

An effective criterion for prescribed block-code extensions

Alec Kriebel

Independent researcher

ORCID: 0009-0001-9320-500X

16 October 2026 Version 1.0

Abstract

Given an onto block self-map of a nonempty two-sided irreducible sofic shift, we give a finite, computable criterion for extending that same map from a containing shift of finite type into the original shift. This answers the specified-map question in Boyle’s Problem 16.3, including its mixing case. The criterion is locally testable separation of the allowed language from a regular bad-word ideal. An irreducibility argument converts the published Place–van Rooijen–Zeitoun separator bound into a single canonical finite-type approximation test. The result covers every choice of extension radius and local-rule completion. It uses that published theorem; it does not assert a new general separation theorem, historical firstness, or a solution of the separate existence-over-all-maps and cellular-automaton stability questions.

1 The question and finite criterion

A shift is a closed, shift-invariant subset of $A^{\mathbb{Z}}$ for a finite alphabet A . Write $\mathcal{L}(T)$ for its finite-word language, including ε . A shift of finite type (SFT) is specified by finitely many forbidden words; a sofic shift has a finite labelled-graph presentation. Irreducibility means that any two allowed words can be joined by an allowed word. All shifts below are nonempty and two-sided.

Boyle’s Problem 16.3 [1, Section 16] asks when a *specified* surjective block code $f : T \rightarrow T$ on a mixing sofic shift has an extension $G : S \rightarrow T$ from some SFT $S \supseteq T$. Containment is non-strict, and the extension must agree with f on T . This property is called steadiness. Earlier terminology and a broader sofic-homomorphism question occur in Dykstra’s thesis [2, Section 3.6, especially Problem 3.6.8]; Ballier credits Barth–Dykstra [3] and develops substantial closing/continuing and cover-based results [4, Sections 2–4]. We retain these attributions.

Present f with memory a , anticipation b and window length $\ell = a + b + 1$. Complete its local rule arbitrarily on the ℓ -words absent from T , obtaining a block map $F : A^{\mathbb{Z}} \rightarrow A^{\mathbb{Z}}$. For a finite word w , let $F_*(w)$ be its consecutive local-rule outputs, or ε if $|w| < \ell$. Set

$$L = \mathcal{L}(T), \quad B = \{w \in A^* : F_*(w) \notin L\}.$$

The language B is regular: a finite automaton stores the most recent $\ell - 1$ input letters and feeds each new output letter into a complete automaton for L . Moreover B is a two-sided ideal. Indeed, $F_*(w)$ is a factor of $F_*(uwv)$ for a bad word w , and L is factorial. Since F extends f , $L \cap B = \emptyset$.

Choose a finite monoid M and morphism $\alpha : A^* \rightarrow M$ recognizing both L and B . Effectively one may enumerate the transition monoid of the disjoint union of their complete deterministic automata. For $n \geq 2$ define

$$T[n] = \{x \in A^{\mathbb{Z}} : \text{every length-}n \text{ block of } x \text{ belongs to } \mathcal{L}_n(T)\}.$$

This is a canonical SFT containing T .

Theorem 1. *Let T be an irreducible sofic shift and $f : T \rightarrow T$ an onto block code. With the data above, put*

$$k = 4(|M| + 1), \quad N = \max\{2, \ell, k\}.$$

The following conditions are equivalent:

1. *f extends to a block map $G : S \rightarrow T$ for some SFT $S \supseteq T$;*
2. *some locally testable language C satisfies $L \subseteq C$ and $C \cap B = \emptyset$;*
3. *$\mathcal{L}(T[N]) \cap B = \emptyset$;*
4. *$F(T[N]) = T$.*

Consequently one finite-automaton reachability test decides the extension question. On success, $T[N]$ and the chosen F are a witness; on failure no different completion, radius or containing alphabet can provide an extension.

Here a language is locally testable if membership is determined by bounded-length factors and bounded prefixes and suffixes. The precise bound in the theorem is imported from Place–van Rooijen–Zeitoun [5, Theorem 4.2]: two regular languages recognized by a common finite monoid M are LT-separable if and only if they are $\text{LT}[k]$ -separable for $k = 4(|M| + 1)$. Their $\text{LT}[k]$ equivalence records the *set* of k -profiles, with no multiplicities. A profile consists of at most $\lfloor k/2 \rfloor$ letters before a position and at most $\lceil k/2 \rceil$ letters beginning there, truncated at endpoints [5, Sections 2–4]. We use the published theorem, without reproving its general algebraic argument.

2 Proof

Normalization. Suppose $G : S \rightarrow T$ is an extension. If its alphabet is larger than A , replace S by $S \cap A^{\mathbb{Z}}$. This remains an SFT containing T . For all sufficiently large n , $T[n] \subseteq S$: choose n at least the lengths of a finite forbidden list for S . Every such forbidden word is absent from T .

Write F and G on one common finite interval. Their rules agree on every word of that interval appearing in T . Increasing n beyond its length makes $F = G$ on $T[n]$. Hence an extension exists if and only if $F(T[n]) \subseteq T$ for some $n \geq 2$. Canonical approximation normalization is also used in [4, Lemma 2.1]. For every n ,

$$F(T[n]) \subseteq T \iff \mathcal{L}(T[n]) \cap B = \emptyset. \tag{1}$$

A bad input word produces a forbidden output block; conversely any forbidden output block has a finite bad input interval. Since $T \subseteq T[n]$ and $f(T) = T$, image inclusion in (1) implies equality. This is the only use of surjectivity.

A witness gives a separator. The n -block overlap graph, with vertices $\mathcal{L}_{n-1}(T)$ and edges $\mathcal{L}_n(T)$, is strongly connected by irreducibility. Thus any word of length at least n whose n -factors are allowed extends in both directions in $T[n]$. The allowed shorter words are exactly the words of L of those lengths. Therefore $\mathcal{L}(T[n])$ is locally testable: for long words check the finite list of forbidden n -factors, and handle the finitely many short words separately. A successful canonical witness supplies condition 2.

Lemma 1 (Saturation). *If an $\text{LT}[k]$ language contains L and avoids the ideal B , then $\mathcal{L}(T[n]) \cap B = \emptyset$ for every $n \geq \max\{2, k\}$.*

Proof. Suppose a bad word w occurs in $T[n]$. Irreducibility provides a word $U \in L$, of length at least $n + k$, containing every word of $\mathcal{L}_n(T)$: concatenate that finite list using allowed connecting words, and extend if necessary. Strong connectivity of the overlap graph supplies connecting words c, d with

$$W = Ucw dU \in \mathcal{L}(T[n]).$$

Since B is an ideal, $W \in B$. Independently, irreducibility of T gives e with $V = UeU \in L$.

Both V and W contain precisely all allowed n -words. They therefore have the same set of length- k factors, namely $\mathcal{L}_k(T)$: use $k \leq n$ and extend each allowed k -word to an allowed n -word. Their prefixes and suffixes of length $k - 1$ agree because they both begin and end with U .

Their profile sets agree under the published convention. Full profiles are precisely the length- k factors split after $\lfloor k/2 \rfloor$ letters. Profiles truncated at either end are determined by the common prefixes and suffixes. There is no profile truncated at both ends, since both words have length at least $2k$. An $\text{LT}[k]$ separator containing V must consequently contain W , contradicting $W \in B$. $\square$

Apply the published LT-separation bound to condition 2 and then Lemma 1 with $n = N$ to obtain condition 3. Equation (1) and surjectivity give condition 4. Its restriction is the extension in condition 1. This proves all equivalences.

Effectivity. All constructions use finite data. Build the overlap graph of $T[N]$, label each edge by its appended letter, and take every vertex as both initial and final. Its finite-path language is exactly $\mathcal{L}(T[N])$: strong connectivity extends every path bi-infinitely, and every occurring word can be read after a preceding $(N - 1)$ -block. Product with the automaton for B reduces condition 3 to reachability of an accepting product state. No empirical word-length or radius cutoff is used. The monoid and graph can be enormous; no useful complexity improvement or implemented arbitrary-input solver is claimed.

3 Controls, scope and verification

If T is SFT, take $S = T$. Conversely an automorphism of T is steady only if T is SFT. Its inverse g is a block map, and the local identity $gF = \text{id}$ on T holds on $T[n]$ for large enough n . If $F(T[n]) \subseteq T$, then $x = gF(x) \in T$ for every $x \in T[n]$, so $T = T[n]$.

For example, the binary even shift (between successive zeros there are evenly many ones) is mixing sofic and not SFT. For any n , choose odd m with $m + 2 > n$. The periodic point $(01^m)^\infty$ belongs to $T[n]$ but not T , so its identity has no extension. Onto also matters for the image equality: a constant code to a fixed point can extend from the full shift while having only a singleton image.

The theorem answers the given-map question 16.3, with irreducibility sufficing in place of mixing. It does not decide whether some suitable self-code exists as in Problem 16.2, classify stable cellular-automaton limit sets, or solve Dykstra's broader question for arbitrary reducible sofic shifts.

The companion archive preserves all 48 files at the audited PR head, two independent mathematical approach families, two independent priority reviews with a later peer challenge, source hashes and locators, and root reproduction receipts. Its finite automata and symbolic controls test examples, saturation profiles and boundary cases; they are not a formal proof of the general theorem or its imported ingredient. The source includes an optimization-safe custody checker replacing reliance on a historical assertion-based checker. The general-purpose decision procedure is specified by the proof, rather than implemented by those example programs.

Bounded primary-source and citation searches through 16 October 2026 identified no full predecessor in the inspected statements. Originality, historical firstness and present worldwide open status are not certified. Source-access and manuscript/version gaps are documented in the archive; in particular the actual Dykstra thesis is distinct from the unread Barth–Dykstra journal body. The contribution claimed here is the complete mathematical criterion and its proof, with the cited prior machinery fully credited.

Research disclosure. AI tools were used extensively in development, source investigation, writing, computation, reproduction and adversarial review. This is an unrefereed preprint; no external human peer review or proof-assistant certification is claimed. The note and authored companion material are licensed CC BY 4.0. No third-party full source bodies are redistributed.

References

- [1] M. Boyle, *Open Problems in Symbolic Dynamics*, Contemporary Mathematics **469** (2008), 69–118. doi:10.1090/conm/469/09161. Problem 16.3, Section 16 (p. 17 of the inspected author manuscript).
- [2] A. Dykstra, *Two Equivalence Relations in Symbolic Dynamics*, Ph.D. thesis, University of Maryland, 2007, Section 3.6, pp. 46–48. Institutional repository.
- [3] J. Barth and A. Dykstra, *Weak equivalence for shifts of finite type*, Indagationes Mathematicae **18**(4) (2007), 495–506. doi:10.1016/S0019-3577(07)80058-9. Terminology attribution via [4]; journal body not inspected.
- [4] A. Ballier, *Limit sets of stable cellular automata*, Ergodic Theory and Dynamical Systems **35**(3) (2015), 673–690; online 2013. doi:10.1017/etds.2013.72. Inspected manuscript: arXiv:1301.3790v1 (2013).
- [5] T. Place, L. van Rooijen and M. Zeitoun, *On Separation by Locally Testable and Locally Threshold Testable Languages*, Logical Methods in Computer Science **10**(3:24) (2014), 1–28, Theorem 4.2. doi:10.2168/LMCS-10(3:24)2014.