JOTA manuscript No.
(will be inserted by the editor)

Asymptotic Linear Convergence of ADMM for Isotropic TV Norm Compressed Sensing

Emmanuel Gil Torres · Matt Jacobs · Xiangxiong Zhang

Received: date / Accepted: date

Abstract We prove an explicit local linear rate for ADMM solving the isotropic Total Variation (TV) norm compressed sensing problem in multiple dimensions, by analyzing the auxiliary variable in the equivalent Douglas-Rachford splitting on a dual problem. Numerical verification on large 3D problems and real MRI data will be shown. Though the proven rate is not sharp, it is close to the observed ones in numerical tests.

Keywords Isotropic TV Norm · Compressed Sensing · ADMM · Asymptotic Linear Convergence

Mathematics Subject Classification (2000) 49J52 · 65K05 · 65K10 · 90C25

1 Introduction

1.1 The isotropic TV norm compressed sensing

The isotropic total variation (TV) norm compressed sensing (CS) [30] is

$$\min_u \|u\|_{TV} \quad \text{subject to} \quad \hat{u}(k) = b_k, \quad \forall k \in \Omega = \{0, i_2, \cdots, i_m\}, \tag{1a}$$

where u is a d-dimensional image of size $n_1 \times n_2 \times \cdots \times n_d = N$, $\hat{u}$ denotes the d-dimensional discrete Fourier transform of u, Ω is a set of observed frequency indices with $m < N$, and $b \in \mathbb{C}^m$ denotes the observed data. In (1a), $0 \in \Omega$ means that given observed data should include the zeroth frequency of u.

We also regard u as a vector $u \in \mathbb{R}^N \simeq \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$. Let $\mathcal{K} : \mathbb{R}^N \to [\mathbb{R}^N]^d$ denote the discrete gradient operator, which will be defined in Section 2. Then

Emmanuel Gil Torres, Purdue University, egiltorr@purdue.edu
Matt Jacobs, University of California, Santa Barbara, majaco@ucsb.edu
Xiangxiong Zhang, Purdue University, zhan1966@purdue.edu

the isotropic TV norm is defined as $\|u\|_{TV} \coloneqq \|\mathcal{K}u\|_{1,2}$ and $\|\cdot\|_{1,2}$ norm is

$$\|v\|_{1,2} = \sum_{j=1}^{N} \sqrt{\sum_{i=1}^{d} |v_j^i|^2}, \quad v = \begin{bmatrix} v^1 \\ \vdots \\ v^d \end{bmatrix} \in [\mathbb{R}^N]^d, \quad v^i = \begin{bmatrix} v_1^i \\ \vdots \\ v_N^i \end{bmatrix} \in \mathbb{R}^N, \tag{1b}$$

which reduces to the classical ℓ^1 norm for $\mathbb{R}^N$ when $d = 1$.

For processing images, the isotropic TV norm was introduced for denoising in [32], and used in many applications such as deconvolution and zooming, image in-painting and motion estimation [7], as well as compressed sensing [6]. TVCS has been used practically in the areas of nuclear medicine and limited view angle tomosynthesis studies [29, 24, 21, 35]. Though in this paper we only focus on the Fourier measurements, e.g., MRI Imaging [28], the algorithm and our analysis may also be also useful for for applications using the Radon Transform [22] and radio interferometry [36] since the sampling process can be modeled as samples of the Fourier transform [22, 36].

1.2 ADMM for TV norm minimization

For solving (1), we focus on the alternating direction method of multipliers (ADMM) [12], and study its asymptotic linear convergence rate. Though the local linear convergence has been established for ADMM solving TV norm minimization [25] and [1], no explicit rates were given for multiple dimensional case due to the fact that $\|\cdot\|_{1,2}$ is no longer locally polyhedral for $d \geq 2$. There are other popular first order splitting methods, such as the primal dual hybrid gradient (PDHG) method [7]. For problem (1), it has been well known [16, 15, 18, 10, 11] that ADMM is also equivalent to quite a few popular first order methods with special choice of parameters including Douglas-Rachford splitting (DRS) [26] and split Bregman method [19]. In Section 3.1, we will show that ADMM is also equivalent to G-prox PDHG method introduced in [20], which was proven and shown to be efficient for very large images.

ADMM can be applied to any problem in the following form:

$$\min_{u \in X} f(\mathcal{K}u) + g(u), \tag{2a}$$

where X, Y are two finite-dimensional real Hilbert spaces, the map $\mathcal{K} : X \to Y$ is a continuous linear operator, $g : X \to \mathbb{R}$ and $f : Y \to \mathbb{R}$ are proper, convex, and lower semi-continuous functions. For the problem (1), we have

$$X = \mathbb{R}^N, \quad \mathcal{K} : \mathbb{R}^N \to [\mathbb{R}^N]^d, \quad f(v) = \|v\|_{1,2}, \quad g(u) = \iota_{\{u: Au=b\}}(u), \tag{2b}$$

where $\iota_C(u) = \begin{cases} 0, & u \in C \\ +\infty, & u \notin C \end{cases}$ is the indicator function of a set C, and $Au = b$ denotes measurements $\hat{u}_k = b_k, k \in \Omega$ in (1). ADMM for (2) is described as:

Algorithm 1 ADMM with step-size γ.

1: $x_{k+1} = \mathrm{argmin}_x\ g(x) + \langle \mathcal{K}x, z_k \rangle + \frac{\gamma}{2}\|\mathcal{K}x - y_k\|^2$

2: $y_{k+1} = \mathrm{argmin}_y\ f(y) - \langle y, z_k \rangle + \frac{\gamma}{2}\|y - \mathcal{K}x_{k+1}\|^2$

3: $z_{k+1} = z_k - \gamma\big(y_{k+1} - \mathcal{K}x_{k+1}\big)$

1.3 The main result: a local linear rate of ADMM

The Fenchel dual problem to (2) can be written as:

$$\min_{p \in \mathbb{R}^{N\times d}} f^*(p) + h^*(-p), \quad h^*(-p) := g^*(-\mathcal{K}^* p), \tag{3}$$

where f^*, g^* are convex conjugates of f, g, and $\mathcal{K}^*$ is the adjoint operator of $\mathcal{K}$. For analyzing Algorithm 1, we will consider the Fenchel dual problem to (3). As shown in Appendix A, the dual problem to (3) can be given as

$$\min_{v \in \mathbb{R}^{N\times d}} f(v) + h(v), \quad f(v) = \|v\|_{1,2}, \quad h(v) = \iota_{\mathcal{K}\{u: Au=b\}}(v), \tag{4}$$

where $\mathcal{K}\{u : Au = b\} := \{v : v = \mathcal{K}u, Au = b\}$. It is well known that the ADMM on (2) with a step size γ is equivalent to DRS on (3) with a step size γ, which is also equivalent to DRS on (4) with a step size $\frac{1}{\gamma}$ as reviewed in Appendix B. Next we describe DRS solving (4) which will be used to analyze Algorithm 1. Let $\mathbb{I}$ be the identity operator. Define the proximal and reflection operators with a step size $\tau > 0$ respectively as

$$\mathrm{Prox}_f^\tau(x) = \mathrm{argmin}_z\, f(z) + \frac{1}{2\tau}\|z - x\|^2, \quad \mathrm{R}_f^\tau = 2\,\mathrm{Prox}_f^\tau - \mathbb{I}. \tag{5}$$

DRS on problem (4) is defined by a fixed point iteration of the operator $H_\tau = \frac{\mathbb{I} + \mathrm{R}_h^\tau \mathrm{R}_f^\tau}{2}$. In particular, in Algorithm 2, q_k is an auxiliary variable and v_k converges to the minimizer to (4). The equivalence between Algorithm 1 and Algorithm 2 will be reviewed in Section 3.

The function $f(v) = \|v\|_{1,2}$ is sparsity promoting [33], and its proximal operator Prox_f^τ is the well known Shrinkage operator in multiple dimensions. Let S_τ denote the shrinkage operator with step size τ. For any $q = [q^1 \ \cdots \ q^d]^T \in [\mathbb{R}^N]^d$ with $q^i = [q_1^i \ \cdots \ q_N^i]^T \in \mathbb{R}^N$, we introduce the notation $q_j = [q_j^1 \ \cdots \ q_j^d] \in \mathbb{R}^d$ and we will call the subscript the *spatial index*. Then the shrinkage operator $\mathrm{Prox}_f^\tau(q) = S_\tau(q) \in [\mathbb{R}^N]^d$ can expressed as

$$\mathrm{Prox}_f^\tau(q)_j = S_\tau(q)_j = \begin{cases} 0, & \text{if } \|q_j\| \le \tau \\ q_j - \tau\frac{q_j}{\|q_j\|}, & \text{otherwise} \end{cases}. \tag{6}$$

We need proper assumptions so that (4) has a unique minimizer.

Algorithm 2 Douglas-Rachford splitting (DRS) on Problem 4 with a step size $\tau > 0$.

1: $q_{k+1} = H_\tau(q_k) = \frac{\mathbb{I}+\mathrm{R}_h^\tau \mathrm{R}_f^\tau}{2}(q_k) = \mathrm{Prox}_h^\tau(\mathrm{R}_f^\tau(q_k)) + q_k - \mathrm{Prox}_f^\tau(q_k)$

2: $v_{k+1} = \mathrm{Prox}_f^\tau(q_{k+1})$

Assumption 1.1 Let u_* be the true image, $s > 0$ be a fixed accuracy parameter, $\mathcal{K}u_*$ be the gradient of the image, and $\mathcal{S}$ be the support of $\mathcal{K}u_*$. Let $|\mathcal{S}|$ denote the number of nonzero entries in $\mathcal{K}u_*$. Assume Ω is chosen uniformly at random from sets of size $|\Omega| = m \geq C_s^{-1} \cdot |\mathcal{S}| \cdot \log(N)$ for some constant C_s.

Theorem 1.1 (Theorem 1.5 in [6]) *Under Assumption 1.1 in which* $C_s \approx \frac{1}{23(s+1)}$ *for* $|\Omega| \leq N/4, s \geq 2,$ and $N \geq 20$*, with probability at least* $1-O(N^{-s})$*, the minimizer* v_* *to* (4) *is unique and* $v_* = \mathcal{K}u_*$*.*

Assume the minimizer v_* to (4) vanishes at r spatial indices, i.e., $(v_*)_j = [(v_*)_j^1 \ \cdots \ (v_*)_j^d] = 0$ for $j = j_1, \cdots, j_r$. Let $e_i \in \mathbb{R}^N$ be the standard basis in $\mathbb{R}^N$. Denote the basis vectors corresponding to zero components in v_* as e_i $(i = j_1, \cdots, j_r)$. Let $B = [e_{j_1}, \ \dots \ , \ e_{j_r}]^T \in \mathbb{R}^{r\times N}$ be selector matrix of the zero components of v_*. Let $\widetilde{B}$ be a block diagonal matrix:

$$\widetilde{B} = \begin{pmatrix} B & & \\ & \ddots & \\ & & B \end{pmatrix} \in \mathbb{R}^{dr} \times \mathbb{R}^{dN}. \tag{7}$$

For the Algorithm 2, its fixed point q_* is not unique, depending on the initial guess q_0, even if the minimizer v_* to Problem (4) is unique. Our main result is a local linear rate of Algorithm 2 solving problem (4) for *standard* fixed points similar to the ones defined in [9], in the sense of the following.

Definition 1.1 *For TVCS* (1) *with measurements denoted as* $Au = b$*, consider its equivalent problem* (4) *with a solution* v_**. Let* $\overline{\mathcal{B}_\tau(0)}$ *be the closed ball in* $\mathbb{R}^d$ *of radius* τ *centered at* 0*, and* $\overline{\mathcal{B}_\tau(0)}^c$ *be its complement. Define*

$$\mathcal{Q}_i = \begin{cases} \overline{\mathcal{B}_\tau(0)}, & \text{if } (v_*)_j = \left[(v_*)_j^1 \cdots (v_*)_j^d\right]^T = 0 \\ \overline{\mathcal{B}_\tau(0)}^c, & \text{if } (v_*)_j = \left[(v_*)_j^1 \cdots (v_*)_j^d\right]^T \neq 0 \end{cases} \subset \mathbb{R}^d,$$

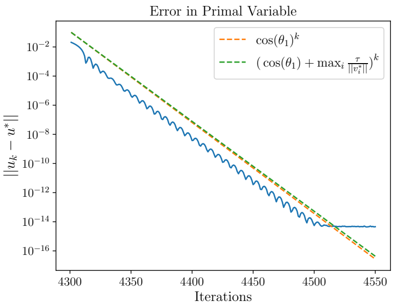
$$\mathcal{Q} = \{v \in [\mathbb{R}^N]^d : S_\tau(v)_j = 0 \iff (v_*)_j = 0\} \backsimeq \mathcal{Q}_1 \oplus \ldots \oplus \mathcal{Q}_N,$$

which is the preimage of the shrinkage operator (6) *on vectors with the same support set as* v_**. Let* q^0 *be the initial value in DRS, and* $q_* = \lim_{k\to\infty} H_\tau^k(q^0)$*. We call* $(b, A; q_0)$ *a standard problem for the DRS if* q_* *belongs to the interior of* $\mathcal{Q}$*. In this case, we call* q_* *an interior fixed point. Otherwise, we say that* $(b, A; q_0)$ *is nonstandard for DRS and that* q_* *is a boundary fixed point.*

Now the main result of this paper can be stated as follows:

Theorem 1.2 *Let θ_1 be the smallest non-zero principal angle between the two linear spaces $\mathcal{K}\mathrm{Kernel}A = \{\mathcal{K}u : u \in \mathrm{Kernel}(A)\}$ and $\mathrm{Kernel}(\widetilde{B})$ with $\widetilde{B}$ defined in (7). Consider ADMM (Algorithm 1) solving (1) with a step size $\gamma = \frac{1}{\tau} > 0$, which is equivalent to DRS (Algorithm 2) solving (4) with a step size τ. The convexity of the problem (4) implies that DRS iterates q_k converges to a fixed point q_*. Assume that q_* is a standard fixed point. Under Assumption 1.1, with probability $1-O(N^{-s})$, for small enough $\tau > 0$, there is an integer K such that for all $k \geq K$, $\|q_k - q_*\| \leq \left[\cos\theta_1 + \max_{j:\|(v_*)_j\|\neq 0} \frac{2\tau}{\|(v_*)_j\|_2}\right]^{k-K} \|q_K - q_*\|$.*

We remark that the local linear rate above looks similar to the one proven for ℓ^1-norm compressed sensing in [9], but with two differences. The first difference is that the angle θ_1 in this paper for the TV-norm is different from the angle in [9] due to the fact that the set $\mathcal{Q}$ is more complicated for TV norm. The second difference is the term $\max_{j:\|(v_*)_j\|\neq 0} \frac{2\tau}{\|(v_*)_j\|}$, which arises only in multiple dimensions, $d \geq 2$. When $d = 1$, this additional term can be removed in the proof and the main result proven in this paper reduces to the same local linear rate convergence rate in [9]. Hence, the novelty comes in providing an estimate for the linear rate of TVCS in higher dimensions albeit it is not sharp. On the other hand, as shown in Figure 1, even though the proven rate is not sharp, it is not far from a sharp rate for 3D problems of a large size.



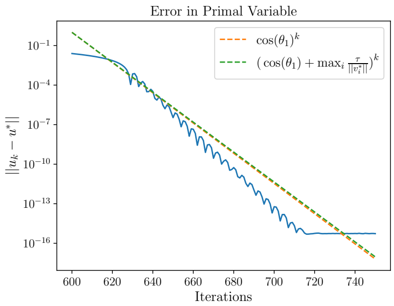


Fig. 1: The local linear rate of $u_k - u_*$ for TVCS. Here u_* is the true image and u_k is the image at k-th iteration of ADMM, i.e., x_k in Algorithm 1. Left: a 2D Shepp–Logan phantom image of size 64×64 with a step size $\gamma = \frac{1}{\tau} = 100$. Right: a 3D Shepp–Logan phantom image of size 512^3, with a step size $\gamma = \frac{1}{\tau} = 10$. In both tests, about 30% of the Fourier frequencies are observed.

1.4 Related work, contributions and outline

Convergence rates of DRS and ADMM have been studied in different settings. In [26], a global linear convergence was shown when one of the two functions is strongly convex with a Lipschitz continuous gradient. In [17,8], local linear convergence was shown under the assumptions of smoothness and strong convexity. For ℓ^1-norm compressed sensing, local linear rate is related to the first principle angle between two subspaces in [9]. In [5], local linear convergence of ADMM was shown for quadratic and linear programs as long as the solution is unique and strict complimentary condition holds. By the idea of partial smoothness developed in [23], the results of [9, 2, 5] can be unified under a general framework in [25], which shows the existence of local linear convergence for many problems, and provides explicit convergence rates if the cost functions are locally polyhedral. In [1], it was proved that applying DR or ADMM to composite problems consisting of a convex function and a convex function composed with an injective linear map yields local linear rates.

The main contribution of this paper is to provide an explicit rate for the local linear convergence of ADMM applied to isotropic TV-norm compressed sensing problem. Our explicit rate, albeit not sharp mathematically, provides some insights into behavior of ADMM for TV-norm minimization. On the other hand, the proven rate matches well with observed rate for ADMM with a large step size γ for large 3D problems as real 3D MRI data. Moreover, while our proof is largely based on the work in [9], we introduce some novel ideas for the istropic TV-norm which might be also useful for other problems such as second order cone programs. Our main techniques include exploiting the specific structure of the DRS fixed points for specific problems, and using the equivalencies of algorithms to study the local linear convergence through the equivalent problem (4). Other contributions consist of adding the recently developed algorithm G-prox PDHG [20], to the already known equivalencies among ADMM, DRS and Split-Bregman method, which will be summarized in Table 1 in Section 3.1 with derivations in the Appendix C.

The rest of this paper is organized as follows. Section 2 contains some preliminaries and notation needed. In Section 3, we provide the equivalence between ADMM and G-prox PDHG for general problems and give an explicit implementation formula for the problem (1). In Section 4, we provide the theorem and proof of our main result. Section 5 includes numerical experiments, which validate the theoretical results and show what performance we can expect for 2D and 3D problems. Section 6 gives concluding remarks.

2 Preliminaries

2.1 Notation and preliminaries

Let $\mathbb{I}$ be the identity operator. Let I be the identity matrix and I_n denote the identity matrix of size $n \times n$. For any matrix A, A^T denotes its transpose, A^* denotes its conjugate transpose and A^+ denotes its pseudo inverse. For a linear operator $\mathcal{K}$, $\mathcal{K}^*$ denotes its adjoint operator. For any $v = [v_1 \ \cdots \ v_N]^T \in [\mathbb{R}^N]^d$, the $\|\cdot\|_{1,2}$ norm is defined in (1b) and its dual norm is $\|v\|_{\infty,2} = \max_{i=1,\ldots,N} \sqrt{\sum_{i=1}^N v_i^T v_i}$. For convenience, we will also regard any $q \in [\mathbb{R}^N]^d$ as a vector in $\mathbb{R}^{Nd}$, then $\|q\|$ denotes the 2-norm in $\mathbb{R}^{Nd}$.

All functions considered in this paper are *closed, convex,* and *proper* [31, 3]. A closed extended function is also a lower semi-continuous function [3, Theorem 2.6]. If C is a closed convex set, the indicator function $\iota_C(x)$ is a closed convex proper function thus also lower semi-continuous. For a function f, its *subgradient* is a set $\partial f(x)$. We summarize a few useful results, see [3].

Theorem 2.1 *A closed convex proper function f satisfies:*

(i) $\operatorname{Prox}^{\tau}_{f\circ(-\mathbb{I})}(x) = -\operatorname{Prox}^{\tau}_{f}(-x)$.
(ii) $f^{**}(x) = f(x)$.
(iii) $\langle x, y\rangle = f(x) + f^*(y) \Leftrightarrow x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x)$.
(iv) $x^* = \operatorname{argmin}_x \langle x, y^*\rangle + f(x) \iff -y^* \in \partial f(x^*)$.
(v) Moreau Decomposition: $\operatorname{Prox}^{\gamma}_{f}(x) + \gamma \operatorname{Prox}^{\frac{1}{\gamma}}_{f^*}\left(\frac{x}{\gamma}\right) = x$.

2.2 Discrete Fourier transform and differential operators

Notation for one dimensional problems: Let $\mathcal{F}$ denote the normalized discrete Fourier transform (DFT) operator, and $\hat{u}$ denote the normalized discrete transform of $u \in \mathbb{R}^N \backsimeq \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$. We also let $\mathcal{F}$ denote the DFT matrix, i.e., $\hat{u} = \mathcal{F}(u) = \mathcal{F}u \in \mathbb{C}^N$. Let $\check{v}$ denote the inverse DFT of v, then $\check{v} = \mathcal{F}^* v$. We have $\langle u, v\rangle_{\mathbb{R}^N} = \langle \mathcal{F}u, \mathcal{F}v\rangle_{\mathbb{C}^N}, \forall u, v \in \mathbb{R}^N$ and $\langle u, v\rangle_{\mathbb{C}^N} = \langle \mathcal{F}^*u, \mathcal{F}^*v\rangle_{\mathbb{C}^N}, \forall u, v \in \mathbb{C}^N$. For the discrete gradient operator, we first consider the 1D periodic case. For $u \in \mathbb{R}^n$, define the forward difference matrix as,

$$K = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ 1 & & & -1 \end{pmatrix}, \tag{8}$$

then its transpose K^T approximates the negative derivative and $D = K^T K$ is the negative discrete Laplacian. For a one-dimensional image u, the operators $\mathcal{K}$ and $\mathcal{K}^*$ can be expressed as $\mathcal{K}u = Ku$ and $\mathcal{K}^*u = K^T u$. Let T be the normalized DFT matrix for 1D, i.e., $\hat{u} = \mathcal{F}u = Tu$, and $T^*T = I$, where T^* is the conjugate transpose of T. Notice that the matrix K in (8) is circulant, thus K can be diagonalized by DFT matrix, i.e., $K = T^*\Lambda T$ where Λ is diagonal.

Notation for multiple dimensional problems: For multiple dimensions, we focus on $d = 2$ as an example of introducing notation. For simplicity, we assume $n_1 = n_2$ for a two-dimensional image. For $U \in \mathbb{R}^{n\times n}$, let $u = \operatorname{vec}(U) \in \mathbb{R}^N$ be the column-wise vectorization of the matrix, then $(A\otimes B)u = \operatorname{vec}(BUA^T), \forall A, B \in \mathbb{C}^{N\times N}$. Define discrete gradient and negative discrete divergence as follows,

$$\nabla_h u = \begin{pmatrix} K\otimes I \\ I\otimes K \end{pmatrix} u = \begin{pmatrix} \operatorname{vec}(UK^T) \\ \operatorname{vec}(KU) \end{pmatrix}, -\nabla_h \cdot \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} K^T\otimes I & I\otimes K^T \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix},$$

where $U, V \in \mathbb{R}^{n\times n}, u = \operatorname{vec}(U), v = \operatorname{vec}(V)$. The operators $\mathcal{K}$ and $\mathcal{K}^*$ can be expressed by $\mathcal{K}u = \nabla_h u \in \mathbb{R}^N \oplus \mathbb{R}^N$, $\forall u \in \mathbb{R}^N$, and $\mathcal{K}^*p = -\nabla_h \cdot p \in \mathbb{R}^N$, $\forall p \in$

$\mathbb{R}^N \oplus \mathbb{R}^N$. Let $\mathcal{F}$ be the DFT matrix for 2D image $u = \mathrm{vec}(U) \in \mathbb{R}^N$ where $U \in \mathbb{R}^{n\times n}$, then $\hat{u} = \mathcal{F}u = (T \otimes T)u = \mathrm{vec}(TUT^T)$. With the fact

$$K \otimes I = (T^*\Lambda T) \otimes (T^* I T) = (T^* \otimes T^*)(\Lambda \otimes I)(T \otimes T) = \mathcal{F}^*(\Lambda \otimes I)\mathcal{F},$$

the operator $\mathcal{K} : \mathbb{R}^N \to \mathbb{R}^{2N} \cong \mathbb{R}^N \oplus \mathbb{R}^N$ can be decomposed as:

$$\mathcal{K} = \nabla_h = \begin{pmatrix} \mathcal{F}^* & 0 \\ 0 & \mathcal{F}^* \end{pmatrix} \begin{pmatrix} \Lambda \otimes I \\ I \otimes \Lambda \end{pmatrix} \mathcal{F} = \begin{pmatrix} \mathcal{F}^* & 0 \\ 0 & \mathcal{F}^* \end{pmatrix} \boldsymbol{\Lambda}\mathcal{F}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \Lambda \otimes I \\ I \otimes \Lambda \end{pmatrix}, \tag{9}$$

$$\mathcal{K}^* = -\nabla_h \cdot = \mathcal{F}^* \left(\Lambda^* \otimes I \;\, I \otimes \Lambda^* \right) \begin{pmatrix} \mathcal{F} & 0 \\ 0 & \mathcal{F} \end{pmatrix} = \mathcal{F}^* \boldsymbol{\Lambda}^* \widetilde{\mathcal{F}}, \quad \widetilde{\mathcal{F}} = \begin{pmatrix} \mathcal{F} & 0 \\ 0 & \mathcal{F} \end{pmatrix}. \tag{10}$$

The d-dimensional case can be defined similarly. We refer to [27, Section 2.4] for how to define $\mathrm{vec}(U)$ for a three-dimensional image U. Let K_n be the matrix in (8) of size $n \times n$, then consider the matrix constructed by one K matrix and $d-1$ identity matrices via Kronecker product:

$$\mathcal{K} = \nabla_h = \begin{pmatrix} \mathcal{K}^1 \\ \vdots \\ \mathcal{K}^d \end{pmatrix}, \quad \mathcal{K}^i = I_{n_1} \otimes \cdots \otimes K_{n_i} \otimes \cdots \otimes I_{n_d} \in \mathbb{R}^{N\times N}.$$

Recall K in (8) has an eigenvalue decomposition $K = T^*\Lambda T$. Let Λ_n be the same diagonal eigenvalue matrix of size $n \times n$. We construct the matrix:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}^1 \\ \vdots \\ \boldsymbol{\Lambda}^d \end{pmatrix}, \quad \boldsymbol{\Lambda}^i = I_{n_1} \otimes \cdots \otimes \Lambda_{n_i} \otimes \cdots \otimes I_{n_d} \in \mathbb{R}^{N\times N},$$

and let λ_k^i $(k = 1, \cdots N)$ be the diagonal entries of $\boldsymbol{\Lambda}^i$.

2.3 The constraint of partially observed Fourier frequencies

For simplicity, we focus on the case $d = 2$ and the discussion for $d \geq 3$ is similar. In (1), the constraint $\hat{u}(k) = b_k, \quad \forall k \in \Omega = \{1, i_2, \cdots, i_m\}$ can be denoted as an affine constraint $Au = b$ by a linear operator $A : \mathbb{R}^N \to \mathbb{C}^m$ with $m < N$, where the linear operator $A = M\mathcal{F}$ is a composition of a mask M and the 2D DFT matrix $\mathcal{F}$ such that $\mathcal{F}\mathcal{F}^* = I$. The mask matrix $M \in \mathbb{R}^{m\times N}$ is the submatrix of the I_N. We define $\Omega = \{1, i_2 \ldots, i_m\} \subset \{1, \ldots, N\}$ to be the indicator of which frequencies we know a priori, then $M = [e_1;\, e_{i_1};\, ...\,;\, e_{i_m}]^T \in \mathbb{R}^{m\times N}$, where e_k are the standard basis vectors in $\mathbb{R}^N$. Notice, $AA^* = I_{m\times m}$, hence its pseudo inverse is $A^+ = A^*$. For convenience, we will use the notation

$$\widetilde{M} = \begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix}, \quad \widetilde{A} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} = \begin{pmatrix} M\mathcal{F} & 0 \\ 0 & M\mathcal{F} \end{pmatrix}. \tag{11}$$

Since M is a submatrix of I_N, $M^* = M^T$. Since $\Lambda \otimes I \in \mathbb{R}^{N\times N}$ is a diagonal matrix, $M^*M(\Lambda\otimes I)$ is a diagonal matrix of size $N \times N$. Therefore, we have $M^*M(\Lambda\otimes I) = [M^*M(\Lambda\otimes I)]^T = (\Lambda\otimes I)M^*M$, thus

$$\widetilde{M^*}\widetilde{M}\boldsymbol{\Lambda} = \begin{pmatrix} M^* & 0 \\ 0 & M^* \end{pmatrix}\begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix}\begin{pmatrix} \Lambda\otimes I \\ I\otimes\Lambda \end{pmatrix} = \boldsymbol{\Lambda}M^*M. \tag{12}$$

Similarly, $\boldsymbol{\Lambda}^*\boldsymbol{\Lambda} = \Lambda^*\Lambda\otimes I + I\otimes\Lambda^*\Lambda$ is a diagonal matrix, thus

$$M^*M(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+ = (\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+M^*M. \tag{13}$$

3 Equivalency to G-prox PDHG and an implementation formula

3.1 The equivalence between ADMM and G-prox PDHG

In this section, we first give an equivalent primal dual formulation of ADMM then provide an implementation formula for TV compressed sensing problem. The G-prox PDHG method introduced in [20] for solving a composite convex minimization problem (2) can be written as Algorithm 3. The equivalence of G-prox PDHG above and ADMM is stated in Theorem 3.1, which will be proven in Appendix C. There are many known equivalent yet seemingly different formulations of the ADMM in Algorithm 1. We provide a summary of the variables that are equivalent in these algorithms in Table 1. These relations can be modified to extend to the generalized forms of these algorithms.

Algorithm 3 G-prox PDHG with step sizes $\tau, \sigma > 0$.
Initial guess $u_0 \in \mathbb{R}^N, v_0, w_0 \in [\mathbb{R}^N]^d$.

1: $u_{k+1} = \operatorname{argmin}_u g(u) + \langle \mathcal{K}u, w_k\rangle + \frac{1}{2\tau}\|\mathcal{K}(u-u_k)\|^2$

2: $v_{k+1} = \operatorname{argmax}_v -f^*(v) + \langle \mathcal{K}u_{k+1}, v\rangle - \frac{1}{2\sigma}\|v-v_k\|^2$

3: $w_{k+1} = 2v_{k+1} - v_k$

Theorem 3.1 *Algorithm 1 (ADMM) with a step size $\gamma > 0$ is equivalent to Algorithm 3 (G-prox PDHG) with $\tau = \frac{1}{\sigma} = \frac{1}{\gamma}$ via the change of variables:* $u_k := x_k, \quad p_k := z_k, \quad w_k := \frac{1}{\tau}Kx_k + z_k - \frac{1}{\tau}y_k$.

3.2 An explicit implementation formula of G-prox PDHG

For any vector $v = \left[v^1 \dots v^d\right]^T \in [\mathbb{R}^N]^d$ with $v^i = \left[v^i_1 \dots v^i_N\right]^T \in \mathbb{R}^N$, let v_j denote $v_j = \left[v_j^1 \cdots v_j^d\right]^T \in \mathbb{R}^d$. Define $|v| := \left[\|v_1\| \dots \|v_N\|\right]^T \in \mathbb{R}^N$, and $\frac{v^i}{\max(1,|v|)} = \left[v^i_1/\max(1,\|v_1\|)\dots v^i_N/\max(1,\|v_N\|)\right]^T \in \mathbb{R}^N$. Then we define,

$$\frac{v}{\max(1,|v|)} := \begin{bmatrix} v^1/\max(1,|v|) \\ \vdots \\ v^d/\max(1,|v|) \end{bmatrix} \in [\mathbb{R}^N]^d.$$

Table 1: A summarization of equivalent variables in ADMM, DRS and Gprox PDHG algorithms with proper step sizes: variables in each row are equivalent.

Method	ADMM for (2)	Douglas-Rachford for (4)	G-prox PDHG for (2)
Formula	Alg. 1 for (2)	Alg. 2 on (4)	Alg. 3 on (2)
Step Size	$\gamma = \frac{1}{\tau}$	τ	$\sigma = \frac{1}{\tau}$
Primal Iterate	$\mathcal{K}x_k$	$q_k - (q_{k-1} - v_{k-1})$	$\mathcal{K}u_k$
Dual Iterate	z_k	$\frac{q_k - v_k}{\tau}$	p_k
Extragradient	$\frac{1}{\tau}\mathcal{K}x_k + z_k - \frac{1}{\tau}y_k$	$\frac{2q_k - q_{k-1}}{\tau} - \frac{2v_k - v_{k-1}}{\tau}$	w_k

Let $\overline{\lambda}$ denote the complex conjugate of λ. For $w_n \in [\mathbb{R}^N]^d$, where n will be the iteration index, we also denote it by $w_n = \left[(w_n)^1 \cdots (w_n)^d\right]^T$ with $(w_n)^i \in \mathbb{R}^N$ and $\widehat{(w_n)^i}$ being the d-dimensional discrete Fourier transform of $(w_n)^i$. With the notation in Section 2.2, for the TV compressed-sensing problem (1), Algorithm 3 can be explicitly implemented in Fourier space as described by Algorithm 4. The derivation of Algorithm 4 will be given in Appendix D.

Algorithm 4 An implementation formula of G-prox PDHG with a step size $\tau > 0$ and $\sigma = \frac{1}{\tau}$ (or equivalently ADMM in Algorithm 1 with $\gamma = \frac{1}{\tau}$) for TV-norm compressed sensing.

Initial guess: $u_0 \in \mathbb{R}^N, v_0, w_0 \in [\mathbb{R}^N]^d$

1: $\begin{cases} \widehat{u_{n+1}}(k) &= b_k, \quad k \in \Omega \\ \widehat{u_{n+1}}(k) &= \widehat{u_n}(k) - \tau \left[\sum_{i=1}^{d} \overline{\lambda_k^i}\,\widehat{(w_n)^i}(k)\right] / \left[\sum_{i=1}^{d} |\lambda_k^i|^2\right], \; k \notin \Omega \end{cases}$

2: $v_{k+1} = \frac{v_k + \sigma \mathcal{K} u_{k+1}}{\max(1, |v_k + \sigma \mathcal{K} u_{k+1}|)}$,

3: $w_{n+1} = 2v_{n+1} - v_n$.

Notation: n is the iteration index and k is the frequency index. For any $w_n \in [\mathbb{R}^N]^d$, let $w_n = \left[(w_n)^1 \cdots (w_n)^d\right]^T$ with $(w_n)^i \in \mathbb{R}^N$, then $\widehat{(w_n)^i}$ denotes the discrete Fourier transform of $(w_n)^i$, and $\widehat{(w_n)^i}(k)$ denotes the component of $\widehat{(w_n)^i}$ at the k-th frequency. As defined in Section 2.2, λ_k^i $(k = 1, \cdots, N)$ are diagonal entries of $\boldsymbol{\Lambda}^i$.

4 The main result on an explicit local linear rate

We prove the main result in this section. For simplicity, we focus on the case $d = 2$, and extensions to higher dimensions are straightforward.

4.1 DRS on the equivalent problem

In order to analyze the local linear convergence of ADMM, we will utilize some of the equivalent formulations. Recall that TVCS problem (1) can be written as the primal formulation (2), and its Fenchel dual formulation is given as (3). The dual formulation of (3) can be written as (4). We first make a few remarks about total duality. We have strong duality between the primal and dual problem due to Slater's conditions, which are satisfied if $\exists$ x s.t. $x \in$

$\mathrm{ri}(\mathrm{dom} f) = \mathbb{R}^N$ and $Ax = b$. For strong duality between (3) and (4), Slater's conditions are satisfied by choosing $p = 0 \in \mathbb{R}^{2N}$ which implies $\|p\|_{\infty,2} < 1$, i.e $p = 0 \in \mathrm{ri}(\mathrm{dom} f^*)$, and $\mathcal{K}^* p \in \mathrm{Range}(A^*)$. To show total duality, we need existence of a solution of (4). By Theorem 1.1, under Assumption 1.1, with high probability, (4) has a unique minimizer. Thus total duality holds.

4.2 The proximal operators

For the two functions f and h in (4), we need their proximal operators for studying DRS. Since the function $h(v) = \iota_{\mathcal{K}\{u:Au=b\}}(v)$ is an indicator function to an affine set, the proximal operator is the Euclidean projection to the set. With derivation shown in Appendix A, the projection formula can be given as

$$\mathrm{Prox}_h^{\tau}(q) = \widetilde{\mathcal{F}}^*(I - \widetilde{M}^*\widetilde{M})\Sigma\widetilde{\mathcal{F}}q + \widetilde{\mathcal{F}}^*\widetilde{M}^*\widetilde{M}\boldsymbol{\Lambda}M^*b, \quad \Sigma = \boldsymbol{\Lambda}(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+\boldsymbol{\Lambda}^*, \tag{14}$$

where $(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+$ is the pseudo inverse of $\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}$. Next, we discuss S_τ.

Definition 4.1 For any $q = [q^1 \cdots q^d]^T \in [\mathbb{R}^N]^d$ with $q^i = [q_1^i \cdots q_N^i]^T \in \mathbb{R}^N$, which can also be represented by $q_j = [q_j^1 \cdots q_j^d] \in \mathbb{R}^d$ with a spatial index $j = 1, \cdots, N$, define an operator $\mathcal{N} : [\mathbb{R}^N]^d \to [\mathbb{R}^N]^d$ via the spatial index as

$$\mathcal{N}(q)_j = \begin{cases} 0, & \text{if } q_j = 0, \\ \frac{q_j}{\|q_j\|} & \text{otherwise} \end{cases} \in \mathbb{R}^d, \quad j = 1, \cdots, N.$$

Recall that we have defined $B = [e_{j_1}, \ \dots \ , \ e_{j_r}]^T \in \mathbb{R}^{r\times N}$ to be the selector matrix of the zero components of v_*. For any $q \in \mathcal{Q}$, with $\mathcal{Q}$ in Definition 1.1, it is straightforward to verify that the shrinkage operator can be written as

$$S_\tau(q) = (I - \widetilde{B}^+\widetilde{B})(q - \tau\mathcal{N}(q)), \quad \forall q \in \mathcal{Q}, \tag{15}$$

in which we regard q and $\mathcal{N}(q)$ as column vectors in $\mathbb{R}^{Nd}$.

Lemma 4.1 *Under Assumption 1.1, with probability* $1 - O(N^{-s})$, $\mathcal{K}\mathrm{Kernel}(A) \cap \mathrm{Kernel}(\widetilde{B}) = \{0\}$, *where* $\mathcal{K}\mathrm{Kernel}(A) = \{v \in \mathbb{R}^{N\times d} : v = \mathcal{K}u, \ \ u \in \mathrm{Kernel}(A)\}$.

Proof Consider any $z \in \mathcal{K}\mathrm{Kernel}(A) \cap \mathrm{Kernel}(\widetilde{B})$. First,

$$z \in \mathcal{K}\mathrm{Kernel}(A) \Rightarrow z = \mathcal{K}u, u \in \mathrm{Kernel}(A).$$

By the fact that $MM^* = I_{m\times m}$ and the notation in (9) and (11),

$$\widetilde{A}\mathcal{K} = \begin{pmatrix} M\mathcal{F} & 0 \\ 0 & M\mathcal{F} \end{pmatrix} \begin{pmatrix} \mathcal{F}^* & 0 \\ 0 & \mathcal{F}^* \end{pmatrix} \boldsymbol{\Lambda}\mathcal{F} = \begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix} \boldsymbol{\Lambda}\mathcal{F} = \widetilde{M}\boldsymbol{\Lambda}\mathcal{F} = (\widetilde{MM^*})\widetilde{M}\boldsymbol{\Lambda}\mathcal{F}.$$

By the property (12) and $u \in \mathrm{Kernel}(A)$, we have

$$\widetilde{A}z = \widetilde{A}\mathcal{K}u = \widetilde{MM^*}\widetilde{M}\boldsymbol{\Lambda}\mathcal{F}u = \widetilde{M}\boldsymbol{\Lambda}M^*M\mathcal{F}u = \widetilde{M}\boldsymbol{\Lambda}M^*Au = 0.$$

Second, $z \in \text{Kernel}(\widetilde{B})$ implies the support of z is contained in the same support, S, as the unique solution v_* to (4). Let A_S denote the partial Fourier Transform restricted to signals with the support included in the set S. Then

$$\begin{pmatrix} A_S & 0 \\ 0 & A_S \end{pmatrix} z = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} z = \widetilde{A} z = 0.$$

By Theorem 3.1 in [6], A_S is injective, which implies $z = 0$. □

Remark 4.1 For ℓ^1-norm compressed sensing, there are necessary [38] and sufficient [14] conditions to ensure a unique solution to (2), and the same techniques can be used to show $\mathcal{K}\text{Kernel}(A) \cap \text{Kernel}(\widetilde{B}) = \{0\}$ for one-dimensional TVCS problem, i.e., Problem (2) with $d = 1$. However, such a proof breaks down for $d \geq 2$. As shown in Lemma 4.1 above, $\mathcal{K}\text{Kernel}(A) \cap \text{Kernel}(\widetilde{B}) = \{0\}$ can be ensured by the robust uncertainty principle in [6].

4.3 Characterization of the fixed points to DRS

For the function $h(v) = \iota_{\mathcal{K}\{u:Au=b\}}(v)$, we have

$$\partial h(q) = \{q : \mathcal{K}^* q \in \text{Range}(A^*)\} = (\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big],$$

where $(\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big]$ denotes the pre-image of $\text{Range}(A^*)$ under the operator $\mathcal{K}^*$. By the optimality condition of (4), its minimizer v_* satisfies $0 \in \partial f(v_*) + (\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big]$, therefore $\partial f(v_*) \cap (\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big] \neq \emptyset$. Any vector $\eta \in \partial f(v_*) \cap (\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big]$ is called a *dual certificate*. The subgradient of $f = \|\cdot\|_{1,2}$ is given below as

$$\partial f(v_*) = \left\{ w \in \mathbb{R}^{Nd} : w_j \in \begin{cases} \frac{(v_*)_j}{||(v_*)_j||} & \text{if } (v_*)_j \neq 0 \\ \mathcal{B}_1(0) & \text{else} \end{cases} \right\}. \tag{16}$$

Theorem 1.1 (Theorem 1.5 in [6]) gives existence and uniqueness of the minimizer v_*, which implies the existence of a dual certificate.

Lemma 4.2 *The set of fixed points of DRS iteration operator* $H_\tau = \frac{\mathbb{I} + \mathrm{R}_h^\tau \mathrm{R}_f^\tau}{2}$ *for the problem* (4) *is given by:*

$$\{v_* + \tau\eta : \eta \in \partial f(v_*) \cap (\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big]\}, \tag{17}$$

and the fixed point is unique if and only if $(\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big] \cap \text{Range}(\widetilde{B}^T) = \{0\}$ *where* $\widetilde{B}^T$ *is the transpose matrix of* $\widetilde{B}$ *with* $\widetilde{B}$ *defined in* (7)*.*

Proof Consider any $\eta \in \partial f(v_*) \cap (\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big]$. First, since S_τ is the proximal operator of $f(v)$, $\eta \in \partial f(v_*)$ implies $S_\tau(v_* + \tau\eta) = v_*$. Second, by (10) and $A = M\mathcal{F}$, we have

$$\eta \in (\mathcal{K}^*)^{-1}\big[\text{Range}(A^*)\big] \Rightarrow \mathcal{K}^*\eta \in \text{Range}(A^*) \Rightarrow \mathcal{F}^* \boldsymbol{\Lambda}^* \widetilde{\mathcal{F}}\eta \in \text{Range}(\mathcal{F}^* M^*),$$

$$\Rightarrow \boldsymbol{\Lambda}^* \widetilde{\mathcal{F}}\eta \in \mathrm{Range}(M^*) \Rightarrow (I - M^*M)\boldsymbol{\Lambda}^* \widetilde{\mathcal{F}}\eta = 0.$$

By (12) and (13), we have

$$(I - \widetilde{M}^*\widetilde{M})\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+(I - M^*M)\boldsymbol{\Lambda}^*$$

$$\Rightarrow (I - \widetilde{M}^*\widetilde{M})\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+\boldsymbol{\Lambda}^*\widetilde{\mathcal{F}}\eta = \boldsymbol{\Lambda}(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+(I - M^*M)\boldsymbol{\Lambda}^*\widetilde{\mathcal{F}}\eta = 0.$$

Since $v_* = \mathcal{K}u_*$ and $Au_* = b$, by (14) and (9), we have

$$\begin{aligned}\mathrm{Prox}_h^\tau(v_* - \tau\eta) =& \widetilde{\mathcal{F}}^*(I - \widetilde{M}^*\widetilde{M})\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+\boldsymbol{\Lambda}^*\widetilde{\mathcal{F}}(v_* - \tau\eta) + \widetilde{\mathcal{F}}^*\widetilde{M}^*\widetilde{M}\boldsymbol{\Lambda}M^*b\\ =& \widetilde{\mathcal{F}}^*(I - \widetilde{M}^*\widetilde{M})\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^*\boldsymbol{\Lambda})^+\boldsymbol{\Lambda}^*\widetilde{\mathcal{F}}\mathcal{K}u_* + \widetilde{\mathcal{F}}^*\widetilde{M}^*\widetilde{M}\boldsymbol{\Lambda}M^*Au_*\\ =& \widetilde{\mathcal{F}}^*(I - \widetilde{M}^*\widetilde{M})\boldsymbol{\Lambda}(\mathcal{F}u_*) + \widetilde{\mathcal{F}}^*\widetilde{M}^*\widetilde{M}\boldsymbol{\Lambda}\mathcal{F}u_* = \widetilde{\mathcal{F}}^*\boldsymbol{\Lambda}\mathcal{F}u_* = \mathcal{K}u_* = v_*.\end{aligned}$$

Moreover, $\mathrm{Prox}_f^\tau(v_* + \tau\eta) = v_*$ implies $\mathrm{R}_f^\tau(v_* + \tau\eta) = v_* - \tau\eta$. Thus,

$$\begin{aligned}H_\tau(v_* + \tau\eta) &= \mathrm{Prox}_h^\tau(\mathrm{R}_f^\tau(v_* + \tau\eta)) + v_* + \tau\eta - \mathrm{Prox}_f^\tau(v_* + \tau\eta)\\ &= \mathrm{Prox}_h^\tau(v_* - \tau\eta) + \tau\eta = v_* + \tau\eta.\end{aligned}$$

Conversely, suppose $H_\tau(q) = q$ and define $\eta = \frac{q-v_*}{\tau}$. We want to show $\eta \in \partial f(v_*) \cap (\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$. By the convergence of the DRS iteration [26], $\mathrm{Prox}_f^\tau(q) = v_*$, which implies that $\eta = \frac{q-v_*}{\tau} \in \partial f(v_*)$. Second, $H_\tau(q) = q$ and $\mathrm{Prox}_f^\tau(q) = v_*$ imply $v_* = \mathrm{Prox}_h^\tau(2v_* - q)$, which gives $-\eta \in \partial h(v_*) = (\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$ thus $\eta \in (\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$.

To discuss uniqueness, let $q_1 = v_* + \tau\eta_1, q_2 = v_* + \tau\eta_2$ be two fixed points of H_τ. Then $q_1 - q_2 = \tau(\eta_1 - \eta_2)$, where $\eta_1, \eta_2 \in \partial f(v_*) \cap (\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$. From (16), note that $\eta_1, \eta_2 \in \partial f(v_*)$ implies that $\pm(\eta_1 - \eta_2) \in \mathrm{Range}(\widetilde{B}^T)$. Hence, $q_1 - q_2 \in (\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big] \cap \mathrm{Range}(\widetilde{B}^T)$, so the fixed point is unique if and only if $(\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big] \cap \mathrm{Range}(\widetilde{B}^T) = \{0\}$. $\square$

4.4 Characterization of the DRS operator H_τ

Next, we estimate the nonlinear DRS operator H_τ.

Lemma 4.3 *For any fixed point q_* of $H_\tau = \frac{\mathbb{I}+\mathrm{R}_h^\tau \mathrm{R}_f^\tau}{2}$, it satisfies*

$$\|(I - \widetilde{B}^+\widetilde{B})\mathcal{N}(q) - \mathcal{N}(q_*)\| \le \max_{j:\|(v_*)_j\|\neq 0} \frac{2}{\|(v_*)_j\|}\|q - q_*\|, \quad \forall q \in [\mathbb{R}^N]^d \backsimeq \mathbb{R}^{Nd},$$

where $\|\cdot\|$ is the 2-norm in $\mathbb{R}^{Nd}$.

Proof By Definition 4.1 and the definition of $\widetilde{B}$ in (7), we have

$$\|(I - \widetilde{B}^+\widetilde{B})\mathcal{N}(q) - \mathcal{N}(q_*)\|^2 = \sum_{i:(v_*)_i\neq 0}\left\|\frac{q_i}{\|q_i\|} - \frac{(q_*)_i}{\|(q_*)_i\|}\right\|^2.$$

For any nonzero $a, b \in \mathbb{R}^d$, we have $\left\| \frac{a}{\|a\|} - \frac{b}{\|b\|} \right\| \leq \left\| \frac{a}{\|a\|} - \frac{b}{\|a\|} \right\| + \left\| \frac{b}{\|a\|} - \frac{b}{\|b\|} \right\| = \frac{1}{\|a\|} \|a - b\| + \|b\| \left| \frac{\|b\| - \|a\|}{\|a\|\|b\|} \right| = \frac{1}{\|a\|} \Big(\|a - b\| + \big| \|b\| - \|a\| \big| \Big) \leq \frac{2}{\|a\|} \|a - b\|$. By Lemma 4.2, $q_* = v_* + \tau\eta$ for a dual certificate η. For any index i satisfying $(v_*)_i \neq 0$, we have $q_i^* = (v_*)_i + \tau \frac{(v_*)_i}{\|(v_*)_i\|}$, which is implied by $\eta \in \partial \|v_*\|_{1,2}$. Hence, $\|(q_*)_i\| \geq \|(v_*)_i\|$. If we also use the inequality above with $a = q_i,\ b = (q_*)_i$, we obtain $\left\| \frac{q_i}{\|q_i\|} - \frac{(q_*)_i}{\|(q_*)_i\|} \right\| \leq \frac{2}{\|(v_*)_i\|} \|q_i - (q_*)_i\|$. □

Lemma 4.4 *For any $q \in \mathcal{Q}$ with $\mathcal{Q}$ and any DRS fixed point q_*,*

$$H_\tau(q) - H_\tau(q_*) = \widetilde{H}(q - q_*) + \tau \Big[(I - 2C)(I - \widetilde{B}^+ \widetilde{B}) \Big] (\mathcal{N}(q) - \mathcal{N}(q_*)),$$

where $\widetilde{H} = \Big[C(I - \widetilde{B}^+ \widetilde{B}) + (I - C)\widetilde{B}^+ \widetilde{B} \Big]$ *and* $C = \widetilde{\mathcal{F}}^* (I - \widetilde{M}^* \widetilde{M}) \boldsymbol{\Lambda} (\boldsymbol{\Lambda}^* \boldsymbol{\Lambda})^+ \boldsymbol{\Lambda}^* \widetilde{\mathcal{F}}$.

Proof By (15), we have $\mathrm{Prox}_f^\tau(q) = S_\tau(q) = (I - \widetilde{B}^+ \widetilde{B})\big(q - \tau \mathcal{N}(q)\big)$, thus $\mathrm{R}_f^\tau(q) = 2(I - \widetilde{B}^+ \widetilde{B})\big(q - \tau \mathcal{N}(q)\big) - q$. By (14) and C in Lemma 4.4, we have $\mathrm{Prox}_h^\tau(q) = Cq + \widetilde{b}, \quad \widetilde{b} = \widetilde{\mathcal{F}}^* \widetilde{M}^* \widetilde{M} \boldsymbol{\Lambda} M^* b$. Hence,

$$\begin{aligned} H_\tau(q) &= \mathrm{Prox}_h^\tau \left(\mathrm{R}_f^\tau(q) \right) + q - \mathrm{Prox}_f^\tau(q) \\ &= C\big[(I - 2\widetilde{B}^+ \widetilde{B})q - 2\tau (I - \widetilde{B}^+ \widetilde{B}) \mathcal{N}(q) \big] + \widetilde{b} + \widetilde{B}^+ \widetilde{B} q + \tau (I - \widetilde{B}^+ \widetilde{B}) \mathcal{N}(q), \end{aligned}$$

$$\begin{aligned} & H_\tau(q) - H_\tau(q_*) \\ =& \Big[C(I - \widetilde{B}^+ \widetilde{B}) + (I - C)\widetilde{B}^+ \widetilde{B} \Big] (q - q_*) + \tau \Big[(I - 2C)(I - \widetilde{B}^+ \widetilde{B}) \Big] (\mathcal{N}(q) - \mathcal{N}(q_*)) \\ =& \widetilde{H}(q - q_*) + \tau \Big[(I - C)(I - \widetilde{B}^+ \widetilde{B}) - C(I - \widetilde{B}^+ \widetilde{B}) \Big] (\mathcal{N}(q) - \mathcal{N}(q_*)). \end{aligned}$$

This concludes the proof. □

Notice that C in Lemma 4.4 is the projection matrix onto $\mathcal{K}\mathrm{Kernel}(A)$, and $I - C$ is the projection matrix onto $(\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$. Since B and $\widetilde{B}$ in (7) are also projection matrices, we may rewrite them as follows. Define C_0 as the $2N \times (N - m)$ matrix whose columns form an orthonormal bases of $\mathcal{K}\mathrm{Kernel}(A)$, and C_1 the $2N \times (N + m)$ matrix whose columns form an orthonormal bases of $(\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$. Similarly we define the $2N \times 2(N - r)$ matrix B_0 and $2N \times 2r$ matrix B_1 to be the matrices whose columns are orthonormal bases of $\mathrm{Kernel}(\widetilde{B})$ and $\mathrm{Range}(\widetilde{B}^*)$ respectively. Therefore, we have

$$C_0 C_0^* + C_1 C_1^* = I, \quad B_0 B_0^* + B_1 B_1^* = I, \tag{18}$$

and we can rewrite expression in Lemma 4.4 as:

$$\begin{aligned} H_\tau(q) - H_\tau(q_*) =& \big[C_0 C_0^* B_0 B_0^* + C_1 C_1^* B_1 B_1^* \big] (q - q_*) \\ &+ \tau \big[C_1 C_1^* B_0 B_0^* - C_0 C_0^* B_0 B_0^* \big] (\mathcal{N}(q) - \mathcal{N}(q_*)). \end{aligned} \tag{19}$$

Definition 4.2 [4] Let $\mathcal{U}$ and $\mathcal{V}$ be two subspaces of a linear space with $\dim(\mathcal{U}) = s \leq \dim(\mathcal{V})$. The principal angles $\theta_k \in [0, \frac{\pi}{2}]$ $(k = 1, \ldots, p)$ between $\mathcal{U}$ and $\mathcal{V}$, and principal vectors vectors u_j and v_j are defined recursively as:

$$\cos\theta_k = \max_{u\in\mathcal{U}} \max_{v\in\mathcal{V}} \langle u_k, v_k\rangle, \|u\| = \|v\| = 1, \ \langle u_k, u_j\rangle = \langle v_k, v_j\rangle = 0, \ \forall j < k.$$

Without loss of generality, assume $N-m \leq 2(N-r)$. Let θ_i $(i = 1, \ldots, N-m)$ be the principal angles between $\mathcal{K}\mathrm{Kernel}(A)$ and $\mathrm{Kernel}(\widetilde{B})$. Then $\theta_1 > 0$ since $\mathcal{K}\mathrm{Kernel}(A) \cap \mathrm{Kernel}(\widetilde{B}) = \{0\}$ by Lemma 4.1. Let $\cos\Theta$ be the $(N-m)\times(N-m)$ diagonal matrix with diagonal entries $(\cos\theta_1, \ldots, \cos\theta_{N-m})$. By [4, Theorem 1], the Singular Value Decomposition (SVD) of the $(N-m)\times 2(N-r)$ matrix $E_0 = C_0^* B_0$ is $E_0 = U_0 \cos\Theta V^*$, with $V^*V = U_0^* U_0 = I_{(N-m)}$, and columns of $C_0 U_0$ and $B_0 V$ give the principal vectors. By the definition of SVD, V is a matrix of size $2(N-r)\times(N-m)$, with orthonormal columns. Let V' be a matrix of size $2(N-r)\times(N-2r+m)$ whose columns are normalized and orthogonal to columns of V. Define $\widetilde{V} = (V|V')$, then $\widetilde{V}$ is a unitary matrix of size $2(N-r)\times 2(N-r)$. Now consider $E_1 = C_1^* B_0$, then by (18) we have

$$\begin{aligned} E_1^* E_1 = B_0^* C_1 C_1^* B_0 = B_0^* B_0 - B_0^* C_0 C_0^* B_0 = I_{(2N-2r)} - E_0^* E_0 \\ = I_{(2N-2r)} - V\cos^2\Theta V^* = \widetilde{V} \begin{pmatrix} \sin^2\Theta & 0 \\ 0 & I_{(N-2r+m)} \end{pmatrix} \widetilde{V}^*, \end{aligned}$$

which implies the SVD $E_1 = U_1 \begin{pmatrix} \sin\Theta & 0 \\ 0 & I_{(N-2r+m)} \end{pmatrix} \widetilde{V}^*$. Thus we have

$$E_0 E_0^* = U_0 \cos^2\Theta U_0^*, \quad E_1 E_1^* = U_1 \begin{pmatrix} \sin^2\Theta & 0 \\ 0 & I_{(N-2r+m)} \end{pmatrix} U_1^*,$$

$$E_1 E_0^* = U_1 \begin{pmatrix} \sin\Theta\cos\Theta \\ 0 \end{pmatrix} U_0^*, \quad E_0 E_1^* = U_0 \begin{pmatrix} \cos\Theta\sin\Theta | 0 \end{pmatrix} U_1^*.$$

Notice that $B_0 = (C_0 C_0^* + C_1 C_1^*) B_0 = C_0 E_0 + C_1 E_1$, so we obtain

$$\begin{aligned} B_0 B_0^* &= (C_0|C_1) \left(\begin{array}{c|c} E_0 E_0^* & E_0 E_1^* \\ \hline E_1 E_0^* & E_1 E_1^* \end{array}\right) (C_0|C_1)^* \\ &= (C_0 U_0 | C_1 U_1) \left(\begin{array}{c|cc} \cos^2\Theta & \cos\Theta\sin\Theta & 0 \\ \hline \sin\Theta\cos\Theta & \sin^2\Theta & 0 \\ 0 & 0 & I_{(N-2r+m)} \end{array}\right) (C_0 U_0 | C_1 U_1)^*. \end{aligned}$$

Define $\widetilde{C}_0 = C_0 U_0$ and $\widetilde{C}_1 = C_1 U_1$, which are $2N\times(N-m)$ and $2N\times 2(N-r)$ matrices respectively. Then the columns of $\widetilde{C}_0$ form a basis of $\mathcal{K}\mathrm{Kernel}(A)$, and columns of $\widetilde{C}_1$ are orthonormal vectors in $(\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$. Let D denote the orthogonal complement of $(\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big] \cap \mathrm{Range}(\widetilde{B}^*)$ in the subspace $(\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$. Then $\dim(D) = 2(N-r)$. Since $\theta_1 > 0$, the largest angle between $(\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$ and $\mathrm{Kernel}(\widetilde{B})$ is less than $\pi/2$. So none of the column vectors of $\widetilde{C}_1$ are orthogonal to $\mathrm{Kernel}(\widetilde{B})$. Hence, by counting the dimension of D and columns of $\widetilde{C}_1$, we conclude that columns of $\widetilde{C}_1$ form an orthonormal basis of D. Define $\widetilde{C}_2$ to be the $2N \times (m+2r-N)$ matrix whose columns form an orthonormal basis of $\mathrm{Range}(\widetilde{B}^*) \cap (\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$. Then,

$$B_0 B_0^* = (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2) \left(\begin{array}{c|cc|c} \cos^2\Theta & \cos\Theta\sin\Theta & 0 & 0 \\ \hline \sin\Theta\cos\Theta & \sin^2\Theta & 0 & 0 \\ 0 & 0 & I_{(N-2r+m)} & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2)^*.$$

By (18), $B_1 B_1^* = I - B_0 B_0^*$, thus

$$B_1 B_1^* = (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2) \left(\begin{array}{c|cc|c} \sin^2\Theta & -\cos\Theta\sin\Theta & 0 & 0 \\ \hline -\sin\Theta\cos\Theta & \cos^2\Theta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & I_{(m+2r-N)} \end{array}\right) (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2)^*,$$

$$\begin{aligned} \widetilde{H} &= C_0 C_0^* B_0 B_0^* + C_1 C_1^* B_1 B_1^* \\ &= (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2) \left(\begin{array}{c|cc|c} \cos^2\Theta & \cos\Theta\sin\Theta & 0 & 0 \\ \hline -\sin\Theta\cos\Theta & \cos^2\Theta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & I_{(m+2r-N)} \end{array}\right) (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2)^*, \end{aligned} \tag{20}$$

$$\begin{aligned} & C_1 C_1^* B_0 B_0^* - C_0 C_0^* B_0 B_0^* \\ =& (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2) \left(\begin{array}{c|cc|c} -\cos^2\Theta & -\cos\Theta\sin\Theta & 0 & 0 \\ \hline \sin\Theta\cos\Theta & \sin^2\Theta & 0 & 0 \\ 0 & 0 & I_{(N-2r+m)} & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2)^*. \end{aligned}$$

We now summarize the discussion above as the following result:

Lemma 4.5 *For any $q \in \mathcal{Q}$ and any DRS fixed point q_*,*

$$\begin{aligned} H_\tau(q) - H_\tau(q_*) = \widetilde{C} & \left(\begin{array}{c|cc|c} \cos^2\Theta & \cos\Theta\sin\Theta & 0 & 0 \\ \hline -\sin\Theta\cos\Theta & \cos^2\Theta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & I_{(m+2r-N)} \end{array}\right) \widetilde{C}^*(q - q_*) \\ + \widetilde{C} & \left(\begin{array}{c|cc|c} -\cos^2\Theta & -\cos\Theta\sin\Theta & 0 & 0 \\ \hline \sin\Theta\cos\Theta & \sin^2\Theta & 0 & 0 \\ 0 & 0 & I_{(N-2r+m)} & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) \widetilde{C}^* \tau\big[\mathcal{N}(q) - \mathcal{N}(q_*)\big], \end{aligned}$$

where $\widetilde{C} = (\widetilde{C}_0|\widetilde{C}_1|\widetilde{C}_2)$ and $\widetilde{C}_2$ is the $2N \times (m+2r-N)$ matrix with columns forming an orthonormal basis of $\mathrm{Range}(\widetilde{B}^*) \cap (\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big]$.

Lemma 4.6 *Assume DRS iterates q_k converge to an interior fixed point q_*. Then there exists $K \in \mathbb{N}$ such that for all $k \geq K$, $\mathbb{P}(q_k - q_*) = 0$, where $\mathbb{P}$ is the Euclidean projection to* $\mathrm{Range}(\widetilde{B}^*) \cap (\mathcal{K}^*)^{-1}[\mathrm{Range}(A^*)]$.

Proof Since q_* is in the interior of $\mathcal{Q}$, there exists K such that $q_k \in \mathcal{Q}$ for all $k \geq K$. Since columns of $\widetilde{C}_2$ span the subspace $\mathrm{Range}(\widetilde{B}^*) \cap (\mathcal{K}^*)^{-1}[\mathrm{Range}(A^*)]$, Lemma 4.5 and Lemma 4.4 imply that $\mathbb{P}(q_k - q_*) = \mathbb{P}(q_K - q_*)$ for all $k \geq K$. If $\mathbb{P}(q_K - q_*) \neq 0$, then $\mathbb{P}(q_k - q_*)$ is a constant for $k \geq K$, which contradicts with $q_k \to q_*$. So $\mathbb{P}(q_K - q_*) = 0$, which implies $\mathbb{P}(q_k - q_*) = 0$ for any $k \geq K$. $\square$

4.5 The proof of the main theorem

Now we are ready to prove Theorem 1.2.

Proof First of all, by Definition 1.1 and Lemma 4.2, any DRS fixed point is in the set $\mathcal{Q}$. The convexity of the problem (4) ensures that DRS iterates converges to the minimizer v_*, i.e., q_k converges to some fixed point q_* to DRS and $S_\tau(q_*) = \mathrm{Prox}_f^\tau(q_*) = v_*$. For a standard problem, q_* is the interior of the set $\mathcal{Q}$. We first discuss a simple case that $\mathrm{Range}(\widetilde{B}^*) \cap (\mathcal{K}^*)^{-1}\left[\mathrm{Range}(A^*)\right] = \{0\}$. By Lemma 4.2 and the definition of $\widetilde{C}_2$, we deduce that the fixed point is unique and $m + 2r = N$. Notice (20) shows that $\|\widetilde{H}\|_2 = \cos\theta_1$, where $\|\widetilde{H}\|_2$ denotes the matrix spectral norm. For any $q \in \mathcal{Q}$, by the fact that C is a projection matrix and Lemma 4.3, we have $\|[(I - 2C)(I - \widetilde{B}^+\widetilde{B})](\mathcal{N}(q) - \mathcal{N}(q_*))\| \leq \max_{j:\|(v_*)_j\|\neq 0} \frac{2}{\|(v_*)_j\|}\|q - q_*\|$, thus by triangle inequality

$$\begin{aligned}\|H_\tau(q) - H_\tau(q_*)\| &\leq \|\widetilde{H}\|_2\|q - q_*\| + \tau\|[(I - 2C)(I - \widetilde{B}^+\widetilde{B})](\mathcal{N}(q) - \mathcal{N}(q_*))\| \\ &\leq \left(\cos\theta_1 + \max_{j:\|(v_*)_j\|\neq 0} \frac{2\tau}{\|(v_*)_j\|}\right)\|q - q_*\|.\end{aligned}$$

Since q_k converges to q_* and q_* is in the interior of $\mathcal{Q}$, there exists K such that for all $k \geq K$, $q_k \in \mathcal{Q}$. Hence, there exists K such that for all $k \geq N$, we have

$$\|H_\tau(q_k) - H_\tau(q_*)\| \leq \left(\cos(\theta_1) + \max_{j:\|(v_*)_j\|\neq 0} \frac{2\tau}{\|(v_*)_j\|}\right)^{k-K}\|q_K - q_*\|.$$

Now we consider the case when $\mathrm{Range}(\widetilde{B}^*) \cap (\mathcal{K}^*)^{-1}\left[\mathrm{Range}(A^*)\right] \neq \{0\}$ for which DRS fixed points are not unique by Lemma 4.2. By Lemma 4.6, there exists K such that $\mathbb{P}_{\mathrm{Range}(\widetilde{B}^*)\cap(\mathcal{K}^*)^{-1}[\mathrm{Range}(A^*)]}(q_k - q_*) = 0, \quad \forall k \geq K$. Since $\widetilde{C}_2$ is the basis for $\mathrm{Range}(\widetilde{B}^*) \cap (\mathcal{K}^*)^{-1}[\mathrm{Range}(A^*)]$, the decomposition in (20) implies $\|\widetilde{H}(q_k - q_*)\| \leq \cos\theta_1\|q_k - q_*\|$ for $k \geq K$, thus

$$\begin{aligned}\|H_\tau(q_k) - H_\tau(q_*)\| &\leq \|\widetilde{H}(q_k - q_*)\| + \tau\|[(I - 2C)(I - \widetilde{B}^+\widetilde{B})](\mathcal{N}(q_k) - \mathcal{N}(q_*))\| \\ &\leq \left(\cos\theta_1 + \max_{j:\|(v_*)_j\|\neq 0} \frac{2\tau}{\|(v_*)_j\|}\right)\|q_k - q_*\|.\end{aligned}$$

This concludes the proof. $\square$

4.6 Remarks on possible further extensions

It is possible to extend the discussion to more general problems and algorithms, but we do not pursue these extensions. The following extensions can be considered:

1. In Theorem 1.2, we only considered the case that $q_* \in \mathcal{Q}$ lies in the interior of $\mathcal{Q}$. For the non-standard cases, iterates q_k converge to a fixed point lying on the boundary of the set $\mathcal{Q}$, and it is possible to have a similar result with a redefined angle θ_1 following the arguments for such non-standard cases in [9]. For whether the converged fixed point is standard or non-standard, it depends on the data (A, b) and initial guess q_0 of the DRS iteration. In our numerical tests, we have not observed non-standard cases.
2. The more general DRS operator can be written as $H_\tau^\lambda = (1-\lambda)\mathbb{I} + \lambda \frac{\mathbb{I} + \mathrm{R}_h^\tau \mathrm{R}_f^\tau}{2}$ with a relaxation parameter $\lambda \in (0, 2)$. Since H_τ^λ is very similar to H_τ, such a discussion is quite straightforward.
3. One can also consider adding regularization [13] to the problem (1). One suitable way of adding regularization is to add ℓ^2 regularization to the equivalent problem (4) with parameter α.
$$\min_{v \in \mathbb{R}^{N \times d}} \|v\|_{1,2} + \iota_{\mathcal{K}\{u: Au=b\}}(v) + \frac{1}{2\alpha}\|v\|^2, \tag{21}$$
where $\|\cdot\|$ is the 2-norm for $\mathbb{R}^{Nd}$. When α is large enough, (21) gives the same minimizer [37]. We refer to [9] for techniques of incorporating the relaxation λ and regularization into analyzing $\widetilde{H}$, which seems possible to be combined with the discussion in the previous subsection.

5 Numerical tests

We report numerical results of implementing Algorithm 4 with step sizes $\sigma = \frac{1}{\tau}$ for solving TVCS problem (1) which is equivalent to ADMM, with step-size $\gamma = \frac{1}{\tau}$ on (2) by the relations in Table 1. We construct TVCS problems using 2D and 3D Shepp-Logan images [34] as well as some 3D MRI data. The 2D tests were performed on a MacBook Air with M1 Chip (8 core) with 16G memory, while the 3D tests were performed on one Nvidia A100 GPU card with 80G memory, implemented in Python with single and double precision. Similar to [27], the Python package JAX was used to achieve a simple implementation on the GPU. Unless stated otherwise, the initial conditions used for all the tests were the given data $u_0 = \mathcal{F}^* M \mathcal{F} u_*$ and $p_0 = 0$, where u^* is the true image (Shepp-Logan or MRI image). The mask matrix M is generated randomly.

5.1 2D Shepp-Logan image

We first study how sharp the estimate in Theorem 1.2 is for small τ. We construct a TVCS testing problems by 2D Shepp-Logan image [34] with 30% frequencies which are chosen randomly with the zeroth frequency included.

5.1.1 Local linear rate validation

Figure 1 (left) shows result $\tau = 0.01$ for 2D image of size 64×64. For computing the angle θ_1, we need the minimizer v^*, to (4), which is approximated by running 10000 iterations of ADMM on (1) and then using Table 1 to transform the ADMM variables into the physical variable for DR on (4). The angle between the subspaces $\mathcal{K}$Kernel(A) and Kernel$(\widetilde{B})$ is then computed by SVD In Figure 1 (left), we observe that $\cos\theta_1$ matches quite well with the actual local linear rate. The estimate in Theorem 1.2 is more conservative, but for $\tau = 0.01$ it still seems a good estimate in practice. On the other hand, the linear convergence regime is not reached until iteration number 4300, and the number of iterations needed to enter the linear convergence regime can be sensitive to τ in practice. A larger τ may give fewer iterations needed to enter the linear convergence regime [25].

5.1.2 The effects of different step size τ

For the same 2D problem, Figure 2 shows that the results for different step sizes ranging from $\tau = 0.01$ to $\tau = 20$, which does not induce a big change in the local linear rate, even though our provable rate does contain τ in the estimate. We remark that the dependence on τ in Theorem 1.2 can be removed in our proof when $d = 1$, i.e., the local linear rate of Algorithm 1 for ℓ^1-norm CS problem does not depend on step size in both analysis and numerical tests [9]. On the other hand, Figure 2 shows that different step sizes significantly affect number of iterations needed to enter the linear convergence regime. As shown in Figure 2, for $\tau = 20.0$, the number of iterations it takes to enter linear convergence regime is $l = 0$, i.e., numerically it seems a global linear convergence.

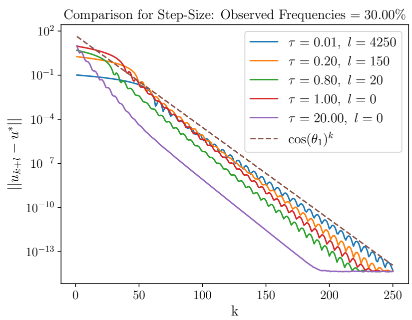


Fig. 2: Algorithm 1 with $\gamma = \frac{1}{\tau}$ for (1) with 30% observed frequencies for a 2D Shepp-Logan image of size 64×64. **Here k is not the iteration number. Instead, $k+l$ is the iteration number where l is the number of iteration needed to enter the linear convergence regime.**

5.2 Effects of regularization and relaxation

Consider a generalized version of ADMM by applying the general DRS operator $H_\tau^\lambda = (1-\lambda)\mathbb{I} + \lambda \frac{\mathbb{I}+\mathrm{R}_h^\tau \mathrm{R}_f^\tau}{2}$ with a relaxation parameter $\lambda \in (0,2)$ to the regularized problem (21) with a regularization parameter α. See Figure 3 for results with different λ and a $\alpha = 100$. For these tests, the 2D Shepp-Logan image is 128×128, the step size is $\gamma = \frac{1}{\tau} = \frac{1}{22}$, and 30% of the frequencies are observed. As proven in [9], special choices of parameters α and λ can speed up the local linear convergence rate for ℓ^1-norm CS problem. Figure 3 shows that this is also the case for TVCS in two dimensions.

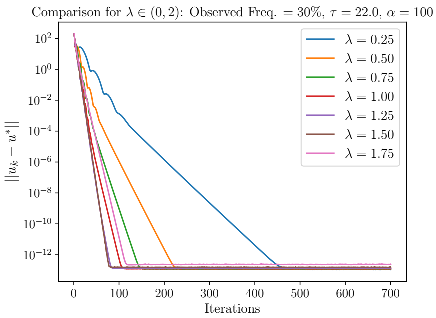


Fig. 3: Local convergence rate of generalized ADMM (corresponding to the general DRS operator H_τ^λ) solving (21) with different parameters λ, $\alpha = 100$. A 2D Shepp-Logan image of size 128×128 with 30% observed frequencies.

5.3 3D Images

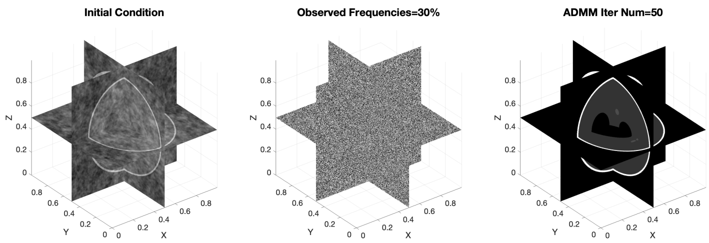


Fig. 4: Left: The initial guess in ADMM. Middle: the non-zero entries of the mask, observed frequencies is 30%. Right: the primal iterate output by ADMM after 50 iterations u_{50} for a 3D Shepp-Logan image of size 512^3.

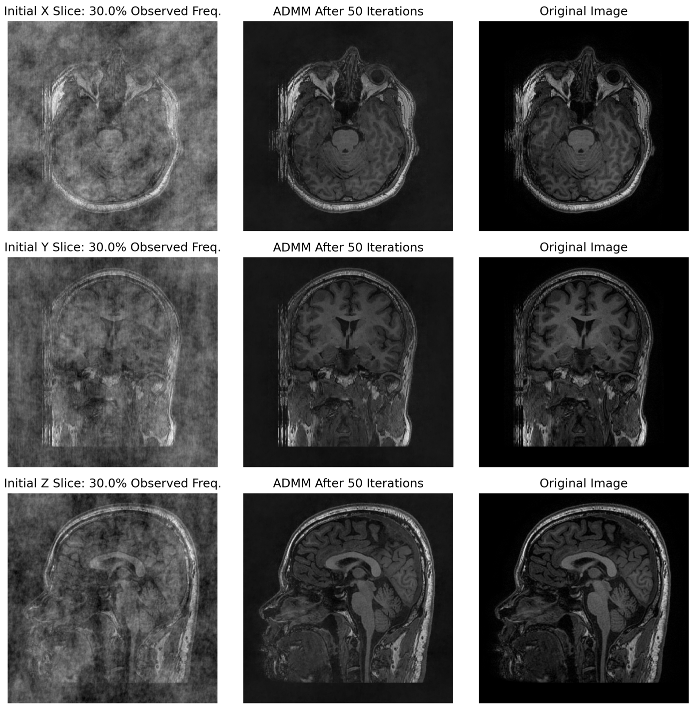


Fig. 5: 3D MRI image of size 512^3. Left: slices of the initial condition for the primal variable in ADMM. Middle: slices of the primal variable of ADMM with $\gamma = \frac{1}{22}$ after 50 iterations. Right: slices of the true MRI image.

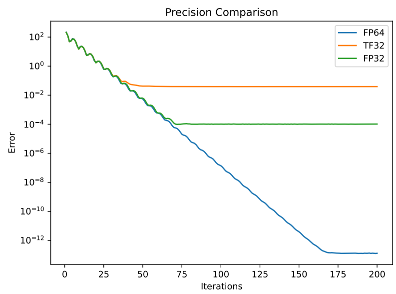


Fig. 6: 3D Shepp-Logan image of size 128^3 with 30% observed frequencies. ADMM with $\gamma = \frac{1}{\tau} = \frac{1}{22}$. Performance of the algorithm using Double Precision vs Single Precision (FP32 and TF32) in Python Jax on Nvidia A100.

Table 2: GPU Time (minutes) vs. relative error $\left(\frac{\|u_k-u_*\|}{\|u_*\|}\right)$ of ADMM (Algorithm 1 with step–size $\gamma = \frac{1}{\tau} = \frac{1}{22}$) solving (1) with 30% observed frequencies for real MRI data of size 512^3. Double precision computation in Python Jax on one Nvidia A100 card with 80G memory.

Iteration Number	1	10	20	80	350
GPU Time (min)	0.02	0.06	0.1	0.33	1.23
Relative Error	6.2×10^{-1}	2.9×10^{-2}	7.2×10^{-3}	8.7×10^{-4}	9.5×10^{-5}

Table 3: Comparison of the computational time (in seconds) of Algorithm 1 with $\gamma = \frac{1}{\tau} = \frac{1}{22}$ to perform 250 iterations of ADMM implemented by Python Jax on one Nvidia A100 80G card: double–precision (FP64) V.S. single–precision (FP32 and TF32). 3D Shepp-Logan of different sizes with 30% observed frequencies. For FP64, memory is not sufficient to compute the problem size 700^3.

Problem Size	128^3	256^3	512^3	700^3
FP64	2.36	7.70	57.14	-
FP32	2.34	4.78	31.14	86.98
TF32	2.30	4.30	23.79	62.22

We test large 3D problems using the 3D Shepp-Logan image as well as some MRI data with 30% observed frequencies. The step size is taken to be $\tau = 0.1$. An estimate of v_* was obtained by running ADMM on 1 for 10,000 iterations and then using the relations in Table 1 to obtain the physical variable of DR v_k. The angle between two subspaces is approximated by the procedure in [9, Appendix B].

First, we consider a 3D Shepp-Logan image of size 512^3, and the performance is shown in Figure 1 (right) and also Figure 4. Next we verify the performance on some MRI image of size 512^3 with 30% frequencies observed. Figure 5 shows that 50 iterations of ADMM with $\gamma = \frac{1}{22}$ produce a result satisfactory to the human eye. Table 2 shows the computational time on GPU, and the reference u^* is the numerical solution after 5000 ADMM iterations.

Finally, we consider single precision computation on GPU, which is sufficient for many imaging purposes. Results in [27] show that single precision computation allows computation of larger problems on one GPU card due to the consumption of less memory. The python package JAX offers two options for single-precision computing with default Float-32 (FP32), and also TensorFloat-32 (TF32), see [27] for technical details. These tests were conducted for 3D Shepp-Logan images with 30% observed frequencies, and ADMM with $\gamma = \frac{1}{\tau} = \frac{1}{22}$. Figure 6 shows shows that single-precision computation does not affect the local linear rate. In Table 3, we see that single-precision computing is not only faster than double-precision (FP64), but it also allows us to compute problems of size 700^3 while double-precision runs out of memory for any problem larger than 512^3 on one Nvidia A100 80G card. Moreover, the difference in speed between double-precision and single-precision is widened as the size of the problem grows larger.

6 Concluding remarks

In this paper, we have provided an asymptotic linear convergence rate of ADMM applied to the Total-Variation Compressed Sensing (TVCS) problem by applying DRS to an equivalent problem. The explicit rate shows the similarities and differences between TVCS and Basis Pursuit. The results were validated with large three-dimensional tests, where a simple but efficient GPU implementation was provided. Among these results, it was shown that the generalized version of ADMM on the regularized TVCS problem has the potential to speed up the convergence rate as in Basis Pursuit. This intuition could shed some light on how to choose parameters for the TVCS problem as well.

Acknowledgements E.T. and X.Z. were supported by NSF grant DMS-2208515, and M.J. was supported by NSF grant DMS-2400641.

Data Availability MRI data was provided with the consent of the individual(s). The source wished to remain unacknowledged, and all data handling complied with applicable privacy and ethical guidelines to maintain confidentiality and respect the source's wishes for anonymity. All other datasets generated during and/or analysed during the current study are available from the authors on reasonable request.

Appendix

A Derivation of dual problems

For $G^* = g^* \circ (-\mathcal{K}^*)$ where $g = \iota_{\{u\in\mathbb{R}^N: Au=b\}}(u)$ as defined in Section 2.3, we derive its convex dual function $G = (G^*)^*$. Let $\mathrm{P}(v)$ denote the projection of $v \in \mathbb{R}^N$ onto the affine set $\{u \in \mathbb{R}^N : Au = b\}$. Recall that $A \in \mathbb{C}^{m\times N}$ defined in Section 2.3 satisfies $AA^* = I$, thus $\mathrm{P}(v) = v + A^*(AA^*)^{-1}(b - Av) = v + A^*(b - Av)$. For any $p \in \mathrm{Range}(A^*)$, let $p = A^*z$, then $z = Ap$ due to the fact $AA^* = I_{m\times m}$. The convex conjugate of $g = \iota_{\{u\in\mathbb{R}^N: Au=b\}}(u)$ is the support function of the affine set, which can be simplified as follows. For $p \in \mathrm{Range}(A^*)$,

$$g^*(p) = \sup_{u:Au=b} \langle u, p\rangle = \sup_{u:Au=b} \langle u, A^*z\rangle = \sup_{u:Au=b} \langle Au, Ap\rangle = \langle b, Ap\rangle = \langle A^*b, p\rangle.$$

Thus $g^*(p) = \begin{cases} \langle p, A^*b\rangle_{\mathbb{R}^N}, & \text{if } p \in \mathrm{Range}(A^*) \\ +\infty, & \text{otherwise} \end{cases}.$

Let $(\mathcal{K}^*)^{-1}\left[\mathrm{Range}(A^*)\right]$ be the pre-image of $\mathrm{Range}(A^*)$ under $\mathcal{K}^*$. By the Lemma above,

$$g^*(-\mathcal{K}^*p) = \begin{cases} -\langle \mathcal{K}^*p, A^*b\rangle & \text{if } \mathcal{K}^*p \in \mathrm{Range}(A^*) \\ +\infty & \text{otherwise.} \end{cases} = -\langle p, \mathcal{K}A^*b\rangle + \iota_{(\mathcal{K}^*)^{-1}\left[\mathrm{Range}(A^*)\right]}(p).$$

Notice that $\langle p, \mathcal{K}A^*b\rangle$ is continuous in p. Since $\mathrm{Range}(A^*)$ is a closed set and K^* is a bounded linear transformation, $(\mathcal{K}^*)^{-1}\left[\mathrm{Range}(A^*)\right]$ is a closed convex set. Since an indicator function of a closed convex set is a closed convex proper function, $G^* = g^* \circ -K^*$ is a closed convex proper function. By the regularity condition $\mathrm{Range}(A^*) \cap \mathrm{Range}(\mathcal{K}^*) \neq \emptyset$, we have

$$-\mathcal{K}\{u : Au = b\} = -\mathcal{K}\partial g^*(-K^* p) = \partial(g^* \circ -\mathcal{K}^*)(p) = \partial G^*(p), \quad \forall p \in (\mathcal{K}^*)^{-1}\big[\mathrm{Range}(A^*)\big].$$

So $G(v) = [G^*]^*(v) = \sup_p [\langle v, p\rangle_{\mathbb{R}^{2N}} - G^*(p)] = \begin{cases} 0 & \text{if } v \in -\mathcal{K}\{u : Au = b\} \\ +\infty & \text{otherwise} \end{cases}$. Define $h(v) := G(-v) = \iota_{\mathcal{K}\{u:Au=b\}}(v)$, then we have derived the formulation (4). Since $G(v)$ is an indicator function, its proximal operator is a projection, which can be written as

$$\mathrm{Prox}^{\gamma}_{G^*}(q) = \mathcal{F}^*\Big[\widetilde{M}^*\widetilde{M} + (I - \widetilde{M}^*\widetilde{M^*})(I - \Lambda(\Lambda^*\Lambda)^+\Lambda^*)\Big]\mathcal{F}(q + \gamma\mathcal{K}A^*b),$$

where $\widetilde{M}$ is defined in (11). Then by Theorem 2.1 (iv), we obtain Prox^{τ}_{h} as (14).

B Equivalence of DRS on primal and dual problems

Consider $(P) \min_x f(x) + g(x)$ and $(D) \min_p f^*(p) + g^*(-p)$ for two closed convex proper functions $f(x)$ and $g(x)$. DRS with a step size $\gamma > 0$ and a relaxation parameter λ for (P) is

$$\begin{cases} s_{k+1} &= s_k - \lambda t_k + \lambda\,\mathrm{Prox}^{\gamma}_{g}(2t_k - s_k), \quad \lambda \in (0,2) \\ t_k &= \mathrm{Prox}^{\gamma}_{f}(s_k) \end{cases}. \tag{22}$$

With the fact $\mathrm{Prox}^{\tau}_{g^*\circ(-\mathbb{I})}(p) = -\mathrm{Prox}^{\tau}_{g^*}(-p)$, DRS with a step size $\frac{1}{\gamma}$ and a relaxation parameter λ for the dual problem can be written as

$$\begin{cases} q_{k+1} &= q_k - \lambda p_k - \lambda\,\mathrm{Prox}^{\frac{1}{\gamma}}_{g^*}(-2p_k + q_k), \quad \lambda \in (0,2) \\ p_k &= \mathrm{Prox}^{\frac{1}{\gamma}}_{f^*}(q_k) \end{cases}. \tag{23}$$

With Moreau Decomposition, (22) is equivalent to (23) via $q_k = \frac{s_k}{\gamma}, p_k = \frac{s_k - t_k}{\gamma}$.

C Proof of equivalence of G-prox PDHG and ADMM

We give the proof of Theorem 3.1 The main tool we will need is the following lemma:

Lemma C.1 *For a closed convex proper function h, $\beta > 0$, and a matrix $\mathcal{K}$,*

$$\hat{p} = \operatorname*{argmin}_p h(p) + \frac{\beta}{2}||\mathcal{K}p - q||^2 \implies \beta(\mathcal{K}\hat{p} - q) = \mathrm{Prox}^{\beta}_{h^*\circ(-\mathcal{K}^*)}(-\beta q).$$

Proof By Theorem 2.1 (iii), we have $0 \in \partial h(\hat{p}) + \beta\mathcal{K}^*(\mathcal{K}\hat{p} - q)$, which holds if and only if $\hat{p} \in \partial h^*\Big(-\beta K^*(K\hat{p} - q)\Big)$. Multiplying both sides by $-\mathcal{K}$, we get $-\mathcal{K}\hat{p} \in -\mathcal{K}\partial h^*\Big[-\beta\mathcal{K}^*(\mathcal{K}\hat{p} - q)\Big]$. Let $y = \beta\Big(\mathcal{K}\hat{p} - q\Big)$ and $g(x) = -\mathcal{K}^*x$. By chain rule, we have

$$-\mathcal{K}\partial h^*\Big[g(y)\Big] = \partial(h^*\circ g)(y) = \partial[h^*\circ(-\mathcal{K}^*)]\Big(\beta[\mathcal{K}\hat{p} - q]\Big) \Rightarrow -\mathcal{K}\hat{p} \in \partial[h^*\circ(-\mathcal{K}^*)]\Big(\beta[\mathcal{K}\hat{p} - q]\Big).$$

By adding $\mathcal{K}\hat{p} - q$ then multiplying β to both sides, we get

$$-\beta q \in \beta(\mathcal{K}\hat{p} - q) + \beta\partial(h^* \circ -\mathcal{K}^*)\Big[\beta(\mathcal{K}\hat{p} - q)\Big] = \Big[I + \beta\partial[h^* \circ (-\mathcal{K}^*)]\Big]\Big(\tau(\mathcal{K}\hat{p} - q)\Big),$$

which implies $\beta(\mathcal{K}\hat{p} - q) = \Big[I + \beta\partial(h^* \circ -\mathcal{K}^*)\Big]^{-1}(-\beta q) = \mathrm{Prox}^{\beta}_{h^*\circ -\mathcal{K}^*}(-\beta q)$. □

The first line of G-prox PDHG with step-size τ in Algorithm 3 can be written as $u_{k+1} = \operatorname{argmin}_u g(u) + \frac{1}{2\tau}||\mathcal{K}u - (\mathcal{K}u_k - \tau w_k)||^2$. Apply Lemma C.1 to the line above with $h = g$, $\widehat{p} = u_{k+1}$, $\beta = \frac{1}{\tau}$, and $q = \mathcal{K}u_k - \tau w_k$, we get $\mathcal{K}u_{k+1} - \mathcal{K}u_k + \tau w_k = \tau \operatorname{Prox}^{\frac{1}{\tau}}_{g^* \circ (-\mathcal{K}^*)}(w_k - \frac{1}{\tau}\mathcal{K}u_k)$. By Moreau Decomposition, the second line of G-prox PDHG with $\tau = \frac{1}{\sigma}$ can be written as

$$v_{k+1} = \operatorname*{argmin}_v f^*(v) + \frac{\tau}{2}||v - (v_k + \mathcal{K}u_{k+1})||^2 = v_k + \frac{1}{\tau}\mathcal{K}u_{k+1} - \frac{1}{\tau}\operatorname{Prox}^{\tau}_f(\tau v_k + \mathcal{K}u_{k+1}).$$

Thus the G-prox PDHG in Algorithm 3 with $\tau = \frac{1}{\sigma}$ gives:

$$\mathcal{K}u_{k+1} - \mathcal{K}u_k + \tau w_k = \tau \operatorname{Prox}^{\frac{1}{\tau}}_{g^* \circ (-\mathcal{K}^*)}(w_k - \frac{1}{\tau}\mathcal{K}u_k) \tag{24a}$$

$$\tau v_{k+1} = \tau v_k + \mathcal{K}u_{k+1} - \operatorname{Prox}^{\tau}_f(\tau v_k + \mathcal{K}u_{k+1}) \tag{24b}$$

$$w_{k+1} = 2v_{k+1} - v_k. \tag{24c}$$

The first line in Algorithm 1 can be written as $x_{k+1} = \operatorname{argmin}_x \; g(x) + \frac{\gamma}{2}||\mathcal{K}x - (y_k - \frac{1}{\gamma}z_k)||^2$. By Lemma C.1 with $h = g$, $\beta = \gamma$, and $\widehat{p} = y_k - \frac{1}{\gamma}z_k$, we get

$$-\gamma\mathcal{K}x_{k+1} - (z_k - \gamma y_k) = \operatorname{Prox}^{\gamma}_{g^* \circ \mathcal{K}^*}\left[\gamma y_k - z_k\right] \Longleftrightarrow \gamma\mathcal{K}x_{k+1} + (z_k - \gamma y_k) = \operatorname{Prox}^{\gamma}_{g^* \circ (-\mathcal{K}^*)}\left[z_k - \gamma y_k\right].$$

By the definition of the proximal operator, the second line of in Algorithm 1 reduces to

$$y_{k+1} = \operatorname*{argmin}_y f(y) - \langle y, z_k\rangle + \frac{\gamma}{2}||y - \mathcal{K}x_{k+1}||^2 = \operatorname{Prox}^{\frac{1}{\gamma}}_f\left[\frac{1}{\gamma}(z_k + \gamma\mathcal{K}x_{k+1})\right].$$

Thus the ADMM in Algorithm 1 is equivalent to

$$\gamma\mathcal{K}x_{k+1} + (z_k - \gamma y_k) = \operatorname{Prox}^{\gamma}_{g^* \circ -\mathcal{K}^*}\left[z_k - \gamma y_k\right] \tag{25a}$$

$$y_{k+1} = \operatorname{Prox}^{\frac{1}{\gamma}}_f\left[\frac{1}{\gamma}(z_k + \gamma\mathcal{K}x_{k+1})\right] \tag{25b}$$

$$z_{k+1} = z_k - \gamma(y_{k+1} - \mathcal{K}x_{k+1}). \tag{25c}$$

Finally, we prove the equivalence between (24) and (25). Define the following variables,

$$\tau := \frac{1}{\gamma}, \quad v_k := z_k, \quad u_k := x_k, \quad \tau w_k := \mathcal{K}x_k + \tau z_k - y_k,$$

then (25a) becomes (24a) by

$$\begin{aligned}
&\frac{1}{\tau}\mathcal{K}x_{k+1} + (z_k - \frac{1}{\tau}y_k) = \operatorname{Prox}^{\frac{1}{\tau}}_{g^* \circ (-\mathcal{K}^*)}\left[z_k - \frac{1}{\tau}y_k\right] \\
\Longleftrightarrow\; &\mathcal{K}x_{k+1} + \tau z_k - y_k = \tau \operatorname{Prox}^{\frac{1}{\tau}}_{g^* \circ (-\mathcal{K}^*)}\left[z_k - \frac{1}{\tau}y_k\right] \\
\Longleftrightarrow\; &\mathcal{K}u_{k+1} + \tau v_k - \left[\mathcal{K}u_k + \tau(v_k - w_k)\right] = \tau \operatorname{Prox}^{\frac{1}{\tau}}_{g^* \circ (-\mathcal{K}^*)}\left[v_k - \frac{1}{\tau}\big(\mathcal{K}u_k + \tau(v_k - w_k)\big)\right] \\
\Longleftrightarrow\; &\mathcal{K}(u_{k+1} - u_k) + \tau w_k = \tau \operatorname{Prox}^{\frac{1}{\tau}}_{g^* \circ (-\mathcal{K}^*)}\left[w_k - \frac{1}{\tau}\mathcal{K}u_k\right],
\end{aligned}$$

(25b) becomes (24b) by

$$y_{k+1} = \operatorname{Prox}^{\frac{1}{\gamma}}_f\left[\frac{1}{\gamma}(z_k + \gamma\mathcal{K}x_{k+1})\right] \iff \mathcal{K}u_{k+1} + \tau(v_{k+1} - w_{k+1}) = \operatorname{Prox}^{\tau}_f\left[\tau v_k + \mathcal{K}u_{k+1}\right],$$

and (25c) becomes (24c) by

$$z_{k+1} = z_k - \frac{1}{\tau}(y_{k+1} - \mathcal{K}x_{k+1}) \iff v_{k+1} = v_k + (w_{k+1} - v_{k+1}) \iff w_{k+1} = 2v_{k+1} - v_k.$$

D Derivation of the explicit implementation formula

With $g(u) = \iota_{\{u\in\mathbb{R}^N:\widehat{u}(k)=b_k,k\in S\}}(u)$, $\quad f^*(v) = \iota_{\{v\in[\mathbb{R}^N]^d:||v||_{\infty,2}\leq 1\}}(v)$, we reformulate (1) into (2), then we apply G-prox PDHG to (2) to obtain:

$$u_{n+1} = \underset{\{u\in\mathbb{R}^N:\widehat{u}(k)=b_k,k\in S\}}{\operatorname{argmin}} \langle \mathcal{K}u, w_n\rangle + \frac{1}{2\tau}||\mathcal{K}(u-u_n)||^2 \tag{26a}$$

$$v_{n+1} = \underset{\{v\in[\mathbb{R}^N]^d:||v||_{\infty,2}\leq 1\}}{\operatorname{argmax}} \langle \mathcal{K}u_{n+1}, v\rangle - \frac{1}{2\sigma}||v-v_n||^2 \tag{26b}$$

$$w_{n+1} = 2v_{n+1} - v_n. \tag{26c}$$

From now on, we focus on the two-dimensional problem and the extension to higher dimensions is straightforward. We first derive an explicit formula of (26a). With the notation in Section 2, let $\mathcal{F}u$ and $\widehat{u}$ be the normalized discrete Fourier transform, i.e., $\mathcal{F}u = \widehat{u}$ and $\langle u, v\rangle_{\mathbb{R}^N} = \langle \mathcal{F}u, \mathcal{F}v\rangle_{\mathbb{C}^N}$. Notice that the matrix K is circulant thus diagonalizable by the 1D normalized DFT matrix T, which implies that the discrete gradient matrix $\mathcal{K}$ and the 2D DFT matrix $\mathcal{F}$ commute. Regard $\mathcal{F}$ as an $N\times N$ matrix, then with (9), we get

$$\begin{aligned}
&\underset{\{u\in\mathbb{R}^N:\widehat{u}(k)=b_k,k\in S\}}{\operatorname{argmin}} \langle \mathcal{K}u, w_n\rangle_{\mathbb{R}^{2N}} + \frac{1}{2\tau}||\mathcal{K}(u-u_n)||^2_{\mathbb{R}^{2N}}\\
=&\underset{\{u:\widehat{u}(k)=b_k,k\in S\}}{\operatorname{argmin}} \langle \begin{pmatrix}\mathcal{F} & 0\\ 0 & \mathcal{F}\end{pmatrix}\mathcal{K}u, \begin{pmatrix}\mathcal{F} & 0\\ 0 & \mathcal{F}\end{pmatrix} w_n\rangle_{\mathbb{C}^{2N}} + \frac{1}{2\tau}|| \begin{pmatrix}\mathcal{F} & 0\\ 0 & \mathcal{F}\end{pmatrix}\mathcal{K}(u-u_n)||^2_{\mathbb{C}^{2N}}\\
=&\underset{\{u:\widehat{u}(k)=b_k,k\in S\}}{\operatorname{argmin}} \langle u, \mathcal{F}^*\boldsymbol{\Lambda}^*\begin{pmatrix}\mathcal{F} & 0\\ 0 & \mathcal{F}\end{pmatrix} w_n\rangle_{\mathbb{C}^{2N}} + \frac{1}{2\tau}||\boldsymbol{\Lambda}\mathcal{F}(u-u_n)||^2_{\mathbb{C}^{2N}}
\end{aligned}$$

Let $\bar{v}$ denote the complex conjugate of v. Since both $\mathcal{F}^*\boldsymbol{\Lambda}^*\begin{pmatrix}\mathcal{F} & 0\\ 0 & \mathcal{F}\end{pmatrix} = \mathcal{K}^*$ and $\mathcal{F}^*\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}\mathcal{F} = \mathcal{K}^*\mathcal{K}$ are real-valued matrices, by taking the derivative with respect to $u\in\mathbb{R}^N$, we get $\tau\mathcal{F}^*\boldsymbol{\Lambda}^*\begin{pmatrix}\mathcal{F} & 0\\ 0 & \mathcal{F}\end{pmatrix} w_n + \mathcal{F}^*\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}\mathcal{F}(u_{n+1}-u_n) = 0$. For $w\in\mathbb{R}^{2N}$, let $w = \begin{pmatrix}w^1\\ w^2\end{pmatrix}$ with $w^1, w^2\in\mathbb{R}^N$. With the notation $\mathcal{F}w = \widehat{w}$, we have $\mathcal{F}^*\boldsymbol{\Lambda}^*\begin{pmatrix}\mathcal{F} & 0\\ 0 & \mathcal{F}\end{pmatrix} w = \mathcal{F}^*\boldsymbol{\Lambda}^*\begin{pmatrix}\widehat{w}^1\\ \widehat{w}^2\end{pmatrix} = \mathcal{F}^*\begin{pmatrix}(\Lambda\otimes I)\widehat{w}^1\\ (I\otimes\Lambda)\widehat{w}^2\end{pmatrix}$. Let λ^1_k $(k=1,\cdots,N)$ be the diagonal entries of $\Lambda\otimes I$ and λ^2_k $(k=1,\cdots,N)$ be the diagonal entries of $I\otimes\Lambda$, we obtain the update rule in Fourier domain:

$$\begin{aligned}
&\widehat{u_{n+1}}(k) = b_k, && k\in S\\
&\widehat{u_{n+1}}(k) = \widehat{u_n}(k) - \tau\frac{\overline{\lambda^1_k}\widehat{w^1_n}(k) + \overline{\lambda^2_k}\widehat{w^2_n}(k)}{|\lambda^1_k|^2 + |\lambda^2_k|^2}, && k\notin S
\end{aligned}$$

Since (26b) can be rewritten as $v_{n+1} = \operatorname{argmin}_{\{v\in[\mathbb{R}^N]^d:||v||_{\infty,2}\leq 1\}} ||v-(v_n+\sigma\mathcal{K}u_{n+1})||^2$, (26b) can be implemented as the projection of $v_k+\sigma\mathcal{K}u_{k+1}$ onto the $\|\cdot\|_{\infty,2}$ ball:

$$v_{k+1} = \text{Projection}_{\{v:\|v\|_{\infty,2}\leq 1\}}(v_k+\sigma\mathcal{K}u_{k+1}) = \frac{v_k+\sigma\mathcal{K}u_{k+1}}{\max(1, |v_k+\sigma\mathcal{K}u_{k+1}|)}.$$

References

1. Timo Aspelmeier, C Charitha, and D Russell Luke. Local linear convergence of the ADMM/Douglas–Rachford algorithms without strong convexity and application to statistical imaging. *SIAM Journal on Imaging Sciences*, 9(2):842–868, 2016.

2. Heinz H Bauschke, JY Bello Cruz, Tran TA Nghia, Hung M Phan, and Xianfu Wang. The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle. *Journal of Approximation Theory*, 185:63–79, 2014.
3. Amir Beck. *First-order methods in optimization.* SIAM, 2017.
4. Åke Björck and Gene H Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of computation*, 27(123):579–594, 1973.
5. Daniel Boley. Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. *SIAM Journal on Optimization*, 23(4):2183–2207, 2013.
6. Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.
7. Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40:120–145, 2011.
8. Damek Davis and Wotao Yin. Convergence rate analysis of several splitting schemes. *Splitting methods in communication, imaging, science, and engineering*, pages 115–163, 2016.
9. Laurent Demanet and Xiangxiong Zhang. Eventual linear convergence of the douglas-rachford iteration for basis pursuit. *Mathematics of Computation*, 85(297):209–238, 2016.
10. Ernie Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. *CAM report*, 9:31, 2009.
11. Ernie Esser, Xiaoqun Zhang, and Tony F Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.
12. Michel Fortin and Roland Glowinski. *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems.* Elsevier, 2000.
13. Michael P Friedlander and Paul Tseng. Exact regularization of convex programs. *SIAM Journal on Optimization*, 18(4):1326–1350, 2008.
14. J-J Fuchs. On sparse representations in arbitrary redundant bases. *IEEE transactions on Information theory*, 50(6):1341–1344, 2004.
15. D Gabay. Applications of the method of multipliers to variational inequalities, augmented lagrangian methods, 1983.
16. Daniel Gabay. *Methodes numeriques pour l'optimisation non-lineaire.* PhD thesis, 1979.
17. Pontus Giselsson and Stephen Boyd. Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. *IEEE Transactions on Automatic Control*, 62(2):532–544, 2016.
18. Roland Glowinski and Patrick Le Tallec. *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics.* SIAM, 1989.
19. Tom Goldstein and Stanley Osher. The split Bregman method for L1-regularized problems. *SIAM journal on imaging sciences*, 2(2):323–343, 2009.
20. Matt Jacobs, Flavien Léger, Wuchen Li, and Stanley Osher. Solving large-scale optimization problems with a convergence rate independent of grid size. *SIAM Journal on Numerical Analysis*, 57(3):1100–1123, 2019.
21. Ville Kolehmainen, Samuli Siltanen, Seppo Järvenpää, Jari P Kaipio, P Koistinen, M Lassas, J Pirttilä, and E Somersalo. Statistical inversion for medical x-ray tomography with few radiographs: Ii. application to dental radiology. *Physics in Medicine & Biology*, 48(10):1465, 2003.
22. Rowan Leary, Zineb Saghi, Paul A Midgley, and Daniel J Holland. Compressed sensing electron tomography. *Ultramicroscopy*, 131:70–91, 2013.
23. Adrian S Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2002.
24. Meihua Li, Haiquan Yang, and Hiroyuki Kudo. An accurate iterative reconstruction algorithm for sparse objects: application to 3d blood vessel reconstruction from a limited number of projections. *Physics in Medicine & Biology*, 47(15):2599, 2002.
25. Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Local convergence properties of Douglas–Rachford and alternating direction method of multipliers. *Journal of Optimization Theory and Applications*, 172:874–913, 2017.

26. Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.
27. Xinyu Liu, Jie Shen, and Xiangxiong Zhang. A Simple GPU Implementation of Spectral-Element Methods for Solving 3D Poisson Type Equations on Rectangular Domains and Its Applications. *Communications in Computational Physics*, 36(5):1157–1185, 2024.
28. Michael Lustig, David Donoho, and John M Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.
29. M Persson, D Bone, and Hakan Elmqvist. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine & Biology*, 46(3):853, 2001.
30. Clarice Poon. On the role of total variation in compressed sensing. *SIAM Journal on Imaging Sciences*, 8(1):682–720, 2015.
31. Ralph Tyrell Rockafellar. *Convex analysis.* Princeton university press, 1970.
32. Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.
33. Fadil Santosa and William W Symes. Linear inversion of band-limited reflection seismograms. *SIAM journal on scientific and statistical computing*, 7(4):1307–1330, 1986.
34. Lawrence A Shepp and Benjamin F Logan. The Fourier reconstruction of a head section. *IEEE Transactions on nuclear science*, 21(3):21–43, 1974.
35. Julia Velikina, Shuai Leng, and Guang-Hong Chen. Limited view angle tomographic image reconstruction via total variation minimization. In *Medical imaging 2007: physics of medical imaging*, volume 6510, pages 709–720. SPIE, 2007.
36. Yves Wiaux, Laurent Jacques, Gilles Puy, Anna MM Scaife, and Pierre Vandergheynst. Compressed sensing imaging techniques for radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 395(3):1733–1742, 2009.
37. Wotao Yin. Analysis and generalizations of the linearized Bregman method. *SIAM Journal on Imaging Sciences*, 3(4):856–877, 2010.
38. Hui Zhang, Wotao Yin, and Lizhi Cheng. Necessary and sufficient conditions of solution uniqueness in 1-norm minimization. *Journal of Optimization Theory and Applications*, 164:109–122, 2015.