A Proofs and derivations

A.1 Proof of Lemma 2.1

Proof. In every iteration, MP-MCMC generates a set of N new samples which, together with the current state, constitute the $N+1$ proposed states $\boldsymbol{y}_{1:N+1}$ from which M samples are drawn in sequence. Based at a current state and last accepted sample $\boldsymbol{x}$ from the current iteration, say $\boldsymbol{x} = \boldsymbol{y}_i \in \mathbb{R}^d$ for some $i = 1, ..., N+1$, the probability of the mth such sample ($m = 1, ..., M$) being in $B \in \mathcal{B}(\mathbb{R}^d)$ can be expressed by the transition kernel $P^{(m)}(\boldsymbol{y}_i, B)$, given by

$$\int \tilde{\kappa}(\boldsymbol{y}_i, \boldsymbol{y}_{\setminus i}) \sum_{i_1,\dots,i_m} A(i, i_1) \prod_{j=1}^{m-1} A(i_j, i_{j+1}) \mathbb{1}_B(\boldsymbol{y}_{i_m}) \mathrm{d}\boldsymbol{y}_{\setminus i}.$$

Any transition from the current iteration into the next has this form. Hence, it is sufficient to prove that updates of this type preserve π. For arbitrary $m = 1, ..., M$, we compute,

$$\begin{aligned}
&\int_{\mathbb{R}^d} P^{(m)}(\boldsymbol{x}, B)\pi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\
&= \frac{1}{N+1} \int_{\mathbb{R}^{(N+1)d}} \sum_{i=1}^{N+1} \tilde{\kappa}(\boldsymbol{y}_i, \boldsymbol{y}_{\setminus i}) \sum_{i_1,\dots,i_m=1}^{N+1} A(i, i_1) \prod_{j=2}^{M} A(i_{j-1}, i_j) \mathbb{1}_B(\boldsymbol{y}_{i_m})\pi(\boldsymbol{y}_i)\mathrm{d}\boldsymbol{y}_{1:N+1} \\
&= \frac{1}{N+1} \int_{\mathbb{R}^{(N+1)d}} \sum_{i_1=1}^{N+1} \sum_{i=1}^{N+1} \pi(\boldsymbol{y}_i)\tilde{\kappa}(\boldsymbol{y}_i, \boldsymbol{y}_{\setminus i}) A(i, i_1) \sum_{i_2,\dots,i_m=1}^{N+1} \prod_{j=2}^{M} A(i_{j-1}, i_j) \mathbb{1}_B(\boldsymbol{y}_{i_m})\mathrm{d}\boldsymbol{y}_{1:N+1} \\
&= \frac{1}{N+1} \int_{\mathbb{R}^{(N+1)d}} \sum_{i_1=1}^{N+1} \pi(\boldsymbol{y}_{i_1})\tilde{\kappa}(\boldsymbol{y}_{i_1}, \boldsymbol{y}_{\setminus i_1}) \sum_{i_2,\dots,i_m=1}^{N+1} \prod_{j=2}^{M} A(i_{j-1}, i_j) \mathbb{1}_B(\boldsymbol{y}_{i_m})\mathrm{d}\boldsymbol{y}_{1:N+1} \\
&= \cdots \\
&= \frac{1}{N+1} \sum_{i_m=1}^{N+1} \int_{\mathbb{R}^{(N+1)d}} \pi(\boldsymbol{y}_{i_m})\tilde{\kappa}(\boldsymbol{y}_{i_m}, \boldsymbol{y}_{\setminus i_m}) \mathbb{1}_B(\boldsymbol{y}_{i_m})\mathrm{d}\boldsymbol{y}_{1:N+1} \\
&= \int_B \pi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.
\end{aligned}$$

where we used condition (2) repeatedly, and that κ is a kernel. □

A.2 Proof of Lemma 3.1

Proof. Using the variance decomposition formula,

$$\begin{aligned}
\mathrm{Var}\left(\hat{\mu}^{(f)}_{L,M,N}\right) &= \mathrm{Var}\left(\mathbb{E}[\hat{\mu}^{(f)}_{L,M,N} | \boldsymbol{y}^{(\ell)}_{1:N+1}, \ell = 1, ..., L]\right) \\
&\quad + \mathbb{E}\left[\mathrm{Var}(\hat{\mu}^{(f)}_{L,M,N} | \boldsymbol{y}^{(\ell)}_{1:N+1}, \ell = 1, ..., L)\right] \\
&\geq \mathrm{Var}\left(\frac{1}{LM} \sum_{\ell=1}^{L} \sum_{m=1}^{M} \mathbb{E}[f(\boldsymbol{x}^{(\ell)}_m) | \boldsymbol{y}^{(\ell)}_{1:N+1}]\right) \\
&= \mathrm{Var}\left(\hat{\mu}^{(f)}_{L,N}\right),
\end{aligned}$$

where we used $\mathrm{Var}(\hat{\mu}^{(f)}_{L,M,N} | \boldsymbol{y}^{(\ell)}_{1:N+1}, \ell = 1, ..., L) \geq 0$, that $f(\boldsymbol{x}^{(\ell)}_m) | \boldsymbol{y}^{(k)}_{1:N+1}, k = 1, ..., L$, is independent of $\boldsymbol{y}^{(k)}_{1:N+1}$ for $k \neq \ell$, and equation (6). □

A.3 Proof of Lemma 3.2

Proof. Ergodicity follows as π is preserved (Lemma 2.1) and the underlying chain is positive Harris (Meyn and Tweedie, 1993, Chapter 13). Since the asymptotic behaviour of the chain is independent of its initial distribution we may assume the stationary distribution π as initial distribution. It follows,

$$\begin{aligned}
\mathbb{E}_p\left[\hat{\mu}_{L,N}^{(f)}\right] &= \sum_{i=1}^{N+1} \mathbb{E}_p\left[w_i f(\boldsymbol{y}_i)\right] \\
&= \sum_{i=1}^{N+1} \int_{\mathbb{R}^{(N+1)d}} w_i f(\boldsymbol{y}_i) \sum_{j=1}^{N+1} \Pr(I=j) p_I(\boldsymbol{y}_{1:N+1}) \mathrm{d}\boldsymbol{y}_{1:N+1} \\
&= \sum_{i=1}^{N+1} \int_{\mathbb{R}^{(N+1)d}} \frac{\pi(\boldsymbol{y}_i)\tilde{\kappa}(\boldsymbol{y}_i, \boldsymbol{y}_{\setminus i})}{\sum_{k=1}^{N+1} \pi(\boldsymbol{y}_k)\tilde{\kappa}(\boldsymbol{y}_k, \boldsymbol{y}_{\setminus k})} f(\boldsymbol{y}_i) \sum_{j=1}^{N+1} \frac{1}{N+1} \pi(\boldsymbol{y}_j)\tilde{\kappa}(\boldsymbol{y}_j, \boldsymbol{y}_{\setminus j}) \mathrm{d}\boldsymbol{y}_{1:N+1} \\
&= \frac{1}{N+1} \sum_{i=1}^{N+1} \int_{\mathbb{R}^d} f(\boldsymbol{y}_i)\pi(\boldsymbol{y}_i) \left(\int_{\mathbb{R}^{Nd}} \tilde{\kappa}(\boldsymbol{y}_i, \boldsymbol{y}_{\setminus i}) \mathrm{d}\boldsymbol{y}_{\setminus i}\right) \mathrm{d}\boldsymbol{y}_i \\
&= \frac{1}{N+1} \sum_{i=1}^{N+1} \int_{\mathbb{R}^d} f(\boldsymbol{y}_i)\pi(\boldsymbol{y}_i) \mathrm{d}\boldsymbol{y}_i \\
&= \mathbb{E}_\pi\left[f(\boldsymbol{y})\right].
\end{aligned}$$

The statement follows by the ergodic theorem. □

A.4 Proof of Proposition 3.3

Proof. Let the ith component of an arbitrary vector $\boldsymbol{x} \in \mathbb{R}^d$ be denoted by $[\boldsymbol{x}]_i$. Due to ergodicity, and since the asymptotic behaviour of the Markov chain is independent of its initial distribution we may assume the stationary distribution π as initial distribution. For $j, k \in \{1, ..., d\}$ we have

$$\begin{aligned}
\mathbb{E}_p\left[\left(\hat{\Sigma}_{L,N}\right)_{j,k}\right] &= \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}\left[\sum_{i=1}^{N+1} w_i^{(\ell)} \left[\left([\boldsymbol{y}_i^{(\ell)}]_j - [\boldsymbol{\mu}]_j\right) - \left([\hat{\boldsymbol{\mu}}_{L,N}]_j - [\boldsymbol{\mu}]_j\right)\right]\right. \\
&\qquad\qquad \left.\cdot \left[\left([\boldsymbol{y}_i^{(\ell)}]_k - [\boldsymbol{\mu}]_k\right) - \left(\hat{\boldsymbol{\mu}}_{L,N}]_k - [\boldsymbol{\mu}]_k\right)\right]\right] \\
&= \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}\left[\sum_{i=1}^{N+1} w_i^{(\ell)} \left\{\left([\boldsymbol{y}_i^{(\ell)}]_j - [\boldsymbol{\mu}]_j\right)\left([\boldsymbol{y}_i^{(\ell)}]_k - [\boldsymbol{\mu}]_k\right)\right.\right. \\
&\qquad - \left([\boldsymbol{y}_i^{(\ell)}]_j - [\boldsymbol{\mu}]_j\right)\left([\hat{\boldsymbol{\mu}}_{L,N}]_k - [\boldsymbol{\mu}]_k\right) - \left([\hat{\boldsymbol{\mu}}_{L,N}]_j - [\boldsymbol{\mu}]_j\right)\left([\boldsymbol{y}_i^{(\ell)}]_k - [\boldsymbol{\mu}]_k\right) \\
&\qquad \left.\left. + \left([\hat{\boldsymbol{\mu}}_{L,N}]_j - [\boldsymbol{\mu}]_j\right)\left([\hat{\boldsymbol{\mu}}_{L,N}]_k - [\boldsymbol{\mu}]_k\right)\right\}\right] \\
&= \frac{1}{L}\sum_{\ell=1}^{L} \left(\mathbb{E}\left[\sum_{i=1}^{N+1} w_i^{(\ell)} \left([\boldsymbol{y}_i^{(\ell)}]_j - [\boldsymbol{\mu}]_j\right)\left([\boldsymbol{y}_i^{(\ell)}]_k - [\boldsymbol{\mu}]_k\right)\right]\right. \\
&\qquad \left. - \mathbb{E}\left[\left([\hat{\boldsymbol{\mu}}_{L,N}]_j - [\boldsymbol{\mu}]_j\right)\left([\hat{\boldsymbol{\mu}}_{L,N}]_k - [\boldsymbol{\mu}]_k\right)\right]\right) \\
&= \mathrm{Cov}_\pi\left([\boldsymbol{x}]_j, [\boldsymbol{x}]_k\right) - \mathrm{Cov}_p\left([\hat{\boldsymbol{\mu}}_{L,N}]_j, [\hat{\boldsymbol{\mu}}_{L,N}]_k\right).
\end{aligned}$$

In the last line we applied Lemma 3.2. For $L \to \infty$, $\hat{\boldsymbol{\mu}}_{L,N}$ converges to the constant mean vector $\boldsymbol{\mu}$. Hence, for any $j, k \in \{1, ..., d\}$,

$$\mathrm{Cov}_p\left([\hat{\boldsymbol{\mu}}_{L,N}]_j, [\hat{\boldsymbol{\mu}}_{L,N}]_k\right) \to 0 \quad \text{for } L \to \infty.$$

Applying the ergodic theorem concludes the proof. □

A.5 Proof of 4.1

Proof. The key is to show ergodicity of the underlying adaptive chain. If it is ergodic, we may assume the stationary distribution π as initial distribution, and proceed analogously to the proof of Lemma 3.2. The ergodicity proof uses (Roberts and Rosenthal, 2007, Theorem 2), which relies on coupling the underlying adaptive chain with another chain that is adaptive only up to a certain iteration. The theorem requires two conditions: the first is diminishing adaptation, that is,

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d} \left\| P^{(1)}_{\Upsilon_{\ell+1}}(\boldsymbol{x},B) - P^{(1)}_{\Upsilon_\ell}(\boldsymbol{x},B) \right\| \to 0 \quad \text{for } \ell \to \infty. \tag{18}$$

The second is called containment condition, which is satisfied due to (Bai et al., 2011) if the family $\{P^{(1)}_\Upsilon : \Upsilon \in \mathcal{Y}\}$ is simultaneous strongly aperiodic geometric ergodic (SSAGE), that is, there is $C \in \mathcal{B}(\mathbb{R}^d)$, $V : \mathbb{R}^d \to [1,\infty)$ and $\delta > 0$, $\rho < 1$, $b < \infty$ so that $\sup_{\boldsymbol{x}\in C} V(\boldsymbol{x}) < \infty$, and

1. $\forall\ \Upsilon \in \mathcal{Y}\ \exists$ a probability measure ν_Υ on C such that $P^{(1)}_\Upsilon(\boldsymbol{x},\cdot) \geq \delta\nu_\Upsilon(\cdot)$ for all $\boldsymbol{x} \in C$, and
2. $P^{(1)}_\Upsilon V(\boldsymbol{x}) \leq \rho V(\boldsymbol{x}) + b\mathbb{1}_C(\boldsymbol{x})$ for all $\Upsilon \in \mathcal{Y}, \boldsymbol{x} \in \mathbb{R}^d$,

where $P^{(1)}_\Upsilon V(\boldsymbol{x}) := \mathbb{E}_{P^{(1)}_\Upsilon}[V(\boldsymbol{x}_1)|\boldsymbol{x}_0 = \boldsymbol{x}]$.

If the chain is SSAGE and if (18) holds, then it is ergodic.

A.5.1 Continuity assumption

We proceed by proving the two above conditions.

(I) Diminishing adaptation: There are two types of adaptation parameters $\Upsilon \in \{\Sigma, (\boldsymbol{\mu},\Sigma)\}$, which result in two cases to consider. The case $\Upsilon = (\boldsymbol{\mu},\Sigma)$, i.e. where κ_Υ states an independence kernel is covered in A.5.2. Hence, let $\Upsilon = \Sigma$.

Let $B \in \mathcal{B}(\mathbb{R}^d)$ and $\boldsymbol{x} \in \mathbb{R}^d$ be arbitrary. Without loss generality, $\boldsymbol{x} = \boldsymbol{y}_{i_0}$ for some $i_0 \in \{1, ..., N+1\}$. Then,

$$\begin{aligned}
&\left\| P^{(1)}_{\Upsilon_{\ell+1}}(\boldsymbol{x},B) - P^{(1)}_{\Upsilon_\ell}(\boldsymbol{x},B) \right\| \\
&= \int_{\mathbb{R}^{Nd}} \left[\tilde{\kappa}_{\Sigma_{\ell+1}}(\boldsymbol{y}_{i_0},\boldsymbol{y}_{\setminus i_0}) - \tilde{\kappa}_{\Sigma_\ell}(\boldsymbol{y}_{i_0},\boldsymbol{y}_{\setminus i_0})\right] \sum_{i_1=1}^{N+1} A(i_0,i_1)\mathbb{1}_B(\boldsymbol{y}_{i_1})\mathrm{d}\boldsymbol{y}_{\setminus i_0} &(19)\\
&\leq (N+1)\int_{\mathbb{R}^{Nd}} \Big| \prod_{n=1}^N \kappa_{(\boldsymbol{0},\Sigma_{\ell+1})}(\boldsymbol{y}_n) - \prod_{n=1}^N \kappa_{(\boldsymbol{0},\Sigma_\ell)}(\boldsymbol{y}_n)\Big|\mathrm{d}\boldsymbol{y}_{1:N} \\
&\leq (N+1)\int_{\mathbb{R}^{Nd}} \int_0^1 \Big|\frac{\mathrm{d}}{\mathrm{d}s}\prod_{n=1}^N \kappa_{(\boldsymbol{0},\Sigma_\ell+s(\Sigma_{\ell+1}-\Sigma_\ell))}(\boldsymbol{y}_n)\Big|\mathrm{d}s\ \mathrm{d}\boldsymbol{y}_{1:N}, &(20)
\end{aligned}$$

where we used that $A(i_0,i_1) \leq 1$. Let us now differentiate between the two choices of proposal distributions.

(A) $\underline{\kappa_\Sigma = \mathcal{N}_\Sigma}$: Setting $A_\ell(s) = \Sigma_\ell + s(\Sigma_{\ell+1} - \Sigma_\ell)$ leads to

$$\prod_{n=1}^N \mathcal{N}_{(\boldsymbol{0},\Sigma_\ell+s(\Sigma_{\ell+1}-\Sigma_\ell))}(\boldsymbol{y}_n) = (2\pi)^{d\frac{N}{2}} \det\left(A_\ell(s)\right)^{-\frac{N}{2}} \exp\left(-\frac{1}{2}\sum_{n=1}^N \boldsymbol{y}_n^T A_\ell(s)^{-1}\boldsymbol{y}_n\right). \tag{21}$$

For the first of the two individual terms of derivatives of the product on the right hand side of (21), we have

$$\begin{aligned}
\Big|\frac{\mathrm{d}}{\mathrm{d}s}\left[\det\left(A_\ell(s)\right)^{-N/2}\right]\Big| &= \Big|\frac{N}{2}\det\left(A_\ell(s)\right)^{-\frac{N}{2}-1}\det(A_\ell(s))\,\mathrm{tr}\left(A_\ell(s)^{-1}(\Sigma_{\ell+1}-\Sigma_\ell)\right)\Big| \\
&\leq \mathrm{const}\,\|\Sigma_{\ell+1}-\Sigma_\ell\|\,, &(22)
\end{aligned}$$

where we used Jacobi's formula, and in the last line we used $0 < c_1 \leq \det(A_\ell(s)) \leq c_2 < \infty$ for any $\ell \in \mathbb{N}$, which is a consequence of $c_1 I \leq \Sigma \leq c_2 I$ for any $\Sigma \in \mathcal{Y}$, i.e. the boundedness of $\mathcal{Y}$. Moreover, we used

$$\mathrm{tr}\left(A_\ell(s)^{-1}(\Sigma_{\ell+1}-\Sigma_\ell)\right) = \langle A_\ell(s)^{-1}, \Sigma_{\ell+1}-\Sigma_\ell\rangle_F \leq \|A_\ell(s)^{-1}\|_F\|\Sigma_{\ell+1}-\Sigma_\ell\|_F \tag{23}$$

where $\langle\cdot,\cdot\rangle_F$ denotes the Frobenius inner product and $\|\cdot\|_F$ the associated Frobenius norm, for which we made use of the Cauchy-Schwarz inequality. Further, $\|A_\ell(s)^{-1}\|_F \leq$ const, since $A_\ell(s) \geq c_1 I$ and hence, $A_\ell^{-1}(s) \leq c_1^{-1} I$.

We do not need to further define the norm used in (22) since all norms are equivalent over finite-dimensional linear spaces. For the second term of derivatives on the right hand side of (21) we have

$$\begin{aligned}
&\left|\frac{\mathrm{d}}{\mathrm{d}s}\left[\exp\left(-\frac{1}{2}\sum_{n=1}^{N}\boldsymbol{y}_n^T A_\ell(s)^{-1}\boldsymbol{y}_n\right)\right]\right| && (24)\\
&= \left|\exp\left(-\frac{1}{2}\sum_{n=1}^{N}\boldsymbol{y}_\ell^T A_\ell(s)^{-1}\boldsymbol{y}_n\right)\sum_{n=1}^{N}\boldsymbol{y}_n^T A_\ell(s)^{-1}(\Sigma_{\ell+1}-\Sigma_\ell)A_\ell(s)^{-1}\boldsymbol{y}_n\right|\\
&\leq \text{const}\cdot\sum_{n=1}^{N}\boldsymbol{y}_n^T\boldsymbol{y}_n\exp\left(-\frac{1}{2c_1^2}\sum_{n=1}^{N}\boldsymbol{y}_n^T\boldsymbol{y}_n\right)\|\Sigma_{\ell+1}-\Sigma_\ell\|, && (25)
\end{aligned}$$

where we used $c_2^{-1}I \leq A_\ell^{-1}(s) \leq c_1^{-1}I$. Finally, using Fubini for interchanging integration and the boundedness of (second) moments of the Normal distribution, we conclude

$$\left\|P^{(1)}_{\Upsilon_{\ell+1}}(\boldsymbol{x},B) - P^{(1)}_{\Upsilon_\ell}(\boldsymbol{x},B)\right\| \leq \text{const}\,\|\Sigma_{\ell+1}-\Sigma_\ell\| \leq \text{const}\cdot\frac{1}{\ell} \longrightarrow 0 \quad \text{for } \ell\to\infty. \tag{26}$$

(B) $\underline{\kappa_\Sigma = t_\Sigma}$: We proceed as in the Gaussian case. Replacing (21) by the respective t-distribution density, using (15), and taking its derivative leads to similar estimates as in (22) and (25). Using the boundedness of moments of the t-distribution leads to the respective equation (26).

(II) SSAGE: Let $C=\mathbb{R}^d$. We need to find $\delta>0$ and ν_Υ such that $P_\Upsilon(\boldsymbol{x},B)\geq\delta\nu_\Upsilon(B)$ for all $B\in\mathcal{B}(\mathbb{R}^d)$ and for all $\boldsymbol{x}\in\mathbb{R}^d, \Upsilon\in\mathcal{Y}$. Since π is continuous on S and $\lambda(S)>0$, there is $\tilde{S}\subset S$ closed, $\tilde{S}\neq S$ and $\lambda(\tilde{S})>0$. Since $\tilde{S}$ is compact, $\mathcal{Y}$ is bounded, and π and the Gaussian and t-distribution PDFs are continuous and positive on $\tilde{S}$, there is a $c_A>0$ such that

$$c_A \leq \frac{\pi(\boldsymbol{y}_i)\tilde{\kappa}_\Upsilon(\boldsymbol{y}_i,\boldsymbol{y}_{\setminus i})}{\sum_{j=1}^{N+1}\pi(\boldsymbol{y}_j)\tilde{\kappa}_\Upsilon(\boldsymbol{y}_j,\boldsymbol{y}_{\setminus j})}, \tag{27}$$

for all $\boldsymbol{y}_1,...,\boldsymbol{y}_{N+1}\in\tilde{S}$ and all $\Upsilon\in\mathcal{Y}$. Similarly, there is a $c_N>0$ such that $c_N\leq\tilde{\kappa}_\Upsilon(\boldsymbol{y}_i,\boldsymbol{y}_{\setminus i})$ for all $\boldsymbol{y}_1,...,\boldsymbol{y}_{N+1}\in\tilde{S}$ and all $\Upsilon\in\mathcal{Y}$. Without loss of generality let $\boldsymbol{x}=\boldsymbol{y}_{i_0}$ for some $i_0\in\{1,...,N+1\}$. Then,

$$\begin{aligned}
P^{(1)}_\Upsilon(\boldsymbol{x},B) &= \int_{\mathbb{R}^{Nd}}\kappa_\Upsilon(\boldsymbol{y}_{i_0},\boldsymbol{y}_{\setminus i_0})\sum_{i_1=1}^{N+1}A(i_0,i_1)\mathbb{1}_B(\boldsymbol{y}_{i_1})\mathrm{d}\boldsymbol{y}_{\setminus i_0}\\
&\geq c_N c_A\sum_{i_1=1}^{N+1}\int_{\tilde{S}}\mathbb{1}_B(\boldsymbol{y}_{i_1})\mathrm{d}\boldsymbol{y}_{\setminus i_0}\\
&\geq c_N c_A(N+1)\lambda^N\left(B\cap\tilde{S}\right).
\end{aligned}$$

Set $\delta := c_N c_A(N+1)\lambda^N(\tilde{S})$ and $\nu_\Upsilon(B)=\nu(B):=\lambda^N(B\cap\tilde{S})/\lambda^N(\tilde{S})$. This concludes the first condition for SSAGE. The second one follows by setting $\rho=1/2$, $b=1$ and $V\equiv 1$.

A.5.2 Independence assumption

(I) Diminishing adaptation: Due to the independence sampler we need to consider $\Upsilon=(\boldsymbol{\mu},\Sigma)$. The analogue to (19) for $\Upsilon=(\boldsymbol{\mu},\Sigma)$ reads as,

$$\begin{aligned}
&\left\|P^{(1)}_{\Upsilon_{\ell+1}}(\boldsymbol{x},B) - P^{(1)}_{\Upsilon_\ell}(\boldsymbol{x},B)\right\|\\
&\leq (N+1)\int_{\mathbb{R}^{Nd}}\int_0^1\left|\frac{\mathrm{d}}{\mathrm{d}s}\prod_{n=1}^{N}\kappa_{(\boldsymbol{\mu}_\ell+s(\boldsymbol{\mu}_{\ell+1}-\boldsymbol{\mu}_\ell),\Sigma_\ell+s(\Sigma_{\ell+1}-\Sigma_\ell))}(\boldsymbol{y}_n)\right|\mathrm{d}s\ \mathrm{d}\boldsymbol{y}_{1:N}.
\end{aligned}$$

(A) $\underline{\kappa_{(\boldsymbol{\mu},\Sigma)} = \mathcal{N}_{(\boldsymbol{\mu},\Sigma)}}$: Setting $\boldsymbol{m}_\ell(s) = \boldsymbol{\mu}_\ell + s(\boldsymbol{\mu}_{\ell+1} - \boldsymbol{\mu}_\ell)$, leads to

$$\begin{aligned}
&\prod_{n=1}^{N} \mathcal{N}_{(\boldsymbol{\mu}_\ell + s(\boldsymbol{\mu}_{\ell+1}-\boldsymbol{\mu}_\ell), \Sigma_\ell + s(\Sigma_{\ell+1}-\Sigma_\ell))}(\boldsymbol{y}_n) \\
&= (2\pi)^{d\frac{N}{2}} \det\left(A_\ell(s)\right)^{-\frac{N}{2}} \exp\left(-\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{y}_n - \boldsymbol{m}_\ell(s))^T A_\ell(s)^{-1}(\boldsymbol{y}_n - \boldsymbol{m}_\ell(s))\right)
\end{aligned}$$

The determinant term can be bounded via (22). Using the product rule, the derivative of the exponential term follows as

$$\begin{aligned}
&\left|\frac{\mathrm{d}}{\mathrm{d}s}\exp\left(-\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{y}_n - \boldsymbol{m}_\ell(s))^T A_\ell(s)^{-1}(\boldsymbol{y}_n - \boldsymbol{m}_\ell(s))\right)\right| \\
&\leq \text{const}\sum_{n=1}^{N}\left|\left(\sum_{i=1}^{d}[\boldsymbol{y}_n]_i - [\boldsymbol{m}_\ell(s)]_i\right) + (\boldsymbol{y}_n - \boldsymbol{m}_\ell(s))^T(\boldsymbol{y}_n - \boldsymbol{m}_\ell(s))\right| \cdot \\
&\quad \exp\left(-\frac{1}{2c_1^2}\sum_{n=1}^{N}(\boldsymbol{y}_n - \boldsymbol{m}_\ell(s))^T(\boldsymbol{y}_n - \boldsymbol{m}_\ell(s))\right)\left(\|\Sigma_{\ell+1} - \Sigma_\ell\| + \|\boldsymbol{\mu}_{\ell+1} - \boldsymbol{\mu}_\ell\|\right).
\end{aligned}$$

Due to the boundedness of first and second moments of the Normal distribution, we have

$$\begin{aligned}
\left\|P^{(1)}_{\Upsilon_{\ell+1}}(\boldsymbol{x}, B) - P^{(1)}_{\Upsilon_\ell}(\boldsymbol{x}, B)\right\| &\leq \text{const}\left(\|\Sigma_{\ell+1} - \Sigma_\ell\| + \|\boldsymbol{\mu}_{\ell+1} - \boldsymbol{\mu}_\ell\|\right) && (28)\\
&\leq \text{const}\cdot\frac{1}{\ell} \longrightarrow 0 \quad \text{for } \ell \to \infty. && (29)
\end{aligned}$$

The case **(B)** $\underline{\kappa_\Upsilon = t_\Upsilon}$ follows again analogously to the computations of derivatives in the Gaussian case and by using the boundedness of moments of the t-distribution.

(II) SSAGE: Set $C = \mathbb{R}^d$ and $V \equiv 1$. Since the proposal distribution is independent of previous samples except for their effect on Υ, we may set $\delta = 1$, $\nu_\Upsilon = P_\Upsilon$. With $\rho = 1/2$ and $b = 1$ we have $P^{(m)}_\gamma V(\boldsymbol{x}) = 1 \leq \rho \cdot 1 + b = \rho V(\boldsymbol{x}) + b\mathbb{1}_C(\boldsymbol{x})$.

$\square$