



Matrix-Valued LogSumExp Approximation for Colour Morphology

Marvin Kahra¹ · Michael Breuß¹ · Andreas Kleefeld^{2,3} · Martin Welk⁴

Received: 8 January 2025 / Accepted: 18 August 2025
© The Author(s) 2025

Abstract

Mathematical morphology is a part of image processing that employs a moving window to modify pixel values through the application of specific operations. The supremum and infimum are pivotal concepts, yet defining them in a general sense for high-dimensional data such as colour is a challenging endeavour. As a result, a number of different approaches have been taken to try to find a solution, with certain compromises being made along the way. In this paper, we present an analysis of a novel approach that replaces the supremum within a morphological operation with the LogExp approximation of the maximum for matrix-valued colours. This approach has the advantage of extending the associativity of dilation from the one-dimensional to the higher-dimensional case. Furthermore, the minimality property is investigated and a relaxation specified to ensure that the approach is continuously dependent on the input data.

Keywords Mathematical morphology · Colour image · Matrix-valued image · Positive definite matrix · Symmetric matrix · Supremum

1 Introduction

Mathematical morphology is a theory used to analyse spatial structures in images. Over the decades, it has developed into a very successful field of image processing, see, for example, [1–3] for an overview. Morphological operators basically

consist of two main components. The first of these is the structuring element (SE), which is characterised by its shape, size and position. These in turn can be divided into two types of SEs, flat and non-flat cf. [4]. A flat SE basically defines a neighbourhood of the central pixel where appropriate morphological operations are performed, while a non-flat SE also contains a mask with finite values that are used as additive offsets. The SE is usually implemented as a window sliding over the image. The second main component is used to perform a comparison of values within an SE. The basic operations in mathematical morphology are dilation and erosion, where a pixel value is set to the maximum and minimum, respectively, of the discrete image function within the SE. Many morphological filtering procedures of practical interest, such as opening, closing or top hats, can be formulated by combining dilation and erosion operations. Since dilation and erosion are dual operations, it is often sufficient to restrict oneself to one of the two when constructing algorithms.

Let us also briefly extend this concept to colour morphology, as it is the underlying concept for our further considerations. As already mentioned, the most important operation in morphology is to perform a comparison of the tonal values or, in our case, the colour values within the SE over certain sets of pixels in an image domain. For the simpler application areas such as binary or grey value morphology, one can act directly on complete lattices in order to obtain

✉ Marvin Kahra
marvin.kahra@b-tu.de

Michael Breuß
breuss@b-tu.de

Andreas Kleefeld
a.kleefeld@fz-juelich.de

Martin Welk
martin.welk@umit-tirol.at

¹ Institute for Mathematics, Brandenburg University of Technology Cottbus-Senftenberg, Platz der Deutschen Einheit 1, 03046 Cottbus, Brandenburg, Germany

² Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, North Rhine-Westphalia, Germany

³ Faculty of Medical Engineering and Technomathematics, University of Applied Sciences Aachen, Heinrich-Mußmann-Str. 1, 52428 Jülich, North Rhine-Westphalia, Germany

⁴ Institute of Biomedical Image Analysis, UMIT TIROL – Private University for Health Sciences and Health, Eduard-Wallnöfer-Zentrum 1, 6060 Hall, Tyrol, Austria

a total order of the colour values, see [3]. In the case of colour morphology, this is no longer the case, as there is no total order of the colour values. For this reason, corresponding semi-orders and different basic structures are used, cf. [5]. The first approach that could be used for this would be to regard each colour channel of an image as an independent image and to perform grey value morphology on each of them. This approach has the serious disadvantage that we lose the correlated information between the colour channels, which could be used to further improve the filtering results. The other approach, which is more popular, uses a vector space structure in which each colour is considered a vector in an underlying colour space. In order to compute a supremum or infimum, it is necessary to have an order for the vector space. However, there are a plethora of ordering approaches for colour morphology. For details of the most commonly used approaches, we refer the reader to the overview provided in [6]. We will take the latter approach, but use symmetric matrices instead of vectors. Since there is also no total order for colour matrices either, we will order the elements by means of a semi-order, namely the Loewner order, see [7]. This means, though, that we need an additional function to select a minimum upper bound, namely the supremum function.

To calculate two of the basic operations of colour morphology, dilation and erosion, it is necessary to determine the supremum or infimum. Because of the duality between these two operations, it is common to consider only one of them. Here we will concentrate on dilation and the construction of the supremum. However, there are several approaches how to choose the supremum of a set of symmetric matrices, based on different norms. To give some examples, we mention here the nuclear norm, the Frobenius norm and the spectral norm. For a comparison of these norms we refer to the work by Welk, Kleefeld and Breuß [8].

Here we want to consider another approach, namely the approximation of the supremum by a so-called LogSumExp approximation of Maslov [9]. This is an approximation which has already given promising results in the work of Kahra, Sridhar and Breuß [10] for greyscale images and for colour images in [11] in conjunction with a fast Fourier transform. However, the latter only represents a one-dimensional channel-wise approach to colour morphology. Another connection worth mentioning is the work [12] of Burgeth, Welk, Feddern and Weickert, where root and power functions were used for symmetric positive semi-definite matrices instead of logarithm and exponential function. However, our approach does not require a positive semi-definite matrix, but works with any colour matrix, and preserves the so-called transitivity of greyscale morphology.

This paper will be the next step from [13] for transferring the LogSumExp approach to colour morphology with tonal vectors/matrices. The goal is to present a clear characterisa-

tion of this approach for tonal value matrices to close the gap in the reasoning of [13] and to extend it with regard to certain properties. In this way, we will end up with a dilation operator that, with a few minor compromises, combines many of the advantageous properties of the other multidimensional approaches while preserving the associativity of the dilation, which, as far as we know, is not the case with the other multidimensional approaches. In addition, we will present a relaxed formulation of the operator, which addresses one of the primary limitations of the operator.

2 General Definitions

To make this paper self-contained, we want to use this section to clarify some basic definitions and terminology, using our previous paper [13] as a guide. This is divided into two subsections, one for the morphological terms and one for the terms related to the Loewner order.

We will start with the morphological concepts of dilation/erosion for greyscale images and then extend this to colour images. In particular, we will discuss how colours can be represented and which algebraic structures we will consider for this paper and describe the one we have chosen in more detail.

In the second subsection, we look at why we need such an order and why we have chosen it. We show what we mean by a minimiser of a convex set with respect to the Loewner order and which properties it must fulfil. This will give us a general explanation for a matrix supremum. Finally, we will give a brief overview of how different norms lead to different matrix suprema.

2.1 Colour Morphology

We begin with a two-dimensional, discrete image domain $\Omega \subseteq \mathbb{Z}^2$ and a single-channel greyscale image, which is described by a discrete function $f : \Omega \rightarrow [0, 255]$. In the case of non-flat morphology, the **structuring element (SE)** can be represented as a function $b : \mathbb{Z}^2 \rightarrow \mathbb{R} \cup \{-\infty\}$ with

$$b(\mathbf{x}) := \begin{cases} \beta(\mathbf{x}), & \mathbf{x} \in B_0, \\ -\infty, & \text{otherwise,} \end{cases} \quad B_0 \subset \mathbb{Z}^2, \quad (1)$$

where β is itself a small greyscale image that should have the same scaling for the greyscale values as the input image. It defines the height of the SE or in other words which pixel will be “prioritised” by the filtering. B_0 is a set centred at the origin. However, the origin of the SE needs not always to be at its centre. It determines the shape and size of the SE, as it specifies which elements are to be compared with each other. Frequently used shapes for this are squares, discs,



Fig. 1 Application of different morphological operators on a 256×256 greyscale image with a 5×5 SE. **From left to right:** Downscaled original image from TAMPERE17 noise-free image database [14], dilation, erosion, opening and closing

diamonds, hexagons and crosses, see [15, 16]. In the case of a flat filter (flat morphology), it is simply the special case $\beta(\mathbf{x}) = 0$. Two of the most elementary operations of mathematical morphology are **dilation**

$$(f \oplus b)(\mathbf{x}) := \max_{\mathbf{u} \in \mathbb{Z}^2} \{f(\mathbf{x} - \mathbf{u}) + b(\mathbf{u})\}, \quad \mathbf{x} \in \Omega, \quad (2)$$

and **erosion**

$$(f \ominus b)(\mathbf{x}) := \min_{\mathbf{u} \in \mathbb{Z}^2} \{f(\mathbf{x} + \mathbf{u}) - b(\mathbf{u})\}, \quad \mathbf{x} \in \Omega, \quad (3)$$

see Figure 1 for an example. In particular, these two operations are dual in the following sense with respect to complementation. Let the range of the greyscale values of f be given by the interval $[f_{\min}, f_{\max}] \subseteq [0, 255]$, where f_{\min} is the lower limit and f_{\max} the upper limit of the greyscale values of f . We define the **complementary image** f^c as

$$f^c(\mathbf{x}) := f_{\max} - f(\mathbf{x}) + f_{\min}, \quad \mathbf{x} \in \Omega. \quad (4)$$

Then one has

$$\begin{aligned} & \left((f^c \oplus \check{b})^c \right)(\mathbf{x}) \\ &= f_{\max} - \max_{\mathbf{u} \in \mathbb{Z}^2} \left\{ f_{\max} - f(\mathbf{x} - \mathbf{u}) + \check{b}(\mathbf{u}) + f_{\min} \right\} \\ & \quad + f_{\min} \\ &= f_{\max} - f_{\max} - \max_{\mathbf{u} \in \mathbb{Z}^2} \left\{ -f(\mathbf{x} - \mathbf{u}) + \check{b}(\mathbf{u}) \right\} \\ & \quad - f_{\min} + f_{\min} \\ &= \min_{\mathbf{u} \in \mathbb{Z}^2} \left\{ f(\mathbf{x} - \mathbf{u}) - \check{b}(\mathbf{u}) \right\} \\ &= \min_{\mathbf{u} \in \mathbb{Z}^2} \{f(\mathbf{x} + \mathbf{u}) - b(\mathbf{u})\} = (f \ominus b)(\mathbf{x}), \quad \mathbf{x} \in \Omega, \end{aligned} \quad (5)$$

where $\check{b}(\mathbf{x}) = b(-\mathbf{x})$. This shows the duality between dilation and erosion.

However, with these two operations, many other operations can be defined that are of great interest in practice, e.g. **opening** $f \circ b = (f \ominus b) \oplus b$ and **closing** $f \bullet b = (f \oplus b) \ominus b$, see Fig. 1. In general, an opening will result in the deletion of minor, protruding components of an object. For example,

this can be observed by the first column on the left side of the entrance in Fig. 1. The closing operation will fill small holes or thin intruding parts of the object. This may entail the destruction of smaller dark areas, such as the shadows of the columns on the outermost left side or the thin flat window illustrated in Fig. 1. These filtering operations can be employed in conjunction with varying sizes of the SE to compute size distributions in binary images, see [1] for further information about granulometry and related applications.

We turn now to our actual area of interest, namely colour morphology. This is similar to the greyscale morphology already shown, with the difference that we no longer have just one channel, but three. There are many useful formats for expressing this, see [17]. A classic approach in this sense is the channel-by-channel processing of an image with the red–green–blue (RGB) colour model, see, for example, [11] for a recent example of channel-wise scheme implementation. However, instead of RGB vectors, we will use symmetric 2×2 matrices.

For this we assume that the colour values are already normalised to the interval $[0, 1]$ for the corresponding channels. First, we transfer this vector into the hue–chroma–luminance (HCL) colour space by means of $M = \max\{R, G, B\}$, $m = \min\{R, G, B\}$, $C = M - m$, $L = \frac{1}{2}(M + m)$ and

$$H = \begin{cases} \frac{G-B}{6C} \bmod 1, & \text{if } M = R, \\ \frac{B-R}{6C} + \frac{1}{3} \bmod 1, & \text{if } M = G, \\ \frac{R-G}{6C} + \frac{2}{3} \bmod 1, & \text{if } M = B. \end{cases}$$

Subsequently, the luminance L is replaced with the modified luminance $\tilde{L} = 2L - 1$ and the quantities C , $2\pi H$ and \tilde{L} are regarded as radial, angular and axial coordinates of a cylindrical coordinate system, respectively. Since the transformation shown here maps each colour from the RGB colour space one-to-one to a colour in the HCL bi-cone, see Figure 2, this represents a bijection onto the bi-cone, which in turn is interpreted with Cartesian coordinates by $x = C \cos(2\pi H)$, $y = C \sin(2\pi H)$ and $z = \tilde{L}$. Finally, we map these Cartesian coordinates onto a symmetric matrix in the following

manner:

$$\mathbf{A} := \frac{\sqrt{2}}{2} \begin{pmatrix} z-y & x \\ x & z+y \end{pmatrix},$$

where the complete transformation process is a bijective mapping, see [7].

2.2 The Loewner Order and the Decision of a Minimiser

In the absence of a conventional ordering of the elements in either \mathbb{R}^3 or $\mathbb{R}^{2 \times 2}$, it becomes necessary to define the relative positions of two elements with respect to one another. In order to respond to this query, it is necessary to utilise a more lenient interpretation of the ordering relation, namely that of a semi-order. For this reason, we resort to the Loewner order and the colour morphological processing based on this, which was already presented in the work [18] by Burgeth and Kleefeld.

Definition 1 We define the **set of symmetric matrices** as

$$\text{Sym}(n) := \{ \mathbf{A} = (a_{ij})_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n} : a_{ij} = a_{ji} \quad \forall i, j = 1, \dots, n \}, \quad n \in \mathbb{N}, \quad (6)$$

and the **set of positive semi-definite symmetric matrices** as

$$\text{Sym}_+(n) := \{ \mathbf{A} \in \text{Sym}(n) : \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n \text{ with } \mathbf{x} \neq \mathbf{0} \}, \quad n \in \mathbb{N}. \quad (7)$$

One can see that $\text{Sym}_+(n)$ is a convex cone, cf. [19]. Thus, this cone induces a partial order in the space of symmetric matrices, which we define as follows.

Definition 2 Let $\mathbf{A}, \mathbf{B} \in \text{Sym}(n)$, $n \in \mathbb{N}$. We define the **Loewner semi-order** \geq_L as follows

$$\mathbf{A} \geq_L \mathbf{B} : \iff \mathbf{A} - \mathbf{B} \in \text{Sym}_+(n). \quad (8)$$

One of the benefits of employing the Loewner order in conjunction with the HCL bi-cone is that the Cartesian coordinates of a given bi-cone vector can be utilised to represent each colour matrix \mathbf{X} as a cone of a specified height within the bi-cone. In this representation, the radius of the cone of \mathbf{X} is given by $\frac{1}{\sqrt{2}} \text{tr} \mathbf{X}$. Therefore, a colour is larger than another colour in the Loewner sense if the cone of the larger colour contains the cone of the smaller colour as a subset. Since the radius is equal to the height of the resulting cones, we only need to find a larger base circle for the corresponding cones, see Figure 3 for a visualising example. For more details, we refer to the paper [18] by Burgeth and Kleefeld.

The problem with this semi-order is that it is not a lattice order [20], and therefore, it is not possible to find a unique maximum or minimum. To get around this problem,

one needs another property to select a uniquely determined maximum from the convex set of symmetric matrices $\mathcal{U}(\mathcal{X})$ that are upper bounds in the Loewner sense for the multi-set $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$ of given symmetric real 2×2 data matrices with

$$\mathcal{U}(\mathcal{X}) := \{ \mathbf{Y} \in \text{Sym}(2) : \mathbf{X} \leq_L \mathbf{Y} \quad \forall \mathbf{X} \in \mathcal{X} \},$$

see [8] for this. In particular, $\mathcal{U}(\mathcal{X})$ is a convex set, since for any $\mathbf{A}, \mathbf{B} \in \mathcal{U}(\mathcal{X})$ one has

$$\alpha \mathbf{A} + (1 - \alpha) \mathbf{B} \geq_L \alpha \mathbf{X} + (1 - \alpha) \mathbf{X} = \mathbf{X} \quad \forall \mathbf{X} \in \mathcal{X}, \alpha \in [0, 1]$$

and the linear combination of two symmetric matrices always result in a symmetric matrix. For representation of the previously mentioned property, we use the function $\varphi : \mathcal{U}(\mathcal{X}) \rightarrow \mathbb{R}$, which should be convex on the set $\mathcal{U}(\mathcal{X})$ and Loewner-monotone, i.e.

$$\varphi(\mathbf{A}) \leq \varphi(\mathbf{B}) \iff \mathbf{A} \leq_L \mathbf{B}.$$

To clarify this terminology of convexity, we introduce the following definition.

Definition 3 Let V_1, V_2 be two real-valued vector spaces, $M \subset V_1$ a convex set and $K \subset V_2$ an order cone, i.e. $K := \{ \mathbf{x} \in V_2 : \mathbf{x} \geq 0 \}$ on the ordered vector space (V_2, \geq) . A function $f : M \rightarrow V_2$ is called **convex on the set M** if and only if

$$\alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) - f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \in K \quad \forall \mathbf{x}, \mathbf{y} \in M, \alpha \in [0, 1]. \quad (9)$$

Furthermore, φ should have a unique minimiser in $\mathcal{U}(\mathcal{X})$; thus, we can define the φ -supremum of \mathcal{X} as said minimiser:

$$\text{Sup}_\varphi(\mathcal{X}) := \arg \min_{\mathbf{Y} \in \mathcal{U}(\mathcal{X})} \varphi(\mathbf{Y}). \quad (10)$$

The matrix supremum introduced in the paper [12] is based on the calculation of the trace and is therefore also called trace supremum. That is, one has $\varphi(\mathbf{Y}) = \text{tr} \mathbf{Y}$ and we get as supremum:

$$\text{Sup}_{\text{tr}}(\mathcal{X}) := \arg \min_{\mathbf{Y} \in \mathcal{U}(\mathcal{X})} \text{tr} \mathbf{Y}.$$

Based on the corresponding norms, the Frobenius supremum

$$\text{Sup}_2(\mathcal{X}) := \arg \min_{\mathbf{Y} \in \mathcal{U}(\mathcal{X})} \sum_{\mathbf{X} \in \mathcal{X}} \|\mathbf{Y} - \mathbf{X}\|_2$$

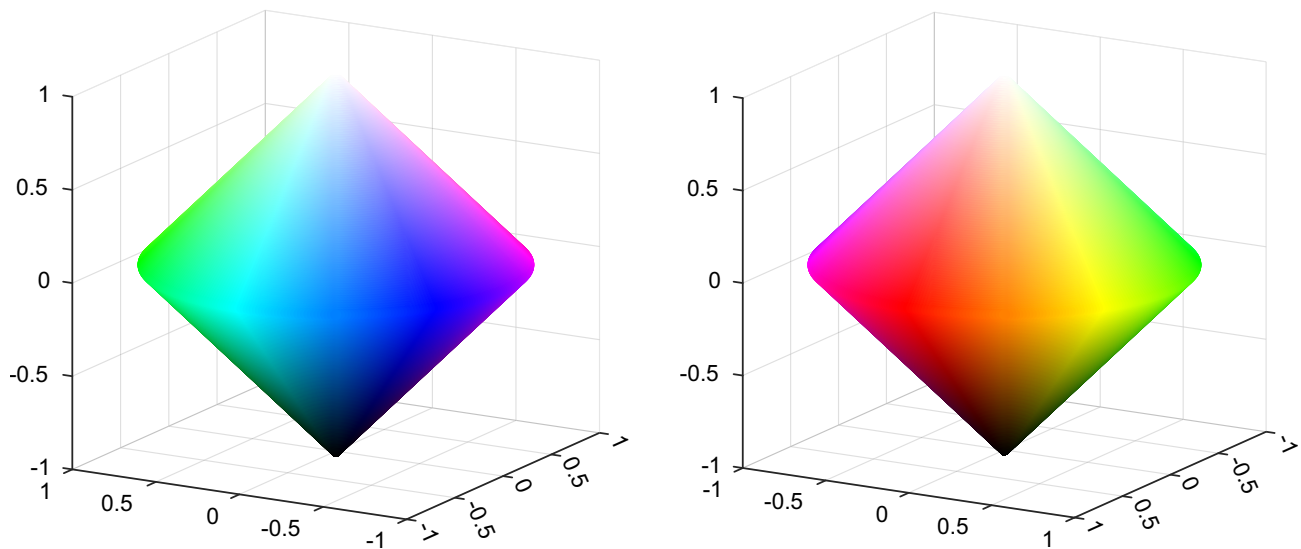


Fig. 2 HCL bi-cone according to [18]

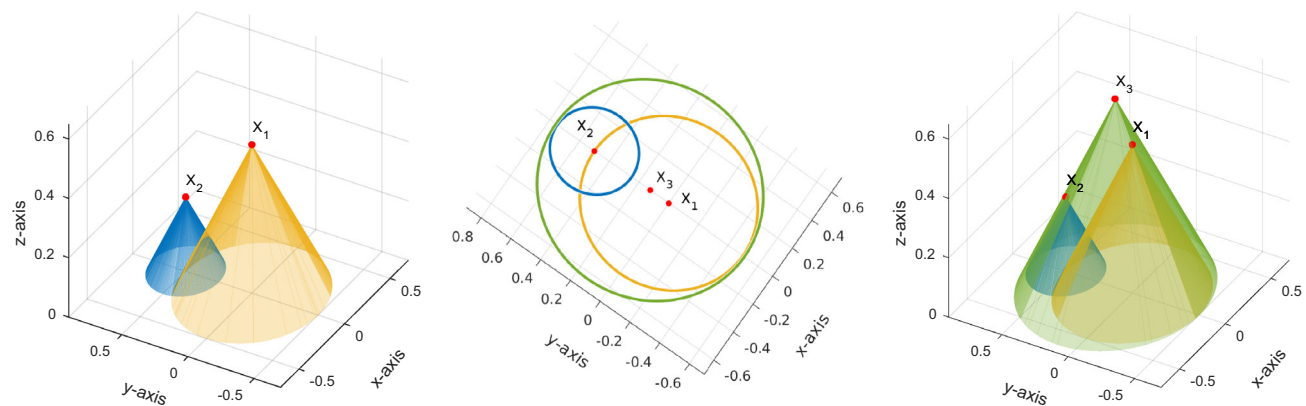


Fig. 3 Visualisation of the Loewner ordering in the HCL bi-cone for the three colour matrices X_i , $i = 1, 2, 3$, with $X_3 \geq_L X_1$ and $X_3 \geq_L X_2$. **From left to right:** The cone representation of a yellow colour X_1 and

a cyan–blue colour X_2 , the base circles of X_1 and X_2 with a third base circle of a green colour X_3 that encompasses both of them, the cone representation of all three colours

and the spectral supremum

$$\text{Sup}_\infty(\mathcal{X}) := \arg \min_{Y \in \mathcal{U}(\mathcal{X})} \sum_{X \in \mathcal{X}} |\lambda_1(Y - X)|,$$

where $\lambda_1(A)$ denotes the largest eigenvalue of A , were derived in [8]. At this point, it should be noted that, in the case of positive semi-definite matrices, all three norms are Schatten norms $\|\cdot\|_p$ for $p \in \{1, 2, \infty\}$.

In contrast to the above approaches utilising norms, which compare matrices that are upper bounds, we will adopt a more direct approach. To accomplish this, we will employ an approximation of the maximum function by Maslov [9] as a matrix-valued function to directly calculate a matrix that acts as an upper bound. The subsequent section aims to present a detailed characterisation of this approach.

3 Characterisation of the Log-Exp-Supremum

In this section, we will construct a characterisation for the log-exp-supremum that depends solely on the input data, albeit not continuously. This characterisation is based on the spectral decomposition of the matrices under consideration and some properties of symmetric 2×2 matrices and the Rayleigh product. Given that this characterisation is based on spectral decomposition, this section will be divided into three parts. The first part will consider the case where there is a unique largest eigenvalue among the input data. The second part will consider the case where there is no unique largest eigenvalue. The third and final part will consider a general case that includes the other two cases and some associated

properties. In particular, the log-exp-supremum is transitive, a property that will be demonstrated at the conclusion of this section.

Definition 4 We define for a multi-set $\mathcal{X} = \{X_1, \dots, X_n\}$, $n \in \mathbb{N}$, of symmetric real 2×2 matrices the **log-exp-supremum (LES)** as

$$S := \text{Sup}_{\text{LE}}(\mathcal{X}) := \lim_{m \rightarrow \infty} \left(\frac{1}{m} \log \sum_{i=1}^n \exp(mX_i) \right). \quad (11)$$

Remark 1 The log-exp-infimum follows by duality:

$$\text{Inf}_{\text{LE}}(\mathcal{X}) := - \lim_{m \rightarrow \infty} \left(\frac{1}{m} \log \sum_{i=1}^n \exp(-mX_i) \right). \quad (12)$$

Furthermore, we explain the spectral decomposition of X_i by

$$\begin{aligned} X_i &= \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \mu_i \mathbf{v}_i \mathbf{v}_i^T, \quad \lambda_i \geq \mu_i, \quad \langle \mathbf{u}_i, \mathbf{v}_i \rangle = 0, \\ |\mathbf{u}_i| &= 1 = |\mathbf{v}_i|, \quad \mathbf{u}_i = (c_i, s_i)^T, \quad \mathbf{v}_i = (-s_i, c_i)^T, \\ c_i &= \cos(\varphi_i), \quad s_i = \sin(\varphi_i), \quad \varphi_i \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \\ i &= 1, \dots, n, \end{aligned} \quad (13)$$

where $\lambda_i, \mu_i \in \mathbb{R}$ are the eigenvalues of X_i and $\mathbf{u}_i \perp \mathbf{v}_i$ are the normalised eigenvectors to the corresponding eigenvalues.

3.1 LES for a Unique Largest Eigenvalue

We will now assume that

$$\lambda_1 \text{ is the unique largest eigenvalue among all eigenvalues of } \mathcal{X} \quad (14)$$

and further that without loss of generality

$$\text{the eigenvectors of } X_1 \text{ have } \varphi_1 = 0, \quad (15)$$

i.e. they are axis-aligned and fulfil $\mathbf{u}_1 = \mathbf{e}_1 := (1, 0)^T$ and $\mathbf{v}_1 = \mathbf{e}_2 := (0, 1)^T$. Using these assumptions, we calculate the matrix exponential according to the general application of functions to diagonalisable matrices, see [21]:

$$\begin{aligned} \exp(mX_i) &= \exp \left(m \lambda_i \begin{pmatrix} c_i^2 & c_i s_i \\ c_i s_i & s_i^2 \end{pmatrix} + m \mu_i \begin{pmatrix} s_i^2 & -c_i s_i \\ -c_i s_i & c_i^2 \end{pmatrix} \right) \\ &= \begin{pmatrix} c_i^2 e^{m\lambda_i} + s_i^2 e^{m\mu_i} & c_i s_i (e^{m\lambda_i} - e^{m\mu_i}) \\ c_i s_i (e^{m\lambda_i} - e^{m\mu_i}) & s_i^2 e^{m\lambda_i} + c_i^2 e^{m\mu_i} \end{pmatrix} \end{aligned}$$

and the summation over all $i = 1, \dots, n$ leads to

$$E_m := \sum_{i=1}^n \exp(mX_i) = \begin{pmatrix} e_{m11} & e_{m12} \\ e_{m21} & e_{m22} \end{pmatrix}$$

with

$$\begin{aligned} e_{m11} &= e^{m\lambda_1} + \sum_{i=2}^n (c_i^2 e^{m\lambda_i} + s_i^2 e^{m\mu_i}), \\ e_{m12} &= \sum_{i=2}^n c_i s_i (e^{m\lambda_i} - e^{m\mu_i}) = e_{m21} \text{ and} \\ e_{m22} &= e^{m\mu_1} + \sum_{i=2}^n (s_i^2 e^{m\lambda_i} + c_i^2 e^{m\mu_i}). \end{aligned}$$

Next, we determine the major eigenvector for this matrix when m tends to infinity.

Lemma 1 Let the conditions (14) and (15) be fulfilled according to (13). Then $\mathbf{u}_1 = (1, 0)^T$ represents the major eigenvector of E_m for $m \rightarrow \infty$.

Proof We declare the Rayleigh product as

$$R_{E_m}(\mathbf{w}) := \langle \mathbf{w}, E_m \mathbf{w} \rangle, \quad \mathbf{w} \in \mathbb{R}^2,$$

and consider it for vectors

$$\mathbf{w}_\varepsilon = \frac{1}{\sqrt{1+\varepsilon^2}} \begin{pmatrix} 1 \\ \varepsilon \end{pmatrix}, \quad |\varepsilon| \ll 1,$$

which vary in the second component with ε around \mathbf{u}_1 . We determine

$$\begin{aligned} R_{E_m}(\mathbf{w}_\varepsilon) &= \frac{1}{1+\varepsilon^2} \left(e^{m\lambda_1} + \varepsilon^2 e^{m\mu_1} + \sum_{i=2}^n ((c_i + \varepsilon s_i)^2 e^{m\lambda_i} + (s_i - \varepsilon c_i)^2 e^{m\mu_i}) \right) \\ &=: \frac{1}{1+\varepsilon^2} (e^{m\lambda_1} + h_m(\varepsilon)). \end{aligned} \quad (16)$$

If $m \rightarrow \infty$, we observe

$$\lim_{m \rightarrow \infty} \frac{h_m(\varepsilon)}{e^{m\lambda_1}} = 0, \quad (17)$$

because we assumed that λ_1 is the unique largest eigenvalue and $|\varepsilon| \ll 1$. By considering

$$R_{E_m}(\mathbf{w}_0) = e^{m\lambda_1} + \sum_{i=2}^n (c_i^2 e^{m\lambda_i} + s_i^2 e^{m\mu_i}),$$

we obtain

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{R_{E_m}(\mathbf{w}_\varepsilon)}{R_{E_m}(\mathbf{w}_0)} &= \lim_{m \rightarrow \infty} \frac{e^{m\lambda_1} + h_m(\varepsilon)}{(1 + \varepsilon^2)(e^{m\lambda_1} + h_m(0))} \\ &= \lim_{m \rightarrow \infty} \left(\frac{1}{1 + \varepsilon^2} \right. \\ &\quad + \frac{\varepsilon^2 e^{m\mu_1} + \sum_{i=2}^n ((2\varepsilon c_i s_i + \varepsilon^2 s_i^2) e^{m\lambda_i})}{(1 + \varepsilon^2)(e^{m\lambda_1} + h_m(0))} \\ &\quad \left. + \frac{\sum_{i=2}^n ((-2\varepsilon c_i s_i + \varepsilon^2 c_i^2) e^{m\mu_i})}{(1 + \varepsilon^2)(e^{m\lambda_1} + h_m(0))} \right) \\ &=: \frac{1}{1 + \varepsilon^2} \lim_{m \rightarrow \infty} \left(1 + \frac{f_m^{(1)}(\varepsilon)}{f_m^{(2)}} \right) \\ &=: \frac{1}{1 + \varepsilon^2} \lim_{m \rightarrow \infty} (1 + f_m(\varepsilon)). \end{aligned}$$

We know that $f_m(\varepsilon) \geq 0$ for all $m > 0$, since $\lambda_i \geq \mu_i$ for $i = 2, \dots, n$. Furthermore, we set

$$g_m(\varepsilon) := \sum_{i=2}^n \left((c_i + \varepsilon s_i)^2 e^{m\lambda_i} + (s_i - \varepsilon c_i)^2 e^{m\mu_i} \right)$$

and estimate $f_m^{(1)}(\varepsilon) \leq \varepsilon^2 e^{m\mu_1} + g_m(\varepsilon)$. We therefore have the following estimation:

$$\begin{aligned} 0 &\leq \lim_{m \rightarrow \infty} f_m(\varepsilon) = \lim_{m \rightarrow \infty} \frac{f_m^{(1)}(\varepsilon)}{f_m^{(2)}} \\ &\leq \lim_{m \rightarrow \infty} \frac{\varepsilon^2 e^{m\mu_1} + g_m(\varepsilon)}{f_m^{(2)}} \stackrel{(17)}{=} 0, \end{aligned}$$

which leads to

$$\lim_{m \rightarrow \infty} \frac{R_{E_m}(\mathbf{w}_\varepsilon)}{R_{E_m}(\mathbf{w}_0)} = \frac{1}{1 + \varepsilon^2} (1 + 0) = \frac{1}{1 + \varepsilon^2}.$$

This means that the function $\lim_{m \rightarrow \infty} \frac{R_{E_m}(\mathbf{w}_\varepsilon)}{R_{E_m}(\mathbf{w}_0)}$ has a local maximum at $\varepsilon = 0$ and thus concludes the proof. \square

To interpret this lemma, we declare the major eigenvector of E_m as $\bar{\mathbf{u}}_m$. Then we obtain from the lemma the equation $\lim_{m \rightarrow 0} \bar{\mathbf{u}}_m = \mathbf{u}_1$, and since the eigenvectors of $\frac{1}{m} \log E_m$ are the same as those of E_m , it also follows that \mathbf{u}_1 is the major eigenvector of S . Since all considered matrices are 2×2 matrices, we can deduce the second normalised eigenvector of S as the vector that is orthogonal to the first eigenvector $\mathbf{u}_1 = (1, 0)^T$, namely $\mathbf{v}_1 = (0, 1)^T$. As the next step, we will determine the eigenvalue corresponding to \mathbf{u}_1 .

Lemma 2 *Let the conditions of Lemma 1 be fulfilled. Then the eigenvector \mathbf{u}_1 of S according to (11) has the corresponding eigenvalue λ_1 .*

Proof Since the largest eigenvalue of E_m is just expressed by the value of the Rayleigh product at the point where it takes its maximum, i.e. at \mathbf{w}_ε , as we have established in the proof of Lemma 2, we obtain as major eigenvalue (16). This means that the major eigenvalue of $\frac{1}{m} \log E_m$ is

$$\begin{aligned} \frac{1}{m} \log R_{E_m}(\mathbf{w}_\varepsilon) &= \frac{1}{m} \log \left[\frac{1}{1 + \varepsilon^2} (e^{m\lambda_1} + h_m(\varepsilon)) \right] \\ &= \frac{1}{m} \log \left[\frac{e^{m\lambda_1}}{1 + \varepsilon^2} (1 + h_m(\varepsilon) e^{-m\lambda_1}) \right] \\ &= \underbrace{\frac{1}{m} \log (1 + \underbrace{h_m(\varepsilon) e^{-m\lambda_1}}_{\rightarrow 0 \text{ (} m \rightarrow \infty)})}_{\rightarrow 0 \text{ (} m \rightarrow \infty)} - \underbrace{\frac{1}{m} \log (1 + \varepsilon^2)}_{\rightarrow 0 \text{ (} m \rightarrow \infty)} \\ &\quad + \frac{1}{m} \log e^{m\lambda_1}, \end{aligned}$$

where the second term in the first log function vanishes in the limit because μ_1, λ_i, μ_i for $i = 2, \dots, n$ are smaller than λ_1 , see (17). Thus, the major eigenvalue of S is just

$$\lim_{m \rightarrow \infty} \left(\frac{1}{m} \log R_{E_m}(\mathbf{w}_\varepsilon) \right) = \lim_{m \rightarrow \infty} \left(\frac{1}{m} \log e^{m\lambda_1} \right) = \lambda_1.$$

\square

The final step in the characterisation of S is to calculate the second eigenvalue corresponding to \mathbf{v}_1 , but for this we need a refinement of our calculations in the proof of Lemma 1. To achieve this, we establish the following lemma:

Lemma 3 *Let the conditions of Lemma 1 be fulfilled and additionally $\varepsilon(m)$ be the value of ε for which $\mathbf{v}_m = \mathbf{w}_\varepsilon$ holds for every $m > 0$. Then $\lim_{m \rightarrow \infty} \varepsilon(m) = 0$ with $\varepsilon(m) \sim e^{m(\lambda - \lambda_1)}$, where $\lambda = \max(\mu_1, \lambda_2, \dots, \lambda_n)$.*

Proof In the proof of Lemma 1, we have already seen that $\varepsilon \rightarrow 0$ for $m \rightarrow \infty$ and that the Rayleigh product fulfils

$$R(\varepsilon) := R_{E_m}(\mathbf{w}_\varepsilon) = \langle \mathbf{w}_\varepsilon, E_m \mathbf{w}_\varepsilon \rangle \in C^\infty(\mathbb{R}).$$

Since the critical point is at $\varepsilon \rightarrow 0$, we want to approximate this point by a Taylor approximation of order two:

$$T_2 R(\varepsilon, 0) = R(0) + R'(0)\varepsilon + \frac{1}{2} R''(0)\varepsilon^2. \quad (18)$$

For this, we determine

$$\begin{aligned} R(0) &= e^{m\lambda_1} + \sum_{i=2}^n \left(c_i^2 e^{m\lambda_i} + s_i^2 e^{m\mu_i} \right), \\ R'(\varepsilon) &= -\frac{2\varepsilon}{(1 + \varepsilon^2)^2} \left(R(\varepsilon)(1 + \varepsilon^2) \right) + \frac{1}{1 + \varepsilon^2} \end{aligned}$$

$$\begin{aligned}
& \left(2\epsilon e^{m\mu_1} + \sum_{i=2}^n \left((2s_i c_i + 2\epsilon s_i^2) e^{m\lambda_i} \right. \right. \\
& \quad \left. \left. + (-2s_i c_i + 2\epsilon c_i^2) e^{m\mu_i} \right) \right) \\
&= -\frac{2\epsilon}{1+\epsilon^2} R(\epsilon) + \frac{2\epsilon}{1+\epsilon^2} e^{m\mu_1} \\
& \quad + \sum_{i=2}^n \frac{2(c_i + \epsilon s_i) s_i}{1+\epsilon^2} e^{m\lambda_i} \\
& \quad + \sum_{i=2}^n \frac{2(-s_i + \epsilon c_i) c_i}{1+\epsilon^2} e^{m\mu_i}, \\
R'(0) &= \sum_{i=2}^n 2c_i s_i e^{m\lambda_i} - \sum_{i=2}^n 2s_i c_i e^{m\mu_i} \\
&= 2 \sum_{i=2}^n c_i s_i (e^{m\lambda_i} - e^{m\mu_i}), \\
R''(\epsilon) &= \frac{-2(1+\epsilon^2) + 4\epsilon^2}{(1+\epsilon^2)^2} R(\epsilon) - \frac{2\epsilon}{1+\epsilon^2} R'(\epsilon) \\
& \quad + \frac{-2(1+\epsilon^2) + 4\epsilon^2}{(1+\epsilon^2)^2} e^{m\mu_1} \\
& \quad + \sum_{i=2}^n \frac{2s_i^2(1+\epsilon^2) - 4\epsilon(c_i + \epsilon s_i) s_i}{(1+\epsilon^2)^2} e^{m\lambda_i} \\
& \quad + \sum_{i=2}^n \frac{2c_i^2(1+\epsilon^2) + 4\epsilon(s_i - \epsilon c_i) c_i}{(1+\epsilon^2)^2} e^{m\mu_i} \\
&= -\frac{2(1-\epsilon^2)}{(1+\epsilon^2)^2} R(\epsilon) - \frac{2\epsilon}{1+\epsilon^2} R'(\epsilon) \\
& \quad + \frac{2(1-\epsilon^2)}{(1+\epsilon^2)^2} e^{m\mu_1} \\
& \quad + \sum_{i=2}^n \frac{2(1-\epsilon^2)s_i^2 - 4\epsilon c_i s_i}{(1+\epsilon^2)^2} e^{m\lambda_i} \\
& \quad + \sum_{i=2}^n \frac{2(1-\epsilon^2)c_i^2 + 4\epsilon c_i s_i}{(1+\epsilon^2)^2} e^{m\mu_i}, \\
R''(0) &= -2R(0) + 2e^{m\mu_1} + \sum_{i=2}^n 2s_i^2 e^{m\lambda_i} \\
& \quad + \sum_{i=2}^n 2c_i^2 e^{m\mu_i} \\
&= -2e^{m\lambda_1} - 2 \sum_{i=2}^n (c_i^2 e^{m\lambda_i} + s_i^2 e^{m\mu_i}) \\
& \quad + 2e^{m\mu_1} + 2 \sum_{i=2}^n (s_i^2 e^{m\lambda_i} + c_i^2 e^{m\mu_i}) \\
&= -2e^{m\lambda_1} + 2e^{m\mu_1} - 2 \sum_{i=2}^n (c_i^2 - s_i^2) e^{m\lambda_i}
\end{aligned}$$

$$- 2 \sum_{i=2}^n (s_i^2 - c_i^2) e^{m\mu_i}$$

and by setting $\lambda := \max(\mu_1, \lambda_2, \dots, \lambda_n)$, we obtain

$$\begin{aligned}
R(0) &= e^{m\lambda_1} \left(1 + \sum_{i=2}^n (c_i^2 e^{m(\lambda_i - \lambda_1)} \right. \\
& \quad \left. + s_i^2 e^{m(\mu_i - \lambda_1)}) \right) \\
&= e^{m\lambda_1} \left(1 + \mathcal{O}(e^{m(\lambda - \lambda_1)}) \right), \\
R'(0) &= \Theta(e^{m\lambda}), \\
R''(0) &= -2e^{m\lambda_1} \left(1 - e^{m(\mu_1 - \lambda_1)} \right. \\
& \quad \left. + \sum_{i=2}^n (c_i^2 - s_i^2) e^{m(\lambda_i - \lambda_1)} \right. \\
& \quad \left. + \sum_{i=2}^n (s_i^2 - c_i^2) e^{m(\mu_i - \lambda_1)} \right) \\
&= -2e^{m\lambda_1} \left(1 + \mathcal{O}(e^{m(\lambda - \lambda_1)}) \right).
\end{aligned}$$

In this context, the symbols \mathcal{O} and Θ represent the corresponding Landau symbols. The term $\mathcal{O}(\cdot)$ symbolises an upper bound for the asymptotic order, while the term $\Theta(\cdot)$ denotes the exact asymptotic order. We will now determine the maximum of (18) by means of differentiation:

$$\begin{aligned}
R'(\epsilon) &= \frac{d}{d\epsilon} (T_2 R(\epsilon, 0) + \mathcal{O}(\epsilon^3)) \\
&= R'(0) + R''(0)\epsilon + 3\mathcal{O}(\epsilon^2) \stackrel{!}{=} 0 \\
\iff \epsilon &= -\frac{R'(0) + 3\mathcal{O}(\epsilon^2)}{R''(0)} \\
&= -\frac{\Theta(e^{m\lambda}) + 3\mathcal{O}(\epsilon^2)}{-2e^{m\lambda_1} (1 + \mathcal{O}(e^{m(\lambda - \lambda_1)}))} \\
\iff \epsilon &= \frac{1}{2\Theta(e^{m(\lambda_1 - \lambda)}) (1 + \mathcal{O}(e^{m(\lambda - \lambda_1)}))} \\
& \quad + \frac{3\mathcal{O}(\epsilon^2)}{2e^{m\lambda_1} (1 + \mathcal{O}(e^{m(\lambda - \lambda_1)}))} \\
\iff \epsilon &= \frac{1}{\Theta(e^{m(\lambda_1 - \lambda)}) + \mathcal{O}(1)} \\
& \quad + \frac{3\mathcal{O}(\epsilon^2)}{2e^{m\lambda_1} (1 + \mathcal{O}(e^{-m(\lambda_1 - \lambda)}))}.
\end{aligned}$$

Since we can add a constant value to all eigenvalues and subtract it again later when we have formed the LES, we can assume without loss of generality that $\lambda_1 = 1$. If we let $m \rightarrow \infty$, then the first fraction will converge to zero and the second fraction will diverge in the denominator to infinity,

because $\lambda_1 > \lambda$. This means, we have $\varepsilon \rightarrow 0$ for $m \rightarrow \infty$. Together with $R''(0) < 0$ it follows that $\varepsilon \rightarrow 0$ represents a maximum of $R(\varepsilon)$, which concludes this proof. \square

With this enhancement of our previous calculation of Lemma 1, we are in the position to determine the second eigenvalue that is needed for the characterisation of \mathbf{S} . To this end, we note the next lemma.

Lemma 4 *Let the conditions of Lemma 3 be fulfilled. Then, the eigenvalue of \mathbf{S} belonging to the eigenvector \mathbf{v}_1 is the largest eigenvalue of any of the matrices \mathbf{X}_i , $i = 1, \dots, n$, whose eigenvector is not aligned with \mathbf{u}_1 .*

Proof 1.) Analogous to the proof of Lemma 1, we start with the Rayleigh product of \mathbf{E}_m , but with an eigenvector that is perpendicular to our first “test eigenvector” \mathbf{w}_ε , namely $\bar{\mathbf{w}}_\varepsilon = \mathbf{w}_\varepsilon^\perp = \frac{1}{\sqrt{1+\varepsilon^2}}(-\varepsilon, 1)^T$:

$$\begin{aligned}\bar{R}(\varepsilon) &= R_{\mathbf{E}_m}(\bar{\mathbf{w}}_\varepsilon) = \langle \bar{\mathbf{w}}_\varepsilon, \mathbf{E}_m \bar{\mathbf{w}}_\varepsilon \rangle \\ &= \frac{\varepsilon^2 e^{m\lambda_1} + e^{m\mu_1}}{1 + \varepsilon^2} \\ &\quad + \sum_{i=2}^n \frac{(s_i - \varepsilon c_i)^2 e^{m\lambda_i} + (c_i + \varepsilon s_i)^2 e^{m\mu_i}}{1 + \varepsilon^2}.\end{aligned}$$

2.) Let us first assume that the second largest eigenvalue is λ_2 and $s_2 \neq 0$, so that the eigenvector \mathbf{u}_2 is not aligned with the first eigenvector \mathbf{u}_1 . Then we can rewrite this equation as

$$\begin{aligned}\bar{R}(\varepsilon) &= \frac{e^{m\lambda_2}}{1 + \varepsilon^2} \left((s_2 - \varepsilon c_2)^2 + \varepsilon^2 e^{m(\lambda_1 - \lambda_2)} \right. \\ &\quad \left. + e^{m(\mu_1 - \lambda_2)} + (c_2 + \varepsilon s_2)^2 e^{m(\mu_2 - \lambda_2)} \right. \\ &\quad \left. + \sum_{i=3}^n \left((s_i - \varepsilon c_i)^2 e^{m(\lambda_i - \lambda_2)} \right. \right. \\ &\quad \left. \left. + (c_i + \varepsilon s_i)^2 e^{m(\mu_i - \lambda_2)} \right) \right).\end{aligned}$$

If we only look at the terms in the big brackets, we find that only $(s_2 - \varepsilon c_2)^2$ is constant with respect to m , while all the other terms, with the exception of $\varepsilon^2 e^{m(\lambda_1 - \lambda_2)}$, involve exponential functions with negative multiples of m , which will approach zero for $m \rightarrow \infty$. We conclude from Lemma 3 that

$$\begin{aligned}\varepsilon \sim e^{m(\lambda_2 - \lambda_1)} &\iff \varepsilon^{-1} \sim e^{m(\lambda_1 - \lambda_2)} \\ &\iff \varepsilon^2 e^{m(\lambda_1 - \lambda_2)} \sim \varepsilon \sim e^{m(\lambda_2 - \lambda_1)}\end{aligned}$$

and we also know that $\varepsilon \rightarrow 0$ for $m \rightarrow \infty$. So will $\varepsilon^2 e^{m(\lambda_1 - \lambda_2)} \rightarrow 0$ for $m \rightarrow \infty$. The remaining terms we

consider are

$$\begin{aligned}\bar{R}(\varepsilon) &= \frac{e^{m\lambda_2}}{1 + \varepsilon^2} \left((s_2 - \varepsilon c_2)^2 \right. \\ &\quad \left. + \mathcal{O}(e^{m(\lambda_2 - \lambda_1)} + e^{m(\mu - \lambda_2)}) \right),\end{aligned}\tag{19}$$

where μ is the next largest eigenvalue after λ_2 . From there, we calculate

$$\begin{aligned}\frac{1}{m} \log \bar{R}(\varepsilon) &= \lambda_2 - \frac{1}{m} \log(1 + \varepsilon^2) + \frac{1}{m} \log \left((s_2 - \varepsilon c_2)^2 \right. \\ &\quad \left. + \mathcal{O}(e^{m(\lambda_2 - \lambda_1)} + e^{m(\mu - \lambda_2)}) \right) \\ &= \lambda_2 - \frac{\log(1 + \varepsilon^2)}{m} + \frac{\log((s_2 - \varepsilon c_2)^2)}{m} \\ &\quad + \frac{\log \left(1 + \mathcal{O}(e^{m(\lambda_2 - \lambda_1)} + e^{m(\mu - \lambda_2)}) \right)}{m}\end{aligned}$$

and for $m \rightarrow \infty$ we obtain that $\frac{1}{m} \log \bar{R}(\varepsilon) \rightarrow \lambda_2$, since ε goes exponentially to 0 for $m \rightarrow \infty$. This proves this case.

3.) Let us now assume that μ_1 is the second largest eigenvalue. In this case, the results of Lemma 3 would be obtained with $\lambda = \mu_1$, and the same procedure as that employed in the current proof would be followed, but with the exclusion of the μ_1 term in place of the λ_2 term. Thereby, we would achieve

$$\bar{R}(\varepsilon) = \frac{e^{m\mu_1}}{1 + \varepsilon^2} \left(1 + \mathcal{O}(e^{m(\mu_1 - \lambda_1)} + e^{m(\mu - \mu_1)}) \right)$$

for the leading terms in (19), where μ is the next largest eigenvalue after μ_1 . The rest is done analogously.

4.) Now we assume that λ_2 has the same eigenvector as λ_1 . Then it follows that $s_2 = 0$, and therefore, $(s_2 - \varepsilon c_2)^2 \rightarrow 0$ for $m \rightarrow \infty$ in equation (19). Because of that we would extract the next smaller eigenvalue and repeat the above calculation until we find an eigenvalue whose eigenvector is not \mathbf{u}_1 . Then we would do the same proof with the term of this eigenvalue instead the λ_2 term.

5.) The final case would be that the second largest eigenvalue λ_2 is not unique. This would result in further terms like $(s_2 - \varepsilon c_2)^2$ in (19) for the other eigenvalues equal to λ_2 . Since these terms remain $\mathcal{O}(1)$, they will not change the result of the proof. \square

We summarise our previous findings on our first theorem:

Theorem 1 *Let $\mathcal{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$, $n \in \mathbb{N}$, be a multi-set of symmetric real 2×2 matrices and λ_1 be the unique largest*

eigenvalue of all matrices of \mathcal{X} with the corresponding normalised eigenvector \mathbf{u}_1 . Then the log-exp-supremum of \mathcal{X} has the representation

$$\begin{aligned} S &:= \lim_{m \rightarrow \infty} \left(\frac{1}{m} \log \sum_{i=1}^n \exp(mX_i) \right) \\ &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \mu_* \mathbf{v}_1 \mathbf{v}_1^T, \end{aligned} \quad (20)$$

where μ_* is the next largest eigenvalue among all matrices of \mathcal{X} whose normalised eigenvector \mathbf{v}_* is not aligned with the eigenvector \mathbf{u}_1 and \mathbf{v}_1 is the normalised eigenvector perpendicular to \mathbf{u}_1 .

Proof A rotation matrix $\mathbf{R}_{-\varphi_1}$ is employed to rotate all eigenvectors in \mathcal{X} , with the rotation angle represented by $-\varphi_1$. This results in the transformation of \mathbf{X}_1 into a diagonal matrix with diagonal entries λ_1 and μ_1 . Thereby, assumption (15) is fulfilled and we can apply Lemma 1–4 to the rotated input matrices $\mathcal{X}_{-\varphi_1}$. For the LES, we get

$$S_{-\varphi_1} := \text{Sup}_{\text{LE}}(\mathcal{X}_{-\varphi_1}) = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \mu_* \mathbf{e}_2 \mathbf{e}_2^T,$$

where μ_* is the next largest eigenvalue among all matrices of \mathcal{X} whose normalised eigenvector \mathbf{v}_* is not aligned with the eigenvector $\mathbf{e}_1 = (1, 0)^T$, and $\mathbf{e}_2 = (0, 1)^T$. We achieve the LES of the original \mathcal{X} by rotating all of the eigenvectors back with \mathbf{R}_{φ_1} :

$$\begin{aligned} S &= \mathbf{R}_{\varphi_1} S_{-\varphi_1} \mathbf{R}_{\varphi_1}^T = \mathbf{R}_{\varphi_1} (\lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \mu_* \mathbf{e}_2 \mathbf{e}_2^T) \mathbf{R}_{\varphi_1}^T \\ &= \lambda_1 (\mathbf{R}_{\varphi_1} \mathbf{e}_1) (\mathbf{R}_{\varphi_1} \mathbf{e}_1)^T + \mu_* (\mathbf{R}_{\varphi_1} \mathbf{e}_2) (\mathbf{R}_{\varphi_1} \mathbf{e}_2)^T \\ &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \mu_* \mathbf{v}_1 \mathbf{v}_1^T. \end{aligned}$$

□

Before we look at the case where the largest eigenvalue is not unique, let us calculate a small example with numerical values for better understanding.

Example 1 We consider as RGB colours blue $\mathbf{C}_1 = (0, 0, 1)$, a medium dark brown $\mathbf{C}_2 = (\frac{3}{5}, \frac{2}{5}, \frac{1}{5})$ and a shade of blue–magenta $\mathbf{C}_3 = (\frac{1}{3}, \frac{1}{3}, \frac{5}{6})$ and calculate the LES for them. To do this, we first convert the RGB colours into symmetric matrices as described in Section 2.1:

$$\begin{aligned} \mathbf{X}_1 &= \frac{1}{2\sqrt{2}} \begin{pmatrix} \sqrt{3} & -1 \\ -1 & -\sqrt{3} \end{pmatrix} \\ &\approx \begin{pmatrix} 0.6124 & -0.3535 \\ -0.3535 & -0.6124 \end{pmatrix}, \\ \mathbf{X}_2 &= \frac{1}{5\sqrt{2}} \begin{pmatrix} -2 & \sqrt{3} \\ \sqrt{3} & 0 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} &\approx \begin{pmatrix} -0.2828 & 0.2450 \\ 0.2450 & 0 \end{pmatrix} \quad \text{and} \\ \mathbf{X}_3 &= \frac{1}{12\sqrt{2}} \begin{pmatrix} 2 + 3\sqrt{3} & -3 \\ -3 & 2 - 3\sqrt{3} \end{pmatrix} \\ &\approx \begin{pmatrix} 0.4240 & -0.1768 \\ -0.1768 & -0.1883 \end{pmatrix}. \end{aligned}$$

Then, we form the spectral decomposition (13) with the eigenvalues λ_i , μ_i and eigenvectors \mathbf{u}_i , \mathbf{v}_i for $i = 1, 2, 3$:

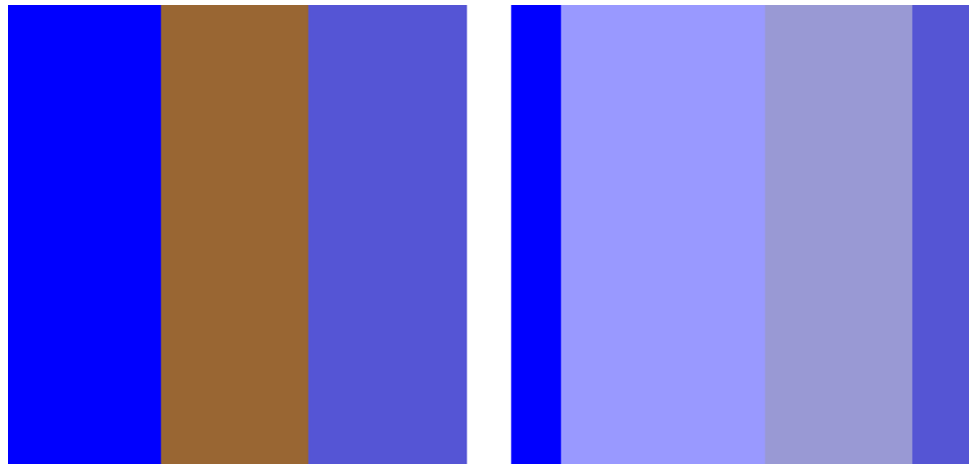
$$\begin{aligned} \lambda_1 &= \frac{1}{\sqrt{2}} \approx 0.7071, \quad \mu_1 = -\lambda_1, \\ \mathbf{u}_1 &= \frac{1}{\sqrt{8 + 4\sqrt{3}}} \begin{pmatrix} -2 - \sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.9659 \\ 0.2588 \end{pmatrix}, \\ \mathbf{v}_1 &= \frac{1}{\sqrt{8 - 4\sqrt{3}}} \begin{pmatrix} 2 - \sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.2588 \\ -0.9659 \end{pmatrix}, \\ \lambda_2 &= \frac{1}{5\sqrt{2}} \approx 0.1414, \quad \mu_2 = -\frac{3}{5\sqrt{2}} \approx -0.4243, \\ \mathbf{u}_2 &= \frac{1}{2} \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix} \approx \begin{pmatrix} 0.5000 \\ 0.8660 \end{pmatrix}, \\ \mathbf{v}_2 &= \frac{1}{2} \begin{pmatrix} -\sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.8660 \\ 0.5000 \end{pmatrix} \quad \text{and} \\ \lambda_3 &= \frac{2}{3\sqrt{2}} \approx 0.4714, \quad \mu_3 = -\frac{1}{3\sqrt{2}} \approx -0.2357, \\ \mathbf{u}_3 &= \frac{1}{\sqrt{8 + 4\sqrt{3}}} \begin{pmatrix} -2 - \sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.9659 \\ 0.2588 \end{pmatrix}, \\ \mathbf{v}_3 &= \frac{1}{\sqrt{8 - 4\sqrt{3}}} \begin{pmatrix} 2 - \sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.2588 \\ 0.9659 \end{pmatrix}. \end{aligned}$$

The largest eigenvalue of this is λ_1 and it is also unique. According to Theorem 1, we take as eigenvectors for \mathbf{S} the vectors \mathbf{u}_1 and \mathbf{v}_1 . The second largest eigenvalue is λ_3 , but $\mathbf{u}_3 = \mathbf{u}_1$, so we do not consider this eigenvalue. Therefore, the next largest eigenvalue is λ_2 , and since it has a different eigenvector direction than λ_1 , we select it as the second eigenvalue for \mathbf{S} . In conclusion, we thus obtain for the LES:

$$\begin{aligned} \mathbf{S} &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{v}_1 \mathbf{v}_1^T \\ &= \frac{1}{5\sqrt{2}} \begin{pmatrix} 3 + \sqrt{3} & -1 \\ -1 & 3 - \sqrt{3} \end{pmatrix} \\ &\approx \begin{pmatrix} 0.6692 & -0.1414 \\ -0.1414 & 0.1793 \end{pmatrix}, \end{aligned}$$

which represents in the RGB space a medium light shade of blue–magenta colour: $(\frac{3}{5}, \frac{3}{5}, 1)$ (Fig. 4).

Fig. 4 Dilation of an 90×90 image based on the colours of example 1 with a centred 1×41 SE using the LES. **From left to right:** original image and dilated image



3.2 LES for Non-unique Largest Eigenvalues

Until now, we have assumed that the largest eigenvalue λ_1 is unique, but we now want to show what we get when this is not the case. To do this, we first show that the largest eigenvalue of S cannot be greater than λ_1 :

Lemma 5 *Let S be defined as in (11) and condition (15) be fulfilled according to (13). Furthermore, let X_1 has the largest not necessarily unique eigenvalue λ_1 of all matrices of \mathcal{X} . Then the largest eigenvalue of S cannot be greater than λ_1 .*

Proof We have already seen that the largest eigenvalue of S arises from the Rayleigh product $R(\varepsilon)$, $|\varepsilon| \ll 1$, which can be estimated as follows:

$$\begin{aligned} R(\varepsilon) &= \frac{1}{1+\varepsilon^2} \left(e^{m\lambda_1} + \varepsilon^2 e^{m\mu_1} \right. \\ &\quad \left. + \sum_{i=2}^n ((c_i + \varepsilon s_i)^2 e^{m\lambda_i} + (s_i - \varepsilon c_i)^2 e^{m\mu_i}) \right) \\ &\leq \frac{1}{1+\varepsilon^2} \left(e^{m\lambda_1} + \varepsilon^2 e^{m\lambda_1} \right. \\ &\quad \left. + \sum_{i=2}^n ((c_i + \varepsilon s_i)^2 e^{m\lambda_1} + (s_i - \varepsilon c_i)^2 e^{m\lambda_1}) \right) \\ &= \frac{e^{m\lambda_1}}{1+\varepsilon^2} \left(1 + \varepsilon^2 + \sum_{i=2}^n ((c_i + \varepsilon s_i)^2 \right. \\ &\quad \left. + (s_i - \varepsilon c_i)^2) \right) \\ &= \frac{e^{m\lambda_1}}{1+\varepsilon^2} \left(1 + \varepsilon^2 + \sum_{i=2}^n (c_i^2 + 2\varepsilon c_i s_i + \varepsilon^2 s_i^2 \right. \\ &\quad \left. + s_i^2 - 2\varepsilon c_i s_i + \varepsilon^2 c_i^2) \right) \end{aligned}$$

$$\begin{aligned} &= \frac{e^{m\lambda_1}}{1+\varepsilon^2} \left(1 + \varepsilon^2 + \sum_{i=2}^n (1 + \varepsilon^2) \right) \\ &= e^{m\lambda_1} \left(1 + \sum_{i=2}^n 1 \right) = n e^{m\lambda_1}. \end{aligned}$$

Thus, the largest eigenvalue of S can be approximated with

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{1}{m} \log R(\varepsilon) &\leq \lim_{m \rightarrow \infty} \frac{1}{m} \log (n e^{m\lambda_1}) \\ &= \lim_{m \rightarrow \infty} \frac{1}{m} (\log (e^{m\lambda_1}) + \log n) \\ &= \lim_{m \rightarrow \infty} \left(\lambda_1 + \frac{\log n}{m} \right) = \lambda_1. \end{aligned}$$

□

In a similar way, one can also show that S is indeed an upper bound for the X_i , $i = 1, \dots, n$. To do this, we prove the following lemma:

Lemma 6 *The LES S according to (11) is an upper bound in the Loewner sense for the given matrices \mathcal{X} .*

Proof Apparently one has

$$\sum_{i=1}^n \exp(mX_i) \geq_L \exp(mX_j) \quad \forall j \in \{1, \dots, n\},$$

which in combination with the fact that the logarithm is an operator-monotone function [22], i.e. $A \leq_L B \implies \log A \leq_L \log B$ for symmetric positive definite matrices A, B , results in $S \geq_L X_j$ for all $j \in \{1, \dots, n\}$. □

We now continue our considerations regarding the other largest eigenvalue. If there is a second eigenvalue which is equal to λ_1 , it can then be either μ_1 or one of the λ_j , since $\lambda_j \geq \mu_j$. If it were μ_1 , then X_1 would have the representation $\lambda_1 I$ and the only matrix with equally aligned eigenvectors

that would be greater than or equal to $\lambda_1 \mathbf{I}$ in the Loewner sense without having an even greater eigenvalue would be $\lambda_1 \mathbf{I}$ itself.

For the other case that $\lambda_j = \lambda_1$ for a fixed $j \in \{2, \dots, n\}$ will we assume that without loss of generality $j = 2$ holds. We summarise the argumentation necessary for this in the following theorem:

Theorem 2 *Let \mathbf{S} be the LES of the multi-set $\mathcal{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$, $n \in \mathbb{N}$, of symmetric real 2×2 matrices with the spectral decomposition (13) and $\lambda_1 = \lambda_2$ the largest eigenvalues of all these matrices. Then one has*

$$\mathbf{S} = \begin{cases} \lambda_1 \mathbf{I}, & \text{if } \mathbf{u}_1 \neq \pm \mathbf{u}_2, \\ \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \mu_* \mathbf{v}_1 \mathbf{v}_1^T, & \text{otherwise,} \end{cases}$$

where $\mu_* \leq \lambda_1$ is the next largest eigenvalue whose eigenvector \mathbf{v}_* is not aligned with \mathbf{u}_1 .

Proof 1.) To ensure assumption (15), we begin this proof with the same rotation as in the proof of Theorem 1. Then, we consider some of the properties of \mathbf{S} . The first property is that the largest eigenvalue of \mathbf{S} is λ_1 . This is a consequence of Lemma 5 and the fact that the Rayleigh product $R^*(\varepsilon)$ for this case is greater than or equal to the $R(\varepsilon)$ in Lemma 1. Consequently, Lemma 2 is also subject to this relation in accordance with

$$\lim_{m \rightarrow \infty} \frac{1}{m} \log R^*(\varepsilon) \geq \lim_{m \rightarrow \infty} \frac{1}{m} \log R(\varepsilon) = \lambda_1.$$

This leads to the second property, namely the spectral decomposition of the rotated LES $\mathbf{S}_{-\varphi_1}$:

$$\begin{aligned} \mathbf{S}_{-\varphi_1} &= \lambda_1 \bar{\mathbf{u}} \bar{\mathbf{u}}^T + \alpha \bar{\mathbf{v}} \bar{\mathbf{v}}^T, \quad \lambda_1 \geq \alpha \in \mathbb{R}, \\ \bar{\mathbf{u}} &= (c, s)^T, \quad \bar{\mathbf{v}} = (-s, c)^T, \quad c = \cos(\varphi), \\ s &= \sin(\varphi), \quad \varphi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. \end{aligned}$$

Lemma 6 gives us the upper bound property, in particular, this satisfies

$$\begin{aligned} \mathbf{0} &\leq_L \mathbf{S}_{-\varphi_1} - \mathbf{X}_1 \\ &= \lambda_1 (\bar{\mathbf{u}} \bar{\mathbf{u}}^T - \mathbf{u}_1 \mathbf{u}_1^T) + \alpha \bar{\mathbf{v}} \bar{\mathbf{v}}^T - \mu_1 \mathbf{v}_1 \mathbf{v}_1^T \\ &= \lambda_1 \begin{pmatrix} c^2 - 1 & cs \\ cs & s^2 \end{pmatrix} + \alpha \begin{pmatrix} s^2 & -cs \\ -cs & c^2 \end{pmatrix} \\ &\quad - \begin{pmatrix} 0 & 0 \\ 0 & \mu_1 \end{pmatrix} \\ &= \begin{pmatrix} \alpha s^2 - \lambda_1 s^2 & cs(\lambda_1 - \alpha) \\ cs(\lambda_1 - \alpha) & \lambda_1 s^2 + \alpha c^2 - \mu_1 \end{pmatrix} \\ &= \begin{pmatrix} s^2(\alpha - \lambda_1) & cs(\lambda_1 - \alpha) \\ cs(\lambda_1 - \alpha) & c^2(\alpha - \lambda_1) + \lambda_1 - \mu_1 \end{pmatrix}, \end{aligned}$$

but this is only fulfilled if

$$\begin{aligned} s^2(\alpha - \lambda_1) &\geq 0 \wedge c^2(\alpha - \lambda_1) + \lambda_1 - \mu_1 \geq 0 \\ \wedge \det(\mathbf{S}_{-\varphi_1} - \mathbf{X}_1) &\geq 0 \end{aligned} \quad (21)$$

apply.

2.) Let us assume that $s^2 \neq 0$, then it follows from the first relation of (21) that $\alpha \geq \lambda_1$. Because of Lemma 5, the eigenvalue α cannot be greater than λ_1 , which leads to $\alpha = \lambda_1$. This in turn leads to the spectral decomposition having the following form:

$$\begin{aligned} \mathbf{S}_{-\varphi_1} &= \lambda_1 (\bar{\mathbf{u}} \bar{\mathbf{u}}^T + \bar{\mathbf{v}} \bar{\mathbf{v}}^T) \\ &= \lambda_1 \begin{pmatrix} c^2 + s^2 & 0 \\ 0 & c^2 + s^2 \end{pmatrix} = \lambda_1 \mathbf{I}. \end{aligned}$$

By rotating all eigenvectors back by \mathbf{R}_{φ_1} , we achieve

$$\mathbf{S} = \mathbf{R}_{\varphi_1} \mathbf{S}_{-\varphi_1} \mathbf{R}_{\varphi_1}^T = \lambda_1 \mathbf{R}_{\varphi_1} \mathbf{I} \mathbf{R}_{\varphi_1}^T = \lambda_1 \mathbf{I}.$$

3.) Now we assume the other case $s^2 = 0$. Therefore, we have $s = 0$ and $c = \pm 1$. If we insert these values into the spectral decomposition, we obtain

$$\mathbf{S}_{-\varphi_1} = \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \alpha \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \alpha \end{pmatrix}.$$

By comparing with equation (21), we see that the only non-trivial condition remaining is $\alpha \geq \mu_1$.

If now $\mathbf{u}_1 = \pm \mathbf{u}_2$ would hold, the matrix $\mathbf{E}_m = (e_{mij})_{i,j=1,2}$ would have the entries

$$\begin{aligned} e_{m11} &= 2e^{m\lambda_1} + \sum_{i=3}^n (c_i^2 e^{m\lambda_i} + s_i^2 e^{m\mu_i}), \\ e_{m12} &= \sum_{i=3}^n c_i s_i (e^{m\lambda_i} - e^{m\mu_i}) = e_{m21} \quad \text{and} \\ e_{m22} &= e^{m\mu_1} + e^{m\mu_2} + \sum_{i=3}^n (s_i^2 e^{m\lambda_i} + c_i^2 e^{m\mu_i}). \end{aligned}$$

Thus, we can repeat all the steps up to Theorem 1 with minor adjustments and get the result from the theorem with $\alpha = \mu_*$, where μ_* is the next largest eigenvalue of \mathcal{X} whose eigenvector \mathbf{v}_* is not aligned with \mathbf{u}_1 . By rotating it back, we achieve as in the proof of Theorem 1:

$$\begin{aligned} \mathbf{S} &= \mathbf{R}_{\varphi_1} \mathbf{S}_{-\varphi_1} \mathbf{R}_{\varphi_1}^T \\ &= \mathbf{R}_{\varphi_1} (\lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \mu_* \mathbf{e}_2 \mathbf{e}_2^T) \mathbf{R}_{\varphi_1}^T \\ &= \lambda_1 \mathbf{R}_{\varphi_1} \mathbf{e}_1 (\mathbf{R}_{\varphi_1} \mathbf{e}_1)^T + \mu_* \mathbf{R}_{\varphi_1} \mathbf{e}_2 (\mathbf{R}_{\varphi_1} \mathbf{e}_2)^T \end{aligned}$$

$$= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \mu_* \mathbf{v}_1 \mathbf{v}_1^T.$$

Otherwise, we return to the argumentation of Lemma 4 with the difference that $\lambda_2 = \lambda_1$ holds. We then obtain for the Rayleigh product in the case $s_2^2 \neq 0$:

$$\begin{aligned} \bar{R}(\varepsilon) &= \frac{e^{m\lambda_1}}{1 + \varepsilon^2} \left((s_2 - \varepsilon c_2)^2 + \varepsilon^2 + e^{m(\mu_1 - \lambda_1)} \right. \\ &\quad + (c_2 + \varepsilon s_2)^2 e^{m(\mu_2 - \lambda_1)} \\ &\quad + \sum_{i=3}^n \left((s_i - \varepsilon c_i)^2 e^{m(\lambda_i - \lambda_1)} \right. \\ &\quad \left. \left. + (c_i + \varepsilon s_i)^2 e^{m(\mu_i - \lambda_1)} \right) \right) \\ &= \frac{e^{m\lambda_1}}{1 + \varepsilon^2} \left((s_2 - \varepsilon c_2)^2 + \varepsilon^2 \right. \\ &\quad \left. + \mathcal{O}(e^{m(\mu_* - \lambda_1)}) \right), \end{aligned}$$

where μ_* is again the next largest eigenvalue of \mathcal{X} whose eigenvector \mathbf{v}_* is not aligned with \mathbf{u}_1 . By determining $\frac{1}{m} \log \bar{R}(\varepsilon)$ again for this and then taking the limit for $m \rightarrow \infty$, the ε terms and the $\mathcal{O}(\cdot)$ term will disappear and only λ_1 remains. Thus, we have $\mathbf{S}_{-\varphi_1} = \lambda_1 \mathbf{I}$ again. \square

At this point, we would also like to give a small example in the case that the largest eigenvalue is not unique.

Example 2 Here we consider the simple example of a bipartite image consisting of the two RGB colours blue $\mathbf{C}_1 = (0, 0, 1)$ and green $\mathbf{C}_2 = (0, 1, 0)$. The corresponding symmetric matrices are

$$\begin{aligned} \mathbf{X}_1 &= \frac{1}{2\sqrt{2}} \begin{pmatrix} \sqrt{3} & -1 \\ -1 & -\sqrt{3} \end{pmatrix} \\ &\approx \begin{pmatrix} 0.6124 & -0.3535 \\ -0.3535 & -0.6124 \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} \mathbf{X}_2 &= \frac{1}{2\sqrt{2}} \begin{pmatrix} -\sqrt{3} & -1 \\ -1 & \sqrt{3} \end{pmatrix} \\ &\approx \begin{pmatrix} -0.6124 & -0.3535 \\ -0.3535 & 0.6124 \end{pmatrix}. \end{aligned}$$

This results in the following eigenvalues and eigenvectors:

$$\begin{aligned} \lambda_1 &= \frac{1}{\sqrt{2}} \approx 0.7071, \quad \mu_1 = -\lambda_1, \\ \mathbf{u}_1 &= \frac{1}{2\sqrt{2 + \sqrt{3}}} \begin{pmatrix} -\sqrt{3} - 2 \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.9659 \\ 0.2588 \end{pmatrix}, \end{aligned}$$

$$\mathbf{v}_1 = \frac{1}{2\sqrt{2 + \sqrt{3}}} \begin{pmatrix} -1 \\ -\sqrt{3} - 2 \end{pmatrix} \approx \begin{pmatrix} -0.2588 \\ -0.9659 \end{pmatrix}$$

and

$$\lambda_2 = \lambda_1, \quad \mu_2 = \mu_1,$$

$$\mathbf{u}_2 = \frac{1}{2\sqrt{2 + \sqrt{3}}} \begin{pmatrix} -1 \\ \sqrt{3} + 2 \end{pmatrix} \approx \begin{pmatrix} -0.2588 \\ 0.9659 \end{pmatrix},$$

$$\mathbf{v}_2 = \frac{1}{2\sqrt{2 + \sqrt{3}}} \begin{pmatrix} -\sqrt{3} - 2 \\ -1 \end{pmatrix} \approx \begin{pmatrix} -0.9659 \\ -0.2588 \end{pmatrix}.$$

Since the largest eigenvalue is not unique and the corresponding eigenvectors $\mathbf{u}_1, \mathbf{u}_2$ are not equal, we obtain by Theorem 2 the LES

$$\mathbf{S} = \lambda_1 \mathbf{I} = \frac{1}{\sqrt{2}} \mathbf{I} \approx \begin{pmatrix} 0.7071 & 0 \\ 0 & 0.7071 \end{pmatrix},$$

which represents the RGB colour white $(1, 1, 1)$.

To clarify the phenomenon that the supremum of blue and green in Fig. 5 of Example 2 is white, we must first bear in mind that both colours lie on the edge of the unit circle of the bi-cone, directly opposite each other in the y-direction, see Figure 6.

This means that in order to maintain the Loewner order, you would have to choose the smallest circle that contains both points. However, this smallest circle depends on the choice of the supremum. If, for example, we choose the trace supremum of Burgeth et al. [23], see Fig. 6, this circle protrudes from the unit circle and the resulting cone from the bi-cone, i.e. it would be an impermissible colour. This is avoided by the warp factor introduced by Burgeth and Kleefeld [18], see Fig. 6, but at the cost that the circle no longer includes blue and green, i.e. is no longer larger according to the Loewner order. The LES basically sees it in such a way that the smallest and only circle within the bi-cone, which still includes blue and green, is the unit circle itself, i.e. white.

Remark 2 The discolouration that occurs in the LES is therefore a fundamental consequence of the Loewner order, which states that the closer the colours are to the edge and the further apart they are, the higher the colour cone of the supremum must be.

3.3 General Characterisation and Properties of the LES

In this subsection, we will give a general characterisation of the LES for any combination of eigenvalues. Based on this, we will show two interesting properties in the form of transitivity and associativity with respect to dilation. The latter

Fig. 5 Dilation of a bivariate image consisting of blue and green from example 2 with a centred quadratic SE using the LES. **From left to right:** original image and dilated image

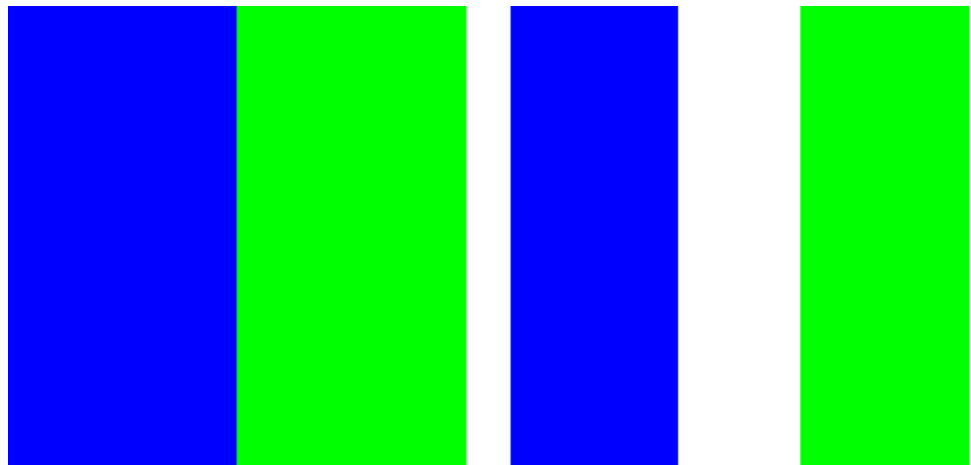
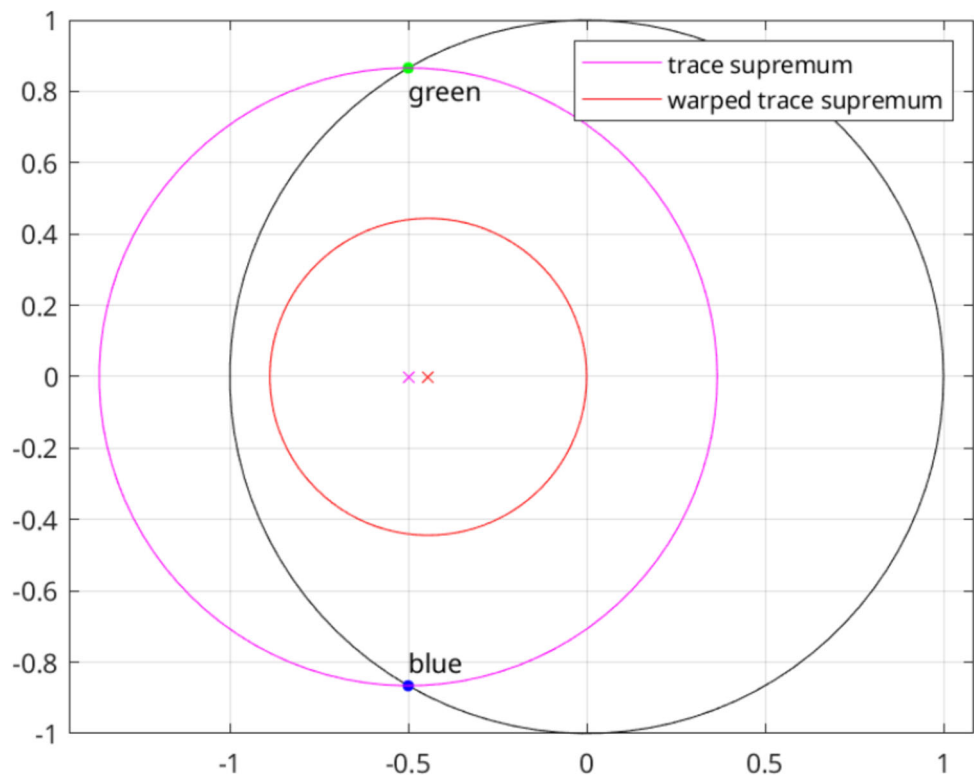


Fig. 6 Comparison of dilation of the colours green and blue with different suprema regarding the Loewner order inside the bi-cone



property in particular sets our approach apart from other multidimensional approaches, since, to our knowledge, it does not exist in colour morphology as opposed to greyscale morphology.

In order to provide greater clarity, we shall now present a summary of the two preceding theorems in the form of a corollary that gives us a general characterisation of the LES.

Corollary 1 Let S be the LES (11) of the multi-set $\mathcal{X} = \{X_1, \dots, X_n\}$, $n \in \mathbb{N}$, of symmetric real 2×2 matrices with the spectral decompositions (13) and λ_1 (one of) the largest eigenvalues of all these matrices. Furthermore, let $\mathcal{V}(\mathcal{X})$ be the set of all eigenvectors of the matrices of \mathcal{X} and $\mathcal{V}_{\sup}^{\lambda_1}(\mathcal{X})$ be the set of corresponding eigenvectors to the largest eigen-

values equal to λ_1 , i.e.

$$\mathcal{V}_{\sup}^{\lambda_1}(\mathcal{X}) := \{v \in \mathcal{V}(\mathcal{X}) : \exists X \in \mathcal{X} : Xv = \lambda v \\ \wedge \lambda = \lambda_1\}.$$

Then, the LES can be characterised as follows:

$$S = \begin{cases} \lambda_1 I, & \text{if } \lambda_1 \text{ is not unique} \\ & \text{and } \exists v \in \mathcal{V}_{\sup}^{\lambda_1}(\mathcal{X}) : \\ & v \neq \pm u_1, \\ \lambda_1 u_1 u_1^T + \mu_* v_1 v_1^T, & \text{otherwise,} \end{cases} \quad (22)$$

where $\mu_* \leq \lambda_1$ is the next largest eigenvalue of \mathcal{X} whose corresponding eigenvector \mathbf{v}_* is not aligned with \mathbf{u}_1 .

We will show in the following proposition that the LES S , as previously characterised by (22), exhibits a transitive property in general.

Proposition 1 *The LES (11) is transitive, i.e. for multi-sets \mathcal{X} and \mathcal{Y} of symmetric 2×2 matrices one has*

$$\begin{aligned} \text{Sup}_{\text{LE}}(\mathcal{X} \cup \mathcal{Y}) \\ = \text{Sup}_{\text{LE}}(\{\text{Sup}_{\text{LE}}(\mathcal{X}), \text{Sup}_{\text{LE}}(\mathcal{Y})\}). \end{aligned} \quad (23)$$

Proof 1.) For the multi-sets, we will use the notation

$$\begin{aligned} \mathcal{X} &= \{\mathbf{X}_1, \dots, \mathbf{X}_n\}, \quad \mathcal{Y} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_m\}, \\ \mathcal{Z} &= \{\mathbf{Z}_1, \dots, \mathbf{Z}_{n+m}\} := \mathcal{X} \cup \mathcal{Y}, \quad m, n \in \mathbb{N}. \end{aligned}$$

The right-hand side of (23) can be calculated as

$$\begin{aligned} S_r &:= \text{Sup}_{\text{LE}}(\{\mathbf{S}_1, \mathbf{S}_2\}) \\ &:= \text{Sup}_{\text{LE}}(\{\text{Sup}_{\text{LE}}(\mathcal{X}), \text{Sup}_{\text{LE}}(\mathcal{Y})\}). \end{aligned}$$

The LES \mathbf{S}_1 and \mathbf{S}_2 are characterised according to (22) as

$$\mathbf{S}_1 = \begin{cases} \lambda_1 \mathbf{I}, & \text{if } \lambda_1 \text{ is not unique} \\ & \text{and } \exists \mathbf{v} \in \mathcal{V}_{\text{sup}}^{\lambda_1}(\mathcal{X}) : \\ & \mathbf{v} \neq \pm \mathbf{u}_1, \\ \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_* \mathbf{v}_1 \mathbf{v}_1^T, & \text{otherwise,} \end{cases}$$

and

$$\mathbf{S}_2 = \begin{cases} \mu_1 \mathbf{I}, & \text{if } \mu_1 \text{ is not unique} \\ & \text{and } \exists \mathbf{v} \in \mathcal{V}_{\text{sup}}^{\mu_1}(\mathcal{Y}) : \\ & \mathbf{v} \neq \pm \tilde{\mathbf{u}}_1, \\ \mu_1 \tilde{\mathbf{u}}_1 \tilde{\mathbf{u}}_1^T + \mu_* \tilde{\mathbf{v}}_1 \tilde{\mathbf{v}}_1^T, & \text{otherwise.} \end{cases}$$

For the characterisation of \mathbf{S}_1 , we assumed that λ_1 is the largest eigenvalue of \mathcal{X} with the corresponding eigenvector \mathbf{u}_1 , λ_* is the next largest eigenvalue of \mathcal{X} whose eigenvector is not aligned with \mathbf{u}_1 and \mathbf{v}_1 is the eigenvector perpendicular to \mathbf{u}_1 . The same can be said of the characterisation of \mathbf{S}_2 , insofar as we use \mathcal{Y} , μ_1 and μ_* instead of λ_1 and λ_* and the eigenvectors $\tilde{\mathbf{u}}_1$ and $\tilde{\mathbf{v}}_1$ as replacement for \mathbf{u}_1 and \mathbf{v}_1 .

Since S_r is again an LES approximation for two symmetric 2×2 matrices, we can characterise it as follows

$$\mathbf{S}_r = \begin{cases} \eta_1 \mathbf{I}, & \text{if } \eta_1 \text{ is not unique} \\ & \text{and } \exists \mathbf{v} \in \mathcal{V}_{\text{sup}}^{\eta_1}(\mathcal{S}) : \\ & \mathbf{v} \neq \pm \bar{\mathbf{u}}_1, \\ \eta_1 \bar{\mathbf{u}}_1 \bar{\mathbf{u}}_1^T + \eta_* \bar{\mathbf{v}}_1 \bar{\mathbf{v}}_1^T, & \text{otherwise,} \end{cases}$$

where η_1 represents the maximum of λ_1 and μ_1 and it is associated with the corresponding normalised eigenvector $\bar{\mathbf{u}}_1$. The term η_* , on the other hand, denotes the next largest eigenvalue within the set $\{\lambda_1, \lambda_*, \mu_1, \mu_*\}$ of possible eigenvalues and, in addition, corresponds to an eigenvector that is not aligned with $\bar{\mathbf{u}}_1$. The eigenvector $\bar{\mathbf{v}}_1$ denotes the eigenvector perpendicular to $\bar{\mathbf{u}}_1$ and \mathcal{S} is defined as $\mathcal{S} := \{\mathbf{S}_1, \mathbf{S}_2\}$.

2.) For the left-hand side of (23), we obtain with the characterisation (22):

$$\begin{aligned} \mathbf{S}_l &:= \text{Sup}_{\text{LE}}(\mathcal{Z}) \\ &= \begin{cases} v_1 \mathbf{I}, & \text{if } v_1 \text{ is not unique} \\ & \text{and } \exists \mathbf{v} \in \mathcal{V}_{\text{sup}}^{v_1}(\mathcal{Z}) : \\ & \mathbf{v} \neq \pm \hat{\mathbf{u}}_1, \\ v_1 \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + v_* \hat{\mathbf{v}}_1 \hat{\mathbf{v}}_1^T, & \text{otherwise,} \end{cases} \end{aligned}$$

where v_1 is (one of) the largest eigenvalues of \mathcal{Z} with the corresponding normalised eigenvector $\hat{\mathbf{u}}_1$, v_* is the next largest eigenvalue of \mathcal{Z} with an eigenvector that is not aligned with v_1 and $\hat{\mathbf{v}}_1$ is the normalised eigenvector perpendicular to $\hat{\mathbf{u}}_1$. Since v_1 is the largest eigenvalue of $\mathcal{Z} = \mathcal{X} \cup \mathcal{Y}$, it is also the largest eigenvalue of \mathcal{X} and \mathcal{Y} and as such it fulfils $v_1 = \eta_1$.

3.) The first case of \mathbf{S}_l will only happen if there are at least two largest eigenvalues in \mathcal{Z} whose eigenvectors are not aligned. As they would be the largest eigenvalues of \mathcal{Z} , they would also be the largest eigenvalues of \mathcal{X} or \mathcal{Y} . In the event that both values correspond to matrices from the multi-set \mathcal{X} , then \mathbf{S}_1 would be a diagonal matrix with these eigenvalues. The same would apply for \mathbf{S}_2 if they would both correspond to matrices from \mathcal{Y} . This implies that in these cases, one of \mathbf{S}_1 and \mathbf{S}_2 would be a diagonal matrix with these eigenvalues. As they are part of the largest eigenvalues, the corresponding matrix would be selected for S_r .

For the case that these two eigenvalues are distributed between \mathcal{X} and \mathcal{Y} , they would nevertheless be selected as either λ_1 , λ_* , μ_1 or μ_* . This is because they remain the largest eigenvalues, and any other largest eigenvalues would have the same eigenvectors as these two, which would then be deemed equivalent to the corresponding one of the two largest eigenvalues with non-aligned eigenvectors. Consequently, the corresponding eigenvectors would be in $\mathcal{V}_{\text{sup}}^{\eta_1}(\mathcal{S})$, and S_r would take the form of a diagonal matrix with these two eigenvalues. In conclusion, The first case of \mathbf{S}_l occurs if and only if it occurs at \mathbf{S}_r and they coincide.

4.) For the second case, we use that we have already identified η_1 with v_1 . This implies that the corresponding eigenvectors are also the same, that is, we have $\hat{\mathbf{u}}_1 = \bar{\mathbf{u}}_1$ and $\hat{\mathbf{v}}_1 = \bar{\mathbf{v}}_1$. Ultimately, it follows from the construction of the corresponding characterisations shown above and a similar reasoning as in

the first case that $v_2 = \eta_2$ must hold. Furthermore, it can be shown that S_l and S_r are identical in this case. \square

To conclude this section, we want to use this property to prove the associativity of LES dilation. For this purpose, we first define what we understand by LES dilation and LES erosion.

Definition 5 Let $\Omega \subseteq \mathbb{Z}^2$ be the two-dimensional, discrete image domain, $f : \Omega \rightarrow \mathbb{R}^3$ a colour image and $b : \mathbb{Z}^2 \rightarrow \mathbb{R}^3 \cup \{(-\infty, -\infty, -\infty)^T\}$ the structuring function with components according to (1) of the set $B_0 \subset \mathbb{Z}^2$. Furthermore, let $\tau : \mathbb{R}^3 \rightarrow \text{Sym}(2)$ be the transformation from the vector of the corresponding colour space to the symmetric 2×2 matrix according to Subsection 2.1. Then we define the **LES dilation** as

$$\begin{aligned} (f \oplus_{\text{LES}} b)(x) \\ := \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\{ \tau(f(x-u)) + \tau(b(u)) : \right. \right. \\ \left. \left. u \in \mathbb{Z}^2 \} \right) \right), \quad x \in \Omega. \end{aligned} \quad (24)$$

In accordance with the duality (5) between dilation and erosion for the grey value case, we can also define the **LES erosion** here by means of LES dilation as follows:

$$(f \ominus_{\text{LES}} b)(x) := (f^c \oplus_{\text{LES}} b)^c(x), \quad x \in \Omega, \quad (25)$$

where the complement of a colour vector is to be understood as an component-wise complementation according to (4):

$$\begin{aligned} f^c(x) &= ((R, G, B)^T)^c := (R^c, G^c, B^c)^T \\ &= \begin{pmatrix} R_{\max} - R + R_{\min} \\ G_{\max} - G + G_{\min} \\ B_{\max} - B + B_{\min} \end{pmatrix}, \quad x \in \Omega. \end{aligned}$$

Here, R_{\max} represents the largest red value of the image f and R_{\min} the smallest; the same applies to the corresponding G and B terms.

We now turn to the mentioned associativity of LES dilation.

Theorem 3 The LES dilation (24) is associative, i.e. for a colour image $f : \mathbb{Z}^2 \rightarrow [0, 1]^3$ and for the structuring elements given by the structuring functions $b_1, b_2 : \mathbb{Z}^2 \rightarrow \mathbb{R}^3 \cup \{(-\infty, -\infty, -\infty)^T\}$ of two sets $B_1, B_2 \subset \mathbb{Z}^2$ applies

$$(f \oplus_{\text{LES}} b_1) \oplus_{\text{LES}} b_2 = f \oplus_{\text{LES}} (b_1 \oplus_{\text{LES}} b_2). \quad (26)$$

Proof To proof this claim, we will compare the considered matrices for both sides of the equation. We use the expression $X(x) := \tau(f(x))$ to notate the matrices of the image f . Similarly, we do this for the structuring functions of B_1 and

B_2 using the matrices $W_1(x) := \tau(b_1(x))$ and $W_2(x) := \tau(b_2(x))$.

For the left-hand side of equation (26), we calculate for the first dilation

$$\begin{aligned} (f \oplus_{\text{LES}} b_1)(x) \\ = \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\{ X(x-y) + W_1(y) : y \in B_1 \} \right) \right) \end{aligned}$$

and by the second dilation

$$\begin{aligned} ((f \oplus_{\text{LES}} b_1) \oplus_{\text{LES}} b_2)(x) \\ = \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\{ \tau((f \oplus_{\text{LES}} b_1)(x-z)) \right. \right. \\ \left. \left. + W_2(z) : z \in B_2 \} \right) \right) \\ = \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\{ \text{Sup}_{\text{LE}} \left(\{ X(x-y-z) \right. \right. \right. \\ \left. \left. + W_1(y) : y \in B_1 \} \right) + W_2(z) : z \in B_2 \} \right) \right) \end{aligned}$$

Since B_2 is a discrete set, we can numerate the elements of B_2 with $z_1, \dots, z_m, m \in \mathbb{N}$, and rewrite the equation as

$$\begin{aligned} ((f \oplus_{\text{LES}} b_1) \oplus_{\text{LES}} b_2)(x) \\ = \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\text{Sup}_{\text{LE}} \left(\{ X(x-y-z_1) \right. \right. \right. \\ \left. \left. + W_1(y) : y \in B_1 \} \right) + W_2(z_1), \dots, \right. \\ \left. \text{Sup}_{\text{LE}} \left(\{ X(x-y-z_m) + W_1(y) : y \in B_1 \} \right) \right. \\ \left. + W_2(z_m) \right) \right) \\ = \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\text{Sup}_{\text{LE}} \left(\{ X(x-y-z_1) \right. \right. \right. \\ \left. \left. + W_1(y) + W_2(z_1) : y \in B_1 \} \right), \dots, \right. \\ \left. \text{Sup}_{\text{LE}} \left(\{ X(x-y-z_m) + W_1(y) \right. \right. \\ \left. \left. + W_2(z_m) : y \in B_1 \} \right) \right) \right), \quad x \in \Omega, \end{aligned}$$

since the summation with a constant matrix $W_2(z_i)$ does not influence the decision of the LES except of a global (for each individual LES) rotation of the eigenvectors and change of the eigenvalues. Then by setting

$$\mathcal{Y}_i(x) := \{ X(x-y-z_i) + W_1(y) + W_2(z_i) : y \in B_1 \}, \quad x \in \Omega,$$

and using the transitivity (23), we obtain for all $x \in \Omega$:

$$((f \oplus_{\text{LES}} b_1) \oplus_{\text{LES}} b_2)(x)$$

$$\begin{aligned}
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\text{Sup}_{\text{LE}} (\mathcal{Y}_1(x)), \dots, \right. \right. \\
&\quad \left. \left. \text{Sup}_{\text{LE}} (\mathcal{Y}_m(x)) \right) \right) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\bigcup_{i=1}^m \mathcal{Y}_i(x) \right) \right) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\{X(x - y - z) + W_1(y) \right. \right. \\
&\quad \left. \left. + W_2(z) : y \in B_1 \wedge z \in B_2\} \right) \right). \quad (27)
\end{aligned}$$

The right-hand side of (26) fulfils for the first dilation with the structuring functions the equality

$$\begin{aligned}
&(b_1 \oplus_{\text{LES}} b_2)(y) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} (\{W_1(y - z) + W_2(z) : \right. \\
&\quad \left. z \in B_2\}) \right), \quad y \in \mathbb{Z}^2.
\end{aligned}$$

However, since b_1 and as such W_1 will “vanish” with $-\infty$ in the sense of the dilation for $y - z \notin B_1$, we can introduce the set

$$B_1 \oplus B_2 := \{x_1 + x_2 : x_1 \in B_1 \wedge x_2 \in B_2\}$$

and replace $y \in \mathbb{Z}^2$ with $y \in B_1 \oplus B_2$. We calculate the second dilation as follows

$$\begin{aligned}
&(f \oplus_{\text{LES}} (b_1 \oplus_{\text{LES}} b_2))(x) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\{X(x - y) \right. \right. \\
&\quad \left. \left. + \text{Sup}_{\text{LE}} (\{W_1(y - z) + W_2(z) : z \in B_2\}) : \right. \right. \\
&\quad \left. \left. y \in B_1 \oplus B_2 \right) \right), \quad x \in \Omega.
\end{aligned}$$

Given that $B_1 \oplus B_2$ is also a discrete set, the same countability trick that was employed previously can be applied, resulting with y_1, \dots, y_k in the following:

$$\begin{aligned}
&(f \oplus_{\text{LES}} (b_1 \oplus_{\text{LES}} b_2))(x) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(X(x - y_1) \right. \right. \\
&\quad \left. \left. + \text{Sup}_{\text{LE}} (\{W_1(y_1 - z) + W_2(z) : z \in B_2\}), \right. \right. \\
&\quad \left. \left. \dots, X(x - y_k) + \text{Sup}_{\text{LE}} (\{W_1(y_k - z) \right. \right. \\
&\quad \left. \left. + W_2(z) : z \in B_2\}) \right) \right) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\text{Sup}_{\text{LE}} (\{X(x - y_1) \right. \right. \\
&\quad \left. \left. + W_1(y_1 - z) + W_2(z) : z \in B_2\}), \dots, \right. \right.
\end{aligned}$$

$$\begin{aligned}
&\text{Sup}_{\text{LE}} (\{X(x - y_k) + W_1(y_k - z) + W_2(z) : \\
&\quad z \in B_2\}) \Big), \quad x \in \Omega.
\end{aligned}$$

By defining

$$\begin{aligned}
Z_i(x) &:= \{X(x - y_i) + W_1(y_i - z) + W_2(z) : \\
&\quad z \in B_2\}, \quad x \in \Omega,
\end{aligned}$$

and using the transitivity (23), we obtain for all $x \in \Omega$:

$$\begin{aligned}
&(f \oplus_{\text{LES}} (b_1 \oplus_{\text{LES}} b_2))(x) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\text{Sup}_{\text{LE}} (Z_1(x)), \dots, \right. \right. \\
&\quad \left. \left. \text{Sup}_{\text{LE}} (Z_k(x)) \right) \right) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\bigcup_{i=1}^k Z_i(x) \right) \right) \\
&= \tau^{-1} \left(\text{Sup}_{\text{LE}} \left(\bigcup_{y \in B_1 \oplus B_2} \{X(x - y) \right. \right. \\
&\quad \left. \left. + W_1(y - z) + W_2(z) : z \in B_2\} \right) \right).
\end{aligned}$$

By substituting $y = \hat{y} + z$ with $\hat{y} \in B_1$ and $z \in B_2$ in the last equation, we see that it equals (27), which concludes the proof. \square

We demonstrate the correctness of this statement again using the colours from example 1, see Fig. 7. To do this, we also use a different SE for b_2 in the form of a centred quadratic 11×11 mask. We can see from this that the two resulting images match visually according to (26). An examination of the numerical colour values also shows an identity between these images.

Remark 3 Analogously, it can be shown for LES erosion (25) that

$$(f \ominus_{\text{LES}} b_1) \ominus_{\text{LES}} b_2 = f \ominus_{\text{LES}} (b_1 \oplus_{\text{LES}} b_2)$$

applies.

4 Minimality of the LES

So far, we have only shown that the LES (11) is an upper bound in the sense of the Loewner order. In this section, we address the question of whether it is also a smallest upper bound or under which conditions it is. To do this, we start from the framework [8] of Welk, Kleefeld and Breuß and add another function to our semi-order to select a smallest upper

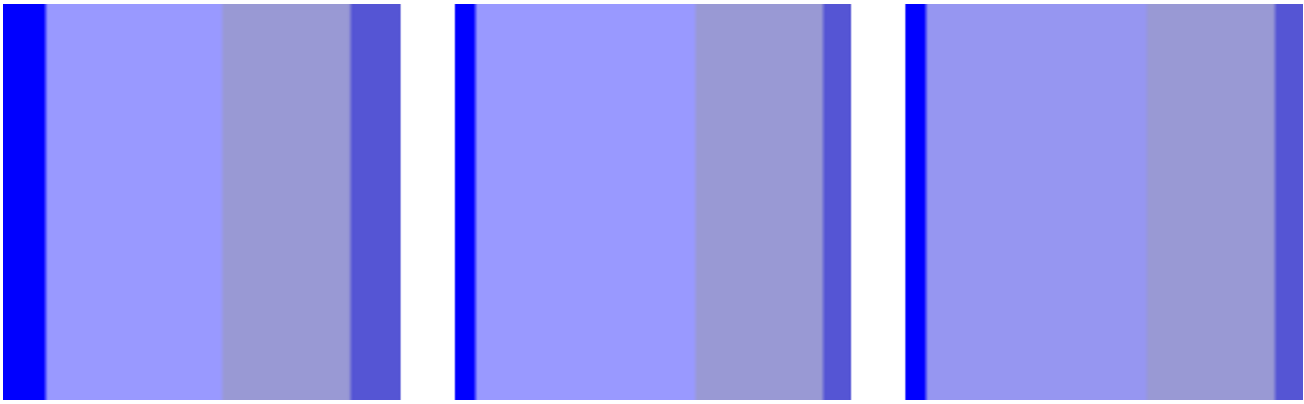


Fig. 7 Comparison of the associativity of dilation using the LES based on example 1. **From left to right:** LES dilation δ_1 of the 90×90 image from example 1 with the same 1×41 SE, LES dilation δ_2 of δ_1 with a

centred 11×11 SE and LES dilation of the image from example 1 with the LES-dilated SE of the 1×41 SE and the 11×11 SE

bound from the set of upper bounds introduced by the semi-order, as already mentioned in Section 2.2. This auxiliary function serves to find a unique minimum within the given set, so that we may have no total order but an auxiliary order. This is the reason why we refer to it as the auxiliary ordering function.

The main problem here is the existence of a total ordering function φ on the set $\mathcal{U}(\mathcal{X})$ of upper bounds in the Loewner sense of the symmetric matrices $\mathcal{X} = \{X_1, \dots, X_n\}$, $n \in \mathbb{N}$, for which the LES S is the unique minimiser. To see this, we first prove the following lemma.

Lemma 7 *Let S be the LES (11) for a given multi-set \mathcal{X} of symmetric 2×2 matrices and let λ_1 be the unique largest eigenvalue of \mathcal{X} . Whenever the second largest eigenvalue λ_2 of \mathcal{X} is unique, and its original eigenvector \mathbf{u}_2 is not perpendicular to the eigenvector \mathbf{u}_1 of λ_1 , there exists some matrix*

$$S' := \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + (\lambda_2 - \varepsilon) \mathbf{v}_1 \mathbf{v}_1^T \in \mathcal{U}(\mathcal{X}), \quad \varepsilon > 0. \quad (28)$$

Proof Due to Theorem 1, $S = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{v}_1 \mathbf{v}_1^T$ holds, where \mathbf{v}_1 is the eigenvector perpendicular to \mathbf{u}_1 . Since we can add a constant value to all eigenvalues and subtract it again later when we have formed the LES, we assume without loss of generality that $\lambda_1 = 1$, and its associated eigenvector is $\mathbf{u}_1 = (1, 0)^T$. This makes $\mathbf{v}_1 = (0, 1)^T$ and we get the representation

$$S = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad \lambda_2 < 1.$$

Let

$$S' = \begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix}, \quad q < \lambda_2.$$

Then we examine for which q the matrix S' remains in $\mathcal{U}(\mathcal{X})$.

For this, we choose an arbitrary $X \in \mathcal{X}$ with a major eigenvalue λ satisfying $\lambda \leq \lambda_2$ and the associated eigenvector $\mathbf{u} = (c, s)^T$ with $c = \cos(\varphi)$ and $s = \sin(\varphi)$. Further, let μ be the minor eigenvalue of X and have the associated eigenvector $\mathbf{v} = (-s, c)^T$. Then X has the following representation

$$X = \begin{pmatrix} \lambda c^2 + \mu s^2 & cs(\lambda - \mu) \\ cs(\lambda - \mu) & \lambda s^2 + \mu c^2 \end{pmatrix}$$

because of its spectral decomposition (13). Note that $Y := S' - X \geq_L \mathbf{0}$ must necessarily be satisfied for $S' \in \mathcal{U}(\mathcal{X})$ to hold. To verify $Y \geq_L \mathbf{0}$, the following relations must apply:

$$R_Y(\mathbf{u}_1) \geq 0 \wedge R_Y(\mathbf{u}_2) \geq 0 \wedge \det(Y) \geq 0.$$

We calculate for the first term

$$\begin{aligned} R_Y(\mathbf{u}_1) &= 1 - \lambda c^2 - \mu s^2 \\ &\geq 1 - \lambda c^2 - \lambda s^2 = 1 - \lambda \\ &\geq 1 - \lambda_2 > 0. \end{aligned} \quad (29)$$

In order for the second term, $R_Y(\mathbf{u}_2) = q - \lambda s^2 - \mu c^2$, to be non-negative, it is always possible to find a $q \in [\lambda s^2 + \mu c^2, \lambda_2]$ that will fulfil this requirement, provided $X \neq X_2$ with $\mathbf{u}_2 = (0, 1)^T \perp \mathbf{u}_1$. Then, the last term can be calculated as

$$\begin{aligned} \det(Y) &= \begin{vmatrix} 1 - \lambda c^2 - \mu s^2 & cs(\lambda - \mu) \\ cs(\lambda - \mu) & q - \lambda s^2 - \mu c^2 \end{vmatrix} \\ &= (1 - \lambda c^2 - \mu s^2)(q - \lambda s^2 - \mu c^2) \\ &\quad - c^2 s^2 (\lambda - \mu)^2 \\ &= q(1 - \lambda c^2 - \mu s^2) - \lambda s^2 + c^2 s^2 (\lambda^2 + \mu^2) \end{aligned}$$

$$\begin{aligned}
& + (c^4 + s^4)\lambda\mu - \mu c^2 - c^2 s^2 (\lambda - \mu)^2 \\
& = q(1 - \lambda c^2 - \mu s^2) - \lambda s^2 - \mu c^2 \\
& + (c^4 + 2c^2 s^2 + s^4)\lambda\mu \\
& = q(1 - \lambda c^2 - \mu s^2) - \lambda s^2 - \mu c^2 \\
& + (c^2 + s^2)^2 \lambda\mu \\
& = q(1 - \lambda c^2 - \mu s^2) - \lambda s^2 - \mu c^2 + \lambda\mu
\end{aligned}$$

which is non-negative if and only if

$$\begin{aligned}
q & \geq \frac{\lambda s^2 + \mu c^2 - \lambda\mu}{1 - \lambda c^2 - \mu s^2} \\
& = \lambda \frac{1 - \lambda c^2 - \mu s^2}{1 - \lambda c^2 - \mu s^2} \\
& + \frac{\lambda(s^2 - 1) + \mu c^2 - \lambda\mu(1 - s^2) + \lambda^2 c^2}{1 - \lambda c^2 - \mu s^2} \\
& = \lambda + \frac{c^2(\mu - \lambda\mu + \lambda^2 - \lambda)}{1 - \lambda c^2 - \mu s^2} \\
& = \lambda + \frac{c^2(1 - \lambda)(\mu - \lambda)}{1 - \lambda c^2 - \mu s^2} \\
& = \lambda - \frac{c^2(1 - \lambda)(\lambda - \mu)}{1 - \lambda c^2 - \mu s^2}. \tag{30}
\end{aligned}$$

Since $c^2 \geq 0$, $\mu \leq \lambda < 1$ and (29), we see that the fraction on the right-hand side is non-negative. This means that the largest term that can occur on the right-hand side would be λ and since λ_2 is our second largest eigenvalue and unique, $\lambda < \lambda_2$ except in the case where we consider the matrix X belonging to λ_2 . However, in this case, because of uniqueness, $\lambda_2 > \mu_2$ and $c^2 \neq 0$, otherwise u_2 would point in the same direction as v_1 , contradicting the condition $u_2 \not\propto u_1$. This means that the fraction on the right-hand side would be positive, and thus, the term on the right-hand side would be strictly less than λ_2 .

The case for X with $\lambda = \lambda_1$ provides the matrix

$$Y = \begin{pmatrix} 0 & 0 \\ 0 & q - \mu_1 \end{pmatrix}$$

with $\mu_1 < \lambda_2$ because λ_2 is the second largest eigenvalue and unique. The relation $Y \geq_L \mathbf{0}$ holds if and only if $q \geq \mu_1$. Since $\mu_1 < \lambda_2$ and $q < \lambda_2$ (otherwise $S' = S$), we can find a q with $\mu_1 \leq q < \lambda_2$ to fulfil the requirement.

In summary, this means that in every case the right-hand side of (30) is strictly less than λ_2 . Thus, if we compute for all $X \in \mathcal{X}$ the bounds of q , which are all strictly less than λ_2 , and then take the largest of them, we obtain the claim (28) of the lemma. \square

Corollary 2 Under the conditions of Lemma 7, S cannot be the unique minimiser of a function φ according to (10).

Proof Suppose S were a minimiser of a function φ according to (10). Then $\varphi(S) \geq \varphi(S')$ because $S \geq_L S'$ and the Loewner monotonicity of φ . Since S is minimal with respect to φ , only $\varphi(S) = \varphi(S')$ can be true. However, this is a contradiction to the uniqueness of the minimiser. \square

This may seem like a big disadvantage at first. However, with the following slight modification of our set $\mathcal{U}(\mathcal{X})$ of possible upper bounds of \mathcal{X} , we can ensure that the LES actually acts as a minimiser there, as we will demonstrate in this section.

Definition 6 For a multi-set \mathcal{X} of symmetric 2×2 matrices with non-negative eigenvalues, let $\mathcal{X}^p := \{X^p : X \in \mathcal{X}\}$ be the element-wise application of the p th power to the multi-set \mathcal{X} . We define the p -power upper bound cone as

$$\begin{aligned}
\mathcal{U}_p(\mathcal{X}) & := (\mathcal{U}(\mathcal{X}^p))^{\frac{1}{p}} \\
& = \{Y \in \text{Sym}(2) : Y^p \in \mathcal{U}(\mathcal{X}^p)\} \tag{31}
\end{aligned}$$

and denote the intersection of all p -power upper bound cones as the **super-upper bound cone**

$$\mathcal{U}_*(\mathcal{X}) := \bigcap_{p>0} \mathcal{U}_p(\mathcal{X}). \tag{32}$$

For a multi-set \mathcal{X} of symmetric 2×2 matrices for which $-c < 0$ is the smallest among all eigenvalues of the matrices in \mathcal{X} , define

$$\mathcal{U}_*(\mathcal{X}) := \mathcal{U}_*(\mathcal{X} + c) - c, \tag{33}$$

where

$$\mathcal{X} + c := \{X + cI : X \in \mathcal{X}\}$$

and I is the identity matrix.

Remark 4 Equation (33) is particularly applicable to the set $\mathcal{U}(\mathcal{X})$ of upper bounds of a multi-set \mathcal{X} , as $\text{Sup}_{\text{LE}}(\mathcal{X} + c) - cI = \text{Sup}_{\text{LE}}(\mathcal{X})$, and thus, $\mathcal{U}(\mathcal{X} + c) - c = \mathcal{U}(\mathcal{X})$ always hold.

To show that S is the unique minimiser of φ in $\mathcal{U}_*(\mathcal{X})$, we first prove that it is contained in it.

Lemma 8 Let \mathcal{X} be a multi-set of symmetric 2×2 matrices, and let S be as in Lemma 7. Then one has $S \in \mathcal{U}_p(\mathcal{X})$ for any $p > 0$.

Proof We use for S the same form as in Lemma 7 and assume again without loss of generality that $\lambda_1 = 1$, and its associated eigenvector is $u_1 = (1, 0)^T$. This implies $v_1 = (0, 1)^T$.

Since we only consider real symmetric and thus diagonalisable matrices, we can simply express the power of such a matrix according to [21] as follows:

$$X^p = \lambda^p \mathbf{u}\mathbf{u}^T + \mu^p \mathbf{v}\mathbf{v}^T.$$

The representation of S in terms of the p th power will then take the following form:

$$S^p = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_2^p \end{pmatrix}.$$

On the basis of $\mathcal{U}(\mathcal{X} + c) - c = \mathcal{U}(\mathcal{X})$, see Remark 4, we can assume without restriction that $\lambda_2 \geq 0$ applies. So we see that the only difference to our previous calculation of the LES is that all eigenvalues are now raised to the power of p . This means that our LES looks the same as before with the difference that we exponentiate its entries, i.e. the two largest permissible eigenvalues, by p and thus obtain S^p . Thus, we have $S^p \in \mathcal{U}(\mathcal{X}^p)$, from which the claim follows. \square

Next, we will define the as yet unspecified function φ and show that it fulfils all the necessary conditions according to Section 2.2. In particular, we will see that for this function S is the unique minimiser in $\mathcal{U}_*(\mathcal{X})$.

Theorem 4 *Let the conditions of Lemma 8 be satisfied and let the function $\varphi : \mathcal{U}(\mathcal{X}) \rightarrow \mathbb{R}^2$ be explained as $\varphi(Y) = (\lambda, \mu)$, where $\lambda > \mu$ are the eigenvalues of $Y \in \mathcal{U}(\mathcal{X})$. Further, let \mathbb{R}^2 be endowed with the lexicographic order*

$$(a, b) < (a', b') : \iff (a < a' \vee (a = a' \wedge b \leq b')), \quad (34)$$

$$a, b, a', b' \in \mathbb{R}.$$

Then it follows that

- (i) the function φ is Loewner-monotone,
- (ii) the function φ is convex on $\mathcal{U}(\mathcal{X})$, and
- (iii) S is the unique minimiser of φ in $\mathcal{U}_*(\mathcal{X})$.

Proof (i) Let $A, B \in \mathcal{U}(\mathcal{X})$ with $B \leq_L A$. The inequality implies that for any vector $\mathbf{w} \in \mathbb{R}^2$ the Rayleigh products must satisfy $R_B(\mathbf{w}) \leq R_A(\mathbf{w})$.

By setting $\mathbf{w} = \mathbf{u}_B$ to the unit eigenvector corresponding to the largest eigenvalue λ_B of B , the Rayleigh product of B will equal the largest eigenvalue according to $R_B(\mathbf{u}_B) = \lambda_B$. The Rayleigh product of A with \mathbf{u}_B will take a value $\alpha \in [\mu_A, \lambda_A]$ according to the min-max theorem (cf. [24], Theorem 4.2.2), where μ_A is the smaller and λ_A the larger eigenvalue of A . Upon inserting this into the inequality for the Rayleigh products, we obtain

$$\lambda_B = R_B(\mathbf{u}_B) \leq R_A(\mathbf{u}_B) = \alpha \leq \lambda_A,$$

from which it follows that the largest eigenvalue of A is greater or equal to the largest eigenvalue of B . If $\lambda_A > \lambda_B$, $\varphi(B) \preceq \varphi(A)$ is already established.

If this is not the case, we need to consider the minor eigenvalues of A and B . By choosing $\mathbf{w} = \mathbf{v}_A$ as the unit eigenvector corresponding to the smallest eigenvalue μ_A of A , the Rayleigh product of A will equal the smallest eigenvalue according to $R_A(\mathbf{v}_A) = \mu_A$. If we also consider the Rayleigh product with B , we again obtain with the min-max theorem a value $\beta \in [\mu_B, \lambda_B]$, where μ_B is the smallest eigenvalue of B , for this and by inserting it into the inequality mentioned at the beginning, we obtain

$$\mu_B \leq \beta = R_B(\mathbf{v}_A) \leq R_A(\mathbf{v}_A) = \mu_A.$$

Consequently, we can guarantee that the smallest eigenvalue of B is less than or equal to the smallest eigenvalue of A , and thus, $\varphi(B) \preceq \varphi(A)$ holds.

(ii) We want to prove here that φ is convex on $\mathcal{U}(\mathcal{X})$ according to Definition 3. For this, we consider the two vector spaces $\text{Sym}(2)$ and \mathbb{R}^2 and the convex set $\mathcal{U}(\mathcal{X}) \subset \text{Sym}(2)$. We have already seen the convexity of the last set in Section 2.2. Further, we need an order cone on the \mathbb{R}^2 equipped with the lexicographic order, which we denote simply as

$$K := \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \succcurlyeq \mathbf{0}\}.$$

That is, the only thing we still need to check is the validity of (9). For this, we consider for all $Y_1, Y_2 \in \mathcal{U}(\mathcal{X})$ and $\alpha \in [0, 1]$:

$$\begin{aligned} \alpha\varphi(Y_1) + (1 - \alpha)\varphi(Y_2) - \varphi(\alpha Y_1 + (1 - \alpha)Y_2) &\in K \\ \iff \mathbf{0} \preceq \alpha\varphi(Y_1) + (1 - \alpha)\varphi(Y_2) \\ &\quad - \varphi(\alpha Y_1 + (1 - \alpha)Y_2) \\ \iff \alpha\varphi(Y_1) + (1 - \alpha)\varphi(Y_2) \\ &\quad \succcurlyeq \varphi(\alpha Y_1 + (1 - \alpha)Y_2) =: \varphi(Z). \end{aligned} \quad (35)$$

Let us assume without loss of generality that Z is a diagonal matrix. Further, let $\lambda_i \geq \mu_i$ be the eigenvalues of Y_i with the corresponding eigenvectors $(c_i, s_i)^T, (-s_i, c_i)^T$ for $i = 1, 2$ according to (13). Then, we have

$$Z = \begin{pmatrix} \lambda_* & 0 \\ 0 & \mu_* \end{pmatrix}$$

with

$$\lambda_* := \alpha(\lambda_1 c_1^2 + \mu_1 s_1^2) + (1 - \alpha)(\lambda_2 c_2^2 + \mu_2 s_2^2)$$

and

$$\mu_* := \alpha(\lambda_1 s_1^2 + \mu_1 c_1^2) + (1 - \alpha)(\lambda_2 s_2^2 + \mu_2 c_2^2).$$

Note that the two eigenvalues λ_* , μ_* do not necessarily have to be ordered in this way. If we substitute this into our inequality (35), we obtain

$$\alpha \begin{pmatrix} \lambda_1 \\ \mu_1 \end{pmatrix} + (1 - \alpha) \begin{pmatrix} \lambda_2 \\ \mu_2 \end{pmatrix} \succcurlyeq \begin{pmatrix} \lambda_* \\ \mu_* \end{pmatrix}. \quad (36)$$

To prove this inequality, we first show that both eigenvalues λ_* , μ_* are less than or equal to $\alpha\lambda_1 + (1 - \alpha)\lambda_2$:

$$\begin{aligned} \lambda_* &\leq \alpha(\lambda_1 c_1^2 + \lambda_1 s_1^2) + (1 - \alpha)(\lambda_2 c_2^2 + \lambda_2 s_2^2) \\ &= \alpha\lambda_1 + (1 - \alpha)\lambda_2, \\ \mu_* &\leq \alpha(\lambda_1 s_1^2 + \lambda_1 c_1^2) + (1 - \alpha)(\lambda_2 s_2^2 + \lambda_2 c_2^2) \\ &= \alpha\lambda_1 + (1 - \alpha)\lambda_2. \end{aligned}$$

If strict inequality applies to both inequalities, we would have proven the statement. Otherwise, one of the two diagonal entries is equal to $\alpha\lambda_1 + (1 - \alpha)\lambda_2$, without loss of generality, let this be λ_* . Then it follows

$$\begin{aligned} \mu_* &= \text{tr}(\mathbf{Z}) - \lambda_* = \text{tr}(\mathbf{Z}) - \alpha\lambda_1 - (1 - \alpha)\lambda_2 \\ &= \alpha\text{tr}(\mathbf{Y}_1) + (1 - \alpha)\text{tr}(\mathbf{Y}_2) - \alpha\lambda_1 \\ &\quad - (1 - \alpha)\lambda_2 \\ &= \alpha(\text{tr}(\mathbf{Y}_1) - \lambda_1) + (1 - \alpha)(\text{tr}(\mathbf{Y}_2) - \lambda_2) \\ &= \alpha(\lambda_1 + \mu_1 - \lambda_1) + (1 - \alpha)(\lambda_2 + \mu_2 - \lambda_2) \\ &= \alpha\mu_1 + (1 - \alpha)\mu_2 \end{aligned}$$

whereby we have exploited in the last line that the trace of a matrix can be understood as the sum of its eigenvalues. This means that in this case the smaller eigenvalue of \mathbf{Z} is just the convex combination of the smaller eigenvalues of \mathbf{Y}_1 and \mathbf{Y}_2 or, in other words, satisfies the second line of inequality (36) with equality.

(iii) For this, we use the argumentation from the proof of Lemma 7 for $(\mathbf{S}')^p$ and \mathbf{X}^p instead of \mathbf{S}' and \mathbf{X} to obtain a necessary condition for $(\mathbf{S}')^p \geq_L \mathbf{X}^p$, using the representation from the proof of Lemma 8 for \mathbf{X}^p . Since in both cases the exponentiation by p does not change the eigenvectors but only the eigenvalues, we can take our conclusion (30) from the proof of Lemma 7 by exponentiating the eigenvalues all by p :

$$q^p \geq \lambda^p - \frac{c^2(1 - \lambda^p)(\lambda^p - \mu^p)}{1 - \lambda^p c^2 - \mu^p s^2}. \quad (37)$$

As in the proof of Lemma 7, we have again that the fraction on the right-hand side of (37) is non-negative for all $\mu \leq \lambda < 1$ and $p > 0$.

In the proof of Lemma 7 we had shown that the right-hand side of the inequality (30) analogous to (37) was always smaller than λ_2 . Thus, we could guarantee that an

$\varepsilon > 0$ exists, so that a similar matrix to the LES \mathbf{S} exists whose second eigenvalue according to $\lambda_2 - \varepsilon < \lambda_2$ is strictly smaller than that of the LES \mathbf{S} , which showed that the LES was not the unique minimiser. So all we have to prove here is that no such ε exists for given λ , μ , c , s and $p \rightarrow \infty$.

Let us assume that $\varepsilon > 0$ exists so that (37) is satisfied for $q = \lambda - \varepsilon$ and arbitrarily large $p > 0$. In particular, we can also assume without restriction of generality that our eigenvalues can be mapped bijectively to the interval $[0, 1]$ or more precisely to

$\left[\frac{\frac{1}{\sqrt{2}} - \lambda_1^{\text{old}}}{\sqrt{2}}, 1 \right]$, where λ_1^{old} is the original largest eigenvalue of \mathcal{X} , so that $\lambda_1 = 1$ and $\mu_1, \lambda_i, \mu_i \in [0, 1)$ for $i = 2, \dots, n$, see the following remark. This would mean that

$$\begin{aligned} 0 &\leq (\lambda - \varepsilon)^p - \lambda^p + \frac{c^2(1 - \lambda^p)(\lambda^p - \mu^p)}{1 - \lambda^p c^2 - \mu^p s^2} \\ &\iff \lambda^p - (\lambda - \varepsilon)^p \leq \frac{c^2(1 - \lambda^p)(\lambda^p - \mu^p)}{1 - \lambda^p c^2 - \mu^p s^2} \\ &\leq \frac{(1 - \lambda^p)(\lambda^p - \mu^p)}{1 - \lambda^p c^2 - \mu^p s^2} \\ &\iff (\lambda - \varepsilon)^p \geq \lambda^p - \frac{(1 - \lambda^p)(\lambda^p - \mu^p)}{1 - \lambda^p c^2 - \mu^p s^2} \\ &\iff \varepsilon \leq \lambda - \left[\lambda^p - \frac{(1 - \lambda^p)(\lambda^p - \mu^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}}, \\ 0 &\leq \mu \leq \lambda < 1, \quad p > 0, \end{aligned} \quad (38)$$

holds. We rewrite the right-hand side of the inequality (38) into

$$\begin{aligned} &\lambda - \left[\lambda^p - \frac{(1 - \lambda^p)(\lambda^p - \mu^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \\ &= \lambda - \left[\lambda^p - \lambda^p \frac{\left(1 - \frac{\mu^p}{\lambda^p}\right)(1 - \lambda^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \\ &= \lambda - \left[\lambda^p \left(1 - \left(1 - \frac{\mu^p}{\lambda^p}\right) \frac{(1 - \lambda^p)}{1 - \lambda^p c^2 - \mu^p s^2}\right) \right]^{\frac{1}{p}} \\ &= \lambda \left\{ 1 - \left[1 - \left(1 - \frac{\mu^p}{\lambda^p}\right) \frac{(1 - \lambda^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \right\}. \end{aligned} \quad (39)$$

If we consider the last $[\cdot]^{\frac{1}{p}}$ term in the equation (39), we can estimate as follows

$$\begin{aligned} & \left[1 - \frac{1}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \\ & \leq \left[1 - \left(1 - \frac{\mu^p}{\lambda^p} \right) \frac{(1 - \lambda^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \leq 1^{\frac{1}{p}} \end{aligned} \quad (40)$$

and by using the limes for $p \rightarrow \infty$ we obtain for the upper and lower limit

$$\lim_{p \rightarrow \infty} 1^{\frac{1}{p}} = 1, \quad (41)$$

$$\begin{aligned} & \lim_{p \rightarrow \infty} \left[1 - \frac{1}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \\ & = \lim_{p \rightarrow \infty} \exp \left(\frac{1}{p} \log \left(1 - \frac{1}{1 - \lambda^p c^2 - \mu^p s^2} \right) \right) \\ & = \exp \left(\lim_{p \rightarrow \infty} \frac{1}{p} \log \left(1 - \frac{1}{1 - \lambda^p c^2 - \mu^p s^2} \right) \right) \\ & = e^0 = 1. \end{aligned} \quad (42)$$

By combining (40) with (41) and (42), we achieve

$$1 \leq \lim_{p \rightarrow \infty} \left[1 - \left(1 - \frac{\mu^p}{\lambda^p} \right) \frac{(1 - \lambda^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \leq 1$$

or

$$\lim_{p \rightarrow \infty} \left[1 - \left(1 - \frac{\mu^p}{\lambda^p} \right) \frac{(1 - \lambda^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} = 1. \quad (43)$$

For the last step, we substitute (39) in (38), let p approach infinity and use the equation (43):

$$\begin{aligned} \varepsilon & \leq \lim_{p \rightarrow \infty} \lambda \left\{ 1 - \left[1 - \left(1 - \frac{\mu^p}{\lambda^p} \right) \frac{(1 - \lambda^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \right\} \\ & = \lambda \left\{ 1 - \lim_{p \rightarrow \infty} \left[1 - \left(1 - \frac{\mu^p}{\lambda^p} \right) \frac{(1 - \lambda^p)}{1 - \lambda^p c^2 - \mu^p s^2} \right]^{\frac{1}{p}} \right\} \\ & = 0, \end{aligned}$$

which is a contradiction to $\varepsilon > 0$. This means that $S' \notin \mathcal{U}_*(\mathcal{X})$. \square

Remark 5 Let $X \in \mathcal{X}$ be arbitrary and $X = \lambda \mathbf{u} \mathbf{u}^T + \mu \mathbf{v} \mathbf{v}^T$ as in (13), where λ, μ are the eigenvalues of X and \mathbf{u}, \mathbf{v} the associated normalised eigenvectors. Since the greatest element in the HCL bi-cone is the colour white with the eigenvalues $\frac{1}{\sqrt{2}}$ and the smallest element is the colour black with the eigenvalues $-\frac{1}{\sqrt{2}}$, see [8], we have $\lambda, \mu \in \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]$. This means

that the transformation

$$\begin{aligned} \bar{X} & := \frac{1}{\sqrt{2}} X + \left(1 - \frac{\lambda_1}{\sqrt{2}} \right) \mathbf{I} \\ & = \frac{1}{\sqrt{2}} \lambda \mathbf{u} \mathbf{u}^T + \frac{1}{\sqrt{2}} \mu \mathbf{v} \mathbf{v}^T + \left(1 - \frac{\lambda_1}{\sqrt{2}} \right) \mathbf{u} \mathbf{u}^T \\ & \quad + \left(1 - \frac{\lambda_1}{\sqrt{2}} \right) \mathbf{v} \mathbf{v}^T \\ & = \left(\frac{\lambda}{\sqrt{2}} + 1 - \frac{\lambda_1}{\sqrt{2}} \right) \mathbf{u} \mathbf{u}^T \\ & \quad + \left(\frac{\mu}{\sqrt{2}} + 1 - \frac{\lambda_1}{\sqrt{2}} \right) \mathbf{v} \mathbf{v}^T \\ & =: \bar{\lambda} \mathbf{u} \mathbf{u}^T + \bar{\mu} \mathbf{v} \mathbf{v}^T \end{aligned}$$

has the eigenvalues $\bar{\lambda}, \bar{\mu} \in [0, 1]$. It follows from the fact that

$$1 + \frac{\lambda - \lambda_1}{\sqrt{2}} \in \left[\frac{\frac{1}{\sqrt{2}} - \lambda_1}{\sqrt{2}}, 1 \right] \subseteq [0, 1],$$

where the lower and upper bound are derived from $\lambda \in \left[-\frac{1}{\sqrt{2}}, \lambda_1 \right]$. Consequently, the addition of the scaled unit matrix affects only the eigenvalues but not the eigenvectors. As a result, this can also be applied directly to the LES:

$$\begin{aligned} \bar{S} & := \text{Sup}_{\text{LE}} (\bar{X}_1, \dots, \bar{X}_n) \\ & = \frac{1}{\sqrt{2}} S + \left(1 - \frac{\lambda_1}{\sqrt{2}} \right) \mathbf{I}. \end{aligned}$$

5 Relaxation of the LES

Now that we have seen several favourable properties of the LES, we want to address a disadvantage of our approach that we have not considered further so far. It is one of the reasons why this approach was not pursued further in favour of the trace supremum Sup_{tr} in the work [12]. This is due to the fact that the LES does not depend continuously on the input data. In this section, we will first look at the cases where this problem occurs and then see that it can be solved using a straightforward relaxation.

To visualise this problem, let us consider the following scenario: We have a multi-set $\mathcal{X} = \{X_1, \dots, X_n\}$, $n \geq 3$, of symmetric 2×2 matrices in which the three largest eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$ with the associated eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ belong to three different matrices $X_1, X_2, X_3 \in \mathcal{X}$. In addition, the eigenvectors have the relationship $\mathbf{u}_1 = \mathbf{u}_2$ and $\mathbf{u}_3 \neq \pm \mathbf{u}_1$ to each other. According to Theorem 1, we obtain for the LES

$$S = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_3 \mathbf{v}_1 \mathbf{v}_1^T, \quad \mathbf{v}_1 \perp \mathbf{u}_1.$$

We will now consider a rotated multi-set $\mathcal{X}^{(\delta)} = \{X_1^{(\delta)}, \dots, X_n^{(\delta)}\}$ by rotating the matrices X_i around the angles $\delta\alpha_i$ for $i \geq 2$ with the factor $\delta > 0$. The angles α_i can be different but $\sin(\delta\alpha_2) \neq 0$, such that the new eigenvector

$$\begin{aligned} \mathbf{u}_2^{(\delta)} &= R_{\delta\alpha_2} \mathbf{u}_2 = \begin{pmatrix} \cos(\delta\alpha_2) & -\sin(\delta\alpha_2) \\ \sin(\delta\alpha_2) & \cos(\delta\alpha_2) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\delta\alpha_2) \\ \sin(\delta\alpha_2) \end{pmatrix} \neq \pm \mathbf{u}_1. \end{aligned}$$

As these are only rotations, the eigenvalues $\lambda_i^{(\delta)}, \mu_i^{(\delta)}$ of $X_i^{(\delta)}$ according $\lambda_i^{(\delta)} = \lambda_i$ and $\mu_i^{(\delta)} = \mu_i$ are retained for all $i \in \{1, \dots, n\}$ and $\delta > 0$. By also applying Theorem 1 to this, we obtain

$$S^{(\delta)} := \text{Sup}_{\text{LE}}(\mathcal{X}^{(\delta)}) = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{v}_1 \mathbf{v}_1^T =: S',$$

which is not equal to S . If we now want to determine the limit value of this for $\delta \rightarrow 0$, we go to step 2 of the proof of Lemma 4 and obtain $\bar{R}(\varepsilon, \delta)$ according to Equation (19) with $(s_2, c_2)^T = \mathbf{u}_2^{(\delta)}$. Then, we calculate

$$\begin{aligned} &\lim_{\delta \rightarrow 0} \lim_{m \rightarrow \infty} \left(\frac{1}{m} \log(\bar{R}(\varepsilon, \delta)) \right) \\ &= \lim_{\delta \rightarrow 0} \lim_{m \rightarrow \infty} \left(\lambda_2 - \frac{\log(1 + \varepsilon^2)}{m} \right. \\ &\quad \left. + \frac{\log((- \sin(\delta\alpha_2) - \varepsilon \cos(\delta\alpha_2))^2)}{m} \right. \\ &\quad \left. + \frac{\log(1 + \mathcal{O}(e^{m(\lambda_2 - \lambda_1)} + e^{m(\mu - \lambda_2)}))}{m} \right) \\ &= \lim_{\delta \rightarrow 0} \lambda_2 = \lambda_2, \end{aligned}$$

because $\varepsilon \rightarrow 0$ for $m \rightarrow \infty$, see proof of Lemma 4, and $\sin(\delta\varphi) \neq 0$. This leads to

$$\lim_{\delta \rightarrow 0} S^{(\delta)} = S' \neq S$$

although $\mathcal{X}^{(\delta)} \rightarrow \mathcal{X}$ for $\delta \rightarrow 0$.

These types of discontinuity are removable, since they only occur in non-generic configurations of \mathcal{X} . To address these kind of issues, we introduce the following definition.

Definition 7 Let $\mathcal{X} = \{X_1, \dots, X_n\}$, $n \in \mathbb{N}$, be a multi-set of symmetric real 2×2 matrices with the spectral decomposition (13). We call the multi-set

$$\begin{aligned} \mathcal{B}(\mathcal{X}) &:= \{\mathcal{B}(X_1), \dots, \mathcal{B}(X_n)\} \\ &:= \{\{\mathbf{u}_1, \mathbf{v}_1\}, \dots, \{\mathbf{u}_n, \mathbf{v}_n\}\} \end{aligned}$$

of orthonormal bases of \mathcal{X} **generic** if and only if no two orthonormal bases $\mathcal{B}(X_i), \mathcal{B}(X_j)$, $i \neq j$, share the same orientation, i.e. \mathbf{u}_i is not aligned with \mathbf{u}_j . Furthermore, we define that the multi-sets $\mathcal{X}^{(\delta)} = \{X_1^{(\delta)}, \dots, X_n^{(\delta)}\}$ **converge planar towards the multi-set** \mathcal{X} , symbolically $\mathcal{X}^{(\delta)} \xrightarrow{\text{pl}} \mathcal{X}$, if $\mathcal{X}^{(\delta)} \rightarrow \mathcal{X}$ and the eigenvalues of $X_i^{(\delta)}$ and X_i remain the same for all $i \in \{1, \dots, n\}$ and δ . Then, we declare the **relaxed log-exp-supremum (RLES)** as

$$\text{Sup}_{\text{RLE}}(\mathcal{X}) := \begin{cases} \text{Sup}_{\text{LE}}(\mathcal{X}), & \text{if } \mathcal{B}(\mathcal{X}) \text{ is generic,} \\ S_{\text{NG}}, & \text{otherwise,} \end{cases} \quad (44)$$

with

$$S_{\text{NG}} := \lim_{\substack{\mathcal{X}^{(\delta)} \xrightarrow{\text{pl}} \mathcal{X} \\ \mathcal{B}(\mathcal{X}^{(\delta)}) \text{ generic}}} \text{Sup}_{\text{LE}}(\mathcal{X}^{(\delta)}).$$

Remark 6 The idea of planar convergence stems from the fact that for

$$\begin{aligned} X &= \lambda \mathbf{u} \mathbf{u}^T + \mu \mathbf{v} \mathbf{v}^T \\ &= \begin{pmatrix} \lambda u_1^2 + \mu v_1^2 & \lambda u_1 u_2 + \mu v_1 v_2 \\ \lambda u_1 u_2 + \mu v_1 v_2 & \lambda u_2^2 + \mu v_2^2 \end{pmatrix}, \\ |\mathbf{u}| = 1 = |\mathbf{v}|, \quad \mathbf{u} &= \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2, \end{aligned}$$

$$\lambda, \mu \in \mathbb{R},$$

and

$$\begin{aligned} Y &= \lambda s s^T + \mu t t^T \\ &= \begin{pmatrix} \lambda s_1^2 + \mu t_1^2 & \lambda s_1 s_2 + \mu t_1 t_2 \\ \lambda s_1 s_2 + \mu t_1 t_2 & \lambda s_2^2 + \mu t_2^2 \end{pmatrix}, \\ |s| = 1 = |t|, \quad s &= \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}, t = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \in \mathbb{R}^2, \end{aligned}$$

their z -component in the bi-cone coordinates can be calculated as follows according to Burgeth and Kleefeld, see [18]:

$$\begin{aligned} z(X) &= \frac{\text{tr}(X)}{\sqrt{2}} = \frac{\lambda u_1^2 + \mu v_1^2 + \lambda u_2^2 + \mu v_2^2}{\sqrt{2}} \\ &= \frac{\lambda |\mathbf{u}|^2 + \mu |\mathbf{v}|^2}{\sqrt{2}} = \frac{\lambda + \mu}{\sqrt{2}}, \\ z(Y) &= \frac{\text{tr}(Y)}{\sqrt{2}} = \frac{\lambda s_1^2 + \mu t_1^2 + \lambda s_2^2 + \mu t_2^2}{\sqrt{2}} \\ &= \frac{\lambda |s|^2 + \mu |t|^2}{\sqrt{2}} = \frac{\lambda + \mu}{\sqrt{2}}. \end{aligned}$$

This means that the z -component remains the same for all colour matrices as long as they have the same eigenvalues, and that all the convergence happens on a particular z -plane.

We can immediately deduce from this definition by means of the following corollary that the RLES does not leave the bi-cone.

Corollary 3 Let $\mathcal{X} = (X_1, \dots, X_n)$, $n \in \mathbb{N}$, be a multi-set of symmetric real 2×2 matrices, which satisfy $-\frac{1}{\sqrt{2}}\mathbf{I} \leq_L X_i \leq_L \frac{1}{\sqrt{2}}\mathbf{I}$ for all $i \in \{1, \dots, n\}$. Then we have

$$-\frac{1}{\sqrt{2}}\mathbf{I} \leq_L \text{Sup}_{\text{RLE}}(\mathcal{X}) \leq_L \frac{1}{\sqrt{2}}\mathbf{I}.$$

Proof The assertion follows from the fact that the LES lies in the bi-cone and the RLES has the same eigenvalues for selection as the LES due to Definition 7. As a result, the largest eigenvalue cannot be greater than $\frac{1}{\sqrt{2}}$, which gives rise to the upper bound. The lower bound results from the transitivity of the Loewner order via $-\frac{1}{\sqrt{2}}\mathbf{I} \leq_L X_i \leq_L \text{Sup}_{\text{RLE}}(\mathcal{X})$ for all $i \in \{1, \dots, n\}$. \square

Following our first two previous theorems, we therefore also give a characterisation for the RLES in this form.

Theorem 5 Let \mathcal{X} be a multi-set of symmetric real 2×2 matrices. Further, let λ_1 be (one of) the largest eigenvalue(s) of \mathcal{X} , and \mathbf{u}_1 the eigenvector associated with the eigenvalue λ_1 and $\mathbf{v}_1 \perp \mathbf{u}_1$ the minor eigenvector of the same matrix from \mathcal{X} , and let $\lambda_2 \leq \lambda_1$ be the next largest eigenvalue of \mathcal{X} . Then the RLES of \mathcal{X} is characterised by

$$\begin{aligned} \text{Sup}_{\text{RLE}}(\mathcal{X}) &= \begin{cases} \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{v}_1 \mathbf{v}_1^T, & \text{if } \lambda_1 \text{ is unique,} \\ \lambda_1 \mathbf{I}, & \text{otherwise,} \end{cases} \end{aligned} \quad (45)$$

where \mathbf{I} is the identity matrix.

Proof The generic case of (44) is a straightforward consequence of Theorem 1 and Theorem 2. So we only need to consider the non-generic case.

Let λ_1 be unique. In this case, we just need to follow the argumentation of the calculation that we used to introduce this section. By the rotation of eigenvectors that were shown there and the fact that the $\mathcal{X}^{(\delta)}$ of Definition 7 fulfils all necessary properties of the mentioned rotated multi-set, we can conclude the first case of Equation (45).

Now we want λ_1 to be not unique. This means that we have

$$\lambda_1 = \lambda_2 = \lambda_1^{(\delta)} = \lambda_2^{(\delta)}, \quad (46)$$

where $\lambda_1^{(\delta)}, \lambda_2^{(\delta)}$ are the largest eigenvalues of $\mathcal{X}^{(\delta)}$ from Definition 7, since the $\mathcal{X}^{(\delta)}$ converge planar towards \mathcal{X} . Furthermore, let the corresponding eigenvectors of λ_1, λ_2 be given by $\mathbf{u}_1 = \mathbf{u}_2$. We can find multi-sets $\mathcal{X}^{(\delta)}$ for $\delta > 0$

which have the same eigenvalues as \mathcal{X} for all δ , generic $\mathcal{B}(\mathcal{X}^{(\delta)})$ and converge to \mathcal{X} for $\delta \rightarrow 0$, e.g. by rotating the eigenvectors of \mathcal{X} as at the beginning of this section. By Theorem 2 and Equation (46), we get for the LES of $\mathcal{X}^{(\delta)}$:

$$\mathbf{S}^{(\delta)} := \text{Sup}_{\text{LE}}(\mathcal{X}^{(\delta)}) = \lambda_1^{(\delta)} \mathbf{I} = \lambda_1 \mathbf{I} \quad \forall \delta > 0.$$

This means $\mathbf{S}^{(\delta_1)} = \mathbf{S}^{(\delta_2)}$ for all $\delta_1, \delta_2 > 0$. For $\delta \rightarrow 0$, it follows

$$\lim_{\delta \rightarrow 0} \mathbf{S}^{(\delta)} = \lim_{\delta \rightarrow 0} \lambda_1 \mathbf{I} = \lambda_1 \mathbf{I},$$

which concludes the proof. \square

However, the question remains as to whether there is another type of problem that can cause the aforementioned discontinuity. We consider the following lemma to partially answer this question.

Lemma 9 Let the conditions of Theorem 5 be fulfilled. Then, the RLES, given by Definition 7, depends continuously on the input data.

Proof Since the LES depends only on the spectral decomposition of the input data, the discontinuities can only occur through transformations of the input data, namely the eigenvalues and eigenvectors. Moreover, these transformations must remain in the bi-cone, so that only a scaling of the eigenvalues in the interval $\left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$ and a rotation of the eigenvectors are possible. We have already shown that the RLES depends continuously on the rotated input data. Therefore, it is sufficient to consider only the case of the scaled eigenvalues.

Since we only want to investigate the case of scaled eigenvalues, we can assume that the eigenvectors will not change in the modified multi-set $\mathcal{Y}^{(\delta)} = \{\mathbf{Y}_1^{(\delta)}, \dots, \mathbf{Y}_n^{(\delta)}\} \rightarrow \mathcal{X}$ with $\mathbf{Y}_i^{(\delta)} = \lambda_i^{(\delta)} \mathbf{u}_i^{(\delta)} (\mathbf{u}_i^{(\delta)})^T + \mu_i^{(\delta)} \mathbf{v}_i^{(\delta)} (\mathbf{v}_i^{(\delta)})^T$ according to $\mathbf{u}_i^{(\delta)} = \mathbf{u}_i$ and $\mathbf{v}_i^{(\delta)} = \mathbf{v}_i$ for all $i \in \{1, \dots, n\}$. As a result, as long as the order of the eigenvalues sorted by size does not change, no discontinuities can occur with regard to the input data, since $\lambda_i^{(\delta)} \rightarrow \lambda_i$ and $\mu_i^{(\delta)} \rightarrow \mu_i$ for all $i \in \{1, \dots, n\}$.

For a multiplicative scaling in the form of $\lambda_i^{(\delta)} := \delta \lambda_i$ for $\delta \rightarrow 1$, there will not occur any discontinuity, since the order of eigenvalues would stay the same. If we assume an additive scaling in form of $\lambda_i^{(\delta)} := \lambda_i + \delta \alpha_i$, $\alpha_i \in \mathbb{R}$, for $\delta \rightarrow 0$, we could change the order of the eigenvalues, e.g. according to $\lambda_1 > \lambda_2 > \lambda_3$ but $\lambda_3^{(\delta)} > \lambda_1^{(\delta)} > \lambda_2^{(\delta)}$ for $\delta \in [a, b] \subset [\varepsilon_1, \infty)$ with a sufficiently small $\varepsilon_1 > 0$. Since the RLES operates only with generic bases, $\delta \rightarrow 0$ and $\lambda_1^{(\delta)} > \lambda_2^{(\delta)} > \lambda_3^{(\delta)}$ for $\delta \in (0, \varepsilon_2)$ with a sufficiently small $\varepsilon_2 < \varepsilon_1$, we would obtain as RLES

$$\lim_{\delta \rightarrow 0} \text{Sup}_{\text{RLE}}(\mathcal{Y}^{(\delta)})$$

$$\begin{aligned}
&= \lim_{\substack{\delta \rightarrow 0 \\ \delta \in (0, \varepsilon_2)}} \lambda_1^{(\delta)} \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2^{(\delta)} \mathbf{v}_1 \mathbf{v}_1^T \\
&= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{v}_1 \mathbf{v}_1^T = \text{Sup}_{\text{RLE}}(\mathcal{X}).
\end{aligned}$$

An analogous reasoning is applicable in the case of non-unique eigenvalues. \square

We would now like to summarise the most important features of the RLES in the following corollary.

Corollary 4 *Let the conditions of Theorem 5 be fulfilled. Then, the RLES (44) can be characterised by (45), depends continuously on its input data and is transitive.*

Proof The first two properties are covered by Theorem 5 and Lemma 9. The eigenvectors in (23) can be rotated, such that the final input data is represented by generic orthonormal bases. Consequently, this constitutes a special case for input data with generic orthonormal bases in Proposition 1, which implies the transitivity. \square

To conclude this section, we will consider a small example for the RLES with originally non-generic input data.

Example 3 We calculate the RLES for the input data of Example 1, namely blue $\mathbf{C}_1 = (0, 0, 1)$, a medium dark brown $\mathbf{C}_2 = (\frac{3}{5}, \frac{2}{5}, \frac{1}{5})$ and a shade of blue–magenta $\mathbf{C}_3 = (\frac{1}{3}, \frac{1}{3}, \frac{5}{6})$ with $\mathbf{X}_i = \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \mu_i \mathbf{v}_i \mathbf{v}_i^T$, $i = 1, 2, 3$, and the corresponding spectral data

$$\begin{aligned}
\lambda_1 &= \frac{1}{\sqrt{2}} \approx 0.7071, \quad \mu_1 = -\lambda_1, \\
\mathbf{u}_1 &= \frac{1}{\sqrt{8+4\sqrt{3}}} \begin{pmatrix} -2-\sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.9659 \\ 0.2588 \end{pmatrix}, \\
\mathbf{v}_1 &= \frac{1}{\sqrt{8-4\sqrt{3}}} \begin{pmatrix} 2-\sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.2588 \\ -0.9659 \end{pmatrix}, \\
\lambda_2 &= \frac{1}{5\sqrt{2}} \approx 0.1414, \quad \mu_2 = -\frac{3}{5\sqrt{2}} \approx -0.4243, \\
\mathbf{u}_2 &= \frac{1}{2} \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix} \approx \begin{pmatrix} 0.5000 \\ 0.8660 \end{pmatrix}, \\
\mathbf{v}_2 &= \frac{1}{2} \begin{pmatrix} -\sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.8660 \\ 0.5000 \end{pmatrix} \quad \text{and} \\
\lambda_3 &= \frac{2}{3\sqrt{2}} \approx 0.4714, \quad \mu_3 = -\frac{1}{3\sqrt{2}} \approx -0.2357, \\
\mathbf{u}_3 &= \frac{1}{\sqrt{8+4\sqrt{3}}} \begin{pmatrix} -2-\sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.9659 \\ 0.2588 \end{pmatrix}, \\
\mathbf{v}_3 &= \frac{1}{\sqrt{8-4\sqrt{3}}} \begin{pmatrix} 2-\sqrt{3} \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.2588 \\ 0.9659 \end{pmatrix}.
\end{aligned}$$

Since $\mathcal{B}(\mathcal{X})$ is non-generic for $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\}$ because of \mathbf{X}_1 and \mathbf{X}_3 , we consider a modified multi-set $\mathcal{X}^{(\delta)} =$

$\{\mathbf{X}_1^{(\delta)}, \mathbf{X}_2^{(\delta)}, \mathbf{X}_3^{(\delta)}\}$ according to Definition 7, which results by rotating the corresponding eigenvectors $\mathbf{u}_3, \mathbf{v}_3$ of the matrix \mathbf{X}_3 of \mathbf{C}_3 by $\delta \in (0, \pi)$:

$$\begin{aligned}
\mathbf{X}_1^{(\delta)} &= \mathbf{X}_1, \quad \mathbf{X}_2^{(\delta)} = \mathbf{X}_2 \quad \text{and} \\
\mathbf{X}_3^{(\delta)} &= \lambda_3^{(\delta)} \mathbf{u}_3^{(\delta)} \left(\mathbf{u}_3^{(\delta)}\right)^T + \mu_3^{(\delta)} \mathbf{v}_3^{(\delta)} \left(\mathbf{v}_3^{(\delta)}\right)^T.
\end{aligned}$$

The components of the spectral decomposition of the rotated matrix $\mathbf{X}_3^{(\delta)}$ are

$$\begin{aligned}
\lambda_3^{(\delta)} &= \lambda_3, \quad \mu_3^{(\delta)} = \mu_3, \\
\mathbf{u}_3^{(\delta)} &= \frac{1}{\sqrt{8+4\sqrt{3}}} \begin{pmatrix} (-2-\sqrt{3})\cos(\delta) - \sin(\delta) \\ (-2-\sqrt{3})\sin(\delta) + \cos(\delta) \end{pmatrix}, \\
\mathbf{v}_3^{(\delta)} &= \frac{1}{\sqrt{8-4\sqrt{3}}} \begin{pmatrix} (2-\sqrt{3})\cos(\delta) - \sin(\delta) \\ (2-\sqrt{3})\sin(\delta) + \cos(\delta) \end{pmatrix}.
\end{aligned}$$

Due to the generic nature of $\mathcal{B}(\mathcal{X}^{(\delta)})$, we obtain for the RLES according to Theorem 5:

$$\begin{aligned}
\mathbf{S}' &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_3 \mathbf{v}_1 \mathbf{v}_1^T \\
&= \frac{1}{12\sqrt{2}} \begin{pmatrix} 10+\sqrt{3} & -1 \\ -1 & 10-\sqrt{3} \end{pmatrix} \\
&\approx \begin{pmatrix} 0.6913 & -0.0589 \\ -0.0589 & 0.4872 \end{pmatrix}
\end{aligned}$$

which represents in the RGB space a light shade of blue–magenta colour: $(\frac{5}{6}, \frac{5}{6}, 1) \approx (0.8333, 0.8333, 1)$.

6 Conclusion and Future Work

Building on the work of Burgeth and his co-authors, see [7, 12, 18], we have investigated a characterisation for a previously unexplored application of the log-exp-supremum, which was introduced by Maslov [9] as an approximation of the maximum in convex optimisation, for colour morphology. To do this, we used Burgeth's and Kleefeld's [18] bijective mapping to assign each colour from the RGB colour space a colour in the form of a symmetric real 2×2 matrix in the HCL bi-cone and used the spectral decomposition of these matrices, the Loewner order and properties of the Rayleigh product to calculate the approximation. In particular, we were able to show that the LES is transitive according to equation (23) and, in combination with the dilation, makes it associative in colour morphology. The latter is a property that is otherwise only known from binary and grey value morphology and, to our knowledge, the only multidimensional dilation in colour morphology with this property.

Moreover, our findings indicate that while the LES is not uniquely minimal with respect to the set of upper bounds of the input data derived from the Loewner ordering, it does, however, exhibit unique minimality when considered in conjunction with the intersection of all p -power upper bounds, as defined in Definition 6. To achieve this, the eigenvalues and the Rayleigh product of the exponentiated matrices were employed to eliminate the problematic cases through the limit as $p \rightarrow \infty$.

Finally, we have also addressed one of the biggest downsides of the LES, namely the fact that it does not depend continuously on the input data. However, among other things, this only applied to non-generic configurations of the orthonormal bases of the input data, which allowed us to remove these by means of a relaxation of the LES, the RLES. The input data was changed slightly so that these cases no longer occurred. This change also meant that not only these discontinuities were removed, but all discontinuities were removed without losing the favourable properties of the LES shown above. Since the duality relationship between dilation and erosion is also intact in our method, all the properties of dilation shown here can be transferred to the corresponding properties of erosion.

In light of these findings, there are several avenues for further investigation of this method in the future. The most notable aspect is the significant increase in brightness of the image. One could compare whether the behaviour is similar to that of the paper by Kahra, Sridhar and BreuSS, see [10]. In addition, given the paper [25] by Kahra and BreuSS, it seems possible to find parameters so that these supposed artefacts disappear from [10]. However, this would likely be a challenging undertaking. Alternatively, one could also consider the differences in supremum formation with respect to other suprema employed in colour morphology, such as the one-dimensional case or other multidimensional cases, including the trace supremum [23] or other suprema [8], and compare them with each other. It should be noted that our method is not an optimal case and could be replaced by selecting a colour space other than the HCL bi-cone, such as Lab, in order to potentially offer additional advantages. The choice of this space was influenced by its graphically clear geometric properties.

Author Contributions M.K. and M.B. were involved in conceptualisation; M.K. contributed the methodology, software, formal analysis and investigation, and writing—original draft preparation; M.K., M.B., A.K. and M.W. were responsible for writing—reviewing and editing, and supervision.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors did not receive support from any organisation for the submitted work.

Data Availability No datasets were generated or analysed during the current study.

Materials Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest The authors declare no Conflict of interest.

Ethical Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Najman, J., Talbot, H.: *Mathematical Morphology: From Theory to Applications*. ISTE-Wiley, London-Hoboken (2010). <https://doi.org/10.1002/9781118600788>
2. Roerdink, J.B.T.M.: Mathematical morphology in computer graphics, scientific visualization and visual exploration. In: Soille, P., Pesaresi, M., Ouzounis, G.K. (eds.) *Mathematical Morphology and Its Applications to Image and Signal and Image Processing*, pp. 367–380. Springer, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21569-8_32
3. Serra, J., Soille, P.: *Mathematical Morphology and Its Applications to Image Processing* vol. 2. Springer, Dordrecht (2012). <https://doi.org/10.1007/978-94-011-1040-2>
4. Haralick, R.M., Sternberg, S.R., Zhuang, X.: Image analysis using mathematical morphology. *IEEE Trans. Pattern Anal. Mach. Intell.* **9**(4), 532–550 (1987). <https://doi.org/10.1109/TPAMI.1987.4767941>
5. Barnett, V.: The ordering of multivariate data. *J. R. Stat. Soc. Ser. A* **139**(3), 318–344 (1976). <https://doi.org/10.2307/2344839>
6. Aptoula, E., Lefèvre, S.: A comparative study on multivariate mathematical morphology. *Pattern Recognit.* **40**(11), 2914–2929 (2007). <https://doi.org/10.1016/j.patcog.2007.02.004>
7. Burgeth, B., Kleefeld, A.: An approach to color-morphology based on Einstein addition and Loewner order. *Pattern Recognit. Lett.* **47**, 29–39 (2014). <https://doi.org/10.1016/j.patrec.2014.01.018>
8. Welk, M., Kleefeld, A., Breuß, M.: Quantile filtering of colour images via symmetric matrices. *MMTA* **1**(1), 136–174 (2016). <https://doi.org/10.1515/mathm-2016-0008>
9. Maslov, V.P.: On a new superposition principle for optimization problems. *Russian Math. Surv.* **42**(3), 39–48 (1987). <https://doi.org/10.1070/RM1987v042n03ABEH001439>
10. Kahra, M., Sridhar, V., Breuß, M.: Fast morphological dilation and erosion for grey scale images using the Fourier transform. In: Elmoataz, A., Fadili, J., Quéau, Y., Rabin, J., Simon, L. (eds.)

- SSVM, pp. 65–77. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75549-2_6
11. Sridhar, V., Breuß, M., Kahra, M.: Fast approximation of color morphology. In: Bebis, G., Athitsos, V., Yan, T., Lau, M., Li, F., Shi, C., Yuan, X., Mousas, C., Bruder, G. (eds.) *Advances in Visual Computing*, pp. 488–499. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-90436-4_39
 12. Burgeth, B., Welk, M., Feddern, C., Weickert, J.: Morphological operations on matrix-valued images. In: Pajdla, T., Matas, J. (eds.) *Computer Vision—ECCV 2004*, pp. 155–167. Springer, Berlin, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24673-2_13
 13. Kahra, M., Breuß, M., Kleefeld, A., Welk, M.: An approach to colour morphological supremum formation using the logsumexp approximation. In: Brunetti, S., Frosini, A., Rinaldi, S. (eds.) *DGMM*, pp. 325–337. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-57793-2_25
 14. Ponomarenko, M., Gapon, N., Voronin, V., Egiastian, K.: Blind estimation of white Gaussian noise variance in highly textured images. *arXiv preprint arXiv:1711.10792* (2017) <https://doi.org/10.48550/arXiv.1711.10792>
 15. Boroujerdi, A.S., Breuß, M., Burgeth, B., Kleefeld, A.: PDE-based color morphology using matrix fields. In: Aujol, J.-F., Nikolova, M., Papadakis, N. (eds.) *SSVM*, pp. 461–473. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18461-6_37
 16. Soille, P.: *Morphological Image Analysis: Principles and Applications* vol. 2. Springer, Berlin, Heidelberg (1999). <https://doi.org/10.1007/978-3-662-03939-7>
 17. Sharma, G., Bala, R.: *Digital Color Imaging Handbook*. CRC Press, Bosa Roca (2017). <https://doi.org/10.1201/9781420041484>
 18. Burgeth, B., Kleefeld, A.: Morphology for color images via Loewner order for matrix fields. In: Hendriks, C.L.L., Borgefors, G., Strand, R. (eds.) *Mathematical Morphology and Its Applications to Signal and Image Processing*, pp. 243–254. Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38294-9_21
 19. Jarre, F.: Convex analysis on symmetric matrices. In: *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, pp. 13–27. Springer, Boston, MA (2000). https://doi.org/10.1007/978-1-4615-4381-7_2
 20. Borwein, J.M., Lewis, A.S.: *Convex Analysis and Nonlinear Optimization*. Springer, New York, NY (2000). <https://doi.org/10.1007/978-0-387-31256-9>
 21. Hackbusch, W., et al: *Hierarchical Matrices: Algorithms and Analysis* vol. 49. Springer, Berlin, Heidelberg (2015). <https://doi.org/10.1007/978-3-662-47324-5>
 22. Löwner, K.: Über monotone Matrixfunktionen. *Math. Z.* **38**(1), 177–216 (1934). <https://doi.org/10.1007/BF01170633>
 23. Burgeth, B., Welk, M., Feddern, C., Weickert, J.: *Mathematical Morphology on Tensor Data Using the Loewner Ordering*. Springer, Berlin, Heidelberg (2006). https://doi.org/10.1007/3-540-31272-2_22
 24. Horn, R.A., Johnson, C.R.: *Matrix Analysis*, 2nd edn. Cambridge University Press, New York, NY (2012). <https://doi.org/10.1017/CBO9781139020411>
 25. Kahra, M., Breuß, M.: Properties of morphological dilation in max-plus and plus-prod algebra in connection with the Fourier transformation. *JMIV* **65**(4), 577–591 (2023). <https://doi.org/10.1007/s10851-022-01138-3>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Marvin Kahra received his Master of Science from the Brandenburg University of Technology Cottbus-Senftenberg in 2020. Since then, he has been employed there as a research assistant in the Department of Applied Mathematics while working on his dissertation. His research interests lie in mathematical image processing, differential geometry and variational calculus.



Michael Breuß received his diploma and Ph.D. degrees in mathematics from the University of Hamburg in 1998 and 2001, respectively. He worked as an assistant professor at the University of Hamburg, as post-doctoral researcher at the University of Bordeaux 1 and as a researcher at the Technical University Braunschweig. From 2006 to 2012, he was a research assistant in the Mathematical Image Analysis group at Saarland University. Since 2012, he has been a professor in the Department of Applied Mathematics at the Brandenburg University of Technology Cottbus-Senftenberg. His research interests are centred around numerical methods for partial differential equations, hyperbolic conservation laws and visual computing.



Andreas Kleefeld received his Master and Ph.D. in mathematics from the University of Wisconsin-Milwaukee in 2006 and 2009, respectively. From 2010 to 2016, he was a member of the Numerical Analysis and Scientific Computing group at Brandenburg University of Technology Cottbus-Senftenberg, where he completed his habilitation in 2015. From 2016 onwards, he has worked at the Jülich Supercomputing Centre, part of Forschungszentrum Jülich GmbH, where he leads the

Algorithms, Tools and Methods Laboratory for Numerical and Statistical Methods. Since 2020, he has also been a professor in the Faculty of Medical Engineering and Technomathematics at the FH Aachen – University of Applied Sciences. His research interests include statistical and numerical methods for partial differential equations, nonlinear eigenvalue problems, scattering problems, and related inverse problems.



Martin Welk received his Ph.D. in mathematics in 1998 from the University of Leipzig. From 2002 till 2010, he worked with the Mathematical Image Analysis group at Saarland University, Saarbrücken, Germany. In 2010, he moved to UMIT TIROL – Private University for Health Sciences and Health Technology, Hall/Tyrol, Austria, where he currently holds an Associate Professor position. His research interests are centered around mathematical principles of image analysis, especially PDE

and variational methods, image deconvolution and mathematical morphology.