Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 135

Median and Related Local Filters for Tensor-Valued Images

Martin Welk, Joachim Weickert, Florian Becker,
Christoph Schnörr, Christian Feddern and
Bernhard Burgeth

Saarbrücken 2005

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 135
submitted: April 6, 2005

Median and Related Local Filters for Tensor-Valued Images

Martin Welk
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Saarland University, Building 27, 66041 Saarbrücken, Germany
welk@mia.uni-saarland.de

Joachim Weickert
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Saarland University, Building 27, 66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Florian Becker
Computer Vision, Graphics, and Pattern Recognition Group
Faculty of Mathematics and Computer Science, University of Mannheim, 68131 Mannheim, Germany
flbecker@rumms.uni-mannheim.de

Christoph Schnörr
Computer Vision, Graphics, and Pattern Recognition Group
Faculty of Mathematics and Computer Science, University of Mannheim, 68131 Mannheim, Germany
schnoerr@uni-mannheim.de

Christian Feddern
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Saarland University, Building 27, 66041 Saarbrücken, Germany
feddern@mia.uni-saarland.de

Bernhard Burgeth
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Saarland University, Building 27, 66041 Saarbrücken, Germany
burgeth@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

Abstract

We develop a concept for the median filtering of tensor data. The main part of this concept is the definition of median for symmetric matrices. This definition is based on the minimisation of a geometrically motivated objective function which measures the sum of distances of a variable matrix to the given data matrices. This theoretically well-founded concept fits into a context of similarly defined median filters for other multivariate data. Unlike some other approaches, we do not require by definition that the median has to be one of the given data values. Nevertheless, it happens so in many cases, equipping the matrix-valued median even with root signals similar to the scalar-valued situation.

Like their scalar-valued counterparts, matrix-valued median filters show excellent capabilities for structure-preserving denoising. Experiments on diffusion tensor imaging, fluid dynamics and orientation estimation data are shown to demonstrate this. The orientation estimation examples give rise to a new variant of a robust adaptive structure tensor which can be compared to existing concepts.

For the efficient computation of matrix medians, we present a convex programming framework.

By generalising the idea of the matrix median filters, we design a variety of other local matrix filters. These include matrix-valued mid-range filters and, more generally, M-smoothers but also weighted medians and α-quantiles. Mid-range filters and quantiles allow also interesting cross-links to fundamental concepts of matrix morphology.

Keywords: tensor image processing, local image filter, median filtering, adaptive structure tensor, convex optimisation

1 Introduction

In contemporary image processing, images whose values are second-order tensors represented by symmetric matrices gain increasing importance. They appear as physical quantities which are measured e.g. by diffusion tensor magnetic resonance imaging DTI [26], or computed, as in computational fluid dynamics, or as derived quantities like the *structure tensor* [16] which plays an important role in fields like motion detectio, texture analysis or segmentation.

Pollution of measured or computed tensor fields by noise makes it an important task in matrix-valued image processing to design filters that denoise tensor data without destroying essential image features. While linear filters

like Gaussian convolution transfer straightforward to matrix-valued data by component-wise application, this is not possible for non-linear filters like median filters which offer a better preservation of structures in the denoising process. In [36], a matrix-valued median filter was introduced. The principal idea of this generalisation is to use a minimisation property instead of the rank order which underlies the standard definition of medians. This approach can be extended to further classes of local image filters. In [37] similar generalisations of mid-range filters and the more general class of M-smoothers were presented. As an important algorithmic improvement for the computation of matrix-valued medians and mid-range filters convex optimisation techniques were introduced in [35] and [6].

In this paper we give an overview over the median filtering techniques introduced in [36, 35, 37]. Moreover, we discuss a number of related filters which include the midrange filters and matrix M-smoothers introduced in [37], weighted matrix median filters and matrix quantiles. Midrange and quantile filters reveal relationships to the supremum and infimum filters established in [11, 12].

The paper is organised as follows. In Sec. 2 matrix-valued medians are introduced, important properties discussed and illustrated by experiments on DTI and fluid dynamics data sets. Sec. 3 is devoted to numerical techniques, presenting two approaches. An application to robust structure estimation is the topic of Sec. 4. The following sections 5–7 describe classes of local matrix filters that are extensions of matrix-valued median filtering. Sec. 8 summarises the results.

Related work. Denoising techniques for tensor data have been under intensive investigation recently, mainly in connection with DTI data, see the linear approaches by Westin et al. [38] or the nonlinear ones by Hahn et al. [20]. Nonlinear filters need to take into account the inherent relations of data matrices, either by channel coupling as in Tschumperlé and Deriche [31], or by working on derived quantities like eigendecompositions [27, 31] or fractional anisotropy [25].

Median filtering in signal processing has been established by Tukey [32] and has now become a standard technique in image processing, see Dougherty and Astola [15] or Klette and Zamperoni [22]. The matrix-valued median definition given by Welk et al. in [36] stands in the context of earlier attempts to vector-valued median filtering, see e.g. [3], [23]. In an image processing context, we mention Astola et al. [1] and Caselles et al. [13]. Both definitions are built on the property of the median to be one of the *given* vectors, with a slight extension by admitting also their arithmetic mean in [1]. This

property is also required in [4] by Barni et al. who otherwise use Euclidean distance sum minimisation similar to [36]. Surprisingly, for 2-D vectors already Austin's 1959 paper [2] proposes the exact analog of the definition given in [36]. Austin also gives a graphical algorithm which can be considered a direct predecessor of the gradient descent algorithm in [36]. Seymour's 1970 reply [29] to Austin discusses algorithmical difficulties and improvements of this procedure. Moreover, vector-valued medians and mid-range values (often by the name of 1-centres) have also been investigated in the literature on facility location problems, see the papers by Megiddo [24], Fekete et al. [17] and the references therein.

The convex programming method discussed in this paper relies on concepts which can e.g. be found in the book by Boyd and Vandenberghe [7]. For applications in image processing contexts we mention Keuchel et al. [21].

The structure tensor has been established by Förstner and Gülch [16]. It is constructed by Gaussian smoothing of the outer product matrices $\nabla u \nabla u^{\mathrm{T}}$ of the image gradient. To adapt better to orientation discontinuities, a modification called nonlinear structure tensor has been proposed by Weickert and Brox [34, 8]. Here, the Gaussian smoothing is replaced by a nonlinear diffusion process. Nonlinear structure tensors have proven their use in texture segmentation [8, 28] and motion analysis [10]. Another way to introduce structure adaptivity into the structure tensor has been opened by van den Boomgaard and van der Weijer [33] who proposed a concept of robust structure tensor which is also linked to matrix-valued medians. For both adaptive structure tensor concepts, see also [9].

2 Median Filtering of Tensor-Valued Data

2.1 Scalar-Valued Median Filters

Given a finite set of real numbers, their median is defined as the middle element in the sequence that contains these numbers ordered according to size. It can be considered as a robust average since it is insensitive to outliers in the given data set. The median operation commutes with monotonic transformations of the data.

Without any reference to ordering, the median of the n-tuple $S = (a_1, \ldots, a_n)$ can be characterised as the minimiser of the convex function

$$E_S(x) := \sum_{i=1}^{n} |x - a_i| \tag{1}$$

where $|x - a|$ is the Euclidean distance of real numbers.

The median concept gives rise to a local image filter with interesting properties. *Median filtering* requires the specification of a neighbourhood for each pixel which is commonly chosen either as a $(2k+1) \times (2k+1)$ square or a discretely approximated disc centred at the pixel. The new grey-value of a pixel is obtained as the median of the old grey-values within the neighbourhood. Median filtering can be iterated and so constitutes a discontinuity-preserving denoising process. The insensitivity of medians to outliers enables median filtering to cope even with extreme types of noise like uniform or impulse noise. Unlike Gaussian smoothing, median filtering possesses non-constant steady states called *root signals*.

For a space-continuous variant of median filtering, Guichard and Morel [19] have proven that it approximates mean curvature flow, thereby establishing a remarkable link between a nonlinear local filter and a PDE-based image evolution.

2.2 Matrix-Valued Median Filters

Pollution of tensor image data with noise makes it a desideratum to provide a discontinuity-preserving and robust denoising filter for such data. To meet this need, we want to generalise median filtering to matrix-valued images. The main task in doing so is to give an appropriate notion of medians for matrices since the construction of the local image filter by applying the median to input values from a neighbourhood transfers straightforward.

While not all properties of scalar-valued medians can be retained by such a generalisation, the following requirements are essential from the modeling viewpoint:

Preservation of symmetry. The median of symmetric matrices must again be a symmetric matrix.

Scaling invariance. For a real number λ, the median med should satisfy

$$\operatorname{med}(\lambda A_1, \ldots, \lambda A_n) = \lambda \operatorname{med}(A_1, \ldots, A_n)$$

for arbitrary input matrices $A_1, \ldots, A_n$.

Rotational invariance. Rotating all input matrices by the same rotation matrix R should result in equal rotation of the median:

$$\operatorname{med}(R^{\mathrm{T}} A_1 R, \ldots, R^{\mathrm{T}} A_n R) = R^{\mathrm{T}} \operatorname{med}(A_1, \ldots, A_n)\, R\,.$$

Embedding of scalar-valued median. If all input matrices are scalar multiples of the same non-zero matrix A, the median should reduce to the scalar median:

$$\operatorname{med}(\lambda_1 A, \ldots, \lambda_n A) = \operatorname{med}(\lambda_1, \ldots, \lambda_n)\, A\ .$$

Preservation of positive semidefiniteness. Since positive semidefiniteness is an indispensable property of some sorts of matrix data, such as DTI or structure tensor fields, a sensible filter for such data should not destroy it.

Since matrices lack a linear ordering, a rank-order approach to defining matrix medians is impractical. Instead, we generalise the minimising property (1).

Definition 1 *Given a matrix tuple $S = (A_1, \ldots, A_n)$, the minimiser of*

$$E_S(X) = \sum_{i=1}^{n} \|X - A_i\| \tag{2}$$

where $\|\cdot\|$ is a matrix norm is called median of S and denoted by $\operatorname{med}(S)$.

While E_S is convex for any norm $\|\cdot\|$, rotational invariance and semidefiniteness preservation restrict the choice of $\|\cdot\|$. Before discussing possibilities, we notice one property of the median which is independent of the norm.

Lemma 1 *Let $X = \operatorname{med}(A_1, \ldots, A_n)$. If each A_i is replaced by $A_i' := X + k_i(A_i - X)$ with real $k_i > 0$, then X is also the median of $A_1', \ldots, A_n'$.*

Accordingly, the matrices A_i can be shifted along the rays from X to A_i without affecting the median. The statement follows directly from the scaling property $\|kA\| = |k|\,\|A\|$, $k \in \mathbb{R}$. It can be considered a restricted form of the independence on outliers known from the scalar-valued median.

2.2.1 Choice of possible norms

We consider three norms for $d \times d$ matrices. All are constructed from the eigenvalues $\lambda_1(A), \ldots, \lambda_d(A)$ of A, which guarantees rotational invariance.

- The first norm is the so called *nuclear norm* which is given by

$$\|A\|_{(1)} = \sum_{j=1}^{d} |\lambda_j(A)|$$

  For positive semidefinite matrices, we have $\|A\|_{(1)} = \mathrm{tr}(A)$. Typically, however, even when the A_i are positive semidefinite, the differences $X - A_i$ aren't.

- Second, we consider the Frobenius norm which can also be computed directly from the matrix entries a_{jk}, $j, k = 1, \ldots, d$, of A,

$$\|A\|_{(2)} = \sqrt{\sum_{j=1}^{d} (\lambda_j(A))^2} = \sqrt{\sum_{j,k=1}^{d} a_{jk}^2} \ .$$

- Third, we have the spectral norm

$$\|A\|_{(\infty)} = \max_{j=1,\ldots,d} |\lambda_j(A)| \ .$$

We remark that these norms are examples of the family of norms

$$\|A\|_{(p)} = \left(\sum_{j=1}^{d} |\lambda_j(A)|^p \right)^{1/p}, \quad p \geq 1$$

which includes the spectral norm as limit case $p \to \infty$.

For brevity, we shall refer to medians defined via these norms as *nuclear median* med_1, *Frobenius median* med_2 and *spectral median* med_∞. We turn now to study their further properties.

2.2.2 Properties of the Frobenius median

The Frobenius norm coincides with the Euclidean norm if the $d \times d$ matrices are interpreted as vectors in $\mathbb{R}^{d^2}$. As a consequence, in this case our median definition is not restricted to symmetric square matrices but works equally on non-square matrices, including vectors. For $\mathbb{R}^2$, Austin's bivariate median [2] is recovered.

Proposition 2 *The Frobenius median* $\mathrm{med}_2(S)$ *of a tuple* $S = (A_1, \ldots, A_n)$ *of matrices is a convex combination of* $A_1, \ldots, A_n$.

Proof. We identify matrices with vectors and denote by $\langle \cdot, \cdot \rangle$ the corresponding Euclidean scalar product. Assume now that X is a matrix outside the convex hull of $A_1, \ldots, A_n$. Then a hyperplane h separates X from $A_1, \ldots, A_n$. Let Y be the orthogonal projection of X onto h, i.e. $X - Y$ is perpendicular to h. Then $\langle X - Y, Y - A_i \rangle$ is positive for $i = 1, \ldots, n$. Hence,

$$\begin{aligned}
&\langle X - A_i, X - A_i \rangle - \langle Y - A_i, Y - A_i \rangle \\
&\quad = \langle X, X \rangle - 2\langle X, A_i \rangle - \langle Y, Y \rangle + 2\langle Y, A_i \rangle \\
&\quad = 2\langle X - Y, Y - A_i \rangle + \langle X - Y, X - Y \rangle > 0
\end{aligned}$$

which proves that X is not the minimiser of (2). □

Since convex combinations of positive semidefinite matrices are positive semidefinite, the following corollary is obvious.

Corollary 3 *The Frobenius median* $\mathrm{med}_2(S)$ *of a tuple* $S = (A_1, \ldots, A_n)$ *of positive semidefinite symmetric* $d \times d$ *matrices is positive semidefinite.*

We use the planar Euclidean case of 2-dimensional vectors to illustrate simple geometric properties of our median concept. Three points in the plane which span a triangle with all angles smaller than 120 degrees have as their median the so-called Fermat–Torricelli or Steiner point. From this point, each connecting line between two of the given points appears under a 120 degree angle. If instead one angle of the triangle is larger or equal 120 degrees, then its vertex is the median. In the case of four points spanning a convex quadrangle, the median is the intersection point of the diagonals. The median of four points whose convex hull is a triangle is the one of the points which is not a corner of this triangle. Combinatorial and geometric complexity prevents similar elementary geometric considerations for more points or higher dimensions. It is evident, though, that although our definition does not force the median to be one of the given data points, this still happens to be true in many generic cases. Only if none of the given values is located sufficiently well in the middle of the data set, a new value is created.

2.2.3 Properties of the nuclear median

With the nuclear norm, the energy (2) displays non-strict convexity in a broader range of configurations, leading to non-unique minimisers. Our result on semidefiniteness is therefore weaker than before.

Proposition 4 *Let a tuple $S = (A_1, \ldots, A_n)$ of positive semidefinite $d \times d$ matrices be given, and consider the objective function E_S with the nuclear norm. If E_S is minimised by a matrix which is not positive semidefinite, then there exists also a positive semidefinite argument for which E_S attains the same value.*

Proof. We consider a symmetric matrix X whose smallest eigenvalue μ is negative. The difference matrix $X - A_i$ for any A_i has two eigenvalues $\lambda_1 \geq \lambda_2$ where $\lambda_2 \leq \mu$. The matrix $X - \mu I - A_i$ has the same eigensystem as $X - A_i$, with both eigenvalues shifted by μ. From $\lambda_2 \leq \mu$ it follows that

$$\begin{aligned} \|X - \mu I - A_i\|_{(1)} &= |\lambda_1 - \mu| + |\lambda_2 - \mu| \\ &\leq |\lambda_1| + |\mu| + |\lambda_2| - |\mu| \\ &= \|X - A_i\|_{(1)} \ . \end{aligned}$$

This proves the statement of the proposition. □

As another remarkable property of the nuclear median, we mention that it reveals an insensitivity w.r.t. outliers which goes beyond the one described in Lemma 1 and is in fact close to the corresponding property of its scalar-valued counterpart.

Lemma 5 *Let $X = \mathrm{med}_1(A_1, \ldots, A_n)$. Assume that for the data matrix A_i, the difference $X - A_i$ is positive or negative definite. If A_i is replaced by some other A_i' for which $X - A_i'$ has the same (positive or negative) definiteness as $X - A_i$, then X is also the nuclear median of $A_1, \ldots, A_i', \ldots, A_n$.*

A disadvantage of this behaviour is that the orientation of the median X depends exclusively on those A_i for which $X - A_i$ is indefinite.

2.2.4 Properties of the spectral median

The spectral median deviates from the previously discussed variants in that it does not always preserve positive semidefiniteness. The three positive definite 2×2 matrices $\begin{pmatrix} 2.95 & 0.42 \\ 0.42 & 0.07 \end{pmatrix}$, $\begin{pmatrix} 2.95 & -0.42 \\ -0.42 & 0.07 \end{pmatrix}$, $\begin{pmatrix} 4 & 0 \\ 0 & 0.01 \end{pmatrix}$ form a counter-example since their spectral median $\begin{pmatrix} 2.95 + 0.07\sqrt{3} & 0 \\ 0 & 0.07 - 0.07\sqrt{3} \end{pmatrix} \approx \begin{pmatrix} 3.0712 & 0 \\ 0 & -0.0512 \end{pmatrix}$ is indefinite.

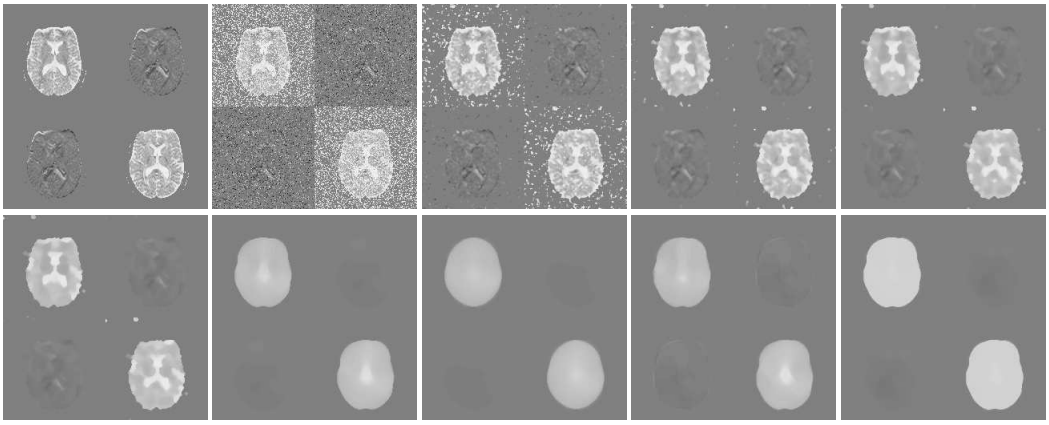

Figure 1: 2-D DTI data median filtering. *Top, left to right:* **(a)** One slice from a DTI brain scan. Only tensor components belonging to the cut plane are shown. **(b)** 30 % of the matrices of (a) have been replaced with uniform noise (uniform in directions and uniform in both eigenvalues). **(c)** Image (b) filtered by Frobenius median, 3×3 stencil, 1 iteration. **(d)** 5 iterations. **(e)** 25 iterations. *Bottom, left to right:* **(f)** Frobenius median, 3×3 stencil, 125 iterations. **(g)** Frobenius median, 7×7 stencil, 25 iterations. **(h)** Frobenius median, 9×9 stencil, 25 iterations. **(i)** Nuclear median, 7×7 stencil, 25 iterations. **(k)** Spectral median, 7×7 stencil, 25 iterations.

For symmetric 2×2 matrices, the following statement can be proven by explicit calculation of the stationarity conditions.

Lemma 6 *Let n symmetric 2×2-matrices $A_1, \ldots, A_n$ be given whose spectral median is X. Assume that $A_i = \begin{pmatrix} a_i & c_i \\ c_i & b_i \end{pmatrix}$, $i = 1, \ldots, n$, and $X = \begin{pmatrix} x & z \\ z & y \end{pmatrix}$. Then the following are true:*

- *The trace of X is a scalar median of the traces of $A_1, \ldots, A_n$ (in the sense that it minimises (1)).*
- *The vector $(x - y, 2z)^{\mathrm{T}}$ is the bivariate median w.r.t. Euclidean norm of the vectors $(a_i - b_i, 2c_i)$, $i = 1, \ldots, n$.*

2.2.5 Examples

To demonstrate the denoising capabilities of median filtering we use two test images. First, we show in Fig. 1 (a, b) a 2-D slice from a diffusion tensor

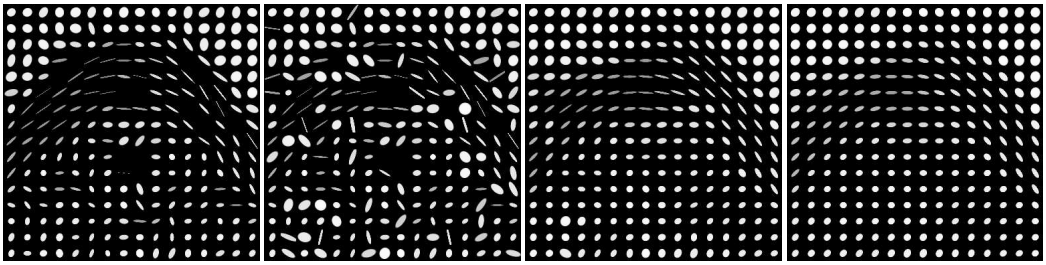

Figure 2: 2-D DTI data and median filtering results visualised by ellipses (see text). *Left to right:* **(a)** Corpus callosum detail from original image. **(b)** Same from noisy image. **(c)** Frobenius median, 3×3 stencil, 1 iteration. **(d)** 5 iterations.

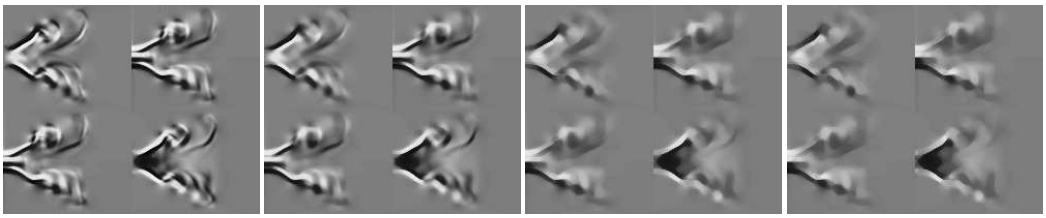

Figure 3: Frobenius median filtering of a tensor field containing indefinite matrices. The data are deformation tensors originating from a fluid dynamics simulation. *Left to right:* **(a)** Initial data, 124×101 pixels. **(b)** 10 iterations, 3×3 stencil. **(c)** 100 iterations. **(d)** 1000 iterations. From [37].

magnetic resonance data set and a noisy version of it where 30 % of all matrices have been replaced by noise matrices. The eigenvector orientations of the noise matrices are uniformly distributed on the circle while the eigenvales are uniformly distributed over an interval $[0, M]$ that covers the eigenvalue range of the original measurements.

In the grey-value visualisation each sub-image shows the values of one matrix component over the whole image. A middle grey-tone represents zero. Note that the upper right and lower left sub-images are equal because of the symmetry of the matrices. For positive semidefinite tensor data such as DTI, the main diagonal entries in the upper left and lower right sub-image contain only nonnegative values while the off-diagonal entries can be of either sign but have smaller variation.

Fig. 1 (c–k) shows results of iterated median filtering. The top row illustrates the effect of increasing numbers of iterations with a 3×3 stencil. Note that the third and fourth image hardly differ, evidence that root signals occur also in matrix-valued median filtering. The second row shows the progressive

simplification of shapes by increasing stencil size. It also demonstrates that the *shape* of objects in the median-filtered images depends not seriously on the particular norm chosen in the median computation. However, under spectral median filtering the matrices tend to higher magnitudes (higher contrast in the main diagonal entries) and faster reduction of anisotropy (low contrast in off-diagonal parts).

An alternative visualisation is used in Fig. 2 which shows detail views from the original and median-filtered images. Here each matrix is represented by an ellipse whose principal axis directions coincide with the eigenvector directions of the matrix. The principal axis lengths are proportional to the eigenvalues. However, each ellipsis as a whole has been rescaled such that the areas are proportional to the third roots of the determinants, instead of the determinants themselves. Compared to a representation with constant scale for all matrices, we achieve so a better representation of matrices with large variations in magnitude within the image, without reducing eccentricities.

Our second example in Fig. 3 uses a deformation tensor set from computational fluid dynamics. Here, the eigenvalues of the tensors are of different signs. The structure-preserving smoothing effect of the median filter is again visible.

3 Algorithms for Computing Matrix Medians

Only in simple cases it is possible to compute matrix medians directly. In general numerical approximation methods are required. We discuss two approaches.

3.1 Computation by Gradient Descent

In computing Frobenius medians of matrices, the convexity of $E_S(X)$ and its differentiability except at $X = A_i$ motivate the use of gradient descent techniques. One difficulty has to be overcome: Since the gradient vector $\nabla \|A_i - X\|_{(2)}$ has equal length for all $X \neq A_i$, it lacks any information about the distance to A_i. This deficiency is inherited by the gradient of E_S. Though clearly indicating the *direction* for descent, it is useless in determining how far to go in one step. A remedy for this is to use an adaptive step-size control

which uses information from the over- and undershoots encountered during iteration.

The algorithm starts by identifying the $A_j \in S$ with the smallest $E_S(A_j)$. If for this A_j, we have $\left\|\nabla \sum_{i \neq j} \|A_i - A_j\|\right\| \leq 1$, then A_j is also the global minimiser, so we stop. Otherwise we choose an arbitrary initial step size $s_0 > 0$ and proceed by gradient descent in the direction of $-\nabla E_S(X)$. After each iteration step, the step size is adapted as follows. Assume that the matrix X_{k-1} has been replaced in step k by $X_k = X_{k-1} - s_k \nabla E_S(X_{k-1})$. We compare now the projection of $\nabla E_S(X_k)$ onto $\nabla E_S(X_{k-1})$ to detect over- and undershoots. Our indicator is the quotient

$$r := \frac{\langle \nabla E_S(X_{k-1}), \nabla E_S(X_k) \rangle}{\langle \nabla E_S(X_k), \nabla E_S(X_k) \rangle} \ .$$

If $r < 0$, an overshoot has occurred while $r > 0$ signals an undershoot. The ratio r is then used in two ways. First, the step size for the next step is adapted by a rule like $s_{k+1} = s_k/(1-r)$. In practice, one limits the adaptation factor per iteration step e.g. to $[1/2, 2]$. Second, in case of extreme overshoots, i.e. $r < r_{\text{crit}}$ with $r_{\text{crit}} \in (-1, 0)$, step k is rolled back and repeated with the new step size.

3.1.1 Adaptation to the nuclear and spectral median

While the Frobenius norm is differentiable everywhere except at zero, the spectral norm displays additional singularities along the hypersurfaces of matrices with multiple eigenvalues. Similarly, the nuclear norm is non-differentiable at singular matrices. As a consequence, also $E_S(X)$ for the nuclear or spectral median possesses these hypersurface-singularities which arise from the maximum operation applied on two differentiable functions $f_1(X)$ and $f_2(X)$ (absolute values of different eigenvalues for the spectral norm, an eigenvalue and its negative for the nuclear norm). By replacing $\max(f_1(X), f_2(X))$ with $wf_1(X)+(1-w)f_2(X)$ where $w = w(f_1(X)-f_2(X))$ is a smoothed Heaviside function, we achieve differentiability everywhere outside the A_i. The gradient descent algorithm then works as before.

3.2 A Convex Optimisation Approach

Another attractive method to compute matrix medians which bypasses elegantly the difficulties of the gradient descent starts directly from the optimisation form of our definition. By a chain of transformations, the median

definition is translated into a convex optimisation problem that admits the use of established and efficient algorithms. This approach has been described in [35] and, more detailed, in [6].

3.2.1 Frobenius median

First we develop the framework in the case of the Frobenius median. In the definition

$$\mathrm{med}_2(A_1,\ldots,A_n) := \operatorname*{argmin}_{X\in\mathcal{S}^d} \sum_{i=1}^{n} \|X - A_i\|_{(2)} \ , \tag{3}$$

we identify again matrices from the space of symmetric $d \times d$ matrices, $\mathcal{S}^d$, with vectors in $\mathbb{R}^{d^2}$. By introducing a vector $t = (t_1,\ldots,t_n)^{\mathrm{T}}$ of additional variables the problem (3) can be rewritten as

$$\inf_{X\in\mathbf{R}^{d^2},t\in\mathbf{R}^n} (t_1+\ldots+t_n), \qquad \|X - A_i\|_{(2)} \le t_i,\ i=1,\ldots,n \ .$$

Note that each vector $(X^{\mathrm{T}}, t_i)^{\mathrm{T}}$ varies within a convex constraint set C_i which can be written as a translated convex cone,

$$C_i = \begin{pmatrix} A_i \\ 0 \end{pmatrix} + \mathcal{L}^{d^2+1} \ , \quad \mathcal{L}^{d^2+1} := \left\{ x \in \mathbb{R}^{d^2+1} \;\middle|\; x_{d^2+1} \ge \sqrt{x_1^2+\ldots+x_{d^2}^2} \right\} \ .$$

With the weight vector $e = (1,\ldots,1)^{\mathrm{T}}$ we finally formalise our optimisation problem as

$$\inf_{X\in\mathbf{R}^{d^2},t\in\mathbf{R}^n} \langle e, t\rangle, \qquad \begin{pmatrix} X \\ t \end{pmatrix} \in \bigcap_{i=1}^{n} C_i \ . \tag{4}$$

Since the intersection of convex sets is again convex, and the objective function is linear, this is a convex optimisation problem.

We continue by transforming this problem into a convex program that allows the application of a numerical interior-point algorithm.

With the $n(d^2+1) \times (d^2+n)$-matrix F and the vector $g \in \mathbb{R}^{n(d^2+1)}$ given by

$$F := \begin{pmatrix} F_1 \\ \vdots \\ F_n \end{pmatrix}, \qquad g := \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix},$$

$$F_i := \begin{pmatrix} I_{d^2} & 0_{d^2\times n} \\ 0_{1\times d^2} & e_i \end{pmatrix}, \qquad g_i := \begin{pmatrix} A_i \\ 0 \end{pmatrix}, \quad i = 1,\ldots,n$$

where I_m denotes a $m \times m$ unit matrix and $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ the i-th unit row vector, we can rewrite (4) as

$$\inf_{X \in \mathbf{R}^{d^2}, t \in \mathbf{R}^n} \langle e, t \rangle, \qquad F \begin{pmatrix} X \\ t \end{pmatrix} - g \in \mathcal{K} \,. \tag{5}$$

Here, the convex cone $\mathcal{K}$ is given by $\mathcal{K} = \left(\mathcal{L}^{d^2+1}\right)^n$. The problem (5) is a convex conic program.

In general, a convex conic program

$$\inf_{x \in \mathbf{R}^m} \langle c, x \rangle \;, \quad Fx - g \in \mathcal{K}$$

with some cone $\mathcal{K} \subset \mathbb{R}^m$ corresponds to the dual conic program

$$\sup_y \langle g, y \rangle \;, \quad F^{\mathrm{T}} y = c, \; y \in \mathcal{K}^*$$

where $\mathcal{K}^*$ is the dual cone for $\mathcal{K}$. If at least one of the problems – the original or the dual one – is bounded and strictly feasible, then $\{x, y\}$ is a pair of optimal solutions if and only if the duality gap $\langle c, x \rangle - \langle g, y \rangle$ is zero.

In (5) the cone is self-dual, $\mathcal{K} = \mathcal{K}^*$. The dual conic problem thus reads

$$\sup_{Y_i \in \mathbf{R}^{d^2}} \sum_{i=1}^n \langle Y_i, A_i \rangle, \quad \sum_{i=1}^n Y_i = 0, \quad \|Y_i\|_{(2)} \leq 1, \; i = 1, \ldots, n \,.$$

By virtue of $\langle \sum_{i=1}^n Y_i, X \rangle = 0$ the objective function can be rewritten as $\sum_{i=1}^n \langle Y_i, A_i - X \rangle$. The vanishing of the duality gap yields

$$\sum_{i=1}^n \|X - A_i\|_{(2)} = \sum_{i=1}^n \langle Y_i, A_i - X \rangle$$

which together with the constraints $\|Y_i\|_{(2)} \leq 1$ directly lead to the solution

$$Y_i = \frac{A_i - X}{\|A_i - X\|_{(2)}}, \quad i = 1, \ldots, n \,,$$

which after substitution into $\sum_{i=1}^n Y_i = 0$ reproduces the stationarity condition known from the previous sections

$$\sum_{i=1}^n \frac{X - A_i}{\|X - A_i\|_{(2)}} = 0 \,.$$

3.2.2 Spectral median

The definition of the spectral median,

$$\mathrm{med}_\infty(A_1,\ldots,A_n) := \operatorname*{argmin}_{X\in\mathcal{S}^d} \sum_{i=1}^n \|X-A_i\|_{(\infty)} \ , \tag{6}$$

is treated analogously by introducing a vector $t \in \mathbb{R}^n$ of auxiliary variables and the corresponding constraints

$$\|X-A_i\|_{(\infty)} \le t_i, \quad i=1,\ldots,n \ .$$

In this case, the constraints can be decomposed into the requirements

$$t_i I_d - (X-A_i) \in \mathcal{S}^d_+, \qquad t_i I_d + (X-A_i) \in \mathcal{S}^d_+, \quad i=1,\ldots,n,$$

that must be satisfied simultaneously. Note that X, A_i are to be read as matrices here. One finds that again each $(X^{\mathrm{T}}, t_i)^{\mathrm{T}}$ is confined to a convex set obtained from intersecting the affine set from the left-hand side with the convex cone of positive definite symmetric $d \times d$-matrices $\mathcal{S}^d_+$. Abbreviating these constraint sets by $C_{i,+}$, $C_{i,-}$, $i=1,\ldots,n$, (6) is transformed into

$$\inf_{X\in\mathcal{S}^d, t\in\mathbf{R}^n} \langle e,t\rangle \ , \quad \begin{pmatrix} X \\ t \end{pmatrix} \in \bigcap_{i=1}^n (C_{i,+}\cap C_{i,-}) \tag{7}$$

which is easily identified as a convex optimisation problem because of its convex constraint set and linear objective function.

If positive semidefinite data are processed, $A_i \in \mathcal{S}^d_+$, $i=1,\ldots,n$, the constraints represented by the $C_{i,-}$ constitute no restriction and can therefore be discarded.

The conditions can again be cast into a conic program formulation

$$\begin{gathered} \inf_{X\in\mathcal{S}^d, t\in\mathbf{R}^n} \langle e,t\rangle \ , \quad F\begin{pmatrix} X \\ t \end{pmatrix} - G \in \mathcal{S}^{n\times d^2}_+ \ , \\ F(X,t) = \mathrm{diag}(\ldots, t_i I_d - X, \ldots, t_i I_d + X, \ldots) \\ G = \mathrm{diag}(\ldots, -A_i, \ldots, +A_i, \ldots) \ . \end{gathered}$$

3.2.3 The nuclear median

Similarly like the Frobenius median, the nuclear median defined by

$$\mathrm{med}_1(A_1,\ldots,A_n) := \operatorname*{argmin}_{X\in\mathcal{S}^d} \sum_{i=1}^n \|X-A_i\|_{(1)} \tag{8}$$

Table 1: Average angular errors (AAE) measured in orientation estimation. Method-specific parameters in brackets include stencil diameter and iteration count for median, m and s (see [33]) for Boomgaard–Weijer tensor.

Method	AAE undisturbed	AAE impulse noise	AAE Gaussian noise
gradient direction	3.387°	20.612°	31.429°
Frobenius median			
w/o normalising	1.591° (7, 1)	1.914° (9, 4)	3.207° (9, 5)
with normalising	1.312° (7, 1)	1.655° (5, 5)	3.434° (15, 4)
Boomgaard–Weijer	1.634° (0.1, 3)	1.489° (0.05, 5)	3.657° (0.05, 9)

is translated into the optimisation problem

$$\inf_{X \in \mathcal{S}^d, t \in \mathbf{R}^n} \langle e, t \rangle \,, \quad \|X - A_i\|_{(1)} \leq t_i, \quad i = 1, \ldots, n \,.$$

The constraint sets are easily checked to be convex again such that we have a convex optimisation problem.

4 Robust Structure Estimation

As a discontinuity-preserving matrix smoother, the matrix median can be used to smooth orientation information that is extracted from textured images via structure tensors. This application has been exposed in [35, 6].

Structure tensors [16] are computed by Gaussian smoothing of the outer product matrices $\nabla u \nabla u^{\mathrm{T}}$ of an image u. They encode local orientation estimation integrated within a neighbourhood on the scale of the Gaussian which is used.

A matrix median filtering step can now be employed for a robust filtering of these structure tensors. This is demonstrated by Fig. 4. Pursuing this idea further, the Gaussian smoothing can even be omitted; one then applies the median filtering directly on the rank one outer product matrices. When processing structure information from images with no or moderate noise, see e.g. Fig. 5 a favourable smoothing is achieved which keeps discontinuities in the orientation field fairly sharp.

For images contaminated with stronger noise, see Fig. 6, the quality of the results is still less satisfactory. Improvements of the orientation estimation are made using two modifications which can be used separately or combined.

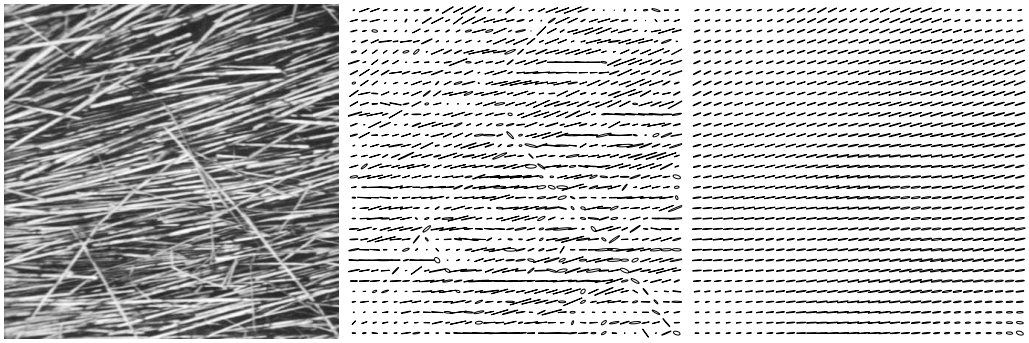

Figure 4: *Left to right:* **(a)** Image containing oriented texture with inhomogeneities. **(b)** Structure tensors computed by smoothing the outer products $\nabla u \nabla u^\top$ with 15×15 Gaussian. Gradients have been calculated by 3×3 derivative-of-Gaussian filtering. The final matrix field has been subsampled for visualisation. **(c)** Result of Frobenius median filtering of (b) with 7×7 stencil, subsampled. From [35].

First, since we are only interested in directional information, the gradients (or, equivalently, the outer product matrices) can be normalised before median filtering. Second, median filtering itself can be iterated. The experiments in Fig. 7 reveal that in case of impulse noise, each of these ideas is capable of sharpening the discontinuity. For Gaussian noise, iterated median filtering gives the greater gain in performance. The combination in this case does not pay off significantly. Table 1 juxtaposes quality measurements based on average angular errors for the different methods.

To end this section, we want to point out another aspect. The classical structure tensor smoothes outer product matrices by means of the Gaussian scale space which is simple and efficient but insensitive to features. In [34], compare also [9], Weickert and Brox have replaced Gaussian smoothing, which is in fact a linear diffusion process, by a feature-preserving nonlinear diffusion process, yielding a *nonlinear* structure tensor. Assigning the role of the smoothing process to iterated median filtering, which also constitutes a scale space, stands in analogy to this procedure and can be seen as construction of a *robust* structure tensor.

The notion of robust structure tensor has also been used by Boomgaard and Weijer in [33], see also [9]. They propose the minimisation of an objective function which leads to a (noniterated) weighted median, compare Sec. 6 below. Since the weights are defined by a Gaussian, the Boomgaard–Weijer tensor in fact combines median and diffusion operations in one filter. We

Figure 5: *Left to right:* **(a)** Synthetic image with oriented textures, inspired by [33]. **(b)** Local orientations computed via derivatives of Gaussians. Orientations have been mapped to grey-values. Note that the orientations represented by black and white are close neighbours. **(c)** Orientations after median filtering of the orientation matrices with Frobenius norm and a disk-shaped structure element of diameter 7. **(d)** Same with structure element of diameter 9. **(e)** Spectral norm median filtering, diameter 9. From [35].

include orientation estimates with the Boomgaard–Weijer tensor in Fig. 7 and Table 1.

5 Matrix M-Smoothers and Mid-Range Filters

The minimisation approach underlying the matrix median definition can easily be extended to a larger class of local image filters, cf. [37].

5.1 Mid-Range Values and Mid-Range Filters

Given a set S of real numbers, its mid-range value is simply the arithmetic mean of their maximum and minimum. A mid-range filter is then obtained by taking the mid-range value of the grey-values within a suitable neighbourhood of a pixel. Mid-range filters are rarely used for denoising purposes since they perform reasonably only in fairly special situations (noise distributions with "thin tails"). They can, however, be used in the construction of more relevant filters.

A generalisation to matrix data is easily derived from the observation that the scalar mid-range value minimises the convex function

$$E_S(x) = \max_{i=1,\dots,n} |x - a_i| \ .$$

The generalisation is then straightforward.

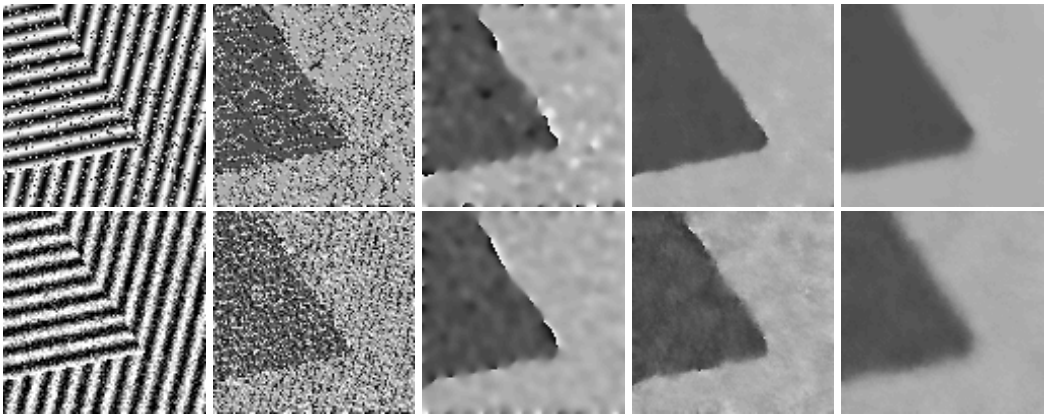

Figure 6: *Top, left to right:* **(a)** Test image with 20 % impulse noise. **(b)** Orientation field of (a). **(c)** Structure tensor orientation obtained by Gaussian smoothing of the outer product matrices with standard deviation 19. **(d)** Same after median filtering with Frobenius norm and disk-shaped structure element of diameter 9. **(e)** Median filtering of (a) with Frobenius norm and disk-shaped structure element of diameter 19. *Bottom, left to right:* **(f)** Test image perturbed by Gaussian noise of standard deviation 0.2 (where grey-values vary between 0 and 1). **(g)** Orientation field of (f). **(h)** Structure tensor orientation as in (c). **(i)** Median filtering as in (d). **(k)** Median filtering as in (e). From [35].

Definition 2 *Given a tuple* $S = (A_1, \ldots, A_n)$ *of symmetric matrices, its mid-range value* $\mathrm{midr}(S)$ *is the minimiser of the convex function*

$$E_S(X) = \max_{i=1,\ldots,n} \|X - A_i\| \tag{9}$$

with a matrix norm $\|\cdot\|$.

Based on similar requirements as for the median, nuclear, Frobenius and spectral norm are once more suitable choices. An example of a mid-range filtered image is shown in Fig. 8.

5.2 M-Estimators and M-Smoothers

Replacing the distances $|x - a_i|$ in the function (1) by their p-th powers, $p > 0$, yields a more general class of nonlinear averages for real numbers. Minimisers of

$$E_S(x) := \sum_{i=1}^{n} |x - a_i|^p$$

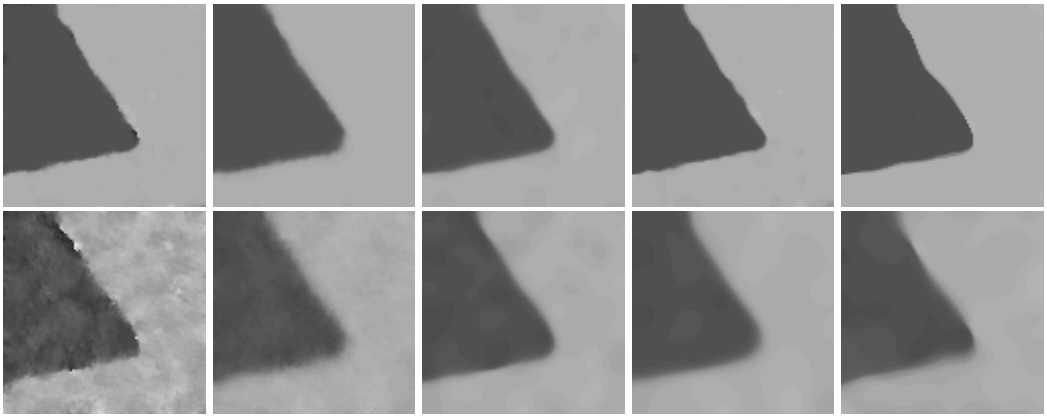

Figure 7: *Top row:* Modified local orientation filtering for the impulse-noise image, Fig. 6 (a). *Left to right:* **(a)** Frobenius median filtering of normalised outer product matrices with disk-shaped stencil of diameter 9. **(b)** As (a) but with stencil of diameter 19. **(c)** Four iterations of the median filter from Fig. 6 (d). **(d)** Five iterations of median filter with normalisation, stencil diameter 5. **(e)** Orientation estimate from the Boomgaard–Weijer robust structure tensor, parameters (see [33]) $m = 0.05$, $s = 5$. **Bottom, left to right:** Filtering of Fig. 6 (f). **(f)** Frobenius median filtering with normalisation and stencil of diameter 9. **(g)** Same with diameter 19. **(h)** Median filtering as in Fig. 6 (i), five iterations. **(i)** Four iterations of median filter with normalisation, stencil diameter 15. **(k)** Boomgaard and Weijer's robust structure tensor, $m = 0.05$, $s = 9$. From [35].

are called *M-estimators* [5]. Like the median and mid-range value, M-estimators give rise to local image filters which are denoted as *M-smoothers* [30, 39].

Special cases of M-estimators include the median for $p = 1$ but also, in the least-squares case $p = 2$, the arithmetic mean. As limit case for $p \to \infty$, the mid-range value as well fits into the framework. M-estimators for $p < 1$ are more difficult to handle since their objective functions are not longer convex – instead, they have local minima at all input values and are strictly concave in the remainder of the real line. Nevertheless, the corresponding M-smoothers display attractive properties for applications since they exceed the median filter in robustness and are able even to *enhance* edges.

Definition 3 *Let $S = (A_1, \ldots, A_n)$ be a tuple of symmetric matrices, and p a positive real number. A symmetric matrix which minimises the convex*

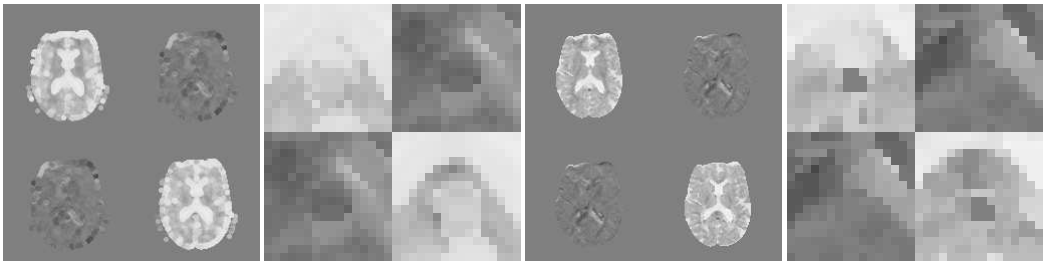

Figure 8: *Left to right:* **(a)** Frobenius mid-range filtering of Fig. 1 (a). **(b)** Corpus callosum detail from (a). **(c)** M-smoothing of Fig. 1 (a) with Frobenius norm and $p = 0.1$. Global minimisers have been searched by grid search. **(d)** Corpus callosum detail from (c).

function

$$E_S(X) := \sum_{i=1}^{n} \|X - A_i\|^p$$

with some matrix norm $\|\cdot\|$ *is a matrix-valued M-estimator for* S*.*

Clearly, the matrix-valued median is recovered for $p = 1$. For $p > 1$, there exists a unique minimiser for E_S because of the strict convexity of that function. As in the scalar case, one faces a more complex situation for $p < 1$ which we discuss here exemplarily for the Frobenius norm. First of all, we note that each of the given matrices A_i is a local minimum of E_S. Second, there can now exist additional minima of the objective function. A remarkable fact is that these keep a minimum distance to the A_i. Since the gradient magnitude $\left|\nabla(\|X - A_j\|^p)\right|$ grows over all limits when S approaches the singularity at A_j, there exists a radius $\varrho = \varrho(p, S)$ which depends on the exponent p and the data set S such that the gradient $\nabla(\|X - A_j\|^p)$ dominates the sum $\sum_{i \neq j} \nabla(\|X - A_j\|^p)$ within the ϱ-neighbourhood of each A_j, thus preventing any additional minimum to come closer than ϱ to any A_j.

Because of the non-convexity, proper selection of the minimum is an important issue for $p < 1$. In Fig. 8 (c,d) we show the result of a grid search for the global minimum. Alternatively one can also think of a down-focussing strategy starting from the unique median, see [37].

5.3 Algorithmic Aspects

For M-smoothers with $1 < p < \infty$, the gradient descent algorithm can be applied similarly as for the median, taking care of the necessary regularisa-

tions in case of the spectral norm. Since the gradient magnitude for $p > 1$ contains information on the distance to the minimum, the step-size control mechanism can be replaced.

In the case of the mid-range filter, the maximum operation applied to the norms $\|X - A_i\|$ induces additional singularities of the objective function that require an analogous regularisation as in the case of the spectral norm. Moreover, the step-size control mechanism is indispensable in this case.

A representation of the mid-range operator by a convex optimisation problem works along the lines described for the spectral median, with the difference that only a scalar auxiliary variable is needed. For details we refer to [6].

6 Matrix-Valued Weighted Median Filters

As demonstrated before, matrix median filtering allows an efficient and edge-preserving denoising. However, fine details which are smaller than the stencil size still experience a degradation even by a single iteration of median filtering. When denoising images which contain only moderate amounts of noise, the preservation of small details can be improved.

We achieve this by using weighted medians. Unweighted scalar median filtering changes each pixel which has not exactly the middle value within its neighbourhood. If instead the central pixel is repeated more than once within the ordered sequence, its value survives even if it is just close to the middle value. Only pixels whose values are close to the extrema within their neighbourhood are treated as outliers and therefore changed.

Definition 4 *Given a matrix tuple $S = (A_1, \ldots, A_n)$, a vector of non-negative weights $w = (w_1, \ldots, w_n)$ and a norm $\|\cdot\|$. The weighted median* $\operatorname{med}(S, w)$ *is defined as the minimiser of*

$$E_{(S,w)}(X) = \sum_{i=1}^{n} w_i \|X - A_i\| \ .$$

Fig. 9 demonstrates denoising of tensor images by weighted matrix median filtering. We use a 3×3 stencil in which the weight of the central pixel is varied. It can be seen that fine structures can be retained that are removed by unweighted median filtering even with small stencils. The admissible weight for the central pixel depends sensitively on the noise level. In our noisy test image, a weight of 2 or slightly above for the central pixel considerably

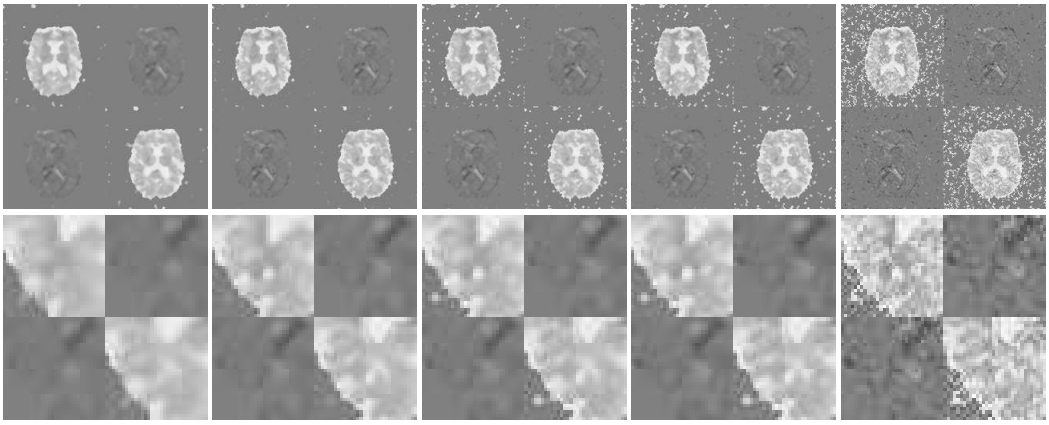

Figure 9: Weighted median filtering. The noisy DTI image has been filtered with a 3×3 stencil, 5 iterations, Frobenius norm. The weight w of the central pixel was varied while that of the remaining pixels was fixed to 1. *Top, left to right:* **(a)** Standard median filtering ($w = 1$). **(b)** $w = 2$. **(c)** $w = 2.2$. **(d)** $w = 2.5$. **(e)** $w = 5$. *Bottom, left to right:* **(f)**–**(k)** Detail views from top row.

enhances structure preservation while higher weights lead directly to stronger noise.

7 Matrix-Valued Quantiles

The possibility to transfer the notion of median, thus a 50%-quantile, to matrix data motivates us to check whether even other quantiles carry over to this type of data. Indeed, the α-quantile $\mathrm{qu}_\alpha(S)$, $0 < \alpha < 1$, of a real data tuple $S = (a_1, \ldots, a_n)$ admits a characterisation by a minimisation property similar to that for the median. Indeed, one has that $\mathrm{qu}_\alpha(S)$ minimises the convex function

$$E_{S,\alpha}(x) := \sum_{i=1}^{n} f_\alpha(x - a_i)$$

where $f_\alpha(z)$ is a piecewise linear but asymmetric (except for $\alpha = 1/2$) function replacing the absolute value,

$$f_\alpha(z) := |z| + (1-2\alpha)z = \begin{cases} (2-2\alpha)\,|z|\,, & z \geq 0, \\ 2\alpha\,|z|\,, & z < 0. \end{cases} \tag{10}$$

In defining matrix-valued quantiles, we require again the properties of scaling and rotational invariance as well as the embedding property for the scalar-

valued quantiles. The way of generalising is mostly analogous to the median case. However, matrix equivalents of f_α has to be used. Formally, such an equivalent is given for

Definition 5 *Let $S = (A_1, \ldots, A_n)$ be a tuple of symmetric matrices, $0 < \alpha < 1$, and let a norm $\|\cdot\|$ be given. The α-quantile $\mathrm{qu}_\alpha(S)$ of S w.r.t. $\|\cdot\|$ is defined as minimiser of the convex function*

$$E_{S,\alpha}(X) := \sum_{i=1}^{n} \|f_\alpha(X - A_i)\|$$

with the function f_α defined in (10).

As usual, the operation of f_α on a symmetric matrix Y is defined by action on the eigenvalues. More explicitly, if $Y = Q\ \mathrm{diag}(\lambda_1, \ldots, \lambda_d)Q^{\mathrm{T}}$ with orthogonal Q, then

$$f_\alpha(Y) := Q\ \mathrm{diag}(f_\alpha(\lambda_1), \ldots, f_\alpha(\lambda_n))Q^{\mathrm{T}}\ .$$

Suitable choices for $\|\cdot\|$ again include nuclear, Frobenius and spectral norm. However, the necessity to apply f_α to the matrices by diagonalisation prevents any generalisation of this quantile definition to other than symmetric square matrices.

7.1 Relation to Matrix Suprema and Infima

Scalar α-quantiles include the minimum and maximum of a tuple of real numbers as special cases for $\alpha = 0$ and $\alpha = 1$. The minimisation characterisation is not fully sufficient in this case since e.g. $E_{S,0}(x)$ is equally minimised by all lower bounds of the given data. Similarly, the characterisation of matrix-valued quantiles from Def. 5 becomes deficient for $\alpha = 0$ ($\alpha = 1$), admitting as minimisers all X for which $X - A_i$ are uniformly negative (positive) semidefinite.

Conditions of this type were also used in [11, 12] where supremum and infimum notions for matrix tuples were defined in order to establish morphological filters. The semi-ordering used in [12] (so-called Loewner ordering) is defined exactly by the definiteness of the $X - A_i$ while the geometrical semi-ordering used in Sec. 4 of [11] is equivalent to the same condition for the difference of squared matrices $X^2 - A_i^2$. By an additional criterion, the supremum (infimum) is then selected among all matrices which are lower

(upper) bounds of the given tuple w.r.t. the semi-ordering. In [11] a lexicographic ordering of eigenvalues plays this role while in [12] the matrix with smallest trace is selected as supremum.

In the quantile framework, the limit process $\alpha \uparrow 1$ ($\alpha \downarrow 0$) lends itself as a way to disambiguate the supremum (infimum). It turns out that for the α-quantiles formed with the nuclear norm, the limit process $\alpha \uparrow 1$ leads to the supremum matrix of minimal trace as in [12]. We remark that a robustness property similar to Lemma 5 ensures the independency of the supremum on those A_i for which $X - A_i$ is positive definite. – The quantiles formed with Frobenius or spectral norm tend for $\alpha \uparrow 1$ to different suprema for which this independency is not guaranteed.

By the following proposition we establish an additional link between matrix suprema and matrix filters defined via minimisation.

Proposition 7 *Let a tuple $S = (A_1, \ldots, A_n)$ of symmetric matrices be given. Provided that the spectral mid-range value $\operatorname{midr}_\infty(S)$ is uniquely determined, we have*

$$\operatorname{midr}_\infty(S) + (\max_{i=1,\ldots,n} (\operatorname{midr}_\infty(S) - A_i))\, I = \sup(S) \tag{11}$$

where E_S is the objective function from (9) with spectral norm and $\sup(S)$ is the matrix X with smallest trace for which all $X - A_i$ are positive semidefinite.

The scalar-valued equivalent of (11) is the equation

$$\operatorname{midr}(a_1, \ldots, a_n) + \max_{i=1,\ldots,n} |\operatorname{midr}(a_1, \ldots, a_n) - a_i| = \max(a_1, \ldots, a_n)\ . \tag{12}$$

An important difference to the scalar case is that (11), with sensible treatment of the non-unique case, could also be used to define and compute matrix suprema since the maximisation in (11) is scalar-valued. A similar "definition" using (12) to introduce scalar maximum via the mid-range value would be circular.

8 Summary

A concept of matrix-valued median filters has been presented which is based on the minimisation of a geometrically motivated objective function. This function measures the sum of distances of a variable matrix to the given data matrices. This median concept is theoretically sound, fits well into the

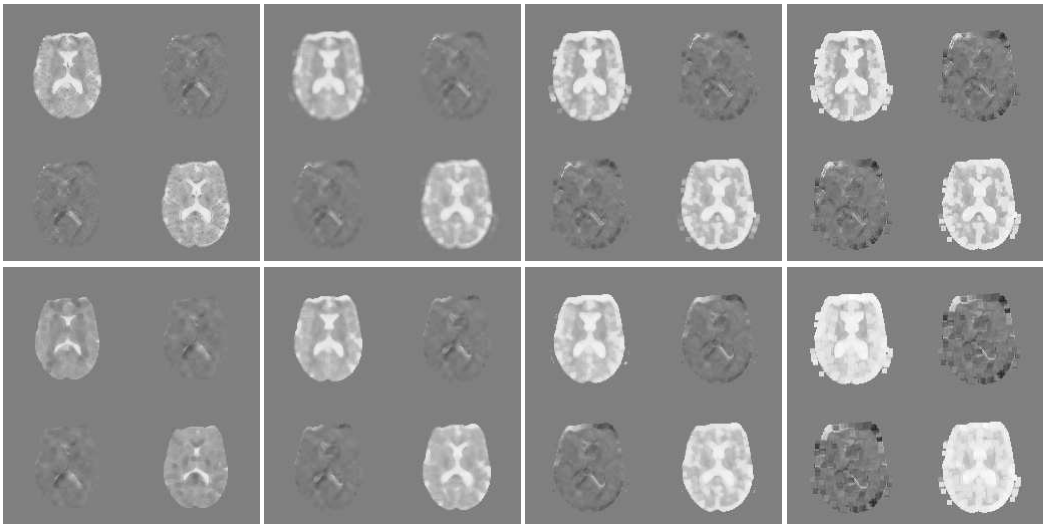

Figure 10: α-quantile filtering of 2-D DTI data with 5×5 stencil. *Top:* With Frobenius norm, *left to right:* **(a)** $\alpha = 0.1$. **(b)** $\alpha = 0.5$ (i.e. median). **(c)** $\alpha = 0.8$. **(d)** $\alpha = 0.99$. *Bottom, left to right:* **(e)**–**(h)** Same as above but with nuclear norm.

context of other multivariate median approaches and possesses favourable mathematical properties.

To compute matrix-valued medians, a convex programming framework has been established which allows an efficient numerical evaluation.

Applied as an image filter to tensor-valued images, the matrix-valued median performs well as a structure-preserving denoising filter. It can also be applied for the smoothing of orientation estimates and thereby gives rise to a variant of an adaptive (robust) structure tensor.

The proposed minimisation idea allows generalisations in a number of directions, yielding a diversified family of further local image filters for matrix-valued images including mid-range filters, M-smoothers, weighted median filters and quantiles. Particularly from mid-range filters (with spectral norm) and quantiles (with nuclear norm) there are close connections to matrix supremum/infimum concepts that have been established as foundation for matrix-valued morphology.

Acknowledgements. We are thankful to Anna Vilanova i Bartrolí and Carola van Pul, both TU Eindhoven, for providing us with the DTI data set and discussing data conversion issues. The fluid dynamics data set is owed to Wolfgang Kollmann (UC Davis) and Gerik Scheuermann (University of

Leipzig). Susanne Biehl and Stephan Zimmer helped us by writing data conversion tools.

References

[1] J. Astola, P. Haavisto, Y. Neuvo. Vector median filters. *Proc. IEEE,* 78(4):678–689, 1990

[2] T. L. Austin. An approximation to the point of minimum aggregate distance. *Metron,* 19:10–21, 1959

[3] V. Barnett. The ordering of multivariate data. *Journal of the Royal Statistical Society A,* 139(3):318–355, 1976

[4] M. Barni, F. Buit, F. Bartolini, V. Cappellini. A quasi-Euclidean norm to speed up vector median filtering. *IEEE Trans. Image Processing,* 9(10):1704–1709, 2000

[5] J. Barral Souto. El modo y otras medias, casos particulares de una misma expresión matemática. *Cuadernos de Trabajo* No. 3, Instituto de Biometria, Universidad Nacional de Buenos Aires, Argentina, 1938

[6] F. Becker. *Matrix-Valued Filters as Convex Programs.* Diploma thesis, CVGPR group, University of Mannheim, 2004

[7] S. Boyd, L. Vandenberghe. *Convex Optimization.* 2004, Cambridge University Press

[8] T. Brox, M. Rousson, R. Deriche, J. Weickert. Unsupervised segmentation incorporating colour, texture, and motion. In N. Petkov, M. A. Westenberg, editors, *Computer Analysis of Images and Patterns, Lecture Notes in Computer Science,* vol. 2756, 353–360, Berlin, 2003, Springer

[9] T. Brox, R. van den Boomgaard, F. Lauze, J. van de Weijer, J. Weickert, P. Mrázek, P. Kornprobst. Adaptive structure tensors and their applications. In J. Weickert, H. Hagen, editors, *Visualization and Processing of Tensor Fields,* Berlin, 2005, Springer, to appear.

[10] T. Brox, J. Weickert. Nonlinear matrix diffusion for optic flow estimation. In L. Van Gool, editor, *Pattern Recognition, Lecture Notes in Computer Science,* vol. 2449, 446–453, Berlin, 2002, Springer

[11] B. Burgeth, M. Welk, C. Feddern, J. Weickert. Morphological operations on matrix-valued images. In T. Pajdla, J. Matas, editors, *Computer Vision - ECCV 2004, Lecture Notes in Computer Science,* vol. 3024, 155–167, Berlin, 2004, Springer

[12] B. Burgeth, M. Welk, C. Feddern, J. Weickert. Mathematical morphology on tensor data using the Loewner ordering. In J. Weickert, H. Hagen, editors, *Visualization and Processing of Tensor Fields,* Berlin, 2005, Springer, to appear

[13] V. Caselles, G. Sapiro, D. H. Chung. Vector median filters, inf-sup operations, and coupled PDE's: Theoretical Connections. *J. Mathematical Imaging and Vision,* 8:109–119, 2000

[14] O. Coulon, D. C. Alexander, S. A. Arridge. A regularization scheme for diffusion tensor magnetic resonance images. In M. F. Insana, R. M. Leahy, editors, *Information Processing in Medical Imaging – IPMI 2001, Lecture Notes in Computer Science,* vol. 2082, 92–105, Berlin, 2001, Springer

[15] E. R. Dougherty, J. Astola, editors. *Nonlinear Filters for Image Processing.* Bellingham, 1999, SPIE Press

[16] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data,* 281–304, Interlaken, Switzerland, June 1987

[17] S. P. Fekete, J. S. B. Mitchell, K. Beurer. On the continuous Fermat–Weber problem. *Operations Research*, 53(1):61–76, 2005..

[18] G. H. Granlund, H. Knutsson. *Signal Processing for Computer Vision.* Dordrecht, 1995, Kluwer

[19] F. Guichard, J. M. Morel. Partial differential equations and image iterative filtering. In: I. S. Duff, G. A. Watson, editors, *The State of the Art in Numerical Analysis, IMA Conference Series (New Series),* no. 63, 525–562, Oxford, 1997, Clarendon Press

[20] K. Hahn, S. Pigarin, B. Pütz. Edge preserving regularization and tracking for diffusion tensor imaging. In W. J. Niessen, M. A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001. Lecture Notes in Computer Science,* vol. 2208, 195–203, Berlin, 2001, Springer

[21] J. Keuchel, C. Schnörr, C. Schellewald, D. Cremers. Binary partitioning, perceptual grouping, and restoration with semidefinite programming. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 25(11):1364–1379, 2003

[22] R. Klette, R. Zamperoni. *Handbook of Image Processing Operators.* New York, 1996, Wiley

[23] A. Koschan, M. Abidi. A comparison of median filter techniques for noise removal in color images. In *Proc. Seventh German Workshop on Color Image Processing,* Erlangen, Germany, 69–79, 2001

[24] N. Megiddo. The weighted Euclidean 1-center problem. *Mathematics of Operations Research,* 8(4):498–504, 1983

[25] G. J. M. Parker, J. A. Schnabel, M. R. Symms, D. J. Werring, G. J. Barker. Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging. *Journal of Magnetic Resonance Imaging,* 11:702–710, 2000

[26] C. Pierpaoli, P. Jezzard, P. J. Basser, A. Barnett, G. Di Chiro. Diffusion tensor MR imaging of the human brain. *Radiology,* 201(3):637–648, Dec. 1996

[27] C. Poupon, J. Mangin, V. Frouin, J. Régis, F. Poupon, M. Pachot-Clouard, D. Le Bihan, I. Bloch. Regularization of MR diffusion tensor maps for tracking brain white matter bundles. In W. M. Wells, A. Colchester, S. Delp, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 1998. Lecture Notes in Computer Science,* vol. 1496, 489–498, Berlin, 1998, Springer

[28] M. Rousson, T. Brox, R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. Technical report no. 4695, Odyssée, INRIA Sophia-Antipolis, France, 2003

[29] D. R. Seymour. Note on Austin's "An approximation to the point of minimum aggregate distance". *Metron,* 28:412–421, 1970

[30] P. L. Torroba, N. L. Cap, H. J. Rabal, W. D. Furlan. Fractional order mean in image processing. *Optical Engineering,* 33(2):528–534, 1994

[31] D. Tschumperlé, R. Deriche. Diffusion tensor regularization with constraints preservation. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Vol. 1, Kauai, HI, Dec. 2001, 948–953, IEEE Computer Society Press

[32] J. W. Tukey. *Exploratory Data Analysis.* Menlo Park, 1971, Addison-Wesley

[33] R. van den Boomgaard, J. van de Weijer. Least squares and robust estimation of local image structure. In L. D. Griffin, M. Lillholm, editors, *Scale Space Methods in Computer Vision, Lecture Notes in Computer Science,*, vol. 2695, 237–254, Berlin, 2003, Springer

[34] J. Weickert, T. Brox. Diffusion and regularization of vector- and matrix-valued images. In M. Z. Nashed, O. Scherzer, editors, *Inverse Problems, Image Analysis, and Medical Imaging,* vol. 313 of *Contemporary Mathematics,* 251–268, Providence, 2002, AMS

[35] M. Welk, F. Becker, C. Schnörr, J. Weickert. Matrix-valued filters as convex programs. In R. Kimmel, N. Sochen, J. Weickert, editors, *Scale Space 2005*, *Lecture Notes in Computer Science*, vol. 3459, Berlin, 2005, Springer, in print

[36] M. Welk, C. Feddern, B. Burgeth, J. Weickert. Median filtering of tensor-valued images. In B. Michaelis, G. Krell, editors, *Pattern Recognition, Lecture Notes in Computer Science*, vol. 2781, 17–24, Berlin, 2003, Springer

[37] M. Welk, C. Feddern, B. Burgeth, J. Weickert. Tensor median filtering and M-smoothing. In J. Weickert, H. Hagen, editors, *Visualization and Processing of Tensor Fields,* Berlin, 2005, Springer, to appear

[38] C. Westin, S. E. Maier, B. Khidhir, P. Everett, F. A. Jolesz, R. Kikinis. Image processing for diffusion tensor magnetic resonance imaging. In C. Taylor, A. Colchester, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 1999, Lecture Notes in Computer Science,* vol. 1679, 441–452, Berlin, 1999, Springer

[39] G. Winkler, V. Aurich, K. Hahn, A. Martin. Noise reduction in images: some recent edge-preserving methods. *Pattern Recognition and Image Analysis,* 9(4):749–766, 1999