SIAM J. MATRIX ANAL. APPL.
Vol. 39, No. 3, pp. 1071–1094

NEWTON CORRECTION METHODS FOR COMPUTING REAL EIGENPAIRS OF SYMMETRIC TENSORS*

ARIEL JAFFE†, ROI WEISS†, AND BOAZ NADLER†

Abstract. Real eigenpairs of symmetric tensors play an important role in multiple applications. In this paper we propose and analyze a fast iterative Newton-based method to compute real eigenpairs of symmetric tensors. We derive sufficient conditions for a real eigenpair to be a stable fixed point for our method and prove that given a sufficiently close initial guess, the convergence rate is quadratic. Empirically, our method converges to a significantly larger number of eigenpairs compared with previously proposed iterative methods, and with enough random initializations typically finds all real eigenpairs. In particular, for a generic symmetric tensor, the sufficient conditions for local convergence of our Newton-based method hold simultaneously for all its real eigenpairs.

Key words. tensor eigenvectors, tensor eigenvalues, symmetric tensor, higher-order power method, Newton-based methods, Newton correction method

AMS subject classifications. 15A69, 15A72, 15A18, 49M15

DOI. 10.1137/17M1133312

Notation. We denote vectors by lowercase boldface letters, as in $\boldsymbol{x}$, matrices by uppercase letters, as in W, and higher-order tensors by caligraphic letters, as in $\mathcal{T}$. We denote $[n] = \{1, \dots, n\}$. I is the identity matrix whose dimension depends on the context and $S_{n-1} = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| = 1\}$ is the unit sphere.

1. Introduction.

Eigenpairs of symmetric tensors have received much attention in recent years because of their applicability in a wide range of disciplines. Introduced by Lim [21] and Qi [26], tensor eigenpairs were used, for example, in the analysis of hypergraphs [20], high-order Markov chains [24], establishing the positive-definiteness of multivariate forms [25], diffusion tensor imaging [28, 27], and data analysis [2, 1].

The focus of this paper is on fast iterative methods to compute the real eigenpairs of symmetric tensors. These methods were recently applied by the authors and collaborators in [16] for learning a binary latent variable model by computing the eigenpairs of a third order moment tensor of the observed data.

There are major differences between tensor eigenpairs, whose formal definition is reviewed in section 2, and their well-studied matrix counterparts. Whereas any symmetric $n \times n$ matrix has exactly n real eigenvalues with corresponding orthogonal eigenvectors, the situation for tensors is fundamentally different. The eigenvectors of a symmetric tensor are not necessarily orthogonal and some may in fact be complex-valued. Furthermore, the number of eigenvalues is in general significantly larger than n. For symmetric real tensors of order m and dimensionality n, [5] proved that the number of complex eigenvalues is at most $((m-1)^n - 1)/(m-2)$. For generic tensors, this is the *exact* number of complex eigenvalues. As a lower bound, it is known that for odd-order tensors at least one real eigenvalue exists [5], while for symmetric

*Received by the editors June 7, 2017; accepted for publication (in revised form) by L. De Lathauwer February 21, 2018; published electronically July 3, 2018.
http://www.siam.org/journals/simax/39-3/M113331.html
Funding: The research of the authors was funded in part by the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI) and by NIH grant 1R01HG008383-01A1.
†Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel (ariel.jy@gmail.com, weiss.roi@gmail.com, boaz.nadler@weizmann.ac.il).

even-ordered tensors at least n real eigenvalues exist [6]. Recently, [4] analyzed the expected number of real eigenvalues of a random Gaussian symmetric tensor.

From a computational perspective, while all matrix eigenpairs can be computed efficiently, Hillar and Lim [15] showed that enumerating all eigenpairs of a symmetric tensor is in general #P. Nonetheless, Cui et al. [8] derived a method to compute *all* eigenpairs sequentially, based on a hierarchy of semidefinite relaxations. Chen et al. [7] proposed a homotopy continuation method for the same purpose. While these algorithms are guaranteed to find all isolated eigenpairs, they are computationally demanding even for moderate tensor dimensions. For example, computing all real eigenpairs of a random $8 \times 8 \times 8 \times 8$ tensor using the `zeig` procedure in the `TenEig` package of [7] takes several hours on a standard PC.

In recent years, several iterative methods were developed to compute at least some tensor eigenpairs. Some methods were specifically designed to compute the largest eigenvalue [24, 22, 14, 11]. Han [13] proposed a method based on unconstrained optimization to compute both the maximal and minimal eigenvalues of an even-order tensor. As described in section 3, adaptations of the popular power method to tensors were suggested in [9, 18, 19, 31]. While these iterative methods are computationally fast, in general they converge to only a strict subset of all eigenpairs.

In this work we present a different iterative approach to compute real eigenpairs of symmetric tensors. As detailed in section 4, our approach is based on adapting the matrix Newton correction method (NCM) to the tensor case. We derive sufficient conditions for local convergence of NCM and prove that its convergence rate is quadratic. Our analysis reveals that NCM may fail to converge to eigenvectors with eigenvalue zero and has small attraction region for eigenvalues close to zero. To overcome this limitation, we next derive a variant, denoted the orthogonal Newton correction method (O–NCM), which enjoys improved run-time and convergence guarantees. We observe that for a generic symmetric tensor, the sufficient conditions for either NCM or O–NCM to converge to all its eigenpairs hold with probability one. In section 5 we illustrate that these sufficient conditions are not necessary.

In section 6 we present numerical simulations that support our theoretical analysis. For random tensors of modest size, multiple random initializations of NCM or O–NCM can find all eigenpairs significantly faster than other methods. For example, on a random $8 \times 8 \times 8 \times 8$ tensor, our methods typically found all eigenpairs within a few seconds. We conclude with a summary and discussion in section 7.

2. The symmetric tensor eigenproblem.

Let $\mathcal{T} \in \mathbb{R}^{n\times\cdots\times n}$ be a tensor of order m and dimension n, with entries $t_{i_1,\dots,i_m}$, where $i_1,\dots,i_m \in [n]$. We assume that $\mathcal{T}$ is symmetric, namely, $t_{i_1,\dots,i_m} = t_{\pi(i_1,\dots,i_m)}$ for all permutations π of the m indices $i_1,\dots,i_m$. A tensor $\mathcal{T}$ can be viewed as a *multilinear* operator: for matrices $W^1,\dots,W^m$ with $W^i \in \mathbb{R}^{n\times d_i}$, the tensor-mode product, denoted by $\mathcal{T}(W^1,\dots,W^m) \in \mathbb{R}^{d_1\times\cdots\times d_m}$, yields a new tensor whose $(i_1,\dots,i_m)$th entry is

$$\left[\mathcal{T}(W^1,\dots,W^m)\right]_{i_1,\dots,i_m} = \sum_{j_1,\dots,j_m\in[n]} W^1_{j_1,i_1}\cdots W^m_{j_m,i_m} t_{j_1,\dots,j_m}.$$

Tensor eigenpairs. Several definitions of tensor eigenpairs appear in the literature. Here we consider the one introduced as Z-eigenpairs in [26] and l^2-eigenpairs in [21].

DEFINITION 2.1. *A pair $(\boldsymbol{x}^*,\lambda^*) \in \mathbb{R}^n \times \mathbb{R}$ is a real eigenpair of $\mathcal{T}$ if*

$$\mathcal{T}(I,\boldsymbol{x}^*,\dots,\boldsymbol{x}^*) = \lambda^*\boldsymbol{x}^* \quad \textit{and} \quad \|\boldsymbol{x}^*\| = 1. \tag{1}$$

Note that if $(\boldsymbol{x}^*, \lambda^*)$ is an eigenpair, then $(-\boldsymbol{x}^*, (-1)^m\lambda^*)$ is an eigenpair as well. Following common practice, we treat these two pairs as belonging to the same equivalence class [5].

Definition 2.1 can be equivalently stated using the following m-degree homogeneous polynomial in $\boldsymbol{x} \in \mathbb{R}^n$,

$$\mu(\boldsymbol{x}) = \mathcal{T}(\boldsymbol{x}, \dots, \boldsymbol{x}) = \sum_{i_1,\dots,i_m \in [n]} t_{i_1,\dots,i_m} x_{i_1} \cdots x_{i_m}. \tag{2}$$

As shown in [21], the real eigenpairs of $\mathcal{T}$ correspond to the *critical points* of $\mu(\boldsymbol{x})$ when constrained to the unit sphere S_{n-1}. Formally, define the Lagrangian

$$L(\boldsymbol{x}, \lambda) = \mu(\boldsymbol{x}) - \frac{m\lambda}{2}(\|\boldsymbol{x}\|^2 - 1), \qquad \lambda \in \mathbb{R}.$$

A constrained critical point $(\boldsymbol{x}^*, \lambda^*)$ of μ satisfies the Karush–Kuhn–Tucker conditions,

$$\frac{1}{m}\nabla_{\boldsymbol{x}^*} L(\boldsymbol{x}^*, \lambda^*) = \mathcal{T}(I, \boldsymbol{x}^*, \dots, \boldsymbol{x}^*) - \lambda^* \boldsymbol{x}^* = 0,$$

where $\lambda^* = \mu(\boldsymbol{x}^*)$ is such that $\|\boldsymbol{x}^*\| = 1$. This is precisely (1). For future use, we denote the *gradient* of $L(\boldsymbol{x}, \lambda)$ at an arbitrary point $\boldsymbol{x} \in S_{n-1}$ by

$$\boldsymbol{g}(\boldsymbol{x}) = \frac{1}{m}\, \nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda)|_{\lambda=\mu(\boldsymbol{x})} = \mathcal{T}(I, \boldsymbol{x}, \dots, \boldsymbol{x}) - \mu(\boldsymbol{x})\boldsymbol{x}. \tag{3}$$

Similarly, we denote the *Hessian matrix* by

$$H(\boldsymbol{x}) = \frac{1}{m}\, \nabla^2_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda)\big|_{\lambda=\mu(\boldsymbol{x})} = (m-1)\mathcal{T}(I, I, \boldsymbol{x}, \dots, \boldsymbol{x}) - \mu(\boldsymbol{x})I. \tag{4}$$

As will become clear in section 4, the spectral structure of $H(\boldsymbol{x}^*)$ plays a fundamental role in the convergence of our proposed Newton-based methods to $(\boldsymbol{x}^*, \lambda^*)$.

3. Power methods for computing tensor eigenpairs. To motivate our approach, it is first instructive to briefly review previous iterative methods, specifically the symmetric higher-order power method (HOPM) [9] and the shifted-HOPM [18, 19]. The HOPM was derived as a way to compute the best rank-1 approximation of a symmetric tensor under the squared error loss,

$$\operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} \|\mathcal{T} - \boldsymbol{x} \otimes \ldots \otimes \boldsymbol{x}\|_F^2 = \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} \sum_{i_1,\dots,i_m=1}^{n} (t_{i_1,\dots,i_m} - x_{i_1} \ldots x_{i_m})^2.$$

Although the above problem is nonconvex and has no closed-form solution, it was shown in [9] that it is equivalent to finding the vector $\boldsymbol{x}^*$ with $\|\boldsymbol{x}^*\| = 1$ that maximizes the objective function $\mu(\boldsymbol{x})$ in (2). To compute $\boldsymbol{x}^*$, the following generalization of the matrix power method to high-order tensors was proposed. Starting from a (random) initial guess $\boldsymbol{x}_{(0)} \in S_{n-1}$, HOPM iterates

$$\boldsymbol{x}_{(k+1)} = \frac{\nabla_{\boldsymbol{x}} \mu(\boldsymbol{x}_{(k)})}{\|\nabla_{\boldsymbol{x}} \mu(\boldsymbol{x}_{(k)})\|} = \frac{\mathcal{T}(I, \boldsymbol{x}_{(k)}, \dots, \boldsymbol{x}_{(k)})}{\|\mathcal{T}(I, \boldsymbol{x}_{(k)}, \dots, \boldsymbol{x}_{(k)})\|}. \tag{5}$$

It was shown in [17, Theorem 4] that for an even-order tensor, if its associated function $\mu(\boldsymbol{x})$ is *convex* or *concave*, then HOPM is guaranteed to converge to a local optimum of $\mu(\boldsymbol{x})$ in S_{n-1}. For general symmetric tensors, however, HOPM has no convergence guarantees and may indeed fail to converge. See [17] for a specific example.

To overcome the limitations of HOPM, [18, 19] proposed the shifted function

$$\mu_\alpha(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \alpha\|\boldsymbol{x}\|^m, \qquad \alpha \in \mathbb{R}.$$

Since on the unit sphere $\mu_\alpha(\boldsymbol{x}) = \mu(\boldsymbol{x})+\alpha$, the critical points of μ and μ_α are identical. Instead of (5), the shifted-HOPM iterates

$$\boldsymbol{x}_{(k+1)} = \frac{\nabla_{\boldsymbol{x}}\mu_\alpha(\boldsymbol{x}_{(k)})}{\|\nabla_{\boldsymbol{x}}\mu_\alpha(\boldsymbol{x}_{(k)})\|} = \frac{\mathcal{T}(I, \boldsymbol{x}_{(k)}, \ldots, \boldsymbol{x}_{(k)}) + \alpha\boldsymbol{x}_{(k)}}{\|\mathcal{T}(I, \boldsymbol{x}_{(k)}, \ldots, \boldsymbol{x}_{(k)}) + \alpha\boldsymbol{x}_{(k)}\|}.$$

Importantly, the value of α can be tuned so that from any starting point $\boldsymbol{x}_{(0)} \in S_{n-1}$, the shifted-HOPM is guaranteed to converge to a critical point of μ_α. [19] further devised an adaptive shifted-HOPM, whereby the value of $\alpha_{(k)}$ is updated at each iteration so that $\mu_{\alpha_{(k)}}(\boldsymbol{x})$ is locally convex or concave around $\boldsymbol{x}_{(k)}$. This avoids the possible slowdown of the shifted-HOPM with a fixed value of α, while maintaining its convergence guarantees.

Convergence, attraction regions, and stable eigenpairs. The adaptive shifted-HOPM converges only to some eigenpairs of a tensor. These may be characterized as follows. For any $\boldsymbol{x} \in S_{n-1}$, let $U_{\boldsymbol{x}} \in \mathbb{R}^{n\times n-1}$ be a matrix with $n-1$ orthonormal columns that span the subspace orthogonal to $\boldsymbol{x}$. Define the *projected Hessian matrix*,

$$H_p(\boldsymbol{x}) = U_{\boldsymbol{x}}^T H(\boldsymbol{x}) U_{\boldsymbol{x}} \in \mathbb{R}^{(n-1)\times(n-1)}, \tag{6}$$

where $H(\boldsymbol{x}) \in \mathbb{R}^{n\times n}$ is the Hessian matrix in (4). In [18], an eigenvector $\boldsymbol{x}^*$ was termed *positive-stable* if $H_p(\boldsymbol{x}^*)$ was positive-definite and *negative-stable* if $H_p(\boldsymbol{x}^*)$ was negative-definite. Otherwise, $\boldsymbol{x}^*$ was termed *unstable*. [18] showed that the shifted HOPM does not converge to unstable eigenvectors but does converge to the stable ones. Further, the convergence is at a linear rate. To distinguish between eigenpairs that are stable for the (adaptive) shifted-HOPM and those that are stable for the Newton-based methods, we henceforth refer to the above as power-stable eigenpairs, power-unstable eigenpairs, etc.

As an example, the left panel of Figure 1 shows the value of $\mu(\boldsymbol{x})$ over the unit sphere for a $3\times3\times3$ symmetric tensor with 7 real eigenvectors. Three eigenvectors,

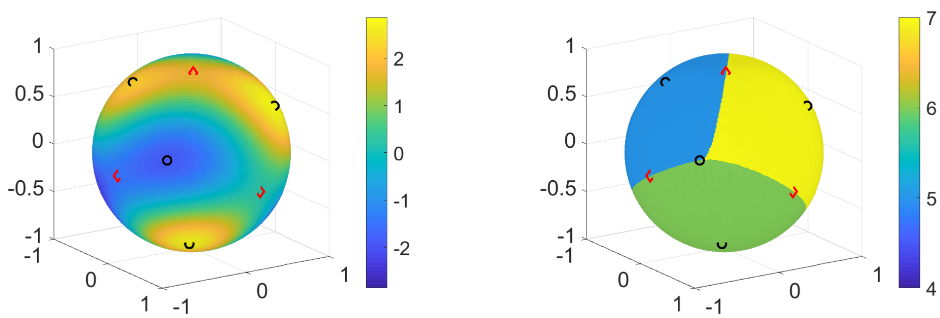

FIG. 1. *Left: The value of $\mu(\boldsymbol{x})$ on the unit sphere for a random tensor of order $m = 3$ and dimension $n = 3$. The circle/diamond shape markers indicate power-stable/power-unstable eigenvectors. Right: Attracting regions for the adaptive HOPM. The shade of each point on the unit sphere represent the eigenvector to which the method converged.*

depicted by diamond shaped markers, are power-unstable, while the remaining four, depicted by circle shaped markers, are power-stable. The right panel of the same figure shows the results of the adaptive shifted-HOPM. The color indicates the eigenvector to which the method converged, starting from various locations on the unit sphere. The figure shows clear convergence regions around 3 of the power-stable eigenpairs. The region around the fourth power-stable eigenpair appears on the back side of the sphere. In agreement with theory, the adaptive HOPM did not converge to any of the three power-unstable eigenpairs.

4. Newton-based methods for the tensor eigenproblem.

Given the limitations of the aforementioned methods, our goal is to derive a fast iterative algorithm that under mild assumptions is able to converge to *all* real eigenpairs of a symmetric tensor. To this end, we develop a Newton-based method.

4.1. NCM.

Several variants of Newton's method were derived for the symmetric matrix eigenproblem; see, for example, [30, Chapter 6]. Here we derive a Newton-based method for computing the eigenpairs of symmetric tensors. Recently, [12] considered a similar approach for finding some nonnegative eigenpairs of a nonnegative tensor.

Let $(\boldsymbol{x}^*, \lambda^*)$ be an eigenpair of a symmetric tensor $\mathcal{T}$ of order m and dimensionality n. Given an approximation $\boldsymbol{x}$ to $\boldsymbol{x}^*$, our goal is to obtain an improved approximation $\boldsymbol{x}'$ (see Figure 2, left). Denote the *exact* unknown correction by $\boldsymbol{y}^* = \boldsymbol{x}^* - \boldsymbol{x}$ and recall $\mu(\boldsymbol{x}) = \mathcal{T}(\boldsymbol{x}, \dots, \boldsymbol{x})$. Since $\boldsymbol{x}^* = \boldsymbol{x} + \boldsymbol{y}^*$ and $\lambda^* = \mu(\boldsymbol{x}^*) = \mu(\boldsymbol{x} + \boldsymbol{y}^*)$, the eigenproblem in (1) can be written as

$$\mathcal{T}(I, \boldsymbol{x}+\boldsymbol{y}^*, \dots, \boldsymbol{x}+\boldsymbol{y}^*) = \mu(\boldsymbol{x}+\boldsymbol{y}^*) \cdot (\boldsymbol{x}+\boldsymbol{y}^*). \tag{7}$$

Recalling the Hessian matrix $H(\boldsymbol{x})$ in (4), we define the matrix $A(\boldsymbol{x}) \in \mathbb{R}^{n \times n}$ by

$$A(\boldsymbol{x}) = H(\boldsymbol{x}) - m\boldsymbol{x}\mathcal{T}(I, \boldsymbol{x}, \dots, \boldsymbol{x})^T. \tag{8}$$

Setting apart the terms that are linear in $\boldsymbol{y}^*$, (7) takes the form

$$A(\boldsymbol{x})\boldsymbol{y}^* = -\boldsymbol{g}(\boldsymbol{x}) + \Delta(\boldsymbol{x}, \boldsymbol{y}^*), \tag{9}$$

where $\boldsymbol{g}(\boldsymbol{x})$ is given in (3). Here, $\Delta(\boldsymbol{x}, \boldsymbol{y}^*)$ accounts for all high order terms in $\boldsymbol{y}^*$,

$$\begin{aligned}\Delta(\boldsymbol{x}, \boldsymbol{y}^*) = {} & \sum_{i=2}^{m} \binom{m}{i} \mathcal{T}(\underbrace{\boldsymbol{x}, \dots, \boldsymbol{x}}_{m-i}, \underbrace{\boldsymbol{y}^*, \dots, \boldsymbol{y}^*}_{i})\boldsymbol{x} + \sum_{i=1}^{m} \binom{m}{i} \mathcal{T}(\underbrace{\boldsymbol{x}, \dots, \boldsymbol{x}}_{m-i}, \underbrace{\boldsymbol{y}^*, \dots, \boldsymbol{y}^*}_{i})\boldsymbol{y}^* \\ & - \sum_{i=2}^{m-1} \binom{m-1}{i} \mathcal{T}(I, \underbrace{\boldsymbol{x}, \dots, \boldsymbol{x}}_{m-i-1}, \underbrace{\boldsymbol{y}^*, \dots, \boldsymbol{y}^*}_{i}).\end{aligned} \tag{10}$$

By definition, the solution $\boldsymbol{y}^*$ to (9) satisfies $\boldsymbol{x} + \boldsymbol{y}^* = \boldsymbol{x}^*$. However, solving (9) exactly for $\boldsymbol{y}^*$ is as difficult as finding the eigenpair $(\boldsymbol{x}^*, \lambda^*)$ of the tensor $\mathcal{T}$ we started from. Instead, we devise an iterative NCM that solves (9) only approximately. Given the approximation $\boldsymbol{x}_{(k)}$ of $\boldsymbol{x}^*$ at the k^{th} iteration, NCM computes a new approximation $\boldsymbol{x}_{(k+1)}$ by neglecting the high order terms $\Delta(\boldsymbol{x}, \boldsymbol{y}^*)$ in (9). This amounts to solving the system of n linear equations

$$A(\boldsymbol{x}_{(k)})\boldsymbol{y}_{(k)} = -\boldsymbol{g}(\boldsymbol{x}_{(k)}). \tag{11}$$

Assuming $A(\boldsymbol{x}_{(k)})$ is invertible, the unique solution to (11) is given by

$$\boldsymbol{y}_{(k)} = -A(\boldsymbol{x}_{(k)})^{-1}\boldsymbol{g}(\boldsymbol{x}_{(k)}). \tag{12}$$

Algorithm 1. Newton correction method.

Input: tensor $\mathcal{T} \in \mathbb{R}^{n\times\cdots\times n}$, tolerance parameter $\delta > 0$
Initialization: Randomly choose $\boldsymbol{x}_{(0)} \in S_{n-1}$
Set $k = 0$ and $\lambda_{(0)} = \mu(\boldsymbol{x}_{(0)})$
while $\|\boldsymbol{x}_{(k)} - \boldsymbol{x}_{(k-1)}\| > \delta$ **do**
  Compute $\boldsymbol{y}_{(k)} = -A(\boldsymbol{x}_{(k)})^{-1}\boldsymbol{g}(\boldsymbol{x}_{(k)})$
  Set $\boldsymbol{x}_{(k+1)} = (\boldsymbol{x}_{(k)} + \boldsymbol{y}_{(k)})/\|\boldsymbol{x}_{(k)} + \boldsymbol{y}_{(k)}\|$
  Set $\lambda_{(k+1)} = \mu(\boldsymbol{x}_{(k+1)})$
  $k \leftarrow k+1$
end while
return $(\boldsymbol{x}_{(k)}, \lambda_{(k)})$

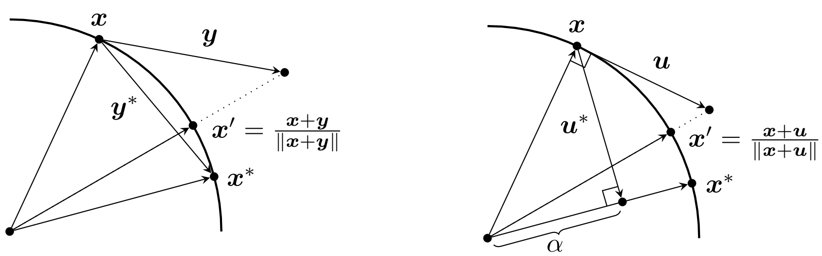

FIG. 2. *Illustration of one iteration of NCM (left) and O–NCM (right).*

The new approximation $\boldsymbol{x}_{(k+1)}$ for $\boldsymbol{x}^*$ is then

$$\boldsymbol{x}_{(k+1)} = \frac{\boldsymbol{x}_{(k)} + \boldsymbol{y}_{(k)}}{\|\boldsymbol{x}_{(k)} + \boldsymbol{y}_{(k)}\|}. \tag{13}$$

Given an initial guess $\boldsymbol{x}_{(0)} \in S_{n-1}$, NCM iterates steps (12) and (13). Once a stopping condition is met, the pair $(\boldsymbol{x}_{(k)}, \mu(\boldsymbol{x}_{(k)}))$ is returned; see Algorithm 1.

The left panel of Figure 3 shows the convergence regions of NCM for the eigenpairs of the same tensor as in Figure 1. In this case, all eigenpairs are attracting points of NCM and can thus be found by running Algorithm 1 multiple times with different (random) initial guesses.

Convergence guarantees. Two questions regarding NCM are (i) which eigenpairs of $\mathcal{T}$ can the method converge to? and (ii) what is the convergence rate? To answer these questions we recall the definition of the Hessian matrix in (4). Given an eigenpair $(\boldsymbol{x}^*, \lambda^*)$, its corresponding Hessian matrix is

$$H(\boldsymbol{x}^*) = (m-1)\mathcal{T}(I, I, \boldsymbol{x}^*, \ldots, \boldsymbol{x}^*) - \lambda^* I.$$

Note that $H(\boldsymbol{x}^*)$ is symmetric and has $\boldsymbol{x}^*$ as an eigenvector with eigenvalue $(m-2)\lambda^*$. Denote by $\mu_1^*, \ldots, \mu_{n-1}^*$ the other $n-1$ eigenvalues of $H(\boldsymbol{x}^*)$. By definition, these are the eigenvalues of the projected Hessian $H_p(\boldsymbol{x}^*)$ in (6).

DEFINITION 4.1. *For $\gamma > 0$, an eigenpair $(\boldsymbol{x}^*, \lambda^*)$ is* γ-Newton-stable *if all eigenvalues of $H_p(\boldsymbol{x}^*)$ in absolute value are at least γ, namely,* $\min_i |\mu_i^*| \geq \gamma$.

Note that $H_p(\boldsymbol{x}^*)$ is full rank if and only if $\boldsymbol{x}^*$ is γ-Newton-stable for some $\gamma > 0$. Similarly, $H(\boldsymbol{x}^*)$ is full rank if and only if $\boldsymbol{x}^*$ is γ-Newton-stable for some $\gamma > 0$ and

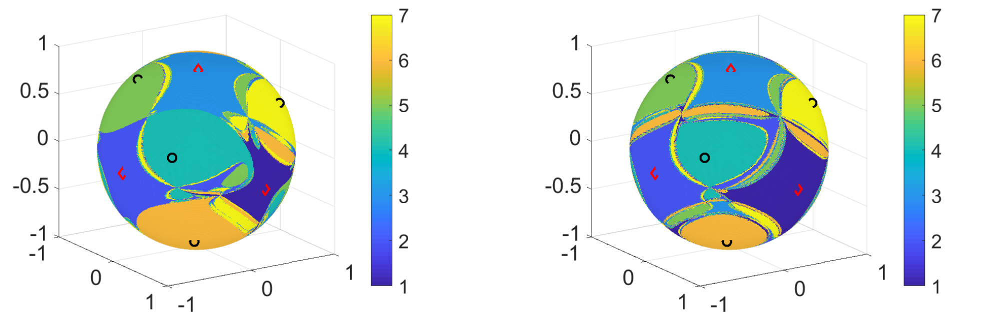

FIG. 3. *Attracting regions for NCM (left) and O–NCM (right) for the same tensor as in Figure* 1.

$\lambda^* \neq 0$. We have the following convergence guarantee for NCM. The proof is given in Appendix A.

THEOREM 4.2. *Let* $(\boldsymbol{x}^*, \lambda^*)$ *be an eigenpair of a symmetric tensor* $\mathcal{T}$. *Suppose that* $(\boldsymbol{x}^*, \lambda^*)$ *is* γ*-Newton-stable and that* $\lambda^* \neq 0$. *Then there exists an* $\varepsilon = \varepsilon(\gamma, \lambda^*) > 0$ *such that for any* $\boldsymbol{x}_{(0)}$ *that satisfies* $\|\boldsymbol{x}_{(0)} - \boldsymbol{x}^*\| < \varepsilon$, *the sequence* $\boldsymbol{x}_{(0)}, \boldsymbol{x}_{(1)}, \ldots,$ *computed by Algorithm* 1 *converges to* $\boldsymbol{x}^*$ *at a quadratic rate.*

4.2. O–NCM. As discussed above, the NCM method may not converge to an eigenvector with eigenvalue $\lambda^* = 0$. To remove this limitation, we now develop an O–NCM variant. Given an approximation $\boldsymbol{x} \in S_{n-1}$ of $\boldsymbol{x}^*$, we first decompose it into its projection onto $\boldsymbol{x}^*$ and a residual (see Figure 2, right),

$$\boldsymbol{x} = \alpha \boldsymbol{x}^* - \boldsymbol{u}^*,$$

where $\alpha = \boldsymbol{x}^T \boldsymbol{x}^*$ and $\boldsymbol{u}^* = \alpha \boldsymbol{x}^* - \boldsymbol{x}$ is the residual. Since $\boldsymbol{x}^*$ and $\boldsymbol{u}^*$ are orthogonal and $\boldsymbol{x} \in S_{n-1}$, $\|\boldsymbol{x}\|^2 = \alpha^2 + \|\boldsymbol{u}^*\|^2 = 1$. For reasons to become clear shortly, we also introduce a correction $\beta^* \equiv \beta^*(\boldsymbol{x}, \boldsymbol{x}^*)$ to the eigenvalue λ^*, defined as

$$\beta^* = \alpha^{m-2}\lambda^* - \mathcal{T}(\boldsymbol{x}, \ldots, \boldsymbol{x}) = \alpha^{m-2}\lambda^* - \mu(\boldsymbol{x}). \tag{14}$$

When $\boldsymbol{x} = \boldsymbol{x}^*$, we have $\boldsymbol{u}^* = 0$, $\alpha = 1$, and $\beta^* = 0$. Since $(\boldsymbol{x}^*, \lambda^*)$ is an eigenpair,

$$\mathcal{T}(I, \alpha\boldsymbol{x}^*, \ldots, \alpha\boldsymbol{x}^*) = \alpha^{m-1}\lambda^*\boldsymbol{x}^*.$$

Inserting $\boldsymbol{x}^* = \frac{1}{\alpha}(\boldsymbol{x} + \boldsymbol{u}^*)$ and β^* into the above equation gives

$$\mathcal{T}(I, \boldsymbol{x} + \boldsymbol{u}^*, \ldots, \boldsymbol{x} + \boldsymbol{u}^*) = (\mu(\boldsymbol{x}) + \beta^*) \cdot (\boldsymbol{x} + \boldsymbol{u}^*).$$

We set apart the terms that are linear in $\boldsymbol{u}^*$ and β^* to obtain

$$H(\boldsymbol{x})\boldsymbol{u}^* - \beta^*\boldsymbol{x} = -\boldsymbol{g}(\boldsymbol{x}) + \tilde{\Delta}(\boldsymbol{x}, \boldsymbol{u}^*, \beta^*), \tag{15}$$

where $\boldsymbol{g}$ and H were defined in (3) and (4), respectively, and $\tilde{\Delta}(\boldsymbol{x}, \boldsymbol{u}^*, \beta^*)$ includes all the remaining higher order terms in $(\boldsymbol{u}^*, \beta^*)$,

$$\tilde{\Delta}(\boldsymbol{x}, \boldsymbol{u}^*, \beta^*) = \beta^*\boldsymbol{u}^* - \sum_{i=2}^{m-1} \binom{m-1}{i} \mathcal{T}(I, \underbrace{\boldsymbol{x}, \ldots, \boldsymbol{x}}_{m-i-1}, \underbrace{\boldsymbol{u}^*, \ldots, \boldsymbol{u}^*}_{i}). \tag{16}$$

Combining (15) with the orthogonality condition, $(\boldsymbol{x}^*)^T\boldsymbol{u}^* \propto (\boldsymbol{x} + \boldsymbol{u}^*)^T\boldsymbol{u}^* = 0$, gives the following set of nonlinear equations in $(\boldsymbol{u}^*, \beta^*)$,

Algorithm 2. Orthogonal Newton correction method.

Input: tensor $\mathcal{T}$, threshold δ
Initialization: Randomly choose $\boldsymbol{x}_{(0)} \in S_{n-1}$
Set $k = 0$ and $\lambda_{(0)} = \mu(\boldsymbol{x}_{(0)})$
while $\|\boldsymbol{x}_{(k)} - \boldsymbol{x}_{(k-1)}\| > \delta$ **do**
  Compute $\boldsymbol{u}_{(k)} = -U_{\boldsymbol{x}_{(k)}} H_p(\boldsymbol{x}_{(k)})^{-1} U_{\boldsymbol{x}_{(k)}}^T \boldsymbol{g}(\boldsymbol{x}_{(k)})$
  Set $\boldsymbol{x}_{(k+1)} = (\boldsymbol{x}_{(k)} + \boldsymbol{u}_{(k)})/\|\boldsymbol{x}_{(k)} + \boldsymbol{u}_{(k)}\|$
  Set $\lambda_{(k+1)} = \mu(\boldsymbol{x}_{(k+1)})$
  $k \leftarrow k + 1$
end while
return $(\boldsymbol{x}_{(k)}, \lambda_{(k)})$

$$\begin{pmatrix} H(\boldsymbol{x}) & -\boldsymbol{x} \\ \boldsymbol{x}^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^* \\ \beta^* \end{pmatrix} = -\begin{pmatrix} \boldsymbol{g}(\boldsymbol{x}) \\ 0 \end{pmatrix} + \begin{pmatrix} \tilde{\Delta}(\boldsymbol{x}, \boldsymbol{u}^*, \beta^*) \\ -\|\boldsymbol{u}^*\|^2 \end{pmatrix}. \tag{17}$$

By construction, the solution $(\boldsymbol{u}^*, \beta^*)$ to (17) satisfies $(\boldsymbol{x} + \boldsymbol{u}^*)/\|\boldsymbol{x} + \boldsymbol{u}^*\| = \boldsymbol{x}^*$. Similarly to the NCM, we neglect the high order terms in the right-hand side of (17) and solve the system of linear equations in the $n + 1$ unknowns $(\boldsymbol{u}, \beta)$,

$$\begin{pmatrix} H(\boldsymbol{x}) & -\boldsymbol{x} \\ \boldsymbol{x}^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \beta \end{pmatrix} = -\begin{pmatrix} \boldsymbol{g}(\boldsymbol{x}) \\ 0 \end{pmatrix}. \tag{18}$$

Because of the extra variable β, (18) seems to be of dimension $n + 1$, as opposed to the n dimensional system in (11). However, as we now show, the system in (18) can be equivalently solved in the $n - 1$ dimensional subspace orthogonal to $\boldsymbol{x}$. More precisely, let $P_{\boldsymbol{x}}^{\perp} = (I - \boldsymbol{x}\boldsymbol{x}^T)$ be the projection matrix into the subspace orthogonal to $\boldsymbol{x}$ and let $U_{\boldsymbol{x}} \in \mathbb{R}^{n \times (n-1)}$ have orthonormal columns such that $P_{\boldsymbol{x}}^{\perp} = U_{\boldsymbol{x}} U_{\boldsymbol{x}}^T$. Recall the projected Hessian matrix $H_p(\boldsymbol{x}) = U_{\boldsymbol{x}}^T H(\boldsymbol{x}) U_{\boldsymbol{x}}$ in (6). The following lemma is an adaptation of [30, Theorem 6.2.2] to our setting. Its proof is given in Appendix B.

LEMMA 4.3. *A vector $\boldsymbol{u} \in \mathbb{R}^n$ satisfies* (18) *if and only if $\boldsymbol{z} \in \mathbb{R}^{n-1}$ satisfies*

$$H_p(\boldsymbol{x})\boldsymbol{z} = -U_{\boldsymbol{x}}^T \boldsymbol{g}(\boldsymbol{x}) \quad \text{and} \quad \boldsymbol{u} = U_{\boldsymbol{x}} \boldsymbol{z}. \tag{19}$$

Assuming $H_p(\boldsymbol{x})$ is invertible, the solution to (19) is $\boldsymbol{u} = -U_{\boldsymbol{x}} H_p(\boldsymbol{x})^{-1} U_{\boldsymbol{x}}^T \boldsymbol{g}(\boldsymbol{x})$. So given the k^{th} approximation $\boldsymbol{x}_{(k)}$ to $\boldsymbol{x}^*$, O–NCM computes

$$\boldsymbol{u}_{(k)} = -U_{\boldsymbol{x}_{(k)}} H_p(\boldsymbol{x}_{(k)})^{-1} U_{\boldsymbol{x}_{(k)}}^T \boldsymbol{g}(\boldsymbol{x}_{(k)}) \tag{20}$$

and the new approximation is $\boldsymbol{x}_{(k+1)} = (\boldsymbol{x}_{(k)} + \boldsymbol{u}_{(k)})/\|\boldsymbol{x}_{(k)} + \boldsymbol{u}_{(k)}\|$. Given an initial $\boldsymbol{x}_{(0)}$, O–NCM iterates these steps until a stopping condition is met; see Algorithm 2.

The right panel of Figure 3 shows the convergence regions of O–NCM for the various eigenpairs of the same tensor in Figure 1. Similarly to NCM, in this case, all eigenpairs are attracting points of O–NCM but with slightly different regions.

Convergence guarantees. We have the following convergence guarantee for O–NCM. It is similar to that of NCM in Theorem 4.2 but with the condition $\lambda^* \neq 0$ removed. The proof is given in Appendix C.

THEOREM 4.4. *Let $(\boldsymbol{x}^*, \lambda^*)$ be a γ-Newton-stable eigenpair of a symmetric tensor $\mathcal{T}$. There exists an $\varepsilon = \varepsilon(\gamma) > 0$ such that for any $\boldsymbol{x}_{(0)} \in S_{n-1}$ that satisfies $\|\boldsymbol{x}_{(0)} - \boldsymbol{x}^*\| < \varepsilon$, the sequence $\boldsymbol{x}_{(0)}, \boldsymbol{x}_{(1)}, \ldots,$ computed by Algorithm* 2 *converges to $\boldsymbol{x}^*$ at a quadratic rate.*

Remark 1. Recall that any eigenpair $(\boldsymbol{x}^*, \lambda^*)$ to which the shifted-HOPM method converges has a Hessian matrix which is either positive definite or negative definite. In either case, this Hessian matrix has full rank and thus, by Theorem 4.4, is a stable fixed point of O–NCM. In other words, O–NCM typically converges to many more tensor eigenpairs than the shifted power method. However, the adaptive shifted-HOPM is guaranteed to converge from *any* initial point, whereas no such global convergence guarantee is currently available for the Newton-based methods.

5. Convergence to eigenpairs with a rank-deficient Hessian. The sufficient condition in Theorem 4.4 for O–NCM to converge to an eigenpair $(\boldsymbol{x}^*, \lambda^*)$ rests on the smallest absolute eigenvalue of the projected Hessian matrix $H_p(\boldsymbol{x}^*)$. In particular, if the eigenpair is γ-Newton-stable for some $\gamma > 0$, an attraction neighborhood around $\boldsymbol{x}^*$ exists. We now illustrate that this sufficient condition for convergence is by no means necessary. To this end, we analyze a simple example. Denote $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^n$. For any $\omega \in \mathbb{R}$, define the following cubic n-dimensional symmetric tensor:

$$\mathcal{T}_\omega = \sum_{i=1}^{n} \boldsymbol{e}_i \otimes \boldsymbol{e}_i \otimes \boldsymbol{e}_i + \omega(\mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1}). \tag{21}$$

When $\omega = 0$, $\mathcal{T}_\omega$ is orthogonal, having the maximally possible number of $2^n - 1$ real eigenpairs. All these are Newton-stable, and among them are the n power-stable eigenpairs $\{(\boldsymbol{e}_i, 1)\}_{i=1}^n$. Assume n is odd and denote $l = \lfloor n/2 \rfloor$. Let $N(\omega)$ be the number of real eigenpairs of $\mathcal{T}_\omega$. The following proposition is proved in Appendix D.

PROPOSITION 5.1. *Define l thresholds, $\omega_i = \frac{1}{4(l-i)(n-l+i)} > 0$, $i \in \{0, \ldots, l-1\}$.*

(i) *The number $N(\omega)$ of real eigenpairs of $\mathcal{T}_\omega$ of* (21) *is*

$$N(\omega) = \begin{cases} 2^n - 1, & \omega < \omega_0, \\ 1 + 2\sum_{j=1}^{l-i-1} \binom{n}{j} + \binom{n}{l-i}, & \omega = \omega_i, \\ 1 + \sum_{j=1}^{l-i-1} \binom{n}{j}, & \omega_i < \omega < \omega_{i+1}, \\ 1, & \omega > \omega_{l-1}. \end{cases}$$

(ii) *For $\omega_i \in \{\omega_0, \ldots, \omega_{l-1}\}$, $\binom{n}{l-i}$ out of the $N(\omega_i)$ real eigenpairs of $\mathcal{T}(\omega_i)$ are not Newton-stable.*

We illustrate the above properties for $\mathcal{T}_\omega$ with $n = 5$. In this case, $l = \lfloor n/2 \rfloor = 2$, and there are two thresholds, $\omega_0 = \frac{1}{4 \cdot 2 \cdot 3} \approx 0.0417$ and $\omega_1 = \frac{1}{4 \cdot 1 \cdot 4} = 0.0625$. Figure 4 shows the number $N(\omega)$ of real eigenpairs (left) and the different eigenvalues of $\mathcal{T}_\omega$ (right) as computed by O–NCM. As expected, at ω_0 and ω_1, the number of real eigenvalues decreases.

Next, we examine the convergence of O–NCM on $\mathcal{T}_{\omega=\omega_0}$ with $n = 3$. According to Proposition 5.1, $\omega_0 = 0.125$, and the number of real eigenpairs is $N(\omega_0) = 1 + \binom{3}{1} = 4$, three of which are not Newton-stable. Figure 5 shows the attraction regions around two of the eigevectors of $\mathcal{T}_{\omega_0}$, as well as the full unit sphere. On the left, the eigenvector is Newton-stable. As expected, O–NCM converged to this eigenvector from any point in its neighborhood. In contrast, the eigenvector on the right is not Newton-stable. In this case, there is a positive probability of converging to a different eigenvector even when the initial guess is arbitrary close. Nonetheless, O–NCM converged to this eigenvector from *some* directions, even though the sufficient condition in Theorem 4.4 does not hold.

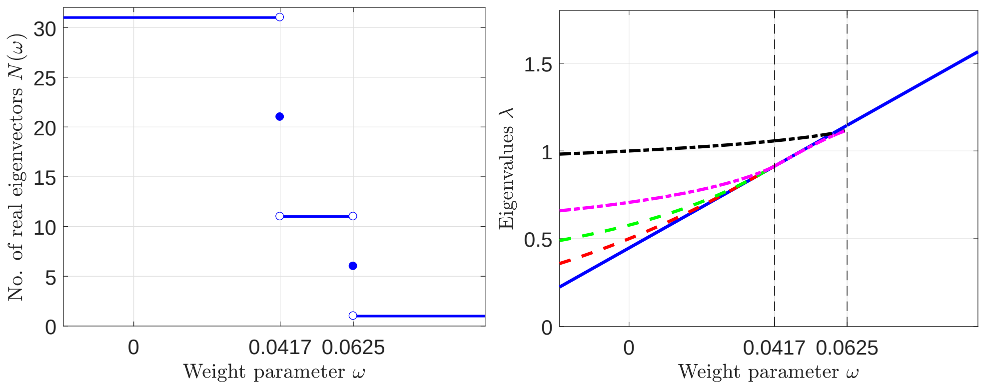

FIG. 4. *Number of real eigenpairs (left) and corresponding eigenvalues (right) for $\mathcal{T}_\omega$ with $n = 5$. Although for small values of ω there are* 31 *distinct eigenpairs, some of the eigenvalues are equal due to the tensor's symmetry.*

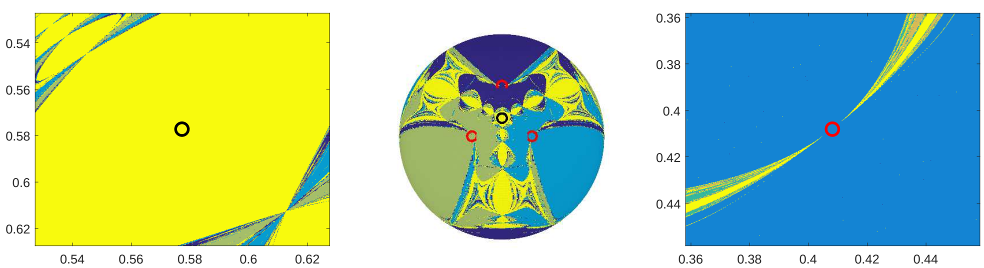

FIG. 5. *O–NCM's attraction region around an eigenvector who is Newton-stable (left), not Newton-stable (right), and the full unit sphere (middle) for $\mathcal{T}_{\omega=\omega_0}$ with dimension $n = 3$.*

In this example, all eigenpairs of $\mathcal{T}_\omega$ are *isolated*; namely, each one of them is the unique eigenpair in a small neighborhood around it. In addition, the eigenvectors which are not Newton-stable have a projected Hessian matrix that is rank deficient but *non-zero*. In Appendix E, we illustrate the behavior of O–NCM near eigenvectors that are either non-isolated or have a projected Hessian matrix equal to zero. In these cases, O–NCM may not converge.

6. Simulation results. In this section we study numerically the performance of NCM and O–NCM for computing the real eigenpairs of symmetric tensors, as compared to the homotopy method [7] and the adaptive shifted-HOPM [19].[1] In the experiments we consider *random Gaussian symmetric tensors*, whose entries are all i.i.d. $\mathcal{N}(0, 1)$ up to the symmetry constraints. All experiments were done on a PC with an Intel i-3820 3.6GHz processor, 16GB RAM, and MATLAB version R2016a.

Finding all real eigenpairs. We examine the time needed to compute *all* eigenpairs for random tensors of order $m = 4$ and various dimensions n. Similar results are obtained for other values of m. For each tensor, we first ran the homotopy method to obtain all its eigenpairs. Next, we ran NCM and O–NCM, initialized repeatedly with random points on the unit sphere, until all eigenpairs were found. Note that

[1]MATLAB code for the NCM methods can be found at https://github.com/arJaffe/NCM, for the homotopy method at http://users.math.msu.edu/users/chenlipi/TenEig.html, and for the shifted HOPM at http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html.

without running the homotopy method first, we would have no criterion to decide whether we actually found all tensors' eigenpairs. The process was sequential, where a new run was initialized only after the previous one ended. This process, however, can be easily parallelized. We stopped the NCM iterations when $\|\boldsymbol{x}_{(k)} - \boldsymbol{x}_{(k-1)}\| < \delta = 10^{-10}$ or if a maximal number of $k = k_{\max} = 200$ iterations was reached. In the latter case we declared that NCM failed to converge. For both NCM and O–NCM, and for all tensor dimensions we considered, only $\approx 0.2\%$ of all random initializations failed to converge within the maximal number of iterations.

Figure 6 (left) shows the number of real eigenpairs as averaged over 10 independent tensors for each value of n. Figure 6 (right) shows on a *logarithmic* scale the average time it took to compute all real eigenpairs via the homotopy, NCM, and O–NCM for the same tensors. These results show that both NCM and O–NCM recovered all eigenpairs faster than the homotopy method by approximately two orders of magnitude. Moreover, O–NCM did so much faster than NCM.

Small eigenvalues. To understand the gap in the runtime of NCM and O–NCM shown in Figure 6, we next examine the dependence of both methods on the eigenpair to which they converge. As suggested by Theorems 4.2 and 4.4, we expect O–NCM to have larger attraction regions for small eigenvalues as compared to NCM.

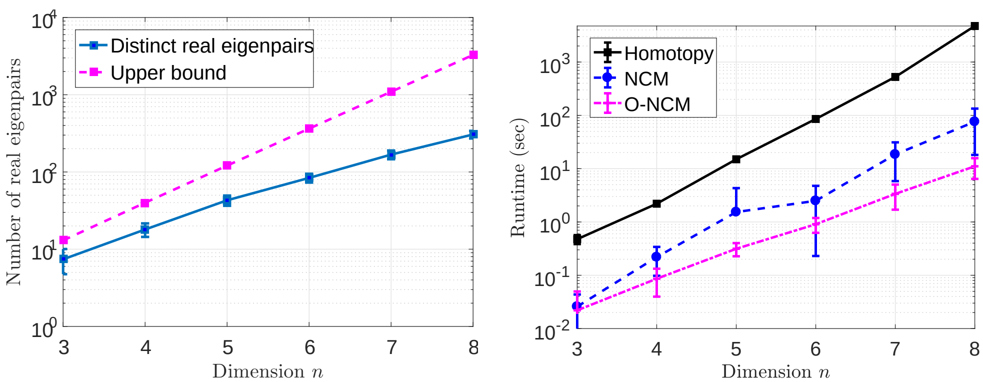

Fig. 6. *Left: The average number of real eigenpairs for random tensors of order $m = 4$. Right: The average time to compute all eigenpairs via the homotopy, NCM, and O–NCM.*

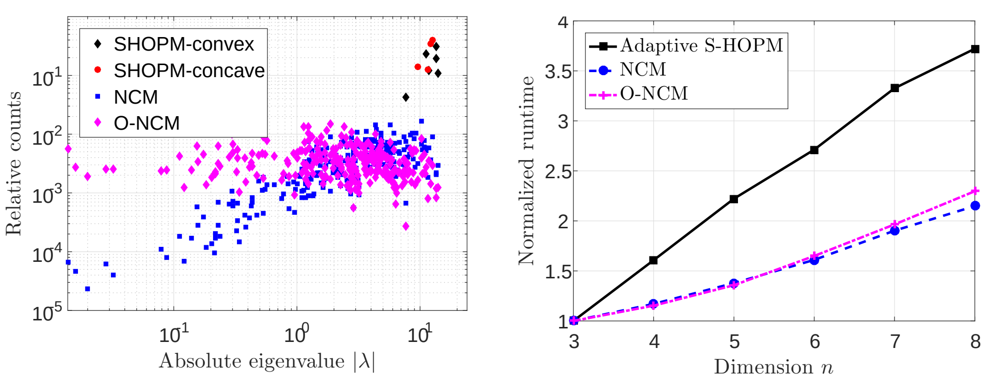

Fig. 7. *Left: Number of times NCM, O–NCM, and adaptive shifted-HOPM converged to each of the eigenvalues of a random tensor with $m = 4$ and $n = 8$, over a total of 10^6 random initial guesses. Right: Median runtime until convergence for NCM, O–NCM and the adaptive shifted-HOPM.*

Figure 7 (left) shows on a log-log scale the relative number of times the two methods converged to each eigenvalue as a function of its absolute value for a typical random tensor of order $m = 4$ and dimension $n = 8$. These counts correspond to a total of 10^6 random initializations uniformly distributed on the unit sphere. As one can see, the probability for NCM to converge to an eigenpair decreases sharply when its eigenvalue becomes small, while for O–NCM this probability seems to be independent of $|\lambda|$. This difference is the source of the gap in the runtime of the two methods for finding all eigenpairs. For completeness, the eigenvalues found by the adaptive S-HOPM are also presented.

Convergence rates. Figure 7 (right) shows the median runtime until convergence of the NCM and the shifted-HOPM. The stopping condition for all methods was set to $\|\boldsymbol{x}_{(k)} - \boldsymbol{x}_{(k+1)}\| < \delta = 10^{-10}$. The experiment was done on 100 random tensors of fourth order with various dimensions. For each tensor, we initialized all methods with 100 random starting points. To avoid the influence of any particular implementation, we normalized the results with the runtime of both methods for $n = 3$. As illustrated in Figure 7, the runtime increase of the NCM or O–NCM is significantly slower than the corresponding increase in the adaptive shifted-HOPM. However, each NCM/O–NCM iteration may be slower, as it requires matrix inversion.

7. Discussion and summary. In this paper, we developed and analyzed a Newton-based iterative approach to compute real eigenpairs of symmetric tensors. We now briefly discuss three important issues: its runtime, its ability to find all tensor eigenpairs, and its optimization point of view.

Runtime. The computational complexity of each NCM or O–NCM iteration is dominated by two operations: computing the Hessian matrix in $\mathcal{O}(n^m)$ time and solving a system of n linear equations in $\mathcal{O}(n^3)$ time. The latter step may be significantly sped up by applying various preconditioning techniques, as done in other iterative methods that solve systems of linear equations [10]. For sparse tensors, the computation of the Hessian can be accelerated as well; see [29].

Optimization point of view. Following a constructive comment by one of the referees, we note that NCM can be seen as an adaptation of the Gauss–Newton method [3]. Recall that $\boldsymbol{g}(\boldsymbol{x}) = \mathcal{T}(I, \boldsymbol{x}, \ldots, \boldsymbol{x}) - \mu(\boldsymbol{x})\boldsymbol{x} = \mathbf{0}$ if and only if $\boldsymbol{x}^* \in S_{n-1}$ is an eigenvector of $\mathcal{T}$, with a corresponding eigenvalue $\lambda^* = \mu(\boldsymbol{x}^*)$. Hence, our goal is to find the *global* minima of the realizable nonlinear least-squares problem

$$\min_{\boldsymbol{x} \in S_{n-1}} \tfrac{1}{2}\|\boldsymbol{g}(\boldsymbol{x})\|^2. \tag{22}$$

Given the current estimate $\boldsymbol{x}_{(k)} \in S_{n-1}$, Gauss–Newton first linearizes $\boldsymbol{g}(\boldsymbol{x})$ at $\boldsymbol{x}_{(k)}$,

$$\boldsymbol{g}(\boldsymbol{x}) \approx \boldsymbol{g}(\boldsymbol{x}_{(k)}) + A(\boldsymbol{x}_{(k)})(\boldsymbol{x} - \boldsymbol{x}_{(k)}) = \boldsymbol{g}(\boldsymbol{x}_{(k)}) + A(\boldsymbol{x}_{(k)})\boldsymbol{y},$$

where $[A(\boldsymbol{x})]_{ij} = \partial g_i(\boldsymbol{x})/\partial x_j$ is the $n \times n$ Jacobian matrix of $\boldsymbol{g}(\boldsymbol{x})$, given in (8). Then, instead of (22), the following approximate linear least-squares problem is solved,

$$\boldsymbol{y}_{(k)} = \operatorname*{argmin}_{\boldsymbol{y} \in \mathbb{R}^n} \tfrac{1}{2}\|\boldsymbol{g}(\boldsymbol{x}_{(k)}) + A(\boldsymbol{x}_{(k)})\boldsymbol{y}\|^2,$$

which is exactly the NCM correction in (11).

Besides NCM, other nonlinear optimization methods can be used to solve (22), such as the Levenberg–Marquardt algorithm [23] and other trust-region and line search

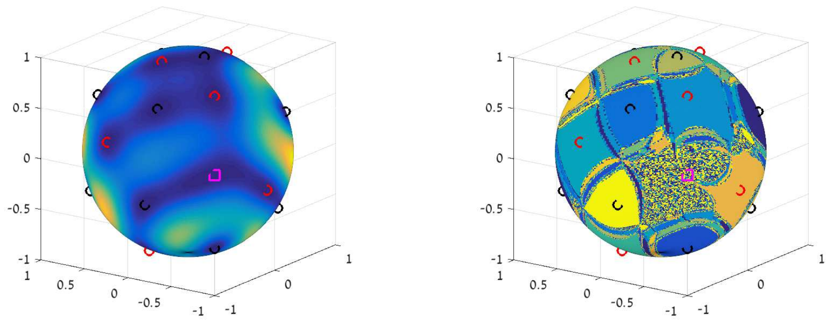

Fig. 8. *Left: Plot of* $f(\boldsymbol{x})$ *for* [17, Example 1]. *The global minima (circles) correspond to eigenpairs. A local minimum is depicted by a square. Right: Attraction regions for NCM.*

algorithms. These methods, among other things, introduce an additional regularization term to better control the direction in which the method proceeds at each iteration, similarly to the role played by the (adaptive) shifted-HOPM as compared to HOPM. Specifically, instead of (11), one solves the following linear system with an appropriate regularization matrix $B_k \in \mathbb{R}^{n\times n}$:

$$(A(\boldsymbol{x}_{(k)})^T A(\boldsymbol{x}_{(k)}) + B_k)\boldsymbol{y} = -A(\boldsymbol{x}_{(k)})^T \boldsymbol{g}(\boldsymbol{x}_{(k)}).$$

While NCM currently has no global convergence guarantees, an appropriate (adaptive) choice of B_k can lead to global convergence guarantees, including to eigenvectors having a zero Hessian. Further studying the role of regularization for the tensor eigenproblem is an interesting direction for future research.

However, in addition to the global minima, $f(\boldsymbol{x}) \equiv \frac{1}{2}\|\boldsymbol{g}(\boldsymbol{x})\|^2$ may have local minima which should be avoided. Interestingly, NCM elegantly avoids such local minima as the following example illustrates. In Figure 8 (left), we plot $f(\boldsymbol{x})$ as a function of $\boldsymbol{x} \in S_2$ for the $3\times 3\times 3\times 3$ symmetric tensor of Example 1 in [17]. Its eigenvectors are depicted by circles while a local minimum $\boldsymbol{x}_{\text{loc}}$ of f is depicted by a square symbol. In Figure 8 (right), we show the attraction regions for NCM starting from various locations on S_{n-1}. As one can see, NCM does not converge to $\boldsymbol{x}_{\text{loc}}$ and in fact is highly unstable around this point; close initial points in this neighborhood may converge to arbitrarily far eigenvectors.

To see why this is so, note that since $\boldsymbol{x}_{\text{loc}}$ is a local minimum, for an initial point $\boldsymbol{x}_{(0)}$ near $\boldsymbol{x}_{\text{loc}}$, NCM may get closer and closer to $\boldsymbol{x}_{\text{loc}}$ at the first few iterations. However, the facts that $\boldsymbol{g}(\boldsymbol{x}_{\text{loc}})$ is bounded away from $\mathbf{0}$ and $\nabla f(\boldsymbol{x}_{\text{loc}}) = A(\boldsymbol{x}_{\text{loc}})^T\boldsymbol{g}(\boldsymbol{x}_{\text{loc}}) = \mathbf{0}$ implies that $\boldsymbol{g}(\boldsymbol{x}_{\text{loc}})$ is in the null space of $A(\boldsymbol{x}_{\text{loc}})^T$. As $\boldsymbol{x}_{(k)}$ gets closer to $\boldsymbol{x}_{\text{loc}}$, $A(\boldsymbol{x}_{(k)})$ becomes close to singular. The result is an *overshoot*, a sharp increase in $\|\boldsymbol{y}_{(k)}\| = \|A(\boldsymbol{x}_{(k)})^{-1}\boldsymbol{g}(\boldsymbol{x}_{(k)})\|$, taking $\boldsymbol{x}_{(k+1)}$ far away from $\boldsymbol{x}_{(k)}$ and $\boldsymbol{x}_{\text{loc}}$.

Finding all eigenpairs of generic tensors. According to our theoretical analysis, NCM and O–NCM converge to eigenpairs whose Hessian matrix is full rank. An interesting question is whether these methods can thus converge to *all* real eigenpairs of a *generic* symmetric tensor [5]. Interpreting generic in the sense of algebraic geometry, an adaptation of [5, Theorem 1.2] to the symmetric tensor case implies the following (proof omitted).

Proposition 7.1. *All real eigenpairs of a generic symmetric tensor are Newton-stable.*

Hence, Theorems 4.2 and 4.4 imply that NCM and O–NCM are guaranteed to find all eigenpairs of a generic symmetric tensor given a sufficiently large number of random initializations.

Appendix A. Convergence of NCM. To prove Theorem 4.2 we shall make use of the following auxiliary lemma.

LEMMA A.1. *Consider one update step of Algorithm* 1, *as in* (13), *starting from an initial* $\boldsymbol{x} \in S_{n-1}$ *and ending with* $\boldsymbol{x}' = (\boldsymbol{x}+\boldsymbol{y})/\|\boldsymbol{x}+\boldsymbol{y}\| \in S_{n-1}$. *Let* $\boldsymbol{y}^* = \boldsymbol{x}^* - \boldsymbol{x}$. *If* $\|\boldsymbol{y}-\boldsymbol{y}^*\| \leq 1/2$, *then*

$$\|\boldsymbol{x}^* - \boldsymbol{x}'\| \leq \frac{2\|\boldsymbol{y}^* - \boldsymbol{y}\|}{1 - \|\boldsymbol{y}^* - \boldsymbol{y}\|}. \tag{23}$$

Proof. By definition,

$$\begin{aligned}\|\boldsymbol{x}^* - \boldsymbol{x}'\| &= \left\|\boldsymbol{x}^* - \frac{\boldsymbol{x}+\boldsymbol{y}}{\|\boldsymbol{x}+\boldsymbol{y}\|}\right\| = \left\|\boldsymbol{x}^* - \frac{\boldsymbol{x}^*-\boldsymbol{y}^*+\boldsymbol{y}}{\|\boldsymbol{x}^*-\boldsymbol{y}^*+\boldsymbol{y}\|}\right\| \\ &= \left\|\boldsymbol{x}^*\left(1 - \frac{1}{\|\boldsymbol{x}^*-\boldsymbol{y}^*+\boldsymbol{y}\|}\right) + \frac{\boldsymbol{y}^*-\boldsymbol{y}}{\|\boldsymbol{x}^*-\boldsymbol{y}^*+\boldsymbol{y}\|}\right\|. \end{aligned} \tag{24}$$

Since $\|\boldsymbol{x}^*\| = 1$, by the triangle inequality,

$$1 - \|\boldsymbol{y}^* - \boldsymbol{y}\| \leq \|\boldsymbol{x}^* - \boldsymbol{y}^* + \boldsymbol{y}\| \leq 1 + \|\boldsymbol{y}^* - \boldsymbol{y}\|.$$

Applying the triangle inequality to (24), combined with the assumption $\|\boldsymbol{y}^* - \boldsymbol{y}\| \leq 1/2$,

$$\begin{aligned}\|\boldsymbol{x}^* - \boldsymbol{x}'\| &\leq \left|1 - \frac{1}{\|\boldsymbol{x}^*-\boldsymbol{y}^*+\boldsymbol{y}\|}\right| + \frac{\|\boldsymbol{y}^*-\boldsymbol{y}\|}{\|\boldsymbol{x}^*-\boldsymbol{y}^*+\boldsymbol{y}\|} \\ &\leq \left(\frac{1}{1-\|\boldsymbol{y}^*-\boldsymbol{y}\|} - 1\right) + \frac{\|\boldsymbol{y}^*-\boldsymbol{y}\|}{1-\|\boldsymbol{y}^*-\boldsymbol{y}\|} = \frac{2\|\boldsymbol{y}^*-\boldsymbol{y}\|}{1-\|\boldsymbol{y}^*-\boldsymbol{y}\|},\end{aligned}$$

hence concluding the proof. ∎

Proof of Theorem 4.2. To prove quadratic convergence, it suffices to show that there exists an $\varepsilon > 0$ and a constant $C > 0$ such that from any initial point $\boldsymbol{x}_{(0)}$ that satisfies $\|\boldsymbol{x}_{(0)} - \boldsymbol{x}^*\| < \varepsilon$,

$$e_k = \frac{\|\boldsymbol{x}^* - \boldsymbol{x}_{(k+1)}\|}{\|\boldsymbol{x}^* - \boldsymbol{x}_{(k)}\|^2} < C \quad \forall k \geq 0.$$

We start by analyzing e_k at the first iteration $k = 0$. Let $\boldsymbol{y}_{(0)}$ be the approximate correction of $\boldsymbol{y}^* = \boldsymbol{x}^* - \boldsymbol{x}_{(0)}$, given by the solution of (11). The new approximation of $\boldsymbol{x}^*$, given by (13), is $\boldsymbol{x}_{(1)} = (\boldsymbol{x}_{(0)} + \boldsymbol{y}_{(0)})/\|\boldsymbol{x}_{(0)} + \boldsymbol{y}_{(0)}\|$. Assume for the moment that the initial guess $\boldsymbol{x}_{(0)}$ is sufficiently close to $\boldsymbol{x}^*$ so that $\|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\| < 1/2$. Then, by Lemma A.1,

$$\|\boldsymbol{x}^* - \boldsymbol{x}_{(1)}\| \leq \frac{2\|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\|}{1 - \|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\|}. \tag{25}$$

Hence, it suffices to bound $\|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\|$. To this end, we view the exact system of nonlinear (9), whose solution is $\boldsymbol{y}^*$, as a perturbation of the approximate system of linear (11), whose solution is $\boldsymbol{y}_{(0)}$. Consider the matrix A of (8) evaluated at the eigenvector $\boldsymbol{x}^*$,

$$A(\boldsymbol{x}^*) = H(\boldsymbol{x}^*) - m\lambda^* \boldsymbol{x}^* (\boldsymbol{x}^*)^T.$$

Note that $A(\boldsymbol{x}^*)$ is symmetric with eigenvalues $(\mu_1^*, \ldots, \mu_{n-1}^*, -2\lambda^*)$. Since $\boldsymbol{x}^*$ is γ-Newton-stable, $|\mu_i^*| \geq \gamma$ for all $i \in [n-1]$. In addition, since $\lambda^* \neq 0$ by assumption, $A(\boldsymbol{x}^*)$ is full rank with smallest singular value

$$\sigma^* = \sigma_{\min}(A(\boldsymbol{x}^*)) \geq \min\{\gamma, 2|\lambda^*|\} > 0. \tag{26}$$

By the continuity of $\sigma_{\min}(A(\boldsymbol{x}))$ in $\boldsymbol{x}$, there exists a $\varepsilon_1 > 0$ such that $\sigma_{\min}(A(\boldsymbol{x})) \geq \sigma^*/2$ for all $\boldsymbol{x}$ with $\|\boldsymbol{x}^* - \boldsymbol{x}\| \leq \varepsilon_1$. In particular, if $\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| < \varepsilon_1$, then $A(\boldsymbol{x}_{(0)})$ is invertible and the solution to (9) satisfies the following implicit equation in $\boldsymbol{y}^*$:

$$\boldsymbol{y}^* = -A(\boldsymbol{x}_{(0)})^{-1}(\boldsymbol{g}(\boldsymbol{x}_{(0)}) - \Delta(\boldsymbol{x}_{(0)}, \boldsymbol{y}^*)).$$

Similarly, the unique solution to the correction equation (11) is, as in (12),

$$\boldsymbol{y}_{(0)} = -A(\boldsymbol{x}_{(0)})^{-1}\boldsymbol{g}(\boldsymbol{x}_{(0)}).$$

Subtracting the last two equations gives

$$\|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\| \leq \|A(\boldsymbol{x}_{(0)})^{-1}\| \cdot \|\Delta(\boldsymbol{x}_{(0)}, \boldsymbol{y}^*)\| \leq \frac{2}{\sigma^*}\|\Delta(\boldsymbol{x}_{(0)}, \boldsymbol{y}^*)\|. \tag{27}$$

To bound $\|\Delta(\boldsymbol{x}_{(0)}, \boldsymbol{y}^*)\|$, first note that for any symmetric tensor $\mathcal{T}$ there exists an $M = M(\mathcal{T}) < \infty$ such that for any $\boldsymbol{x} \in S_{n-1}$, $\boldsymbol{y} \in \mathbb{R}^n$ and $j \leq m-1$,

$$\big\|\mathcal{T}(I, \underbrace{\boldsymbol{x}, \ldots, \boldsymbol{x}}_{m-j-1}, \underbrace{\boldsymbol{y}, \ldots, \boldsymbol{y}}_{j \text{ times}})\big\| \leq M\|\boldsymbol{y}\|^j. \tag{28}$$

Similar bounds hold for $\mathcal{T}(\boldsymbol{x}, \ldots, \boldsymbol{x}, \boldsymbol{y}, \ldots, \boldsymbol{y})\boldsymbol{x}$ and $\mathcal{T}(\boldsymbol{x}, \ldots, \boldsymbol{x}, \boldsymbol{y}, \ldots, \boldsymbol{y})\boldsymbol{y}$ according to their powers in $\boldsymbol{y}$. Bounding each term of $\Delta(\boldsymbol{x}_{(0)}, \boldsymbol{y}^*)$ in (10) separately by (28), there are less than $3m^2$ terms involving $M\|\boldsymbol{y}^*\|^2$ and, at most, $3 \cdot 2^m$ terms involving $M\|\boldsymbol{y}^*\|^j$ with $j \in \{3, \ldots, m\}$. Assuming $\|\boldsymbol{y}^*\| = \|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| < m^2/2^m \leq 1$ implies

$$\|\Delta(\boldsymbol{x}_{(0)}, \boldsymbol{y}^*)\| \leq 3m^2 M\|\boldsymbol{y}^*\|^2 + 3 \cdot 2^m M\|\boldsymbol{y}^*\|^3 \leq 6m^2 M\|\boldsymbol{y}^*\|^2.$$

Inserting this bound into (27),

$$\|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\| \leq \frac{12m^2 M}{\sigma^*}\|\boldsymbol{y}^*\|^2. \tag{29}$$

Note that if $\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| = \|\boldsymbol{y}^*\| \leq \varepsilon_2 = (\sigma^*/24m^2 M)^{1/2}$, then $\|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\| \leq 1/2$ as required by Lemma A.1. Under this condition, by (25), it follows that

$$e_0 = \frac{\|\boldsymbol{x}^* - \boldsymbol{x}_{(1)}\|}{\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\|^2} \leq \frac{1}{\|\boldsymbol{y}^*\|^2}\frac{2\|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\|}{1 - \|\boldsymbol{y}^* - \boldsymbol{y}_{(0)}\|} \leq \frac{48m^2 M}{\sigma^*}. \tag{30}$$

As an interim summary, if $\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| \leq \min\{\varepsilon_1, \varepsilon_2, m^2/2^m\} = \varepsilon_0$, then (30) holds. We conclude the proof for a general iteration $k \geq 1$ by induction. For the first induction step to work, it required that if $\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| \leq \varepsilon < \varepsilon_0$, then $\|\boldsymbol{x}^* - \boldsymbol{x}_{(1)}\| < \varepsilon$ as well. By (30), this is satisfied for $\varepsilon = \min\{\varepsilon_0, \sigma^*/48m^2 M\}$, and the proof for a general k holds similarly. The quadratic convergence of Algorithm 1 follows. ∎

Appendix B. Proof of Lemma 4.3. We show that a vector $\boldsymbol{u}$ satisfies (18) if and only if it satisfies

$$P_{\boldsymbol{x}}^{\perp} H(\boldsymbol{x}) P_{\boldsymbol{x}}^{\perp} \boldsymbol{u} = -P_{\boldsymbol{x}}^{\perp} \boldsymbol{g}(\boldsymbol{x}) \quad \text{and} \quad \boldsymbol{x}^T \boldsymbol{u} = 0. \tag{31}$$

Lemma 4.3 then follows by recalling that $P_{\boldsymbol{x}}^{\perp} = U_{\boldsymbol{x}} U_{\boldsymbol{x}}^T$ and multiplying the first equation in (31) by $U_{\boldsymbol{x}}^T$ from the left.

To prove the first direction, note that by the last row of (18), the solution $\boldsymbol{u}$ to (18) is perpendicular to $\boldsymbol{x}$, so $\boldsymbol{x}^T\boldsymbol{u} = 0$ and $P_{\boldsymbol{x}}^{\perp}\boldsymbol{u} = \boldsymbol{u}$. Multiplying the first row of (18) by $P_{\boldsymbol{x}}^{\perp}$ from the left and noting that $P_{\boldsymbol{x}}^{\perp}\boldsymbol{x} = \boldsymbol{0}$, we find that the left-hand side is given by

$$P_{\boldsymbol{x}}^{\perp}(H(\boldsymbol{x})\boldsymbol{u} - \beta\boldsymbol{x}) = P_{\boldsymbol{x}}^{\perp} H(\boldsymbol{x})\boldsymbol{u} = P_{\boldsymbol{x}}^{\perp} H(\boldsymbol{x}) P_{\boldsymbol{x}}^{\perp}\boldsymbol{u}.$$

In addition, one can easily check that $\boldsymbol{g}(\boldsymbol{x})$ is perpendicular to $\boldsymbol{x}$, so the right-hand side of the equality in (31) is $-P_{\boldsymbol{x}}^{\perp}\boldsymbol{g}(\boldsymbol{x}) = -\boldsymbol{g}(\boldsymbol{x})$ and (31) follows.

To prove the other direction, suppose $\boldsymbol{u}$ satisfies (31). So $\boldsymbol{u}^T\boldsymbol{x} = 0$ and $P_{\boldsymbol{x}}^{\perp}\boldsymbol{u} = \boldsymbol{u}$. Define $\beta = \boldsymbol{x}^T H(\boldsymbol{x})\boldsymbol{u}$ and write the left-hand side of (31) as

$$P_{\boldsymbol{x}}^{\perp} H(\boldsymbol{x}) P_{\boldsymbol{x}}^{\perp}\boldsymbol{u} = (I - \boldsymbol{x}\boldsymbol{x}^T) H(\boldsymbol{x}) P_{\boldsymbol{x}}^{\perp}\boldsymbol{u} = H(\boldsymbol{x})\boldsymbol{u} - \boldsymbol{x}\boldsymbol{x}^T H(\boldsymbol{x})\boldsymbol{u} = H(\boldsymbol{x})\boldsymbol{u} - \beta\boldsymbol{x}.$$

Since $-P_{\boldsymbol{x}}^{\perp}\boldsymbol{g}(\boldsymbol{x}) = -\boldsymbol{g}(\boldsymbol{x})$, it follows that $(\boldsymbol{u}, \beta)$ satisfies (18) as required.

Appendix C. Convergence of O–NCM. The proof of Theorem 4.4 is similar to that of Theorem 4.2 and makes use of the following auxiliary lemma.

LEMMA C.1. *Consider one update step of Algorithm* 2*, as in* (20)*, starting from an initial* $\boldsymbol{x} \in S_{n-1}$ *and ending with* $\boldsymbol{x}' = (\boldsymbol{x}+\boldsymbol{u})/\|\boldsymbol{x}+\boldsymbol{u}\| \in S_{n-1}$*. Let* $\alpha = \boldsymbol{x}^T\boldsymbol{x}^*$ *and* $\boldsymbol{u}^* = \alpha\boldsymbol{x}^* - \boldsymbol{x}$*. If* $\alpha \geq 1/2$ *and* $\|\boldsymbol{u}^* - \boldsymbol{u}\| \leq 1/4$*, then*

$$\|\boldsymbol{x}^* - \boldsymbol{x}'\| \leq \frac{2\|\boldsymbol{u}^* - \boldsymbol{u}\|}{\alpha - \|\boldsymbol{u}^* - \boldsymbol{u}\|} \leq 8\|\boldsymbol{u}^* - \boldsymbol{u}\|. \tag{32}$$

Proof. By definition,

$$\begin{aligned} \|\boldsymbol{x}^* - \boldsymbol{x}'\| &= \left\| \boldsymbol{x}^* - \frac{\boldsymbol{x}+\boldsymbol{u}}{\|\boldsymbol{x}+\boldsymbol{u}\|} \right\| = \left\| \boldsymbol{x}^* - \frac{\alpha\boldsymbol{x}^* - \boldsymbol{u}^* + \boldsymbol{u}}{\|\alpha\boldsymbol{x}^* - \boldsymbol{u}^* + \boldsymbol{u}\|} \right\| \\ &= \left\| \boldsymbol{x}^* \left(1 - \frac{\alpha}{\|\alpha\boldsymbol{x}^* - \boldsymbol{u}^* + \boldsymbol{u}\|}\right) + \frac{\boldsymbol{u}^* - \boldsymbol{u}}{\|\alpha\boldsymbol{x}^* - \boldsymbol{u}^* + \boldsymbol{u}\|} \right\|. \end{aligned} \tag{33}$$

Since $\|\boldsymbol{x}^*\| = 1$ and $\alpha > 0$, by the triangle inequality,

$$\alpha - \|\boldsymbol{u}^* - \boldsymbol{u}\| \leq \|\alpha\boldsymbol{x}^* - \boldsymbol{u}^* + \boldsymbol{u}\| \leq \alpha + \|\boldsymbol{u}^* - \boldsymbol{u}\|.$$

Applying the triangle inequality to (33), combined with the assumption $\|\boldsymbol{u}^* - \boldsymbol{u}\| \leq \alpha/2$,

$$\begin{aligned} \|\boldsymbol{x}^* - \boldsymbol{x}'\| &\leq \left| 1 - \frac{\alpha}{\|\alpha\boldsymbol{x}^* - \boldsymbol{u}^* + \boldsymbol{u}\|} \right| + \frac{\|\boldsymbol{u}^* - \boldsymbol{u}\|}{\|\alpha\boldsymbol{x}^* - \boldsymbol{u}^* + \boldsymbol{u}\|} \\ &\leq \left(\frac{\alpha}{\alpha - \|\boldsymbol{u}^* - \boldsymbol{u}\|} - 1 \right) + \frac{\|\boldsymbol{u}^* - \boldsymbol{u}\|}{\alpha - \|\boldsymbol{u}^* - \boldsymbol{u}\|} = \frac{2\|\boldsymbol{u}^* - \boldsymbol{u}\|}{\alpha - \|\boldsymbol{u}^* - \boldsymbol{u}\|}, \end{aligned}$$

hence concluding the proof. ∎

Proof of Theorem 4.4. We show that there exists an $\varepsilon > 0$ and a constant $C > 0$, such that for any initial point $\boldsymbol{x}_{(0)}$ that satisfies $\|\boldsymbol{x}_{(0)} - \boldsymbol{x}^*\| < \varepsilon$,

$$e_k = \frac{\|\boldsymbol{x}^* - \boldsymbol{x}_{(k+1)}\|}{\|\boldsymbol{x}^* - \boldsymbol{x}_{(k)}\|^2} < C \quad \forall k \geq 0.$$

We start by analyzing e_k at the first iteration $k = 0$. Let $\boldsymbol{u}_{(0)} = U_{\boldsymbol{x}_{(0)}} \boldsymbol{z}_{(0)}$ be the approximate correction of $\boldsymbol{u}^* = \alpha \boldsymbol{x}^* - \boldsymbol{x}_{(0)}$, given by the solution of (19). The new approximation of $\boldsymbol{x}^*$ is $\boldsymbol{x}_{(1)} = (\boldsymbol{x}_{(0)} + \boldsymbol{u}_{(0)})/\|\boldsymbol{x}_{(0)} + \boldsymbol{u}_{(0)}\|$. Since $\boldsymbol{u}^*$ is orthogonal to $\boldsymbol{x}^*$, the denominator of e_0 satisfies

$$\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\|^2 = \|\boldsymbol{x}^* - (\alpha \boldsymbol{x}^* - \boldsymbol{u}^*)\|^2 = (1-\alpha)^2 + \|\boldsymbol{u}^*\|^2 \geq \|\boldsymbol{u}^*\|^2. \tag{34}$$

To bound the numerator of e_0, assume for the moment that $\boldsymbol{x}_{(0)}$ is sufficiently close to $\boldsymbol{x}^*$ so that $\alpha = \boldsymbol{x}_{(0)}^T \boldsymbol{x}^* \geq 1/2$ and $\|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\| \leq \alpha/2$. Then, by Lemma C.1,

$$\|\boldsymbol{x}^* - \boldsymbol{x}_{(1)}\| \leq \frac{2\|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\|}{\alpha - \|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\|} \leq 8\|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\|. \tag{35}$$

Hence, it suffices to bound $\|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\|$. Define $\boldsymbol{z}^* = U_{\boldsymbol{x}_{(0)}}^T \boldsymbol{u}^*$ and note that since $\boldsymbol{x}^*$ and $\boldsymbol{u}^*$ are orthogonal, $\boldsymbol{x}_{(0)}^T \boldsymbol{u}^* = (\alpha \boldsymbol{x}^* - \boldsymbol{u}^*)^T \boldsymbol{u}^* = -\|\boldsymbol{u}^*\|^2$. Writing $I = U_{\boldsymbol{x}_{(0)}} U_{\boldsymbol{x}_{(0)}}^T + \boldsymbol{x}_{(0)} \boldsymbol{x}_{(0)}^T$, we thus have

$$\boldsymbol{u}^* = (U_{\boldsymbol{x}_{(0)}} U_{\boldsymbol{x}_{(0)}}^T + \boldsymbol{x}_{(0)} \boldsymbol{x}_{(0)}^T) \boldsymbol{u}^* = U_{\boldsymbol{x}_{(0)}} \boldsymbol{z}^* - \|\boldsymbol{u}^*\|^2 \boldsymbol{x}_{(0)}. \tag{36}$$

Since $\|\boldsymbol{x}_{(0)}\| = 1$,

$$\|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\| = \left\| U_{\boldsymbol{x}_{(0)}} \boldsymbol{z}^* - \|\boldsymbol{u}^*\|^2 \boldsymbol{x}_{(0)} - U_{\boldsymbol{x}_{(0)}} \boldsymbol{z}_{(0)} \right\| \leq \|\boldsymbol{z}^* - \boldsymbol{z}_{(0)}\| + \|\boldsymbol{u}^*\|^2. \tag{37}$$

To bound $\|\boldsymbol{z}^* - \boldsymbol{z}_{(0)}\|$, we view the exact system of nonlinear equations (17), whose solution is $\boldsymbol{u}^*$, as a perturbation of the approximate system of linear equations (19), whose solution is $\boldsymbol{u}_{(0)} = U_{\boldsymbol{x}_{(0)}} \boldsymbol{z}_{(0)}$. By (17), $\boldsymbol{u}^*$ solves the nonlinear equation

$$H(\boldsymbol{x}_{(0)}) \boldsymbol{u}^* = -\boldsymbol{g}(\boldsymbol{x}_{(0)}) + \beta^* \boldsymbol{x}_{(0)} + \tilde{\Delta}(\boldsymbol{x}_{(0)}, \boldsymbol{u}^*, \beta^*). \tag{38}$$

We multiply (38) by $U_{\boldsymbol{x}_{(0)}}^T$ from the left and plug in (36) to obtain the set of nonlinear equations in $\boldsymbol{z}^*$ (and $\beta^*, \boldsymbol{u}^*$),

$$H_p(\boldsymbol{x}_{(0)}) \boldsymbol{z}^* = -U_{\boldsymbol{x}_{(0)}}^T (\boldsymbol{g}(\boldsymbol{x}_{(0)}) - \tilde{\Delta}(\boldsymbol{x}_{(0)}, \boldsymbol{u}^*, \beta^*)). \tag{39}$$

Since $\boldsymbol{x}^*$ is γ-Newton-stable, the projected Hessian $H_p(\boldsymbol{x}^*)$ is full rank with smallest singular value $\sigma_{\min}(H_p(\boldsymbol{x}^*)) = \gamma > 0$. By the continuity of $\sigma_{\min}(H_p(\boldsymbol{x}))$ in $\boldsymbol{x}$, there exists an $\varepsilon_1 > 0$ such that $\sigma_{\min}(H_p(\boldsymbol{x}_{(0)})) \geq \gamma/2$ for all $\boldsymbol{x}$ with $\|\boldsymbol{x}^* - \boldsymbol{x}\| \leq \varepsilon_1$. In particular, if $\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| < \varepsilon_1$, then $H_p(\boldsymbol{x}_{(0)})$ is invertible and the solution to (39) satisfies the following implicit equation in $\boldsymbol{z}^*$,

$$\boldsymbol{z}^* = -H_p(\boldsymbol{x}_{(0)})^{-1} U_{\boldsymbol{x}_{(0)}}^T (\boldsymbol{g}(\boldsymbol{x}_{(0)}) - \tilde{\Delta}(\boldsymbol{x}_{(0)}, \boldsymbol{u}^*, \beta^*)).$$

Similarly, the unique solution to (19) is

$$\boldsymbol{z}_{(0)} = -H_p(\boldsymbol{x}_{(0)})^{-1} U_{\boldsymbol{x}_{(0)}}^T \boldsymbol{g}(\boldsymbol{x}_{(0)}).$$

Subtracting the last two equations gives

$$\|\boldsymbol{z}^* - \boldsymbol{z}_{(0)}\| \leq \|H_p(\boldsymbol{x}_{(0)})^{-1}\| \cdot \|\tilde{\Delta}(\boldsymbol{x}_{(0)}, \boldsymbol{u}^*, \beta^*)\| \leq \frac{2}{\gamma}\|\tilde{\Delta}(\boldsymbol{x}_{(0)}, \boldsymbol{u}^*, \beta^*)\|. \tag{40}$$

We bound the norm of $\tilde{\Delta}(\boldsymbol{x}_{(0)}, \boldsymbol{u}^*, \beta^*)$ in (16) by

$$\|\tilde{\Delta}(\boldsymbol{x}, \boldsymbol{u}^*, \beta^*)\| \leq |\beta^*| \cdot \|\boldsymbol{u}^*\| + \sum_{i=2}^{m-1} \tbinom{m-1}{i} \Big\|\mathcal{T}(I, \underbrace{\boldsymbol{x}, \dots, \boldsymbol{x}}_{m-i-1}, \underbrace{\boldsymbol{u}^*, \dots, \boldsymbol{u}^*}_{i})\Big\|. \tag{41}$$

To bound the terms in the sum, note that there exists an $M = M(\mathcal{T}) < \infty$ such that for any $\boldsymbol{x} \in S_{n-1}$, $\boldsymbol{u} \in \mathbb{R}^n$ and $j \leq m-1$,

$$\big\|\mathcal{T}(I, \underbrace{\boldsymbol{x}, \dots, \boldsymbol{x}}_{m-j-1}, \underbrace{\boldsymbol{u}, \dots, \boldsymbol{u}}_{j \text{ times}})\big\| \leq M\|\boldsymbol{u}\|^j. \tag{42}$$

Bounding each term in the sum in (41) by (42), there are at most m^2 terms involving $M\|\boldsymbol{u}^*\|^2$ and at most 2^m terms involving $M\|\boldsymbol{u}^*\|^i$ with $i \in \{3, \dots, m-1\}$. Assuming $\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| \leq m^2/2^m \leq 1$ and recalling that by (34), $\|\boldsymbol{u}^*\| \leq \|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\|$,

$$\sum_{i=2}^{m-1} \tbinom{m-1}{i} \Big\|\mathcal{T}(I, \underbrace{\boldsymbol{x}, \dots, \boldsymbol{x}}_{m-i-1}, \underbrace{\boldsymbol{u}^*, \dots, \boldsymbol{u}^*}_{i})\Big\| \leq m^2 M\|\boldsymbol{u}^*\|^2 + 2^m M\|\boldsymbol{u}^*\|^3 \leq 2m^2 M\|\boldsymbol{u}^*\|^2.$$

For the first term in (41), recalling the definition of β^* in (14),

$$\begin{aligned} |\beta^*| &= |\mu(\boldsymbol{x}) - \alpha^{m-2}\lambda^*| = |\mathcal{T}(\alpha\boldsymbol{x}^* - \boldsymbol{u}^*, \dots, \alpha\boldsymbol{x}^* - \boldsymbol{u}^*) - \alpha^{m-2}\lambda^*| \\ &\leq |\alpha^{m-2}\lambda^*(\alpha^2 - 1)| + \sum_{j=1}^{m} \alpha^{m-j} \tbinom{m}{j} \Big|\mathcal{T}(\underbrace{\boldsymbol{x}^*, \dots, \boldsymbol{x}^*}_{m-j}, \underbrace{\boldsymbol{u}^*, \dots, \boldsymbol{u}^*}_{j \text{ times}})\Big|. \end{aligned} \tag{43}$$

For the first term in (43), note that $\alpha^2 - 1 = \|\boldsymbol{u}^*\|^2$, $|\alpha| \leq 1$, and $|\lambda^*| \leq M$, hence

$$|\alpha^{m-2}\lambda^*(\alpha^2 - 1)| \leq M\|\boldsymbol{u}^*\|^2.$$

Since $\boldsymbol{x}^*$ is an eigenvector and $(\boldsymbol{u}^*)^T\boldsymbol{x}^* = 0$, all terms in the sum in (43) with $j = 1$ vanish:

$$\mathcal{T}(\boldsymbol{x}^*, \dots, \boldsymbol{x}^*, \boldsymbol{u}^*) = \lambda^*(\boldsymbol{u}^*)^T\boldsymbol{x}^* = 0.$$

Bounding each term in the sum in (43) with $j \geq 2$ by (42), there are, at most, m^2 terms involving $M\|\boldsymbol{u}^*\|^2$ and, at most, 2^m terms involving $M\|\boldsymbol{u}^*\|^j$ with $j \in \{3, \dots, m\}$. Since $\|\boldsymbol{u}^*\| \leq m^2/2^m$, the first term in (41) is thus bounded by

$$|\beta^*| \cdot \|\boldsymbol{u}^*\| \leq M\|\boldsymbol{u}^*\|^3 + 2^m M\|\boldsymbol{u}^*\|^3 \leq 2Mm^2\|\boldsymbol{u}^*\|^2.$$

It follows that

$$\|\tilde{\Delta}(\boldsymbol{x}_{(0)}, \boldsymbol{u}^*, \beta^*)\| \leq 4Mm^2\|\boldsymbol{u}^*\|^2.$$

Inserting this bound into (40),

$$\|\boldsymbol{z}^* - \boldsymbol{z}_{(0)}\| \leq \frac{8Mm^2}{\gamma}\|\boldsymbol{u}^*\|^2.$$

Inserting this into (37),

$$\|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\| \le \|\boldsymbol{z}^* - \boldsymbol{z}_{(0)}\| + \|\boldsymbol{u}^*\|^2 \le \Big(\frac{8Mm^2}{\gamma} + 1\Big)\|\boldsymbol{u}^*\|^2. \tag{44}$$

Note that if $\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| \le \varepsilon_2 = \min\{1, (4(\frac{8Mm^2}{\gamma} + 1))^{-1/2}\}$, then $\alpha \ge 1/2$. By (34), $\|\boldsymbol{u}^*\| \le \varepsilon_2$ as well. Thus, (44) implies $\|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\| \le 1/4$ as required by Lemma C.1. Under this condition, (35) implies

$$\|\boldsymbol{x}^* - \boldsymbol{x}_{(1)}\| \le 8\|\boldsymbol{u}^* - \boldsymbol{u}_{(0)}\| \le 8\Big(\frac{8Mm^2}{\gamma} + 1\Big)\|\boldsymbol{u}^*\|^2.$$

Combining the last two bounds, we obtain

$$e_0 = \frac{\|\boldsymbol{x}^* - \boldsymbol{x}_{(1)}\|}{\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\|^2} \le 8\Big(\frac{8Mm^2}{\gamma} + 1\Big). \tag{45}$$

As an interim summary, if $\|\boldsymbol{x}^* - \boldsymbol{x}_{(0)}\| \le \min\{\varepsilon_1, \varepsilon_2, m^2/2^m\} = \varepsilon_0$, then (45) holds. The rest of the proof follows by induction as in the proof of Theorem 4.2. ∎

Appendix D. Proof of Proposition 5.1. First, we prove an auxiliary lemma concerning the structure of the eigenvectors of $\mathcal{T}_\omega$. Recall $l = \lfloor n/2 \rfloor$. For any subset $\mathbb{A} \subseteq \{1, \dots, l\}$ define the following two n-dimensional vectors,

$$\mathbf{1}_{\mathbb{A}} = \sum_{i \in \mathbb{A}} \boldsymbol{e}_i \quad \text{and} \quad \mathbf{1}_{\mathbb{A}^c} = \sum_{i \notin \mathbb{A}} \boldsymbol{e}_i.$$

LEMMA D.1. *There is a function* $\alpha(\omega, |\mathbb{A}|) : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$ *such that all eigenvectors of* $\mathcal{T}_\omega$ *are of the form*

$$\boldsymbol{x}^*(\omega, \mathbb{A}) \propto \alpha(\omega, |\mathbb{A}|)\mathbf{1}_{\mathbb{A}} + \mathbf{1}_{\mathbb{A}^c}.$$

Proof. Let $\boldsymbol{x}^* = \sum_{i=1}^n \alpha_i \boldsymbol{e}_i$ be an eigenvector of $\mathcal{T}_\omega$ with eigenvalue λ^*. To prove the lemma it suffices to show that the coefficients $\alpha_1, \dots, \alpha_n$ can attain at most two distinct values. Applying mode product to $\mathcal{T}_\omega$ with $\boldsymbol{x}^*$,

$$\mathcal{T}_\omega(I, \boldsymbol{x}^*, \boldsymbol{x}^*) = \sum_{i=1}^n \alpha_i^2 \boldsymbol{e}_i + \omega\Big(\sum_{i=1}^n \alpha_i\Big)^2 \mathbf{1} = \sum_{i=1}^n (\alpha_i^2 + \omega\bar{\alpha}^2)\boldsymbol{e}_i, \tag{46}$$

where $\bar{\alpha} = \sum_{i=1}^n \alpha_i$. Since $(\boldsymbol{x}^*, \lambda^*)$ is an eigenpair, it satisfies

$$\sum_{i=1}^n (\alpha_i^2 + \omega\bar{\alpha}^2)\boldsymbol{e}_i = \lambda^* \sum_{i=1}^n \alpha_i \boldsymbol{e}_i. \tag{47}$$

Multiplying both sides of (47) with $\boldsymbol{e}_i^T$ gives

$$(\alpha_i^2 + \omega\bar{\alpha}^2) = \lambda^*\alpha_i \qquad \forall i \in \{1, \dots, n\}. \tag{48}$$

Subtracting (48) with $j \ne i$,

$$\alpha_i^2 - \alpha_j^2 = (\alpha_i + \alpha_j)(\alpha_i - \alpha_j) = \lambda^*(\alpha_i - \alpha_j).$$

We thus conclude that for any $j \ne i$, either $\alpha_j = \alpha_i$ or $\alpha_j = \lambda^* - \alpha_i$. It follows that the set $\{\alpha_1, \dots, \alpha_n\}$ contains up to two distinct values satisfying (48). ∎

The first part of Proposition 5.1 determines the number of real eigenvectors for $\mathcal{T}_\omega$. Following Lemma D.1, let $\boldsymbol{x}^* = \mathbf{1}_{\mathbb{A}^c} + \alpha \mathbf{1}_{\mathbb{A}}$ be proportional to some eigenvector of $\mathcal{T}_\omega$. By (46),

$$\mathcal{T}_\omega(I, \boldsymbol{x}^*, \boldsymbol{x}^*) = (\alpha^2 + \omega\bar{\alpha}^2)\mathbf{1}_{\mathbb{A}} + (1 + \omega\bar{\alpha}^2)\mathbf{1}_{\mathbb{A}^c}.$$

Since $\boldsymbol{x}^* = \alpha\mathbf{1}_{\mathbb{A}} + \mathbf{1}_{\mathbb{A}^c}$ is proportional to an eigenvector of $\mathcal{T}_\omega$,

$$\alpha = \frac{\alpha^2 + \omega\bar{\alpha}^2}{1 + \omega\bar{\alpha}^2},$$

or equivalently,

$$\alpha(1-\alpha) = \omega(1-\alpha)\bar{\alpha}^2, \tag{49}$$

One solution to (49) is $\alpha = 1$, which corresponds to the eigenvector $\boldsymbol{x} = \frac{1}{\sqrt{n}}\mathbf{1}$. For $\alpha \neq 1$ we replace $\bar{\alpha}$ with

$$\bar{\alpha} = \sum_{i=1}^{n} \alpha_i = \alpha|\mathbb{A}| + (n - |\mathbb{A}|).$$

The result is the following quadratic equation:

$$\omega|\mathbb{A}|^2\alpha^2 + (2\omega|\mathbb{A}|(n - |\mathbb{A}|) - 1)\alpha + \omega(n - |\mathbb{A}|)^2 = 0. \tag{50}$$

The solutions to (50) determine, up to a normalizing factor, the eigenvectors (both real and complex) of $\mathcal{T}_\omega$. Because of the problem's symmetry, we may charaterize *all* real eigenvectors by computing the solutions to (50) only for subsets $\mathbb{A}$ with $0 \le |\mathbb{A}| \le l$. Consider the discriminant $\mathcal{D}(\omega, |\mathbb{A}|)$ of the quadratic equation (50):

$$\mathcal{D}(\omega, |\mathbb{A}|) = (2\omega|\mathbb{A}|(n - |\mathbb{A}|) - 1)^2 - 4\omega^2|\mathbb{A}|^2(n - |\mathbb{A}|)^2 = 1 - 4\omega|\mathbb{A}|(n - |\mathbb{A}|).$$

For a given $\mathbb{A}$, the number of real solutions to (50) is

$$\begin{cases} 2, & \omega < \frac{1}{4|\mathbb{A}|(n-|\mathbb{A}|)}, \\ 1, & \omega = \frac{1}{4|\mathbb{A}|(n-|\mathbb{A}|)}, \\ 0, & \omega > \frac{1}{4|\mathbb{A}|(n-|\mathbb{A}|)}. \end{cases}$$

Hence, the number of real eigenpairs decreases at specific thresholds. The smallest threshold corresponds to $|\mathbb{A}| = l$ and is given by $\omega_0 = \frac{1}{4l(n-l)}$. When $\omega < \omega_0$, there are two real solutions to (50) for all subsets $1 \le |\mathbb{A}| \le l$. So the total number of solutions is equal to two times the number of distinct subsets,

$$N(\omega < \omega_0) = 1 + 2\sum_{i=1}^{l} \binom{n}{i} = 2^n - 1,$$

where we add one to account for $\frac{1}{\sqrt{n}}\mathbf{1}$, corresponding to $\mathbb{A} = \emptyset$. Note that this is also the bound on the number of eigenvectors of a generic cubic tensor; see [5]. When $\omega = \omega_0$, $\mathcal{D}(\omega_0, |\mathbb{A}| = l) = 0$. In this case, $N(\omega)$ is composed of two eigenvectors for all subsets $1 \le |\mathbb{A}| \le l - 1$ and one eigenvector for each subset of size $|\mathbb{A}| = l$,

$$N(\omega = \omega_0) = 1 + 2\sum_{j=1}^{l-1} \binom{n}{j} + \binom{n}{l}.$$

For $\omega_0 < \omega < \omega_1$, there are no real solutions of (50) for subsets of size $|\mathbb{A}| = l$. The number of real solutions is therefore

$$N(\omega_0 < \omega < \omega_1) = 1 + 2\sum_{j=1}^{l-1}\binom{n}{j}.$$

Repeating the argument for increasing values of ω, we obtain $N(\omega)$ as given in the proposition's statement.

We now prove the second part of the proposition, stating that at the thresholds $\omega_i = \frac{1}{4(l-i)(n-l+i)}$, $\binom{n}{l-i}$ of the eigenvectors are not Newton-stable. In this case, $\mathcal{D}(\omega_i, l-i) = 0$ and only one (real) solution to (50) exists for each $\mathbb{A}$ with $|\mathbb{A}| = l - i$. Solving (50) for $\omega = \omega_i$, we find that the $\binom{n}{l-i}$ eigenpairs $(\boldsymbol{x}^*(\mathbb{A}), \lambda^*(\mathbb{A}))$ with $|\mathbb{A}| = l-i$ are

$$\boldsymbol{x}^*(\mathbb{A}) = \sqrt{\frac{1}{n|\mathbb{A}|(n-|\mathbb{A}|)}}\big((n-|\mathbb{A}|)\mathbf{1}_{\mathbb{A}} + |\mathbb{A}|\mathbf{1}_{\mathbb{A}^c}\big), \qquad \lambda^*(\mathbb{A}) = \sqrt{\frac{n}{|\mathbb{A}|(n-|\mathbb{A}|)}}.$$

We show that each such eigenpair is not Newton-stable. To do so, we prove that the projected Hessian $H_p(\boldsymbol{x}^*)$ is rank deficient. First we compute the Hessian $H(\boldsymbol{x}^*)$. Abbreviate $b = \sqrt{\frac{1}{n|\mathbb{A}|(n-|\mathbb{A}|)}}$ and note that $\lambda^* = nb$. Then,

$$H(\boldsymbol{x}^*) = 2\mathcal{T}(I, I, \boldsymbol{x}^*) - \lambda^* I = b\left((n-2|\mathbb{A}|)\,\mathrm{diag}(\mathbf{1}_{\mathbb{A}}) + (2|\mathbb{A}|-n)\,\mathrm{diag}(\mathbf{1}_{\mathbb{A}^c}) + \mathbf{1}\mathbf{1}^T\right).$$

Consider the vector $\boldsymbol{v} = \mathbf{1}_{\mathbb{A}} - \mathbf{1}_{\mathbb{A}^c}$. A simple calculation yields $\boldsymbol{v}^T H(\boldsymbol{x}^*)\boldsymbol{v} = 0$. Since $\boldsymbol{v}$ is orthogonal to $\boldsymbol{x}^*$, $H_p(\boldsymbol{x}^*)$ is rank deficient and $(\boldsymbol{x}^*, \lambda^*)$ is not Newton-stable.

Appendix E. Convergence to eigenvectors which are not Newton-stable. In this section we present a detailed empirical study of the convergence properties of O–NCM. As discussed in section 4, the main property that governs the convergence of O–NCM to an eigenpair $(\boldsymbol{x}^*, \lambda^*)$ is the spectral structure of the projected Hessian at the eigenvector, $H_p(\boldsymbol{x}^*)$. As shown in Theorem 4.4, when $H_p(\boldsymbol{x}^*)$ is full rank, O–NCM converges in a quadratic rate to $\boldsymbol{x}^*$ given a sufficiently close initial point. When $\boldsymbol{x}^*$ is isolated but $1 \le \mathrm{rank}(H_p(\boldsymbol{x}^*)) < n$, the convergence rate may be less than quadratic. When $H_p(\boldsymbol{x}^*) = 0$ and/or $\boldsymbol{x}^*$ is nonisolated, full convergence to $\boldsymbol{x}^*$ is not always observed. These properties are summarized in Table 1.

We illustrate these convergence properties via two examples.

(a) Consider the tensor $\mathcal{T}$ with order $m = 3$ and dimensionality $n = 6$ of Example 5.8 in [5], corresponding to the homogeneous polynomial

$$\mu(\boldsymbol{x}) = x_1^4x_2^2 + x_1^2x_2^4 + x_3^6 - x_1^2x_2^2x_3^2.$$

This tensor has a total of 17 real eigenpairs. Six of them correspond to a $\lambda = 0$ eigenvalue, two of which are not Newton-stable with $\mathrm{rank}(H_p(\boldsymbol{x}^*)) = 2$. The

Table 1
O–NCM convergence properties to an eigenvector $\boldsymbol{x}^*$.

	$H_p(\boldsymbol{x}^*)$ full rank	$H_p(\boldsymbol{x}^*)$ rank deficient	$H_p(\boldsymbol{x}^*) = 0$
Isolated	Quadratic convergence	Slow convergence	No guarantees
Nonisolated	—	No guarantees	No guarantees

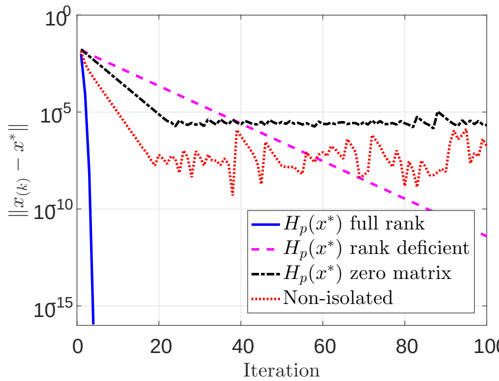

FIG. 9. *Convergence properties of O–NCM to eigenvectors with different stability properties.*

rest are Newton-stable. Figure 9 shows the value of $\|\boldsymbol{x}_{(k)} - \boldsymbol{x}^*\|$ as a function of the iteration k for one eigenvector that is Newton-stable and one that is not. While the convergence to the stable eigenvector is quadratic, the convergence to the point which is not Newton-stable point is much slower.

(b) Consider the tensor $\mathcal{T} \in \mathbb{R}^{6\times6\times6\times6}$ of example 6.4 in [20], corresponding to the homogeneous polynomial

$$\mu(\boldsymbol{x}) = \sum_{i=1}^{6} \sum_{j>i} (x_j - x_i)^4.$$

There are a total of 42 isolated eigenvectors, including one that corresponds to an eigenvalue $\lambda = 0$. The projected Hessian $H_p(\boldsymbol{x}^*)$ for this vector is equal to a zero matrix. As can be seen in Figure 9, in this case the O–NCM does not fully converge.

In addition, there are also infinitely many nonisolated eigenvectors corresponding to an eigenvalue $\lambda = 4.5$. The projected Hessian of these eigenpairs is a rank deficient (though nonzero) matrix. For example, any vector of the form

$$\boldsymbol{x}^* = [a,\ a,\ b,\ b,\ -(a+b),\ -(a+b)]^T, \qquad a, b \in \mathbb{R}, \tag{51}$$

is proportional to a nonisolated eigenvector. Note that the vectors corresponding to (51) form a two-dimensional subspace. Since these vectors are nonisolated, in this case we measure $\|(I_n - P_{\boldsymbol{x}^*})\boldsymbol{x}_{(k)}\|$ instead of $\|\boldsymbol{x}^* - \boldsymbol{x}_{(k)}\|$ where $P_{\boldsymbol{x}^*} \in R^{n\times n}$ is the projection matrix onto that subspace. As can be seen in Figure 9 in this case the O–NCM does not converge.

Trivial eigenvectors. In some cases, the tensor fibers are spanned by a low dimension subspace. Any vector orthogonal to this subspace is an eigenvector corresponding to an eigenvalue $\lambda = 0$, and a projected Hessian $H_p(\boldsymbol{x}^*)$ equal to a zero matrix. This is the case, for instance, in example (b), where all fibers are orthogonal to $\boldsymbol{x}^* = [1\ldots,1]^T$. As we have seen, this may cause the O–NCM to slow down, since the iterations do not converge to these points.

A simple preprocessing step is to find these eigenvectors by calculating the subspace of the tensor fibers, namely $\mathcal{T}_{:,i_2,\ldots,i_m}, i_2,\ldots,i_m \in [n]$. As a second step, the O–NCM can easily be constrained to that subspace.

Acknowledgments. We thank Lek–Heng Lim, Meirav Galun, Haim Avron, and Gal Binyamini for fruitful discussions. We also thank the anonymous referees for multiple constructive suggestions that significantly improved the manuscript.

REFERENCES

[1] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky, *Tensor decompositions for learning latent variable models*, J. Mach. Learn. Res., 15 (2014), pp. 2773–2832.

[2] A. Anandkumar, D. Hsu, and S. M. Kakade, *A method of moments for mixture models and hidden Markov models*, in Proceedings of the 25th conference on Learning Theory, Edinburgh, UK, 2012, pp. 33–1.

[3] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[4] P. Breiding, *The Average Number of Critical Rank-One-Approximations to a Symmetric Tensor*, preprint, arXiv:1701.07312, 2017.

[5] D. Cartwright and B. Sturmfels, *The number of eigenvalues of a tensor*, Linear Algebra Appl., 438 (2013), pp. 942–952, https://doi.org/10.1016/j.laa.2011.05.040.

[6] K. C. Chang, K. Pearson, and T. Zhang, *On eigenvalue problems of real symmetric tensors*, J. Math. Anal. Appl., 350 (2009), pp. 416–422, https://doi.org/10.1016/j.jmaa.2008.09.067.

[7] L. Chen, L. Han, and L. Zhou, *Computing tensor eigenvalues via homotopy methods*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 290–319, https://doi.org/10.1137/15M1010725.

[8] C.-F. Cui, Y.-H. Dai, and J. Nie, *All real eigenvalues of symmetric tensors*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1582–1601, https://doi.org/10.1137/140962292.

[9] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-1 and rank-$(R_1, R_2, \cdots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342, https://doi.org/10.1137/S0895479898346995.

[10] R. S. Dembo, S. C. Eisenstat, and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408, https://doi.org/10.1137/0719025.

[11] A. Gautier and M. Hein, *Tensor norm and maximal singular vectors of nonnegative tensors—a Perron-Frobenius theorem, a Collatz-Wielandt characterization and a generalized power method*, Linear Algebra Appl., 505 (2016), pp. 313–343, https://doi.org/10.1016/j.laa.2016.04.024.

[12] C.-H. Guo, W.-W. Lin, and C.-S. Liu, *A Modified Newton Iteration for Finding Nonnegative Z-Eigenpairs of a Nonnegative Tensor*, preprint, arXiv:1705.07487, 2017.

[13] L. Han, *An unconstrained optimization approach for finding real eigenvalues of even order symmetric tensors*, Numer. Algebra Control Optim., 3 (2013), pp. 583–599, https://doi.org/10.3934/naco.2013.3.583.

[14] C. L. Hao, C. F. Cui, and Y. H. Dai, *A sequential subspace projection method for extreme Z-eigenvalues of supersymmetric tensors*, Numer. Linear Algebra Appl., 22 (2015), pp. 283–298, https://doi.org/10.1002/nla.1949.

[15] C. J. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, J. ACM, 60 (2013), pp. 45:1–45:39, https://doi.org/10.1145/2512329.

[16] A. Jaffe, R. Weiss, S. Carmi, Y. Kluger, and B. Nadler, *Learning binary latent variable models: A tensor eigenpair approach*, in Proceedings of the International Conference on Machine Learning (ICML), to appear.

[17] E. Kofidis and P. A. Regalia, *On the best rank-1 approximation of higher-order supersymmetric tensors*, SIAM J. Matrix Anal. Appl., 23 (2001/02), pp. 863–884, https://doi.org/10.1137/S0895479801387413.

[18] T. G. Kolda and J. R. Mayo, *Shifted power method for computing tensor eigenpairs*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1095–1124, https://doi.org/10.1137/100801482.

[19] T. G. Kolda and J. R. Mayo, *An adaptive shifted power method for computing generalized tensor eigenpairs*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1563–1581, https://doi.org/10.1137/140951758.

[20] G. Li, L. Qi, and G. Yu, *The Z-eigenvalues of a symmetric tensor and its application to spectral hypergraph theory*, Numer. Linear Algebra Appl., 20 (2013), pp. 1001–1029, https://doi.org/10.1002/nla.1877.

[21] L.-H. Lim, *Singular values and eigenvalues of tensors: a variational approach*, in Proceedings of the 2005 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Puerto Vallarta, Mexico, IEEE, 2005, pp. 129–132.

[22] Y. Liu, G. Zhou, and N. F. Ibrahim, *An always convergent algorithm for the largest eigenvalue of an irreducible nonnegative tensor*, J. Comput. Appl. Math., 235 (2010), pp. 286–292, https://doi.org/10.1016/j.cam.2010.06.002.

[23] D. W. Marquardt, *An algorithm for least-squares estimation of nonlinear parameters*, J. Soc. Ind. Appl. Math., 11 (1963), pp. 431–441.

[24] M. Ng, L. Qi, and G. Zhou, *Finding the largest eigenvalue of a nonnegative tensor*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1090–1099, https://doi.org/10.1137/09074838X.

[25] Q. Ni, L. Qi, and F. Wang, *An eigenvalue method for testing positive definiteness of a multivariate form*, IEEE Trans. Automat. Control, 53 (2008), pp. 1096–1107, https://doi.org/10.1109/TAC.2008.923679.

[26] L. Qi, *Eigenvalues of a real supersymmetric tensor*, J. Symbolic Comput., 40 (2005), pp. 1302–1324, https://doi.org/10.1016/j.jsc.2005.05.007.

[27] L. Qi, Y. Wang, and E. X. Wu, *D-eigenvalues of diffusion kurtosis tensors*, J. Comput. Appl. Math., 221 (2008), pp. 150–157, https://doi.org/10.1016/j.cam.2007.10.012.

[28] T. Schultz, A. Fuster, A. Ghosh, R. Deriche, L. Florack, and L.-H. Lim, *Higher-Order Tensors in Diffusion Imaging*, in Visualization and Processing of Tensors and Higher Order Descriptors for Multi-valued Data, Springer, Basel, Switzerland, 2014, pp. 129–161.

[29] S. Smith, N. Ravindran, N. D. Sidiropoulos, and G. Karypis, *Splatt: Efficient and parallel sparse tensor-matrix multiplication*, in 2015 IEEE International Parallel and Distributed Processing Symposium, May 2015, pp. 61–70, https://doi.org/10.1109/IPDPS.2015.27.

[30] G. W. Stewart, *Matrix Algorithms. Vol.* II *: Eigensystems*, SIAM, Philadelphia, 2001, https://doi.org/10.1137/1.9780898718058.

[31] X. Zhang, C. Ling, and L. Qi, *The best rank*-1 *approximation of a symmetric tensor and related spherical optimization problems*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 806–821, https://doi.org/10.1137/110835335.