

On the spectral problem for trivariate functions

Behnam Hashemi · Yuji Nakatsukasa

Received: date / Accepted: date

Abstract Using a variational approach applied to generalized Rayleigh functionals, we extend the concepts of singular values and singular functions to trivariate functions defined on a rectangular parallelepiped. We also consider eigenvalues and eigenfunctions for trivariate functions whose domain is a cube. For a general finite-rank trivariate function, we describe an algorithm for computing the canonical polyadic (CP) decomposition, provided that the CP factors are linearly independent in two variables. All these notions are computed using Chebfun3; a part of Chebfun for numerical computing with 3D functions. Application in finding the best rank-1 approximation of trivariate functions is investigated. We also prove that if the function is analytic and two-way orthogonally decomposable (odeco), then the CP values decay geometrically, and optimal finite-rank approximants converge at the same rate.

Keywords Trivariate functions · Singular values · Higher order power method (HOPM) · Low-rank approximation · CP decomposition

Mathematics Subject Classification (2000) 65K10 · 15A18 · 15A69

1 Introduction

The singular value decomposition (SVD) of a bivariate function was first introduced by Schmidt [39]. He proved existence, uniqueness and the best low rank approxi-

The work of the first author is supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/20072013)/ERC grant agreement no. 291068. The views expressed in this article are not those of the ERC or the European Commission, and the European Union is not liable for any use that may be made of the information contained here. The second author is supported by JSPS as an Overseas Research Fellow.

B. Hashemi
Department of Mathematics, Shiraz University of Technology, Modarres BLVD., 71555-313 Shiraz , Iran
E-mail: hashemi@sutech.ac.ir

Y. Nakatsukasa

mation property of the SVD. Hammerstein and Lipschitz proved theorems regarding the uniform and absolute convergence of the singular functions. The name ‘‘singular value’’ actually comes from integral equations, i.e., in connection to the continuous SVD. For more on the history see [43] and [49]. Bounds on the decay rate of singular values of analytic bivariate functions are given in [25, p. 313] and [49]. Hille and Tamarkin [20] summarized different decay rates for eigenvalues of bivariate functions. See also [32, 33, 47] and [42, p. 379].

This paper considers the spectral problem for trivariate functions, by examining their singular values and eigenvalues. For the discrete case, a huge amount of work has been presented in the literature to discuss the spectral problems for tensors. Beginning from some of their continuous counterparts, we present the notions of singular values and eigenvalues of trivariate functions, discuss the connection to the canonical polyadic (CP) decomposition and prove decay rates for the CP values of certain trivariate functions. We also present numerical algorithms for numerically computing the aforementioned notions.

The spectral theory of discrete tensors was initiated by Lim [31] and Qi [35] independently. We start by extending Lim’s ideas to develop a spectral framework for trivariate functions. Our work involves functionals $\phi : L^2([a, b]) \times L^2([c, d]) \times L^2([e, l]) \rightarrow \mathbb{R}$. Assuming $g(x) \in L^2([a, b])$, $h(y) \in L^2([c, d])$, and $w(z) \in L^2([e, l])$, the action of ϕ on g, h and w is denoted by $\phi(g, h, w)$. Since Lim’s derivation is based on a variational approach, we first present the variational derivative for trivariate functions.

Definition 1.1 The functional (variational) derivative of ϕ is

$$\partial\phi := \begin{bmatrix} \partial_g\phi \\ \partial_h\phi \\ \partial_w\phi \end{bmatrix},$$

where $\partial_g\phi$, $\partial_h\phi$ and $\partial_w\phi$ denote the partial functional derivatives of ϕ with respect to g , h and w , respectively [17, p. 49]. Here, e.g., $\partial_g\phi$ is a function of x, y and z that satisfies

$$\phi(g + \delta g, h, w) - \phi(g, h, w) = \int_a^b \partial_g\phi(x, y, z) \delta g(x) dx + \mathcal{O}(\delta g^2).$$

Analogous formulas hold for $\partial_h\phi$ and $\partial_w\phi$. See e.g., [1, p. 224], [7], [15, p. 405] and [41].

Note that the kernel of the integral on the right-hand side defines the partial variational derivative of the functional.

Given $f : [a, b] \times [c, d] \times [e, l] \rightarrow \mathbb{R}$, the problem of finding the best rank-1 approximation of f is to compute the functions $g(x)$, $h(y)$, and $w(z)$ which have unit 2-norms and a scalar σ that solve

$$\min \int_a^b \int_c^d \int_e^l (f(x, y, z) - \sigma g(x) h(y) w(z))^2 dz dy dx. \quad (1.1)$$

Throughout the paper $\|\cdot\|$ is the spectral norm of discrete tensors and multivariate functions, i.e.,

$$\|f(x, y, z)\| := \sup \frac{\int_a^b \int_c^d \int_e^l f(x, y, z) g(x) h(y) w(z) dz dy dx}{\|g\|_2 \|h\|_2 \|w\|_2},$$

where $\|g\|_2^2 := \int_a^b |g(x)|^2 dx$, $\|h\|_2^2 := \int_c^d |h(y)|^2 dy$, and $\|w\|_2^2 := \int_e^l |w(z)|^2 dz$. As in the 2D case, the spectral norm turns out to be equal to the largest singular value, and its corresponding singular factors g, h, w give the best rank-1 approximation (1.1). See Section 2.

One of the extensions of the SVD to higher-order tensors is the canonical polyadic (CP) decomposition, which is a decomposition of a discrete tensor into a sum of rank-one tensors. The natural continuous analogue for trivariate functions is

$$f(x, y, z) = \sum_{i=1}^{\infty} \alpha_i g_i(x) h_i(y) w_i(z). \quad (1.2)$$

The minimal number of terms in (1.2) is called the rank in which case it is called the canonical polyadic (CP) decomposition or expansion of f . Even if f does not have a finite rank, in practice one only keeps a finite number of terms to approximate f . In this work we derive a practical numerical algorithm for computing an approximate finite-rank CP decomposition of trivariate functions.

As mentioned above, it is known that for bivariate functions the singular values decay geometrically if the function is analytic. We prove that for trivariate functions that are analytic and two-way odeco (which is less restrictive than (three-way) odeco), the CP-values α_i of admit a similar rate of decay, and best low-rank approximations can be obtained by truncating the CP decomposition.

This paper is organized as follows. In Section 2 we present the singular values and singular functions of trivariate functions. Section 3 deals with eigenvalues and eigenfunctions for trivariate functions of cube domain. In Section 4 we discuss the computation of the CP decomposition. Section 5 treats the decay of CP values for analytic trivariate functions that are (two-way) odeco, and their low-rank approximation. Section 6 presents numerical experiments to illustrate the notions and results.

2 The singular value problem

Here we consider the singular value problem for trivariate functions. We first review the counterpart for matrices and bivariate functions, then discrete order-three tensors, and then present the singular value problem for trivariate functions as a natural extension.

2.1 Review of the 2D case

The singular vectors of a discrete matrix $A \in \mathbb{R}^{m \times n}$ are the stationary points of the following functional

$$\begin{cases} r_A : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}, \\ r_A(u, v) := \frac{u^T A v}{\|u\|_2 \|v\|_2}. \end{cases} \quad (2.1)$$

Let $f(x, y)$ be a bivariate function defined on the rectangular domain $[a, b] \times [c, d]$. One way to define the singular values and singular functions of f is to consider the stationary points of the following continuous analog of (2.1):

$$\begin{cases} r_f : L^2([a, b]) \times L^2([c, d]) \rightarrow \mathbb{R} \\ r_f(g, h) := \frac{g^T f h}{\|g\|_2 \|h\|_2} = \frac{\int_a^b \int_c^d f(x, y) g(x) h(y) dy dx}{\|g\|_2 \|h\|_2}. \end{cases} \quad (2.2)$$

Closely related is the maximization problem

$$\max \int_a^b \int_c^d f(x, y) g(x) h(y) dy dx, \quad (2.3)$$

subject to the constraint that

$$\|g(x)\|_2^2 = \|h(y)\|_2^2 = 1.$$

To find a stationary point of the functional $r_f(g, h)$ we first write down the Lagrangian $L_f(g, h, \mu_1, \mu_2)$, or $L_f(g, h)$ for short, of the above constrained optimization problem

$$\begin{aligned} L_f(g, h) &:= \int_a^b \int_c^d f(x, y) g(x) h(y) dy dx \\ &\quad - \frac{\mu_1}{2} \int_a^b (g(x)^2 - 1) dx - \frac{\mu_2}{2} \int_c^d (h(y)^2 - 1) dy. \end{aligned}$$

It is clear that

$$L_f(g + \delta g, h) - L_f(g, h) = \int_a^b \int_c^d f(x, y) h(y) dy \delta g(x) dx - \mu_1 \int_a^b g(x) \delta g(x) dx + \mathcal{O}(\|\delta g\|^2).$$

Hence, the partial Frechét (or functional) derivative of L_f with respect to g is

$$\frac{\partial L_f}{\partial g}(x) = \int_c^d f(x, y) h(y) dy - \mu_1 g(x)$$

Writing down a similar formula for $\frac{\partial L_f}{\partial h}(y)$ and setting them equal to zero gives

$$\begin{cases} \int_c^d f(x, y) h(y) dy = \sigma g(x) \\ \int_a^b f(x, y) g(x) dx = \sigma h(y), \end{cases}$$

where $\sigma = \mu_1 = \mu_2 = r_f(g, h)$. Therefore, the singular functions of f satisfy the following two equations

$$\begin{cases} fh = \sigma g, \\ g^T f = \sigma h^T. \end{cases}$$

Once these equations are satisfied, the value of the functional r_f at (g, h) is called a singular value of f . A bivariate function can have an infinite number of singular values and singular functions. Putting each right singular functions h into a column of an infinite quasimatrix¹ V , each left singular function g into a column of an infinite

¹ Stewart [44] coined the name *quasimatrix* for a ‘matrix’ whose ‘columns’ are functions rather than

quasimatrix U , and the corresponding singular value into a diagonal of an infinite matrix Σ , we obtain the familiar SVD (see e.g., [25])

$$f(x, y) = U \Sigma V^T = \sum_{k=1}^{\infty} \sigma_k u_k(x) v_k(y).$$

In practice we usually work with bivariate functions $f(x, y)$ of finite rank (degenerate); this is possible because singular values of smooth functions decay rapidly, so they can be approximated well by finite-rank bivariate functions. This is a key observation in the construction of a `chebfun2` object [48].

2.2 3D case

The singular vectors of a given discrete tensor $A \in \mathbb{R}^{m \times n \times p}$ are the stationary points of the following functional [31], [16, p. 739]:

$$\begin{cases} r_A : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}, \\ r_A(u, v, w) := \frac{\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^p A(i, j, k) u(i) v(j) w(k)}{\|u\|_2 \|v\|_2 \|w\|_2}. \end{cases} \quad (2.4)$$

If u , v , and w are unit vectors, then from the gradient of $r_A(u, v, w)$ we get the nonlinear system of equations

$$\nabla r_A = \begin{bmatrix} A_{(1)}(w \otimes v) \\ A_{(2)}(w \otimes u) \\ A_{(3)}(v \otimes u) \end{bmatrix} - r_A(u, v, w) \begin{bmatrix} u \\ v \\ w \end{bmatrix} = 0. \quad (2.5)$$

Here $A_{(i)}$ is the mode- i unfolding [23, Sec. 2.4] of A and \otimes is the Kronecker product. The above equation can be solved by an iterative method like the power iteration. Applications include finding the nearest rank-1 tensor to A .

Next, let $f(x, y, z)$ be a trivariate function defined on the rectangular parallelepiped $[a, b] \times [c, d] \times [e, l]$. Analogous to (2.4), we define the singular values and singular functions of f as the stationary points of the generalized Rayleigh functional

$$\begin{cases} r_f : L^2([a, b]) \times L^2([c, d]) \times L^2([e, l]) \rightarrow \mathbb{R} \\ r_f(g, h, w) := \frac{\int_a^b \int_c^d \int_e^l f(x, y, z) g(x) h(y) w(z) dz dy dx}{\|g\|_2 \|h\|_2 \|w\|_2}. \end{cases} \quad (2.6)$$

As before, we consider

$$\max \int_a^b \int_c^d \int_e^l f(x, y, z) g(x) h(y) w(z) dz dy dx, \quad (2.7)$$

subject to the constraint that

$$\|g(x)\|_2^2 = \|h(y)\|_2^2 = \|w(z)\|_2^2 = 1. \quad (2.8)$$

vectors.

The Lagrangian $L_f(g, h, w)$ for the constraint optimization problem (2.7) is

$$\begin{aligned} \int_a^b \int_c^d \int_e^l f(x, y, z) g(x) h(y) w(z) dz dy dx - \frac{\mu_1}{2} \int_a^b (g(x)^2 - 1) dx \\ - \frac{\mu_2}{2} \int_c^d (h(y)^2 - 1) dy \\ - \frac{\mu_3}{2} \int_e^l (w(z)^2 - 1) dz. \end{aligned}$$

Since

$$\begin{aligned} L_f(g + \delta g, h, w) - L_f(g, h, w) &= \int_a^b \left(\int_c^d \int_e^l f(x, y, z) w(z) h(y) dz dy \right) \delta g(x) dx \\ &\quad - \mu_1 \int_a^b g(x) \delta g(x) dx + O(\|\delta g\|^2), \end{aligned}$$

the partial functional derivative of $L_f(g, h, w)$ with respect to g is

$$\int_c^d \int_e^l f(x, y, z) w(z) h(y) dz dy - \mu_1 g(x).$$

By computing the partial functional derivatives of L_f with respect to h and w and setting them to zero, we obtain the following system at a critical point:

$$\begin{cases} \int_c^d \int_e^l f(x, y, z) w(z) h(y) dz dy = \mu_1 g(x), \\ \int_a^b \int_e^l f(x, y, z) w(z) g(x) dz dx = \mu_2 h(y), \\ \int_a^b \int_c^d f(x, y, z) h(y) g(x) dy dx = \mu_3 w(z), \end{cases}$$

which together with the three equality constraints (2.8) form the KKT system. We now apply $g(x)$, $h(y)$, and $w(z)$ to the first, second and last equations above, respectively to get

$$\mu_1 = \mu_2 = \mu_3 = \int_a^b \int_c^d \int_e^l f(x, y, z) g(x) h(y) w(z) dz dy dx = r_f(g, h, w).$$

Setting $\sigma := r_f(g, h, w)$ we have

$$\begin{cases} \int_c^d \int_e^l f(x, y, z) w(z) h(y) dz dy = \sigma g(x), \\ \int_a^b \int_e^l f(x, y, z) w(z) g(x) dz dx = \sigma h(y), \\ \int_a^b \int_c^d f(x, y, z) h(y) g(x) dy dx = \sigma w(z), \\ \int_a^b \int_c^d \int_e^l f(x, y, z) g(x) h(y) w(z) dz dy dx = \sigma. \end{cases} \quad (2.9)$$

We call (σ, g, h, w) satisfying (2.8) and (2.9) a *singular quadruplet* of f . Note from (2.9) that (σ, g, h, w) is a singular quadruplet of f if and only if the same is true for $(-\sigma, -g, -h, -w)$. We do not distinguish these and assume $\sigma \geq 0$ unless otherwise mentioned.

The singular quadruplet $(\sigma_1, g_1, h_1, w_1)$ with the largest σ_1 is called the largest singular quadruplet. Below is a natural continuous analogue of the well-known best rank-1 approximation property of the largest singular quadruplets for discrete tensors.

Theorem 2.1 *The largest singular quadruplet $(\sigma_1, g_1, h_1, w_1)$ is the best rank-1 approximation to f , i.e., it is the solution for (1.1).*

Proof As shown in [54] for the discrete case, we can verify that when g, h , and w are fixed, the objective function in (1.1) under the constraints (2.8) is minimized when $\sigma = r_f(g, h, w)$, with the minimum value

$$\min \int_a^b \int_c^d \int_e^l (f(x, y, z))^2 dz dy dx - (r_f(g, h, w))^2. \quad (2.10)$$

Since this is clearly minimized when $|r_f(g, h, w)|$ is maximized, it follows that (1.1) is equivalent to the dual problem of maximizing the generalized Rayleigh functional (2.7).

2.3 Algorithms for computing a singular value

The higher order power method (HOPM) for discrete tensors is a generalization of the power method for discrete matrices [9, 54]. Each iteration of the HOPM for an order three discrete tensor has three steps. At each step one of the three singular vectors is updated while the remaining two are kept unchanged and this process continues alternatively in the next two steps of each iteration. It is shown in [36] that the sequence of values of the functional (2.4) generated in this way is monotonic and therefore the HOPM converges to a singular value². See also [52] for further references and a recent convergence analysis of the singular factors.

Algorithm 1 is a continuous extension of the HOPM based on (2.9) that can be used to compute a singular quadruplet of $f(x, y, z)$. Essentially it is obtained by replacing tensor-vector products with integration. As in the discrete case [52], it can be viewed as an alternating least-squares algorithm.

Algorithm 1 HOPM for 3D functions

1: Initialize the three functions $g(x), h(y), w(z)$.

2: Repeat until convergence:

- $\hat{g}(x) := \int_c^d \int_e^l f(x, y, z) w(z) h(y) dz dy$, $g(x) = \hat{g}(x) / \|\hat{g}(x)\|$,
 - $\hat{h}(y) := \int_a^b \int_e^l f(x, y, z) w(z) g(x) dz dx$, $h(y) = \hat{h}(y) / \|\hat{h}(y)\|$,
 - $\hat{w}(z) := \int_a^b \int_c^d f(x, y, z) h(y) g(x) dy dx$, $w(z) = \hat{w}(z) / \|\hat{w}(z)\|$,
 - $\sigma = \int_a^b \int_c^d \int_e^l f(x, y, z) w(z) h(y) g(x) dz dy dx$.
-

The classical convergence analysis for the singular value σ can be extended to HOPM for (trivariate) functions as follows: for fixed h and g , consider the functional $(\sigma(w) =) \int_a^b \int_c^d \int_e^l f(x, y, z) w(z) h(y) g(x) dz dy dx$. By the Cauchy-Schwartz inequality, $\sigma(w)$ is maximized over $\|w\| = 1$ by taking w to be a positive scalar multiple of $\int_a^b \int_c^d f(x, y, z) h(y) g(x) dy dx$, which is precisely the w obtained by the update in HOPM. Therefore each HOPM update results in a monotonic nondecrease of σ , hence σ converges to a local maximum.

² Unless in the exceptional case that the initial vectors correspond to a saddle point of (2.4) [21, Thm. 2].

It is easy to implement Algorithm 1 in Chebfun. A chebfun3 object [19] constructed from a 3D function f has the Tucker representation

$$f(x, y, z) \approx T \times_1 A(x) \times_2 B(y) \times_3 C(z), \quad (2.11)$$

where T is a discrete tensor of size $r_x \times r_y \times r_z$, $A(x)$ is a quasimatrix of size $\infty \times r_x$ defined over $[a, b]$, $B(y)$ is a quasimatrix of size $\infty \times r_y$ over $[c, d]$ and $C(z)$ is a quasimatrix of size $\infty \times r_z$ defined over $[e, g]$. The above formula involves modal contractions of the core tensor T and the quasimatrices $A(x)$, $B(y)$ and $C(z)$ and can be rewritten as

$$f(x, y, z) \approx \sum_{i=1}^{r_x} \sum_{j=1}^{r_y} \sum_{k=1}^{r_z} t_{i,j,k} a_i(x) b_j(y) c_k(z). \quad (2.12)$$

Either (r_x, r_y, r_z) or $r := \max\{r_x, r_y, r_z\}$ is called the numerical Tucker rank³ of f . Assuming that the number of Chebyshev coefficients needed to represent each univariate function $a_i(x)$, $b_j(y)$, and $c_k(z)$ is m, n and p , respectively, either (m, n, p) or $\max\{m, n, p\}$ is called the *length* of a chebfun3 object. Assuming for notational simplicity that $n = \max\{m, n, p\}$, the cost of constructing a chebfun3 object from $f(x, y, z)$ is $\mathcal{O}(nr^4 + rn \log n)$ [19].

We claim that the cost of each iteration of Algorithm 1 is $\mathcal{O}(r^3 + rn \log n)$. This can be verified as follows. The dominant cost of each iteration is the computation of the three double integrals. For example the first one can be computed as follows.

$$\hat{g}(x) = T \times_1 A(x) \times_2 \underbrace{\left(\int_c^d B(y) h(y) dy \right)}_{\text{an } r \times 1 \text{ vector}} \times_3 \underbrace{\left(\int_e^l C(z) w(z) dz \right)}_{\text{another } r \times 1 \text{ vector}}$$

Each column of $B(y)$ has length n and the inner product of two 1D chebfuns using Clenshaw-Curtis quadrature needs $\mathcal{O}(n \log n)$ operations [51, Chap. 19] (the initial functions in step 1 are assumed to have lengths bounded by n ; then the degrees are bounded by $n - 1$ throughout the iterations). Since $B(y)$ has r columns, applying $B(y)$ to $h(y)$ costs $\mathcal{O}(rn \log n)$. Similarly, the second integral above needs $\mathcal{O}(rn \log n)$ arithmetic operations. The modal multiplication of the $r \times r \times r$ core tensor T with an $r \times 1$ discrete vector needs $\mathcal{O}(r^3)$ operations and produces a matrix of size $r \times r$. We then multiply the matrix with the vector resulting from the second integral above. Finally the resulting vector is used to form a linear combination of the columns of $A(x)$ to compute $\hat{g}(x)$. Hence the overall cost is $\mathcal{O}(r^3 + rn \log n)$.

Therefore, the cost of forming a chebfun3 object plus that of using Algorithm 1 in Chebfun is quartic in r and only log-linear in the length of the chebfun3 object. It is therefore clear that Algorithm 1 is particularly efficient for chebfun3 object that are already constructed from low rank functions. Note that the cost of each iteration of HOPM for discrete full tensors of size $n \times n \times n$ is $\mathcal{O}(n^3)$ since this is the cost of tensor-vector contractions. The cost reduces to $\mathcal{O}(r^3 + rn)$ provided that the Tucker decomposition of the discrete tensor is already available; this is comparable to the cost of working with a chebfun3 object.

³ relative to a tolerance to which (2.12) holds; machine epsilon by default

If we apply the continuous analogue of the higher order SVD [10] to the above Tucker representation of f , then we get

$$f(x, y, z) \approx \hat{f}(x, y, z) = \hat{T} \times_1 \hat{A}(x) \times_2 \hat{B}(y) \times_3 \hat{C}(z), \quad (2.13)$$

in which the columns of the factor quasimatrices $\hat{A}(x)$, $\hat{B}(y)$ and $\hat{C}(z)$ are orthonormal and the discrete core tensor \hat{T} is all orthogonal, i.e., its horizontal, lateral and frontal slices are orthogonal. An alternative to Algorithm 1 for computing a singular quadruplet of a chebfun3 object is as follows. Using Chebfun's command `hosvd` we compute the higher-order SVD of the chebfun3 object where the factor quasimatrices have unit 2-norm. Then we simply apply the discrete version of HOPM, based on (2.5), to the core tensor \hat{T} of the HOSVD of f in (2.13). See Algorithm 2.

Algorithm 2 HOSVD-based computation of a singular quadruplet of f

- 1: Let f be a trivariate function and \hat{f} be its HOSVD as in (2.13).
 - 2: Apply the HOPM to the discrete tensor \hat{T} to compute a singular value σ and the corresponding discrete singular vectors u, v , and w . See (2.5).
 - 3: Compute the singular functions $g(x) = \hat{A}(x)u$, $h(y) = \hat{B}(y)v$ and $w(z) = \hat{C}(z)w$.
-

In Step 2 of Algorithm 2, the discrete HOPM can be replaced by any other algorithm that computes the singular values and singular vectors of a discrete tensor. Assuming that we first construct a chebfun3 object from the input function handle in both Algorithm 1 and Algorithm 2, the latter has the practical benefit that one can directly work with a discrete tensor which can be more efficient than manipulating continuous functions.

3 The eigenvalue problem

Eigenvalue problems are defined for matrices that are square. Analogously, for the eigenvalue problem to make sense, the domain of f has to be a cube $[a, b]^3$.

Definition 3.1 $f(x, y, z)$ is called a square function, if its domain is a cube, i.e., if x, y , and z are all defined on a single interval $[a, b]$, i.e., $f : [a, b] \times [a, b] \times [a, b] \rightarrow \mathbb{R}$.

We first consider the case where f is symmetric and then treat the general non-symmetric case. Different types of eigenvalues are defined for a discrete tensor. Our definitions here correspond to the Z-eigenvalues of discrete tensors as in [31, 46]. Such tensor eigenvalue problems have applications for example in random walks [2, 29], blind source separation [21] and MRI [40], and below we present its analogue in infinite dimensions for order three tensors.

3.1 Symmetric functions

Definition 3.2 A square function $f(x, y, z)$ is called symmetric if it is invariant under any permutation of variables, i.e., if $f(x, y, z) = f(y, x, z) = f(z, y, x) = f(x, z, y)$.

Definition 3.3 (λ, g) is called an eigenpair of $f(x, y, z)$ if

$$\begin{cases} \int_a^b \int_a^b f(x, y, z) g(y) g(z) dz dy = \lambda g(x), \\ \int_a^b g(x)^2 dx = 1. \end{cases} \quad (3.1)$$

Taking advantage of the symmetry results in the symmetric HOPM that is slightly cheaper but might not converge to an eigenvalue. This motivated Kolda and Mayo [24] to consider a *shifted* symmetric higher-order power method (SS-HOPM), which is guaranteed to converge to a stable eigenpair of discrete tensors of any order. Algorithm 3 is an extension of the SS-HOPM for trivariate functions. It uses a shift parameter s whose appropriate choice guarantees convergence to an eigenvalue of f .

Algorithm 3 SS-HOPM for symmetric 3D functions with shift s

- 1: Initialize $g(z)$ to be a function defined on $[a, b]$ with unit 2-norm.
 - 2: Repeat until convergence:
 - Compute $\hat{g}(x) := \int_a^b \int_a^b f(x, y, z) g(y) g(z) dy dz$,
 - Compute $\beta := \|\hat{g} + sg(x)\|_2$ and $g(x) = \hat{g}(x)/\beta$,
 - 3: $\lambda := \beta - s$ is an eigenvalue and $g(x)$ is the corresponding eigenfunction.
-

Let $f(x, y, z)$ be a symmetric chebfun3 of length n and numerical Tucker rank r and $g(x)$ be a 1D chebfun of length n . Similar to the discussion of the complexity of Algorithm 1, we can show that each iteration of the above continuous analogue of the SS-HOPM costs $\mathcal{O}(r^3 + rn \log n)$ operations. Recall that the SS-HOPM for discrete full tensors of size $n \times n \times n$ requires $\mathcal{O}(n^3)$ arithmetic operations (as before, this reduces to $\mathcal{O}(r^3 + rn)$ if a Tucker decomposition is available).

3.2 Nonsymmetric functions

A bivariate function defined on a square has left and right eigenfunctions which are not necessarily the same (unless the function is symmetric). Equivalently, the nonsymmetric 3D function $f(x, y, z)$ has modal eigenpairs, i.e., there are x -eigenpairs, y -eigenpairs, and z -eigenpairs.

Definition 3.4 $(\mu_1, g(x))$ is called an x -eigenpair of $f(x, y, z)$ if

$$\begin{cases} \int_a^b \int_a^b f(x, y, z) g(y) g(z) dz dy = \mu_1 g(x), \\ \int_a^b g(x)^2 dx = 1. \end{cases} \quad (3.2)$$

A y -eigenpair $(\mu_2, h(y))$ and a z -eigenpair $(\mu_3, w(z))$ can be defined analogously. It is easy to see from (3.2) that $(\mu_1, g(x))$ is an x -eigenpair if and only if $(-\mu_1, -g(x))$ is an x -eigenpair. (3.1) implies that the same relation holds for eigenpairs of symmetric functions.

4 CP decomposition of trivariate functions

Given a trivariate function $f(x, y, z)$, in this section we consider the problem of finding a CP decomposition (1.2). This is one of the canonical factorizations of a tensor of

order > 2 . We note that the construction of Chebfun3 [19] is based on the so-called Tucker decomposition, which is of a different nature. CP-type decompositions are explored in the literature as a tool to combat the curse of dimensionality for approximating multivariate functions, most notably in [3,4]. See also [6] for a least-square approach exploiting sparsity. Let us emphasize also that our main problem in this paper is not function approximation. Our interest here is more on the *connections* between the spectral problem and CP decomposition for trivariate functions. Even though our results can perhaps be extended to higher dimensional functions, we have focused on 3D functions for which Chebfun3 is available as a convenient exploratory numerical environment.

Generally, computing the CP decomposition of a discrete tensor is NP-hard. For example, an $n \times n \times n$ tensor can have a CP rank $\mathcal{O}(n^2)$: a trivial upper bound is n^2 , and the bound $2\lceil \frac{n^3}{3n-2} \rceil \approx \frac{2}{3}n^2$ is given in [5, Ex. 12], building upon [27,30,45]. To obtain a computationally tractable algorithm for CP decomposition of trivariate functions, we shall invoke a key tool in the theory of computing the CP decomposition of a three-way tensor. Namely, it is sufficient for the direct computation of the CP decomposition

$$T = \sum_{\ell=1}^r \mathbf{a}_\ell \circ \mathbf{b}_\ell \circ \mathbf{c}_\ell, \quad (4.1)$$

of a discrete tensor T of size $m \times n \times \ell$ that the following two conditions are satisfied: (i) $r \leq \min(m, n, \ell)$ (or a slightly weaker condition), and (ii) two of the matrices $A = [\mathbf{a}_1, \dots, \mathbf{a}_r]$, $B = [\mathbf{b}_1, \dots, \mathbf{b}_r]$ and $C = [\mathbf{c}_1, \dots, \mathbf{c}_r]$ have full column rank. Specifically, the CP can be computed exactly via a matrix decomposition, namely a simultaneous diagonalization of the slices, which in turn can be computed via a generalized eigenvalue problem.

The essence is that two random slices will have a generalized eigenvalue decomposition where the matrices of left and right eigenvector have columns equal to the factors in the CP decomposition. This fact was reported in [38], and independently in [28], and this idea has been refined and generalized in [8, 11]. In [13] it is shown that the CP decomposition can be computed using standard linear algebra provided that Kruskal's condition [26] (which guarantees uniqueness of CP decomposition) is satisfied. Since the algorithms in these recent papers are rather intricate, here we consider the continuous analogues of the classical result [28],[38].

This result has remained of relatively limited applicability in the context of discrete tensors, primarily due to the condition "CP rank $r \leq \min(m, n, \ell)$ " being too restrictive. However, consider our main subject of trivariate functions. Here, the dimension, if any, is $\infty \times \infty \times \infty$. Crucially, this means that the above requirement reduces merely to that the trivariate function is of *finite rank*. This is not much of a restriction, since in practice, we usually work with functions of finite rank (this is surely the case in Chebfun3). This suggests that perhaps for a wide class of trivariate functions of practical interest, the CP decomposition could be computed via linear algebra algorithms, provided that the linear independence condition holds among the CP factors (this does pose a practical limitation in the applicability).

We describe the process for trivariate functions. Let us introduce the *factor quasi-matrices* $\mathbf{G}(x) = [g_1(x), \dots, g_r(x)] \in \mathbb{R}^{\infty \times r}$, $\mathbf{H}(y) = [h_1(y), \dots, h_r(y)] \in \mathbb{R}^{\infty \times r}$, and

$\mathbf{W}(z) = [w_1(z), \dots, w_r(z)] \in \mathbb{R}^{\infty \times r}$. Then, we can write f in (1.2) as

$$C \times_1 \mathbf{G}(x) \times_2 \mathbf{H}(y) \times_3 \mathbf{W}(z), \quad (4.2)$$

where C is an $r \times r \times r$ diagonal core tensor with i th diagonal element α_i .

We now assume that $\mathbf{G}(x)$ and $\mathbf{H}(y)$ have full column rank r . The $z = z_0$ slice $f(x, y, z_0)$, for any $z_0 \in [e, l]$, is the bivariate function

$$f(x, y, z_0) = \mathbf{H}(y) \begin{bmatrix} \alpha_1 w_1(z_0) & & & \\ & \alpha_2 w_2(z_0) & & \\ & & \ddots & \\ & & & \alpha_r w_r(z_0) \end{bmatrix} \mathbf{G}(x)^T. \quad (4.3)$$

Another slice $z = z_1$ gives

$$f(x, y, z_1) = \mathbf{H}(y) \begin{bmatrix} \alpha_1 w_1(z_1) & & & \\ & \alpha_2 w_2(z_1) & & \\ & & \ddots & \\ & & & \alpha_r w_r(z_1) \end{bmatrix} \mathbf{G}(x)^T. \quad (4.4)$$

From these we see that $(f(x, y, z_0), f(x, y, z_1))$, regarded as a continuous analogue of matrix pairs, has the eigenvalue equations

$$f(x, y, z_1)(\mathbf{G}(x)^T)^\dagger = f(x, y, z_0)(\mathbf{G}(x)^T)^\dagger \text{diag}\left(\frac{w_1(z_1)}{w_1(z_0)}, \dots, \frac{w_r(z_1)}{w_r(z_0)}\right). \quad (4.5)$$

Here $(\mathbf{G}(x)^T)^\dagger \in \mathbb{R}^{\infty \times r}$ is the quasimatrix pseudoinverse of $\mathbf{G}(x)^T$ such that $\mathbf{G}(x)^T (\mathbf{G}(x)^T)^\dagger = I_r$ and $\sigma_i((\mathbf{G}(x)^T)^\dagger) = 1/\sigma_i(\mathbf{G}(x))$. The pair $(f(x, y, z_0), f(x, y, z_1))$ also has the analogous left eigenvalue equation

$$\mathbf{H}(y)^\dagger f(x, y, z_1) = \text{diag}\left(\frac{w_1(z_1)}{w_1(z_0)}, \dots, \frac{w_r(z_1)}{w_r(z_0)}\right) \mathbf{H}(y)^\dagger f(x, y, z_0). \quad (4.6)$$

As in standard matrix theory [12, §. 4.5], the equations (4.3) and (4.4) also imply the simultaneous diagonalization property

$$\begin{aligned} \mathbf{H}(y)^\dagger f(x, y, z_0)(\mathbf{G}(x)^T)^\dagger &= \begin{bmatrix} \alpha_1 w_1(z_0) & & & \\ & \ddots & & \\ & & \alpha_r w_r(z_0) & \\ & & & \end{bmatrix}, \\ \mathbf{H}(y)^\dagger f(x, y, z_1)(\mathbf{G}(x)^T)^\dagger &= \begin{bmatrix} \alpha_1 w_1(z_1) & & & \\ & \ddots & & \\ & & \alpha_r w_r(z_1) & \\ & & & \end{bmatrix}. \end{aligned} \quad (4.7)$$

In order to obtain the decomposition (4.7) and hence the desired factors $\mathbf{G}(x)$ and $\mathbf{H}(y)$ in practice, we reduce the task to a matrix eigenvalue problem. First, compute the SVD of the bivariate functions

$$f(x, y, z_0) = \mathbf{U}_0(y) \mathbf{S}_0 \mathbf{V}_0(x)^T, \quad f(x, y, z_1) = \mathbf{U}_1(y) \mathbf{S}_1 \mathbf{V}_1(x)^T. \quad (4.8)$$

By (4.3) and (4.4), unless $w_i(z_j) = 0$ for some (i, j) , \mathbf{U}_0 and \mathbf{U}_1 have the same column space spanned by those of $\mathbf{H}(y)$. In practice, we compute the numerical column space by finding the nonnegligible singular values larger than a tolerance, and find the corresponding columns $\hat{\mathbf{U}}_0(y)$ of $\mathbf{U}_0(y)$ (its first \hat{r} columns). We do the same to the row space by extracting the first \hat{r} rows $\hat{\mathbf{V}}_0(x)^T$ of $\mathbf{V}_0(x)^T$, and project both $f(x, y, z_0)$ and $f(x, y, z_1)$ onto the spaces to obtain two matrices of size $\hat{r} \times \hat{r}$:

$$\begin{aligned} A_0 &= \hat{\mathbf{U}}_0(y)^T f(x, y, z_0) \hat{\mathbf{V}}_0(x) = (\hat{\mathbf{U}}_0(y)^T \mathbf{U}_0(y)) S_0(\mathbf{V}_0(x)^T \hat{\mathbf{V}}_0(x)), \\ A_1 &= \hat{\mathbf{U}}_0(y)^T f(x, y, z_1) \hat{\mathbf{V}}_0(x) = (\hat{\mathbf{U}}_0(y)^T \mathbf{U}_1(y)) S_1(\mathbf{V}_1(x)^T \hat{\mathbf{V}}_0(x)). \end{aligned} \quad (4.9)$$

One then finds the generalized eigenvalue decomposition for the matrix pair (A_0, A_1) as

$$Y^T A_0 X = \Lambda_0, \quad Y^T A_1 X = \Lambda_1 \quad (4.10)$$

where Λ_0, Λ_1 are diagonal matrices. We then obtain the right eigenfunctions in (4.5), (4.6) as $(\hat{\mathbf{G}}(x)^T)^{(\dagger)} := \hat{\mathbf{V}}_0(x) X$ (which approximates $(\mathbf{G}(x)^T)^\dagger$) and the left eigenfunctions as $\hat{\mathbf{H}}(y)^{(\dagger)} := Y^T \hat{\mathbf{U}}_0(y)^T$ (which approximates $\mathbf{H}(y)^\dagger$). We will also use approximants $\hat{\mathbf{G}}(x), \hat{\mathbf{H}}(y)$ to $\mathbf{G}(x), \mathbf{H}(y)$, which are obtained as

$$\hat{\mathbf{G}}(x) = \hat{\mathbf{V}}_0(x) X^{-T}, \quad \hat{\mathbf{H}}(y) = \hat{\mathbf{U}}_0(y) Y^{-T}. \quad (4.11)$$

Finally, to obtain an approximant to $\mathbf{W}(z)$, we see from (4.2) that

$$c_i w_i(z) = \int_a^b \int_c^d f(x, y, z) g_i^\dagger(x) h_i^\dagger(y) dy dx. \quad (4.12)$$

Here $g_i^\dagger(x), h_i^\dagger(y)$ are the i th column of the quasimatrices $\mathbf{G}(x)^\dagger, \mathbf{H}(y)^\dagger$, which we approximate via those of $\hat{\mathbf{G}}(x)^{(\dagger)}, \hat{\mathbf{H}}(y)^{(\dagger)}$. We can retrieve $c_i, w_i(z)$ from (4.12) simply by the normalization $\|w_i(z)\| = 1$, that is, $c_i = \|\int_a^b \int_c^d f(x, y, z) g_i(x) h_i(y) dy dx\|$, $w_i(z) = \frac{1}{c_i} \int_a^b \int_c^d f(x, y, z) g_i(x) h_i(y) dy dx$. Putting these together, we obtain an approximate CP decomposition $f \approx \sum_{i=1}^r \alpha_i \hat{g}_i(x) \hat{h}_i(y) \hat{w}_i(z)$. We summarize the process in Algorithm 4.

Algorithm 4 CP decomposition for 3D function $f : [a, b] \times [c, d] \times [e, l] \rightarrow \mathbb{R}$ for which the CP factors in the x, y -directions are linearly independent

- 1: Take random slices $f(x, y, z_0)$ and $f(x, y, z_1)$ for $z_0, z_1 \in [e, l]$.
 - 2: Compute the SVD (4.8), and define $\hat{\mathbf{U}}_0(y)$ and $\hat{\mathbf{V}}_0(x)^T$ as the singular functions corresponding to the singular values larger than a prescribed tolerance.
 - 3: Compute discrete matrices A_0 and A_1 as in (4.9).
 - 4: Find the generalized eigenvalue decomposition $Y^T A_0 X = \Lambda_0, Y^T A_1 X = \Lambda_1$ as in (4.10).
 - 5: Set $(\hat{\mathbf{G}}(x)^T)^{(\dagger)} := \hat{\mathbf{V}}_0(x) X, \hat{\mathbf{H}}(y)^{(\dagger)} := Y^T \hat{\mathbf{U}}_0(y)^T$.
 - 6: Compute $\hat{\mathbf{G}}(x) = \hat{\mathbf{V}}_0(x) X^{-T} = [\hat{g}_1(x), \dots, \hat{g}_r(x)]$, and $\hat{\mathbf{H}}(y) = \hat{\mathbf{U}}_0(y) Y^{-T} = [\hat{h}_1(y), \dots, \hat{h}_r(y)]$.
 - 7: For $i = 1, \dots, r$, compute $\hat{w}_i(z) = \int_a^b \int_c^d f(x, y, z) \hat{g}_i^\dagger(x) \hat{h}_i^\dagger(y) dy dx$, $\alpha_i = \|\hat{w}_i(z)\|$, and $\hat{w}_i(z) = \hat{w}_i(z) / \alpha_i$.
 - 8: $f(x, y, z) \approx \sum_{i=1}^r \alpha_i \hat{g}_i(x) \hat{h}_i(y) \hat{w}_i(z)$ is an approximate CP decomposition.
-

For improved reliability, instead of starting from two specific slices (4.3) and (4.4), it is recommended that one take a number of randomly sampled slices and take

their average (or integrate with respect to two different weight functions). Then the diagonal matrices in (4.3), (4.4) are modified accordingly, but the rest of the argument remains the same. This avoids the defective case where $w_i(z_j) = 0$ for $j = 0, 1$, for some i (in which case the corresponding elements h, g cannot be retrieved).

The success of Algorithm 4 can be measured by $\|f(x, y, z) - \sum_{i=1}^r \alpha_i \hat{g}_i(x) \hat{h}_i(y) \hat{w}_i(z)\|$, which is hoped to be sufficiently small. Failure of this can be caused (and predicted within the algorithm) by the condition numbers of $\hat{\mathbf{G}}(x)$ and $\hat{\mathbf{H}}(y)$, and by examining whether $\mathbf{H}(y)^\dagger f(x, y, z_2) (\mathbf{G}(x)^T)^\dagger$ is close to diagonal for another random slice $f(x, y, z_2)$.

When \mathbf{H} or \mathbf{G} is rank deficient, Algorithm 4 will break down; numerically this means the matrices X and Y become highly ill-conditioned. If two factors among \mathbf{H} , \mathbf{G} and \mathbf{W} have full column rank, then we can still compute the CP decomposition by working e.g. with $g(x, y, z) = f(y, z, x)$. If two or more of the three quasimatrices are rank-deficient, then this remedy will still not help. Computing the CP in such cases is left an open problem—algorithms that are applicable under relaxed conditions have been proposed for discrete tensors, e.g. in [8] and [13], but the algorithm is expensive and its continuous analogue appears to be nontrivial.

5 Two-way odeco functions and decay of CP values

In this section we explore the CP-values of trivariate functions. For bivariate functions, the singular values decay geometrically if the function is analytic in the domain. We suspect that an analogous result might hold for the trivariate case; here we give a partial answer by proving the same holds for CP-values of trivariate functions possessing a property called two-way odeco, which is a natural relaxation of odeco tensors/functions. Odeco tensors were introduced in [54], see also [22] for its analysis.

Definition 5.1 $f(x, y, z)$ is called orthogonally decomposable, or *odeco*, if it can be written in the polyadic form (1.2) such that the functions $g_j(x)$, $h_j(y)$, and $w_j(z)$ are orthogonal, i.e., if

$$\int_a^b g_i(x)g_j(x)dx = \int_c^d h_i(y)h_j(y)dy = \int_e^l w_i(z)w_j(z)dz = \delta_{i,j},$$

for every fixed $i, j = 1, 2, \dots$. $f(x, y, z)$ is called *two-way odeco*, if two of the set of functions $g_j(x)$, $h_j(y)$, and $w_j(z)$ are orthogonal.

Suppose that we know the CP decomposition of f , and the factors may or may not be orthonormal. Note that the CP decomposition of f is a special case of the Tucker format in which the discrete core tensor is diagonal. Now consider converting the CP to an odeco expansion of f . To do so, we can always orthonormalize factors $g_j(x)$, $h_j(y)$ and $w_j(z)$ of the CP decomposition of f e.g., using the QR decomposition [50]. In order to replace the factors with the new orthonormalized factors, the diagonal core tensor has to be multiplied by three discrete matrices which are triangular; hence the core tensor will not necessarily be kept diagonal. A sufficient condition for the core tensor to be kept diagonal is that the given function f is originally odeco in

which case the matrices R from the QR factorizations are actually diagonal. In this case, every term in the CP expansion of f is a singular quadruplet of f ; this can be seen straightforwardly by verifying (2.9). In other words, for an odeco function, the triples $(g_j(x), h_j(y), w_j(z))$ corresponding to the rank-1 terms are singular function tuples and the corresponding coefficient is a singular value⁴. The converse is not true—singular quadruplets are not necessarily CP quadruplets; we return to this in Section 5.2.

For odeco trivariate functions, Algorithm 4 can always compute the CP decomposition: the linearly independence assumption is trivially satisfied because the three quasimatrices $\mathbf{G}(x)$, $\mathbf{H}(y)$, $\mathbf{W}(z)$ are all orthonormal, hence have full column rank r . This also gives an algorithm to check whether a trivariate function is odeco: use Algorithm 4 to compute the CP decomposition, and check if the quasimatrices $\mathbf{G}(x)$, $\mathbf{H}(y)$, $\mathbf{W}(z)$ are all orthonormal. The function is not odeco if Algorithm 4 fails or one of the quasimatrices is not orthonormal.

Clearly, two-way odeco is a more general class of trivariate functions than odeco (although admittedly still a very special class). In this section we show that the CP values decay geometrically for two-way odeco analytic functions, and therefore be approximated by a finite-rank tensor with exponential accuracy.

Unless otherwise mentioned, in this section we suppose f is two-way odeco (in y, z directions, for definiteness):

$$f(x, y, z) = \sum_{i=1}^{\infty} \alpha_i g_i(x) h_i(y) w_i(z), \quad (5.1)$$

with g_i, h_i, w_i all real-valued, having unit norms $\langle h_i, h_j \rangle = \langle w_i, w_j \rangle = \delta_{ij}$ (in the L^2 inner products $\langle h_i, h_j \rangle = \int_c^d h_i(y) h_j(y) dy$ and $\langle w_i, w_j \rangle = \int_e^l w_i(z) w_j(z) dz$), and $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \dots \geq 0$; there is no loss of generality in assuming positivity.

5.1 Geometric decay of CP values for two-way odeco functions

Here we show for two-way odeco functions that the CP-values decay geometrically. The analysis below closely follows that in [32] for bivariate functions, but with some modifications to handle the trivariate case (e.g. the $\sqrt{r+1}$ factor).

Lemma 5.1 *Let $f : [a, b] \times [c, d] \times [e, l] \rightarrow \mathbb{R}$ be a trivariate function that is two-way odeco (in y, z) with CP-values $\alpha_1 \geq \alpha_2 \geq \dots$ as in (5.1). Then for any trivariate function \hat{f} of the form*

$$\sum_{j=1}^r s_j(y) p_j(x, z),$$

we have

$$\alpha_{r+1} \leq \sqrt{r+1} \|f - \hat{f}\|. \quad (5.2)$$

⁴ See e.g. [46] for an analogous discussion regarding eigenvalues of symmetric discrete tensors.

Proof Since $[g_1, \dots, g_r]$ spans a subspace of dimension r , we can choose

$$\phi(y) = \sum_{j=1}^{r+1} c_j h_j(y)$$

so that $\|\phi\| = 1$ and $\hat{f} \times_2 \phi := \int_c^d \hat{f}(x, y, z) \phi(y) dy = 0$; specifically, we choose $[c_1, \dots, c_{r+1}]$ so that it has unit norm and $\int_c^d \phi(y) s_j(y) dy = 0$ for $j = 1, \dots, r$. Therefore

$$\|(f - \hat{f}) \times_2 \phi\| = \|f \times_2 \sum_{j=1}^{r+1} c_j h_j\| = \left\| \sum_{j=1}^{r+1} c_j \alpha_j g_j(x) w_j(z) \right\| =: \|G(x, z)\|.$$

For each $\ell \in \{1, \dots, r+1\}$, we have

$$G(x, z) \times_1 g_\ell(x) = \sum_{j=1}^{r+1} c_j \alpha_j \langle g_\ell, g_j \rangle w_j(z). \quad (5.3)$$

Now let $\ell_* = \operatorname{argmax}_{1 \leq j \leq r+1} |c_j|$. Then we have $|c_{\ell_*}| \geq \frac{1}{\sqrt{r+1}}$, since $\|c\| = 1$. Taking $\ell = \ell_*$ in the above equation and integrating with $w_{\ell_*}(z)$ we obtain

$$|(G(x, z) \times_1 g_{\ell_*}(x)) \times_3 w_{\ell_*}(z)| = |c_{\ell_*} \alpha_{\ell_*}| \geq \frac{\alpha_{\ell_*}}{\sqrt{r+1}}.$$

Thus $|(f - \hat{f}) \times_1 g_{\ell_*} \times_2 \phi \times_3 w_{\ell_*}| \geq \frac{\alpha_{\ell_*}}{\sqrt{r+1}} \geq \frac{\alpha_{r+1}}{\sqrt{r+1}}$ with $\|g_{\ell_*}\| = \|\phi\| = \|w_{\ell_*}\| = 1$, establishing (5.2).

Now we establish the geometric decay of CP values. In the following theorem, E_ρ denotes an open Bernstein ellipse with foci at ± 1 and semiminor and semimajor axis lengths summing to $\rho > 1$. Specifically, $E_\rho := \{\frac{1}{2}(\rho z + (\rho z)^{-1}) : |z| < 1, z \in \mathbb{C}\}$.

Theorem 5.1 *Let $f(x, y, z) : [-1, 1]^3 \rightarrow \mathbb{R}$ be two-way odeco in y, z and analytic, and for every fixed $y_0, z_0 \in [-1, 1]^2$, suppose that $f(x, y_0, z_0) \in C[-1, 1]$ is analytically continuable to the Bernstein ellipse E_ρ where it is uniformly bounded by M . Then the CP-values of f decay as*

$$\alpha_{r+1} \leq \frac{2M\sqrt{r+1}}{\rho-1} \rho^{-r}. \quad (5.4)$$

Proof Write f in Chebyshev expansion with respect to x as

$$f(x, y, z) = \frac{a_0(y, z)}{2} + \sum_{j=1}^{\infty} a_j(y, z) T_j(x).$$

Specifically, we have

$$a_j(y, z) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\cos \theta, y, z) e^{ij\theta} d\theta. \quad (5.5)$$

The imposed analyticity assumptions guarantee that [51, Thm. 8.1] $|a_j(y, z)| \leq 2M\rho^{-j}$, uniformly in $y \in [c, d]$ and $z \in [e, l]$. By taking

$$\hat{f}(x, y, z) := \frac{a_0(y, z)}{2} + \sum_{j=1}^r a_j(y, z) T_j(x),$$

together with (5.2) we obtain

$$\alpha_{r+1} \leq \sqrt{r+1} \|f - \hat{f}\| = \sqrt{r+1} \left\| \sum_{j=r+1}^{\infty} a_j(y, z) T_j(x) \right\| \leq \frac{2M\sqrt{r+1}}{\rho-1} \rho^{-r},$$

as required.

If f is entire and (two-way) odeco, then the decay of the CP-values is even faster, i.e., super-geometric. Moreover, employing an appropriate change of variables, we can establish the geometric decay to the case where the domain of f is a general rectangular parallelepiped $[a, b] \times [c, d] \times [e, l]$.

In the case that f is only (ν times) differentiable, we can prove the decay rate of CP values to be algebraic, like $r^{-(\nu+1)}$ by invoking Theorem 7.1 instead of 8.1 in [51].

While two-way odeco functions are more general than odeco functions, a slightly wider class (for discrete tensors) is treated in [53, Def. 3.5], called functions admitting a *strong two-orthogonal decomposition*, to identify a sufficient condition for a Schmidt-Eckart-Young decomposition (a decomposition for which extracting the leading k terms gives the best rank- k approximation for all k) to exist. For example, a function that is not two-way odeco but has a (continuous analogue of the) strong two-orthogonal decomposition is $g_1(x)h_1(y)w_1(z) + g_2(x)h_2(y)w_2(z) + g_3(x)h_3(y)w_3(z)$, where $[w_1, w_2, w_3]$ is orthogonal and so are $\langle g_1, g_2 \rangle = \langle g_1, g_3 \rangle = \langle h_2, h_3 \rangle = 0$, but otherwise arbitrary, in particular $\langle g_2, g_3 \rangle \neq 0$. Examining whether the geometric decay of the CP-values extends to the strong two-orthogonal case or to more general trivariate functions is left as an open problem.

5.2 Low-rank approximation via CP for two-way odeco functions

We have derived decay bounds for CP-values. We next explore the best rank- r approximation to f , along with the connection between the CP decomposition and singular quadruplets.

Proposition 5.1 *Let $f(x, y, z)$ be a two-way odeco function (in y, z as in the CP decomposition (5.1)). Then the following hold:*

1. For every $r \in \mathbb{N}$, $\hat{f}_r := \sum_{i=1}^r \alpha_i g_i(x) h_i(y) w_i(z)$ is the best rank- r approximant of f .
2. For every $i \in \mathbb{N}$, $(\alpha_i, g_i, h_i, w_i)$ is a singular quadruplet satisfying (2.9), and $(\alpha_1, g_1, h_1, w_1)$ is the largest singular quadruplet of f .
3. For any \hat{f} of CP-rank r , $\|f - \hat{f}\| \geq \alpha_{r+1}$, with equality attained when $\hat{f} = \hat{f}_r$.
4. If f further satisfies the assumptions in Theorem 5.1, it can be approximated by a rank- r function with exponential accuracy, namely $\|f - \hat{f}_r\| \leq \frac{2M\sqrt{r+1}}{\rho-1} \rho^{-r}$.

Proof The proof of the first statement is a straightforward adaptation of [53, Thm. 3.6]: We can obviously rewrite f as

$$f(x, y, z) = \sum_{i=1}^{\infty} \alpha_i G_i(x, y) w_i(z), \quad (5.6)$$

where $G_i(x, y) := g_i(x) h_i(y)$ is a bivariate function (of rank one). Consider the Euclidean inner product of bivariate functions (see the definition of the tensor space isomorphism and the following discussion in [18, pp. 59-61])

$$\langle G(x, y), H(x, y) \rangle := \int_a^b \int_c^d G(x, y) H(x, y) dy dx.$$

Note that (5.6) is the SVD of f regarded as a bivariate function of z from one side and (x, y) from the other, because

$$\begin{aligned} \langle G_i(y, z), G_j(y, z) \rangle &= \int_c^d \int_e^f g_i(x) h_i(y) g_j(x) h_j(y) dz dy \\ &= \langle g_i(x), g_j(x) \rangle \langle h_i(y), h_j(y) \rangle \\ &= \delta_{ij}. \end{aligned}$$

Now since the SVD provides an optimal truncation of bivariate functions, the same holds for (5.1), i.e., $\sum_{i=1}^r \alpha_i g_i(x) h_i(y) w_i(z)$ is the best rank- r approximation of f .

We next treat the second statement. It is straightforward to verify that $i \in \mathbb{N}$, $(\alpha_i, g_i, h_i, w_i)$ satisfies (2.9). To see that $(\alpha_1, g_1, h_1, w_1)$ is the largest singular quadruplet of f , let $g(x)$ be a univariate function of unit norm and consider

$$f_1(y, z) := \int_a^b f(x, y, z) g(x) dx = \langle g_1, g \rangle \alpha_1 h_1(y) w_1(z) + \sum_{i=2}^{\infty} \langle g_i, g \rangle \alpha_i h_i(y) w_i(z).$$

By the orthonormality assumption on h_i and w_i , we see that the largest singular value of $f_1(y, z)$ is the largest value among $\langle g_i, g \rangle \alpha_i$ for $i = 1, 2, \dots$. By the normalization this is easily seen to be bounded by α_1 . Since this holds for any $g(x)$, it follows that α_1 is an upper bound for the largest singular value of f . Furthermore, equality holds by taking $g(x) := g_1(x)$, so $(\alpha_1, g_1, h_1, w_1)$ is the largest singular quadruplet of f .

The third statement is a trivial consequence of the first two. The fourth follows directly from combining the third statement and (5.4).

This result tells us that in order to compute the best rank- r approximant to f we can either truncate the leading CP terms, or successively find and subtract the best rank-1 approximant. Another consequence of this result is that if we perturb a rank-1 function f by a function that is orthogonal to f in two ways, then its best rank-1 approximant remains intact (until the perturbation is large enough to change the first term in the CP decomposition).

As the proof suggests, the first statement still holds if we relax the assumption on f to having a strong two-orthogonal decomposition.

Regarding the second statement of Proposition 5.1, one might suspect that the converse could also hold: every singular quadruplet of f would be a CP quadruplet, i.e., CP and SVD are identical. It turns out that this is not true, even for (three-way) odeco functions. For example, consider an odeco function $f = \sigma_1 g_1 h_1 w_1 + \sigma_2 g_2 h_2 w_2$, where $\sigma_1, \sigma_2 > 0$ and $\langle g_1, g_2 \rangle = \langle h_1, h_2 \rangle = \langle w_1, w_2 \rangle = 0$. Note that this is the unique CP decomposition of f by Kruskal's sufficient condition [26]. However, f has other singular quadruplets: Let $s, t \in \mathbb{R}$ be such that $s^2 + t^2 = 1$ and $s\sigma_1 = t\sigma_2$, and define $\tilde{g} = sg_1 + tg_2$, $\tilde{h} = sh_1 + th_2$, $\tilde{w} = sw_1 + tw_2$, and $\tilde{\sigma} = s\sigma_1$. Then we can verify

that $(\tilde{\sigma}, \tilde{g}, \tilde{h}, \tilde{w})$ is a singular quadruplet, satisfying (2.9)⁵. In any case, by the first statement of Proposition 5.1, the largest singular quadruplet is a CP quadruplet.

The above example, together with the second statement of Proposition 5.1, shows that the CP quadruplets are a strict subset of the singular quadruplets for (two-way) odeco functions. However, in view of the first statement of Proposition 5.1, we expect the existence of singular quadruplets that are not CP quadruplets to be of marginal significance, at least in the context of approximating f by a low-rank function.

6 Numerical examples

Unless otherwise stated, 3D functions in this section are defined over $[-1, 1]^3$. All the experiments were done in MATLAB building upon the Chebfun package [14].

6.1 Best rank-1 approximation

By the power method we can find the largest singular quadruplet, which by Theorem 2.1, gives the best rank-1 approximant to f . Figure 6.1 shows the convergence history of power method applied to the function $f(x, y, z) = \exp(\sin(x + y + z))$, measured by the error of the Rayleigh functional $|\hat{\sigma} - \sigma|$ and the residual, computed by (recalling (2.9))

$$\left\| \begin{bmatrix} \int_c^d \int_e^f f(x, y, z) \hat{w}(z) \hat{h}(y) dz dy - \hat{\sigma} \hat{g}(x) \\ \int_a^b \int_e^f f(x, y, z) \hat{w}(z) \hat{g}(x) dz dx - \hat{\sigma} \hat{h}(y) \\ \int_a^b \int_c^d f(x, y, z) \hat{h}(y) \hat{g}(x) dy dx - \hat{\sigma} \hat{w}(z) \end{bmatrix} \right\|_2 \quad (6.1)$$

where the values wearing a hat indicate computed quantities. Here the norm is the standard ℓ_2 norm. Observe that the error of the Rayleigh quotient scales like the squared of the norm of the residual; this is a known phenomenon [54]. Since this function is not two-way odeco, the CP and singular value decompositions are not identical, in particular, $\sigma_1 \neq \alpha_1$.

Sometimes the power method converges not to the largest singular quadruplet but to another one. Experiments suggest this happens with nonnegligible probability (see Section 6.2), although most of the time, the power method finds the largest singular quadruplet. For the purpose of finding the best rank-1 approximant, a simple remedy is to run the power method several times with multiple initial guesses and adopt the one with the largest singular value. Alternatively, one can use the first term of the multivariate adaptive cross approximation (MACA) or HOSVD of f as the initial guess instead of using random ones.

6.1.1 Low-rank approximation via CP vs. greedy, general case

It is known that the greedy algorithm of approximating a tensor by successively finding and subtracting the best rank-1 approximant generally does not give the CP de-

⁵ The singular quadruplets for odeco tensors is fully studied in [37], which shows that the number of

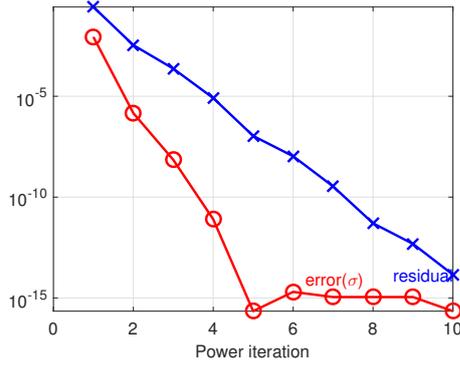


Fig. 6.1 Typical convergence of power method for the largest singular quadruplet. $f(x, y, z) = \exp(\sin(x + y + z))$.

composition [23]. Indeed, the result in [53] can be seen to carry over to functions, to show that generic trivariate functions do not have a Schmidt-Eckart-Young decomposition (such functions need to have a weak two-orthogonal decomposition [53, Thm. 3.3]). We illustrate this in Figure 6.2, where the residual $\|f - \hat{f}\|$ is shown for an approximant \hat{f} of varying rank, obtained by either the greedy approach or truncating the CP decomposition.

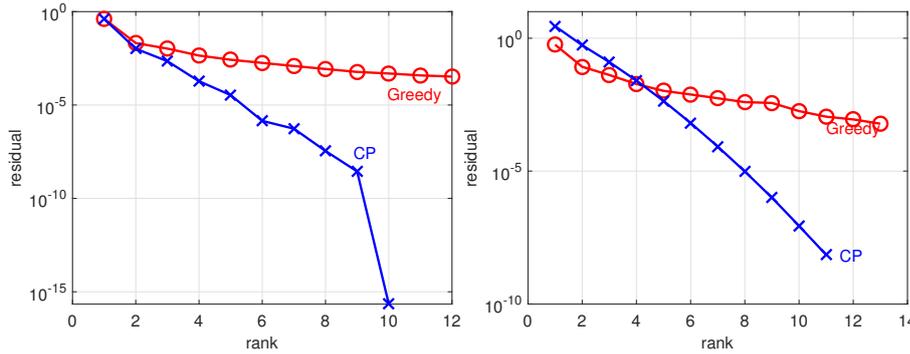


Fig. 6.2 Convergence of residual $\|f - \hat{f}\|$ via greedy rank-1 approximation and CP for non-odeco (in any direction) functions. Left: random rank-10 function with $\alpha_j = 10^{-j}$. Right: $f(x, y, z) = \exp(xyz)$.

We verify as expected that the greedy approach does not give the optimal rank- r approximants, except when $r = 1$. With a rank-one or low-rank approximation, on the other hand, the greedy approach usually gives a smaller residual; this reflects the fact that it is indeed the best rank-1 approximant, which is not necessarily true for CP (unless the function is two-way odeco).

singular values grows rapidly with both the dimension and size, and that the singular vector tuples form a positive-dimensional variety. Extending such results to the continuous case is an interesting open problem.

Note that the CP residual does not always go down to working precision $\approx 10^{-16}$; for example, for $f(x, y, z) = \exp(xyz)$ and the computed CP decomposition has residual $\approx 10^{-8}$, and this does not improve by changing the SVD tolerance in Algorithm 4; we observed similar behaviors for many functions. This seems to be caused by conditioning of the eigenvalue problem, in particular, the quasimatrix of eigenvectors (step 5 of Algorithm 4). For this reason we do not always recommend finding a compressed representation of a trivariate function by the CP decomposition.

6.1.2 Low-rank approximation via CP vs. greedy, two-way odeco case

For two-way odeco functions, we have shown in Proposition 5.1 that the leading r CP factors give the best rank- r approximation. A corollary of this result is that a greedy approach of successively subtracting the largest CP quadruplet (equal to the largest singular quadruplet) gives the CP decomposition and hence the best rank- r approximation. We illustrate this in Figure 6.3. Here we generate a two-way odeco function as in (2.11), where we take T to be a diagonal tensor of diagonal elements 10^{-j} , $j = 1, 2, \dots, 10$ and $B := PQ_B$ and $C := PQ_C$ where P is an orthonormal quasimatrix of Legendre polynomials and Q_B and Q_C are 10×10 random orthonormal matrices. The equivalence between the CP and greedy curves (up to roundoff errors) illustrates the fact that for odeco or two-way odeco functions, the greedy approach and truncated CP both give the best rank- r approximant for every r .

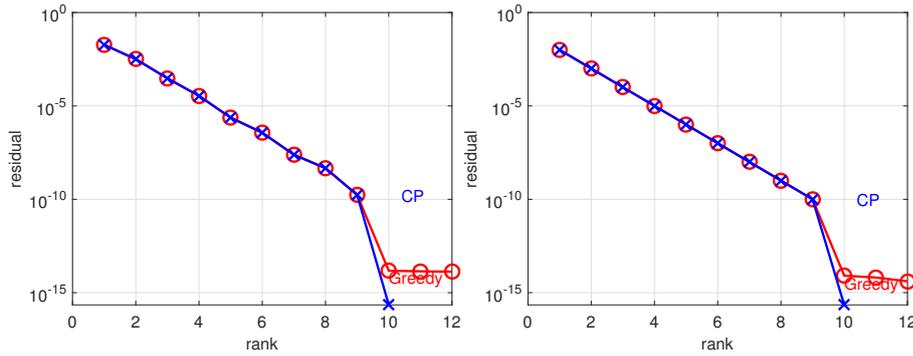


Fig. 6.3 For two-way odeco (left) and three-way odeco (right) functions, the SVD and CP are the same, so power method gives CP (up to floating-point errors).

6.1.3 Finding a rank-one function in a subspace

As a natural continuous analogue of the problem considered in [34], consider the following problem. Let f, g, h be rank-one trivariate functions, and let $\hat{f} = f + c_1g + c_2h$, which is of rank three. Given \hat{f}, g, h , we would like to find (or "recover") the function f such that

1. f is rank one, and

2. \hat{f}, g, h and f, g, h span the same subspace.

Put another way, we are looking for a basis consisting of rank-one functions for the subspace spanned by \hat{f}, g , and h . In the discrete case, this problem has applications in data compression and image separation.

As also described in the Chebfun Example

<http://www.chebfun.org/examples/approx3/FindingRankOne.html>

a simple algorithm executes alternating projection between rank-one functions and the subspace of trivariate functions spanned by \hat{f}, g , and h . In that Example, projection to rank-one functions was done via the MACA process. Now equipped with a method for the best rank-one approximation, we replace MACA with HOPM; Figure 6.4 shows the results, which plot the error $\hat{f}_{it} - f$ in the x -mode after one through seven iterations.

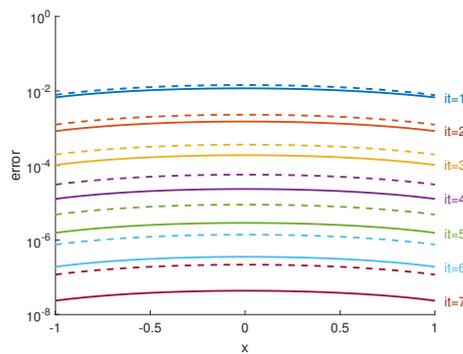


Fig. 6.4 Convergence of x -mode as iteration proceeds for finding a rank-one trivariate function in a subspace of trivariate functions. The dashed line shows the performance when the rank-1 extraction is done in an MACA manner.

We observe from the fact that the solid curves (HOPM) lie below the dashed ones (MACA), that by finding the correct best rank-1 approximant in each step, we improve the convergence.

6.2 Computing more than one singular quadruplet with HOPM

The HOPM usually converges to the dominant singular quadruplet. However, convergence to a different singular quadruplet is also possible. For example for the function⁶ $f(x, y, z) = |((x + iy) \exp(iz))^2 - 1|^2$ defined over $[-1.5, 1.5] \times [-1.5, 1.5] \times [-6, 6]$ we have computed the singular values $\sigma_1 = 55.74$ and $\sigma_2 = 22.46$. The corresponding singular functions in each mode are depicted in Figure 6.5. In this example HOPM converged to the largest singular quadruplet with $\sigma_1 = 55.74$ about 80% of the random runs, the remaining runs all converging to $\sigma_2 = 22.46$. In most other examples that we tried, HOPM converged to the largest singular quadruplet with (often much) higher probability.

⁶ This is the function `doublehelix` from the Chebfun gallery.

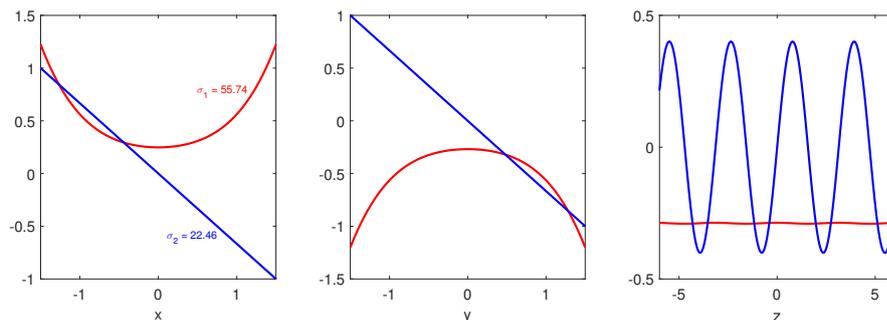


Fig. 6.5 Two leading singular functions of the doublehelix function, in the x, y, z modes.

6.3 Eigenvalues of symmetric functions

Kolda and Mayo [24] introduce a shifted power method for computing eigenvalues and eigenvectors of a symmetric discrete tensor. The shift parameter, if appropriately chosen, reduces the underlying optimization problem to a convex problem. This has the benefit of achieving convexity for certain values of the shift. For symmetric functions, the infinite-dimensional counterpart, we observe similar phenomena. Figure 6.7 shows a typical convergence history, which is again linear with the eigenvalues converging like the square of the residual.

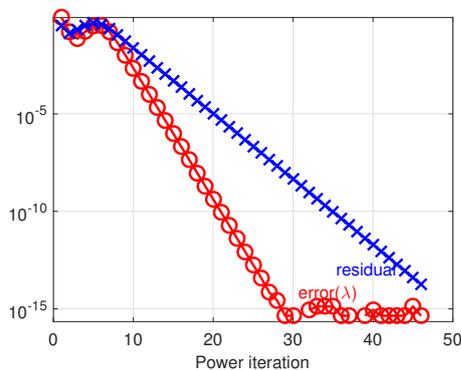


Fig. 6.6 Typical convergence of power method for an eigenpair of a symmetric function $f(x, y, z) = \sin(x + y + z)$ with shift 1 (shift 0 often results in nonconvergence, see Figure 6.7). As for the singular values, the eigenvalue accuracy scales like the residual squared.

The shift also gives the flexibility to compute eigenpairs other than the largest one, unlike the counterpart for singular values presented above. We give an illustration in Figure 6.7, which shows the history of the Rayleigh quotient for different values of the shift parameter. Note that without a shift, the power method did not converge.

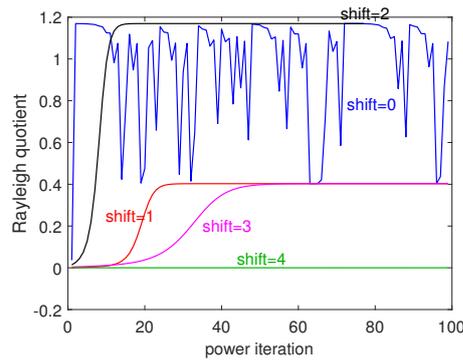


Fig. 6.7 Typical convergence of power method for an eigenpair of a symmetric function $f(x, y, z) = \sin(x + y + z)$ with varying shifts

It would be of interest to find the singular quadruplets that are not the largest by a similar shifting technique.

6.4 Decay of CP values

We have proved in Section 5 that for two-way odeco functions, the CP values decay exponentially, which represent the quality of the best rank- r approximation, and that singular values and CP values are equivalent for three-way odeco functions.

One difficulty to illustrate the geometric decay of CP values is that it is nontrivial to generate a two-way odeco function without specifying the CP values. For example, $f(x, y, z)$ as in (2.11) with T being the diagonal tensor with 1s on the diagonal, and B and C orthonormal, is a (y, z) two-way odeco function, but its CP values are specified explicitly by construction: they are the 2-norms of the columns of the quasimatrix A . Another example of a two-way odeco function is $f(x, y, z) := G(x, y)w(z)$, which involves only one univariate factor in one of the variables. Then the existence of SVD for the bivariate function $G(x, y)$ guarantees that f is two-way odeco. In this case the CP values of f are the singular values of G multiplied by the 2-norm of w .

Meanwhile, we do not know whether the CP values would decay similarly for general functions. Nonetheless, experiments suggest that CP-values decay exponentially for general smooth functions, even if they are not two-way odeco, as illustrated in Figure 6.8. Making this precise is left as an open problem.

Acknowledgements The authors would like to thank Alex Townsend and André Uschmajew for helpful comments on the manuscript.

References

1. Atkinson, K., Han, W.: Theoretical Numerical Analysis. Springer (2005)
2. Benson, A.R., Gleich, D.F., Lim, L.H.: The spacey random walk: A stochastic process for higher-order data. SIAM Rev. **59**(2), 321–345 (2017)

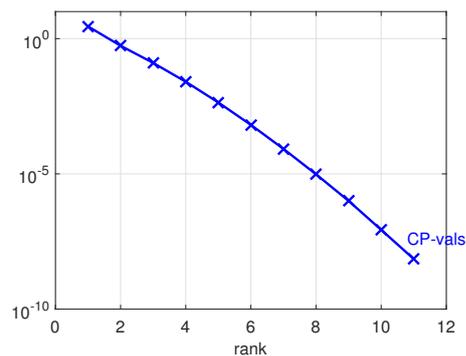


Fig. 6.8 CP values of analytic (entire) function $\exp(xyz)$.

3. Beylkin, G., Mohlenkamp, M.J.: Numerical operator calculus in higher dimensions. *Proc. Natl. Acad. Sci.* **99**(16), 10,246–10,251 (2002)
4. Beylkin, G., Mohlenkamp, M.J.: Algorithms for numerical analysis in high dimensions. *SIAM J. Sci. Comp.* **26**(6), 2133–2159 (2005)
5. Blekherman, G., Teitler, Z.: On maximum, typical and generic ranks. *Mathematische Annalen* **362**(3–4), 1021–1031 (2015)
6. Chevreuil, M., Lebrun, R., Nouy, A., Rai, P.: A least-squares method for sparse low rank approximation of multivariate functions. *SIAM/ASA J. Uncertainty Quantification* **3**(1), 897–921 (2015)
7. Courant, R., Hilbert, D.: *Methods of Mathematical Physics*, vol. 1. CUP Archive (1965)
8. De Lathauwer, L.: A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J. Matrix Anal. Appl.* **28**(3), 642–666 (electronic) (2006).
9. De Lathauwer, L., Comon, P., De Moor, B., Vandewalle, J.: Higher-order power method. *Nonlinear Theory and its Applications, NOLTA95* **1** (1995)
10. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21**, 1253–1278 (2000)
11. De Lathauwer, L., De Moor, B., Vandewalle, J.: Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM J. Matrix Anal. Appl.* **26**(2), 295–327 (electronic) (2004).
12. Demmel, J.: *Applied Numerical Linear Algebra*. SIAM, Philadelphia, USA (1997)
13. Domanov, I., Lathauwer, L.D.: Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition. *SIAM J. Matrix Anal. Appl.* **35**(2), 636–660 (2014)
14. Driscoll, T.A., Hale, N., Trefethen, L.N.: *Chebfun Guide*. Pafnuty Publications, Oxford (2014)
15. Engel, E., Dreizler, R.M.: *Density Functional Theory: An Advanced Course*. Springer (2011)
16. Golub, G.H., Van Loan, C.F.: *Matrix Computations*. Johns Hopkins University Press (2013)
17. van Groesen, E.: *Applied Analytical Method*. Lecture notes for Applied Analysis & Mathematical Physics (March 2001)
18. Hackbusch, W.: *Tensor Spaces and Numerical Tensor Calculus*, *Springer Series in Computational Mathematics*, vol. 42. Springer, Heidelberg (2012).
19. Hashemi, B., Trefethen, L.N.: *Chebfun in three dimensions*. To appear in *SIAM J. on Sci. Comp.*, available at https://people.maths.ox.ac.uk/trefethen/Chebfun3_revised.pdf
20. Hille, E., Tamarkin, J.D.: On the characteristic values of linear integral equations. *Acta Math.* **57**(1), 1–76 (1931).
21. Kofidis, E., Regalia, P.A.: On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM J. Matrix Anal. Appl.* **23**(3), 863–884 (2002)
22. Kolda, T.: Orthogonal tensor decompositions. *SIAM J. Matrix Anal. Appl.* **23**(1), 243–255 (2001)
23. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM Rev.* **51**(3), 455–500 (2009)
24. Kolda, T.G., Mayo, J.R.: Shifted power method for computing tensor eigenpairs. *SIAM J. Matrix Anal. Appl.* **32**(4), 1095–1124 (2011)
25. Kress, R.: *Linear Integral Equations*. Springer (2014)
26. Kruskal, J.B.: Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **18**(2), 95–138 (1977)

27. Kruskal, J.B.: Rank, decomposition, and uniqueness for 3-way and n-way arrays. In: *Multiway Data Analysis*, R. Coppi and S. Bolasco, eds., North-Holland, p. 718 (1989)
28. Leurgans, S.E., Ross, R.T., Abel, R.B.: A decomposition for three-way arrays. *SIAM J. Matrix Anal. Appl.* **14**(4), 1064–1083 (1993).
29. Li, W., Ng, M.K.: On the limiting probability distribution of a transition probability tensor. *Linear Multilinear A.* **62**(3), 362–385 (2014)
30. Lickteig, T.: Typical tensorial rank. *Linear Algebra Appl.* **69**, 95–120 (1985)
31. Lim, L.H.: Singular values and eigenvalues of tensors: a variational approach. In: *Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP '05)*, pp. 129–132 (2005)
32. Little, G., Reade, J.: Eigenvalues of analytic kernels. *SIAM J. Math. Anal.* **15**, 133–136 (1984)
33. Mercer, J.: Functions of positive and negative type and their connection with the theory of integral equations. *Phil. Trans. R. Soc. Lond. A* **209**, 415–446 (1909)
34. Nakatsukasa, Y., Soma, T., Uschmajew, A.: Finding a low-rank basis in a matrix subspace. *Math. Prog.* **162**(1–2), 325–361 (2017)
35. Qi, L.: Eigenvalues of a real supersymmetric tensor. *J. Symb. Comput.* **40**(6), 1302–1324 (2005)
36. Regalia, P.A., Kofidis, E.: The higher-order power method revisited: convergence proofs and effective initialization. In: *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, vol. 5, pp. 2709–2712. IEEE (2000)
37. robeva, E., Seigal, A.: Singular vectors of orthogonally decomposable tensors. *Linear Multilinear A.* pp. 1–15 (2017)
38. Sanchez, E., Kowalski, B.R.: Tensorial resolution: a direct trilinear decomposition. *Journal of Chemometrics* **4**(1), 29–45 (1990)
39. Schmidt, E.: Zur Theorie der linearen und nichtlinearen Integralgleichungen. I Teil. Entwicklung willkuerlichen Funktionen nach System vorgeschriebener. *Math. Ann.* **63**, 433–476 (1907)
40. Schultz, T., Seidel, H.P.: Estimating crossing fibers: A tensor decomposition approach. *IEEE Trans. Vis. Comput. Graph.* **14**(6) (2008)
41. Spatschek, R., Eidel, B.: Driving forces for interface kinetics and phase field models. *Int. J. Solids Struct.* **50**(14), 2424–2436 (2013)
42. Stakgold, I., Holst, M.J.: *Green's Functions and Boundary Value Problems*. John Wiley & Sons (2011)
43. Stewart, G.W.: On the early history of the singular value decomposition. *SIAM Rev.* **35**(4), 551–566 (1993)
44. Stewart, G.W.: *Afternotes Goes to Graduate School: Lectures on Advanced Numerical Analysis*. SIAM (1998)
45. Strassen, V.: Rank and optimal computation of generic tensors. *Linear Algebra and its Applications* **52**, 645–685 (1983)
46. Sturmfels, B.: Tensors and their eigenvectors. *Notices Amer. Math. Soc.* **63**(6) (2016)
47. Townsend, A.: *Computing with functions in two dimensions*. Ph.D. thesis, University of Oxford (2014)
48. Townsend, A., Trefethen, L.N.: An extension of Chebfun to two dimensions. *SIAM J. Sci. Comp.* **35**(6), C495–C518 (2013)
49. Townsend, A., Trefethen, L.N.: Continuous analogues of matrix factorizations. *Proc. Roy. Soc. A* **471**(2173), 20140,585 (2015)
50. Trefethen, L.N.: Householder triangularization of a quasimatrix. *IMA J. Numer. Anal.* **30**(4), 887–897 (2010)
51. Trefethen, L.N.: *Approximation Theory and Approximation Practice*. SIAM (2013)
52. Uschmajew, A.: A new convergence proof for the higher-order power method and generalizations. *Pac. J. Optim.* **11**, 309–321 (2015)
53. Vannieuwenhoven, N., Nicaise, J., Vandebril, R., Meerbergen, K.: On generic nonexistence of the Schmidt–Eckart–Young decomposition for complex tensors. *SIAM J. Matrix Anal. Appl.* **35**(3), 886–903 (2014)
54. Zhang, T., Golub, G.H.: Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.* **23**(2), 534–550 (2001)