A GENERALIZED KRYLOV SUBSPACE METHOD
FOR \(\ell_p\)-\(\ell_q\) MINIMIZATION

A. LANZA\(^*\), S. MORIGI\(^1\), L. REICHEL\(^1\), AND F. SGALLARI\(^5\)

Abstract. This paper presents a new efficient approach for the solution of the \(\ell_p\)-\(\ell_q\) minimization problem based on the application of successive orthogonal projections onto generalized Krylov subspaces of increasing dimension. The subspaces are generated according to the iteratively reweighted least-squares strategy for the approximation of \(\ell_p/\ell_q\)-norms by weighted \(\ell_2\)-norms. Computed image restoration examples illustrate that it suffices to carry out only a few iterations to achieve high-quality restorations. The combination of a low iteration count and a modest storage requirement makes the proposed method attractive.

Key words. \(\ell_p\)-\(\ell_q\) minimization; generalized Krylov subspaces; iteratively reweighted least-squares; half-quadratic; image restoration.

AMS subject classifications. 68U10, 65K10, 65F22, 65F10

1. Introduction. This paper is concerned with the computation of approximate solutions of \(\ell_p\)-\(\ell_q\) minimization problems of the form

\[
\min_{x \in \mathbb{R}^n} J(x), \quad J(x) = \frac{1}{p} \|Ax - b\|_p^p + \frac{\mu}{q} \|\Phi(x)\|_q^q,
\]

where \(0 < p, q \leq 2, \mu > 0, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, x \in \mathbb{R}^n\), and \(\Phi : \mathbb{R}^n \to \mathbb{R}^s\) is a linear or nonlinear operator of interest in image restoration problems with particular properties. Examples of such operators will be given below. The general minimization problem (1.1) encompasses a wide variety of problems that have been extensively studied in many different research areas, including numerical linear algebra [1, 33], image restoration [7, 28], pattern recognition [9, 21], and compressive sensing [6, 7, 10, 22].

Different choices of the parameters \(p\) and \(q\), the matrix \(A\), and the operator \(\Phi\) yield a variety of popular models that have been successfully used in many application fields. For instance, the model (1.1) with \(p = 2, m < n, 0 < q \leq 1\), and \(\Phi\) the identity matrix can be used to compute sparse solutions of undetermined linear systems; when \(p = 2, 0 < q \leq 1\), and \(A\) is a sampling operator, the model (1.1) can be applied for compressive sensing.

In this paper we focus on the efficient solution of the \(\ell_p\)-\(\ell_q\) problem (1.1) with application to image restoration. However, the proposed method easily can be used for the solution of all the other above mentioned applications that use the model (1.1). In the context of image restoration, the vector \(x\) represents the unknown uncorrupted image that is to be estimated, 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 have

\(^*\) Department of Mathematics, University of Bologna, Bologna, Italy. Email: alessandro.lanza2@unibo.it.
\(^1\) Department of Mathematics, University of Bologna, Bologna, Italy. Email: serena.morigi@unibo.it.
\(^5\) Department of Mathematical Sciences, Kent State University, Kent, USA. Email: reichel@math.kent.edu.
\(^5\) Department of Mathematics, University of Bologna, Bologna, Italy. Email: fiorella.sgallari@unibo.it.
the same size (that is, \( m = n \)); however, this is not a requirement of the proposed method.

The first term in the definition of \( J(x) \) is commonly referred to as the fidelity term and the second term as the regularization term. The operator \( \Phi \) is chosen to yield a computed solution with some known desired features. The scalar \( \mu > 0 \) is a regularization parameter. Its purpose is to balance the influence of the fidelity and regularization terms on the computed solution in a suitable manner.

In image restoration applications, the matrix \( A \) is the identity operator or a blurring operator. In the former case, the purpose of solving (1.1) is to denoise an available noise-contaminated image \( b \). Thus, the computed solution \( x \) is a denoised version of \( b \). When \( A \) is a blurring operator, the solution \( x \) of (1.1) is a deblurred and denoised version of an available blur- and noise-contaminated image \( b \). 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 (1.1) typically yields a meaningless computed solution of very large norm. The purpose of the regularization term is to be able to determine a useful solution of (1.1) of moderate norm.

When \( p = q = 2 \) and \( \Phi \) is a linear operator that is represented by a regularization matrix \( L \in \mathbb{R}^{s \times n} \), the minimization problem (1.1) turns into a classical Tikhonov-regularized linear least-squares problem of the form

\[
\min_{x \in \mathbb{R}^n} \left\{ \|Ax - b\|_2^2 + \mu \|Lx\|_2^2 \right\}.
\]

Assuming that

\[
\text{rank} \begin{bmatrix} A \\ L \end{bmatrix} = n,
\]

the problem (1.2) has the unique solution

\[
x^* = (A^T A + \mu L^T L)^{-1} A^T b
\]

for any \( \mu > 0 \). Here and below the superscript \( ^T \) denotes transposition. When \( L \) is the identity matrix, denoted by \( I \), the Tikhonov regularization problem (1.2) is said to be in standard form and the solution can be efficiently computed by partial Golub–Kahan bidiagonalization of \( A \); see, e.g., [3, 4, 5]. Golub–Kahan bidiagonalization also can be applied when \( L \neq I \), provided that the regularization problem can be transformed to standard form without too much effort by applying the A-weighted pseudoinverse of \( L \); see [11] for details. For the situation when the A-weighted pseudoinverse of \( L \) is complicated and unattractive to use, other methods have been proposed in the literature. A scheme that projects \( L \) into a Krylov subspace

\[
K_l(A^T A, A^T b) = \text{span}\{A^T b, (A^T A)A^T b, \ldots, (A^T A)^{l-1} A^T b\}
\]

for some \( l \geq 1 \), and determines an approximate solution of (1.2) in this subspace is described in [14]. Methods based on reducing both \( A \) and \( L \) by generalized Golub–Kahan-type or Arnoldi-type methods are discussed in [15, 25], and a scheme that computes a partial generalized singular value decomposition of the matrix pair \{\( A, L \)\} is advocated in [16]. An efficient iterative algorithm, based on the nonlinear Arnoldi framework developed by Voss [30], for the solution of large-scale Tikhonov regularization problems (1.2) with automatic selection of the regularization parameter \( \mu \) is
proposed in [18]. The method relies on successive orthogonal projections onto generalized Krylov subspaces of increasing and low dimension. We refer to this method as the generalized Krylov subspace (GKS) method and will present an extension that can be applied to the solution of (1.1).

The popular total variation (TV) regularization method is obtained by setting \( q = 1 \) in (1.1) and letting \( \Phi = \Phi_{\text{TV}} : \mathbb{R}^n \to \mathbb{R}^n \) be the (nonlinear) map from the image \( x \) to the \( \ell_2 \)-norm of its gradient field. Thus,

\[
\|x\|_{\text{TV}} := \|\Phi_{\text{TV}}(x)\|_1 = \sum_{i=1}^{n} \|(\nabla x)_i\|_2 = \sum_{i=1}^{n} \sqrt{(D_1x)_i^2 + (D_2x)_i^2}.
\]

(1.3)

Each element of \( x \) represents a pixel and \( (\nabla x)_i \in \mathbb{R}^2 \) denotes the discrete gradient of \( x \) at pixel \( i \). The linear operators \( D_1, D_2 \in \mathbb{R}^{n \times n} \) are finite difference approximations of first-order horizontal and vertical partial derivatives, respectively. TV-norm regularization has been applied extensively and successively to image restoration [19, 20, 28, 29] because of its ability to preserve edges. Substituting (1.3) into (1.1) yields the \( \ell_p \)-TV restoration model

\[
\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{p} \|Ax - b\|_p^p + \mu \|x\|_{\text{TV}} \right\},
\]

(1.4)

where the choice of \( 0 < p \leq 2 \) depends on the type of noise in the available contaminated image \( b \). For additive Gaussian noise one generally chooses \( p = 2 \) and for impulse noise the standard choice is \( p = 1 \). These \( p \)-values give the popular \( \ell_2 \)-TV and \( \ell_1 \)-TV models, respectively. The \( \ell_1 \)-TV model also has been applied successfully in the contexts of image cartoon-texture decomposition [31] and multiscale image decomposition [32]. Notice that, except in some special cases, such as for the above mentioned TV model, the method proposed in this paper requires that the nonlinear operator \( \Phi \) be linearized.

Many algorithms have been proposed for the efficient solution of problems of the form (1.4) for certain \( p \) values. An ad hoc proposal for image restoration, which uses the coherence map of the image, has been introduced in [2]. However, to the best of our knowledge, the only algorithm that can be applied with no or little modification to the solution of the general \( \ell_p \)-\( \ell_q \) minimization problem (1.1) with \( 0 < p, q \leq 2 \) is the iteratively reweighted norm (IRN) approach introduced in [27]. This scheme is shown to be equivalent to the half-quadratic (HQ) method [7] and to the gradient linearization iterative procedure [23]. As will be illustrated in Section 2, the IRN method consists of iteratively solving a sequence of penalized weighted least-squares problems that differ from each other only by the choice of the diagonal weighting matrices.

This paper presents a novel efficient numerical method for the solution of the general \( \ell_p \)-\( \ell_q \) minimization problem (1.1). Our method generalizes the GKS approach described in [18] for the solution of Tikhonov-regularized least-squares problems (1.2). The proposed approach significantly improves the computational efficiency of the IRN method. We will refer to our approach as the GKS\( p,q \) method.

The organization of this paper is as follows. Sections 2 and 3 introduce the main ingredients of the GKS\( p,q \) method by reviewing the IRN and GKS approaches. Our new GKS\( p,q \) method is described in Section 4 and its convergence is discussed in Section 5. Numerical examples are presented in Section 6 and concluding remarks can be found in Section 7.
2. The IRN method. In this section we express the $\ell_p$-$\ell_q$ minimization problem (1.1) as an equivalent sequence of weighted $\ell_2$-$\ell_2$ minimization problems following the HQ technique. The latter has been applied successfully to image restoration; see, e.g., [7]. Our formulation of the problem (1.1) gives a solution method that is equivalent to the one proposed in [27].

Our discussion of the problem (1.1) relies on the following expression for the $p$th power of $t \in \mathbb{R}\{0\}$, $0 < p < 2$, shown in [7, Lemma 1],

$$|t|^p = \min_{w > 0} \left\{ w t^2 + \frac{1}{\beta_p w^{\alpha_p}} \right\},$$

(2.1)

where $\alpha_p$ and $\beta_p$ are positive scalars defined by

$$\alpha_p = \frac{p}{2 - p} \quad \text{and} \quad \beta_p = \frac{2^\frac{2-p}{2-p}}{(2-p) \cdot p^\frac{p}{2-p}}.$$  

(2.2)

The minimum is achieved at

$$w^* = \frac{p}{2} |t|^{p-2}.$$  

(2.3)

Note that the function inside the curly bracket in (2.1) is quadratic in the parameter $t$, but not in the optimization variable $w$; hence the name *half-quadratic*. In order to be able to apply (2.1)-(2.3) for $t = 0$, we need to use the smoothed version $|t|_\epsilon := \sqrt{t^2 + \epsilon}$ for any $t \in \mathbb{R}$ and some $\epsilon > 0$. For simplicity of notation, we will suppress the subscript $\epsilon$.

The functional $J$ in (1.1) can be written component-wise as

$$J(x) = \frac{1}{p} \sum_{i=1}^m |(Ax)_i - b_i|^p + \frac{\mu}{q} \sum_{j=1}^s |(\Phi(x))_j|^q,$$

(2.4)

where $(Ax)_i$, $b_i$, and $(\Phi(x))_j$ denote the $i$th and $j$th entries of the vectors $Ax, b \in \mathbb{R}^m$ and $\Phi(x) \in \mathbb{R}^s$, respectively. The HQ approach presented in [7] exploits the relations (2.1)-(2.3) to minimize (2.4). Using (2.1) and (2.4), the problem (1.1) can be expressed as the following constrained minimization problem

$$\min_{x,v>0,z>0} \mathcal{L}(x,v,z),$$

(2.5)

where

$$\mathcal{L}(x,v,z) := \frac{1}{p} \sum_{i=1}^m \left( v_i |(Ax)_i - b_i|^2 + \frac{1}{\beta_p v_i^{\alpha_p}} \right) + \frac{\mu}{q} \sum_{j=1}^s \left( z_j |(\Phi(x))_j|^2 + \frac{1}{\beta_q z_j^{\alpha_q}} \right),$$

and the constraints $v > 0$ and $z > 0$ are considered component-wise. We have

$$\min_{x \in \mathbb{R}^n} J(x) = \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{p} \sum_{i=1}^m \min_{v_i > 0} \left( v_i |(Ax)_i - b_i|^2 + \frac{1}{\beta_p v_i^{\alpha_p}} \right) + \frac{\mu}{q} \sum_{j=1}^s \min_{z_j > 0} \left( z_j |(\Phi(x))_j|^2 + \frac{1}{\beta_q z_j^{\alpha_q}} \right) \right\}$$

(2.6)

$$= \min_{x,v>0,z>0} \left\{ \frac{1}{p} \sum_{i=1}^m \left( v_i |(Ax)_i - b_i|^2 + \frac{1}{\beta_p v_i^{\alpha_p}} \right) + \frac{\mu}{q} \sum_{j=1}^s \left( z_j |(\Phi(x))_j|^2 + \frac{1}{\beta_q z_j^{\alpha_q}} \right) \right\}.$$
which is (2.5).

To solve the constrained nonlinear problem (2.6), we apply an alternating minimization iterative procedure, namely, for \( k = 0, 1, \ldots \), we solve successively

\[
    v^{(k+1)} = \arg \min_{v > 0} \mathcal{L}(x^{(k)}, v, z^{(k)}), \tag{2.7}
\]
\[
    z^{(k+1)} = \arg \min_{z > 0} \mathcal{L}(x^{(k)}, v^{(k+1)}, z), \tag{2.8}
\]
\[
    x^{(k+1)} = \arg \min_x \mathcal{L}(x, v^{(k+1)}, z^{(k+1)}). \tag{2.9}
\]

The minimization problems (2.7) and (2.8) are separable into one-dimensional subproblems for each component. Using (2.1) and (2.3), we see that these subproblems have the explicit solutions

\[
v^{(k+1)}_i = \frac{p}{2} \left( Ax^{(k)} \right)_i - b_i \right|^{p-2}, \quad i = 1, 2, \ldots, m, \tag{2.10}
\]
\[
z^{(k+1)}_j = \frac{q}{2} \left( \Phi(x^{(k)}) \right)_j \right|^{q-2}, \quad j = 1, 2, \ldots, s. \tag{2.11}
\]

We solve the minimization problem (2.9) by first defining the diagonal weighting matrices \( W_F^{(k+1)} \in \mathbb{R}^{m \times m} \) and \( W_R^{(k+1)} \in \mathbb{R}^{s \times s} \) according to

\[
    W_F^{(k+1)} = \text{diag}(w_F^{(k+1)}), \quad w_F^{(k+1)} = \frac{2}{p} v^{(k+1)} = \left| Ax^{(k)} - b \right|^{p-2}, \tag{2.12}
\]
\[
    W_R^{(k+1)} = \text{diag}(w_R^{(k+1)}), \quad w_R^{(k+1)} = \frac{2}{q} z^{(k+1)} = \left| \Phi(x^{(k)}) \right|^{q-2}. \tag{2.13}
\]

Then we rewrite (2.6) in a more compact form, in which terms that do not depend on \( x \) are omitted,

\[
    x^{(k+1)} = \min_{x \in \mathbb{R}^n} \left\{ \left\| \left( W_F^{(k+1)} \right)^{1/2} (Ax - b) \right\|_2^2 + \mu \left\| \left( W_R^{(k+1)} \right)^{1/2} \Phi(x) \right\|_2^2 \right\}. \tag{2.14}
\]

This minimization problem is a weighted regularized least-squares problem. The problem is linear when the operator \( \Phi \) is linear, and nonlinear otherwise. Note that in all the cases of interest described in Section 1, including nonlinear TV-norm regularization, the operator \( \Phi \) is represented in (2.14) by a fixed matrix; see, e.g., [27]. We therefore from now on assume \( \Phi \) to be linear, and represent \( \Phi \) by the matrix \( L \in \mathbb{R}^{s \times n} \).

Introduce the \( n \times n \) matrix

\[
    T(W_F, W_R) := A^T W_F A + \mu L^T W_R L. \tag{2.15}
\]

The normal equations associated with the minimization problem (2.14) read as

\[
    T(W^{(k+1)}_F, W^{(k+1)}_R) x = A^T W^{(k+1)}_F b. \tag{2.16}
\]

At each (outer) iteration \( k = 0, 1, \ldots \), the standard IRN method [27] solves the linear system of equations (2.16) with a symmetric positive definite matrix for (an approximation of) \( x^{(k+1)} \) with the aid of the conjugate gradient (CG) algorithm, using as initial iterate the approximate solution \( x^{(k)} \) obtained from the previous iteration, and terminating the CG (inner) iterations as soon as a specified stopping criterion is
Algorithm 1 The IRN method [27] for the solution of the $\ell_p$-$\ell_q$ problem (1.1)

**Inputs:** $A \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{s \times n}$, $b \in \mathbb{R}^m$, $0 < p, q \leq 2$, $\mu > 0$

**Output:** Approximate solution $x^*$ of (1.1)

1. Initialize: $x^{(0)} = b$ or $x^{(0)} = (A^T A + \mu L^T L)^{-1} A^T b$

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

3. Compute weighting matrices $W_F^{(k+1)}$ and $W_R^{(k+1)}$ by (2.12) and (2.13)

4. Compute approximate solution $x^{(k+1)}$ by solving the linear system (2.16)

5. end for

6. $x^* = x^{(k+1)}$

satisfied; see Section 6 for more details. Algorithm 1 illustrates the main steps of the standard IRN method. Details of the algorithm are commented on below.

The $k$th (outer) iteration of Algorithm 1 consists of two main steps. The first step (line 3) computes the diagonal weighting matrices by (2.12) and (2.13). The cost of this computation is dominated by the evaluation of one matrix-vector product (MVP) with $A$ in (2.12) and one with $L$ in (2.13). The second step (line 4) solves the linear system of equations (2.16) by the CG algorithm. First, the right-hand side vector of (2.16) is formed with a dominating cost of one MVP with the matrix $A^T$. Then the cost of each CG (inner) iteration is dominated by one MVP evaluation with the matrix $T(W_F, W_R)$ defined in (2.15). This matrix is not explicitly formed. Instead, each MVP with $T(W_F, W_R)$ is computed by evaluating one MVP with each one of the matrices $L, L^T, A, A^T$; cf. (2.15). The cost of the two MVP evaluations with the diagonal weighting matrices $W_F$ and $W_R$ is negligible.

Summarizing, the computational cost of $K$ iterations with Algorithm 1 is dominated by $3K + 4 \text{CGits}_K$ MVP evaluations, with $\text{CGits}_K$ denoting the total number of CG iterations carried out.

3. The GKS method. This section reviews the GKS method proposed in [18]. The method is designed for the solution of large-scale Tikhonov regularization problems (1.2) with a general regularization matrix $L \in \mathbb{R}^{s \times n}$. We notice that (1.2) represents a very specific instance of the $\ell_p$-$\ell_q$ problem in (1.1) obtained for $p = q = 2$.

The method described in [18] determines the regularization parameter $\mu > 0$ by the discrepancy principle [12]; however, many other approaches to compute this parameter can be used; see below. The GKS method computes a sequence of orthogonal projections of generalized Krylov subspaces onto subspaces of low dimension.

The iterative method starts with a user-chosen subspace $V_0 \subset \mathbb{R}^n$ of low dimension $l \ll n$. In [18], one lets $V_0 = K_l(A^T A, A^T b)$ for $l \leq 5$. The columns of the matrix $V_0 \in \mathbb{R}^{n \times l}$ form an orthonormal basis for the space $V_0$.

Assume for the moment that an estimate $\delta > 0$ of the $\ell_2$-norm of the noise in $b$ is available. The discrepancy principle then prescribes that the regularization parameter $\mu$ in (1.2) be chosen so that the solution $x^*$ satisfies $\|Ax^* - b\|_2 = \eta \delta$ for some user-chosen constant $\eta \geq 1$ that is independent of $\delta$. It is applied in [18] to
determine $\mu = \mu_k$ for each iteration $k = 0, 1, \ldots$ and to compute the associated solution $x^{(k)}$ of the Tikhonov minimization problem (1.2) restricted to the subspace $V_k$. This subspace is spanned by the orthonormal columns of the matrix $V_k$. Thus, $x^{(k)}$ is determined as follows:

$$y^{(k)} = \arg\min_{y \in \mathbb{R}^l} \left\{ \| AV_k y - b \|_2^2 + \mu_k \| LV_k y \|_2^2 \right\}, \quad x^{(k)} = V_k y^{(k)}. \quad (3.1)$$

Define the $n \times n$ matrix

$$T(\mu) := A^T A + \mu L^T L.$$ 

The solution $y^{(k)}$ of the reduced minimization problem (3.1) can be determined by solving the associated normal equations,

$$V_k^T T(\mu_k) V_k y = V_k^T A^T b, \quad (3.2)$$

and the corresponding approximate solution $x^{(k)}$ of the original unreduced problem (1.2) is given by $x^{(k)} = V_k y^{(k)}$. Note that QR factorizations of the matrices $AV_k$ and $LV_k$ with small $(l + k) \times (l + k)$ upper triangular matrices are available and can be used in the solution process; see [18] for details.

Following the approach of the nonlinear Arnoldi method proposed by Voss [30], the subspace $V_k$ is expanded to $V_{k+1}$ by adding a new basis vector $v_{\text{new}}$ to $V_k$. This basis vector is determined by normalizing the residual $r^{(k)}$ of the unreduced problem (1.2). Thus,

$$v_{\text{new}} = \frac{r^{(k)}}{\| r^{(k)} \|_2}, \quad r^{(k)} = T(\mu_k) x^{(k)} - A^T b. \quad (3.3)$$

Note that the residual vector $r^{(k)}$ is parallel to the gradient of the functional minimized in the original unreduced problem (1.2) evaluated at $x^{(k)}$. In the absence of round-off errors, $r^{(k)}$ is orthogonal to the search space $V_k$. To enforce orthogonality in the presence of round-off errors, the residual $r^{(k)}$ can be reorthogonalized against $V_k$ before normalization.

We remark that the space $V_k$ is a Krylov subspace only in very special situations. Since the $\mu_k$ are updated during the iterations, so is the matrix $T(\mu_k)$. Therefore the space $V_k$, in general, is not a Krylov subspace when $L \neq I$. In particular, the new basis vector $v_{\text{new}}$ cannot be computed with a short recurrence relation. For this reason, we refer to the search space $V_k$ as a generalized Krylov subspace. Algorithm 2 describes the main steps of the GKS method.

We note that Algorithm 2 also can be applied when no estimate of the norm of the noise in $b$ is available. In this situation, the regularization parameter $\mu_k$ has to be determined by a so-called heuristic parameter choice rule. Recent discussions on heuristic parameter choice rules and computed examples that illustrate their performance can be found, e.g., in [17, 24] and references therein.

4. The GKS method. It is our aim to improve the efficiency of the standard IRN approach [27] described by Algorithm 1 for the solution of the $\ell_p$-$\ell_q$ minimization problem (1.1). By using the HQ formulation, the $\ell_p$-$\ell_q$ minimization problem can be expressed in the form (2.14) and then be solved by applying ideas that form the basis for the GKS approach. Thus, our solution subspaces are chosen to be nested generalized Krylov subspaces of increasing dimension. They are constructed using the
Algorithm 2 The GKS method [18] for the solution of the $\ell_2$-$\ell_2$ problem (1.2)

**Inputs:** $A \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{s \times n}$, $b \in \mathbb{R}^m$, $\delta > 0$

**Output:** Approximate solution $x^*$ of (1.2)

1. Initialize: $V_0 \in \mathbb{R}^{n \times l}$ such that $V_0^T V_0 = I_l$

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

3. Compute regularization parameter $\mu_k$ by the discrepancy principle (see [18])

4. Compute solution $y^{(k)}$ of reduced problem in (3.1) by solving (3.2)

5. Compute residual $r^{(k)}$ of unreduced problem (1.2) by (3.3)

6. Compute new basis vector $v_{\text{new}}$ by normalizing residual $r^{(k)}$

7. Enlarge subspace: $V_{k+1} = [V_k, v_{\text{new}}]$

8. end for

9. Determine approximate solution $x^* = V_k y^{(k)}$

solutions of (2.14) for increasing values of $k$. This defines the GKS$pq$ method. The initial subspace $V_0$ is generated in the same manner as for the GKS method and at each iteration $k = 0, 1, \ldots$, we first determine the new weighting matrices $W_F^{(k+1)}$ and $W_R^{(k+1)}$ in the same way as in the IRN algorithm. Having at our disposal the matrix $V_k$, whose orthonormal columns form a basis for the generalized Krylov subspace $V_k$, the solution $y^{(k+1)}$ of the current weighted least-squares problem (2.14) restricted to the subspace $V_k$ satisfies

$$
\min_{y \in \mathbb{R}^{k+1}} \left\{ \left\| \left( W_F^{(k+1)} \right)^{1/2} (AV_k y - b) \right\|_2^2 + \mu \left\| \left( W_R^{(k+1)} \right)^{1/2} LV_k y \right\|_2^2 \right\}. \tag{4.1}
$$

The normal equations associated with this reduced least-squares problem are given by

$$
V_k^T T(W_F^{(k+1)}, W_R^{(k+1)}) V_k y = V_k^T A^T W_F^{(k+1)} b, \tag{4.2}
$$

with the matrix $T(W_F^{(k+1)}, W_R^{(k+1)})$ defined by (2.15). We note that the matrix $V_k^T T(W_F^{(k+1)}, W_R^{(k+1)}) V_k$ is of small order $k + l \ll n$.

The solution $y^{(k+1)}$ of (4.1) is computed by solving the least-squares problem

$$
\min_{y \in \mathbb{R}^{k+1}} \left\{ \left\| \begin{bmatrix} W_F^{1/2} AV \\ \mu^{1/2} W_R^{1/2} LV \\ \end{bmatrix} y - \begin{bmatrix} W_F^{1/2} b \\ 0 \\ \end{bmatrix} \right\|_2^2, \tag{4.3}
$$

where we for notational simplicity suppress the subscript $k$ of the subspace matrix $V$ and the superscripts $k + 1$ of the weighting matrices. Let $V \in \mathbb{R}^{n \times d}$, $d = k + l \ll n$, ...
and introduce the QR factorizations

\[ W^{1/2}_F AV = Q_A R_A \quad \text{with} \quad Q_A \in \mathbb{R}^{m \times d}, \quad R_A \in \mathbb{R}^{d \times d}, \]  
\[ W^{1/2}_R LV = Q_L R_L \quad \text{with} \quad Q_L \in \mathbb{R}^{s \times d}, \quad R_L \in \mathbb{R}^{d \times d}. \]  

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 (4.3) yields the minimization problem

\[
\min_{y \in \mathbb{R}^{k+1}} \left\| \begin{bmatrix} R_A & \mu^{1/2} R_L \end{bmatrix} y - \begin{bmatrix} Q_A W^{1/2}_F b \\ 0 \end{bmatrix} \right\|_2^2
\]

with the associated normal equations

\[
(R_A^T R_A + \mu R_L^T R_L) y = R_A^T Q_A^T W_F b.
\]

Recalling the definition (2.15) of the matrix \( T(W_R, W_F) \) and using the fact that \( x^{(k+1)} = V_k y^{(k+1)} \), the residual \( r^{(k+1)} \) of (2.16) can be computed according to

\[
r^{(k+1)} = T(W_F, W_R) x - A^T W_F b = (A^T W_F A + \mu L^T W_R L) x - A^T W_F b = A^T W_F (A V y - b) + \mu L^T W_R (L V y). \]  

The subspace \( V_k \) is then expanded to \( V_{k+1} \) in the same fashion as in the GKS method, that is by adding the new unit vector \( v_{\text{new}} \) obtained by normalizing the residual \( r^{(k+1)} \) in (4.7). The main steps of the GKSpq algorithm are described by Algorithm 3.

We briefly comment on the computational cost of Algorithm 3. First, we note that the computation of an orthonormal basis for the initial subspace \( V_0 := K_l(A^T A, A^T b) \) does not require the matrix \( A^T A \) to be explicitly formed; it suffices to carry out \( l - 1 \) steps of Golub–Kahan bidiagonalization of the matrix \( A \) with initial vector \( b \). Moreover, for efficiency reasons, the matrices \( AV \) and \( LV \) are stored and updated during the iterations. The cost for computing the diagonal weighting matrices in step 3 of Algorithm 3 is negligible. In fact, the vectors \( A V y^{(k+1)} - b \) and \( L V y^{(k+1)} \) have already been computed in (4.7) when evaluating the residual \( r^{(k+1)} \) in the previous iteration and are reused. The computation of the QR factorizations (4.4) and (4.5) required in step 4 of Algorithm 3 demands about \( 2m(k + l)^2 \) and \( 2s(k + l)^2 \) arithmetic floating point operations (flops), respectively. Since \( k + l \ll n \), both flop counts are negligible in comparison with the cost for evaluating MVPs with the large matrices \( A, L \), and their transposes. The computational effort required in step 5 to determine the residual by (4.7) is dominated by two MVP evaluations, one with \( A^T \) and one with \( L^T \). Finally, in step 7 of the algorithm, the matrices \( AV \) and \( LV \) are updated for use in the next iteration with a computational cost of two MVPs, one with \( A \) and one with \( L \). In summary, the overall computational cost for \( K \) iterations with the GKSpq algorithm is dominated by the work required to evaluate \( 4K \) MVPs.

5. **Convergence analysis.** In this section we discuss the convergence of the GKSpq method. Its main steps are described by Algorithm 3. Our discussion relies on convergence results reported in [7] for the IRN/HQ method.

To secure that the GKSpq algorithm does not break down, we require that the solution \( y^{(k+1)} \) of the reduced least-squares problem (4.1) exists and is unique or,
Algorithm 3 The GKSpq method for the solution of the $\ell_p$-$\ell_q$ problem (1.1)

Inputs: $A \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{s \times n}$, $b \in \mathbb{R}^m$, $0 < p, q \leq 2$, $\mu > 0$

Output: Approximate solution $x^*$ of (1.1)

1. Initialize: $x^{(0)} = b$ or $x^{(0)} = (A^T A + \mu L^T L)^{-1} A^T b$
   
   \[ V_0 \in \mathbb{R}^{n \times l} \text{ such that } V_0^T V_0 = I_l \]

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

3. Compute weighting matrices $W_{F(k+1)}^{(k+1)}$ and $W_{R(k+1)}^{(k+1)}$ by (2.12) and (2.13)

4. Compute solution $y^{(k+1)}$ of reduced problem (4.6)

5. Compute residual $r^{(k+1)}$ of unreduced problem (2.14) by (4.7)

6. Compute new basis vector $v_{\text{new}}$ by normalizing residual $r^{(k+1)}$

7. Enlarge subspace and update matrices:
   
   \[ V_{k+1} = [V_k, v_{\text{new}}], \quad AV_{k+1} = [AV_k, Av_{\text{new}}], \quad LV_{k+1} = [LV_k, Lv_{\text{new}}] \]

8. end for

9. Determine approximate solution $x^* = V_k y^{(k+1)}$

equivalently, that the coefficient matrix of the normal equations (4.2) is of full rank.

We recall that the diagonal weighting matrices $W_{F(k+1)}^{(k+1)}$ and $W_{R(k+1)}^{(k+1)}$, defined by (2.12) and (2.13), have strictly positive diagonal entries and that the rectangular subspace matrix $V_k$ has full rank. Therefore, in order for (4.2) to be solvable for any such matrices $W_{F(k+1)}^{(k+1)}$, $W_{R(k+1)}^{(k+1)}$, and $V_k$, we have to assume that

\[ \text{Ker}(A^T A) \cap \text{Ker}(L^T L) = \{0\} , \tag{5.1} \]

where Ker($M$) denotes the null space of the matrix $M$. This requirement typically is satisfied. For instance, in image restoration problems $A$ represents a blurring operator, which is a low-pass filter, whereas the regularization matrix $L$ usually is a difference operator and, hence, is a high-pass filter. Chan and Liang [7] point out that the property (5.1) also guarantees that the IRN method does not break down.

Both the IRN and GKSpq methods rely on the three main computational steps (2.7)-(2.9), but differ in how the solution of the step (2.9) is computed. For the purpose of discussing convergence, it is convenient to consider the computations of (2.7)-(2.9) as one step. Towards this end, we substitute the expressions for $v^{(k+1)}$ and $z^{(k+1)}$, given by (2.10)-(2.11), into (2.9). Then the minimization problem (2.5) can be expressed by substituting the solutions of the problems (2.7) and (2.8). We obtain that at iteration $k$, both the IRN and GKSpq algorithms compute the next iterate, $x^{(k+1)}$, by minimizing the following functional over $x \in \mathbb{R}^n$, with $x^{(k)} \in \mathbb{R}^n$ a constant
parameter vector:

\[ Q(x; x^{(k)}) = \frac{1}{p} \sum_{i=1}^{m} \left( \frac{p}{2} |Ax^{(k)}_i| - b_i - b_i^p - 2 - \frac{q}{2} |(Ax^{(k)}_i - b_i)^p| + \mu \right) \]

\[ + \frac{q}{2} \sum_{j=1}^{s} \left( \frac{q}{2} |Lx^{(k)}_j| - (Lx^{(k)}_j)^q - 2 - \frac{q}{2} |(Lx^{(k)}_j)^q| \right). \]  

This functional has some important properties, which are summarized in the following proposition.

**Proposition 5.1.** Let \( 0 < p, q \leq 2 \) and assume that (5.1) holds. Then, for any \( x^{(k)} \in \mathbb{R}^n \), the functional \( Q(x; x^{(k)}) \) defined in (5.2) is a strictly convex tangent majorant function at \( x^{(k)} \) of the \( \ell_p-\ell_q \) functional \( J(x) \) defined in (2.4) with \( \Phi = L \), that is,

- \( Q(x; x^{(k)}) \geq J(x) \quad \forall x \in \mathbb{R}^n \),
- \( Q(x; x^{(k)}) = J(x) \quad \text{at} \ x = x^{(k)} \),
- \( \nabla_x Q(x; x^{(k)}) = \nabla J(x) \quad \text{at} \ x = x^{(k)} \),
- \( \nabla^2_x Q(x; x^{(k)}) \) is positive definite \( \forall x \in \mathbb{R}^n \).

**Proof.** The stated properties can be shown by using results from [7, Lemma 2]. \( \square \)

Thus, in iteration \( k \), the IRN and GKSpq methods replace the minimization of the original convex or nonconvex \( \ell_p-\ell_q \) functional \( J(x) \) by minimizing the simpler surrogate functional \( Q(x; x^{(k)}) \), which is convex and majorizes \( J(x) \).

Using the definition (5.2) of \( Q(x; x^{(k)}) \), the IRN algorithm can be written as

\[ x^{(k+1)} = \arg \min_{x \in \mathbb{R}^n} Q(x; x^{(k)}), \quad k = 0, 1, \ldots , \]

while the GKSpq algorithm can be expressed as

\[ x^{(k+1)} = \begin{cases} 
\arg \min_{x \in V_k} Q(x; x^{(k)}) & \text{for } k = 0, 1, \ldots , n - l - 1, \\
\arg \min_{x \in \mathbb{R}^n} Q(x; x^{(k)}) & \text{for } k = n - l, n - l + 1, \ldots . \end{cases} \]  

(5.3)

Here \( l \geq 1 \) is the dimension of the user-specified initial subspace \( V_0 \), and \( V_k \) is the generalized Krylov subspace generated at iteration \( k \). To justify (5.3), we recall that the subspace \( V_k \), in which in the \( k \)th iteration the GKSpq algorithm searches for the new approximate solution \( x^{(k+1)} \), is of dimension \( l + k \). Hence, from iteration \( n - l \) and onwards, the subspace \( V_k \) is equivalent to the whole space \( \mathbb{R}^n \) and cannot be enlarged further. Therefore, for iterations \( k \geq n - l \), the GKSpq method is equivalent to the IRN method with initial approximate solution \( x^{(n-l-1)} \). This implies that the convergence behavior of the GKSpq and IRN algorithms is the same as \( k \to \infty \).

In the following two subsections, we analyze the convergence of the GKSpq algorithm and first show that the sequence \( \{ J(x^{(k)}) \} \) of values achieved by the functional (1.1) is monotonically decreasing and convergent for any \( 0 < p, q \leq 2 \). Subsequently, we discuss the convergence of the sequence of computed approximate solutions \( \{ x^{(k)} \} \) for \( 1 \leq p, q \leq 2 \).  

### 5.1. Monotonicity and convergence of \( J(x^{(k)}) \) for \( 0 < p, q \leq 2 \)

We first show the following result.

**Lemma 5.2.** Let \( 0 < p, q \leq 2 \) and assume that (5.1) holds. Let \( \{ x^{(k)} \} \) denote the sequence of approximate solutions generated by the GKSpq method using (5.3).
Then, for any initial approximate solution \( x^{(0)} \in \mathbb{R}^n \) and any \( k \geq 1 \), we have
\[
Q(x^{(k+1)}; x^{(k)}) \leq Q(x^{(k)}; x^{(k)}) .
\]

**Proof.** We show (5.4) by considering the cases \( 1 \leq k < n - l \) and \( k \geq n - l \) separately, starting with the former. Recall that generically at iteration \( k - 1 \), the approximate solution \( x^{(k)} \) computed by the GKS pq algorithm (5.3) lives in the subspace \( V_{k-1} \). Moreover, by construction \( V_{k-1} \subseteq V_k \subseteq \mathbb{R}^n \). Given \( x^{(k)} \), the new approximate solution \( x^{(k+1)} \) is determined by minimizing \( Q(x; x^{(k)}) \) with the constraint that \( x \) belongs to the subspace \( V_k \); cf. (5.3). Since \( x^{(k)} \in V_k \) and \( x^{(k+1)} \) is the unique minimizer of \( Q(x; x^{(k)}) \) for \( x \in V_k \), it follows that (5.4) holds.

We turn to the case when \( k \geq n - l \). Then, given \( x^{(k)} \), the GKS pq algorithm (5.3) computes the new approximate solution \( x^{(k+1)} \) as the unique unconstrained global minimizer of \( Q(x; x^{(k)}) \), and (5.4) follows. \( \square \)

**Theorem 5.3.** Let \( 0 < p, q \leq 2 \), assume that (5.1) holds, and let \( \{x^{(k)}\} \) denote the sequence of approximate solutions generated by GKS pq method (5.3). Then, for any initial approximate solution \( x^{(0)} \in \mathbb{R}^n \), the sequence \( \{J(x^{(k)})\} \) of values is monotonically decreasing and convergent where \( J \) is defined by (2.4).

**Proof.** The sequence \( \{J(x^{(k)})\} \) is bounded from below by zero and is monotonically decreasing, since we can write
\[
J(x^{(k+1)}) \leq Q(x^{(k+1)}; x^{(k)}) \leq Q(x^{(k)}; x^{(k)}) = J(x^{(k)}) .
\]

Here the first inequality and last equality follow from the fact that \( Q(x; x^{(k)}) \) is a tangent majorant of \( J(x) \), i.e., they follow from the first two properties of Lemma 5.1. The second inequality is a consequence of Lemma 5.2. Finally, since the sequence \( \{J(x^{(k)})\} \) is monotonically decreasing and bounded from below, it is convergent. \( \square \)

**5.2. Convergence of \( x^{(k)} \) for \( 1 \leq p, q \leq 2 \).** When \( 0 < p, q < 1 \), the \( \ell_p-\ell_q \) minimization problem (1.1) is not guaranteed to be convex. We therefore analyze the convergence of the sequence \( x^{(k)} \), \( k = 1, 2, \ldots \), generated by the GKS pq method for \( 1 \leq p, q \leq 2 \) only. Since after \( n - l \) iterations the GKS pq method turns into the IRN method, the convergence behavior of the sequence \( \{x^{(k)}\} \) is the same as for sequences generated by the IRN method. Therefore the convergence results shown in [7] for the IRN method also hold for the GKS pq method. We state the following convergence result from [7].

**Theorem 5.4.** Let \( 1 \leq p, q \leq 2 \), assume that (5.1) holds, and let \( \{x^{(k)}\} \) denote the sequence of approximate solutions determined by the GKS pq method (5.3). Then, for any initial approximate solution \( x^{(0)} \in \mathbb{R}^n \), we have
(a) \( \lim_{k \to \infty} \|x^{(k)} - x^{(k-1)}\|_2 = 0 \),
(b) The sequence \( \{x^{(k)}\} \) converges to the unique minimizer of the \( \ell_p-\ell_q \) functional (2.4).

An analysis of the behavior of \( x^{(k)} \) for \( k \leq n - l \) is beyond the scope of this paper and will be considered in future work. The following section provides some illustrations of this behavior.

**6. Numerical examples.** We compare the performances of the GKS pq method described by Algorithm 3 and of the IRN method outlined by Algorithm 1 when applied to the restoration of two gray-scale test images cameraman and brain depicted
in Figures 6.1(b) and 6.2(b). 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 $n = 65536$. 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 [13]. This function has the parameters `band` and `sigma`, which determine the half-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 different types and amounts of noise to obtain the image $b \in \mathbb{R}^n$. This image is assumed to be known. It is our aim to determine an accurate approximation $\bar{x}$ of $\bar{x}$, given $A$ and $b$, by using the $\ell_p$-$\ell_q$ model (1.1).

We compare the GKSpq and IRN methods both in terms of accuracy and efficiency. The former is measured by the relative error $e$ and by the Signal-to-Noise Ratio (SNR) defined by

$$e(x^*, \bar{x}) := \frac{\|x^* - \bar{x}\|_2}{\|\bar{x}\|_2}, \quad \text{SNR}(x^*, \bar{x}) := 10 \log_{10} \left( \frac{\|ar{x} - E(\tilde{x})\|_2^2}{\|x^* - \bar{x}\|_2^2} \right) \text{ (dB)}, \quad (6.1)$$

where $E(\tilde{x})$ denotes the mean gray-level of the uncontaminated image $\tilde{x}$. These quantities provide quantitative measures of the quality of the restored image: a large SNR-value and a small $e$-value indicate that $x^*$ is an accurate approximation of $\bar{x}$.

Computational efficiency is measured in terms of the total number of MVP evaluations with the matrices $A$, $A^T$, $L$, $L^T$ required by the algorithms to satisfy the following stopping criterion. The (outer) iterations of the IRN and GKSpq algorithms are terminated as soon as the relative error of the computed approximate solution $x^{(k)}$ drops below a user-specified threshold $\tau > 0$, i.e., when

$$e(x^{(k)}, \bar{x}) < \tau. \quad (6.2)$$

Results for the IRN algorithm are computed by using the freely available implementation in the NUMIPAD library [26]. This code implements the variable auto-adapted CG tolerance policy that yields the best timings [27]. The IRN implementation requires the user to specify the regularization parameter. We therefore choose the regularization parameter empirically to give the most accurate restorations, i.e., restorations with the highest SNR-values. To be able to compare the restorations determined by the IRN and GKSpq methods, we choose the regularization parameter in the same manner for the latter method. The GKSpq algorithm uses the initial search space $V_0 = K_1(A^TA, A^Tb)$. All computations are carried out in MATLAB with about 15 significant decimal digits.

**Example 1.** We consider the restoration of contaminated test images cameraman and brain that have been degraded by Gaussian blur with different parameters `band` and `sigma`, 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 popular TV-regularized $\ell_p$-TV restoration model (1.4), which corresponds to letting $q = 1$, $\Phi = \Phi_{TV}$, and $0 < p \leq 2$ in (1.1). Specifically, we consider the two models $\ell_1$-TV and $\ell_{0.1}$-TV. The $\ell_1$-TV model represents the standard choice for the restoration of images corrupted by salt-and-pepper noise, see [7], since it forces sparsity of the residual while keeping the functional convex. The $\ell_{0.1}$-TV model is not convex, but holds the potential to better induce
Tables 6.1 and 6.2 report quantitative results of the performances of the IRN and GKSpq methods when applied to the $\ell_1$-TV and $\ell_{0.1}$-TV models for different choices of the parameters $\text{band}$ and $\text{sigma}$ that define the Gaussian blur and for sparsity. We are interested in comparing the two models in terms of restoration quality and computational effort required by the IRN and GKSpq minimization algorithms.

Figures 6.1 and 6.2 show restorations obtained by the GKSpq algorithm when applied to the test images. The degraded images in Figures 6.1(b) and 6.2(b) were obtained from the original blur- and noise-free images in Figures 6.1(a) and 6.2(a), respectively, by first applying Gaussian blur with parameters $\text{band} = 5$ and $\text{sigma} = 1.5$, and then corrupting 30% of the pixels by salt-and-pepper noise. Figures 6.1(c), 6.2(c) and Figures 6.1(d), 6.2(d) depict restorations obtained by using the GKSpq method to solve the $\ell_1$-TV and $\ell_{0.1}$-TV models, respectively. For both test images, the latter model yields a significantly more accurate restoration than the former both in terms of visual quality and SNR-value. The SNR-values are displayed in Tables 6.1 and 6.2.

Tables 6.1 and 6.2 report quantitative results of the performances of the IRN and GKSpq methods when applied to the $\ell_1$-TV and $\ell_{0.1}$-TV models for different choices of the parameters $\text{band}$ and $\text{sigma}$ that define the Gaussian blur and for...
Table 6.2

<table>
<thead>
<tr>
<th>blur</th>
<th>noise</th>
<th>parameters</th>
<th>MVPs (outer its)</th>
<th>SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>band</td>
<td>sigma</td>
<td>%</td>
<td>μ</td>
<td>τ</td>
</tr>
</tbody>
</table>

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.150</td>
<td>0.0594</td>
<td>200 (6)</td>
<td>60 (15)</td>
<td>21.59</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>20</td>
<td>0.250</td>
<td>0.0774</td>
<td>354 (9)</td>
<td>112 (28)</td>
<td>19.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.300</td>
<td>0.0975</td>
<td>398 (9)</td>
<td>168 (42)</td>
<td>17.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.015</td>
<td>0.0901</td>
<td>208 (4)</td>
<td>36 (9)</td>
<td>17.76</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>20</td>
<td>0.035</td>
<td>0.0977</td>
<td>206 (5)</td>
<td>44 (11)</td>
<td>17.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.060</td>
<td>0.1092</td>
<td>212 (6)</td>
<td>52 (13)</td>
<td>16.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.010</td>
<td>0.1104</td>
<td>280 (4)</td>
<td>52 (13)</td>
<td>16.35</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>20</td>
<td>0.020</td>
<td>0.1178</td>
<td>216 (4)</td>
<td>52 (13)</td>
<td>15.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.040</td>
<td>0.1265</td>
<td>206 (5)</td>
<td>56 (14)</td>
<td>15.00</td>
</tr>
</tbody>
</table>

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.0347</td>
<td>392 (6)</td>
<td>128 (32)</td>
<td>26.16</td>
<td>26.38</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>20</td>
<td>0.0440</td>
<td>356 (6)</td>
<td>188 (47)</td>
<td>24.27</td>
<td>24.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.0559</td>
<td>402 (7)</td>
<td>228 (57)</td>
<td>21.98</td>
<td>23.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.0568</td>
<td>558 (5)</td>
<td>108 (27)</td>
<td>21.99</td>
<td>22.40</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>20</td>
<td>0.0702</td>
<td>610 (5)</td>
<td>124 (31)</td>
<td>20.17</td>
<td>20.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.0813</td>
<td>566 (5)</td>
<td>156 (39)</td>
<td>18.74</td>
<td>19.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.0828</td>
<td>686 (5)</td>
<td>168 (42)</td>
<td>18.90</td>
<td>18.97</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>20</td>
<td>0.0908</td>
<td>614 (5)</td>
<td>164 (41)</td>
<td>18.12</td>
<td>18.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.1265</td>
<td>606 (5)</td>
<td>164 (41)</td>
<td>18.12</td>
<td>18.18</td>
</tr>
</tbody>
</table>

Table 6.2

Example 1. Comparison of the IRN and GKSpq algorithms applied to the minimization of the $\ell_1$-TV and $\ell_0.1$-TV functionals for the restoration of brain test images corrupted by different kinds of Gaussian blur and salt-and-pepper noise.
$\ell_{0,1}$-TV model in terms of accuracy of the restoration and the faster convergence of the GKS$\text{pq}$ method.

Figure 6.1. Example 1. Restoration results obtained by the GKS$\text{pq}$ algorithm applied to the cameraman image that has been contaminated by Gaussian blur with $\text{band} = 5$ and $\text{sigma} = 1.5$, and salt-and-pepper noise corrupting 30% of the pixels.

Figure 6.3 shows the computational efficiency of the IRN and GKS$\text{pq}$ methods. We plot efficiency curves for some of the parameter combinations of Table 6.1 for the IRN and GKS$\text{pq}$ methods. Each sub-plot of Figure 6.3 displays for both methods $\log_{10}(e^{(k)})$ as a function of the number of MVP evaluations, where $e^{(k)}$ denotes the
Fig. 6.2. Example 1. Restoration results obtained by the GKSpq algorithm applied to a test MRI brain image that has been contaminated by Gaussian blur with band = 5 and sigma = 1.5, and salt-and-pepper noise corrupting 30% of the pixels. The graphs of Figure 6.3 show the GKSpq method to require fewer MVP evaluations than the IRN method to reduce the error in all contaminated images. In fact, the GKSpq method requires about one third of the MVP evaluations needed by the IRN method to reduce the error in the available contaminated images in all examples with $\ell_1$-TV and $\ell_{0.1}$-TV models.
Fig. 6.3. Example 1. Logarithm of the relative error versus the number of matrix-vector product evaluations performed by the IRN and GKSpq algorithms when applied to the minimization of the $\ell_1$-TV (left-hand side column) and $\ell_{0.1}$-TV (right-hand side column) functionals for the restoration of cameraman test images that have been contaminated by Gaussian blur with band = 5 and sigma = 1.5, and salt-and-pepper noise corrupting 10% (first row), 20% (second row), and 30% (third row) of the pixels.

Finally, in Figure 6.4 we provide empirical evidence of the numerical convergence of the proposed GKSpq algorithm applied to the $\ell_1$-TV and $\ell_{0.1}$-TV models for the restoration of the cameraman test image perturbed by Gaussian blur with band = 5 and sigma = 1.5, and salt-and-pepper noise corrupting 20% of the pixels (see the fifth rows in Table 6.1). In particular, the graphs reported in the first and second rows of Figure 6.4 display, respectively, the quantities $\log_{10}(-\frac{J^{(k)} - J^{(k-1)}}{J^{(k-1)}})$ and $\log_{10}(\frac{\|x^{(k)} - x^{(k-1)}\|_2}{\|x^{(k-1)}\|_2})$ as functions of the iteration count $k$, where $J^{(k)}$
Fig. 6.4. Example 1. Convergence results for the GKSpq method applied to $\ell_1$-TV (left-hand side column) and $\ell_{0.1}$-TV (right-hand side column) functionals for the restoration of the cameraman test image that has been contaminated by Gaussian blur with $\text{band} = 5$ and $\text{sigma} = 1.5$, and salt-and-pepper noise corrupting 20% of the pixels.

(a) $\ell_1$-TV, 20%
(b) $\ell_{0.1}$-TV, 20%

denotes the value $J(x^{(k)})$ achieved after $k$ iterations by the $\ell_p$-$\ell_q$ functional defined in (2.4). These graphs show that numerical convergence of the GKSpq algorithm is fast both in terms of the computed approximate solutions $x^{(k)}$ and in terms of the corresponding functional values $J(x^{(k)})$. Moreover, the plots in the first row of Figure 6.4 show that the sequence of function values $J(x^{(k)})$, $k = 1, 2, \ldots$, is monotonically decreasing, thus confirming experimentally the theoretical result in Theorem 5.3. Similar convergence behaviour can be observed for all the other examples.

Example 2. We consider the restoration of the test image cameraman corrupted by Gaussian blur with $\text{band} = 5$ and $\text{sigma} = 1.5$, and by additive zero-mean white Gaussian noise with different standard deviations $\sigma$. It is well known that for this kind of noise accurate restorations can be achieved with the $\ell_p$-$\ell_q$ restoration model (1.1) for $p = 2$. We refer to this model as the $\ell_2$-$\ell_q$ model and compare the IRN and the GKSpq methods when applied to the minimization of the $\ell_2$-$\ell_q$ functional for several values of $q$. Specifically, we let $q \in \{\frac{1}{2}, \frac{3}{2}, 2\}$. The regularization matrix $L$ is chosen to be the first-order discrete derivative operator for two space dimensions,

$$
L = \begin{bmatrix}
L_1 \otimes I_d \\
I_d \otimes L_1
\end{bmatrix}
$$

with $L_1 = \begin{bmatrix}
-1 & 1 & \cdots & \cdots & -1 \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
-1 & 1
\end{bmatrix} \in \mathbb{R}^{(d-1) \times d},$

where $I_d$ denotes the identity matrix of order $d$, $d$ is the number of pixels in each row and column of the image considered, the matrix $L_1$ is bidiagonal, and $\otimes$ denotes the
### Table 6.3

<table>
<thead>
<tr>
<th>q</th>
<th>σ</th>
<th>µ</th>
<th>τ</th>
<th>IRN</th>
<th>GKSpq</th>
<th>SNR</th>
<th>SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10</td>
<td>0.050</td>
<td>0.1293</td>
<td>12  (1)</td>
<td>12  (3)</td>
<td>11.60</td>
<td>11.60</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.190</td>
<td>0.1191</td>
<td>8   (1)</td>
<td>8   (2)</td>
<td>10.50</td>
<td>10.50</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.410</td>
<td>0.1049</td>
<td>14  (1)</td>
<td>8   (2)</td>
<td>9.79</td>
<td>9.79</td>
</tr>
<tr>
<td>1.5</td>
<td>10</td>
<td>0.012</td>
<td>0.0990</td>
<td>40  (2)</td>
<td>24  (6)</td>
<td>12.11</td>
<td>12.11</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.038</td>
<td>0.1128</td>
<td>64  (2)</td>
<td>24  (6)</td>
<td>10.97</td>
<td>10.97</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.080</td>
<td>0.1225</td>
<td>80  (2)</td>
<td>20  (5)</td>
<td>10.26</td>
<td>10.26</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
<td>0.00025</td>
<td>0.1025</td>
<td>272 (2)</td>
<td>148 (37)</td>
<td>11.87</td>
<td>12.03</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.00100</td>
<td>0.1171</td>
<td>372 (2)</td>
<td>180 (45)</td>
<td>10.65</td>
<td>10.93</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.00250</td>
<td>0.1259</td>
<td>560 (2)</td>
<td>312 (78)</td>
<td>10.02</td>
<td>10.37</td>
</tr>
</tbody>
</table>

**Example 2.** Comparison of the IRN and GKSpq algorithms applied to the minimization of the $\ell_2$-$\ell_q$ functional for the restoration of a cameraman test image contaminated by Gaussian blur with $\text{band} = 5$ and $\text{sigma} = 1.5$, and additive zero-mean white Gaussian noise with standard deviation $\sigma$.

Kronecker product.

Table 6.3 reports results for the IRN and GKSpq methods for several $q$-values (first column), and for additive white Gaussian noise of different standard deviations $\sigma$ (second column). The other columns display the value of the regularization parameter $\mu$ used for each test, the value of the threshold $\tau$ for the relative error used in the stopping criterion (6.2), and the total number of MVP evaluations needed to satisfy the stopping criterion. The last column of Table 6.3 reports the SNR-values for the restorations obtained by the IRN and GKSpq methods. Analogously to Example 1, the dimension of the Krylov subspace generated by the GKSpq algorithm at convergence is the sum of the number of iterations carried out and the dimension of the initial solution subspace. The latter is equal to one.

The GKSpq method is seen to require fewer MVP evaluations than the IRN method. The quality of the restoration is higher for $q < 2$ than for $q = 2$. We remark that $q = 0.5$ corresponds to a non-convex functional [8]. Therefore, the IRN and GKSpq methods are not guaranteed to converge to the same restoration in this situation.

### 7. Conclusions

We presented a new method for the efficient solution of the $\ell_p$-$\ell_q$ minimization problem based on successive orthogonal projections onto generalized Krylov subspaces of increasing dimensions. The subspaces are generated according to the iteratively reweighted least-squares strategy for the approximation of $\ell_p/\ell_q$-norms by weighted $\ell_2$-norms. Computed examples with application to image restoration indicate that subspaces of fairly small dimension suffice for the determination of high-quality restorations. The numerical experiments demonstrate that the proposed method requires a smaller number of matrix-vector product evaluations to yield restorations of required accuracy than the standard implementation of the iteratively reweighted least-squares method.

**Acknowledgements.** We would like to thank the referees for comments that lead to improvements of the presentation. Research by LR was supported in part by
REFERENCES


[26] P. Rodríguez and B. Wohlberg, Efficient minimization method for a generalized total variation

NSF grant DMS-1115385.


