) - H_{pq}(Y))] \\
&= \mathbb{E}[\widetilde{\Sigma}_{ij}(Y)\widetilde{\Sigma}_{pq}(Y)] - \mathbb{E}[\widetilde{\Sigma}_{ij}(Y)H_{pq}(Y)] \\
&\quad - \mathbb{E}[\widetilde{\Sigma}_{pq}(Y)H_{ij}(Y)] + \mathbb{E}[H_{ij}(Y)H_{pq}(Y)]
\end{aligned}$$

and since $\mathbb{E}[H_{ij}(Y)] = \mathbb{E}[H_{pq}(Y)] = 0$, we can continue

$$\begin{aligned}
&= \mathbb{E}[\widetilde{\Sigma}_{ij}(Y)\widetilde{\Sigma}_{pq}(Y)] - \mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}H_{pq}(Y)] \\
&\quad - \mathbb{E}[\widetilde{Y_p}\widetilde{Y_q}H_{ij}(Y)] + \mathbb{E}[H_{ij}(Y)H_{pq}(Y)].
\end{aligned}$$

Next,

$$\mathbb{E}[\widetilde{\Sigma}_{ij}(Y)\widetilde{\Sigma}_{pq}(Y)] = \mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_p}\widetilde{Y_q}] - (\Sigma_{YY})_{ij}(\Sigma_{YY})_{pq},$$

and

$$\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}H_{pq}(Y)] = \sum_{r=1}^{k}\sum_{s=1}^{k}\mathbb{E}[\widetilde{Y_p}\widetilde{Y_q}\widetilde{Y_r}](\Sigma_{YY}^{-1})_{rs}\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_s}].$$

Similarly,

$$\mathbb{E}[\widetilde{Y_p}\widetilde{Y_q}H_{ij}(Y)] = \sum_{r=1}^{k}\sum_{s=1}^{k}\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_r}](\Sigma_{YY}^{-1})_{rs}\mathbb{E}[\widetilde{Y_p}\widetilde{Y_q}\widetilde{Y_s}], \tag{48}$$

and

$$\mathbb{E}[H_{ij}(Y)H_{pq}(Y)] = \sum_{r=1}^{k}\sum_{s=1}^{k}\sum_{r'=1}^{k}\sum_{s'=1}^{k}\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_r}](\Sigma_{YY}^{-1})_{rs}\mathbb{E}[\widetilde{Y_p}\widetilde{Y_q}\widetilde{Y_{r'}}](\Sigma_{YY}^{-1})_{r's'}\mathbb{E}[\widetilde{Y_s}\widetilde{Y_{s'}}].$$

Since $\mathbb{E}[\widetilde{Y_s}\widetilde{Y_{s'}}] = (\Sigma_{YY})_{ss'}$, and

$$\sum_{s=1}^{k}\sum_{s'=1}^{k}(\Sigma_{YY}^{-1})_{rs}(\Sigma_{YY}^{-1})_{r's'}(\Sigma_{YY})_{ss'} = \sum_{s=1}^{k}(\Sigma_{YY}^{-1})_{rs}\delta_{sr'} = (\Sigma_{YY}^{-1})_{rr'},$$

we have

$$\mathbb{E}[H_{ij}(Y)H_{pq}(Y)] = \sum_{r=1}^{k}\sum_{r'=1}^{k}\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_r}]\mathbb{E}[\widetilde{Y_p}\widetilde{Y_q}\widetilde{Y_{r'}}](\Sigma_{YY}^{-1})_{rr'} = \mathbb{E}[\widetilde{Y_p}\widetilde{Y_q}H_{ij}(Y)],$$

by (48).

As a result, we obtain (14).

This finishes the proof of Theorem 1.

6.3 Beyong mean field expansion: linear term

For $Z = (Z_1, \ldots, Z_n)^T \in \mathbb{R}^n$, we have

$$\mathbb{E}[Z_i] = q(Z)_i + \varepsilon^d p(Z)_i + o(\varepsilon^d), \tag{49}$$

$$\mathrm{Var}(Z_i) = \frac{1}{\delta}q(Z)_i + \varepsilon^d \frac{1}{\delta}p(X)_i + o(\varepsilon^d), \tag{50}$$

$$\mathrm{Cov}(Z_i, Z_j) = \varepsilon^d g(Z, Z)_{ij} + o(\varepsilon^d). \tag{51}$$

We consider matrices $Q, P, G \in \mathcal{M}_{n,n}$, given by

$$Q = \operatorname{diag}(q_1(Z), \dots, q_n(Z)), \tag{52}$$
$$P = \operatorname{diag}(p_1(Z), \dots, p_n(Z)), \tag{53}$$
$$G_{ij} = g(Z)_{ij}, \quad i \neq j, \qquad G_{ii} = 0. \tag{54}$$

Then,

$$\mathbb{E}[Z] = Q + \varepsilon^d P + o(\varepsilon^d), \tag{55}$$
$$\operatorname{Cov}(Z, Z) = \delta^{-1} Q + \varepsilon^d (\delta^{-1} P + G) + o(\varepsilon^d). \tag{56}$$

Suppose that

$$Y = UZ, \qquad U \in \mathcal{M}_{k,n}, \qquad \operatorname{rank}(U) = k. \tag{57}$$

Then

$$\mathbb{E}[Y] = UQ + \varepsilon^d UP + o(\varepsilon^d), \tag{58}$$

and since

$$\begin{aligned}\operatorname{Cov}(Y, Y) &= \mathbb{E}[\widetilde{Y}\widetilde{Y}^T] \\ &= \mathbb{E}[U\widetilde{Z}(U\widetilde{Z})^T] = U\mathbb{E}[\widetilde{Z}\widetilde{Z}^T]U^T = U \operatorname{Cov}(Z, Z) U^T,\end{aligned}$$

and

$$\begin{aligned}\operatorname{Cov}(Z, Y) &= \mathbb{E}[\widetilde{Z}\widetilde{Y}^T] \\ &= \mathbb{E}[\widetilde{Z}(U\widetilde{Z})^T] = \mathbb{E}[\widetilde{Z}\widetilde{Z}^T]U^T = \operatorname{Cov}(Z, Z)U^T,\end{aligned}$$

we have

$$\Sigma_{YY} = \delta^{-1} UQU^T + \varepsilon^d (\delta^{-1} UPU^T + UGU^T) + o(\varepsilon^d), \tag{59}$$
$$\Sigma_{ZY} = \delta^{-1} QU^T + \varepsilon^d (\delta^{-1} PU^T + GU^T) + o(\varepsilon^d). \tag{60}$$

In particular, Σ_{YY} is invertible. To calculate its inverse, we note that, for any $B_0, B_1 \in \mathcal{M}_{k,k}$, with invertible B_0, one has

$$(B_0 + \varepsilon^d B_1)^{-1} = B_0^{-1} - \varepsilon^d B_0^{-1} B_1 B_0^{-1} + o(\varepsilon).$$

If Q does not contain zeroes on the main diagonal, then UQU^T is invertible, and therefore,

$$\Sigma_{YY}^{-1} = \delta(UQU^T)^{-1} - \varepsilon^d (UQU^T)^{-1}(\delta UPU^T + \delta^2 UGU^T)(UQU^T)^{-1} + o(\varepsilon^d). \tag{61}$$

As a result, the matrix C from (11) has the form

$$\begin{aligned}C &= \Sigma_{ZY}\Sigma_{YY}^{-1} \\ &= QU^T(UQU^T)^{-1} \\ &\quad + \varepsilon^d \Big((P + \delta G)U^T(UQU^T)^{-1} \\ &\qquad - QU^T(UQU^T)^{-1}U(P + \delta G)U^T(UQU^T)^{-1} \Big) + o(\varepsilon^d).\end{aligned} \tag{62}$$

6.4 Beyong mean field expansion: quadratic term

Recall that matrix D in (10) has form (12) similar to the form of C, with Y replaced by $s(Y)$. Matrix Σ_{SS} has entries given by (14). Normally, we expect that

$$\Sigma_{SS} = O(1) = \alpha + \varepsilon^d \beta + o(\varepsilon^d)$$

Next, note that

$$\mathbb{E}[\|s(Y)\|^2] = \mathbb{E}[s(Y)^T s(Y)] = \operatorname{tr} \Sigma_{SS}$$

(here the norm is taken in $\mathbb{R}^{k^2}$). Thus $\Sigma_{SS} = O(1)$ implies $s(Y) = O(1)$.

Therefore, the order in ε of

$$Ds(Y) = \Sigma_{ZS}(\Sigma_{SS})^{-1} s(Y)$$

is the same as the order of Σ_{ZS}.

(Note also that $\Sigma_{SS} = O(\varepsilon^\alpha)$, $\alpha > 0$ would imply then that $s(Y) = O(\varepsilon^{\frac{\alpha}{2}})$, so $(\Sigma_{SS})^{-1} s(Y) = O(\varepsilon^{-\frac{\alpha}{2}})$, and Σ_{ZS} would need to have an additional smallness in ε.)

The problems is, however, that, by (13), Σ_{ZS} may have order $O(1)$. In principle, it depends on U. If U is such that $C = O(1)$, i.e. if $QU^T(UQU^T)^{-1}$ does not "disappear" then since $\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_r}]$ has always a diagonal term when $i = j = r$ it has also the order $O(1)$ with the leading term $\delta^{-2} Q_i$. Therefore, $\Sigma_{ZS} = O(1)$ unless some cancellation happen. Therefore, $r_2(Y) = O(1)$ as well.

6.4.1 Case study 1

Let $U = (\mathbf{O}\ \mathbf{I})$, where $\mathbf{O}$ is the zero matrix from $\mathcal{M}_{n,n-k}$, and $\mathbf{I}$ is the unit matrix from $\mathcal{M}_{k,k}$, we have that $Z = (XY)^T$, where X is the unobserved pat and Y is the observed part. Then

$$\mathbb{E}[Z \mid Y] = (\mathbb{E}[X \mid Y]\ Y)^T,$$

and to calculate $\mathbb{E}[X \mid Y]$ we can use the same formulas as above, with the crucial difference that $\operatorname{Cov}(X, Y)$ now does not contain the main diagonal elements as X and Y are "disjoint", i.e.

$$\Sigma_{XY} = \varepsilon^d G + o(\varepsilon^d).$$

Then the corresponding

$$C = \Sigma_{XY}\Sigma_{YY}^{-1} = \varepsilon^d G(\delta Q^{-1} - \varepsilon^d(\ldots)) = \varepsilon^d \delta G Q^{-1} + o(\varepsilon^d),$$

hence,

$$\begin{aligned} \mathbb{E}[X \mid Y] &= \mathbb{E}[X] + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mathbb{E}[Y]) + r_2(Y) + \ldots \\ &= Q + \varepsilon^d P + \varepsilon^d \delta G Q^{-1} + o(\varepsilon^d) + r_2(Y) + \ldots . \end{aligned} \tag{63}$$

Note that here $C = O(\varepsilon^d)$ because in (62)

$$QU^T(UQU^T)^{-1} = (0 I_k)^T.$$

Next, by (13),

$$(\Sigma_{XS})_{l,i+j(k-1)} = o(\varepsilon) \tag{64}$$

Indeed,

$$\mathbb{E}[\widetilde{X_l}\widetilde{Y_i}\widetilde{Y_j}] = \delta_{i,j}\delta^{-1}\operatorname{Cov}(X_l, Y_i) + o(\varepsilon^d) = \delta_{i,j}\varepsilon^d\delta^{-1}G_{li} + o(\varepsilon^d)$$

and

$$\mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_r}] = \delta_{i,j,r}\frac{Q_i + \varepsilon^d P_i}{\delta^2} + O(\varepsilon^d).$$

Since, by the above,

$$C_{lr} = \varepsilon^d \delta (GQ^{-1})_{lr} + o(\varepsilon^d) = \varepsilon^d \delta G_{lr} Q_r^{-1} + o(\varepsilon^d),$$

we obtain:

$$\sum_{r=1}^{k} \mathbb{E}[\widetilde{Y_i}\widetilde{Y_j}\widetilde{Y_r}] C_{lr} = \delta_{i,j}\varepsilon^d \delta^{-1} G_{li} + o(\varepsilon^d),$$

that exactly cancels $\mathbb{E}[\widetilde{X_l}\widetilde{Y_i}\widetilde{Y_j}]$; this proves (64). As a result, $r_2(Y) = o(\varepsilon)$.