Hindawi Publishing Corporation
Fixed Point Theory and Applications
Volume 2007, Article ID 41930, 69 pages
doi:10.1155/2007/41930

Research Article

Fixed Points of Two-Sided Fractional Matrix Transformations

David Handelman

Received 16 March 2006; Revised 19 November 2006; Accepted 20 November 2006

Recommended by Thomas Bartsch

Let C and D be $n \times n$ complex matrices, and consider the densely defined map $\phi_{C,D}$: $X \mapsto (I - CXD)^{-1}$ on $n \times n$ matrices. Its fixed points form a graph, which is generically (in terms of (C,D)) nonempty, and is generically the Johnson graph $J(n,2n)$; in the nongeneric case, either it is a retract of the Johnson graph, or there is a topological continuum of fixed points. Criteria for the presence of attractive or repulsive fixed points are obtained. If C and D are entrywise nonnegative and CD is irreducible, then there are at most two nonnegative fixed points; if there are two, one is attractive, the other has a limited version of repulsiveness; if there is only one, this fixed point has a flow-through property. This leads to a numerical invariant for nonnegative matrices. Commuting pairs of these maps are classified by representations of a naturally appearing (discrete) group. Special cases (e.g., $CD - DC$ is in the radical of the algebra generated by C and D) are discussed in detail. For invertible size two matrices, a fixed point exists for all choices of C if and only if D has distinct eigenvalues, but this fails for larger sizes. Many of the problems derived from the determination of harmonic functions on a class of Markov chains.

Copyright © 2007 David Handelman. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Contents

1. Introduction 2
2. Preliminaries 3
3. New fixed points from old 8
4. Local matrix units 10
5. Isolated invariant subspaces 13
6. Changing solutions 17

7. Graphs of solutions 18
8. Graph fine structure 22
9. Graph-related examples 27
10. Inductive relations 30
11. Attractive and repulsive fixed points 32
12. Commutative cases 35
13. Commutative modulo the radical 39
14. More fixed point existence results 41
15. Still more on existence 43
16. Positivity 49
17. Connections with Markov chains 58
Appendices 59
A. Continua of fixed points 59
B. Commuting fractional matrix transformations 62
C. Strong conjugacies 66
Acknowledgment 69
References 69

1. Introduction

Let C and D be square complex matrices of size n. We obtain a densely defined mapping from the set of $n \times n$ matrices (denoted $\mathrm{M}_n\mathbf{C}$) to itself, $\phi_{C,D} : X \mapsto (\mathrm{I} - CXD)^{-1}$. We refer to this as a *two-sided matrix fractional linear transformation,* although these really only correspond to the denominator of the standard fractional linear transformations, $z \mapsto (az + b)/(cz + d)$ (apparently more general transformations, such as $X \mapsto (CXD + E)^{-1}$, reduce to the ones we study here). These arise in the determination of harmonic functions of fairly natural infinite state Markov chains [1].

Here we study the fixed points. We show that if $\phi_{C,D}$ has more than $\binom{2n}{n}$ fixed points, then it has a topological continuum of fixed points. The set of fixed points has a natural graph structure. Generically, the number of fixed points is exactly $\binom{2n}{n}$. When these many fixed points occur, the graph is the Johnson graph $J(n, 2n)$. When there are fewer (but more than zero) fixed points, the graphs that result can be analyzed. They are graph retractions of the generic graph, with some additional properties (however, except for a few degenerate situations, the graphs do not have uniform valence, so the automorphism group does not act transitively). We give explicit examples (of matrix fractional linear transformations) to realize all the possible graphs arising when $n = 2$: (a) 6 fixed points, the generic graph (octahedron); (b) 5 points (a "defective" form of (a), square pyramid); (c) 4 points (two graph types); (d) 3 points (two graph types); (e) 2 points (two graph types, one disconnected); and (f) 1 point.

We also deal with attractive and repulsive fixed points. If $\phi_{C,D}$ has the generic number of fixed points, then generically, it will have both an attractive and a repulsive fixed point, although examples with neither are easily constructed. If $\phi_{C,D}$ has fewer than the generic number of fixed points, it can have one but not the other, or neither, but usually has both.

In all cases of finitely many fixed points and CD invertible, there is at most one attractive fixed point and one repulsive fixed point.

We also discuss entrywise positivity. If C and D are entrywise nonnegative and CD is irreducible (in the sense of nonnegative matrices), then $\phi_{C,D}$ has at most two nonnegative fixed points. If there are two, then one of them is attractive, and the other is a rank one perturbation of it; the latter is not repulsive, but satisfies a limited version of repulsivity. If there is exactly one, then $\phi_{C,D}$ has no attractive fixed points at all, and the unique positive one has a "flow-through" property (inspired by a type of tea bag). This leads to a numerical invariant for nonnegative matrices, which, however, is difficult to calculate (except when the matrix is normal).

There are three appendices. The first deals with consequences of and conditions guaranteeing continua of fixed points. The second discusses the unexpected appearance of a group whose finite dimensional representations classify commuting pairs $(\phi_{C,D}, \phi_{A,B})$ (it is not true that $\phi_{A,B} \circ \phi_{C,D} = \phi_{C,D} \circ \phi_{A,B}$ implies $\phi_{A,B} = \phi_{C,D}$, but modulo rational rotations, this is the case). The final appendix concerns the group of densely defined mappings generated by the "elementary" transformations, $X \mapsto X^{-1}$, $X \mapsto X + A$, and $X \mapsto RXS$ where RS is invertible. The sets of fixed points of these (compositions) can be transformed to their counterparts for $\phi_{C,D}$.

2. Preliminaries

For $n \times n$ complex matrices C and D, we define the *two-sided matrix fractional linear transformation,* $\phi \equiv \phi_{C,D}$ via $\phi_{C,D}(X) = (\mathrm{I} - CXD)^{-1}$ for $n \times n$ matrices X. We observe that the domain is only a dense open set of $\mathrm{M}_n\mathbf{C}$ (the algebra of $n \times n$ complex matrices); however, this implies that the set of X such that $\phi^k(X)$ are defined for all positive integers k is at least a dense G_δ of $\mathrm{M}_n\mathbf{C}$.

A square matrix is *nonderogatory* if it has a cyclic vector (equivalently, its characteristic polynomial equals its minimal polynomial, equivalently it has no multiple geometric eigenvectors,…, and a host of other characterizations).

Throughout, the spectral radius of a matrix A, that is, the maximum of the absolute values of the eigenvalues of A, is denoted $\rho(A)$.

If W is a subset of $\mathrm{M}_n\mathbf{C}$, then the *centralizer of* W,

$$\{M \in \mathrm{M}_n\mathbf{C} \mid MB = BM \ \forall B \in W\}, \tag{2.1}$$

is denoted W', and of course, the double centralizer is denoted W''. Typically, $W = \{C, D\}$ for two specific matrices C and D, so the notation will not cause confusion with other uses of primes. The transpose of a matrix A is denoted A^T, and the conjugate transpose is denoted A^*.

Our main object of study is the set of fixed points of ϕ. If we assume that ϕ *has* a fixed point (typically called X), then we can construct all the other fixed points, and in fact, there is a natural structure of an undirected graph on them. For generic choices of C and D, a fixed point exists (Proposition 15.1); this result is due to my colleague, Daniel Daigle.

The method of describing all the other fixed points yields some interesting results. For example, if ϕ has more than $C(2n, n) = \binom{2n}{n}$ fixed points, then it has a topological

continuum of fixed points, frequently an affine line of them. On the other hand, it is generic that ϕ have exactly $C(2n,n)$ fixed points.

(For X and Y in $\mathrm{M}_n\mathbf{C}$, we refer to $\{X+zY \mid z\in\mathbf{C}\}$ as an *affine line.*)

Among our tools (which are almost entirely elementary) are the two classes of linear operators on $\mathrm{M}_n\mathbf{C}$. For R and S in $\mathrm{M}_n\mathbf{C}$, define the maps $\mathcal{M}_{R,S}, \mathcal{J}_{R,S} : \mathrm{M}_n\mathbf{C}\to\mathrm{M}_n\mathbf{C}$ via

$$\begin{aligned}\mathcal{M}_{R,S}(X) &= RXS,\\ \mathcal{J}_{R,S}(X) &= RX-XS.\end{aligned}\tag{2.2}$$

As a mnemonic device (at least for the author), $\mathcal{M}$ stands for multiplication. By identifying these with the corresponding elements of the tensor product $\mathrm{M}_n\mathbf{C}\otimes\mathrm{M}_n\mathbf{C}$, that is, $R\otimes S$ and $R\otimes \mathrm{I}-\mathrm{I}\otimes S$, we see immediately that the (algebraic) spectra are easily determined—$\operatorname{spec}\mathcal{M}_{R,S}=\{\lambda\mu \mid (\lambda,\mu)\in \operatorname{spec}R\times\operatorname{spec}S\}$ and $\operatorname{spec}\mathcal{J}_{R,S}=\{\lambda-\mu \mid (\lambda,\mu)\in \operatorname{spec}R\times\operatorname{spec}S\}$. Every eigenvector decomposes as a sum of rank one eigenvectors (for the same eigenvalue), and each rank one eigenvector of either operator is of the form vw where v is a right eigenvector of R and w is a left eigenvector of S. The Jordan forms can be determined from those of R and S, but the relation is somewhat more complicated (and not required in almost all of what follows).

Before discussing the fixed points of maps of the form $\phi_{C,D}$, we consider a notion of equivalence between more general maps. Suppose that $\phi,\psi : \mathrm{M}_n\mathbf{C}\to\mathrm{M}_n\mathbf{C}$ are both maps defined on a dense open subset of $\mathrm{M}_n\mathbf{C}$, say given by formal rational functions of matrices, that is, a product

$$X\longmapsto p_1(X)\bigl(p_2(X)\bigr)^{-1}p_3(X)\bigl(p_4(X)\bigr)^{-1}\cdots,\tag{2.3}$$

where each $p_i(X)$ is a noncommutative polynomial. Suppose there exists γ of this form, but with the additional conditions that it has $\mathrm{GL}(n,\mathbf{C})$ in its domain and maps it onto itself (i.e., $\gamma \mid \mathrm{GL}(n,\mathbf{C})$ is a self-homeomorphism), and moreover, $\phi\circ\gamma=\gamma\circ\psi$. Then we say that ϕ and ψ are *strongly conjugate,* with the conjugacy implemented by γ (or γ^{-1}). If we weaken the self-homeomorphism part merely to $\mathrm{GL}(n,\mathbf{C})$ being in the domain of both γ and γ^{-1}, then γ induces a *weak conjugacy* between ϕ and ψ.

The definition of strong conjugacy ensures that invertible fixed points of ϕ are mapped bijectively to invertible fixed points of ψ. While strong conjugacy is obviously an equivalence relation, weak conjugacy is not transitive, and moreover, weakly conjugate transformations need not preserve invertible (or any) fixed points (Proposition 15.7(a)). Nonetheless, compositions of weak conjugacies (implementing the transitive closure of weak conjugacy) play a role in what follows. These ideas are elaborated in Appendix C.

Choices for γ include $X\mapsto RXS+T$ where RS is invertible (a self-homeomorphism of $\mathrm{M}_n\mathbf{C}$)) and $X\mapsto X^{-1}$ with inverse $X\mapsto X^{-1}$ (a self-homeomorphism of $\mathrm{GL}(n,\mathbf{C})$). In the first case, $\gamma : X\mapsto RXS+T$ is a weak conjugacy, and is a strong conjugacy if and only if T is zero. (Although translation $X\mapsto X+T$ is a self-homeomorphism of $\mathrm{M}_n\mathbf{C}$, it only implements a weak conjugacy.) The map $X\mapsto X^{-1}$ is a strong conjugacy.

Lemma 2.1. *Suppose that C and D lie in* $\mathrm{GL}(n,\mathbf{C})$. *Then one has the following:*

(i) *$\phi_{C,D}$ is strongly conjugate to each of $\phi_{D,C}^{-1}$, ϕ_{D^T,C^T}, ϕ_{D^*,C^*};*

(ii) *if A and B are in $\mathrm{M}_n\mathbf{C}$ and E is in $\mathrm{GL}(n,\mathbf{C})$, then $\psi : X \mapsto (E - AXB)^{-1}$ is strongly conjugate to $\phi_{AE^{-1},BE^{-1}}$;*

(iii) *if A, B, and F are in $\mathrm{M}_n\mathbf{C}$, and E, $EAE^{-1} + F$, and $B - AE^{-1}F$ are in $\mathrm{GL}(n,\mathbf{C})$, then $\psi : X \mapsto (AX + B)(EX + F)^{-1}$ is weakly conjugate to $\phi_{C,D}$ for some choice of C and D.*

Proof. (i) In the first case, set $\tau(X) = (CXD)^{-1}$ and $\alpha(X) = (1 - X^{-1})^{-1}$ (τ implements a strong conjugacy, but α does not), and form $\alpha \circ \tau$, which of course is just $\phi_{C,D}$. Now $\tau \circ \alpha(X) = D^{-1}(\mathrm{I} - X^{-1})C^{-1}$, and it is completely routine that this is $\phi_{D,C}^{-1}(X)$. Thus $\alpha \circ \tau = \phi_{C,D}$ and $\tau \circ \alpha = \phi_{D,C}^{-1}$. Set $\gamma = \tau^{-1}$ (so that $\gamma(X) = (DXC)^{-1}$).

For the next two, define $\gamma(X) = X^T$ and X^*, respectively, and verify $\gamma^{-1} \circ \phi_{C,D} \circ \gamma$ is what it is supposed to be.

(ii) Set $\gamma(X) = E^{-1}X$ and calculate $\gamma^{-1}\psi\gamma = \phi_{AE^{-1},BE^{-1}}$.

(iii) Set $S = AE^{-1}$ and $R = B - AE^{-1}F$. First define $\gamma_1 : X \mapsto RX + S$. Then $\gamma_1^{-1}\psi\gamma_1(X) = (ESR + FR + CRXR)^{-1}$; this will be of the form described in (ii) if $ESR + FR$ is invertible, that is, $ES + F$ is invertible. This last expression is $EAE^{-1} + F$. Hence we can define $\gamma_2 : X \mapsto R^{-1}(ES + F)^{-1}X$, so that by (ii), $\gamma_2^{-1}\gamma_1^{-1}\psi\gamma_1\gamma_2 = \phi_{C,D}$ for appropriate choices of C and D. Now $\gamma := \gamma_1 \circ \gamma_2 : X \mapsto RZX + S$ where R and Z are invertible, so γ is a homeomorphism defined on all of $\mathrm{M}_n\mathbf{C}$, hence implements a weak conjugacy. □

In the last case, a more general form is available, namely, $X \mapsto (AXG + B)(EXG + F)^{-1}$ (the repetition of G is *not* an error) is weakly conjugate to a $\phi_{C,D}$ under some invertibility conditions on the coefficients. We discuss this in more generality in Appendix C.

Lemma 2.1 entails that when CD is invertible, then $\phi_{C,D}$ is strongly conjugate to $\phi_{D,C}^{-1}$. A consequence of the definition of strong conjugacy is that the structure and quantity of fixed points of $\phi_{C,D}$ is the same as that of $\phi_{D,C}$ (since fixed points are necessarily invertible, the mapping and its inverse is defined on the fixed points, hence acts as a bijection on them). However, attractive fixed points—if there are any—are converted to repulsive fixed points. Without invertibility of CD, there need be no bijection between the fixed points of $\phi_{C,D}$ and those of $\phi_{D,C}$; Example 2.4 exhibits an example wherein $\phi_{C,D}$ has exactly one fixed point, but $\phi_{D,C}$ has two.

We can then ask, if CD is invertible, is $\phi_{C,D}$ strongly conjugate to $\phi_{D,C}$? By Lemma 2.1, this will be the case if either both C and D are self-adjoint or both are symmetric. However, in Section 9, we show how to construct examples with invertible CD for which $\phi_{C,D}$ has an attractive but no repulsive fixed point. Thus $\phi_{D,C}^{-1}$ has an attractive but no repulsive fixed point, whence $\phi_{D,C}$ has a repulsive fixed point, so cannot be conjugate to $\phi_{C,D}$.

We are primarily interested in fixed points of $\phi_{C,D}$ (with CD invertible). Such a fixed point satisfies the equation $X(\mathrm{I} - CXD) = \mathrm{I}$. Post-multiplying by D and setting $Z = XD$, we deduce the quadratic equation

$$Z^2 + AZ + B = \mathbf{0}, \tag{q}$$

where $A = -C^{-1}Z$ and $B = C^{-1}D$. Of course, invertibility of A and B allows us to reverse the procedure, so that fixed points of $\phi_{C,D}$ are in bijection with matrix solutions to (q), where $C = -A^{-1}$ and $D = -A^{-1}B$. If one prefers ZA rather than AZ, a similar result applies, obtained by using $(\mathrm{I} - CXD)X = \mathrm{I}$ rather than $(\mathrm{I} - CXD)X = \mathrm{I}$.

The seemingly more general matrix quadratic

$$Z^2 + AZ + ZA' + B = \mathbf{0} \tag{qq}$$

can be converted into (q) via the simple substitution, $Y = Z + A'$. The resulting equation is $Y^2 + (A - A')Y + B - AA' = \mathbf{0}$.

This yields limited results about fixed points of other matrix fractional linear transformations. For example, the mapping $X \mapsto (XA + B)(EX + F)^{-1}$ is a plausible one-sided generalization of fractional linear transformations. Its fixed points X satisfy $X(EX + F) = (XA + B)$. Right multiplying by E and substituting $Z = XE$, we obtain $Z^2 + Z(E^{-1}F - E^{-1}AE) - BE = \mathbf{0}$, and this can be converted into the quadratic (q) via the simple substitution described above.

A composition of one-sided denominator transformations can also be analyzed by this method. Suppose that $\phi : X \mapsto (\mathrm{I} - RX)^{-1}$ and $\phi_0 : X \mapsto (\mathrm{I} - XS)^{-1}$, where RS is invertible (note that R and S are on opposite sides). The fixed points of $\phi \circ \phi_0$ satisfy $(\mathrm{I} - R + S - XS)X = \mathrm{I}$. Right multiplying by S and substituting $Z = XS$, we obtain the equation $Z^2 + (R - S - \mathrm{I})Z + S = \mathbf{0}$, which is in the form (q).

If we try to extend either of these last reductions to more general situations, we run into a roadblock—equations of the form $Z^2 + AZB + C = \mathbf{0}$ do not yield to these methods, even when C does not appear.

However, the Riccati matrix equation in the unknown X,

$$XVX + XW + YX + A = \mathbf{0}, \tag{2.4}$$

does convert to the form in (q) when V is invertible—premultiply by V and set $Z = VX$. We obtain $Z^2 + ZW + VYV^{-1}Z + VA = \mathbf{0}$, which is of the form described in (qq).

There is a large literature on the Riccati equation and quadratic matrix equations. For example, [2] deals with the Riccati equation for rectangular matrices (and on Hilbert spaces) and exhibits a bijection between isolated solutions (to be defined later) and invariant subspaces of 2×2 block matrices associated to the equation. Our development of the solutions in Sections 4–6 is different, although it can obviously be translated back to the methods in [op cit]. Other references for methods of solution (not including algorithms and their convergence properties) include [3, 4].

The solutions to (q) are tractible (and will be dealt with in this paper); the solutions to $Z^2 + AZB + C = \mathbf{0}$ at the moment seem to be intractible, and certainly have different properties. The difference lies in the nature of the derivatives. The derivative of $Z \mapsto Z^2 + AZ$ (and similar ones), at Z, is a linear transformation (as a map sending $\mathrm{M}_n\mathbf{C}$ to itself) all of whose eigenspaces are spanned by rank one eigenvectors. Similarly, the derivative of $\phi_{C,D}$ and its conjugate forms have the same property at any fixed point. On the other hand, this fails generically for the derivatives of $Z \mapsto Z^2 + AZB$ and also for the general fractional linear transformations $X \mapsto (AXB + E)(FXG + H)^{-1}$.

The following results give classes of degenerate examples.

Proposition 2.2. *Suppose that $DC = \mathbf{0}$ and define $\phi : X \mapsto (\mathrm{I} - CXD)^{-1}$.*

(a) *Then ϕ is defined everywhere and $\phi(X) - \mathrm{I}$ is square zero.*

(b) *If $\rho(C) \cdot \rho(D) < 1$, then ϕ admits a unique fixed point, X_0, and for all matrices X, $\{\phi^N(X)\} \to X_0$.*

Proof. Since $(CXD)^2 = CXDCXC = \mathbf{0}$, $(\mathrm{I} - CXD)^{-1}$ exists and is $\mathrm{I} + CXD$, yielding (a).

(b) If $\rho(C) \cdot \rho(D) < 1$, we may replace (C, D) by $(\lambda C, \lambda^{-1} D)$ for any nonzero number λ, without affecting ϕ. Hence we may assume that $\rho(C) = \rho(D) < 1$. It follows that in any algebra norm (on $\mathrm{M}_n\mathbf{C}$), $\|C^N\|$ and $\|D^N\|$ go to zero, and do so exponentially. Hence $X_0 := \mathrm{I} + \sum_{j=1}^{\infty} C^j D^j$ converges.

We have that for any $X, \phi(X) = \mathrm{I} + CXD$; iterating this, we deduce that $\phi^N(X) = \mathrm{I} + \sum_{j=1}^{N-1} C^j D^j + C^N X D^N$. Since $\{C^N X D^N\} \to \mathbf{0}$, we deduce that $\{\phi^N(X)\} \to X_0$. Necessarily, the limit of all iterates is a fixed point. □

If we arrange that $DC = \mathbf{0}$ and $\rho(D)\rho(C) < 1$, then $\phi_{C,D}$ has exactly one fixed point (and it is attractive). On the other hand, we can calculate fixed points for special cases of $\phi_{D,C}$; we show that for some choices of C and D, $\phi_{C,D}$ has one fixed point, but $\phi_{D,C}$ has two.

Lemma 2.3. *Suppose that R and S are rank one. Set $r = \mathrm{tr}\, R$, $s = \mathrm{tr}\, S$, and denote $\phi_{R,S}$ by ϕ. Let $\{H\}$ be a (one-element) basis for $R\mathrm{M}_n\mathbf{C}S$, and let u be the scalar such that $RS = uH$.*

(a) *Suppose that $rs\,\mathrm{tr}\, H = 0$.*

(i) *There is a unique fixed point for ϕ if and only if $1 - rs + u\,\mathrm{tr}\, H \neq 0$.*

(ii) *There is an affine line of fixed points for ϕ if and only if $1 - rs + u\,\mathrm{tr}\, H = u = 0$; in this case, there are no other fixed points.*

(iii) *There are no fixed points if and only if $1 - rs + u\,\mathrm{tr}\, H = 0 \neq u$.*

(b) *Suppose $rs\,\mathrm{tr}\, H \neq 0$.*

(i) *If $(1 + u\,\mathrm{tr}\, H - rs)^2 \neq -4urs\,\mathrm{tr}\, H$, ϕ has two fixed points, while if $(1 + u\,\mathrm{tr}\, H - rs)^2 = -4urs\,\mathrm{tr}\, H$, it has exactly one.*

Proof. Obviously, $R\mathrm{M}_n\mathbf{C}S$ is one dimensional, so is spanned by a single nonzero matrix H. For a rank one matrix Z, $(\mathrm{I} - Z)^{-1} = \mathrm{I} + Z/(1 - \mathrm{tr}\, Z)$; thus the range of ϕ is contained in $\{\mathrm{I} + zH \mid z \in \mathbf{C}\}$. From $R^2 = rR$ and $S^2 = sS$, we deduce that if X is a fixed point, then $\phi(X) = \phi(\mathrm{I} + tH) = (\mathrm{I} - RS - tRHS)^{-1}$ and this simplifies to $(\mathrm{I} - H(rst - u))^{-1} = \mathrm{I} + H(rst - u)/(1 - (rst - u\,\mathrm{tr}\, H))$. It follows that $t = rst - u/(1 - (rst - u)\,\mathrm{tr}\, H)$, and this is also sufficient for $\mathrm{I} + tH$ to be a fixed point.

This yields the quadratic in t,

$$t^2(rs\,\mathrm{tr}\, H) - t(1 - rs + u\,\mathrm{tr}\, H) - u = 0. \tag{2.5}$$

All the conclusions follow from analyzing the roots. □

Example 2.4. A mapping $\phi_{C,D}$ having exactly one fixed point, but for which $\phi_{D,C}$ has two.

Set $C = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ and $D = (1/2)\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then $DC = \mathbf{0}$ and $\rho(C) \cdot \rho(D) < 1$, so $\phi_{C,D}$ has a unique fixed point. However, with $R = D$ and $S = C$, we have that R and S are rank one, $u = 0$, $H = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}\right)$, so $\mathrm{tr}\, H \neq 0$, and the discriminant of the quadratic is not zero—hence

$\phi_{D,C}$ has exactly two fixed points. In particular, $\phi_{C,D}$ and $\phi_{D,C}$ have different numbers of fixed points.

In another direction, it is easy to construct examples with no fixed points. Let N be an $n \times n$ matrix with no square root. For example, over the complex numbers, this means that N is nilpotent, and in general a nilpotent matrix with index of nilpotence exceeding $n/2$ does not have a square root. Set $C = (1/4)\mathrm{I} + N$ and define the transformation $\phi_{C,\mathrm{I}}(X) = (\mathrm{I} - CX)^{-1}$. This has no fixed points—just observe that if X is a fixed point then $Y = CX$ must satisfy $Y^2 - Y = -C$. This entails $(Y - (1/2)\mathrm{I})^2 = -N$, which has no solutions.

On the other hand, a result due to my colleague, Daniel Daigle, shows that for every C, the set of D such that $\phi_{C,D}$ admits a fixed point contains a dense open subset of $\mathrm{GL}(n, \mathbf{C})$ (see Proposition 15.1). For size 2 matrices, there is a complete characterization of those matrices D such that for every C, $\phi_{C,D}$ has a fixed point, specifically that D have distinct eigenvalues (see Proposition 15.5).

A fixed point is *isolated* if it has a neighborhood which contains no other fixed points. Of course, the following result, suitably modified, holds for more general choices of ϕ.

Lemma 2.5. *The set of isolated fixed points of $\phi \equiv \phi_{C,D}$ is contained in the algebra $\{C, D\}''$.*

Proof. Select Z in the group of invertible elements of the subalgebra $\{C, D\}'$; if X is a fixed point of ϕ, then so is ZXZ^{-1}. Hence the group of invertible elements acts by conjugacy on the fixed points of ϕ. Since the group is connected, its orbit on an isolated point must be trivial, that is, every element of the group commutes with X, and since the group is dense in $\{C, D\}'$, every element of $\{C, D\}'$ commutes with X, that is, X belongs to $\{C, D\}''$. □

The algebra $\{C, D\}''$ cannot be replaced by the (generally) smaller one generated by $\{C, D\}$ (see Example 15.11). Generically, even $\langle C, D\rangle$ will be all of $\mathrm{M}_n\mathbf{C}$, so Lemma 2.5 is useless in this case. However, if, for example, $CD = DC$ and one of them has distinct eigenvalues, then an immediate consequence is that all the isolated fixed points are polynomials in C and D. Unfortunately, even when $CD = DC$ and both have distinct eigenvalues, it can happen that not all the fixed points are isolated (although generically this is the case) and need not commute with C or D (see Example 12.6). This yields an example of $\phi_{C,D}$ with commuting C and D whose fixed point set is topologically different from that of any one-sided fractional linear transformation, $\phi_{E,\mathrm{I}} : X \mapsto (\mathrm{I} - EX)^{-1}$.

3. New fixed points from old

Here and throughout, C and D will be $n \times n$ complex matrices, usually invertible, and $\phi \equiv \phi_{C,D} : X \mapsto (\mathrm{I} - CXD)^{-1}$ is the densely defined transformation on $\mathrm{M}_n\mathbf{C}$. As is apparent from, for example, the power series expansion, the derivative $\mathcal{D}\phi$ is given by $(\mathcal{D}\phi)(X)(Y) = \phi(X)CYD\phi(X) = \mathcal{M}_{\phi(X)C,D\phi(X)}(Y)$, that is, $(\mathcal{D}\phi)(X) = \mathcal{M}_{\phi(X)C,D\phi(X)}$. We construct new fixed points from old, and analyze the behavior of $\phi : X \mapsto (\mathrm{I} - CXD)^{-1}$ along nice trajectories.

Let X be in the domain of ϕ, and let v be a right eigenvector for $\phi(X)C$, say with eigenvalue λ. Similarly, let w be a left eigenvector for $D\phi(X)$ with eigenvalue μ. Set $Y = vw$; this is an $n \times n$ matrix with rank one, and obviously Y is an eigenvector of $\mathcal{M}_{\phi(X)C,\phi(X)D}$ with

eigenvalue $\lambda\mu$. For z a complex number, we evaluate $\phi(X+zY)$,

$$\begin{aligned}\phi(X+zY) &= (\mathrm{I}-CXD-zCYD)^{-1}\\ &= \big((\mathrm{I}-CXD)(\mathrm{I}-z\phi(X)CYD)\big)^{-1}\\ &= (\mathrm{I}-z\lambda YD)^{-1}\phi(X).\end{aligned}\tag{3.1}$$

If Z is rank one, then $\mathrm{I}-Z$ is invertible if and only if $\operatorname{tr}Z\neq 1$, and the inverse is given by $\mathrm{I}+Z/(1-\operatorname{tr}Z)$. It follows that except for possibly one value of z, $(\mathrm{I}-z\lambda YD)^{-1}$ exists, and is given by $\mathrm{I}+YDz\lambda/(1-z\lambda\operatorname{tr}YD)$. Thus

$$\begin{aligned}\phi(X+zY) &= \phi(X)+\frac{z\lambda\mu}{1-z\lambda\operatorname{tr}YD}Y\\ &= \phi(X)+\psi(z)Y,\end{aligned}\tag{3.2}$$

where $\psi : z\mapsto z\lambda\mu/(1-z\lambda\operatorname{tr}YD)$ is an ordinary fractional linear transformation, corresponding to the matrix $\left(\begin{smallmatrix}\lambda\mu & 0\\ -\lambda\operatorname{tr}YD & 1\end{smallmatrix}\right)$. The apparent asymmetry is illusory; from the observation that $\operatorname{tr}(\phi(X)CYD)=\operatorname{tr}(CYD\phi(X))$, we deduce that $\lambda\operatorname{tr}YD=\mu\operatorname{tr}CY$.

Now suppose that X is a fixed point of ϕ. Then $X+zY$ will be a fixed point of ϕ if and only if z is a fixed point of ψ. Obviously, $z=0$ is one fixed point of ψ. Assume that $\lambda\mu\neq 0$ (as will occur if CD is invertible). If $\operatorname{tr}YD\neq 0$, there is exactly one other (finite) fixed point.

If $\operatorname{tr}YD=0$, there are no other (finite) fixed points when $\lambda\mu\neq 1$, and the entire affine line $\{X+zY\}_z$ consists of fixed points when $\lambda\mu=1$.

The condition $\operatorname{tr}YD\neq 0$ can be rephrased as $d:=wDv\neq 0$ (or $wCv\neq 0$), in which case, the new fixed point is $X+vw(1-\lambda\mu)/d\lambda$. Generically of course, each of XC and DX will have n distinct eigenvalues, corresponding to n choices for each of v and w, hence n^2 new fixed points will arise (generically—but not in general—e.g., if $CD=DC$, then either there are at most n new fixed points, or a continuum, from this construction).

Now suppose that X is a fixed point, and Y is a rank one matrix such that $X+Y$ is also a fixed point. Expanding the two equations $X(\mathrm{I}-CXD)=\mathrm{I}$ and $(X+Y)(\mathrm{I}-C(X+Y)D)=\mathrm{I}$, we deduce that $Y=(X+Y)CYD+YCXD$, and then observing that $CXD=\mathrm{I}-X^{-1}$ and post-multiplying by X, we obtain $Y=XCYDX+YCYDX$. Now using the identities with the order-reversed ($(\mathrm{I}-CXD)X=\mathrm{I}$ etc.), we obtain $Y=XCYDX+CYDXY$, in particular, Y commutes with $CYDX$. Since Y is rank one, the product $YCYDX=CYDXY$ is also rank one, and since it commutes with Y, it is of the form tY for some t. Hence $XCYDX=(1-t)Y$, and thus Y is an eigenvector of $\mathcal{M}_{XC,DX}$. Any rank one eigenvector factors as vw where v is a right eigenvector of XC and w is a left eigenvector of DX—so we have returned to the original construction. In particular, if X and X_0 are fixed points with $X-X_0$ having rank one, then $X-X_0$ arises from the construction above.

We can now define a graph structure on the set of fixed points. We define an edge between two fixed points X and X_0 when the rank of the difference is one. We will discuss the graph structure in more detail later, but one observation is immediate: if the number of fixed points is finite, the valence of any fixed point in this graph is at most n^2.

Under some circumstances, it is possible to put a directed graph structure on the fixed points. For example, if the eigenvalues of XC and DX are real and all pairs of products

are distinct from 1 (i.e., 1 is not in the spectrum of $\mathcal{M}_{XC,(DX)^{-1}}$), we should have a directed arrow from X to X_0 if $X_0 - X$ is rank one and $\lambda\mu < 1$. We will see (see Section 12) that the spectral condition allows a directed graph structure to be defined. (The directed arrows will point in the direction of the attractive fixed point, if one exists.)

Of course, it is easy to analyze the behaviour of ϕ along the affine line $X + zY$. Since $\phi(X + zY) = \phi(X) + \psi(z)Y$, the behaviour is determined by the ordinary fractional linear transformation ψ. Whether the nonzero fixed point is attractive, repulsive (with respect to the affine line, not globally) or neither, it is determined entirely by ψ.

4. Local matrix units

Here we analyze in considerably more detail the structure of fixed points of $\phi \equiv \phi_{C,D}$, by relating them to a single one. That is, we assume there is a fixed point X and consider the set of differences $X_0 - X$ where X_0 varies over all the fixed points.

It is convenient to change the equation to an equivalent one. Suppose that X and $X + Y$ are fixed points of ϕ. In our discussion of rank one differences, we deduced the equation (Section 3) $Y = XCYDX + YCYDX$ (without using the rank one hypothesis). Left multiplying by C and setting $B = (DX)^{-1}$ (we are assuming CD is invertible) and $A = CX$, and with $U = CY$, we see that U satisfies the equation

$$U^2 = UB - AU. \tag{4.1}$$

Conversely, given a solution U to this, that $X + C^{-1}U$ is a fixed point, follows from reversing the operations. This yields a rank-preserving bijection between $\{X_0 - X\}$ where X_0 varies over the fixed points of ϕ and solutions to (4.1). It is much more convenient to work with (4.1), although we note an obvious limitation: there is no such bijection (in general) when CD is not invertible.

Let $\{e_i\}_{i=1}^k$ and $\{w_i\}_{i=1}^k$ be subsets of $\mathbf{C}^n = \mathbf{C}^{n\times 1}$ and $\mathbf{C}^{1\times n}$, respectively, with $\{e_i\}_{i=1}^k$ linearly independent. Form the $n \times n$ matrix $M := \sum_{i=1}^k e_i w_i$; we also regard as an endomorphism of $\mathbf{C}^{n\times 1}$ via $Mv = \sum e_i(w_i v)$, noting that the parenthesized matrix products are scalars. Now we have some observations (not good enough to be called lemmas).

(i) The range of M is contained in the span of $\{e_i\}_{i=1}^k$, obviously.

(ii) The following are equivalent:

(a) rk $M = k$,

(b) $\{w_i\}_{i=1}^k$ is linearly independent,

(c) range $M = \sum e_i \mathbf{C}$.

Proof. (c) implies (a). Trivial by (i). (a) implies (b). Suppose $\sum \lambda_i w_i = \mathbf{0}$ and relabel so that $\lambda_k \neq 0$. Then there exist scalars $\{\mu_i\}_{i=1}^{k-1}$ such that $w_k = \sum_{i=1}^{k-1} \mu_i w_i$. Thus

$$\begin{aligned} M &= \sum_{i=1}^{k-1} e_i w_i + e_k \Big(\sum \mu_i \lambda_i w_i \Big) \\ &= \sum_{i=1}^{k-1} (e_i + \mu_i e_k) w_i. \end{aligned} \tag{4.2}$$

Hence, by (i) applied to the set $\{e_i + \mu_i e_k\}_{i=1}^{k-1}$, the range of M is in the span of the set, hence the rank of M is at most $k-1$, a contradiction.

(b) implies (c). Enlarge w_i to a basis of $\mathbf{C}^{1\times n}$ (same notation); let $\{v_i\}$ be a dual basis, which we can view as a basis for $\mathbf{C}^n$, so that $w_i v_j = \delta_{ij}$. Then $Mv_j = e_j$, and so e_j belongs to the range of M.

(iii) The column e_j belongs to the range of M if and only if w_j is not in the span of $\{w_i\}_{i\neq j}$.

Proof. If w_j is not the span, there exists a linear functional v on $\mathbf{C}^{1\times n}$, which we view as an element of $\mathbf{C}^n$, such that $w_i v = 0$ if $i \neq j$ but $w_j v = 1$. Then $Mv = e_j$.

Conversely, suppose that for some v, $Mv = e_j$, that is, $e_j = \sum e_i w_i v$. There exist W_l in $\mathbf{C}^{1\times n} = (\mathbf{C}^n)*$ such that $W_l e_i = \delta_{lj}$. Thus $w_j v = 1$ but $w_i v = 0$ if $i \neq j$. Thus w_j is not in the span of the other ws. □

Now suppose that A and B are square matrices of size n and we wish to solve the matrix equation (4.1). Let k be a number between 1 and n; we try to determine all solutions U of rank k. We first observe that A leaves $\mathrm{Rg}\,U$ (a subspace of $\mathbf{C}^n$ of dimension k) invariant, and similarly, the *left range of* U, $\ell\mathrm{Rg}U := \{wU \mid w \in \mathbf{C}^{1\times n}\}$, is invariant under B (acting on the right). Select a basis $\{e_i\}_{i=1}^k$ for $\mathrm{Rg}\,U$ and for convenience, we may suppose that with respect to this basis, the matrix of $A \mid \mathrm{Rg}\,U$ is in Jordan normal form.

Similarly, we may pick a basis for $\ell\mathrm{Rg}U$, $\{f_j\}$, such that the matrix of $\ell\mathrm{Rg}U \mid B$ (the action of B is on the right, hence the notation) is also in Jordan normal form.

Extend the bases so that A and B themselves are put in Jordan normal form (we take upper triangular rather than lower triangular; however, since B is acting on the other side, it comes out to be the transpose of its Jordan form, i.e., lower triangular; of course, generically both A and B are diagonalizable).

Let $M = U$ be a rank k solution to (4.1). Since $\{e_i f_j\}$ is a basis of $\mathrm{M}_n\mathbf{C}$, there exist scalars μ_{ij} such that $M = \sum \mu_{ij} e_i f_j$. We wish to show that $\mu_{ij} = 0$ if either i or j exceeds k.

We have that $\mathrm{Rg}M$ is spanned by $\{e_i\}_{i\leq k}$. Write $M = \sum_{i=1}^n e_i w_i$ where $w_i = \sum_j \mu_{ij} f_j$. For any $l > k$, find a vector W in $\mathbf{C}^{n\times 1}$ such that $We_1 = We_2 = \cdots = We_k = 0$ but $We_l = 1$. Thus $WM = w_l$, and if the latter were not zero, we would obtain a contradiction. Hence $w_l = 0$ for $l > k$; linear independence of $\{f_j\}$ yields that $\mu_{ij} = 0$ if $j > k$. The same argument may be applied on the left to yield the result.

Next, we claim that the $k \times k$ matrix $(\mu_{ij})_{i,j=1}^k$ is invertible. The rank of M is k, and it follows easily that $\{w_i = \sum_j \mu_{ij} f_j\}_{i=1}^k$ is linearly independent. The map $f_l \mapsto \sum_j \mu_{lj} f_j$ is implemented by the matrix, and since the map is one to one and onto (by linear independence), the matrix is invertible.

Now we can derive a more tractible matrix equation. Write $M = \sum \mu_{ij} e_i f_j$, so that

$$M^2 = \sum_{l,m\leq k} e_l f_m \left(\sum_{j,p\leq k} (f_j e_p) \mu_{pm} \right). \tag{4.3}$$

Define the $k \times k$ matrices, $T = (\mu_{ij})$ and $\mathcal{F} := (f_i e_j)$. Let J_B be the Jordan normal form of B restricted to $\ell\mathrm{Rg}U$. Calculating the coefficient of $e_i f_j$ when we expand MB, we obtain $MB = \sum e_i f_j (TJ_B^T)_{ij}$. Similarly, $AM = \sum e_i f_j (J_A T)$. From the expansion for M^2 and the

equality $M^2 = MB - AM$, we deduce an equation involving only $k \times k$ matrices,

$$T\mathcal{F}T = TJ_B^T - J_A T. \tag{4.4}$$

Since T is invertible, say with inverse V, we may pre- and post-multiply by V and obtain the equation (in V)

$$\mathcal{F} = J_B^T V - V J_A. \tag{4.5}$$

In other words, the matrix $\mathcal{F}$ is in the range of $\mathcal{J}_{J_B^T, J_A}$ (on $\mathrm{GL}(k)$).

A rank k solution to (4.1) thus yields an invertible solution to (4.5). However, it is important to note that the Jordan forms are of the restrictions to the pair of invariant subspaces. In particular, if we begin with a pair of equidimensional left A- and right B-invariant spaces, form the matrix $\mathcal{F}$ (determined by the restrictions A and B), then we will obtain a solution to (4.1), provided we can solve (4.5) with an *invertible* V. The invertibility is a genuine restriction, for example, if the spectra of A and B are disjoint, (4.5) has a unique solution, but it is easy to construct examples wherein the solution is not invertible. It follows that there is no solution to (4.1) with the given pair of invariant subspaces.

We can give a sample result, showing what happens at the other extreme. Suppose that the spectra of A and B consist of just one point, which happens to be the same and there is just one eigenvector (i.e., the Jordan normal forms each consist of a single block). We will show that either there is just the trivial solution to (4.1) ($U = \mathbf{0}$), or there is a line of solutions, and give the criteria for each to occur. First, subtracting the same scalar matrix from A and B does not affect (4.1), so we may assume that the lone eigenvalue is zero, and we label the eigenvectors e and f, so $Ae = \mathbf{0}$ and $fB = \mathbf{0}$.

The invariant subspaces of A form an increasing family of finite dimensional vector spaces, $(\mathbf{0}) = V_0 \subset V_1 \subset \cdots \subset V_n$, exactly one of each dimension, and V_1 is spanned by $e \equiv e_1$. The corresponding generalized eigenvectors e_j satisfy $Ae_j = e_{j-1}$ (of course, we have some flexibility in choosing them), and V_k is spanned by $\{e_i\}_{i \le k}$. Similarly, we have left generalized eigenvectors for B, f_i, and the only k-dimensional left invariant subspace of B is spanned by $\{f_j\}_{j \le k}$.

Next, the Jordan forms of A and B are the single block, J with zero on the diagonal. Suppose that $fe \neq 0$. We claim that there are no invertible solutions to (4.5) if $k > 0$. Let J be the Jordan form of the restriction of A to the k-dimensional subspace. Of course, it must be the single block with zero along the main diagonal, and similarly, the restriction of B has the same Jordan form. We note that $(\mathcal{F})_{11} = fe \neq 0$; however, $(J^T V - VJ)_{11}$ *is* zero for any V, as a simple computation reveals.

The outcome is that if $fe \neq 0$, there are no nontrivial solutions to (4.5), hence to (4.1).

We can extend this result to simply require that the spectra of A and B consist of the same single point (i.e., dropping the single Jordan block hypothesis), but we have to require that $fe \neq 0$ for *all* choices of left eigenvectors f of B and right eigenvectors e of A.

Corollary 4.1. *If A and B have the same one point spectrum, then either the only solution to (4.1) is trivial, or there is a line of rank one solutions. The latter occurs if and only if for some left eigenvector f of B and right eigenvector e of A, $fe = 0$.*

On the other hand, if any $fe = 0$, then there is a line of rank one solutions, as we have already seen.

5. Isolated invariant subspaces

Let A be an $n \times n$ matrix. An A-invariant subspace, H_0, is *isolated* (see [5]) if there exists $\delta > 0$ such that for all other invariant subspaces, H, $d(H, H_0) > \delta$, where $d(\cdot, \cdot)$ is the usual metric on the unit spheres, that is, $\inf \|h - h_0\|$ where h varies over the unit sphere of H and h_0 over the unit sphere of H_0, and the norm (for calculating the unit spheres and for the distance) is inherited from $\mathbf{C}^n$. There are several possible definitions of isolated (or its negation, nonisolated), but they all agree.

If $H_\alpha \to H_0$ (i.e., H is not isolated), then a cofinal set of H_αs are A-module isomorphic to H_0, and it will follow from the argument below (but is easy to see directly) that if we have a Jordan basis for H_0, we can simultaneously approximate it by Jordan bases for the H_α.

We use the notation $J(z,k)$ for the Jordan block of size k with eigenvalue z.

Lemma 5.1. *Suppose that A has only one eigenvalue, z. Let V be an isolated A-invariant subspace of $\mathbf{C}^n$. Then $V = \ker(A - z\mathrm{I})^r$ for some integer r. Conversely, all such kernels are isolated invariant subspaces.*

Proof. We may suppose that $A = \bigoplus_s J(z, n(s))$, where $\sum n(s) = n$. Let V_s be the corresponding invariant subspaces, so that $\mathbf{C}^n = \oplus V_s$ and $A \mid V_s = J(z, n(s))$. We can find an A-module isomorphism from V to a submodule of $\mathbf{C}^n$ so that the image of V is $\oplus W_s$ where each $W_s \subseteq V_s$ (this is standard in the construction of the Jordan forms). We may assume that V is already in this form.

Associate to V the tuple $(m(s) := \dim W_s)$. We will show that V is isolated if and only if

(1) $m(s) \neq n(s)$ implies that $m(s) \geq m(t)$ for all t.

Suppose (1) fails. Then there exist s and t such that $m(s) < m(t), n(s)$. We may find a basis for V_s, $\{e_i\}_{i=1}^{n(s)}$ such that $Ae_i = ze_i + e_{i-1}$ (with usual convention that $e_0 = 0$). Since W_s is an invariant subspace of smaller dimension, $\{e_i\}_{i=1}^{m(s)}$ is a basis of W_s ($A \mid V_s$ is a single Jordan block, so there is a unique invariant subspace for each dimension). Similarly, we find a Jordan basis $\{e_i^o\}_{i=1}^{m(t)}$ for W_t.

Define a map of vector spaces $\psi : W_t \to V_s$ sending $e_i^o \mapsto e_{i-m(t)+m(s)+1}$ (where $e_{<0} = e_0 = 0$). Then it is immediate (from $m(t) > m(s) < n(t)$) that ψ is an A-module homomorphism with image $W_s + e_{m(s)+1}\mathbf{C}$. Extend ψ to a map on W by setting it to be zero on the other direct summands. For each complex number α, define $\phi_\alpha : W \to V$ as $\mathrm{id} + \alpha\psi$. Each is an A-module homomorphism, moreover, the kernels are all zero (if $\alpha \neq 0$, then $w = -\alpha\psi(w)$ implies $w \in V_s$, hence $\psi(w) = 0$, so w is zero). Thus $\{H_\alpha := \mathrm{Rg}\,\phi_\alpha\}$ is a family of A-invariant subspaces, and as $\alpha \to 0$, the corresponding subspaces converge to $H_0 = W$, and moreover, the obvious generalized eigenvectors in H_α converge to their counterparts in W (this is a direct way to prove convergence of the subspaces).

Now we observe that the H_α are distinct. If $H_\alpha = H_\beta$ with $\alpha \neq \beta$, then $(\beta - \alpha)e_{m(s)+1}$ is a difference of elements from each, hence belongs to both. This forces $e_{m(s)+1}$ to belong

to H_α; by A-invariance, each of e_i $(i \le m(s))$ do as well, but it easily follows that the dimension of H_α is too large by at least one.

Next, we show that (1) entails $V = \ker(A - zI)^r$ for some nonnegative integer r. We may write $V = \oplus Z_s$ where $Z_s \subset Y_s$ are indecomposable invariant subspaces and $\mathbf{C}^n = \oplus Y_s$. Now $(A - zI)^r$ on each block Y_s simply kills the first r generalized eigenvectors and shifts the rest down by r. Hence $\ker(A - zI)^r \cap Z_s$ is the invariant subspace of dimension r or if $r > \dim Z_s$, $Z_s \subseteq \ker(A - zI)^r$. In particular, set $r = \max m(s)$; the condition (1) says that $W_s = V_s$ if $\dim W_s < r$ and $\dim W_s = r$ otherwise. Hence $W \subseteq \ker(A - zI)^r$, but has the same dimension. Hence $W = \ker(A - zI)^r$. It follows easily that $V \subseteq \ker(A - zI)^r$ (from being isomorphic to the kernel), and again by dimension, they must be equal.

Conversely, the module $\ker(A - zI)^r$ cannot be isomorphic to any submodule of $\mathbf{C}^n$ other than itself, so it cannot be approximated by submodules. □

When there is more than one eigenvalue, it is routine to see that the isolated subspaces are the direct sums over their counterparts for each eigenvalue.

Corollary 5.2. *Let A be an $n \times n$ matrix with minimal polynomial $p = \prod(x - z_i)^{m(i)}$. Then the isolated invariant subspaces of $\mathbf{C}^n$ are of the form $\ker(\prod(A - z_i I)^{r(i)})$ where $0 \le r(i) \le m(i)$, and these give all of them (and different choices of $(r(1), r(2), \ldots)$ yield different invariant subspaces).*

In [5], convergence of invariant subspaces is developed, and this result also follows from their work.

An obvious consequence (which can be proved directly) is that all A-invariant subspaces are isolated if and only if A is nonderogatory. In this case, if the Jordan block sizes are $b(i)$, the number of invariant subspaces is $\prod(b(i) + 1)$, and if A has distinct eigenvalues (all blocks are size 1), the number is 2^n. In the latter case, the number of invariant subspaces of dimension k is $C(n,k)$ (standard shorthand for $\binom{n}{k}$), but in the former case, the number is a much more complicated function of the block sizes. It is however, easy to see that for *any* choice of A, the number of isolated invariant subspaces of dimension k is at most $C(n,k)$, with equality if and only if A has distinct eigenvalues.

Now we can discuss the sources of continua of solutions to (4.1). Pick a (left) B-invariant subspace of $\mathbf{C}^{1\times n}$, W, and an A-invariant subspace, V, of $\mathbf{C}^n$, and suppose that $\dim V = \dim W = k$. Let $A_V = A \mid V$ and $B_W = W \mid B$, and select Jordan bases for W and V as we have done earlier (with $W = \ell\mathrm{Rg}U$ and $V = \mathrm{Rg}\,U$), and form the matrices $\mathcal{F} = (f_i e_j)$, and J_A, J_B, the Jordan normal forms of A_V and B_W, respectively. Let $\mathcal{R}$ denote the operator $\mathcal{R} : \mathbf{C}^k \to \mathbf{C}^k$ sending Z to $J_B^T Z - Z J_A$. There are several cases.

(i) If there are no invertible solutions Z to $\mathcal{R}(Z) = \mathcal{F}$, there is no solution U to (4.1) with $W = \ell\mathrm{Rg}U$ and $V = \mathrm{Rg}\,U$.

(ii) If $\mathrm{spec}\,A_V \cap \mathrm{spec}\,B_W = \emptyset$, then there is exactly one solution to $\mathcal{R}(Z) = \mathcal{F}$; however, if it is not invertible, (i) applies; otherwise, there is exactly one solution U to (4.1) with $W = \ell\mathrm{Rg}U$ and $V = \mathrm{Rg}\,U$.

(iii) If $\mathrm{spec}\,A_V \cap \mathrm{spec}\,B_W$ is not empty, and there is an invertible solution to $\mathcal{R}(Z) = \mathcal{F}$, then there is an open topological disk (i.e., homeomorphic to the open unit disk in $\mathbf{C}$) of such solutions, hence a disk of solutions U to (4.1) with $W = \ell\mathrm{Rg}U$ and $V = \mathrm{Rg}\,U$.

The third item is a consequence of the elementary fact that a sufficiently small perturbation of an invertible matrix is invertible. There is another (and the only other) source of continua of solutions.

(iv) Suppose that either W or V is not isolated (as a left B- or right A-invariant subspace, resp.), and also suppose that $\mathcal{R}(Z) = \mathcal{F}$ has an invertible solution. Then there exists a topological disk of solutions to (4.1) indexed by a neighborhood of subspaces that converge to the space that is not isolated.

To see this, we note that if (say) V is the limit (in the sense we have described) of invariant V_α (with $\alpha \to 0$, then in the construction of Lemma 5.1 (to characterize the isolated subspaces), the index set was $\mathbf{C}$, and the corresponding Jordan bases converged as well. Thus the matrices $\mathcal{F}_\alpha$ (constructed from the Jordan bases) will also converge. Since the solution at $\alpha = 0$ is invertible, we can easily find a neighbourhood of the origin on which each of $\mathcal{R}(V) = \mathcal{F}_\alpha$ can be solved, noting that the Jordan matrices do not depend on α.

We can rephrase these results in terms of the mapping $\Psi : U \mapsto (\ell\mathrm{Rg}\, U, \mathrm{Rg}\, U)$ from solutions of (4.1) to the set of ordered pairs of equidimensional left B- and right A-invariant subspaces.

Corollary 5.3. *If* $\mathrm{spec}\, A \cap \mathrm{spec}\, B = \emptyset$, *then* Ψ *is one to one.*

Proposition 5.4. *Suppose that for some integer* k, *(4.1) has more than* $C(n,k)^2$ *solutions of rank* k. *Then (4.1) has a topological disk of solutions. In particular, if (4.1) has more than* $C(2n,n)$ *solutions, then it has a topological disk of solutions.*

Proof. If (W,V) is in the range of Ψ but $\mathrm{spec}\, A_V \cap \mathrm{spec}\, B_W$ is not empty, then we are done by (iii). So we may assume that for every such pair in the range of Ψ, $\mathrm{spec}\, A_V \cap \mathrm{spec}\, B_W$ is empty. There are at most $C(n,k)$ A-invariant isolated subspaces of dimension k, and the same for B. Hence there are at most $C(n,k)^2$-ordered pairs of isolated invariant subspaces of dimension k. By (ii) and the spectral assumption, there are at most $C(n,k)^2$ solutions that arise from the pairs of isolated invariant subspaces. Hence there must exist a pair (W,V) in the range of Ψ such that at least one of W and V is not isolated. By (iv), there is a disk of solutions to (4.1).

Vandermonde's identities include $\sum C(n,k)^2 = C(2n,n)$; hence if the number of solutions exceeds $C(2n,n)$, there must exist k for which the number of solutions of rank k exceeds $C(n,k)^2$. □

This numerical result is well known in the theory of quadratic matrix equations.

In case C and D commute, the corresponding numbers are 2^n (in place of $C(2n,n) \sim 4^n/\sqrt{\pi n}$) and $C(n,k)$ (in place of $C(n,k)^2$). Of course, $2^n = \sum C(n,k)$ and $C(2n,n) = \sum C(n,k)^2$. The numbers $C(2n,n)$ are almost as interesting as their close relatives, the Catalan numbers $(C(2n,n)/(n+1))$; in particular, their generating function, $\sum C(2n,n)x^n$, is easier to remember—it is $(1-4x)^{-1/2}$, and so $\sum_{k=0}^{n} C(2k,k)C(2(n-k),n-k) = 4^n$.

Proposition 5.5. *Let* A *and* B *be invertible matrices of size* n. *Consider the following conditions.*

(a) A *has no algebraic multiple eigenvalues.*
(b) B *has no algebraic multiple eigenvalues.*
(c) $\mathrm{spec}\, A \cap \mathrm{spec}\, B = \emptyset$.

If all of (a)–(c) hold, then $U^2 = UB - AU$ has at most $C(2n,n)$ solutions.

Conversely, if the number of solutions is finite but at least as large as $3C(2n,n)/4$, then each of (a)–(c) must hold.

Proof. Condition (c) combined with (ii) entails that the solutions are a subset of the pairs of equidimensional invariant subspaces. However, (a) and (b) imply that the number of invariant subspaces of dimension k is at most $C(n,k)$, and the result follows from the simplest of Vandermonde's identities, $\sum C(n,k)^2 = C(2n,n)$.

Finiteness of the solutions says that there is no solution associated to a pair of invariant subspaces with either one being nonisolated. So solutions only arise from pairs of isolated invariant subspaces. If there were more than one solution arising from a single pair, then there would be a continuum of solutions by (ii) and (iii). Hence there can be at most one solution from any permissible pair of isolated subspaces, and moreover, when a solution does yield a solution, the spectra of the restrictions are disjoint.

As a consequence, there are at least $3C(2n,n)/4$ pairs of equidimensional invariant isolated subspaces on which the restrictions of the spectra are disjoint. Suppose that A has an algebraic multiple eigenvalue. It is easy to check that the largest number of isolated invariant subspaces of dimension k that can occur arises when it has one Jordan block of size two, and all the other blocks come from distinct eigenvalues (distinct from each other and the eigenvalue in the 2-block), and the number is $C(n-2,k-2) + C(n-2,k-1) + C(n-2,k)$ (with the convention $C(m,t) = 0$ if $t \notin \{0,1,\ldots,m\}$). The largest possible number of invariant isolated subspaces for B is $C(n,k)$ (which occurs exactly when B has no multiple eigenvalues), so we have at most $\sum C(n,k)(C(n-2,k-2) + C(n-2,k-1) + C(n-2,k))$ pairs of equidimensional isolated invariant subspaces. Of course

$$\begin{aligned}
\sum_{2\le k\le n} C(n,k)C(n-2,k-2) &= C\big(2(n-1),n-2\big),\\
\sum_{1\le k\le n-1} C(n,k)C(n-2,k-1) &= C\big(2(n-1),n-1\big),\\
\sum_{0\le k\le n-2} C(n,k)C(n-2,k) &= C\big(2(n-1),n\big),
\end{aligned} \tag{5.1}$$

which are the middle three terms of the even rows of Pascal's triangle. The sum of these terms divided by $C(2n,n)$ is exactly $(3n-2)/(4n-2)$, which is less than $3/4$. This yields that A must have distinct eigenvalues. Obviously, this also applies to B as well.

If μ belongs to $\operatorname{spec} A \cap \operatorname{spec} B$, then the left eigenvector of B and the right eigenvector of A for μ cannot simultaneously appear as elements of the pair of invariant supspaces giving rise to a solution of (4.1), that is, if the left B-invariant subspace is Z and the right A-invariant subspace is Y, we cannot simultaneously have the left eigenvector (of B) in Z and the right eigenvector (of A) in Y (because the only contributions to solutions come from pairs of isolated subspaces on which the restrictions have disjoint spectra). As both A and B have distinct eigenvectors, their subspaces of dimension k are indexed by the $C(n,k)$ subsets of k elements in a set with n elements (specifically, let the n-element set consist of n eigenvectors for the distinct eigenvalues, and let the invariant subspace be the span of the k-element subspace).

However, we must exclude the situation wherein both invariant subspaces contain specific elements. The number of such pairs of k element sets is $C(n,k)^2 - C(n-1,k-1)^2$. Summing over k, we obtain at most $C(2n,n) - C(2(n-1),n-1)$ which is (again, just barely) less than $3C(2n,n)/4$. (The ratio $C(2n-2,n-1)/C(2n,n)$ is $n/(4n-2) > 1/4$, which is just what we need here, but explains why simply bounding the sum of three terms above by the middle one does not work.) □

From the computation of the ratio in the last line of the proof and the genericity, 3/4 is sharp asymptotically, but for specific n we may be able to do slightly better.

6. Changing solutions

Begin with the problem (4.1), $U^2 = UB - AU$, and let U_0 be a solution. Consider the new problem

$$U^2 = U(B - U_0) - (A + U_0)U. \tag{2_0}$$

The pair (A,B) has been replaced by the pair $(A + U_0,\ B - U_0)$. In terms of our original problem, the new equation corresponds to referring to the fixed points from $X + Y_0$ ($U_0 = CY_0$), rather than from the original X. In other words, when we translate back to our original fixed point problem, we are using a different fixed point to act as the start-up point, to the same $\phi_{C,D}$. Specifically, if U_1 is also a solution to (4.1), then the difference $U_1 - U_0$ is a solution of (2_0) (direct verification). Thus the affine mapping $U_1 \mapsto U_1 - U_0$ is a bijection from the set of solutions to (4.1) to the set of solutions of (2_0).

We will see that this leads to another representation of the fixed points as a subset of size n of a set of size $2n$ (recalling the bound on the number of solutions is $C(2n,n)$ which counts the number of such subsets).

First, we have the obvious equation $(A + U_0)U_0 = U_0B$. This means that U_0 implements a "partial" isomorphism between left invariant subspaces for $A + U_0$ and B, via $Z \mapsto ZU_0$ for Z a left-invariant $A + U_0$-module—if $Z(A + U_0) \subset Z$, then $ZU_0B = Z(A + U_0) \subseteq ZU_0$. If we restrict the Z to those for which $Z \cap \ell\ker U_0 = \emptyset$, then it is an isomorphism with image the invariant subsets of B that lie in the left range of U_0. On the other hand, $A + U_0$ restricted to $\ell\ker U_0$ agrees with A, and of course, $(\ell\ker U_0)A \subseteq (\ell\ker U_0)$. In particular, the spectrum of $A + U_0$ agrees with that of A on the left A-invariant subspace $\ell\ker U_0$, and acquires part of the spectrum of B (specifically, the spectrum of $\ell\mathrm{Rg}U_0|B$). It is not generally true that $\ell\ker U_0 + \ell\mathrm{Rg}U_0 = \mathbf{C}^{1\times n}$, even if $\operatorname{spec}A \cap \operatorname{spec}B = \emptyset$.

However, suppose that $\operatorname{spec}A \cap \operatorname{spec}B = \emptyset$. Let $k = \operatorname{rank}U_0$. Then including algebraic multiplicities, $\ell\ker U_0 \mid A + U_0$ has $n - k$ eigenvalues of A, and we also obtain k eigenvalues of B in the spectrum of $A + U_0$ from the intertwining relation. Since the spectra of A and B are assumed disjoint, we have accounted for all n (algebraic) eigenvalues of $A + U_0$. So the spectrum (algebraic) of $A + U_0$ is obtained from the spectra of A and B, and the "new" algebraic eigenvalues, that is, those from B, are obtained from the intertwining relation.

Now we attempt the same thing with $B - U_0$. We note the relation $U_0(B - U_0) = AU_0$; if Z is a right $B - U_0$-invariant subspace, then U_0Z is an A-invariant subspace, so that $A \mid \mathrm{Rg}\,U_0$ (the latter is an A-invariant subspace) is similar to B suitably restricted. Obviously, $\ker U_0$ is right B-invariant, and $(B - U_0) \mid \ker U_0$ agrees with $B \mid \ker U_0$. So again

the algebraic spectrum of $B - U_0$ is a hybrid of the spectra of A and B, and $B - U_0$ has acquired k of the algebraic eigenvalues of A (losing a corresponding number from B, of course).

If we assume that the eigenvalues of A are distinct, as are those of B, in addition to being disjoint, then we can attach to U_0 a pair of subsets of size k (or one of size k, the other of size $n-k$) of sets of size n. Namely, take the k eigenvalues of $A + U_0$ that are not in the algebraic spectrum of A (the first set), and the k eigenvalues of $B - U_0$ that are not in the algebraic spectrum of B.

If we now assume that there are at most finitely many solutions to (4.1), from cardinality and the sources of the eigenvalues, then different choices of solutions U_0 yield different ordered pairs. One conclusion is that if there are the maximum number of solutions to (4.1) (which forces exactly the conditions we have been imposing, neither A nor B has multiple eigenvalues, and their spectra have empty intersection), then every possible pair of k-subsets arises from a solution. To explain this, index the eigenvalues of A as $\{\lambda_i\}$ and those of B as $\{\mu_j\}$ where the index set for both is $\{1,2,\ldots,n\}$. Pick two subsets R, S of size k of $\{1,2,\ldots,n\}$. Create a new pair of sets of eigenvalues by interchanging $\{\lambda_i \mid i \in S\}$ with $\{\mu_j \mid j \in R\}$ (i.e., remove the λs in S from the first list and replace by the μs in R, and vice versa). Overall, the set of λs and μ is the same, but has been redistributed in the eigenvalue list. Then there is a solution to (4.1) for which $A + U_0$ and $B - U_0$ have, respectively, the new eigenvalue list.

7. Graphs of solutions

For each integer $n \geq 2$, we describe a graph $\mathcal{G}_n$ with $C(2n,n)$ vertices. Then we show that if there are finitely many fixed points of $\phi_{C,D}$, there is a saturated graph embedding from the graph of the fixed points to $\mathcal{G}_n$ (an embedding of graphs $\Xi : \mathcal{G} \to \mathcal{H}$ is *saturated* if whenever h and h' are vertices in the image of Ξ and there is an edge in $\mathcal{H}$ from h to h', then there is an edge between the preimages). In particular, $\mathcal{G}_n$ is the generic graph of the fixed points.

Define the vertices in $\mathcal{G}_n$ to be the members of

$$\{(R,S) \mid R,S \subseteq \{1,2,3,\ldots,n\},\ |R| = |S|\}. \tag{7.1}$$

If (R,S) is such an element, we define its *level* to be the cardinality of R. There is only one level zero element, obviously $(\emptyset,\emptyset)$, and only one level n element, $(\{1,2,3,\ldots,n\}, \{1,2,3,\ldots,n\})$, and of course there are $C(n,k)^2$ elements of level k.

The edges are defined in three ways: moving up one level, staying at the same level, or dropping one level. Let (R,S) and (R',S') be two vertices in $\mathcal{G}_n$. There is an edge between them if and only if one of the following hold:

(a) there exist $r_0 \notin R$ and $s_0 \notin S$ such that $R' = R \cup \{r_0\}$ and $S' = S \cup \{s_0\}$;

(bi) $S' = S$ and there exist $r \in R$ and $r_0 \notin R$ such that $R' = (R \setminus r) \cup \{r_0\}$;

(bii) $R' = R$ and there exist $s \in S$ and $s_0 \notin S$ such that $S' = (S \setminus s) \cup \{s_0\}$;

(c) there exist $r \in R$ and $s \in S$ such that $R' = R \setminus \{r\}$ and $S' = S \setminus \{s\}$.

Note that if (R,S) is of level k, there are $(n-k)^2$ choices for (R',S') of level $k+1$ (a), k^2 of level $k-1$ (c), and $2k(n-k)$ of the same level (bi) & (bii). The total is n^2, so this is the valence of the graph (i.e., the valence of every vertex happens to be the same).

For $n=2$, $\mathcal{G}_2$ is the graph of vertices and edges of the regular octahedron. When $n=3$, $\mathcal{G}_3$ has 20 vertices and valence 9 is the graph of (the vertices and edges of) a 5-dimensional polytope (not regular in the very strong sense) is relatively easy to be described as a graph (the more explicit geometric realization comes later). The zeroth level consists of a single point, and the first level consists of 9 points arranged in a square, indexed as (i,j). The next level consists of 9 points listed as $(\hat{i},\hat{j})$ where $\hat{i}$ is the complement of the singleton set $\{i\}$ in $\{1,2,3\}$. The fourth level of course again consists of a singleton. The edges from the point (i,j) terminate in the points (k,l) in either the same row or the same column (i.e., either $i=k$ or $j=1$) and in the points $(\hat{p},\hat{q})$ where $p\neq i$ and $q\neq j$, and finally the bottom point. The graph is up-down symmetric.

The graph $\mathcal{G}_n$ is a special case of a *Johnson graph,* specifically $J(n,2n)$ [6] which in this case can be described as the set of subsets of $\{1,2,3,\dots,2n\}$ of cardinality n, with two such subsets connected by an edge if their symmetric difference has exactly two elements. Spectra of all the Johnson graphs and their relatives are worked out in [7]. We can map $\mathcal{G}_n$ to this formulation of the Johnson graph via $(R,S)\mapsto(\{1,2,3,\dots,n\}\setminus R)\cup(n+S)$. The (R,S) formulation is easier to work with in our setting.

Now let $\mathcal{G}\equiv\mathcal{G}_{A,B}$ denote the graph of the solutions to (4.1). Recall that the vertices are the solutions, and there is an edge between two solutions, U_0 and U_1, if the difference U_0-U_1 is a rank one matrix. Assume to begin with that both A and B have distinct eigenvalues, and their spectra have nothing in common. Pick complete sets of n eigenvectors for each of A and B (left eigenvectors for B, right for A), and index them by $\{1,2,\dots,n\}$. Every invariant subspace of A (B) is spanned by a unique set of eigenvectors. So to each solution U_0 of (4.1), we associate the eigenvectors appearing in $\operatorname{Rg}U_0$ and $\ell\operatorname{Rg}U_0$; this yields two equicardinality subsets of $\{1,2,\dots,n\}$, hence the pair (R,S). We also know that as a map on sets, this is one to one, and will be onto provided the number of solutions is $C(2n,n)$.

Next we verify that the mapping associating (R,S) to U_0 preserves the edges. The first observation is that if U_1 is the other end of an edge in $\mathcal{G}$, then the rank of U_1 can only be one of $\operatorname{rank}U_0-1$, $\operatorname{rank}U_0$, and $\operatorname{rank}U+1$, which means that the level of the vertex associated to U_1 either equals or is distance one from that associated to U_0. Now let us return to the formalism of Section 4.

We can reconstruct U_0 as $\sum_{(i,j)\in R\times S}\mu_{ij}e_if_j$ for some coefficients μ_{ij}, where we recall that e_i are the right eigenvectors of A and f_j are the left eigenvectors of B. Similarly, $U_1=\sum_{(i,j)\in R'\times S'}\mu'_{ij}e_if_j$. We wish to show that if U_0-U_1 has rank one, then (R,S) and (R',S') are joined by an edge in $\mathcal{G}_n$.

As we did earlier, we can write $U_0=\sum e_iw_i$ (where $w_i=\sum_i\mu_{ij}f_j$) and $U_1=\sum e_iw'_i$. Then U_1-U_0 breaks up as

$$\sum_{i\in R\cap R'}e_i(w'_i-w_i)+\sum_{i\in R'\setminus R}e_iw'_i-\sum_{i\in R\setminus R'}e_iw_i. \tag{7.2}$$

Since the set $\{e_i\}$ is linearly independent and U_1-U_0 is rank one, all of w'_i-w_i ($i\in R\cap R'$), w'_i ($i\in R'\setminus R$), and w_i ($i\in R\setminus R'$) must be multiples of a common vector (apply

(i)–(iii) of Section 4 to any pair of them). However, we note that the w_i' are the "columns" of the matrix (μ_{ij}'), hence constitute a linearly independent set. It follows immediately that $R' \setminus R$ is either empty or consists of one element. Applying the same reasoning to w_i, we obtain that $R \setminus R'$ is either empty or has just one element. Of course, similar considerations apply to S and S'.

We have $|R| = |S|$ and $|R'| = |S'|$. First consider the case that $R = R'$. Then $|S'| = |S|$ and the symmetric difference must consist of exactly two points, whence (R,S) is connected to (R',S'). Similarly, if $S = S'$, the points are connected.

Now suppose $|R| = |R'|$. We must exclude the possibility that both symmetric differences (of R, R' and S, S') consist of two points. Suppose that $k \in R \setminus R'$ and $l \in R' \setminus R$. Then the set of vectors $\{w_i - w_i'\}_{i \in R \cap R'} \cup \{w_k, w_l'\}$ span a rank one space. Since w_k and w_l' are nonzero (they are each columns of invertible matrices), this forces $w_k = r w_l'$ for some nonzero scalar r, and $w_i - w_i' = r_i w_l'$ for some scalars r_i. Hence the span of $\{w_j\}$ is contained in the span of $\{w_j'\}$. By dimension, the two spans are equal.

However, $\operatorname{span}\{w_j\}$ is spanned by the eigenvectors affiliated to S, while $\operatorname{span}\{w_j'\}$ is spanned by the eigenvectors affiliated to S'. Hence we must have $S = S'$.

Next suppose that $|R| < |R'|$. As each of $R \setminus R'$ and $R' \setminus R$ can consist of at most one element, we must have $R' = R \cup \{k\}$ for some $k \notin R$. Also by $|S| = |R| < |R'| = |S'|$, we can apply the same argument to S and S', yielding that S' is S with one element adjoined. Hence (R,S) is connected to (R',S').

Finally, the case that $|R| > |R'|$ is handled by relabelling and applying the preceding paragraph.

This yields that the map from the graph of solutions to $\mathcal{G}_n$, $U_0 \mapsto (R,S)$ is a graph embedding. Next we show that it is saturated, meaning that if $U_0 \mapsto (R,S)$ and $U_1 \mapsto (R_1,S_1)$, and (R,S) is connected to (R_1,S_1) in $\mathcal{G}_n$, then $\operatorname{rank}(U_1 - U_0) = 1$. This is rather tricky, since the way in which rank one matrices are added to U_0 to create new solutions is complicated. Note, however, if the valence of every point in the graph of solutions is n^2 (i.e., there exists the maximum number of eigenvectors for both matrices with nonzero inner products), then the mapping is already a graph isomorphism.

We remind the reader that the condition $|\operatorname{spec} A \cup \operatorname{spec} B| = 2n$ remains in force. First we observe

$$\begin{aligned} \operatorname{rank} U_0 &= \left| \operatorname{spec} A \setminus \operatorname{spec}(A + U_0) \right| \\ &= \left| \operatorname{spec}(A + U_0) \setminus \operatorname{spec} A \right| \\ &= \left| \operatorname{spec} B \setminus \operatorname{spec}(B - U_0) \right| \\ &= \left| \operatorname{spec}(B - U_0) \setminus \operatorname{spec} B \right|. \end{aligned} \tag{7.3}$$

The first two equalities follow from the fact that the spectrum of $A + U_0$ is that of A with a subset removed and replaced by an equicardinal subset of B; what was removed from the spectrum of A appears in the spectrum of $B - U_0$.

Now suppose that (R,S) is connected to (R',S') in $\mathcal{G}_n$, and suppose that $U_0 \mapsto (R,S)$ and $U_1 \mapsto (R',S')$ for U_0 and U_1 in $\mathcal{G}$. We show that $|\operatorname{spec}(A + U_0) \setminus \operatorname{spec}(A + U_1)| = 1$. Without loss of generality, we may assume that $R = S = \{1,2,\ldots,k\} \subset \{1,2,\ldots,n\}$. Index the eigenvalues λ_i, μ_j, respectively, for the e_i, f_j right and left eigenvectors of A, B.

In particular, $\operatorname{spec}(A+U_0)=\{\mu_1,\mu_2,\ldots,\mu_k,\lambda_{k+1},\ldots,\lambda_n\}$, obtained by replacing $\{\lambda_i\}_{i=1}^k$ by $\{\mu_i\}_{i=1}^k$.

(i) $R=R'$. Without loss of generality, we may assume that $S'=(S\setminus\{k\})\cup\{k+1\}$. Then $\operatorname{spec}(A+U_1)$ is obtained by swapping the eigenvalues corresponding to R with those corresponding to S', that is, $\operatorname{spec}(A+U_1)=\{\mu_1,\mu_2,\ldots,\mu_{k-1},\mu_{k+1},\lambda_{k+1},\ldots,\lambda_n\}$. Then $\operatorname{spec}(A+U_0)\setminus\operatorname{spec}(A+U_1)=\{\mu_k\}$, and so $|\operatorname{spec}(A+U_0)\setminus\operatorname{spec}(A+U_1)|=1$.

(ii) $S=S'$. Without loss of generality, we may assume that $R'=(R\setminus\{k\})\cup\{k+1\}$. Then $\{\lambda_1,\ldots,\lambda_{k-1},\lambda_{k+1}\}$ is swapped with $\{\mu_i\}_{i=1}^k$, and so $\operatorname{spec}(A+U_1)=\{\mu_1,\mu_2,\ldots,\mu_k,\lambda_k,\lambda_{k+2},\ldots,\lambda_n\}$. Thus $\operatorname{spec}(A+U_0)\setminus\operatorname{spec}(A+U_1)=\{\lambda_{k+1}\}$, and again $|\operatorname{spec}(A+U_0)\setminus\operatorname{spec}(A+U_1)|=1$.

(iii) $R\neq R'$ & $S\neq S'$. By interchanging the roles of the primed/unprimed sets if necessary, and then relabelling, we may assume that $R'=R\cup\{k+1\}$ and $S'=S\cup\{k+1\}$. Then $\operatorname{spec}(A+U_1)=\{\mu_1,\mu_2,\ldots,\mu_{k+1},\lambda_{k+2},\ldots,\lambda_n\}$ and thus $\operatorname{spec}(A+U_0)\setminus\operatorname{spec}(A+U_1)=\{\lambda_{k+1}\}$, and once more $|\operatorname{spec}(A+U_0)\setminus\operatorname{spec}(A+U_1)|=1$.

Now the equation $U^2=U(B-U_0)-(A+U_0)U$ has solution U_1-U_0 and $|\operatorname{spec}(A+U_0)\cup\operatorname{spec}(B-U_0)|=2n$, so $\operatorname{rank}(U_1-U_0)|\operatorname{spec}(A+U_0)\setminus\operatorname{spec}(A+U_1)|=1$. Thus U_1 is connected to U_0 within $\mathcal{G}$. □

Theorem 7.1. *If $|\operatorname{spec}A\cup\operatorname{spec}B|=2n$, then the map $\mathcal{G}\to\mathcal{G}_n$ given by $U_0\mapsto(R,S)$ is well defined and a saturated graph is embedding.*

Now we will show some elementary properties of the graph $\mathcal{G}$.

Proposition 7.2. *Suppose that $|\operatorname{spec}A\cup\operatorname{spec}B|=2n$.*

(a) *Every vertex in $\mathcal{G}$ has valence at least n.*

(b) *If one vertex in $\mathcal{G}$ has valence exactly n, then A and B commute, and $\mathcal{G}$ is the graph (vertices and edges) of the n-cube. In particular, all vertices have valence n, and there are $C(n,k)$ solutions of* rank k.

Proof. (a) Let e_i, f_j be right, left A, B eigenvectors. Let $\{\epsilon_j\}\subset\mathbf{C}^{1\times n}$ be the dual basis for $\{e_i\}$, that is, $\epsilon_j(e_i)=\delta_{ij}$. We may write $f_j=\sum_i r_{jk}\epsilon_k$; of course the $k\times k$ matrix (r_{jk}) is invertible, since it transforms one basis to another. Therefore $\det(r_{jk})\neq 0$, so there exists a permutation on the n-element set, π, such that $\prod_j r_{j,\pi(j)}$ is not zero. Therefore

$$f_j e_{\pi(j)}=\sum_k r_{jk}\epsilon_k(e_{\pi(j)})=r_{j,\pi(j)}\neq 0. \tag{7.4}$$

Hence there exist nonzero scalars t_j such that $t_j e_j f_j$ are all nonzero solutions to $U^2=UB-AU$. Thus the solution $\mathbf{0}$ has valence at least n. However, this argument applies equally well to any solution U_0, by considering the modified equation $U^2=U(B-U_0)-(A+U_0)U$.

(b) Without loss of generality, we may assume the solution $\mathbf{0}$ has valence exactly n (by again considering $U^2=U(B-U_0)-(A+U_0)U$). From the argument of part (a), by relabelling the e_i, we may assume that $f_i e_i\neq 0$. Since there are exactly n and no more solutions, we must have $f_i e_j=0$ if $i\neq j$. By replacing each e_i by suitable scalar multiples of itself, we obtain that $\{f_i\}$ is the dual basis of $\{e_i\}$.

Now let U_1 be any solution. Then there exist subsets R and S of $\{1,2,\ldots,k\}$ such that $U_1 = \sum_{(i,j)\in R\times S} e_i f_j \mu_{ij}$ for some invertible matrix $\{\mu_{ij}\}$. From the dual basis property, we have $U_1^2 = \sum e_i f_l \mu_{ij}\mu_{jl}$, and so (4.1) yields (comparing the coefficients of $e_i f_l$) $M^2 = MD_1 - D_2M$ where D_1 is the diagonal matrix with entries the eigenvalues of B indexed by S, and D_2 corresponds to the eigenvalues of A indexed by R.

Write $A = \sum e_i f_j a_{ij}$; from $Ae_i = \lambda_i e_i$, we deduce A is diagonal with respect to this basis. Similarly, B is with respect to $\{f_j\}$, and since the latter is the dual basis, we see that they are simultaneously diagonalizable, in particular, they commute. It suffices to show that each solution U_1 is diagonal, that is, $\mu_{ij} = 0$ if $i \neq j$.

For $M^2 = MD_1 - D_2M$, we have as solutions diagonal matrices whose ith entry is either zero or $\mu_i - \mu_j$, yielding $C(n,k)$ solutions of rank k, and it is easy to see that the graph they form (together) is the graph of the n-cube. It suffices to show there are no other solutions. However, this is rather easy, because of the dual basis property—in the notation above, we cannot have an invertible $k \times k$ solution if $R \neq S$. □

8. Graph fine structure

If we drop the condition $|\operatorname{spec} A \cup \operatorname{spec} B| = 2n$, we can even have the number of solutions being 2^n without A and B commuting (or even close to commuting). This will come as a very special case from the analysis of the "nondefectivegraphs that can arise from a pair of $n \times n$ matrices (A,B).

Let $\mathbf{a} := a(1),a(2),\ldots,$ be an ordered partition of n, that is, $a(i)$ are positive integers and $\sum a(i) = n$. Let $\Lambda := (\lambda_i)$ be distinct complex numbers, in bijection with $a(i)$. Define block $(\mathbf{a},\Lambda)$ to be the Jordan matrix given as the direct sum of elementary Jordan blocks of size $a(i)$ with eigenvalue λ_i. When Λ is understood or does not need to be specified, we abbreviate block $(\mathbf{a},\Lambda)$ to block $(\mathbf{a})$.

Now let $\alpha := \{\alpha(i)\}_{i\in I}$ be an unordered partition of $2n$, and $L := \{t_i\}$ be a set of distinct nonzero complex numbers with the same index set. Pick a subset J of I with the property that $\sum_{j\in J}\alpha(j) \geq n$ but for which there exists an element j_0 in J such that $\sum_{j\in J\setminus\{j_0\}}\alpha(j) < n$. For each, such j_0 form the two partitions of n, the first one $\mathbf{a} = \{\alpha(i)\}_{j\in J\setminus\{j_0\}}, \{n - \sum_{j\in J\setminus\{j_0\}} a(j)\}$; the second one, $\mathbf{b}$ to be the partition given by the rest of $\alpha(j_0)$ and $\{\alpha(j)\}_{j\notin J}$. In particular, if $\sum_{j\in J}\alpha(j) = n$, the "rest of $\alpha(j_0)$" is empty.

For example, if $n = 6$ and $\alpha(i) = 3,5,3,1$, respectively, we can take $J = 1,2$, and have 2 left over; the two partitions are then $\mathbf{a} = 3,3$ and $\mathbf{b} = 2,3,1$. Of course, we can do this in many other ways, since we do not have to respect the order, except that if there is overlap, it is continued as the first piece of the second partition.

Now associate the pair of Jordan matrices by assigning t_i to the corresponding α_i, with the proviso that whichever t_{j_0} is assigned to both the terminal entry of the first partition of n and the "rest of it" in the second. Continuing our example, if $t_i = e,\pi,1,i$, the left Jordan matrix would consist of two blocks of size 3 with eigenvalues e and π, respectively, and the second would consist of three blocks of sizes 2, 3, 1 with corresponding eigenvalues π, 1, i.

Now suppose that each matrix A and B is nonderogatory (to avoid a trivial continuum of solutions).

A function $\mathbf{c} : \mathbf{C} \setminus \{0\} \to \mathbf{N}$ is called a *labelled partition of* N if $\mathbf{c}$ is zero almost everywhere, and $\sum c(\lambda) = N$. From a labelled partition, we can obviously extract an (ordinary) partition of N simply by taking the list of nonzero values of $\mathbf{c}$ (with multiplicities). This partition is the *type* of $\mathbf{c}$.

If $\mathbf{a}$ and $\mathbf{b}$ are labelled partitions of n, then $\mathbf{a}+\mathbf{b}$ is a labelled partition of $2n$. We consider the set of ordered pairs of labelled partitions of n, say $(\mathbf{a},\mathbf{b})$, and define an equivalence relation on them given by $(\mathbf{a},\mathbf{b}) \sim (\mathbf{a}',\mathbf{b}')$ if $\mathbf{a}+\mathbf{b} = \mathbf{a}'+\mathbf{b}'$.

Associated to a nonderogatory $n \times n$ matrix A is a labelled partition of n; assign to the matrix A the function $\mathbf{a}$ defined by

$$\mathbf{a}(\lambda) = \begin{cases} 0 & \text{if } \lambda \notin \operatorname{spec} A, \\ k & \text{if } A \text{ has Jordan block of size } k \text{ at } \lambda. \end{cases} \tag{8.1}$$

Analogous things can also be defined for derogatory matrices (i.e., with multiple geometric eigenvalues), but this takes us a little beyond where we want to go, and in particular heads towards the land of continua of solutions to (4.1).

To the labelled partition $\mathbf{c}$ of $2n$, we attach a graph $\mathcal{G}_{\mathbf{c}}$. Its vertices are the ordered pairs $(\mathbf{a},\mathbf{b})$ of ordered partitions of n such that $\mathbf{a}+\mathbf{b} = \mathbf{c}$, and there is an edge between $(\mathbf{a},\mathbf{b})$ and $(\mathbf{a}',\mathbf{b}')$ if $\sum |\mathbf{a}(\lambda) - \mathbf{a}'(\lambda)| = 2$. This suggests the definition of distance between two equivalent ordered pairs, $d((\mathbf{a},\mathbf{b}),(\mathbf{a}',\mathbf{b}')) = \sum |\mathbf{a}(\lambda) - \mathbf{a}'(\lambda)|$. The distance is always an even integer.

For example, if the type of $\mathbf{c}$ is the partition $(1,1,1,\ldots,1)$ with $2n$ ones (abbreviated 1^{2n}), then the ordered pairs of labelled partitions of size n correspond to the pairs of subsets (λ_r), (μ_s) each of size n, where the complex numbers λ_t, μ_s are distinct. Two such are connected by an edge if we can obtain one from the other by switching one of the λ_r with one of the μ_s. This yields the graph $\mathcal{G}_n$ constructed earlier in the case that A and B were diagonalizable and with no overlapping eigenvalues—the difference is that instead of concentrating on what subsets where altered (as previously, in using the solutions U_0), we worry about the spectra of the pair $(A+U_0,\ B-U_0)$.

If the type of $\mathbf{c}$ is the simple partition $2n$, then the only corresponding bitype is the pair of identical constant functions with value n, and the graph has just a single point. This corresponds to the pair of matrices A and B where each has just a single Jordan block (of size n) and equal eigenvalue. Slightly less trivial is the graph associated to the labelled partition whose type is $(2n-1,1)$. The unlabelled bitypes to which this corresponds can be written as

$$\begin{matrix} (n & 0), & (n-1 & 1), \\ (n-1 & 1), & (n & 0), \end{matrix} \tag{8.2}$$

each of which is to be interpreted as a pair of functions, for example, in the left example, the first function sends $\lambda_1 \mapsto n$ and $\lambda_2 \mapsto 0$, and the second sends $\lambda_1 \mapsto n-1$ and $\lambda_2 \mapsto 1$. The right object reverses the roles of the functions. The column sums are yield the original partition of $2n$, and the row sums are n. There are just two points in the graph, which has an edge joining them. This corresponds to the situation in which $|\operatorname{spec} A \cup \operatorname{spec} B| = 2$,

that is, one of the pair has a Jordan block of size n, the other has a Jordan block of size $n-1$ with the same eigenvalue as that of the other matrix, and another eigenvalue.

It is easy to check that if the type of $\mathbf{c}$ is $n+k, n-k$ for some $0 \le k < n$, then the graph is just a straight line, that is, vertices $v_0, v_1, \ldots, v_k$ with edges joining v_i to v_{i+1}. A particularly interesting case arises when the type is $(n, 1^n)$ (corresponding to A diagonalizable and B having a single Jordan block, but with eigenvalue not in the spectrum of A). Consider the bitypes

$$\begin{pmatrix} n-k & \cdots & 0 & \cdots & 1 & \cdots & \cdots \end{pmatrix}, \\ \begin{pmatrix} k & \cdots & 1 & \cdots & 0 & \cdots & \cdots \end{pmatrix}, \tag{8.3}$$

where there are k ones to the right of $n-k$ in the top row, and the ones in the bottom row appear only where zero appears above. These all yield the partition $n, 1^n$, so they are all equivalent, and it is easy to see that there are $C(n,k)$ different ones for each k. There are thus 2^n vertices in the corresponding graph. However, this graph is rather far from the graph of the power set of an n-element set, as we will see later (it has more edges).

Assume that (4.1) has a finite number of solutions for specific A and B. To each solution U_0, form $A + U_0$ and $B - U_0$, and associate the Jordan forms to each. We can think of the Jordan form as a labelled partition as above. We claim that the assignment that sends the solution U_0 to the pair of labelled partitions is a graph homomorphism from $\mathcal{G}$ (the graph of solutions of (4.1), edges defined by the difference being of rank one) to the graph of $\mathbf{c}$, where $\mathbf{c}$ is the sum of the of two labelled partitions arising from A and B.

For example, if $|\operatorname{spec} A \cup \operatorname{spec} B| = 2n$ as we had before, this assigns to the solution U_0 the pair consisting of the spectrum of $A + U_0$ and the spectrum of B_0, which differs from our earlier graph homomorphism. Notice, however, that the target graph is the same, a complicated thing with $C(2n,n)$ vertices and uniform valence n^2. (Valence is easily computed in all these examples, Proposition 9.3.)

Fix the labelled partition of $2n$, called $\mathbf{c}$. The graph associated to $\mathbf{c}$ is the collection of pairs of labelled partitions of n, $(\mathbf{a}, \mathbf{b})$ with constraint that $\mathbf{c} = \mathbf{a} + \mathbf{b}$. We define the distance between two such pairs in the obvious way

$$d((\mathbf{a},\mathbf{b}),(\mathbf{a}',\mathbf{b}')) = \sum_{\lambda \in \operatorname{supp} \mathbf{c}} |\mathbf{a}(\lambda) - \mathbf{a}'(\lambda)|. \tag{8.4}$$

Obviously, the values of the distance are even integers, with maximum value at most $2n$. We impose a graph structure by declaring an edge between $(\mathbf{a},\mathbf{b})$ and $(\mathbf{a}',\mathbf{b}')$ whenever $d((\mathbf{a},\mathbf{b}),(\mathbf{a}',\mathbf{b}')) = 2$; we use the notation $(\mathbf{a},\mathbf{b}) \approx (\mathbf{a}',\mathbf{b}')$. This is the same as saying that for two distinct complex numbers λ, μ, in the support of $\mathbf{c}$, $\mathbf{a}' = \mathbf{a} + \delta_\lambda - \delta_\mu$ (automatically, $\mathbf{b}' = \mathbf{b} + \delta_\mu - \delta_\lambda$). Note, however, that if $(\mathbf{a},\mathbf{b})$ is a pair of labelled partitions of n which add to $\mathbf{c}$, in order that $(\mathbf{a} + \delta_\lambda - \delta_\mu, \mathbf{b} + \delta_\mu - \delta_\lambda)$ be a pair of labelled partitions, we require that $\mathbf{a}(\mu) > 0$ and $\mathbf{b}(\lambda) > 0$.

Lemma 8.1. *Suppose that* $(\mathbf{a},\mathbf{b})$ *and* $(\mathbf{a}',\mathbf{b}')$ *are pairs of labelled partitions of* n *with* $\mathbf{a}+\mathbf{b} = \mathbf{a}'+\mathbf{b}' := \mathbf{c}$. *Suppose that* $d((\mathbf{a},\mathbf{b}),(\mathbf{a}',\mathbf{b}')) = 2k$. *Then there exist pairs of labelled partitions*

of n, $(\mathbf{a}_i, \mathbf{b}_i)$ *with* $i = 0, 1, \ldots, k$ *such that*

(0) $\mathbf{a}_i + \mathbf{b}_i = \mathbf{c}$ *for* $i = 0, 1, \ldots, k$;

(a) $(\mathbf{a}_0, \mathbf{b}_0) = (\mathbf{a}, \mathbf{b})$;

(b) $(\mathbf{a}_i, \mathbf{b}_i) \approx (\mathbf{a}_{i+1}, \mathbf{b}_{i+1})$ *for* $i = 0, 1, \ldots, k-1$;

(c) $(\mathbf{a}_k, \mathbf{b}_k) = (\mathbf{a}', \mathbf{b}')$.

Proof. Since $\mathbf{a}$ and $\mathbf{a}'$ are labelled partitions of the same number n, there exist distinct complex numbers λ and μ such that $\mathbf{a}(\mu) > \mathbf{a}'(\mu)$ and $\mathbf{a}(\lambda) < \mathbf{a}'(\lambda)$. Set $\mathbf{a}_1 = \mathbf{a} + \delta_\lambda - \delta_\mu$, and define $\mathbf{b}_1 = \mathbf{c} - \mathbf{a}_1$. It is easy to check that $\mathbf{a}_1$ and $\mathbf{b}_1$ are still nonnegative valued (so together define a pair of labelled partitions of n adding to $\mathbf{c}$) and moreover, $d((\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}', \mathbf{b}')) = d((\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}')) - 2 = 2(k-1)$. Now proceed by induction on k. □

We need a hypothesis that simplifies things, namely, we insist that all the matrices of the form $A + U_0$ and $B - U_0$ (where U_0 varies over all the solutions) are nonderogatory. This avoids multiple geometric eigenvalues, which tend to (but need not) yield continua of solutions. With this hypothesis, it is easy to see that the set map from solutions has values in the graph of $\mathbf{c}$—the result about spectra of $A + U_0$ and $B - U_0$ means that the algebraic multiplicities always balance, and our assumption about nonderogatory means that eigenvalues with multiplicity appear only in one Jordan block. In order to establish a graph homomorphism, we vary an earlier lemma.

Proposition 8.2. *Suppose that A and B are* $n \times n$ *matrices. Let* U_0 *be a nonzero solution to* $U^2 = UB - AU$, *and suppose that* $\operatorname{spec}(A \mid \operatorname{Rg} U_0) \cap \operatorname{spec}(\ell\operatorname{Rg} U_0 \mid B)$ *is nonempty. Then there is a topological continuum of matrices* $\{U_z\}_{z \in \mathbf{C}}$ *such that* $\operatorname{rank} U_z = U_0$ *for almost all z and* U_z *is a solution to* $U^2 = UB - AU$.

Proof. From (4.5) in Section 4, some solutions are in bijection with invertible solutions V to $\mathcal{F} = VJ_1^T - J_2 V$, where J_i are the Jordan normal forms of $\ell\operatorname{Rg} U_0 \mid B$ and $A \mid \operatorname{Rg} U_0$, respectively. By hypothesis (the existence of the solution U_0 to the original equation), there is at least one such V. Since the spectra overlap, the operator on $k \times k$ matrices (where $k = \operatorname{rank} U_0$) given by $Z \mapsto VJ_1^T - J_2 V$ has a nontrivial kernel, hence there exist V_0 and V_1 such that V_0 is an invertible solution and $V_0 + zV_1$ are solutions for all complex z. Multiplying by V_0^{-1}, we see that $V_0 + zV_1$ is not invertible only when $-1/z$ belongs to $\operatorname{spec} V_1 V_0^{-1}$, and there are at most n such values. For all other values of z, $(V_0 + zV_1)^{-1}$, after change of basis, yield solutions to (4.1). □

Now we want to show that the mapping from solutions of (4.1) to the pairs of labelled partitions is a graph homomorphism (assuming finiteness of the set solutions). We see that (from the finiteness of the solutions), the algebraic eigenvalues that are swapped by U_0 cannot have anything in common. It follows easily that the map is one to one, and moreover, if the rank of U_0 is one, then exactly one pair of distinct eigenvalues is swapped, hence the distance of the image pair from the original is 2. Thus it is a graph homomorphism. Finally, if the distance between the images of solutions is $2k$, then U_0 has swapped sets of k eigenvalues (with nothing in common), hence it has rank k. In particular, if $k = 1$, then U_0 has rank one, so the map is saturated.

Proposition 8.3. *If* $U^2 = UB - AU$ *has only finitely many solutions, then the map* $\mathcal{G}_{A,B} \to \mathcal{G}_{\mathbf{c}}$ *is a one-to-one saturated graph homomorphism.*

We can determine the valence of $(\mathbf{a},\mathbf{b})$; summing these over all the elements and dividing by 2 yields the number of edges. The vertices adjacent to $(\mathbf{a},\mathbf{b})$ in $\mathcal{G}_{\mathbf{c}}$ are exactly those of the form

$$\{(\mathbf{a}+\delta_\lambda-\delta_\mu,\mathbf{b}-\delta_\lambda+\delta_\mu)\mid \lambda\neq\mu;\ \mathbf{b}(\lambda)>0;\ \mathbf{a}(\mu)>0\}. \tag{8.5}$$

So the valence of $(\mathbf{a},\mathbf{b})$ is $\#\{(\lambda,\mu)\mid(\lambda-\mu)\cdot\mathbf{b}(\lambda)\cdot\mathbf{a}(\mu)\neq 0\}$. For example, if the pair is given by $\left(\begin{smallmatrix}3&3&0&0\\0&0&3&3\end{smallmatrix}\right)$, the valence is merely 4; however, the valence of one of its adjacent points, $\left(\begin{smallmatrix}2&3&1&0\\1&0&2&3\end{smallmatrix}\right)$ is 7, while that of *its* adjacent point $\left(\begin{smallmatrix}2&2&1&1\\1&1&2&2\end{smallmatrix}\right)$ is the maximum possible (within the graph), 12. The graph itself has 45 vertices, and the four nearest neighbours to the original point form a lozenge. There are 9 vertices of distance four, 17 of distance 6, and then 9, 4, 1 of respective distances 8, 10, and 12. (This symmetry is generic—the relevant involution is $(\mathbf{a},\mathbf{b})\mapsto(\mathbf{b},\mathbf{a})$.) Proposition 9.3 contains more general results on valence.

Suppose $\mathbf{c}_0=(1^{2n})$ is the standard labelled partition of $2n$ consisting entirely of 1s, and let $\mathbf{c}$ be any other partition of $2n$. Then there are graph homomorphisms $\psi:\mathcal{G}_{\mathbf{c}_0}\to\mathcal{G}_{\mathbf{c}}$ and $\phi:\mathcal{G}_{\mathbf{c}}\to\mathcal{G}_{\mathbf{c}_0}$ with the property that $\psi\circ\phi$ is the identity on $\mathcal{G}_{\mathbf{c}}$, that is, the latter is a retract of the former. This holds in somewhat more generality, as we now show.

Let $\mathbf{c}$ and $\mathbf{c}'$ be labelled partitions of $2n$. We say $\mathbf{c}'$ is subordinate to $\mathbf{c}$, denoted $\mathbf{c}'\prec\mathbf{c}$, if there is a partition $\{U_\alpha\}_{\alpha\in A}$ of $\operatorname{supp}\mathbf{c}$ and a reindexing $\{\lambda_\alpha\}_{\alpha\in A}$ of $\operatorname{supp}\mathbf{c}'$ such that for all α in A,

$$\mathbf{c}'(\lambda_\alpha)=\sum_{\lambda\in U_\alpha}\mathbf{c}(\lambda). \tag{8.6}$$

We are dealing with loopless graphs, so graph homomorphisms (as usually defined) that are not one-to-one are impossible in our context. Hence we redefine a *graph homomorphism* to be a pair of functions (both denoted ψ) on vertices and edges such that if v and v' are vertices and $\psi(v)\neq\psi(v')$, then the edge (if it exists) $\{v,v'\}$ is mapped to the edge $\{\psi(v),\psi(v')\}$. (Alternatively, we can redefine the graphs to include loops on all vertices, so that the ordinary definition of graph homomorphism will do.)

Lemma 8.4. *If $\mathbf{c}'\prec\mathbf{c}$, then there exist graph homomorphisms $\psi:\mathcal{G}_{\mathbf{c}}\to\mathcal{G}_{\mathbf{c}'}$ and $\phi:\mathcal{G}_{\mathbf{c}'}\to\mathcal{G}_{\mathbf{c}}$ such that $\psi\circ\phi$ is the identity on $\mathcal{G}_{\mathbf{c}'}$.*

Proof. For each subset U_α of $\operatorname{supp}\mathbf{c}$, pick a total ordering $\lambda_{\alpha,1}<\lambda_{\alpha,2}<\cdots$ on the members of U_α (this has nothing to do with the numerical values of the λs, it is simply a way of indexing them). Consider the set

$$V_0:=\{(\mathbf{a},\mathbf{b})\in\mathcal{G}_{\mathbf{c}}\mid\forall\alpha,\ \mathbf{a}(\lambda_{\alpha,i})\neq 0\text{ implies }\mathbf{a}(\lambda_{\alpha,j})=\mathbf{c}(\lambda_{\alpha,j})\ \forall\, j<i\}. \tag{8.7}$$

We see immediately that

(*) if $(\mathbf{a},\mathbf{b})$ and $(\mathbf{a}_1,\mathbf{b}_1)$ belong to V_0 and $\mathbf{a}(\lambda_{\alpha,i})=\mathbf{a}_1(\lambda_{\alpha,i})\neq 0$, then $\mathbf{a}(\lambda_{\alpha,j})=\mathbf{a}_1(\lambda_{\alpha,j})$ for all $j<i$.

Let H denote the subgraph of $\mathcal{G}_{\mathbf{c}}$ whose set of vertices is V_0 and whose edges are inherited from $\mathcal{G}_{\mathbf{c}}$. Define ψ on the vertices by $(\mathbf{a},\mathbf{b})\mapsto(\mathbf{a}',\mathbf{b}')$ where $\mathbf{a}'(\lambda_\alpha)=\sum_i\mathbf{a}(\lambda_{\alpha,i})$ (and $\mathbf{b}'$ is defined as $\mathbf{c}'-\mathbf{a}'$). If $(\mathbf{a},\mathbf{b})$ and $(\mathbf{a}_1,\mathbf{b}_1)$ are connected by an edge, then there are distinct λ and μ in $\operatorname{supp}\mathbf{c}$ such that $\mathbf{a}(\lambda)=\mathbf{a}_1(\lambda)+1$, $\mathbf{a}(\mu)=\mathbf{a}_1(\mu)-1$, and $\mathbf{a}(\rho)=\mathbf{a}_1(\rho)$ for all ρ

not in $\{\lambda,\mu\}$. If λ and μ belong to the same U_α, then $\psi(\mathbf{a},\mathbf{b}) = \psi(\mathbf{a}_1,\mathbf{b}_1)$ (the extra pair of parentheses is suppressed). If λ and μ belong to different U_α, then it is immediate that $d(\psi(\mathbf{a},\mathbf{b}),\psi(\mathbf{a}_1,\mathbf{b}_1)) = 2$. In particular, ψ preserves edges (to the extent that loops are considered edges).

Next, by (*), $\psi \mid V_0$ is one-to-one.

Now define ϕ on vertices. Pick $(\mathbf{a}',\mathbf{b}')$ in $\mathcal{G}_{\mathbf{c}'}$. For each λ_α in $\operatorname{supp}\mathbf{c}'$, there exists a unique $i \equiv i(\alpha)$ such that $\mathbf{a}'(\lambda_\alpha) = \sum_{j<i}\mathbf{c}(\lambda_{\alpha,j}) + \rho_\alpha$ for a unique ρ_α with $0 \le \rho_\alpha < \mathbf{c}(\lambda_{\alpha,i})$. Define $\mathbf{a}$ via

$$\mathbf{a}(\lambda_{\alpha,k}) = \begin{cases} \mathbf{c}(\lambda_{\alpha,k}) & \text{if } k < i(\alpha), \\ \rho & \text{if } k = i(\alpha), \\ 0 & \text{else.} \end{cases} \tag{8.8}$$

This yields a labelled partition of n, so the resulting pair $(\mathbf{a},\mathbf{c}-\mathbf{a})$ is an element of $\mathcal{G}_{\mathbf{c}}$, and we define it to be the image of $(\mathbf{a}',\mathbf{b}')$ under ϕ. It is obvious that $\psi\circ\phi$ is the identity on $\mathcal{G}_{\mathbf{c}'}$, and easy to check that ϕ preserves edges. Also ϕ is one-to-one and its range lies in V_0. A simple cardinality argument yields that $\psi \mid V_0$ is onto

$$\left|V_0\right| = \left|\psi(V_0)\right| \le \left|\mathcal{G}_{\mathbf{c}'}\right| = \left|\phi(\mathcal{G}_{\mathbf{c}'})\right| \le \left|V_0\right|. \tag{8.9}$$

□

If $\mathbf{c} = (1^{2n})$ and $\mathbf{c}'$ is any labelled partition of $2n$, then $\mathbf{c}' \prec \mathbf{c}$, and the result applies. If $\mathbf{c} = (k, 1^{2n-k})$ for some $1 < k \le 2n$, then $\mathbf{c}' \prec \mathbf{c}$ if and only if there exists λ in $\operatorname{supp}\mathbf{c}'$ such that $\mathbf{c}'(\lambda) \ge k$. One extreme occurs when $k = 2n$, which however, does not yield anything of interest; in this case, $\mathcal{G}_{\mathbf{c}}$ consists of one point.

9. Graph-related examples

To a pair of $n\times n$ matrices A and B, we have associated the graph $\mathcal{G}_{A,B}$ whose vertices are the solutions to (4.1)

$$U^2 = UB - AU. \tag{9.1}$$

Assume that only finitely many solutions exist to (4.1). Recall that $\mathbf{c} : \operatorname{spec}A \cup \operatorname{spec}B \to \mathbf{N}$ is the map which associates to an element of the domain, λ, the sum of its algebraic multiplicities in A and B. This permits us to define a mapping $\mathcal{G}_{A,B} \to \mathcal{G}_{\mathbf{c}}$ which is a saturated embedding of graphs. We call $\mathcal{G}_{A,B}$ *defective* if the map is not onto, that is, if there are vertices in $\mathcal{G}_{\mathbf{c}}$ that do not arise from solutions to (4.1). The results, Lemma 9.1, Propositions 9.2 and 9.3 at the end of this section are useful for calculating the examples.

For example, if $n = 2$, the possible choices for $\mathcal{G}_{\mathbf{c}}$ are those arising from the partitions of $2n$, here (1^4), $(2,1,1)$, $(2,2)$, $(3,1)$, (4); these have, respectively, 6, 4, 3, 2, and 1 vertices. So if $\mathcal{G}_{A,B}$ has exactly five points (i.e., 5 solutions to (4.1)), then it is automatically defective. We construct examples to illustrate all possible defective graphs when $n = 2$. It does not seem feasible (at the moment) to analyze all possible defective graphs when $n = 3$.

Consider the case $n = 2$.

(a) $\mathbf{c} = (1^4)$. Then $\mathcal{G}_{\mathbf{c}}$ has 6 vertices (subpartitions of 1^4 that add to 2), every point has valence 4, and the graph is that of the edges and vertices of an octahedron. (For future

reference, "the graph is the polyhedron P," means that the graph is the graph consisting of the vertices and edges of the compact convex polyhedron P.) Since the automorphism group of the octahedron acts transitively, the graph resulting from removing a point and its corresponding edges is the same independently of the choice of point. The resulting graph is a pyramid with square base, having 5 vertices, and all elements but one have valence 3, the nadir having valence 4. As a graph, this is known as the 4-wheel.

Let $\lambda_1, \lambda_2, \mu_1, \mu_2$ be four distinct complex numbers, and set $B = \operatorname{diag}(\lambda_1, \lambda_2)$ and $A = \left(\begin{smallmatrix} \mu_1 & 1 \\ 0 & \mu_2 \end{smallmatrix}\right)$. Right eigenvectors of A are $e_1 = (1,0)^t$ and $e_2 = (\mu_2 - \mu_1, 1)^t$. Left eigenvectors of B are $f_1 = (1,0)$ and $f_2 = (0,1)$. We see that $f_2 e_1 = 0$, but all other $f_i e_j$ are not zero. It follows that the valence of the solution $U = \mathbf{0}$ is 3, and thus there are at least 4 but fewer than 6 solutions.

As A and B do not commute but have disjoint spectra with no multiple eigenvalues, it follows from Proposition 7.2 that every element in $\mathcal{G}_{A,B}$ has valence at least 3. If there were only 4 solutions, the graph would thus have to be a tetrahedron (complete graph on four points). This contradicts Proposition 9.2 (below). Hence there must be five solutions, and because the map on the graphs is saturated, $\mathcal{G}_{A,B}$ is the pyramid with square base.

Doubly defective subgraphs of $\mathcal{G}_{\mathbf{c}}$ can arise. For example, if A and B commute (and as here, have distinct eigenvalues with no multiples), then $\mathcal{G}_{A,B}$ has four points, and consists of the lozenge (every element has valence 2). Since the valence of any vertex in $\mathcal{G}_{A,B}$ cannot drop below two, we cannot remove a third point—triply defective examples do not exist.

(b) $\mathbf{c} = (2,1,1)$. Here $\mathcal{G}_{\mathbf{c}}$ consists of four points arranged in a lozenge, but with a cross bar joining the middle two points; there are two points of valence two and two points of valence 3. There are two possible singly defective subgraphs, obtained by deleting a point of valence 2 (resulting in the triangle, i.e., the complete graph on 3 points) or deleting a point of valence 3 (resulting in a linear graph •–•–• of length 2). Both of these can be realized.

To obtain the linear graph, set $A = \left(\begin{smallmatrix} \mu & 1 \\ 0 & \mu \end{smallmatrix}\right)$ and $B = \operatorname{diag}(\lambda_1, \lambda_2)$ where $\lambda_1, \lambda_2, \mu$ are distinct complex numbers. The valence of the solution $\mathbf{0}$ is one (rather than two, as we would obtain from the nondefective graph), so there are at least two points in the graph, but no more than three. On the other hand, by Proposition 9.2 below, there is a point at distance two from $\mathbf{0}$, so there are at least three points, and thus exactly three, and it follows from the valence of the bottom point being one that the graph must be the line segment.

To obtain the triangle, a slightly more difficult form is required. As before, let λ_1, λ_2, μ be distinct complex numbers. Define $B = \left(\begin{smallmatrix} \mu & 1 \\ 0 & \mu \end{smallmatrix}\right)$ and $A = \left(\begin{smallmatrix} 0 & 1 \\ -\lambda_1\lambda_2 & \lambda_1+\lambda_2 \end{smallmatrix}\right)$. The latter is the companion matrix of the polynomial $(x-\lambda_1)(x-\lambda_2)$. Then we can take as eigenvectors for A, $e_i = (1, \lambda_i)^t$ and for B, $f_1 = (1,0)$ and generalized eigenvector $f_2 = (1,1)$. Form the matrix $\mathcal{F} = (f_i e_j)$, which here is $\left(\begin{smallmatrix} 1 & \lambda_2 \\ \lambda_1+1 & \lambda_2+1 \end{smallmatrix}\right)$. We will choose the three eigenvalues so that the equation

$$\mathcal{F} = B^T V - V \operatorname{diag}(\lambda_1, \lambda_2) \tag{9.2}$$

has no invertible solution (note that B is already in Jordan normal form, and the diagonal matrix is a Jordan form of A). By Section 4 (see (4.5)), this prevents there from being a point in $\mathcal{G}_{A,B}$ at graph distance two from $\mathbf{0}$, in other words, the apex of the lozenge has

been deleted. The valence of $\mathbf{0}$ is clearly two, so the three remaining points of $\mathcal{G}_{\mathbf{c}}$—forming the triangle clearly survive in $\mathcal{G}_{A,B}$.

By disjointness of the spectra, there is a unique solution V; it suffices to choose the parameters so that the determinant of V is zero and the parameters are distinct. By brute force (setting $V = (v_{ij})$), we find that the determinant of V is $(1 - \lambda_1\lambda_2 + \lambda_2/(\mu - \lambda_1) - 1/(\mu - \lambda_2))(\mu - \lambda_1)^{-1}(\mu - \lambda_2)^{-1}$. One solution (determinant zero) is obtained by setting $\lambda_1 = 2$, $\lambda_2 = 1/2$ and $\mu = 34/5$. There are plenty of solutions.

(c) $\mathbf{c} = (2,2)$. This time $\mathcal{G}_{\mathbf{c}}$ consists of the line segment •–•–•. Deleting one of the endpoints will result in a shorter segment, and is easy to do. More interesting is what happens when the middle point is deleted, creating two isolated points. This is the first nonconnected example, and is also easy to implement, because we just have to make sure that $fe = 0$ but there still a second solution.

Pick λ and μ distinct, and set A and B to be the (upper triangular) Jordan matrices of block size two with eigenvalue μ and λ, respectively. The right eigenvector of A is $e_1 = (1,0)^t$, the left eigenvector of B is $f_1 = (0,1)$, so $f_1e_1 = 0$ and the valence of $\mathbf{0}$ is thus zero. On the other hand, since A and B commute, $\mathrm{I} = BV - VA$ has an invertible solution (by Proposition 9.2), so the other endpoint of the line segment appears in the image of $\mathcal{G}_{A,B}$.

(d) $\mathbf{c} = (3,1)$. Here $\mathcal{G}_{\mathbf{c}}$ consists of two vertices and one edge •–•. If defective, there would be just one solution (necessarily the trivial one), and this is routine to arrange. Let B have Jordan form $\left(\begin{smallmatrix}\lambda & 1\\ 0 & \lambda\end{smallmatrix}\right)$ and $A = \mathrm{diag}(\lambda,\mu)$, where $\mu \neq \lambda$. We just have to alter B so that its right eigenvector is orthogonal to the eigenvector for μ.

(e) $\mathbf{c} = (4)$. Here $\mathcal{G}_{\mathbf{c}}$ consists of one point, necessarily the trivial solution.

Generic realization. To realize $\mathcal{G}_{1^{2n}}$ as the graph of a solution space for specific matrices A and B, begin with $2n$ distinct complex numbers $\{\mu_1,\dots,\mu_n;\ \lambda_1,\dots,\lambda_n\}$. Let $B = \mathrm{diag}(\mu_j)$ and set A to be the companion matrix (with 1 in the $(2,1)$, not the $(1,2)$ entry) of $p_A(x) = \prod(x - \lambda_i)$. The right eigenvector of A for λ_i is $e_i = (1,\lambda_i,\lambda_i^2,\dots,\lambda_i^{n-1})^t$, and of course $f_j = (0,0,\dots,0,1,0,\dots)$ (1 in the jth position) is the left eigenvector of B for μ_j.

Pick k-element subsets R, S, respectively, of $\{e_i\}$ and of $\{f_i\}$, and form the $k \times k$ matrix $\mathcal{F} = (f_ie_j)_{(i,j)\in S\times R}$. The equation in V, $\mathcal{F} = \Delta_R V - V\Delta_S$ (where the Δs represent the corresponding diagonal matrices) has a unique solution given by $V = (\lambda_j^{i-1}/(\mu_i - \lambda_j))$ $((i,j) \in S \times R)$. Consider $\det V \cdot \prod(\mu_i - \lambda_j)$. This is a polynomial in $2k$ variables $(\{\mu_i,\lambda_j\})$, and if it does not vanish identically, its zero set is (at worst) a finite union of varieties, hence is nowhere dense in $\mathbf{C}^{S\times R}$. For each k and each choice of pair of k-element subsets, we can embed the space in $\mathbf{C}^{2n}$, and then take the intersection of the nonzero sets. This is a finite intersection of dense open sets (in fact, complements of lower dimensional varieties), hence is dense and open. Thus for almost choices of $\{\mu_1,\dots,\mu_n;\ \lambda_1,\dots,\lambda_n\}$, each of the Vs will be invertible, and thus corresponds to a solution to (4.1).

It is routine to verify that $\det V \cdot \prod(\mu_i - \lambda_j)$ does not vanish identically; it is only required to show a single corresponding solution exists to (4.1).

Other $\mathcal{G}_{\mathbf{c}}$ can be realized, without computing "V" explicitly, along the same lines.

Lemma 9.1. *Suppose that (4.1) has only finitely many solutions for a pair of $n \times n$ matrices (A,B). Then $\mathcal{G}_{A,B}$ cannot contain a subgraph isomorphic to the $n+1$-simplex (complete graph on $n+2$ elements).*

Proof. By replacing (A,B) by $(A+U_0,\ B-U_0)$ if necessary, we may assume that one of the points of the subgraph is the zero solution, and all the others in the simplex are rank 1. Hence the other $n+1$ solutions in the simplex must be of the form ef where e is a right eigenvector of A and f is a left eigenvector of B. Since every one of these solutions is connected to every other one, we must have that all the differences are also rank one. List the left eigenvectors of B, f_i $(i=1,\dots,k\le n)$ and the right eigenvectors of A, e_j $(j=1,\dots,l\le n)$; then the solutions are all of the form $\alpha_{ij}e_if_j$, at most one for each pair, for some complex numbers α_{ij}. It is a routine exercise that $\operatorname{rank}(a_{ij}e_if_j+a_{i'j'}e_{i'}f_{j'})$ is two if a_{ij}, $a_{i'j'}$ are not zero and $i\neq i'$ and $j\neq j'$. It easily follows that there are at most n choices for the e_if_j. □

Proposition 9.2. *Let R, S be $n\times n$ matrices, and let $\mathcal{R}$ denote the unital subalgebra of $\mathrm{M}_n\mathbf{C}$ generated by R and S. Suppose that $\operatorname{spec}R\cap\operatorname{spec}S=\emptyset$, and $\mathcal{R}$ is commutative modulo its (Jacobson) radical. Let T be an element of $\mathcal{R}$. If V is a solution to $T=RV-VS$, then V is invertible if and only if T is.*

Proof. The map $\mathbf{R}_{R,-S}$ restricts to an endomorphism of $\mathcal{R}$. The spectral condition ensures that it is one-to-one, hence onto. Thus there exists unique V in $\mathcal{R}$ solving the equation. Modulo the radical, we have $t=rv-vs$ (using lower case for their images). Since r and s commute with each other and v, we have $v(r-s)=t$. If t is invertible, then v is and thus its lifting, to V, is invertible (since the Jacobson radical has been factored).

If t is not invertible, we note that in any case $\operatorname{spec}r\subseteq\operatorname{spec}R$ and $\operatorname{spec}s\subseteq\operatorname{spec}S$, so, since the factor algebra is commutative, $r-s$ is invertible, whence $v=(r-s)^{-1}t$ is not invertible. Hence its preimage V is not invertible.

By the spectral condition, $\mathbf{R}_{R,-S}$ as an endomorphism of $\mathrm{M}_n\mathbf{C}$ is one-to-one, whence the solution V is unique as a solution in $\mathrm{M}_n\mathbf{C}$. □

Proposition 9.3. *Suppose that $(\mathbf{a},\mathbf{b})$ is an element of $\mathcal{G}_\mathbf{c}$, and one sets $k=|\operatorname{supp}\mathbf{a}|$, $l=|\operatorname{supp}\mathbf{b}|$, and $m=|\operatorname{supp}\mathbf{a}\cap\operatorname{supp}\mathbf{b}|$. Then the valence of $(\mathbf{a},\mathbf{b})$ (within $\mathcal{G}_\mathbf{c}$) is $kl-m$.*

Proof. Obviously, $|\operatorname{supp}\mathbf{c}|=k+l-m$. For λ in $\operatorname{supp}\mathbf{a}\setminus\operatorname{supp}\mathbf{b}$, we can subtract 1 from $\mathbf{a}(\lambda)$ and add one to $\mathbf{a}(\mu)$ for each μ in $\operatorname{supp}\mathbf{b}$ (the process subtracts 1 from $\mathbf{b}(\mu)$). This yields $(k-m)l$ edges. For λ in $\operatorname{supp}\mathbf{a}\cap\operatorname{supp}\mathbf{b}$, we can subtract 1 from $\mathbf{a}(\lambda)$ and add 1 to $\mathbf{b}(\mu)$, provided that μ is in $\operatorname{supp}\mathbf{b}\setminus\{\lambda\}$. This yields $m(l-1)$ edges, and the total is $(k-m)l+m(l-1)=kl-m$. □

10. Inductive relations

For any $\mathbf{c}$, a formula for number of vertices in $\mathcal{G}_\mathbf{c}$ is easily derivable from the inclusion-exclusion principle (the number of solutions to $\sum\mathbf{a}(\lambda)=n$ subject to the constraints that $\mathbf{a}(\lambda)$ are integers and $0\le\mathbf{a}(\lambda)\le\mathbf{c}(\lambda)$ for all λ in the support of $\mathbf{c}$. The resulting formula, however, is generally unwieldy, as it is an alternating sum of sums of combinatorial expressions. Moreover, it says very little about the graph structure.

We can say something about the graph structure in terms of "predecessors," at least in case n is relatively small, by exploiting some natural maps between the $\mathcal{G}_\mathbf{c}$. For example,

pick λ in the support of $\mathbf{c}$ and define $\mathbf{c}' := \mathbf{c} + 2\delta_\lambda$ (where $\delta_\lambda : \mathbf{C} \to \mathbf{N}$ is the characteristic/indicator function of the singleton set $\{\lambda\}$). There is a natural map

$$\begin{aligned} \Phi_\lambda : \mathcal{G}_{\mathbf{c}} &\longrightarrow \mathcal{G}_{\mathbf{c}'} \\ (\mathbf{a}, \mathbf{b}) &\longmapsto (\mathbf{a} + \delta_\lambda, \mathbf{b} + \delta_\lambda). \end{aligned} \tag{10.1}$$

It is obviously well defined, and satisfies the following properties.

(0) Φ_λ is a saturated embedding of graphs. (This is straightforward.)

(a) If $\mathbf{c}(\lambda) \geq n$, then Φ_λ is a graph isomorphism (i.e., it maps the vertices onto the vertices of $\mathcal{G}_{\mathbf{c}'}$).

Proof. By (0), it is sufficient to show that Φ_λ maps onto the vertices. Suppose that $(\widetilde{\mathbf{a}}, \widetilde{\mathbf{b}})$ is an element of $\mathcal{G}_{\mathbf{c}'}$. We have that $\sum_{\mu \neq \lambda} \mathbf{c}'(\mu) = \sum_{\mu \neq \lambda} \leq n$; thus $\widetilde{\mathbf{a}}(\lambda) + \widetilde{\mathbf{b}}(\lambda) \geq n + 2$. If either $\widetilde{\mathbf{a}}(\lambda)$ or $\widetilde{\mathbf{b}}(\lambda)$ were zero, then the other one of the pair would be at least $n+1$; however, the sum of the values of both $\widetilde{\mathbf{a}}$ and $\widetilde{\mathbf{b}}$ is $n+1$, a contradiction. Hence both $\widetilde{\mathbf{a}}(\lambda) \geq 1$ and $\widetilde{\mathbf{b}}(\lambda) \geq 1$, whence $(\widetilde{\mathbf{a}} - \delta_\lambda,\ \widetilde{\mathbf{b}} - \delta_\lambda)$ is in the preimage of $(\widetilde{\mathbf{a}}, \widetilde{\mathbf{b}})$. □

The remaining properties are proved by similar means.

(b) If $\mathbf{c}(\lambda) = n - 1$), then $\mathcal{G}_{\mathbf{c}'}$ is obtained from the image of $\mathcal{G}_{\mathbf{c}}$ under Φ_λ by adjoining the two points $(\mathbf{a}_0 = \mathbf{c} - c(\lambda)\delta_\lambda,\ \mathbf{b}_0 = (\mathbf{c}(\lambda) + 2)\delta_\lambda)$ and $(\mathbf{a}_1 = (\mathbf{c}(\lambda) + 2)\delta_\lambda,\ \mathbf{b}_1 = \mathbf{c} - c(\lambda)\delta_\lambda)$. The edges joining $(\mathbf{a}_0, \mathbf{b}_0)$ to points in the image of the graph have as their other endpoints precisely the points $\Phi(\mathbf{a}, \mathbf{b})$ where $\mathbf{a}(\lambda) = 0$, and similarly, $(\mathbf{a}_1, \mathbf{b}_1)$ is joined only to the points $\Phi(\mathbf{a}, \mathbf{b})$ with $\mathbf{b}(\lambda) = 0$.

(c) If $\mathbf{c}(\lambda) = n - 2$, then $\mathcal{G}_{\mathbf{c}'}$ is obtained from the image of $\mathcal{G}_{\mathbf{c}}$ by adding two copies ("up" and "down") of a $k-1$-simplex (i.e., the complete graph on k points), where $k = |\operatorname{supp} \mathbf{c}|$.

(d) If $\mathbf{c}(\lambda) = n - s$ where $s \geq 1$, then $|\mathcal{G}_{\mathbf{c}'}| \leq |\mathcal{G}_{\mathbf{c}}| + 2\binom{k+s-3}{s-1}$.

For example, $\mathcal{G}_{1,1,1,1}$ ($n = 2$, $s = 1$) is the octohedron (6 points, 12 edges, uniform valence four) and $\mathcal{G}_{3,1,1,1}$ is obtained by adjoining two vertices to the upper left and lower right, respectively, and joining them to the vertices of the nearest triangular face. This creates a graph with 10 points, 18 edges, and all but the two added points have valence 4. Going one step further, however, $\mathcal{G}_{3,1,1,1} = \mathcal{G}_{5,1,1,1}$ by property (a).

There are other types of such maps. For example, suppose that λ is in $\operatorname{supp} \mathbf{c}$ but μ is not. Create the new partition of $2(n+1)$, $\mathbf{c}' := \mathbf{c} + \delta_\lambda + \delta_\mu$ (enlarging the support by one element). There are two possibilities for maps $\mathcal{G}_{\mathbf{c}} \to \mathcal{G}_{\mathbf{c}'}$, either $(\mathbf{a}, \mathbf{b}) \mapsto (\mathbf{a} + \delta_\mu,\ \mathbf{b} + \delta_\lambda)$ or $(\mathbf{a}, \mathbf{b}) \mapsto (\mathbf{a} + \delta_\lambda,\ \mathbf{b} + \delta_\mu)$. These are both saturated graph embeddings, obviously with disjoint images. Under some circumstances, the union of the images gives all the vertices of $\mathcal{G}_{\mathbf{c}'}$.

For example, this occurs if $\mathbf{c}(\lambda) = n$. In this case, the number of vertices doubles, and it is relatively easy to draw the edges whose endpoints lie in different copies. For example, with $\mathbf{c} = (2,2)$, the graph is a line with three points; two copies of this line joined by shifting yield the triangular latticework, the graph of $\mathbf{c}' = (3,2,1)$; the hypothesis is of course preserved, so we can apply the same process to obtain the 12-point graph of $(4,2,1,1)$ (difficult to visualize, let alone attempt to draw), and continue to obtain the 24-point graph of $(5,2,1,1,1)$, and so forth. Unfortunately, the situation is much more

complicated when $\mathbf{c}(\lambda) < n$ (e.g., the graph corresponding to $(4,3,1,1,1)$ obtained from $(4,2,1,1)$ using the second coordinate) has 30 vertices, not 24.

11. Attractive and repulsive fixed points

A fixed point X of a map ϕ is *attractive* if there exist $\delta > 0$ and $c < 1$ such that whenever $\|Z - X\| \le \delta$, it follows that $\|\phi(Z) - X\| \le c\|Z - X\|$. Similarly, X is *repulsive* if there exist $\delta > 0$ and $c > 1$ such that $\|Z - X\| < \delta$ entails $\|\phi(Z) - X\| \ge c\|Z - X\|$. If $\phi = \phi_{C,D}$ and CD is invertible, the conjugacy of $\phi_{D,C}^{-1}$ with $\phi_{C,D}$ (see Section 2) yields a graph and metric isomorphism between the fixed points of $\phi_{D,C}$ and of $\phi_{C,D}$, which however, reverses orientations of the trajectories (see Section 3)—in particular, attractive fixed points are sent to repulsive ones, and vice versa.

Suppose that X is a fixed point of $\phi_{C,D}$. There is a simple criterion for it to be attractive: $\rho(XC)\cdot\rho(DX) < 1$; this can be rewritten as $\rho(\mathcal{D}\phi_{C,D}(X)) < 1$ (recall that ρ denotes the spectral radius, and $\mathcal{D}$ the derivative). For more general systems, this last condition is sufficient but not necessary; however, for fractional matrix transformations, the criterion is necessary and sufficient.

To see the necessity, select a right eigenvector v for XC with eigenvalue λ, and a left eigenvector w for DX. As in Section 3, set $Y = vw$ and consider $\phi(X + zY) = X + \psi(z)Y$, where $\psi : z \mapsto \lambda\mu z/(1 - z\lambda\,\mathrm{tr}\,YD)$ is the ordinary fractional linear transformation corresponding to the matrix $\left(\begin{smallmatrix}\lambda\mu & 0\\ -\lambda\,\mathrm{tr}\,YD & 1\end{smallmatrix}\right)$. Around the fixed point (of ψ) $z = 0$, ψ is attractive if and only if $|\lambda\mu| < 1$. Thus X is attractive entails that $|\lambda\mu| < 1$ and $\rho(XC)\cdot\rho(DX) = \max|\lambda\mu|$, where λ varies over the eigenvalues of XC and μ over the eigenvalues of DX. The same argument also yields that if X is repulsive, then $|\lambda\mu| > 1$ for all choices of λ and μ.

If we assume that $\rho(XC)\cdot\rho(DX) < 1$, then X is attractive by the Hartman-Grobman theorem [5, Theorem 2.2.1], once we observe that $\rho(\mathcal{D}\phi_{C,D}(X)) = \rho(XC)\cdot\rho(DX)$. I am indebted to my colleague Victor Leblanc for telling me about this. It can also be proved directly in an elementary but somewhat tedious way in our context.

A less uninteresting question arises, suppose that $\phi_{C,D}$ has a fixed point; when does it admit an attractive (or a repulsive) one? How about uniqueness, and what is the relation between attractive and repulsive fixed points, if they both exist? We can answer these questions, more or less.

First, assume that $\phi_{C,D}$ has a fixed point X, not assumed to be attractive. Form $B = (DX)^{-1}$ and $A = CX$ as in our earlier reduction, but this time, we refer to the eigenvalues of CX and XD (note that since X is invertible, XD and DX are conjugate). List the (algebraic) eigenvalues with multiplicities of CX and XD as $\lambda_1,\lambda_2,\ldots,\lambda_n$ and $\mu_1,\mu_2,\ldots,\mu_n$, where we have ordered them so that $|\lambda_i| \le |\lambda_{i+1}|$ and $|\mu_i| \le |\mu_{i+1}|$ for all $i = 1,2,\ldots,n$. Keep in mind that the corresponding algebraic spectrum of B is (μ_i^{-1}).

Proposition 11.1. *Suppose $\phi_{C,D}$ that admits an attractive or a repulsive fixed point.*

(a) *Then for all k, $|\lambda_k\mu_k| \ne 1$;*

(b) *if there are only finitely many fixed points, then there is at most one attractive one and one repulsive one.*

Proof. If the condition holds, then there exists k_0 in $\{l+1/2 \mid l=0,1,\dots,n\}$ such that $|\lambda_k\mu_k|<1$ if $k<k_0$ and $|\lambda_{k+1}\mu_{k+1}|>1$ if $k>k_0$. Create the new lists, $\Lambda' := \lambda_1,\dots,\lambda_{\lfloor k_0\rfloor}$, $\mu^{-1}_{\lceil k_0\rceil},\dots,\mu_n^{-1}$, and $M' := \mu_1,\dots,\mu_{\lfloor k_0\rfloor},\lambda^{-1}_{\lceil k_0\rceil},\dots,\lambda_n^{-1}$. If we abbreviate $\max_{s\in\Lambda'}|s|$ by $\max|\Lambda'|$, then we see immediately that $\max|\Lambda'|\max|M'|<1$.

Drop the condition on the products and let k be the smallest integer such that $|\lambda_i\mu_i|>1$; if no such i exists, set $k=n+1$. Set $I_0=J_0=\{k,k+1,\dots,n\}$ (if $k=n+1$, these are the null set).

Next, we show that if $\phi_{C,D}$ has an attractive fixed point X_0, then the algebraic spectra (with multiplicities) of DX_0 and CX_1 must be M' and Λ'. If X_1 is any fixed point of ϕ, the algebraic spectra of DX_1 and CX_1 are obtained from the original lists μ_i and λ_j by a swap of the following form. Select $I,J\subset\{1,2,\dots,n\}$ such that $|I|=|J|$ and replace the original lists by $M_1 := (\mu_i)_{i\notin J},(\lambda_i^{-1})_{i\in I}$ and $\Lambda_1 := (\lambda_i)_{i\notin I},(\mu_i^{-1})_{i\in J}$; there is the additional restriction that if $(i,j)\in I\times J$, then $\lambda_i\mu_j\neq 1$. (This follows from our results on the equation $U^2=UB-AU$.)

We show that if $\max|M_1|\cdot\max|\Lambda_1|<1$, then $I=J=I_0$. This of course forces $M_1=M'$ and $\Lambda_1=\Lambda'$.

Suppose that l belongs to I and $l<k$. Then λ_l^{-1} belongs to M_1; this forces λ_{l+t}^{-1} to also belong to M_1 (the alternative, that any λ_{l+t} belongs to Λ_1 yields a product $|\lambda_l^{-1}\lambda_{l+t}|$, i.e., at least one, since $|\lambda_l|\le|\lambda_{l+t}|$). Hence $I=\{l_0,l+1,l+2,\dots,n\}$ for some $l_0\le l$. Also, since $|\mu_l|\cdot|\lambda_l|\le 1$, we must have μ_l in M_1 (else the product $|\mu_l^{-1}|\cdot|\lambda_l^{-1}|$ is at least one). Again, this forces $\mu_1,\mu_2,\dots,\mu_{l-1}$ to belong to M_1. Together we have $n-l_0+1+l>n$ elements in M_1, a contradiction. Thus $I\subseteq I_0$, and the same arguments also show that I is an interval.

If k is not in I, then λ_k belongs to Λ_1; necessarily μ_k^{-1} belongs to Λ_1 (as the product $|\lambda_k|\cdot|\mu_k|$ exceeds 1). However, this forces μ_{k+t}^{-1} to belong to Λ_1 as well. Also, λ_{k-t} must belong to Λ_1 for $t\ge 1$ (as the product $|\lambda_{k-t}^{-1}|\cdot|\lambda_k|$ is at least one). This yields too many elements in Λ_1, so we have that $I=I_0$.

The symmetric argument yields that $J=J_0$. Now instead of doing this from the point of view of k, define l to be the largest integer i such that $|\lambda_i|\cdot|\mu_i|<1$, and define $I^0=J^0=\{1,2,\dots,l\}$. Symmetric arguments to those above show that the complements of I and J are I^0 and J^0, respectively. This implies that $l=k-1$, which of course is exactly the conclusion for attractive fixed points. For a repulsive fixed point, all products of the eigenvalues have absolute value exceeding one, and we just reverse the roles of C and D (as in Section 2). This yields (a).

(b) There is only one swapping of the eigenvalues that will yield a pair of sets of eigenvalues with the attractive or repulsive property. By Proposition 8.3, the map from the graph of fixed points to $\mathcal{G}_{\mathbf{c}}$ is one-to-one; that is, the algebraic spectrum determines the fixed point. Hence there is at most one attractive or repulsive fixed point. □

We can remove the finiteness hypothesis in part (b). Any repulsive or attractive fixed point must correspond to a pair (as in Section 5) of left/right invariant vector spaces, each of which is isolated. The corresponding pairs of algebraic spectra determine uniquely the fixed point.

Suppose that the mapping $\mathcal{G}_{A,B}\to\mathcal{G}_{\mathbf{c}}$ (where $A=XC$ etc.) is onto, that is, the graph of $\mathcal{G}_{A,B}$ is nondefective. Then it is easy to give necessary and sufficient conditions so that $\phi_{C,D}$ have an attractive or a repulsive fixed point. The first observation is that the existence

of one implies the existence of the other. The flip, $(\mathbf{a},\mathbf{b}) \mapsto (\mathbf{b},\mathbf{a})$ implemented on $\mathcal{G}_{\mathbf{c}}$, reverses the roles of the matrices, in particular, swaps the sets of eigenvalues. (If the graph is defective, this argument fails, and indeed, there are examples with an attractive but no repulsive fixed points.)

A second observation is that the partition corresponding to $\mathbf{c}$ limits the possibility of having an attractive or repulsive fixed point. For example, if $\sum \mathbf{c}(\lambda) = 2n$ but there exist λ_0 such that $\mathbf{c}(\lambda_0) > n$, then the corresponding ϕ can have neither an attractive nor a repulsive fixed point—the corresponding spectra (after converting from A to DX) always have a λ_0 on one side and a λ_0^{-1} on the other, so the necessary condition above fails. If $\mathbf{c}(\lambda_0) = n$, then we must have either $|\lambda_i| > |\lambda_0|$ for all λ_i in $\operatorname{supp}\mathbf{c} \setminus \{\lambda_0\}$, or $|\lambda_i| < |\lambda_0|$ for all such λ_i. In the first example, nonexistence depended only on the partition corresponding to $\mathbf{c}$, while in the second one, existence occurs only under drastic conditions on the support of $\mathbf{c}$, not simply its corresponding partition.

If $\mathbf{c}(\lambda)$ is always less than or equal to one (i.e., $|\operatorname{spec}A \cup \operatorname{spec}B| = n$), and the map is full, then there is an attractive and a repulsive fixed point if and only if for all choices of λ_i and μ_j, $|\lambda_i\mu_j| \neq 1$. The existence of an attractive fixed point in this case implies the existence of a repulsive fixed point, since every point in the graph has an antipode.

The first paragraph of the proof (of Proposition 11.1) shows that if the condition on the spectra holds, then there is a swap so that the lists satisfy the property needed for the eigenvalues of an attractive fixed point. In the nondefective case, we see that the pair obtained from the swap corresponds to an element of $\mathcal{G}_{\mathbf{c}}$, and (being nondefective) there thus exists a fixed point satisfying the sufficient conditions to be attractive.

However, in the defective case, there is no reason why the element of $\mathcal{G}_{\mathbf{c}}$ should be realizable by a fixed point, and thus there is no guarantee that there is an attractive (or repulsive) fixed point.

If X_0 is an attractive fixed point and X_1 is repulsive, then they correspond to a pair $(\mathbf{a},\mathbf{b})$ and its flip $(\mathbf{b},\mathbf{a})$; however, this is not sufficient (e.g., if $\mathbf{a} = \mathbf{b}$, as can certainly happen). It is also straightforward that $\operatorname{rank}(X_0 - X_1) = n$. A particular consequence is that if $\mathbf{c}(\lambda) > 1$ for all λ in $\operatorname{supp}\mathbf{c}$, there are only two points in the graph that can correspond to attractive or repulsive fixed points.

If the graph is the 3-point defective form of $2,1,1$ ($n = 2$; Section 10) in the form of a triangle, we see that any ϕ to which this corresponds cannot have both an attractive and a repulsive fixed point, since the rank of the difference between any two fixed points is one. If we construct such a $\phi_{C,D}$ (with, as usual, CD invertible), then it cannot be conjugate to $\phi_{D,C}$, since $\phi_{D,C}^{-1}$ *is* conjugate to $\phi_{C,D}$ and the orientation is reversed.

If the graph is the 5-point defective form of 1^4, then the one point with valence 4 is connected to everything else, while the other points have antipodes (maximal distance apart). So if the valence 4 point corresponds to an attractive fixed point, the system cannot have a repulsive one (and conversely). Again, such an example would have the property that $\phi_{C,D}$ is not conjugate to $\phi_{D,C}$.

Under some circumstances, we can define a directed graph structure on the fixed points. Suppose that X and X' are fixed points connected by an edge; then, the eigenvalue list for (XC, DX) is obtained by swapping one pair (inverting the second coordinate) from the list for $(X'C, DX')$. Point the edge towards the point (X or X') for which

the product of swapped eigenvalues has absolute value less than one (if the product has absolute value equalling one, the directed structure is not possible to construct). If this is defined and the graph is connected, and there is an attractive fixed point, the arrows will always point towards it.

12. Commutative cases

We can analyze the fixed point structure if $CD = DC$ or if in terms of (4.1), $AB = BA$. The first point is that if $CD = DC$, then $\phi_{C,D}$ has a fixed point which commutes with both C and D under very general conditions. Having such a fixed point, the other fixed points are obtained from our reduction to (4.1) with $AB = BA$. Then we can analyze one important case of the latter.

The following is elementary.

Lemma 12.1. *Suppose that R is an invertible $n \times n$ matrix. There exists a polynomial p such that $S := p(R)$ satisfies $S^2 = R$. In particular, R has a square root that commutes with whatever commutes with R.*

Proof. Let $\{\lambda_i\}$ be the set of *distinct* eigenvalues of R. By conjugating R, we may write it as the matrix direct sum of matrices of the form $\lambda_i \mathrm{I}_i + N_i$ where I_i are identity matrices (each of size equalling the algebraic multiplicity of λ_i) and N_i are nilpotent matrices. This is not of course the Jordan normal form (unless R is nonderogatory). It is routine to see that each $\lambda_i \mathrm{I}_i + N_i$ direct sum with the zero matrices of the appropriate sizes is a polynomial in the conjugated R. Since each λ_i is not zero, we can use the power series expansion for $(1+x)^{1/2}$ to obtain a square root for these (the power series terminates since N_i is nilpotent). Adding these square roots, we obtain a square root of the conjugate, and since each of the summands is a polynomial in the conjugate, so is the sum. □

If we replace "R is invertible" by "R admits a square root," the result fails—it is easy to construct a nilpotent 4×4 matrix which has a square root, but has no square root which commutes with whatever commutes with the original.

Now assume that $CD = DC$ and CD is invertible. Assume that 1/4 is not in the spectrum of CD. With $R = (\mathrm{I} - 4CD)/4$, we find S, a polynomial in R, such that $S^2 = R$; what is important is that S commutes with both C and D. Set $Y = \mathrm{I}/2 + S$. Then X defined by $CXD = Y$ (i.e., $X = C^{-1}YD^{-1}$) commutes with C and D, and it is easy to check that X is a fixed point of $\phi_{C,D}$ (the equation $Y^2 - Y + CD = \mathbf{0}$ yields $X(\mathrm{I} - CXD) = \mathrm{I}$). It follows that $A = CX$ and $B = (DX)^{-1}$ commute with each other. The remaining fixed points of ϕ can be found by analyzing (4.1) with $AB = BA$.

Proposition 12.2. *If $CD = DC$ and $\operatorname{spec} CD \cap \{0, 1/4\} = \varnothing$, then $\phi_{C,D}$ admits a fixed point which commutes with both C and D. The remaining fixed points can be found by analyzing the corresponding equation $U^2 = UB - AU$ where $AB = BA$ commute.*

The condition that 1/4 not belong to $\operatorname{spec} CD$ cannot be dropped, because of our example (see Section 2) with $C = \mathrm{I}$ in which $\phi_{C,D}$ had no fixed points whatsoever. Of course, scalar examples show that this condition is not necessary for the existence of a fixed point.

Now we analyze (4.1) with invertible $AB = BA$. The first and simplest case occurs when the algebra generated by A and B, denoted $\mathcal{A} = \langle A, B\rangle$, decomposes in a particularly nice way.

Proposition 12.3. *Suppose that A and B are commuting nonderogatory matrices. If* $\operatorname{spec} A \cap \operatorname{spec} B = \emptyset$, *then $U^2 = UB - AU$ has exactly 2^k solutions, where k is the number of minimal idempotents in $\mathcal{A} := \langle A, B\rangle$; all the solutions lie in $\mathcal{A}$.*

Proof. From the commuting property and that the matrices are both nonderogatory, each is a polynomial in the other one; moreover, a Jordan subspace of one is also a Jordan subspace of the other, so the number of Jordan blocks, k, is the same for both. Let $\{Y_i\}_{i=1}^{k}$ be a listing of the Jordan subspaces of $\mathbf{C}^{n\times 1}$. We notice that if $S \subset \{1,2,\ldots,n\}$, then $Y_S := \sum_{i\in S} Y_i$ is a direct sum of $\mathcal{A}$-invariant subspaces, admitting a complementary invariant subspace, $Y_{S^c} := \sum_{i\notin S} Y_i$. We first obtain 2^k solutions, one for each subset of $\{1,2,\ldots,n\}$, and then show that there are at most 2^n.

For each subset S, set $U_S = (B - A)|Y_S \oplus \mathbf{0}|Y_{S^c}$. It is obvious that each of these are solutions, and of course there are 2^k; moreover, they commute with each other, and with all the elements of $\mathcal{A}$. The minimal idempotents of $\mathcal{A}$ are given by $\mathrm{I}|Y_{\{i\}} \oplus \mathbf{0}|Y_{i^c}$, and thus there are k of them.

Now we show that 2^k is an upper bound on the number of solutions. To this end, we calculate the matrix $\mathcal{F}$ that appeared in (4.5) of Section 4. Without loss of generality, we may put A in Jordan normal form. The blocks correspond to the Y_i, and we note that A and B leave invariant exactly the same subspaces, both on the left and the right. When we calculate the matrix $\mathcal{F}$, we see immediately that it is a matrix direct sum of matrices of the form

$$Z_l := \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 & 2 \\ 0 & 0 & \cdots & 1 & 2 & 3 \\ & & \cdots & & & \\ 0 & 1 & \cdots & l-3 & l-2 & l-1 \\ 1 & 2 & \cdots & l-2 & l-1 & l \end{pmatrix}, \tag{12.1}$$

the decomposition corresponding to the block decomposition of A. The equation $\mathcal{F} = J_B^T V - V J_A$ (see the relevant section) obtained by restricting to invariant subspaces (left and right) can only be solved by an invertible when the subspace is a direct sum of Y_i (as is easy to see from the Jordan forms), and so there are at most 2^k solutions. We have already written down 2^k obvious solutions, so these must be all of them. □

There are two sources of difficulty in attempting to extend this or a similar result to possibly derogatory matrices. One is the presence of multiple eigenvectors; the other is the possibility that there exists a (left) A-invariant subspace of $\mathbf{C}^{n\times 1}$ that is not B-invariant, or a (right) B-invariant subspace of $\mathbf{C}^{1\times n}$ that is not A-invariant. Here is an elementary example showing what can happen.

Example 12.4. Let $A = s\mathrm{I}$ and B be $n \times n$ matrices such that s does not belong to $\operatorname{spec} B \cup \{0\}$, and B has distinct eigenvalues.

(a) Equation (4.1) has exactly 2^n solutions if and only if $2s \notin \{\mu_1 + \mu_2 \mid \mu_i \in \operatorname{spec} B, \mu_1 \neq \mu_2\}$;
(b) otherwise, (4.1) has an affine line of solutions.

Proof. Equation (4.1) reduces to $U^2 - U(B - s\mathrm{I}) = \mathbf{0}$, which we rewrite as $(2U - (B - s\mathrm{I}))^2 = (B - s\mathrm{I})^2$. The right-hand side is diagonalizable, with spectrum $\{(\mu_i - s)^2\}$. Its eigenvalues are thus distinct if and only if the condition of (a) holds, and of course, this means that it has exactly 2^n square roots, so there are 2^n solutions.

If the condition fails, then the right-hand side has a multiple eigenvector, and thus the right-hand side has an affine line of square roots, yielding (b).

If we permit s to belong to $\operatorname{spec} B$, then in the presence of the hypothesis of (a), there will be exactly 2^{n-1} solutions; again, if (a) fails, there will be an affine line of solutions. □

We can completely analyze the situation in the generic commutative case, that is, when C and D are simultaneously diagonable. We do this directly, that is, without recourse to any of the preceding methods.

Proposition 12.5. *Let C and D be invertible simultaneously diagonable $n \times n$ matrices, and let $\{w\}$ be a basis for the vector space on which they act, consisting of eigenvectors, with corresponding eigenvalues $\{c_w\}$, $\{d_w\}$, respectively. Let T denote the set of sums $\{r_w + r'_w\}$ where r_w and r'_w are solutions to the quadratic equations $\lambda^2 - \lambda/d_w + d_w/c_w = 0$ and $\lambda^2 - \lambda/d_{w'} + d_{w'}/c_{w'} = 0$, respectively, with $w \neq w'$. If $\operatorname{spec} D^{-1} \cap T$ is empty, then $\phi_{C,D}$ has at most 2^n fixed points, these commute with C and D and are simultaneously diagonable with each other and C and D.*

Proof. Let X be a fixed point. From $X - XCXD = \mathrm{I}$, we have $Y^2D - Y + C = \mathbf{0}$, where $Y = CX$. Hitting the quadratic (in matrices) with each of the eigenvectors w, we obtain $(Y^2 - Y/d_w) = -c_w w/d_w$.

It is an exercise involving Jordan normal forms to show that if Z is a square matrix and α is a scalar, then $Z^2 - \alpha Z$ generates a proper subalgebra of that generated by Z (i.e., Z is not in the algebra generated by $Z^2 - \alpha Z$ if and only if at least one of the following conditions holds):

(i) for some eigenvalue r of Z with a Jordan block of size exceeding one, $\alpha = 2r$;
(ii) there exist distinct eigenvalues r and r' of Z such that $\alpha = r + r'$.

In either event, at least one of the eigenvalues of $Z^2 - \alpha Z$ will have multiplicity exceeding one, and all eigenvalues are of the form $r^2 - \alpha r$ for some eigenvalue r of Z.

The product $\prod_w (Y^2 - Y/d_w + c_w\mathrm{I}/d_w)$ is the zero matrix, so that each eigenvalue of Y is a root of at least one of the quadratic equations, $\lambda^2 - \lambda/d_w + c_w/d_w = 0$; moreover, for each w, one of the two roots of the quadratic is an eigenvalue of Y. In particular, $\operatorname{spec} Y$ with multiplicities is obtainable by selecting one of the two roots of the quadratic as w is allowed to vary.

Set $\alpha = 1/d_w$ and $Z = Y$. The hypothesis on the spectrum of D^{-1} ensures that condition (ii) fails, and (i) fails automatically. Hence Y is a polynomial in $Y^2 - Y/d_w$ for

each w. We may thus find polynomials p_w such that for each w, $Yw = p_w(Y^2 - Y/d_w)w = p_w(-c_w/d_w)w$. Thus the w are eigenvectors for Y, so that Y is simultaneously diagonable with C and D. Since $X = C^{-1}Y$, X is similarly simultaneously diagonable.

With respect to the basis $\{w\}$, X, D, and C are diagonal, and the fixed points reduce to the fixed points of the n uncoupled ordinary fractional linear transformations, each of which admits only two fixed points. □

What happens if the hypotheses on the spectrum fail? Then there will be a continuum of fixed points, which do not commute with each other. The condition on the spectrum fails only under very rare circumstances. Here is about the smallest example possible.

Example 12.6. The transformation ϕ defined by $\phi(X) = (\mathrm{I} - CXD)^{-1}$ for $n = 2$, with the following properties:

(a) the matrices C and D are invertible diagonal nonnegative matrices with distinct eigenvalues, and each of the corresponding uncoupled one-dimensional fractional linear transformations has an attractive fixed point;
(b) ϕ has a continuum of fixed points, but no isolated ones;
(c) for at least one fixed point X, $\|X\|^2\|CD\| = 1/2$ for some operator algebra norms, but for all fixed points, $\|X\|^2\|C\|\|D\| > 1$ in all operator algebra norms;
(d) if E is any invertible 2×2 matrix, then the set of fixed points of $\psi : X \mapsto (\mathrm{I} - EX)^{-1}$ is not homeomorphic to that of ϕ.

Proof. Let s be a real or complex number, and set

$$D = \begin{bmatrix} 1 & 0 \\ 0 & \frac{9}{4} \end{bmatrix}, \qquad C = \begin{bmatrix} \frac{2}{9} & 0 \\ 0 & \frac{1}{12} \end{bmatrix}, \qquad X_s = \begin{bmatrix} \frac{3}{2} & s \\ 0 & \frac{4}{3} \end{bmatrix}. \tag{12.2}$$

It is easy to verify that X_s is a fixed point of ϕ for any value of s. All other fixed points are upper triangular with either the same eigenvalues or are obtained by (independently) replacing 3/2 by 3 or 4/3 by 4.

To check that it has no isolated fixed points, by Proposition 2.2, any such would belong to the algebra generated by $\{C, D\}$, that is, would have to be diagonal, and it is easy to check there are four such diagonal fixed points, which already appear as part of the four continua (the ones with $s = 0$).

To prove (d), just note that if X is a fixed point of ψ, then it must commute with E and satisfy $X^2E - X + \mathrm{I} = \mathbf{0}$; we can assume E is invertible. If E is nonderogatory, then X commuting with E entails that X is a polynomial in E, and thus $(X - E^{-1}/2)^2 = E^{-2}/4 - E^{-1}$. The latter has either zero, two, or four square roots, so there are at most four fixed points, necessarily isolated.

If E is derogatory, being of size two, it must be scalar. It is easy to check there can be at most three copies of punctured $\mathbf{C}$-planes among the fixed points after deleting any finite set of points, which is not the case for ϕ, as it has four. □

In particular, this example cannot be realized (up to almost any reasonable notion of equivalence) by a transformation with a matrix on one side only.

13. Commutative modulo the radical

Here we study a special case of a generalization of commuting C and D. Suppose that $\mathcal{A}$ denotes the algebra generated by C and D, and that C commutes with D modulo the radical of $\mathcal{A}$. The first step is to deal with the single block case, that is, when C and D have just one eigenvalue. Even here, the argument is surprisingly tedious and does not yield the optimal result, largely because of its dependence on a fixed point theorem.

The following lemma is ancient and obvious.

Lemma 13.1. *Fix a positive integer n. For each positive real ϵ, there exists an algebra norm $\|\cdot\|$ on $\mathrm{M}_n\mathbf{C}$ such that for all strictly upper triangular matrices M with $|M_{ij}| \le 1$, $\|M\| \le \epsilon$.*

Proof. For $\delta > 0$, let D_δ be the diagonal matrix $\mathrm{diag}(1, \delta, \delta^2, \dots, \delta^{n-1})$. Define the norm via $\|A\| = \|D_\delta A D_\delta^{-1}\|_1$ (where the subscript 1 denotes the 1-1 operator norm; any p-p norm would do). For δ sufficiently large, the norm will have the desired property. □

Lemma 13.2. *Suppose that $C = c\mathrm{I} + N_c$ and $D = d\mathrm{I} + N_d$ are $n \times n$ complex matrices and each of N, N_c, and N_d is strictly upper triangular. Suppose that a is a root of $a(1 - cad) = 1$ and $|a^2 cd| \ne 1$. Then $\phi_{C,D}$ has at least two fixed points which are of the form $b\mathrm{I} + N_b$ where b varies over the two roots of $a(1 - cad) = 1$, and each N_b is strictly upper triangular. Moreover, $\phi_{C,D}$ has both an attractive fixed point and a repulsive one, and they are in this form.*

Proof. First we assume that we can choose a root of $a(1 - cad) = 1$ with $|a^2 cd| < 1$, and show that $\phi \equiv \phi_{C,D}$ has a fixed point in the set $S_a := \{a\mathrm{I} + N \mid N \text{ is strictly upper triangular}\}$, and this fixed point is attractive. After this, we observe that if instead we can choose the root to satisfy $|a^2 cd| > 1$, the methods used in the first case can be applied to the inverse of ϕ, and this will yield a repulsive fixed point for $\phi_{C,D}$. Finally, we show that if ϕ has either an attractive or a repulsive fixed point, then it also has a fixed point with the other property, via a more or less explicit construction.

To begin, let $a(1 - cad) = 1$. Form S_a as above; it is a translation of $\mathbf{C}^{n(n-1)/2}$ and a subset of $\mathrm{M}_n\mathbf{C}$. Since a is a root of the quadratic, it follows that S_a is stable under the action of ϕ (i.e., $\phi(S_a) \subseteq S_a$), and of ϕ^{-1}, and ϕ restricted to S_a is a homeomorphism.

Now assume that $|a^2 cd| < 1$. We show that ϕ leaves stable a closed ball of S_a (with respect to an appropriate algebra norm inherited from $\mathrm{M}_n\mathbf{C}$), and so we may apply Brouwer's fixed point theorem. Let $\|\cdot\|$ be a norm on $\mathrm{M}_n\mathbf{C}$ with the property described in the previous lemma, with ϵ to be prescribed later. Let u be a positive real number, and let B_u be the closed ball of radius u restricted to S_a centred at $a\mathrm{I}$. We determine conditions on a so that $\phi(B_u) \subseteq B_u$, that is, B_u is stable with respect to ϕ.

Let N be a strictly upper triangular matrix, and consider $X = a\mathrm{I} + N$, that is, X is an arbitrary point of S_a; now we impose the condition that $\|N = X - a\mathrm{I}\| \le u$, and consider $\phi(X) = a(\mathrm{I} - aN')^{-1}$ where $N' = adN_c + acN_d + aN_cN_d + cdN + dN_cN + cNN_d + N_cNN_d$.

Obviously, $\|N'\| \le \|N\|(|cd| + t) + |a|t$, where $t = |c|\|N_d\| + |d|\|N_c\| + \|N_c\| \cdot \|N_d\|$. Since c, d, and N_c, and N_d are fixed and the latter two are strictly upper triangular matrices, by reducing ϵ (and changing the norm), we can make t arbitrarily small (once). We

have, since N' is nilpotent,

$$\begin{aligned}\|\phi(X) - a\mathrm{I}\| &= \left\| a\sum_{1}^{n-1} a^k (N')^k \right\| \le |a|^2 \|N'\| \left\| \sum_{0}^{n-2} a^k (N')^k \right\| \\ &\le \frac{|a|^2\|N'\|}{1-|a|\|N'\|} \le \frac{|a|^2\big(\|N\|\big(|cd|+t\big)+|a|t\big)}{1-\big(\|N\|\big(|acd|+|a|t\big)+|a|^2 t\big)},\end{aligned} \tag{13.1}$$

provided that $\|N\|(|cd|+t)+|a|t<1/|a|$ (this forces a restriction on the choice of u). Set $s=\|N\|$. We have to choose t sufficiently small so that for some u and every s in $[0,u]$, $f(s) := s(|a^2cd|+|a^2|t)+|a|t/(1-(s(|acd|+|a|t)+|a|^2 t)) \le u$, and at the same time $s(|cd|+t)+|a|t<1/|a|$. Of course, $f=(Ls+M)/(Ps+Q)$ is a fractional linear transformation (with real coefficients), and it is easy to check to determine its critical point. We note that for all sufficiently small t, $LQ>MP$, so that the critical point, $(MP-LQ)/LQ$, is negative. Thus, for all sufficiently small values of t, f is strictly increasing on $[0,u]$. Its maximum thus occurs at $s=u$.

Now $f(u)\le u$ if and only if $Pu^2+(Q-M)u-L\ge 0$, and since $N<0$, this boils down to

$$u^2\big(|cd|+t\big)-u\big(1-\big|a^2cd\big|-|a|t-\big|a^2\big|t\big)+\big|a^3\big|t\le 0. \tag{13.2}$$

Now we exploit that $1>|a^2cd|$. If $t=0$, the two solutions are simply 0 and $u_0=1-|a^2cd|/|cd|$. It easily follows that for sufficiently small t, any positive choice of u less than u_0 will do. To deal with the auxiliary condition $\|N\|(|cd|+t)+|a|t<1/|a|$, it suffices to require that $u(|cd|+t)+|a|t<1/|a|$, that is, $u<(|a|(|cd|+t)+|a|t)$. At $t=0$, this reduces to simply $u<|acd|$. This condition is thus sufficient for all sufficiently small t, so we take $u<\min\{u_0,|acd|\}$.

With this choice for u, the ball B_u in a copy of (affine) Euclidean space (S_a) is stable with respect to ϕ. By Brouwer's fixed point theorem, ϕ admits a fixed point, X_0, in B_u. The derivative, $\mathcal{M}_{X_0C,DX_0}$, has only one eigenvalue, a^2cd, whose absolute value is less than one, and thus the fixed point is attractive.

If instead $|a^2cd|>1$, then we apply a similar process (mercifully abbreviated here) to the inverse of ϕ, $X\mapsto C^{-1}(\mathrm{I}-X^{-1})D^{-1}$. This is again defined on all of S_a, and the calculation of $\|\phi^{-1}(X)-a\mathrm{I}\|$ is simpler, and a parallel argument works, except that we derive a fixed point for ϕ^{-1} that is attractive, that is, it is repulsive as a fixed point of ϕ.

To show that ϕ has both an attractive and a repulsive fixed point, we have to develop the notion (not new) of an antipode. If X is a fixed point of $\phi_{C,D}$ (for general C and D), an *antipode of* X is a fixed point Y of $\phi_{C,D}$ such that $Y-X$ is invertible, that is, of rank n. Not all fixed points have antipodes (e.g., if the graph of the fixed point set is not defective and not $\mathcal{G}_{1^{2n}}$, then not all fixed points have an antipode, or if the graph is a defective version of $\mathcal{G}_n$ with an odd number of points, then, again not all fixed points have antipodes). The antipode of an attractive fixed point is repulsive (and vice versa), as follows from spectrum swapping.

The existence of an antipode is relatively easy to check. If X is a fixed point, an antipode exists if and only if we can solve the equation $U^2=UB-AU$ (where $A=CX$

and $B = (DX)^{-1}$) with invertible U—as in Section 4, $Y = U + X$ will be a fixed point. Pre- and post-multiplying by U^{-1} yields $\mathrm{I} = BU^{-1} - U^{-1}A$. That is, as in the earlier section, an antipode will exist if and only if we can find an invertible solution V to $\mathrm{I} = BV - VA$ (without the invertibility hypothesis, this is known as Sylvester's equation). If $\operatorname{spec} B \cap \operatorname{spec} A = \varnothing$, a solution (not necessarily invertible) exists, and it is unique. If A and B generate (or simply belong to) an algebra which is commutative modulo its radical, any solution to Sylvester's equation is invertible (see Proposition 9.3). The fact that the solution obtained by the fixed point argument above means that both XC and DX generate such an algebra means that antipodes exist. □

The condition, $|a^2cd| \neq 1$, is equivalent to the much simpler condition, that cd not belong to the ray $\{r \in \mathbf{R} \mid r \geq 1/4\}$. This is straightforward. Suppose that both roots, a_i, of $x^2cd - x + 1 = 0$, satisfy $|a_i^2 cd| = 1$. From the quadratic, $a_1a_2cd = 1$, whence $|a_1|^2 = |a_2|^2 = 1/|cd|$. Write $1 - 4cd = r^2e^{i\theta}$ where $r \geq 0$ and $0 \leq \theta < 2\pi$. Then we can take $2a_1cd = 1 + re^{i\theta/2}$ and $2a_2cd = 1 - re^{i\theta/2}$. Taking absolute values and equating them, we obtain $4r\cos\theta/2 = 0$. Thus either $r = 0$ or $\theta = \pi$.

In either event, we have that $4|a_i|^2|cd|^2 = 1 + r^2$. On the other hand, the left-hand side is $4|cd|$, whence $r^2 = 4|cd| - 1$. If $r = 0$, then $cd = 1/4$. If $\theta = \pi$, then $r^2 = 4cd - 1$, forcing $cd \geq 1/4$. Conversely, if cd is real and at least as large as 1/4, the two roots are complex conjugates of each other, thus have equal modulus, hence $|a^2cd| = 1$.

Corollary 13.3. *If C and D are scalar plus strictly upper triangular with eigenvalue c and d, respectively, and $cd \notin \{r \in \mathbf{R} \mid r \geq 1/4\}$, then ϕ has an attractive and a repulsive fixed point, and they are upper triangular.*

We have already seen an example wherein $cd = 1/4$ and ϕ has no fixed points. In the commutative case, sufficient for the existence of a fixed point is that $1/4 \notin \operatorname{spec} C \cdot \operatorname{spec} D$; the presence of an attractive or repulsive point is equivalent to the additional condition that no real number at least as large as 1/4 appears in the set of products. Unfortunately, the argument here depends on a fixed point property, and does not work when the latter condition fails.

14. More fixed point existence results

This section gives weak sufficient criteria for the existence of fixed points, simply by exploiting the contraction mapping theorem and Brouwer's fixed point theorem. The existence theorems here are useful in the case of (entrywise) positive matrices, but in general are a faint shadow of what the final results ought to be.

Let $\mathscr{L}$ be an element of $\operatorname{End} \mathrm{M}_n\mathbf{C}$. Define the following generalization of spectral radius for $\mathscr{L}$. For each norm $\|\cdot\|$ on $\mathrm{M}_n\mathbf{C}$, define the corresponding operator norm $|||\cdot|||$ in the usual way, that is, $|||\mathscr{L}||| = \sup_{Y \neq \mathbf{0}} \|\mathscr{L}(Y)\|/\|Y\|$. Now define $\rho_n(\mathscr{L}) = \inf\{|||\mathscr{L}||| \mid \|\cdot\| \text{ is an algebra norm on } \mathrm{M}_n\mathbf{C}\}$. Recall that an *algebra norm* (in this case on $\mathrm{M}_n\mathbf{C}$) is a Banach space norm with the additional property that $\|YZ\| \leq \|Y\| \cdot \|Z\|$ for all pairs of elements Y and Z. We can always renormalize so that $\|\mathrm{I}\| = 1$, and so incorporate this into the definition.

An easy consequence of Jordan normal form is that for any $k \times k$ matrix A,

$$\rho(A) = \inf\{\|A\| \mid \|\cdot\| \text{ is an algebra norm on } \mathrm{M}_n\mathbf{C}\}. \tag{14.1}$$

Applying this to $\mathrm{End}\,\mathrm{M}_n\mathbf{C} \cong \mathrm{M}_{n^2}\mathbf{C}$, we see that all the norms on the latter arising in the display above are algebra norms, hence $\rho(\mathcal{L}) \leq \rho_n(\mathcal{L})$. Now $\rho(\mathcal{M}_{C,D}) = \rho(C)\cdot\rho(D)$ (use the natural tensor product decomposition), and so it is easy to calculate the former. Unfortunately, for proving results on the existence of fixed points, it is $\rho_n(\mathcal{M}_{C,D})$ that matters, and in many cases, $\rho_n(\mathcal{M}_{C,D}) > \rho(C)\cdot\rho(D)$.

For a square matrix C, C^* will denote conjugate transpose.

Proposition 14.1. *For C in $\mathrm{M}_n\mathbf{C}$, $\rho_n(\mathcal{M}_{C,C^*}) = \rho(CC^*)$.*

Proof. Since $\mathcal{M}_{C,C^*}(\mathrm{I}) = CC^*$ and for every algebra norm on $\mathrm{M}_n\mathbf{C}$ we have $\|CC^*\| \geq \rho(CC^*)$, it follows that $\rho_n(\mathcal{M}_{C,C^*}) \geq \rho(CC^*)$. On the other hand, for any algebra norm, $\|CYC^*\| \leq \|C\|\cdot\|C^*\|\cdot\|Y\|$, whence $\rho_n(\mathcal{M}_{C,C^*}) \leq \inf\{\|C\|\cdot\|C^*\|\}$, where the norm varies over all algebra norms. If $\|\cdot\|_2$ denotes the usual (2-2) operator norm on $\mathrm{M}_n\mathbf{C}$ (acting on $l^2(1,2,\ldots,n)$ in the usual way), we have $\rho_n(\mathcal{M}_{C,C^*}) \leq \|C\|_2\cdot\|C^*\|_2 = \|CC^*\|_2 = \rho(CC^*)$. □

If, in the context of this result, C is not normal, then $\rho(CC^*)$ can be strictly bigger than $\rho(C)^2$. For example, this occurs if C is nilpotent, or consists of Jordan blocks with at least one being of size exceeding one, and corresponding to an eigenvalue of maximal modulus. These yield (by small perturbations) examples where C has only strictly positive entries. An obvious inequality is that $\rho_n(\mathcal{M}_{C,D}) \leq \inf\{\|C\|\cdot\|D\|\}$ (restricted to normalized algebra norms).

On the other hand, when C and D are both normal (but not necessarily related to each other in any way), $\rho_n(\mathcal{M}_{C,D}) = \rho(C)\cdot\rho(D)$. In fact, $\mathcal{M}_{C,D}$ satisfies a stronger property. For $\mathcal{L}$ in $\mathrm{End}\,\mathrm{M}_n\mathbf{C}$, we say $\mathcal{L}$ *achieves* $\rho_n(\mathcal{L})$ if there is an algebra norm $\|\cdot\|$ on $\mathrm{M}_n\mathbf{C}$ such that $\rho_n(\mathcal{L}) = |||\mathcal{L}|||$ where $|||\cdot|||$ is the operator norm induced by $\|\cdot\|$.

Proposition 14.2. *If C and D are normal matrices of size n, then $\rho_n(\mathcal{M}_{C,D}) = \rho(C)\cdot\rho(D)$ and moreover, this is achieved by the 2-2 norm on $\mathrm{M}_n\mathbf{C}$.*

Proof. Obviously $\rho_n(\mathcal{M}_{C,D}) \leq \|C\|_2\|D\|_2$, and since the matrices are normal, this equals $\rho(C)\cdot\rho(D)$. Hence equality occurs. □

Lemma 14.3. *If $C = \mathrm{I}$ or if $D = \mathrm{I}$, or if $CD = DC$ and one of C, D is diagonalizable with distinct eigenvalues, then $\rho_n(\mathcal{M}_{C,D}) = \rho(C)\rho(D)$. In the last case, $\mathcal{M}_{C,D}$ achieves ρ_n. In the former cases, $\mathcal{M}_{C,D}$ achieves ρ_n if and only if the nonidentity matrix of the pair has the property that for every eigenvalue of modulus equaling the spectral radius, all corresponding Jordan blocks are of size one.*

Proof. If one of the pair is the identity, all the results about $\mathcal{M}_{C,D}$ are routine and follow from the earlier observation about algebra norms on $\mathrm{M}_n\mathbf{C}$. If $CD = DC$ and (say) C is diagonable with distinct eigenvalues, then diagonalizing C automatically diagonalizes D (since the centralizer of a diagonal matrix with distinct eigenvalues consists of diagonal matrices). If A does the diagonalizing, take the norm $\|A\cdot A^{-1}\|_2$ on $\mathrm{M}_n\mathbf{C}$ (or simply

observe that ρ_n is invariant under $\mathcal{L} \mapsto \mathcal{M}_{A,A}\mathcal{L}\mathcal{M}_{A^{-1},A^{-1}}$, and applying this to $\mathcal{M}_{C,D}$ yields diagonal, hence normal, matrices, so the preceding applies). □

The function ρ_n is introduced in order to use the contraction mapping and Brouwer's fixed point theorems.

Proposition 14.4. *Suppose that C and D are square matrices of size n. Let $\phi : X \mapsto (\mathrm{I} - CXD)^{-1}$ be the corresponding fractional linear transformation.*

(a) *If $\rho_n(\mathcal{M}_{C,D}) < 1/4$, then ϕ is contractive on the 2-ball of $\mathrm{M}_n\mathbf{C}$ for some algebra norm, and there exists an attractive fixed point of norm less than 2. Moreover, $\{\phi^N(\mathbf{0})\}$ converges to the fixed point.*

(b) *If $\rho_n(\mathcal{M}_{C,D}) = 1/4$ and $\mathcal{M}_{C,D}$ achieves ρ_n, then ϕ has a fixed point of spectral radius at most 2.*

Proof. In the first case, there exists an algebra norm $\|\cdot\|$ on $\mathrm{M}_n\mathbf{C}$ such that $|||\mathcal{M}_{C,D}||| < 1/4$ in the corresponding operator norm, and thus for every Y in the 2-ball (using $\|\cdot\|$), we have $\|\mathcal{M}_{C,D}(Y)\| < \|Y\|/4 \le 1/2$. Hence $\|(\mathrm{I} - CYD)^{-1}\| < 1/(1 - 1/2) = 2$. So ϕ acts as a (strict) contraction on the 2-ball, and thus by the contraction theorem, all results in (a) follow.

In the second case, using the norm that achieves ρ_n, the same calculation as in (a) yields that the closed 2-ball is stable under ϕ, so by Brouwer's fixed point theorem, ϕ has a fixed point therein. □

15. Still more on existence

We can reformulate the existence of fixed points in terms of solutions to (everywhere defined) matrix equations. As usual, suppose that C and D are invertible. Then X is a fixed point of $\phi_{C,D}$ if and only if $X(\mathrm{I} - CXD) = \mathrm{I}$, that is, $XCXD = X + \mathrm{I}$. Premultiply by C and set $Y = CX$, so the equation becomes $Y^2D - Y = C$, that is, $Y^2 - YD^{-1} = CD^{-1}$. Set $\mathbf{D} = D^{-1}$. For D fixed, $\phi_{C,D}$ will have a fixed point for every choice of invertible C if the map $Y \mapsto Y^2 - Y\mathbf{D}$ from $\mathrm{M}_n\mathbf{C}$ to itself, is onto. It is slightly more convenient to look at $Y^2 - \mathbf{D}Y$, and ontoness of this map is equivalent to that of the former. (Invertibility of C and D is necessary to go from a solution to $Y^2 - Y\mathbf{D} = CD^{-1}$ to a fixed point of $\phi_{C,D}$.)

As Daniel Daigle pointed out to me, for any matrix $\mathbf{D}$, the map $\Psi_{\mathbf{D}} : \mathrm{M}_n\mathbf{C} \to \mathrm{M}_n\mathbf{C}$ given by $\Psi_{\mathbf{D}}(Y) = Y^2 - \mathbf{D}Y$ always has dense range, as a consequence of Chevalley's theorem (which we will explain shortly). In particular, it follows that the set of (C, D) in $\mathrm{GL}(n, \mathbf{C}) \times \mathrm{GL}(n, \mathbf{C})$ for which $\phi_{C,D}$ has a fixed point, is dense. It is easy to see that if $\mathbf{D}$ is a scalar matrix, then $\Psi_{\mathbf{D}}$ is not onto, and similarly, if $\mathbf{D}$ has an eigenvalue with algebraic multiplicity exceeding one, but geometric multiplicity equalling one, then $\Psi_{\mathbf{D}}$ is not onto, and this is very likely true for every $\mathbf{D}$ with an algebraic multiplicity exceeding one.

We show that when $n = 2$, $\Psi_{\mathbf{D}}$ is onto if and only if $\mathbf{D}$ has distinct eigenvalues (which of course is equivalent to D having distinct eigenvalues), and then note some difficulties encountered in higher n, particularly with upper triangular matrices. The size-two matrix result is based on a technique suggested by my colleague Daniel Daigle.

Proposition 15.1 (Daigle). *For any $n \times n$ complex matrix $\mathbf{D}$, the map from $\mathrm{M}_n\mathbf{C}$ to itself given by $\Psi_{\mathbf{D}} : Y \mapsto Y^2 - \mathbf{D}Y$ has dense range.*

Remark 15.2. An immediate consequence is that for every D, the set of C such that $\phi_{C,D}$ has a fixed point is dense. Combined with our earlier results, this means that generically $\phi_{C,D}$ has $\binom{2n}{n}$ fixed points.

Proof. We first observe that the derivative of the map is given by $\mathbf{R}_{Y-\mathbf{D},Y}$ (left multiplication by $Y-\mathbf{D}$ plus right multiplication by Y), and this is nonsingular if $\operatorname{spec}(\mathbf{D}-Y)\cap \operatorname{spec} Y=\emptyset$; it is routine to show that this is true for at least one Y (and hence for almost all), for example, if $Y=\mathbf{0}$ and $\mathbf{D}$ is invertible. In particular, the image of $\Psi_{\mathbf{D}}$ contains an open set.

Next, $\Psi_{\mathbf{D}}$ is obviously a polynomial map when viewed as a map from $\mathbf{C}^{n^2}$ to itself, hence by Chevalley's theorem, the range is a constructible subset of $\mathbf{C}^{n^2}$ (since the spectrum of algebraic geometry, i.e., the maximal ideal space of $\mathbf{C}[x_{ij}]$, omitting the point at infinity, is just $\mathbf{C}^{n^2}$). A constructible set that contains an open subset (in the usual topology) is dense (usual topology). □

Lemma 15.3. *Suppose that $\alpha\mathrm{I}+\mathbf{D}^2/4$ has no square root for some complex number α. Then the range of $\Psi_{\mathbf{D}}$ does not contain at least one scalar multiple of the identity.*

Remark 15.4. This yields a lot of nononto results. If $\mathbf{D}$ is 2×2 and consists of a single Jordan block, say with eigenvalue r, then $-r^2\mathrm{I}/4+\mathbf{D}^2/4$ is nonzero and nilpotent, hence has no square root. If $\mathbf{D}$ is $n\times n$ and has an eigenvalue, r, with only one Jordan block and the size of the block exceeds one, then again $-r^2\mathrm{I}/4+\mathbf{D}^2/4$ has no square roots. When there are multiple blocks (for the same eigenvalue), at least one of which is larger than one, the situation is more complicated.

Proof. If $\Psi_{\mathbf{D}}(Y)=\alpha\mathrm{I}$ for a scalar α, then $(Y-\mathbf{D})Y=\alpha\mathrm{I}$, from which it easily follows that $Y\mathbf{D}=\mathbf{D}Y$. Thus $\alpha\mathrm{I}=(Y-\mathbf{D}/2)^2-\mathbf{D}^2/4$, so that $\alpha\mathrm{I}+\mathbf{D}^2/4$ admits a square root. □

Proposition 15.5. *Let $n=2$ and $\mathbf{D}$ be a matrix with distinct eigenvalues. Then $\Psi_{\mathbf{D}}:\mathrm{M}_2\mathbf{C}\to\mathrm{M}_2\mathbf{C}$ is onto.*

Remark 15.6. The method in this argument was suggested by Daniel Daigle. The converse is easily seen to be true (for $n=2$)—by Lemma 15.3, we need only to consider the case that $\mathbf{D}$ is scalar, but then Y commutes with $\mathbf{D}$, and completing the square yields matrices not in the range.

Proof. Without loss of generality, we may assume that $\mathbf{D}$ is diagonal. Let $\mathbf{D}=\operatorname{diag}(d_1,d_2)$ with $d_1\neq d_2$. Let $Y=\left(\begin{smallmatrix}x_{11}&x_{12}\\x_{21}&x_{22}\end{smallmatrix}\right)$ be the unknown matrix, and let $F=\left(\begin{smallmatrix}f_{11}&f_{12}\\f_{21}&f_{22}\end{smallmatrix}\right)$ be the target matrix, that is, we wish to solve for Y so that $Y^2-\mathbf{D}Y=F$. We abbreviate $x_{11}+x_{22}=t$ and $(t-d_1)(t-d_2)=s$. Expanding, we obtain the four equations,

$$\begin{aligned} f_{11}&=x_{11}^2-d_1x_{11}+x_{12}x_{21}, & f_{12}&=x_{12}(t-d_1),\\ f_{21}&=x_{21}(t-d_2), & f_{22}&=x_{22}^2-d_2x_{22}+x_{12}x_{21}. \end{aligned} \tag{15.1}$$

From the equations on the right, we obtain expressions for x_{12} and x_{21}. Substituting these

into the left equations, we obtain

$$\begin{aligned} sf_{11} &= sx_{11}^2 - sd_1x_{11} + f_{12}f_{21}, \\ sf_{22} &= sx_{22}^2 - sd_2x_{22} + f_{12}f_{21}. \end{aligned} \tag{15.2}$$

Set $P = s(f_{22} + d_2^2/4) - f_{12}f_{21}$, so the second equation is equivalent to $s(x_{22} - d_2/2)^2 = P$. Similarly, we have $s(x_{11} - d_1/2)^2 = Q$, where $Q = s(f_{11} + d_1^2/4) - f_{12}f_{21}$. We have $s(t - x_{11} - d_2/2)^2$, so $x_{11} + d_2/2 = t \pm \sqrt{P/s}$. Rewrite $x_{11} - d_1/2 = (x_{11} + d_2/2) - (d_1 + d_2)/2)$ and substitute this into the expression $\cdots = Q$. We obtain

$$\begin{aligned} \frac{Q}{s} &= \left(x_{11} + \frac{d_2}{2}\right)^2 - (d_1 + d_2)\left(x_{11} + \frac{d_2}{2}\right) + \frac{(d_1 + d_2)^2}{4} \\ &= t^2 + \frac{P}{s} \pm 2t\sqrt{\frac{P}{s}} - (d_1 + d_2)\left(t \pm \sqrt{\frac{P}{s}}\right) + \frac{(d_1 + d_2)^2}{4}. \end{aligned} \tag{15.3}$$

Isolating the surds and noting that the $\pm$ terms are concordant (both are negative, or both are positive), we have

$$\pm\sqrt{\frac{P}{s}}2\left(t + \frac{d_1 - d_2}{2}\right) = \frac{Q - P}{s} - t^2 - \frac{(d_1 + d_2)^2}{4} + (d_1 + d_2)t. \tag{15.4}$$

Before we square to eliminate the surd (and resolve the ambiguity in the $\pm$ terms), we note that $(Q - P)/s = f_{11} - f_{22} + (d_1^2 - d_2^2)/4$, that is, it belongs to $\mathbf{C}[f_{ij}]$ already. We also rewrite $-t^2 + (d_1 + d_2)t = -(t - (d_1 + d_2)/2)^2 + (d_1 + d_2)^2/4$, so the right-hand side becomes $(Q - P)/s - T^2$ where $T = t - (d_1 + d_2)/2$. We obtain

$$s\left(\frac{Q - P}{s} - T^2\right)^2 = 4P \cdot T^2. \tag{15.5}$$

Since $s = (t - d_1)(t - d_2) = T^2 - (d_1 - d_2)^2/4$, the last displayed equation yields a monic sextic (bicubic) equation for $T = t - (d_1 + d_2)/2$ with coefficients from $\mathbf{C}[f_{ij}]$ (in particular, t itself satisfies a monic sextic). It is convenient to obtain the corresponding cubic that is satisfied by s, simply by replacing T^2 by $s + (d_1 - d_2)^2/4$,

$$p_s := s\left(\frac{Q - P}{s} - \left(s + \left(\frac{d_1 - d_2}{2}\right)^2\right)\right)^2 - 4P \cdot \left(\frac{d_1 - d_2}{2}\right)^2\Bigg). \tag{15.6}$$

To determine polynomials over $\mathbf{C}[f_{ij}]$ (not necessarily monic) satisfied by the x_{ij}, we make use of the following observation. Suppose that $R \subset A$ is an inclusion of commutative domains, and a, b are elements of A such that a is integral over R satisfying a monic polynomial q with nonzero constant term and $r := ab$ belongs to R. Then b satisfies a polynomial with coefficients from R whose leading coefficient is $q(0)$ (the constant term of q). The proof is elementary—if $a^n + \sum a^i r_i = 0$, on multiplying by b^n, we obtain $r_0 b^n + \sum r^i r_i b^{n-i} = 0$.

In our situation, $A = \mathbf{C}[x_{ij}]$ and $B = \mathbf{C}[f_{ij}]$. The equation $s(x_{11} - d_1/2)^2 = Q$ yields that $(x_{11} - d_1/2)^2$ (and thus $x_{11} - d_1/2$) satisfies a polynomial with leading term $p_s(0) = -f_{12}f_{21}(d_1 - d_2)^2$ (which is nonzero—an essential hypothesis—since $d_1 \neq d_2$, the first time we use this). By interchanging 1 and 2, we obtain that $x_{22} - d_2/2$ satisfies a polynomial with leading term $-f_{12}f_{21}(d_1 - d_2)^2$ (the same one).

We also apply this technique to $(t - d_2)x_{21} = f_{21}$. We have a sextic satisfied by t (in the form of a bicubic in $t - (d_1 + d_2)/2$), and to determine the constant term in the corresponding equation satisfied by d_2, we simply evaluate at $t \mapsto d_2$. Under this map, $s \mapsto 0$, so we obtain $p_s(0) = -f_{12}f_{21}$ again. By symmetry, this is also the leading coefficient in an equation satisfied by x_{12}. Obviously $\mathbf{C}[x_{i,j}] = \mathbf{C}[x_{ii} - d_i, x_{12}, x_{21}]$, so that the four generators of this ring have leading terms (dividing) $f_{12}f_{21}$.

Now we look at the entries of the target matrix; let F be a specific matrix obtained by evaluating the four entries at a point of $\mathbf{C}^{2\times 2}$. If $f_{12}f_{21} \mapsto z \neq 0$, there exists an algebra map from $\mathbf{C}[x_{ij}]$ extending this evaluation (easy to check from the leading terms of the polynomials), and the matrix resulting from evaluating Y is a desired preimage of F. (Generically, there will be six such algebra maps, which correspond to the generic six fixed points of the fractional linear matrix transformation.)

If under this evaluation $f_{12}f_{21} \mapsto 0$, then the target matrix is either upper or lower triangular, so the problem reduces to showing that all upper triangular and all lower triangular matrices are in the range of $Y \mapsto Y^2 - \mathbf{D}Y$. In fact, symmetry yields that it is sufficient to do this for all upper triangular matrices, if we can do this for all diagonal matrices with distinct eigenvalues, by conjugating with the nontrivial permutation matrix. Now we proceed to the upper triangular case.

Here $F = \left(\begin{smallmatrix} f_{11} & f_{12} \\ 0 & f_{22} \end{smallmatrix}\right)$ and we look for a solution to $Y^2 - \mathbf{D}Y = F$ of the form $Y = \left(\begin{smallmatrix} x_{11} & x_{12} \\ 0 & x_{22} \end{smallmatrix}\right)$. The matrix equation leads to the equations

$$x_{11}^2 - d_1x_{11} = f_{11}, \qquad x_{22}^2 - d_2x_{22} = f_{22}, \qquad x_{12}(x_{11} + x_{22} - d_1) = f_{12}. \tag{15.7}$$

Let $x_{11} = x$ and $x_{22} = y$ be respective solutions to the first two equations. If $x + y \neq d_1$, then we can set $x_{12} = f_{12}/(x + y - d_1)$ and we are done. If $x + y = d_1$, and if either equation has distinct roots, we can replace one of the values for x or y by the other root, forcing $x + y \neq d_1$. This leaves the possibility that both quadratics have double roots and $x + y = d_1$, which we now show is impossible.

If the equations have double roots, then $x = d_1/2$ and $y = d_2/2$, which yields $d_1 = x + y = d_1/2 + d_2/2$, that is, $d_1 = d_2$, a contradiction. □

Upper triangular matrices. Let $\mathcal{A}$ denote the algebra of $n \times n$ upper triangular matrices. We can get a great deal of information on $\Psi_{\mathbf{D}}(\mathcal{A})$, the range of $\Psi_{\mathbf{D}}$ restricted to $\mathcal{A}$. Even in the case that $\mathbf{D}$ is diagonal with distinct eigenvalues, it is not true that $\Psi_{\mathbf{D}}(\mathrm{M}_n\mathbf{C}) \cap \mathcal{A} = \Psi_{\mathbf{D}}(\mathcal{A})$ (the left-hand side can be strictly larger than the right), that is, there can be an upper triangular matrix A for which the equation $\Psi_{\mathbf{D}}(Y) = A$ can be solved, but not with Y in $\mathcal{A}$.

The sufficient condition on the eigenvalues is similar to that appearing in the commutative case.

Proposition 15.7. *Let* $\mathbf{D}$ *be an upper triangular matrix with diagonal part* ∂ *and diagonal entries* ∂_i. *Let* F *be an upper triangular matrix with diagonal part* Δ *and diagonal entries* f_i. *Suppose that* d_i $(i = 1,\dots,n)$ *satisfying* $d_i^2 - \partial_i d_i = f_i$ *can be chosen such that for all* $j > i$, $d_j + d_i \neq \partial_i$. *Then there exists upper triangular* Y *such that* $\Psi_{\mathbf{D}}(Y) = F$.

Remark 15.8. Generically there will be two solutions for each of the n quadratics, so in principal one would have to try 2^n selections of the roots to obtain the inequalities. In practice, far fewer choices are needed. The inequalities are necessary to the extent that if there is no such choice of roots, then there exists an upper triangular F with f_i along the diagonal that is not of the form $\Psi_{\mathbf{D}}(Y)$ with Y upper triangular.

If d_i satisfies $d_i^2 - \delta_i d_i = f_i$, then the other root is just $\partial_i - d_i$. The condition in the inequalities is just that d_j and d_i are not "conjugate" roots (we are dealing here with complex coefficients, so conjugate is not the right word). If the quadratic has just one root, then $d_i = \partial_i/2$.

Remark 15.9. This result yields an existence result for fixed points of fractional linear matrix transformations, in the case that C and D can be simultaneously upper triangularized. It involves only the eigenvalues of $F = CD^{-1}$, which here are determined from the eigenvalues of C and D.

Proof. The proof is via an elementary modification of back substitution. Write $\mathbf{D} = \partial + Z$, $F = S + M$, $Y = \Delta + N$, where the second letter is the strictly upper triangular part of the matrix. We do not assume that any of Z, M, or Y is not zero. We are trying to solve for Y so that $Y^2 - \mathbf{D}Y = F$. We write $N_{ij} = x_{ij}$ and F_{ii} is just f_i. Expanding the product $Y^2 - \mathbf{D}Y$, we obtain

$$(\Delta^2 - \partial\Delta) + (N\Delta + \Delta N - \partial N - Z\Delta) + (N^2 - ZN). \tag{15.8}$$

Calculating the (i,i) coordinates, we obtain the equations $d_i^2 - \partial_i d_i = f_i$, and by hypothesis, we can choose roots d_i to satisfy the inequalities $d_j + d_i \neq \partial_i$ for all $j > i$. Now we calculate the $(i, i+1)$ coordinates, and we obtain $x_{i,i+1}(d_{i+1} + d_i - \partial_i) - Z_{i,i+1}d_{i+1} = f_i$. The d_i having already been determined and the coefficient of $x_{i,i+1}$ being nonzero, we can solve for each $x_{i,i+1}$. Now we proceed by induction on $s = j - i$ to solve for $x_{i,j}$.

The $(i, i+s)$ coordinate of $Y^2 - \mathbf{D}Y$ is

$$x_{i,i+s}(d_{i+s} + d_i - \partial_i) - Z_{i,i+s}d_{i+s} + (N^2 - ZN)_{i,i+s}. \tag{15.9}$$

Since both Z and N are strictly upper triangular, the coefficient of $N^2 - ZN$ is a polynomial in terms of the form $x_{l,m}$ where $m - l < s$, which have already been determined. Since the coefficient of $x_{i,i+s}$ is thus nonzero, we can solve the equation $(Y^2 - \mathbf{D}Y)_{i,i+s} = F_{i,i+s}$, that is, solve for the $x_{i,i+s}$. This completes the induction, and it continues until the very last case, when $i = 1$ and $s = n - 1$. All of the equations are reversible, that is, solving them means solving the original matrix equation, and we are done. □

In fact, the proof shows that for every sequence (d_i) of roots of the quadratics satisfying the various inequalities, there is a unique upper triangular solution to the original matrix equation.

Corollary 15.10. *Suppose that* $\mathbf{D} = \delta \mathrm{I} + Z$ *where* Z *is strictly upper triangular. If* F *is upper triangular and* $-\delta^2/4$ *appears at most once along its diagonal, then there exists an upper triangular* Y *such that* $\Psi_{\mathbf{D}}(Y) = F$.

Proof. The quadratics are now of the form $d_i^2 - \delta d_i = f_i$. If any $f_i = -\delta^2/4$, the corresponding $d_i = \delta/2$, and this is not a root of the quadratic if the corresponding f_j is not $-\delta^2/4$. In fact, distinct values of f_i give rise to disjoint sets of roots (since the sum of the roots of any one of the quadratics is δ). Partition the set $\{1,2,\ldots,n\}$ via $i \sim j$ if $f_i = f_j$. For each equivalence class, choose one of the two roots, d_i, and use the same choice for each member of the equivalence class (if one of the $f_i = -\delta^2/2$, its equivalence class consists of one element anyway). It easily follows that the inequalities hold. □

The following example shows that in Lemma 2.5, $\{C,D\}''$ cannot be replaced by $\langle C,D\rangle$, the algebra generated by C and D.

Example 15.11. A size-three example wherein $\mathcal{A} \subset \Psi_{\mathbf{D}}(\mathrm{M}_n\mathbf{C})$, but $\Psi_{\mathbf{D}}(\mathcal{A}) \neq \mathcal{A}$ ($\Psi_{\mathbf{D}}(\mathcal{A}) \subseteq \mathcal{A}$ since $\mathbf{D}$ is upper triangular). In this example, $\mathbf{D}$ is diagonal with distinct eigenvalues, and there exists an upper triangular matrix F for which $\Psi_{\mathbf{D}}(Y) = F$ has no solutions Y that are upper triangular, but there is a solution that is not upper triangular. There exist, invertible upper triangular C,D with the property that fixed points of $\phi_{C,D}$ exist, are isolated, but do not lie in $\langle C,D\rangle$.

Set

$$\mathbf{D} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 6 \end{pmatrix}, \qquad F = \begin{pmatrix} -4 & 1 & 0 \\ 0 & -16 & 1 \\ 0 & 0 & -8 \end{pmatrix}. \tag{15.10}$$

The eigenvalues have been designed so that the roots of the first two quadratics $d_i^2 - \partial_i d_i = f_i$ (recall in the notation above that $\partial_i = 4,8,6$, resp.) have unique roots, 2 and 4, respectively, and the third one has roots $\{2,4\}$. The set of inequalities cannot be satisfied by any choice of d_i, but more importantly, by brute force substitution, one can check directly that there are no upper triangular Y such that $\Psi_{\mathbf{D}}(Y) = F$. (More elegantly, but less compactly, one can also use the form of the solution in the inductive proof to show that there is no solution because of the off-diagonal ones.)

On the other hand, if we set

$$Y = \begin{pmatrix} 3 & \frac{1}{4} & -\frac{1}{16} \\ -4 & 5 & \frac{3}{4} \\ 0 & 0 & 4 \end{pmatrix}, \tag{15.11}$$

then we can verify that $\Psi_{\mathbf{D}}(Y) = F$.

To show that Y is an isolated solution, we reconstruct one of the fractional linear transformations whose fixed points satisfy the quadratic $\psi_{\mathbf{D}}(Z) = F$. That is, we find C and D (which will turn out to be upper triangular) such that if $(\mathrm{I} - CXD)X = \mathrm{I}$, then Z defined as XD will satisfy the quadratic. Premultiplying by C^{-1} and post-multiplying by D yields $(XD)^2 - C^{-1}XD = C^{-1}D$, so with $Z = XD$, we set $C = \mathbf{D}^{-1}$ and $D = -\mathbf{D}^{-1}F$;

then C is diagonal and D is upper triangular. So $X_0 := YD^{-1}$ is a fixed point of $\phi_{C,D}$. Now we convert the equation into (4.1) of Section 4, $U^2 = UB - AU$. Here $A = CX_0$ and $B^{-1} = DX_0$.

We calculate

$$X_0 = YD^{-1} = \begin{pmatrix} 1 & \frac{1}{8} & \frac{3}{256} \\ 0 & \frac{1}{2} & \frac{3}{64} \\ 0 & 0 & \frac{3}{4} \end{pmatrix}, \qquad A = \mathbf{D}^{-1}X_0 = \begin{pmatrix} \frac{1}{4} & \frac{1}{32} & \frac{3}{1024} \\ 0 & \frac{1}{16} & \frac{3}{512} \\ 0 & 0 & \frac{1}{8} \end{pmatrix},$$
$$B^{-1} = \mathbf{D}^{-1}FX_0 = \begin{pmatrix} 1 & 0 & \frac{-45}{256} \\ 0 & \frac{1}{4} & \frac{9}{32} \\ 0 & 0 & 1 \end{pmatrix}. \tag{15.12}$$

Thus A has distinct eigenvalues 2^{-2}, 2^{-3}, 2^{-4}, and B^{-1} has 2^{-2} as an eigenvalue and a Jordan block of size two for the eigenvalue 1. Hence both A and B are nonderogatory (one Jordan block per eigenvalue), hence each has only finitely many invariant subspaces. Since the spectra of A and B are disjoint, there are only finitely many solutions. In particular, all solutions to $\psi_{\mathbf{D}}(Y) = F$ are isolated. It follows easily that the corresponding fixed points of $\phi_{C,D}$ are isolated, and it is easy (now) to see that the latter cannot be upper triangular.

16. Positivity

In this section, we look for fixed points of $\phi_{C,D}$ that are positive (in the sense that all entries are nonnegative) when C and D are. For the most part, the matrices discussed here will have real entries.

A matrix $M = (M_{ij})$ is *nonnegative* if $M_{ij} \geq 0$ for all i and j. A square matrix M is *irreducible* if it is nonnegative and for all i and j, there exists $k \equiv k(i,j)$ such that $(M^k)_{ij} > 0$. The square matrix M is *primitive* if there exists k such that for all i and j, $(M^k)_{ij} > 0$. Finally, M is *strictly positive* if for all i and j, $M_{ij} > 0$. We use the notation $\mathbf{1}$ to denote the matrices all of whose entries are 1 (the dimensions will be made clear in the context). For matrices, $M \leq N$ means that $N - M$ is nonnegative, and notation $M \ll N$ means that for all i and j, $(N - M)_{ij} > 0$.

In this section, the norm on $n \times n$ matrices will always be the usual operator norm (the 2-2 norm); $\|C\| = \|C^T\| = \rho(C^T C)^{1/2}$.

A *Perron* eigenvector for a primitive or irreducible matrix is the eigenvector for the spectral radius which is strictly positive (up to scalar multiple, exactly one exists on each side, by the Perron-Frobenius theorem).

Among other things, we show that if CD is irreducible, then $\phi_{C,D}$ has at most two positive fixed points. If it has two positive fixed points, one (X_0) is attractive, and the other (X_1) is obtained from the attractive one by adding a positive multiple of vw where v and w are right and left Perron eigenvectors of DX_0 and X_0C, respectively. In particular, the nonattractive one is connected to the attractive one in the graph of fixed points. The

nonattractive one is not repulsive, but is repulsive in a more restricted sense, that is, if $Y \leq X_1$ or $Y \geq X_1$, then there exists a positive real number δ such that either not all $\phi^N(Y)$ exist, or for all sufficiently large N, $\|\phi^N(Y) - X_1\| \geq \delta$. A fixed point with this property is *positively repulsive*.

In case there is only one positive fixed point, it satisfies $\rho(DX) \cdot \rho(XC) = 1$, and thus is neither attractive nor repulsive. However, it satisfies a "flow through" property (à la Lipton$^{\text{TM}}$ tea bags), if $0 \leq Y \leq X$, then $\phi^N(Y) \to X$; however, if $X \leq Y$ and $\|Y - X\|$ is sufficiently small, then Y is repelled from X (in the sense of the previous paragraph for X_1).

It is not entirely obvious that the latter situation (a single positive fixed point) can occur; even if $C_n \to C$, $D_n \to D$ with CD invertible and each ϕ_{C_n,D_n} has an attractive fixed point, it does not follow that $\phi_{C,D}$ has a fixed point (the example with $C = \mathrm{I}$ in Section 2 and no fixed points can be constructed in this fashion). Fortunately, using positivity, we can show that under some circumstances, the limiting $\phi_{C,D}$ does have a fixed point (not generally attractive), and this yields examples with a single positive fixed point.

We note that when C and D are nonnegative, $\phi_{C,D}$ is order preserving, in the sense that if $\mathbf{0} \leq Y \leq Z$ and $\rho(CZD) < 1$, then $\mathrm{I} = \phi(\mathbf{0}) \leq \phi(Y) \leq \phi(Z)$. This follows immediately from the power series expansions of $(\mathrm{I} - CYD)^{-1}$ and $(\mathrm{I} - CZD)^{-1}$. (We cannot conclude anything if, e.g., $\rho(CZD) \geq 1$, or if $Y \leq Z$ but Y is not nonnegative.)

We recall a standard result from the theory of nonnegative matrices: if $\mathbf{0} \leq Z$ and $(\mathrm{I} - Z)^{-1}$ is nonnegative, then $\rho(Z) < 1$. Thus if C and D are nonnegative and X is a nonnegative fixed point, from $X = (\mathrm{I} - CXD)^{-1}$, we deduce that $\rho(CXD) < 1$. Since ρ is monotone on nonnegative matrices, we deduce that if $\mathbf{0} \leq Y \leq X$, then $\rho(CYD) < 1$, and thus ϕ maps the set $\{Y \mid 0 \leq Y \leq X\}$ (called the *interval generated by* X) into itself. If $\mathbf{0} \leq Z \leq Y \leq X$, then $\mathrm{I} \leq \phi(Z) \leq \phi(B) \leq \phi(X) = X$. If $Y \leq \phi(Y)$ and $\mathbf{0} \leq Y \leq X$, then $\phi^N(Y) \leq \phi^{N+1}(Y) \leq X$ for all positive integers N and it follows that $\{\phi^N(Y)\}$ converges, necessarily to a fixed point.

A particular case occurs when $Y = \mathbf{0}$. Then $\mathbf{0} \leq \mathrm{I} = \phi(\mathbf{0})$, and so (if a nonnegative fixed point X exists) $X_0 := \lim \phi^N(\mathbf{0})$ exists and is a nonnegative fixed point, and moreover, if X' is *any* nonnegative fixed point, then $X_0 \leq X'$. We notice, in particular, that $\phi^2(\mathbf{0}) = (\mathrm{I} - CD)^{-1}$ exists and is nonnegative, and thus $\rho(CD) < 1$. If we additionally impose the condition that CD (or what amounts to the same thing, DC) be irreducible, then $(\mathrm{I} - CD)^{-1}$ is strictly positive (from its power series expansion), and this entails that all nonnegative fixed points are strictly positive.

For any fixed point X (not necessarily nonnegative), we abbreviate $\rho(XC) \cdot \rho(DX)$ to $r(X)$. Thus X is attractive if and only if $r(X) < 1$.

We will show that when CD is irreducible, then $r(X_0) \leq 1$; if the inequality is strict, there is only one other fixed point, obtained from the Perron eigenvectors as above, and if the inequality is equality, there are no other fixed points (as though X_0 and X_1 merged).

Theorem 16.1. *Let C and D be nonnegative $n \times n$ matrices such that CD is irreducible. Suppose that $\phi \equiv \phi_{C,D}$ admits a nonnegative fixed point. Then $X_0 := \lim_{N \to \infty} \phi^N(\mathbf{0})$ exists*

and is a strictly positive fixed point for ϕ. Moreover, one of the following is true:

(a) *ϕ admits exactly two nonnegative fixed points, X_0 and $X := X_0 + \alpha vw$ (where α is a positive real number, and v, w are the right, left Perron eigenvectors for X_0C and DX_0, resp.), $r(X_0) < 1$, $r(X) = 1/r(X_0)$ and X_0 is attractive. Furthermore, if $\mathbf{0} \le Y \le X$ but $Y \ne X$, then $\phi^N(Y) \to X_0$. In addition, X is positively repulsive or*

(b) *ϕ admits only one nonnegative fixed point, specifically, X_0, and $r(X_0) = 1$, and if $0 \le Y \le X_0$, then $\phi^N(Y) \to X_0$. Moreover, X_0 satisfies the flow-through property.*

Remark 16.2. The statement of the theorem is so crammed with details that the proof has to be complicated. However, the ideas underlying the proof are fairly simple. First, we show that if X is a fixed point with $r(X) > 1$, for either $0 \le Y \le X$ or $Y \ge X$ with $Y \ne X$, then $\phi^N(Y)$ does not converge to X (this also includes the possibility that $\phi^N(Y)$ is not defined for some N when $Y \ge X$). This is a modest version of positive repulsiveness. Then we show that if $Y \le X$ but $Y \ne X$, then $\{\phi^N(Y)\}$ converges to a fixed point, X_{00}, and by the previous part, this fixed point must have $r(X) \le 1$. Then we show that $r(X_{00}) = 1$ implies one half of the repulsiveness property, that it cannot be a limit of the form $\phi^N(Y)$ for some $Y \ge X_{00}$. This severely reduces the possible configurations of the positive fixed points, and then we conclude with the elementary observation that if in the construction of Section 3, the eigenvectors v and w are used to create a new fixed point, X', they are still eigenvectors (with reciprocal eigenvalues) for DX' and $X'C$ (it is not true that *all* the eigenvectors carry over, only the ones involved in the construction).

Remark 16.3. Provided that n is at least two, the X (nonattractive fixed point) in part (a) is not repulsive—in fact, there always exist $Y \ne X$ such that $\phi^N(Y) \to X$; however, the positive repulsiveness asserts that any such Y cannot be comparable to X with respect to the entrywise ordering.

Proof. First, let X be a nonnegative fixed point. Then $X_0 = \lim \phi^N(\mathbf{0})$ exists (since $\mathbf{0} \le \phi(\mathbf{0}) \le \cdots \le \phi^N(\mathbf{0}) \le \cdots \le X$), is strictly positive, and $X_0 \le X$. Assume for now that $r(X) > 1$. Let v, w be the right, left Perron eigenvectors of XC and DX, respectively, obviously $wv > 0$, and thus is not zero; this means that we obtain another fixed point on the ray $\{X_z := X + zvw\}$, as in Section 3. We recall that $\phi(X_z) = X + \psi(z)vw$ where ψ is an ordinary fractional linear transformation. It is easy to see that $r(X) > 1$ implies that for z real and negative, $|\psi(z)| > |z|$, and thus for sufficiently small $|z| < t_0$, $\mathbf{0} \le \phi(X_z) \le X_z$ (this uses the fact that X is strictly positive). Hence for $|z| < t_0$ and z negative, $\{\phi^N(X_z)\}$ is a descending (entrywise) sequence of positive matrices, and thus has a limit, X', which is obviously a fixed point and equally obviously, $X' \ge X_0$. It is also easy to see that provided $|z|$ is sufficiently small, the limit is independent of the choice of negative z, and in fact is of the form $X - \alpha vw$ for some positive number α.

Now suppose that $\mathbf{0} \le Y \le X$ but $Y \ne X$. Obviously $\phi(Y) \le X$, but in fact, $\phi(Y) \ll X$. To see this, note that $\phi(Y) = (\mathrm{I} - CYD)^{-1} = \sum (CYD)^k$. We observe that CYD is primitive and $CYD \le CXD$, but equality cannot hold because C and D are invertible. As the spectral radius is strictly monotone on primitive matrices, $\rho(CYD) < \rho(CXD)$. Since $\{(CXD/\rho(CXD))^k\}$ converges to a strictly positive matrix, but $\{(CYD/\rho(CXD))^k\} \to \mathbf{0}$, we conclude that for all sufficiently large N, $(CYD)^N \ll (CXD)^N$. Thus $\phi(Y) \ll \sum (CXD)^k = (\mathrm{I} - CXD)^{-1} = X$.

Thus there exists $\epsilon > 0$ such that $\phi(Y) \leq X_{-\epsilon}$. Applying powers of ϕ, the right-hand side converges to $X' \neq X$, while all the limit points of the iterates of the left-hand side are positive, but less than or equal to X'. A conclusion is that if $0 \leq Y \leq X$ and $Y \neq X$, then there exists $\delta > 0$ such that for all sufficiently large N, $\|\phi^N(Y) - X\| \geq \delta$.

Now we work on the other side of X. Consider X_z, this time with z positive. Since $\rho(CXD) < 1$, there exists $u_0 > 0$ such that $\rho(CX_{u_0}D) < 1$, and thus for $0 \leq u \leq u_0$, it follows that $\phi(X_u)$ is positive and $\phi(X_u) = X + \psi(u)vw \gg X + uvw$ (it is easy to check that the condition $r(X) > 1$ translates to $\psi(u) > u$). If we iterate this, we find that $\phi^N(X_u)$ is defined for all N, and the distance to X increases to infinity. Now suppose that $X \leq Y \leq X_{u_0}$; then $X \leq \phi(Y) \leq \phi(X_{u_0})$, but more importantly, $\phi(Y) \gg X$ (the power series argument analogous to the preceding one can be used), whence there exists $u_1 < u_0$ such that $X_{u_1} \leq \phi(Y)$. It easily follows that $\phi^N(Y)$ is either eventually undefined or not positive or is positive but its coordinates go off to infinity. Hence $\phi^N(Y)$ does not converge to X. In particular, X is positively repulsive, and cannot be a limit of an iterated sequence $\{\phi^N(Y)\}$ with Y nonnegative and comparable to X.

A particular consequence is that the limit fixed point constructed above, X', must satisfy $r(X') \leq 1$. (This can also be deduced from the rank one construction method and the Perron-Frobenius theorem, in fact, $r(X') = 1/r(X)$.)

Now we consider what happens when there is a positive fixed point X_{00} with $r(X_{00}) = 1$—we verify the flow-through property. As before, define $X^t = X_{00} + tVW$ (where V and W are the right and left Perron eigenvectors of CX_{00} and $X_{00}D$, resp.) for real t. For real t, $t < 0$ and $r(X_{00}) = 1$ entails that $|\psi(t)| < |t|$, and thus $\phi^N(X^t) \to X_{00}$. On the other hand, for $t > 0$, $\psi(t) > t$ and thus $\phi^N(X^t)$ wanders away from X_{00}.

If $0 \leq Y \leq X_{00}$, $Y \neq X_{00}$, and $\|Y - X_{00}\|$ is sufficiently small, there exists δ such that $\mathbf{0} \leq X_{00} - \delta VW \leq Y$, and it follows that $\phi^N(Y) \to X_{00}$ (this will be improved subsequently).

On the other hand, if $X_{00} \leq Y$ and $\|Y - X_{00}\|$ is sufficiently small, the power series argument again yields an ϵ such that $Y \geq X_{00} + \epsilon VW$, and it follows that $\phi^N(Y)$ does not converge to X_{00}.

In particular, $r(X') \neq 1$, so that $r(X') < 1$, and thus X' is attractive. However, $X_0 \leq X'$ and so $r(X_0) \leq r(X') < 1$ and thus X_0 is also attractive. Since there is only one attractive fixed point, $X' = X_0$. Since X' is in the form $X - \alpha vw$, it follows (easily) that v and w are the Perron eigenvectors (on the appropriate sides) for X_0C and DX_0, and in fact, $X = X_0 + \alpha vw$. If there are any other positive fixed points, say X_1, with $r(X_1)$, we find the corresponding Perron eigenvectors, apply the construction, and end up the same matrix X_0, and thus the Perron eigenvectors are the same. It follows that X_1 is connected to X_0 (in the graph sense) by the same pair of eigenvectors, and thus $X_1 = X$. Thus far, we have that there can be at most one positive fixed point X with $r(X) > 1$ and it is of the form $X_0 + \alpha vw$.

Next, it follows from Proposition 11.1 that if there exists a fixed point X_{00} with $r(X_{00}) = 1$, then ϕ cannot have an attractive fixed point (since, in the notation of that result, $\lambda_n \cdot \mu_n = 1$). Since $X_0 \leq X_{00}$ and X_0 is primitive, strict monotonicity of the spectral radius yields that if $X_0 \neq X_{00}$, then $r(X_0) < 1$, and so an attractive fixed point would exist, a contradiction. Thus $X_{00} = X_0$ and there are no other positive fixed points (any positive fixed point with r value equalling one would have to equal X_0; any positive fixed point

whose r value exceeds one yields a positive fixed point with value less than one which is thus attractive). □

This suggests an algorithm for calculating the attractive fixed point (in the positive case), if it exists—$\{\phi^N(\mathbf{0})\}$ should converge exponentially fast if it converges to an attractive fixed point; if $r(X_0) = 1$, we would expect only polynomial convergence. In the nonpositive case, if $\phi_{C,D}$ has an attractive fixed point, then $\{\phi^N(\mathbf{0})\}$ need not even converge.

Under very restrictive conditions, if $\{\phi_k\}$ is a sequence of transformations that converges, uniformly on compact sets, to a transformation ϕ, and X_k is a fixed point of ϕ_k, then $\{X_k\}$ will converge, necessarily to a fixed point of ϕ. (In general, such convergence fails; ϕ need not have any fixed points at all.)

Lemma 16.4. *Suppose that $\{C_k\}$ and $\{D_k\}$ are invertible $n \times n$ matrices such that $C_k^{\pm 1} \to C^{\pm 1}$ and $D_k^{\pm 1} \to D^{\pm 1}$. If X_k are fixed points of $\phi_k \equiv \phi_{C_k,D_k}$ and X is a limit point of $\{X_k\}$, then X is a fixed point of $\phi_{C,D}$.*

Proof. Without loss of generality, we may assume that $X_k \to X$. As $X_k(\mathrm{I} - C_k X_k D_k) = \mathrm{I}$, so $X_k^{-1} = \mathrm{I} - C_k X_k D_k$, and thus X_k^{-1} converges, necessarily to X^{-1}, and obviously $X^{-1} = \mathrm{I} - CXD$, so X is a fixed point of ϕ. □

The condition $C_k^{\pm 1} \to C^{\pm 1}$ is simply a short way of saying $C_k \to C$ and C is invertible.

Lemma 16.5. *Suppose that $\{E_k\}$ is a set of nonnegative $n \times n$ matrices and δ is a positive real number such that $E_k \gg \delta\mathbf{1}$ for all k. If $\{\rho(E_k)\}$ is bounded above, then $\{E_k\}$ is bounded above entrywise.*

Proof. Let N_k be the maximal entry of E_k. Then $(E_k^3)_{ij} \geq \delta^2 N_k$ for all i and j. Hence $\rho(E_k^3) \geq n\delta^2 N_k$, and thus $\rho(E_k) \geq (n\delta^2 N_k)^{1/3}$. By hypothesis, $\{\rho(E_k)\}$ is bounded above, and thus so is $\{N_k\}$. Hence the conclusion. □

The condition $E_k \gg \delta\mathbf{1}$—a strong type of boundedness below—is essential; the sequence $\{\left(\begin{smallmatrix} 2 & k \\ k^{-1} & 2 \end{smallmatrix}\right)\}$ is clearly not bounded, but the spectral radius of every member is 3.

Lemma 16.6. *Suppose that $\{C_k\}$ and $\{D_k\}$ are nonnegative invertible $n \times n$ matrices with the following properties:*

(a) *$D_k C_k$ is irreducible for all k;*
(b) *$C_k^{\pm 1} \to C^{\pm 1}$ and $D_k^{\pm 1} \to D^{\pm 1}$;*
(c) *DC is irreducible.*

Suppose that X_k are attractive nonnegative fixed points of $\phi_k \equiv \phi_{C_k,D_k}$ and there exist w_k, v_k, the left and right Perron eigenvectors respectively of $D_k X_k$ and $X_k C_k$, such that $\{w_k\}$ and $\{v_k\}$ are bounded, and moreover, $\liminf w_k C v_k > 0$ *and* $\liminf w_k D v_k > 0$.

Then X_k contains a convergent subsequence which converges to a strictly positive fixed point of $\phi \equiv \phi_{C,D}$.

Proof. Without loss of generality, $w_k \to w$ and $v_k \to v$. Moreover, since $C_k \to C$, we have $wCv > 0$, and (in particular) w and v are nonzero, nonnegative vectors. Calculate

$w_k D_k X_k C_k v_k$ in two different ways:

$$\rho(D_k X_k) w_k C_k v_k = w_k D_k X_k C_k v_k = \rho(X_k C_k) w_k D_k v_k. \tag{16.1}$$

We obtain

$$\frac{\rho(D_k X_k)}{\rho(X_k C_k)} = \frac{w_k D_k v_k}{w_k C_k v_k}. \tag{16.2}$$

Taking the limit on the right (which exists and is not zero), we deduce that $\rho(D_k X_k)/\rho(X_k C_k)$ converges to a nonzero positive number. Since X_k is attractive, we have that $\rho(D_k X_k) \cdot \rho(X_k C_k) < 1$. It follows immediately that both sequences $\{\rho(D_k X_k)\}$ and $\{\rho(X_k C_k)\}$ are bounded above and below (away from zero).

Next, we observe that $X_k = \phi_k^2(X_k) \geq \phi_k^2(\mathbf{0})$, and the latter is just $(\mathrm{I} - C_k D_k)^{-1}$. Since it is nonnegative, we must have $\rho(C_k D_k) < 1$, and thus $(\mathrm{I} - C_k D_k)^{-1} = \sum (C_k D_k)^l$, which is entrywise greater than $\mathrm{I} + \sum_{l=1}^{n} (C_k D_k)^l$, and for sufficiently large k, this is greater than $\mathrm{I} + (1/2) \sum_{l=1}^{n} (CD)^l$. Since CD is irreducible, this last is strictly positive. Hence there exists $\delta > 0$ such that for all sufficiently large k, $X_k \geq \delta \mathbf{1}$. No row or column of C can consist entirely of zeros, so $X_k C_k \geq \delta' \mathbf{1}$ for some δ' for sufficiently large k. Since $\rho(X_k C_k)$ is bounded, we conclude from the previous lemma that $\{X_k C_k\}$ itself is bounded, and thus has a limit point, call it E. Multiplying by $C_k^{-1} \to C^{-1}$, we deduce that $X_k \to E C^{-1} := X$. Obviously X is strictly positive (as $X_k \geq \delta \mathbf{1}$), and it is also obviously a fixed point of ϕ. □

The difficulty in the proof of this result and its three corollaries below is that we cannot control the limit of the v_k well enough to show that it is strictly positive (except after the fact). Instead, we rely on conditions on the limiting matrices.

Corollary 16.7. *Suppose that $\{C_k\}$ and $\{D_k\}$ are nonnegative invertible $n \times n$ matrices with the following properties:*

(b) *$C_k^{\pm 1} \to C^{\pm 1}$ and $D_k^{\pm 1} \to D^{\pm 1}$;*

(c) *both C and D are strictly positive.*

Suppose that X_k are attractive nonnegative fixed points of $\phi_k \equiv \phi_{C_k, D_k}$.

Then X_k contains a convergent subsequence which converges to a strictly positive fixed point of $\phi \equiv \phi_{C,D}$.

Proof. In view of the preceding, we need only t show that the left and right Perron eigenvectors can be chosen to have the requisite properties. Select Perron eigenvectors w_k, v_k for $D_k X_k$ and $X_k C_k$, respectively, and normalize each of them so that the sum of the coefficients (of each) is 1. Then obviously $\{w_k\}$ and $\{v_k\}$ contain convergent subsequences. Without loss of generality, we may assume $w_k \to w$ and $v_k \to v$. Since the sums of the coefficients of the limits are each 1, both w and v are nonnegative vectors with coefficients adding to 1. Since C and D are strictly positive matrices, wCv and wDv both exceed zero. Thus the conditions on the eigenvectors is satisfied. Condition (a) holds for sufficiently large k (in fact, for all sufficiently large k, each of C_k and D_k is strictly positive). □

Corollary 16.8. *Suppose that $\{C_k\}$ and $\{D_k\}$ are nonnegative invertible $n \times n$ matrices with the following properties:*

(a) *$D_k = C_k^T$ and $C_k C_k^T$ is irreducible for all k;*

(b) $C_k^{\pm 1} \to C^{\pm 1}$;

(c) CC^T *is irreducible and all diagonal entries of* C *are nonzero.*

Suppose that X_k *is an attractive nonnegative fixed point of* $\phi_k \equiv \phi_{C_k, C_k^T}$.

Then $\{X_k\}$ *contains a convergent subsequence which converges to a strictly positive fixed point of* $\phi \equiv \phi_{C,C^T}$.

Proof. We verify the eigenvector condition. We first observe that each ϕ_k preserves self-adjointness (on the set of Y such that $\|C_k Y C_k^T\| < 1$). Since X_k is the limit of $\phi_k^N(\mathbf{0})$, it follows that X_k is self-adjoint (in fact, positive definite, but, i.e., another story), and therefore $(X_k C_k)^T = C_k^T X_k$. It follows that we can choose $w_k = v_k^T$. Normalize v_k so the sum of its coefficients is one, set $w_k = v_k^T$, and as in the preceding, we may assume that $v_k \to v$ and $w_k \to w = v^T$. Now $v^T C v = 0$ forces *every* term $v(i)c_{ij}v(j)$ to be zero (since all the entries of v and C are at least nonnegative); in particular, $v(i)c_{ii}v(i) = 0$. As c_{ii} are all nonzero, v must be the zero vector, a contradiction. □

For nonnegative matrices, the notation $C_k \uparrow C$ means that $C_k \to C$ and $C_k \le C_{k+1}$ with respect to the entrywise ordering.

Corollary 16.9. *Suppose that* $\{C_k\}$ *and* $\{D_k\}$ *are nonnegative* $n \times n$ *matrices such that* $C_k \uparrow C$ *and* $D_k \uparrow D$ *and* CD *is invertible and irreducible. If* X_k *are attractive nonnegative fixed points of* $\phi_k \equiv \phi_{C_k, D_k}$, *then* X_k *contains a convergent subsequence which converges upward to a strictly positive fixed point,* X, *of* $\phi \equiv \phi_{C,D}$ *such that* $r(X) \le 1$.

Proof. As CD is invertible and irreducible, the same is true for $C_k D_k$ for all sufficiently large k. It again suffices to verify the eigenvector condition. Let w_k and v_k be the left Perron and right Perron eigenvector of $X_k D_k$ and $C_k X_k$, respectively, each normalized so that the sum of its coefficients is one. By choosing the appropriate subsequence, we may assume that $v_k \to v$ and $w_k \to w$, and the limits obviously have their sums of coefficients equalling one. We may obviously assume that $X_k \ne X_{k+1}$.

If $l > k$, then $C_k X_l D_k \le C_l X_l D_l$, and since $X_l^{-1} = (\mathrm{I} - C_l X_l D_l)$ is nonnegative, $\rho(C_l X_l D_l) < 1$. Thus $\rho(C_k X_l D_k) < 1$, and now the power series expansion yields that $\phi_k(X_l) \le X_l$. Hence $\{\phi_k X_l\}$ is a descending sequence of positive matrices; its limit must have r value less than or equal one, and it easily follows that the limit is X_k. In particular, $X_k \le X_l$.

As $\rho(X_k C) \cdot \rho(D X_k) \le 1$, it follows easily that $\rho(C_k X_k)$ is bounded above, and obviously $\{C_k X_k\}$ is entrywise increasing. By Lemma 16.5, $\{C_k X_k\}$ converges upward to a fixed point, E; multiplying by $C_k^{-1} \to C^{-1}$, we obtain $X_k \to X$. Since $X_k \le X_{k+1}$, $X_k \uparrow X$. Obviously XC is strictly positive, and v is a right eigenvector whose eigenvalue is the Perron eigenvalue—hence v is a scalar multiple of the Perron eigenvector, and being nonnegative, it must be strictly positive. Similarly w is strictly positive, and it follows immediately that $vCw > 0$ and $vDw > 0$. Now Lemma 16.6 applies. □

Set $C_\alpha = \sqrt{\alpha}\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and $D_\alpha = C_\alpha^T$, and $\phi_\alpha = \phi_{C_\alpha, C_\alpha^T}$. Since $C_1 C_1^T$ has γ^2 (the square of the golden ratio) as its large eigenvalue, for $\alpha < 1/4\gamma^2$, ϕ_α has an attractive fixed point (by Proposition 14.4). By Corollary 16.9, $\phi_{1/4\gamma^2}$ has a positive fixed point that is a limit of the attractive fixed points of ϕ_α for $\alpha < 1/4\gamma^2$. It turns out (below) to be attractive!

More generally, let C be any $n \times n$ nonnegative invertible matrix such that CC^T is irreducible. For each positive real number α, let ϕ_α (obviously depending on C, but the

notation would be too cumbersome) denote the transformation $X \mapsto (I - \alpha CXC^T)^{-1}$. Define

$$\alpha_0(C) \equiv \alpha_0 := \sup\{0 < \alpha \mid \phi_\alpha \text{ has an attractive positive fixed point}\}. \tag{16.3}$$

In order to obtain the crucial result Theorem 16.12(iv) below about α_0, we require a brief elementary discussion about norms of products.

Lemma 16.10. *Let X and Y be positive semidefinite $n \times n$ matrices. Then $\|XY\| = \|X\| \cdot \|Y\|$ if and only if X and Y have a common eigenvector for their largest eigenvalues.*

Proof. If they have a common eigenvector, then the spectral radius of XY is at least as large as the product of the spectral radii of X and Y, that is, $\|X\| \cdot \|Y\|$. Hence $\|XY\| \geq \|X\| \cdot \|Y\|$, and the reverse inequality is trivial.

Suppose that $\|XY\| = \|X\| \cdot \|Y\|$. Without loss of generality, we may assume that X and Y both have norm one. There exists an element of norm one in $\mathbf{C}^n$, w, such that $\|XYw\| = \|XY\|$. Decompose $w = v + w_1$ where v is in the eigenspace of the largest eigenvalue of Y, and w_1 is orthogonal to that eigenspace. Since Y is positive semidefinite, $\|Yw_1\| < \|w_1\|$, and of course Yw_1 is still orthogonal to v. From $Yw = v + Yw_1$, we have $\|Yw\|^2 = \|v\|^2 + \|Yw_1\|^2$. If w_1 is not zero, $\|Yw\|^2 < \|v\|^2 + \|w_1\|^2 \leq \|Y\|^2$. Thus $\|XYw\| \leq \|X\|\|Yw\| < \|X\|\|Y\|$, a contradiction. Hence w_1 is zero, and so w is an eigenvector of Y for its largest eigenvalue.

Therefore, $XYw = Xv$, and $v = w$ is now known to be a unit vector. Thus $\|Xv\| = \|X\|$. The same argument as in the preceding paragraph yields that v is an eigenvector for X for the latter's largest eigenvalue. □

With normal matrices (in place of positive semidefinite), the situation is more complicated, because after renormalizing, a normal can be rewritten as a unitary direct sum with a normal contraction. More generally, a problem arises when either X or Y has more than one distinct eigenvalue with absolute value equalling the norm. Fortunately, there is a result that covers all matrices. Recall that $*$ denotes conjugate transpose.

Proposition 16.11. *Let X and Y be $n \times n$ complex matrices. Then $\|XY\| = \|X\| \cdot \|Y\|$ if and only if YY^* and X^*X have a common eigenvector for their largest eigenvalues.*

Proof. Suppose that $\|XY\| = \|X\| \cdot \|Y\|$. The spectral radius of XYY^*X^* is the same as that of YY^*X^*X, hence $\|XY\|^2 = \rho(XYY^*X^*) = \rho(YY^*X^*X)$, whence $\|YY^*X^*X\| \geq \|XY\|^2 = \|X\|^2 \cdot \|Y\|^2 = \|YY^*\| \cdot \|X^*X\|$. Now the preceding applies.

Conversely, if YY^* and X^*X have a common eigenvector for their largest eigenvalue, then $\rho(YY^*X^*X) \geq \|X\|^2 \cdot \|Y\|^2$, whence $\rho(XYY^*X^*) \geq \|X\|^2 \cdot \|Y\|^2$, so that $\|XY\|^2 \geq \|X\|^2 \cdot \|Y\|^2$. □

Theorem 16.12. *Let C be a nonnegative invertible $n \times n$ matrix such that CC^T is irreducible. The set $S(C)$ defined as*

$$S(C) = \{\alpha \in \mathbf{R}^{++} \mid \phi_\alpha \text{ has an attractive fixed point}\} \tag{16.4}$$

is a nonempty, bounded open interval. Set $\alpha_0 = \sup S(C)$. Then

(i) $\infty > \alpha_0(C) \geq 1/4\rho(CC^T)$;

(ii) *ϕ_{α_0} has a positive fixed point, X_{00}, with $r(X_{00}) = 1$;*
(iii) *if $\alpha > \alpha_0$, then ϕ_α has no positive fixed points;*
(iv) *if $\alpha_0 = 1/4\rho(CC^T)$, then C^TC and CC^T have a common right Perron eigenvector; if additionally $n = 2$, then CC^T commutes with C^TC.*

Proof. Obviously $\phi_\alpha = \phi_{\sqrt{\alpha}C,\sqrt{\alpha}C^T}$. If $\alpha < 1/4\rho(C^TC)$, then $\|\sqrt{\alpha}C\sqrt{\alpha}C^T\| < 1/4$, and so ϕ_α acts as a strict contraction on the 2-ball (see Propositions 14.1 and 14.4). In particular, $\{\phi_\alpha^N(\mathbf{0})\}$ must converge to the unique fixed point in the 2-ball, and of course, the limit of this sequence, denoted X_α, is positive, with $\|X_\alpha\| < 2$. Thus $\|\sqrt{\alpha}CX_\alpha\| < 1$, so $\rho(\sqrt{\alpha}CX_\alpha) < 1$, and obviously $\rho(X_\alpha\sqrt{\alpha}C^T) = \rho(\sqrt{\alpha}CX_\alpha)$, so that $r(X_\alpha) < 1$, and thus X_α is attractive. In particular, the open interval $(0, 1/4\rho(C^TC))$ is contained in $S(C)$.

Next, we show that if $0 < \beta < \alpha$ and ϕ_α admits a positive fixed point (not necessarily attractive), then ϕ_β admits an attractive positive fixed point. Let X_1 be a positive fixed point of ϕ_α. If $r(X_1) > 1$, then by Theorem 16.1, there exists a fixed point whose r value is less than 1; we may thus assume that $r(X_1) \le 1$. We notice that $\phi_\beta(X_1) \le X_1$ (from the power series expansion), but $\phi_\beta(X_1) \ne X_1$. Hence $\{\phi_\beta^N(\mathbf{0})\}$ is a descending sequence of positive matrices, and thus converges to a nonnegative fixed point, X', of ϕ_β, and moreover, $X' \le X_1$. From strict monotonicity of the spectral radius, $r(X') < r(X_1) \le 1$. Hence X' is an attractive nonnegative (and thus positive) fixed point of ϕ_β. This shows that $S(C)$ is a union of an increasing family of intervals, hence is an interval.

That $S(C)$ is open is a routine consequence of a much more general fact, that if ϕ has an attractive fixed point, then any transformation close enough also has an attractive fixed point—this can be expressed (in this situation) in terms of closeness of the pairs of matrices implementing the transformations.

To see that $S(C)$ is bounded (above), we recall that if (for nonnegative matrices B and D) $\phi_{B,D}$ has a positive fixed point, then $\rho(BD) < 1$, so that $\alpha_0 \le 1/\rho(CC^T)$ (see the discussion prior to Theorem 16.1). This completes the proof of the first assertion and part (i).

(ii) Corollary 16.9 yields that ϕ_{α_0} admits a positive fixed point, X, which is a limit of $\{X_\alpha\}$ (with $\alpha < \alpha_0$, each of which are attractive. Thus $\rho(\sqrt{\alpha_0}CX)$ is the limit of $\{\rho(\sqrt{\alpha}C)X_\alpha\}$). Each term in the latter sequence is less than 1, and thus $r(X) \le 1$. However, if $r(X) < 1$, then α_0 would belong to $S(C)$, contradicting the openness of the set. Hence $r(X) = 1$. We denote X by X_{00}, to be consistent with Theorem 16.1.

(iii) If $\beta > \alpha_0$ and ϕ_β has a positive fixed point, then it admits a positive fixed point X with $r(X) \le 1$. It is easy to check that $\phi_{\alpha_0}^N(X)$ converges down to X_{00}, and it follows immediately that $r(X) > 1$, a contradiction.

(iv) Denote $\sqrt{\alpha_0}C$ by C_0. Suppose that $\alpha_0 = 1/4\rho(CC^T)$. Then $\|C_0\| = 1/2$ and $\rho(C_0X_{00}) = 1$. We have

$$1 = \rho(C_0X_{00}) \le ||C_0X_{00}|| \le ||C_0|| \cdot ||X_{00}|| = \frac{||X_{00}||}{2}. \tag{16.5}$$

If the second inequality were strict, $\|X_{00}\| > 2$. However, X_{00} is the limit of $\{X_\alpha\}$ with $\alpha < \alpha_0$, and as we saw in the contraction argument, when $\alpha < 1/4\rho(CC^T)$, each $\|X_\alpha\| < 2$. Hence $\|X_{00}\| = 2$, and thus $\|C_0X_{00}\| = \|C_0\| \cdot \|X_{00}\|$. By Lemma 16.10, this forces $C_0C_0^T$ to have common right Perron eigenvector with X^2, and thus (by uniqueness) with X. Since

ρ ignores the order of multiplication (for a product of two matrices), we similarly obtain the same with the transposes; this yields that C^TC has the same right Perron eigenvector as X, and thus the same as that of CC^T.

If the matrix size is two and two self-adjoint matrices have a common eigenvector, then all their eigenvectors are in common, because the other eigenvector is the unique vector orthogonal to the original. □

Unexploited in this argument is that $\rho(C_0X_{00}) = \|C_0X_{00}\|$ also follows from the assumption that $\alpha_0 = 1/4\rho(CC^T)$. This condition yields that C_0X_{00} and its transpose has the same right (and left) Perron eigenvectors (for a general matrix A, the condition $\rho(A) = \|A\|$ implies that for A the right eigenvectors corresponding to eigenvalues less in absolute value than the spectral radius are orthogonal to the remaining eigenvectors; the condition is obviously left-right symmetric, so it applies to left eigenvectors as well). It was not clear how to express this condition in terms of C alone.

If we take $C = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, obviously C^TC and CC^T do not commute, so that $\alpha_0(C) > 1/4\gamma^2$. The argument above gives a hint at what $\alpha_0(C)$ might be (in terms of ratios of norms), but computing α_0 seems impossible at the moment. In fact, even computing (exactly) some of the fixed points for any $\alpha < \alpha_0$ (there are generically six, as follows from a brute force computation leading to a polynomial of degree six, distinct roots of which yield distinct fixed points) seems beyond hope.

17. Connections with Markov chains

The fixed point problems discussed in the rest of this paper were originally motivated by examples related to [1, Section 3]. Here is a (very) simplified account. Let Γ be a countable discrete set, partitioned as $\Gamma = \bigcup_{i\in\mathbf{N}}\Gamma_i$, where each Γ_i is an n-element set (n fixed throughout this discussion). We denote by $\mathbf{R}\Gamma$ the set of real-valued functions on Γ that are zero almost everywhere; this admits a positive cone, consisting of the nonnegative functions that are zero almost everywhere. Let B, C, and D denote nonnegative $n \times n$ matrices. We define a positive map $P : \mathbf{R}\Gamma \to \mathbf{R}\Gamma$ via the block matrix (corresponding to the partition)

$$\begin{bmatrix} B & D & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ C & B & D & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & C & B & D & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & C & B & D & \cdots \\ & & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{17.1}$$

We do not require any sort of normalization (e.g., column sums adding to one), so this is not strictly speaking a Markov chain; however, for every nonnegative left eigenvector—meaning a nonzero positive linear functional on $\mathbf{R}\Gamma$, with no constraints on entries, ν, such that $\nu P = \lambda\nu$ for some positive real λ, we can convert P into a Markov chain by conjugating in the standard way with the obvious diagonal matrix.

Under modest conditions on the B, C, and D, the nonnegative left eigenvectors of P will correspond exactly to the faithful extremal harmonic functions (not generally bounded) on the generalized Markov chain corresponding to P. By a special case of

[1, page 78 & Theorem 3.2], the nonnegative left eigenvectors of P with eigenvalue λ (if any exist) are obtainable from the nonnegative fixed points of the transformation $X \mapsto (\mathrm{I} - B/\lambda - DXC/\lambda^2)^{-1}$. With $E_\lambda = \mathrm{I} - B/\lambda$, provided $\lambda > \rho(B)$ (which it will be if the nonnegative eigenvector exists), E_λ^{-1} will exist and be nonnegative, and as in Section 2, this transformation is conjugate to $\phi_{DE_\lambda^{-1}/\lambda, CE_\lambda^{-1}/\lambda}$), and although E is not positive, there will still be a bijection between the positive fixed roots.

The infimum of the λ for which there is a nonnegative left eigenvector is called the spectral radius of P; in the case that $B = \mathbf{0}$, D is replaced by C, and C replaced by C^T, after factoring out the obvious two periodicity, the spectral radius of P is $1/\sqrt{\alpha_0(C)}$ (in the notation of the previous section). (It is unfortunate that the roles of C and D have been reversed.)

In fact, [1] does not require constant choices for B, C, and D, and there are perturbation results available, which reduce the problem of determining nonnegative left eigenvectors of P to fixed point problems, or to eventual convergence to a fixed point.

Appendices

A. Continua of fixed points

This section discusses continua of fixed points—what they look like, and some properties.

Since $\phi_{C,D}$ is (generally) only densely defined, it is not immediate that its fixed point set is closed in $\mathrm{M}_n\mathbf{C}$. However, it happens to be true.

Lemma A.1. *Suppose that $\{Z_n\}$ is a sequence of fixed points of $\phi \equiv \phi_{C,D}$, and that $\{Z_n\}$ converges to Z. Then Z is in the domain of ϕ and is a fixed point.*

Proof. We note that for all n, $Z_n^{-1} = \mathrm{I} - CZ_nD$; thus $\{Z_n^{-1}\} \to \mathrm{I} - CZD$. Thus $\mathrm{I} = \lim Z_nZ_n^{-1} = \lim Z_n(\mathrm{I} - CZD) = Z(\mathrm{I} - CZD)$. In particular, $\mathrm{I} - CZD$ is invertible, so that Z is in the domain of ϕ, and obviously it is a fixed point. □

Although the following was deduced previously—when CD is invertible—it is useful to give a more general and elementary proof.

Corollary A.2. *If Z is a limit point of fixed points of $\phi \equiv \phi_{C,D}$, then $\mathcal{M}_{ZC,DZ}$ has 1 as an eigenvalue.*

Proof. Suppose that $\lim Z_n = Z$ where the Z_n are each fixed points. Set $T_n = (Z_n - Z)/\|Z_n - Z\|$ (where $\|\cdot\|$ is any fixed algebra norm on $\mathrm{M}_n\mathbf{C}$). Then $\{T_n\}$ is a sequence of matrices of norm one, so it has a subsequence which converges to another element of norm one. By reducing to a subsequence, we may assume that $\lim T_n = T$ exists (and is not zero), and moreover that $\|Z_n - Z\| < 1/(2\|Z\| \cdot \|C\| \cdot \|D\|)$. Of course, $\mathrm{I} - CZD = Z^{-1}$ and $\mathrm{I} - CZ_nD = Z_n^{-1}$. Set $V_n = ZC(Z_n - Z)D$, so that $\|V_n\| < 1/2$, and thus $\|\sum_{i=0}^{\infty} V_n^i\| < 2$.

Now

$$\mathrm{I} - CZ_nD = \mathrm{I} - CZD - C(Z_n - Z)D = (\mathrm{I} - CZD)(\mathrm{I} - ZC(Z_n - Z)D), \tag{A.1}$$

so

$$\begin{aligned} Z_n &= \left(\mathrm{I} - CZ_nD\right)^{-1} = \left(\mathrm{I} - ZC(Z_n - Z)D\right)^{-1} \cdot Z \\ &= \left(I + V_n + V_n^2 \sum_{i=0}^{\infty} V_n^i\right) \cdot Z = Z + V_nZ + V_n^2 Y_n, \end{aligned} \tag{A.2}$$

where $\|Y_n\| < 2\|Z\|$. Thus

$$T_n = \frac{Z_n - Z}{||Z_n - Z||} = \frac{ZC(Z_n - Z)DZ}{||Z_n - Z||} + \frac{V_n^2 Y_n}{||Z_n - Z||}. \tag{A.3}$$

Since $\|V_n^2\| = \boldsymbol{O}(\|Z_n - Z\|^2)$, the rightmost summand is $\boldsymbol{O}(\|Z_n - Z\|)$, so on taking limits, we deduce that $T = ZCTDZ$. Thus T is an eigenvector of $\mathcal{M}_{ZC,DZ}$ with eigenvalue 1. □

The converse fails—we can have exactly one fixed point even though 1 is the only eigenvalue of $\mathcal{M}_{ZC,DZ}$.

If $\phi_{C,D}$ has three fixed points in an (affine) line, then the entire line consists of fixed points; another way to put this is that any one-dimensional affine subspace (i.e., translate of a subspace) containing three fixed points must be composed entirely of fixed points. For k-dimensional affine subspaces, it is easy to construct examples containing exactly 2^k fixed points (via C and D diagonal and with a mild assumption on their eigenvalues), but whether $2^k + 1$ is the critical value (i.e., if it contains $2^k + 1$ fixed points, it contains a continuum or better still, an affine line, of fixed points) is still unclear.

Lemma A.3 (Three in a row). *Suppose that C, D, X, Z are $n \times n$ matrices with $\{X, Z\}$ linearly independent. For a complex number α, define $X_\alpha = \alpha X + (1 - \alpha)Z$. Let ϕ denote $\phi_{C,D}$.*

(a) *If there exist three distinct complex values of α such that X_α is a fixed point of ϕ, then X_α is a fixed point of ϕ for all values of α.*

(b) *If for all α, X_α is a fixed point of ϕ, then $M := Z - X$ satisfies the following:*
   (i) $(MX^{-1})^2 = \mathbf{0}$,
   (ii) *M is an eigenvector of $\mathcal{M}_{XC,DX}$ for the eigenvalue* 1,
   (iii) $MCMD = \mathbf{0}$.

(c) *If X is a fixed point of ϕ and M is an eigenvector of $\mathcal{M}_{XC,DX}$ for the eigenvalue* 1 *and $(MX^{-1})^2 = \mathbf{0}$, then for all complex α, $X + (1 - \alpha)M$ is a fixed point of ϕ.*

Proof. (a) Since $\{X_\alpha\}$ is an affine line in $\mathrm{M}_n\mathbf{C}$, we can assume (by relabelling) that X_α is a fixed point for $\alpha = 0, 1$, and some other value, β. Thus X and Z are fixed points, so $X^{-1} = \mathrm{I} - CXD$ and $Z^{-1} = \mathrm{I} - CZD$. Thus

$$\begin{aligned} \left(\beta X + (1 - \beta)Z\right)^{-1} = X_\beta^{-1} &= \mathrm{I} - CX_\beta D = \beta(\mathrm{I} - CXD) + (1 - \beta)(\mathrm{I} - CZD) \\ &= \beta X^{-1} + (1 - \beta)Z^{-1}, \end{aligned} \tag{A.4}$$

and thus

$$\begin{aligned} \mathrm{I} &= (\beta X + (1-\beta)Z)(\beta X^{-1} + (1-\beta)Z^{-1}) \\ &= (\beta^2 + (1-\beta)^2)\mathrm{I} + \beta(1-\beta)(ZX^{-1} + XZ^{-1}). \end{aligned} \tag{A.5}$$

We may divide the last equation by $\beta(1-\beta)$ (which is nonzero), and thus obtain $ZX^{-1} + XZ^{-1} = 2\mathrm{I}$. Multiply this by $\gamma(1-\gamma)$; by reversing the implications, we find that $\gamma X + (1-\gamma)Z$ is a fixed point for every γ.

(b) Set $E = ZX^{-1}$; from the argument in (a), we have $E + E^{-1} = 2\mathrm{I}$. Thus $(E-\mathrm{I})^2 = \mathbf{0}$. As $Z = EX^{-1}$, it follows that $M = (E-\mathrm{I})X$, and thus $MX^{-1} = E - \mathrm{I}$, yielding (i).

Now set $N = E - \mathrm{I}$, so that $N^2 = \mathbf{0}$, $M = NX$, and $Z = (\mathrm{I}+N)X$. As $\phi(Z) = Z$, we have that $CZD = \mathrm{I} - Z^{-1}$, and thus $C(X+NX)D = \mathrm{I} - X^{-1}(\mathrm{I}-N) = \mathrm{I} - X^{-1}N - X^{-1}$. Since $CXD = \mathrm{I} - X^{-1}$, we infer $CNXD = X^{-1}N$. Pre- and post-multiplying by X, we obtain $XC \cdot NX \cdot DX = NX$, which is the conclusion of (ii).

As $CNXD = X^{-1}D = X^{-1}N$, on premultiplying by $M = NX$, we deduce $MCMD = N^2 = \mathbf{0}$ (iii).

(c) Set $Z = M + X$. It suffices to show, as in the proof of (a), that $ZX^{-1} + XZ^{-1} = 2\mathrm{I}$. Set $N = MX^{-1}$ so that $Z = (\mathrm{I}+N)X$, and thus $Z^{-1} = X^{-1}(\mathrm{I}-N)$. Hence $ZX^{-1} + XZ^{-1} = \mathrm{I} + N + \mathrm{I} - N = 2\mathrm{I}$. □

For λ a nonzero complex number, consider the one-parameter family of maps $\phi_\lambda : X \mapsto (\lambda\mathrm{I} - CXD)^{-1}$. We note that each ϕ_λ is strongly conjugate to $\phi^\lambda : X \mapsto (\mathrm{I} - \lambda^{-2}CXD)^{-1}$ via the function $\Psi_\lambda : \mathrm{M}_n\mathbf{C} \to \mathrm{M}_n\mathbf{C}$ defined by $\Psi(Z) = \lambda Z$, that is, $\Psi \circ \phi_\lambda = \phi^\lambda \circ \Psi$. Since $\phi^\lambda = \phi_{C/\lambda, D/\lambda}$, each ϕ_λ comes within the purview of this article, and we expect, for example, $\binom{2n}{n}$ fixed points generically. Assume that CD is invertible. As $\lambda \to 0$, it is clear that $\{\phi_\lambda\}$ converges to the function $\rho_{C,D} : X \mapsto (-CXD)^{-1}$. This is obviously defined only on $\mathrm{GL}(n,\mathbf{C})$. As a limiting case of the maps studied here, we expect it to have similar generic properties. It turns out that this map is rather special—it behaves like the commutative case discussed earlier.

Proposition A.4. *For CD invertible, the fixed points of $\rho_{C,D} : X \mapsto -(CXD)^{-1}$ satisfy the following.*

(a) *If $D^{-1}C$ is nonderogatory and has exactly k Jordan blocks in its normal form, then there are exactly 2^k fixed points, and they all commute with each other and $D^{-1}C$.*

(b) *If $D^{-1}C$ has a multiple eigenvector, then there is an affine line of fixed points.*

Proof. Let Z be a fixed point, so that $ZCZD = \mathrm{I}$. Set $Y = ZC$, so we obtain the equation $Y^2C^{-1}D = -\mathrm{I}$, yielding $Y^2 = -D^{-1}C$; conversely, for any Y satisfying such an equation, YC^{-1} is a fixed point.

If $D^{-1}C$ is nonderogatory, it is straightforward that all the square roots of $-D^{-1}C$ in $\mathrm{M}_n\mathbf{C}$ lie in the algebra generated by $D^{-1}C$, hence commute with each other and $D^{-1}C$. Putting $D^{-1}C$ in Jordan form and noting that it is invertible, it is readily verifiable that each block contributes exactly two square roots in the block, and thus there are 2^k square roots.

On the other hand, if $D^{-1}C$ has a multiple eigenvector, we obtain a line of fixed square roots by noticing that a piece of the Jordan form looks like $\left(\begin{smallmatrix}\lambda & 0\\ 0 & \lambda\end{smallmatrix}\right)$, and off-diagonal square roots of this abound. □

The mappings $X \mapsto -(CXD)^{-1}$ have no attractive fixed points—from $CXD = -X^{-1}$, it follows that $XCXD = -\mathrm{I}$, so that $XC = -(XD)^{-1}$, and thus $\rho((XC)^{-1}) = \rho(XD) = \rho(DX)$. However, $\mathfrak{D}\phi(X)(Y) = -(DX)^{-1}Y(CX)^{-1} = \mathcal{M}_{-(DX)^{-1},(CX)^{-1}}(Y)$. If X was an attractive fixed point, then $\rho(\mathfrak{D}\phi(X)) < 1$; however, this is just $\rho((DX)^{-1}) \cdot \rho((CX)^{-1}) = \rho((DX)^{-1}) \cdot \rho(DX)$, which is always at least one. Since the inverse of the map is $X \mapsto -(DXC)^{-1}$, that is, of the same form, neither can have any repulsive fixed points either.

B. Commuting fractional matrix transformations

Suppose that A, B, C, and D are (invertible) $n \times n$ matrices. What conclusion can we draw from $\phi_{A,B} \circ \phi_{C,D} = \phi_{C,D} \circ \phi_{A,B}$ (where we take the common domain)? If $n = 1$, this implies that $\phi_{A,B} = \phi_{C,D}$ as is easy to see from the 2×2 matrix realization of fractional linear transformations. When $n \geq 2$, some additional phenomena can occur.

Suppose that λ is a complex number unequal to 1, and $\{A, B, C, D\}$ is a set of four $n \times n$ invertible matrices. We say that the quadruple (A, B, C, D) is *λ-linked* if the following three equations hold:

$$A = \lambda CAC^{-1}, \qquad B = \lambda DBD^{-1}, \qquad AB = CD. \tag{B.1}$$

If $\lambda \neq 1$, then $\lambda^n = 1$ (take determinants); moreover, if λ is a primitive kth root of unity, there exists a change of basis so that $A = \mathrm{diag}(1, \lambda, \lambda^2, \ldots, \lambda^{k-1}) \otimes R$ and $C = P_k \otimes S$, P_k is the (permutation) matrix of the k-cycle, and R and S are $n/k \times n/k$ commuting invertible matrices. (If n is prime, then R and S are scalars.) In particular, $\mathrm{tr}\, A = \mathrm{tr}\, B = \mathrm{tr}\, C = \mathrm{tr}\, D = 0$. Consequences of the definition include the following:

$$C = \lambda A^{-1}CA, \qquad D = \lambda B^{-1}DB, \qquad BA = DC. \tag{B.2}$$

(The last follows from substituting $\lambda BD^{-1} = D^{-1}B$ into $AB = CD$.)

We require an elementary lemma. The converse is trivial.

Lemma B.1. *Suppose that $\{R, S, T, U\} \subset \mathrm{GL}(n, \mathbf{C})$. Suppose that for all X in $\mathrm{M}_n\mathbf{C}$, $RXS = TXU$. Then there exists a nonzero scalar λ such that $T = \lambda R$ and $U = \lambda^{-1}S$.*

Proof. We have $(T^{-1}R) \cdot X \cdot (SU^{-1}) = X$ for all X. Let v be a right eigenvector with eigenvalue a for $T^{-1}R$, and let w be a left eigenvector for SU^{-1} with eigenvalue b. With $X = vw$, we deduce $abvw = vw$. Hence $ab = 1$. It follows immediately that both $T^{-1}R$ and SU^{-1} have just one eigenvalue. If $T^{-1}R$ is not a scalar multiple of the identity, then there exists a column v_1 such that $T^{-1}Rv_1 = av_1 + v$. Now set $X = v_1w$. Then $v_1w = X = T^{-1}Rv_1wSU^{-1} = b(av_1 + v)w = ab(v_1w) + bvw$. As $ab = 1$, we have $vw = 0$, which is obviously impossible. So $T^{-1}R$ is a scalar multiple a of the identity, and thus $R = aT$. Similarly, $U = bS$, but $b = 1/a$, and we are done. □

Proposition B.2. *Suppose that* $\{A,B,C,D\} \subset \mathrm{GL}(n,\mathbf{C})$. *The condition,* $\phi_{A,B} \circ \phi_{C,D} = \phi_{C,D} \circ \phi_{A,B}$, *is equivalent to*

$$AB = CD, \qquad A = \lambda CAC^{-1}, \qquad B = \lambda DBD^{-1}, \tag{B.3}$$

for some scalar λ.

Proof. There exists an open neighbourhood of $\mathbf{0}$ in $\mathrm{M}_n\mathbf{C}$ such that for all X therein, both left and right sides are defined on X. For such X, we have $(\mathrm{I} - A(\mathrm{I} - CXD)^{-1}B)^{-1} = (\mathrm{I} - C(\mathrm{I} - AXB)^{-1}D)^{-1}$. Inverting and subtracting the results from I, then inverting again, we obtain

$$B^{-1}(\mathrm{I} - CXD)A^{-1} = D^{-1}(\mathrm{I} - AXB)C^{-1} \tag{B.4}$$

for all such X. Since the relation is linear in X and is equality on an open set, we deduce that it is true for X in $\mathrm{M}_n\mathbf{C}$. When X is set to $\mathbf{0}$, we obtain $(AB)^{-1} = (CD)^{-1}$, that is,

$$AB = CD. \tag{B.5}$$

We deduce that for all X,

$$B^{-1}C \cdot X \cdot DA^{-1} = D^{-1}A \cdot X \cdot BC^{-1}. \tag{B.6}$$

From Lemma B.1, there exists λ such that

$$B^{-1}C = \lambda D^{-1}A, \tag{B.7}$$

$$DA^{-1} = \lambda^{-1}BC^{-1}. \tag{B.8}$$

Multiplying (B.5) by (B.7), we obtain $AB \cdot B^{-1}C = \lambda CD \cdot D^{-1}A$, whence $AC = \lambda CA$, and thus $A = \lambda CAC^{-1}$.

Similarly, (B.8) multiplied by (B.5) gives us $DA^{-1} \cdot AB = \lambda^{-1}BC^{-1} \cdot CD$, which yields $B = \lambda DBD^{-1}$.

To show that the converse holds, it suffices to establish (B.7) and (B.8). From $AB = CD$, we deduce $CA^{-1} = DB^{-1}$; from $B = \lambda DBD^{-1}$, we have $B^{-1} = \lambda{-}1DB^{-1}D^{-1}$, so $DB^{-1} = \lambda B^{-1}D$. Therefore $C^{-1}A = \lambda B^{-1}D$, and thus $BC^{-1} = \lambda DA^{-1}$, which is (B.8).

Similarly, $AB = CD$ entails $BD^{-1} = A^{-1}C = \lambda^{-1}CA^{-1}$, so that $\lambda D^{-1}A = B^{-1}C$. □

This situation can arise with $\lambda \neq 1$—set $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) = -D$ and $B = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) = C$. With $n = 2$ and $\lambda = -1$, we have $AB = CD$, $A = \lambda CAC^{-1}$, and $B = \lambda^{-1}DBD^{-1}$. Of course, $\phi_{A,B} \neq \phi_{C,D}$.

When $\lambda = 1$, there is another characterization.

Recall that if Q is a subset of $\mathrm{M}_n\mathbf{C}$, then Q' will denote its centralizer, $\{A \in \mathrm{M}_n\mathbf{C} \mid Aq = qA \text{ for all } q \text{ in } Q\}$, Q'' its double centralizer, and so forth.

Lemma B.3. *Let* A, B, C, D *be a subset of* $\mathrm{GL}(n,\mathbf{C})$. *Then*

$$AB = CD, \qquad AC = CA, \qquad BD = DB, \tag{B.9}$$

if and only if $\phi_{C,D}$ *agrees with* $\phi_{A,B}$ *on* $\{A,B,C,D\}''$.

Proof. Suppose $AB = CD$. It suffices to show that $C^{-1}A \cdot X \cdot BD^{-1} = X$ for a relatively open set of X in $\{A,B,C,D\}''$. Set $E = C^{-1}A$; then E belongs to $\{A,C\}'$. Since $E = DB^{-1}$, E also belongs to $\{B,D\}'$. Hence E belongs to $\{A,B,C,D\}' = \{A,B,C,D\}'''$. Thus for all X in $\{A,B,C,D\}''$, $EXE^{-1} = X$, so that $C^{-1}A \cdot X \cdot BD^{-1} = X$.

Conversely, from $AXB = CXD$ for a dense set of X in $\{A,B,C,D\}''$, we have equality for all X in $\{A,B,C,D\}''$, in particular for all X in the algebra generated by $\{A,B,C,D\}$. Setting $X = \mathrm{I}$, we deduce $AB = CD$. Setting $X = (BA)^{-1}$, we have $\mathrm{I} = C(BA)^{-1}D$, whence $BA = DC$.

Setting $X = B^{-1}$, we have $A = CB^{-1}D$, whence (right multiplying by C) $AC = CB^{-1}DC$; thus $AC = CB^{-1} \cdot BA = CA$. Finally, with $X = D^{-1}$, we have $AD^{-1}B = C$, and right multiplying by D yields $AD^{-1}BD = CD = AB$, so $D^{-1}BD = B$ and thus $BD = DB$. □

In particular, if $\lambda = 1$ and $\{A,B,C,D\}$ generates all of $\mathrm{M}_n\mathbf{C}$, then $\phi_{A,B} \circ \phi_{C,D} = \phi_{C,D} \circ \phi_{A,B}$ entails that $\phi_{A,B} = \phi_{C,D}$. This corresponds to an irreducible representation of a group we are going to define.

If λ is a *primitive* nth root of unity (where n is the matrix size, as usual), there is a complete description of the λ-linking quadruples, up to conjugacy. Let P denote the cyclic permutation matrix (ones in the $(i,i+1)$ and $(n,1)$ positions). Let p be any complex polynomial which does not have any nth roots of unity (including 1) as a root, and let α and γ be nonzero complex numbers. Up to change of basis of $\mathbf{C}^n$, all possible quadruples are obtained from the formula below:

$$\begin{aligned} A &= \operatorname{diag}(1,\lambda,\lambda^2,\ldots,\lambda^{n-1})\alpha, \\ C &= P\gamma, \\ B &= C \cdot p(A^{-1}C), \\ D &= A^{-1} \cdot p(A^{-1}C)\lambda. \end{aligned} \tag{$\dagger$}$$

This follows from the simple observation that $B = A^{-1}CD$; on plugging this into $B = \lambda DBD^{-1}$, we obtain $A^{-1}CD = \lambda DA^{-1}CDD^{-1}$, so that $D \cdot A^{-1}C \cdot D^{-1} = \lambda^{-1}A^{-1}C$.

The last two remarks are part of a more general theory, namely, the irreducible finite dimensional representations of groups with appropriate generators and relations. Let G be the group with generators $\{a,b,c,d,\Lambda\}$ and relations

$$[\Lambda,a] = [\Lambda,b] = [\Lambda,c] = [\Lambda,d] = 1, \tag{B.10}$$

$$ab = cd, \tag{B.11}$$

$$a = \Lambda cac^{-1}, \tag{B.12}$$

$$b = \Lambda dbd^{-1}. \tag{B.13}$$

The first line merely says that Λ is in the centre of G. We want to record the finite dimensional irreducible representations of G. Let $\pi : G \to \mathrm{GL}(n,\mathbf{C})$ be an irreducible representation, and denote $\pi(\Lambda) = \lambda\mathrm{I}$, $\pi(a) = A$, $\pi(b) = B$, and so forth.

We require a few properties of G.

(a) $ba = dc$.

Proof. Equation (B.12) implies that $ac^{-1} = \Lambda^{-1}c^{-1}a$, so $c^{-1}a = \Lambda ac^{-1}$. From (B.11), $c^{-1}a = db^{-1}$; thus $\Lambda ac^{-1} = db^{-1}$. From (B.13), $bd^{-1} = \Lambda^{-1}d^{-1}b$, and thus $\Lambda ac^{-1} = \Lambda b^{-1}d$, that is, $ba = dc$. □

(b) The group $H := \langle c^{-1}a, \Lambda\rangle$ is a normal subgroup of G.

Proof. Set $e = c^{-1}a = db^{-1}$. Then $aea^{-1} = ac^{-1} = \Lambda^{-1}c^{-1}a = \Lambda^{-1}e$. Since e and a normalize H, so does $c = a^{-1}e^{-1}$. Similarly, $b^{-1}eb = b^{-1}d = \Lambda^{-1}db^{-1} = \Lambda^{-1}e$, and obviously d also normalizes H. □

Now we observe that G has generators $\{\Lambda, e = c^{-1}a, a, b\}$ with relations, Λ is central, and $[a,e] = [b^{-1},e] = \Lambda^{-1}$. There is an obvious normal form for elements of G, namely,

$$(\text{word in } a \text{ and } b)e^k\Lambda^l, \tag{B.14}$$

and since H is a normal subgroup, it follows that G/H is the free group on two generators (the images of a and b).

Suppose that $\lambda = 1$ (where $\pi(\Lambda) = \lambda\mathrm{I}$—since Λ is central and π is irreducible, the image of Λ is a scalar matrix). Then $E = C^{-1}A = DB^{-1}$ is the image of $\pi(e)$ and is thus in the centre of $\pi(G)$. Hence $E = \alpha\mathrm{I}$ for some nonzero scalar α. This entails that $C = \alpha A$ and $B = \alpha^{-1}D$, or what amounts to the same thing, $\phi_{A,B} = \phi_{C,D}$. In order to be irreducible, necessary and sufficient is that $\{A,B\}$ generate $\mathrm{M}_n\mathbf{C}$ as an algebra—this is equivalent to A and B having no nontrivial invariant subspace in common.

If instead $\lambda \neq 1$, then it is necessarily that an nth root of unity and (A,B,C,D) is λ-linked; moreover E^n is a scalar matrix. We have already described the irreducible representations $\pi : G \to \mathrm{GL}(n,\mathbf{C})$ such that $\pi(\Lambda) = \lambda\mathrm{I}$ where λ is a *primitive* nth root of unity. They are given by $a \mapsto A$, $b \mapsto B$, and so forth as in (†). We see that this is irreducible, since $\{A,C\}$ already generates $\mathrm{M}_n\mathbf{C}$. The parameters are given by the $\phi(n)$ possible values of λ, the primitive nth root of unity (since $n\lambda$ is the value of the trace of $\pi(\Lambda)$, different choices for λ yield inequivalent representations), α, γ, and the coefficients of p.

To decide when two such representations are equivalent, it is desirable to simplify the form of π, by using the alternative generators and relations, $\{a,b,e,\Lambda\}$ for G. Then π restricted to the subgroup generated by $\{a,e,\Lambda\}$ is irreducible (and induces a projective representation modulo the parameters α and γ of $\mathbf{Z}_n \times \mathbf{Z}_n$), and we can write it in the form

$$\begin{aligned} a \longmapsto A &:= \mathrm{diag}\,(1,\lambda,\lambda^2,\ldots,\lambda^{n-1})\,\alpha, \\ e \longmapsto E &:= P\gamma. \end{aligned} \tag{B.15}$$

Now ba commutes with e, so $\pi(b)A$ must commute with E. As E has distinct eigenvalues (E is size n, with eigenvalues $\{\gamma\exp(2\pi k/n)\}$), $\pi(b)A$ is a polynomial in E, so we have to set $\pi(b) = p(E)A^{-1}$ for some polynomial p. In order to guarantee that $p(E)$ is invertible, we are required to have that $p(\gamma\exp(2\pi ik/n)) \neq 0$ for all k—but it is easy to see that this is the only condition required of p for π to be a representation (of course, if we insist on special types of representations, such as unitary or of determinant one, more conditions will be imposed).

Let tr be the usual trace on $M_n\mathbf{C}$, and let $\mathrm{Tr} \equiv \mathrm{Tr}_\pi$ be the character associated to the representation, that is, $\mathrm{Tr}(g) = \mathrm{tr}\,\pi(g)$.

Once λ is fixed, we see that $\alpha^n = \mathrm{Tr}(a^n)$ and $\gamma^n = \mathrm{Tr}(e^n)$ are invariants of π. Now write $p(t) = \sum_{j=0}^{k-1} p_j t^j$ (we obviously can assume p has degree less than k). Then $p(E)A^{-1} = \sum p_j E^j A^{-1}$; we calculate the $\mathrm{Tr}(e^{-j}ba)$—we simply obtain np_j. Hence p itself is an invariant of the representation. Thus λ, α^n, γ^n, p are invariants of π, and it is now easy to see that these are complete (i.e., matching invariants implies equivalent representations).

What about irreducible representations of G for which $\lambda^k = \mathrm{I}$ but $k \neq 1, n$? If we do not require that, for example, $\pi(a^k)$ be scalar, then there are plenty.

C. Strong conjugacies

In this section, we discuss how far the conjugacy results of Lemma 2.1 can be extended, and what sort of mappings they apply to, by adapting the notion of elementary transformation to our context.

Define the following three types of mappings, densely defined on $M_n\mathbf{C}$; A, B, and C are $n \times n$ matrices, the latter two invertible,

$$T_A : X \longmapsto X + A, \qquad \mathcal{M}_{B,C} : X \longmapsto BXC, \qquad \mathcal{J} : X \longmapsto X^{-1}. \tag{C.1}$$

The first two types are globally defined (as are their inverses, which are also of the same types), and the third is its own inverse, with domain $\mathrm{GL}(n,\mathbf{C})$. Obviously, we have $T_A \circ T_Z = T_{A+Z}$ and $\mathcal{M}_{B,C} \circ \mathcal{M}_{Z,Y} = \mathcal{M}_{BZ,YC}$. More interesting are the following:

$$\mathcal{M}_{B,C} \circ T_A = T_{BAC} \circ \mathcal{M}_{B,C}, \qquad \mathcal{M}_{B,C} \circ \mathcal{J} = \mathcal{J} \circ \mathcal{M}_{C^{-1},B^{-1}}. \tag{C.2}$$

(In what follows, we often suppress the composition symbol $\circ$, repetitions of them stretch formulas out.) In other words, the group consisting of left and right multiplication operators is normalized by the translation and inversion operators. This permits us to describe all possible compositions of these operators. From the normalization result, every composition can be written in the form $T_1\mathcal{J}T_2\mathcal{J}\cdots\mathcal{J}T_k\mathcal{M}_{R,S}$ or $T_1\mathcal{M}_{R,S}$, where $T_i \equiv T_{A_i}$ and some invertible R and S, multiplication operators can be moved to the right and then composed with each other. In fact, we will soon see that a much shorter word in the Ts and $\mathcal{J}$ is required. In the following, the repetitions of U in the displayed formulas is not a typographical error.

Lemma C.1. *Let $\mathcal{F}$ denote the set of densely defined operators of the form*

$$\{X \longmapsto (R+SXU)(T+VXU)^{-1} \mid R,S,T,U,V \in M_n\mathbf{C},\ U,V,R-SV^{-1}T \in \mathrm{GL}(n,\mathbf{C})\}. \tag{C.3}$$

Then $\mathcal{F}$ is closed with respect to composing to the left by all operators of the form T_A and $\mathcal{M}_{B,C}$ (where BC is invertible). The set

$$\mathcal{G} := \{X \longmapsto (R+SXU)(T+VXU)^{-1} \mid R,S,T,U,V \in M_n\mathbf{C},\ U \in \mathrm{GL}(n,\mathbf{C})\} \tag{C.4}$$

is invariant under composition to the left by $\mathcal{J}$.

Proof. If ϕ is in $\mathcal{F}$, then $T_A \circ \phi(X) = (R+SXU)(T+VXU)^{-1} + A$; we rewrite this as $(R + SXU + AT + AVXU)(T + VXU)^{-1}$, which is of the same form, and the only invertibility that is not obvious is that of $R + AT - (S + AV)V^{-1}T = R - SV^{-1}T$. Similarly, $\mathcal{M}_{B,C} \circ \phi(X) = (BR + BSXU)(C^{-1}T + C^{-1}VXU)$ (and we verify that $BR - BSV^{-1}CC^{-1}T = B(R - SV^{-1}T)$ is invertible, which follows from invertibility of B). Finally, $\mathcal{J} \circ \phi(X) = (T + VXU)(R + SXU)^{-1}$. No nonobvious invertibility condition has to hold according to the definition of $\mathcal{G}$ (this is fortunate, because no such condition appears to hold). □

Lemma C.2. *Suppose that R, S, T, U, and V are in $\mathrm{M}_n\mathbf{C}$ and define $\phi : X \mapsto (R + SXU)(T + VXU)^{-1}$. If U, V, and $R - SV^{-1}T$ are invertible, then $\phi = T_B \circ \mathcal{J} \circ T_A \circ \mathcal{M}_{E,F}$ where EF is invertible.*

Proof. For all invertible Z, we have $\phi(X) = (RZ + SXUZ) \cdot (TZ + VXUZ)^{-1}$. We will solve for A, B, and invertible E, F, and Z for the factorization to hold. We calculate for any choice of the parameters

$$T_B \circ \mathcal{J} \circ T_A(X) = (A + X)^{-1} + B = (BA + \mathrm{I} + BX)(A + X)^{-1} \tag{C.5}$$

whence

$$T_B \circ \mathcal{J} \circ T_A \circ \mathcal{M}_{E,F}(X) = (BA + \mathrm{I} + BEXF)(A + EXF)^{-1}. \tag{C.6}$$

Now we show that we can solve the resulting equations,

$$BA + \mathrm{I} = RZ, \qquad S = BE, \qquad UZ = F, \qquad TZ = A, \qquad V = E. \tag{C.7}$$

Obviously, we can set $E = V$ and $B = SV^{-1}$ (the latter since V is invertible). Plugging the fourth equation into the first, we obtain $(R - SV^{-1}T)Z = \mathrm{I}$, so we set $Z = (R - SV^{-1}T)^{-1}$, and A is thus determined.

Now we check that the assignments $A = T(R - SV^{-1}T)^{-1} = RT^{-1} - SV^{-1}$, $B = SV^{-1}$, $E = V$, $F = U(R - SV^{-1}T)^{-1}$, and $Z = (R - SV^{-1}T)^{-1}$ do satisfy the equations, and moreover, E and F are invertible. □

An immediate consequence is that the ϕ in the statement of the theorem is weakly conjugate to $\mathcal{J}T_A\mathcal{M}_{E,F}T_B = \mathcal{J}T_{A+EBF}\mathcal{M}_{E,F}$, which is the denominator map, $X \mapsto (A + EBF + EXF)^{-1}$; by Lemma 2.1, this in turn will be strongly conjugate to $\phi_{C,D}$ (for some C and D) if $A + EBF$ is invertible. The last is equivalent to invertibility of $T + VSV^{-1}U$ (or $V^{-1}TU^{-1} + SV^{-1}$), but this need not to occur.

Lemma C.3. *Suppose that A, B, E, F are $n \times n$ matrices, the latter two invertible. Define $R = BA + \mathrm{I}$, $S = BE$, $T = A$, $U = F$, and $V = E$. Then $\phi : X \mapsto (R + SXU)(T + VXU)^{-1}$ equals $T_B\mathcal{J}T_A\mathcal{M}_{E,F}$, and U and V are invertible, and moreover, $R - SV^{-1}T = \mathrm{I}$.*

Proof. Direct computation. □

Lemmas C.1–C.3 together yield that every element of the group generated by $\mathcal{J}$, T_A, $\mathcal{M}_{B,C}$ (with all choices for A in $\mathrm{M}_n\mathbf{C}$ and B, C in $\mathrm{GL}(n,\mathbf{C})$) is of the form described in the second displayed formula in Lemma C.1, and strongly conjugate to one in $\mathcal{F}$.

At this point, we can discuss some of the pathological properties of weak conjugacy. These are not really pathologies, but are generic.

Example C.4. (a) Two transformations, ϕ and ψ, that are weakly conjugate yet the set of fixed points of ψ is connected and misses $\mathrm{GL}(n,\mathbf{C})$, and that of ϕ consists of 2^n matrices, is discrete, and is contained in $\mathrm{GL}(n,\mathbf{C})$.

(b) A composition of a strong conjugacy with a weak conjugacy that is not itself a weak conjugacy.

Proof. (a) Let C be in $\mathrm{GL}(n,\mathbf{C})$ such that $C^2 - C$ has distinct eigenvalues, none of which are zero or one. Set $\phi : X \mapsto X(\mathrm{I} - CX)^{-1}$. The set of fixed points of ϕ is $\{X \in \mathrm{M}_n\mathbf{C} \mid (CX)^2 = \mathbf{0}\}$; in particular, the set of fixed points is connected, but contains no invertibles.

Now conjugate ψ with the translate by the identity matrix, $T_{\mathrm{I}} : X \mapsto X + \mathrm{I}$; this defines $\phi = T_{\mathrm{I}}^{-1}\psi T_{\mathrm{I}} : X \mapsto (X + CX + C)(\mathrm{I} - C - CX)^{-1}$. Then $\phi(X) = X$ if and only if $X(\mathrm{I} - C - CX) = X + CX + C$. Simplifying and premultiplying this expression by C and setting $Z = CX$, we derive the equation $Z^2 + CZ + ZC + C = \mathbf{0}$; this yields $(Z + C)^2 = C^2 - C$.

Since $C^2 - C$ has distinct eigenvalues, it has exactly 2^n square roots in $\mathrm{M}_n\mathbf{C}$, and all of them are polynomials in $C^2 - C$, hence in C. Hence there are exactly 2^n solutions for Z, hence for X. Moreover, all of the solutions Z must be invertible—since Z is a polynomial in C, if λ is an eigenvalue of Z, it must satisfy $(\lambda + \mu)^2 = \mu^2 - \mu$ for an eigenvalue μ of C, hence $\lambda \neq 0$ ($\mu \neq 0$). Thus the 2^n fixed points of ϕ are all invertible.

(b) Pick nonzero B in $\mathrm{M}_n\mathbf{C}$, and consider $\gamma := \mathcal{J} \circ T_B : X \mapsto (X + B)^{-1}$; this is a composition of a strong conjugacy ($\mathcal{J}$) with a weak conjugacy, but ϕ is not even a weak conjugacy, since $\mathrm{GL}(n,\mathbf{C})$ is not contained in its domain. Obviously, γ is strongly conjugate to $T_B \circ \mathcal{J} : X \mapsto X^{-1} + B$, via $\mathcal{J}$. □

Despite these drawbacks, weak conjugacy (or its transitive closure) can still provide some information about the mappings, even the fixed points. For example, suppose $\phi \circ \gamma = \gamma \circ \psi$ (this notion is analogous to birational equivalence in algebraic geometry, and is finite equivalence in dynamical systems). If X is a fixed point of ψ, then there are three possibilities: (a) X is not in the domain of γ; (b) X is in the domain of γ, but $\gamma(X)$ is not in the domain of ϕ; (c) X is in the domain of γ and $\gamma(X)$ is in the domain of ϕ. Case (c) entails that $\gamma(X)$ is a fixed point of ϕ, so we have a partially defined map—possibly with empty domain—from the set of fixed points of ψ to the set of those of ϕ (and a corresponding partially defined map in the reverse direction, implemented by γ^{-1}, which we normally assume to exist as a densely defined function). Cases (a) and (b) tell us where to look for missing fixed points of ψ (if we know those of ϕ), and if $\gamma^{\pm 1}$ are defined on all of $\mathrm{M}_n\mathbf{C}$ (as is the case for the operators $T_A \circ \mathcal{M}_{B,C}$), then (a) is vacuous.

In contrast to Example C.4(a), here is a situation in which this idea is completely successful. It says that to study the fixed points of maps of the form $T_B\mathcal{J}T_A\mathcal{M}_{-C,D}$, it is sufficient to deal with the pure denominator forms, $\mathcal{J}T_{A-CBD}\mathcal{M}_{-C,D} : X \mapsto (A - CBD - CXD)^{-1}$. If $A - CBD$ is invertible, this is strongly conjugate (hence there is a natural graph isomorphism on the sets of fixed points) to a denominator form $\phi_{CE^{-1},DE^{-1}} : X \mapsto (\mathrm{I} - CE^{-1}XDE^{-1})^{-1}$ (where $E = (A - CBD)^{-1}$) by Proposition 2.2.

Proposition C.5. *Suppose $\psi = T_B \mathcal{J} T_A \mathcal{M}_{-C,D} : X \mapsto B + (A - CXD)^{-1}$ where CD is invertible. The weak conjugacy yielding $\phi := T_B^{-1} \psi T_B : X \mapsto (A - CBD - CXD)^{-1}$ induces a graph isomorphism from the graph of the fixed points of ϕ to that of ψ.*

Proof. We have $T_B \phi = \psi T_B$ and $\phi = \mathcal{J} T_A \mathcal{M}_{-C,D} T_B = \mathcal{J} T_{A-CBD} \mathcal{M}_{-C,D}$. Suppose $\phi(X) = X$; it suffices to show that $T_B(X) = X + B$ is in the domain of ψ. However, $\phi(X) = X$ entails that $X = (A - CBD - CXD)^{-1}$ which expands to $(A - C(B+X)D)^{-1}$; invertibility of $A - C(B+X)D$ entails that $\psi(B+X)$ is defined. Hence T_B is defined on the set of all fixed points of ϕ, and necessarily sends it to the set of fixed points of ψ.

Now we check that the analogous result holds for $T_B^{-1} = T_{-B}$. We have $\phi T_{-B} = T_{-B} \psi$; if $\psi(X) = X$, then $B + (A - CXD)^{-1} = X$, so $X - B = (A - CXD)^{-1}$. It suffices to show that ϕ is defined at $X - B$, that is, that $A - CBD - C(X - B)D$ is invertible. However, $A - CBD - C(X - B)D = A - CXD$, which is invertible. Thus T_{-B} induces a map from the fixed points of ψ to those of ϕ, and is clearly the inverse of the map in the previous paragraph.

Finally, $T_B(X) - T_B(Y) = X - Y$, so the rank of the differences between fixed points is preserved, so T_B preserves the graph structure (and obviously so does its inverse). □

If E, A, B are invertible, then $\psi : X \mapsto (E - AXB)^{-1}$ is strongly conjugate to $\phi_{C,D}$, where $C = AE^{-1}$ and $D = BE^{-1}$ by Proposition 2.2. If E is not invertible (but A and B remain so), is it true that ψ is not strongly conjugate to any map of the form $\phi_{C,D}$? The fixed points of ψ satisfy the same type of quadratic equations as those of $\phi_{C,D}$, which can be converted to (q) of Section 2, and by a further change of variables, can be converted to one for which the coefficients are invertible.

Acknowledgment

This work was supported in part by an operating grant from the Natural Sciences and Engineering Research Council of Canada.

References

[1] D. Handelman, "Eigenvectors and ratio limit theorems for Markov chains and their relatives," *Journal d'Analyse Mathématique*, vol. 78, pp. 61–116, 1999.
[2] J. Daughtry, "Isolated solutions of quadratic matrix equations," *Linear Algebra and Its Applications*, vol. 21, no. 1, pp. 89–94, 1978.
[3] J. E. Potter, "Matrix quadratic solutions," *SIAM Journal on Applied Mathematics*, vol. 14, no. 3, pp. 496–501, 1966.
[4] A. N. Beavers Jr. and E. D. Denman, "A new solution method for quadratic matrix equations," *Mathematical Biosciences*, vol. 20, no. 1-2, pp. 135–143, 1974.
[5] R. G. Douglas and C. Pearcy, "On a topology for invariant subspaces," *Journal of Functional Analysis*, vol. 2, no. 3, pp. 323–341, 1968.
[6] C. Godsil, "Association schemes," http://quoll.uwaterloo.ca/pstuff/assoc.
[7] D. Handelman, "A less complicated formula for Johnson graphs," in preparation.

David Handelman: Mathematics Department, University of Ottawa, Ottawa, ON, Canada K1N 6N5
Email address: dehsg@uottawa.ca

Mathematical Problems in Engineering

Special Issue on
Modeling Experimental Nonlinear Dynamics and Chaotic Scenarios

Call for Papers

Thinking about nonlinearity in engineering areas, up to the 70s, was focused on intentionally built nonlinear parts in order to improve the operational characteristics of a device or system. Keying, saturation, hysteretic phenomena, and dead zones were added to existing devices increasing their behavior diversity and precision. In this context, an intrinsic nonlinearity was treated just as a linear approximation, around equilibrium points.

Inspired on the rediscovering of the richness of nonlinear and chaotic phenomena, engineers started using analytical tools from "Qualitative Theory of Differential Equations," allowing more precise analysis and synthesis, in order to produce new vital products and services. Bifurcation theory, dynamical systems and chaos started to be part of the mandatory set of tools for design engineers.

This proposed special edition of the *Mathematical Problems in Engineering* aims to provide a picture of the importance of the bifurcation theory, relating it with nonlinear and chaotic dynamics for natural and engineered systems. Ideas of how this dynamics can be captured through precisely tailored real and numerical experiments and understanding by the combination of specific tools that associate dynamical system theory and geometric tools in a very clever, sophisticated, and at the same time simple and unique analytical environment are the subject of this issue, allowing new methods to design high-precision devices and equipment.

Authors should follow the Mathematical Problems in Engineering manuscript format described at http://www.hindawi.com/journals/mpe/. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/ according to the following timetable:

Manuscript Due	December 1, 2008
First Round of Reviews	March 1, 2009
Publication Date	June 1, 2009

Guest Editors

José Roberto Castilho Piqueira, Telecommunication and Control Engineering Department, Polytechnic School, The University of São Paulo, 05508-970 São Paulo, Brazil; piqueira@lac.usp.br

Elbert E. Neher Macau, Laboratório Associado de Matemática Aplicada e Computação (LAC), Instituto Nacional de Pesquisas Espaciais (INPE), São Josè dos Campos, 12227-010 São Paulo, Brazil ; elbert@lac.inpe.br

Celso Grebogi, Center for Applied Dynamics Research, King's College, University of Aberdeen, Aberdeen AB24 3UE, UK; grebogi@abdn.ac.uk