CrossMark

BIT

Majorization–minimization generalized Krylov subspace methods for ℓ_p–ℓ_q optimization applied to image restoration

**G. Huang[1] · A. Lanza[2] · S. Morigi[2] ·
L. Reichel[3] · F. Sgallari[2]**

Abstract A new majorization–minimization framework for ℓ_p–ℓ_q image restoration is presented. The solution is sought in a generalized Krylov subspace that is build up during the solution process. Proof of convergence to a stationary point of the minimized ℓ_p–ℓ_q functional is provided for both convex and nonconvex problems. Computed examples illustrate that high-quality restorations can be determined with a modest number of iterations and that the storage requirement of the method is not very large. A comparison with related methods shows the competitiveness of the method proposed.

Communicated by Lars Eldén.

✉ A. Lanza
   alessandro.lanza2@unibo.it

   G. Huang
   huangx@cdut.edu.cn

   S. Morigi
   serena.morigi@unibo.it

   L. Reichel
   reichel@math.kent.edu

   F. Sgallari
   fiorella.sgallari@unibo.it

[1]   Geomathematics Key Laboratory of Sichuan, College of Management Science, Chengdu University of Technology, Chengdu 610059, People's Republic of China

[2]   Department of Mathematics, University of Bologna, Bologna, Italy

[3]   Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA

⚛ Springer

1 Introduction

This paper is concerned with the efficient computation of approximate solutions of ℓ_p–ℓ_q optimization problems of the form

$$\min_{x \in \mathbb{R}^n} \mathcal{J}(x), \quad \mathcal{J}(x) := \Phi_{\text{fid}}(x) + \mu \, \Phi_{\text{reg}}(x), \tag{1}$$

where $\Phi_{\text{fid}}(x)$ is referred to as the *fidelity term*, $\Phi_{\text{reg}}(x)$ as the *regularization term*, and $\mu > 0$ is a regularization parameter that controls the trade-off between these terms. The fidelity and regularization terms are defined by

$$\Phi_{\text{fid}}(x) := \frac{1}{p} \, \|Ax - b\|_p^p = \frac{1}{p} \sum_{i=1}^r \phi_p((Ax - b)_i) \tag{2}$$

and

$$\Phi_{\text{reg}}(x) := \frac{1}{q} \, \|Lx\|_q^q = \frac{1}{q} \sum_{j=1}^s \phi_q((Lx)_j), \tag{3}$$

respectively, with the function $\phi_z : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ given by

$$\phi_z(t) := |t|^z, \quad z \in \mathbb{R}, \tag{4}$$

where $0 < p, q \leq 2$, $A \in \mathbb{R}^{r \times n}$, $b \in \mathbb{R}^r$, $x \in \mathbb{R}^n$, and $L \in \mathbb{R}^{s \times n}$. We note that the model (1)–(4) is convex and smooth when $1 < p, q \leq 2$, and nonconvex nonsmooth when $0 < p < 1$ or $0 < q < 1$.

Minimization problems of the form (1)–(4) arise in a wide variety of applications and have been studied in several different research areas, including numerical linear algebra [1,28], image restoration [5,14,24], pattern recognition [7,16], and compressive sensing [3–5,8,17]. Different choices of the parameters p and q and of the matrices A and L yield a variety of popular models that have been applied successfully in many fields. For instance, the model (1)–(4) with $p = 2$, $0 < q \leq 1$, $r < n$, and L the identity matrix has been used to compute sparse solutions of undetermined linear systems of equations, while the model with $p = 2$, $0 < q \leq 1$, and A a sampling operator has been applied to compressive sensing; see, e.g., Candes et al. [3,4] for a discussion of the latter.

In this paper we are interested in the application of the model (1)–(4) to image restoration. In this context, the entries of the vector x are the pixel values of the unknown true image that we would like to determine, while the available noise- and possibly blur-contaminated image is represented by the vector b. Typically, both x and b are column-major representations of the corresponding two-dimensional images and are of the same size (that is, $r = n$); however, this is not a requirement of the proposed

method. The matrix A is the identity operator in image denoising problems and a blurring operator when the available image is contaminated by blur. Blurring operators generally are severely ill-conditioned and may be singular. Due to the ill-conditioning of A and the presence of noise in b, minimization of only the fidelity term in (1) typically yields a meaningless computed solution of very large norm. The presence of the regularization term makes it possible to determine an accurate approximation of the unknown true image. The matrix $L \in \mathbb{R}^{s \times n}$ generally is chosen as a discrete gradient operator. For instance, we may let

$$L := \begin{bmatrix} L_1 \\ L_2 \end{bmatrix},$$

where L_1, $L_2 \in \mathbb{R}^{(s/2) \times n}$ are first order difference operators along the horizontal and vertical directions, respectively.

We refer to the image restoration model (1)–(4) as *anisotropic*, because the penalization of Lx depends on the orientation of the image represented by x. This differs from the *isotropic* model, considered in, e.g., [2,13], where the penalization of Lx is independent of the orientation of the image. Specifically, in the isotropic model the penalty terms $\phi_q(\|(Lx)_j\|_2)$, $j = 1, \cdots, s/2$, are used, where

$$(Lx)_j := \begin{bmatrix} (L_1 x)_j \\ (L_2 x)_j \end{bmatrix} \in \mathbb{R}^2, \quad j = 1, \ldots, s/2.$$

One of the most popular and effective approaches for the solution of the ℓ_p–ℓ_q minimization problem defined in (1)–(4) in the general case $0 < p, q \leq 2$ is the iteratively reweighted norm (IRN) algorithm [23], also known as the iteratively reweighted least-squares (IRLS) algorithm [26]. This solution approach is shown to be equivalent to the (multiplicative) half-quadratic method [5] and to the gradient linearization iterative procedure [20]. The IRN method consists of iteratively solving a sequence of penalized weighted least-squares problems that differ from each other only by diagonal weighting matrices. Each one of these least-squares problem is solved by the conjugate gradient (CG) algorithm.

Based on the observation that the weighting matrices generated by the IRN method do not change very quickly during the iterations, the GKSpq approach has been recently proposed in [13]. It represents an extension from the case $p, q = 2$ to the case $0 < p, q \leq 2$ of the GKS method in [12]. Instead of generating a new Krylov subspace at each iteration with the CG method as in the IRN scheme [23], the GKSpq method solves a sequence of reweighted least-squares problems using generalized Krylov subspaces of increasing (and low) dimension. Computed examples in [13] illustrate that this approach may require significantly fewer matrix–vector product evaluations than the IRN scheme.

The idea of this paper is to solve ℓ_p–ℓ_q problems of the form (1)–(4) for $0 < p, q \leq 2$ by coupling the efficient GKS-based minimization procedure proposed in [12] and used in [13] with the robust majorize-minimize (MM) optimization strategy [11]. The resulting approach will be referred to as MM-GKS. The MM strategy applied to the solution of (1)–(4) consists of replacing the original, possibly nonconvex, ℓ_p–ℓ_q

problem by a sequence of simpler convex problems. Specifically, the kth iteration of the standard MM approach relies on two main computational steps: a majorization step which generates a surrogate convex function that majorizes (i.e., bounds above) the $\ell_p-\ell_q$ functional, and a minimization step which determines a minimizer of this majorant function:

- *Majorization:* Generate a majorant $\mathcal{M}(x, x^{(k)})$ for $\mathcal{J}(x)$ near $x^{(k)}$.

- *Minimization:* Compute the next iterate by solving

$$x^{(k+1)} = \arg \min_{x \in \mathbb{R}^n} \mathcal{M}(x, x^{(k)}). \tag{5}$$

Commonly used majorants proposed in the literature include quadratic [23,26], piecewise affine [4,28] and convex-nonconvex [15] functions, which correspond to the so-called ℓ_2, ℓ_1 and CNC majorization strategies. In general, the tighter the majorant is to the objective functional, the fewer iterations are needed for the MM approach to converge, but the higher is the computational cost for carrying out the minimization in (5) at each iteration. A comparison of the performance of methods based on these different majorization strategies is a research problem in its own right. Moreover, the usage of ℓ_1 and CNC majorants is limited to more specific classes of objectives than the quadratic ones.

In this paper, we consider quadratic majorants such that (5) takes the form of a regularized linear least-squares problem, which allows for application of generalized Krylov subspaces. In particular, we consider two different—and to some extent opposite—quadratic majorization strategies, which will be referred to as *adaptive* and *fixed*. Accordingly, two versions of the MM-GKS approach will be proposed and compared, referred to as AMM-GKS (adaptive), and FMM-GKS (fixed). We are particularly (but not only) interested in the robust and efficient solution of the $\ell_p-\ell_q$ model (1)–(4) in its nonconvex (and nonsmooth) regime, that is for $\min\{p, q\} < 1$. In fact, it is well known that image restoration methods based on the use of nonconvex functions Φ_{fid} and/or Φ_{reg} hold the potential for yielding restorations of high quality with sharp edges and homogeneous regions; see, e.g., [5,13,20] for illustrations.

Summarizing, the key contributions of this paper are the following:

(a) Presentation of a new unified formulation of MM methods for $\ell_p-\ell_q$ optimization problems of the form (1)–(4) in the general case $0 < p, q \leq 2$. Our formulation presents a common framework for adaptive and fixed quadratic majorization strategies.
(b) Development of an efficient algorithm for minimizing the (convex) surrogate function in (5), named FMM-GKS strategy.
(c) Analysis of convergence of the discussed MM-GKS methods: we demonstrate that both the AMM-GKS and the FMM-GKS algorithms converge to a stationary point of the $\ell_p-\ell_q$ objective functional in (1)–(4) for any $0 < p, q \leq 2$.

We remark that the AMM-GKS method is mathematically and numerically equivalent to the GKSpq method proposed in [13]. Our derivation in the present paper, based on the MM approach, is new.

This paper is organized as follows. Sections 2 and 3 discuss in detail the majorization and minimization steps of the proposed MM-GKS approach applied to the $\ell_p-\ell_q$ problem (1)–(4). More precisely, Sect. 2 describes the strategy used to generate adaptive or fixed quadratic majorants, whereas Sect. 3 expresses the minimization steps as penalized least-squares problems. The efficient solution of these problems by a suitable implementation of the GKS strategy is considered in Sect. 4. The overall MM-GKS algorithm is summarized in Sect. 5, where the computational cost of both the adaptive and fixed MM strategies is discussed. Convergence of the MM-GKS approach is analyzed in Sect. 6. Numerical examples are presented in Sect. 7 and concluding remarks can be found in Sect. 8.

1.1 Related work

We conclude this section with some comments on alternative approaches to the solution of the minimization problem (1)–(4). In the special case of $p = q = 2$, the minimization problem (1)–(4) is a Tikhonov-regularized linear least-squares problem with regularization matrix L. An efficient iterative algorithm, based on the framework developed by Voss [25] for the solution of nonlinear eigenvalue problems, for the solution of large-scale Tikhonov regularization problems with automatic selection of the regularization parameter μ is proposed in [12]. This method determines successive orthogonal projections onto generalized Krylov subspaces of increasing (and low) dimension. We refer to this kind of solution approach as generalized Krylov subspace (GKS) methods. Candes et al. [4] introduced the iteratively reweighted ℓ_1 (IRL1) method to solve the problem (1)–(4) when $L = I$ and $p = 2$. The aim of this work is to approximate the "0-norm" regularization term. Ramlau and Zarzer [21] describe how, when $p = 2$ in (2) and $0 < q < 1$ in (3), the minimization problem (1) can be transformed into a nonlinear system of equations, which can be solved by a Newton method. The concave–convex procedure described in, e.g., [10,27], is based on expressing the functional (1) as the sum of a convex and a concave function. A solution can be computed with the aid of a majorization–minimization method. The concave–convex procedure can be applied to a variety of optimization problems. A careful comparison of these approaches is outside the scope of the present paper. Here we only note that the fixed majorization–minimization approach of this paper is competitive with the adaptive majorization–minimization described in [13], which in turn is faster than the popular IRN method proposed in [23]. The latter method is related to the attractive half-quadratic algorithm recently discussed in [5].

2 The majorization step

This section introduces fixed and adaptive majorants for the $\ell_p-\ell_q$ functional $\mathcal{J}(x)$ defined by (1)–(4). We are particularly interested in quadratic majorants, for which we give the following definition.

Definition 1 Let $\mathcal{G}(x) : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then the function $\mathcal{Q}(x, v) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is said to be a quadratic tangent majorant for $\mathcal{G}(x)$ if and only if for any $v \in \mathbb{R}^n$ all the following conditions hold:
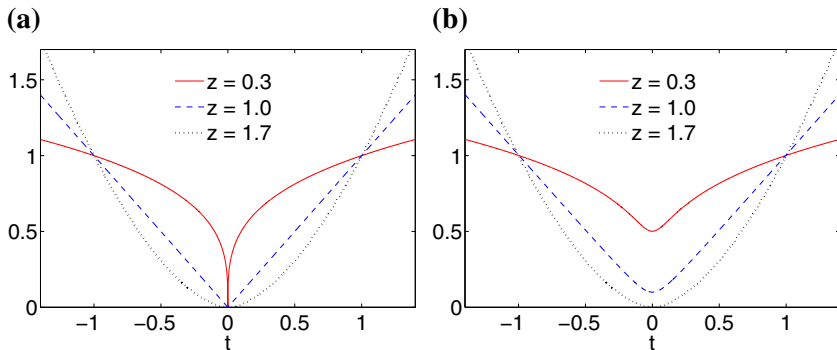
(a)   **(b)**

Fig. 1 Plots of the penalty function $\phi_z(t)$ defined in (4) for some different z-values (**a**), and of the associated smoothed functions $\phi_{z,\varepsilon}(t)$ defined in (6) with $\varepsilon = 0.1$ (**b**)

(c1) $\mathcal{Q}(x, v)$ is quadratic in x,

(c2) $\mathcal{Q}(v, v) = \mathcal{G}(v)$,

(c3) $\nabla_x \mathcal{Q}(v, v) = \nabla_x \mathcal{G}(v)$,

(c4) $\mathcal{Q}(x, v) \geq \mathcal{G}(x) \quad \forall x \in \mathbb{R}^n$,

where ∇_x denotes the gradient with respect to the variable x.

Since the fidelity and regularization terms of $\mathcal{J}(x)$ consist of the sum of r terms with the function ϕ_p and s terms with the function ϕ_q, respectively, majorization of $\mathcal{J}(x)$ reduces to finding majorants for the functions ϕ_p and ϕ_q, that is, in general, for functions ϕ_z in (4) with $z \in\ ]0, 2]$. Note that if $p \in\ ]0, 1]$ (or $q \in\ ]0, 1]$), then the functional $\mathcal{J}(x)$ in (1)–(4) is nonsmooth and all the occurrences of ϕ_p (or ϕ_q) in $\mathcal{J}(x)$ are nonsmooth functions both of their scalar arguments $(Ax - b)_i$ (or $(Lx)_j$) and of the optimization variable x. This follows from the definition of ϕ_z in (4) and is illustrated by the plots of Fig. 1a: if $z \in\ ]1, 2]$, then ϕ_z is smooth and convex, whereas ϕ_z is nonsmooth and convex for $z = 1$, and nonsmooth and nonconvex for $z \in\ ]0, 1[$.

According to Definition 1, in particular condition (c3), quadratic majorization is possible only for continuously differentiable functions $\mathcal{J}(x)$. Therefore the ℓ_p–ℓ_q functional $\mathcal{J}(x)$ in (1)–(4) admits a quadratic tangent majorant for $p, q \in\ ]1, 2]$, but not for $p \in\ ]0, 1]$ and/or $q \in\ ]0, 1]$. For this reason, one commonly smooths the penalty function ϕ_z to make it differentiable also for $z \in\ ]0, 1]$. A popular smoothed version of ϕ_z is given by

$$\phi_{z,\varepsilon}(t) := \left(\sqrt{t^2 + \varepsilon^2}\right)^z \quad \text{with} \quad \begin{cases} \varepsilon > 0 \text{ for } z \in]0, 1], \\ \varepsilon = 0 \text{ for } z \in]1, 2]. \end{cases} \tag{6}$$

To allow quadratic majorization, we thus turn the possibly nonsmooth original ℓ_p–ℓ_q minimization problem (1)–(4) into the following smoothed minimization problem

$$\min_{x\in\mathbb{R}^n} \mathcal{J}_\varepsilon(x), \quad \mathcal{J}_\varepsilon(x) := \frac{1}{p}\sum_{i=1}^{r}\phi_{p,\varepsilon}((Ax-b)_i) + \frac{\mu}{q}\sum_{j=1}^{s}\phi_{q,\varepsilon}((Lx)_j), \quad (7)$$

where a common value $\varepsilon > 0$ of the smoothing parameter is used for the penalty functions in the fidelity and regularization terms. In the remainder of this section, we first discuss how to majorize the one-variable scalar penalty function $\phi_{z,\varepsilon}(t)$ defined by (6), and then present candidate majorants for the smoothed $\ell_p - \ell_q$ functional $\mathcal{J}_\varepsilon(x)$ in (7).

2.1 Majorization of the smoothed penalty function $\phi_{z,\varepsilon}$

The penalty function $\phi_{z,\varepsilon}$ defined by (6) is continuously differentiable with derivative

$$\phi'_{z,\varepsilon}(t) = zt\left(\sqrt{t^2+\varepsilon^2}\right)^{z-2} = zt\phi_{z-2,\varepsilon}(t).$$

Figure 1a, b show the original and smoothed penalty functions ϕ_z and $\phi_{z,\varepsilon}$, respectively, for the parameter z in the interval $]\,0,\,2\,]$ of interest in this paper. Note that the function ϕ_z is not smoothed when $z \in\,]\,1, 2\,]$, i.e., $\phi_{z,\varepsilon}$ coincides with ϕ_z in this case. The following proposition describes the family of all possible quadratic majorants for $\phi_{z,\varepsilon}$.

Proposition 1 *Let $\phi_{z,\varepsilon}(t) : \mathbb{R} \to \mathbb{R}_+$ be the smoothed penalty function defined by (6) with $z \in\,]\,0, 2\,]$. Then any function $m_{z,\varepsilon}(t, v) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ belonging to the a_v-parameterized family determined by*

$$m_{z,\varepsilon}(t, v) := a_v\,(t-b_v)^2 + c_v \qquad (8)$$

with

$$a_v\in[\underline{a_v}, +\infty[, \quad \underline{a_v} = \frac{\phi'_{z,\varepsilon}(v)}{2v} = \frac{z}{2}\phi_{z-2,\varepsilon}(v), \qquad (9)$$

$$b_v = v - \frac{\phi'_{z,\varepsilon}(v)}{2a_v} = v\left(1 - \underline{a_v}/a_v\right), \qquad (10)$$

$$c_v = \phi_{z,\varepsilon}(v) - \frac{\left(\phi'_{z,\varepsilon}(v)\right)^2}{4a_v} = \phi_{z,\varepsilon}(v) - v^2\,\underline{a_v}^2/a_v, \qquad (11)$$

is a quadratic tangent majorant for $\phi_{z,\varepsilon}(t)$, i.e., according to Definition 1:

$$m_{z,\varepsilon}(v, v) = \phi_{z,\varepsilon}(v) \quad \forall v \in \mathbb{R}, \qquad (12)$$

$$m'_{z,\varepsilon}(v, v) = \phi'_{z,\varepsilon}(v) \quad \forall v \in \mathbb{R}, \qquad (13)$$

$$m_{z,\varepsilon}(t, v) \geq \phi_{z,\varepsilon}(t) \quad \forall v \in \mathbb{R}, \quad \forall t \in \mathbb{R}. \qquad (14)$$

Proof Substituting t by v in (8), with a_v, b_v, c_v defined in (9)–(11) gives

$$m_{z,\varepsilon}(v, v) = a_v \left(v - v + v\underline{a_v}/a_v \right)^2 + \phi_{z,\varepsilon}(v) - v^2\underline{a_v}^2/a_v$$
$$= a_v\, v^2\underline{a_v}^2/a_v^2 + \phi_{z,\varepsilon}(v) - v^2\underline{a_v}^2/a_v$$
$$= \phi_{z,\varepsilon}(v),$$

which shows (12). The first-order derivative of the majorant function $m_{z,\varepsilon}$ in (8) with respect to t evaluated at v is given by

$$m'_{z,\varepsilon}(v, v) = 2a_v(v - b_v) = 2a_v \left(v\underline{a_v}/a_v \right) = 2v\underline{a_v}$$
$$= 2v\frac{z}{2}\phi_{z-2,\varepsilon}(v) = \phi'_{z,\varepsilon}(v).$$

This shows (13).

To complete the proof, it suffices to demonstrate that (14) holds if for any $v \in \mathbb{R}$ we take the quadratic majorant (i.e., the parabola) that has the maximum possible aperture, that is if we take $a_v = \underline{a_v}$ with $\underline{a_v}$ defined in (9). Let $m^{(A)}_{z,\varepsilon}(t, v)$ denote the majorant obtained in this way, where the superscript A will be justified below. We notice that all the other parabolas defined by $a_v > \underline{a_v}$ in (9) have a smaller aperture and, hence, are majorants for $m^{(A)}_{z,\varepsilon}(t, v)$. It is immediate to verify that replacing a_v by $\underline{a_v}$ in (10) yields $b_v = 0$, that is $m^{(A)}_{z,\varepsilon}(t, v)$ is an even function of the variable t. Since $\phi_{z,\varepsilon}(t)$ is even as well, to prove (14) it suffices to demonstrate that, for any $v > 0$, $m^{(A)}_{z,\varepsilon}(t, v)$ is a majorant for $\phi_{z,\varepsilon}(t)$ for any $t > 0$. Moreover, since both $m^{(A)}_{z,\varepsilon}(t, v)$ and $\phi_{z,\varepsilon}(t)$ are continuously differentiable functions, we can write

$$m^{(A)}_{z,\varepsilon}(t, v) = m^{(A)}_{z,\varepsilon}(v, v) + \int_v^t m'^{(A)}_{z,\varepsilon}(\tau, v)\, d\tau, \tag{15}$$

$$\phi_{z,\varepsilon}(t) = \phi_{z,\varepsilon}(v) + \int_v^t \phi'_{z,\varepsilon}(\tau)\, d\tau. \tag{16}$$

Subtracting (16) from (15) and recalling (12), we obtain

$$m^{(A)}_{z,\varepsilon}(t, v) - \phi_{z,\varepsilon}(t) = \int_v^t (m'^{(A)}_{z,\varepsilon}(\tau, v) - \phi'_{z,\varepsilon}(\tau))\, d\tau$$
$$= \int_v^t (2\underline{a_v}\,\tau - \phi'_{z,\varepsilon}(\tau))\, d\tau$$
$$= \int_v^t \left(\frac{\phi'_{z,\varepsilon}(v)}{v}\tau - \phi'_{z,\varepsilon}(\tau) \right) d\tau$$
$$= \int_v^t \tau \left(\frac{\phi'_{z,\varepsilon}(v)}{v} - \frac{\phi'_{z,\varepsilon}(\tau)}{\tau} \right) d\tau$$
$$= z \int_v^t \tau (\phi_{z-2,\varepsilon}(v) - \phi_{z-2,\varepsilon}(\tau))\, d\tau\ . \tag{17}$$
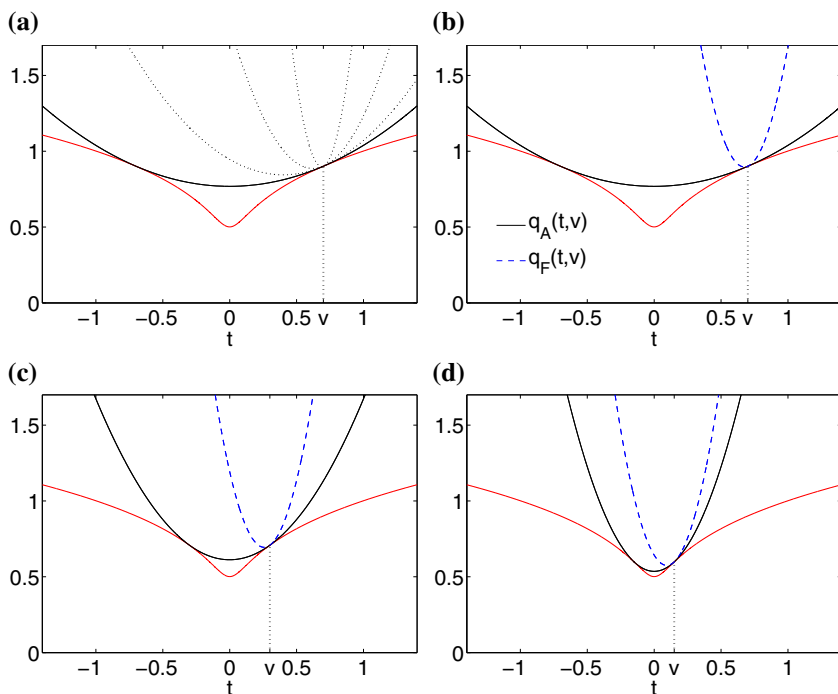
Fig. 2 Plots of some quadratic majorants $m_{z,\varepsilon}(t, v)$ at $v = 0.7$ for $\phi_{z,\varepsilon}(t)$ (*solid red graph*) (**a**), and plots of the adaptive quadratic majorants $m_{z,\varepsilon}^{(A)}(t, v)$ defined by (19) (*solid black graph*) and of the fixed quadratic majorants $m_{z,\varepsilon}^{(F)}(t, v)$ defined by (21) (*dashed blue graph*) for the majorization points $v = 0.7$ (**b**), $v = 0.3$ (**c**), $v = 0.15$ (**d**)

Now rewriting (17), taking the limits of integration into account, yields

$$
m_{z,\varepsilon}^{(A)}(t, v) - \phi_{z,\varepsilon}(t) =
\begin{cases}
z \displaystyle\int_v^t \tau(\phi_{z-2,\varepsilon}(v) - \phi_{z-2,\varepsilon}(\tau))d\tau & \text{if } t > v, \\
0 & \text{if } t = v, \\
z \displaystyle\int_t^v \tau\,(\phi_{z-2,\varepsilon}(\tau) - \phi_{z-2,\varepsilon}(v))d\tau & \text{if } t < v.
\end{cases}
\tag{18}
$$

Since $z \in\,]0, 2\,]$, it follows that the function

$$
\phi_{z-2,\varepsilon}(t) = \frac{1}{\left(t^2 + \varepsilon^2\right)^{1-z/2}}
$$

is monotonically decreasing in t for $t > 0$ for any fixed $\varepsilon \geq 0$. Hence, the integrands in (18) are nonnegative for any τ in their domains of integration with $t, v > 0$. Therefore, the two integrals in (18) are nonnegative and $m_{z,\varepsilon}^{(A)}(t, v)$ is a majorant for $\phi_{z,\varepsilon}(t)$.

Figure 2a displays several majorants at $v = 0.7$ for the nonconvex smoothed penalty function $\phi_{z,\varepsilon}$ with $z = 0.3$ and $\varepsilon = 0.1$, which is depicted in solid red. The solid black

curve in Fig. 2a shows the maximum-aperture majorant parabola $m_{z,\varepsilon}^{(A)}(t, v)$ obtained by setting $a_v = \underline{a_v}$ in (8) with $\underline{a_v}$ defined in (9), while the dotted black curves display three majorant parabolas at the same point $v = 0.7$ with $a_v > \underline{a_v}$. The following definition introduces two important alternative quadratic majorization strategies.

Definition 2 (Adaptive and fixed quadratic majorants) Let $\phi_{z,\varepsilon}(t)$ be the smoothed penalty function defined in (6) with $z \in \,]0, 2\,]$ and let $m_{z,\varepsilon}(t, v)$ be the family of associated quadratic tangent majorants defined in (8)–(11).

– If for every $v \in \mathbb{R}$, the parameter a_v is chosen as the lower limit of the admissible interval given in (9), that is $a_v = \underline{a_v}$, then the quadratic majorant (8) takes the **adaptive** form:

$$
\begin{aligned}
m_{z,\varepsilon}^{(A)}(t, v) &= \underline{a_v}t^2 + \phi_{z,\varepsilon}(v) - v^2\underline{a_v} \\
&= \frac{z}{2}\phi_{z-2,\varepsilon}(v)t^2 + \underbrace{\phi_{z,\varepsilon}(v) - v^2\frac{z}{2}\phi_{z-2,\varepsilon}(v)}_{\text{independent of } t}.
\end{aligned} \tag{19}
$$

– If for every $v \in \mathbb{R}$ the parameter a_v is chosen independently of v as follows

$$
a_v = \bar{a}_{z,\varepsilon}, \quad \bar{a}_{z,\varepsilon} = \max_{v\in\mathbb{R}}\underline{a_v} =, \max_{v\in\mathbb{R}}\left(\frac{\phi'_{z,\varepsilon}(v)}{2v}\right) = \frac{z}{2}\varepsilon^{z-2}, \tag{20}
$$

then the quadratic majorant (8) takes the **fixed** form:

$$
\begin{aligned}
m_{z,\varepsilon}^{(F)}(t, v) &= \bar{a}_{z,\varepsilon}(t - v(1 - \underline{a_v}/\bar{a}_{z,\varepsilon}))^2 + \phi_{z,\varepsilon}(v) - v^2\underline{a_v}^2/\bar{a}_{z,\varepsilon} \\
&= \bar{a}_{z,\varepsilon}(t^2 - 2v(1 - \underline{a_v}/\bar{a}_{z,\varepsilon}t) \\
&\quad + \phi_{z,\varepsilon}(v) + v^2(\bar{a}_{z,\varepsilon}(1 - \underline{a_v}/\bar{a}_{z,\varepsilon})^2 - \underline{a_v}^2/\bar{a}_{z,\varepsilon}) \\
&= \bar{a}_{z,\varepsilon}t^2 - 2\,v(\bar{a}_{z,\varepsilon} - \underline{a_v})t + \phi_{z,\varepsilon}(v) - v^2(2\underline{a_v} - \bar{a}_{z,\varepsilon}) \\
&= \frac{z}{2}(\varepsilon^{z-2}t^2 - 2\,v(\varepsilon^{z-2} - \phi_{z-2,\varepsilon}(v))t) \\
&\quad + \underbrace{\phi_{z,\varepsilon}(v) - v^2\frac{z}{2}(2\phi_{z-2,\varepsilon}(v) - \varepsilon^{z-2})}_{\text{independent of } t}.
\end{aligned} \tag{21}
$$

The last equality in (20) follows from the fact that the maximum is achieved at $v = 0$. Therefore, among all possible adaptive majorants (one for each majorization point $v \in \mathbb{R}$), the one that is tangent to the penalty function $\phi_{z,\varepsilon}(t)$ at $t = 0$ has the smallest aperture. This can be observed in Fig. 2b–d, where we show the adaptive and fixed quadratic majorants $m_{z,\varepsilon}^{(A)}(t, v)$ and $m_{z,\varepsilon}^{(F)}(t, v)$ for the same nonconvex smoothed penalty function $\phi_{z,\varepsilon}(t)$ of Fig. 2a for different majorization points v. One can notice that both majorants $m_{z,\varepsilon}^{(A)}(t, v)$ and $m_{z,\varepsilon}^{(F)}(t, v)$ vary with v. However, $m_{z,\varepsilon}^{(F)}(t, v)$ only translates with fixed aperture, whereas $m_{z,\varepsilon}^{(A)}(t, v)$ also adapts its aperture, hence the names fixed and adaptive.

2.2 Majorization of the smoothed $\ell_p - \ell_q$ functional

Based on the results in the previous section, we introduce candidate quadratic majorants for the smoothed $\ell_p-\ell_q$ functional $\mathcal{J}_\varepsilon(x)$ in (7). More precisely, at each iteration k, we seek to construct a quadratic function $\mathcal{Q}(x, x^{(k)})$ that is a tangent majorant for $\mathcal{J}_\varepsilon(x)$ near the previously computed iterate $x^{(k)}$, so that the new iterate $x^{(k+1)}$ can be sought by minimizing $\mathcal{Q}(x, x^{(k)})$. Introduce the majorization points for the penalty functions $\phi_{p,\varepsilon}$ and $\phi_{q,\varepsilon}$ of the fidelity and regularization terms of $\mathcal{J}_\varepsilon(x)$, respectively:

$$v_i^{(k)} := \left(Ax^{(k)} - b \right)_i, \quad i = 1, \ldots, r, \tag{22}$$

$$u_j^{(k)} := \left(Lx^{(k)} \right)_j, \quad j = 1, \ldots, s. \tag{23}$$

The scalars $v_i^{(k)}$ and $u_j^{(k)}$ should be considered constants, since the previous iterate $x^{(k)}$ is fixed. We can now construct the quadratic majorant $\mathcal{Q}(x, x^{(k)})$ for $\mathcal{J}_\varepsilon(x)$ in (7) by replacing all penalty functions $\phi_{p,\varepsilon}$ and $\phi_{q,\varepsilon}$ by their associated quadratic tangent majorants at the points $v_i^{(k)}$ and $u_j^{(k)}$, and in (22) and (23) using the formulas of Definition 2. In the following subsections, we derive expressions for the adaptive and fixed quadratic majorants used in the computed examples.

2.2.1 Adaptive quadratic majorants

Replacing the functions $\phi_{p,\varepsilon}$ and $\phi_{q,\varepsilon}$ in (7) by the associated adaptive quadratic majorants defined in (19) at the points $v_i^{(k)}$ and $u_j^{(k)}$ given in (22) and (23), respectively, we obtain the following adaptive quadratic majorant $\mathcal{Q}^{(A)}(x, x^{(k)})$ for $\mathcal{J}_\varepsilon(x)$ at the point $x^{(k)}$:

$$\begin{aligned}
\mathcal{Q}^{(A)}(x, x^{(k)}) &= \frac{1}{p} \sum_{i=1}^{r} m_{p,\varepsilon}^{(A)}((Ax - b)_i, v_i^{(k)}) + \frac{\mu}{q} \sum_{j=1}^{s} m_{q,\varepsilon}^{(A)}((Lx)_j, u_j^{(k)}) \\
&= \frac{1}{2} \sum_{i=1}^{r} \phi_{p-2,\varepsilon}(v_i^{(k)})(Ax - b)_i^2 + \frac{\mu}{2} \sum_{j=1}^{s} \phi_{q-2,\varepsilon}(u_j^{(k)})(Lx)_j^2 + c,
\end{aligned} \tag{24}$$

where we collect all terms that are independent of x in the term c.

Define the vectors $w_{\text{fid}}^{(k)} \in \mathbb{R}^r$ and $w_{\text{reg}}^{(k)} \in \mathbb{R}^s$ of majorization weights for the fidelity and regularization terms, respectively, as follows:

$$w_{\text{fid}}^{(k)} = \phi_{p-2,\varepsilon}(v^{(k)}) = ((v^{(k)})^2 + \varepsilon^2)^{p/2-1}, \tag{25}$$

$$w_{\text{reg}}^{(k)} = \phi_{q-2,\varepsilon}(u^{(k)}) = ((u^{(k)})^2 + \varepsilon^2)^{q/2-1}, \tag{26}$$

where all the operations in (25), (26) are component-wise and the last equalities in (25), (26) follow from the definition of the smoothed penalty function $\phi_{z,\varepsilon}$ in (6). Introduce the diagonal matrices

$$W_{\mathrm{fid}}^{(k)} = \mathrm{diag}(w_{\mathrm{fid}}^{(k)}) \in \mathbb{R}^{r \times r}, \quad W_{\mathrm{reg}}^{(k)} = \mathrm{diag}(w_{\mathrm{reg}}^{(k)}) \in \mathbb{R}^{s \times s}.$$

They allow us to express the adaptive quadratic majorant in (24) in the following compact form:

$$\mathcal{Q}^{(A)}(x, x^{(k)}) = \frac{1}{2} \|(W_{\mathrm{fid}}^{(k)})^{1/2}(Ax - b)\|_2^2 + \frac{\mu}{2} \|(W_{\mathrm{reg}}^{(k)})^{1/2}Lx\|_2^2 + c. \quad (27)$$

2.2.2 Fixed quadratic majorants

Replacing the functions $\phi_{p,\varepsilon}$ and $\phi_{q,\varepsilon}$ in (7) by the associated fixed quadratic majorants defined in (21) at the points $v_i^{(k)}$ and $u_j^{(k)}$ given in (22) and (23), respectively, we obtain the following fixed quadratic majorant $\mathcal{Q}^{(F)}(x, x^{(k)})$ for $\mathcal{J}_\varepsilon(x)$ at the point $x^{(k)}$:

$$\begin{aligned}
\mathcal{Q}^{(F)}(x, x^{(k)}) &= \frac{1}{p} \sum_{i=1}^{r} m_{p,\varepsilon}^{(F)}((Ax - b)_i, v_i^{(k)}) + \frac{\mu}{q} \sum_{j=1}^{s} m_{q,\varepsilon}^{(F)}((Lx)_j, u_j^{(k)}) \\
&= \frac{\varepsilon^{p-2}}{2} \sum_{i=1}^{r} \left[(Ax - b)_i^2 - 2v_i^{(k)} \left(1 - \frac{\phi_{p-2,\varepsilon}(v_i^{(k)})}{\varepsilon^{p-2}} \right) (Ax - b)_i \right] \\
&\quad + \frac{\mu \varepsilon^{q-2}}{2} \sum_{j=1}^{s} \left[(Lx)_j^2 - 2u_j^{(k)} \left(1 - \frac{\phi_{q-2,\varepsilon}(u_j^{(k)})}{\varepsilon^{q-2}} \right) (Lx)_j \right] + c,
\end{aligned}$$
$$(28)$$

where similarly as above, we collect all terms that are independent of x in the term c. Define the two vectors $w_{\mathrm{fid}}^{(k)} \in \mathbb{R}^r$ and $w_{\mathrm{reg}}^{(k)} \in \mathbb{R}^s$ of majorization weights for the fidelity and regularization terms by:

$$w_{\mathrm{fid}}^{(k)} = v^{(k)} \left(1 - \frac{\phi_{p-2,\varepsilon}(v^{(k)})}{\varepsilon^{p-2}} \right) = v^{(k)} \left(1 - \left(\frac{(v^{(k)})^2 + \varepsilon^2}{\varepsilon^2} \right)^{p/2-1} \right), \quad (29)$$

$$w_{\mathrm{reg}}^{(k)} = u^{(k)} \left(1 - \frac{\phi_{q-2,\varepsilon}(u^{(k)})}{\varepsilon^{q-2}} \right) = u^{(k)} \left(1 - \left(\frac{(u^{(k)})^2 + \varepsilon^2}{\varepsilon^2} \right)^{q/2-1} \right). \quad (30)$$

It follows that the fixed quadratic majorant (28) can be expressed in the compact form:

$$\begin{aligned}
\mathcal{Q}^{(F)}(x, x^{(k)}) &= \frac{\varepsilon^{p-2}}{2} (\|Ax - b\|_2^2 - 2\langle w_{\mathrm{fid}}^{(k)}, Ax \rangle) \\
&\quad + \frac{\mu \varepsilon^{q-2}}{2} (\|Lx\|_2^2 - 2\langle w_{\mathrm{reg}}^{(k)}, Lx \rangle) + c.
\end{aligned}$$
$$(31)$$

3 The minimization step

Recalling the definitions (27) and (31) of the adaptive and fixed quadratic majorants $Q^{(A)}(x, x^{(k)})$ and $Q^{(F)}(x, x^{(k)})$, respectively, the minimization steps in the kth iteration of the adaptive and fixed MM-GKS approaches can be written as

$$x^{(k+1)} = \arg \min_{x \in \mathbb{R}^n} \left[\left\| \left(W_{\text{fid}}^{(k)}\right)^{1/2} (Ax - b) \right\|_2^2 + \mu \left\| \left(W_{\text{reg}}^{(k)}\right)^{1/2} Lx \right\|_2^2 \right] \qquad (32)$$

and

$$x^{(k+1)} = \arg \min_{x \in \mathbb{R}^n} \left[\|Ax - b\|_2^2 - 2 \left\langle w_{\text{fid}}^{(k)}, Ax \right\rangle + \eta \left(\|Lx\|_2^2 - 2 \left\langle w_{\text{reg}}^{(k)}, Lx \right\rangle \right) \right], \quad (33)$$

respectively, where the constant terms and the factor $1/2$ have been omitted in both (32) and (33), and where we have introduced the constant

$$\eta := \mu \frac{\varepsilon^{q-2}}{\varepsilon^{p-2}}$$

in (33). The optimization problems (32) and (33) are regularized (or penalized) least-squares problems. Introduce the $n \times n$ matrices

$$T^{(A)}\left(W_{\text{fid}}, W_{\text{reg}}\right) := A^T W_{\text{fid}} A + \mu L^T W_{\text{reg}} L, \qquad (34)$$

$$T^{(F)} := A^T A + \eta L^T L. \qquad (35)$$

The normal equations associated with the adaptive and fixed quadratic minimization problems (32) and (33) can be written as

$$T^{(A)}\left(W_{\text{fid}}^{(k)}, W_{\text{reg}}^{(k)}\right) x = A^T W_{\text{fid}}^{(k)} b, \qquad (36)$$

$$T^{(F)} x = A^T \left(b + w_{\text{fid}}^{(k)}\right) + \eta L^T w_{\text{reg}}^{(k)}. \qquad (37)$$

Note that the matrix $T^{(F)}$ of the normal equations (37) associated with the fixed majorization strategy does not depend on the majorization weights, that is, it is fixed during the MM-GKS iterations. We will see in the next section how this property allows for efficient solution.

The linear systems of Eqs. (36) and (37) have unique solutions if the matrices $T^{(A)}$ and $T^{(F)}$ are nonsingular for all iterations, that is, if the following conditions hold for all k:

$$\text{Ker}(A^T W_{\text{fid}}^{(k)} A) \cap \text{Ker}(L^T W_{\text{reg}}^{(k)} L) = \{0\}, \qquad (38)$$

$$\text{Ker}(A^T A) \cap \text{Ker}(L^T L) = \{0\}, \qquad (39)$$

where $\mathrm{Ker}(M)$ denotes the null space of the matrix M. We notice that the constants μ in (34) and η in (35) are positive. Since the majorization weights are positive as well, the diagonal matrices $W_{\mathrm{fid}}^{(k)}$ and $W_{\mathrm{reg}}^{(k)}$ are positive definite. Therefore, condition (38) is equivalent to condition (39). The latter condition is typically satisfied for image restoration problems, because A represents a blurring operator, which is a low-pass filter, while the regularization matrix L usually is a difference operator and, hence, is a high-pass filter.

4 Efficient solution of the normal equations

When (39) holds, the matrices $T^{(A)}$ and $T^{(F)}$ in (34), (35) are symmetric positive definite. They generally also are structured and quite sparse, depending on the size of the blur kernel. Therefore, the normal equations (36), (37) can be solved quite efficiently iteratively by the conjugate gradient (CG) method.

The popular IRN approach [23] consists of an adaptive quadratic majorization step and a CG-based minimization step: at each (outer) iteration $k = 0, 1, \ldots$ of the MM iterative procedure (5), the new iterate $x^{(k+1)}$ is obtained by applying the CG method to solve the normal equations in (36) with the matrix $T^{(A)}$ defined in (34) and the adaptive majorization weights given in (25), (26). For each k, the CG method generates a new Krylov subspace, because the matrix $T^{(A)}$ varies with k due to adaptivity of the majorization weights.

The basic idea behind the GKSpq approach described in [13] is to determine generalized Krylov subspaces of increasing (and not very large) dimension instead of determining several (standard) Krylov subspaces. We use generalized Krylov subspaces also in the method proposed in the present paper. They are constructed as follows. First, an initial user-chosen Krylov subspace $\mathcal{V}_0 \subset \mathbb{R}^n$ is generated. Similarly as in [12], we define the subspace $\mathcal{V}_0 = \mathcal{K}_l(A^T A, A^T b)$ for some $l \leq 5$. Let the columns of the matrix $V_0 \in \mathbb{R}^{n \times l}$ form an orthonormal basis for the space $\mathcal{V}_0$.

Let $\mathcal{V}_k$ denote the generalized Krylov subspace at step k of our solution method. It is of dimension $k + l$. Let the columns of the matrix V_k form an orthonormal basis for this subspace. We will now describe how the columns are determined. Let $x^{(k+1)} = V_k y^{(k+1)}$ denote the solution of (36) restricted to the subspace $\mathcal{V}_k$. The vector $y^{(k+1)}$ is computed by solving the reduced least-squares problem

$$\min_{y \in \mathbb{R}^{k+l}} \left\| \begin{bmatrix} \left(W_{\mathrm{fid}}\right)^{1/2} A V_k \\ \mu^{1/2} \left(W_{\mathrm{reg}}\right)^{1/2} L V_k \end{bmatrix} y - \begin{bmatrix} \left(W_{\mathrm{fid}}\right)^{1/2} b \\ 0 \end{bmatrix} \right\|_2^2. \tag{40}$$

Similarly, the solution $x^{(k+1)} = V_k y^{(k+1)}$ of (37) restricted to $\mathcal{V}_k$ is computed by determining the solution $y^{(k+1)}$ of the least-squares problem

$$\min_{y \in \mathbb{R}^{k+l}} \left\| \begin{bmatrix} A V_k \\ \eta^{1/2} L V_k \end{bmatrix} y - \begin{bmatrix} b + w_{\mathrm{fid}}^{(k)} \\ \eta^{1/2} w_{\mathrm{reg}}^{(k)} \end{bmatrix} \right\|_2^2. \tag{41}$$

4.1 GKS for the fixed case

Let $V_k \in \mathbb{R}^{n \times d}$, $d = k + l \ll n$, and introduce the QR factorizations

$$AV_k = Q_A R_A \quad \text{with} \quad Q_A \in \mathbb{R}^{r \times d}, \quad R_A \in \mathbb{R}^{d \times d},$$
$$LV_k = Q_L R_L \quad \text{with} \quad Q_L \in \mathbb{R}^{s \times d}, \quad R_L \in \mathbb{R}^{d \times d},$$

where the matrices Q_A and Q_L have orthonormal columns and the matrices R_A and R_L are upper triangular. Substituting these factorizations into (41) yields the low-dimensional minimization problem

$$\min_{y \in \mathbb{R}^{k+l}} \left\| \begin{bmatrix} R_A \\ \eta^{1/2} R_L \end{bmatrix} y - \begin{bmatrix} Q_A^T (b + w_{\text{fid}}^{(k)}) \\ \eta^{1/2} Q_L^T w_{\text{reg}}^{(k)} \end{bmatrix} \right\|_2^2 \tag{42}$$

with the associated normal equations

$$(R_A^T R_A + \eta R_L^T R_L) y = R_A^T Q_A^T (b + w_{\text{fid}}^{(k)}) + \eta R_L^T Q_L^T w_{\text{reg}}^{(k)}.$$

The residual $r^{(k+1)}$ of (37) can be computed according to

$$\begin{aligned} r^{(k+1)} &= T^{(F)} x^{(k+1)} - A^T (b + w_{\text{fid}}^{(k)}) - \eta L^T w_{\text{reg}}^{(k)} \\ &= A^T (AV y^{(k+1)} - b - w_{\text{fid}}^{(k)}) + \eta L^T (LV y^{(k+1)} - w_{\text{reg}}^{(k)}). \end{aligned} \tag{43}$$

Following Voss [25], the subspace $\mathcal{V}_k$ is expanded to $\mathcal{V}_{k+1}$ by adding a new basis vector v_{new} to $\mathcal{V}_k$, with the latter chosen as the normalized residual $r^{(k+1)}$. Thus,

$$v_{\text{new}} := \frac{r^{(k+1)}}{\|r^{(k+1)}\|_2}, \quad V_{k+1} := [V_k, v_{\text{new}}].$$

To enforce orthogonality in the presence of round-off errors, the residual $r^{(k+1)}$ is reorthogonalized against V_k before normalization.

When the new vector v_{new} is added to the solution subspace, the matrices AV_k and LV_k are updated to obtain AV_{k+1} and LV_{k+1}, respectively. We can implement these updates as follows

$$AV_{k+1} := [AV_k, Av_{\text{new}}], \quad LV_{k+1} := [LV_k, Lv_{\text{new}}].$$

This requires the evaluation of the matrix–vector products Av_{new} and Lv_{new}. Then the QR factorizations $AV_k = Q_A R_A$ and $LV_k = Q_L R_L$ are updated according to

$$A[V_k, v_{\text{new}}] = [Q_A, \tilde{q}_{A,k+1}] \begin{bmatrix} R_K & r_{K,k+1} \\ 0 & \tau_{K,k+1} \end{bmatrix},$$

$$L[V_k, v_{\text{new}}] = [Q_L, \tilde{q}_{L,k+1}] \begin{bmatrix} R_L & r_{L,k+1} \\ 0 & \iota_{L,k+1} \end{bmatrix};$$

see Daniel et al. [6] for a detailed discussion on updating and downdating of the QR factorization. The new vectors in the above factorizations are obtained by

$$
\begin{aligned}
r_{A,k+1} &= Q_A^T(Av_{\mathrm{new}}), \quad q_{A,k+1} = Av_{\mathrm{new}} - r_{A,k+1}, \\
\tau_{A,k+1} &= \|q_{A,k+1}\|, \quad \tilde{q}_{A,k+1} = q_{A,k+1}/\tau_{A,k+1}, \\
r_{L,k+1} &= Q_L^T(Lv_{\mathrm{new}}), \quad q_{L,k+1} = Lv_{\mathrm{new}} - r_{L,k+1}, \\
\iota_{L,k+1} &= \|q_{L,k+1}\|, \quad \tilde{q}_{L,k+1} = q_{L,k+1}/\iota_{L,k+1}.
\end{aligned}
\tag{44}
$$

4.2 GKS for the adaptive case

Let $V_k \in \mathbb{R}^{n \times d}$, $d = k + l \ll n$, and introduce the QR factorizations

$$
W_{\mathrm{fid}}^{1/2} A V_k = Q_A R_A \quad \text{with} \quad Q_A \in \mathbb{R}^{r \times d}, \quad R_A \in \mathbb{R}^{d \times d}, \tag{45}
$$

$$
W_{\mathrm{reg}}^{1/2} L V_k = Q_L R_L \quad \text{with} \quad Q_L \in \mathbb{R}^{s \times d}, \quad R_L \in \mathbb{R}^{d \times d}. \tag{46}
$$

Thus, the matrices Q_A and Q_L have orthonormal columns and the matrices R_A and R_L are upper triangular. Substituting these factorizations into (40) yields the low-dimensional minimization problem

$$
\min_{y \in \mathbb{R}^{k+l}} \left\| \begin{bmatrix} R_A \\ \mu^{1/2} R_L \end{bmatrix} y - \begin{bmatrix} Q_A^T W_{\mathrm{fid}}^{1/2} b \\ 0 \end{bmatrix} \right\|_2^2 \tag{47}
$$

with the associated normal equations

$$
\left(R_A^T R_A + \mu R_L^T R_L \right) y = R_A^T Q_A^T W_{\mathrm{fid}} b.
$$

Recalling the definition (34) of the matrix $T^{(A)}(W_R, W_F)$ and using the fact that $x^{(k+1)} = V_k y^{(k+1)}$, the residual $r^{(k+1)}$ of (36) can be computed according to

$$
\begin{aligned}
r^{(k+1)} &= T(W_{\mathrm{fid}}, W_{\mathrm{reg}}) \, x^{(k+1)} - A^T W_{\mathrm{fid}} b \\
&= (A^T W_{\mathrm{fid}} A + \mu L^T W_{\mathrm{reg}} L) \, x^{(k+1)} - A^T W_{\mathrm{fid}} b \\
&= A^T W_{\mathrm{fid}} (A V_k y^{(k+1)} - b) + \mu L^T W_{\mathrm{reg}} (L V_k y^{(k+1)}) .
\end{aligned}
\tag{48}
$$

The subspace $\mathcal{V}_k$ is then expanded to $\mathcal{V}_{k+1}$ in the same fashion as above, that is by adding the new unit vector v_{new} obtained by normalizing the residual $r^{(k+1)}$ in (48).

5 The MM-GKS algorithm

Algorithm 1 shows the main steps of the MM-GKS iterative method. Both the adaptive and fixed approaches are described by Algorithm 1, but clearly they have been implemented as separated algorithms that we will denote by AMM-GKS and FMM-GKS, respectively.

Algorithm 1 Main computational steps of the proposed MM-GKS method for the solution of the (smoothed) ℓ_p–ℓ_q optimization problem in (6), (7)

---

inputs: $A \in \mathbb{R}^{r \times n}$, $L \in \mathbb{R}^{s \times n}$ s.t. (39) holds, $b \in \mathbb{R}^r$, $0 < p, q \leq 2$, $\mu > 0$
output: approximate solution x^* of (6), (7)

1. Initialization: • set $x^{(0)} = b$

   • generate the initial subspace: $V_0 \in \mathbb{R}^{n \times l}$ s.t. $V_0^T V_0 = I$

2. **for** $k = 0, 1, \ldots$ until convergence **do**

   • Majorization:

3. • generate the quadratic majorant $\mathcal{Q}^{(A)}(x, x^{(k)})$ (or $\mathcal{Q}^{(F)}(x, x^{(k)})$)
   by computing the weights $w_{\text{fid}}^{(k)}$, $w_{\text{reg}}^{(k)}$ in (25), (26) (or (29), (30))

   • Minimization:

4a. • compute (or update) the QR factorizations by (45)–(46) (or (45))

4b. • compute the solution $y^{(k+1)}$ of reduced problem (47) (or (42))

   • GKS update:

5a. • compute the residual $r^{(k+1)}$ of (36) by (48) (or of (37) by (43))

5b. • compute the new basis vector v_{new} by normalizing $r^{(k+1)}$

5c. • enlarge the GKS and update matrices AV and LV:

   $$V_{k+1} = \left[V_k, v_{\text{new}} \right], \quad AV_{k+1} = \left[AV_k, Av_{\text{new}} \right], \quad LV_{k+1} = \left[LV_k, Lv_{\text{new}} \right]$$

6. **end for**

7. Compute the approximate solution $x^* = V_k y^{(k+1)}$

The solution of the minimization step is determined in a generalized Krylov subspace. We comment on a few details of the algorithm starting with the computational cost. An orthonormal basis for the initial solution subspace $\mathcal{V}_0 = \mathcal{K}_l(A^T A, A^T b)$ is determined by carrying out l steps of Golub–Kahan bidiagonalization of the matrix A with initial vector b. This requires the evaluation of $l - 1$ matrix–vector products (MVPs) with the matrix A and l MVPs with A^T. The matrix $A^T A$ does not have to be formed. The computation of an orthonormal basis for $\mathcal{K}_l(A^T A, A^T b)$ is fairly inexpensive when l is small. For efficiency reasons, we store and update the matrices AV_k and LV_k during the iterations.

We choose the initial approximate solution $x_0 = b$ in Algorithm 1. It therefore would appear natural to let the initial Krylov subspace be $\mathcal{K}_l(A^T A, A^T r_0)$ with $r_0 = b - Ax_0$ instead of $\mathcal{K}_l(A^T A, A^T b)$. However, numerical experiments show that the latter subspace generally gives more accurate approximations of the desired image in fewer iterations than the former. We therefore define the initial Krylov subspace $\mathcal{V}_0$ as described.

The cost of the majorization Step 3 reduces to the computation of the vectors of majorization weights $w_{\text{fid}}^{(k)}$ and $w_{\text{reg}}^{(k)}$ which is negligible. In fact, the vectors $A V_k y^{(k+1)} - b$ and $L V_k y^{(k+1)}$ already have been computed in (48) when evaluating the residual $r^{(k+1)}$ in the previous iteration and are reused. Concerning the minimization step, the cost in Step 4b for solving the normal equations projected into the generalized Krylov subspace $\mathcal{V}_k$, is negligible for $k \ll n$, which is always the case.

The main computational difference between the AMM-GKS and FMM-GKS algorithms is in Step 4a of Algorithm 1. In the AMM-GKS algorithm, the computation of the QR factorizations (45), (46) demands about $2r(k+l)^2$ and $2s(k+l)^2$ arithmetic floating point operations (flops), respectively. The FMM-GKS algorithm does not require the recomputation of the QR factorizations. Instead, these factorizations are updated according to (45). The cost is dominated by two MVPs, one with the matrix $Q_A^T \in \mathbb{R}^{(l+k) \times r}$ and one with the matrix $Q_L^T \in \mathbb{R}^{(l+k) \times s}$. The computational effort is negligible in comparison with the MVPs required in Step 4a of AMM-GKS.

As far as the GKS updating step is concerned, the computational effort required to determine the residual in Step 5a is dominated by one MVP with each one of the matrices A^T and L^T for both the AMM-GKS and FMM-GKS algorithms. The cost of Step 5b is clearly negligible, whereas the updates of the matrices $A V_k$ and $L V_k$ in Step 5c requires two MVPs, one with A and one with L.

In summary, the overall computational cost for k iterations of the AMM-GKS and FMM-GKS algorithms is dominated by the work required to evaluate $4k$ MVPs with matrices A, A^T, L, and L^T, and for the AMM-GKS algorithm in addition to the computations required for Step 4a. The latter computations are not required by the FMM-GKS algorithm.

6 Convergence analysis

In this section we analyze the convergence of the MM-GKS approach, whose main steps are given in Algorithm 1, in both its adaptive and fixed forms. The presented results extend the convergence analysis carried out in [13], which was limited to the AMM-GKS method and to the convex case $1 \le p, q \le 2$.

First, we notice that for securing that the MM-GKS approach does not break down, we have to require that at each iteration k the solutions $y^{(k+1)}$ of the reduced least-squares problems (40) or (41) exist and are unique. We therefore assume that condition (39) is satisfied; cf. the discussion at the end of Sect. 3.

Using the MM formulation, our MM-GKS approach can be written as follows:

$$x^{(k+1)} = \begin{cases} \arg\min_{x \in \mathcal{V}_k} Q(x, x^{(k)}) \text{ for } k = 0, 1, \ldots, n-l-1, \\ \arg\min_{x \in \mathbb{R}^n} Q(x, x^{(k)}) \text{ for } k = n-l, n-l+1, \ldots, \end{cases} \quad (49)$$

where $l \ge 1$ is the dimension of the user-specified initial subspace $\mathcal{V}_0$, $\mathcal{V}_k$ is the generalized Krylov subspace used at iteration k, and $Q(x, x^{(k)})$ is either the adaptive

quadratic majorant $Q^{(A)}(x, x^{(k)})$ or the fixed quadratic majorant $Q^{(F)}(x, x^{(k)})$. To justify (49), we recall that the subspace $\mathcal{V}_k$, in which in the kth iteration the MM-GKS algorithm computes the new approximate solution $x^{(k+1)}$, is of dimension $l+k$. Hence, from iteration $n - l$ and onwards, the subspace $\mathcal{V}_k$ is equivalent to the whole space $\mathbb{R}^n$ and cannot be enlarged further.

In the following, we denote by $\{x^{(k)}\}_{k\geq 1}$ the sequence of iterates generated by the MM-GKS algorithm applied to the minimization of the smoothed ℓ_p–ℓ_q functional defined in (6), (7) with $0 < p, q \leq 2$. Properties of the sequence of values $\mathcal{J}_\varepsilon(x^{(k)})$, $k = 0, 1, \ldots$, are described by the following theorem, whose proof can be found in [13].

Theorem 2 *Let condition* (39) *hold. Then, for any initial guess* $x^{(0)} \in \mathbb{R}^n$, *the sequence* $\{\mathcal{J}_\varepsilon(x^{(k)})\}_{k\geq 0}$ *is monotonically nonincreasing and convergent.*

In the remainder of this section, we analyze the behavior of the sequence of iterates $\{x^{(k)}\}_{k\geq 1}$.

Proposition 3 *For any initial guess* $x^{(0)} \in \mathbb{R}^n$ *and for any* $k \geq 0$, *the majorization error function* $\mathcal{E}(x, x^{(k)})$, *defined by*

$$\mathcal{E}(x, x^{(k)}) := Q(x, x^{(k)}) - \mathcal{J}_\varepsilon(x), \tag{50}$$

has the following properties:

$$\mathcal{E}(x, x^{(k)}) \in \mathcal{C}^1(\mathbb{R}^n), \tag{51}$$
$$\mathcal{E}(x, x^{(k)}) \geq 0 \quad \forall x \in \mathbb{R}^n, \tag{52}$$
$$\mathcal{E}(x^{(k)}, x^{(k)}) = 0, \tag{53}$$
$$\nabla_x \mathcal{E}(x^{(k)}, x^{(k)}) = 0, \tag{54}$$
$$\nabla_x \mathcal{E}(x^{(k+1)}, x^{(k)}) = -\nabla_x \mathcal{J}_\varepsilon(x^{(k+1)}). \tag{55}$$

Moreover, $\nabla_x \mathcal{E}(x, x^{(k)})$ *is L-Lipschitz continuous, i.e., there exists a constant $L > 0$ such that*

$$\left\| \nabla_x \mathcal{E}(x_1, x^{(k)}) - \nabla_x \mathcal{E}(x_2, x^{(k)}) \right\|_2 \leq L \|x_1 - x_2\|_2 \quad \forall x_1, x_2 \in \mathbb{R}^n. \tag{56}$$

Proof Property (51) follows immediately from the definition of the majorization error function (50), and from $Q(x, x^{(k)})$ and $\mathcal{J}_\varepsilon(x)$ being continuously differentiable functions of x. Properties (52)–(54) follow from definition (50) and from $Q(x, x^{(k)})$ being a quadratic tangent majorant for $\mathcal{J}_\varepsilon(x)$ at $x^{(k)}$ according to Definition 1. Turning to property (55), we obtain from definition (50) that

$$\nabla_x \mathcal{E}(x^{(k+1)}, x^{(k)}) = \nabla_x Q(x^{(k+1)}, x^{(k)}) - \nabla_x \mathcal{J}_\varepsilon(x^{(k+1)}). \tag{57}$$

Since $\mathcal{Q}(x, x^{(k)})$ is a strictly convex quadratic function of x and $x^{(k+1)}$ represents its global minimizer, $\nabla_x \mathcal{Q}(x^{(k+1)}, x^{(k)}) = 0$, and (55) is a consequence of (57). Finally, property (56) follows from the gradients of the functions $\mathcal{Q}(x, x^{(k)})$ and $\mathcal{J}_\varepsilon(x)$ being L-Lipschitz continuous.

Definition 3 A convex (not necessarily differentiable) function $f(x)$ is said to be δ-*strongly convex* if and only if there exists a constant $\delta > 0$, called the modulus of strong convexity of f, such that the function $f(x) - \frac{\delta}{2} \|x\|_2^2$ is convex.

The following result is shown in [18,19].

Lemma 4 *Let* $f(x) : \mathbb{R}^n \to \mathbb{R}$ *be a* δ-*strongly convex function, and let* $x^* \in \mathbb{R}^n$ *be a minimizer of* $f(x)$. *Then*

$$\frac{\delta}{2}\|x - x^*\|_2^2 \leq f(x) - f(x^*) \quad \forall x \in \mathbb{R}^n. \tag{58}$$

Theorem 5 *Let condition* (39) *hold. Then, for any initial guess* $x^{(0)} \in \mathbb{R}^n$, *the sequence* $\{x^{(k)}\}_{k \geq 1}$ *converges to a stationary point of* $\mathcal{J}_\varepsilon(x)$. *Thus,*

   a. $\lim_{k \to \infty} \| x^{(k+1)} - x^{(k)} \|_2 = 0,$

   b. $\lim_{k \to \infty} \nabla_x \mathcal{J}_\varepsilon(x^{(k)}) = 0.$

Proof Let us consider the quadratic majorant function $\mathcal{Q}(x, x^{(k)})$ generated by the MM-GKS approach at iteration k. Notice that the global minimizer of $\mathcal{Q}(x, x^{(k)})$ is the next iterate $x^{(k+1)}$ of the MM-GKS algorithm. Since $\mathcal{Q}(x, x^{(k)})$ is δ-strongly convex, we can apply Lemma 4. In particular, inequality (58) with the function $\mathcal{Q}(\cdot, x^{(k)})$ in place of $f(\cdot)$ and $x^{(k+1)}$ in place of x^* yields

$$\frac{\delta}{2}\|x - x^{(k+1)}\|_2^2 \leq \mathcal{Q}(x, x^{(k)}) - \mathcal{Q}(x^{(k+1)}, x^{(k)}) \quad \forall x \in \mathbb{R}^n, \quad \forall k. \tag{59}$$

Substituting the iterate $x^{(k)}$ for x in (59), we obtain

$$\frac{\delta}{2}\|x^{(k)} - x^{(k+1)}\|_2^2 \leq \mathcal{Q}(x^{(k)}, x^{(k)}) - \mathcal{Q}(x^{(k+1)}, x^{(k)}) \tag{60}$$

$$= \mathcal{J}_\varepsilon(x^{(k)}) + \cancel{\mathcal{E}(x^{(k)}, x^{(k)})} - \mathcal{J}_\varepsilon(x^{(k+1)}) - \mathcal{E}(x^{(k+1)}, x^{(k)}) \tag{61}$$

$$\leq \mathcal{J}_\varepsilon(x^{(k)}) - \mathcal{J}_\varepsilon(x^{(k+1)}) \quad \forall k. \tag{62}$$

Properties (53) and (52) of the majorization error function allowed us to cancel out the term in (61) and to bound (61) by (62).

Summing the inequalities (60)–(62) over k yields

$$
\sum_{k=0}^{\infty} \| x^{(k+1)} - x^{(k)} \|_2^2 \leq \frac{2}{\delta} \sum_{k=0}^{\infty} \left[\mathcal{J}_\varepsilon(x^{(k)}) - \mathcal{J}_\varepsilon(x^{(k+1)}) \right]
$$
$$
= \frac{2}{\delta} \big(\underbrace{\mathcal{J}_\varepsilon(x^{(0)}) - \mathcal{J}_\varepsilon(x^{(1)})}_{k=0} + \underbrace{\mathcal{J}_\varepsilon(x^{(1)}) - \mathcal{J}_\varepsilon(x^{(2)})}_{k=1} + \cdots \big)
$$
$$
= \frac{2}{\delta} (\mathcal{J}_\varepsilon(x^{(0)}) - \mathcal{J}_\varepsilon^*), \tag{63}
$$

where $\mathcal{J}_\varepsilon^*$ denotes the limit point of the sequence $\{\mathcal{J}_\varepsilon(x^{(k)})\}_{k\geq 0}$, which is monotonically nonincreasing and convergent according to Theorem 2. Hence, $\mathcal{J}_\varepsilon(x^{(0)}) - \mathcal{J}_\varepsilon^*$ is a finite nonnegative number. It follows that the series on the left-hand side of inequality (63) is convergent. We conclude that statement a. holds.

Regarding the gradient of the smoothed ℓ_p–ℓ_q functional $\mathcal{J}_\varepsilon(x)$, we obtain

$$
\left\| \nabla_x \mathcal{J}_\varepsilon(x^{(k+1)}) \right\|_2 = \left\| \nabla_x \mathcal{E}(x^{(k+1)}, x^{(k)}) \right\|_2 \tag{64}
$$
$$
= \left\| \nabla_x \mathcal{E}(x^{(k+1)}, x^{(k)}) - \nabla_x \mathcal{E}(x^{(k)}, x^{(k)}) \right\|_2 \tag{65}
$$
$$
\leq L \left\| x^{(k+1)} - x^{(k)} \right\|_2 \quad \forall k, \tag{66}
$$

where (64)–(66) follow from properties (55)–(56) of the majorization error function, respectively. Since the elements of the sequence $\left\{ \|x^{(k+1)} - x^{(k)}\|_2 \right\}_{k\geq 0}$ converge to zero as k increases, it follows from (64)–(66) that the elements of the sequence $\left\{ \|\nabla \mathcal{J}_\varepsilon(x^{(k+1)})\|_2 \right\}_{k\geq 0}$ also converge to zero. This shows statement b. $\qquad\square$

Corollary 6 *Let condition* (39) *hold. Then, in case that* $p \geq 1$ *and* $q \geq 1$, *for any initial guess* $x^{(0)} \in \mathbb{R}^n$, *the sequence* $\left\{ x^{(k)} \right\}_{k\geq 0}$ *converges towards the unique global minimizer of the smoothed* ℓ_p–ℓ_q *functional defined in* (7).

Proof The proof is immediate by recalling that when $p, q \geq 1$, the smoothed ℓ_p–ℓ_q functional in (7) is strictly convex and, therefore, admits a unique stationary point at its global minimizer.

An analysis of the behavior of $\left\{ x^{(k)} \right\}$ for $k \leq n - l$ is beyond the scope of this paper and will be considered in future work. In the following section we will provide some empirical evidence of the theoretical results shown above.

7 Numerical examples

We compare the performances of the adaptive MM method (AMM-GKS) and the fixed MM method (FMM-GKS) described by Algorithm 1 with the IRN algorithm when applied to the restoration of two gray-scale test images `cameraman` and `QRcode` depicted in Figs. 3a and 5a. These images are synthetically corrupted by blur and noise. They are represented by arrays of 256×256 pixels stored column-wise in vectors in $\mathbb{R}^n$ with $n = 65{,}536$. Let $\bar{x} \in \mathbb{R}^n$ represent the original blur- and noise-free image. This

image is assumed not to be available. A block Toeplitz with Toeplitz blocks blurring matrix $A \in \mathbb{R}^{n \times n}$ is generated with the function `blur` from [9]. This function has the parameters `band` and σ, that determine the bandwidth of each Toeplitz block in A and the standard deviation of the underlying Gaussian point spread function, respectively. The blurred image $A\bar{x}$ is corrupted by impulsive noise to obtain the image $b \in \mathbb{R}^n$. This image is assumed to be known. It is our aim to determine an as accurate as possible approximation x^* of $\bar{x}$, given A and b, by using the ℓ_p–ℓ_q model (7). We recall that the AMM-GKS method is equivalent to the GKSpq method described in [13].

Quantitative and qualitative measures of accuracy and efficiency are investigated. The accuracy is measured by the signal-to-noise ratio (SNR) defined by

$$\text{SNR}(x^*, \bar{x}) := 10 \log_{10} \frac{\|\bar{x} - E(\bar{x})\|_2^2}{\|x^* - \bar{x}\|_2^2} \text{ (dB)},$$

where $E(\bar{x})$ denotes the mean gray-level of the uncontaminated image $\bar{x}$. This quantity provides a quantitative measure of the quality of the restored image: a large SNR-value indicates that x^* is an accurate approximation of $\bar{x}$. Computational efficiency is measured in terms of the total computational time and the total number of matrix–vector product (MVP) evaluations required by the algorithms to reach convergence. The (outer) iterations of the three algorithms considered are terminated as soon as one of the following stopping criteria is satisfied:

1. The relative change of the computed approximate solution $x^{(k)}$ drops below a user-specified threshold, i.e., we terminate the iterations as soon as

$$\|x^{(k+2)} - x^{(k)}\|_2 / \|x^{(k)}\|_2 < 10^{-4}.$$

2. The number of (outer) iterations is 1000.

Results for the IRN algorithm are computed by using the freely available implementation in the NUMIPAD library [22]. We chose the regularization parameter empirically to give the most accurate restorations, i.e., restorations with the highest SNR-values. To allow for a fair comparison, for each restoration example, we chose the same regularization parameter for all the algorithms. The AMM-GKS and FMM-GKS algorithms use the initial search space $\mathcal{V}_0 = \mathcal{K}_1(A^T A, A^T b)$. The examples were run on an Intel®Core™i3 Quad-Core 2.27 GHz computer with 4 GB of RAM, using Windows 7 OS and 64 bits arithmetic. All computations are carried out in MATLAB with about 15 significant decimal digits.

Example 1 We consider the restoration of a contaminated test image `cameraman` that has been degraded by Gaussian blur with different values of the parameters `band` and σ, and by salt-and-pepper noise of different intensity. Recall that salt-and-pepper noise corrupts images by changing a given percentage of pixels into either the minimum or maximum possible gray-level-value with equal probability. The other pixels are left unchanged. We consider the ℓ_p–ℓ_1 restoration model (7) for $0 < p \leq 1$. Specifically, we consider the two models ℓ_1–ℓ_1 and $\ell_{0.7}$–ℓ_1. The ℓ_1-norm for the fidelity term is a standard choice for the restoration of images that are corrupted by salt-and-pepper
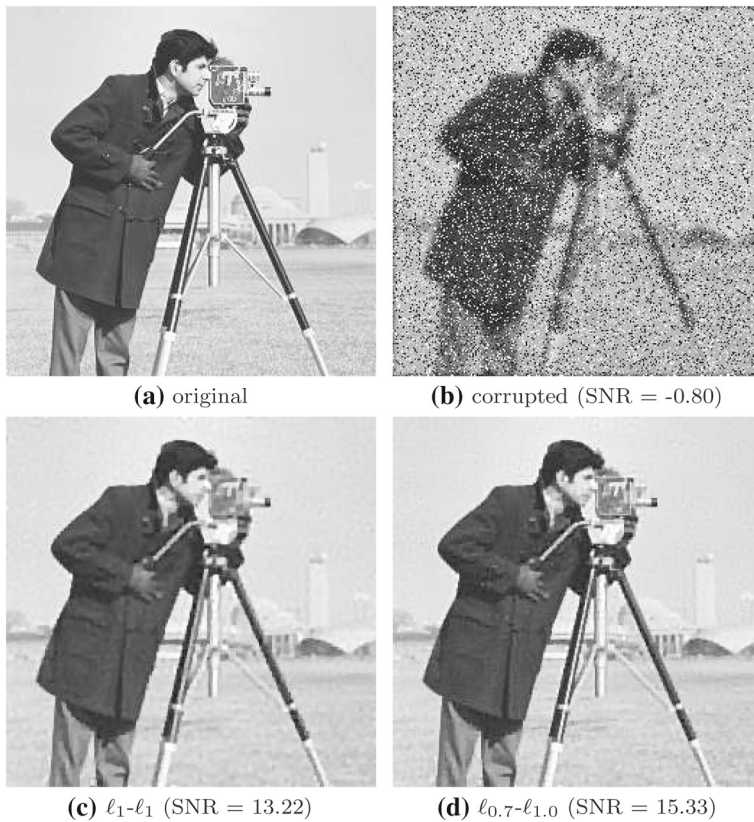
(a) original   **(b)** corrupted (SNR = -0.80)

(c) ℓ_1-ℓ_1 (SNR = 13.22)   **(d)** $\ell_{0.7}$-$\ell_{1.0}$ (SNR = 15.33)

Fig. 3 Example 1. Restoration results (**c**, **d**) obtained by the FMM-GKS algorithm applied to the `cameraman` image (**a**) that has been contaminated by Gaussian blur with `band = 7` and $\sigma = 2.0$, and salt-and-pepper noise corrupting 20% of the pixels (**b**)

noise, see [5], since it forces sparsity of the residual while keeping the functional convex. The $\ell_{0.7}$-norm induces sparsity better, but leads to a nonconvex optimization problem. We are interested in comparing the three algorithms in terms of restoration quality and computational effort required for the computation of the approximate solutions of the ℓ_1–ℓ_1 and $\ell_{0.7}$–ℓ_1 models.

Figure 3 shows the restorations determined by the FMM-GKS algorithm when applied to the test image. The degraded image in Fig. 3b was obtained from the original blur- and noise-free image in Fig. 3a, by first applying Gaussian blur with parameters `band = 7` and $\sigma = 2.0$, and then corrupting 20% of the pixels by salt-and-pepper noise. Figure 3c, d depict restorations obtained by the FMM-GKS method when solving the $\ell_1 - \ell_1$ and $\ell_{0.7}$–ℓ_1 models, respectively. The latter model yields a significantly more accurate restoration than the former both in terms of visual quality and SNR-value.

Table 1 reports quantitative results of the performance of the IRN, AMM-GKS, and FMM-GKS algorithms when applied to the ℓ_1–ℓ_1 and $\ell_{0.7}$–ℓ_1 models for differ-

Table 1 Example 1

Blur		Noise		Efficiency: time (iterations, MVPs)			Accuracy: SNR		
Band	σ	%	μ	IRN	AMM	FMM	IRN	AMM	FMM
$\ell_1 - \ell_1$									
7	2.0	10	0.004	235.66	144.20	40.35	13.86	13.91	13.98
				(36,7368)	(168,672)	(197,788)			
		20	0.010	115.00	71.20	33.38	13.21	13.22	13.22
				(40,3484)	(125,500)	(177,708)			
		30	0.020	101.79	53.71	35.75	12.56	12.56	12.55
				(48,3004)	(111,1464)	(177,708)			
9	2.5	10	0.004	303.12	163.08	42.83	13.00	12.99	12.98
				(39,6182)	(177,708)	(202,808)			
		20	0.005	291.40	155.69	44.63	12.01	12.01	12.05
				(42,5892)	(174,696)	(203,812)			
		30	0.020	180.29	65.89	29.23	11.64	11.65	11.69
				(55,3586)	(123,492)	(162,648)			
$\ell_{0.7}-\ell_1$									
7	2.0	10	0.004	273.11	327.59	65.69	16.31	16.31	16.31
				(31,8658)	(225,900)	(251,1004)			
		20	0.007	214.51	216.74	55.26	15.34	15.34	15.33
				(34,6768)	(192,768)	(245,980)			
		30	0.013	165.11	136.22	74.48	14.71	14.70	14.67
				(36,5164)	(165,660)	(293,1172)			
9	2.5	10	0.004	497.87	427.96	70.49	15.20	15.19	15.15
				(34,10292)	(256,1024)	(274,1096)			
		20	0.006	430.04	300.07	69.08	14.29	14.28	14.26
				(37,8838)	(224,896)	(265,1060)			
		30	0.010	365.49	224.34	67.20	13.47	13.47	13.43
				(41,7450)	(201,804)	(266,1064)			

Comparison of the IRN, AMM-GKS, and FMM-GKS algorithms applied to the $\ell_1 - \ell_1$ and $\ell_{0.7}-\ell_1$ models for restoration of the cameraman test images corrupted by Gaussian blur and salt-and-pepper noise

ent Gaussian blurs, defined by the parameters band and σ, and for salt-and-pepper noise, defined by the percentage of corrupted pixels. The fourth column of Table 1 reports the value of the regularization parameter μ used. Efficiency results in terms of computational time (in seconds), number of (outer) iterations, and number of MVPs required by the algorithms to reach convergence are reported in columns 5–7, while the accuracy in terms of SNR-values is shown in columns 8–10.

Table 1 shows the MM-GKS approaches to typically require significantly fewer MVP evaluations and less computing time than the IRN algorithm. The AMM-GKS method requires fewer iterations (and consequently fewer MVPs) than the FMM-GKS method. This depends on the fact that the adaptive majorants used by AMM-GKS are tighter approximations of the $\ell_p-\ell_q$ functional than the fixed majorants of

FMM-GKS. However, an iteration with the AMM-GKS method is more demanding computationally than an iteration with the FMM-GKS method. This is mainly due to the necessity to refactor the reduced least-squares problem at each iteration. The benefit of the FMM-GKS method is illustrated by the lower computational times required than by the AMM-GKS method. Table 1 shows the nonconvex $\ell_{0.7}-\ell_1$ model to give restorations of higher quality than the convex $\ell_1-\ell_1$ model.

Finally, Fig. 4 shows the convergence of the FMM-GKS algorithm when applied to the convex $\ell_1-\ell_1$ and the nonconvex $\ell_{0.7}-\ell_1$ models for the restoration of the cameraman image perturbed by Gaussian blur with parameters band $= 9$, $\sigma =$
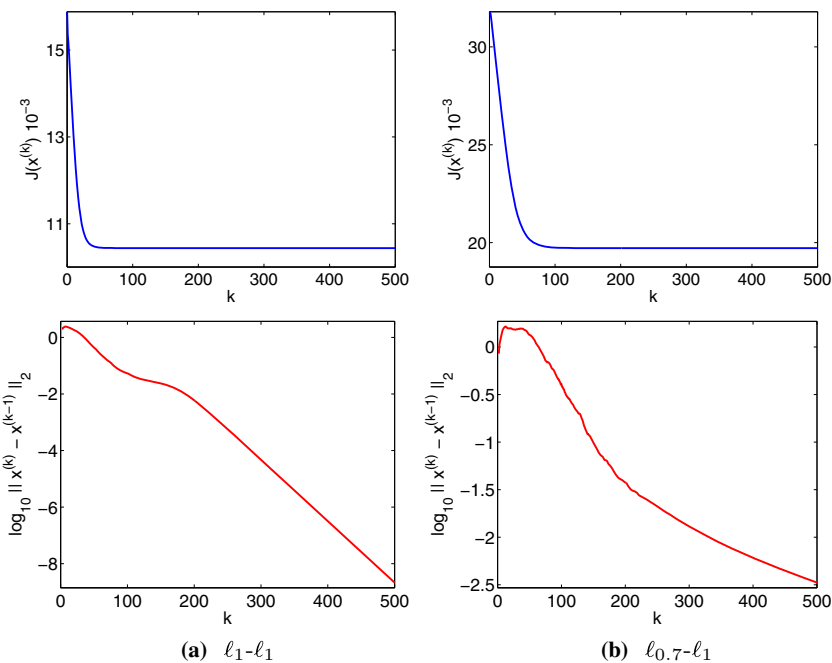

(a) ℓ_1-ℓ_1 **(b)** $\ell_{0.7}$-ℓ_1

Fig. 4 Example 1. Empirical convergence plots for the FMM-GKS algorithm applied to the restoration of the cameraman test image

Table 2 Example 2

p	q	μ	SNR
1.0	1.0	0.06	25.61
0.3	1.0	0.20	31.91
0.1	1.0	0.50	32.28
0.1	0.8	0.35	32.72
0.1	0.5	0.05	34.83

SNR-values obtained by the FMM-GKS algorithm for different values of p and q when restoring the QRcode test image corrupted by Gaussian blur defined by band $= 5$ and $\sigma = 2.0$, and 20% salt-and-pepper noise

2.5, and salt-and-pepper noise that corrupts 30% of the pixels. The two examples correspond to the last rows of Table 1. The graphs shown in the first and second row of Fig. 4 display, respectively, the quantities $\mathcal{J}_\varepsilon(x^{(k)})$ and $\log_{10}(\|x^{(k)} - x^{(k-1)}\|_2)$ as functions of the iteration number k. These graphs illustrate the theoretical convergence

(a) original

(b) corrupted (SNR = 0.90)

(c) ℓ_1-ℓ_1 (SNR = 25.61)

(d) $\ell_{0.1}$-$\ell_{0.5}$ (SNR = 34.83)

(e) error by ℓ_1-ℓ_1
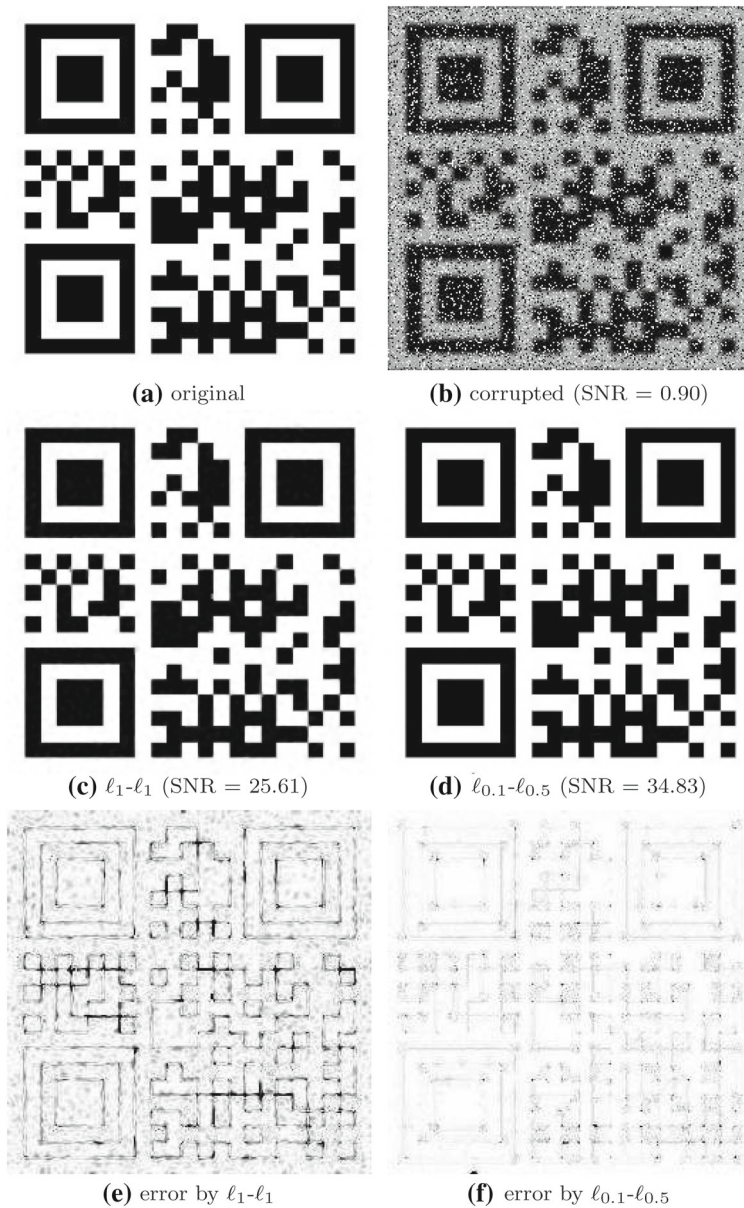
(f) error by $\ell_{0.1}$-$\ell_{0.5}$

Fig. 5 Example 2. Restoration results obtained by the FMM-GKS algorithm applied to a QRcode image that has been contaminated by Gaussian blur with band = 5 and $\sigma = 2.0$, and salt-and-pepper noise corrupting 20% of the pixels

results presented in Sect. 6. The two sequences of functional values $\mathcal{J}_\varepsilon(x^{(k)})$, $k = 0, 1, \ldots$, shown in the first row of Fig. 4 are monotonically nonincreasing and seem to converge, thus confirming the theoretical results of Theorem 2. Moreover, the two graphs reported in the second row of Fig. 4 indicate that the iterates $x^{(k)}$, $k = 0, 1, \ldots$, converge, thus illustrating Theorem 5.

Example 2 We focus on the FMM-GKS method, which is the most efficient method according to Example 1. We are interested in investigating the usefulness of the method when applied to nonconvex models defined by p- and q-values smaller than unity. Table 2 reports SNR-values obtained by the FMM-GKS method when applied to the ℓ_p–ℓ_q models for different choices of p and q. The improvement in the quality of the restoration in terms of SNR-values that can be achieved by letting p be smaller than one (than when p is one) already has been pointed out in Example 1 and can also be observed in this example. The last two rows of Table 2 illustrate that when the image has sparse gradients, nonconvex regularization improves the quality of the restoration.

Figure 5 shows restorations obtained by the FMM-GKS algorithm when applied to the test image QRcode. The degraded image in Fig. 5b was obtained from the original blur- and noise-free image in Fig. 5a, by first applying Gaussian blur with parameters band $= 5$ and $\sigma = 2.0$, and then corrupting 20% of the pixels by salt-and-pepper noise. Figure 5c, d depict restorations obtained by using the FMM-GKS method to solve the ℓ_1–ℓ_1 and $\ell_{0.1} - \ell_{0.5}$ models. Figure 5e, f display the error images, which have been obtained by subtracting the original image in Fig. 5a from the computed restorations. Ideally, the error image vanishes. The nonconvex model $\ell_{0.1} - \ell_{0.5}$ yields a significantly more accurate restoration than the convex ℓ_1–ℓ_1 model both in terms of visual quality and SNR-values. The SNR-values, as well as the regularization parameter μ used for the restorations are displayed in Table 2.

8 Conclusions

We presented a novel approach for the solution of ℓ_p–ℓ_q models applied to image restoration. Combining the majorize-minimize optimization framework with generalized Krylov subspace methods, we derived MM-GKS methods that allow fast solution of the ℓ_p–ℓ_q problem, especially in the nonsmooth and nonconvex case. Computed examples demonstrate that the MM-GKS approach, in both its adaptive and fixed versions, yields high-quality restorations with less computational effort than the state-of-the-art IRN algorithm.

References

1. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**, 183–202 (2009)

2. Birkholz, H.: A unifying approach to isotropic and anisotropic total variation denoising models. J. Comput. Appl. Math. **235**, 2502–2514 (2011)
3. Candes, E.J., Romberg, J., Tao, T.: Robust uncertainty principle: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**, 489–509 (2006)
4. Candes, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted ℓ_1 minimization. J. Fourier Anal. Appl. **14**, 877–905 (2008)
5. Chan, R.H., Liang, H.X.: Half-quadratic algorithm for ℓ_p-ℓ_q problems with applications to TV-ℓ_1 image restoration and compressive sensing. In Efficient Algorithms for Global Optimization Methods in Computer Vision, Lecture Notes in Computer Science, vol. 8293, pp. 78–103. Springer, Berlin (2014)
6. Daniel, J.W., Gragg, W.B., Kaufman, L., Stewart, G.W.: Reorthogonalization and stable algorithms for updating the Gram–Schmidt QR factorization. Math. Comput. **30**, 772–795 (1976)
7. Do, T.-M.-T., Artières, T.: Regularized bundle methods for convex and non-convex risks. J. Mach. Learn. Res. **13**, 3539–3583 (2012)
8. Donoho, D.: Compressed sensing. IEEE Trans. Inf. Theory **52**, 1289–1306 (2006)
9. Hansen, P.C.: Regularization tools version 4.0 for Matlab 7.3. Numer. Algorithms **46**, 189–194 (2007)
10. Horst, R., Thoai, N.V.: DC programming: overview. J. Optim. Theory Appl. **103**, 1–43 (1999)
11. Hunter, D., Lange, K.: A tutorial on MM algorithms. Am. Stat. **58**, 30–37 (2004)
12. Lampe, J., Reichel, L., Voss, H.: Large-scale Tikhonov regularization via reduction by orthogonal projection. Linear Algebra Appl. **436**, 2845–2865 (2012)
13. Lanza, A., Morigi, S., Reichel, L., Sgallari, F.: A generalized Krylov subspace method for ℓ_p-ℓ_q minimization. SIAM J. Sci. Comput. **37**, S30–S50 (2015)
14. Lanza, A., Morigi, S., Sgallari, F.: Constrained TV$_p$-ℓ_2 model for image restoration. J. Sci. Comput. **68**, 64–91 (2016)
15. Lanza, A., Morigi, S., Seleznik, I., Sgallari, F.: Nonconvex nonsmooth optimization via convex–nonconvex majorization–minimization. Numer. Math. (2017). doi:10.1007/s00211-016-0842-x
16. Laporte, L., Flamary, R., Canu, S., Déjean, S., Mothe, J.: Nonconvex regularizations for features selection in ranking with sparse SVM. IEEE Trans. Neural Netw. Learn. Syst. **25**, 1118–1130 (2014)
17. Liu, Z., Wei, Z., Sun, W.: An iteratively approximated gradient projection algorithm for sparse signal reconstruction. Appl. Math. Comput. **228**, 454–462 (2014)
18. Mairal, J.: Incremental majorization-minimization optimization with application to large-scale machine learning. SIAM J. Optim. **25**, 829–855 (2015)
19. Nesterov, Y.: Gradient methods for minimizing composite objective functions. Math. Program. **140**, 125–161 (2012)
20. Nikolova, M., Chan, R.H.: The equivalence of the half-quadratic minimization and the gradient linearization iteration. IEEE Trans. Image Process. **16**, 5–18 (2007)
21. Ramlau, R., Zarzer, C.A.: On the minimization of a Tikhonov functional with a non-convex sparsity constraint. Electron. Trans. Numer. Anal. **39**, 476–507 (2012)
22. Rodríguez, P., Wohlberg, B.: Numerical methods for inverse problems and adaptive decomposition (NUMIPAD), software library. http://numipad.sourceforge.net/. Accessed 19 Feb 2015
23. Rodríguez, P., Wohlberg, B.: Efficient minimization method for a generalized total variation functional. IEEE Trans. Image Process. **18**, 322–332 (2009)
24. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
25. Voss, H.: An Arnoldi method for nonlinear eigenvalue problems. BIT Numer. Math. **44**, 387–401 (2004)
26. Wolke, R., Schwetlick, H.: Iteratively reweighted least squares: algorithms, convergence analysis, and numerical comparisons. SIAM J. Sci. Stat. Comput. **9**, 907–921 (1988)
27. Yuille, A.L., Rangarajan, A.: The convex–concave procedure. Neural Comput. **15**, 915–936 (2003)
28. Zhao, Y., Li, D.: Reweighted ℓ_1-minimization for sparse solutions to undetermined linear systems. SIAM J. Optim. **22**, 1065–1088 (2012)