



PERSISTENCE OF SPECTRAL PROJECTIONS FOR STOCHASTIC OPERATORS ON LARGE TENSOR PRODUCTS

ROBERT S. MACKAY ,* *University of Warwick*

Abstract

It is proved that for families of stochastic operators on a countable tensor product, depending smoothly on parameters, any spectral projection persists smoothly, where smoothness is defined using norms based on ideas of Dobrushin. A rigorous perturbation theory for families of stochastic operators with spectral gap is thereby created. It is illustrated by deriving an effective slow two-state dynamics for a three-state probabilistic cellular automaton.

Keywords: Stochastic operator; probabilistic cellular automaton; spectral projection

2020 Mathematics Subject Classification: Primary 60J05

Secondary 60G60

1. Introduction

Persistence of spectral projections under perturbations, for linear operators on large tensor products, is crucial in many domains. Perhaps the most significant one is many-particle quantum systems, but a parallel problem occurs for stochastic systems with many components, like interacting particle systems and probabilistic cellular automata (PCA). In particular, we would like to know whether the subspace corresponding to an isolated part of the spectrum is robust to small perturbations, meaning that it has a smooth continuation as the subspace corresponding to the part of the spectrum near the original. We call this ‘persistence of spectral projections’.

As a toy example, consider a PCA with three states $\{+, 0, -\}$ per site of a graph with N nodes and bounded degree. For parameter $\varepsilon = 0$, suppose the state on each site evolves independently of the others, with $\{+, -\}$ being absorbing and 0 going to $\{+, -\}$ with probabilities $\frac{1}{2}, \frac{1}{2}$. Then the subset $\{+, -\}^N$ consists of absorbing states, so produces an eigenspace of the transition operator (on functions on the state space), with eigenvalue $+1$ and dimension 2^N . The subset where at least one node is in state 0 produces an eigenspace of eigenvalue 0 and the complementary dimension $3^N - 2^N$. When an interaction of strength $\varepsilon > 0$ along the edges of the graph is introduced, for example favouring alignment of $\{+, -\}$, do these two spectral projections persist? That is, are there nearby invariant subspaces for the perturbed transition operator, with spectra contained in neighbourhoods of $+1$ and 0 respectively?

Liggett’s statement that ‘total variation convergence essentially never occurs for particle systems’ [11, p. 70] makes the persistence of spectral projections look unlikely, but our solution is to use a different metric than total variation.

Received 13 April 2022; accepted 24 February 2024.

* Postal address: Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK. Email: R.S.MacKay@warwick.ac.uk

© The Author(s), 2024. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper specialises to Markov processes, mostly in discrete time, such as PCA. The transition operator T for updating a probability distribution in one time-step acts dually on the space of real-valued continuous functions of the state of the whole system. This space can be considered to be the tensor product of the spaces of real-valued continuous functions of the state of the individual units. To see this, recall that the tensor product $\otimes_{s \in S} V_s$ of a finite set of vector spaces V_s is the set of multilinear forms on the product $\times_{s \in S} V_s^*$ of the dual spaces V_s^* . Taking $V_s = C(X_s, \mathbb{R})$ (continuous functions from X_s to \mathbb{R}) then (modulo assumptions that we'll specify) V_s^* is the space of measures \mathcal{M}_s on X_s . Define δ_{x_s} to be the atomic probability at $x_s \in X_s$. To a continuous function $f: \times_{s \in S} X_s \rightarrow \mathbb{R}$ associate the element \hat{f} of the tensor product $\otimes_{s \in S} V_s$ defined by $\hat{f}((\delta_{x_s})_{s \in S}) = f(x)$ for all $x = (x_s)_{s \in S} \in \times_{s \in S} X_s$. Thus, the problem is about stochastic operators on large tensor products.

Even if the system is geometrically ergodic (meaning there is a unique stationary probability and it attracts every probability exponentially in an appropriate metric) and the update of each unit is independent of the state outside a bounded neighbourhood, when we change parameters in a reasonable way the stationary probability may move at a speed going to infinity with the number of units if distances between probability distributions are measured in any of the standard ways (e.g. total variation, Jensen–Shannon, Hellinger, Kantorovich, Fisher information [13], and Prokhorov [14]). The solution proposed in [13] was to introduce a new metric for probabilities on large product systems, christened the ‘Dobrushin metric’ as most of the ingredients were already in Dobrushin’s work [7], but credit should also be given to Vasershtein [22] (more commonly transliterated now as Wasserstein). Furthermore, as neither paper actually introduced it, I will now just call it the D -metric. After the publication of [13], I found that Steif had proposed a solution 20 years earlier [19]. Although his metric is defined differently and in a restricted context, it is in fact equal to mine in the finite case [14, Appendix], and with Armstrong–Goodall we have proved they are equal under the general conditions for definition of mine [3]; it is a case of ‘strong duality’.

With respect to the D -metric, the stationary probability of a family of PCA with non-degenerate stationary probability varies smoothly, uniformly in the size of the system [13]; the proof there is defective but it is rectified here in Appendix B. A slightly more sophisticated way of viewing this result is as persistence of a (rank-1) projection P onto the space of stationary measures, and its complementary projection $Q = I - P$ to the subspace of neutral measures (for the terminology, see Definition 3), between which there is a spectral gap. Thus, P is a spectral projection (a projection operator onto a subspace corresponding to a closed subset of the spectrum of T , whose complementary projection $Q = I - P$ is onto a complementary subspace corresponding to the disjoint closed complement of the spectrum).

A question that the work of [13] prompted is whether other spectral projections for stochastic operators might also persist uniformly smoothly in the size of the system. Here, suitable conditions are formulated and a proof of persistence is given. Furthermore, a perturbation theory is developed to derive effective dynamics on the image of a spectral projection.

Here is an outline of the paper. First, the D -metric is reviewed (Section 2). Then, the continuation problem for spectral projections of a class of stochastic operators is formulated and solved (Section 3). An illustration of the result is given (Section 4), followed by a general development of second-order perturbation theory for stochastic operators (Section 5) and then a discussion of further potential applications, including to metastability (Section 6). The paper ends with a short summary (Section 7) and two Appendices.

2. The D -metric

The exposition of this section follows [13] but fills some gaps.

Let X be the product of a set of metric spaces (X_s, d_s) , for ‘sites’ s in a countable set S (countable includes finite). The spaces (X_s, d_s) are assumed to be Polish (complete separable metric spaces) with bounded diameter, $\Omega = \sup_{s \in S} \text{diam}_s(X_s) < \infty$, and to have at least two points each. The product X is endowed with product topology and the resulting Borel sets. By a *measure* I mean a finite signed real Borel measure.

With the above assumptions, X is compact and the space \mathcal{M} of measures is the dual of the space $C(X, \mathbb{R})$ of continuous functions from X to \mathbb{R} with supremum norm $\|\cdot\|_\infty$. Given a measure m and a continuous function $f: X \rightarrow \mathbb{R}$, $m(f)$ (or just mf) denotes the integral of f with respect to m . Define $\|m\|_1 = \sup\{m(f): f \in C(X, \mathbb{R}), \|f\|_\infty \leq 1\}$.

The set \mathcal{P} of *probabilities* on X consists of the measures p satisfying $p(X) = 1$ and $p(Y) \geq 0$ for all Borel subsets Y . In particular, $p \in \mathcal{P}$ implies $p(\mathbf{1}) = 1$, where $\mathbf{1}$ is the function taking the value 1 everywhere on X . Also, $\|p\|_1 = 1$.

Definition 1. For a function $f: X \rightarrow \mathbb{R}$, its *Lipschitz constant* with respect to variations on site $s \in S$ is

$$\Delta_s(f) = \sup \frac{f(x) - f(x')}{d_s(x_s, x'_s)}$$

over $x, x' \in X$ differing at site s and agreeing elsewhere.

Note that $\Delta_s(f) \geq 0$ and may be $+\infty$.

Definition 2. The set F of *Dobrushin smooth functions* consists of those $f: X \rightarrow \mathbb{R}$ for which the semi-norm $\|f\|_F = \sum_{s \in S} \Delta_s(f)$ is finite.

Note that Dobrushin smooth functions are automatically continuous: firstly,

$$|f(x) - f(y)| \leq \sum_s \Delta_s(f) d_s(x_s, y_s);$$

secondly, given $\varepsilon > 0$, there exists a finite (possibly empty) subset $K \subset S$ such that $\sum_{s \in S \setminus K} \Delta_s(f) < \varepsilon / \Omega$ and $\Delta_s(f) > 0$ for $s \in K$. For $s \in K$, let $A_s \subset X_s$ be the open ball of radius $\varepsilon / (|K| \Delta_s(f))$ about y_s , and for $s \in S \setminus K$, let $A_s = X_s$. Then $\prod_{s \in S} A_s$ is an open neighbourhood of y in X , and $x \in \prod_{s \in S} A_s$ implies $|f(x) - f(y)| < \varepsilon$.

Also, letting C be the set of constant functions, then $\|\cdot\|_F$ is a norm on the quotient space F/C (F modulo addition of constants).

Definition 3. The space Z of *neutral (or zero-charge) measures* on X is the set of measures μ for which $\mu(X) = 0$. Equivalently, $\mu(\mathbf{1}) = 0$.

Definition 4. The *norm* $\|\mu\|_Z$ of $\mu \in Z$ is $\|\mu\|_Z = \sup\{\mu(f): \|f\|_F \leq 1\}$.

It is a norm on Z ; see [13]. With this norm, the dual space Z^* is F/C (proved in Appendix A).

Definition 5. The D -metric on \mathcal{P} is defined by $D(\rho, \sigma) = \|\rho - \sigma\|_Z$ for any two $\rho, \sigma \in \mathcal{P}$.

The D -metric makes \mathcal{P} into a complete metric space, with diameter $\Omega = \sup_{s \in S} \text{diam}_s(X_s)$. The proof of completeness in [13] is wrong (the direction of duality above was mistaken), but a proof is provided here in Appendix B.

Definition 6. For $m \in \mathcal{M}$, define $|m|_{\mathcal{M}} = \sup\{m(f) : |f|_F \leq 1, |f|_{\infty} \leq \Omega\}$.

It is a norm on \mathcal{M} , and satisfies $|m|_{\mathcal{M}} \leq \Omega|m|_1$. Note that for $\mu \in Z$, $|\mu|_{\mathcal{M}} = |\mu|_Z$, because if $|f|_F \leq 1$, we can add a constant to f to achieve $|f|_{\infty} \leq \Omega$ and it does not change $\mu(f)$.

Definition 7. A transition operator T is a bounded linear operator on \mathcal{M} , written as acting to the left, such that $p \in \mathcal{P}$ implies $pT \in \mathcal{P}$.

Note that a transition operator maps Z to Z (write $\mu \in Z$ as $\mu^+ - \mu^-$ with μ^{\pm} non-negative; $\mu\mathbf{1} = 0$ so $\mu^{\pm}\mathbf{1}$ are equal; if $\mu \neq 0$ then the common value $k > 0$; then $\mu^{\pm}/k \in \mathcal{P}$ and it follows that $(\mu T)\mathbf{1} = 0$). Furthermore, to check T is bounded it is enough to check that its restriction $T_Z : Z \rightarrow Z$ is bounded. This is because, choosing a $p \in \mathcal{P}$, any $m \in \mathcal{M}$ can be written as $kp + \mu$ for some $k \in \mathbb{R}$, $\mu \in Z$ (simply let $k = m\mathbf{1}$ and $\mu = m - kp$). Then, for $|f|_F \leq 1, |f|_{\infty} \leq \Omega$,

$$(mT)f = k(pT)f + (\mu T)f \leq |k||f|_{\infty} + |\mu T|_Z|f|_F \leq k\Omega + |\mu|_Z|T_Z|_Z.$$

Definition 8. For $f \in F$, let $\|f\| = \max(|f|_F, |f|_{\infty}/\Omega)$.

It is a complete norm on F .

A transition operator T has a dual acting on F that we denote by the same symbol but acting to the right.

Proposition 2.1. Given a transition operator T on \mathcal{M} , there is a unique linear operator on F such that $p(Tf) = (pT)f$ for all $f \in F, p \in \mathcal{M}$.

Proof. First, prove the uniqueness of Tf . If there are $g_1, g_2 \in F$ such that, for all $p \in \mathcal{M}$, $p(g_1) = (pT)f = p(g_2)$, let $g = g_1 - g_2$. Then, for all $\mu \in Z$, $\mu(g) = 0$, which implies that g is constant. Then, choosing a $\pi \in \mathcal{P}$, $\pi(\mathbf{1}) = 1$, so $\pi(g) = 0$ implies $g = 0$.

Next, prove the existence of Tf . Given $x \in X$, let $g(x) = (\delta_x T)f$ (recall that δ_x is the atomic probability at x). We have to check that the function g is in F . Making a change in x at a single site s , we obtain $g(x') - g(x) \leq |\delta_{x'} - \delta_x|_Z|T|_Z|f|_F$. Noting that $|\delta_{x'} - \delta_x|_Z = d_s(x', x_s)$, we obtain $\Delta_s(g) \leq |T|_Z|f|_F$. To bound the sum of the $\Delta_s(g)$, let $\gamma \in (0, 1)$. For each $s \in S$, let $x^{(s)}, y^{(s)}$ be points of X differing only at site s and that achieve $g(x^{(s)}) - g(y^{(s)}) \geq \gamma \Delta_s(g) d_s(x^{(s)}, y^{(s)})$. Let $\mu \in Z$ be the sum of ‘dipoles’

$$\frac{\delta_{x^{(s)}} - \delta_{y^{(s)}}}{d_s(x^{(s)}, y^{(s)})}$$

over a finite subset $S' \subset S$. Then $|\mu|_Z = 1$ and

$$\gamma \sum_{s \in S'} \Delta_s(g) \leq \mu(g) = (\mu T)f \leq |\mu T|_Z|f|_F \leq |T|_Z|f|_F.$$

This is true for all finite $S' \subset S$, hence $\gamma \sum_{s \in S} \Delta_s(g) \leq |T|_Z|f|_F$. So $g \in F$. □

Elaborating on the proof, we can obtain useful bounds. Specifically, taking $\gamma \rightarrow 1$ we obtain $|Tf|_F \leq |T|_Z|f|_F$. Also, $|Tf|_{\infty} = \sup_x |(\delta_x T)f| \leq |f|_{\infty}$. We quantify the size of a transition operator T by its operator norm on F :

$$\|T\| = \sup\{\|Tf\|; \|f\| \leq 1\}. \tag{1}$$

Note that $\|T\| \leq \max(|T|_Z, 1)$.

It will be convenient in much of this paper to consider the action of T on F rather than \mathcal{M} , because F is complete. The property that T preserves probability can be written as $T\mathbf{1} = \mathbf{1}$.

Definition 9. A parametrised family of transition operators T_ν , $\nu \in \mathbb{R}^m$, is *smooth* if it depends C^r on ν for some $r \geq 1$, using the operator norm (1). A family of probability distributions p_ν is *smooth* if it depends C^r on ν , using the norm $|\cdot|_Z$ on differences of probabilities.

The reason for introducing the Dobrushin metric and the concept of smoothness in [13] was to derive the following theorem.

Theorem 1. *Let T be a transition operator. If $(I - T)$ is invertible on Z with bounded inverse then T has a unique stationary probability $p \in \mathcal{P}$ (i.e. $pT = p$) and it varies smoothly with respect to smooth changes in T , with $p' = pT'(I - T)_Z^{-1}$, where $'$ denotes the derivative with respect to parameters and $(I - T)_Z$ is the restriction of $I - T$ to Z .*

A different norm on transition operators was used in [13], but it is equivalent to the operator norm (1). Also, it was mistakenly assumed in [13] that Z is complete and hence that if a bounded operator on Z is invertible then its inverse is bounded; it is not clear that this always holds for $(I - T)_Z$ so we include the requirement for its inverse to be bounded in the statement of the theorem. With this addition, the proof in [13] goes through. There are sufficient conditions for bounded invertibility of $(I - T)_Z$ in terms of Dobrushin's 'dependency matrix' [7], which are verifiable in relevant classes of system, so this is a useable result [13].

3. Spectral projections

We begin the discussion of spectral projections in the general context of bounded linear operators on Banach spaces.

Definition 10. Given a bounded linear operator T on a Banach space F , a *spectral projection* for T is a bounded linear operator P on F such that $P^2 = P$, $PT = TP$, and the parts of the spectrum of T corresponding to the restrictions of T to the image (commonly called the 'range', but we reserve 'range' for the whole space into which P is defined to map) of P and the image of $Q = I - P$ (which are invariant under T) are disjoint.

Because the two parts of the spectrum are closed and bounded, it follows that the distance between them is positive, called a *spectral gap*.

An example of a spectral projection is $P = \mathbf{1}p$ for the stationary probability p for a geometrically ergodic transition operator T . This is because $p\mathbf{1} = 1$ for any probability p , $T\mathbf{1} = \mathbf{1}$ by the definition of a transition operator, $pT = p$ for a stationary measure, and geometric ergodicity implies the eigenspace for eigenvalue 1 is one-dimensional and the rest of the spectrum is in a disk of radius less than 1 around 0. Theorem 1 gives smooth dependence of p on T . Using the norm (1) on spectral projections we find that $\|\mathbf{1}p'\| = |p'|_Z/\Omega$, so we see that Theorem 1 gives smooth dependence of the spectral projection $\mathbf{1}p$ on T . The goal of this section is to extend this result to arbitrary spectral projections for T .

Returning to the general context of bounded linear operators on a Banach space F , the set \mathcal{M} of (bounded) projections on F is a smooth manifold (a variant of a Grassmann manifold). One way to see this is that projections are equivalent to direct sum decompositions $R \oplus K$ of F into two closed subspaces, namely the projection onto R along K . The space of closed subspaces is the Grassmann manifold. The space of complementary pairs of closed subspaces is also a manifold.

Here is an outline proof that M is a manifold, from the definition as the set of projections, because the ingredients are useful later. Let $B(F)$ be the space of bounded linear operators P on the space F , with operator norm, which makes $B(F)$ a Banach space. Define Φ on $B(F)$ by $\Phi(P) = P^2 - P$, so $M = \Phi^{-1}(0)$. Given $P_0 \in M$, let $R = \text{im } P_0$ and $K = \text{ker } P_0$. Both are closed. Then $F = R \oplus K$. Relative to this direct-sum decomposition,

$$P_0 = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}. \quad (2)$$

For an arbitrary operator

$$\pi = \begin{bmatrix} \pi_1 & \pi_2 \\ \pi_3 & \pi_4 \end{bmatrix} \text{ on } F, \quad D\Phi_{P_0}(\pi) = \begin{bmatrix} \pi_1 & 0 \\ 0 & -\pi_4 \end{bmatrix}.$$

With respect to the same direct-sum decomposition, let $T_{P_0}M$ (note the reuse of the symbol T , which will signify tangent space instead of transition operator) be the space of operators on F of the form

$$\begin{bmatrix} 0 & \pi_2 \\ \pi_3 & 0 \end{bmatrix}, \quad (3)$$

and $N_{P_0}M$ be those of the form

$$\begin{bmatrix} \pi_1 & 0 \\ 0 & \pi_4 \end{bmatrix}.$$

Consequently, $D\Phi_{P_0}$ maps $N_{P_0}M$ to itself by $(\pi_1, \pi_4) \mapsto (\pi_1, -\pi_4)$, which has bounded inverse, namely itself. So the implicit function theorem shows that the set of P for which the diagonal blocks of $P^2 - P$ in the decomposition $R \oplus K$ are zero is locally the graph of a C^1 function $\psi: T_{P_0}M \rightarrow N_{P_0}M$.

To complete the outline proof, we have to check that the off-diagonal blocks of $P^2 - P$ automatically evaluate to zero too. The way we found to do this is via an explicit formula for ψ , as follows. Writing $P = P_0 + \pi$, the equations for the diagonal blocks of $P^2 - P$ to be zero are $\pi_1 + \pi_1^2 = -\pi_2\pi_3$, $\pi_4 - \pi_4^2 = \pi_3\pi_2$. The unique solution ψ for π small is

$$\pi_1 = -\frac{1}{2}I + \sqrt{\frac{1}{4}I - \pi_2\pi_3}, \quad \pi_4 = \frac{1}{2}I - \sqrt{\frac{1}{4}I - \pi_3\pi_2},$$

where the square roots are defined by their binomial expansions. Then the off-diagonal blocks $\pi_1\pi_2 + \pi_2\pi_4$, $\pi_3\pi_1 + \pi_4\pi_3$ can be seen to evaluate to 0. Note that this explicit formula for ψ gives a direct proof that M is a manifold, rendering use of the implicit function theorem redundant. Perhaps there is a simpler way to show the off-diagonal blocks are zero.

Note that $\Phi(P_0 + \tilde{\pi} + \psi(\tilde{\pi})) = 0$ for $\tilde{\pi} \in T_{P_0}M$ near 0 (or the explicit formula for ψ) implies $D\psi(0) = 0$, so $T_{P_0}M$ is the tangent space to M at P_0 (hence the notation). Note also that the manifold M of projections has many components of different dimensions, corresponding to the similarity class (rank in finite dimensions) of the projection.

Then we have the following classical theorem. It is often proved by contour integration of the resolvent operator, e.g. [10], but we prefer our own proof.

Theorem 2. *Given a family of bounded linear operators T_μ on a Banach space F , depending C^r ($r \geq 1$) on $\mu \in \mathbb{R}^n$, and P_0 a spectral projection for T_0 , there is a neighbourhood of P_0 containing a unique spectral projection P_μ for T_μ for all μ near 0 and it depends C^r on μ .*

Proof. Define a map $G: M \times B(F) \rightarrow B(F)$ by the commutator $G(P, T) = [P, T] = PT - TP$. Note that the image lies in $T_P M$, because in the direct sum decomposition $F = R \oplus K$ above, where P has the form (2), then $[P, T]$ has the form (3). Thus G can be considered as a map from $B(F)$ to sections of TM .

Now we apply the implicit function theorem to G . We have $G(P_0, T_0) = 0$, and G is C^r . It remains to check that the derivative $\partial G/\partial P$ at (P_0, T_0) , mapping $T_{P_0} M$ to itself, is invertible with bounded inverse. We denote it for short by L .

Applied to a tangent vector $P' \in T_P M$, $L(P') = P'T - TP' \in T_P M$. In the direct sum decomposition $F = R \oplus K$,

$$T \text{ has the form } \begin{bmatrix} T_P & 0 \\ 0 & T_Q \end{bmatrix} \text{ and } P' \text{ has the form } \begin{bmatrix} 0 & U \\ V & 0 \end{bmatrix}.$$

So

$$L(P') = \begin{bmatrix} 0 & UT_Q - T_P U \\ VT_P - T_Q V & 0 \end{bmatrix}.$$

A general element of $T_P M$ has the form

$$\begin{bmatrix} 0 & C \\ D & 0 \end{bmatrix}.$$

We claim that L considered as mapping $T_P M$ to itself is invertible, because inverting L reduces to solving two ‘Sylvester equations’ $UT_Q - T_P U = C$, $VT_P - T_Q V = D$, for operators U, V . There is a unique bounded solution to each of the Sylvester equations if and only if the spectra of T_P and T_Q are disjoint [4]. It is automatic that the resulting inverse operators $C \mapsto U$, $D \mapsto V$ are bounded, because the inverse of an invertible bounded linear operator between Banach spaces is bounded [10].

Thus, L has bounded inverse and the result follows by the implicit function theorem. \square

Note that the result does not preclude existence of other spectral projections further away from P_0 , for example of different rank.

Denoting derivatives with respect to μ by $'$, we note the useful formula $P' = -L^{-1}[P, T']$. This results from applying the chain rule to the equation $G(P_\mu, T_\mu) = 0$ at $\mu = 0$, but holds more generally if L is evaluated at (P_μ, T_μ) . If we have bounds on L^{-1} and T' , it allows us to deduce bounds on $P_\mu - P_0$.

We now apply Theorem 2 to the case of F being the Dobrushin smooth functions on a countable product of Polish spaces with bounded diameter, with the norm of Definition 8 and the associated norm (1) on linear operators on F , such as transition operators T , projections P , and infinitesimal changes to each, and smoothness as in Definition 9.

Theorem 3. *Given a smooth family of transition operators T_μ , for each spectral projection P_0 of T_0 there is $\varepsilon > 0$ such that P_0 persists smoothly to a spectral projection P_μ for all $|\mu| < \varepsilon$.*

For practical purposes, we need a bound on the solutions of the Sylvester equations.

Definition 11. For bounded linear maps $A: K \rightarrow K$, $B: H \rightarrow H$ between normed spaces, their *separation* is

$$\text{sep}(A, B) = \inf_{X \neq 0} \frac{\|AX - XB\|}{\|X\|}$$

over bounded linear $X: H \rightarrow K$.

This generalises the concept of the separation of two matrices A, B , which is usually defined using the Frobenius norm [20, 21]. Note that $\text{sep}(B, A)$ is not necessarily equal to $\text{sep}(A, B)$. From the definition, $\text{sep}(A, B) > 0$ if and only if the Sylvester equation $AX - XB = C$ has a unique solution X for each C , and by the theory of Sylvester equations already used, this is true if and only if the spectra of A and B are disjoint. Then $\|X\| \leq \|C\|/\text{sep}(A, B)$. Thus, taking $A = T_P$ and $B = T_Q$,

$$\|L^{-1}\| \leq \gamma = (\text{sep}(A, B))^{-1} + (\text{sep}(B, A))^{-1}.$$

In some cases we can give explicit lower bounds on $\text{sep}(A, B)$. For example, if the spectral radii $\rho_A, \rho_{B^{-1}}$ of A and B^{-1} satisfy $\rho_A \rho_{B^{-1}} < 1$ then for $\lambda > 1$ there exist $C_A, C_{B^{-1}}$ (dependent on λ) such that, for $n \geq 0$, $\|A^n\| \leq C_A(\lambda \rho_A)^n$ and $\|B^{-n}\| \leq C_{B^{-1}}(\lambda \rho_{B^{-1}})^n$. The explicit (convergent) solution $X = -\sum_{n=0}^{\infty} A^n C B^{-n-1}$ of the Sylvester equation shows that

$$\text{sep}(A, B) \geq \frac{1 - \lambda^2 \rho_A \rho_{B^{-1}}}{C_A C_{B^{-1}} \lambda \rho_{B^{-1}}}.$$

Similarly, for $YA - BY = D$ the explicit solution $Y = -\sum_{n \geq 0} B^{-n-1} D A^n$ provides the same bound on $\text{sep}(B, A)$. A method to obtain similar bounds for some more general forms of separation of the spectra is described in [16].

The important point is that for a family of examples with growing system size N , if the separation is bounded away from zero then the continuation in Theorem 3 is uniform in N . Furthermore, it can apply to infinite systems.

The same can be done in continuous time, with the transition operator replaced by a transition generator, but we do not spell it out here, save to mention that if the spectra of A and B lie in $\text{Re } z \geq r_A, \text{Re } z \leq r_B$ respectively, with $r_A > r_B$, the explicit (convergent) solution $X = \int_0^{\infty} e^{(-A+r)t} C e^{(B-r)t} dt$ for $r \in (r_B, r_A)$ provides a bound of the form

$$\text{sep}(A, B) \geq \frac{r_A - r_B - 2\varepsilon}{c_A c_B}$$

for any $\varepsilon > 0$, with c_A, c_B depending on ε .

4. An illustration

As an example, consider a three-state PCA, with local state space $\{+, 0, -\}$ at each site of a finite undirected graph with N nodes and bounded degree, say by m . The transition probabilities for the state at a given site are taken to be

$$\begin{bmatrix} 1 - (1 + \alpha n_-)\varepsilon & (1 + \alpha n_-)\varepsilon & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & (1 + \alpha n_+)\varepsilon & 1 - (1 + \alpha n_+)\varepsilon \end{bmatrix} \tag{4}$$

in basis $(+, 0, -)$, where n_{\pm} are the numbers of neighbours in states \pm respectively, $\alpha \geq 0$, and $\varepsilon \in [0, (1 + \alpha m)^{-1}]$. For $\varepsilon = 0$ the system has eigenvalue 1 with multiplicity 2^N and eigenvalue 0 with multiplicity $3^N - 2^N$.

From the above theory, the spectral projection to the subspace for eigenvalue 1 has a continuation for ε small. The continuation is uniform in N because the separation of the relevant operators can be bounded away from 0 uniformly in N . The dynamics on the image of the

projection is slow because the spectrum moves from $\{1\}$ by at most $O(\varepsilon)$. It is still Markovian. It might loosely be considered as an effective PCA with two states $\{+, -\}$ on each site but $R = \text{im } P$ is not spanned by δ -functions on states, so such a description requires interpretation analogous to quasiparticles in quantum mechanics.

If $\alpha < 1/m$ and $\varepsilon > 0$, the dynamics on R is geometrically ergodic, because the whole system is. The latter can be proved using Dobrushin's dependency matrix, as follows (compare an example of its use in [13]). Take the discrete metric on the local state spaces. Then the update probability distributions $p(\sigma)$ for the state at a site given its current state σ and the numbers n_{\pm} are the rows of the matrix (4) and so the variation distance for a change of state on the given site is at most $(1 - \varepsilon)$, and for a change on a neighbouring site is at most $\alpha\varepsilon$. Thus the dependency matrix has ℓ_{∞} -norm at most $1 - (1 - m\alpha)\varepsilon$. If $\alpha < 1/m$ and $\varepsilon > 0$, this is less than 1 and so the system is geometrically ergodic.

5. Second-order perturbation theory

We might ask what the dynamics of the above example look like on the image of the projection corresponding to the spectrum near 1. This can be answered by analogy to second-order perturbation theory computations in quantum mechanics, e.g. the derivation of the $t - J$ model from the Hubbard model [17]. In the above example, the 0 state mediates interactions between the \pm states. Here, a general treatment of second-order perturbation theory for families of stochastic operators is given and applied to this example.

Under the spectral gap condition at $\varepsilon = 0$ and with the norm (1), it has been proved above that for a smooth family of stochastic operators T_{ε} and a spectral projection P_0 for T_0 , there is an open neighbourhood in ε for which P_0 continues smoothly to a spectral projection P_{ε} for T_{ε} . It is convenient to write $P_{\varepsilon} = \psi_{\varepsilon} P_0 \psi_{\varepsilon}^{-1}$ for an ε -dependent invertible bounded linear map ψ , with ψ_0 the identity. There is a lot of freedom in the choice of ψ , the only constraints being that $\psi_{\varepsilon}(\text{im } P_{\varepsilon}) = \text{im } P_0$ and $\psi_{\varepsilon}(\ker P_{\varepsilon}) = \ker P_0$, but it is good also to take ψ to be as smooth in ε as is T . Then we can define $S = \psi' \psi^{-1}$, where $'$ denotes $d/d\varepsilon$, and the above conditions reduce to

$$[S, P] = P', \quad (5)$$

where $[\cdot, \cdot]$ again denotes the commutator. A convenient solution is $S = P'(P - Q)$, which can be checked to satisfy the condition (5).

Then it is desired to compute $\hat{T} = \psi^{-1} T \psi$ on $\text{im } P_0$. This is the stochastic operator that represents the dynamics of T_{ε} on $\text{im } P_{\varepsilon}$, using the coordinate system ψ_{ε} .

Start from $\hat{T}_0 = T_0$. The first derivative is $\hat{T}' = \psi^{-1}(T' + [T, S])\psi$. Using the above choice of S , we obtain $\hat{T}' = \psi^{-1} J \psi$, where $J = PT'P + QT'Q$. The second derivative can be evaluated to $\hat{T}'' = \psi^{-1}(PT''P + QT''Q + [[T, P'], P'])\psi$.

Evaluating these at $\varepsilon = 0$, we obtain \hat{T} to second order in ε as

$$\hat{T}_{\varepsilon} = P_0 T_{\varepsilon} P_0 + Q_0 T_{\varepsilon} Q_0 + \frac{\varepsilon^2}{2} [[T_0, P'_0], P'_0]. \quad (6)$$

It is perhaps more useful to substitute $[T, P'] = [T', P]$, since the right-hand side of this is readily computable, but the second occurrence of P' above means that solution of this equation for P' is required.

For the example of Section 4, there is already an effect at first order in ε , so second-order perturbation theory is perhaps unnecessary, but it still serves as an illustration of the procedure.

$$T_0 = P_0 = \otimes_{s \in S} \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}, \quad T'_0 = \otimes_{s \in S} \begin{bmatrix} -\beta_- & \beta_- & 0 \\ 0 & 0 & 0 \\ 0 & \beta_+ & -\beta_+ \end{bmatrix},$$

where $\beta_{\pm} = 1 + \alpha n_{\pm}$. Thus

$$P_0 T_{\varepsilon} P_0 + Q_0 T_{\varepsilon} Q_0 = \otimes_{s \in S} \begin{bmatrix} 1 - \beta_- \varepsilon / 2 & 0 & \beta_- \varepsilon / 2 \\ \frac{1}{2} + \beta_+ \varepsilon / 2 & (\beta_+ + \beta_-) \varepsilon / 2 & \frac{1}{2} + \beta_- \varepsilon / 2 \\ \beta_+ \varepsilon / 2 & 0 & 1 - \beta_+ \varepsilon / 2 \end{bmatrix}.$$

To compute the second-order term, begin with

$$[T'_0, P_0] = \otimes_{s \in S} \begin{bmatrix} \frac{1}{2} \beta_- & -\beta_- & \frac{1}{2} \beta_- \\ \frac{1}{2} \beta_- & -\frac{1}{2}(\beta_- + \beta_+) & \frac{1}{2} \beta_+ \\ \frac{1}{2} \beta_+ & -\beta_+ & \frac{1}{2} \beta_+ \end{bmatrix}.$$

Using $P_0 P'_0 = P'_0 Q_0$, P'_0 has just four independent parameters on each site, and solving $[T, P'] = [T', P]$ for them yields

$$P'_0 = \otimes_{s \in S} \begin{bmatrix} \frac{1}{2} \beta_- & -\beta_- & \frac{1}{2} \beta_- \\ \frac{1}{2} \beta_+ & -\frac{1}{2}(\beta_- + \beta_+) & \frac{1}{2} \beta_- \\ \frac{1}{2} \beta_+ & -\beta_+ & \frac{1}{2} \beta_+ \end{bmatrix}.$$

To compute the ε^2 term in (6), we can first subtract $[T'_0, P_0]$ from P'_0 , leaving, site-wise,

$$[[T_0, P'_0], P'_0] = \frac{\beta_+ - \beta_-}{2} \left[[T'_0, P_0], \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \right] = \frac{\beta_+ - \beta_-}{2} \begin{bmatrix} -\beta_- & 0 & \beta_- \\ -\beta_- & \beta_- - \beta_+ & \beta_+ \\ -\beta_+ & 0 & \beta_+ \end{bmatrix}.$$

Thus, the final result to second order, using the (+, -) part of the ψ_{ε} basis, is the effective PCA

$$\hat{T}_{\varepsilon} = \otimes_{s \in S} \begin{bmatrix} 1 - \varepsilon \beta_- / 2 + \varepsilon^2 \beta_- (\beta_- - \beta_+) / 4 & \varepsilon \beta_- / 2 - \varepsilon^2 \beta_- (\beta_- - \beta_+) / 4 \\ \varepsilon \beta_+ / 2 - \varepsilon^2 \beta_+ (\beta_+ - \beta_-) / 4 & 1 - \varepsilon \beta_+ / 2 + \varepsilon^2 \beta_+ (\beta_+ - \beta_-) / 4 \end{bmatrix}.$$

The outcome is a two-state model in which to leading order there is a small probability $\varepsilon \beta_{\mp} / 2 = \varepsilon (1 + \alpha n_{\mp}) / 2$ per timestep for transition from state + to -, respectively - to +.

Perhaps a reader can suggest (and apply the method to) a more significant example.

6. Further potential applications

One area to which the above results might be usefully applied is metastability. Metastability is the phenomenon that an ergodic process may spend long times exploring restricted subsets of the support of the probability distribution, switching between them infrequently but sufficiently to achieve ergodicity in the long run. For an infinite limit, there may be more than one stationary distribution. One reference is [5].

As an application of the result, we can start from [6] in which the hypothesis is a reversible Markov process in continuous time with spectrum in $[-\varepsilon, 0] \cup [-\infty, -1]$ such that each function in the image of the spectral projection P for $[-\varepsilon, 0]$ is bounded, and it is proved that there is a partition of the state spaces into ‘metastable regions’. It is not clear to me, however, whether there are relevant examples satisfying the hypotheses. We might think that Glauber dynamics of the 2D Ising model below the critical temperature would qualify, but I am not aware that it is proved to have a spectral gap. Nevertheless, if there are examples on product spaces then the result of the present paper proves robustness of spectral gap and hence of the phenomenon of metastability.

We could envisage the result also being useful to treat perturbation of Markov dynamics with more than one stationary distribution, for example with more than one communicating component. Addition of some interaction between the communicating components typically reduces the system to a single communicating component, but the result of this paper shows there is a continuation of the spectral projection to a spectral projection with spectrum contained near 1 and the same rank (equal to the original number of communicating components), and it gives strong control over the resulting continuation. More substantially, we would like to deduce something about metastability in ergodic finite versions of infinite PCA with non-unique stationary distribution.

Another possible application is to perturbations of product systems in which the units all have simple eigenvalue $+1$ and isolated spectrum near some λ in the open unit disk. The conclusion is that there persists an invariant subspace with decay constant near λ .

7. Discussion

It has been proved here that every spectral projection of a stochastic operator on a product space persists C^r -smoothly with respect to the norm (1) for C^r -smooth changes in the transition operator, again measured using (1). This generalises the case of the rank-one projection onto a stationary distribution, treated in [13].

Second-order perturbation theory has been developed for families of such operators and an example treated. Potential applications have also been suggested to robustness of metastability and some other uses for multi-component stochastic processes.

A reviewer raised the question of relations to some other literature. In particular, linear operators on a general class of Banach spaces with a positive cone and a multiple eigenvalue $+1$ are considered in [8, 15]. It appears, however, that our space does not satisfy the additivity property that they require for the norm on vectors in the positive cone. Furthermore, those papers do not consider the question of persistence of the spectral projection for the multiple eigenvalue, nor treat any other spectral projections. Nor do they address the difficulties with making useful norms on large tensor products.

Looking to the future, an additional benefit of the work of this paper is to gain some understanding of spectral projections on large tensor products with a view to tackling the quantum case. In addition to its clear role in condensed matter physics, the quantum case has taken on enhanced interest because of the problem of designing quantum registers for quantum computing. In many quantum contexts, we want the subspace corresponding to an isolated part of the spectrum to be robust to small perturbations. Yet, the Anderson orthogonalisation catastrophe [1] apparently precludes this. The approach here suggests that with a suitable new norm, we could nonetheless prove persistence of spectral projections for Hermitian operators on large tensor products.

Appendix A. Dual of space Z of neutral measures

Here it is proved that the dual space Z^* of the space Z of neutral measures with the $|\cdot|_Z$ -norm can be regarded naturally as the quotient F/C of the space F of Dobrushin smooth functions by the constant ones C .

Given $f \in F$, define $\hat{f}: Z \rightarrow \mathbb{R}$ by $\hat{f}(\mu) = \mu(f)$. Then \hat{f} is linear. Also, it is bounded: $\hat{f}(\mu) = \mu(f) \leq |\mu|_Z |f|_F$. Thus, $\hat{f} \in Z^*$ with $|\hat{f}|_{Z^*} \leq |f|_F$. Adding a constant to f does not change \hat{f} so we can consider $F/C \subset Z^*$.

Conversely, given $u \in Z^*$, choose a reference point $a \in X$ and define $\check{u}: X \rightarrow \mathbb{R}$ by $\check{u}(x) = u(\delta_{xa}) + c$ for arbitrary $c \in \mathbb{R}$, where, for $x, y \in X$, δ_{xy} is the dipole measure $\delta_x - \delta_y \in Z$. Then $\check{u}(x) - \check{u}(y) = u(\delta_{xy}) \leq |u|_{Z^*} |\delta_{xy}|_Z$. Taking x, y to differ at only site s , we obtain that $|\delta_{xy}|_Z = \sup_{f \in F/C} (f(x) - f(y))/|f|_F$ is attained for $f(z) = d_s(z_s, y_s)$, giving value $d_s(x_s, y_s)$. Hence $\Delta_s(\check{u}) \leq |u|_{Z^*}$. Given $\gamma \in (0, 1)$, for each $s \in S$ there exist $x^{(s)}, y^{(s)}$, differing only on site s , such that $\check{u}(x^{(s)}) - \check{u}(y^{(s)}) \geq \gamma \Delta_s(\check{u}) d_s(x_s^{(s)}, y_s^{(s)})$. Choose

$$\mu = \sum_{s \in S'} \frac{\delta_{x^{(s)}, y^{(s)}}}{d_s(x_s^{(s)}, y_s^{(s)})}$$

where S' is any finite subset of S . Then $\mu \in Z$ and $|\mu|_Z \leq 1$ by using the Lipschitz bounds for $f \in F$ on $\mu(f)$; furthermore, $|\mu|_Z \geq 1$ by choosing $f(x) = \sum_{s \in S'} \alpha_s d_s(x_s, y_s^{(s)})$, any $\alpha_s > 0$. Thus $|\mu|_Z = 1$. It follows that $\sum_{s \in S'} \Delta_s(\check{u}) \leq \gamma |u|_{Z^*}$ for all $\gamma < 1$ and $S' \subset S$. So $\check{u} \in F$ and $|\check{u}|_F \leq |u|_{Z^*}$. So we can consider $Z^* \subset F/C$.

From the bounds on the norms in the two directions, we deduce equality: $|\hat{f}|_{Z^*} = |f|_F$ and $|\check{u}|_F = |u|_{Z^*}$.

Note, however, that the dual of F/C is strictly larger than Z . It contains limits of sequences of dipoles, for example [2].

Appendix B. Completeness of \mathcal{P} in the D -metric

We prove this via a few artificial constructions for which precursors occur in [9, 18].

We first metrize product-topology on X , by choosing an enumeration of S (i.e. label the sites by non-negative integers) and an $\eta > 1$, and using $d_\eta(x, y) = \sum_j \eta^{-j} d_j(x_j, y_j)$ for some $\eta > 1$. The sum converges because the diameters of the X_j are bounded.

Next, we define the Kantorovich–Rubinstein (KR) metric on \mathcal{P} with respect to d_η on X :

$$d_{KR}(\mu, \nu) = \sup\{(\mu - \nu)(f); f: X \rightarrow \mathbb{R}, f(x) - f(y) \leq d_\eta(x, y) \text{ for all } x, y \in X\}.$$

As a countable product of complete separable metric spaces, X is a complete separable metric space. For any complete separable metric space X , the space $\mathcal{M}(X)$ of real-valued Borel measures on X is the dual of the space of continuous functions $X \rightarrow \mathbb{R}$ with sup-norm $|\cdot|_\infty$, and the KR metric metrises weak* convergence on $\mathcal{P}(X)$ [23], denoted by \rightharpoonup (confusingly, weak* convergence on \mathcal{M} is often called weak convergence).

Then we define an auxiliary metric ρ on \mathcal{P} :

$$\rho(\mu, \nu) = \sup \left\{ (\mu - \nu)(f); f \in F, \sum_j \eta^j \Delta_j(f) \leq 1 \right\}.$$

This is clearly a semi-metric. To prove it is a metric, if $\mu \neq \nu$ there exists a continuous f with $(\mu - \nu)f \neq 0$; approximate f by g with $\sum_j \eta^j \Delta_j g < \infty$ and $|f - g|_\infty$ small enough that $(\mu - \nu)g \neq 0$.

We show that $\rho \leq d_{\text{KR}}$. If $\sum_j \eta^j \Delta_j f \leq 1$ then, for any $x, y \in X$,

$$\begin{aligned} f(x) - f(y) &\leq \sum_j \Delta_j(f) d_j(x_j, y_j) \leq \sum_j \eta^j \Delta_j(f) \eta^{-j} d_j(x_j, y_j) \\ &\leq \left(\sum_j \eta^j \Delta_j(f) \right) \left(\sum_j \eta^{-j} d_j(x_j, y_j) \right) \leq d_\eta(x, y). \end{aligned}$$

So the set of f in the definition of ρ is a subset of that for d_{KR} , so $\rho(\mu, \nu) \leq d_{\text{KR}}(\mu, \nu)$.

It follows that if $\mu_n \rightarrow \mu$ then $\rho(\mu_n, \mu) \rightarrow 0$. Although not needed, the converse is also true: if $\rho(\mu_n, \mu) \rightarrow 0$ then, given f continuous and $\varepsilon > 0$, there exists $K \subset S$ such that $f = f_K + \tilde{f}$ with f_K independent of coordinates outside K and $|\tilde{f}| \leq \varepsilon/4$. So there exists N such that $(\mu_n - \mu)(f_K) < \varepsilon/2$ for $n \geq N$. Also, for all n , $(\mu_n - \mu)(\tilde{f}) < \varepsilon/2$. Thus, $(\mu_n - \mu)(f) < \varepsilon$ for $n \geq N$. So $\mu_n \rightarrow \mu$. So ρ metrises weak* convergence in \mathcal{P} .

It follows that \mathcal{P} is complete in ρ , because suppose μ_n is Cauchy in ρ . Given $\varepsilon > 0$ there exists N such that $n, m \geq N$ imply $\rho(\mu_n, \mu_m) < \varepsilon$. \mathcal{P} is weak* sequentially compact in the usual topology. Thus there is a weak*-convergent subsequence to some $\mu \in \mathcal{P}$: $\mu_{n_k} \rightarrow \mu$. Then $\rho(\mu_{n_k}, \mu) \rightarrow 0$ so, given $\varepsilon > 0$, $\rho(\mu_n, \mu) \leq \rho(\mu_n, \mu_{n_k}) + \rho(\mu_{n_k}, \mu) < 2\varepsilon$ for $n \geq N$ and sufficiently large k . Thus, $\rho(\mu_n, \mu) \rightarrow 0$.

We also have $\rho \leq D$: if $\sum_j \eta^j \Delta_j(f) \leq 1$ then

$$\sum_j \Delta_j(f) \leq \sum_j \eta^j \Delta_j(f) - \sum_j (\eta^j - 1) \Delta_j(f) \leq 1.$$

So, given a Cauchy sequence μ_n in D , it is Cauchy in ρ , thus ρ -converges to some $\mu \in \mathcal{P}$. Given $\varepsilon > 0$ there exists $N(\varepsilon)$ such that $D(\mu_n, \mu_m) < \varepsilon/4$ for $n, m \geq N(\varepsilon)$; it also follows that $\rho(\mu_m, \mu) < \varepsilon/4$ for $m \geq N(\varepsilon)$. Given $f \in F$, a reference point $0 \in X$ and a finite subset $K \subset S$, let $f_K(x) = f(x_K, 0_{S \setminus K})$ and $\tilde{f} = f - f_K$. Then $\Delta_j f_K \leq \Delta_j f$ for $j \in K$ and is 0 otherwise. There exists $K \subset S$ such that $|\tilde{f}| \leq \frac{1}{4} \varepsilon |f|_F$. So

$$\begin{aligned} (\mu_n - \mu)f &= (\mu_n - \mu_m)f + (\mu_m - \mu)f_K + (\mu_m - \mu)\tilde{f} \\ &\leq D(\mu_n, \mu_m)|f|_F + \rho(\mu_m, \mu) \sum_{j \in K} \eta^j \Delta_j(f) + \frac{1}{2} \varepsilon |f|_F \\ &\leq (\varepsilon/4 + \varepsilon/4 + \varepsilon/2)|f|_F \end{aligned}$$

for $n \geq N(\varepsilon)$ and $m \geq \max(N(\varepsilon), N(\varepsilon \eta^{-k}))$, where k is the largest label of K . Thus, $D(\mu_n, \mu) \rightarrow 0$ as $n \rightarrow \infty$. So (\mathcal{P}, D) is complete.

Acknowledgements

I am grateful to Nick Higham for pointers to the literature on Sylvester equations. I dedicate this paper to his memory. I am also grateful to a reviewer who insisted that I give more complete proofs, which led me to realise that my statement in [13] that \mathcal{P} is complete with respect to the D -metric required proof (given here in Appendix B).

Funding information

There are no funding bodies to thank relating to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- [1] ANDERSON, P. W. (1967). Infrared catastrophe in Fermi gases with local scattering potentials. *Phys. Rev. Lett.* **18**, 1049–1051.
- [2] ARENS, R. F. AND EELLS, J. (1956). On embedding uniform and topological spaces. *Pacific J. Math.* **6**, 397–403.
- [3] ARMSTRONG-GOODALL, J. AND MACKAY, R. S. (2021). Dobrushin and Steif metrics are equal. Preprint, [arXiv:2104.08365](https://arxiv.org/abs/2104.08365).
- [4] BHATIA, R. AND ROSENTHAL, P. (1997). How and why to solve the operator equation $AX - XB = Y$? *Bull. Lond. Math. Soc.* **29**, 1–21.
- [5] BOVIER, A. AND DEN HOLLANDER, F. (2015). *Metastability*. Springer, New York.
- [6] DAVIES, E. B. (1982). Metastable states of symmetric Markov semigroups II. *J. London Math. Soc.* **26**, 541–550.
- [7] DOBRUSHIN, R. L. (1971). Markov processes with a large number of locally interacting components: Existence of a limit process and its ergodicity. *Problems Inform. Transm.* **7**, 149–164.
- [8] ERKUŞUN-ÖZCAN, N. AND MUKHAMEDOV, F. (2021). Spectral conditions for uniform P -ergodicities of Markov operators on abstract state spaces. *Glasgow Math. J.* **63**, 682–696.
- [9] FÖLLMER, H. AND HORST, U. (2001). Convergence of locally and globally interacting Markov chains. *Stoch. Process. Appl.* **96**, 99–121.
- [10] KATO, T. (1980). *Perturbation Theory for Linear Operators*. Springer, New York.
- [11] LIGGETT, T. M. (2005). *Interacting Particle Systems*. Springer, New York.
- [12] MACKAY, R. S. (2000). Discrete breathers: Classical and quantum. *Physica A* **288**, 174–198.
- [13] MACKAY, R. S. (2011). Robustness of Markov processes on large networks. *J. Difference Equat. Appl.* **17**, 1155–1167.
- [14] MACKAY, R. S. (2018). Management of complex systems. *Nonlinearity* **31**, R52–65.
- [15] MUKHAMEDOV, F. AND AL-RAWASHDEH, A. (2020). Generalized Dobrushin ergodicity coefficient and uniform ergodicity of Markov operators. *Positivity* **24**, 855–890.
- [16] NEVANLINNA, O. (2019). Sylvester equations and polynomial separation of spectra. Preprint, [arxiv:1904.07549](https://arxiv.org/abs/1904.07549).
- [17] SPALEK, J. (2007). $t - J$ model then and now: A personal perspective from the pioneering times. *Acta Phys. Polon. A* **111**, 409–424.
- [18] STEIF, J. E. (1988). The ergodic structure of interacting particle systems. PhD thesis, Stanford University.
- [19] STEIF, J. E. (1991). Convergence to equilibrium and space-time Bernoullicity for spin systems in the $M < \varepsilon$ case. *Ergod. Theory Dyn. Syst.* **11**, 547–575.
- [20] STEWART, G. W. (1973). Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Rev.* **15**, 727–764.
- [21] VARAH, J. M. (1979). On the separation of two matrices. *SIAM J. Numer. Anal.* **16**, 216–222.
- [22] VASERSHTEIN, L. N. (1969). Markov processes over denumerable products of spaces describing large systems of automata. *Problemy Peredachi Informatskii* **5**, 64–72.
- [23] VILLANI, C. (2009). *Optimal Transport*. Springer, New York.