Denoising Sphere-Valued Data by relaxed Total Variation Regularization

Robert Beinert Jonas Bresch

TECHNISCHE UNIVERSITÄT BERLIN

- **Contribution.** Denoising technique via rewriting and relaxing the objective function and constraints for highly nonconvex problems. For data, which is originally located on the sphere. We regularize the squared-L_2 norm with the anisotropic TV-norm, i.e. L_1 norm.
- **Prior work.** Regularization with squared-L_2 norm and similar rewriting and relaxation technique [2].

Problem setting

We are interested in the **restoration** of a sphere valued signal $\boldsymbol{x} := (\boldsymbol{x}_n)_{n \in V}$ with

$$\boldsymbol{x}_n \in \mathbb{S}_{d-1} := \{\xi \in \mathbb{R}^d : \|\xi\|_2 = 1\} \subset \mathbb{R}^d$$

from a **disturbed signal** $\boldsymbol{y} = (\boldsymbol{y}_n)_{n \in V}$ with $\boldsymbol{y}_n \in \mathbb{S}_{d-1}$ or, more generally, $\boldsymbol{y}_n \in \mathbb{R}^d$.
The data is supported on a connected, undirected graph $G = (V, E)$.

- $V := \{1, \dots, N\}$ denotes the set of vertices with $N := |V|$ elements
- $E := \{(n, m) : n < m\} \subset V \times V$ the set of edges with $M := |E|$ elements, encoding the data structure.

Regularizing with the squared-L_2 norm [2]

$$\operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{S}_{d-1}^N} \frac{1}{2} \sum_{n \in V} \|\boldsymbol{x}_n - \boldsymbol{y}_n\|_2^2 + \frac{\lambda}{2} \sum_{(n,m) \in E} \|\boldsymbol{x}_n - \boldsymbol{x}_m\|_2^2, \tag{1}$$

Rewriting and relaxation

By the property of the squared-L_2 norm, we have for both parts in (1) respectively

$$\|\boldsymbol{v} - \boldsymbol{w}\|_2^2 = \begin{cases} -2\langle \boldsymbol{v}, \boldsymbol{w} \rangle + \text{const} & : (\boldsymbol{v}, \boldsymbol{w}) \in \mathbb{S}_{d-1}^2, \\ -2\langle \boldsymbol{v}, \boldsymbol{w} \rangle + \text{const}(\boldsymbol{w}) & : (\boldsymbol{v}, \boldsymbol{w}) \in \mathbb{S}_{d-1} \times \mathbb{R}^d. \end{cases} \tag{2}$$

Rewritten optimization problem with linear objectiive

Incorporating (2) and using parameters $\ell_{(n,m)} \in \mathbb{R}$ yields

$$\mathcal{L} : \mathbb{R}^{d \times N} \times \mathbb{R}^M \mapsto \mathbb{R}, (\boldsymbol{x}, \boldsymbol{\ell}) \mapsto -\sum_{n \in V} \langle \boldsymbol{x}_n, \boldsymbol{y}_n \rangle - \lambda \sum_{(n,m) \in E} \ell_{(n,m)},$$

Moreover

$$(1) = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{S}_{d-1}^N, \boldsymbol{\ell} \in \mathbb{R}^M} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\ell}) \quad \text{s.t.} \quad \ell_{(n,m)} = \langle \boldsymbol{x}_n, \boldsymbol{x}_m \rangle \quad \text{for all} \quad (n, m) \in E.$$

Remark. The latter optimization problem is still non-convex.

Proposition [2, Lem 7]. It holds $\boldsymbol{x}_n, \boldsymbol{x}_m \in \mathbb{S}_{d-1}$ and $\ell_{(n,m)} = \langle \boldsymbol{x}_n, \boldsymbol{x}_m \rangle$ if and only if

$$\boldsymbol{Q}_{(n,m)} \succcurlyeq 0 \quad \text{and} \quad \operatorname{rk}(\boldsymbol{Q}_{(n,m)}) = d, \quad \text{where} \quad \boldsymbol{Q}_{(n,m)} := \begin{bmatrix} \boldsymbol{I}_d & \boldsymbol{x}_n & \boldsymbol{x}_m \\ \boldsymbol{x}_n^\top & 1 & \ell_{(n,m)} \\ \boldsymbol{x}_m^\top & \ell_{(n,m)} & 1 \end{bmatrix} \in \mathbb{R}^{d+2 \times d+2}.$$

Relaxed convex optimization problem [2]

Neglecting the rank constraints in the Proposition yields **our relaxed convex model**:

$$\operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^{d \times N}, \boldsymbol{\ell} \in \mathbb{R}^M} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\ell}) \quad \text{s.t.} \quad \boldsymbol{Q}_{(n,m)} \succcurlyeq 0 \quad \text{for all} \quad (n, m) \in E.$$

Regularizing with the L_1 norm [1]

- The squared-L_2 (Tikhonov) regularization, is suitable for smooth signals.
- We want to **recover piecewise constant signals** from noisy measurements.
- We **replace the squared-L_2 norm** of the regularizer in (1) **by the 1-norm** yielding the total variation (TV) regularization:

$$\operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{S}_{d-1}^N} \frac{1}{2} \sum_{n \in V} \|\boldsymbol{x}_n - \boldsymbol{y}_n\|_2^2 + \lambda \mathrm{TV}(\boldsymbol{x}) \quad \text{with} \quad \mathrm{TV}(\boldsymbol{x}) := \sum_{(n,m) \in E} \|\boldsymbol{x}_n - \boldsymbol{x}_m\|_1. \tag{3}$$

Rewritten optimization problem with linear data fidelity term

Using (2) again, we rewrite the objective of (3) into

$$\mathcal{K} : \mathbb{R}^{d \times N} \to \mathbb{R}, \boldsymbol{x} \mapsto -\sum_{n \in V} \langle \boldsymbol{x}_n, \boldsymbol{y}_n \rangle + \lambda \mathrm{TV}(\boldsymbol{x}) \quad \text{with} \quad (3) = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{S}_{d-1}^N} \mathcal{K}(\boldsymbol{x}).$$

Convexifying the sphere-valued domain $\mathbb{S}_{d-1}^N$, we propose **our relaxed convex problem**:

$$\operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^{d \times N}} \mathcal{K}(\boldsymbol{x}) \quad \text{s.t.} \quad \boldsymbol{x}_n \in \mathbb{B}_d \quad \text{for all} \quad n \in V. \tag{4}$$

References

[1] Robert Beinert and Jonas Bresch. *Denoising Sphere-Valued Data by Relaxed Total Variation Regularization*. 2024. arXiv: 2404.13181 [math.NA]. URL: https://arxiv.org/abs/2404.13181.

[2] Robert Beinert, Jonas Bresch, and Gabriele Steidl. "Denoising of sphere- and SO(3)-valued data by relaxed Tikhonov regularization". In: *Inverse Problems and Imaging* (2024). ISSN: 1930-8337. DOI: 10.3934/ipi.2024026. URL: https://www.aimsciences.org/article/id/66728cba5a42b314c5c01e82.

[3] Ronny Bergmann et al. "Second Order Differences of Cyclic Data and Applications in Variational Denoising". In: *SIAM J. Imaging Sci.* 7.4 (Jan. 2014), pp. 2916–2953. DOI: 10.1137/140969993. URL: https://doi.org/10.1137%2F140969993.

[4] Rustum Choksi, Yves van Gennip, and Adam Oberman. "Anisotropic Total Variation Regularized L^1-Approximation and Denoising/Deblurring of 2D Bar Codes". In: *Inverse Probl. Imaging* 5.3 (2011), pp. 591–617. ISSN: 1930-8337. DOI: 10.3934/ipi.2011.5.591. URL: https://www.aimsciences.org/article/id/497753d8-5d6d-447d-8952-fe9f3da0c125.

[5] Laurent Condat. "A Direct Algorithm for 1D Total Variation Denoising". In: *IEEE Signal Process. Lett.* 20.11 (2013), pp. 1054–1057. DOI: 10.1109/LSP.2013.2278339. URL: https://hal.science/hal-00675043.

[6] Robin Kenis, Emanuel Laude, and Panagiotis Patrinos. *arXiv:2308.00079*. 2023. arXiv: 2308.00079 [math.OC].

Tightness for the binary convexification

We introduce the characteristic χ_η regarding the level $\eta \in [-1, 1]$ of a signal $\boldsymbol{x} := (\boldsymbol{x}_n)_{n \in V} \in \mathbb{B}_1^N, \mathbb{B}_1 := \operatorname{conv}(\mathbb{S}_0) = [-1, 1]$ by

$$\chi_\eta(\boldsymbol{x}) := (\chi_\eta(\boldsymbol{x}_n))_{n \in V} \quad \text{with} \quad \chi_\eta(\boldsymbol{x}_n) := \begin{cases} 1 & \text{if } \boldsymbol{x}_n > \eta, \\ -1 & \text{if } \boldsymbol{x}_n \leq \eta. \end{cases}$$

For any $\boldsymbol{x} := (\boldsymbol{x}_n)_{n \in V} \in \mathbb{B}_1^N$, the characteristic $\chi_\eta(\boldsymbol{x})$ is a binary signal, i.e. $\chi_\eta(\boldsymbol{x}) \in \mathbb{S}_0^N$.

Lemma. For $\boldsymbol{x}_n, \boldsymbol{x}_m \in \mathbb{B}_1$, it holds $|\boldsymbol{x}_n - \boldsymbol{x}_m| = \frac{1}{2} \int_{-1}^{1} |\chi_\eta(\boldsymbol{x}_n) - \chi_\eta(\boldsymbol{x}_m)| \, \mathrm{d}\eta$.

Theorem. (Tightness of the binary problem)

Let $\boldsymbol{x}^* \in \mathbb{B}_1^N$ be a solution of (4) for $d = 1$. Then $\chi_\eta(\boldsymbol{x}^*)$ is a solution of (3) for almost all $\eta \in [-1, 1]$.

Numerical experiments — Binay-, circle- and SO(3)-valued signals

For our specific setting, ADMM reads as below, where the proximation of TV is defined as

$$\operatorname{prox}_{\mathrm{TV},\gamma}(\boldsymbol{z}) := \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^{d \times N}} \Big\{ \mathrm{TV}(\boldsymbol{x}) + \frac{1}{2\gamma} \sum_{n \in V} \|\boldsymbol{x}_n - \boldsymbol{z}_n\|_2^2 \Big\}, \tag{5}$$

and $\operatorname{proj}_{\mathbb{B}_d^N}$ denotes to orthogonal projection onto $\mathbb{B}_d^N$.

Algorithm: ADMM-TV (ADMM to solve (4)).
Choose: $\boldsymbol{x}^{(0)} = \boldsymbol{u}^{(0)} = \boldsymbol{z}^{(0)} = \boldsymbol{0} \in \mathbb{R}^{d \times N}$, step size $\rho > 0$ and TV parameter $\lambda > 0$.
For $i \in \mathbb{N}$ **do:**

$$\boldsymbol{x}^{(i+1)} = \operatorname{prox}_{\mathcal{K}, \frac{1}{\rho}}(\boldsymbol{u}^{(i)} - \boldsymbol{z}^{(i)}) = \operatorname{prox}_{\mathrm{TV}, \frac{\lambda}{\rho}}(\boldsymbol{u}^{(i)} - \boldsymbol{z}^{(i)} + \boldsymbol{y}\rho^{-1}),$$

$$\boldsymbol{u}^{(i+1)} = \operatorname{prox}_{\iota_{\mathbb{B}_d^N}}(\boldsymbol{x}^{(i+1)} + \boldsymbol{z}^{(i)}) = \operatorname{proj}_{\mathbb{B}_d^N}(\boldsymbol{x}^{(i+1)} + \boldsymbol{z}^{(i)}), \qquad \boldsymbol{z}^{(i+1)} = \boldsymbol{z}^{(i)} + \boldsymbol{x}^{(i+1)} - \boldsymbol{u}^{(i+1)}.$$

Table 1. Averages for 50 randomly generated QR codes for different noise levels. One specific instance is illustrated in Fig. 1 (ANISO-TV [4], fast-TV [5]).

$\sigma =$	Algorithm	signal errors MSE	signal errors MIoU	λ	time (sec.)	distance to sphere
$\sqrt{2}\frac{5}{10}$	fast-TV	**0.00036**	**0.99734**	0.7	< 0.1	0.07879
	ANISO-TV	0.00112	0.97495	1.9	1.8	0.00122
	ADMM-TV	**0.00036**	**0.99734**	0.7	1.1	**0.00000**
$\sqrt{2}\frac{7}{10}$	fast-TV	0.00069	0.99018	1.0	< 0.1	0.10021
	ANISO-TV	0.00190	0.92942	2.4	1.9	0.00732
	ADMM-TV	**0.00069**	**0.99030**	1.0	1.1	**0.00000**
$\sqrt{2}\frac{9}{10}$	fast-TV	0.00106	0.97741	1.7	< 0.1	0.12263
	ANISO-TV	0.00263	0.86956	2.9	2.1	0.01445
	ADMM-TV	**0.00106**	**0.97753**	1.7	1.2	**0.00000**

Table 2. Averages for 20 randomly generated noisy instances of the ground truth in Fig. 2 for different noise levels (CPPA-TV [3], fast-TV [5]).

$\kappa =$	Algorithm	signal error MSE	λ	time (sec.)	distance to sphere
50	CPPA-TV	0.00076	0.15	128.2	–
	fast-TV	0.00056	0.25	< 0.1	0.00055
	ADMM-TV	**0.00043**	0.2	4.3	**0.00000**
20	CPPA-TV	0.00198	0.25	166.1	–
	fast-TV	0.00199	0.25	< 0.1	0.00098
	ADMM-TV	**0.00107**	0.3	6.2	**0.00000**
10	CPPA-TV	0.00409	0.55	191.7	–
	fast-TV	0.00388	0.25	< 0.1	0.00155
	ADMM-TV	**0.00218**	0.45	10.5	**0.00000**

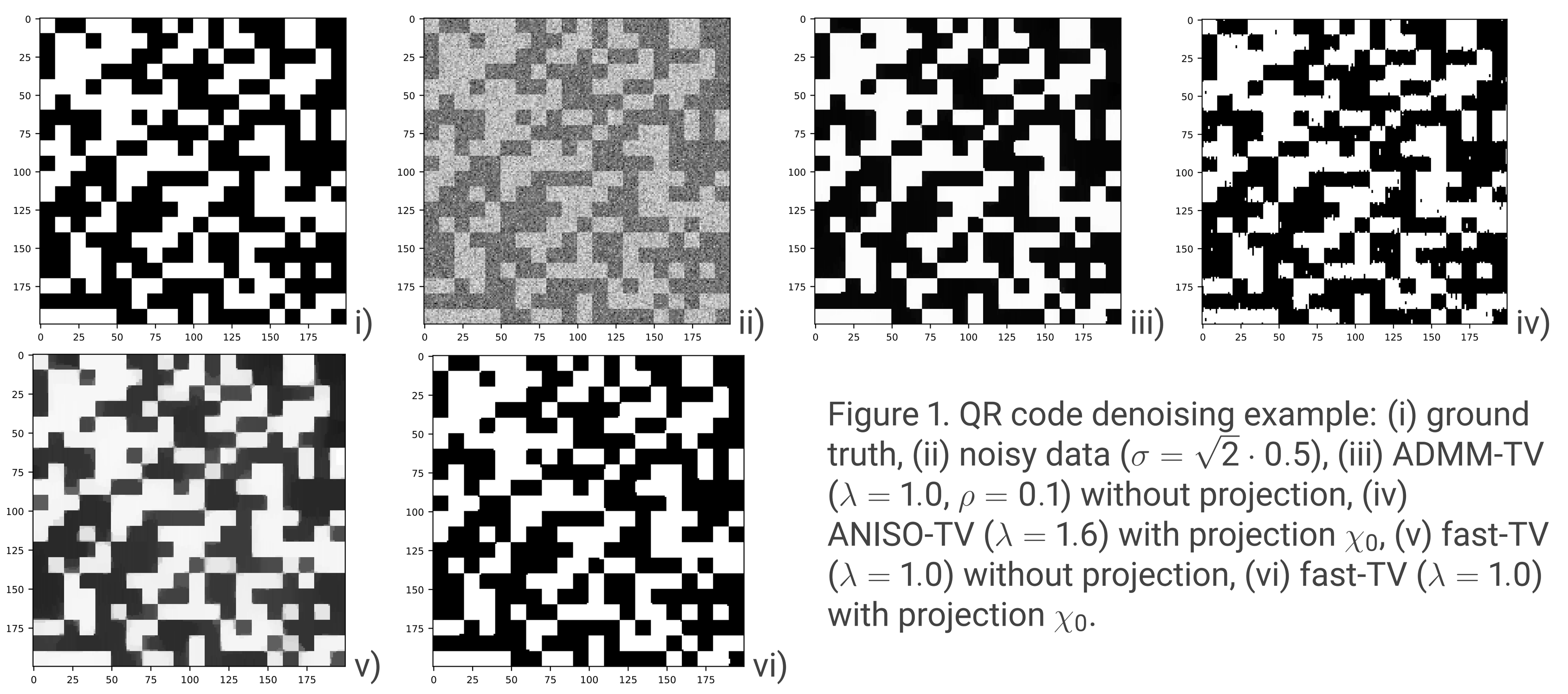

Figure 1. QR code denoising example: (i) ground truth, (ii) noisy data ($\sigma = \sqrt{2} \cdot 0.5$), (iii) ADMM-TV ($\lambda = 1.0$, $\rho = 0.1$) without final projection, (iv) ANISO-TV ($\lambda = 1.6$) with projection χ_0, (v) fast-TV ($\lambda = 1.0$) without projection, (vi) fast-TV ($\lambda = 1.0$) with projection χ_0.

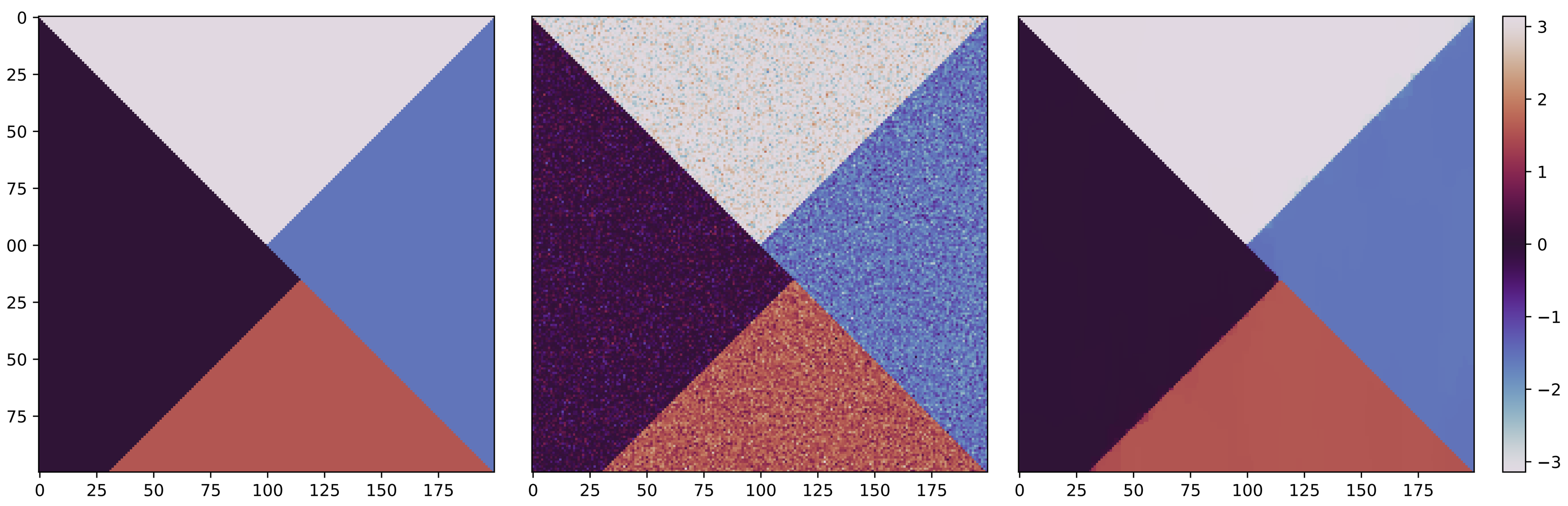

Figure 2. Toy-data example following [6] for $\mathbb{S}_1$-image denoising (from left to right): (i) ground truth, (ii) noisy measurement generated by the von Mises–Fisher distribution with $\kappa = 10$, (iii) solution via ADMM-TV ($\lambda = 0.55$, $\rho = 10$) without final projection.

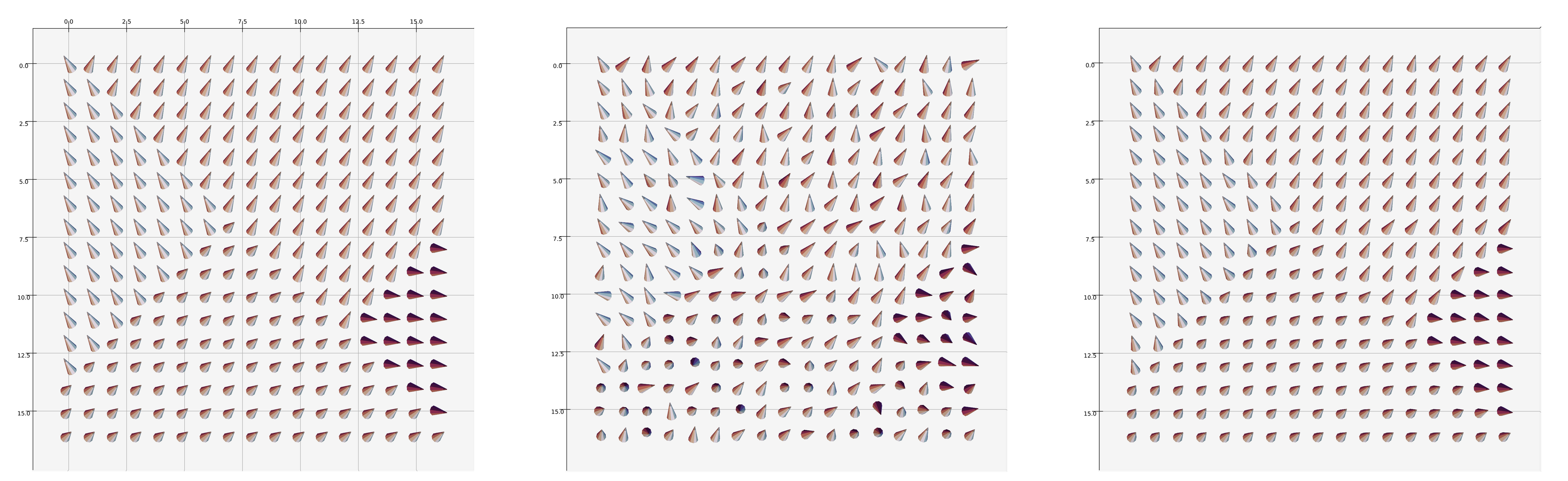

Figure 3. Denoising of a synthetic SO(3)-image (90 × 90 pixels) inspired by [6], following [2] (from left to right, downsampled results): (i) ground truth, (ii) noisy measurment generated by the von Mises–Fisher distribution with capacity $\kappa_1 = 10$ for the rotation angles and $\kappa_2 = 10$ for the rotation axis. (iii) solution via ADMM-TV ($\lambda = 0.10$, $\rho = 100$) without final projection and with MSE $5.889 \cdot 10^{-3}$ and averaged distance to the unit quaternions $1.45 \cdot 10^{-7}$.

Conclusion

- Fast and efficient approach with an incrementally speed-up.
- Tightness for the convexification, if $d = 1$.