

In press in the *Journal of Biological Dynamics*
Vol. 00, No. 00, Month 2016, 1–5

COMMENT

Norm Statement Considered Harmful: Comment on “Evolution of unconditional dispersal in periodic environments”

Lee Altenberg^{a*}

^a*The Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria*

(May 23, 2016)

The mathematical symbol for the norm, which is heavily overloaded with multiple definitions that have both universal and specific properties, lends itself to confusion. This is manifest in the proof of an important theorem for population dynamics by Schreiber and Li on how dispersal increases population growth in a periodic environment. Here the theorem is placed in context, the proof is clarified, and the confusing but inconsequential errors corrected.

Keywords: Reduction principle, modifier theory, evolution of dispersal, operator norm, induced norm, spectral radius, Perron root, perturbations of linear operators

AMS Subject Classification: 15A18; 15A60; 15B51; 47A55; 92D15

In a classic paper in computer science, “Go To Statement Considered Harmful”, Dijkstra [7] points out that while there is nothing *incorrect* in using `goto` in computer programs, “The go to statement as it stands is just too primitive; it is too much an invitation to make a mess of one’s program.” Ponder and Bush’s homage to Dijkstra, “Polymorphism considered harmful” [24] considers the problem of *operator overloading* — using a single symbol to represent multiple different operations — which creates a situation of *polymorphism*: “Polymorphism is a powerful and flexible programming mechanism, but like many others it presents opportunities for abuses. These abuses can ruin program understandability.”

Understandability in mathematics is similarly at risk from operator overloading. The *norm symbol*, $\|\cdot\|$, is heavily overloaded due to the large family of functions used as norms, and has the added complication that some mathematical results hold for all norms while others hold only for specific norms. The hazards of norm notation prompted Mathias to write, in *The Handbook of Linear Algebra* [22, p. 24.6]:

Warning: There is potential for considerable confusion. For example, $\|A\|_2 = \|A\|_{K,1} = \|A\|_{S,\infty}$, while $\|\cdot\|_\infty \neq \|\cdot\|_{S,\infty}$ (unless $m = 1$), and generally $\|A\|_2$, $\|A\|_{S,2}$ and $\|A\|_{K,2}$ are all different, as are $\|A\|_1$, $\|A\|_{S,1}$ and $\|A\|_{K,1}$.

This brings us to the paper by Schreiber and Li, “Evolution of unconditional dispersal in periodic environments” [26]. Schreiber and Li prove a theorem of broad interest to population biology, and their method of proof thus merits study. Norm notation in their proof, however, causes some confusion.

First, let us have some context for their theorem. Karlin [17] proved a deep theorem

*Email: Lee.Altenberg@kli.ac.at

on the asymptotic growth rate of populations that combine (1) a Markov chain with (2) heterogeneous growth rates:

Theorem 1.1 (Theorem 5.2 in [17]) Let \mathbf{P} be an irreducible stochastic matrix and \mathbf{D} be a nonnegative nonscalar matrix. Let $\mathbf{M}(\alpha) := (1 - \alpha)\mathbf{I} + \alpha\mathbf{P}$ represent a family of stochastic matrices parameterized by a mixing value $\alpha \in [0, 1]$. Then the spectral radius of $\mathbf{M}(\alpha)\mathbf{D}$ strictly decreases in α .

Karlin’s motivation for the theorem was to analyze the effect of dispersal in structured populations upon the protection of genetic diversity. In the theorem, α is the dispersal rate, and \mathbf{P} is the matrix of probabilities P_{ij} of moving from deme j to deme i given that an organism disperses. \mathbf{D} is the diagonal matrix of growth rates of the rare allele in each deme. The spectral radius $r(\mathbf{M}(\alpha)\mathbf{D})$ is the asymptotic growth rate of a rare allele in the metapopulation, and the increase in $r(\mathbf{M}(\alpha)\mathbf{D})$ as α decreases means that reducing dispersal gives rare alleles greater protection against extinction.

The generality of Karlin’s theorem allows it to be employed towards a seemingly unrelated problem — the evolution of information transmission in organisms — to prove that evolution would favor reductions in migration rates, mutation rates, recombination rates, and even rates of cultural change [1, 4], producing a unification of the “Reduction Principle” found for specific models by Feldman and coworkers (cf. [10, 11]). Karlin’s theorem and its application to the evolution of dispersal rates were independently rediscovered by Kirkland, Li, and Schreiber [18]. The theorem’s extension to linear operators on Banach spaces is provided in [3], which unifies the result that “the slower diffuser wins” found in a number of reaction diffusion models for the evolution of dispersal ([12]; [14, Lemma 5.2]; [8, Lemma 2.1]; [15]; [5]).

Empirically, it is clear that dispersal, mutation, and recombination rates have not evolved to zero in organisms, so one is directed to look for mathematical sources of departure from the reduction principle. The characterization of these conditions remains largely an open question. Just as the reduction principle appears in many context, we are seeing new situations in which departures from reduction hold, for example, for linear stochastic differential equations [9] (S. Schreiber, personal communication).

For linear systems, departure from reduction means that the spectral radius *increases* with the mixing rate α . One such departure occurs when the variation in the stochastic matrix has the general form $\mathbf{M}(\alpha) = \mathbf{B}[(1 - \alpha)\mathbf{I} + \alpha\mathbf{P}]$, where \mathbf{B} and \mathbf{P} are specially related stochastic matrices representing multiple transformation processes [2].

Schreiber and Li [26] prove a general result for another source of departure: temporally changing environments, where the growth rate matrix \mathbf{D} alternates every time step with its inverse \mathbf{D}^{-1} . They generalize, to arbitrary $n \times n$ transition matrices of reversible Markov chains, the behavior found for 2×2 matrices in [16] [19] [25] [20, Result 2], and 4×4 matrices in [6].

Theorem 1.2 ([26, Appendix 1])

Let $\mathbf{D} = \mathbf{diag}[d_1, \dots, d_n]$ with $d_1, \dots, d_n \in (0, \infty)$, and let \mathbf{S} be an $n \times n$ column stochastic matrix such that $\mathbf{R}\mathbf{S}\mathbf{R}^{-1}$ is symmetric for some diagonal matrix \mathbf{R} . For $t \in [0, 1]$, denote by $r(\mathbf{F}(t))$ the Perron (largest) eigenvalue of

$$\mathbf{F}(t) = \mathbf{D}[(1 - t)\mathbf{I} + t\mathbf{S}] \mathbf{D}^{-1}[(1 - t)\mathbf{I} + t\mathbf{S}],$$

where \mathbf{I} is the order- n identity matrix. Then $r(\mathbf{F}(t))$ is either an increasing function on $[0, 1]$ or a constant function on $[0, 1]$.

To prove this theorem, they provide another theorem in which norm notation enters:

Theorem 1.3 ([26, Appendix 2]) Denote by $\|\mathbf{A}\|$ the operator norm of the matrix \mathbf{A} . Suppose $\mathbf{A} \in \mathbf{M}_n$ is nonzero and satisfies $\|\mathbf{I} + \mathbf{A}\| \geq 1$. Then $\|\mathbf{I} + t\mathbf{A}\| \geq \|\mathbf{I} + \mathbf{A}\|$ for all $t \geq 1$.

The term “operator norm” is polymorphic in the literature. In some uses (e.g. [23]) “operator norm” is synonymous with the *spectral norm*, $\|\mathbf{A}\|_2 := r(\mathbf{A}^* \mathbf{A})^{1/2}$, where \mathbf{A}^* is the conjugate transpose of \mathbf{A} , and $r(\cdot)$ is the spectral radius.

Others use “operator norm” more generally as the norm of a matrix induced by a chosen vector norm $\|\cdot\|$ on \mathbb{C}^n (e.g. [21], [13]):

$$\|\mathbf{A}\| := \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}. \tag{1}$$

Throughout the main text of Schreiber and Li [26] and in Theorem 1.2, the vector norm used is $\|\mathbf{x}\| = \sum_{i=1}^m x_i$ for $x_i \geq 0$. However, in Theorem 1.3, used to prove Theorem 1.2, we find this inequality:

$$\|(\mathbf{I} + \mathbf{A})\mathbf{u}\| = \|(1 + \alpha)\mathbf{u} + \beta\mathbf{v}\| = |1 + \alpha|^2 + |\beta|^2 \geq 1, \tag{2}$$

where \mathbf{u} and \mathbf{v} are vectors such that $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$, $\|\mathbf{I} + \mathbf{A}\| = \|(\mathbf{I} + \mathbf{A})\mathbf{u}\|$, $\mathbf{A}\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{v}$, and $\{\mathbf{u}, \mathbf{v}\}$ is an orthonormal family.

Two things are clear from (2):

- (1) The vector norm is no longer $\|\mathbf{x}\| = \sum_{i=1}^m x_i$ but is now $\|\mathbf{x}\| := \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i x_i^*}$, so that $\|\mathbf{A}\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$.
- (2) Squares are missing, and (2) should read:

$$\|(\mathbf{I} + \mathbf{A})\mathbf{u}\|^2 = \|(1 + \alpha)\mathbf{u} + \beta\mathbf{v}\|^2 = |1 + \alpha|^2 + |\beta|^2 \geq 1. \tag{3}$$

The squares were also dropped from the next sequence of calculations, which should read

$$\|\mathbf{I} + t\mathbf{A}\|^2 \geq \|(\mathbf{I} + t\mathbf{A})\mathbf{u}\|^2 = |1 + t\alpha|^2 + |t\beta|^2 = \dots \geq \|\mathbf{I} + \mathbf{A}\|^2.$$

Fortunately, the final inequality $\|\mathbf{I} + t\mathbf{A}\|^2 \geq \|\mathbf{I} + \mathbf{A}\|^2$ remains true when the squares are dropped, so the errors are inconsequential to the proof.

Norm notation allows Theorem 1.3 to be stated quite simply, but non-transparently, and the polymorphism in the literature for the usage of “operator norm” makes us unsure of exactly which vector norm is being used (It is, however, hard to imagine a vector norm that would provide a counterexample to Theorem 1.3). The theorem could have been expressed by unpacking the spectral norm explicitly as $r((\mathbf{I} + t\mathbf{A})(\mathbf{I} + t\mathbf{A})^*) \geq r((\mathbf{I} + \mathbf{A})(\mathbf{I} + \mathbf{A})^*)$ for all $t \geq 1$, which leaves no room for ambiguity or error, and segues directly into Schreiber and Li’s sagacious decomposition (p. 134), $r(\mathbf{F}(t)) = r(\mathbf{D}^{-1/2}\mathbf{F}(t)\mathbf{D}^{1/2}) = r(\mathbf{B}(t)\mathbf{B}(t)^\top)$.

Norms are of course fundamental to mathematical analysis. The “information hiding” achieved by $\|\cdot\|$ makes norm notation concise, but both the reader — and the writer — may benefit from always explicitly stating its content.

Acknowledgements

This work was supported by the Konrad Lorenz Institute for Evolution and Cognition Research, Klosterneuburg, Austria, and the Mathematical Biosciences Institute at The Ohio State University, USA, through National Science Foundation Award #DMS 0931642.

References

- [1] L. Altenberg, *A Generalization of Theory on the Evolution of Modifier Genes*, ProQuest, 1984 October, <http://search.proquest.com/docview/303425436/abstractProQuest> document ID: 303425436.
- [2] L. Altenberg, *The evolution of dispersal in random environments and the principle of partial control*, Ecological Monographs 82 (2012), pp. 297–333.
- [3] L. Altenberg, *Resolvent positive linear operators exhibit the reduction phenomenon*, Proceedings of the National Academy of Sciences U.S.A. 109 (2012), pp. 3705–3710.
- [4] L. Altenberg and M.W. Feldman, *Selection, generalized transmission, and the evolution of modifier genes. I. The reduction principle*, Genetics 117 (1987), pp. 559–572.
- [5] R. Cantrell, C. Cosner, and Y. Lou, *Evolution of dispersal in heterogeneous landscapes*, in *Spatial Ecology*, R. Cantrell, C. Cosner, and S. Ruan, eds., Mathematical and Computational Biology, Chapman & Hall/CRC Press, London, 2010, pp. 213–229.
- [6] O. Carja, U. Liberman, and M.W. Feldman, *Evolution in changing environments: Modifiers of mutation, recombination, and migration*, Proceedings of the National Academy of Sciences 111 (2014), pp. 17935–17940.
- [7] E.W. Dijkstra, *Go to statement considered harmful*, Communications of the ACM 11 (1968), pp. 147–148.
- [8] J. Dockery, V. Hutson, K. Mischaikow, and M. Pernarowski, *The evolution of slow dispersal rates: A reaction diffusion model*, Journal of Mathematical Biology 37 (1998), pp. 61–83.
- [9] S.N. Evans, P.L. Ralph, S.J. Schreiber, and A. Sen, *Stochastic population growth in spatially heterogeneous environments*, Journal of Mathematical Biology 66 (2013), pp. 423–476.
- [10] M.W. Feldman, *Selection for linkage modification: I. Random mating populations*, Theoretical Population Biology 3 (1972), pp. 324–346.
- [11] M.W. Feldman, F.B. Christiansen, and L.D. Brooks, *Evolution of recombination in a constant environment*, Proceedings of the National Academy of Sciences U.S.A. 77 (1980), pp. 4838–4841.
- [12] A. Hastings, *Can spatial variation alone lead to selection for dispersal?*, Theoretical Population Biology 24 (1983), pp. 244–251.
- [13] R.A. Horn and C.R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge, 2013.
- [14] V. Hutson, J. López-Gómez, K. Mischaikow, and G.T. Vickers, *Limit behavior for a competing species problem with diffusion*, in *Dynamical Systems and Applications*, R.P. Agarwal, ed., World Scientific, Singapore, 1995, pp. 343–358.
- [15] V. Hutson, S. Martinez, K. Mischaikow, and G.T. Vickers, *The evolution of dispersal*, Journal of Mathematical Biology 47 (2003), pp. 483–517.
- [16] K. Ishii, H. Matsuda, Y. Iwasa, and A. Sasaki, *Evolutionarily stable mutation rate in a periodically changing environment*, Genetics 121 (1989), pp. 163–174.
- [17] S. Karlin, *Classifications of selection–migration structures and conditions for a protected polymorphism*, in *Evolutionary Biology*, M.K. Hecht, B. Wallace, and G.T. Prance, eds., Vol. 14, Plenum Publishing Corporation, New York, 1982, pp. 61–204.
- [18] S. Kirkland, C.-K. Li, and S.J. Schreiber, *On the evolution of dispersal in patchy landscapes*, SIAM Journal on Applied Mathematics 66 (2006), pp. 1366–1382.
- [19] M. Lachmann and E. Jablonka, *The inheritance of phenotypes: an adaptation to fluctuating environments*, Journal of Theoretical Biology 181 (1996), pp. 1–9.

- [20] U. Liberman, J. Van Cleve, and M.W. Feldman, *On the evolution of mutation in changing environments: recombination and phenotypic switching*, *Genetics* 187 (2011), pp. 837–851.
- [21] P.L. Loh, M.J. Wainwright, *et al.*, *High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity*, *The Annals of Statistics* 40 (2012), pp. 1637–1664.
- [22] R. Mathias, *Singular values and singular value inequalities*, in *Handbook of Linear Algebra*, L. Hogben, ed., 2nd ed., chap. 24, Chapman and Hall, 2014, pp. 24-1–24-17.
- [23] L. Molnár and P. Szokol, *Transformations on positive definite matrices preserving generalized distance measures*, *Linear Algebra and Its Applications* 466 (2015), pp. 141–159.
- [24] C. Ponder and B. Bush, *Polymorphism considered harmful*, *ACM SIGSOFT Software Engineering Notes* 19 (1994), pp. 35–37.
- [25] M. Salathé, J. Van Cleve, and M.W. Feldman, *Evolution of stochastic switching rates in asymmetric fitness landscapes*, *Genetics* 182 (2009), pp. 1159–1164.
- [26] S.J. Schreiber and C.-K. Li, *Evolution of unconditional dispersal in periodic environments*, *Journal of Biological Dynamics* 5 (2011), pp. 120–134.