On Vector and Matrix Median Computation

S. Setzer*, G. Steidl† and T. Teuber†

September 16, 2010

Abstract

The aim of this paper is to gain more insight into vector and matrix medians and to investigate algorithms to compute them. We prove relations between vector and matrix means and medians, particularly regarding the classical structure tensor. Moreover, we examine matrix medians corresponding to different unitarily invariant matrix norms for the case of symmetric 2×2 matrices which frequently arise in image processing. Our findings are explained and illustrated by numerical examples. To solve the corresponding minimization problems we propose different algorithms. Weiszfeld's algorithm was developed for the ℓ_2 vector median computation and semi-definite programming, in particular second order cone programming, was used for the matrix median computation in the literature. We show that also two splitting methods, the alternating direction method of multipliers and the parallel proximal algorithm, can be applied for generalized vector and matrix median computations. Furthermore, we adapt Weiszfeld's algorithm for our setting. We compare these algorithms numerically and apply them within local median filters.

1 Introduction

While medians of one-dimensional data are well known in image processing, vector or matrix medians are not that common. The reason is that the generalization of the one-dimensional median to higher dimensions is not straightforward and, in contrast to the one-dimensional case, there exists in general no analytical expression. Instead, these medians have to be computed as solutions of certain minimization problems. Literature on this topic can for example be found in [2, 21, 32, 34] and the references therein. Moreover, theoretical connections between vector median filters, morphology and PDEs are given in [7]. There exist various applications of multidimensional medians. Indeed, our interest in this topic comes from a dithering algorithm in [33] where a generalized vector median in $\mathbb{R}^2$ has to be computed in an intermediate step of the algorithm. We use the notation 'generalized' because, in contrast to the usual median, there appears an additional squared ℓ_2 term in the functional to be minimized.

In [42], a concept for matrix median computation was proposed and further extended in [44]. The authors suggest to apply semi-definite programming and second order cone programming (SOCP) to find the sought minimizers. Matrix medians of special rank-1 matrices are

*Saarland University, Dept. of Mathematics and Computer Science, Campus E1.1, 66041 Saarbrücken, Germany

†University of Mannheim, Dept. of Mathematics and Computer Science, A5, 68131 Mannheim, Germany

of interest in connection with the so-called structure tensor of Förstner and Gülch [17] which can be used to approximate directions of constant gray values in images like edges. Recently, SL(2) invariant shape medians were considered in [4].

In this paper, we present a collection of theoretical results on vector and matrix medians. In particular, we investigate matrix medians for different unitarily invariant matrix norms for the case of symmetric 2×2 matrices and show relations between certain vector and matrix problems. The findings are illustrated by numerical examples and compared to the results of the classical structure tensor.
Beyond that, we propose different algorithms to solve the involved minimization problems. As a first approach, we consider the alternating direction methods of multipliers (ADMM) for the generalized vector as well as matrix median computation. We introduce the algorithms for both problems systematically starting with the vector median computation and use for the matrix median computation a relation between the proximum with respect to a unitarily invariant matrix norm and the proximum with respect to its related gauge function. Next, we apply a relative of the ADMM algorithm, namely the parallel proximal algorithm (PPXA) from [9] which appears to be indeed slightly faster than ADMM. Besides, we briefly introduce second order cone programming (SOCP) and Weiszfeld's algorithm for our generalized ℓ_2 vector median problem. Next, we give a comparison of the computation times required by the different algorithms. Although ADMM and PPXA are slower than Weiszfeld's algorithm it should be noted that ADMM and PPXA can be parallelized to a high degree so that we expect a significant speed-up for a parallel implementation, e.g., on a GPU.

The paper is organized as follows: In Section 2, we recall special proximation problems with vector norms and unitarily invariant matrix norms which we need for our median computations. Then, Section 3 deals with vector median computations. After collecting a number of theoretical results we propose the above mentioned algorithms for the ℓ_p vector median computation, namely ADMM and PPXA. Furthermore, SOCP and Weiszfeld's algorithm are applied for the ℓ_2 vector norm. The proof of the convergence of Weiszfeld's algorithm for our slightly more general setting is given in the appendix. In Section 4, we are interested in matrix median computations with respect to different unitarily invariant matrix norms. In particular, we deduce the ADMM algorithm and PPXA for matrix median computations in Subsection 4.1. Then, in Subsection 4.2, we prove several relations for the matrix mean/medians of 2×2 rank-1 matrices $P_i = p_i p_i^{\mathrm{T}}$ and show connections to special vector mean/medians appearing from the vectors p_i. Numerical experiments are reported in Section 5. In Subsection 5.1, ADMM, PPXA and Weiszfeld's algorithm are compared with respect to their computation time. It appears that SOCP implemented in the optimization toolbox MOSEK cannot compare to these algorithms for our settings. Subsection 5.2 illustrates the behavior of the matrix/vector mean and medians within local filters and illuminates the results obtained in Subsection 4.2. The paper ends with conclusions in Section 6.

2 Proximation with Vector and Matrix Norms

In this section, we recall some special proximation problems which we need for our median computations. First, we are interested in

$$\hat{x} = \operatorname*{argmin}_{x\in\mathbb{R}^d}\{\frac{1}{2}\|f-x\|_2^2 + \lambda\|x\|_p\}, \quad 1 \le p \le \infty \tag{1}$$

for a given data vector $f \in \mathbb{R}^d$. The Fenchel conjugates of the ℓ_p-norms in $\mathbb{R}^d$ are given by

$$\|x\|_p^* := \sup_{y\in\mathbb{R}^d}\{\langle y,x\rangle - \|y\|_p\} = \begin{cases} \infty & \text{if } \|x\|_q > 1, \\ 0 & \text{if } \|x\|_q \le 1, \end{cases} \tag{2}$$

where $\frac{1}{p}+\frac{1}{q}=1$ and as usual $p=1$ corresponds to $q=\infty$ and conversely. Now the minimizer can be found by $\hat{x} = f - \hat{v}$, where $\hat{v}$ is the solution of the dual problem

$$\hat{v} = \operatorname*{argmin}_{v\in\mathbb{R}^d}\{\frac{1}{2}\|f-v\|_2^2 + \lambda\|v/\lambda\|_p^*\}.$$

By (2) this can be rewritten as the constrained problem

$$\|f-v\|_2 \to \min_v \quad \text{subject to} \quad \|v\|_q \le \lambda.$$

Hence, $\hat{v} = \Pi_{B_{q,\lambda}}(f)$ is the orthogonal projection of f onto the ℓ_q–ball $B_{q,\lambda}$ with radius λ and center 0 and

$$\hat{x} = f - \Pi_{B_{q,\lambda}}(f).$$

For $p = 1, 2, \infty$ the orthogonal projections onto $B_{q,\lambda}$ are given by

$p = 1$: $\Pi_{B_{\infty,\lambda}}(f) = (P_\lambda(f_i))_{i=1}^d$ and $\hat{x} = (S_\lambda(f_i))_{i=1}^d$, where

$$P_\lambda(f_i) := \begin{cases} f_i & \text{if } |f_i| \le \lambda, \\ \lambda\,\mathrm{sgn}(f_i) & \text{if } |f_i| > \lambda, \end{cases} \quad \text{and} \quad S_\lambda(f_i) = \begin{cases} 0 & \text{if } |f_i| \le \lambda, \\ f_i - \lambda\,\mathrm{sgn}(f_i) & \text{if } |f_i| > \lambda. \end{cases}$$

The function S_λ is known as *soft-shrinkage*, see [12].

$p = 2$:

$$\Pi_{B_{2,\lambda}}(f) = \begin{cases} f & \text{if } \|f\|_2 \le \lambda, \\ \lambda\frac{f}{\|f\|_2} & \text{if } \|f\|_2 > \lambda \end{cases} \quad \text{and} \quad \hat{x} = \begin{cases} 0 & \text{if } \|f\|_2 \le \lambda, \\ f(1-\frac{\lambda}{\|f\|_2}) & \text{if } \|f\|_2 > \lambda. \end{cases}$$

The function which produces $\hat{x}$ is sometimes called *coupled shrinkage*, see [30, 43, 38]. Orthogonal projections onto the ℓ_1-ball were fewer used in the literature so that we give an explanation of the following formula in Remark 2.1.

$p = \infty$:

$$\Pi_{B_{1,\lambda}}(f) = \begin{cases} f & \text{if } \|f\|_1 \le \lambda, \\ (S_\mu(f_i))_{i=1}^d & \text{if } \|f\|_1 > \lambda \end{cases} \quad \text{and} \quad \hat{x} = \begin{cases} 0 & \text{if } \|f\|_1 \le \lambda, \\ f - (S_\mu(f_i))_{i=1}^d & \text{if } \|f\|_1 > \lambda \end{cases}$$

with $\mu := \frac{|f_{\pi(1)}|+\ldots+|f_{\pi(m)}|-\lambda}{m}$, where $|f_{\pi(1)}| \geq \ldots \geq |f_{\pi(d)}| \geq 0$ are the sorted absolute values of the components of f and $m \leq d$ is the largest index such that $|f_{\pi(m)}| > 0$ and

$$\frac{|f_{\pi(1)}| + \ldots + |f_{\pi(m)}| - \lambda}{m} \leq |f_{\pi(m)}|.$$

The computation of $\Pi_{B_{1,\lambda}}(f)$ requires $\mathcal{O}(d \log d)$ operations due to the sorting procedure. Note that

$$f_i - S_\mu(f_i) = \begin{cases} f_i & \text{if} \quad |f_i| \leq \mu, \\ \mu \operatorname{sgn}(f_i) & \text{if} \quad |f_i| > \mu. \end{cases}$$

Remark 2.1. (**Projection onto the ℓ_1-ball**)
We briefly explain orthogonal projections onto the ℓ_1-ball from a geometrical point of view. For $d = 2$ see Fig. 1 for an illustration. Let f have $n \leq d$ nonzero components ordered by π as above and let $\|f\|_1 = |f_{\pi(1)}| + \ldots + |f_{\pi(n)}| > \lambda$. Further, we denote by $f|_s$ the orthogonal projection of f onto $\operatorname{span}\{e_{\pi(j)} : j = 1, \ldots, s\}$, i.e., $f|_s$ has components $f_{\pi(j)}$ for $j \leq s$ and zero components elsewhere. For the sake of brevity, we write $v := \Pi_{B_{1,\lambda}}(f)$ which clearly implies that $\|v\|_1 = \lambda$. Obviously, $\operatorname{sgn}(f_i) = \operatorname{sgn}(v_i)$ or $v_i = 0$ for all $i = 1, \ldots, n$. First, we consider the case $\operatorname{sgn}(v) = \operatorname{sgn}(f)$. Since in this case $\operatorname{sgn}(f)$ is a multiple of the normal vector to $B_{1,\lambda}$ at the point v, cf., Fig. 1, we see that

$$\operatorname{sgn}(v) = \operatorname{sgn}(f) \;\Leftrightarrow\; \exists \mu_n > 0 : \, v = f - \mu_n \operatorname{sgn}(f) \text{ and } |f_{\pi(n)}| > \mu_n. \tag{3}$$

Thus, we can conclude from

$$\lambda = \|v\|_1 = \operatorname{sgn}(v)^T v = \operatorname{sgn}(f)^T (f - \mu_n \operatorname{sgn}(f))$$

that

$$\mu_n = \frac{\|f\|_1 - \lambda}{n}.$$

It follows from (3) *that*

$$\operatorname{sgn}(v) = \operatorname{sgn}(f) \;\Leftrightarrow\; \mu_n = \frac{\|f\|_1 - \lambda}{n} < |f_{\pi(n)}| \tag{4}$$

and in this case we can set $v = f - \mu_n \operatorname{sgn}(f)$. Assume the right-hand side of (4) *does not hold true. Then, it follows that $v_{\pi(n)} = 0$ and we apply the same arguments as above to $f|_{n-1}$. If*

$$\mu_{n-1} := \frac{\|f|_{n-1}\|_1 - \lambda}{n-1} < |f_{\pi(n-1)}|$$

we can set $v = f|_{n-1} - \mu_{n-1} \operatorname{sgn}(f|_{n-1})$. Otherwise we have to continue with $f|_{n-2}$ and so on. Note that $\mu_i > |f_{\pi(i)}|$ implies $\mu_{i-1} > |f_{\pi(i)}|$ so that $\mu_{i-1} > 0$ and $\|f|_{i-1}\|_1 > \lambda$. Moreover, $\mu_{i-1} > |f_{\pi(i)}|$ implies further $\mu_{i-1} > \mu_i$. This leads finally to the above ℓ_1-ball projection. For another deduction of the projection we refer to [10]. *After finishing this paper we became aware of the paper [13] on this topic.*

Next we deal with proximation problems involving matrix norms. For $F \in \mathbb{R}^{m,n}$, we are looking for

$$\hat{X} = \operatorname*{argmin}_{X \in \mathbb{R}^{m,n}} \{\frac{1}{2}\|F - X\|_{\mathcal{F}}^2 + \lambda \|X\|_\bullet\}, \tag{5}$$

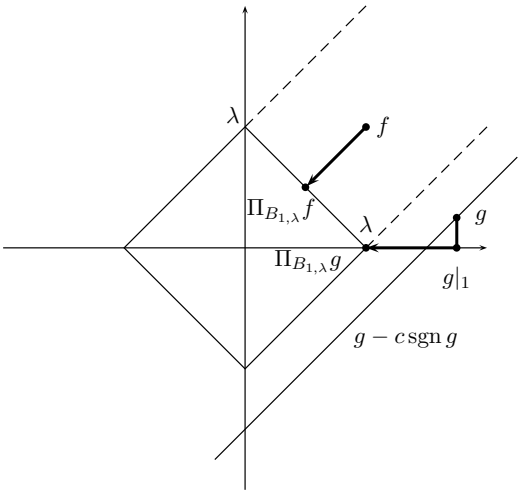


Figure 1: Illustration of orthogonal projections onto $B_{1,\lambda}$ in two dimensions.

where $\|\cdot\|_\bullet$ is a unitarily invariant matrix norm, i.e., $\|X\|_\bullet = \|UXV^T\|_\bullet$ for all unitary matrices $U \in \mathbb{R}^{m,m}, V \in \mathbb{R}^{n,n}$. Von Neumann (1937) has characterized the unitarily invariant matrix norms as those matrix norms which can be written in the form $\|X\|_\bullet = g_\bullet(\sigma(X))$, where $\sigma(X)$ is the vector of singular values of X and g is a symmetric gauge function, see [39]. An analogous result was given by Davis [11] for symmetric matrices, where V^{T} is replaced by U^{T}, the singular values by the eigenvalues and gauge functions by symmetric, convex functions.
We are interested in the *Schatten-p norms* for $p = 1, 2, \infty$ which are defined for $X \in \mathbb{R}^{m,n}$ and $t := \min\{m, n\}$ by

$$\begin{aligned}
\|X\|_* &:= \sum_{i=1}^{t} \sigma_i(X) = g_*(\sigma(X)) = \|\sigma(X)\|_1, && \text{(Nuclear norm)} \\
\|X\|_{\mathcal{F}} &:= (\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2)^{\frac{1}{2}} = (\sum_{i=1}^{t} \sigma_i(X)^2)^{\frac{1}{2}} = g_{\mathcal{F}}(\sigma(X)) = \|\sigma(X)\|_2, && \text{(Frobenius norm)} \\
\|X\|_2 &:= \max_{i=1,\ldots,t} \sigma_i(X) = g_2(\sigma(X)) = \|\sigma(X)\|_\infty, && \text{(Spectral norm).}
\end{aligned}$$

The following proposition describes the solution of (5). Another proof for the special case of the nuclear norm can be found in [6].

Proposition 2.2. *Let $F = U_F \Sigma_F V_F^T$ be the singular value decomposition of F. Then the minimizer of (5) is given by $\hat{X} = U_F \Sigma_{\hat{X}} V_F^T$, where the singular values $\sigma(\hat{X})$ in $\Sigma_{\hat{X}}$ are determined by*

$$\sigma(\hat{X}) := \operatorname*{argmin}_{\sigma \in \mathbb{R}^t} \{\frac{1}{2}\|\sigma(F) - \sigma\|_2^2 + \lambda g_\bullet(\sigma)\} \tag{6}$$

with the symmetric gauge function $g_\bullet$ corresponding to $\|\cdot\|_\bullet$.

Proof: By Fermat's rule we know that the solution $\hat{X}$ of (5) is determined by

$$0 \in \hat{X} - F + \lambda \partial \|\hat{X}\|_\bullet \tag{7}$$

and from [39] that

$$\partial\|X\|_\bullet = \operatorname{conv}\{UDV^{\mathrm{T}} : X = U\Sigma V^{\mathrm{T}},\, D = \operatorname{diag}(d),\, d \in \partial g_\bullet(\sigma(X))\}. \tag{8}$$

We now construct the solution $\hat{X}$ of (7) which is unique since (5) is strictly convex. Let $\hat{\sigma}$ be the unique solution of (6). By Fermat's rule $\hat{\sigma}$ satisfies $0 \in \hat{\sigma} - \sigma(F) + \lambda\partial g_\bullet(\hat{\sigma})$ and consequently there exists $d \in \partial g_\bullet(\hat{\sigma})$ such that

$$0 = U_F\big(\operatorname{diag}(\hat{\sigma}) - \Sigma_F + \lambda\operatorname{diag}(d)\big)V_F^{\mathrm{T}} \quad \Leftrightarrow \quad 0 = U_F \operatorname{diag}(\hat{\sigma})\, V_F^{\mathrm{T}} - F + \lambda U_F \operatorname{diag}(d)\, V_F^{\mathrm{T}}.$$

By (8) we see that $\hat{X} := U_F \operatorname{diag}(\hat{\sigma})\, V_F^{\mathrm{T}}$ is a solution of (7). This completes the proof. □

For our special matrix norms this means that the minimizer of (5) is given by $\hat{X} = U_F\Sigma_{\hat{X}}V_F^{\mathrm{T}}$ with

- $\bullet = *:$ $\quad \sigma_{\hat{X}} := \sigma_F - \Pi_{B_{\infty,\lambda}}(\sigma_F)$,
- $\bullet = \mathcal{F}:$ $\quad \sigma_{\hat{X}} := \sigma_F - \Pi_{B_{2,\lambda}}(\sigma_F)$,
- $\bullet = 2:$ $\quad \sigma_{\hat{X}} := \sigma_F - \Pi_{B_{1,\lambda}}(\sigma_F)$.

In case of the Frobenius norm it is straightforward that the minimizer can also be obtained without singular value decomposition of F by

$$\hat{X} := \begin{cases} 0 & \text{if} \quad \|F\|_{\mathcal{F}} \le \lambda, \\ (1 - \frac{\lambda}{\|F\|_{\mathcal{F}}})F & \text{if} \quad \|F\|_{\mathcal{F}} > \lambda. \end{cases}$$

3 Vector Median Computation

For given pairwise different points $P := \{p_i \in \mathbb{R}^d : i = 1, ..., M\}$, positive weights w_i, $i = 1, ..., M$, $f \in \mathbb{R}^{\tilde{d}}$ and a linear operator $K \in \mathbb{R}^{\tilde{d},d}$, we are interested in minimizing

$$E_v(x) := \frac{\lambda}{2}\|Kx - f\|_2^2 + \sum_{i=1}^{M} w_i\|x - p_i\|_p, \quad \lambda \ge 0 \tag{9}$$

for $p = 1, 2, \infty$.

3.1 Theoretical Results

Let us briefly recall the one-dimensional setting $d = 1$.

Proposition 3.1. (Analytical solution in 1D)
Let $p_1 \le \ldots \le p_M$ be ordered real numbers and w_i, $i = 1, \ldots, M$ positive weights. Then the minimizer $\hat{x}$ of

$$\frac{\lambda}{2}(x - f)^2 + \sum_{i=1}^{M} w_i|x - p_i|$$

is given for $\lambda = 0$ by

$$\hat{x} = \begin{cases} p_k & \text{if} \quad 0 \in \big(\sum\limits_{i=k+1}^{M} w_i - \sum\limits_{i=1}^{k} w_i, \sum\limits_{i=k}^{M} w_i - \sum\limits_{i=1}^{k-1} w_i\big), \\ [p_k, p_{k+1}] & \text{if} \quad \sum\limits_{i=1}^{k} w_i = \sum\limits_{i=k+1}^{M} w_i \end{cases}$$

and for $\lambda > 0$ *by*

$$\hat{x} = \text{median}(p_1, \dots, p_M, a_0, \dots, a_M), \tag{10}$$

where

$$a_0 := f + \frac{1}{\lambda}\sum_{i=1}^{M} w_i, \quad a_k := f + \frac{1}{\lambda}\left(\sum_{i=k+1}^{M} w_i - \sum_{i=1}^{k} w_i\right),\ k = 1, \dots, M.$$

The proof can be found in [27]. Note that $a_0 \geq \ldots \geq a_M$ and that the above median (10) can be computed for ordered p_i with $\mathcal{O}(M)$ operations, see [27]. Of course, the ordering of the p_i requires $\mathcal{O}(M \log(M))$ operations.

Let us turn to higher dimensions $d \geq 2$. In some special cases, the computation can be reduced to the one-dimensional setting: For $p = 1$ and $K = I$, the functional (9) can be minimized componentwise using Proposition 3.1, i.e.,

$$\hat{x}_k = \operatorname*{argmin}_{x_k \in \mathbb{R}} \left\{ \frac{\lambda}{2}(x_k - f_k)^2 + \sum_{i=1}^{M} w_i |x_k - p_{i,k}| \right\}, \qquad k = 1, \dots, d. \tag{11}$$

Let $p = \infty$, $K = I$ and $d = 2$. For $x \in \mathbb{R}^2$ we have that $\|x\|_\infty = \frac{1}{\sqrt{2}}\|y\|_1$, where $y = Qx$ and $Q = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Note that, conversely, $x = Qy$. Then we obtain

$$\frac{\lambda}{2}\|x - f\|_2^2 + \sum_{i=1}^{M} w_i \|x - p_i\|_\infty \;=\; \frac{\lambda}{2}\|y - Qf\|_2^2 + \sum_{i=1}^{M} \frac{w_i}{\sqrt{2}} \|y - Qp_i\|_1.$$

Now the minimizer $\hat{y}$ of the functional on the right-hand side can be computed componentwise using Proposition 3.1.

For general p the following property was observed in [44]. We add the simple proof.

Proposition 3.2. *Let* $\lambda = 0$. *Then,* $\hat{x} \in \operatorname{argmin} E_v$ *implies that* $\hat{x} \in \operatorname{argmin} \tilde{E}_v$, *where* $\tilde{E}_v$ *denotes the functional* (9) *with respect to the points* $\tilde{p}_i := \hat{x} + \alpha_i(p_i - \hat{x})$, $\alpha_i > 0$, $i = 1, \dots, M$.

Proof: From Fermat's rule we obtain

$$\hat{x} \in \operatorname{argmin} E_v \quad \Leftrightarrow \quad 0 \in \partial E_v(\hat{x}) = \sum_{i=1}^{M} w_i \partial \|\cdot\|(\hat{x} - p_i).$$

Since we have for arbitrary norms in $\mathbb{R}^d$ (more generally for homogeneous functions) and $\alpha > 0$ that

$$\partial \|\cdot\|(\alpha x_0) = \{z : \|\alpha x_0\| + \langle z, y - \alpha x_0 \rangle \leq \|y\| \ \forall y \in \mathbb{R}^d\} = \partial \|\cdot\|(x_0)$$

this yields

$$0 \in \sum_{i=1}^{M} w_i \partial \|\cdot\|(\alpha_i(\hat{x} - p_i)) = \partial \tilde{E}_v(\hat{x}).$$

Using Fermat's rule again we see that $\hat{x}$ is also a minimizer of $\tilde{E}_v$. □

Results for ℓ_2 Median Computation

The case $p = 2$, i.e.,

$$E_v(x) := \frac{\lambda}{2}\|Kx - f\|_2^2 + \sum_{i=1}^{M} w_i \|x - p_i\|_2, \quad \lambda \geq 0 \tag{12}$$

is of special interest.

Remark 3.3. *For $\lambda = 0$ and weights $w_i = 1$, $i = 1, \ldots, M$, problem* (12) *is known as* spatial median problem*. It has several names including Steiner problem, generalized Weber problem or general Fermat problem. In two dimensions ($d = 2$) the explicit solutions for three or four input points are well-known: For $M = 3$ points spanning a triangle with angles smaller than $120°$ the median is the Steiner point from which the given points can be seen under an angle of $120°$. If one angle of the triangle is larger than $120°$, the median is just this point. For $M = 4$ points spanning a convex quadrangle the median is the intersection of its diagonals. If their convex hull is a triangle, the median is the inner point.*
In general the median does not necessarily coincide with one of the vectors p_i. In contrast, the coincidence with one of the given vectors is a key requirement of diverse modified spatial vector median definitions as the '(extended) vector median filter' in [2]*, the '(generalized) vector directional filter' and the 'directional distance filter' in* [21, 34].

The functional (12) has a unique minimizer if one of the following assumptions is fulfilled:

- the points p_i are not aligned,
- the points are aligned and $\sum_{i=1}^{k} w_i \neq \sum_{i=k+1}^{M} w_i$, $k = 1, \ldots, M-1$,
- $\lambda \neq 0$ and K is invertible.

For convenience we prove the first assertion for $\lambda = 0$. If the points lie on a line, the assertion follows from Proposition 3.1.

Proposition 3.4. *If the points p_i, $i = 1, ..., M$ are not collinear, then functional* (12) *with $\lambda = 0$ is strictly convex.*

Proof: For $z_1 \neq z_2$ and $z := \lambda z_1 + (1-\lambda) z_2$, $\lambda \in (0, 1)$, we have

$$\begin{aligned} E_v(z) &= \sum_{i=1}^{M} w_i \|\lambda(z_1 - p_i) + (1-\lambda)(z_2 - p_i)\|_2 \\ &\leq \sum_{i=1}^{M} \lambda w_i \|z_1 - p_i\|_2 + (1-\lambda) w_i \|z_2 - p_i\|_2 \\ &= \lambda E_v(z_1) + (1-\lambda) E_v(z_2). \end{aligned}$$

Equality holds true if and only if $z_1 - p_i$ and $z_2 - p_i$ are parallel vectors for all $i = 1, ..., M$, i.e., if all points p_i lie on a line. Otherwise, the functional is strictly convex. □

By the separation theorem for convex sets one can prove the following proposition, see [44].

Proposition 3.5. *For $\lambda = 0$, any minimizer of* (12) *is in the convex hull of $\{p_1, \ldots, p_M\}$. For $\lambda > 0$ and an orthogonal matrix K any minimizer of* (12) *is in the convex hull of $\{K^T f, p_1, \ldots, p_M\}$.*

3.2 Algorithms

A general algorithm to solve (9) is the *alternating direction method of multipliers* (ADMM) which we will introduce in the following. The ADMM can be used to solve constrained minimization problems of the form

$$\min_{x\in\mathbb{R}^d, v\in\mathbb{R}^D} \{G_1(x) + G_2(v)\} \quad \text{subject to} \quad Ax = v, \tag{13}$$

where G_i, $i = 1, 2$ are proper closed convex functions and $A \in \mathbb{R}^{D,d}$ is a linear operator, as follows:

ADMM Algorithm:
Initialization: $v^{(0)} \in \mathbb{R}^D$, $b^{(0)} \in \mathbb{R}^D$ and $\gamma > 0$.
For $r = 0, 1, ...$ repeat until a convergence criterion is reached

$$\begin{aligned} x^{(r+1)} &= \operatorname*{argmin}_{x\in\mathbb{R}^d} \big\{G_1(x) + \frac{1}{2\gamma}\|b^{(r)} + Ax - v^{(r)}\|_2^2\big\}, \\ v^{(r+1)} &= \operatorname*{argmin}_{v\in\mathbb{R}^D} \big\{G_2(v) + \frac{1}{2\gamma}\|b^{(r)} + Ax^{(r+1)} - v\|_2^2\big\}, \\ b^{(r+1)} &= b^{(r)} + Ax^{(r+1)} - v^{(r+1)}. \end{aligned}$$

For the above problem, the ADMM coincides with the alternating split Bregman method [20] and with the Douglas-Rachford splitting method applied to the dual problem of (13), see [14, 15, 18, 31]. The convergence of the algorithm is ensured by the following theorem, see, e.g., [31].

Theorem 3.6. *Let G_i, $i = 1, 2$ be proper closed convex functions and $A \in \mathbb{R}^{D,d}$ a linear operator. Then, for any starting values and any $\gamma > 0$, the ADMM sequences $\{b^{(r)}\}$ and $\{v^{(r)}\}$ converge to some $\hat{b}$ and $\hat{v}$, respectively. The sequence $\{x^{(r)}\}$ converges to a solution of (13) if one of the following conditions is fulfilled:*
i) The problem has a unique solution.
ii) The problem $\operatorname{argmin}_{x\in\mathbb{R}^d}\{G_1(x) + \frac{1}{\gamma}\|\hat{b} + Ax - \hat{v}\|_2^2\}$ has a unique solution.

Note that $\frac{1}{\gamma}\hat{b}$ is a solution of the dual problem $\operatorname*{argmin}_{p\in\mathbb{R}^D}\{G_1^*(-A^*p) + G_2^*(p)\}$.

To solve our generalized median problem (9), we apply the algorithm to the equivalent constrained problem

$$\min_{x\in\mathbb{R}^d, v\in\mathbb{R}^{Md}} \left\{\frac{\lambda}{2}\|Kx - f\|_2^2 + \sum_{i=1}^{M} w_i\|v_i - p_i\|_p\right\} \quad \text{subject to} \quad x = v_i,\ i = 1, \ldots, M,$$

i.e., we set $G_1(x) := \frac{\lambda}{2}\|Kx - f\|^2$, $G_2(x) := \sum_{i=1}^{M} w_i\|v_i - p_i\|_p$ and $A := 1_M \otimes I$, where 1_M denotes the vector consisting of M entries 1 and $A \otimes B$ represents the *Kronecker product* of

A and B. Then the ADMM steps read as follows:

$$\begin{aligned} x^{(r+1)} &= \operatorname*{argmin}_{x\in\mathbb{R}^d} \{\frac{\lambda}{2}\|Kx-f\|_2^2 + \frac{1}{2\gamma}\|b^{(r)}+Ax-v^{(r)}\|_2^2\}, \\ v^{(r+1)} &= \operatorname*{argmin}_{v\in\mathbb{R}^D} \{\sum_{i=1}^{M} w_i\|v_i-p_i\|_p + \frac{1}{2\gamma}\|b^{(r)}+Ax^{(r+1)}-v\|_2^2\}, \\ b^{(r+1)} &= b^{(r)} + Ax^{(r+1)} - v^{(r+1)}. \end{aligned}$$

The minimizer in the first step is the solution of

$$(\lambda\gamma K^{\mathrm{T}}K + MI)x^{(r+1)} = \lambda\gamma K^{\mathrm{T}}f + \sum_{i=1}^{M}(v_i^{(r)} - b_i^{(r)}). \tag{14}$$

Note that the matrix $\lambda\gamma K^{\mathrm{T}}K + MI$ is symmetric and positive definite. The minimizer in the second step can be computed componentwise by solving for $i = 1, \ldots, M$ the proximation problems

$$y_i^{(r)} := \operatorname*{argmin}_{y_i}\{w_i\|y_i\|_p + \frac{1}{2\gamma}\|s_i - y_i\|_2^2\}, \quad \text{with} \quad s_i := b_i^{(r)} + x^{(r+1)} - p_i$$

via orthogonal projection of s_i onto $B_{q,\gamma w_i}$ as described in Section 2 and setting $v_i^{(r)} := y_i^{(r)} + p_i$. Note that in some applications, especially if (14) is hard to solve, the *primal-dual hybrid gradient method* (PDHG) might be useful, cf., [8, 16, 45] and the references therein. It is a version of ADMM which usually needs more iterations but works without the matrix inversion in (14).

As we have mentioned above, ADMM can be interpreted as a Douglas-Rachford splitting algorithm. Alternatively, we can also apply another derivative of the Douglas-Rachford splitting algorithm, the *parallel proximal algorithm* (PPXA) proposed in [9]. To deduce this algorithm, we consider the general problem

$$\operatorname*{argmin}_{x\in\mathbb{R}^d}\{g_1(x) + \cdots + g_N(x)\}, \tag{15}$$

where g_i, $i = 1, \ldots, N$ are proper, closed and convex functions. We write $D := dN$. Then, PPXA has the following form:

PPXA Algorithm:

Initialization: $y^{(0)} \in \mathbb{R}^D$, $\sigma_i > 0$ with $\sum_{i=1}^N \sigma_i = 1$, $x^{(0)} = \sum_{i=1}^N \sigma_i y_i^{(0)}$, $\gamma > 0$ and $\mu \in (0, 2)$. For $r = 0, 1, \ldots$ repeat until a convergence criterion is reached

$$\begin{aligned} v_i^{(r+1)} &= \operatorname*{argmin}_{v_i\in\mathbb{R}^D}\{\frac{\gamma}{\sigma_i}g_i(v_i) + \frac{1}{2}\|v_i - y_i^{(r)}\|_2^2\}, \quad i = 1, \ldots, N, \\ z^{(r+1)} &= \sum_{i=1}^{N}\sigma_i v_i^{(r+1)}, \\ y_i^{(r+1)} &= y_i^{(r)} + \mu(2z^{(r+1)} - x^{(r)} - v_i^{(r+1)}), \quad i = 1, \ldots, N, \\ x^{(r+1)} &= x^{(r)} + \mu(z^{(r+1)} - x^{(r)}). \end{aligned} \tag{16}$$

The following convergence result was proved in [9].

Theorem 3.7. *Let g_i, $i = 1, \dots, N$, be proper closed convex functions such that a solution of* (15) *exists. Furthermore, suppose that $\sigma_i > 0$ with $\sum_{i=1}^{N} \sigma_i = 1$, $\gamma > 0$ and $\mu \in (0, 1)$. Then, for any starting value $y^{(0)} \in \mathbb{R}^D$, the sequence $\{x^{(r)}\}$ generated by PPXA converges to a solution of problem* (15).

For our vector median problem we set $g_i(x) := w_i \|x - p_i\|_p$, $i = 1, \dots, M$ and $g_{M+1}(x) = \frac{\lambda}{2}\|Kx - f\|_2^2$. Then the subproblems of (16) are again proximal problems which can be solved similarly as those in the second step of the ADMM algorithm.

Further Algorithms for ℓ_2 Median Computation

In the case $p = 2$ at least two other methods were applied to minimize (12) with $\lambda = 0$, namely

- Second Order Cone Programming (SOCP) [3, 41, 44],
- Weiszfeld's algorithm [22, 23, 25, 26, 40].

In the following, we briefly explain these methods for our slightly more general setting with $\lambda > 0$. In Subsection 5.1 we will compare the methods numerically.

SOCP is a special case of semi-definite programming which can be implemented efficiently due to special constraints. More precisely, SOCP [28] amounts to minimize a linear objective function subject to the constraints that several affine functions of the variables have to lie in a *second-order cone* $\mathcal{C}^{n+1} \subset \mathbb{R}^{n+1}$ defined by the convex set

$$\mathcal{C}^{n+1} = \left\{ \begin{pmatrix} x \\ \bar{x}_{n+1} \end{pmatrix} = (x_1, \dots, x_n, \bar{x}_{n+1})^{\mathrm{T}} : \|x\|_2 \le \bar{x}_{n+1} \right\}.$$

With this notation, the general form of a SOCP is given by

$$\inf_{x \in \mathbb{R}^n} a^{\mathrm{T}} x \quad \text{subject to } \begin{pmatrix} A_i x + b_i \\ c_i^{\mathrm{T}} x + d_i \end{pmatrix} \in \mathcal{C}^{n+1}, \ i = 1, \dots, r. \tag{17}$$

Alternatively, one can also use the rotated version of the standard cone:

$$\mathcal{K}^{n+2} := \left\{ \left(x, \bar{x}_{n+1}, \bar{x}_{n+2}\right)^{\mathrm{T}} \in \mathbb{R}^{n+2} : \|x\|_2^2 \le 2\, \bar{x}_{n+1} \bar{x}_{n+2} \right\},$$

which allows us to incorporate quadratic constraints. There are efficient, large scale solvers for (17) available [1, 28, 29]. Generally preconditioned Newton steps are applied within a primal-dual interior point program. In our numerical examples we use the software package MOSEK for SOCP computations.

SOCP for ℓ_2 median computation (12) reads:

$$\begin{aligned} \min_{s \in \mathbb{R}, t \in \mathbb{R}^M} \quad & \left\langle \begin{pmatrix} \lambda \\ w \end{pmatrix}, \begin{pmatrix} s \\ t \end{pmatrix} \right\rangle \\ \text{subject to} \quad & Kx - f = \bar{x}, \ x - p_i = y_i, \ i = 1, \dots, M, \\ & \|\bar{x}\|_2^2 \le 2s, \quad \|y_i\|_2 \le t_i \quad i = 1, \dots, M. \end{aligned}$$

Finally, let us explain Weiszfeld's algorithm. If $\hat{x} \notin P$, then E_v is differentiable at $\hat{x}$ and

$$\begin{aligned} 0 &= \nabla E_v(\hat{x}) = \lambda K^{\mathrm{T}} K\hat{x} - \lambda K^{\mathrm{T}} f + \sum_{i=1}^{M} w_i \frac{\hat{x} - p_i}{\|\hat{x} - p_i\|_2}, \\ \hat{x} &= (\lambda K^{\mathrm{T}} K + \sum_{i=1}^{M} w_i \frac{1}{\|\hat{x} - p_i\|_2} I)^{-1} \left(K^{\mathrm{T}} f + \sum_{i=1}^{M} w_i \frac{p_i}{\|\hat{x} - p_i\|_2} \right). \end{aligned} \tag{18}$$

In this case, Weiszfeld's algorithm can be considered in the more general context of quadratic (Taylor) approximation of a twice differentiable functional, see [37].
By Fermat's rule $\hat{x} \in \operatorname{argmin}_x E_v(x)$ if and only if $0 \in \partial E_v(\hat{x})$, i.e., if and only if either $\hat{x} \notin P$ and

$$0 = \lambda K^{\mathrm{T}}(K\hat{x} - f) + \sum_{i=1}^{M} w_i \frac{\hat{x} - p_i}{\|\hat{x} - p_i\|_2} \tag{19}$$

or $\hat{x} = p_k \in P$ and

$$0 \in \lambda K^{\mathrm{T}}(Kp_k - f) + \sum_{\substack{i=1 \\ i \neq k}}^{M} w_i \frac{p_k - p_i}{\|p_k - p_i\|_2} + B_{2,w_k}(0). \tag{20}$$

The last inclusion is fulfilled if and only if $\|G_k\|_2 \leq w_k$, where

$$G_k := \lambda K^{\mathrm{T}}(Kp_k - f) + \sum_{\substack{i=1 \\ i \neq k}}^{M} w_i \frac{p_k - p_i}{\|p_k - p_i\|_2}.$$

Suppose that $M \geq 2$ or $\lambda \neq 0$ and K is invertible. We define

$$G(x) := \begin{cases} \lambda K^{\mathrm{T}}(Kx - f) + \sum_{i=1}^{M} w_i \frac{x - p_i}{\|x - p_i\|_2} & \text{if } x \notin P, \\ 0 & \text{if } x = p_k \in P,\ \|G_k\|_2 \leq w_k, \\ G_k - w_k \frac{G_k}{\|G_k\|_2} & \text{if } x = p_k \in P,\ \|G_k\|_2 > w_k \end{cases}$$

and

$$S(x) := \begin{cases} \lambda K^{\mathrm{T}} K + \sum_{i=1}^{M} \frac{w_i}{\|x - p_i\|_2} I & \text{if } x \notin P, \\ \lambda K^{\mathrm{T}} K + \sum_{\substack{i=1 \\ i \neq k}}^{M} \frac{w_i}{\|p_k - p_i\|_2} I & \text{if } x = p_k \in P. \end{cases}$$

Note that $S(x)$ is symmetric and positive definite. Now the generalized Weiszfeld algorithm is defined as follows:

Weiszfeld's Algorithm:
Initialization: $x^{(0)} \in \mathbb{R}^d, \quad c_r \in [1, 2)$.
For $r = 0, 1, ...$ repeat until a convergence criterion is reached

$$x^{(r+1)} = T_{c_r}(x^{(r)}) := x^{(r)} - c_r S(x^{(r)})^{-1} G(x^{(r)}).$$

Note that in the case $c_r = 1$ and $x^{(r)} \notin P$ for all $r = 0, 1, \ldots$ Weiszfeld's algorithm is just the Picard iteration of (18).

Remark 3.8. *The above algorithm is based on the assumption that P contains only distinct points p_i. In case that some points are equal the algorithm can be adapted as follows: Suppose that $\lambda \neq 0$ and K is invertible or there exist at least two points p_i, p_j such that $p_i \neq p_j$. Let $I_k := \{i \in \{1, \ldots, M\} : p_k = p_i\}$. Then, we simply set*

$$G_k := \lambda K^T(Kp_k - f) + \sum_{\substack{i=1 \\ i \notin I_k}}^{M} w_i \frac{p_k - p_i}{\|p_k - p_i\|_2},$$

$$G(x) := \begin{cases} \lambda K^T(Kx - f) + \sum_{i=1}^{M} w_i \frac{x - p_i}{\|x - p_i\|_2} & \text{if } x \notin P, \\ 0 & \text{if } x = p_k \in P,\ \|G_k\|_2 \leq \sum_{i \in I_k} w_i, \\ G_k - \sum_{i \in I_k} w_i \frac{G_k}{\|G_k\|_2} & \text{if } x = p_k \in P,\ \|G_k\|_2 > \sum_{i \in I_k} w_i \end{cases}$$

and

$$S(x) := \begin{cases} \lambda K^T K + \sum_{i=1}^{M} \frac{w_i}{\|x - p_i\|_2} I & \text{if } x \notin P, \\ \lambda K^T K + \sum_{\substack{i=1 \\ i \notin I_k}}^{M} \frac{w_i}{\|p_k - p_i\|_2} I & \text{if } x = p_k \in P. \end{cases}$$

Remark 3.9. *Weiszfeld's algorithm with $c_r = 1$ and starting vector $x^{(0)} = 0$ ($K = I$, $M = 1$ and $p_1 = 0$) applied to the proximation problem (1) with $p = 2$ terminates after one step which computes the coupled shrinkage of f.*

Theorem 3.10. *Let $K = I$. Assume that E_v has a unique minimizer $\hat{x}$. For $1 \leq c_r < 2$, the sequence $\{x^{(r)}\}_{r \in \mathbb{N}}$ generated by the Weiszfeld algorithm converges for any $x^{(0)}$ to the minimizer of E_v.*

The proof which follows mainly the lines of [26], where the case $\lambda = 0$ was considered, is given in the appendix. Indeed local linear convergence of the algorithm can be shown.

4 Matrix Median Computation

For given pairwise different matrices $P_i \in \mathbb{R}^{m,n}, i = 1, ..., M$, positive weights w_i, $i = 1, ..., M$ and a matrix $F \in \mathbb{R}^{m,n}$ we are interested in minimizing

$$E_m(X) := \frac{\lambda}{2}\|X - F\|_{\mathcal{F}}^2 + \sum_{i=1}^{M} w_i \|X - P_i\|_\bullet, \quad \lambda \geq 0, \tag{21}$$

where $\bullet \in \{*, \mathcal{F}, 2\}$.
The Frobenius norm plays a special role here: If $\bullet = \mathcal{F}$, a columnwise reordering of the matrix components into corresponding vectors leads to the vector median problem (12). Then we know by Proposition 3.5 that $\hat{X} = \operatorname{argmin}_X E_m(X)$ is in the convex hull of $\{F, P_1, \ldots, P_M\}$. Therefore, if $F, P_1, \ldots, P_M$ are symmetric, positive semi-definite matrices, then noting that

the symmetric, positive semi-definite matrices form a convex cone, $\hat{X}$ is also symmetric and positive semi-definite. Furthermore, we can use the algorithms from Subsection 3.2 to compute the minimizer of (21) with respect to the Frobenius norm.
In the general case we can again apply the ADMM algorithm and its 'relatives' as described in the following subsection.

4.1 Algorithms

An ADMM can be deduced using the same ideas as for the generalized vector median problem (21) which leads to the following algorithm:

ADMM for Matrix Median Computation:
Initialization: $V^{(0)} \in \mathbb{R}^{Mm,n}$, $B^{(0)} \in \mathbb{R}^{Mm,n}$ and $\gamma > 0$.
For $r = 0, 1, ...$ repeat until a convergence criterion is reached

$$\begin{aligned} X^{(r+1)} &= \operatorname*{argmin}_{X\in\mathbb{R}^{m,n}} \{\frac{\lambda}{2}\|X - F\|_{\mathcal{F}}^2 + \frac{1}{2\gamma}\|B^{(r)} + 1_M \otimes X - V^{(r)}\|_{\mathcal{F}}^2\}, \\ V^{(r+1)} &= \operatorname*{argmin}_{V\in\mathbb{R}^{Mm,n}} \{\sum_{i=1}^{M} w_i\|V_i - P_i\|_\bullet + \frac{1}{2\gamma}\|B^{(r)} + 1_M \otimes X^{(r+1)} - V\|_{\mathcal{F}}^2\}, \\ B^{(r+1)} &= B^{(r)} + 1_M \otimes X^{(r+1)} - V^{(r+1)}. \end{aligned}$$

Here, $V = (V_1^{\mathrm{T}}, \ldots, V_n^{\mathrm{T}})^{\mathrm{T}}$ with $V_i \in \mathbb{R}^{m,n}$.
Again, the second step can be computed separately for each V_i, resp., $Y_i = V_i - P_i$, where the proximation problems

$$Y_i^{(r)} := \operatorname*{argmin}_{Y_i\in\mathbb{R}^{m,n}}\{w_i\|Y_i\|_\bullet + \frac{1}{2\gamma}\|S_i - Y_i\|_{\mathcal{F}}^2\}, \quad \text{with} \quad S_i := B_i^{(r)} + X^{(r+1)} - P_i$$

can be solved as shown in Section 2. Similarly, we obtain a corresponding PPXA method for the matrix median problem:

PPXA for Matrix Median Computation:
Initialization: $Y^{(0)} \in \mathbb{R}^{Mm,n}$, $\sigma_i > 0$ with $\sum_{i=1}^{M+1}\sigma_i = 1$, $X^{(0)} = \sum_{i=1}^{M+1}\sigma_i Y_i^{(0)}$, $\gamma > 0$ and $\mu \in (0, 2)$.
For $r = 0, 1, ...$ repeat until a convergence criterion is reached

$$\begin{aligned} V_i^{(r+1)} &= \operatorname*{argmin}_{V_i\in\mathbb{R}^{m,n}}\{\frac{\gamma}{\sigma_i}w_i\|V_i - P_i\|_\bullet + \frac{1}{2}\|V_i - Y_i^{(r)}\|_{\mathcal{F}}^2\}, \quad i = 1, \ldots, M, \\ V_{M+1}^{(r+1)} &= (1 + \frac{\gamma\lambda}{\sigma_{M+1}})^{-1}(F + Y_{M+1}^{(r)}), \\ Z^{(r+1)} &= \sum_{i=1}^{M+1}\sigma_i V_i^{(r+1)}, \\ Y_i^{(r+1)} &= Y_i^{(r)} + \mu(2Z^{(r+1)} - X^{(r)} - V_i^{(r+1)}), \quad i = 1, \ldots, M+1, \\ X^{(r+1)} &= X^{(r)} + \mu(Z^{(r+1)} - X^{(r)}). \end{aligned}$$

If $F, P_1, \ldots, P_M$ are symmetric with $m = n$ and if we start with symmetric matrices, then the both ADMM and PPXA produce in each step again a symmetric matrix $X^{(r)}$ so that we

end up with a symmetric matrix as the minimizer of (21). In the following, we denote by $\mathrm{Sym}_n(\mathbb{R})$ the *space of symmetric $n \times n$ matrices with real components.*
Finally we note that it is also possible to solve (21) by semi-definite programming as it was proposed in [3, 41, 44].

4.2 Median Computation for Symmetric 2×2 Matrices

The weighted medians of symmetric 2×2 matrices have interesting properties which we will consider in this subsection. In image processing medians of this type are of interest, since they are closely related to structure tensors of images. To motivate our interest in this topic let us briefly recall the definition of the classical structure tensor of Förstner and Gülch [17].

Remark 4.1. (**Classical structure tensor**)
Assume that a given image $u : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^2$, has nearly constant values along a single direction v *with $\|\mathrm{v}\|_2 = 1$ in a neighborhood $B_{\tilde{\rho}}(x_0) \subset \Omega$ of x_0. Then $0 \approx \frac{\partial}{\partial \mathrm{v}} u(x) = \mathrm{v}^T \nabla u(x)$ for $x \in B_{\tilde{\rho}}(x_0)$ and we obtain for a nonnegative weight function $w : \Omega \to \mathbb{R}$ with support in $B_{\tilde{\rho}}(0)$ that*

$$0 \approx \int_\Omega w(y - x_0)(\mathrm{v}^T \nabla u(y))^2 \, dy = \mathrm{v}^T \underbrace{\int_\Omega w(y - x_0) \nabla u(y) \nabla u(y)^T \, dy}_{\mathcal{J}(x_0)} \mathrm{v}.$$

Hence, the direction v *of constant gray values can be obtained by computing the eigenvector of the smallest eigenvalue of the matrix $\mathcal{J}(x_0)$. The usual choice for w is a truncated Gaussian $w = K_\rho$ with mean 0, standard deviation ρ and support in $B_{\lfloor 3\rho \rfloor}(0)$. If we apply this idea for every image point $x \in \Omega$ we end up with a tensor field*

$$\mathcal{J}_\rho(x) := (K_\rho * \nabla u \nabla u^T)(x), \qquad x \in \Omega \tag{22}$$

which is called structure tensor *of u. Often u is first pre-smoothed by convolving it with another Gaussian of small standard deviation σ before computing the gradients ∇u_σ. If the eigenvalues of $\mathcal{J}_\rho(x_0)$ fulfill $\lambda_1 \gg \lambda_2$ we can assume that x_0 is in a region with homogenous gradient directions, e.g. in the neighborhood of an edge and the corresponding eigenvectors $v_1 = v = \mathrm{v}^\perp$ and $v_2 = \mathrm{v}$ approximate the gradient direction and the isophote direction, respectively.*
To get a discrete version of the structure tensor, we consider $\Omega := \{1, \dots, n\} \times \{1, \dots, n\}$ and an image $u = (u_{i,j})_{(i,j) \in \Omega}$. Set

$$P_{i,j} := \nabla u_{i,j} \nabla u_{i,j}^T, \quad (i,j) \in \Omega, \tag{23}$$

where ∇ is a discrete version of the gradient now. Let $\mathcal{N}(i_0, j_0) \subset \Omega$ be the neighborhood of $x_0 = (i_0, j_0)$ and let $(w_{i,j} / \sum w_{i,j})_{i,j}$ denote the sampled and normalized truncated Gaussian with mean 0 and standard deviation ρ. Then, we see that (22) corresponds to

$$\begin{aligned} \mathcal{J}_\rho(i_0, j_0) &= \frac{\sum\limits_{(i,j) \in \mathcal{N}(i_0,j_0)} w_{i_0 - i, j_0 - j} P_{i,j}}{\sum\limits_{(i,j) \in \mathcal{N}(i_0,j_0)} w_{i_0 - i, j_0 - j}} && (24) \\ &= \operatorname*{argmin}_{X \in \mathbb{R}^{2,2}} \sum_{(i,j) \in \mathcal{N}(i_0,j_0)} w_{i_0 - i, j_0 - j} \|X - P_{i,j}\|_{\mathcal{F}}^2. && (25) \end{aligned}$$

Hence, the discrete structure tensor at (i_0, j_0) is the solution of the weighted least squares problem (25)*, i.e., the weighted mean of the matrices $P_{i,j}$ in the neighborhood of (i_0, j_0). It has nearly the form of* (21) *with $\lambda = 0$ except that the Frobenius norm is squared now. To find gradients in images with similar orientations one is in the discrete setting, too, interested in the eigenvectors of $\mathcal{J}_\rho$.*

For the subsequent considerations it is useful to consider the common mapping $T : \mathrm{Sym}_2(\mathbb{R}) \to \mathbb{R}^3$ given by

$$T(X) := \frac{1}{\sqrt{2}}(x_{1,1} - x_{2,2},\ 2x_{1,2},\ x_{1,1} + x_{2,2})^{\mathrm{T}} = (x, y, z)^{\mathrm{T}}. \tag{26}$$

This mapping is an isometry from $\mathrm{Sym}_2(\mathbb{R})$ equipped with the Frobenius norm onto $\mathbb{R}^3$ with the ℓ_2-norm. In particular we have that

$$T^{-1}\left((x, y, z)^{\mathrm{T}}\right) = \frac{1}{\sqrt{2}} \begin{pmatrix} x + z & y \\ y & z - x \end{pmatrix}.$$

Note that the set of positive semi-definite matrices forms a convex cone in $\mathrm{Sym}_2(\mathbb{R})$ which can be illustrated using the mapping T as in Fig. 2.

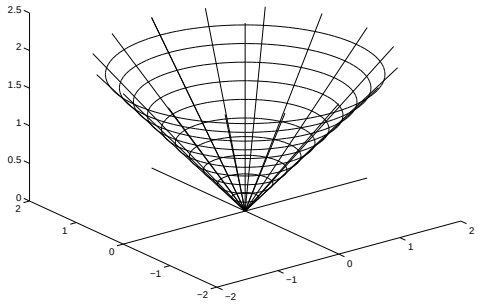


Figure 2: Cone $T(\mathrm{Sym}_2(\mathbb{R}))$ of symmetric, positive semi-definite matrices in $\mathrm{Sym}_2(\mathbb{R})$ visualized in $\mathbb{R}^3$ via (26).

The eigenvalues $\lambda_1 \geq \lambda_2$ of X are given by

$$\lambda_1 = \frac{1}{2}(x_{1,1} + x_{2,2} + \sqrt{(x_{1,1} - x_{2,2})^2 + 4x_{1,2}^2}) = \frac{1}{\sqrt{2}}(z + \|(x, y)^{\mathrm{T}}\|_2), \tag{27}$$

$$\lambda_2 = \frac{1}{2}(x_{1,1} + x_{2,2} - \sqrt{(x_{1,1} - x_{2,2})^2 + 4x_{1,2}^2}) = \frac{1}{\sqrt{2}}(z - \|(x, y)^{\mathrm{T}}\|_2), \tag{28}$$

so that, conversely, $\|(x, y)^{\mathrm{T}}\|_2 = \frac{1}{\sqrt{2}}(\lambda_1 - \lambda_2)$ and $z = \frac{1}{\sqrt{2}}(\lambda_1 + \lambda_2)$. If $\lambda_1 > \lambda_2$, a short calculation shows that the eigenspace of λ_1 is spanned by

$$v = \begin{pmatrix} x + \|(x, y)^{\mathrm{T}}\|_2 \\ y \end{pmatrix}$$

which is independent of z.
Finally, our three matrix norms can be rewritten as

$$\|X\|_* = \sqrt{2}\max\{\|(x, y)^{\mathrm{T}}\|_2, |z|\}, \quad \|X\|_{\mathcal{F}} = \|(x, y, z)^{\mathrm{T}}\|_2, \quad \|X\|_2 = \frac{1}{\sqrt{2}}\left(\|(x, y)^{\mathrm{T}}\|_2 + |z|\right). \tag{29}$$

This immediately implies the following proposition. In particular, the proposition shows that the matrix median of 2×2 matrices with respect to the spectral norm can be computed via vector medians, see also [44].

Proposition 4.2. *Let* $T(\hat{X}) := (\hat{x}, \hat{y}, \hat{z})^T$, $T(F) := (x_F, y_F, z_F)^T$ *and* $T(P_i) := (x_i, y_i, z_i)^T$.
i) *The matrix* $\hat{X}$ *is the minimizer of* $\sum_{i=1}^{M} w_i \|X - P_i\|_{\mathcal{F}}^2$, *i.e.,* $\hat{X} = \sum_{i=1}^{M} w_i \, P_i / \sum_{i=1}^{M} w_i$ *if and only if*

- $\hat{z}$ *is the minimizer of* $\sum_{i=1}^{M} w_i (z - z_i)^2$ *and*
- $(\hat{x}, \hat{y})^T$ *is the minimizer of* $\sum_{i=1}^{M} w_i \|(x, y)^T - (x_i, y_i)^T\|_2^2$.

ii) *The matrix* $\hat{X}$ *is a minimizer of* (21) *with the spectral norm* $\bullet = 2$ *if and only if*

- $\hat{z}$ *is a minimizer of* $\frac{\lambda}{2}(z - z_F)^2 + \frac{1}{\sqrt{2}} \sum_{i=1}^{M} w_i \, |z - z_i|$ *and*
- $(\hat{x}, \hat{y})^T$ *is a minimizer of*

$$\frac{\lambda}{2} \|(x, y)^T - (x_F, y_F)^T\|_2^2 + \frac{1}{\sqrt{2}} \sum_{i=1}^{M} w_i \|(x, y)^T - (x_i, y_i)^T\|_2. \tag{30}$$

If $\mathrm{tr}(F) = \mathrm{tr}(P_i)$ for all $i = 1, \ldots, M$, by the following proposition the minimizers of (21) coincide if λ is substituted by $\sqrt{2}\lambda$ and $\lambda/\sqrt{2}$ for the nuclear and spectral norm, respectively. In particular, we have that the traces of rank-1 matrices $P_i = p_i p_i^{\mathrm{T}}$ as those in (23) are equal if we use normalized directions $\|p_i\|_2 = 1$ for all $i = 1, \ldots, M$.

Proposition 4.3. i) *Suppose that* $\mathrm{tr}(X) = 0$, *then* $\|X\|_{\mathcal{F}} = \frac{1}{\sqrt{2}} \|X\|_* = \sqrt{2} \|X\|_2 = \|(x, y)^T\|_2$.
ii) *Suppose that* $\mathrm{tr}(F) = \mathrm{tr}(P_i)$ *for all* $i = 1, \ldots, M$. *Let* $w_{\mathcal{F},i} = w_i$, $w_{*,i} = \frac{1}{\sqrt{2}} w_i$, $= w_{2,i} = \sqrt{2} w_i$. *Then, the minima*

$$\min_X \big\{ \frac{\lambda}{2} \|X - F\|_{\mathcal{F}}^2 + \sum_{i=1}^{M} w_{\bullet,i} \|X - P_i\|_\bullet \big\} \tag{31}$$

are the same for all $\bullet \in \{*, \mathcal{F}, 2\}$ *and there exists a common minimizer* $\hat{X}$.

Proof: Part i) follows directly from (29) and the definition $z = \mathrm{tr}(X)$.
To prove ii) we set $T(F) := (x_F, y_F, z_F)^{\mathrm{T}}$ and $T(P_i) := (x_i, y_i, z_i)^{\mathrm{T}}$. Since $\mathrm{tr}(F) = \mathrm{tr}(P_i)$, $i = 1, \ldots, M$, we obtain by i) that

$$\begin{aligned}
&\min_{X \in \mathbb{R}^{2,2}} \big\{ \frac{\lambda}{2} \|X - F\|_{\mathcal{F}}^2 + \sum_{i=1}^{M} w_{\bullet,i} \|X - P_i\|_\bullet \big\} \\
&\leq \min_{\mathrm{tr}(X) = \mathrm{tr}(F)} \big\{ \frac{\lambda}{2} \|X - F\|_{\mathcal{F}}^2 + \sum_{i=1}^{M} w_{\bullet,i} \|X - P_i\|_\bullet \big\} \\
&= \min_{x,y} \big\{ \frac{\lambda}{2} \|(x - x_F, y - y_F)^{\mathrm{T}}\|_2^2 + \sum_{i=1}^{M} w_i \|(x - x_i, y - y_i)^{\mathrm{T}}\|_2 \big\}.
\end{aligned} \tag{32}$$

On the other hand we have by (29) that

$$\begin{aligned}&\min_{X\in\mathbb{R}^{2,2}}\{\frac{\lambda}{2}\|X-F\|_{\mathcal{F}}^2+\sum_{i=1}^{M}w_{\bullet,i}\|X-P_i\|_\bullet\}\\ &\geq\ \min_{x,y}\frac{\lambda}{2}\big\{\|(x-x_F,y-y_F)^{\mathrm{T}}\|_2^2+\sum_{i=1}^{M}w_i\|(x-x_i,y-y_i)^{\mathrm{T}}\|_2\big\}.\end{aligned}$$

Hence, (32) is the minimum of (31) for all $\bullet\in\{*,\mathcal{F},2\}$. A common minimizer is given by $\hat{X}$ with $T(\hat{X})=(\hat{x},\hat{y},\frac{1}{\sqrt{2}}\mathrm{tr}(F))^{\mathrm{T}}$, where $(\hat{x},\hat{y})^{\mathrm{T}}$ is the unique minimizer of (32), cf. Subsection 3.1. □

Finally, we will deal again with rank-1 matrices $P_i=p_ip_i^{\mathrm{T}}$ as those used in the computation of the structure tensor, where $\|p_i\|_2$ may vary for $i=1,\dots,M$ now. Then, the natural question arises if the eigenvectors $v=v(\hat{X})$ belonging to the largest eigenvalue of

$$\hat{X}:=\operatorname*{argmin}_{X\in\mathbb{R}^{2,2}}\big\{\sum_{i=1}^{M}w_i\|X-P_i\|_{\mathcal{F}}^2\big\}\tag{33}$$

are related to the vector

$$\hat{p}:=\operatorname*{argmin}_{p\in\mathbb{R}^2}\big\{\sum_{i=1}^{M}w_i\|p-g(p_i)\|_2^2\big\}\tag{34}$$

for an appropriate function g. Of course we cannot expect a relation for $g(p_i):=p_i$, since P_i does not take the orientation of p_i into account, i.e., $P_i=(-p_i)(-p_i)^{\mathrm{T}}$. However, we can define an appropriate function g which is invariant to a sign change. To this end, let us first write p_i in the form $p_i=\|p_i\|_2(\cos\alpha_i,\sin\alpha_i)^{\mathrm{T}}$ and let $R(\alpha):=\begin{pmatrix}\cos\alpha&-\sin\alpha\\ \sin\alpha&\cos\alpha\end{pmatrix}$ denote the rotation matrix by some angle α. Now observe that $R(\alpha_i)p_i=R(\alpha_i+180°)(-p_i)$ and $-p_i=\|p_i\|_2(\cos(\alpha_i+180°),\sin(\alpha_i+180°))^{\mathrm{T}}$. We will see that

$$g(p):=\frac{\|p\|_2}{\sqrt{2}}R(\alpha_p)p,\tag{35}$$

where $p=\|p\|_2(\cos\alpha_p,\sin\alpha_p)^{\mathrm{T}}$, is a good choice.
The following proposition shows the desired relation for the above squared Frobenius norm and for the spectral median.

Proposition 4.4. *Let $F:=ff^T$, $P_i:=p_ip_i^T$ and $p_i=\|p_i\|_2(\cos\alpha_i,\sin\alpha_i)^T$, $i=1,\dots,M$.*
i) *Assume the eigenvalues of the minimizer $\hat{X}$ of (33) satisfy $\lambda_1\neq\lambda_2$. If v is the eigenvector belonging to the largest eigenvalue, then*

$$v=const\cdot R\left(-\frac{\alpha_{\hat{p}}}{2}\right)\hat{p},\tag{36}$$

where $\hat{p}$ is the minimizer of (34) with g as defined in (35). We have $\hat{p}=0$ if and only if the eigenvalues of $\hat{X}$ are equal. Furthermore, it holds that $\|\hat{p}\|_2=\frac{1}{\sqrt{2}}(\lambda_1-\lambda_2)$.
ii) *Let $\hat{X}$ be a minimizer of (21) with the spectral norm $\bullet=2$. If the eigenvalues of $\hat{X}$ are*

not equal, then the eigenvectors v belonging to the largest eigenvalue are given by (36)*, where $\hat{p}$ is a minimizer of*

$$\frac{\lambda}{2}\|p-f\|_2^2 + \frac{1}{\sqrt{2}}\sum_{i=1}^{M} w_i \|p - g(p_i)\|_2. \tag{37}$$

If $\hat{p}=0$, then there exists a minimizer $\hat{X}$ of (21) *with equal eigenvalues. Morover, $\|\hat{p}\|_2 = \frac{1}{\sqrt{2}}(\lambda_1 - \lambda_2)$.*

Note that we have seen in Subsection 3.1 that (37) has in general a unique minimizer.

Proof: i) Let $\hat{X}$ be the minimizer of (33) and let $T(\hat{X}) = (\hat{x},\hat{y},\hat{z})^{\mathrm{T}}$ and $(\hat{x},\hat{y})^{\mathrm{T}} = \|(\hat{x},\hat{y})^{\mathrm{T}}\|_2(\cos\alpha,\sin\alpha)^{\mathrm{T}}$. Then we know that the eigenspace of the largest eigenvalue of $\hat{X}$ is spanned by

$$\begin{aligned} v &= \begin{pmatrix} \hat{x} + \|(\hat{x},\hat{y})^{\mathrm{T}}\|_2 \\ \hat{y} \end{pmatrix} = \|(\hat{x},\hat{y})^{\mathrm{T}}\|_2 \begin{pmatrix} \cos\alpha + 1 \\ \sin\alpha \end{pmatrix} \\ &= 2\|(\hat{x},\hat{y})^{\mathrm{T}}\|_2 \cos\frac{\alpha}{2} \begin{pmatrix} \cos\frac{\alpha}{2} \\ \sin\frac{\alpha}{2} \end{pmatrix} = const \cdot R\left(-\frac{\alpha}{2}\right)\begin{pmatrix}\hat{x}\\ \hat{y}\end{pmatrix}. \end{aligned} \tag{38}$$

By definition of P_i, we obtain that $T(P_i) = (p_{i,1}^2 - p_{i,2}^2, 2p_{i,1}p_{i,2}, p_{i,1}^2 + p_{i,2}^2)^{\mathrm{T}}/\sqrt{2}$ and

$$\begin{pmatrix} p_{i,1}^2 - p_{i,2}^2 \\ 2p_{i,1}p_{i,2} \end{pmatrix} = \|p_i\|^2 \begin{pmatrix} \cos^2\alpha_i - \sin^2\alpha_i \\ 2\cos\alpha_i \sin\alpha_i \end{pmatrix} = \|p_i\|^2 \begin{pmatrix} \cos(2\alpha_i) \\ \sin(2\alpha_i) \end{pmatrix} = \|p_i\| R(\alpha_i) p_i.$$

Hence by Proposition 4.2 i) we conclude that $(\hat{x},\hat{y})^{\mathrm{T}}$ is the minimizer of (34) with $\|(\hat{x},\hat{y})^{\mathrm{T}}\|_2 = \frac{1}{\sqrt{2}}(\lambda_1 - \lambda_2)$. By (38) this implies (36). Furthermore we have that $\hat{p}=0 \Leftrightarrow (\hat{x},\hat{y})^{\mathrm{T}} = 0 \Leftrightarrow \lambda_1 = \lambda_2$.
ii) The proof of ii) follows the same line as i), where we finally use Proposition 4.2 ii). □

This proposition shows that instead of computing the eigenvectors to the largest eigenvalue of the structure tensor, we can take the vectors p_i, rotate them by their own angle, scale them by $1/\sqrt{2}$ times their length and compute the weighted vector mean of these vectors. The result must then be rotated back by half of its angle to obtain the orientation of the eigenvectors. The length of the weighted vector mean can be used as a measure for the dominance of the found orientation. Similarly, instead of computing the eigenvectors of the spectral matrix median (21), we can apply the same procedure and solve the vector median problem (37).

5 Numerical Experiments

The purpose of this section is twofold. In the next subsection, we compare the different algorithms proposed in Subsection 3.2 for the ℓ_2 vector median computation with respect to their efficiency if sequential programming is used. In Subsection 5.2, we compute locally matrix means (= structure tensors) and matrix medians for noisy images and compare the directions of their eigenvectors. Moreover, we show their relation to special vector medians considered in Subsection 4.2. Depending on the kind of noise in the images and, consequently, in the image gradients, we will see that matrix means and medians with respect to different norms show a different behavior.

5.1 Comparison of Algorithms for ℓ_2 Vector Median Computation

In this subsection, we compare the computation time of Weiszfeld's algorithm, ADMM and PPXA introduced in Subsection 3.2 for the solution of the two-dimensional ℓ_2 vector median problem

$$\operatorname*{argmin}_{x\in\mathbb{R}^2}\{\sum_{i=1}^{M} w_i\|x-p_i\|_2\}, \tag{39}$$

with given points $p_i \in \mathbb{R}^2$ and weights $w_i > 0$, $i = 1,\dots,M$. We also solved (39) via SOCP using the commercial software MOSEK 6.0. The reason for not showing the corresponding running times here is that they turn out to be much higher than for the other three algorithms.

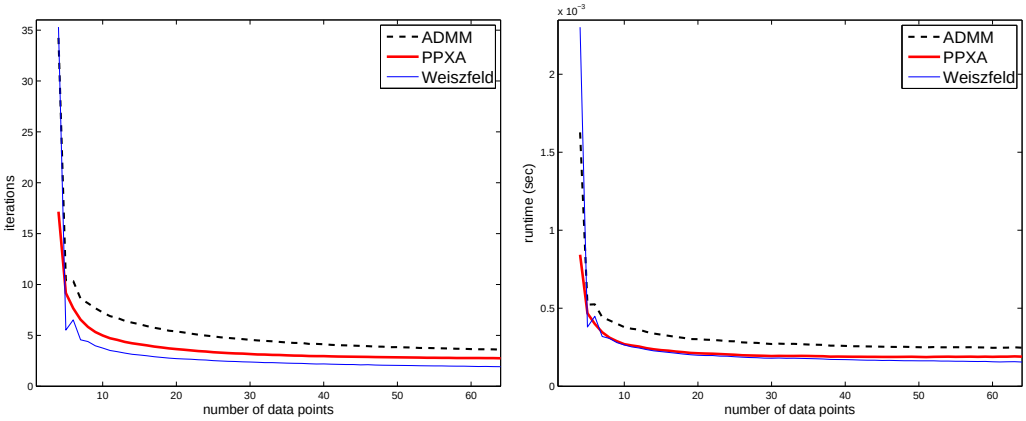


Figure 3: Performance of Weiszfeld's algorithm, ADMM and PPXA for problem (39) applied to random points on the unit circle and unit weights. Average number of iterations (left) and average computation time (right) in dependence on the number of points.

In the **first experiment**, the points p_i, $i = 1,\dots,M$ are chosen randomly on the unit circle and we use unit weights. This is repeated $10,000$ times for any number $M = 4,\dots,64$ of points. Fig. 3 shows the average number of iterations and the average computation time in dependence on the number M of data points. As stopping criterion, we use that the maximal difference of each component of $x^{(r)}$ with respect to a reference solution obtained after sufficiently many iterations must be smaller than 0.001. Note that we have determined the optimal parameters for the different algorithms by hand. The best parameters of Weiszfeld's algorithm and of ADMM turn out to be the same for all M, namely $c_r = 1.8$ and $\gamma = 1.1$, respectively. For the PPXA, we use $\sigma_i = 1/M$. The parameter μ can be set to 1.8 but for optimal performance it is necessary to increase the parameter γ from 9.0 to 64.0 as the number of data points increases from 4 to 64. Concerning the initial values, we set in Weiszfeld's algorithm $x^{(0)} = 0$, in ADMM $b_i^{(0)} = -p_i$, $v^{(0)} = 0$ and in PPXA $x^{(0)} = 0$, $y_i^{(0)} = 0$.
Interestingly, in this experiment all three algorithms become faster as the number of data points increases. The reason for this might be that for our data both median and mean tend to zero as the number of points goes to infinity. Moreover, we see that Weiszfeld's algorithm performs best, followed by PPXA and ADMM. It is important to note, however, that in contrast to Weiszfeld's algorithm, ADMM and PPXA can be parallelized to a high degree so

that a significant speed-up of a parallel implementation, e.g., on a GPU, can be expected for these two algorithms.

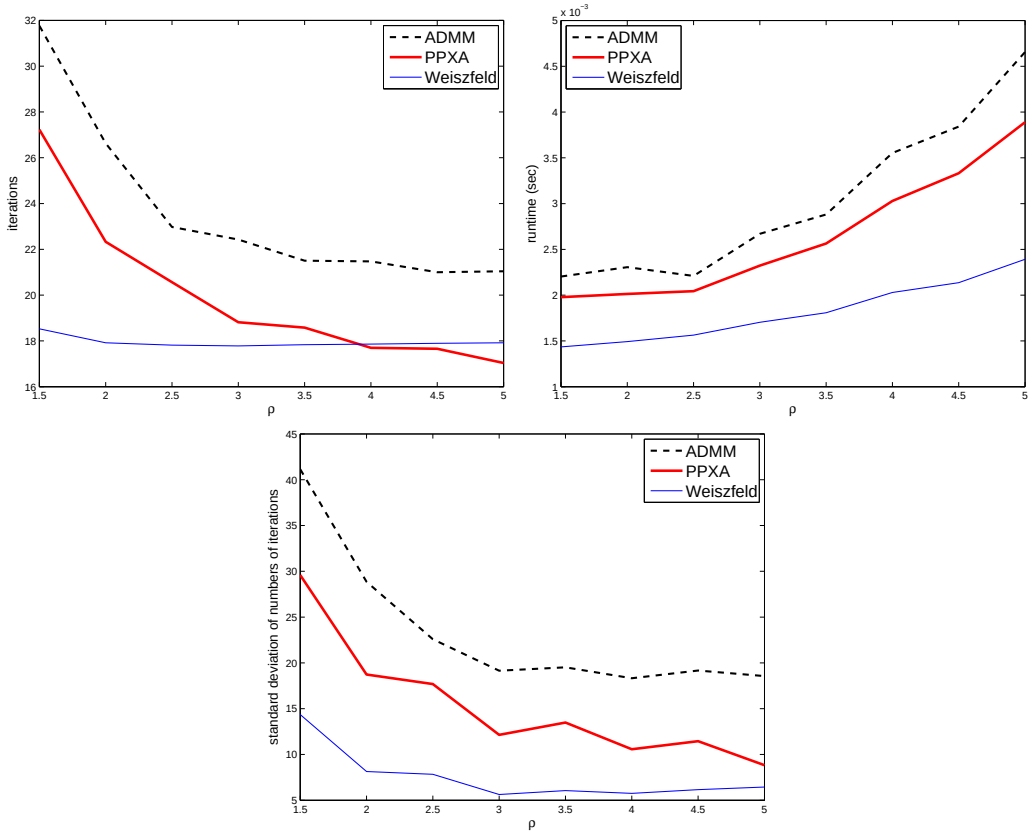


Figure 4: Comparison of Weiszfeld's algorithm, ADMM and PPXA for problem (39) applied to the gradients of the noisy image on the left-hand side of Fig. 7 and Gaussian weights. Top: average number of iterations and average computation time for each pixel in dependence on ρ, the standard deviation of the Gaussian weights. Bottom: standard deviation in the number of iterations per pixel.

In our **second experiment**, we show the computation time for (39) applied to the noisy image on the left-hand side of Fig. 7. More precisely, for each pixel our data points are now the gradients of neighboring pixels which result from a central difference discretization. The weights w_i are given by a sampled Gaussian and we vary the corresponding standard deviation ρ in steps of 0.5 from 1 to 5 with $M = (2\lfloor 3\rho \rfloor + 3)^2$ data points. The stopping criterion is the same as in the first experiment, i.e., the maximal error in each component of $x^{(r)}$ must be smaller than 0.001. Note that, in contrast to the first experiment, the norm of the data points can vary substantially because of the impulse noise. The optimal parameter for Weiszfeld's algorithm is again $c_r = 1.8$. The parameters μ and σ_i in PPXA is set to 1.8 and $1/M$, respectively, for all ρ. However, for both ADMM and PPXA we have to adapt the parameters γ as ρ runs from 1.5 to 5: The parameter γ in ADMM and PPXA is increased from 540 to 6,000 and from 36,000 to 4,400,000, respectively. We use the same initial values as described in the first experiment.
The results are shown in Fig. 4. Weiszfeld's algorithm again needs the fewest number of

iterations with the exception of relatively large ρ where PPXA comes in first. ADMM needs the most iterations for all ρ. In contrast to the first experiment, the computation time now increases for all three algorithms as ρ increases. Furthermore, it is interesting to see that there is a high standard deviation of the number of iterations, especially for ADMM and to a lesser degree for PPXA. For all three algorithms the standard deviation is larger for small ρ.

5.2 Eigenvectors of Structure Tensors and Matrix Medians and Vector Means and Medians

In this section, we investigate the properties of

- eigenvectors of structure tensors (= matrix means) and corresponding vector means,
- eigenvectors of matrix medians for different matrix norms and corresponding ℓ_2 vector medians.

In particular, we want to demonstrate consequences of the results of Subsection 4.2 in image processing.
For this purpose, we apply our medians as local filters to the image gradients in the neighborhood of image pixels. Thus, $M = M(i_0, j_0)$ is the number of pixels in the neighborhood of (i_0, j_0). If not stated otherwise, we use the input matrices

$$P_{i,j} = \nabla u_{i,j} \nabla u_{i,j}^{\mathrm{T}} = p_{i,j} p_{i,j}^{\mathrm{T}}$$

for the structure tensor and the matrix medians and the vectors $p_{i,j}$, resp., $\tilde{g}(p_{i,j}) = R(\alpha_{ij})\nabla u_{ij}$ with $p_{ij} = \|p_{ij}\|_2(\cos\alpha_{i,j}, \sin\alpha_{i,j})^{\mathrm{T}}$ for vector mean and median computations. Note that in contrast to Proposition 4.4 we do not apply a scaling for the modified vectors $\tilde{g}(p_{i,j})$ since the results do not show visual differences. For the modified vectors $\tilde{g}(p_{i,j})$ we finally rotate the result $\hat{v}$ back by half of its angle as described in Proposition 4.4. The gradients are discretized by central differences. Moreover, we set $\lambda = 0$ and the weights w_{ij} to be a sampled Gaussian centered at the center of the neighborhood for given standard deviation ρ. In general, no presmoothing of the corrupted image is used. In those cases, where we presmoothed the image by convolving it with a Gaussian, the standard deviation σ of the Gaussian is specified.

The results of our matrix/vector median filters are compared to those of the structure tensor and the corresponding local weighted vector means. To visualize the results, we plot the angles of the solutions of the vector filters as well as the angles of the eigenvectors to the largest eigenvalue of the results of the matrix filters. To have a consistent coloring of the angles, all angles are displayed modulus 180°.

First Example. Our first example in Fig. 5 shows a test image with sharp edges and regions with constant gradients, which is corrupted by additive Gaussian noise. In Fig. 6 the results of the structure tensor are compared to our matrix median filter with the spectral norm as a representative for all median filters. Here, we can see that the structure tensor is especially well-suited to restore the gradient orientations at the edges of the objects. After a slight presmoothing, it is also able to restore the gradients within the objects to a certain degree. For this noisy image the results of our matrix median filter are slightly worse, since all gradients are corrupted by the noise. As one may expect, the results are completely different if the image is corrupted by impulse noise instead of Gaussian noise.

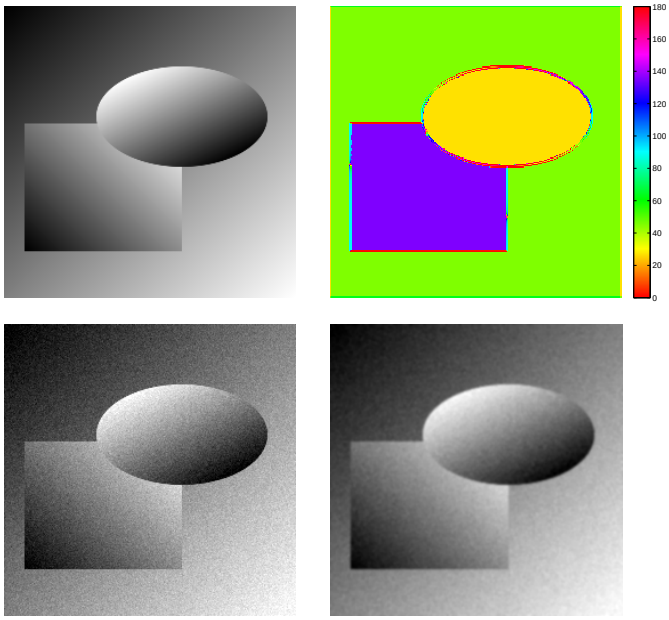
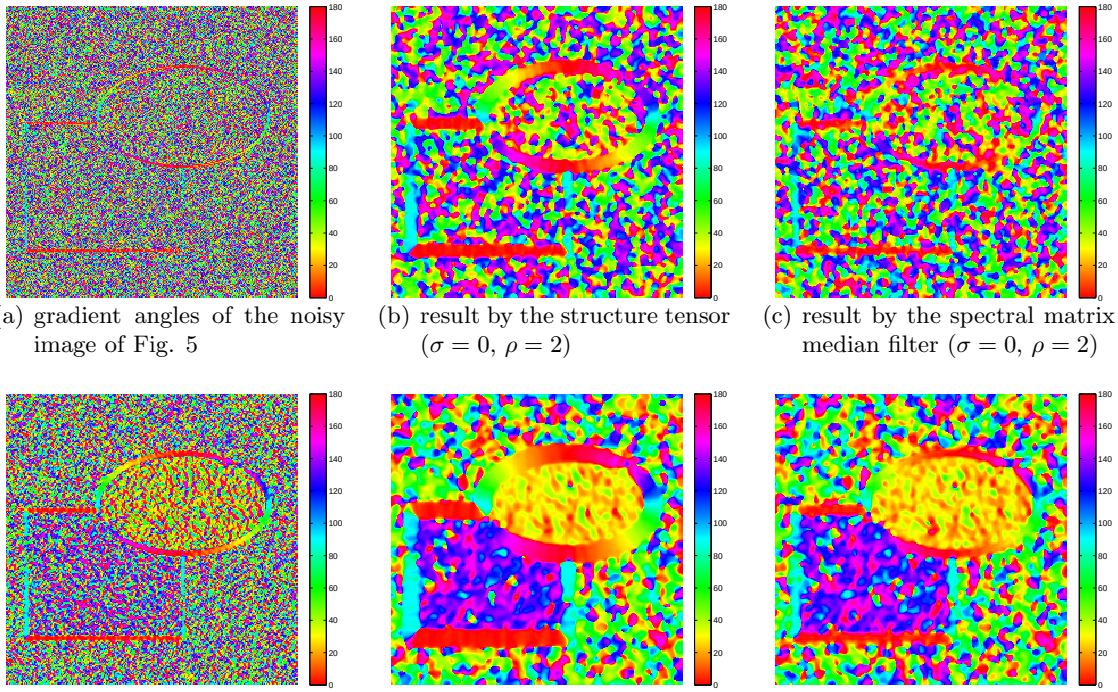


Figure 5: Top: Test image of size 256×256 (left) and the angles of the gradients (right). Bottom: Image corrupted by additive Gaussian noise of standard deviation 10 (left) and smoothed image by a Gaussian kernel of standard deviation $\sigma = 1$ (right).

Second Example. Now the initial image in Fig. 5 (top) has been corrupted by 20% impulse noise as it is displayed in Fig. 7. The results in Fig. 8 show that all median filters (second and third row) give much better results than the mean filters with $p_{i,j}$ and $\tilde{g}(p_{i,j})$ and the structure tensor in the first row. The best results are obtained by the spectral matrix median and its relatives in the second row. Especially the gradient angles at regions where the original image has constant gradients are much better restored. The mean filter as well as the structure tensor have quite big difficulties with these regions. As indicated by Proposition 4.4 the results for the modified vector mean and the structure tensor, resp., for the modified vector median and the spectral matrix median are nearly the same. The fact that we did not scale $\tilde{g}(p_{ij})$ hardly plays a role for the results.
With our matrix median filters the angles of the eigenvectors are almost the same with the nuclear norm and the spectral norm. Only for the Frobenius norm the result is worse. To show also an example with normalized gradients we include Fig. 8 (i). This result was generated by the matrix median with the spectral norm. However, by Proposition 4.3 ii) it follows that the minimizers with the nuclear, the Frobenius and the spectral norm are the same in this case. Compared to the results by the matrix median filters without normalized gradients it is significantly worse than the one with the spectral and nuclear norm, but better than the result with the Frobenius norm.

To investigate the *similar eigenvectors of the matrix median filters with the nuclear and spectral norm*, we include Fig. 9. By Proposition 4.2 we know that if we set $T(P_i) = (x_i, y_i, z_i)^{\mathrm{T}}$ and $T(\hat{X}) = (\hat{x}, \hat{y}, \hat{z})^{\mathrm{T}}$, the vector $(\hat{x}, \hat{y})^{\mathrm{T}}$ and thus the eigenvectors of a minimizer



(a) gradient angles of the noisy image of Fig. 5

(b) result by the structure tensor ($\sigma = 0$, $\rho = 2$)

(c) result by the spectral matrix median filter ($\sigma = 0$, $\rho = 2$)

(d) gradients of the presmoothed image of Fig. 5 ($\sigma = 1$)

(e) result by the structure tensor ($\sigma = 1$, $\rho = 2.5$)

(f) result by the spectral matrix median filter ($\sigma = 1$, $\rho = 2.5$)

Figure 6: Results for the Gaussian noise corrupted and presmoothed image of Fig. 5.

$\hat{X}$ of the matrix median filter with the spectral norm can be obtained by

$$\min_{(x,y)^{\mathrm{T}}} \frac{1}{\sqrt{2}} \sum_{i=1}^{M} w_i \|(x,y)^{\mathrm{T}} - (x_i,y_i)^{\mathrm{T}}\|_2.$$

On the other hand, a minimizer $\hat{X}$ of the matrix median filter with the nuclear norm is given by

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \operatorname*{argmin}_{(x,y,z)^{\mathrm{T}}} \frac{1}{\sqrt{2}} \sum_{i=1}^{M} w_i \max\{\|(x,y)^{\mathrm{T}} - (x_i,y_i)^{\mathrm{T}}\|_2, |z - z_i|\}.$$

Hence, if $\hat{X}$ fulfills $\|(\hat{x},\hat{y})^{\mathrm{T}} - (x_i,y_i)^{\mathrm{T}}\|_2 \geq |\hat{z} - z_i|$ for all $i = 1,\ldots,M$, the eigenvectors of the solutions of both problems coincide. This is the case for the patch shown in Fig. 9 where only small differences appear, cf. the difference image shown Fig. 9 (b). Note that we have for the eigenvalues λ_1 and λ_2 of $\hat{X} - P_i$ that

$$\|(x,y)^{\mathrm{T}} - (x_i,y_i)^{\mathrm{T}}\|_2 = \frac{1}{\sqrt{2}}(\lambda_1 - \lambda_2) \quad \text{and} \quad |z - z_i| = \frac{1}{\sqrt{2}}|\lambda_1 + \lambda_2|.$$

Third Example. Our last example in Figs. 10 and 11 shows the results for a second noisy test image. In this image the gradient angles change steadily. As we can see, this time all results of our median filters look quite similar. Only the ordinary vector median filter is

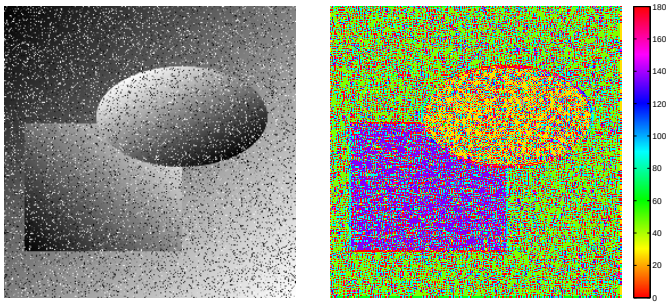


Figure 7: Original image of Fig. 5 corrupted by 20% impulse noise (left) and corresponding gradient directions (right).

different at pixels where the sign of the gradients changes in the noise-free image. Moreover, the result with the Frobenius norm differs at some of the pixels where the original image was constant.

6 Conclusions

In this paper, we have applied various algorithms for the computation of generalized vector and matrix medians. We have shown how these algorithms look like for our problems and compared their efficiency numerically. Furthermore, we have investigated the relations between means and medians of symmetric 2×2 matrices arising in image processing. A connection between the eigenvectors of these matrices and corresponding vector means and medians have been established. The application of these algorithms in local median filters leads to the smoothing of directions. However, the purpose of this paper was to understand the differences between the matrix mean and the matrix medians with respect to the nuclear norm, Frobenius norm and spectral norm and not the smoothing of directions itself as for example done in [19, 24, 35, 36]. Other applications such as colorization or topics like tensor median computation are also subjects of future research.

A Appendix

We prove the convergence of the generalized Weiszfeld algorithm. Throughout the appendix the norm $\|\cdot\|$ denotes the Euclidean norm.

Lemma A.1. *Let $K = I$ and $0 < c_r \leq 2$. Then we have that $x^{(r)} \in \operatorname{argmin} E_v$ if and only if $x^{(r+1)} = x^{(r)}$. For $x^{(r)} \notin \operatorname{argmin} E_v$ it holds $E_v(x^{(r+1)}) < E_v(x^{(r)})$ if $0 < c_r < 2$ or if $c_r = 2$ and there exist $d+1$ affine independent points in P.*

Proof: 1. For $x^{(r)} \in \operatorname{argmin} E_v$ we have by (19) and (20) that $G(x^{(r)}) = 0$ and consequently that $x^{(r+1)} = x^{(r)}$. Conversely, if $x^{(r+1)} = x^{(r)}$, then $G(x^{(r)}) = 0$ which is by (19) and (20) only possible if $x^{(r)} \in \operatorname{argmin} E_v$.
2. Assume that $x^{(r)} \notin \operatorname{argmin} E_v$ so that $G(x^{(r)}) \neq 0$. Since $S(x^{(r)})$ is symmetric and positive

definite, we obtain

$$\begin{aligned} x^{(r+1)} - x^{(r)} &= -c_r S(x^{(r)})^{-1} G(x^{(r)}) \\ \langle x^{(r+1)} - x^{(r)}, G(x^{(r)})\rangle &= -c_r \langle S(x^{(r)})^{-1} G(x^{(r)}), G(x^{(r)})\rangle < 0. \end{aligned}$$

Hence it follows for $c_r \leq 2$ that

$$\begin{aligned} 0 &\geq (2 - c_r)\langle x^{(r+1)} - x^{(r)}, G(x^{(r)})\rangle \\ &= \langle x^{(r+1)} - x^{(r)}, 2G(x^{(r)}) - c_r G(x^{(r)})\rangle \\ &= \langle x^{(r+1)} - x^{(r)}, 2G(x^{(r)}) + S(x^{(r)})(x^{(r+1)} - x^{(r)})\rangle. \end{aligned} \tag{40}$$

Case 1: If $x^{(r)} \notin P$, we get by definition of G and S that

$$\begin{aligned} 0 \geq\ & \langle x^{(r+1)} - x^{(r)}, 2\lambda K^{\mathrm{T}} K x^{(r)} - 2\lambda K^{\mathrm{T}} f + 2\sum_{i=1}^{M} w_i \frac{x^{(r)} - p_i}{\|x^{(r)} - p_i\|} \\ & + \lambda K^{\mathrm{T}} K (x^{(r+1)} - x^{(r)}) + \sum_{i=1}^{M} w_i \frac{1}{\|x^{(r)} - p_i\|}\,(x^{(r+1)} - x^{(r)})\rangle \\ =\ & \lambda \langle x^{(r+1)} - x^{(r)}, K^{\mathrm{T}} K (x^{(r+1)} - x^{(r)})\rangle + 2\lambda \langle x^{(r+1)} - x^{(r)}, K^{\mathrm{T}} K x^{(r)} - K^{\mathrm{T}} f\rangle \\ & + \sum_{i=1}^{M} w_i \frac{\|x^{(r+1)}\|^2 - \|x^{(r)}\|^2 - 2\langle x^{(r+1)}, p_i\rangle + 2\langle x^{(r)}, p_i\rangle}{\|x^{(r)} - p_i\|} \end{aligned}$$

and since the first two summands add to $\lambda\|Kx^{(r+1)} - f\|^2 - \lambda\|Kx^{(r)} - f\|^2$ we further have

$$0 \geq \lambda\|Kx^{(r+1)} - f\|^2 - \lambda\|Kx^{(r)} - f\|^2 + \sum_{i=1}^{M} w_i \frac{\|x^{(r+1)} - p_i\|^2 - \|x^{(r)} - p_i\|^2}{\|x^{(r)} - p_i\|}. \tag{41}$$

By definition of E_v this implies

$$\begin{aligned} E_v(x^{(r)}) &\geq \lambda\|Kx^{(r+1)} - f\|^2 - \frac{\lambda}{2}\|Kx^{(r)} - f\|^2 + \sum_{i=1}^{M} w_i \frac{\|x^{(r+1)} - p_i\|^2}{\|x^{(r)} - p_i\|}, \\ 2E_v(x^{(r)}) &\geq 2E_v(x^{(r+1)}) + \sum_{i=1}^{M} w_i \frac{(\|x^{(r+1)} - p_i\| - \|x^{(r)} - p_i\|)^2}{\|x^{(r)} - p_i\|}, \\ E_v(x^{(r)}) - E_v(x^{(r+1)}) &\geq \frac{1}{2}\sum_{i=1}^{M} w_i \frac{(\|x^{(r+1)} - p_i\| - \|x^{(r)} - p_i\|)^2}{\|x^{(r)} - p_i\|} \ \geq\ 0. \end{aligned} \tag{42}$$

The left-hand side is positive if $c_r < 2$. If $c_r = 2$ and there exist $d + 1$ points in P which are affine independent, then there exits a unique point having the same distance from these points, [5, p. 127]. Therefore the left-hand side is also positive in this case.

Case 2: If $x^{(r)} = p_k \in P$, we can rewrite (40) by definition of G and S as follows:

$$\begin{aligned} 0 \;\geq\; & \lambda\|K(x^{(r+1)} - p_k)\|^2 + 2\lambda\langle K(x^{(r+1)} - p_k), Kp_k - f\rangle - 2w_k\langle x^{(r+1)} - p_k, \frac{G_k}{\|G_k\|}\rangle \\ & + \sum_{\substack{i=1 \\ i\neq k}}^{M} \frac{w_i}{\|p_k - p_i\|}\left(2\langle x^{(r+1)} - p_k, p_k - p_i\rangle + \langle x^{(r+1)} - p_k, x^{(r+1)} - p_k\rangle\right) \\ = \; & \lambda\left(\|Kx^{(r+1)} - f\|^2 - \|Kp_k - f\|^2\right) - 2w_k\langle x^{(r+1)} - p_k, \frac{G_k}{\|G_k\|}\rangle \\ & + \sum_{\substack{i=1 \\ i\neq k}}^{M} \frac{w_i}{\|p_k - p_i\|}\left(\|x^{(r+1)} - p_i\|^2 - \|p_k - p_i\|^2\right). \end{aligned} \tag{43}$$

Since

$$x^{(r+1)} - p_k = -c_r S(p_k)^{-1} G(p_k) = -c_r S(p_k)^{-1}(\|G_k\| - w_k)\frac{G_k}{\|G_k\|}$$

and

$$S(p_k)(x^{(r+1)} - p_k) = -c_r(\|G_k\| - w_k)\frac{G_k}{\|G_k\|}$$

we obtain with $\|G_k\| - w_k > 0$ and $c_r > 0$ that $\|S(p_k)(x^{(r+1)} - p_k)\| = c_r(\|G_k\| - w_k)$ and consequently

$$\langle x^{(r+1)} - p_k, \frac{G_k}{\|G_k\|}\rangle = -\frac{\langle x^{(r+1)} - p_k, S(p_k)(x^{(r+1)} - p_k)\rangle}{\|S(p_k)(x^{(r+1)} - p_k)\|}. \tag{44}$$

Adding $E_v(p_k)$ on both sides of (43) and using (44) yields

$$\begin{aligned} E_v(p_k) \;\geq\; & \lambda\|Kx^{(r+1)} - f\|^2 - \frac{\lambda}{2}\|Kp_k - f\|^2 + 2w_k\frac{\langle x^{(r+1)} - p_k, S(p_k)(x^{(r+1)} - p_k)\rangle}{\|S(p_k)(x^{(r+1)} - p_k)\|} \\ & + \sum_{\substack{i=1 \\ i\neq k}}^{M} w_i\frac{((\|x^{(r+1)} - p_i\| - \|p_k - p_i\|) + \|p_k - p_i\|)^2}{\|p_k - p_i\|}, \\ 2E_v(p_k) \;>\; & 2E_v(x^{(r+1)}) + \sum_{\substack{i=1 \\ i\neq k}}^{M} w_i\frac{(\|x^{(r+1)} - p_i\| - \|p_k - p_i\|)^2}{\|p_k - p_i\|}, \end{aligned}$$

where the last inequality takes into account that $x^{(r)} = p_k \neq x^{(r+1)}$ and $\frac{\langle x^{(r+1)} - p_k, S(p_k)(x^{(r+1)} - p_k)\rangle}{\|S(p_k)(x^{(r+1)} - p_k)\|} - \|x^{(r+1)} - p_k\| = 0$ by the special structure of S. □

Lemma A.2. *Let $p_k \notin \operatorname{argmin} E_v$ and $c_r \geq 1$. Then there exists $\epsilon_k > 0$ such that*

$$\liminf_{x \to p_k} \frac{\|T_{c_r}(x) - p_k\|}{\|x - p_k\|} \geq 1 + \epsilon_k.$$

In particular, $\{T_c(x^{(r)})\}_{r\in\mathbb{N}}$ cannot converge to p_k.

Proof: For $x \neq p_k$, we obtain

$$T_{c_r}(x) = (1 - c_r)x + c_r S(x)^{-1}\left(\lambda K^{\mathrm{T}} f + \sum_{i=1}^{M} \frac{w_i p_i}{\|x - p_i\|}\right)$$

and further

$$\begin{aligned} T_{c_r}(x) - p_k &= (1-c_r)(x-p_k) + c_r S(x)^{-1}\tilde{G}(x), \\ &= (1-c_r)(x-p_k) + c_r\|x-p_k\|(\|x-p_k\|S(x))^{-1}\tilde{G}(x) \end{aligned}$$

where

$$\tilde{G}(x) := \lambda K^{\mathrm{T}}(f - Kp_k) + \sum_{\substack{i=1 \\ i\neq k}}^{M} \frac{w_i(p_i - p_k)}{\|x - p_i\|}.$$

Since $\tilde{G}(p_k) = -G_k$ and $\|x-p_k\|S(x) = \|x-p_k\|\big(\lambda K^{\mathrm{T}}K + \sum_{\substack{i=1 \\ i\neq k}}^{M} \frac{w_i}{\|x-p_i\|} I\big) + w_k I$ we conclude that

$$\liminf_{x\to p_k} \frac{\|T_{c_r}(x) - p_k\|}{\|x - p_k\|} \geq \left| |1-c_r| - c_r \frac{\|G_k\|}{w_k} \right|$$

Since $p_k \notin \operatorname{argmin} E_v$ there exists $\epsilon_k > 0$ such that $\frac{\|G_k\|}{w_k} \geq 1 + \epsilon_k \geq 1 + \frac{\epsilon_k}{c_r}$, cf. (20). Using $c_r \geq 1$, this implies

$$\liminf_{x\to p_k} \frac{\|T_{c_r}(x) - p_k\|}{\|x - p_k\|} \geq 1 + c_r(\frac{\|G_k\|}{w_k} - 1) = 1 + \epsilon_k$$

and we are done. □

Theorem A.3. *Let $K = I$. Assume that E_v has a unique minimizer $\hat{x}$. Then, for $1 \leq c_r < 2$ and any $x^{(0)} \in \mathbb{R}^d$, the sequence $\{T_{c_r}(x^{(r)})\}_{r\in\mathbb{N}}$ converges to $\hat{x}$.*

Proof: If $x^{(r)} = \hat{x}$ for some r we are done by Lemma A.1. Assume now that $\{x^{(r)}\}$ is an infinite sequence. By Lemma A.1 and since E_v is coercive we know that $\{x^{(r)}\}$ is bounded. Hence there exists a convergent subsequence $\{x^{(r_j)}\}$ of $\{x^{(r)}\}$. Let $\lim_{j\to\infty} x^{(r_j)} = \tilde{x}$. It remains to show that $\tilde{x} = \hat{x}$. Since $\{E_v(x^{(r)})\}$ is monotone decreasing and bounded, it is convergent, say

$$\lim_{r\to\infty} E_v(x^{(r)}) = M.$$

Thus,

$$M = \lim_{j\to\infty} E_v(x^{(r_j)}) = \lim_{j\to\infty} E_v(x^{(r_j+1)}) = \lim_{j\to\infty} E_v(T_{c_r}(x^{(r_j)})).$$

If $\tilde{x} \notin P$, the T_{c_r} is continuous at $\tilde{x}$, i.e. $\lim_{j\to\infty} T_{c_r}(x^{(r_j)}) = T_{c_r}(\tilde{x})$. By continuity of E_v we obtain $M = E_v(\tilde{x}) = E_v(T_{c_r}(\tilde{x}))$. By Lemma A.1 this is only possible if $\tilde{x} = \hat{x}$.

If $\tilde{x} = p_k \in P$ and $\tilde{x} \neq \hat{x}$. By Lemma A.2, the whole sequence $\{x^{(r)}\}$ cannot converge to p_k. Therefore, there exist (one or more) subsequences converging to some point $p_j \neq p_k$. (Convergence of subsequences to points not contained in P is not possible since by the above considerations the only other cluster point could be $\hat{x}$, but $E_v(\hat{x}) < E_v(p_k)$ which contradicts Lemma A.1.) Thus, for a small enough constant $\varepsilon > 0$, all but a finite number of points lie within ε-balls around some points $p_j \in P$. Here, we choose $\varepsilon < \min_{i\neq j} \|p_i - p_j\|/3$ which implies that there is an index $n(\varepsilon) \in \mathbb{N}$ such that all $x^{(r)}$, $r \geq n(\varepsilon)$, lie within these balls. Then,

there exits an infinite number of indices $s \geq n(\varepsilon)$ with $\|x^{(s+1)} - p_k\| > 2\varepsilon$ and $\|x^{(s)} - p_k\| < \varepsilon$. By (42) in the proof of Lemma A.1 we have that

$$E_v(x^{(s)}) - E_v(x^{(s+1)}) > \frac{w_k}{2} \frac{(\|x^{(s+1)} - p_k\| - \|x^{(s)} - p_k\|)^2}{\|x^{(s)} - p_k\|} \geq \frac{w_k}{2} \frac{\varepsilon^2}{2\varepsilon}.$$

This is a contradiction to the convergence of $E_v(x^{(s)})$. □

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3 – 51, 2003.

[2] J. Astola, P. Haavisto, and Y. Neuvo. Vector median filters. *Proceedings of the IEEE*, 78(4):678 – 689, 1990.

[3] F. Becker. Matrix-valued filters as convex programs. *Diplomarbeit, Universität Mannheim*, 2004.

[4] B. Berkels, G. Linkmann, and M. Rumpf. An SL(2) invariant shape median. *Journal of Mathematical Imaging and Vision*, 37(2):85 – 97, 2010.

[5] K. Borsuk. *Multidimensional Analytic Geometry*. Panstwowe Wydawnictwo Naukowe, Warszawa, Poland, 1969.

[6] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, 20(4):1956 – 1982, 2010.

[7] V. Caselles, G. Sapiro, and D. Chung. Vector median filters, inf-sup operations, and coupled pde's: Theoretical connections. *Journal of Mathematical Imaging and Vision*, 12(2):109–119, 2000.

[8] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Preprint, University of Graz, May 2010.

[9] P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6), 2008.

[10] I. Daubechies, M. Fornasier, and I. Loris. Accelerated projected gradient methods for linear inverse problems with sparsity constraints. *The Journal of Fourier Analysis and Applications*, 14(5-6):764–792, 2008.

[11] C. Davis. All convex invariant functions of Hermitian matrices. *Archive in Mathematics*, 8:276–278, 1957.

[12] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *Annals of Statistics*, 26(3):879–921, 1998.

[13] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the l1-ball for learning in high dimensions. In *ICML '08 Proceedings of the 25th International Conference on Machine Learning*, ACM New York, 2008.

[14] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293 – 318, 1992.

[15] E. Esser. Applications of Lagrangian based alternating direction methods and connections to split Bregman. *CAM Report 09-31, UCLA*, 2009.

[16] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for TV minimization. CAM Report 09-67, UCLA, 2009.

[17] 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*, pages 281–305, Interlaken, Switzerland, June 1987.

[18] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*, chapter IX, pages 299 – 340. North-Holland, Amsterdam, 1983.

[19] D. Goldfarb, Z. Wen, and W. Yin. A curvilinear search method for p-harmonic flows on spheres. *SIAM Journal on Imaging Sciences*, 2(1):84 – 109, 2009.

[20] T. Goldstein and S. Osher. The split Bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323 – 343, 2009.

[21] D. Karakos and P. E. Trahanias. Generalized multichannel image-filtering structures. *IEEE Transactions on Image Processing*, 6(7):1038 – 1044, 1997.

[22] T. Kärkkäinen and S. Äyrämö. On computation of spatial median for robust data mining. In R. Schilling, W. Haase, J. Periaux, H. Baier, and G. Bugeda, editors, *Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems, EUROGEN*, Munich, 2005.

[23] I. N. Katz. Local convergence in Fermat's problem. *Mathematical Programming*, 6(1):89 – 104, 1974.

[24] R. Kimmel and N. Sochen. Orientation diffusion or how to comb a porcupine. *Journal of Visual Communication and Image Representation*, 13:238–248, 2002.

[25] H. Kuhn. A note on Fermat's problem. *Mathematical Programming*, 4:98 – 107, 1973.

[26] M. Lawrence and J. Ostresh. On the convergence of a class of iterative methods for solving the Weber location problem. *Operations Research*, 26(4):597 – 609, 1987.

[27] Y. Li and S. Osher. A new median formula with applications to PDE based denoising. *Communications in Mathematical Sciences*, 7(3):741 – 754, 2009.

[28] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193 – 228, 1998.

[29] H. Mittelmann. An independent benchmarking of SDP and SOCP solvers. *Mathematical Programming*, 95(2):407 – 430, 2003.

[30] P. Mrázek and J. Weickert. Rotationally invariant wavelet shrinkage. In B. Michaelis and G. Krell, editors, *Pattern Recognition*, volume 2781 of *Lecture Notes in Computer Science*, pages 156–163, Berlin, 2003. Springer.

[31] S. Setzer. Operator splittings, Bregman methods and frame shrinkage in image processing. *International Journal of Computer Vision*, 2010 to appear.

[32] C. Small. A survey to multidimensional medians. *International Statistical Review*, 58(3):263–277, 1990.

[33] T. Teuber, G. Steidl, P. Gwosdek, C. Schmaltz, and J. Weickert. Dithering by differences of convex functions. *SIAM Journal on Imaging Science*, to appear, 2010.

[34] P. E. Trahanias and A. N. Venetsanopoulos. Vector directional filters - a new class of multichannel image processing filters. *IEEE Transactions on Image Processing*, 2(4):528 – 534, 1993.

[35] D. Tschumperlé and R. Deriche. Regularization of orthonormal vector sets using coupled PDEs. In *Proceedings of the first IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pages 3 – 10. 2001.

[36] L. A. Vese and S. J. Osher. Numerical methods for p-harmonic flows and applications to image processing. *SIAM Journal on Numerical Analysis*, 40(6):2085–2104, 2002.

[37] H. Voss and U. Eckhardt. Linear convergence of generalized Weiszfeld's method. *Computing*, 25:243 – 251, 1980.

[38] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[39] G. A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:33–45, 1992.

[40] E. Weiszfeld. Sur le point pour lequel les sommes des distances de n points donnés et minimum. *Tôhoku Mathematical Journal*, 43:355 – 386, 1937.

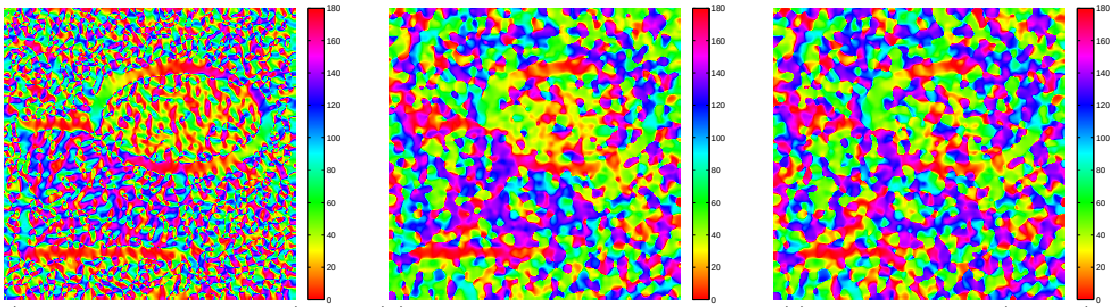
[41] M. Welk, F. Becker, C. Schnörr, and J. Weickert. Matrix-valued filters as convex programs. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale Space and PDE Methods in Computer Vision*, volume 3459 of *Lecture Notes in Computer Science*, pages 204–216, Berlin, 2005. Springer.

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

[43] M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. *Applied and Computational Harmonic Analysis*, 24:195–224, 2008.

[44] M. Welk, J. Weickert, F. Becker, C. Schnörr, C. Feddern, and B. Burgeth. Median and related local filters for tensor-valued images. *Signal Processing*, 87(2):291 – 308, 2007.

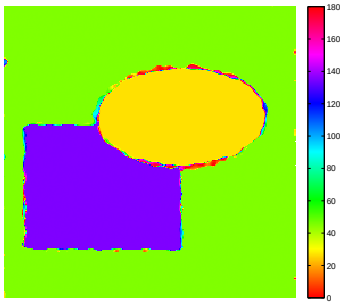
[45] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, UCLA, 2008.



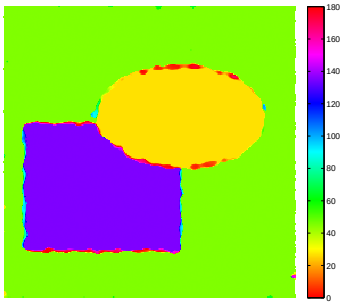
(a) vector mean with $p_{i,j}$ ($\rho = 2.5$)

(b) vector mean with modified vectors $\tilde{g}(p_{i,j})$ ($\rho = 2.5$)

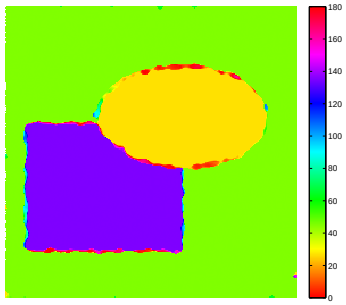
(c) structure tensor ($\rho = 2.5$)



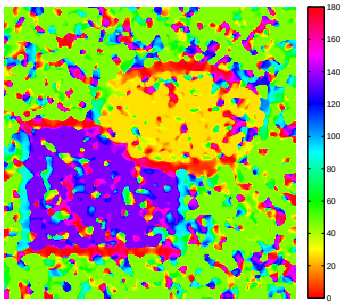
(d) vector median with $p_{i,j}$ ($\rho = 2.5$)



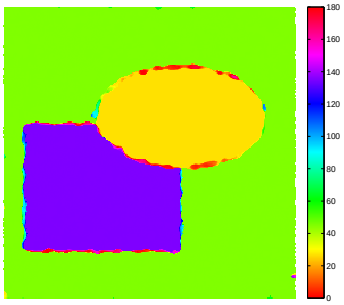
(e) vector median with modified vectors $\tilde{g}(p_{i,j})$ ($\rho = 2.5$)



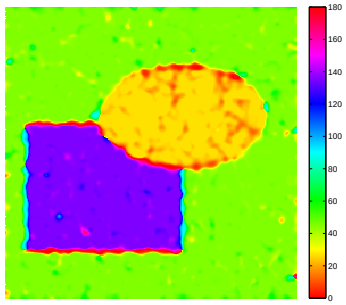
(f) spectral matrix median ($\rho = 2.5$)



(g) Frobenius matrix median ($\rho = 2.5$)

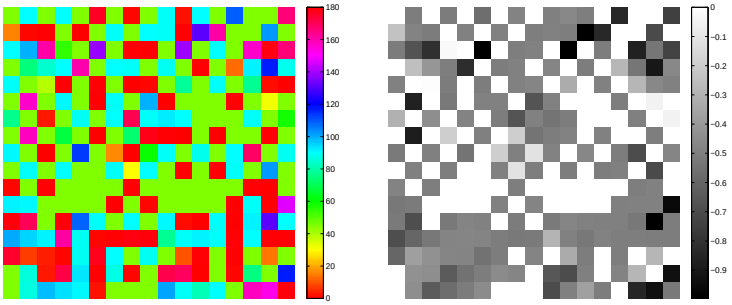


(h) nuclear matrix median ($\rho = 2.5$)



(i) spectral matrix median with normalized gradients ($\rho = 2.5$)

Figure 8: Results for the impulse noise corrupted image of Fig. 7.



(a) zoom into Fig. 7 (right) at pixel $(45, 45)$

(b) $|\lambda_1+\lambda_2|-(\lambda_1-\lambda_2)$ for $\hat{X}-P_{i,j}$

Figure 9: Details of the matrix median filter with the nuclear norm for the filtering of pixel $(45, 45)$ of Fig. 7: The figure shows that for all $P_{i,j}$ the eigenvalues λ_1, λ_2 of $\hat{X} - P_{i,j}$ fulfill $\lambda_1 - \lambda_2 > |\lambda_1 + \lambda_2|$ or $\lambda_1 - \lambda_2 \approx |\lambda_1 + \lambda_2|$.

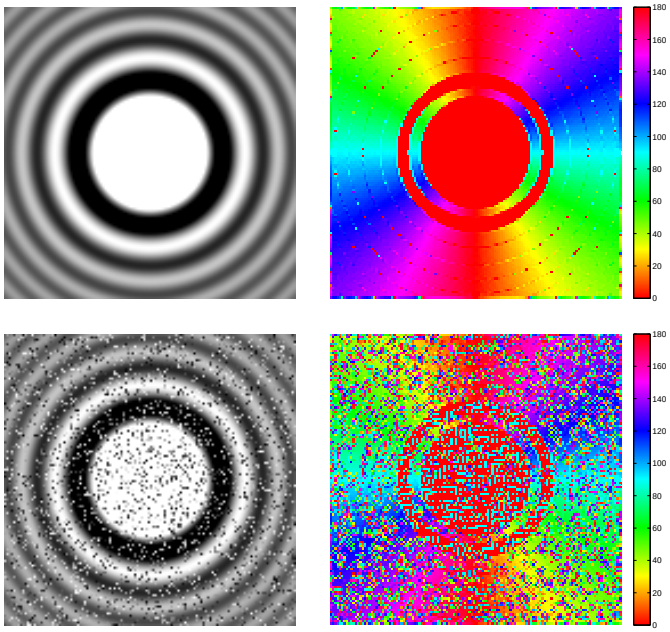
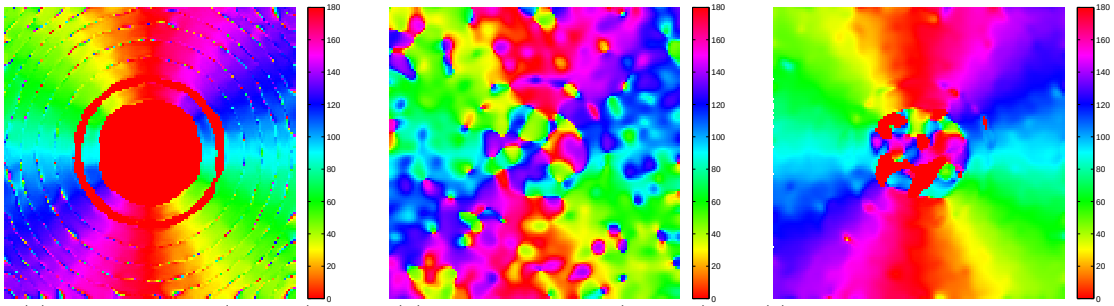


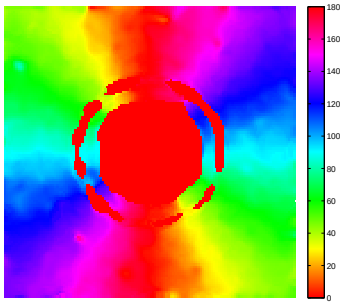
Figure 10: Top: Test image of size 128×128 (left) and the angles of the gradients (right). Bottom: Test image corrupted by 20% impulse noise (left) and corresponding gradient directions (right). In constant areas the gradient angle is set to 0.



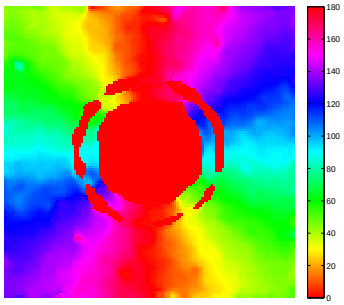
(a) vector median ($\rho = 2.5$)

(b) structure tensor ($\rho = 2.5$)

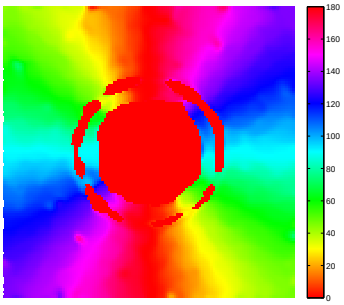
(c) Frobenius matrix median ($\rho = 2.5$)



(d) nuclear matrix median ($\rho = 2.5$)



(e) spectral matrix median ($\rho = 2.5$)



(f) spectral matrix median with normalized gradients ($\rho = 2.5$)

Figure 11: Results for the impulse noise corrupted image of Fig. 10 (bottom).