On Sparsity Maximization in Tomographic Particle Image Reconstruction[⋆]

Stefania Petra[1], Andreas Schröder[2], Bernhard Wieneke[3], Christoph Schnörr[1]

[1] Image and Pattern Analysis Group, Heidelberg Collaboratory for Image Processing, University of Heidelberg, Germany, {petra,schnoerr}@math.uni-heidelberg.de
[2] Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Göttingen, Germany, andreas.schroeder@dlr.de
[3] LaVision GmbH, Göttingen, Germany, bwieneke@lavision.de

Abstract. This work focuses on tomographic image reconstruction in experimental fluid mechanics (TomoPIV), a recently established 3D particle image velocimetry technique. Corresponding 2D image sequences (projections) and the 3D reconstruction via tomographical methods provides the basis for estimating turbulent flows and related flow patterns through image processing. TomoPIV employs undersampling to make the high-speed imaging process feasible, resulting in an ill-posed image reconstruction problem. We address the corresponding basic problems involved and point out promising optimization criteria for reconstruction based on sparsity maximization, that perform favorably in comparison to classical algebraic methods currently in use for TomoPIV.

1 Introduction

Recent developments of *particle image velocimetry (PIV)* techniques [14] allow to capture the flow velocity of large and unsteady flow fields instantaneously. Among the different 3D techniques presently available for measuring velocities of fluids, *tomographic particle image velocimetry (TomoPIV)* [6] has recently received most attention, due to its increased seeding density with respect to other 3D PIV methods. This, in turn, enables high-resolution velocity field estimates of turbulent flows by means of a *cross correlation technique* [15]. TomoPIV is based on a multiple camera-system, three-dimensional volume illumination and subsequent 3D reconstruction, cf. [6]. In this paper we consider the essential step of this technique, the 3D reconstruction of the particle volume functions from few projections, which amounts to solve an underdetermined system of linear equations of the form

$$Ax = b. \tag{1}$$

Such systems, disregarding for the moment the inconsistent case, have infinitely many solutions. Classical regularization approaches with strictly convex objective functions are therefore applied – cf. [6] and section 3 below – in order to

single out a particular solution. However, commonly choices such as maximizing entropy or minimizing the ℓ_2 norm can be problematic within the TomoPIV setting, since both approaches provide much too dense "solutions", i.e. solutions x with most entries different from zero, resulting in blurry "ghost particles". This is clearly detrimental for subsequent motion estimation. The objective of this paper is to point out better alternatives.

Organization, After sketching the imaging process in section 2, we recall the classical solution concepts that are currently in use for TomoPIV, in section 3. Better alternatives based on sparsity-based regularization are the subject of section 4. Our arguments are confirmed and illustrated by numerical experiments in section 5.

2 Discretization

A basic assumption in image reconstruction is that the image I to be reconstructed can be approximated by a linear combination of *basis functions* $\mathcal{B}_j$,

$$I(z) \approx \hat{I}(z) = \sum_{j=1}^{n} x_j \mathcal{B}_j(z), \quad \forall z \in \Omega \subset \mathbb{R}^3 \,, \tag{2}$$

where Ω denotes the volume of interest and $\hat{I}$ is the *digitization* of I. The main task is to estimate the weights x_j from the recorded 2D images, corresponding to basis functions located at a Cartesian equidistant 3D grid p_j, $j = 1, \ldots, n$. We consider Gaussian-type basis functions ("blobs"), an alternative to the classical voxels, of the form

$$\mathcal{B}_j(z) = e^{-\frac{\|z - p_j\|_2^2}{2\sigma^2}}, \quad \text{for } z \in \mathbb{R}^3 : \|z - p_j\|_2 \leq r \,, \tag{3}$$

or value 0, if $\|z - p_j\|_2 > r$. See Fig. 1, left, for the 2D case.

The choice of a Gaussian-type basis function is justified in the TomoPIV setting, since a particle projection in all directions results in a so-called *diffraction spot*. Figure 1 (right) shows a typical projection of a 3D particle distribution.

Based on geometrical optics, the recorded pixel intensity is the object intensity integrated along the corresponding line of sight, obtained from a calibration procedure. Thus the i-th measurement obeys

$$b_i :\approx \int_{L_i} I(z)dz \approx \int_{L_i} \hat{I}(z)dz = \sum_{j=1}^{n} x_j \underbrace{\int_{L_i} \mathcal{B}_j(z)dz}_{:=a_{ij}} \,, \tag{4}$$

where L_i is a line or more realistically a thin cone of light. Compare Fig. 1, left. Due to inaccuracies of the physical measurements the set of all projection "rays" yields an approximate linear system (1), where A is the *projection* matrix and b the *measurement* vector.
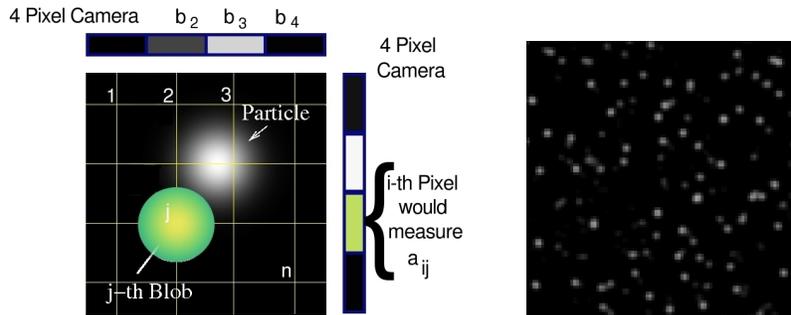
Fig. 1. Left: Discretization by covering the domain Ω with compactly supported basis functions. **Right:** Projection of a 3D particle distribution. The volume function has to be reconstructed from 4–6 such measurements.

3 Classical Solution Concepts

Throughout this paper we concentrate on ill-posedness of (1). We have to handle a underdetermined system with $m << n$, having a rank-deficient and very sparse coefficient matrix $A \in \mathbb{R}^{m \times n}$. The sparsity structure of A is due to the fact that only a few basis function are on the line of sight of each particular camera pixel. By the nature of the problem, we always have $a_{ij} \geq 0$ and $b_i \geq 0, \forall i, j$. Typical TomoPIV measurements b_i are nonzero to $\approx 5\%$, see Fig. 1, left.

To restrict the feasible set (1) we can add nonnegativity constraints and look for elements in the polyhedral set

$$\mathcal{F} := \{x : Ax = b, x \geq 0\}. \tag{5}$$

If $b_i = 0$, then we can remove all columns of A, whose i-th entry is positive, as well as the i-th row. This procedure will lead to a "equivalent" feasible set of reduced dimensionality $\mathcal{F}_r := \{x \in \mathbb{R}^{n_r} : A_r x = b_r, x \geq 0\}$, where $A_r \in \mathbb{R}^{m_r \times n_r}$ is the reduced projection matrix and $b_r > 0$ the new data vector. All x_j variables corresponding to removed columns in A can be set to zero.

The classical minimum energy approach for regularizing (1) computes a least-squares solution by solving the constrained minimization

$$(P_{LS}) \qquad \min \|x\|_2^2 \quad \text{s.t.} \quad Ax = b, \tag{6}$$

where $\| \cdot \|_2$ denotes the Euclidean ℓ_2 norm. A common and successful class of algorithms called *row-action* methods [2] exists to solve this problem and its variants. They are well suited for parallel computation and therefore particularly attractive for our application.

A further established approach is the linearly constrained entropy maximization problem

$$(P_E) \qquad \min \sum_{j=1}^{n} x_j \log(x_j) \quad \text{s.t.} \quad Ax = b, \ x \geq 0, \tag{7}$$

where $E(x) := -\sum_{j=1}^{n} x_j \log(x_j)$, $x \geq 0$, is the Boltzmann-Shannon entropy measure. Adding the nonnegativity constraint is necessary for the maximum entropy approach, since the log function is defined only for positive values and $0 \log 0 := 0$. After the removal of redundant equations as described at the beginnig of this section, the relative interior of the feasible set is typically nonempty, and the unique solution to (P_E) is strictly positive, except for those variables corresponding to removed columns in A.

Algebraic reconstruction techniques (ART, MART) [8] are classical row-action methods for solving (6) and (7). For details, we refer to, e.g., [2]. In connection with TomoPIV, MART *(multiplicative algebraic reconstruction technique)* has been advocated in [6]. The behavior of MART in the case of inconsistent equations, is not known. In contrast, ART converges even when applied to inconsistent systems to the least-square solution. ART can be adapted to involve nonnegativity constraints by including certain constraining strategies [9] in the iterative process.

4 Regularization via Sparsity

Since the particles are sparsely spread in the 3D volume, we are interested to look for solutions of (1) with as many components equal to zero as possible. This leads us to the following optimization problem

$$(P_0) \quad \min \|x\|_0 \quad \text{s.t.} \quad Ax = b, \tag{8}$$

providing a *minimal-support* solution. We denote the support of a vector by $\text{supp}(x) := \{i : x_i \neq 0\}$ and by $\|x\|_0 := \#\text{supp}(x)$ the sparsity measure.

Problem (P_0) is nonsmooth with a highly nonconvex objective function. Thus many local optima may occur. Moreover its complexity grows exponentially with the number of variables n and if P$\neq$NP there is no polynomial time algorithm, that for every instance A and b computes the optimal solution of (P_0), see [12] for this NP-hardness result. In addition to this negative results the regularization attempt via sparsity may be inappropriate in case of nonuniqueness of the sparsest solution of (P_0). Fortunately previous work has shown that if a sparse enough solution to (P_0) exists than it will be necessarily unique. In what follows we will give a flavor of this results. They involve the measure $\text{spark}(A)$ which equals the minimal number of linearly dependent columns of A (see [4, 5]) and the signature of a matrix $A \in \mathbb{R}^{m \times n}$. This is defined as the discrete function $\text{sig}_A(k) \in [0, 1]$, for $k \in \{2, \ldots, n\}$, that equals the number of k column combinations in A which are linearly dependent divided by the number of all k columns from the n existing ones. We have

$$2 \leq \text{spark}(A) \leq \text{rank}(A) + 1,$$

where $\text{rank}(A)$ is at most m. By definition $\text{sig}_A(k) = 0$, for all $k < \text{spark}(A)$. In contrast to $\text{rank}(A)$, *spark(A)* as well as $\text{sig}_A(k)$ is NP-hard to compute. Fortunately bounds on this measures can be derived [4].

The following result is surprisingly elementary and can be found in [4].

Theorem 1. *(Uniqueness) Let x be a solution of (1) with $\|x\|_0 < \frac{\mathrm{spark}(A)}{2}$. Then x is the unique solution of (P_0).*

In [5] Elad adopts a probabilistic point of view to study uniqueness of sparse solutions of (P_0) beyond the worst-case scenario.

Theorem 2. *[5, Th. 6, Th. 5] Let $\sigma := \mathrm{spark}(A) \leq \mathrm{rank}(A) =: r$ and x be a solution of (1). Assume the locations of the nonzero entries in x are chosen at random with equal and independent probability. If $1/2\sigma \leq \|x\|_0 =: k \leq r$, then the probability that x is the sparsest solution of (1) is $1 - \mathrm{sig}_A(k)$ and the probability to find a solution of (1) of the same cardinality k is*

(a) $\sum_{j=0}^{k-\sigma} (k-j)(n-k+j) \binom{k}{j} \mathrm{sig}_A(k-j)$ *or lower, if $\|x\|_0 \geq \sigma$;*
(b) 0, *if $1/2\sigma \leq \|x\|_0 < \sigma$.*

Hence uniqueness of the sparsest solution with cardinality less then $\mathrm{spark}(A)$ can be claimed with probability 1. Some of our previous experiments have shown that for projection matrices A as arising in our application, but based on voxel distretization, $\mathrm{rank}(A)$ is approaching m, even though A is rank-deficient. In contrast $\mathrm{spark}(A)$ remains small even for increasing m. However, in the case of blob-based discretizations we made better experiences regarding the value of $\mathrm{spark}(A)$.

An upper bound on the signature was derived via arguments from matroid theory [1], under the assumption that the spark is known. Compare also Fig. 2, left.

4.1 Minimum Support Solutions by Solving Polyhedral Concave Programs

In this section we present an adaption of a general method due to Mangasarian [11] to solve the NP-hard problem (8) and its counterpart

$$(P_0^+) \quad \min_{x \in \mathcal{F}} \|x\|_0 \,, \tag{9}$$

which involves the nonnegativity constraints and amounts to find the nonnegative sparsest solution of (1). The core idea is to replace for $x \geq 0$, $\|x\|_0$ by the exponential approximation $f_\alpha(x) := \sum_{i=1}^{n}(1 - e^{-\alpha x_i})$, for some $\alpha > 0$. f_α is also smooth and concave. Note that $f_\alpha(x) \leq \|x\|_0 = \lim_{\alpha \to \infty(x)} f_\alpha(x)$ holds, for all $x \geq 0$. Consider now the problem

$$(P_0^\alpha) \quad \min_{x \in \mathcal{F}} f_\alpha(x) \,, \tag{10}$$

for some fixed parameter $\alpha > 0$. This problem solves the minimal-support problem (9) exactly for a finite value of the smoothing parameter α.

Theorem 3. *Under the assumption that $\mathcal{F} \neq \emptyset$, problem (9) has a vertex of $\mathcal{F}$ as a solution, which is also a vertex solution of (10) for some sufficiently large positive but finite value α_0 of α.*

The proof of the above statement rely on the fact that (P_0^α) has a vertex solution, since the concave objective function f_α is bounded bellow on the polyhedral convex set $\mathcal{F}$, which contains no straight lines that go to infinity in both directions (it contains the constraint $x \geq 0$).

Note that Theorem 3 guarantees a vertex solution of $\mathcal{F}$ despite the nonconcavity of $\|\cdot\|_0$. Moreover it states that by solving problem (P_0^α) for a sufficiently large but finite α we have solved problem (P_0^+). Similar statements can be made for sparse solutions of polyhedral convex sets, see [11].

We stress however that there is no efficient way of computing α as long P$\neq$NP.

4.2 SLA

In this section we turn our attention to a computational algorithm which at every step solves a linear program and terminates at a stationary point, i.e. a point satisfying the minimum principle necessary optimality condition [10]. It is a successive linear approximation method of minimizing a concave function on a polyhedral set due to Mangasarian [11]. The method can also be seen as a stepless Franke-Wolfe algorithm [7] with finite termination or a DC algorithm for a particular decomposition of the concave objective function as a *difference of convex* functions [13]. We state now the algorithm for problem (P_0^α) which has a differentiable concave objective function.

Algorithm 1 *(Successive Linearization Algorithm - SLA)*

(S.0) Choose $x^0 \in \mathbb{R}^n$ and set $l := 0$.
(S.1) Set $c^l = \alpha e^{-\alpha x^l}$ and compute x^{l+1} as a vertex solution of the following linear program

$$\min_{x \in \mathcal{F}} (c^l)^T x\,. \tag{11}$$

(S.2) If $x^l \in \mathcal{F}$ and $x^{l+1} = x^l$ is satisfied within the tolerance level: STOP.
Otherwise, increase the iteration counter $l \leftarrow l + 1$ and continue with (S.1).

Note that the finite termination criteria in step $(S.2)$ is nothing else but the minimal principle necessary optimality condition [10] for this particular choice of objective function with $\nabla f_\alpha(x^l) > 0$. The additional condition $x^l \in \mathcal{F}$ treats the case of x^0 not belonging to $\mathcal{F}$.

Theorem 4. *[10, Th. 4.2] Algorithm 1 is well defined and generates a finite sequence of iterates $\{x^l\}$ with strictly decreasing objective function values: $f_\alpha(x^0) > f_\alpha(x^1) > \cdots > f_\alpha(x^f)$ such that the final iterate x^f is a local optima of (P_0^α).*

5 Some Experiments and Discussion

We demonstrate the feasibility and the typical behavior of the proposed approach on one medium size example in order to facilitate the evaluation of the measure

spark and to enable visualization. We consider 1 to 40 particles in a 2D volume $V = [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$. The grid refinement was chosen $d = 0.0154$, resulting in 4356 gridpoints. At these gridpoints we centre a Gaussian-type basis function, where $\sigma = d$. Particle positions were chosen randomly but at grid positions, to avoid discretization errors. Four $50-$pixel cameras are measuring the 2D volume from angles $45^o, 15^o, -15^o, -45^o$, according to a fan beam geometry. The screen and focal length of each camera is 0.5.

For the obtained projection matrix, with about 9% nonzero entries, we obtained $rank(A) = 172$ and a tight upper bound on $\text{spark}(A) \leq 25$, by computing a sparse kernel vector from the reduced row echelon form of A. In order to estimate sig_A, we assumed that $\text{spark}(A) = 25$, although a computed lower bound derived via the *mutual coherence* [4] is by far not so optimistic.

For each number of particles $k = 1, \ldots, 40$, we generated 1000 random distributions, yielding each time an original weighting vector of support length $\|x_{orig}\|_0 = k$. The pixel intensities in the measurement vector b are computed according to (4), integrating the particle image exactly along each line of sight. Two randomly selected 30 and 40 particle distribution are depicted in Fig. 3.

We compared the results of SLA, see Tab. 1 and Fig. 3, to the classical algebraic methods. Besides ART and MART, both without relaxation, we considered also a modification of ART, which we further call $ART+$, based on constraining strategies proposed in [9]. This method amounts to project the current iterate on the positive orthant after each complete sweep of the ART, through all equations.

As a preprocessing step we reduce system $Ax = b$ according to the methodology proposed in section 3. The reduced dimensionalities are summarized in Tab. 1. As starting point for ART(+) and SLA we chose the all zero vector, whereas $x^0 := e^{-\mathbb{1}}$ for MART.

We terminate the iteration of the main algorithm if the appropriate termination is satisfied within tolerance level 10^{-4} or if the maximum number of outer iterations was reached, i.e. 1000 complete sweeps through the rows of the reduced matrix A in the case of (M)ART(+) or 100 linear programs solved in $(S.1)$ of SLA.

The linear programs we solved using the primal simplex method (since we are interested in vertex solutions) that come along with MOSEK Version 5.0.0.45. Having obtained with the SLA a stationary point x^{f_0} different from our original solution we applied a heuristic method to further reduce the support. Starting with x^{f_0}, we randomly chose $\mathcal{T}_i \subset \mathcal{S}_i = \text{supp}(x^{f_i})$, with $\#\mathcal{T}_i = 10$, then deleted the corresponding columns in A and restarted SLA with this reduced matrix. Following this and increasing $i \leftarrow i + 1$ it often helped to recover the original solution. The appropriate line in Tab. 1 reports the 74 restarts of SLA along with the average number (bold) of linear programs solved.

More significant are the reconstructed particle images, depicted in Fig. 3. The "smearing" of the particles in the lines of the projections is typical for minimum energy reconstructions. This phenomenon is preserved by ART. The MART and ART+ reconstructions show more distinct particles. However, additional spots

Table 1. SLA vs. (M)ART(+) for the examples in Fig. 3

$\|x_{orig}\|_0$	$m_r \times n_r$	Method	#Outer Iter.	$\|x_{orig} - x^f\|_2$	$\|Ax^f - b\|_\infty$
30	145×3239	SLA	2	$7.21 \cdot 10^{-11}$	$2.20 \cdot 10^{-10}$
		MART	145	3.88	$1.73 \cdot 10^{-2}$
		ART	145	5.18	$1.12 \cdot 10^{-3}$
		ART+	145	46.7	$1.52 \cdot 10^{-2}$
40	142×3352	SLA	**24**(74)	$6.31 \cdot 10^{-12}$	$5.32 \cdot 10^{-12}$
		MART	142	5.70	$2.45 \cdot 10^{-3}$
		ART	142	6.02	$3.12 \cdot 10^{-3}$
		ART+	142	68.6	$5.98 \cdot 10^{-3}$

are visible. SLA was able to reconstruct the original image, at considerably increased costs for the 40 particles considered.

An interesting behavior of SLA is worth to mention. Very often (83% of the 1000 test runs) it takes only two outer iteration until convergence to the original solution. Since the last iteration is just a convergence test, this amounts to solve a single linear program with equally weighted linear objective, in view of our starting point $x^0 = 0$, to obtain x_{orig}. This phenomenon is superficially tackled by counting those test runs with respect to an increasing particle density, whereby SLA succeeds to find the original distribution by solving one linear program only. The results in Fig. 2, right, complement the received opinion [3–5] that sufficiently sparse solution to underdetermined systems can be found by ℓ_1 norm minimization.
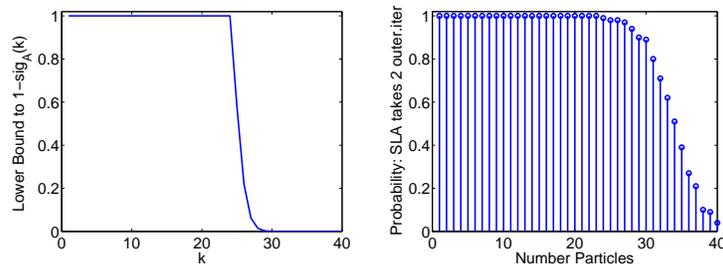

Fig. 2. Left: Lower bound, derived via [5, Th. 7], on the probability that the underlying k particle distribution is the sparsest of all those satisfying the measurements. **Right:** The probability (must be read with caution) of success of the SLA performs only one LP to find the original particle distribution, with respect to increasing density.
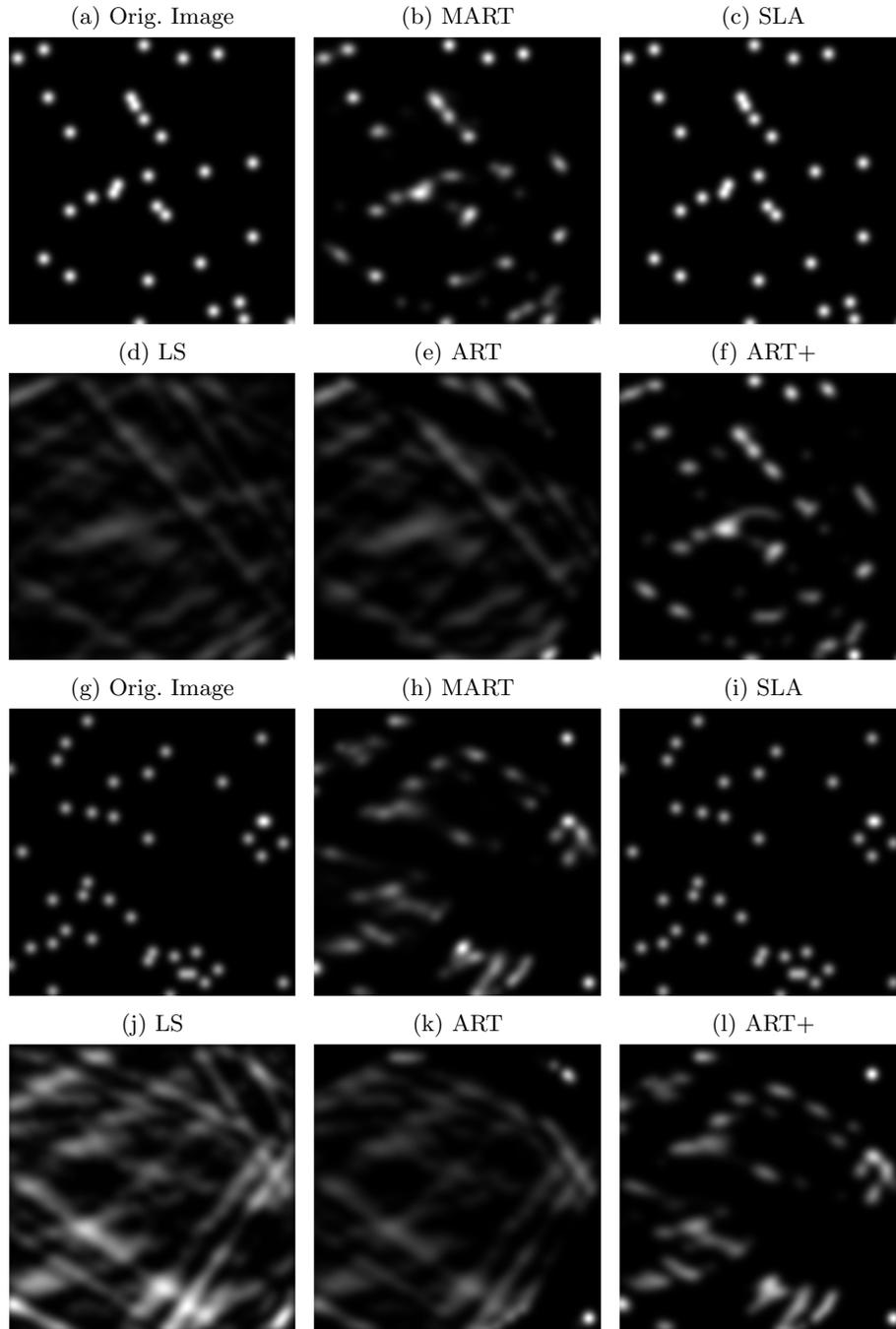
(a) Orig. Image   (b) MART   (c) SLA

(d) LS   (e) ART   (f) ART+

(g) Orig. Image   (h) MART   (i) SLA

(j) LS   (k) ART   (l) ART+

Fig. 3. Top: (a)-(f) 30 particles and their reconstruction. **Bottom:** (g)-(l) 40 particles and their reconstruction. (a),(g) The original particle distribution contains 30 resp. 40 particles. (b),(h) Reconstruction using MART. (c),(i) Reconstruction using SLA. (d),(j) Minimal norm solution of (P_{LS}) obtained via the Moore-Penrose pseudoinverse A^+, without reducing A. (e),(k) Reconstruction using ART. (f),(l) Reconstruction using ART+.

6 Conclusion and Further Work

A classical algebraic reconstruction approach that is currently in use, together with closely related variants, were recently re-considered by the authors in some detail to reveal pros and cons from the perspective of TomoPIV. While pursuing these goals, an regularization alternative was developed, which amounts to find positive sparse solutions of underdetermined systems. Promising research directions were outlined to accomplish more efficiently this task in terms of computational time. Still model extensions have to be developed to the problem in the presence of noise and discretization errors.

References

1. A. Björner: Some matroid inequalities. Discrete Math. **31** (1980) 101–103
2. Y. Censor: Parallel Optimization: Theory, Algorithms and Applications. Oxford University Press, Inc., New York, 1997
3. S.S. Chen, D.l. Donoho and M.A. Saunders: Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. **20** (1998) 33–61
4. D.L. Donoho and M. Elad: Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization. Proc. Natl. Acad. of Sci. U.S.A. **100** (2003) 2197–2202
5. M. Elad: Sparse Representations Are Most Likely to Be the Sparsest Possible. EURASIP JASP (2006) 1–12.
6. G. Elsinga, F. Scarano, B. Wieneke,B. van Oudheusden: Tomographic particle image velocimetry. Exp. Fluids **41** (2006) 933–947
7. M. Frank and P. Wolfe: An algorithm for quadratic programming. Nav. Res. Log. Quart. **3** (1956) 95–110
8. R. Gordon, R. Bender and G.T. Herman : Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. J. Theor. Biol. **29** (1970) 471–481
9. I. Koltracht and P. Lancaster: Contraining Strategies for Linear Iterative Processes. IMA J. Num. Anal. **10** (1990) 55–567
10. O.L. Mangasarian: Machine Learning via Polyhedral Concave Minimization. Applied Mathematics and Parallel Computing – Festschrift for Klaus Ritter, H. Fischer, B. Riedmüller, S. Schäffler, editors, Physica-Verlag, Heidelberg, 1996, 175–188
11. O.L. Mangasarian: Minimum-support solutions of polyhedral concave programs. Optimization **45** (1999) 149–162
12. K. Natarajan: Sparse approximate solutions to linear systems. SIAM J. Comput. **24** (1995) 227–234
13. P.D. Tao and L.T.H. An: A D.C. Optimization Algorithm for Solving the Trust-Region Subproblem. SIAM J. Optim. **8** (1998) 476–505.
14. M. Raffel, and C. Willert, and J. Kompenhans: Particle Image Velocimetry, 2nd edition, Springer, 2001
15. F. Scarano and M.L. Riethmüller: Advances in iterative multigrid PIV image proccesing. Exp. Fluids **29** (2000) 51–60