Enhancing Sparsity by Constraining Strategies: Constrained SIRT versus Spectral Projected Gradient Methods

Stefania Petra*, **Constantin Popa**** and **Christoph Schnörr***

* "Ruprecht-Karls" University of Heidelberg, Germany; e-mail: {petra, schnoerr}@math.uni-heidelberg.de; the authors gratefully acknowledge financial support by the DFG Grant SCHN 457/10-1;
**"Ovidius" University of Constanta, Romania; e-mail: cpopa@univ-ovidius.ro;

Abstract

We investigate a constrained version of simultaneous iterative reconstruction techniques (SIRT) from the general viewpoint of projected gradient methods. This connection enable us to assess the computational merit of this algorithm class. We borrow a leaf from numerical optimization to cope with the slow convergence of projected gradient methods and propose an acceleration procedure based on the spectral gradient choice of steplength along with a nonmonotone strategy. We compare these schemes and present numerical experiments on some algebraic image reconstruction models with sparsity constraints, with particular attention to tomographic particle image reconstruction. The performance of both constrained SIRT and nonmonotone spectral projected gradient approach is illustrated for several constraining strategies.

1 Introduction

Successfully employed at the solution of huge and sparse systems of linear algebraic equations which arise in many application areas (most notably discrete models of computerized tomography) *Simultaneous Iterative Reconstruction Techniques (SIRT)* [10, 9] continue to receive great attention due to their low memory requirements and extreme simplicity. The SIRT are inherently parallel schemes which after each (possibly relaxed) reflection or projection of a current approximation with respect to each hyperplan (described by each equation of the linear algebraic system) take a convex combination of these intermediate points as the next iterate. The convergence to a (weighted) least-squares solution is guaranteed even in case of inconsistency. In order to deal with limited-data linear inverse problems or with noise corrupted data a *regularization* technique is required. Regularization techniques try, as much as possible, to take advantage of prior knowledge one may have about the nature of the "true" solution. This can be modeled by assuming that the solution is contained in a (compact) set $\mathcal{B}$. If this set is convex and exhibits a simple structure one may (orthogonally) project the iterates generated by SIRT onto the range within the components of an acceptable reconstruction vector must lie. These projection techniques traditionally termed as *constraining strategies* were generalized by the authors in [16] and applied to the sequential reconstruction technique ART [14]. Inter alia we show in the present work that such constraining strategies can be applied also to SIRT, see Section 3.

However, the approach in this paper is tailored to the case when the object (image I) to be reconstructed can be represented by a sparse expansion, i.e., when I can be represented by a series expansion with respect to a basis with only a small number of nonzero coefficients x. Moreover this number, say k, is device-controlled and thus known a priori in the application area in focus. Hence $\mathcal{B}$ may be written as the *union* of all subsets of $\mathbb{R}^n$ with at most k nonzero components, thus a union of linear subspaces. Together with the nonconvexity of such $\mathcal{B}$, the number of such subspaces, which grows exponentially with n and k, make "projection" onto $\mathcal{B}$ unrealistic. Fortunately this complicated set $\mathcal{B}$ can be replaced by a nice convex set, e.g. a ℓ_1-ball or even the positive orthant, provided that the underlying solution is sufficiently sparse and positive. Successive orthogonal projections on this "new" feasible region, which are now nonexpensive operations, lend themselves to constraining strategies, see Section 3.2, and constrained versions of SIRT emerge as classical gradient projection methods, see Section 4.

It is well known that these methods may exhibit very slow convergence if not combined with appropriate steplength selections. In order to accelerate the projected gradient method we exploit the spectral steplength introduced by Barzilai and Borwein in [2] for the unconstrained case. We consider a nonmonotone spectral projected gradient method developed in [4], see Section 4.1, and present extensive numerical experiments in Section 5 on image reconstruction problems motivated by the work [13].

The authors introduced a new 3D technique, called *Tomographic Particle Image Velocimetry (TomoPIV)* for imaging turbulent fluids with high speed cameras. The technique is based on the instantaneous reconstructions of particle volume functions from few and simultaneous projections (2D images) of the tracer particles within the fluid. Since the relationship between observable 2D images and interesting 3D images is (approximately) linear, as detailed in [13, 19], the situation can be modeled mathematically by

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

The projection matrix A is underdetermined since, in contrast to medical imaging, the object to be reconstructed is acquired under a tiny range of angles, i.e. the image to be reconstructed is highly undersampled. As a consequence, the reconstruction problem becomes severely ill-posed and a regularization approach has to be applied to estimate the weights x from the recorded 2D images.

2 Regularization via Sparsity Maximization, ℓ_1-Minimization or Positivity Constraints

The original 3D light intensity distribution I can be well approximated by only a very small number of active basis functions, see [19], relative to the number of possible particle positions in a 3D domain, since the particles are sparsely spread in the 3D volume. This leads us to the following regularization principle: find an (approximative) solution of (1) with as many components equal to zero as possible, i.e.,

$$\min \|x\|_0 \text{ s.t. } Ax = b \ , \tag{2}$$

where $\|x\|_0$ counts the nonzero components in $x \in \mathbb{R}^n$. In general the search for the sparsest solution is intractable (NP-hard), however. The newly founded theory of *Compressed Sensing* [6, 7] showed that one can compute via ℓ_1-minimization the sparsest solution for underdetermined systems of equations provided certain properties [8] are satisfied, which unfortunately do not hold for our particular scenario. The authors in [20] showed empirically that there are thresholds on sparsity (i.e. density of the particles) depending on the numbers of measurements (recording pixel in the CCD arrays) which resemble the known thresholds for the idealized mathematical setups. ℓ_1-Minimization methods yield (near) perfect reconstructions below these sparsity thresholds and above they fail with high probability, similar to the results of Candès and Tao [7]. These authors showed that there is a constant C such that for a signal $\widetilde{x}$ with at most k nonzero entries, $b \approx A\widetilde{x}$ and $m \geq Ck \log(\frac{n}{k})$, the solution of

$$\min \|x\|_1 \text{ s.t. } Ax = b \tag{3}$$

will be exactly the original signal $\widetilde{x}$ with overwhelming probability, provided the rows of A are randomly chosen Gaussian distributed vectors, which guarantees the favorable properties of A, like incoherence, see [8]. Even for coherent matrices A ℓ_1-minimization seems to lead to promising results, see [20].

When the sparsity parameter k of the solution of (2) is known a priori it is possible to consider instead of problem (2) and (3) the least-squares problem

$$\min \frac{1}{2}\|Ax - b\|^2 \quad \text{s.t. } \|x\|_0 \leq k \tag{4}$$

imposing a sparsity constraint. For consistent systems $Ax = b$, in particular when A is a full rank underdetermined matrix, problems (2) and (4) are. As already discussed before the nonconvexity and the structure of the constraint set make problem (4) a difficult combinatorial problem. Similar to the developments in the compressed sensing literature a relaxed model may be considered

$$\min \frac{1}{2}\|Ax - b\|^2 \quad \text{s.t. } \|x\|_1 \leq r \ , \tag{5}$$

known as the *LASSO* problem in the statistical community. Again, problems (4) and (5) are equivalent, under an appropriate correspondence of parameters k and r and certain properties of A. Moreover, problem (5) is tractable since the feasible set is the convex ℓ_1-ball of radius r and can be recast as an quadratic program with linear constraints.

An even simpler regularization approach, much less perceived in the sparse regression literature, is a least-squares fit subject to simple positive constraints, i.e.,

$$\min \frac{1}{2}\|Ax - b\|^2 \quad \text{s.t. } x \geq 0 \ , \tag{6}$$

if the original solution is known to be sparse and positive. It was shown in [19] that one can reduce the original linear system $Ax = b$ by eliminating the i-th row of A corresponding to a zero measurement $b_i = 0$ as well as all columns in A whose i-th entries are positive, provided that the entries in b and A are nonnegative. If the reduced system has an overdetermined coefficient matrix of full rank then the original (positive) solution must be the unique positive solution of the underdetermined system. Even beyond the thresholds on sparsity of an original positive solution generating such an "overdetermined" reduction a sufficiently sparse positive solution might be unique, provided that A satisfies some (difficult to check) properties, see [23]. Additionally, it can be shown that the unique positive solution of an underdetermined system is also the solution of minimal ℓ_1 norm.

Combining (5) and (6) we obtain

$$\min \frac{1}{2}\|Ax - b\|^2 \quad \text{s.t. } \mathbf{1}^T x \leq r, \ x \geq 0 \ , \tag{7}$$

which is a quadratic problem subject to simplex constraints. On the other hand problem (3) can be solved by linear programming. Nevertheless, general-purpose LP and QP solvers involve solution of full $n \times n$ linear systems, an operation costing order $O(n^3)$ flops. Therefore, there is a need to find a more efficient algorithm that requires only matrix-vector products involving A and A^T and therefore adapts to the difficulty that the matrix A is huge and not explicitly available.

3 Constrained Simultaneous Iterative Reconstruction Techniques

3.1 Simultaneous Iterative Reconstruction Techniques

The well-known *Algebraic Reconstruction Techniques (ART)* [14], for solving least-squares problems, orthogonally projects the current approximation x^k onto the hyperplanes $H_i = \{x | A^T_{i,\bullet} x = b_i\}, \quad i = 1, \ldots, m$, not simultaneously but sequentially. The projection onto the n-th hyperplane is taken as the new approximation x^{k+1}, and the process is repeated. Such a method can converge only if the right-hand side b lies in the span of the matrix. For perturbed right-hand sides one may therefore not expect convergence.

Simultaneous Iterative Reconstruction Techniques (SIRT) are designed to give convergence in this case. They distinguish themselves from ART methods in that they do not update the iterated vector after each equation, but after an entire sweep through all the equations, and thus, during one sweep, they use the same residual vector for each equation.

Given the current iterate x^k, it is first projected on all hyperplans H_i and then the next iterate is

$$x^{k+1} = x^k + \alpha_k \left(\sum_{i=1}^{m} \omega_i \Pi_{H_i}(x^k) - x^k \right) , \tag{8}$$

where ω_i are fixed positive weights summing up to 1, $\alpha_k \in [\varepsilon, 2 - \varepsilon]$ is a relaxation parameter, with $\varepsilon > 0$ fixed but arbitrary tiny and P_{H_i} is the orthogonal projection onto the i-th hyperplan H_i. In short, x^{k+1} is a weighted average of relaxed projections of x^k.

If the relaxation parameters satisfy $\alpha_k = 2$ for all k we obtain Cimmino's method of simultaneous reflections [10]. Cimmino takes the weighted average of all reflections $y^{k,i} := (2\Pi_{H_i} - I)x^k$ of x^k with respect to all hyperplanes H_i. In view of the explicit form of the projection onto a hyperplan equation (8) can be written in matrix notation

$$x^{k+1} = x^k - \alpha_k A^T D(Ax^k - b) , \tag{9}$$

where D is a positive definite diagonal matrix defined by

$$D := \mathrm{diag}\left(\frac{\omega_1}{\|A_{1,\bullet}\|^2}, \ldots, \frac{\omega_m}{\|A_{m,\bullet}\|^2} \right) . \tag{10}$$

SIRT (8) iteratively approximates a weighted least-squares solution

$$\min \|Ax - b\|_D , \tag{11}$$

even in the inconsistent case, see e.g. the result due to Combettes [11, Th. 4]. When the weights in (8) are given by

$$\omega_i = \frac{\|A_{i,\bullet}\|^2}{\sum_{j=1}^m \|A_{j,\bullet}\|^2} \tag{12}$$

the sequence $\{x^k\}_k$ always converges (also in the inconsistent case) to a least squares solution. We can replace the fixed weights ω_i in (8) by ω_i^k with $\omega_i^k > 0$ and $\sum_{i=1}^m \omega_i^k = 1$ for all k and still have a convergent algorithm in the consistent case, i.e. when $Ax = b$ has an exact solution, see [1, Th. 1].

Defining the matrices $T := I - \alpha_k A^T DA$ and $R := \alpha_k A^T D$ with D from equation (10) we can rewrite the iteration in (8) as

$$x^{k+1} = Tx^k + Rb \; . \tag{13}$$

It turns out that $\mathcal{R}(A^T)$ is an invariant subspace on which operator T is contractive and $Rb \in \mathcal{R}(A^T)$ for every right-hand side b. Thus (linear) convergence of the sequence $\{x^k\}_k$ towards an $x^* \in \mathcal{R}(A^T)$ can be obtained by Banach-like arguments, provided that $x^0 \in \mathcal{R}(A^T)$. Denoting $\widetilde{T} = T|_{\mathcal{R}(A^T)}$ we obtain the following convergence result, see [21] for a proof.

Theorem 1 *For any initial approximation $x^0 \in \mathbb{R}^n$, the sequence $\{x^k\}$ generated by SIRT* (13) *converges and*

$$\lim_{k\to\infty} x^k = P_{\mathcal{N}(A)}(x^0) \; + \; x_{LS} + \Delta \quad \text{with} \quad \Delta = (I - \widetilde{T})^{-1} R P_{\mathcal{N}(A^T)}(b) \; . \tag{14}$$

In particular, $\Delta = 0$ when the system (1) *is consistent. Moreover, we have the a priori estimate*

$$\|x^k - x^*\| \leq \frac{\kappa^k}{1-\kappa} \|x^0 - x^1\| \tag{15}$$

and the a posteriori estimate

$$\|x^{k+1} - x^*\| \leq \frac{\kappa}{1-\kappa} \|x^{k+1} - x^k\| \; , \tag{16}$$

where $\kappa = \|\widetilde{T}\|$ and x^* be the limit point in (14)*. In particular, the convergence rate of sequence $\{x^k\}_k$ is linear.*

However, the minimum norm solution of $Ax = b$ (in the consistent case) or of the weighted least-squares problem (11) is in general a dense vector and may considerably differ from the true sparse solution. As discussed in Section 2 we usually have a priori information about the range within the values of the solution components must lie, e.g. $\|x\|_1 \leq r$ etc. This should be exploited by the iterative method (8).

3.2 Constraining Strategies

In this section we are interested in techniques able to steer the approximations x^k generated by SIRT in some given set $\mathcal{B}$. In particular we are interested in the choices $\mathcal{B} = \mathbb{R}^n_+$, $\mathcal{B} = \{x \mid \|x\|_1 \leq r\} =: \mathcal{B}_{\ell_1,r}$ or $\mathcal{B} = \{x \mid \mathbf{1}^T x \leq r, x \geq 0\} =: \Delta_{n,r}$, compare Section 2.

Such techniques traditionally termed as *constraining strategies* were investigated in [16] and applied to ART iterations of the form (13). Following the authors in [16] we consider a constraining function $C : \mathbb{R}^n \to \mathbb{R}^n$ with a closed image $\mathcal{I}(C) \subset \mathbb{R}^n$ and the properties

$$\|C(x) - C(y)\| \leq \|x - y\| \ , \tag{17}$$

$$\text{if} \ \ \|C(x) - C(y)\| = \|x - y\| \ \ \text{then} \ \ C(x) - C(y) = x - y \ , \tag{18}$$

$$\text{if} \ \ y \in \mathcal{I}(C) \ \ \text{then} \ \ y = C(y) \ , \tag{19}$$

and propose the constrained SIRT

$$x^{k+1} = C(Tx^k + Rb) \ , \tag{20}$$

were T and R are defined as in the previous section.

Similar to [16, Th. 3], we obtain the following convergence result for the constrained SIRT (20).

Theorem 2 *Let us suppose that all rows of the matrix A are nonzero,* $\operatorname{rank}(A) \geq 2$*, the constraining function C satisfies* (17) – (19) *and the set $\mathcal{V}$ defined by*

$$\mathcal{V} = \{y \in \mathcal{I}(C), \ y - \Delta \in LSS(A, b)\} \tag{21}$$

is nonempty, where Δ is defined in 2. Then, for any $x^0 \in \mathcal{I}(C)$ the sequence $\{x^k\}$ generated by (20) *converges and its limit belongs to the set $\mathcal{V}$. If the problem* (1) *is consistent, then the above limit is one of its constrained solutions.*

Orthogonal projections onto convex sets $\mathcal{K}$ are constraining strategies, since $\Pi_{\mathcal{K}}$ satisfies (17) – (19), compare [21].

3.3 Projections onto the ℓ_1-Ball, the Simplex or the Positive Orthant

While projection onto the positive orthant $\mathbb{R}^n_+$ is simply

$$[\Pi_{\mathbb{R}^n_+}(x)]_i = \max\{x_i, 0\} \ , \quad i \in \{1, \dots, n\} \ , \tag{22}$$

projection onto the simplex or the ℓ_1-ball is more involved. The next result shows how we can perform Euclidean projection onto the positive simplex. See, e.g. [19, 12] for a proof.

Proposition 1 *Let $x_{(i)}$ denote the i-th order statistics of x, that is, $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(n)}$ and denote the positive simplex by $\Delta_{n,r} := \{y \mid \sum_{i=1}^{n} y_i = r, y \geq 0\}$. Then*

$$[\Pi_{\Delta_{n,r}}(x)]_i = \begin{cases} \frac{1}{|I(x)|}\left(r - \sum_{j \in I(x)}(x_j - x_i)\right), & i \in I(x) \ , \\ 0, & otherwise \ , \end{cases} \tag{23}$$

where $I(x)$ contains the indexes of the $m := |I(x)|$ largest components of x such that $\sum_{j=1}^{m}(x_{(j)} - x_{(i)}) < 1$.

The next result [21] says that finding the orthogonal projection of a vector $x \in \mathbb{R}^n$ onto the ℓ_1-ball of radius r can be reduced to the problem of finding the projection onto the simplex.

Proposition 2 *Let y^* be the (unique) solution of*

$$\min \frac{1}{2}\|y - |x|\|^2 \quad \text{s.t. } \|y\|_1 \leq r, \ y \geq 0 \ , \tag{24}$$

where $|x|$ denotes the vector of absolute values $|x| := (|x_1|, \ldots, |x_n|)^T$. Then $\operatorname{sign}(x) \cdot |y^*| := (\operatorname{sign}(x_1)|y_1^*|, \ldots, \operatorname{sign}(x_n)|y_n^*|)^T$ *solves*

$$\min \frac{1}{2}\|y - x\|^2 \quad \text{s.t. } \|y\|_1 \leq r \ . \tag{25}$$

We stress that exact projection onto the ℓ_1-ball can also be performed in $O(n)$ linear time, see [12], by avoiding sorting the vector first.

4 Projected Gradient Method

For the particular choice $C = \Pi_{\mathcal{K}}$ with $\mathcal{K}$ some nonempty closed convex set, it turns out that iteration (20) is the basic gradient descent iteration with damping parameter α_k,

$$x^{k+1} = \Pi_{\mathcal{K}}(x^k - \alpha_k \nabla f(x^k)) \ , \tag{26}$$

applied to the convex and differentiable function

$$f(x) = \frac{1}{2}\|Ax - b\|_D^2 \ , \tag{27}$$

where $\|x\|_D$ denotes the energy norm $\langle x, Dx\rangle^{1/2}$. For $\mathcal{K} = \mathbb{R}^n$ we obtain SIRT.

Iteration (26) converges if $\alpha_k < \frac{2}{L}$ with L being the Lipschitz constant of the gradient ∇f of f in (27), see [17, Th. 5.1]. Since $\nabla f(x) = A^T D(Ax-b)$, the Lipschitz constant L is obviously the largest eigenvalue of the matrix $A^T DA$. A simple upper bound is given by

$$\|A^T DA\| = \left\| \sum_{i=1}^{m} \omega_i \frac{A_{i,\bullet} A_{i,\bullet}^T}{\|A_{i,\bullet}\|^2} \right\| \leq \sum_{i=1}^{m} \omega_i \underbrace{\left\| \frac{A_{i,\bullet} A_{i,\bullet}^T}{\|A_{i,\bullet}\|^2} \right\|}_{=1} = 1 \; . \tag{28}$$

Hence iteration (26) converges to a solution of

$$\min_{x \in K} f(x) \tag{29}$$

provided that $\alpha_k \leq 2$ and a solution to (29) exists. When ∇f is Lipschitz continuous in $\mathcal{K}$ with known Lipschitz constant L, the iteration (26) generates for the every stepsize $\alpha_k \leq \frac{2}{L}$ a sequence in $\mathcal{K}$ for which f decreases towards its minimal value on $\mathcal{K}$. If the stepsize α_k in (26) is chosen to be

$$\alpha_k = \operatorname{argmin}_\alpha f(x^k - \alpha \nabla f(x^k))$$

it can be computed explicitly since f is a quadratic function. However it is not guaranteed that the function value f at $x^{k+1} = \Pi_{\mathcal{K}}(x^k - \alpha_k \nabla f(x^k))$ will decrease for this particular α_k.

In the safeguard approach proposed by Bertsekas [3, p. 226], we search from each iterate x^k along the negative gradient $-\nabla f(x^k)$, projecting onto $\mathcal{K}$, and performing a backtracking line search (referred as "Armijo rule along the projection arc") until a sufficient decrease is attained in f. Several trail steps are projected on the convex set and at each f has to be evaluated, making this process expansive even if projection is inexpensive, as in the case of simple positive constraints.

4.1 Spectral Projected Gradient

The method proposed in [4] combines the classical projected gradient method (26) with the spectral gradient choice of steplength [2] and a nonmonotone line search strategy [15] to avoid additional trial projections during the one dimensional search process. The *Spectral Projected Gradient (SPG)* method [4] proposed for the minimization of a smooth nonlinear function f subject to convex constraints calculates at each step an approximation to the Hessian H_k of f at x^k by $\eta_k I$ based on the secant condition

$$\nabla f(x^{k+1}) - \nabla f(x^k) \approx \eta_k (x^{k+1} - x^k) \; ,$$

following Barzilai and Borwein [2].

Algorithms 1 and 2 from [4] describe how to obtain x^{k+1} and α_k in the constrained case. The algorithms use an integer $m \geq 1$; a tiny parameter $\alpha_{min} > 0$; a large parameter $\alpha_{max} > \alpha_{min}$; a sufficient decrease parameter $\gamma \in (0, 1)$; and safeguarding parameters $0 < \sigma_1 < \sigma_2 < 1$. Initially, $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$ is arbitrary.

Algorithm 1 *(Spectral Projected Gradient Method - SPG)*

(S.0) *Choose $x^0 \in \mathcal{K}$ and set $k := 0$.*

(S.1) *If $\|\Pi_{\mathcal{K}}(x^k - \nabla f(x^k)) - x^k\| = 0$ is satisfied within the tolerance level: STOP. Otherwise, continue with (S.2).*

(S.2) *Compute $d^k = \Pi_{\mathcal{K}}(x^k - \alpha_k \nabla f(x^k)) - x^k$, λ_k using the line search algorithm below and $x^{k+1} = x^k + \lambda_k d^k$.*
Compute $s^k = x^{k+1} - x^k$, $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ and $\beta_k = \langle s^k, y^k \rangle$.
If $\beta_k \leq 0$ set $\alpha_{k+1} = \alpha_{max}$. Otherwise,
compute $\alpha_{k+1} = \min\{\lambda_{max}, \max\{\alpha_{min}, \frac{\langle s_k, s_k \rangle}{\beta_k}\}\}$

(S.3) *Increase the iteration counter $k \leftarrow k + 1$ and goto (S.1).*

The line search procedure below is based on a safeguarded quadratic interpolation.

Algorithm 2 *(Line Search)*

(S.2.0) *Compute $d^k = \Pi_{\mathcal{K}}(x^k - \alpha_k \nabla f(x^k)) - x^k$, $\delta = \langle \nabla f(x^k), d^k \rangle$ and set $\lambda := 1$.*

(S.2.1) *Set $x_+ = x^k + \lambda d^k$.*

(S.2.2) *If*

$$f(x_+) \leq \max_{0 \leq j \leq \{k, m-1\}} f(x^{k-j}) + \gamma \lambda \delta, \tag{30}$$

then define $\lambda_k = \lambda$ and goto (S.2.1).
If (30) does not hold define $\lambda_{new} = -\frac{1}{2}\lambda^2 \delta / (f(x_+) - f(x^k) - \lambda\delta)$. If $\lambda_{new} \in [\sigma_1, \sigma_2 \lambda]$ set $\lambda = \lambda_{new}$. Otherwise, compute $\lambda = \lambda/2$ and goto (S.2.1).

Convergence of SPG method follow directly from the results of Birgin, Martinez, and Raydan [4].

Theorem 3 *[4, Th. 2.2] The sequence of iterates $\{x^k\}_k$ generated by the SPG algorithm 1 is well defined and either terminates at a solution of $\min_{x \in \mathcal{K}} f(x)$, or else converges to a constrained minimizer of f at an R-linear rate, provides such minimizer exists.*

5 Numerical Results

5.1 Test Data

We consider a 2D model inspired by a real-world TomoPIV application, see for details [19], and stress that 3D models are direct extensions of the present one. We consider 10 and 20 particles in a 2D volume $\Omega = [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$, see Fig. 1(a) and 3(a). Particle positions were chosen randomly in Ω for 10 particles examples $I^{ex,2}$ and for $I^{ex,1}$ randomly but at grid positions, to avoid discretization errors. Thus, $x^{ex,1}$ corresponding to $I^{ex,1}$ is a binary vector in $\mathbb{R}^{4356}$ having 10 nonzero components. 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, see [19, p. 17]. The pixel intensities in the measurement vector b are computed by integrating the particle image exactly along each line of sight [19] and perturbing the result according to (31) in Section 5.2.

5.2 General Considerations

We applied the algorithms constrained SIRT (20) from Section 3.2 and the SPG method (1) from Section 4.1 to the perturbed system $Ax = b_\varepsilon$, where $b_\varepsilon = b + e$ and b is obtained by integrating exactly along the pixels line of sight. The error vector $e = e(\varepsilon) \in \mathbb{R}^m$ is defined by

$$e(\varepsilon) := \varepsilon \frac{v}{\|v\|} \|b\| \ , \tag{31}$$

where the components of v are chosen at random drawn from a uniform distribution on the unit interval. We have chosen three different values for ε, i.e. $\varepsilon \in \{0, 0.05, 0.1\}$. The bigger is ε, the bigger will be $\|\Delta\| = \|A^+ P_{\mathcal{N}(A^T)}(b)\|$, see 3.2 and Th. 1 for the constrained SIRT (20).

The constraining function used in all computations was either the orthogonal projection onto the positive orthant, i.e. $C = \Pi_{\mathbb{R}^n_+}$ from (22), the orthogonal projection onto the simplex $\Delta_{n,r}$ or the ℓ_1-ball of radius r, computed cf. Section 3.3. In both procedures we sorted the vector v to be projected first (an $O(n \log(n))$ operation), hence being significantly more involved than just taking the positive components of v. Exact projection onto the ℓ_1-ball can also be performed in $O(n)$ linear time according to [12].

As a preprocessing step we reduce system $Ax = b$ according to the methodology described in Section 2. For all considered examples the reduced coefficient matrices are full-ranked but still underdetermined. Hence, all reduced systems (denoted by $A_r x = b_r$) are consistent. Interestingly, for the second example or when the data is perturbed ($\varepsilon \in \{0.05, 0.1\}$) there is no positive solution that satisfy $A_r x = b_r$ (as well as $A_r^T A_r x = A_r^T b_r$) exactly. This findings we verified by using Farkas's lemma. For instance to verify that $Ax = b, x \geq 0$ has no solution we solved $A^T y >= 0, b^T y < 0$. This situation is reflected also by the high value of the (relative) normal residual (34) at the final iterate, compare the results presented in the next section.

Note that $x^{ex,1}$ is the unique positive solution of $Ax = b_\varepsilon$, for $\varepsilon = 0$, due to its high enough sparsity, as well the solution of minimal ℓ_1-norm. In this cases, $\mathcal{V}$ from (21) consists of only one point for $\varepsilon = 0$ and constrained SIRT will converge according to Theorem 2 to $x^{ex,1}$ in the noiseless (and consistent) case. Otherwise, $\mathcal{V}$ from (21) will be empty. Constrained SIRT will still converge to a global optimum of

$$\min \frac{1}{2}\|Ax - b\|^2 \quad \text{s.t.} \quad x \in \mathcal{B} \ , \tag{32}$$

since we have chosen the weights ω_i as in (12). Value $\|\Delta\|$ represents the distance between this limit point and the least-squares solutions set $LSS(A, b)$.

In all computations we have chosen the steplength $\alpha_k = 2$ (closer to the optimal value $\frac{2}{L}$ then other values of α_k, see Section 4, p. 16) obtaining the constrained classical Cimmino algorithm.

In all computations we used $x^0 = 0$ as an initial approximation and terminating if the relative error at the current iterate x^k is small enough, i.e.,

$$\frac{\|x^k - x^S\|}{\|x^S\|} < 10^{-3}$$

or if the maximum iteration number is reached, i.e. $k \geq k_{max}$, where $k_{max} = 10^4 m_r$. Since a ground truth is not available for all considered examples x^S is chosen to be the solution of (32) for $\mathcal{B} \in \{R_+^n, \Delta_{n,r}, \mathcal{B}_{\ell_1,r}\}$ and $\varepsilon = 0$ obtained by recasting (32) as a linearly constrained quadratic program (QP) and solving it with MOSEK [18]. All radii r are chosen to be the ℓ_1-norms of the minimal ℓ_1-norm solutions of $Ax = b$, (3). Interestingly, all r (approximately) equals the number of particles even in the case of example $I^{ex,2}$. Additionally to the above mentioned criteria we test if

$$K(x^k) = \|x^k - \Pi_{\mathcal{B}}(x^k - \nabla f(x^k))\|_\infty < 10^{-5} \ . \tag{33}$$

This criterion is motivated by the fact that K is continuous in x and zero if and only if x^k is optimal for the constrained problem (32) provided $\mathcal{B}$ is convex. A last criterion involves the relative normal residual

$$\frac{\|A^T(Ax^k - b)\|}{\|A^T b\|} < 10^{-6} \ . \tag{34}$$

Within the implementation of the SPG method we used exactly the same termination criteria. In the experiments presented in the next section we chose the parameters recommended in [5]: $m = 10$, $\alpha_{min} = 10^{-3}$, $\alpha_{max} = 10^3$, $\alpha_0 = \min(\alpha_{max}, \max(\alpha_{min}, 1/\|x^k - \Pi_{\mathcal{B}}(x^k - \nabla f(x^k))\|_\infty))$, $\gamma = 10^{-4}$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.9$.

5.3 Results

Here we summarize the results obtained by the proposed constrained SIRT (20) and the SPG algorithm 1, for all three levels of perturbation. Table 1 and 2 show the results for both example $I^{ex,1}$ and $I^{ex,2}$, for both methods of choice, whereas the reconstructed images are presented in Fig. 1–4. Although numbers and pictures speak for themselves several remarks are in order.

The SPG algorithm clearly outperforms the constrained SIRT in terms of speed (i.e. # iterations). Constraining has different effects onto the number of iterations. For all considered examples and both methods of choice projection onto the simplex yield the lowest number of iterations. This becomes evident especially in the case of $I^{ex,2}$. In these two cases adding positivity constraints seem to be relevant and also have a nice denoising effect which is not given for projection onto $\mathcal{B}_{\ell_1,r}$.

In order to avoid the excessive computation involved in finding an overly accurate solution we also investigated the question when the support of the current iteration is (approximately) equal to that of x^S. This seem to happen only for the limit point. However less iterations are sufficient to yield a fairly reconstruction.

Table 1: Results of SIRT and SPG applied to $I^{ex,1}$

	SIRT			SPG		
	$+$	Δ	ℓ_1	$+$	Δ	ℓ_1
ε	0	0	0	0	0	0
#Iter.	1280000	1280000	1280000	107173	75509	74965
$\frac{\|A^T(Ax^k-b)\|}{\|A^Tb\|}$	2.82e-02	1.99e-02	2.64e-02	2.68e-03	1.71e-03	2.55e-03
$K(x^k)$	4.43e-04	4.64e-04	4.41e-04	9.92e-06	9.89e-06	9.94e-06
$\frac{\|x^k-x^{ex}\|}{\|x^{ex}\|}$	5.78e-01	5.70e-01	5.82e-01	1.21e-01	1.08e-01	1.24e-01
ε	0.05	0.05	0.05	0.05	0.05	0.05
#Iter.	1280000	1280000	1280000	74223	23949	23458
$\frac{\|A^T(Ax^k-b)\|}{\|A^Tb\|}$	2.74e-02	6.82e-01	6.82e-01	2.15e-02	6.82e-01	6.82e-01
$K(x^k)$	3.00e-04	3.33e-04	3.33e-04	9.73e-06	9.36e-06	9.95e-06

Table 1: Results of SIRT and SPG applied to $x^{ex,1}$ (continued)

	SIRT			SPG		
	$+$	Δ	ℓ_1	$+$	Δ	ℓ_1
$\frac{\Vert x^k - x^{ex}\Vert}{\Vert x^{ex}\Vert}$	6.66e-01	5.46e-01	5.46e-0	5.24e-01	1.99e-01	2.00e-01
ε	0.10	0.10	0.10	0.10	0.10	0.10
#Iter.	1280000	1280000	1280000	35935	16985	17291
$\frac{\Vert A^T(Ax^k - b)\Vert}{\Vert A^T b\Vert}$	4.25e-02	1.32e+00	1.32e+00	3.91e-02	1.32e+00	1.32e+00
$K(x^k)$	2.74e-04	4.22e-04	4.22e-04	9.97e-06	9.95e-06	9.99e-06
$\frac{\Vert x^k - x^{ex}\Vert}{\Vert x^{ex}\Vert}$	7.53e-01	5.69e-01	5.69e-01	7.18e-01	3.90e-01	3.90e-01

Table 2: Results of SIRT and SPG applied to $I^{ex,2}$

	SIRT			SPG		
	$+$	Δ	ℓ_1	$+$	Δ	ℓ_1
ε	0	0	0	0	0	0
#Iter.	950000	950000	950000	5246	3512	18686
$\frac{\Vert A^T(Ax^k - b)\Vert}{\Vert A^T b\Vert}$	1.54e-01	1.26e-01	1.59e-04	1.54e-01	1.26e-01	7.01e-05
$K(x^k)$	4.44e-04	2.00e-04	2.60e-05	9.66e-06	5.79e-06	1.00e-05
$\frac{\Vert x^k - x^{ex}\Vert}{\Vert x^{ex}\Vert}$	1.61e-01	4.75e+00	7.33e-01	6.93e-02	4.74e-02	7.37e-01
ε	0.05	0.05	0.05	0.05	0.05	0.05
#Iter.	950000	950000	950000	11798	9559	37348
$\frac{\Vert A^T(Ax^k - b)\Vert}{\Vert A^T b\Vert}$	8.38e-02	6.62e-01	8.17e-05	8.11e-02	6.62e-01	1.46e-05
$K(x^k)$	2.16e-04	3.63e-04	9.81e-05	8.90e-06	9.98e-06	9.96e-06
$\frac{\Vert x^k - x^{ex}\Vert}{\Vert x^{ex}\Vert}$	2.99e-01	2.19e-01	9.99e-01	2.39e-02	1.15e-02	9.92e-01
ε	0.10	0.10	0.10	0.10	0.10	0.10
#Iter.	950000	950000	950000	12175	8993	101476
$\frac{\Vert A^T(Ax^k - b)\Vert}{\Vert A^T b\Vert}$	7.80e-02	1.23e+00	1.54e-04	7.53e-02	1.23e+00	1.67e-05
$K(x^k)$	2.56e-04	1.06e-04	1.90e-04	1.00e-05	9.92e-06	9.99e-06
$\frac{\Vert x^k - x^{ex}\Vert}{\Vert x^{ex}\Vert}$	2.54e-01	4.91e+00	9.99e-01	2.34e-02	7.77e-02	9.81e-01

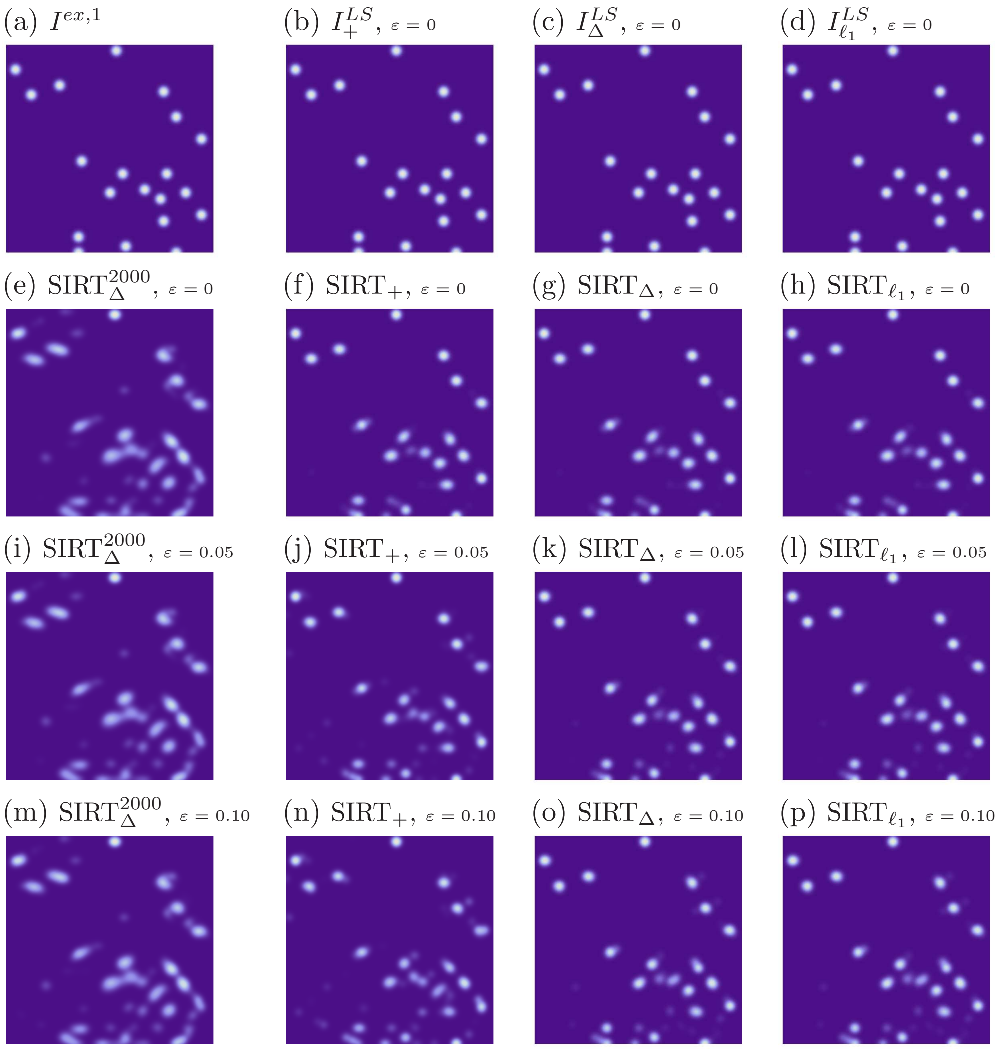

Figure 1. Reconstruction results for image $I^{ex,1}$ (20 particles located randomly at grid positions): (a): Original image. (b)–(d): The reconstructions corresponding to the solutions x^S of (32) obtained via the QP solver of MOSEK [18] for $\varepsilon = 0$ and the three constraining sets, $\mathbb{R}^n_+$, $\Delta_{n,r}$ and $\mathcal{B}_{\ell_1,r}$ respectively, equal $I^{ex,1}$ exactly. (e)–(p): Reconstruction using constrained SIRT algorithm for different perturbation levels. (e),(i),(m): Reconstruction using constrained SIRT algorithm after only 2000 iterations for $\varepsilon \in \{0, 0.05, 0.5\}$ and $\mathcal{B} = \Delta_{n,r}$.

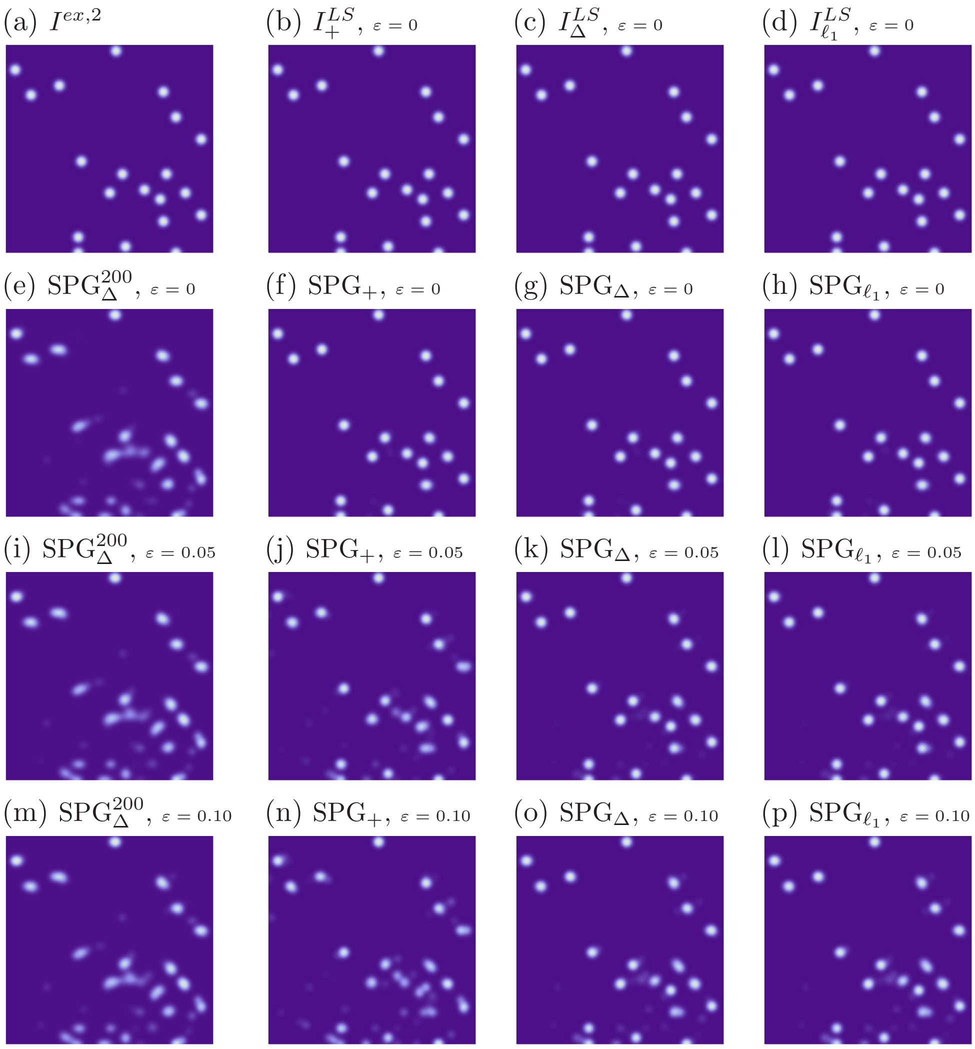

Figure 2. Reconstruction results for image $I^{ex,2}$ (20 particles located randomly at grid positions): (a): Original image. (b)–(d): The reconstructions corresponding to the solutions x^S of (32) obtained via the QP solver of MOSEK [18] for $\varepsilon = 0$ and the three constraining sets, $\mathbb{R}_+^n$, $\Delta_{n,r}$ and $\mathcal{B}_{\ell_1,r}$ respectively, equal $I^{ex,2}$ exactly. (e)–(p): Reconstruction using the SPG algorithm for different perturbation levels. (e),(i),(m): Reconstruction after only 200 iterations of the SPG algorithm for $\varepsilon \in \{0, 0.05, 0.5\}$ and $\mathcal{B} = \Delta_{n,r}$.

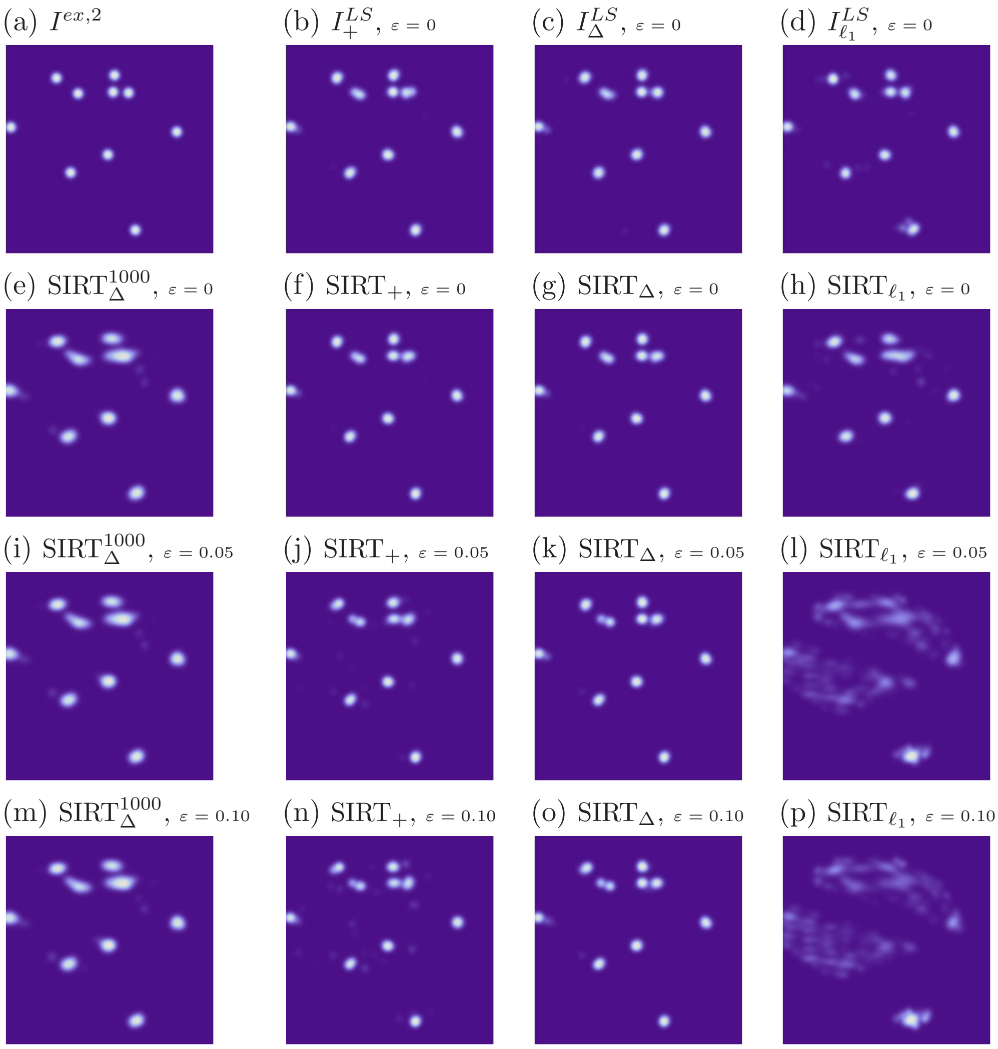

Figure 3. Reconstruction results for image $I^{ex,2}$ (10 particles located randomly in Ω): (a): Original image. (b)–(d): The reconstructions corresponding to the solutions x^S of (32) obtained via the QP solver of MOSEK [18] for $\varepsilon = 0$ and the three constraining sets, $\mathbb{R}^n_+$, $\Delta_{n,r}$ and $\mathcal{B}_{\ell_1,r}$ respectively. (e)–(p): Reconstruction using the constrained SIRT for different perturbation levels. (e),(i),(m): Reconstruction after only 1000 iterations of the constrained SIRT for $\varepsilon \in \{0, 0.05, 0.5\}$ and $\mathcal{B} = \Delta_{n,r}$.

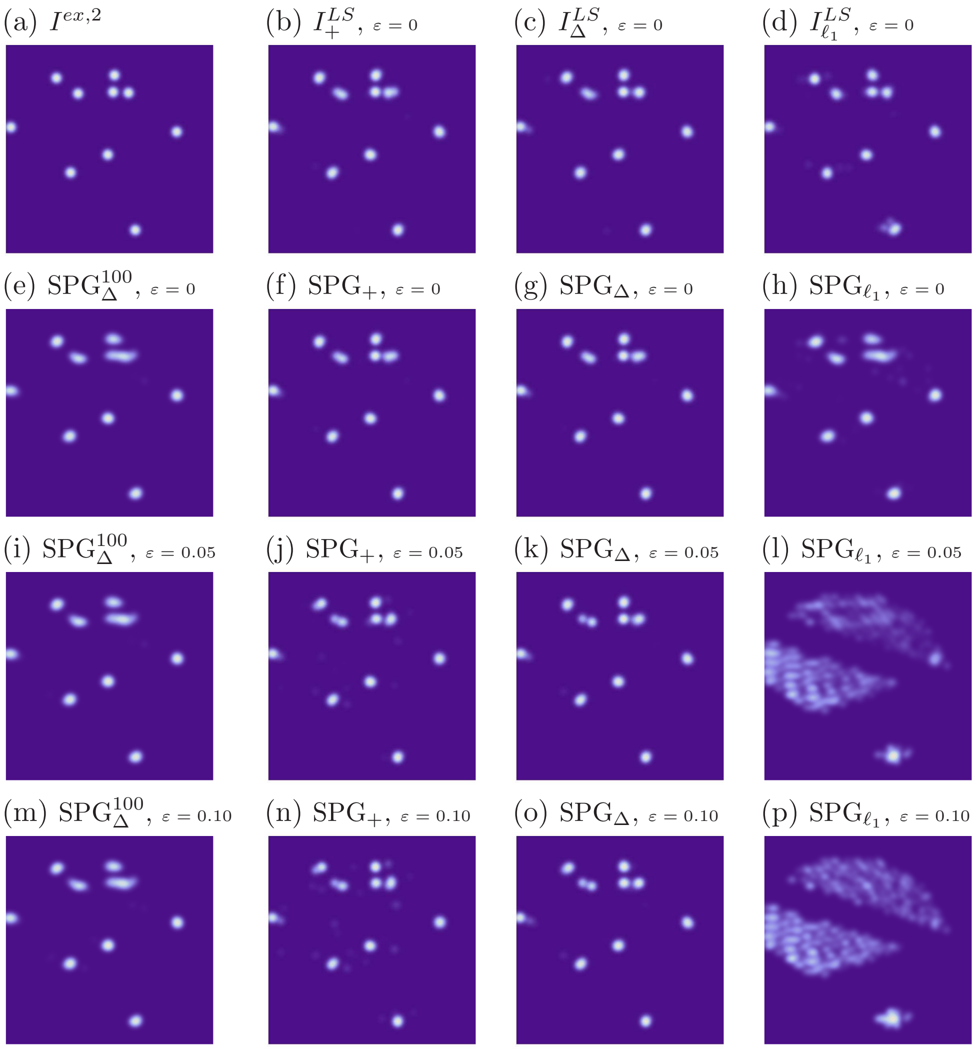

Figure 4. Reconstruction results for image $I^{ex,2}$ (10 particles located randomly in Ω): (a): Original image. (b)–(d): The reconstructions corresponding to the solutions x^S of (32) obtained via the QP solver of MOSEK [18] for $\varepsilon = 0$ and the three constraining sets, $\mathbb{R}_+^n$, $\Delta_{n,r}$ and $\mathcal{B}_{\ell_1,r}$ respectively. (e)–(p): Reconstruction using the SPG algorithm for different perturbation levels. (e),(i),(m): Reconstruction after only 100 iterations of the SPG algorithm for $\varepsilon \in \{0, 0.05, 0.5\}$ and $\mathcal{B} = \Delta_{n,r}$.

6 Conclusion and Further Work

We presented a constrained version of the classical SIRT along with a corresponding convergence analysis for iteratively computing a least-squares solution subject to sparsity constraints. This setting is especially useful when the system matrix is huge and not explicitly available and a solution with high degrees of sparsity is desirable. When the original solution is sparse enough one may use a least-squares fit subject to an ℓ_1-norm constraint on the coefficients. This results in a tractable problem, even though the problem of finding sparse (least-squares) solutions has been cataloged as belonging to a class of combinatorial optimization problems. Successive orthogonal projections onto the (convex) ℓ_1 constraints lend themselves to constraining strategies for the SIRT iterations. Intriguingly, also simple projections of the SIRT iterates onto the positive orthant promote sparsity when the original solution is known to be sparse and positive. A combination of both (thus simplex projections) seem to outperform both in term of quality of the reconstruction.

Moreover, it turns out that constrained SIRT is just a classical gradient projected method. This ensures linear convergence. In practice convergence is very slow. In order to speed up the constrained SIRT we propose choosing larger stepsizes based on the Barzilai-Borwein [2] approach. From the performance viewpoint, this spectral steplength, coupled with a nonmonotone linesearch strategy that accepts the corresponding iterate as frequently as possible, is as a successful idea to accelerate the convergence rate. Its efficiency is then shown on several test problems simulating a challenging real-world application, where it clearly outperforms constrained SIRT. This confirms the received opinion that the spectral steplength is an essential feature for accelerating gradient projection schemes.

References

[1] Aharoni, R. and Censor, Y., *Block-iterative projection methods for parallel computation of solutions to convex feasibility problems*, Linear Algebra Appl. **120**(1989), 165–175.

[2] Barzilai, J. and Borwein, J., *Two point step size gradient methods*, IMA J. Num. Anal., **8**(1988), 141–148.

[3] Bertsekas, D. P., *Nonlinear Programming*, 2nd ed., Athena Scientific, Boston, 1999.

[4] Birgin, E. G., Martinez, J. M. and Raydan, M., *Nonmonotone spectral projected gradient methods on convex sets*, SIAM J. Optimiz., **10**(2000), 1196–1211.

[5] Birgin, E. G., Martinez, J. M. and Raydan, M., *SPG: Software for Convex-Constrained Optimization*, ACM Trans. Math. Softw., **27**(2001), 340–349.

[6] Candès, E., Romberg, J. and Tao, T., *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE T. Inform. Theory, **52**(2006), 489–509.

[7] Candès, E.. and Tao, T., *Near optimal signal recovery from random projections: Universal encoding strategies?*, IEEE T. Inform. Theory, **52**(2006), 5406–5425.

[8] Candès, E. and Romberg, J., *Sparsity and incoherence in compressive sampling*, Inverse Probl., **23**(2007), 969–985.

[9] Censor, Y. and Zenios, S.A., *Parallel Optimization: Theory, Algorithms and Applications*, Oxford University Press, New York, 1997.

[10] Cimmino, G., *Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari*, Ric. Sci. progr. tecn. econom. naz. **1**(1938), 326 – 333.

[11] Combettes, P.L., *Inconsistent signal feasibility problems: least-squares solutions in a product space*, IEEE Trans. Signal Process., **42**(1994), 2955–2966.

[12] Duchi, J., Shalev-Shwartz, S., Singer, Y. and Chandra, T., *Efficient projections onto the ℓ_1-ball for learning in high dimensions*, in Proc. ICML, (2008), 272–279.

[13] Elsinga, G., Scarano, F., Wieneke, B. and van Oudheusden, B., *Tomographic particle image velocimetry*, Exp. Fluids, **41**(2006), 933–947.

[14] Gordon, R., Bender, R. and Herman G.T., *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography*, J. Theor. Biol. **29**(1970), 471–481.

[15] Grippo, L., Lampariello, F. and Lucidi, S., *A nonmonotone line search technique for Newtons method*, SIAM J. Numer. Anal., **23**(1986), 707–716.

[16] Koltracht, I. and Lancaster, P., *Contraining Strategies for Linear Iterative Processes*, IMA J. Numer. Anal. **10**(1990), 55–567.

[17] Levitin, E.S. and Polyak, B.T., *Constrained Minimization Methods*, USSR Comp. Math. and Math. Physics, **6**(1966), 1–50.

[18] Mosek 5.0 - `http://www.mosek.com/`

[19] Petra, S., Schnörr, C., Schröder, A. and Wieneke, B., *Tomographic Image Reconstruction in Experimental Fluid Dynamics: Synopsis and Problems*, In *Proc. 6th Work. Math. Model. Envir. Life Sci. Problems*, Sept 5-9, 2007, Constanta, Romania; Ed Acad Romane, Bucuresti, 2009, ?-?.

[20] Petra, S., Schröder, A., Wieneke, B. and Schnörr, C., *3D Tomography from Few Projections in Experimental Fluid Dynamics*, In *Notes on Numerical Fluid Mechanics and Multidisciplinary Design (NNFM) – Imaging Measurement Methods for Flow Analysis*, Springer, 2009.

[21] Petra, S., Popa, C. and Schnörr, C., *Accelerating Constrained SIRT with Applications in Tomographic Image Reconstruction*, IWR preprint 2009, `http://www.ub.uni-heidelberg.de/archiv/????/`

[22] Tanabe, K., *Projection Method for Solving a Singular System of Linear Equations and its Applications*, Numer. Math., **17**(1971), 203 – 214.

[23] Donoho, D. L. and Tanner, J., *Sparse nonnegative solutions of underdetermined linear equations by linear programming*, Proc. Nat. Acad. Sci. USA, **102**(2005), 9446–9451.