Computing Tensor Z-Eigenvectors with Dynamical Systems

Austin R. Benson\textsuperscript{\dagger} and David F. Gleich\textsuperscript{\ddagger}

Abstract. We present a new framework for computing Z-eigenvectors of general tensors based on numerically integrating a dynamical system that can only converge to a Z-eigenvector. Our motivation comes from our recent research on spacey random walks, where the long-term dynamics of a stochastic process are governed by a dynamical system that must converge to a Z-eigenvector of a transition probability tensor. Here, we apply the ideas more broadly to general tensors and find that our method can compute Z-eigenvectors that algebraic methods like the higher-order power method cannot compute.

Key words. tensors, eigenvectors, dynamical systems

AMS subject classifications. 15A69, 65F15, 65P99

DOI. 10.1137/18M1229584

1. Preliminaries on tensor eigenvectors. Computing matrix eigenvalues is a classic problem in numerical linear algebra and scientific computing. Given a square matrix $A$, the goal is to find a vector-scalar pair $(x, \lambda)$ with $x \neq 0$ satisfying
\begin{equation}
Ax = \lambda x.
\end{equation}
The pair $(x, \lambda)$ is called the eigenpair, $x$ the eigenvector, and $\lambda$ the eigenvalue. After several decades of research and development, we have, by and large, reliable methods and software for computing all eigenpairs of a given matrix $A$. (Experts will, of course, be aware of exceptions, but we hope they would agree with the general sentiment of the statement.)

In numerical multilinear algebra, there are analogous eigenvector problems (note the plurality). For example, given a three-mode cubic tensor $T$ (here meaning that $T$ is a multidimensional $n \times n \times n$ array with entries $T_{i,j,k}$, $1 \leq i, j, k \leq n^3$), the two most common tensor eigenvector problems are the following:

- Z-eigenvectors (Qi, 2005) \[ \sum_{j,k} T_{i,j,k} x_j x_k = \lambda x_i, \quad 1 \leq i \leq n \]
- $l^2$-eigenvectors (Lim, 2005) \[ \|x\|_2 = 1 \]
- $l^k$-eigenvectors (Lim, 2005) \[ \sum_{j,k} T_{i,j,k} x_j x_k = \lambda x_i^k, \quad 1 \leq i \leq n \]
- H-eigenvectors (Qi, 2005) \[ \sum_{j,k} T_{i,j,k} x_j x_k = \lambda x_i, \quad 1 \leq i \leq n \]

We use the “Z” and “H” terminology instead of “$l^2$” and “$l^k$.” Both Z- and H-eigenvectors are defined for tensors with the dimension equal in all modes (such a...
tensor is called cubic (Comon et al., 2008). The definitions can be derived by showing that the eigenpairs are KKT points for a generalization of a Rayleigh quotient to tensors (Lim, 2005). One key difference between the types is that $H$-eigenvectors are scale-invariant, while $Z$-eigenvectors are not—this is why we put a norm constraint on the vector. Specifically, if we ignore the norm constraint and scale $\mathbf{x}$ by a constant, the corresponding $Z$-eigenvalue would change; for $H$-eigenpairs, this is not the case.

If $\mathbf{T}$ is symmetric, then it has a finite set of $Z$-eigenvalues, and, moreover, there must be a real eigenpair when the order of the tensor (i.e., the number of modes or indices) is odd (Cartwright and Sturmfels, 2013).

This paper presents a new framework for computing $Z$-eigenpairs. Tensor $Z$-eigenvectors show up in a variety of applications, including evolutionary biology (Bini, Meini, and Poloni, 2011; Meini and Poloni, 2011), low-rank factorizations and compression (De Lathauwer, De Moor, and Vandewalle, 2000; Kofidis and Regalia, 2002; Anandkumar et al., 2014), signal processing (De Lathauwer, 1997; Kofidis and Regalia, 2001), quantum geometry (Wei and Goldbart, 2003; Hu, Qi, and Zhang, 2016), medical imaging (Qi, Wang, and Wu, 2008), and data mining (Benson, Gleich, and Leskovec, 2015; Gleich, Lim, and Yu, 2015; Wu, Benson, and Gleich, 2016; Benson, 2019). All real eigenpairs can be computed with a Lassere-type semidefinite programming hierarchy (Cui, Dai, and Nie, 2014; Nie and Wang, 2014; Nie and Zhang, 2017), but the problem of computing them remains NP-hard (Hillar and Lim, 2013), and the scalability of such methods is limited (see subsection 3.3 for experiments). Thus, we still lack robust and scalable general-purpose methods for computing these eigenvectors.

We introduce two special cases of tensor contractions that will be useful:

1. The tensor apply takes a cubic tensor and a vector and produces a vector, akin to Qi’s notation (Qi, 2005):

   - three-mode tensor $\mathbf{y} = \mathbf{T} \mathbf{x}^2$
   - $m$-mode tensor $\mathbf{y} = \mathbf{T} \mathbf{x}^{m-1}$

   $y_i = \sum_{j,k} T_{i,j,k} x_j x_k$

2. The tensor collapse takes a cubic tensor and a vector and produces a matrix:

   - three-mode tensor $\mathbf{Y} = \mathbf{T}[\mathbf{x}]$
   - $m$-mode tensor $\mathbf{Y} = \mathbf{T}[\mathbf{x}]^{m-2}$

   $Y_{ij} = \sum_k T_{i,j,k} x_k$

For the tensor collapse operator, the “:)” symbol signifies taking all entries along that index, so $\mathbf{T}_{::,k}$ is a square matrix. The tensor may not be symmetric, but we are always contracting onto the first mode (tensor apply) or first and second modes (tensor collapse); we assume that $\mathbf{T}$ has been permuted in the appropriate manner for the problem at hand. With this notation, the $Z$-eigenvector problem can be written as

$$\mathbf{T} \mathbf{x}^{m-1} = \lambda \mathbf{x}, \quad \|\mathbf{x}\|_2 = 1.$$  

The crux of our computational method is based on the following observation that relates tensor and matrix eigenvectors.

**Observation 1.1.** A tensor $Z$-eigenvector $\mathbf{x}$ of an $m$-mode tensor must be a matrix eigenvector of the collapsed matrix $\mathbf{T}[\mathbf{x}]^{m-2}$, i.e.,

$$\mathbf{T} \mathbf{x}^{m-1} = \lambda \mathbf{x} \iff \mathbf{T}[\mathbf{x}]^{m-2} \mathbf{x} = \lambda \mathbf{x}.$$
The catch, of course, is that the matrix itself depends on the tensor eigenvector we want to compute, which we do not know beforehand. Therefore, we still have a nonlinear problem.

2. A dynamical system’s framework for computing $Z$-eigenvectors. Observation 1.1 provides a new perspective on the tensor $Z$-eigenvector problem. Specifically, tensor $Z$-eigenvectors are matrix eigenvectors, just for some unknown matrix. Our computational approach is based on the following continuous-time dynamical system, for reasons that we will make clear in subsection 2.2:

$$\frac{dx}{dt} = \Lambda(T[x]^{m-2}) - x.$$  

(2.1)

Here, $\Lambda$ is some fixed map that takes as input a matrix and produces as output some prescribed eigenvector of the matrix with unit norm. For example, on an input $M$, $\Lambda$ could be defined to compute several objects:

1. the eigenvector of $M$ with $k$th smallest/largest magnitude eigenvalue;
2. the eigenvector of $M$ with $k$th smallest/largest algebraic eigenvalue;
3. the eigenvector of $M$ closest in distance to a given vector $v$.

We resolve the ambiguity in the sign of the eigenvector by picking the sign based on the first element. In the case of multiple eigenvectors sharing an eigenvalue, we propose using the closest eigenvector to $x$, although we have not evaluated this technique.

**Proposition 2.1.** Let $\Lambda$ be a prescribed map from a matrix to one of its eigenvectors. Then, if the dynamical system in (2.1) converges to a nonzero solution, it must converge to a tensor $Z$-eigenvector.

**Proof.** If the dynamical system converges, then it converges to a stationary point. Any stationary point has zero derivative, so

$$\frac{dx}{dt} = 0 \iff \Lambda(T[x]^{m-2}) = x \iff T[x]^{m-2}x = \lambda x \text{ for some } \lambda \text{ that depends on } \Lambda \iff Tx^{m-1} = \lambda x.$$

One must be a bit careful with the input and output values of $\Lambda$. If $T$ is not symmetric, then $T[x]^{m-2}$ might not be diagonalizable, and we may have to deal with complex eigenvalues. To keep the dynamical system real-valued, one could always modify the map $\Lambda$ to output the real part. However, the tensor need not be symmetric (not $T[x]^{m-2}$ normal for all $x$) for the dynamical system to maintain real values. In fact, our motivation for this dynamical system comes from a tensor that is not necessarily symmetric, which we will discuss in subsection 2.2.

**Proposition 2.1** leads to a broad framework for computing $Z$-eigenvectors:

1. choose a map $\Lambda$ and a numerical integration scheme;
2. numerically integrate (2.1).

Different choices of $\Lambda$ may converge to different $Z$-eigenvectors, and different numerical integration schemes may lead to different convergence properties. Figure 1 shows a concrete example, where $\Lambda$ picks the eigenvector corresponding to eigenvalue with largest magnitude real part, along with the forward Euler numerical integration scheme.

The dynamical system in (2.1) has no dependence on the time $t$. Thus, the system might be a good candidate for an explicit solution; however, this would require integrating the map $\Lambda$, for which an explicit solution is unclear in general. Thus, we focus on numerical integration.
using LinearAlgebra

function tensor_apply(T::Array{Float64,3}, x::Vector{Float64})
    n = length(x)
    y = zeros(Float64, n)
    for k in 1:n; y += T[:, :, k] * x * x[k]; end
    return y
end

function tensor_collapse(T::Array{Float64,3}, x::Vector{Float64})
    n = length(x)
    Y = zeros(Float64, n, n)
    for k in 1:n; Y += T[:, :, k] * x[k]; end
    return Y
end

function dynsys_forw_eul(T::Array{Float64,3}, h::Float64, niter::Int64)
    function dx_dt(u::Vector{Float64})  # Derivative
        F = eigen(tensor_collapse(T, u))
        ind = sortperm(abs.(real(F.values)))[1]
        v = F.vectors[:, ind]
        return sign(v[1]) * v - u  # sign consistency
    end
    x = normalize(ones(Float64, size(T, 1)), 1)  # starting point
    eval_hist = [x' * tensor_apply(T, x)]
    for _ = 1:niter
        x += h * dx_dt(x)  # forward Euler
        push!(eval_hist, x' * tensor_apply(T, x))  # Rayleigh quotient
    end
    return (x, eval_hist)  # guess at evec and history of evals
end

Fig. 1. Julia implementation of the dynamical system for a 3-mode tensor with a map $\Lambda$ that picks the largest magnitude real eigenvalue and numerical integration with the forward Euler method. Code snippet is available at https://gist.github.com/arbenson/f28d1b2de9aa72882735e1be24d05a7f. A more expansive code is available at https://github.com/arbenson/TZE-dynsys.

2.1. Forward Euler and diagonal tensors. As an illustrative example, we consider the special case of using the forward Euler numerical integration scheme for computing the tensor eigenvalues of an $n$-dimensional, $m$-mode diagonal tensor $T$. Without loss of generality, assume that the diagonal entries of $T$ are decreasing in order so that $T_{i,i} < T_{j,j}$ if $i > j$. This tensor has at least $n Z$-eigenpairs: $(e_i, T_{i,i})$ for $1 \leq i \leq n$, where $e_i$ is the $i$th standard basis vector. Suppose that we want to compute the $i$th eigenvector and set $\Lambda$ to select the unit-norm eigenvector closest to $e_i$ in angle. Since $T[x]^{m-2}$ is diagonal, its eigenvectors are the standard basis vectors, and $\Lambda(T[x]^{m-2}) = e_i$. Let $r_k = x_k - e_i$ be the residual at the $k$th iteration. If the step size is $h$, then

$$
\|r_{k+1}\| = \|x_{k+1} - e_i\| = \|x_k + h(e_i - x_k) - e_i\| = (1 - h)\|x_k - e_i\| = (1 - h)^k \|r_0\|.
$$

Thus, the forward Euler scheme converges if $h \leq 1$ and converges in one step if $h = 1$. Figure 2 (left) illustrates the dynamics for an example tensor $T$ ($n = 3, m = 3$) with $T_{1,1,1} = 5, T_{2,2,2} = 2$, and $T_{1,1,1} = 1$. In this case, the entire surface of the three-dimensional sphere is a basin of attraction for this eigenvector, which is consistent with our convergence analysis. In fact, for this specific case, we have the closed-form solution

$$
x(t) = e^{-t}[x(0) - e_i] + e_i,
$$

and we have exponential convergence to a solution, consistent with Figure 2 (left).
However, our analysis relies on the particular choice of $\Lambda$. Suppose instead that we choose $\Lambda$ to select the eigenvector corresponding to the smallest algebraic eigenvalue, and we are trying to compute the eigenvector $e$ of a $3 \times 3 \times 3$ diagonal tensor $T$ with strictly decreasing diagonal entries. Moreover, suppose we have a starting iterate $x_0 = [\varepsilon/2 \quad \varepsilon/2 \quad 1 - \varepsilon]^T$, which is close to the $Z$-eigenvector $e_3$. Then

$$\Lambda(T[x_0]) = \Lambda \left( \begin{bmatrix} \frac{\varepsilon}{2} T_{1,1,1} & 0 & 0 \\ 0 & \frac{\varepsilon}{2} T_{2,2,2} & 0 \\ 0 & 0 & (1 - \varepsilon) T_{3,3,3} \end{bmatrix} \right) = e_2$$

if $\varepsilon$ is sufficiently small. Forward Euler integration with step size $h$ gives the next iterate $x_1 = x_0 + h(\Lambda(T[x_0]) - x_0) = [(1 - h)\varepsilon \quad (1 - h)\varepsilon + h \quad 1 - \varepsilon - h]^T$, which is further away from the $Z$-eigenvector $e_3$ than $x_0$. Thus, there is no basin of attraction for the $Z$-eigenvector $e_3$ with this particular choice of map $\Lambda$.

It turns out that this is a case where the dynamical system does not converge. The system is ill-defined for some points, and, moreover, these points are attractors (specifically, the eigenvector $(x_1, x_2, x_3) \approx (0.18, 0.44, 0.88)$ in the example above; see Figure 2, right). In general, for some time, the dynamical system will evolve in the direction of $e_i$, where $T_{i,i,i} x_i < \min_j T_{j,j,j} x_j$ for $i \neq j$, $i,j \in \{1, 2, 3\}$. Along this direction, the $i$th coordinate of the vector increases until $T_{i,i,i} x_i = T_{j,j,j} x_j$ for some $j \neq i$. At this point, the map is ill-defined since the eigenspace corresponding to the smallest eigenvalue of $T[x]$ has dimension at least two. Since the diagonal tensor entries are distinct by assumption, this is not a fixed point. We can disambiguate the map at these ambiguous points. However, any way of doing so besides artificially mapping the vector to a $Z$-eigenvector of $T$ would result in immediate attraction back to one of these ambiguous points.

2.2. Spacey random walks motivation for the dynamical system. The motivation for the dynamical system comes from our previous analysis of a stochas-
tic process known as the “spacey random walk” that relates tensor eigenvectors of a particular class of tensors to a stochastic process (Benson, Gleich, and Lim, 2017). Specifically, the class of tensors are irreducible transition probability tensors (any irreducible tensor $\mathbf{P}$ with $\sum_{i_1=1}^{n} P_{i_1,i_2,...,i_m} = 1$ for $1 \leq i_2,...,i_m \leq n$). For simplicity, we discuss a three-mode transition probability tensor $\mathbf{P}$, where the entries can be interpreted as coming from a second-order Markov chain—the entry $P_{i,j,k}$ is the probability of transitioning to state $i$ given that the last two states were $j$ and $k$. Due to the theory of Li and Ng (2014), there exists a tensor $Z$-eigenvector $\mathbf{x}$ with eigenvalue 1 satisfying

\begin{equation}
\mathbf{Px}^2 = \mathbf{x}, \quad \sum_{i=1}^{n} x_i = 1, \quad x_i \geq 0.
\end{equation}

The vector $\mathbf{x}$ is stochastic, but it does not represent the stationary distribution of a Markov chain. Instead, we showed that $\mathbf{x}$ is the limiting distribution of a non-Markovian, generalized vertex-reinforced random walk (Benaïm, 1997) that we called the spacey random walk (Benson, Gleich, and Lim, 2017). In the $n$th step of a spacey random walk, after the process has visited states $X_1,...,X_n$, it spaces out and forgets its second last state (that is, the state $X_{n-1}$). It then invents a new history state $Y_n$ by randomly drawing a past state $X_1,...,X_n$. Finally, it transitions to $X_{n+1}$ via the second-order Markov chain represented by $\mathbf{P}$ as if its last two states were $X_n$ and $Y_n$; i.e., it transitions to $X_{n+1}$ with probability $P_{X_{n+1},X_n,Y_n}$. (In contrast, a true second-order Markov chain would transition with probability $P_{X_{n+1},X_n,X_{n-1}}$.)

Using results from Benaïm (1997), we showed that the long-term dynamics of the spacey random walk for an $m$-mode transition probability tensor are governed by the following dynamical system (Benson, Gleich, and Lim, 2017):

\begin{equation}
\frac{d\mathbf{x}}{dt} = \Pi(\mathbf{P[\mathbf{x}]^m}) - \mathbf{x},
\end{equation}

where $\Pi$ is a map that takes a column-stochastic transition matrix and maps it to the Perron vector of the matrix. In other words, if the spacey random walk converges, it must converge to an attractor of the dynamical system in (2.3). The dynamical system in (2.3) is a special case of the more general system in (2.1), where the map $\Lambda$ picks the eigenvector with the largest algebraic eigenvalue (the Perron vector) and the tensor has certain structural properties (it is an irreducible transition probability tensor).

To summarize, our prior work studied a specific case of the general dynamical system in (2.1) to understand the stochastic process behind principal $Z$-eigenvectors of transition probability tensors. The general dynamical system provides a new framework for computing general tensor eigenvectors—if the dynamical system in (2.1) converges, then it converges to a tensor $Z$-eigenvector. The dynamical system may not have an attractor (Peterson, 2018), but it usually does in practice (see section 3).

2.3. Relationship to the Perron iteration. Bini, Meini, and Poloni derived a Perron iteration to compute the minimal nonnegative solution of the equation

\begin{equation}
\mathbf{x} = \mathbf{a} + \mathbf{Bx}^2,
\end{equation}

where $\mathbf{a}$ and $\mathbf{B}$ are nonnegative and the all-ones vector $\mathbf{e}$ is a (nonminimal) nonnegative solution (Bini, Meini, and Poloni, 2011; Meini and Poloni, 2011). The Perron iteration for computing the minimal nonnegative solution is

\begin{equation}
\mathbf{x}_{k+1} = \Pi(\mathbf{F} + \mathbf{B}[\mathbf{e}] - \mathbf{B}[\mathbf{x}_k]),
\end{equation}

\text{where} \ \mathbf{F} = \Pi(\mathbf{a}).
where $F = \sum F_{i,j,:}$, $e$ is the vector of all ones, and $\Pi$ maps a nonnegative matrix to its Perron vector with unit 1-norm. Suppose that $x_0 \geq 0$ and $e^T x_0 = 1$. Then every iterate $x_k$ is stochastic and

$$
(2.6) \quad F = \mathcal{W}[x_k], \quad W_{i,j,\ell} = F_{i,j}, \quad \mathcal{B}[e] = \mathcal{Z}[x_k], \quad Z_{i,j,\ell} = [\mathcal{B}[e]]_{i,j}.
$$

Thus, we can rewrite the Perron iteration in (2.5) as $x_{k+1} = \Pi(\mathcal{T}[x_k])$, where $\mathcal{T} = \mathcal{W} + \mathcal{Z} - \mathcal{B}$. These iterates are equivalent to forward Euler integration of the dynamical system in (2.1) with unit step size and eigenvector map $\Lambda = \Pi$.

Meini and Poloni (2018) derived a similar Perron iteration for the solution to (2.2) for the case of a 3-mode transition probability tensor. The algorithm first computes a minimal substochastic nonnegative vector $m$ satisfying $m = Pm^2$ using a Newton method. The Perron iteration for the transition probability tensor is then $x_{k+1} = \Pi(\mathcal{T}[x_k])$ for $\mathcal{T}_{i,j,\ell} = P_{i,j,\ell}(1 + m_j + m_\ell)$. The iterates are again equivalent to forward Euler integration of (2.1) with unit step size.

### 2.4. Relationship to the shifted higher-order power method

The shifted higher-order power method (Kolda and Mayo, 2011) can be derived by noticing that

$$
(2.7) \quad (1 + \gamma) \lambda x = \mathcal{T} x^{m-1} + \gamma \lambda x
$$

for any eigenpair. This yields the iteration

$$
(2.8) \quad x_{k+1} = \frac{1}{1 + \gamma} \left( \frac{\mathcal{T} x_k^{m-1} + \gamma x_k}{\Pi(\mathcal{T} x_k^{m-1} + \gamma x_k)} \right)
$$

for any shift parameter $\gamma$ (the case where $\gamma = 0$ is just the classical “higher-order power method” (De Lathauwer, De Moor, and Vandewalle, 2000; Regalia and Kofidis, 2000; Kofidis and Regalia, 2002)). Kolda and Mayo showed that when $T$ is symmetric, the iterates in (2.8) converge monotonically to a tensor eigenvector given an appropriate shift $\gamma$.

If $T = P$ for some transition probability tensor $P$ and we are interested in the case when $\lambda = 1$ and we normalize via $\|x\|_1 = 1$, then one can also derive these iterates by the dynamical system

$$
(2.9) \quad \frac{dx}{dt} = Px^{m-1} - x
$$

(cf. (2.3)). If this dynamical system converges ($dx/dt = 0$), then $x = Px^{m-1}$, and $x$ is a tensor Z-eigenvector with eigenvalue $1$. If we numerically integrate (2.9) using the forward Euler method with step size $h = 1/(1 + \gamma)$ and any starting vector $x_0$ satisfying $x_0 \geq 0$ and $\|x_0\|_1 = 1$, then the iterates are

$$
(2.10) \quad x_{k+1} = x_k + \frac{1}{1 + \gamma} \left( Px_k^{m-1} - x_k \right)
$$

$$
(2.11) \quad = \frac{1}{1 + \gamma} \left( Px_k^{m-1} + \gamma x_k \right) = \frac{1}{1 + \gamma} \left( Px_k^{m-1} + \gamma x_k \right) || P x_k^{m-1} + \gamma x_k ||_1,
$$

which are the same as the shifted higher-order power method iterates in (2.8). The last equality follows from the fact that $\|x_k\|_1 = 1$ and $x_k \geq 0$, which is true by a simple induction argument. The base case holds by the initial conditions and

$$
(2.12) \quad || Px_k^{m-1} + \gamma x_k ||_1 = || P x_k^{m-1} ||_1 + \gamma = 1 + \gamma
$$

since $Px_k^{m-1}$ and $x_k$ are both stochastic vectors.
With a general tensor $\mathbf{T}$, we can either enforce normalization by evolving the dynamical system over, say, a unit sphere, or we can let the vector $\mathbf{x}$ be unnormalized. The latter case gives a more direct connection to the shifted symmetric higher-order power method (SS-HOPM). In this case, any vector $\mathbf{x}$ where $\mathbf{Tx} = \|\mathbf{x}\|_2 \mathbf{x}$ is a tensor $\mathbf{Z}$-eigenvector. This leads to the following dynamical system:

$$
\frac{d\mathbf{x}}{dt} = \mathbf{Tx} - \|\mathbf{x}\|_2 \mathbf{x}.
$$

If $d\mathbf{x}/dt = 0$, then $\|\mathbf{x}\|_2 \mathbf{x} = \mathbf{Tx}$, so $\mathbf{x}$ is a $\mathbf{Z}$-eigenvector of $\mathbf{T}$ with eigenvalue $\|\mathbf{x}\|_2$. Now suppose that we numerically integrate the dynamical system in (2.13) by

1. taking a forward Euler step to produce the iterate $\mathbf{x}'_{k+1}$;
2. projecting $\mathbf{x}'_{k+1}$ onto the unit sphere by $\mathbf{x}_{k+1} = \mathbf{x}'_{k+1}/\|\mathbf{x}'_{k+1}\|_2$.

If the step size of the forward Euler method is $h = 1/(1 + \gamma)$, then

$$
\mathbf{x}'_{k+1} = \mathbf{x}_k + \frac{1}{1 + \gamma} (\mathbf{Tx}_k - \|\mathbf{x}_k\|_2 \mathbf{x}_k) = \frac{1}{1 + \gamma} (\mathbf{Tx}_k + \gamma \mathbf{x}_k)
$$

since $\|\mathbf{x}_k\|_2 = 1$. The projection onto the unit sphere then gives the shifted higher-order power method iterates in (2.8).

3. Numerical examples. We now show that our method works on two test tensors used in prior work. Subsection 3.1 shows that our approach can compute all eigenvalues of a specific tensor, while the (shifted) higher-order power method cannot compute all of the eigenvalues. Subsection 3.2 verifies that our approach can compute all eigenvalues of a tensor whose eigenvalues were found with semidefinite programming (SDP). Finally, subsection 3.3 shows that our method is faster than SS-HOPM and the SDP method.

3.1. Example 3.6 from Kolda and Mayo (2011). Our first test case is a $3 \times 3 \times 3$ symmetric tensor from Kolda and Mayo (2011, Example 3.6):

$$
\mathbf{T}_{:, :, 1} = \begin{bmatrix}
-0.1281 & 0.0516 & -0.0954 \\
0.0516 & -0.1958 & -0.179 \\
-0.0954 & -0.179 & -0.2676
\end{bmatrix}, \quad \mathbf{T}_{:, :, 2} = \begin{bmatrix}
0.0516 & -0.1958 & -0.179 \\
-0.1958 & 0.3251 & 0.2513 \\
-0.179 & 0.2513 & 0.1773
\end{bmatrix},
$$

$$
\mathbf{T}_{:, :, 3} = \begin{bmatrix}
-0.0954 & -0.179 & -0.2676 \\
-0.179 & 0.2513 & 0.1773 \\
-0.2676 & 0.1773 & 0.0338
\end{bmatrix}.
$$

The tensor has 7 eigenvalues, which Kolda and Mayo classify as “positive stable,” “negative stable,” or “unstable” (see Figure 3, top), corresponding to positive definiteness, negative definiteness, or indefiniteness of the projected Hessian of the Lagrangian of their optimization function (Kolda and Mayo, 2011). (Since the tensor has an odd number of modes, we only consider eigenvalues up to sign.) Kolda and Mayo showed that their SS-HOPM, a generalization of the symmetric higher-order power method (S-HOPM) (De Lathauwer, De Moor, and Vandewalle, 2000; Regalia and Kofidis, 2000; Kofidis and Regalia, 2002), only converges to eigenvectors of the positive or negative stable eigenvalues. An adaptive version of SS-HOPM has the same shortcoming (Kolda and Mayo, 2014). A recently proposed Newton iteration can converge to eigenpairs where the projected Hessian has eigenvalues bounded away from 0 (Jaffe, Weiss, and Nadler, 2018).
<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Type</th>
<th>S-HOPM</th>
<th>SS-HOPM</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0180</td>
<td>Neg. stable</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>0.4306</td>
<td>Neg. stable</td>
<td>38</td>
<td>29</td>
<td>38</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.8730</td>
<td>Neg. stable</td>
<td>62</td>
<td>40</td>
<td>62</td>
<td>0</td>
<td>47</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.0006</td>
<td>Pos. stable</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>19</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.0018</td>
<td>Unstable</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>0.0033</td>
<td>Unstable</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>0.2294</td>
<td>Unstable</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Map $M$ to the eigenvector with largest magnitude eigenvalue;</td>
</tr>
<tr>
<td>2</td>
<td>Map $M$ to the eigenvector with smallest magnitude eigenvalue;</td>
</tr>
<tr>
<td>3</td>
<td>Map $M$ to the eigenvector with largest algebraic eigenvalue;</td>
</tr>
<tr>
<td>4</td>
<td>Map $M$ to the eigenvector with smallest algebraic eigenvalue;</td>
</tr>
<tr>
<td>5</td>
<td>Map $M$ to the eigenvector with second smallest algebraic eigenvalue.</td>
</tr>
</tbody>
</table>

We used the forward Euler method with step size set to 0.5 in order to compute the eigenvalues. Empirically, convergence is fast, requiring fewer than 10 iterations (Figure 3, bottom row). One can also also compute these eigenvectors with semidefinite programming (Cui, Dai, and Nie, 2014), although the scalability of such methods is limited (see subsection 3.3). We next provide numerical results from a tensor in this literature.

3.2. Example 4.11 from Cui, Dai, and Nie (2014). Our second test case is a $5 \times 5 \times 5$ symmetric tensor from Cui, Dai, and Nie (2014, Example 4.11):

$$T_{i,j,k} = \frac{(-1)^i}{i} + \frac{(-1)^j}{j} + \frac{(-1)^k}{k}, \quad 1 \leq i, j, k \leq 5. \tag{3.1}$$

The tensor has 3 eigenvalues (again, the tensor has an odd number of modes, so...
Table 1. Eigenvalues and corresponding eigenvectors for the test tensor.

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>SDP</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.9779</td>
<td>✓</td>
<td>94</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>4.2876</td>
<td>✓</td>
<td>6</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.0000</td>
<td>✓</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Fig. 4. (Top) The 3 eigenvalues of the test tensor from Cui, Dai, and Nie (2014, Example 4.11) and the number of random trials (out of 100) that converge to the eigenvalue for 5 variations of our dynamical systems approach. Variation 1 selects the largest magnitude eigenvalue, variation 2 selects the smallest magnitude eigenvalue, variation 3 selects the largest algebraic eigenvalue, variation 4 selects the smallest algebraic eigenvalue, and variation 5 selects the second smallest algebraic eigenvalue. Our algorithm is able to compute all of the eigenvalues, which the SDP approach is guaranteed to compute. (Bottom) Convergence plots for the three eigenvalues from different variations of our algorithm in terms of the Rayleigh quotient, where \( \mathbf{x}_k \) is the \( k \)th iterate.

The eigenvalues are only defined up to sign. We use the same 5 variations of our algorithm to compute the eigenpairs (Figure 4). Again, we are able to compute all of the eigenvalues of the tensor, and convergence is rapid.

3.3. Scalability experiments. Now we compare the performance of our algorithm to both SS-HOPM, as implemented in the Tensor Toolbox for MATLAB,\(^2\) (Bader and Kolda, 2006; Bader et al., 2017)), and the SDP method, also implemented in MATLAB.\(^3\) Our implementation is written in Julia and is also publicly available.\(^4\)

The following order-\( m \), \( n \)-dimensional tensor (which is a generalization of (3.1)) serves as the test case for our experiments:

\[
T_{i_1, \ldots, i_m} = \sum_{r=1}^m \frac{(-1)^{i_r}}{i_r}, \quad 1 \leq i_1, \ldots, i_m \leq n.
\]

With SS-HOPM, we use a tolerance of \( 10^{-6} \), a shift of 1, and 100n random initializations. With the SDP method, we use the default parameter settings. With our dynamical systems method, we use a stopping tolerance of \( 10^{-6} \), the forward Euler

\(^2\)https://www.tensortoolbox.org
\(^3\)http://www.math.ucsd.edu/~njw/CODES/reigymtensor/areigstsrweb.html
\(^4\)https://github.com/arbenson/TZE-dynsys
integration scheme with step size $= 0.5$, and maps $\Lambda$ corresponding to $k$th largest algebraic and magnitude eigenvalue, $k = 1, \ldots, n$, each with 50 trials of random initial starting points. With this setup, SS-HOPM and our approach use the same number of randomly initialized trials. We performed all experiments on a 3.1-GHz Intel Core i7 MacBook Pro with 16 GB of RAM.

Figure 5 shows the running times of the algorithms for $m = 3, 4, 5$ and $n = 5, 6, \ldots, 15$. The main takeaway is that the SDP method is much slower than the other two methods—this is the price we pay for being able to compute all of the real eigenvalues and dealing with NP-hardness. Our dynamical systems approach is faster than SS-HOPM, which is somewhat surprising since we require an eigendecomposition of an $n \times n$ matrix at each iteration. However, the performance difference is a result of rapid convergence, as observed in Figures 3 and 4. Finally, although the tensors here are relatively small, our method has been used in recent work to compute eigenvectors of tensors of orders 3, 4, and 5 with dimensions in the tens of thousands (Benson, 2019).

4. Stochastics as a guide. Scalable methods for computing tensor eigenvectors remain a challenge. Our new framework for computing $Z$-eigenvectors offers insights through three observations. First, a tensor $Z$-eigenvector is a matrix eigenvector of some matrix, where the matrix is obtained by applying the tensor collapse operator with the $Z$-eigenvector itself. Second, for a certain class of tensors where eigenvectors have a stochastic interpretation, the dynamical system in (2.1) is the one that governs the long-term dynamics of the stochastic process. Third, the same type of dynamical system seems to work for more general tensors. This framework can compute tensor eigenvectors that other scalable methods, such as SS-HOPM, cannot.

The dynamical system framework is a flexible setup to create solvers for tensor eigenvector problems, and dynamical systems have also been used in matrix eigenvector problems (Chu, 1984; Golub and Liao, 2006). The difference between SS-HOPM and our proposed framework, for instance, is essentially that SS-HOPM takes a single step of the power method on the matrix $T[x]$, whereas we converge to an eigenvector of $T[x]$. There is a rich space to interpolate between these positions. Straightforward ideas include low-degree polynomial filters that target specific eigenvectors.

Indeed, one major challenge is knowing what map $\Lambda$ to choose—different choices lead to different eigenvectors, and there is no immediate relationship between them for general tensors. A second class of open questions relates to convergence theory. At the
moment, we have demonstrated that there are both convergent and nonconvergent cases (Figure 2). We can alleviate this problem by normalizing the iterates of the integration scheme to have unit 2-norm after each iteration (Figure 6): however, we do not have a good theory for why this works. Finally, our method is not immediately applicable to $H$-eigenvectors because Observation 1.1 no longer holds. Adapting our methodology to this class of eigenvectors is an area for future research.

Our framework came from relating tensor eigenvectors to stochastic processes. This is quite different from the core ideas in the tensor literature, which are firmly rooted in algebraic generalizations. We hope that these results encourage further development of the relationships between stochastics and tensor problems.

Acknowledgments. We thank Brad Nelson for providing valuable feedback.

REFERENCES


