Geometric Ergodicity in Hidden Markov Models

François Le Gland, Laurent Mevel

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Geometric Ergodicity in Hidden Markov Models

François LE GLAND, Laurent MEVEL

N ° 2991

Septembre 1996

THÈME 4

Rapport de recherche

ISSN 0249-6399

INRIA RENNES

Geometric Ergodicity in Hidden Markov Models

François LE GLAND*, Laurent MEVEL**

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet SIGMA2 (σ^2)

Rapport de recherche n° 2991 — Septembre 1996 — 54 pages

Abstract: We consider an hidden Markov model with multidimensional observations, and with misspecification, i.e. the *assumed* coefficients (transition probability matrix, and observation conditional densities) are possibly different from the *true* coefficients. Under mild assumptions on the coefficients of both the true and the assumed models, we prove that : (i) the prediction filter, and its gradient w.r.t. some parameter in the model, forget almost surely their initial condition exponentially fast, and (ii) the extended Markov chain, whose components are : the unobserved Markov chain, the observation sequence, the prediction filter, and its gradient, is geometrically ergodic and has a unique invariant probability distribution.

Key-words: HMM, misspecified model, prediction filter, exponential forgetting, geometric ergodicity, product of random matrices, Birkhoff contraction coefficient

(Résumé : tsvp)

This work was partially supported by the Commission of the European Communities, under the SCIENCE project *System Identification*, project number SC1*–CT92–0779, and under the HCM project *Statistical Inference for Stochastic Processes*, project number CHRX–CT92–0078, and by the Army Research Office, under grant DAAH04–95–1–0164.

* IRISA / INRIA, Campus de Beaulieu, 35042 Rennes Cédex, France. — legland@irisa.fr
** IRMAR and IRISA / INRIA, Campus de Beaulieu, 35042 Rennes Cédex, France. — lmevel@irisa.fr

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : (33) 99 84 71 00 – Télécopie : (33) 99 84 71 71

Ergodicité Géométrique
dans les Modèles de Markov Cachés

Résumé : On considère un modèle de Markov caché avec observations vectorielles, et mal spécifié, c'est–à–dire que les coefficients *supposés* (matrice des probabilités de transition, et densités conditionnelles des observations) sont éventuellement différents des *vrais* coefficients. Sous des hypothèses assez faibles portant sur les coefficients du vrai modèle et du modèle supposé, on montre que : (i) la distribution conditionnelle de la prédiction, et son gradient par rapport à un coefficient quelconque du modèle, oublient presque sûrement leur condition initiale, avec une vitesse exponentielle, et (ii) la chaîne de Markov étendue, dont les composantes sont : la chaîne de Markov non–observée, la suite des observations, la distribution conditionnelle de la prédiction, et son gradient, est géométriquement ergodique, et possède une unique distribution de probabilité invariante.

Mots-clé : modèle de Markov caché, modèle mal spécifié, prédiction, oubli exponentiel, ergodicité géométrique, produit de matrices aléatoires, coefficient de contraction de Birkhoff

Contents

1 Introduction

Let $\{X_n \, , \, n \geq 0\}$ and $\{Y_n \, , \, n \geq 0\}$ be two random sequences defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P}_\bullet)$, with values in the finite set $S = \{1, \cdots, N\}$ and in $\mathbf{R}^d$ respectively. It is assumed that :

- The unobserved state sequence $\{X_n \, , \, n \geq 0\}$ is a time–homogeneous Markov chain with transition probability matrix $Q_\bullet = (q_\bullet^{i,j})$, i.e. for any integer $n \geq 0$, and for any $i, j \in S$

$$\mathbf{P}_\bullet[X_{n+1} = j \mid X_n = i] = q_\bullet^{i,j} \; ,$$

  and initial probability distribution $p_\bullet = (p_\bullet^i)$, i.e. for any $i \in S$

$$\mathbf{P}_\bullet[X_0 = i] = p_\bullet^i \; .$$

- The observations $\{Y_n \, , \, n \geq 0\}$ are mutually independent given the sequence of states of the Markov chain, i.e. for any integer $n \geq 0$, and for any $i_0, \cdots, i_n \in S$

$$\mathbf{P}_\bullet[Y_n \in dy_n, \cdots, Y_0 \in dy_0 \mid X_n = i_n, \cdots, X_0 = i_0] = \prod_{k=0}^{n} \mathbf{P}_\bullet[Y_k \in dy_k \mid X_k = i_k] \; .$$

  For any integer $n \geq 0$, and for any $i \in S$, the conditional probability distribution of the observation Y_n given that $(X_n = i)$ is absolutely continuous with respect to a non–negative and σ–finite measure λ on $\mathbf{R}^d$, i.e.

$$\mathbf{P}_\bullet[Y_n \in dy \mid X_n = i] = b_\bullet^i(y) \; \lambda(dy) \; ,$$

  with a λ–a.e. positive density. For any $y \in \mathbf{R}^d$, let

$$b_\bullet(y) = [b_\bullet^1(y), \cdots, b_\bullet^N(y)]^* \qquad \text{and} \qquad B_\bullet(y) = \text{diag}\,[b_\bullet^1(y), \cdots, b_\bullet^N(y)] \; .$$

  Here and throughout the paper, the notation * denotes the transpose of a matrix.

Example 1.1 [conditionally Gaussian observations] Assume that the observations are of the form

$$Y_n = h_\bullet(X_n) + V_n \; ,$$

for any integer $n \geq 0$, where $\{V_n \, , \, n \geq 0\}$ is a Gaussian white noise sequence, with identity covariance matrix. The mapping $h_\bullet$ from S to $\mathbf{R}^d$ is equivalently defined as $h_\bullet = (h_\bullet^i)$ where $h_\bullet^i \in \mathbf{R}^d$ for any $i \in S$. In this case, λ is the Lebesgue measure on $\mathbf{R}^d$, the mutual independence condition is satisfied, and

$$b_\bullet^i(y) = (2\pi)^{-d/2} \, \exp\{\, -\tfrac{1}{2} \, |y - h_\bullet^i|^2 \,\} \; ,$$

for any $i \in S$, and any $y \in \mathbf{R}^d$. Here and throughout the paper, the notation $|\cdot|$ denotes the Euclidean norm.

Throughout the paper, we make the following assumption :

Assumption A : The stochastic matrix $Q_\bullet = (q_\bullet^{i,j})$ is primitive (or equivalently, irreducible and aperiodic).

Remark 1.2 Under Assumption A, there exist constants $0 < \rho_\bullet < 1$ and $K_\bullet > 1$ such that, for any integer $n \geq 1$

$$\tfrac{1}{2} \max_{i,i' \in S} \sum_{j \in S} |\, q_\bullet^{i,j}(n) - q_\bullet^{i',j}(n) \,| \leq K_\bullet \, \rho_\bullet^n \; , \tag{1}$$

and the Markov chain $\{X_n \, , \, n \geq 0\}$ is ergodic, with a unique invariant probability distribution $\mu_\bullet = (\mu_\bullet^i)$ on S. Here and throughout the paper $(q_\bullet^{i,j}(n))$ denote the entries of the stochastic matrix $Q_\bullet^n$, i.e. for any $i, j \in S$

$$\mathbf{P}_\bullet[X_n = j \mid X_0 = i] = q_\bullet^{i,j}(n) \; .$$

For any integer $n \geq 1$, let $p_n^\bullet = (p_n^i)$ denote the *prediction filter*, i.e. the conditional probability distribution under $\mathbf{P}_\bullet$ of the state X_n given observations $(Y_0, \cdots, Y_{n-1})$: for any $i \in S$

$$p_n^i = \mathbf{P}_\bullet[X_n = i \mid Y_0, \cdots, Y_{n-1}] \; .$$

The random sequence $\{p_n^\bullet \, , \, n \geq 0\}$ takes values in the set $\mathcal{P}(S)$ of probability distributions over the finite set S, and satisfies the forward Baum equation

$$p_{n+1}^\bullet = \frac{Q_\bullet^* B_\bullet(Y_n) \, p_n^\bullet}{b_\bullet^*(Y_n) \, p_n^\bullet} \; , \tag{2}$$

for any integer $n \geq 0$, with initial condition $p_0^\bullet = p_\bullet$.

In practice, the transition probability matrix $Q_\bullet$ and the initial probability distribution $p_\bullet$ of the unobserved Markov chain $\{X_n \, , \, n \geq 0\}$, and the conditional densities $b_\bullet(\cdot)$ of the observation sequence $\{Y_n \, , \, n \geq 0\}$ are possibly unknown. For this reason, we consider instead of (2) the more general equation

$$p_{n+1} = \frac{Q^* B(Y_n) \, p_n}{b^*(Y_n) \, p_n} = f[Y_n, p_n] \; , \tag{3}$$

for any integer $n \geq 0$, with initial condition $p_0 = p$, where $Q = (q^{i,j})$ is a $N \times N$ stochastic matrix, $p = (p^i)$ is a probability vector on S, and $b(\cdot) = (b^i(\cdot))$ are λ–a.e. positive densities on $\mathbf{R}^d$.

To make explicit the dependency w.r.t. the initial condition and the observations, we introduce the notation

$$p_{n+1} = f[Y_n, \cdots, Y_m, p_m] \; ,$$

for any integers n, m such that $n \geq m$.

Under the *true* probability measure $\mathbf{P}_\bullet$, the extended Markov chain $\{Z_n = (X_n, Y_n, p_n) \, , \, n \geq 0\}$ with values in $S \times \mathbf{R}^d \times \mathcal{P}(S)$, has the following transition probability matrix / kernel

$$\begin{aligned} \Pi^{i,j}(y, p, dy', dp') &= \mathbf{P}_\bullet[X_{n+1} = j, Y_{n+1} \in dy', p_{n+1} \in dp' \mid X_n = i, Y_n = y, p_n = p] \\ &= q_\bullet^{i,j} \; b_\bullet^j(y') \, \lambda(dy') \, \delta_{f[y,\,p]}(dp') \; . \end{aligned}$$

Notice that both

- the initial condition for the prediction filter,
- the transition matrix, and the observation conditional densities,

are possibly unknown. However these two misspecification issues are of a different nature :

- we expect that a wrong initial condition for the prediction filter is rapidly forgotten, so that we could use any initial condition with practically the same effect,
- on the other hand, we expect that two different transition probability matrices, and two different observation conditional densities will have a significantly different effect, i.e. will produce significantly different values for some related Kullback–Leibler information, so that we could estimate the unknown transition probability matrix and the unknown observation conditional densities accurately, by accumulating observations — this estimation problem will be addressed in subsequent papers.

Indeed, it can be shown that the log–likelihood function for the estimation of the unknown transition probability matrix and the unknown observation conditional densities, can be easily expressed as an additive functional of the extended Markov chain $\{Z_n = (X_n, Y_n, p_n) \, , \, n \geq 0\}$. Consequently, an asymptotic expression can be obtained for the corresponding Kullback–Leibler information provided some ergodicity property holds.

These statements are made rigorous in this paper, with the proof of the following two properties :

- the exponential forgetting of the initial condition for the sequence $\{p_n \, , \, n \geq 0\}$ defined in (3),
- and the geometric ergodicity of the extended Markov chain $\{Z_n = (X_n, Y_n, p_n) \, , \, n \geq 0\}$.

In the case of conditionally Gaussian observations, the exponential forgetting property has been investigated by Atar and Zeitouni [2] in terms of Lyapunov exponents, and the existence of a unique invariant probability distribution for a related extended Markov chain (based on the filter itself instead of the prediction filter) has been proved by DiMasi and Stettner [4] and by Stettner [12], under some additional restrictive assumption. In this paper, we adopt the approach initiated, in a special case of HMM with observations in a finite set, by Arapostathis and Marcus [1] — see also LeGland and Mevel [9] for a generalization.

Inspired by Arapostathis and Marcus [1], and motivated by the problem of recursive identification of the model, we consider the following more general problem : Assume that the coefficients of (3), i.e. the transition probability matrix Q or the vector $b(\cdot) = (b^i(\cdot))$ of observation densities, depend on some one–dimensional parameter. Differentiating (3) w.r.t. this parameter yields

$$\partial p_{n+1} = \frac{Q^* B(Y_n)}{b^*(Y_n)\, p_n} \, [\, I - \frac{p_n\, b^*(Y_n)}{b^*(Y_n)\, p_n} \,] \; \partial p_n + \partial f[Y_n, p_n] \; ,$$

where

$$\partial f[Y_n, p_n] = \frac{\partial Q^* \, B(Y_n)\, p_n}{b^*(Y_n)\, p_n} + \frac{Q^* B(Y_n)}{b^*(Y_n)\, p_n} \, [\, I - \frac{p_n\, b^*(Y_n)}{b^*(Y_n)\, p_n} \,] \; \partial \Lambda(Y_n)\, p_n \; ,$$

and for any $y \in \mathbf{R}^d$

$$\partial \Lambda(y) = \operatorname{diag}\,[\partial \log[\, b^1(y)\,], \cdots, \partial \log[\, b^N(y)\,]\,] \; .$$

Example 1.3 If the stochastic matrix Q is parametrized by its off–diagonal entries, and the one–dimensional parameter is precisely the (i,j)–th entry $q^{i,j}$ with $i \neq j$, then

$$\partial Q^* = (e_j - e_i)\, e_i^* \; .$$

Example 1.4 In the case of conditionally Gaussian observations, if the one–dimensional parameter is precisely the k–th coordinate h_k^i of the i–th mean vector, then for any $y \in \mathbf{R}^d$

$$\partial \Lambda(y) = (y_k - h_k^i)\; e_i\, e_i^* \; .$$

More generally, we consider the following equation

$$\begin{aligned} w_{n+1} \quad &= \quad \frac{Q^* B(Y_n)}{b^*(Y_n)\, p_n} \, [\, I - \frac{p_n\, b^*(Y_n)}{b^*(Y_n)\, p_n} \,] \; w_n + u[Y_n, p_n] \\ &= \quad \Phi[Y_n, p_n]\; w_n + u[Y_n, p_n] = F[Y_n, p_n, w_n] \; , \end{aligned} \tag{4}$$

with values in

$$\Sigma = \{ w \in \mathbf{R}^N \; : \; e^* \, w = 0 \} \; ,$$

where $e = (1, \cdots, 1)^*$ denotes the N–dimensional vector with all entries equal to 1. To make explicit the dependency w.r.t. the initial condition and the observations, we introduce the notation

$$w_{n+1} = F[Y_n, \cdots, Y_m, p_m, w_m] \; ,$$

for any integers n, m such that $n \geq m$.

Under the *true* probability measure $\mathbf{P}_\bullet$, the extended Markov chain $\{ Z_n' = (X_n, Y_n, p_n, w_n) \; , \; n \geq 0 \}$ with values in $S \times \mathbf{R}^d \times \mathcal{P}(S) \times \Sigma$, has the following transition probability matrix / kernel

$$\begin{aligned} & \Pi^{i,j}(y, p, w, dy', dp', dw') = \\ = \quad & \mathbf{P}_\bullet[X_{n+1} = j, Y_{n+1} \in dy', p_{n+1} \in dp', w_{n+1} \in dw' \mid X_n = i, Y_n = y, p_n = p, w_n = w] \\ = \quad & q_\bullet^{i,j}\; b_\bullet^j(y')\, \lambda(dy')\; \delta_{f[y,\,p]}(dp')\; \delta_{F[y,\,p,\,w]}(dw') \; . \end{aligned}$$

We will also consider the time–dependent equations

$$p_{n+1} = \frac{Q_n^* B_n(Y_n)\; p_n}{b_n^*(Y_n)\; p_n} = f_n[Y_n, p_n] \; , \tag{5}$$

and

$$\begin{aligned} w_{n+1} &= \frac{Q_n^* B_n(Y_n)}{b_n^*(Y_n)\, p_n} \, [\, I - \frac{p_n\, b_n^*(Y_n)}{b_n^*(Y_n)\, p_n} \,] \, w_n + u_n[Y_n, p_n] \\ &= \Phi_n[Y_n, p_n] \, w_n + u_n[Y_n, p_n] = F_n[Y_n, p_n, w_n] \; , \end{aligned} \tag{6}$$

for any integer $n \geq 0$, where $Q_n = (q_n^{i,j})$ is a stochastic matrix, and $b_n(\cdot) = (b_n^i(\cdot))$ are λ–a.e. positive densities. To make explicit the dependency w.r.t. the initial condition and the observations, we introduce the notations

$$p_{n+1} = f_{n,m}[Y_n, \cdots, Y_m, p_m] \; ,$$

and

$$w_{n+1} = F_{n,m}[Y_n, \cdots, Y_m, p_m, w_m] \; ,$$

for any integers n, m such that $n \geq m$.

Here is a short overview of our results in the time–independent case, under the assumption that the stochastic matrix Q is primitive :

- We obtain in Theorem 2.3 a bound for the difference between the solutions of equation (3) starting from two different initial conditions. As a corollary, we obtain in Proposition 2.5, an upper bound for the $\mathbf{P}_\bullet$–a.s. exponential rate of forgetting of the initial condition for equation (3).
- Using the estimate of Theorem 2.3, we prove in Theorem 3.6, under some integrability assumption on the observation densities $b(\cdot)$ and $b_\bullet(\cdot)$, the geometric ergodicity of the extended Markov chains $\{Z_n = (X_n, Y_n, p_n) \, , \, n \geq 0\}$.
- We obtain in Theorem 4.6 a bound for the difference between the solutions of equation (4) starting from two different initial conditions. As a corollary, we obtain in Proposition 4.7, under some integrability assumption on the observation densities $b(\cdot)$ and $b_\bullet(\cdot)$, and on the function u, an upper bound for the $\mathbf{P}_\bullet$–a.s. exponential rate of forgetting of the initial condition for equation (4).
- Using the estimate of Theorem 4.6, we prove in Theorem 5.4, under some integrability assumption on the observation densities $b(\cdot)$ and $b_\bullet(\cdot)$, and on the function u, the geometric ergodicity of the extended Markov chains $\{Z'_n = (X_n, Y_n, p_n, w_n) \, , \, n \geq 0\}$.

And in the time–dependent case, under the assumption that for any integer $n \geq 0$, the stochastic matrix $Q_n = (q_n^{i,j})$ is positive, and $\varepsilon_n = \min_{i,j \in S} q_n^{i,j} > 0$ does not go to zero too rapidly :

- We obtain in Proposition 2.1 and in Proposition 4.1 respectively, a bound for the difference between the solutions of equation (5) and (6) starting from two different initial conditions.

We notice that

$$f_{n,m}[y_n, \cdots, y_m, p] = \frac{Q_n^* B(y_n) \cdots Q_m^* B(y_m)\; p}{e^* \; [\, Q_n^* B(y_n) \cdots Q_m^* B(y_m)\; p\,]} = \frac{M_{n,m}\; p}{e^* \; M_{n,m}\; p} \; ,$$

if we define

$$M_{n,m} = Q_n^* B(y_n) \cdots Q_m^* B(y_m) \; ,$$

and our results will be based on auxiliary estimates for products of non–negative matrices, which improve earlier estimates in Furstenberg and Kesten [6]. These results are stated with proofs in Appendix A.

Remark 1.5 Most of this work could be generalized easily, under suitable assumptions, to the case where the state space of the Markov chain $\{X_n \, , \, n \geq 0\}$ is compact.

2 Exponential forgetting for the prediction filter

Throughout the paper, $\|\cdot\|$ will denote the L_1–norm, i.e. for any $u = (u^i)$ in $\mathbf{R}^N$

$$\|u\| = \sum_{i \in S} |u^i| \; .$$

Time–dependent case

Recall that in the time–dependent case $\{p_n \, , \, n \geq 0\}$ satisfies equation (5), i.e.

$$p_{n+1} = \frac{Q_n^* B_n(Y_n) \, p_n}{b_n^*(Y_n) \, p_n} = f_n[Y_n, p_n] = f_{n,m}[Y_n, \cdots, Y_m, p_m] \; ,$$

for any integers n, m such that $n \geq m$, where the stochastic matrix $Q_n = (q_n^{i,j})$ is positive, and $b_n(\cdot) = (b_n^i(\cdot))$ are λ–a.e. positive densities, and we define for any $y \in \mathbf{R}^d$

$$\delta_n(y) = \frac{\max_{i \in S} b_n^i(y)}{\min_{i \in S} b_n^i(y)} < \infty \qquad \text{and} \qquad \varepsilon_n = \min_{i,j \in S} q_n^{i,j} > 0 \; .$$

Proposition 2.1 *For any $p, p' \in \mathcal{P}(S)$, any integers n, m such that $n \geq m$, and any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$*

$$\| f_{n,m}[y_n, \cdots, y_m, p] - f_{n,m}[y_n, \cdots, y_m, p'] \| \leq 2 \, \varepsilon_m^{-1} \, \delta_m(y_m) \prod_{k=m}^{n} (1 - \varepsilon_k) \, \| p - p' \| \; .$$

This is an immediate application of Proposition A.4.

If $\prod_{k=m}^{\infty} (1 - \varepsilon_k) = 0$, then for any $p, p' \in \mathcal{P}(S)$, and for any infinite sequence $y_m, \cdots, y_n, \cdots \in \mathbf{R}^d$, the difference $\| f_{n,m}[y_n, \cdots, y_m, p] - f_{n,m}[y_n, \cdots, y_m, p'] \|$ goes to zero. If in addition $\limsup_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \log(1 - \varepsilon_k) < 0$, then the rate is exponential

$$\limsup_{n \to \infty} \frac{1}{n} \, \log \| f_{n,m}[y_n, \cdots, y_m, p] - f_{n,m}[y_n, \cdots, y_m, p'] \| \leq \limsup_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \log(1 - \varepsilon_k) < 0 \; ,$$

provided $p \neq p'$.

Time–independent case

In the time–independent case $\{p_n \, , \, n \geq 0\}$ satisfies (3), i.e.

$$p_{n+1} = \frac{Q^* B(Y_n) \, p_n}{b^*(Y_n) \, p_n} = f[Y_n, p_n] = f[Y_n, \cdots, Y_m, p_m] \; ,$$

for any integers n, m such that $n \geq m$. For the stochastic matrix $Q = (q^{i,j})$, and for any $y \in \mathbf{R}^d$, we define

$$\delta(y) = \frac{\max_{i \in S} b^i(y)}{\min_{i \in S} b^i(y)} < \infty \qquad \text{and} \qquad \varepsilon = \min_{i,j \in S}{}^{+} \, q^{i,j} > 0 \; ,$$

where the notation $\min^+$ denotes the minimum over positive elements.
By particularizing to the time–independent case the result of the more general Proposition 2.1 above, we obtain the following result.

Proposition 2.2 *If the stochastic matrix Q is positive, then for any $p, p' \in \mathcal{P}(S)$, any integers n, m such that $n \geq m$, and any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$*

$$\| f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p'] \| \leq 2 \, \varepsilon^{-1} \, \delta(y_m) \, (1 - \varepsilon)^{n-m+1} \, \| p - p' \| \; .$$

For any $p, p' \in \mathcal{P}(S)$ such that $p \neq p'$, and for any infinite sequence $y_m, \cdots, y_n, \cdots \in \mathbf{R}^d$, the difference $\| f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p'] \|$ goes to zero at exponential rate

$$\limsup_{n \to \infty} \frac{1}{n} \, \log \| f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p'] \| \leq \log(1 - \varepsilon) \; .$$

Theorem 2.3 *If the stochastic matrix Q is primitive, with index of primitivity r, then for any $p, p' \in \mathcal{P}(S)$, any integers n, m such that $n \geq m + r - 1$, and any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$*

$$\| f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p'] \| \leq$$

$$\leq 2\, \varepsilon^{-r}\, \delta(y_m) \cdots \delta(y_{m+r-1}) \prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r\, [\, \delta(y_{m+\kappa r+1}) \cdots \delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\, \|p - p'\| \ ,$$

where $[n, m] = \lfloor \dfrac{n - m + 1}{r} \rfloor - 1.$

This is an immediate application of Proposition A.5.

For any $p, p' \in \mathcal{P}(S)$, such that $p \neq p'$, and for any infinite sequence $y_m, \cdots, y_n, \cdots \in \mathbf{R}^d$, such that

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{\kappa=0}^{[n,m]} \log\,(1 - \varepsilon^r\, [\, \delta(y_{m+\kappa r+1}) \cdots \delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,) < 0 \ ,$$

the difference $\| f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p'] \|$ goes to zero at exponential rate

$$\limsup_{n \to \infty} \frac{1}{n}\ \log\, \| f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p'] \| \leq$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \sum_{\kappa=0}^{[n,m]} \log\,(1 - \varepsilon^r\, [\, \delta(y_{m+\kappa r+1}) \cdots \delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,) < 0 \ .$$

To obtain an estimate of the almost sure exponential rate of forgetting in this case, we define

$$\Delta_{-1} = \min_{i \in S} \int_{\mathbf{R}^d} \delta^{-1}(y)\, b^i_\bullet(y)\, \lambda(dy) \ .$$

Notice that $0 < \Delta_{-1} \leq 1$, hence for any sequence $i_1, \cdots, i_r \in S$

$$\begin{aligned} 0 \ &\leq \ \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} (1 - \varepsilon^r\, [\, \delta(y_2) \cdots \delta(y_r)\,]^{-1}\,) \\ &\qquad\qquad b^{i_1}_\bullet(y_1) \cdots b^{i_r}_\bullet(y_r)\, \lambda(dy_1) \cdots \lambda(dy_r) = \\ &= \ 1 - \varepsilon^r \prod_{k=2}^{r} \int_{\mathbf{R}^d} \delta^{-1}(y)\, b^{i_k}_\bullet(y)\, \lambda(dy) \leq 1 - \varepsilon^r\ \Delta_{-1}^{r-1} = 1 - R < 1 \ , \end{aligned} \tag{7}$$

with $R = \varepsilon^r\ \Delta_{-1}^{r-1} > 0$. Notice that when the matrix Q is positive, i.e. when $r = 1$, then R reduces to $R = \varepsilon$.

Remark 2.4 For any integers n, m such that $n \geq m + 2r - 1$

$$r\,[n, m] = r\, \lfloor \frac{n - m + 1}{r} \rfloor - r > n - m + 1 - 2r \ ,$$

hence

$$(1 - R)^{[n,m]} \leq \rho_*^{n-m+1-2r} \ , \tag{8}$$

with $\rho_* = (1 - R)^{1/r}$.

Proposition 2.5 *If the stochastic matrix Q is primitive, with index of primitivity r, and if Assumption A holds, then for any $p, p' \in \mathcal{P}(S)$ such that $p \neq p'$, and any integer m*

$$\limsup_{n \to \infty} \frac{1}{n}\ \log\, \| f[Y_n, \cdots, Y_m, p] - f[Y_n, \cdots, Y_m, p'] \| \leq \frac{1}{r}\ \log(1 - R) \ , \qquad \mathbf{P}_\bullet\text{-}a.s.$$

where $R = \varepsilon^r\ \Delta_{-1}^{r-1}$.

PROOF. The estimate of Theorem 2.3 above yields

$$\|f[Y_n,\cdots,Y_m,p]-f[Y_n,\cdots,Y_m,p']\| \leq$$

$$\leq 2\,\varepsilon^{-r}\,\delta(Y_m)\cdots\delta(Y_{m+r-1})\prod_{\kappa=0}^{[n,m]}(1-\varepsilon^r\,[\,\delta(Y_{m+\kappa r+1})\cdots\delta(Y_{m+(\kappa+1)r-1})\,]^{-1})\,\|p-p'\|\;,$$

hence

$$\limsup_{n\to\infty}\frac{1}{n}\,\log\|f[Y_n,\cdots,Y_m,p]-f[Y_n,\cdots,Y_m,p']\| \leq$$

$$\leq \frac{1}{r}\,\limsup_{\ell\to\infty}\frac{1}{\ell}\sum_{\kappa=0}^{\ell}\log(1-\varepsilon^r\,[\,\delta(Y_{m+\kappa r+1})\cdots\delta(Y_{m+(\kappa+1)r-1})\,]^{-1})\;.$$

If Assumption A holds, then the Markov chain $\{(X_n,Y_n)\,,\,n\geq 0\}$ has, under the true probability measure $\mathbf{P}_\bullet$, a unique invariant probability distribution $\nu_\bullet=(\nu_\bullet^i)$ on $S\times\mathbf{R}^d$ and for any $i\in S$

$$\nu_\bullet^i(dy)=\mu_\bullet^i\,b_\bullet^i(y)\,\lambda(dy)\;.$$

Therefore, $\mathbf{P}_\bullet$–a.s

$$\lim_{\ell\to\infty}\frac{1}{\ell}\sum_{\kappa=0}^{\ell}\log(1-\varepsilon^r\,[\,\delta(Y_{m+\kappa r+1})\cdots\delta(Y_{m+(\kappa+1)r-1})\,]^{-1})=$$

$$=\sum_{i_1,\cdots,i_r\in S}\int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d}\log(1-\varepsilon^r\,[\,\delta(y_2)\cdots\delta(y_r)\,]^{-1})$$

$$\mu_\bullet^{i_1}\,q_\bullet^{i_1,i_2}\cdots q_\bullet^{i_{r-1},i_r}\,b_\bullet^{i_1}(y_1)\cdots b_\bullet^{i_r}(y_r)\,\lambda(dy_1)\cdots\lambda(dy_r)\;.$$

Using the Jensen inequality, and the estimate (7) above, the limit is bounded by $\log(1-R)<0$. □

Remark 2.6 In Atar and Zeitouni [2, Corollary 2.1] the exact exponential rate of forgetting is expressed, in the case of an HMM with conditionally Gaussian observations, without misspecification, as the difference between the two top Lyapunov exponents of the equation (2). Since no explicit expression is available in general for these Lyapunov exponents, estimates for the exponential rate are given in Theorems 1.3 and 1.4 of [2] when the signal–to–noise ratio is large or small respectively.

In this paper, we not only obtain explicit estimates for the exponential rate of forgetting, but we also obtain non–logarithmic and non–asymptotic bounds, in the more general case of a misspecified HMM with arbitrary observation conditional densities, and with primitive transition probability matrices (for both the true and the assumed models). Our proof of the geometric ergodicity of the extended Markov chain $\{Z_n=(X_n,Y_n,p_n)\,,\,n\geq 0\}$ is based on these explicit bounds.

3 Geometric ergodicity of the Markov chain $\{X_n,Y_n,p_n\}$

Under the probability measure $\mathbf{P}_\bullet$ corresponding to the *true* transition probability matrix $Q_\bullet$ and the *true* observation densities $b_\bullet(\cdot)$, the extended Markov chain $\{Z_n=(X_n,Y_n,p_n)\,,\,n\geq 0\}$ has the following transition probability matrix / kernel

$$\begin{aligned}\Pi^{i,j}(y,p,dy',dp') &= \mathbf{P}_\bullet[X_{n+1}=j,Y_{n+1}\in dy',p_{n+1}\in dp'\mid X_n=i,Y_n=y,p_n=p]\\ &= q_\bullet^{i,j}\,b_\bullet^j(y')\,\lambda(dy')\,\delta_{f[y,\,p]}(dp')\;.\end{aligned}$$

For any real–valued function g defined on $S\times\mathbf{R}^d\times\mathcal{P}(S)$, which is equivalently defined as a collection $g=(g^i)$ of real–valued functions defined on $\mathbf{R}^d\times\mathcal{P}(S)$, we have

$$(\Pi\,g)^i(y,p) = \mathbf{E}_\bullet[g(X_{n+1},Y_{n+1},p_{n+1})\mid X_n=i,Y_n=y,p_n=p]$$

$$
\begin{aligned}
&= \sum_{j \in S} \mathbf{E}_\bullet [g^j(Y_{n+1}, p_{n+1}) \, \mathbf{1}_{[X_{n+1} = j]} \mid X_n = i, Y_n = y, p_n = p] \\
&= \sum_{j \in S} \int_{\mathbf{R}^d} g^j(y', f[y,p]) \, q_\bullet^{i,j} \, b_\bullet^j(y') \, \lambda(dy') \\
&= \sum_{j \in S} \int_{\mathbf{R}^d \times \mathcal{P}(S)} g^j(y', p') \, \Pi^{i,j}(y, p, dy', dp') \ ,
\end{aligned}
$$

for any $i \in S$, any $y \in \mathbf{R}^d$, and any $p \in \mathcal{P}(S)$. More generally, for any integer $n \geq 2$ we have

$$
\begin{aligned}
(\Pi^n \, g)^i(y,p) &= \mathbf{E}_\bullet [g(X_n, Y_n, p_n) \mid X_0 = i, Y_0 = y, p_0 = p] \\
&= \sum_{i_n \in S} \mathbf{E}_\bullet [g^{i_n}(Y_n, p_n) \, \mathbf{1}_{[X_n = i_n]} \mid X_0 = i, Y_0 = y, p_0 = p] \\
&= \sum_{i_1, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} g^{i_n}(y_n, f[y_{n-1}, \cdots, y_1, y, p]) \\
&\qquad\qquad q_\bullet^{i,i_1} \cdots q_\bullet^{i_{n-1},i_n} \, b_\bullet^{i_1}(y_1) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_1) \cdots \lambda(dy_n) \ ,
\end{aligned}
$$

for any $i \in S$, any $y \in \mathbf{R}^d$, and any $p \in \mathcal{P}(S)$.

In addition to the Assumption A on the *true* transition probability matrix $Q_\bullet$, we shall need the following *integrability* assumption on the observation densities $b(\cdot)$ and $b_\bullet(\cdot)$:

$$
\Delta = \max_{i \in S} \int_{\mathbf{R}^d} \delta(y) \, b_\bullet^i(y) \, \lambda(dy) < \infty \ .
$$

More generally, we introduce the following definition.

Definition 3.1 *For any $s \geq 0$, let*

$$
\Delta_s = \max_{i \in S} \int_{\mathbf{R}^d} \delta^s(y) \, b_\bullet^i(y) \, \lambda(dy) \ .
$$

Notice that $\Delta_1 = \Delta$, and that the mapping $s \mapsto \Delta_s$ is nondecreasing on $[0, \infty)$, since $\delta(y) \geq 1$ for any $y \in \mathbf{R}^d$.

Example 3.2 If the observation conditional densities are Gaussian for both the true and the assumed models, i.e. in particular

$$
b^i(y) = (2\pi)^{-d/2} \, \exp\{ -\tfrac{1}{2} |y - h^i|^2 \} \ ,
$$

for any $i \in S$, and any $y \in \mathbf{R}^d$, then Δ_s is finite for any $s \geq 0$. Indeed, for any $i, j \in S$, and any $y \in \mathbf{R}^d$

$$
\frac{b^i(y)}{b^j(y)} = \exp\{ -\tfrac{1}{2} |y - h^i|^2 + \tfrac{1}{2} |y - h^j|^2 \} = \exp\{ y^* \, (h^i - h^j) - \tfrac{1}{2} (h^i + h^j)^* \, (h^i - h^j) \} \ ,
$$

hence

$$
\delta(y) \leq \exp\{ \max_{i,j \in S} |h^i - h^j| \, [\, |y| + \max_{i \in S} |h^i| \,] \} \ .
$$

Definition 3.3 *Let L denote the set of functions $g = (g^i)$ defined on $S \times \mathbf{R}^d \times \mathcal{P}(S)$, such that for any $i \in S$, and any $y \in \mathbf{R}^d$ the partial mapping $p \mapsto g^i(y,p)$ is Lipschitz continuous (hence bounded since $\mathcal{P}(S)$ is compact), i.e.*

$$
|g^i(y,p) - g^i(y,p')| \leq \mathrm{Lip}(g^i, y) \, \|p - p'\| \ ,
$$

for any $p, p' \in \mathcal{P}(S)$, and such that

$$
\mathrm{Lip}(g) = \max_{i \in S} \int_{\mathbf{R}^d} \mathrm{Lip}(g^i, y) \, b_\bullet^i(y) \, \lambda(dy) < \infty \ ,
$$

$$\mathrm{K}(g) = \max_{i \in S} \int_{\mathbf{R}^d} \mathrm{K}(g^i, y)\, b^i_\bullet(y)\, \lambda(dy) < \infty \;,$$

where by definition

$$\mathrm{K}(g^i, y) = \sup_{p \in \mathcal{P}(S)} |g^i(y,p)| \;.$$

Example 3.4 [log–likelihood function] If Δ is finite, and if

$$\max_{i \in S} \int_{\mathbf{R}^d} [\max_{j \in S} |\log b^j(y)|\,]\, b^i_\bullet(y)\, \lambda(dy) < \infty \;,$$

then the function g defined by

$$g(y,p) = \log[\, b^*(y)\, p\,] \;,$$

for any $y \in \mathbf{R}^d$, and any $p \in \mathcal{P}(S)$, i.e. constant over S, belongs to the set L. Indeed, for any $y \in \mathbf{R}^d$, and any $p \in \mathcal{P}(S)$

$$\min_{j \in S} b^j(y) \leq b^*(y)\, p \leq \max_{j \in S} b^j(y) \;,$$

hence

$$\begin{aligned}
|\log[\, b^*(y)\, p\,]\,| &= \log^+[\, b^*(y)\, p\,] + \log^-[\, b^*(y)\, p\,] \\
&\leq \log^+[\, \max_{j \in S} b^j(y)\,] + \log^-[\, \min_{j \in S} b^j(y)\,] \\
&\leq \max_{j \in S}\, \log^+[\, b^j(y)\,] + \max_{j \in S}\, \log^-[\, b^j(y)\,] \leq 2\, \max_{j \in S}\, |\log b^j(y)| \;,
\end{aligned}$$

whereas for any $y \in \mathbf{R}^d$, and any $p, p' \in \mathcal{P}(S)$

$$\log[\, b^*(y)\, p\,] - \log[\, b^*(y)\, p'\,] = \log[\, 1 + \frac{b^*(y)\, (p-p')}{b^*(y)\, p'}\,] \leq \frac{b^*(y)\, (p-p')}{b^*(y)\, p'} \leq \delta(y)\, \|p - p'\| \;,$$

hence

$$|\log[\, b^*(y)\, p\,] - \log[\, b^*(y)\, p'\,]\,| \leq \delta(y)\, \|p-p'\| \;.$$

Example 3.5 [prediction error] Consider the $N \times d$ matrix $\phi = (\phi^i)$, where for any $i \in S$, the mean vector $\phi^i \in \mathbf{R}^d$ is defined by

$$\phi^i = \int_{\mathbf{R}^d} y\, b^i(y)\, \lambda(dy) \;,$$

assuming the integral exist. If

$$\max_{i \in S} \int_{\mathbf{R}^d} |y|^2\, b^i_\bullet(y)\, \lambda(dy) < \infty \;,$$

then the function g defined by

$$g(y,p) = \tfrac{1}{2}\, |y - \phi^*\, p|^2 \;,$$

for any $y \in \mathbf{R}^d$, and any $p \in \mathcal{P}(S)$, i.e. constant over S, belongs to the set L. Indeed, for any $y \in \mathbf{R}^d$, and any $p \in \mathcal{P}(S)$

$$\tfrac{1}{2}\, |y - \phi^*\, p|^2 \leq |y|^2 + \max_{i \in S} |\phi^i|^2 \;,$$

whereas for any $y \in \mathbf{R}^d$, and any $p, p' \in \mathcal{P}(S)$

$$\tfrac{1}{2}\, |y - \phi^*\, p|^2 - \tfrac{1}{2}\, |y - \phi^*\, p'|^2 = -(\phi\, y)^*\, (p - p') + \tfrac{1}{2}\, [\, \phi\, \phi^*\, (p + p')\,]^*\, (p - p') \;,$$

hence

$$|\, \tfrac{1}{2}\, |y - \phi^*\, p|^2 - \tfrac{1}{2}\, |y - \phi^*\, p'|^2\, | \leq \max_{i \in S} |\phi^i|\, [\, |y| + \max_{i \in S} |\phi^i|\,]\, \|p - p'\| \;.$$

Theorem 3.6 *If the stochastic matrix Q is primitive, with index of primitivity r, if Assumption A holds, and if Δ is finite, then there exist constants $0<\rho<1$ and $C>0$ such that, for any $z,z' \in S \times \mathbf{R}^d \times \mathcal{P}(S)$, and for any function $g=(g^i)$ in L*

$$|\Pi^n\, g(z) - \Pi^n\, g(z')| \leq C\, \varepsilon^{-r}\ [\,\mathrm{Lip}(g) + \mathrm{K}(g)\,]\ \rho^n \ ,$$

where the constant C depends only on r, Δ, and $K_\bullet$.

The following corollary holds, whose proof is similar to the proof of Proposition 2 in Benveniste, Métivier and Priouret [3, Part II, Chapter 2].

Corollary 3.7 *With the assumptions of Theorem 3.6, the Markov chain $\{Z_n = (X_n, Y_n, p_n)\ ,\ n \geq 0\}$ has, under the true probability measure $\mathbf{P}_\bullet$, a unique invariant probability distribution $\mu = (\mu^i)$ on $S \times \mathbf{R}^d \times \mathcal{P}(S)$. For any $z \in S \times \mathbf{R}^d \times \mathcal{P}(S)$, and for any function $g=(g^i)$ in L*

$$|\Pi^n\, g(z) - \lambda| \leq C\, \varepsilon^{-r}\ [\,\mathrm{Lip}(g) + \mathrm{K}(g)\,]\ \frac{\rho^n}{1-\rho}\ ,$$

and there exist a unique solution $V=(V^i)$ defined on $S \times \mathbf{R}^d \times \mathcal{P}(S)$ of the Poisson equation

$$[\,I - \Pi\,]\, V(z) = g(z) - \lambda\ ,$$

where the constant λ is defined as

$$\lambda = \sum_{i \in S}\ \int_{\mathbf{R}^d \times \mathcal{P}(S)} g^i(y,p)\, \mu^i(dy,dp)\ .$$

The proof of the theorem is based on the next proposition.

Proposition 3.8 *If the stochastic matrix Q is primitive, with index of primitivity r, and if Δ is finite, then for any $p,p' \in \mathcal{P}(S)$, for any integers n,m such that $n \geq m+2r-1$, and for any function $g=(g^i)$ in L*

$$\max_{i_m,\cdots,i_{n+1} \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} |g^{i_{n+1}}(y_{n+1}, f[y_n,\cdots,y_m,p]) - g^{i_{n+1}}(y_{n+1}, f[y_n,\cdots,y_m,p'])\,|$$
$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1}) \leq$$
$$\leq C\, \varepsilon^{-r}\ \mathrm{Lip}(g)\ \rho_*^{n-m+1-2r}\ .$$

where $\rho_* = (1-R)^{1/r}$, and where the constant C depends only on r, and Δ.

Proof. For any sequence $i_m, \cdots, i_{n+1} \in S$

$$\int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} |g^{i_{n+1}}(y_{n+1}, f[y_n,\cdots,y_m,p]) - g^{i_{n+1}}(y_{n+1}, f[y_n,\cdots,y_m,p'])\,|$$
$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1}) \leq$$
$$\leq \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \mathrm{Lip}(g^{i_{n+1}}, y_{n+1})\ \|f[y_n,\cdots,y_m,p] - f[y_n,\cdots,y_m,p']\,\|$$
$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1})$$
$$\leq \mathrm{Lip}(g) \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \|f[y_n,\cdots,y_m,p] - f[y_n,\cdots,y_m,p']\,\|$$
$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_m) \cdots \lambda(dy_n)\ .$$

Recall from Theorem 2.3, that

$$\| f[y_n , \cdots , y_m , p] - f[y_n , \cdots , y_m , p'] \| \leq$$

$$\leq 2\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,\prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\,\|p - p'\|$$

$$\leq 2\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,\prod_{\kappa=1}^{[n,m]} (1 - \varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\,\|p - p'\|\;.$$

Notice that the last expression is a product of factors which span disjoint blocks of length r, hence integration is straightforward, and the lower bound (2.4) yields

$$\int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} \| f[y_n , \cdots , y_m , p] - f[y_n , \cdots , y_m , p'] \|\, b_\bullet^{i_m}(y_m)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_m)\cdots\lambda(dy_n) \leq$$

$$\leq 2\,\varepsilon^{-r}\,\Delta^r\,(1-R)^{[n,m]}\,\|p - p'\| \leq 4\,\varepsilon^{-r}\,\Delta^r\,(1-R)^{[n,m]} \leq 4\,\varepsilon^{-r}\,\Delta^r\,\rho_*^{n-m+1-2r}\;,$$

for any sequence $i_m, \cdots , i_n \in S$. □

Proof of Theorem 3.6. Let $\rho_{\max} = \max(\rho_*, \rho_\bullet)$, where $\rho_* = (1-R)^{1/r}$, and where the constants $\rho_\bullet$ and R are defined in Remark 1.2 and in Proposition 2.5 respectively. Recall that

$$(\Pi^n\, g)^i(y,p) =$$

$$= \sum_{i_1,\cdots,i_n \in S} \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} g^{i_n}(y_n, f[y_{n-1}, \cdots , y_1, y, p])\, q_\bullet^{i,i_1}\cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_1)\cdots\lambda(dy_n)\;,$$

for any $i \in S$, any $y \in \mathbf{R}^d$, and any $p \in \mathcal{P}(S)$. The following decomposition holds

$$(\Pi^n\, g)^i(y,p) - (\Pi^n\, g)^{i'}(y',p') = (\Pi^n\, g)^i(y,p) - (\Pi^n\, g)^i(y',p') + (\Pi^n\, g)^i(y',p') - (\Pi^n\, g)^{i'}(y',p')\;,$$

and we will estimate separately the two terms in the right–hand side.

□ To estimate the first term, we use the exponential forgetting of the prediction filter. Using Proposition 3.8 yields

$$|\,(\Pi^n\, g)^i(y,p) - (\Pi^n\, g)^i(y',p')\,| \leq$$

$$\leq \sum_{i_1,\cdots,i_n \in S} \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} |\,g^{i_n}(y_n, f[y_{n-1}, \cdots , y_1, y, p]) - g^{i_n}(y_n, f[y_{n-1}, \cdots , y_1, y', p'])\,|$$

$$q_\bullet^{i,i_1}\cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_1)\cdots\lambda(dy_n)$$

$$\leq \sum_{i_1,\cdots,i_n \in S} \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} |\,g^{i_n}(y_n, f[y_{n-1}, \cdots , y_1, f[y,p]\,]) - g^{i_n}(y_n, f[y_{n-1}, \cdots , y_1, f[y',p']\,])\,|$$

$$q_\bullet^{i,i_1}\cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_1)\cdots\lambda(dy_n)$$

$$\leq C'\,\varepsilon^{-r}\,\mathrm{Lip}(g)\,\rho_*^{n-1-2r}\;,$$

where the constant C' depends only on r, and Δ.

□ To estimate the second term, we use the geometric convergence of the *true* transition probabilities of the chain $\{X_n\;,\; n \geq 0\}$, which is a consequence of Assumption A, and we use again the exponential forgetting of the prediction filter. Recall that

$$(\Pi^n\, g)^i(y',p') - (\Pi^n\, g)^{i'}(y',p') =$$

$$= \sum_{i_1,\cdots,i_n\in S} \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} g^{i_n}(y_n, f[y_{n-1},\cdots,y_1,p_1])\; [q_\bullet^{i,i_1} - q_\bullet^{i',i_1}]$$

$$q_\bullet^{i_1,i_2}\cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_1)\cdots\lambda(dy_n)\ ,$$

where the notation $p_1 = f[y',p']$ is used.
Notice that, for any integer m such that $m \leq n-1$, and for any sequence $z_1,\cdots,z_m \in \mathbf{R}^d$

$$\begin{aligned} g^{i_n}(y_n, f[y_{n-1},\cdots,y_1,p_1]) \quad = \quad & \sum_{k=1}^m [\, g^{i_n}(y_n, f[y_{n-1},\cdots,y_k,z_{k-1},\cdots,z_1,p_1]) \\ & \qquad - g^{i_n}(y_n, f[y_{n-1},\cdots,y_{k+1},z_k,\cdots,z_1,p_1])\,] \\ & + g^{i_n}(y_n, f[y_{n-1},\cdots,y_{m+1},z_m,\cdots,z_1,p_1])\ . \end{aligned}$$

Therefore

$$\begin{aligned} &(\Pi^n\, g)^i(y',p') - (\Pi^n\, g)^{i'}(y',p') = \\ &= \sum_{k=1}^m \sum_{i_1,\cdots,i_n\in S} \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} [\, g^{i_n}(y_n, f[y_{n-1},\cdots,y_k,z_{k-1},\cdots,z_1,p_1]) \\ &\qquad - g^{i_n}(y_n, f[y_{n-1},\cdots,y_{k+1},z_k,\cdots,z_1,p_1])\,]\; [q_\bullet^{i,i_1} - q_\bullet^{i',i_1}] \\ &\qquad q_\bullet^{i_1,i_2}\cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_1)\cdots\lambda(dy_n) \\ &\quad + \sum_{i_1,\cdots,i_n\in S} \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} g^{i_n}(y_n, f[y_{n-1},\cdots,y_{m+1},z_m,\cdots,z_1,p_1])\; [q_\bullet^{i,i_1} - q_\bullet^{i',i_1}] \\ &\qquad q_\bullet^{i_1,i_2}\cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_1)\cdots\lambda(dy_n) \\ &= \sum_{k=1}^m \sum_{i_k,\cdots,i_n\in S} \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} [\, g^{i_n}(y_n, f[y_{n-1},\cdots,y_k,z_{k-1},\cdots,z_1,p_1]) \\ &\qquad - g^{i_n}(y_n, f[y_{n-1},\cdots,y_{k+1},z_k,\cdots,z_1,p_1])\,]\; [\, q_\bullet^{i,i_k}(k) - q_\bullet^{i',i_k}(k)\,] \\ &\qquad q_\bullet^{i_k,i_{k+1}}\cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_k}(y_k)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_k)\cdots\lambda(dy_n) \\ &\quad + \sum_{i_{m+1},\cdots,i_n\in S} \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} g^{i_n}(y_n, f[y_{n-1},\cdots,y_{m+1},z_m,\cdots,z_1,p_1]) \\ &\qquad [\, q_\bullet^{i,i_{m+1}}(m+1) - q_\bullet^{i',i_{m+1}}(m+1)\,] \\ &\qquad q_\bullet^{i_{m+1},i_{m+2}}\cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_{m+1}}(y_{m+1})\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_{m+1})\cdots\lambda(dy_n)\ . \end{aligned}$$

This holds for any integer m such that $m \leq n-1$. Taking now $m = n-2r-1$, using Proposition 3.8 again, and using estimate (1) yields

$$|\,(\Pi^n\, g)^i(y',p') - (\Pi^n\, g)^{i'}(y',p')\,| \leq$$

$$\leq \sum_{k=1}^{m} \sum_{i_k, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} | \, g^{i_n}(y_n, f[y_{n-1}, \cdots, y_k, z_{k-1}, \cdots, z_1, p_1])$$

$$- g^{i_n}(y_n, f[y_{n-1}, \cdots, y_{k+1}, z_k, \cdots, z_1, p_1]) \, | \, | \, q_\bullet^{i,i_k}(k) - q_\bullet^{i',i_k}(k) \, |$$

$$q_\bullet^{i_k, i_{k+1}} \cdots q_\bullet^{i_{n-1}, i_n} \, b_\bullet^{i_k}(y_k) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_k) \cdots \lambda(dy_n)$$

$$+ \sum_{i_{m+1}, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} | \, g^{i_n}(y_n, f[y_{n-1}, \cdots, y_{m+1}, z_m, \cdots, z_1, p_1]) \, |$$

$$| \, q_\bullet^{i,i_{m+1}}(m+1) - q_\bullet^{i',i_{m+1}}(m+1) \, |$$

$$q_\bullet^{i_{m+1}, i_{m+2}} \cdots q_\bullet^{i_{n-1}, i_n} \, b_\bullet^{i_{m+1}}(y_{m+1}) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_{m+1}) \cdots \lambda(dy_n)$$

$$\leq \sum_{k=1}^{m} \sum_{i_k, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} | \, g^{i_n}(y_n, f[y_{n-1}, \cdots, y_{k+1}, f[y_k, z_{k-1}, \cdots, z_1, p_1]])$$

$$- g^{i_n}(y_n, f[y_{n-1}, \cdots, y_{k+1}, f[z_k, z_{k-1}, \cdots, z_1, p_1]]) \, |$$

$$| \, q_\bullet^{i,i_k}(k) - q_\bullet^{i',i_k}(k) \, | \, q_\bullet^{i_k, i_{k+1}} \cdots q_\bullet^{i_{n-1}, i_n} \, b_\bullet^{i_k}(y_k) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_k) \cdots \lambda(dy_n)$$

$$+ \sum_{i_{m+1}, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} | \, g^{i_n}(y_n, f[y_{n-1}, \cdots, y_{m+1}, z_m, \cdots, z_1, p_1]) \, |$$

$$| \, q_\bullet^{i,i_{m+1}}(m+1) - q_\bullet^{i',i_{m+1}}(m+1) \, |$$

$$q_\bullet^{i_{m+1}, i_{m+2}} \cdots q_\bullet^{i_{n-1}, i_n} \, b_\bullet^{i_{m+1}}(y_{m+1}) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_{m+1}) \cdots \lambda(dy_n)$$

$$\leq C' \, \varepsilon^{-r} \, \mathrm{Lip}(g) \, \sum_{k=1}^{m} \rho_*^{n-1-k-2r} \, [\, 2 \, K_\bullet \, \rho_\bullet^k \,] + \mathrm{K}(g) \, [\, 2 \, K_\bullet \, \rho_\bullet^{m+1} \,] \ ,$$

where the constant C' depends only on r, and Δ, hence (using $m = n - 2r - 1$)

$$| \, (\Pi^n \, g)^i(y', p') - (\Pi^n \, g)^{i'}(y', p') \, | \leq$$

$$\leq 2 \, C' \, K_\bullet \, \varepsilon^{-r} \, \mathrm{Lip}(g) \, (n - 2r - 1) \, \rho_{\max}^{n-2r-1} + 2 \, K_\bullet \, \mathrm{K}(g) \, \rho_{\max}^{n-2r} \leq C'' \, \varepsilon^{-r} \, [\, \mathrm{Lip}(g) + \mathrm{K}(g) \,] \, \rho^n \ ,$$

for any ρ such that $\rho_{\max} < \rho < 1$, where the constant C'' depends only on r, Δ, and $K_\bullet$.

Combining the above estimates finishes the proof. □

Remark 3.9 In DiMasi and Stettner [4], the existence of a unique invariant probability distribution is proved for another extended Markov chain based on the filter itself instead of the prediction filter, in the case of a misspecified HMM with conditionally Gaussian observations, with positive transition probability matrices (for both the true and the assumed models), and under the somewhat restrictive assumption that the mapping $h_\bullet = (h_\bullet^i)$ from S to $\mathbf{R}^d$ is *injective*.

4 Exponential forgetting for the gradient

We start with some general remarks, which are formulated in the time–dependent case. Recall that $\{p_n \, , \, n \geq 0\}$ and $\{w_n \, , \, n \geq 0\}$ satisfy equations (5) and (6) respectively, i.e.

$$p_{n+1} = \frac{Q_n^* B_n(Y_n) \; p_n}{b_n^*(Y_n) \; p_n} = f_n[Y_n, p_n] = f_{n,m}[Y_n, \cdots, Y_m, p_m] \; ,$$

and

$$\begin{aligned} w_{n+1} &= \frac{Q_n^* B_n(Y_n)}{b_n^*(Y_n) \, p_n} \, [\, I - \frac{p_n \, b_n^*(Y_n)}{b_n^*(Y_n) \, p_n} \,] \; w_n + u_n[Y_n, p_n] \\ &= \Phi_n[Y_n, p_n] \; w_n + u_n[Y_n, p_n] = F_n[Y_n, p_n, w_n] = F_{n,m}[Y_n, \cdots, Y_m, p_m, w_m] \; , \end{aligned}$$

for any integers n, m such that $n \geq m$, where $Q_n = (q_n^{i,j})$ is a stochastic matrix, and $b_n(\cdot) = (b_n^i(\cdot))$ are λ–a.e. positive densities.

The *variation of parameters* formula yields

$$w_{n+1} = \prod_{k=m}^{n} \Phi_k[Y_k, p_k] \; w_m + \sum_{l=m}^{n-1} \prod_{k=l+1}^{n} \Phi_k[Y_k, p_k] \; u_l[Y_l, p_l] + u_n[Y_n, p_n] \; .$$

Throughout the end of this section, we use the following notation : For any $p, p' \in \mathcal{P}(S)$, any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$, and any integers m, l, n such that $m \leq l \leq n$, let

$$p_{l+1} = f_{l,m}[y_l, \cdots, y_m, p] \qquad \text{and} \qquad p'_{l+1} = f_{l,m}[y_l, \cdots, y_m, p'] \; ,$$

and recall that

$$M_{n,l} = Q_n^* B(y_n) \cdots Q_l^* B(y_l) \; .$$

For any $p, p' \in \mathcal{P}(S)$, any $w, w' \in \Sigma$, and any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$, the following decomposition holds

$$\begin{aligned} & F_{n,m}[y_n, \cdots, y_m, p, w] - F_{n,m}[y_n, \cdots, y_m, p', w'] = \\ = \; & \prod_{k=m}^{n} \Phi_k[y_k, p_k] \; w + \sum_{l=m}^{n-1} \prod_{k=l+1}^{n} \Phi_k[y_k, p_k] \; u_l[y_l, p_l] + u_n[y_n, p_n] \\ & - \prod_{k=m}^{n} \Phi_k[y_k, p'_k] \; w' - \sum_{l=m}^{n-1} \prod_{k=l+1}^{n} \Phi_k[y_k, p'_k] \; u_l[y_l, p'_l] - u_n[y_n, p'_n] \\ = \; & \prod_{k=m}^{n} \Phi_k[y_k, p_k] \; (w - w') + \prod_{k=m}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \; w' \\ & + \sum_{l=m}^{n-1} \prod_{k=l+1}^{n} \Phi_k[y_k, p_k] \; [\, u_l[y_l, p_l] - u_l[y_l, p'_l] \,] + [\, u_n[y_n, p_n] - u_n[y_n, p'_n] \,] \\ & + \sum_{l=m}^{n-1} \prod_{k=l+1}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \; u_l[y_l, p'_l] \; . \end{aligned} \tag{9}$$

As noticed in Arapostathis and Marcus [1, pp.24–25], for any integers l, n such that $n \geq l$, and any $z \in \mathbf{R}^N$

$$\begin{aligned} \prod_{k=l}^{n} \Phi_k[y_k, p_k] \; z &= \Phi_n[y_n, p_n] \; \frac{M_{n-1,l} \; z}{e^* \; M_{n-1,l} \; p_l} \\ &= \frac{e^* \; M_{n,l} \; z}{e^* \; M_{n,l} \; p_l} \; [\, f_{n,l}[y_n, \cdots, y_l, z] - f_{n,l}[y_n, \cdots, y_l, p_l] \,] \; . \end{aligned}$$

The proof is based on the identity

$$[\, I - \frac{p\, b_n^*(y)}{b_n^*(y)\, p} \,]\; p = 0 \; ,$$

which holds for any $y \in \mathbf{R}^d$, and any $p \in \mathcal{P}(S)$. Notice that any $z \in \mathbf{R}^N$ can be written as $z = z^+ - z^-$, where z^+ and z^- are positive, and $||z|| = ||z^+|| + ||z^-||$, hence

$$\begin{aligned}\prod_{k=l}^{n} \Phi_k[y_k, p_k]\; z \;&=\; \frac{e^*\, M_{n,l}\, z^+}{e^*\, M_{n,l}\, p_l}\, [\, f_{n,l}[y_n, \cdots, y_l, \frac{z^+}{||z^+||}] - f_{n,l}[y_n, \cdots, y_l, p_l]\,] \\ &\quad - \frac{e^*\, M_{n,l}\, z^-}{e^*\, M_{n,l}\, p_l}\, [\, f_{n,l}[y_n, \cdots, y_l, \frac{z^-}{||z^-||}] - f_{n,l}[y_n, \cdots, y_l, p_l]\,] \; .\end{aligned} \tag{10}$$

From this identity, we also have

$$\begin{aligned}&\prod_{k=l}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k]\,]\; z = \\ &\quad = \frac{e^*\, M_{n,l}\, z}{e^*\, M_{n,l}\, p_l}\, [\, f_{n,l}[y_n, \cdots, y_l, z] - f_{n,l}[y_n, \cdots, y_l, p_l]\,] \\ &\qquad - \frac{e^*\, M_{n,l}\, z}{e^*\, M_{n,l}\, p'_l}\, [\, f_{n,l}[y_n, \cdots, y_l, z] - f_{n,l}[y_n, \cdots, y_l, p'_l]\,] \\ &\quad = \frac{e^*\, M_{n,l}\, z}{e^*\, M_{n,l}\, p_l}\, [\, f_{n,l}[y_n, \cdots, y_l, p'_l] - f_{n,l}[y_n, \cdots, y_l, p_l]\,] \\ &\qquad + \frac{e^*\, M_{n,l}\, (p'_l - p_l)}{e^*\, M_{n,l}\, p_l}\; \frac{e^*\, M_{n,l}\, z}{e^*\, M_{n,l}\, p'_l}\, [\, f_{n,l}[y_n, \cdots, y_l, z] - f_{n,l}[y_n, \cdots, y_l, p'_l]\,] \; ,\end{aligned}$$

hence

$$\begin{aligned}&\prod_{k=l}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k]\,]\; z = \\ &\quad = \frac{e^*\, M_{n,l}\, z}{e^*\, M_{n,l}\, p_l}\, (p'_{n+1} - p_{n+1}) + \frac{e^*\, M_{n,l}\, (p'_l - p_l)}{e^*\, M_{n,l}\, p_l} \prod_{k=l}^{n} \Phi_k[y_k, p'_k]\; z \; .\end{aligned} \tag{11}$$

After these preliminary remarks, we consider the situation in more details, in various different cases.

Time–dependent case

Assume that for any integers n, m such that $n \geq m$, the stochastic matrix $Q_n = (q_n^{i,j})$ is positive, and define for any $y \in \mathbf{R}^d$

$$\delta_n(y) = \frac{\max_{i \in S} b_n^i(y)}{\min_{i \in S} b_n^i(y)} < \infty \qquad \text{and} \qquad \varepsilon_n = \min_{i,j \in S} q_n^{i,j} > 0 \; .$$

Assumption C : For any integer $n \geq 0$ and any $y \in \mathbf{R}^d$, the partial mapping $p \mapsto u_n[y, p]$ is Lipschitz continuous (hence bounded since $\mathcal{P}(S)$ is compact), i.e.

$$||u_n[y, p] - u_n[y, p']\,|| \leq \mathrm{Lip}(u_n, y)\; ||p - p'|| \; ,$$

for any $p, p' \in \mathcal{P}(S)$, and by definition

$$\mathrm{K}(u_n, y) = \sup_{p \in \mathcal{P}(S)} ||u_n[y, p]\,|| \; .$$

Proposition 4.1 *If Assumption C holds, then for any $p, p' \in \mathcal{P}(S)$, any $w, w' \in \Sigma$, any integers n, m such that $n \geq m$, and any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$*

$$\| F_{n,m}[y_n, \cdots, y_m, p, w] - F_{n,m}[y_n, \cdots, y_m, p', w'] \| \leq$$

$$\leq 6\, [\varepsilon_m^{-1}\, \delta_m(y_m)]^3 \prod_{k=m}^{n} (1 - \varepsilon_k)\, [\, \|w - w'\| + \|p - p'\|\, (1 + \|w'\| + \|w\|\,)\,]$$

$$+ 10\, \varepsilon_m^{-1}\, \delta_m(y_m)\ \Big[\sum_{l=m}^{n-1} (1 - \varepsilon_l)^{-1}\, [\varepsilon_{l+1}^{-1}\, \delta_{l+1}(y_{l+1})\,]^2\, \mathrm{Lip}(u_l, y_l) + (1 - \varepsilon_n)^{-1}\, \mathrm{Lip}(u_n, y_n)$$

$$+ \sum_{l=m}^{n-1} [\varepsilon_{l+1}^{-1}\, \delta_{l+1}(y_{l+1})\,]^3\, \mathrm{K}(u_l, y_l)\ \Big] \prod_{k=m}^{n} (1 - \varepsilon_k)\, \|p - p'\|\ .$$

The proof of Proposition 4.1 is given in Appendix D.

If the sequence $\{\varepsilon_n\ ,\ n \geq 0\}$ is nonincreasing, if $\limsup_{n\to\infty} \frac{1}{n} \log \varepsilon_n^{-1} < 0$, and if $\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} \log(1 - \varepsilon_k) < 0$, then for any $p, p' \in \mathcal{P}(S)$, and any $w, w' \in \Sigma$, such that $(p, w) \neq (p', w')$, and for any infinite sequence $y_m, \cdots, y_n, \cdots \in \mathbf{R}^d$, such that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{l=m}^{n} \delta_{l+1}^2(y_{l+1})\, \mathrm{Lip}(u_l, y_l) < \infty \qquad \text{and} \qquad \limsup_{n\to\infty} \frac{1}{n} \sum_{l=m}^{n} \delta_{l+1}^3(y_{l+1})\, \mathrm{K}(u_l, y_l) < \infty\ ,$$

the difference $\| F_{n,m}[y_n, \cdots, y_m, p, w] - F_{n,m}[y_n, \cdots, y_m, p', w'] \|$ goes to zero at exponential rate

$$\limsup_{n\to\infty} \frac{1}{n}\ \log \| F_{n,m}[y_n, \cdots, y_m, p, w] - F_{n,m}[y_n, \cdots, y_m, p', w'] \| \leq \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} \log(1 - \varepsilon_k) < 0\ .$$

Time–independent case

In the time–independent case $\{p_n\ ,\ n \geq 0\}$ and $\{w_n\ ,\ n \geq 0\}$ satisfy equations (3) and (4) respectively, i.e.

$$p_{n+1} = \frac{Q^*\, B(Y_n)\; p_n}{b^*(Y_n)\; p_n} = f[Y_n, p_n] = f[Y_n, \cdots, Y_m, p_m]\ ,$$

and

$$\begin{aligned} w_{n+1} &= \frac{Q^*\, B(Y_n)}{b^*(Y_n)\, p_n}\, [\, I - \frac{p_n\, b^*(Y_n)}{b^*(Y_n)\, p_n}\,]\, w_n + u[Y_n, p_n] \\ &= \Phi[Y_n, p_n]\, w_n + u[Y_n, p_n] = F[Y_n, p_n, w_n] = F[Y_n, \cdots, Y_m, p_m, w_m]\ , \end{aligned}$$

for any integers n, m such that $n \geq m$. For the stochastic matrix $Q = (q^{i,j})$, and for any $y \in \mathbf{R}^d$, we define

$$\delta(y) = \frac{\max_{i \in S} b^i(y)}{\min_{i \in S} b^i(y)} < \infty \qquad \text{and} \qquad \varepsilon = \min_{i,j \in S}{}^{+}\, q^{i,j} > 0\ ,$$

where the notation $\min^+$ denotes the minimum over positive elements.

Assumption C (revisited) : For any $y \in \mathbf{R}^d$, the partial mapping $p \mapsto u[y, p]$ is Lipschitz continuous (hence bounded since $\mathcal{P}(S)$ is compact), i.e.

$$\| u[y, p] - u[y, p'] \| \leq \mathrm{Lip}(u, y)\, \|p - p'\|\ ,$$

for any $p, p' \in \mathcal{P}(S)$, and by definition

$$\mathrm{K}(u, y) = \sup_{p \in \mathcal{P}(S)} \| u[y, p] \|\ .$$

Definition 4.2 *Under Assumption C, let*

$$\begin{aligned}
\mathrm{Lip}(u) &= \max_{i\in S} \int_{\mathbf{R}^d} \mathrm{Lip}(u,y)\, b^i_\bullet(dy)\, \lambda(dy) \ , \\
\mathrm{K}(u) &= \max_{i\in S} \int_{\mathbf{R}^d} \mathrm{K}(u,y)\, b^i_\bullet(dy)\, \lambda(dy) \ , \\
[\Delta\cdot\mathrm{K}](u) &= \max_{i\in S} \int_{\mathbf{R}^d} \delta(y)\, \mathrm{K}(u,y)\, b^i_\bullet(dy)\lambda(dy) \ .
\end{aligned}$$

Notice that $\mathrm{K}(u) \leq [\Delta\cdot\mathrm{K}](u)$, since $\delta(y)\geq 1$ for any $y\in\mathbf{R}^d$.

Example 4.3 If Δ is finite, then the function u defined by

$$u[y,p] = \frac{\partial Q^*\, B(y)\, p}{b^*(y)\, p} \ ,$$

for any $y\in\mathbf{R}^d$, and any $p\in\mathcal{P}(S)$, satisfies Assumption C, and moreover $\mathrm{Lip}(u)$ and $[\Delta\cdot\mathrm{K}](u)$ are finite. Indeed, for any $y\in\mathbf{R}^d$, and any $p\in\mathcal{P}(S)$

$$\| \frac{\partial Q^*\, B(y)\, p}{b^*(y)\, p} \| \leq \max_{i\in S} \sum_{j\in S} |\partial q^{i,j}| \ ,$$

whereas for any $y\in\mathbf{R}^d$, and any $p,p'\in\mathcal{P}(S)$

$$\frac{\partial Q^*\, B(y)\, p}{b^*(y)\, p} - \frac{\partial Q^*\, B(y)\, p'}{b^*(y)\, p'} = \frac{\partial Q^*\, B(y)\, (p-p')}{b^*(y)\, p} - \frac{b^*(y)\, (p-p')}{b^*(y)\, p}\, \frac{\partial Q^*\, B(y)\, p'}{b^*(y)\, p'} \ ,$$

hence

$$\| \frac{\partial Q^*\, B(y)\, p}{b^*(y)\, p} - \frac{\partial Q^*\, B(y)\, p'}{b^*(y)\, p'} \| \leq 2\, [\max_{i\in S} \sum_{j\in S} |\partial q^{i,j}|\,]\, \delta(y)\, \|p-p'\| \ .$$

By particularizing to the time–independent case the result of the more general Proposition 4.1 above, we obtain the following result.

Proposition 4.4 *If the stochastic matrix Q is positive, and if Assumption C holds, then for any $p,p'\in\mathcal{P}(S)$, any $w,w'\in\Sigma$, any integers n,m such that $n\geq m$, and any sequence $y_m,\cdots,y_n\in\mathbf{R}^d$*

$$\begin{aligned}
&\|F[y_n,\cdots,y_m,p,w] - F[y_n,\cdots,y_m,p',w']\,\| \leq \\
&\qquad \leq 6\, [\varepsilon^{-1}\, \delta(y_m)\,]^3\, (1-\varepsilon)^{n-m+1}\, [\, \|w-w'\| + \|p-p'\|\, (1+\|w\|+\|w'\|\,)\,] \\
&\qquad\quad + 10\, \varepsilon^{-1}\, \delta(y_m)\ \Big[\ (1-\varepsilon)^{-1} \sum_{l=m}^{n-1} [\varepsilon^{-1}\, \delta(y_{l+1})\,]^2\, \mathrm{Lip}(u,y_l) + (1-\varepsilon)^{-1}\, \mathrm{Lip}(u,y_n) \\
&\qquad\qquad + \sum_{l=m}^{n-1} [\varepsilon^{-1}\, \delta(y_{l+1})\,]^3\, \mathrm{K}(u,y_l)\ \Big]\ (1-\varepsilon)^{n-m+1}\, \|p-p'\| \ .
\end{aligned}$$

For any $p,p'\in\mathcal{P}(S)$, and any $w,w'\in\Sigma$, such that $(p,w)\neq(p',w')$, and for any infinite sequence $y_m,\cdots,y_n,\cdots\in\mathbf{R}^d$, such that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{l=m}^{n} \delta^2(y_{l+1})\, \mathrm{Lip}(u,y_l) < \infty \qquad \text{and} \qquad \limsup_{n\to\infty} \frac{1}{n} \sum_{l=m}^{n} \delta^3(y_{l+1})\, \mathrm{K}(u,y_l) < \infty \ ,$$

the difference $\|F[y_n,\cdots,y_m,p,w] - F[y_n,\cdots,y_m,p',w']\,\|$ goes to zero at exponential rate

$$\limsup_{n\to\infty} \frac{1}{n}\, \log\, \|F[y_n,\cdots,y_m,p,w] - F[y_n,\cdots,y_m,p',w']\,\| \leq \log(1-\varepsilon) \ .$$

Proposition 4.5 *If the stochastic matrix Q is positive, if Assumptions A and C hold, and if Δ_3,* $\mathrm{Lip}(u)$ *and* $\mathrm{K}(u)$ *are finite, then for any $p, p' \in \mathcal{P}(S)$, and any $w, w' \in \Sigma$, such that $(p, w) \neq (p', w')$*

$$\limsup_{n\to\infty} \frac{1}{n} \, \log \| F[Y_n, \cdots, Y_m, p, w] - F[Y_n, \cdots, Y_m, p', w'] \| \leq \log(1-\varepsilon) \qquad \mathbf{P}_\bullet\text{–}a.s.$$

Proof. If Assumption A holds, then the Markov chain $\{(X_n, Y_n)\,,\, n \geq 0\}$ has, under the true probability measure $\mathbf{P}_\bullet$, a unique invariant probability distribution $\nu_\bullet = (\nu_\bullet^i)$ on $S \times \mathbf{R}^d$ and for any $i \in S$

$$\nu_\bullet^i(dy) = \mu_\bullet^i \, b_\bullet^i(y) \, \lambda(dy) \ .$$

If Δ_3, $\mathrm{Lip}(u)$ and $\mathrm{K}(u)$ are finite, then $\mathbf{P}_\bullet$–a.s.

$$\lim_{n\to\infty} \frac{1}{n} \, \sum_{l=m}^{n} \delta^2(Y_{l+1}) \, \mathrm{Lip}(u, Y_l) =$$

$$= \sum_{i,j\in S} \mu_\bullet^i \, q_\bullet^{i,j} \, \int_{\mathbf{R}^d} \delta^2(y') \, b_\bullet^j(y') \, \lambda(dy') \, \int_{\mathbf{R}^d} \mathrm{Lip}(u,y) \, b_\bullet^i(y) \, \lambda(dy) \leq \Delta_2 \, \mathrm{Lip}(u) < \infty \ ,$$

and similarly

$$\lim_{n\to\infty} \frac{1}{n} \, \sum_{l=m}^{n} \delta^3(Y_{l+1}) \, \mathrm{K}(u, Y_l) =$$

$$= \sum_{i,j\in S} \mu_\bullet^i \, q_\bullet^{i,j} \, \int_{\mathbf{R}^d} \delta^3(y') \, b_\bullet^j(y') \, \lambda(dy') \, \int_{\mathbf{R}^d} \mathrm{K}(u,y) \, b_\bullet^i(y) \, \lambda(dy) \leq \Delta_3 \, \mathrm{K}(u) < \infty \ . \qquad \square$$

Theorem 4.6 *If the stochastic matrix Q is primitive, with index of primitivity r, and if Assumption C holds, then for any $p, p' \in \mathcal{P}(S)$, any $w, w' \in \Sigma$, any integers n, m such that $n \geq m + r - 1$, and any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$*

$$\| F[y_n, \cdots, y_m, p, w] - F[y_n, \cdots, y_m, p', w'] \| \leq$$

$$\leq 6 \, [\, \varepsilon^{-r} \, \delta(y_m) \cdots \delta(y_{m+r-1}) \,]^3 \, \prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r \, [\, \delta(y_{m+\kappa r+1}) \cdots \delta(y_{m+(\kappa+1)r-1}) \,]^{-1} \,)$$

$$[\, \|w - w'\| + \|p - p'\| \, (1 + \|w\| + \|w'\| \,) \,]$$

$$+ \, 10 \, \varepsilon^{-r} \, \delta(y_m) \cdots \delta(y_{m+r-1}) \, (1 - \varepsilon^r)^{-1}$$

$$[\, \sum_{l=m}^{n-1} [\, \varepsilon^{-r} \, \delta(y_{l+1}) \cdots \delta(y_{\min(l+r,n)}) \,]^2 \, \mathrm{Lip}(u, y_l) + \mathrm{Lip}(u, y_n)$$

$$+ \sum_{l=m}^{n-1} [\, \varepsilon^{-r} \, \delta(y_{l+1}) \cdots \delta(y_{\min(l+r,n)}) \,]^3 \, \mathrm{K}(u, y_l) \,]$$

$$\prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r \, [\, \delta(y_{m+\kappa r+1}) \cdots \delta(y_{m+(\kappa+1)r-1}) \,]^{-1} \,) \, \|p - p'\| \ .$$

The proof of Theorem 4.6 is given in Appendix D.

For any $p, p' \in \mathcal{P}(S)$, and any $w, w' \in \Sigma$, such that $(p, w) \neq (p', w')$, and for any infinite sequence $y_m, \cdots, y_n, \cdots \in \mathbf{R}^d$, such that

$$\limsup_{n\to\infty} \frac{1}{n} \, \sum_{l=m}^{n} [\, \delta(y_{l+1}) \cdots \delta(y_{l+r}) \,]^2 \, \mathrm{Lip}(u, y_l) < \infty \ ,$$

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{l=m}^{n} [\, \delta(y_{l+1}) \cdots \delta(y_{l+r})\,]^3 \, \mathrm{K}(u, y_l) < \infty \ ,$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{\kappa=0}^{[n,m]} \log\,(1 - \varepsilon^r \, [\, \delta(y_{m+\kappa r+1}) \cdots \delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,) < 0 \ ,$$

the difference $\|F[y_n, \cdots, y_m, p, w] - F[y_n, \cdots, y_m, p', w']\,\|$ goes to zero at exponential rate

$$\limsup_{n\to\infty} \frac{1}{n} \, \log \|F[y_n, \cdots, y_m, p, w] - F[y_n, \cdots, y_m, p', w']\,\| \leq$$

$$\leq \limsup_{n\to\infty} \frac{1}{n} \sum_{\kappa=0}^{[n,m]} \log\,(1 - \varepsilon^r \, [\, \delta(y_{m+\kappa r+1}) \cdots \delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,) < 0 \ .$$

Proposition 4.7 *If the stochastic matrix Q is primitive, with index of primitivity r, if Assumptions A and C hold, and if Δ_3, $\mathrm{Lip}(u)$ and $\mathrm{K}(u)$ are finite, then for any $p, p' \in \mathcal{P}(S)$, and any $w, w' \in \Sigma$, such that $(p, w) \neq (p', w')$*

$$\limsup_{n\to\infty} \frac{1}{n} \, \log \|F[Y_n, \cdots, Y_m, p, w] - F[Y_n, \cdots, Y_m, p', w']\,\| \leq \frac{1}{r} \, \log(1 - R) \ , \qquad \mathbf{P}_\bullet\text{-}a.s.$$

where $R = \varepsilon^r \, \Delta_{-1}^{r-1}$.

Proof. If Assumption A holds, then the Markov chain $\{(X_n, Y_n)\,,\, n \geq 0\}$ has, under the true probability measure $\mathbf{P}_\bullet$, a unique invariant probability distribution $\nu_\bullet = (\nu_\bullet^i)$ on $S \times \mathbf{R}^d$ and for any $i \in S$

$$\nu_\bullet^i(dy) = \mu_\bullet^i \, b_\bullet^i(y) \, \lambda(dy) \ .$$

It as already been proved in Proposition 2.5, that $\mathbf{P}_\bullet$–a.s.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\kappa=0}^{[n,m]} \log\,(1 - \varepsilon^r \, [\, \delta(Y_{m+\kappa r+1}) \cdots \delta(Y_{m+(\kappa+1)r-1})\,]^{-1}\,) = \frac{1}{r} \, \log(1 - R) \ .$$

If Δ_3, $\mathrm{Lip}(u)$ and $\mathrm{K}(u)$ are finite, then $\mathbf{P}_\bullet$–a.s.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=m}^{n} [\, \delta(Y_{l+1}) \cdots \delta(Y_{l+r})\,]^2 \, \mathrm{Lip}(u, Y_l) =$$

$$= \sum_{i_0, \cdots, i_r \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} [\, \delta(y_1) \cdots \delta(y_r)\,]^2 \, \mathrm{Lip}(u, y_0)$$

$$\mu_\bullet^{i_0} \, q_\bullet^{i_0, i_1} \cdots q_\bullet^{i_{r-1}, i_r} \, b_\bullet^{i_0}(y_0) \cdots b_\bullet^{i_r}(y_r) \, \lambda(dy_0) \cdots \lambda(dy_r)$$

$$\leq \Delta_2^r \, \mathrm{Lip}(u) < \infty \ ,$$

and similarly

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=m}^{n} [\, \delta(Y_{l+1}) \cdots \delta(Y_{l+r})\,]^3 \, \mathrm{K}(u, Y_l) =$$

$$= \sum_{i_0, \cdots, i_r \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} [\, \delta(y_1) \cdots \delta(y_r)\,]^3 \, \mathrm{K}(u, y_0)$$

$$\mu_\bullet^{i_0} \, q_\bullet^{i_0, i_1} \cdots q_\bullet^{i_{r-1}, i_r} \, b_\bullet^{i_0}(y_0) \cdots b_\bullet^{i_r}(y_r) \, \lambda(dy_0) \cdots \lambda(dy_r)$$

$$\leq \Delta_3^r \, \mathrm{K}(u) < \infty \ . \qquad \Box$$

5 Geometric ergodicity of the Markov chain $\{X_n, Y_n, p_n, w_n\}$

Under the probability measure $\mathbf{P}_\bullet$ corresponding to the *true* transition probability matrix $Q_\bullet$ and the *true* observation densities $b_\bullet(\cdot)$, the extended Markov chain $\{Z'_n = (X_n, Y_n, p_n, w_n)\,,\, n \geq 0\}$ has the following transition probability matrix / kernel

$$\begin{aligned}
\Pi^{i,j}(y, p, w, dy', dp', dw') &= \\
&= \mathbf{P}_\bullet[X_{n+1} = j, Y_{n+1} \in dy', p_{n+1} \in dp', w_{n+1} \in dw' \mid X_n = i, Y_n = y, p_n = p, w_n = w] \\
&= q_\bullet^{i,j}\, b_\bullet^j(y')\, \lambda(dy')\, \delta_{f[y,\,p]}(dp')\, \delta_{F[y,\,p,\,w]}(dw')\ .
\end{aligned}$$

For any real–valued function g defined on $S \times \mathbf{R}^d \times \mathcal{P}(S) \times \Sigma$, which is equivalently defined as a collection $g = (g^i)$ of real–valued functions defined on $\mathbf{R}^d \times \mathcal{P}(S) \times \Sigma$, we have

$$\begin{aligned}
(\Pi\, g)^i(y, p, w) &= \mathbf{E}_\bullet[g(X_{n+1}, Y_{n+1}, p_{n+1}, w_{n+1}) \mid X_n = i, Y_n = y, p_n = p, w_n = w] \\
&= \sum_{j \in S} \mathbf{E}_\bullet[g^j(Y_{n+1}, p_{n+1}, w_{n+1})\, \mathbf{1}_{[X_{n+1} = j]} \mid X_n = i, Y_n = y, p_n = p, w_n = w] \\
&= \sum_{j \in S} \int_{\mathbf{R}^d} g^j(y', f[y,p], F[y,p,w])\, q_\bullet^{i,j}\, b_\bullet^j(y')\, \lambda(dy') \\
&= \sum_{j \in S} \int_{\mathbf{R}^d \times \mathcal{P}(S) \times \Sigma} g^j(y', p', w')\, \Pi^{i,j}(y, p, w, dy', dp', dw')\ ,
\end{aligned}$$

for any $i \in S$, any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$. More generally, for any integer $n \geq 2$ we have

$$\begin{aligned}
(\Pi^n\, g)^i(y, p, w) &= \mathbf{E}_\bullet[g(X_n, Y_n, p_n, w_n) \mid X_0 = i, Y_0 = y, p_0 = p, w_0 = w] \\
&= \sum_{i_n \in S} \mathbf{E}_\bullet[g^{i_n}(Y_n, p_n, w_n)\, \mathbf{1}_{[X_n = i_n]} \mid X_0 = i, Y_0 = y, p_0 = p, w_0 = w] \\
&= \sum_{i_1, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} g^{i_n}(y_n, f[y_{n-1}, \cdots, y_1, y, p], F[y_{n-1}, \cdots, y_1, y, p, w]) \\
&\qquad q_\bullet^{i,i_1} \cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_1) \cdots \lambda(dy_n)\ ,
\end{aligned}$$

for any $i \in S$, any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$.

For any $p \in \mathcal{P}(S)$, any $w \in \Sigma$, and any $y \in \mathbf{R}^d$, let

$$f \otimes F[y, p, w] = (f[y, p], F[y, p, w])\ .$$

More generally, for any $p \in \mathcal{P}(S)$, any $w \in \Sigma$, and any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$, let

$$f \otimes F[y_n, \cdots, y_m, p, w] = (f[y_n, \cdots, y_m, p], F[y_n, \cdots, y_m, p, w])\ .$$

With this definition, we can write the transition probability matrix / kernel equivalently as

$$\Pi^{i,j}(y, p, w, dy', dp', dw') = q_\bullet^{i,j}\, b_\bullet^j(y')\, \lambda(dy')\, \delta_{f \otimes F[y,\,p,\,w]}(dp', dw')\ ,$$

and

$$\begin{aligned}
(\Pi^n\, g)^i(y, p, w) &= \sum_{i_1, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_1, y, p, w]) \\
&\qquad q_\bullet^{i,i_1} \cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_1) \cdots \lambda(dy_n)\ ,
\end{aligned}$$

for any $i \in S$, any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$.

Definition 5.1 *Let L' denote the set of functions $g = (g^i)$ defined on $S \times \mathbf{R}^d \times \mathcal{P}(S) \times \Sigma$, such that for any $i \in S$, and any $y \in \mathbf{R}^d$ the partial mapping $(p, w) \mapsto g^i(y, p, w)$ is locally Lipschitz continuous, in the sense that*

$$|g^i(y,p,w) - g^i(y,p',w')| \leq \mathrm{Lip}(g^i, y)\; [\,\|w - w'\| + \|p - p'\|\; (1 + \|w\| + \|w'\|\,)\,]$$

for any $p, p' \in \mathcal{P}(S)$, and any $w, w' \in \Sigma$, that

$$|g^i(y,p,w)| \leq \mathrm{K}(g^i, y)\; (1 + \|w\|\,)\ ,$$

for any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$, and such that

$$\mathrm{Lip}(g) = \max_{i \in S} \int_{\mathbf{R}^d} \mathrm{Lip}(g^i, y)\; b^i_\bullet(y)\, \lambda(dy) < \infty\ ,$$

$$\mathrm{K}(g) = \max_{i \in S} \int_{\mathbf{R}^d} \mathrm{K}(g^i, y)\; b^i_\bullet(y)\, \lambda(dy) < \infty\ .$$

Example 5.2 [score function] If Δ_2 is finite, then the function g defined by

$$g(y,p,w) = \frac{b^*(y)\, w}{b^*(y)\, p}\ ,$$

for any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$, i.e. constant over S, belongs to the set L'. Indeed, for any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$

$$|\frac{b^*(y)\, w}{b^*(y)\, p}| \leq \delta(y)\; \|w\|\ ,$$

whereas for any $y \in \mathbf{R}^d$, any $p, p' \in \mathcal{P}(S)$, and any $w, w' \in \Sigma$

$$\frac{b^*(y)\, w}{b^*(y)\, p} - \frac{b^*(y)\, w'}{b^*(y)\, p'} = \frac{b^*(y)\, (w - w')}{b^*(y)\, p} - \frac{b^*(y)\, (p - p')}{b^*(y)\, p}\; \frac{b^*(y)\, w'}{b^*(y)\, p'}$$

hence

$$\begin{aligned} |\frac{b^*(y)\, w}{b^*(y)\, p} - \frac{b^*(y)\, w'}{b^*(y)\, p'}| \;&\leq\; \delta(y)\; \|w - w'\| + \delta^2(y)\; \|p - p'\|\; \|w'\| \\ &\leq\; \delta^2(y)\; [\,\|w - w'\| + \|p - p'\|\, (1 + \|w\| + \|w'\|\,)\,]\ . \end{aligned}$$

Example 5.3 As in Example 3.5 above, consider the $N \times d$ matrix $\phi = (\phi^i)$, where for any $i \in S$, the mean vector $\phi^i \in \mathbf{R}^d$ is defined by

$$\phi^i = \int_{\mathbf{R}^d} y\; b^i(y)\, \lambda(dy)\ ,$$

assuming the integral exist. If

$$\max_{i \in S} \int_{\mathbf{R}^d} |y|\; b^i_\bullet(y)\, \lambda(dy) < \infty\ ,$$

then the function g defined by

$$g(y,p,w) = [\,\phi\, (y - \phi^*\, p)\,]^*\; w\ ,$$

for any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$, i.e. constant over S, belongs to the set L'. Indeed, for any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$

$$|\,[\,\phi\, (y - \phi^*\, p)\,]^*\; w\,| \leq \max_{i \in S} |\phi^i|\; [\,|y| + \max_{i \in S} |\phi^i|\,]\; \|w\|\ ,$$

whereas for any $y \in \mathbf{R}^d$, any $p, p' \in \mathcal{P}(S)$, and any $w, w' \in \Sigma$

$$[\,\phi\, (y - \phi^*\, p)\,]^*\; w - [\,\phi\, (y - \phi^*\, p')\,]^*\; w' = [\,\phi\, (y - \phi^*\, p)\,]^*\; (w - w') - [\,\phi\, \phi^*\, (p - p')\,]^*\; w'\ ,$$

hence

$$\begin{aligned} &|\,[\,\phi\, (y - \phi^*\, p)\,]^*\; w - [\,\phi\, (y - \phi^*\, p')\,]^*\; w'\,| \leq \\ &\leq \max_{i \in S} |\phi^i|\; [\,|y| + \max_{i \in S} |\phi^i|\,]\; \|w - w'\| + \max_{i \in S} |\phi^i|^2\; \|p - p'\|\; \|w'\| \\ &\leq \max_{i \in S} |\phi^i|\; [\,|y| + \max_{i \in S} |\phi^i|\,]\; [\,\|w - w'\| + \|p - p'\|\, (1 + \|w\| + \|w'\|\,)\,]\ . \end{aligned}$$

Theorem 5.4 *If the stochastic matrix Q is primitive, with index of primitivity r, if Assumptions A, and C hold, and if Δ_4, $\mathrm{Lip}(u)$, and $[\Delta \cdot \mathrm{K}](u)$ are finite, then there exist constants $0 < \rho < 1$ and $C > 0$ such that, for any $z, z' \in S \times \mathbf{R}^d \times \mathcal{P}(S) \times \Sigma$, and for any function $g = (g^i)$ in L'*

$$|(\Pi^n\, g)(z) - (\Pi^n\, g)(z')| \leq C\, \varepsilon^{-6r}\, [\,\mathrm{Lip}(g) + \mathrm{K}(g)\,]\; (1 + \|F[y,p,w]\,\| + \|F[y',p',w']\,\|\,)\; \rho^n \;,$$

where the constant C depends only on r, Δ_4, $[\Delta \cdot \mathrm{K}](u)$, and $K_\bullet$.

Notice that

$$\sum_{j \in S} \int_{\mathbf{R}^d} \|F[y',p',w']\,\|\; \Pi^{i,j}(y,p,w,dy',dp',dw') \leq 2\, \Delta_2\, \|F[y,p,w]\,\| + \mathrm{K}(u,y) \;,$$

for any $i \in S$, any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$, hence the following corollary holds, whose proof is similar to the proof of Proposition 2 in Benveniste, Métivier and Priouret [3, Part II, Chapter 2].

Corollary 5.5 *With the assumptions of Theorem 5.4, the Markov chain $\{Z'_n = (X_n, Y_n, p_n, w_n)\,,\; n \geq 0\}$ has, under the true probability measure $\mathbf{P}_\bullet$, a unique invariant probability distribution $\mu = (\mu^i)$ on $S \times \mathbf{R}^d \times \mathcal{P}(S) \times \Sigma$. For any $z \in S \times \mathbf{R}^d \times \mathcal{P}(S) \times \Sigma$, and for any function $g = (g^i)$ in L'*

$$|\Pi^n\, g(z) - \lambda| \leq C\, \varepsilon^{-6r}\, [\,\mathrm{Lip}(g) + \mathrm{K}(g)\,]\; (1 + \|F[y,p,w]\,\| + \mathrm{K}(u,y)\,)\; \frac{\rho^n}{1-\rho} \;,$$

and there exist a unique solution $V = (V^i)$ defined on $S \times \mathbf{R}^d \times \mathcal{P}(S) \times \Sigma$ of the Poisson equation

$$[\,I - \Pi\,]\; V(z) = g(z) - \lambda \;,$$

where the constant λ is defined as

$$\lambda = \sum_{i \in S} \int_{\mathbf{R}^d \times \mathcal{P}(S) \times \Sigma} g^i(y,p,w)\, \mu^i(dy,dp,dw) \;.$$

The proof of the theorem is based on the next two propositions.

Proposition 5.6 *If the stochastic matrix Q is primitive, with index of primitivity r, if Assumption C holds, and if Δ_4, $\mathrm{Lip}(u)$, and $[\Delta \cdot \mathrm{K}](u)$ are finite, then there exist a constant $C > 0$ such that, for any $p, p' \in \mathcal{P}(S)$, and any $w, w' \in \Sigma$, for any integers n, m such that $n \geq m + 4r$, and for any function $g = (g^i)$ in L'*

$$\begin{aligned}\max_{i_m, \cdots, i_{n+1} \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \; & |g^{i_{n+1}}(y_{n+1}, f \otimes F[y_n, \cdots, y_m, p, w]) \\ & \qquad - g^{i_{n+1}}(y_{n+1}, f \otimes F[y_n, \cdots, y_m, p', w'])\,| \\ & \qquad b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1}) \leq \\ \leq C\, \varepsilon^{-4r}\; \mathrm{Lip}(g)\; (n - m + 1)\, \rho_*^{n-m-4r}\; & (1 + \|w\| + \|w'\|\,) \;,\end{aligned}$$

where $\rho_* = (1 - R)^{1/r}$, and where the constant C depends only on r, Δ_4, $\mathrm{Lip}(u)$, and $[\Delta \cdot \mathrm{K}](u)$.

The proof of Proposition 5.6 is given in Appendix D.

Proposition 5.7 *If the stochastic matrix Q is primitive, with index of primitivity r, if Assumption C holds, and if Δ_2 and $\mathrm{K}(u)$ are finite, then there exist a constant $C > 0$ such that, for any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$, for any integers n, m such that $n \geq m + 1$, and for any function $g = (g^i)$ in L'*

$$\begin{aligned}\max_{i_m, \cdots, i_{n+1} \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \; & |g^{i_{n+1}}(y_{n+1}, f \otimes F[y_n, \cdots, y_m, p, w])\,| \\ & \qquad b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1}) \\ \leq C\, \varepsilon^{-2r}\; \mathrm{K}(g)\; (n - m + 1)\; & (1 + \|w\|\,) \;,\end{aligned}$$

where the constant C depends only on r, Δ_2 and $\mathrm{K}(u)$.

Proof. For any sequence $i_m, \cdots, i_{n+1} \in S$

$$\int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} |g^{i_{n+1}}(y_{n+1}, f \otimes F[y_n, \cdots, y_m, p, w])|$$

$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1}) \leq$$

$$\leq \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \mathrm{K}(g^{i_{n+1}}, y_{n+1})\, (1 + \|F[y_n, \cdots, y_m, p, w]\|)$$

$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1})$$

$$\leq \mathrm{K}(g) \sum_{i_m, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} (1 + \|F[y_n, \cdots, y_m, p, w]\|)$$

$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_m) \cdots \lambda(dy_n)\ .$$

The rough estimate (21) derived in the proof of Proposition 5.6 (see Appendix D), yields

$$\begin{aligned} &\int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \|F[y_n, \cdots, y_m, p, w]\|\, b_\bullet^{i}(y_m) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_m) \cdots \lambda(dy_n) \leq \\ &\leq 4\, \varepsilon^{-2r}\, \Delta_2^r\, \|w\| + 4\,(n-m)\, \varepsilon^{-2r}\, \Delta_2^r\, \mathrm{K}(u) + \mathrm{K}(u) \\ &\leq C'\, \varepsilon^{-2r}\, (n-m+1)\, (1 + \|w\|)\ , \end{aligned} \tag{12}$$

where the constant C' depends only on r, Δ_2 and $\mathrm{K}(u)$. □

Proof of Theorem 5.4. The proof follows the same lines as the proof of Theorem 3.6. Let $\rho_{\max} = \max(\rho_*, \rho_\bullet)$, where $\rho_* = (1-R)^{1/r}$, and where the constants $\rho_\bullet$ and R are defined in Remark 1.2 and in Proposition 2.5 respectively. Recall that

$$(\Pi^n\, g)^i(y, p, w) = \sum_{i_1, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_1, y, p, w])$$

$$q_\bullet^{i, i_1} \cdots q_\bullet^{i_{n-1}, i_n}\, b_\bullet^{i_1}(y_1) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_1) \cdots \lambda(dy_n)\ ,$$

for any $i \in S$, any $y \in \mathbf{R}^d$, any $p \in \mathcal{P}(S)$, and any $w \in \Sigma$. The following decomposition holds

$$\begin{aligned} (\Pi^n\, g)^i(y, p, w) - (\Pi^n\, g)^{i'}(y', p', w') &= (\Pi^n\, g)^i(y, p, w) - (\Pi^n\, g)^i(y', p', w') \\ &\quad + (\Pi^n\, g)^i(y', p', w') - (\Pi^n\, g)^{i'}(y', p', w')\ , \end{aligned}$$

and we will estimate separately the two terms in the right–hand side.

□ To estimate the first term, we use the exponential forgetting of the prediction filter and its gradient. Using Proposition 5.6 yields

$$|\, (\Pi^n\, g)^i(y, p, w) - (\Pi^n\, g)^i(y', p', w')\, | \leq$$

$$\leq \sum_{i_1, \cdots, i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} |\, g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_1, y, p, w])$$

$$- g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_1, y', p', w'])\, |$$

$$q_\bullet^{i, i_1} \cdots q_\bullet^{i_{n-1}, i_n}\, b_\bullet^{i_1}(y_1) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_1) \cdots \lambda(dy_n)$$

$$\leq \sum_{i_1,\cdots,i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} |\, g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_1, f \otimes F[y,p,w]\,])$$

$$- g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_1, f \otimes F[y',p',w']\,])\,|$$

$$q_\bullet^{i,i_1} \cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_1) \cdots \lambda(dy_n)$$

$$\leq C'\, \varepsilon^{-4r}\ \mathrm{Lip}(g)\ (n-1)\, \rho_*^{n-2-4r}\ (1 + \|F[y,p,w]\,\| + \|F[y',p',w']\,\|)\ ,$$

where the constant C' depends only on r, Δ_4, $\mathrm{Lip}(u)$, and $[\Delta \cdot \mathrm{K}](u)$, hence

$$|\, (\Pi^n\, g)^i(y,p,w) - (\Pi^n\, g)^i(y',p',w')\,| \leq C''\, \varepsilon^{-4r}\ \mathrm{Lip}(g)\ \rho^n\ (1 + \|F[y,p,w]\,\| + \|F[y',p',w']\,\|)\ ,$$

for any ρ such that $\rho_{\max} < \rho < 1$, where the constant C'' depends only on r, Δ_4, $\mathrm{Lip}(u)$, and $[\Delta \cdot \mathrm{K}](u)$.

□ To estimate the second term, we use the geometric convergence of the *true* transition probabilities of the chain $\{X_n\ ,\ n \geq 0\}$, which is a consequence of Assumption A, and we use again the exponential forgetting of the prediction filter and its gradient. Recall that

$$(\Pi^n\, g)^i(y',p',w') - (\Pi^n\, g)^{i'}(y',p',w') =$$

$$= \sum_{i_1,\cdots,i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_1, p_1, w_1])\, [q_\bullet^{i,i_1} - q_\bullet^{i',i_1}]$$

$$q_\bullet^{i_1,i_2} \cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_1}(y_1) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_1) \cdots \lambda(dy_n)\ ,$$

where the notation $(p_1, w_1) = f \otimes F[y',p',w'] = (f[y',p'], F[y',p',w'])$ is used.
As in the proof of Theorem 3.6 above, for any integer m such that $m \leq n-1$, and for any sequence $z_1, \cdots, z_m \in \mathbf{R}^d$

$$(\Pi^n\, g)^i(y',p',w') - (\Pi^n\, g)^{i'}(y',p',w') =$$

$$= \sum_{k=1}^{m} \sum_{i_k,\cdots,i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} [\, g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_k, z_{k-1}, \cdots, z_1, p_1, w_1])$$

$$- g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_{k+1}, z_k, \cdots, z_1, p_1, w_1])\,]$$

$$[\, q_\bullet^{i,i_k}(k) - q_\bullet^{i',i_k}(k)\,]\, q_\bullet^{i_k,i_{k+1}} \cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_k}(y_k) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_k) \cdots \lambda(dy_n)$$

$$+ \sum_{i_{m+1},\cdots i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_{m+1}, z_m, \cdots, z_1, p_1, w_1])$$

$$[\, q_\bullet^{i,i_{m+1}}(m+1) - q_\bullet^{i',i_{m+1}}(m+1)\,]$$

$$q_\bullet^{i_{m+1},i_{m+2}} \cdots q_\bullet^{i_{n-1},i_n}\, b_\bullet^{i_{m+1}}(y_{m+1}) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_{m+1}) \cdots \lambda(dy_n)\ .$$

This holds for any integer m such that $m \leq n-1$. Taking now $m = n - 4r - 2$, using Propositions 5.6 and 5.7, and using estimate (1) yields

$$|(\Pi^n\, g)^i(y',p',w') - (\Pi^n\, g)^{i'}(y',p',w')| \leq$$

$$\leq \sum_{k=1}^{m} \sum_{i_k,\cdots,i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} |\, g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_k, z_{k-1}, \cdots, z_1, p_1, w_1])$$

$$- g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_{k+1}, z_k, \cdots, z_1, p_1, w_1]) \, |$$

$$| \, q_\bullet^{i,i_k}(k) - q_\bullet^{i',i_k}(k) \, | \, q_\bullet^{i_k,i_{k+1}} \cdots q_\bullet^{i_{n-1},i_n} \, b_\bullet^{i_k}(y_k) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_k) \cdots \lambda(dy_n)$$

$$+ \sum_{i_{m+1},\cdots i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} | \, g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_{m+1}, z_m, \cdots, z_1, p_1, w_1]) \, |$$

$$| \, q_\bullet^{i,i_{m+1}}(m+1) - q_\bullet^{i',i_{m+1}}(m+1) \, |$$

$$q_\bullet^{i_{m+1},i_{m+2}} \cdots q_\bullet^{i_{n-1},i_n} \, b_\bullet^{i_{m+1}}(y_{m+1}) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_{m+1}) \cdots \lambda(dy_n)$$

$$\leq \sum_{k=1}^m \sum_{i_k,\cdots,i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} | \, g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_{k+1}, f \otimes F[y_k, z_{k-1}, \cdots, z_1, p_1, w_1]\,])$$

$$- g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_{k+1}, f \otimes F[z_k, z_{k-1}, \cdots, z_1, p_1, w_1]\,]) \, |$$

$$| \, q_\bullet^{i,i_k}(k) - q_\bullet^{i',i_k}(k) \, | \, q_\bullet^{i_k,i_{k+1}} \cdots q_\bullet^{i_{n-1},i_n} \, b_\bullet^{i_k}(y_k) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_k) \cdots \lambda(dy_n)$$

$$+ \sum_{i_{m+1},\cdots i_n \in S} \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} | \, g^{i_n}(y_n, f \otimes F[y_{n-1}, \cdots, y_{m+1}, f \otimes F[z_m, \cdots, z_1, p_1, w_1]\,]) \, |$$

$$| \, q_\bullet^{i,i_{m+1}}(m+1) - q_\bullet^{i',i_{m+1}}(m+1) \, |$$

$$q_\bullet^{i_{m+1},i_{m+2}} \cdots q_\bullet^{i_{n-1},i_n} \, b_\bullet^{i_{m+1}}(y_{m+1}) \cdots b_\bullet^{i_n}(y_n) \, \lambda(dy_{m+1}) \cdots \lambda(dy_n)$$

$$\leq C' \, \varepsilon^{-4r} \, \mathrm{Lip}(g) \, \sum_{k=1}^m (n-1-k) \, \rho_*^{n-k-2-4r}$$

$$\sum_{i_k \in S} \int_{\mathbf{R}^d} (1 + \|F[z_k, z_{k-1}, \cdots, z_1, p_1, w_1]\| + \|F[y_k, z_{k-1}, \cdots, z_1, p_1, w_1]\|)$$

$$| \, q_\bullet^{i,i_k}(k) - q_\bullet^{i',i_k}(k) \, | \, b_\bullet^{i_k}(y_k) \, \lambda(dy_k)$$

$$+ C'' \, \varepsilon^{-2r} \, \mathrm{K}(g) \, (n-1-m) \sum_{i_{m+1} \in S} | \, q_\bullet^{i,i_{m+1}}(m+1) - q_\bullet^{i',i_{m+1}}(m+1) \, |$$

$$(1 + \| \, F[z_m, \cdots, z_1, p_1, w_1] \, \|) \; ,$$

where the constant C' depends only on r, Δ_4, $\mathrm{Lip}(u)$, and $[\Delta \cdot \mathrm{K}](u)$, and the constant C'' depends only on r, Δ_2, and $\mathrm{K}(u)$. Recall that

$$\|F[y_k, z_{k-1}, \cdots, z_1, p_1, w_1]\| \leq 2 \, \delta^2(y_k) \, \|F[z_{k-1}, \cdots, z_1, p_1, w_1]\| + \mathrm{K}(u, y_k) \; ,$$

hence

$$\int_{\mathbf{R}^d} \|F[y_k, z_{k-1}, \cdots, z_1, p_1, w_1]\| \, b_\bullet^{i_k}(y_k) \, \lambda(dy_k) \leq 2 \, \Delta_2 \, \|F[z_{k-1}, \cdots, z_1, p_1, w_1]\| + \mathrm{K}(u) \; .$$

Therefore, using estimate (1) yields

$$|(\Pi^n \, g)^i(y', p', w') - (\Pi^n \, g)^{i'}(y', p', w')| \leq$$

$$\leq C' \, \varepsilon^{-4r} \, \mathrm{Lip}(g) \, \sum_{k=1}^m (n-1-k) \, \rho_*^{n-k-2-4r} \, [\, 2 \, K_\bullet \, \rho_\bullet^k \,]$$

$$(1 + \|F[z_k, z_{k-1}, \cdots, z_1, p_1, w_1]\| + 2\,\Delta_2\,\|F[z_{k-1}, \cdots, z_1, p_1, w_1]\| + \mathrm{K}(u)\,)$$

$$+\, C''\,\varepsilon^{-2r}\;\mathrm{K}(g)\;(n-1-m)\;[\,2\,K_\bullet\,\rho_\bullet^{m+1}\,]\;(1 + \|\,F[z_m, \cdots, z_1, p_1, w_1]\,\|\,)\;.$$

This estimate holds for any sequence $z_1, \cdots, z_m \in \mathbf{R}^d$. Integrating and using the estimate (12) obtained in the proof of Proposition 5.7 yields, for any sequence $i_1, \cdots, i_m \in S$

$$|(\Pi^n\,g)^i(y', p', w') - (\Pi^n\,g)^{i'}(y', p', w')| \leq$$

$$\leq 2\,C'\,K_\bullet\,\varepsilon^{-4r}\;\mathrm{Lip}(g)\;\sum_{k=1}^{m}(n-1-k)\;\rho_{\max}^{n-2-4r}$$

$$(1 + \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d}\|F[z_k, z_{k-1}, \cdots, z_1, p_1, w_1]\|\;b_\bullet^{i_1}(z_1)\cdots b_\bullet^{i_k}(z_k)\,\lambda(dz_1)\cdots\lambda(dz_k)$$

$$+\,2\,\Delta_2\;\int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d}\|F[z_{k-1}, \cdots, z_1, p_1, w_1]\|\;b_\bullet^{i_1}(z_1)\cdots b_\bullet^{i_{k-1}}(z_{k-1})\,\lambda(dz_1)\cdots\lambda(dz_{k-1})$$

$$+\,\mathrm{K}(u)\,)$$

$$+\,2\,C''\,K_\bullet\,\varepsilon^{-2r}\;\mathrm{K}(g)\;(n-1-m)\;\rho_{\max}^{m+1}$$

$$(1 + \int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d}\|\,F[z_m, \cdots, z_1, p_1, w_1]\,\|\;b_\bullet^{i_1}(z_1)\cdots b_\bullet^{i_m}(z_m)\,\lambda(dz_1)\cdots\lambda(dz_m)\,)$$

$$\leq 2\,C'\,C'''\,K_\bullet\,\varepsilon^{-6r}\;\mathrm{Lip}(g)\;[\sum_{k=1}^{m}(n-1-k)\,k\,]\;\rho_{\max}^{n-2-4r}\;(1 + \|w_1\|\,)$$

$$+\,2\,C''\,C'''\,K_\bullet\,\varepsilon^{-4r}\;\mathrm{K}(g)\;(n-1-m)\,m\;\rho_{\max}^{m+1}\;(1 + \|w_1\|\,)\;,$$

where the constant C''' depends only on r, Δ_2 and $\mathrm{K}(u)$, hence (using $m = n - 4r - 2$)

$$|(\Pi^n\,g)^i(y', p', w') - (\Pi^n\,g)^{i'}(y', p', w')| \leq$$

$$\leq 2\,C'\,C'''\,K_\bullet\,\varepsilon^{-6r}\;\mathrm{Lip}(g)\;[\;\sum_{k=1}^{n-4r-1}(n-1-k)\,k\,]\;\rho_{\max}^{n-2-4r}\;(1 + \|F[y', p', w']\|\,)$$

$$+\,2\,C''\,C'''\,K_\bullet\,\varepsilon^{-4r}\;\mathrm{K}(g)\;(4r+1)\,(n-4r-2)\;\rho_{\max}^{n-4r-1}\;(1 + \|F[y', p', w']\|\,)$$

$$\leq C''''\,\varepsilon^{-6r}\;[\,\mathrm{Lip}(g) + \mathrm{K}(g)\,]\;\rho^n\;(1 + \|F[y', p', w']\|\,)\;,$$

for any ρ such that $\rho_{\max} < \rho < 1$, where the constant C'''' depends only on r, Δ_4, $\mathrm{Lip}(u)$, $[\Delta\cdot\mathrm{K}](u)$, and $K_\bullet$.

Combining the above estimates finishes the proof. □

Acknowledgement

The authors gratefully acknowledge Jan van Schuppen for bringing the paper [1] to their attention, Ofer Zeitouni for providing a preprint of the paper [2], and Dan Ocone and Amarjit Budhiraja for interesting discussion, in particular for their suggestion that using the Birkhoff contraction coefficient could improve our earlier estimate (15).

References

[1] A. ARAPOSTATHIS and S.I. MARCUS. Analysis of an identification algorithm arising in the adaptive estimation of Markov chains. *Mathematics of Control, Signals, and Systems*, 3(1):1–29, 1990.

[2] R. ATAR and O. ZEITOUNI. Lyapunov exponents for finite state nonlinear filtering. *SIAM Journal on Control and Optimization*. (to appear).

[3] A. BENVENISTE, M. MÉTIVIER, and P. PRIOURET. *Adaptive Algorithms and Stochastic Approximations*, volume 22 of *Applications of Mathematics*. Springer Verlag, New York, 1990.

[4] G. DI MASI and Ł. STETTNER. On adaptive control of a partially observed Markov chain. *Applicationes Mathematicæ*, 22(2):165–180, 1994.

[5] J.L. DOOB. *Stochastic Processes*. Wiley, New York, 1953.

[6] H. FURSTENBERG and H. KESTEN. Product of random matrices. *The Annals of Mathematical Statistics*, 31(2):457–469, June 1960.

[7] T. KAIJSER. A limit theorem for partially observed Markov chains. *The Annals of Probability*, 3(4):677–696, August 1975.

[8] T. KAIJSER. A limit theorem for Markov chains with applications to products of random matrices. *Duke Mathematical Journal*, 45(2):311–349, June 1978.

[9] F. LE GLAND and L. MEVEL. Recursive identification of HMM's with observations in a finite set. In *Proceedings of the 34th Conference on Decision and Control, New Orleans 1995*, pages 216–221. IEEE–CSS, December 1995.

[10] E. SENETA. Coefficients of ergodicity : structure and applications. *Advances in Applied Probability*, 11:576–590, 1979.

[11] E. SENETA. *Non–negative Matrices and Markov Chains*. Springer Series in Statistics. Springer Verlag, New York, 2nd edition, 1981.

[12] Ł. STETTNER. Invariant measures for the pair : state, approximate filtering process. *Colloquium Mathematicum*, LXII(2):347–351, 1991.

A Product of non–negative matrices

Let $S = \{1, \cdots, N\}$, and let $\| \cdot \|$ denote the L_1–norm, i.e. for any $u = (u^i)$ in $\mathbf{R}^N$

$$\|u\| = \sum_{i \in S} |u^i| \; .$$

We recall some results on coefficients of ergodicity, taken from Seneta [10] and [11, Chapter 3].

For any column–allowable nonnegative $N \times N$ matrix $T = (T^{j,i})$ (i.e. with at least one positive element in each column) the Birkhoff contraction coefficient $\tau_{\rm B}(T)$ is defined by

$$\tau_{\rm B}(T) = \frac{1 - \sqrt{\Phi(T)}}{1 + \sqrt{\Phi(T)}} \leq 1 - \sqrt{\Phi(T)} \; ,$$

where

- if each row of T is either zero–free or all–zero

$$\Phi(T) = \min_{i_1, i_2, j_1, j_2 \in S} \frac{T^{j_1, i_1} \, T^{j_2, i_2}}{T^{j_1, i_2} \, T^{j_2, i_1}} \; ,$$

  where the indices $j_1, j_2 \in S$ are restricted to the zero–free rows,

- otherwise (i.e. if there is a row with both zero and positive elements), $\Phi(T) = 0$,

see Seneta [11, Theorems 3.10 and 3.12].

For any stochastic matrix $P = (P^{i,j})$, and any probability vectors $p, q \in \mathcal{P}(S)$

$$\|P^* \, (p - q)\| \leq \tau_1(P) \, \|p - q\| \; ,$$

where the coefficient of ergodicity $\tau_1(P)$ is defined by

$$\tau_1(P) = \tfrac{1}{2} \max_{i, i' \in S} \sum_{j \in S} |P^{i,j} - P^{i',j}| \; ,$$

see Seneta [10]. Moreover $\tau_1(P) \leq \tau_{\rm B}(P^*)$, and if in addition the stochastic matrix P is allowable (i.e. both row– and column–allowable), then $\tau_1(P) \leq \tau_{\rm B}(P^*) = \tau_{\rm B}(P)$, see Seneta [11, Theorem 3.13].

For any integer $n \geq 0$, let $A_n = (A_n^{i,j})$ be a row–allowable nonnegative $N \times N$ matrix. For any integers n, m such that $n \geq m$, define the backward product

$$A_{m,n} = (A_{m,n}^{i,j}) = A_m \cdots A_n \; ,$$

and the forward product

$$M_{n,m} = (M_{n,m}^{j,i}) = M_n \cdots M_m = A_{m,n}^* \; ,$$

and for any $i \in S$, define the sum of the i–th column entries of $M_{n,m}$, or equivalently the sum of the i–th row entries of $A_{m,n}$, as

$$M_{n,m}^{\bullet,i} = \sum_{j \in S} M_{n,m}^{j,i} = \sum_{j \in S} A_{m,n}^{i,j} = A_{m,n}^{i,\bullet} \; .$$

Define also the ratio

$$\omega(M_{n,m}) = \frac{\displaystyle\max_{i \in S} M_{n,m}^{\bullet,i}}{\displaystyle\min_{i \in S} M_{n,m}^{\bullet,i}} \; .$$

Let $e = (1, \cdots, 1)^*$ denote the N–dimensional vector with all entries equal to 1.

Proposition A.1 *For any integers n, m such that $n \geq m$, and any $p, q \in \mathcal{P}(S)$*

$$\| \, \frac{M_{n,m} \, p}{e^* \, M_{n,m} \, p} - \frac{M_{n,m} \, q}{e^* \, M_{n,m} \, q} \, \| \leq 2 \, \omega(M_{n,m}) \; \tau_{\rm B}(M_{n,m}) \; \|p - q\| \; .$$

The proof follows the same lines as the proof of Lemma 2.2 in Arapostathis and Marcus [1], see also Lemma 6.2 in Kaijser [7], and Lemma 8.1 in Kaijser [8].

Proof. For any integers n, m such that $n \geq m$, and any $p, q \in \mathcal{P}(S)$

$$\begin{aligned} \| \frac{M_{n,m}\, p}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^*\, M_{n,m}\, q} \| &= \sum_{j \in S} | \sum_{i \in S} [\, \frac{M_{n,m}^{j,i}\, p^i}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}^{j,i}\, q^i}{e^*\, M_{n,m}\, q} \,] \, | \\ &= \sum_{j \in S} | \sum_{i \in S} \frac{M_{n,m}^{j,i}}{M_{n,m}^{\bullet,i}} \, [\, \frac{M_{n,m}^{\bullet,i}\, p^i}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}^{\bullet,i}\, q^i}{e^*\, M_{n,m}\, q} \,] \, | \\ &= \sum_{j \in S} | \sum_{i \in S} \bar{M}_{n,m}^{j,i}\, (\bar{p}^i - \bar{q}^i) \, | = \| \bar{M}_{n,m}\, (\bar{p} - \bar{q}) \, \| \ , \end{aligned}$$

where $\bar{M}_{n,m} = (\bar{M}_{n,m}^{j,i})$, $\bar{p} = (\bar{p}^i)$ and $\bar{q} = (\bar{q}^i)$ are defined by

$$\bar{M}_{n,m}^{j,i} = \frac{M_{n,m}^{j,i}}{M_{n,m}^{\bullet,i}} = \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} = \bar{A}_{m,n}^{i,j} \ , \tag{13}$$

for any $i, j \in S$, and

$$\bar{p}^i = \frac{M_{n,m}^{\bullet,i}\, p^i}{e^*\, M_{n,m}\, p} \qquad \text{and} \qquad \bar{q}^i = \frac{M_{n,m}^{\bullet,i}\, q^i}{e^*\, M_{n,m}\, q} \ ,$$

for any $i \in S$, respectively. Notice that $\bar{A}_{m,n} = \bar{M}_{n,m}^*$ is a stochastic matrix, and $\bar{p}, \bar{q} \in \mathcal{P}(S)$ are probability vectors, hence

$$\| \bar{M}_{n,m}\, (\bar{p} - \bar{q}) \, \| \leq \tau_1(\bar{A}_{m,n}) \; \| \bar{p} - \bar{q} \| \ ,$$

and

$$\tau_1(\bar{A}_{m,n}) \leq \tau_{\mathrm{B}}(\bar{M}_{n,m}) = \tau_{\mathrm{B}}(M_{n,m}) \ .$$

Notice also that

$$\bar{p} = \frac{M\, p}{e^*\, M\, p} \qquad \text{and} \qquad \bar{q} = \frac{M\, q}{e^*\, M\, q} \ ,$$

where the diagonal matrix $M = (M^{j,i})$ is defined by

$$M^{j,i} = M_{n,m}^{\bullet,i}\, \delta^{j,i} \ ,$$

for any $i, j \in S$, and the rough estimate of Lemma C.2 gives

$$\| \bar{p} - \bar{q} \| \leq 2\, \omega(M) \; \| p - q \| = 2\, \omega(M_{n,m}) \; \| p - q \| \ ,$$

which finishes the proof. □

For any integer $n \geq 0$, assume that the matrix A_n is positive, and define

$$1 \leq \delta_n = \frac{\displaystyle\max_{i,j \in S} A_n^{i,j}}{\displaystyle\min_{i,j \in S} A_n^{i,j}} < \infty \ .$$

In this special case, the estimate of Proposition A.1 can be expressed as follows.

Proposition A.2 *For any integers n, m such that $n \geq m$, and any $p, q \in \mathcal{P}(S)$*

$$\| \frac{M_{n,m}\, p}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^*\, M_{n,m}\, q} \| \leq 2\, \delta_m \prod_{k=m}^{n} (1 - \delta_k^{-1}) \; \| p - q \| \ .$$

Proof. For any $i, i' \in S$

$$\frac{M_{n,m}^{\bullet,i}}{M_{n,m}^{\bullet,i'}} = \frac{\displaystyle\sum_{j \in S} M_{n,m+1}^{\bullet,j}\, M_m^{j,i}}{\displaystyle\sum_{j \in S} M_{n,m+1}^{\bullet,j}\, M_m^{j,i'}} \leq \delta_m \ ,$$

(if $m = n$, then $M_{n,m+1} = M_{n,n+1} = I$ by convention) hence $\omega(M_{n,m}) \leq \delta_m$.

Moreover, the multiplicative property of the Birkhoff contraction coefficient yields

$$\tau_{\mathrm{B}}(M_{n,m}) = \tau_{\mathrm{B}}(A_{m,n}) \leq \prod_{k=m}^{n} \tau_{\mathrm{B}}(A_k) = \prod_{k=m}^{n} \frac{1-\delta_k^{-1}}{1+\delta_k^{-1}} \leq \prod_{k=m}^{n} (1-\delta_k^{-1}) \; . \qquad \square$$

Remark A.3 Using the alternate bound

$$\tau_1(\bar{A}_{m,n}) = \tfrac{1}{2} \max_{i,i' \in S} \sum_{j \in S} |\bar{A}^{i,j}_{m,n} - \bar{A}^{i',j}_{m,n}| \leq \tfrac{1}{2} \sum_{j \in S} [\, \max_{i \in S} \bar{A}^{i,j}_{m,n} - \min_{i \in S} \bar{A}^{i,j}_{m,n} \,] \; ,$$

and the estimate

$$[\, \max_{i \in S} \bar{A}^{i,j}_{m,n} - \min_{i \in S} \bar{A}^{i,j}_{m,n} \,] = [\, \max_{i \in S} \frac{A^{i,j}_{m,n}}{A^{i,\bullet}_{m,n}} - \min_{i \in S} \frac{A^{i,j}_{m,n}}{A^{i,\bullet}_{m,n}} \,] \leq \prod_{k=m}^{n-1} (1-\delta_k^{-2}) \; , \tag{14}$$

one would obtain

$$\| \frac{M_{n,m}\, p}{e^* \, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^* \, M_{n,m}\, q} \| \leq N \, \delta_m \prod_{k=m}^{n-1} (1-\delta_k^{-2}) \, \|p-q\| \; . \tag{15}$$

which has the unpleasant feature that it depends explicitly on the dimension N. For the sake of completeness, the estimate (14) will be proved in Lemma B.3 below, and a direct proof of the bound

$$\|\bar{M}_{n,m}\, (\bar{p} - \bar{q})\,\| \leq \tfrac{1}{2} \sum_{j \in S} [\, \max_{i \in S} \bar{M}^{j,i}_{n,m} - \min_{i \in S} \bar{M}^{j,i}_{n,m} \,] \; \|\bar{p} - \bar{q}\| \; .$$

will be given in Lemma C.1.

We consider now two important examples, for which we can improve or extend the general estimate of Lemma A.1 above.

Example 1

For any integer $n \geq 0$, assume that $A_n = (A^{i,j}_n) = B_n \, Q_n$, where $Q_n = (q^{i,j}_n)$ is an $N \times N$ positive stochastic matrix, and $B_n = \mathrm{diag}\,[b^1_n, \cdots, b^N_n]$ is a diagonal $N \times N$ matrix with positive diagonal entries. For any integer $n \geq 0$, define

$$1 \leq \delta^B_n = \frac{\max_{i \in S} b^i_n}{\min_{i \in S} b^i_n} < \infty \qquad \text{and} \qquad \varepsilon_n = \min_{i,j \in S} q^{i,j}_n > 0 \; .$$

In this special case

$$\delta_n = \frac{\max_{i,j \in S} A^{i,j}_n}{\min_{i,j \in S} A^{i,j}_n} \leq \varepsilon_n^{-1} \, \delta^B_n \; ,$$

and

$$\tau_{\mathrm{B}}(A_n) = \tau_{\mathrm{B}}(Q_n) = \frac{1-\varepsilon_n}{1+\varepsilon_n} \leq 1 - \varepsilon_n \; ,$$

for any integer $n \geq 0$, and the estimate of Proposition A.1 can be improved as follows.

Proposition A.4 *For any integers n, m such that $n \geq m$, and any $p, q \in \mathcal{P}(S)$*

$$\| \frac{M_{n,m}\, p}{e^* \, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^* \, M_{n,m}\, q} \| \leq 2\, \varepsilon_m^{-1} \, \delta^B_m \prod_{k=m}^{n} (1-\varepsilon_k) \, \|p-q\| \; .$$

Example 2

For any integer $n \geq 0$, assume that $A_n = (A_n^{i,j}) = B_n\, Q$, where $Q = (q^{i,j})$ is a primitive (or equivalently, irreducible and aperiodic) $N \times N$ stochastic matrix, and $B_n = \mathrm{diag}\,[b_n^1, \cdots, b_n^N]$ is a diagonal $N \times N$ matrix with positive diagonal entries. Define

$$1 \leq \delta_n^B = \frac{\max\limits_{i \in S} b_n^i}{\min\limits_{i \in S} b_n^i} < \infty \qquad \text{and} \qquad \varepsilon = \min_{i,j \in S}{}^{+}\, q^{i,j} > 0 \;,$$

where the notation $\min^+$ denotes the minimum over positive elements. Let r denote the index of primitivity of Q, i.e. the smallest integer r such that the stochastic matrix $Q^r = (q^{i,j}(r))$ is positive. Then

$$\varepsilon(r) = \min_{i,j \in S} q^{i,j}(r) \geq \varepsilon^r > 0 \;,$$

and for any product of r matrices, the Birkhoff contraction coefficient is stictly less than 1, i.e.

$$\begin{aligned} \tau_{\mathrm{B}}(A_{n,n+r-1}) = \tau_{\mathrm{B}}(B_n\, Q\, A_{n+1,n+r-1}) \;&=\; \tau_{\mathrm{B}}(Q\, A_{n+1,n+r-1}) \\ &=\; \frac{1 - \varepsilon^r\, [\delta_{n+1}^B \cdots \delta_{n+r-1}^B]^{-1}}{1 + \varepsilon^r\, [\delta_{n+1}^B \cdots \delta_{n+r-1}^B]^{-1}} \leq 1 - \varepsilon^r\, [\delta_{n+1}^B \cdots \delta_{n+r-1}^B]^{-1} \;, \end{aligned}$$

for any integer $n \geq 0$, and the following lower bound holds

$$1 - \varepsilon^r \leq 1 - \varepsilon^r\, [\delta_{n+1}^B \cdots \delta_{n+r-1}^B]^{-1} \;. \tag{16}$$

In this special case, the estimate of Proposition A.1 can be extended as follows.

Proposition A.5 *For any $p, q \in \mathcal{P}(S)$, and any integers n, m such that $n \geq m + r - 1$, the following estimates hold :*

$$\begin{aligned} &\| \frac{M_{n,m}\, p}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^*\, M_{n,m}\, q} \| \leq \\ &\qquad \leq 2\, \varepsilon^{-r}\, \delta_m^B \cdots \delta_{m+r-1}^B \prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r\; [\delta_{m+\kappa r+1}^B \cdots \delta_{m+(\kappa+1)r-1}^B]^{-1})\; \|p - q\| \;, \end{aligned} \tag{17}$$

and

$$\begin{aligned} &\| \frac{M_{n,m}\, p}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^*\, M_{n,m}\, q} \| \leq \\ &\qquad \leq 2\, \varepsilon^{-r}\, \delta_m^B \cdots \delta_{m+r-1}^B \prod_{\kappa=1}^{[n,m]+1} (1 - \varepsilon^r\; [\delta_{n-\kappa r+2}^B \cdots \delta_{n-(\kappa-1)r}^B]^{-1})\; \|p - q\| \;, \end{aligned}$$

where $[n,m] = \lfloor \dfrac{n-m+1}{r} \rfloor - 1$.

Remark A.6 If $m \leq n \leq m + r - 2$, the rough estimate of Lemma C.3 yields

$$\| \frac{M_{n,m}\, p}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^*\, M_{n,m}\, q} \| \leq 2\, \delta_m^B \cdots \delta_n^B\; \|p - q\| \;,$$

for any $p, q \in \mathcal{P}(S)$.

Remark A.7 To make sure that the factors in the product on the right–hand side span disjoint blocks, the following estimate will be frequently used instead of (17)

$$\begin{aligned} \| \frac{M_{n,m}\, p}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^*\, M_{n,m}\, q} \| &\leq \\ &\leq 2\, \varepsilon^{-r}\, \delta^B_m \cdots \delta^B_{m+r-1} \prod_{\kappa=1}^{[n,m]} (1 - \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1})\, \|p - q\| \,, \end{aligned} \tag{18}$$

for any $p, q \in \mathcal{P}(S)$, and any integers n, m such that $n \geq m + 2r - 1$. If $m \leq n \leq m + 2r - 2$, the rough estimate of Lemma C.3 yields

$$\| \frac{M_{n,m}\, p}{e^*\, M_{n,m}\, p} - \frac{M_{n,m}\, q}{e^*\, M_{n,m}\, q} \| \leq 2\, \varepsilon^{-r}\, \delta^B_m \cdots \delta^B_{\min(m+r-1,n)}\, \|p - q\| \,,$$

for any $p, q \in \mathcal{P}(S)$.

Proof of Proposition A.5. It follows from Lemma C.3 below, that if $n \geq m + r - 1$ then

$$\omega(M_{n,m}) \leq \varepsilon^{-r}\, \delta^B_m \cdots \delta^B_{m+r-1} \;.$$

There are several different ways of splitting the set $\{m, \cdots, n\}$ into $([n, m] + 1)$ blocks of length r. Two examples are presented in Figure 1.

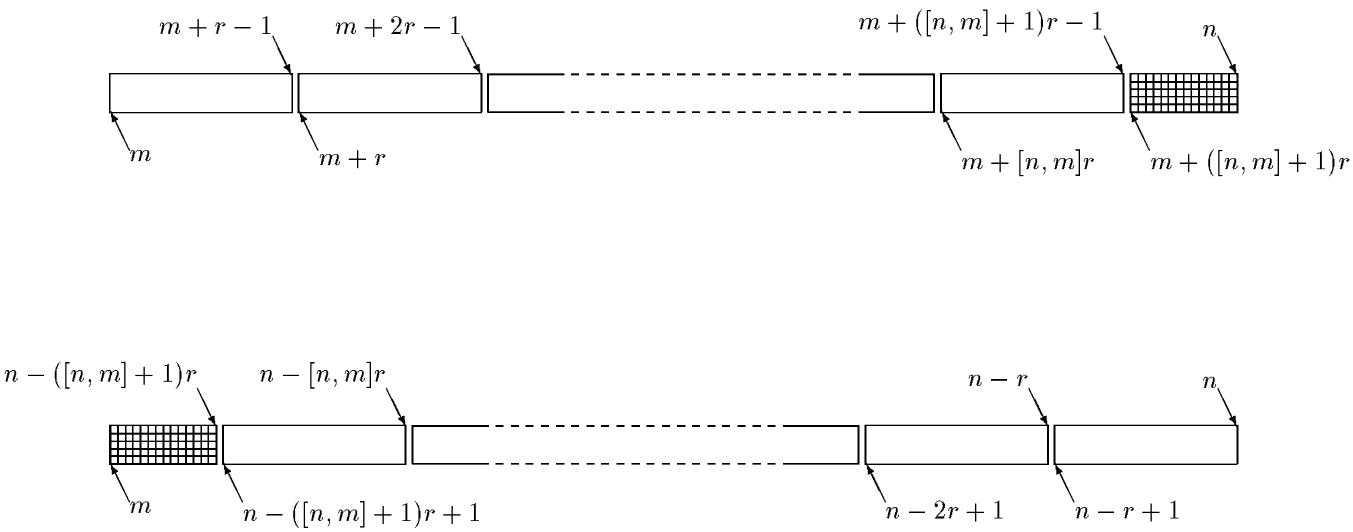

Figure 1: Two different splittings of the set $\{m, \cdots, n\}$

In the first case (forward splitting)

$$A_{m,n} = [\, \prod_{\kappa=0}^{[n,m]} A_{m+\kappa r, m+(\kappa+1)r-1}\,]\; A_{m+([n,m]+1)r, n} \;,$$

and the multiplicative property of the Birkhoff contraction coefficient yields

$$\begin{aligned} \tau_{\mathrm{B}}(M_{n,m}) = \tau_{\mathrm{B}}(A_{m,n}) &\leq \prod_{\kappa=0}^{[n,m]} \tau_{\mathrm{B}}(A_{m+\kappa r, m+(\kappa+1)r-1}) \\ &\leq \prod_{\kappa=0}^{[n,m]} \frac{1 - \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}}{1 + \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}} \end{aligned}$$

$$\leq \prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}\,)\ .$$

In the second case (backward splitting)

$$A_{m,n} = A_{m,n-([n,m]+1)r}\ [\ \prod_{\kappa=1}^{[n,m]+1} A_{n-\kappa r+1,n-(\kappa-1)r}\ ]\ ,$$

and the multiplicative property of the Birkhoff contraction coefficient yields

$$\begin{aligned} \tau_{\mathrm{B}}(M_{n,m}) = \tau_{\mathrm{B}}(A_{m,n}) &\leq \prod_{\kappa=1}^{[n,m]+1} \tau_{\mathrm{B}}(A_{n-\kappa r+1,n-(\kappa-1)r}) \\ &\leq \prod_{\kappa=1}^{[n,m]+1} \frac{1 - \varepsilon^r\, [\, \delta^B_{n-\kappa r+2} \cdots \delta^B_{n-(\kappa-1)r}\,]^{-1}}{1 + \varepsilon^r\, [\, \delta^B_{n-\kappa r+2} \cdots \delta^B_{n-(\kappa-1)r}\,]^{-1}} \\ &\leq \prod_{\kappa=1}^{[n,m]+1} (1 - \varepsilon^r\, [\, \delta^B_{n-\kappa r+2} \cdots \delta^B_{n-(\kappa-1)r}\,]^{-1}\,)\ . \qquad \square \end{aligned}$$

Remark A.8 The two estimates of Proposition A.5 are sharp, in the sense that for $r = 1$ they both reduce to the estimate obtained directly in Proposition A.4 for a positive stochastic matrix Q.

Lemma A.9 *For any integers n, l, m such that $n \geq m + r - 1$, and $m + 1 \leq l \leq n - 1$*

$$\tau_{\mathrm{B}}(M_{n,l+1})\ \tau_{\mathrm{B}}(M_{l-1,m}) \leq (1 - \varepsilon^r)^{-1} \prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}\,)\ .$$

The same bound holds also for $l = m$ and $l = n$, i.e.

$$\tau_{\mathrm{B}}(M_{n,m+1}) \leq (1 - \varepsilon^r)^{-1} \prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}\,)\ .$$

and

$$\tau_{\mathrm{B}}(M_{n-1,m}) \leq (1 - \varepsilon^r)^{-1} \prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}\,)\ .$$

Proof. For any integer l such that $m + 1 \leq l \leq n - 1$, there exist a splitting of the set $\{l + 1, \cdots, n\}$ into blocks of length r, which is compatible with the forward splitting of the set $\{m, \cdots, n\}$. Indeed, if l belongs to the κ_0–th block, i.e. $m + (\kappa_0 - 1)r \leq l \leq m + \kappa_0 r - 1$, hence $\kappa_0 = \lfloor \dfrac{l-m}{r} \rfloor + 1 = [l - 1, m] + 2$, then

$$\begin{aligned} A_{l+1,n} &= A_{l+1,m+\kappa_0 r-1}\ A_{m+\kappa_0 r,m+([n,m]+1)r-1}\ A_{m+([n,m]+1)r,n} \\ &= A_{l+1,m+\kappa_0 r-1}\ [\ \prod_{\kappa=\kappa_0}^{[n,m]} A_{m+\kappa r,m+(\kappa+1)r-1}\ ]\ A_{m+([n,m]+1)r,n}\ , \end{aligned}$$

(if $l = m + \kappa_0 r - 1$, then $A_{l+1,m+\kappa_0 r-1} = A_{m+\kappa_0 r,m+\kappa_0 r-1} = I$ by convention), and the multiplicative property of the Birkhoff contraction coefficient yields

$$\begin{aligned} \tau_{\mathrm{B}}(M_{n,l+1}) = \tau_{\mathrm{B}}(A_{l+1,n}) &\leq \prod_{\kappa=\kappa_0}^{[n,m]} \tau_{\mathrm{B}}(A_{m+\kappa r,m+(\kappa+1)r-1}) \\ &\leq \prod_{\kappa=\kappa_0}^{[n,m]} \frac{1 - \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}}{1 + \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}} \\ &\leq \prod_{\kappa=\kappa_0}^{[n,m]} (1 - \varepsilon^r\, [\, \delta^B_{m+\kappa r+1} \cdots \delta^B_{m+(\kappa+1)r-1}\,]^{-1}\,)\ . \end{aligned}$$

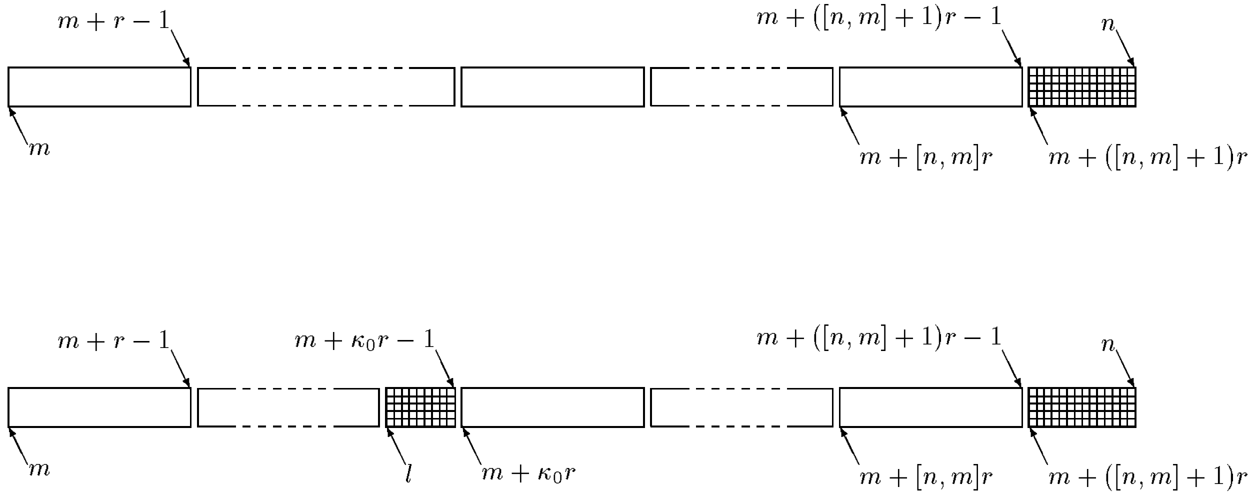

Figure 2: Compatible splittings of the sets $\{m,\cdots,n\}$ and $\{l+1,\cdots,n\}$

Therefore

$$\begin{aligned}
&\tau_{\mathrm{B}}(M_{n,l+1})\ \tau_{\mathrm{B}}(M_{l-1,m}) \leq \\
&\leq \Big[\prod_{\kappa=[l-1,m]+2}^{[n,m]} \frac{1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1}}{1+\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1}}\Big]\Big[\prod_{\kappa=0}^{[l-1,m]} \frac{1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1}}{1+\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1}}\Big] \\
&\leq \prod_{\kappa=[l-1,m]+2}^{[n,m]} (1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1}) \prod_{\kappa=0}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1}) \\
&= \prod_{\substack{\kappa=0\\ \kappa\neq\lfloor\frac{l-m}{r}\rfloor}}^{[n,m]} (1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1})\ ,
\end{aligned}$$

since $[l-1,m]+1=\lfloor\dfrac{l-m}{r}\rfloor$, i.e. the block containing l is ruled out of the product, and the lower bound (16) yields

$$\tau_{\mathrm{B}}(M_{n,l+1})\ \tau_{\mathrm{B}}(M_{l-1,m}) \quad\leq\quad (1-\varepsilon^r)^{-1}\prod_{\kappa=0}^{[n,m]}(1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1})\ .$$

If $l=m$, then

$$\begin{aligned}
\tau_{\mathrm{B}}(M_{n,m+1})\leq\tau_{\mathrm{B}}(M_{n,m+r}) \quad&\leq\quad \prod_{\kappa=1}^{[n,m]}(1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1}) \\
&\leq\quad (1-\varepsilon^r)^{-1}\prod_{\kappa=0}^{[n,m]}(1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1})\ ,
\end{aligned}$$

and if $l=n$, then

$$\tau_{\mathrm{B}}(M_{n-1,m})\leq\tau_{\mathrm{B}}(M_{m+[n,m]r-1,m}) \quad\leq\quad \prod_{\kappa=0}^{[n,m]-1}(1-\varepsilon^r\,[\,\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}\,]^{-1})$$

$$\leq \quad (1-\varepsilon^r)^{-1} \prod_{\kappa=0}^{[n,m]} (1-\varepsilon^r\,[\delta^B_{m+\kappa r+1}\cdots\delta^B_{m+(\kappa+1)r-1}]^{-1})\ . \qquad \square$$

B Backward products

The purpose of this section is to prove the estimate (14). With the notations of the Appendix A, the following estimate holds.

Lemma B.1 *For any integers n, m such that $n \geq m$, and any $j \in S$*

$$\frac{\max\limits_{i\in S} A^{i,j}_{m,n}}{\min\limits_{i\in S} A^{i,j}_{m,n}} \leq \delta_m\ .$$

This estimate improves the estimate given in Lemma 2 of Furstenberg and Kesten [6].
Proof. The estimate obviously holds for $n = m$. For any $n \geq m+1$, and any $i_1, i_2, j \in S$

$$\frac{A^{i_2,j}_{m,n}}{A^{i_1,j}_{m,n}} = \frac{\sum\limits_{i\in S} A^{i_2,i}_m\ A^{i,j}_{m+1,n}}{\sum\limits_{i\in S} A^{i_1,i}_m\ A^{i,j}_{m+1,n}} \leq \max_{i\in S} \frac{A^{i_2,i}_m}{A^{i_1,i}_m} \leq \delta_m\ . \qquad \square$$

Corollary B.2 *For any integers n, m such that $n \geq m+1$, and any $j \in S$*

$$\frac{\max\limits_{i\in S} A^{i,\bullet}_{m,n}}{\min\limits_{i\in S} A^{i,\bullet}_{m,n}} \leq \delta_m\ .$$

Lemma B.3 *For any integers n, m such that $n \geq m+1$, and any $j \in S$*

$$[\ \max_{i\in S} \frac{A^{i,j}_{m,n}}{A^{i,\bullet}_{m,n}} - \min_{i\in S} \frac{A^{i,j}_{m,n}}{A^{i,\bullet}_{m,n}}\ ] \leq \prod_{k=m}^{n-1} (1-\delta_k^{-2})\ .$$

This estimate improves the estimate given in Lemma 3 of Furstenberg and Kesten [6], and the proof follows the same lines, see also Doob [5, pp. 173–174].
Proof. For $n = m$, the following estimate obviously holds

$$[\ \max_{i\in S} \frac{A^{i,j}_m}{A^{i,\bullet}_m} - \min_{i\in S} \frac{A^{i,j}_m}{A^{i,\bullet}_m}\ ] \leq 1\ .$$

For any $n \geq m+1$, and any $i_1, i_2, j \in S$

$$\begin{aligned}
\frac{A^{i_1,j}_{m,n}}{A^{i_1,\bullet}_{m,n}} - \frac{A^{i_2,j}_{m,n}}{A^{i_2,\bullet}_{m,n}} &= \sum_{i\in S} [\ \frac{A^{i_1,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_1,\bullet}_{m,n}} - \frac{A^{i_2,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_2,\bullet}_{m,n}}\ ]\ \frac{A^{i,j}_{m+1,n}}{A^{i,\bullet}_{m+1,n}} \\
&= \sum_{i\in S} \frac{A^{i,j}_{m+1,n}}{A^{i,\bullet}_{m+1,n}}\ [\ \frac{A^{i_1,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_1,\bullet}_{m,n}} - \frac{A^{i_2,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_2,\bullet}_{m,n}}\ ]^+ \\
&\quad - \sum_{i\in S} \frac{A^{i,j}_{m+1,n}}{A^{i,\bullet}_{m+1,n}}\ [\ \frac{A^{i_1,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_1,\bullet}_{m,n}} - \frac{A^{i_2,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_2,\bullet}_{m,n}}\ ]^- \\
&\leq [\ \max_{i\in S} \frac{A^{i,j}_{m+1,n}}{A^{i,\bullet}_{m+1,n}}\ ] \sum_{i\in S} [\ \frac{A^{i_1,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_1,\bullet}_{m,n}} - \frac{A^{i_2,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_2,\bullet}_{m,n}}\ ]^+ \\
&\quad - [\ \min_{i\in S} \frac{A^{i,j}_{m+1,n}}{A^{i,\bullet}_{m+1,n}}\ ] \sum_{i\in S} [\ \frac{A^{i_1,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_1,\bullet}_{m,n}} - \frac{A^{i_2,i}_m\ A^{i,\bullet}_{m+1,n}}{A^{i_2,\bullet}_{m,n}}\ ]^-\ .
\end{aligned}$$

Notice that

$$0 = \sum_{i \in S} \Big[\frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} - \frac{A_m^{i_2,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_2,\bullet}} \Big] \;=\; \sum_{i \in S} \Big[\frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} - \frac{A_m^{i_2,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_2,\bullet}} \Big]^+ - \sum_{i \in S} \Big[\frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} - \frac{A_m^{i_2,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_2,\bullet}} \Big]^- \; .$$

Therefore

$$\frac{A_{m,n}^{i_1,j}}{A_{m,n}^{i_1,\bullet}} - \frac{A_{m,n}^{i_2,j}}{A_{m,n}^{i_2,\bullet}} \;\leq\; \Big[\max_{i \in S} \frac{A_{m+1,n}^{i,j}}{A_{m+1,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m+1,n}^{i,j}}{A_{m+1,n}^{i,\bullet}} \Big] \sum_{i \in S} \Big[\frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} - \frac{A_m^{i_2,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_2,\bullet}} \Big]^+ \; .$$

From the Corollary B.2 above

$$\frac{A_m^{i_1,i}}{A_{m,n}^{i_1,\bullet}} = \frac{A_m^{i_1,i}}{A_m^{i_2,i}}\, \frac{A_{m,n}^{i_2,\bullet}}{A_{m,n}^{i_1,\bullet}}\, \frac{A_m^{i_2,i}}{A_{m,n}^{i_2,\bullet}} \leq \delta_m^2\, \frac{A_m^{i_2,i}}{A_{m,n}^{i_2,\bullet}} \; . \tag{19}$$

Therefore

$$\frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} - \frac{A_m^{i_2,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_2,\bullet}} \leq (1 - \delta_m^{-2})\, \frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} \; ,$$

hence

$$\Big[\frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} - \frac{A_m^{i_2,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_2,\bullet}} \Big]^+ \leq (1 - \delta_m^{-2})\, \frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} \; ,$$

and

$$\sum_{i \in S} \Big[\frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} - \frac{A_m^{i_2,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_2,\bullet}} \Big]^+ \leq (1 - \delta_m^{-2}) \sum_{i \in S} \frac{A_m^{i_1,i}\, A_{m+1,n}^{i,\bullet}}{A_{m,n}^{i_1,\bullet}} = (1 - \delta_m^{-2}) \; .$$

This yields

$$\frac{A_{m,n}^{i_1,j}}{A_{m,n}^{i_1,\bullet}} - \frac{A_{m,n}^{i_2,j}}{A_{m,n}^{i_2,\bullet}} \leq (1 - \delta_m^{-2})\, \Big[\max_{i \in S} \frac{A_{m+1,n}^{i,j}}{A_{m+1,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m+1,n}^{i,j}}{A_{m+1,n}^{i,\bullet}} \Big] \; .$$

Since this holds for any $i_1, i_2 \in S$

$$\Big[\max_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} \Big] \leq (1 - \delta_m^{-2})\, \Big[\max_{i \in S} \frac{A_{m+1,n}^{i,j}}{A_{m+1,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m+1,n}^{i,j}}{A_{m+1,n}^{i,\bullet}} \Big] \; ,$$

and by iterating the estimate

$$\Big[\max_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} \Big] \leq \prod_{k=m}^{n-1} (1 - \delta_k^{-2}) \; . \qquad \square$$

Example 1 (revisited)

For any integer $n \geq 0$, assume that $A_n = (A_n^{i,j}) = B_n\, Q_n$, where $Q_n = (q_n^{i,j})$ is an $N \times N$ matrix (not necessarily stochastic) with positive entries, and $B_n = \mathrm{diag}\,[b_n^1, \cdots, b_n^N]$ is a diagonal $N \times N$ matrix with positive diagonal entries. For any integer $n \geq 0$, define

$$1 \leq \delta_n^B = \frac{\max_{i \in S} b_n^i}{\min_{i \in S} b_n^i} < \infty \qquad \text{and} \qquad 1 \leq \gamma_n = \frac{\max_{i,j \in S} q_n^{i,j}}{\min_{i,j \in S} q_n^{i,j}} < \infty \; .$$

Under these assumptions, the result of Lemma B.3 can be improved as follows.

Lemma B.4 *For any integers n, m such that $n \geq m+1$, and any $j \in S$*

$$\Big[\max_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} \Big] \leq \prod_{k=m}^{n-1} (1 - \gamma_k^{-2}) \; .$$

Proof. In this special case

$$\frac{A_m^{i_1,i}}{A_m^{i_2,i}} = \frac{b_m^{i_1}\ q_m^{i_1,i}}{b_m^{i_2}\ q_m^{i_2,i}} \leq \gamma_m\ \frac{b_m^{i_1}}{b_m^{i_2}}\ ,$$

and

$$\frac{A_{m,n}^{i_2,j}}{A_{m,n}^{i_1,j}} = \frac{\displaystyle\sum_{i \in S} A_m^{i_2,i}\ A_{m+1,n}^{i,j}}{\displaystyle\sum_{i \in S} A_m^{i_1,i}\ A_{m+1,n}^{i,j}} = \frac{b_m^{i_2}\ \displaystyle\sum_{i \in S} q_m^{i_2,i}\ A_{m+1,n}^{i,j}}{b_m^{i_1}\ \displaystyle\sum_{i \in S} q_m^{i_1,i}\ A_{m+1,n}^{i,j}} \leq \frac{b_m^{i_2}}{b_m^{i_1}}\ \max_{i \in S} \frac{q_m^{i_2,i}}{q_m^{i_1,i}} \leq \gamma_m\ \frac{b_m^{i_2}}{b_m^{i_1}}\ ,$$

hence

$$\frac{A_{m,n}^{i_2,\bullet}}{A_{m,n}^{i_1,\bullet}} \leq \gamma_m\ \frac{b_m^{i_2}}{b_m^{i_1}}\ .$$

Therefore

$$\frac{A_m^{i_1,i}}{A_{m,n}^{i_1,\bullet}} = \frac{A_m^{i_1,i}}{A_m^{i_2,i}}\ \frac{A_{m,n}^{i_2,\bullet}}{A_{m,n}^{i_1,\bullet}}\ \frac{A_m^{i_2,i}}{A_{m,n}^{i_2,\bullet}} \leq \gamma_m^2\ \frac{A_m^{i_2,i}}{A_{m,n}^{i_2,\bullet}}\ .$$

The result follows by using this estimate in place of estimate (19) in the proof of the Lemma B.3. □

Remark B.5 If in addition, for any integer $n \geq 0$, the positive matrix Q_n is a stochastic matrix, with

$$\varepsilon_n = \min_{i,j \in S} q_n^{i,j} > 0\ ,$$

the estimate of Lemma B.4 above reduces to

$$[\ \max_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}}\ ] \leq \prod_{k=m}^{n-1} (1 - \varepsilon_k^2)\ .$$

Example 2 (continued)

Under the assumptions of Example 2 in Appendix A, the result of Lemma B.3 can be extended as follows.

Lemma B.6 *For any integers n, m such that $n \geq m + r - 1$, and for any $j \in S$*

$$[\ \max_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}}\ ] \leq \prod_{\kappa=0}^{[n,m]-1} (1 - \varepsilon^{2r}\ [\delta_{m+\kappa r+1}^B \cdots \delta_{m+(\kappa+1)r-1}^B]^{-2})\ .$$

Proof. The following decomposition into a product of positive matrices holds

$$A_{m,n} = [\ \prod_{\kappa=0}^{[n,m]-1} A_{m+\kappa r, m+(\kappa+1)r-1}\ ]\ A_{m+[n,m]r,n} = \bar{A}_0 \cdots \bar{A}_{\ell-1}\ A_{m+[n,m]r,n}\ ,$$

where $\ell = [n,m]$, and for any $\kappa = 0, \cdots, \ell - 1$

$$\bar{A}_\kappa = A_{m+\kappa r, m+(\kappa+1)r-1} = B_{m+\kappa r}\ [Q\, B_{m+\kappa r+1}\, Q \cdots B_{m+(\kappa+1)r-1}\, Q] = \bar{B}_\kappa\, \bar{Q}_\kappa\ .$$

This is the situation of Example 1 above, since $\bar{Q}_\kappa = (\bar{q}_\kappa^{i,j})$ is positive for any $\kappa = 0, \cdots, \ell - 1$, hence

$$\begin{aligned}
[\ \max_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m,n}^{i,j}}{A_{m,n}^{i,\bullet}}\ ] &\leq \prod_{\kappa=0}^{\ell-1} (1 - \bar{\gamma}_\kappa^{-2})\ [\ \max_{i \in S} \frac{A_{m+[n,m]r,n}^{i,j}}{A_{m+[n,m]r,n}^{i,\bullet}} - \min_{i \in S} \frac{A_{m+[n,m]r,n}^{i,j}}{A_{m+[n,m]r,n}^{i,\bullet}}\ ] \\
&\leq \prod_{\kappa=0}^{\ell-1} (1 - \bar{\gamma}_\kappa^{-2})\ ,
\end{aligned}$$

and it is immediate to check that

$$1 \leq \bar{\gamma}_\kappa = \frac{\displaystyle\max_{i,j \in S} \bar{q}_\kappa^{i,j}}{\displaystyle\min_{i,j \in S} \bar{q}_\kappa^{i,j}} \leq \varepsilon^{-r}\ [\delta_{m+\kappa r+1}^B \cdots \delta_{m+(\kappa+1)r-1}^B]\ ,$$

for any $\kappa = 0, \cdots, \ell - 1$. □

C Rough estimates

Let $M = (M^{i,j})$ be an $N \times N$ matrix with nonnegative entries.

Lemma C.1 *For any $z \in \mathbf{R}^N$ such that $e^* z = 0$*

$$\|M z\| \leq \tfrac{1}{2} \sum_{j \in S} \Big[\max_{i \in S} M^{j,i} - \min_{i \in S} M^{j,i} \Big] \; \|z\| \; .$$

Proof. By definition

$$\|M z\| = \sum_{j \in S} | \sum_{i \in S} M^{j,i} \; z_i \, | \; .$$

Since

$$e^* z = \sum_{i \in S} z_i = 0 \; ,$$

then, for any $j \in S$, and any scalar λ^j

$$\sum_{i \in S} M^{j,i} \; z_i = \sum_{i \in S} (M^{j,i} - \lambda^j) \; z_i \; ,$$

hence

$$| \sum_{i \in S} M^{j,i} \; z_i | \leq \sum_{i \in S} |M^{j,i} - \lambda^j| \; |z_i| \; .$$

Selecting

$$\lambda^j = \tfrac{1}{2} \Big[\max_{i \in S} M^{j,i} + \min_{i \in S} M^{j,i} \Big] \; ,$$

yields

$$\sum_{i \in S} |M^{j,i} - \lambda^j| \; |z_i| \leq \tfrac{1}{2} \Big[\max_{i \in S} M^{j,i} - \min_{i \in S} M^{j,i} \Big] \; \|z\| \; ,$$

hence

$$| \sum_{i \in S} M^{j,i} \; z_i \, | \leq \tfrac{1}{2} \; \Big[\max_{i \in S} M^{j,i} - \min_{i \in S} M^{j,i} \Big] \; \|z\| \; . \qquad \square$$

For any $i \in S$, define the sum of the i–th column entries as

$$M^{\bullet,i} = \sum_{j \in S} M^{j,i} \; ,$$

and define the ratio

$$\omega(M) = \sup_{p,q \in \mathcal{P}(S)} \frac{\displaystyle\sum_{i \in S} M^{\bullet,i} \, q^i}{\displaystyle\sum_{i \in S} M^{\bullet,i} \, p^i} = \frac{\displaystyle\max_{i \in S} M^{\bullet,i}}{\displaystyle\min_{i \in S} M^{\bullet,i}} \; .$$

Notice that for any $p \in \mathcal{P}(S)$, and any $z \in \mathbf{R}^N$

$$|\frac{e^* \, M \, z}{e^* \, M \, p}| \leq \frac{\|M \, z\|}{e^* \, M \, p} \leq \frac{\displaystyle\sum_{i \in S} M^{\bullet,i} \, |z^i|}{\displaystyle\sum_{i \in S} M^{\bullet,i} \, p^i} \leq \omega(M) \; \|z\| \; . \tag{20}$$

Lemma C.2 *For any $p, q \in \mathcal{P}(S)$, the following* rough *estimate holds*

$$\|\frac{M \, p}{e^* \, M \, p} - \frac{M \, q}{e^* \, M \, q}\| \leq 2 \, \omega(M) \; \|p - q\| \; .$$

Proof. From the decomposition

$$\frac{M \, p}{e^* \, M \, p} - \frac{M \, q}{e^* \, M \, q} = \frac{M \, (p-q)}{e^* \, M \, p} - \frac{e^* \, M \, (p-q)}{e^* \, M \, p} \; \frac{M \, q}{e^* \, M \, q} \; ,$$

and the bounds in (20), it follows that

$$\|\frac{M \, p}{e^* \, M \, p} - \frac{M \, q}{e^* \, M \, q}\| \leq \frac{\|M \, (p-q)\|}{e^* \, M \, p} + |\frac{e^* \, M \, (p-q)}{e^* \, M \, p}| \leq 2 \, \omega(M) \; \|p - q\| \; . \qquad \square$$

Example 2 (continued)

Under the assumptions of Example 2 in Appendix A, the following estimates hold.

Lemma C.3 *For any integers n, l such that $l \leq n$*

$$\omega(M_{n,l}) \leq \delta_l^B \cdots \delta_n^B \ .$$

If $l \leq n-r+1$, then

$$\omega(M_{n,l}) \leq \varepsilon^{-r} \ \delta_l^B \cdots \delta_{l+r-1}^B \ .$$

Proof. In the general case, for any $i, i' \in S$

$$\frac{M_{n,l}^{\bullet,i}}{M_{n,l}^{\bullet,i'}} = \frac{\displaystyle\sum_{j \in S} M_{n,l}^{j,i}}{\displaystyle\sum_{j \in S} M_{n,l}^{j,i'}} \quad = \quad \frac{\displaystyle\sum_{i_{l+1},\cdots,i_n,j \in S} q^{i,i_{l+1}} \, b_l^i \cdots q^{i_n,j} \, b_n^{i_n}}{\displaystyle\sum_{i_{l+1},\cdots,i_n,j \in S} q^{i',i_{l+1}} \, b_l^{i'} \cdots q^{i_n,j} \, b_n^{i_n}}$$

$$\leq \quad \delta_l^B \cdots \delta_n^B \ \frac{\displaystyle\sum_{i_{l+1},\cdots,i_n,j \in S} q^{i,i_{l+1}} \cdots q^{i_n,j}}{\displaystyle\sum_{i_{l+1},\cdots,i_n,j \in S} q^{i',i_{l+1}} \cdots q^{i_n,j}} = \delta_l^B \cdots \delta_n^B \ .$$

In the particular case where $l \leq n-r+1$, for any $i, i' \in S$

$$\frac{M_{n,l}^{\bullet,i}}{M_{n,l}^{\bullet,i'}} = \frac{\displaystyle\sum_{j \in S} M_{n,l+r}^{\bullet,j} \, M_{l+r-1,l}^{j,i}}{\displaystyle\sum_{j \in S} M_{n,l+r}^{\bullet,j} \, M_{l+r-1,l}^{j,i'}} \leq \frac{\displaystyle\max_{i,j \in S} M_{l+r-1,l}^{j,i}}{\displaystyle\min_{i,j \in S} M_{l+r-1,l}^{j,i}} \leq \varepsilon^{-r} \ \delta_l^B \cdots \delta_{l+r-1}^B$$

(if $l = n-r+1$, then $M_{n,l+r} = M_{n,n+1} = I$ by convention). □

Remark C.4 These estimates can be merged into the following single estimate

$$\omega(M_{n,l}) \leq \varepsilon^{-r} \ \delta_l^B \cdots \delta_{\min(l+r-1,n)}^B \ ,$$

which holds for any integers n, l such that $l \leq n$.

D Technical proofs

Proof of Proposition 4.1

For any $z \in \mathbf{R}^N$, the identity (10) and the estimate in Proposition 2.1 yield

$$\begin{aligned}
\| \prod_{k=l}^{n} \Phi_k[y_k, p_k] \, z \| \quad &\leq \quad \frac{e^* \, M_{n,l} \, z^+}{e^* \, M_{n,l} \, p_l} \, \| \, f_{n,l}[y_n, \cdots, y_l, \frac{z^+}{\|z^+\|}] - f_{n,l}[y_n, \cdots, y_l, p_l] \, \| \\
&\quad + \frac{e^* \, M_{n,l} \, z^-}{e^* \, M_{n,l} \, p_l} \, \| \, f_{n,l}[y_n, \cdots, y_l, \frac{z^-}{\|z^-\|}] - f_{n,l}[y_n, \cdots, y_l, p_l] \, \| \\
&\leq \quad [\varepsilon_l^{-1} \, \delta_l(y_l) \, \|z^+\|] \, [2 \, \varepsilon_l^{-1} \, \delta_l(y_l) \, \prod_{k=l}^{n} (1-\varepsilon_k) \, \| \frac{z^+}{\|z^+\|} - p_l \|] \\
&\quad + [\varepsilon_l^{-1} \, \delta_l(y_l) \, \|z^-\|] \, [2 \, \varepsilon_l^{-1} \, \delta_l(y_l) \, \prod_{k=l}^{n} (1-\varepsilon_k) \, \| \frac{z^-}{\|z^-\|} - p_l \|] \\
&\leq \quad 4 \, [\varepsilon_l^{-1} \, \delta_l(y_l)]^2 \, \prod_{k=l}^{n} (1-\varepsilon_k) \, \|z\| \ ,
\end{aligned}$$

hence

$$\| \prod_{k=l}^{n} \Phi_k[y_k, p_k] \| \leq 4 \, [\varepsilon_l^{-1} \, \delta_l(y_l)]^2 \prod_{k=l}^{n} (1 - \varepsilon_k) \; .$$

For any $z \in \mathbf{R}^N$, the identity (11) and the estimate in Proposition 2.1 yield :

- If $l \geq m+1$, then

$$\| \prod_{k=l}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \, z \| \leq$$

$$\leq [\varepsilon_l^{-1} \, \delta_l(y_l) \, \|z\|] \, [2 \, \varepsilon_m^{-1} \, \delta_m(y_m) \prod_{k=m}^{n} (1 - \varepsilon_k) \, \|p - p'\|]$$

$$+ [\varepsilon_l^{-1} \, \delta_l(y_l)] \, [2 \, \varepsilon_m^{-1} \, \delta_m(y_m) \prod_{k=m}^{l-1} (1 - \varepsilon_k) \, \|p - p'\|] \, [4 \, [\varepsilon_l^{-1} \, \delta_l(y_l)]^2 \prod_{k=l}^{n} (1 - \varepsilon_k) \, \|z\|]$$

$$\leq 2 \, \varepsilon_m^{-1} \, \delta_m(y_m) \, \varepsilon_l^{-1} \, \delta_l(y_l) \, [1 + 4 \, [\varepsilon_l^{-1} \, \delta_l(y_l)]^2] \prod_{k=m}^{n} (1 - \varepsilon_k) \, \|p - p'\| \, \|z\| \; ,$$

hence

$$\| \prod_{k=l}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \| \leq 10 \, \varepsilon_m^{-1} \, \delta_m(y_m) \, [\varepsilon_l^{-1} \, \delta_l(y_l)]^3 \prod_{k=m}^{n} (1 - \varepsilon_k) \, \|p - p'\| \; .$$

- Otherwise if $l = m$, then

$$\| \prod_{k=m}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \, z \| \leq$$

$$\leq [\varepsilon_m^{-1} \, \delta_m(y_m) \, \|z\|] \, [2 \, \varepsilon_m^{-1} \, \delta_m(y_m) \prod_{k=m}^{n} (1 - \varepsilon_k) \, \|p - p'\|]$$

$$+ [\varepsilon_m^{-1} \, \delta_m(y_m) \, \|p - p'\|] \, [4 \, [\varepsilon_m^{-1} \, \delta_m(y_m)]^2 \prod_{k=m}^{n} (1 - \varepsilon_k) \, \|z\|]$$

$$\leq 2 \, [\varepsilon_m^{-1} \, \delta_m(y_m)]^2 \, [1 + 2 \, \varepsilon_m^{-1} \, \delta_m(y_m)] \prod_{k=m}^{n} (1 - \varepsilon_k) \, \|p - p'\| \, \|z\| \; ,$$

hence

$$\| \prod_{k=m}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \| \leq 6 \, [\varepsilon_m^{-1} \, \delta_m(y_m)]^3 \prod_{k=m}^{n} (1 - \varepsilon_k) \, \|p - p'\| \; .$$

With the decomposition introduced in (9) above

$$\|F_{n,m}[y_n, \cdots, y_m, p, w] - F_{n,m}[y_n, \cdots, y_m, p', w'] \| \leq$$

$$\leq \| \prod_{k=m}^{n} \Phi_k[y_k, p_k] \| \, \|w - w'\| + \| \prod_{k=m}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \| \, \|w'\|$$

$$+ \sum_{l=m}^{n-1} \mathrm{Lip}(u_l, y_l) \, \|p_l - p'_l\| \, \| \prod_{k=l+1}^{n} \Phi_k[y_k, p_k] \| + \mathrm{Lip}(u_n, y_n) \, \|p_n - p'_n\|$$

$$+ \sum_{l=m}^{n-1} \mathrm{K}(u_l, y_l) \, \| \prod_{k=l+1}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \|$$

$$\le [\,4\,[\varepsilon_m^{-1}\,\delta_m(y_m)\,]^2 \prod_{k=m}^{n}(1-\varepsilon_k)\,]\,\|w-w'\| + [\,6\,[\varepsilon_m^{-1}\,\delta_m(y_m)\,]^3 \prod_{k=m}^{n}(1-\varepsilon_k)\,\|p-p'\|\,]\,\|w'\|$$

$$+ \mathrm{Lip}(u_m, y_m)\,\|p-p'\|\,[\,4\,[\varepsilon_{m+1}^{-1}\,\delta_{m+1}(y_{m+1})\,]^2 \prod_{k=m+1}^{n}(1-\varepsilon_k)\,]$$

$$+ \sum_{l=m+1}^{n-1} \mathrm{Lip}(u_l, y_l)\,[\,2\,\varepsilon_m^{-1}\,\delta_m(y_m) \prod_{k=m}^{l-1}(1-\varepsilon_k)\,\|p-p'\|\,]\,[\,4\,[\varepsilon_{l+1}^{-1}\,\delta_{l+1}(y_{l+1})\,]^2 \prod_{k=l+1}^{n}(1-\varepsilon_k)\,]$$

$$+ \mathrm{Lip}(u_n, y_n)\,[\,2\,\varepsilon_m^{-1}\,\delta_m(y_m) \prod_{k=m}^{n-1}(1-\varepsilon_k)\,\|p-p'\|\,]$$

$$+ \sum_{l=m}^{n-1} \mathrm{K}(u_l, y_l)\,[\,10\,\varepsilon_m^{-1}\,\delta_m(y_m)\,[\varepsilon_{l+1}^{-1}\,\delta_{l+1}(y_{l+1})\,]^3 \prod_{k=m}^{n}(1-\varepsilon_k)\,\|p-p'\|\,]$$

$$\le 6\,[\varepsilon_m^{-1}\,\delta_m(y_m)\,]^3 \prod_{k=m}^{n}(1-\varepsilon_k)\,[\,\|w-w'\| + \|p-p'\|\,(1+\|w\|+\|w'\|\,)\,]$$

$$+ 10\,\varepsilon_m^{-1}\,\delta_m(y_m)\;[\,\sum_{l=m}^{n-1}(1-\varepsilon_l)^{-1}\,[\varepsilon_{l+1}^{-1}\,\delta_{l+1}(y_{l+1})\,]^2\,\mathrm{Lip}(u_l, y_l) + (1-\varepsilon_n)^{-1}\,\mathrm{Lip}(u_n, y_n)$$

$$+ \sum_{l=m}^{n-1}[\varepsilon_{l+1}^{-1}\,\delta_{l+1}(y_{l+1})\,]^3\,\mathrm{K}(u_l, y_l)\,]\prod_{k=m}^{n}(1-\varepsilon_k)\,\|p-p'\|\ .$$

This concludes the proof. □

Proof of Theorem 4.6

Recall the following estimate, which is an immediate application of Proposition A.1 : For any $p, p' \in \mathcal{P}(S)$, any integers n, m such that $n \ge m$, and any sequence $y_m, \cdots, y_n \in \mathbf{R}^d$

$$\|f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p']\,\| \le 2\,\omega(M_{n,m})\;\tau_{\mathrm{B}}(M_{n,m})\;\|p-q\|\ ,$$

where the Birkhoff contraction coefficient $\tau_{\mathrm{B}}(M_{n,m})$ is defined in Appendix A, and the coefficient $\omega(M_{n,m})$ satisfies the following estimate, which is an immediate application of Lemma C.3.

Lemma D.1 *If the stochastic matrix Q is primitive, with index of primitivity r, then for any integers n, l such that $l \le n$, any $p \in \mathcal{P}(S)$, and any $z \in \mathbf{R}^N$*

$$|\frac{e^*\,M_{n,l}\,z}{e^*\,M_{n,l}\,p}| \le \frac{\|M_{n,l}\,z\|}{e^*\,M_{n,l}\,p} \le \omega(M_{n,l})\;\|z\|\ ,$$

and

$$\omega(M_{n,l}) \le \delta(y_l)\cdots\delta(y_n)\ .$$

If $l \le n-r+1$, then

$$\omega(M_{n,l}) \le \varepsilon^{-r}\,\delta(y_l)\cdots\delta(y_{l-r+1})\ .$$

Remark D.2 These estimates can be merged into the following single estimate

$$\omega(M_{n,l}) \le \varepsilon^{-r}\;\delta(y_l)\cdots\delta(y_{\min(l+r-1,n)})\ ,$$

which holds for any integers n, l such that $l \le n$.

For any $z \in \mathbf{R}^N$, the identity (10) and the estimates in Proposition 2.1 and in Lemma D.1 yield

$$\|\prod_{k=l}^{n}\Phi[y_k, p_k]\;z\| \quad\le\quad \frac{e^*\,M_{n,l}\,z^+}{e^*\,M_{n,l}\,p_l}\,\|\,f[y_n, \cdots, y_l, \frac{z^+}{\|z^+\|}] - f[y_n, \cdots, y_l, p_l]\,\|$$

$$
\begin{aligned}
& + \frac{e^* \, M_{n,l} \, z^-}{e^* \, M_{n,l} \, p_l} \, \| \, f[y_n, \cdots, y_l, \frac{z^-}{\|z^-\|}] - f[y_n, \cdots, y_l, p_l] \, \| \\
\leq \quad & [\, \omega(M_{n,l}) \, \|z^+\| \,] \, [2 \, \omega(M_{n,l}) \, \tau_{\mathrm{B}}(M_{n,l}) \, \| \frac{z^+}{\|z^+\|} - p_l \| \,] \\
& [\, \omega(M_{n,l}) \, \|z^-\| \,] \, [2 \, \omega(M_{n,l}) \, \tau_{\mathrm{B}}(M_{n,l}) \, \| \frac{z^-}{\|z^-\|} - p_l \| \,] \\
\leq \quad & 4 \, \omega^2(M_{n,l}) \, \tau_{\mathrm{B}}(M_{n,l}) \, \|z\| \; ,
\end{aligned}
$$

hence

$$
\| \prod_{k=l}^{n} \Phi[y_k, p_k] \, \| \leq 4 \, \omega^2(M_{n,l}) \, \tau_{\mathrm{B}}(M_{n,l}) \; .
$$

For any $z \in \mathbf{R}^N$, the identity (11) and the estimates in Proposition 2.1 and in Lemma D.1 yield :

- If $l \geq m+1$, then

$$
\begin{aligned}
& \| \prod_{k=l}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \, z \| \leq \\
& \quad \leq [\omega(M_{n,l}) \, \|z\| \,] \, [2 \, \omega(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m}) \, \|p - p'\| \,] \\
& \quad\quad + \omega(M_{n,l}) \, [2 \, \omega(M_{l-1,m}) \, \tau_{\mathrm{B}}(M_{l-1,m}) \, \|p - p'\| \,] \, [4 \, \omega^2(M_{n,l}) \, \tau_{\mathrm{B}}(M_{n,l}) \, \|z\| \,] \\
& \quad \leq 2 \, \omega(M_{n,l}) \, [\omega(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m}) \\
& \quad\quad\quad + 4 \, \omega^2(M_{n,l}) \, \omega(M_{l-1,m}) \, \tau_{\mathrm{B}}(M_{n,l}) \, \tau_{\mathrm{B}}(M_{l-1,m}) \,] \, \|p - p'\| \, \|z\| \; ,
\end{aligned}
$$

  hence

$$
\begin{aligned}
& \| \prod_{k=l}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \, \| \leq \\
& \quad \leq 2 \, \omega(M_{n,l}) \, [\omega(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m}) \\
& \quad\quad\quad + 4 \, \omega^2(M_{n,l}) \, \omega(M_{l-1,m}) \, \tau_{\mathrm{B}}(M_{n,l}) \, \tau_{\mathrm{B}}(M_{l-1,m}) \,] \, \|p - p'\| \; .
\end{aligned}
$$

- Otherwise if $l = m$, then

$$
\begin{aligned}
& \| \prod_{k=m}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \, z \| \leq \\
& \quad \leq [\omega(M_{n,m}) \, \|z\| \,] \, [2 \, \omega(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m}) \, \|p - p'\| \,] \\
& \quad\quad + \omega(M_{n,m}) \, \|p - p'\| \, [4 \, \omega^2(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m}) \, \|z\| \,] \\
& \quad \leq 2 \, \omega^2(M_{n,m}) \, [1 + 2 \, \omega(M_{n,m}) \,] \, \tau_{\mathrm{B}}(M_{n,m}) \, \|p - p'\| \, \|z\| \; ,
\end{aligned}
$$

  hence

$$
\| \prod_{k=m}^{n} [\, \Phi_k[y_k, p_k] - \Phi_k[y_k, p'_k] \,] \, \| \leq 6 \, \omega^3(M_{n,m}) \; \tau_{\mathrm{B}}(M_{n,m}) \, \|p - p'\| \; .
$$

Making use of the above estimates yields

$$\sum_{l=m}^{n-1} \| \prod_{k=l+1}^{n} \Phi[y_k, p_k] \| \, \mathrm{Lip}(u, y_l) \, \|p_l - p'_l\| \leq$$

$$\leq \mathrm{Lip}(u, y_m) \, [4\,\omega^2(M_{n,m+1}) \, \tau_{\mathrm{B}}(M_{n,m+1})] \, \|p - p'\| \,]$$

$$+ \sum_{l=m+1}^{n-1} \mathrm{Lip}(u, y_l) \, [4\,\omega^2(M_{n,l+1}) \, \tau_{\mathrm{B}}(M_{n,l+1}) \,] \, [2\,\omega(M_{l-1,m}) \, \tau_{\mathrm{B}}(M_{l-1,m}) \, \|p - p'\| \,]$$

$$\leq 4 \, \mathrm{Lip}(u, y_m) \, \omega^2(M_{n,m+1}) \, \tau_{\mathrm{B}}(M_{n,m+1}) \, \|p - p'\|$$

$$+ 8 \sum_{l=m+1}^{n-1} \mathrm{Lip}(u, y_l) \, \omega^2(M_{n,l+1}) \, \omega(M_{l-1,m}) \, \tau_{\mathrm{B}}(M_{n,l+1}) \, \tau_{\mathrm{B}}(M_{l-1,m}) \, \|p - p'\| \; ,$$

and similarly

$$\sum_{l=m}^{n-1} \| \prod_{k=l+1}^{n} [\, \Phi[y_k, p_k] - \Phi[y_k, p'_k] \,] \, \| \, \mathrm{K}(u, y_l) \leq$$

$$\leq 2 \sum_{l=m}^{n-1} \mathrm{K}(u, y_l) \, \omega(M_{n,l+1}) \, [\omega(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m})$$

$$+ 4\,\omega^2(M_{n,l+1}) \, \omega(M_{l,m}) \, \tau_{\mathrm{B}}(M_{n,l+1}) \, \tau_{\mathrm{B}}(M_{l,m}) \,] \, \|p - p'\| \; .$$

With the decomposition introduced in (9) above

$$\|F[y_n, \cdots, y_m, p, w] - F[y_n, \cdots, y_m, p', w'] \, \| \leq$$

$$\leq \| \prod_{k=m}^{n} \Phi[y_k, p_k] \, \| \, \|w - w'\| + \| \prod_{k=m}^{n} [\, \Phi[y_k, p_k] - \Phi[y_k, p'_k] \,] \, \| \, \|w'\|$$

$$+ \sum_{l=m}^{n-1} \| \prod_{k=l+1}^{n} \Phi[y_k, p_k] \, \| \, \mathrm{Lip}(u, y_l) \, \|p_l - p'_l\| + \mathrm{Lip}(u, y_n) \, \|p_n - p'_n\|$$

$$+ \sum_{l=m}^{n-1} \| \prod_{k=l+1}^{n} [\, \Phi[y_k, p_k] - \Phi[y_k, p'_k] \,] \, \| \, \mathrm{K}(u, y_l)$$

$$\leq 4\,\omega^2(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m}) \, \|w - w'\|$$

$$+ 6\,\omega^3(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m}) \, \|p - p'\| \, \|w'\|$$

$$+ 4 \, \mathrm{Lip}(u, y_m) \, \omega^2(M_{n,m+1}) \, \tau_{\mathrm{B}}(M_{n,m+1}) \, \|p - p'\|$$

$$+ 8 \sum_{l=m+1}^{n-1} \mathrm{Lip}(u, y_l) \, \omega^2(M_{n,l+1}) \, \omega(M_{l-1,m}) \, \tau_{\mathrm{B}}(M_{n,l+1}) \, \tau_{\mathrm{B}}(M_{l-1,m}) \, \|p - p'\|$$

$$+ 2 \, \mathrm{Lip}(u, y_n) \, \omega(M_{n-1,m}) \, \tau_{\mathrm{B}}(M_{n-1,m}) \, \|p - p'\|$$

$$+ 2 \sum_{l=m}^{n-1} \mathrm{K}(u, y_l) \, \omega(M_{n,l+1}) \, [\omega(M_{n,m}) \, \tau_{\mathrm{B}}(M_{n,m})$$

$$+ 4\,\omega^2(M_{n,l+1})\,\omega(M_{l,m})\,\tau_{\rm B}(M_{n,l+1})\,\tau_{\rm B}(M_{l,m})\,]\,\|p-p'\|$$

$$\leq 6\,\omega^3(M_{n,m})\;\tau_{\rm B}(M_{n,m})\;[\,\|w-w'\|+\|p-p'\|\;(1+\|w\|+\|w'\|\,)\,]$$

$$+\;[\;4\,\mathrm{Lip}(u,y_m)\;\omega^2(M_{n,m+1})\;\tau_{\rm B}(M_{n,m+1})$$

$$+\,8\sum_{l=m+1}^{n-1}\;\mathrm{Lip}(u,y_l)\;\omega^2(M_{n,l+1})\,\omega(M_{l-1,m})\,\tau_{\rm B}(M_{n,l+1})\,\tau_{\rm B}(M_{l-1,m})$$

$$+\,2\,\mathrm{Lip}(u,y_n)\;\omega(M_{n-1,m})\;\tau_{\rm B}(M_{n-1,m})$$

$$+\,2\sum_{l=m}^{n-1}\;\mathrm{K}(u,y_l)\;\omega(M_{n,l+1})\;[\omega(M_{n,m})\;\tau_{\rm B}(M_{n,m})$$

$$+\,4\,\omega^2(M_{n,l+1})\,\omega(M_{l,m})\,\tau_{\rm B}(M_{n,l+1})\,\tau_{\rm B}(M_{l,m})\,]\,]\;\;\|p-p'\|\;.$$

The estimates obtained in Lemma A.9 yield

$$\|F[y_n,\cdots,y_m,p,w]-F[y_n,\cdots,y_m,p',w']\,\|\leq$$

$$\leq 6\,\omega^3(M_{n,m})\;\prod_{\kappa=0}^{[n,m]}(1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)$$

$$[\,\|w-w'\|+\|p-p'\|\;(1+\|w\|+\|w'\|\,)\,]$$

$$+\;[\;4\,(1-\varepsilon^r)^{-1}\;\mathrm{Lip}(u,y_m)\;\omega^2(M_{n,m+1})$$

$$+\,8\,(1-\varepsilon^r)^{-1}\;\sum_{l=m+1}^{n-1}\;\mathrm{Lip}(u,y_l)\;\omega^2(M_{n,l+1})\,\omega(M_{l-1,m})$$

$$+\,2\,(1-\varepsilon^r)^{-1}\;\mathrm{Lip}(u,y_n)\;\omega(M_{n-1,m})$$

$$+\,2\sum_{l=m}^{n-1}\;\mathrm{K}(u,y_l)\;\omega(M_{n,l+1})\;[\,\omega(M_{n,m})+4\,(1-\varepsilon^r)^{-1}\;\omega^2(M_{n,l+1})\,\omega(M_{l,m})\,]\,]$$

$$\prod_{\kappa=0}^{[n,m]}(1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\;\|p-p'\|$$

$$\leq 6\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]^3\;\prod_{\kappa=0}^{[n,m]}(1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)$$

$$[\,\|w-w'\|+\|p-p'\|\;(1+\|w\|+\|w'\|\,)\,]$$

$$+\,10\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,(1-\varepsilon^r)^{-1}$$

$$[\,\sum_{l=m}^{n-1}\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{\min(l+r,n)})\,]^2\;\mathrm{Lip}(u,y_l)+\mathrm{Lip}(u,y_n)$$

$$+ \sum_{l=m}^{n-1} [\, \varepsilon^{-r}\, \delta(y_{l+1}) \cdots \delta(y_{\min(l+r,n)})\,]^3\, \mathrm{K}(u, y_l)\,]$$

$$\prod_{\kappa=0}^{[n,m]} (1 - \varepsilon^r\, [\, \delta(y_{m+\kappa r+1}) \cdots \delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\, \|p - p'\| \; .$$

This finishes the proof. □

Proof of Proposition 5.6

For any sequence $i_m, \cdots, i_{n+1} \in S$

$$\int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} |g^{i_{n+1}}(y_{n+1}, f \otimes F[y_n, \cdots, y_m, p, w]) - g^{i_{n+1}}(y_{n+1}, f \otimes F[y_n, \cdots, y_m, p', w'])\,|$$

$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1}) \leq$$

$$\leq \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \mathrm{Lip}(g^{i_{n+1}}, y_{n+1})\, [\, \|F[y_n, \cdots, y_m, p, w] - F[y_n, \cdots, y_m, p', w']\,\|$$

$$+ \|f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p']\,\|$$

$$(1 + \|F[y_n, \cdots, y_m, p, w]\| + \|F[y_n, \cdots, y_m, p', w']\|\,)\,]$$

$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_{n+1}}(y_{n+1})\, \lambda(dy_m) \cdots \lambda(dy_{n+1})$$

$$\leq \mathrm{Lip}(g) \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \|F[y_n, \cdots, y_m, p, w] - F[y_n, \cdots, y_m, p', w']\,\|$$

$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_m) \cdots \lambda(dy_n)$$

$$+ \mathrm{Lip}(g) \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \|f[y_n, \cdots, y_m, p] - f[y_n, \cdots, y_m, p']\,\|$$

$$(1 + \|F[y_n, \cdots, y_m, p, w]\,\| + \|F[y_n, \cdots, y_m, p', w']\,\|\,)$$

$$b_\bullet^{i_m}(y_m) \cdots b_\bullet^{i_n}(y_n)\, \lambda(dy_m) \cdots \lambda(dy_n) \; .$$

□ To estimate the first term, recall from the proof of Theorem 4.6 above, that

$$\|F[y_n, \cdots, y_m, p, w] - F[y_n, \cdots, y_m, p', w']\,\| \leq$$

$$\leq 6\, \omega^3(M_{n,m})\; \tau_{\mathrm{B}}(M_{n,m})\; [\, \|w - w'\| + \|p - p'\|\; (1 + \|w\| + \|w'\|\,)\,]$$

$$+ \Big[\; 4\, \mathrm{Lip}(u, y_m)\; \omega^2(M_{n,m+1})\, \tau_{\mathrm{B}}(M_{n,m+1})$$

$$+ 8 \sum_{l=m+1}^{n-1} \mathrm{Lip}(u, y_l)\; \omega^2(M_{n,l+1})\, \omega(M_{l-1,m})\; \tau_{\mathrm{B}}(M_{n,l+1})\; \tau_{\mathrm{B}}(M_{l-1,m})$$

$$+ 2\, \mathrm{Lip}(u, y_n)\; \omega(M_{n-1,m})\; \tau_{\mathrm{B}}(M_{n-1,m})$$

$$+\,2 \sum_{l=m}^{n-1} \mathrm{K}(u,y_l)\,\omega(M_{n,l+1})\,[\omega(M_{n,m})\,\tau_{\mathrm{B}}(M_{n,m})$$

$$+\,4\,\omega^2(M_{n,l+1})\,\omega(M_{l,m})\,\tau_{\mathrm{B}}(M_{n,l+1})\,\tau_{\mathrm{B}}(M_{l,m})\,]\,\big]\ \|p-p'\|$$

$$\leq 6\,\omega^3(M_{n,m})\ \tau_{\mathrm{B}}(M_{n,m})\ [\,\|w-w'\| + \|p-p'\|\ (1+\|w\|+\|w'\|\,)\,]$$

$$+\ \big[\ 4\,\mathrm{Lip}(u,y_m)\ \omega^2(M_{n,m+1})\,\tau_{\mathrm{B}}(M_{n,m+1})$$

$$+\,8 \sum_{l=m+1}^{n-1} \mathrm{Lip}(u,y_l)\ \omega^2(M_{n,l+1})\,\omega(M_{l-1,m})\,\tau_{\mathrm{B}}(M_{n,l+1})\,\tau_{\mathrm{B}}(M_{l-1,m})$$

$$+\,2\,\mathrm{Lip}(u,y_n)\ \omega(M_{n-1,m})\ \tau_{\mathrm{B}}(M_{n-1,m})$$

$$+\,2\,\mathrm{K}(u,y_m)\ \omega(M_{n,m+1})\ [\omega(M_{n,m}) + 4\,\omega^2(M_{n,m+1})\,\omega(M_m)\,]\ \tau_{\mathrm{B}}(M_{n,m+1})$$

$$+\,2 \sum_{l=m+1}^{n-1} \mathrm{K}(u,y_l)\ \omega(M_{n,l+1})\ [\omega(M_{n,m})$$

$$+\,4\,\omega^2(M_{n,l+1})\,\omega(M_{l,m})\,]\ \tau_{\mathrm{B}}(M_{n,l+1})\,\tau_{\mathrm{B}}(M_{l-1,m})\ \big]\ \|p-p'\|\ .$$

If $n \geq m+4r$, and if the estimates obtained in Proposition A.5, and especially in (18), are used in place of the estimates obtained in Lemma A.9, then depending on the position of l relative to m and n :

- If $m+1 \leq l \leq m+2r-1$, then

$$\begin{aligned}
\tau_{\mathrm{B}}(M_{l-1,m}) &\leq 1\ ,\\
\tau_{\mathrm{B}}(M_{n,l+1}) &\leq \prod_{\kappa=1}^{[n,l+1]} (1-\varepsilon^r\,[\,\delta(y_{l+\kappa r+2})\cdots\delta(y_{l+(\kappa+1)r})\,]^{-1}\,)\ .
\end{aligned}$$

- Else if $m+2r \leq l \leq n-2r$, then

$$\begin{aligned}
\tau_{\mathrm{B}}(M_{l-1,m}) &\leq \prod_{\kappa=1}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\ ,\\
\tau_{\mathrm{B}}(M_{n,l+1}) &\leq \prod_{\kappa=1}^{[n,l+1]} (1-\varepsilon^r\,[\,\delta(y_{l+\kappa r+2})\cdots\delta(y_{l+(\kappa+1)r})\,]^{-1}\,)\ .
\end{aligned}$$

- Otherwise if $n-2r+1 \leq l \leq n$, then

$$\begin{aligned}
\omega(M_{n,l+1}) &\leq \delta(y_{l+1})\cdots\delta(y_{\min(l+r,n)})\ ,\\
\tau_{\mathrm{B}}(M_{l-1,m}) &\leq \prod_{\kappa=1}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\ ,\\
\tau_{\mathrm{B}}(M_{n,l+1}) &\leq 1\ .
\end{aligned}$$

These estimates yield

$$\|F[y_n,\cdots,y_m,p,w] - F[y_n,\cdots,y_m,p',w']\,\| \leq$$

$$\leq 6\,[\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]^3 \prod_{\kappa=1}^{[n,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)$$

$$[\,\|w-w'\|+\|p-p'\|\,(1+\|w\|+\|w'\|\,)\,]$$

$$+\,\Big[\,4\,\mathrm{Lip}(u,y_m)\,[\,\varepsilon^{-r}\,\delta(y_{m+1})\cdots\delta(y_{m+r})\,]^2 \prod_{\kappa=1}^{[n,m+1]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+2})\cdots\delta(y_{m+(\kappa+1)r})\,]^{-1}\,)$$

$$+\,8\sum_{l=m+1}^{m+2r-1} \mathrm{Lip}(u,y_l)\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{l+r})\,]^2\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{\min(m+r-1,l-1)})\,]$$

$$\prod_{\kappa=1}^{[n,l+1]} (1-\varepsilon^r\,[\,\delta(y_{l+\kappa r+2})\cdots\delta(y_{l+(\kappa+1)r})\,]^{-1}\,)$$

$$+\,8\sum_{l=m+2r}^{n-2r} \mathrm{Lip}(u,y_l)\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{l+r})\,]^2\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]$$

$$\prod_{\kappa=1}^{[n,l+1]} (1-\varepsilon^r\,[\,\delta(y_{l+\kappa r+2})\cdots\delta(y_{l+(\kappa+1)r})\,]^{-1}\,)$$

$$\prod_{\kappa=1}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)$$

$$+\,8\sum_{l=n-2r+1}^{n-1} \mathrm{Lip}(u,y_l)\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{\min(l+r,n)})\,]^2\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]$$

$$\prod_{\kappa=1}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)$$

$$+\,2\,\mathrm{Lip}(u,y_n)\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]$$

$$\prod_{\kappa=1}^{[n-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)$$

$$+\,2\,\mathrm{K}(u,y_m)\,[\,\varepsilon^{-r}\,\delta(y_{m+1})\cdots\delta(y_{m+r})\,]\,[\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]$$

$$+\,4\,[\,\varepsilon^{-r}\,\delta(y_{m+1})\cdots\delta(y_{m+r})\,]^2\,\delta(y_m)\,]$$

$$\prod_{\kappa=1}^{[n,m+1]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+2})\cdots\delta(y_{m+(\kappa+1)r})\,]^{-1}\,)$$

$$+\,2\sum_{l=m+1}^{m+2r-1} \mathrm{K}(u,y_l)\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{l+r})\,]\,[\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]$$

$$+\,4\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{l+r})\,]^2\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{\min(m+r-1,l)})\,]\,]$$

$$\prod_{\kappa=1}^{[n,l+1]} (1-\varepsilon^r\,[\,\delta(y_{l+\kappa r+2})\cdots\delta(y_{l+(\kappa+1)r})\,]^{-1}\,)$$

$$+\,2 \sum_{l=m+2r}^{n-2r} \mathrm{K}(u,y_l)\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{l+r})\,]\,[\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]$$

$$+\,4\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{l+r})\,]^2\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]\,]$$

$$\prod_{\kappa=1}^{[n,l+1]} (1-\varepsilon^r\,[\,\delta(y_{l+\kappa r+2})\cdots\delta(y_{l+(\kappa+1)r})\,]^{-1}\,)$$

$$\prod_{\kappa=1}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)$$

$$+\,2 \sum_{l=n-2r+1}^{n-1} \mathrm{K}(u,y_l)\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{\min(l+r,n)})\,]\,[\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]$$

$$+\,4\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{\min(l+r,n)})\,]^2\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]\,]$$

$$\prod_{\kappa=1}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\;\Big]\;\|p-p'\|\,.$$

Notice that each term in the sum above is a product of factors which span disjoint blocks of length r, with the only exception of the third sum from the end : indeed, the sets $\{m,\cdots,m+r-1\}$ and $\{l,l+1,\cdots,l+r\}$ are disjoint if $l\geq m+r$, but they do overlap if $m+1\leq l\leq m+r-1$. Except for this sum which requires special attention, integration is straightforward, and yields for any sequence $i_m,\cdots,i_n\in S$

$$\int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} \|F[y_n,\cdots,y_m,p,w]-F[y_n,\cdots,y_m,p',w']\,\|$$

$$b_\bullet^{i_m}(y_m)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_m)\cdots\lambda(dy_n)\leq$$

$$\leq 6\,\varepsilon^{-3r}\,\Delta_3^r\,(1-R)^{[n,m]}\,[\,\|w-w'\|+\|p-p'\|\,(1+\|w\|+\|w'\|\,)\,]$$

$$+\,\Big[\;4\,\mathrm{Lip}(u)\,\varepsilon^{-2r}\,\Delta_2^r\,(1-R)^{[n,m+1]}$$

$$+\,8\,\mathrm{Lip}(u)\,\varepsilon^{-3r}\,\Delta_2^r\,\Delta^r \sum_{l=m+1}^{m+2r-1} (1-R)^{[n,l+1]}$$

$$+\,8\,\mathrm{Lip}(u)\,\varepsilon^{-3r}\,\Delta_2^r\,\Delta^r \sum_{l=m+2r}^{n-2r} (1-R)^{[n,l+1]+[l-1,m]}$$

$$+\,8\,\mathrm{Lip}(u)\,\varepsilon^{-3r}\,\Delta_2^r\,\Delta^r \sum_{l=n-2r+1}^{n-1} (1-R)^{[l-1,m]}$$

$$+\,2\,\mathrm{Lip}(u)\,\varepsilon^{-r}\,\Delta^r\,(1-R)^{[n-1,m]}$$

$$+\,2\,[\Delta\cdot\mathrm{K}](u)\,[\,\varepsilon^{-2r}\,\Delta_2^{r-1}\,\Delta+4\,\varepsilon^{-3r}\,\Delta_3^r\,]\,(1-R)^{[n,m+1]}$$

$$+\,2\,[\Delta\cdot\mathrm{K}](u) \sum_{l=m+1}^{m+r-2} \Delta^{l-m}\,[\,\varepsilon^{-2r}\,\Delta_2^{m+r-l-1}\,\Delta^{l-m+1}$$

$$+\,4\,\varepsilon^{-4r}\,\Delta_4^{m+r-l-1}\,\Delta_3^{l-m+1}\,]\,(1-R)^{[n,l+1]}$$

$$
\begin{aligned}
&+ 2\,[\Delta \cdot \mathrm{K}](u)\,\Delta^{r-1}\,[\,\varepsilon^{-2r}\,\Delta^{r} + 4\,\varepsilon^{-4r}\,\Delta_3^{r}\,]\,(1-R)^{[n,m+r]} \\
&+ 2\,\mathrm{K}(u)\,[\,\varepsilon^{-2r}\,\Delta^{2r} + 4\,\varepsilon^{-4r}\,\Delta_3^{r}\,\Delta^{r}\,] \sum_{l=m+r-1}^{m+2r-1} (1-R)^{[n,l+1]} \\
&+ 2\,\mathrm{K}(u)\,[\,\varepsilon^{-2r}\,\Delta^{2r} + 4\,\varepsilon^{-4r}\,\Delta_3^{r}\,\Delta^{r}\,] \sum_{l=m+2r}^{n-2r} (1-R)^{[n,l+1]+[l-1,m]} \\
&+ 2\,\mathrm{K}(u)\,[\,\varepsilon^{-2r}\,\Delta^{2r} + 4\,\varepsilon^{-4r}\,\Delta_3^{r}\,\Delta^{r}\,] \sum_{l=n-2r+1}^{n-1} (1-R)^{[l-1,m]}\,]\;\|p-p'\|\;.
\end{aligned}
$$

Using the lower bound (2.4) yields

$$
\begin{aligned}
&\int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \|F[y_n,\cdots,y_m,p,w] - F[y_n,\cdots,y_m,p',w']\,\| \\
&\qquad\qquad b_\bullet^{i_m}(y_m)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_m)\cdots\lambda(dy_n) \leq \\
&\leq 6\,\varepsilon^{-3r}\,\Delta_3^{r}\,\rho_*^{n-m+1-2r}\,[\,\|w-w'\| + \|p-p'\|\,(1+\|w\|+\|w'\|\,)\,] \\
&\quad + 8\,\varepsilon^{-3r}\,\Delta_3^{r}\,\mathrm{Lip}(u)\,(n-m+1)\,\rho_*^{n-m-4r}\,\|p-p'\| \\
&\quad + 10\,\varepsilon^{-4r}\,\Delta_4^{r}\,[\Delta\cdot\mathrm{K}](u)\,(n-m+1)\,\rho_*^{n-m-4r}\,\|p-p'\| \\
&\leq C'\,\varepsilon^{-4r}\,(n-m+1)\,\rho_*^{n-m-4r}\,(1+\|w\|+\|w'\|\,)\;,
\end{aligned}
$$

where the constant C' depends only on r, Δ_4, $\mathrm{Lip}(u)$, and $[\Delta\cdot\mathrm{K}](u)$.

□ To estimate the second term, recall from the proof of Theorem 2.3 above, that

$$
\|f[y_n,\cdots,y_m,p] - f[y_n,\cdots,y_m,p']\,\| \leq 2\,\omega(M_{n,m})\,\tau_{\mathrm{B}}(M_{n,m})\,\|p-p'\|\;,
$$

for any $p,p' \in \mathcal{P}(S)$, and any sequence $y_m,\cdots,y_n \in \mathbf{R}^d$, whereas the variation of parameters formula yields directly the following rough estimate

$$
\begin{aligned}
\|F[y_n,\cdots,y_m,p,w]\,\| &\leq \|\prod_{k=m}^{n} \Phi[y_k,p_k]\,\|\,\|w\| + \sum_{l=m}^{n-1} \|\prod_{k=l+1}^{n} \Phi[y_k,p_k]\,\|\,\mathrm{K}(u,y_l) + \mathrm{K}(u,y_n) \\
&\leq 4\,\omega^2(M_{n,m})\,\tau_{\mathrm{B}}(M_{n,m})\,\|w\| \\
&\qquad + 4\sum_{l=m}^{n-1} \omega^2(M_{n,l+1})\,\tau_{\mathrm{B}}(M_{n,l+1})\,\mathrm{K}(u,y_l) + \mathrm{K}(u,y_n) \\
&\leq 4\,\omega^2(M_{n,m})\,\|w\| + 4\sum_{l=m}^{n-1} \omega^2(M_{n,l+1})\,\mathrm{K}(u,y_l) + \mathrm{K}(u,y_n)\;,
\end{aligned}
\tag{21}
$$

for any $p \in \mathcal{P}(S)$, any $w \in \Sigma$, and any sequence $y_m,\cdots,y_n \in \mathbf{R}^d$. Therefore

$$
\begin{aligned}
&\|f[y_n,\cdots,y_m,p] - f[y_n,\cdots,y_m,p']\,\|\;\|F[y_n,\cdots,y_m,p,w]\,\| \leq \\
&\leq [\;8\,\omega^3(M_{n,m})\,\tau_{\mathrm{B}}(M_{n,m})\,\|w\|
\end{aligned}
$$

$$+ 4\,\omega(M_{n,m}) \sum_{l=m}^{n-1} \mathrm{K}(u,y_l)\,\omega^2(M_{n,l+1})\,\tau_{\mathrm{B}}(M_{n,l+1})\,\tau_{\mathrm{B}}(M_{l-1,m})$$

$$+ 2\,\mathrm{K}(u,y_n)\,\omega(M_{n,m})\,\tau_{\mathrm{B}}(M_{n-1,m})\,]\;\|p-p'\|$$

$$\leq [\,8\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]^3 \prod_{\kappa=1}^{[n,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\;\|w\|$$

$$+ 4\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,] \sum_{l=m}^{m+2r-1} \mathrm{K}(u,y_l)\;[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{l+r})\,]^2$$

$$\prod_{\kappa=1}^{[n,l+1]} (1-\varepsilon^r\,[\,\delta(y_{l+\kappa r+2})\cdots\delta(y_{l+(\kappa+1)r})\,]^{-1}\,)\,]$$

$$+ 4\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,] \sum_{l=m+2r}^{n-2r} \mathrm{K}(u,y_l)\;[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{l+r})\,]^2$$

$$\prod_{\kappa=1}^{[n,l+1]} (1-\varepsilon^r\,[\,\delta(y_{l+\kappa r+2})\cdots\delta(y_{l+(\kappa+1)r})\,]^{-1}\,)\,]$$

$$\prod_{\kappa=1}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\,]$$

$$+ 4\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,] \sum_{l=n-2r+1}^{n-1} \mathrm{K}(u,y_l)\,[\,\varepsilon^{-r}\,\delta(y_{l+1})\cdots\delta(y_{\min(l+r,n)})\,]^2$$

$$\prod_{\kappa=1}^{[l-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\,]$$

$$+ 2\,\mathrm{K}(u,y_n)\,[\,\varepsilon^{-r}\,\delta(y_m)\cdots\delta(y_{m+r-1})\,]$$

$$\prod_{\kappa=1}^{[n-1,m]} (1-\varepsilon^r\,[\,\delta(y_{m+\kappa r+1})\cdots\delta(y_{m+(\kappa+1)r-1})\,]^{-1}\,)\;]\;\;\|p-p'\|\;.$$

Notice again that each term in the sum above is a product of factors which span disjoint blocks of length r, with the only exception of the fourth sum from the end : indeed, the sets $\{m,\cdots,m+r-1\}$ and $\{l,l+1,\cdots,l+r\}$ are disjoint if $l \geq m+r$, but they do overlap if $m+1 \leq l \leq m+r-1$. Except for this sum which requires special attention, integration is straightforward, and yields for any sequence $i_m,\cdots,i_n \in S$

$$\int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} \|f[y_n,\cdots,y_m,p]-f[y_n,\cdots,y_m,p']\|\;\|F[y_n,\cdots,y_m,p,w]\|$$

$$b_\bullet^{i_m}(y_m)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_m)\cdots\lambda(dy_n) \leq$$

$$\leq 8\,\varepsilon^{-3r}\,\Delta_3^r\,(1-R)^{[n,m]}\;\|p-p'\|\;\|w\|$$

$$+ [\;4\,[\Delta\cdot\mathrm{K}](u)\,\varepsilon^{-3r} \sum_{l=m}^{m+r-2} \Delta^{l-m}\,\Delta_3^{m+r-1-l}\,\Delta_2^{l-m+1}\,(1-R)^{[n,l+1]}$$

$$+ 4\,[\Delta\cdot\mathrm{K}](u)\,\varepsilon^{-3r}\,\Delta_2^r\,\Delta^{r-1}\,(1-R)^{[n,m+r]}$$

$$+ 4\,\mathrm{K}(u)\,\varepsilon^{-3r}\,\Delta_2^r\,\Delta^r \sum_{l=m+r}^{m+2r-1} (1-R)^{[n,l+1]}$$

$$+ 4\,\mathrm{K}(u)\,\varepsilon^{-3r}\,\Delta_2^r\,\Delta^r \sum_{l=m+2r}^{n-2r} (1-R)^{[n,l+1]+[l-1,m]}$$

$$+ 4\,\mathrm{K}(u)\,\varepsilon^{-3r}\,\Delta_2^r\,\Delta^r \sum_{l=n-2r+1}^{n-1} (1-R)^{[l-1,m]}$$

$$+ 2\,\mathrm{K}(u)\,\varepsilon^{-r}\,\Delta^r\,(1-R)^{[n-1,m]}\,\Big]\;\|p-p'\|\;.$$

Using the lower bound (2.4) yields

$$\int_{\mathbf{R}^d}\cdots\int_{\mathbf{R}^d} \|f[y_n,\cdots,y_m,p] - f[y_n,\cdots,y_m,p']\|\;\|F[y_n,\cdots,y_m,p,w]\|$$

$$b_\bullet^{i_m}(y_m)\cdots b_\bullet^{i_n}(y_n)\,\lambda(dy_m)\cdots\lambda(dy_n) \leq$$

$$\leq 8\,\varepsilon^{-3r}\,\Delta_3^r\,\rho_*^{n-m+1-2r}\;\|p-p'\|\;\|w\|$$

$$+ 4\,\varepsilon^{-3r}\,\Delta_3^r\,[\Delta\cdot\mathrm{K}](u)\;(n-m+1)\,\rho_*^{n-m-4r}\;\|p-p'\|$$

$$\leq C''\,\varepsilon^{-3r}\;(n-m+1)\,\rho_*^{n-m-4r}\;(1+\|w\|+\|w'\|)\;,$$

where the constant C'' depends only on r, Δ_3, and $[\Delta\cdot\mathrm{K}](u)$.

Combining the above estimates finishes the proof. □

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399