Accelerating Convergence of Proximal Methods for Compressed Sensing using Polynomials with Application to MRI

Siddharth Srinivasan Iyer, Frank Ong, Xiaozhi Cao, Congyu Liao, Luca Daniel, Jonathan I. Tamir*, and Kawin Setsompop*

Abstract—This work accelerates the convergence of iterative proximal methods used to solve linear inverse problems that arise from compressed sensing applications. This is achieved by designing a polynomial-based preconditioner that target the eigenvalue spectrum of the normal operator derived from the linear measurement matrix in a manner that does not assume any explicit structure. The resulting preconditioner can thus be deployed in diverse applications of interest. The efficacy of the preconditioner is validated on four different MRI applications, where it is seen to achieve faster convergence while achieving similar reconstruction quality.

Index Terms—compressed sensing, linear inverse problems, iterative reconstruction, polynomial preconditioner, proximal gradient descent, MRI

I. INTRODUCTION

Linear inverse problems derived from compressed sensing formulations are typically posed as the following optimization problem [1], [2]:

$$x^\star = \arg\min_x \{ g(x) : Ax = b \} \tag{1}$$

Here, $A : \mathbb{C}^n \rightarrow \mathbb{C}^m$ is the linear forward model or the measurement matrix, $b \in \mathbb{C}^m$ is the acquired data or measurement vector and $g : \mathbb{C}^n \rightarrow \mathbb{R}$ is a prior regularization function, often the $l_1-$norm, to enforce sparsity. In practice, due to measurement errors (such as noise), the following optimization problem is solved instead [1], [3]–[6]:

$$x^\star = \arg\min_x \left\{ \frac{1}{2} \| Ax - b \|_2^2 + \lambda g(x) \right\} \tag{2}$$

The Karush-Kuhn-Tucker (KKT) conditions can be used to verify that for any $\epsilon$, there is an appropriate choice of $\lambda$ such that the solutions of (2) and (3) coincide.

Iterative proximal methods [8] have emerged as the workhorse algorithms to solve linear inverse problems that are posed in the form of (2) or (3). In particular, these algorithms are heavily used for compressed sensing MRI [3]–[6], [9], where matrix-free implementations of the forward-model $A$ are leveraged for computationally efficient processing of high-dimensional problems. In practice, MRI reconstructions are typically posed as in (3) and are solved using the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [10] which enjoys theoretically optimal convergence guarantees. However, the convergence of FISTA and other iterative proximal methods is largely limited by the conditioning of $A$ as determined by the eigenvalues of $A^*A$, where $A^*$ is the adjoint of $A$.

This is particularly true for Non-Cartesian MRI acquisition methods, where the corresponding measurement operator $A$ can be highly ill-conditioned, resulting in slow iterative convergence and hence long reconstruction times. To accelerate said convergence, Density Compensation [11] and preconditioning methods [12]–[15] are often used, but as discussed in [16], these methods often result in worse reconstruction accuracy and per-iteration computational complexity, respectively. To overcome both these problems, the work in [16] presented a Frobenius-norm-optimized diagonal preconditioner for the dual variables of the primal-dual hybrid gradient (PDHG) algorithm [17]. This reduced reconstruction times for non-Cartesian MRI while not compromising on reconstruction accuracy. However, the use of a primal-dual style algorithm requires storing one or more instances of the dual variable (of
the same size as $b$), which is often significantly larger and computationally slower to compute compared to algorithms which require the primal variable alone. For this reason, the preconditioner in [16] is difficult to calculate and employ in applications with high-dimensional measurements.

This is particularly true for recent contrast-resolved MRI methods such as Echo Planar Time Resolved Imaging [18] and Magnetic Resonance Fingerprinting (MRF) [19], where the large measurement sizes along with the inherent ill-conditioning of the corresponding $A$ operator result in long reconstruction times, which in turn limit the translation of such methods into clinical practice. Additionally, PDHG cannot leverage more efficient implementations of the normal operator $(A^*A)$ that arise in specific applications to reduce the per-iteration time. For example, non-Cartesian reconstructions can leverage the Toeplitz structure of the normal operator of the non-uniform Fourier transform to avoid expensive gridding operations [20]. Temporal subspace methods such as $T_2$-Shuffling can use the “spatio-temporal” kernel to avoid expanding into the “echo” dimension at each iteration to significantly reduce the number of Fast Fourier Transforms need [21].

Lastly, the Frobenius norm formulation used in [16] and circulant preconditioner design proposed in [15] do not generalize well to an arbitrary $A$. In particular, the former requires that the preconditioner is re-calculated each time new data ($b$ in (3)) are acquired, and the latter explicitly utilizes circulant properties inherent in non-Cartesian acquisitions.

This work aims to alleviate these issues by designing a polynomial-based preconditioner for faster convergence of (2) and (3). The evolution of the iterates in proximal gradient descent (PGD) [8] is analyzed. A cost function for polynomial optimization is derived such that the optimized polynomial yields a preconditioner that directly improves the convergence rate of PGD. The derived preconditioner does not assume any structure on $A$ and instead relies solely on calculating matrix-vector products of $A^*A$, allowing the preconditioner to be directly used with algorithms like PGD/FISTA without dual variables. In other words, explicit knowledge of the entries of $A$ is not required. By only requiring $A^*A$ matrix-vector products (and not the values themselves), the preconditioner is highly generalize-able and can be efficiently applied to various linear inverse problems of interest. Once the polynomial is calculated, it can be applied to any arbitrary $A$ as long as the maximum eigenvalue of $A^*A$ can be estimated. While the proposed preconditioner does increase the per-iteration computational cost, the inclusion of the preconditioner results in significantly fewer iterations required to converge for overall faster processing times.

Similarly to Density Compensation, the proposed preconditioner re-weights the least squares objective in (3). However, instead of weighting the coordinates of $b$ directly, the polynomial preconditioner in PGD amplifies the contribution to the overall objective of the components of the iterands that correspond to the small singular values of $A$. Uniquely, in the absence of regularization, the polynomial-preconditioned reconstruction matches the non-preconditioned result, which is not necessarily true for Density Compensation.

The cost function for polynomial optimization mentioned above has a natural connection to [22], where a similar objective function was proposed as a means of accelerating the convergence of the Conjugate Gradient Algorithm. However, [22] focuses explicitly on ordinary least squares without regularization and motivates the polynomial design as a means of approximating the inverse of $A^*A$. This work arrives at the same objective function for polynomial design, but from a completely different perspective. Specifically, the objective function is derived from analyzing error propagation of PGD iterates when solving ill-posed problems that leverage prior regularization.

The structure of the paper is as follows: Section II derives the preconditioner from PGD and analyzes its application to (3). Section III demonstrates how the polynomial preconditioner can be applied to ADMM [8] with multiple regularizations. In Section IV, the efficacy of the preconditioner is validated on four varied MRI applications, where it is seen to achieve faster convergence at comparable image quality.

II. POLYNOMIAL PRECONDITIONED COMPRESSED SENSING

A. Formulation

To motivate the use of the polynomial preconditioner, first consider the exact formulation (1) with $b \in \text{range}(A)$. Without loss of generality, let the induced norm of $A$ be unitary, which then implies the eigenvalues of $A^*A$ and $AA^*$ lie in the interval $[0, 1]$.

The singular value decomposition of $A$ in dyadic form is as follows:

$$ A(\cdot) = \sum_{i=1}^{j} \sigma_i(\cdot, v_i)u_i $$

(4)

Here, $j \leq n \geq \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_j > 0$ and $\langle \cdot, \cdot \rangle$ denotes the standard Euclidean complex inner product. Consequently,

$$ A^*A(\cdot) = \sum_{i=1}^{j} \sigma_i^2(\cdot, v_i)v_i $$

$$ AA^*(\cdot) = \sum_{i=1}^{j} \sigma_i^2(\cdot, u_i)u_i $$

(5)

Let $p(z)$ be a polynomial of degree $d$ such that $p(z) > 0$ for $z \in (0, 1]$, and let $P = p(AA^*)$. It follows that

$$ P(\cdot) = \sum_{i=1}^{j} p(\sigma_i^2)(\cdot, u_i)u_i $$

(6)

As an ansatz, let $P^{\frac{1}{2}}$ be defined as the square root operator of $P$:

$$ P^{\frac{1}{2}}(\cdot) = \sum_{i=1}^{j} \left[p(\sigma_i^2)\right]^{\frac{1}{2}}(\cdot, u_i)u_i $$

(7)

Note that $P$ and $P^{\frac{1}{2}}$ are self-adjoint. The condition $p(z) > 0$ for $z \in (0, 1]$ implies $P$ and $P^{\frac{1}{2}}$ are injective when the domain and co-domain for both operators are restricted to range$(A)$. Thus, the condition $Ax = b$ in (1) is equivalently enforced
by the constraint $P^\perp Ax = P^\perp b$, yielding the following “preconditioned” compressed sensing formulation.

$$x^* = \begin{cases} \arg\min_x g(x) \\ \text{subject to } P^\perp Ax = P^\perp b \end{cases}$$ (8)

Note that the injectivity of $P^\perp$ implies that the solutions to (1) and (8) are identical.

To account for model and measurement errors, (8) can be relaxed to the following constrained formulation:

$$x^* = \begin{cases} \arg\min_x \frac{1}{2}||P^\perp (Ax - b)||^2 + \lambda_p g(x) \\ \text{subject to } ||P^\perp (Ax - b)||_2 \leq \epsilon_p \end{cases}$$ (9)

The sequel will focus on the unconstrained formulations (3) and (10) as the following experiments (Section IV) leverage the smoothness within the respective objective functions for faster convergence via Nesterov acceleration [10], [23].

At first glance, the inclusion of $P^\perp$ into (10) results in an objective function with a weighted least squares term that is similar to density compensation methods [16]. However, instead of using a diagonal matrix to weight to the entries (or coordinates) of the measurement $b$, $P^{1/2}$ “spectrally” weights $b$. In other words, $P^{1/2}$ is a diagonal matrix with respect to the left singular vectors of $A$ (or the eigen-basis of $AA^*$).

Interestingly, the spectral interpretation of (10) shows that when $\lambda = \lambda_p = 0$, the minimum $l_2$-norm solutions to (3) and (10) are identical with the assumption $p(z) > 0$ for $z \in (0, 1]$ as shown below. This is not necessarily true for density compensation methods.

$$x \text{ solves (3)} \iff \sigma_i(x, v_i) = \langle b, u_i \rangle$$

$$\iff p(\sigma_i^2)^2 \sigma_i \langle x, v_i \rangle = p(\sigma_i^2)^2 \langle b, u_i \rangle$$ (11)

$$\iff x \text{ solves (10)}$$

When $\lambda = \lambda_p$, Subsection II-D derives an error bound between the solutions of (3) and (10) under the assumptions that $A^*A$ is injective, and demonstrates that this error is small for applications of interest. In practice, since $\epsilon \neq \epsilon_p$, $\lambda_p$ will need to be tuned for applications of interest.

B. Algorithm and Convergence Analysis

For simplicity, the following analysis focuses on traditional PGD. That being said, the improved conditioning of (10) naturally translates into faster convergence compared to (3) when using FISTA [10] for both problems.

Let $\text{prox}_g$ be the proximal operator of $g$ as defined in [8]:

$$\text{prox}_{\alpha g}(v) = \arg\min_x \frac{1}{2}||v - x||^2 + \alpha g(x)$$ (12)

The PGD iterations to solve (3) are as below with $k$ denoting the iteration number. Additionally, note that the optimal solution $x^*$ is the fixed point of the iterations [8].

$$x_{k+1} = \text{prox}_{\lambda g}(x_k - A^* (Ax_k - b))$$

$$x^* = \text{prox}_{\lambda g}(x^* - A^* (Ax^* - b))$$ (13)

Let $e_k = x^* - x_k$. Subtracting the two equations of (13), taking the $l_2$-norm and utilizing the firm non-expansiveness property of proximal operators [8] yields:

$$||e_{k+1}||^2 \leq ||(I - A^*A)e_k||^2$$ (14)

Splitting $e_k$ into $s_k + t_k$ where $s_k \in \text{null}(A)$ and $t_k \in \text{null}(A)^\perp$ results in:

$$||e_{k+1}||^2 \leq ||(I - A^*A)t_k||^2 + ||s_k||^2$$ (15)

The decrease in error of $e_k$ within $\text{null}(A)^\perp$ (i.e. $t_k$) is determined by the eigenvalue spectrum of $I - A^*A$. Thus, inverse problems involving an $A^*A$ with small eigenvalues will suffer from slow iterative convergence.

The PGD iterations to solve (10) are as below:

$$y_{k+1} = \text{prox}_{\lambda p g}(y_k - A_P^* (A_P y_k - b_P))$$

$$y^* = \text{prox}_{\lambda p g}(y^* - A_P^* (A_P y^* - b_P))$$ (16)

where $A_P = P^\perp A$ and $b_P = P^\perp b$

The iteration variable name has been changed to reflect which algorithm is being used.

As $p$ is a spectral function of $A$, $p(AA^*)$ and $p(A^*A)$ can easily be evaluated with respect to the singular value decomposition of $A$ as defined in (4) and (5). This can be used to show that:

$$A^*p(AA^*)A = p(A^*A) A^* A$$

$$A^* p(A^*A) = p(A^*A) A^*$$ (17)

Incorporating the above into (16) significantly simplifies the iterations. Redefining $P = p(A^*A)$ yields the following:

$$y_{k+1} = \text{prox}_{\lambda p g}(y_k - PA_P^* (A_P y_k - b_P))$$

$$y^* = \text{prox}_{\lambda p g}(y^* - PA_P^* (A_P y^* - b_P))$$ (18)

By utilizing the permute-ability provided by polynomials and linear operators, the polynomial preconditioner can be applied to the iterates directly instead of in the range space of $A$, thus significantly reducing the computational requirements (assuming $n \ll m$, which is often the case in MRI). Additionally, it enables the use of the faster $A^*A$ evaluations, if applicable.

Redefining $e_k = y_k - y^*$, splitting $e_k$ into $s_k + t_k$ where $s_k \in \text{null}(A)$ and $t_k \in \text{null}(A)^\perp$, and applying a similar process to how (15) was derived to (18) shows:

$$||e_{k+1}||^2 \leq ||(I - PA_P^*A)t_k||^2 + ||s_k||^2$$

with $P = p(A^*A)$ (19)

This motivates finding a polynomial $p$ such that $I - PA_P^*A$, with $P = p(A^*A)$, is as close to zero as possible.
C. Polynomial Optimization

As the dimensions of $A$ are typically very large, it is not computationally feasible (in terms of processing time) to perform an eigenvalue decomposition of $A^*A$ to use as prior information for polynomial design. To avoid this, the coefficients of the polynomial $p$ can be found by optimizing the following continuous approximation of the induced norm:

$$ p = \arg\min_q \int_{z=0}^{1} (1 - q(z))z^2dz. \quad (20) $$

The polynomial that minimizes the above cost function in-turn optimizes the induced $l_2$-norm of $I - PA^*A$ in (19), where $P = p(A^*A)$. In other words, the component $t_k$ in $v_i$ (as in (19) and (4)) is upper-bounded by $|1 - p(\sigma_j^2)|$, which according to (20), is minimized to be close to zero. Note that the larger the degree $d$ of polynomial $p$, the better $(1 - p(x))$ can approximate the zero function.

Additionally, priors on the spectrum can be easily incorporated into (20) as follows:

$$ p = \arg\min_q \int_{z=0}^{1} w(z)(1 - q(z))z^2dz \quad (21) $$

Here, $w$ can weight the cost to prioritize certain components of the spectrum.

At first glance, minimizing the following objective instead of (20) is preferable as it directly translates into minimizing the appropriate induced norm of (19):

$$ p = \arg\min_q \max_{z \in [0,1]} |1 - q(z)z| \quad (22) $$

It is well known that Chebyshev polynomials of the first kind can be derived to derive the optimal polynomials such that the maximum absolute value of that polynomial over a specified interval is minimized. However, defining $r(z) = 1 - q(z)z$, and using Chebyshev polynomials to determine $r$ yields a polynomial with $r(z) = 1$ for multiple values of $z \in [0,1]$ due to the constraint $r(0) = 1$, which implies the components of $t_k$ in (19) corresponding to eigenvalues $\sigma^2$ such that $r(\sigma^2) = 1$ will not decrease. However, if the minimum non-zero eigenvalue $\mu = \sigma_j^2$ of $A^*A$ is known a-priori, the following polynomial minimizes (22) over the interval $[\mu, 1]$ (See [22], [24]–[26]):

$$ r(z) = \frac{\frac{1+\mu-2z}{1-\mu}}{T_{d+1} \left( \frac{1+\mu}{1-\mu} \right)} \quad (23) $$

$$ p(z) = \frac{1 - r(z)}{z} $$

Here, $T_{d+1}$ is the Chebyshev polynomial of the first kind of degree $d+1$.

In practice, (20) is preferred as it is often computationally expensive to estimate $\mu$ unless $A^*A$ happens to be injective, in which case $\mu$ can be estimated by performing power-iteration on $I - A^*A$.

Rather interestingly, while (20) and (22) were motivated by studying the evolution of iterates in PGD when solving regularized optimization problems in the form of (10), the exact formulation for polynomial optimization was studied in [22] as a means of accelerating the convergence of Conjugate Gradient for ordinary least squares optimization, where (20) and (22) were optimized to construct an incomplete inverse of $(A^*A)$ to use a preconditioner. Theorem 4 of [22] additionally proves that, for the optimized polynomial $p$, $p(z) > 0$ for $z \in (0,1]$, which implies the optimized polynomial $p$ satisfies the assumption $p(z) > 0$ for $z \in (0,1]$ as desired.

D. Error Bound

With the simplifying assumptions that $A^*A$ is injective and $\lambda_p = \lambda$, it is possible to bound the difference between the solutions of (3) and (10). Let $x^*$ be the optimal solution to (3) and let $y_k$ be the iterations of (18).

$$ y_{k+1} = \text{prox}_{\lambda g}(y_k - PA^*(Ay_k - b)) $$

$$ x^* = \text{prox}_{\lambda g}(x^* - A(Ax^* - b)) \quad (24) $$

Define $e_k = y_k - x^*$. Subtracting the two above equations, taking the $l_2$-norm of the difference and utilizing the firm non-expansiveness property of proximal operators [8] with the triangle inequality yields:

$$ ||e_{k+1}||_2 \leq ||(I - PA^*A)e_k||_2 + ||(I - P)A^*(Ax^* - b)||_2 \quad (25) $$

Let $\gamma$ be the $l_2$-induced norm of $I - PA^*A$ with $\delta$ as defined:

$$ \delta = ||(I - P)A^*(Ax^* - b)||_2 \quad (26) $$

Then,

$$ ||e_{k+1}||_2 \leq \gamma ||e_k||_2 + \delta \quad (27) $$

As $A^*A$ is injective, it follows that, after optimizing for $P$ via (20) or (22), $\gamma < 1$. This in-turn implies (27) converges and the limit $e_{\infty}$ satisfies the following:

$$ ||e_{\infty}||_2 \leq \frac{\delta}{1 - \gamma} \quad (28) $$

Thus, the $l_2$-difference of the solutions to (3) and (10) is bounded by (28). Additionally, if $\kappa$ is the $l_2$-induced norm of $I - P$, note that,

$$ \delta \leq \kappa ||Ax^* - b||_2 \quad (29) $$

For most compressed sensing applications, $||Ax^* - b||$ is small, which in turn implies the error between (3) and (10) is also small.

III. POLYNOMIAL ADMM

If the precise solution to (2) or (3) without regularization tuning is desired, the above polynomial analysis can be applied to ADMM in order to solve for the exact formulation in a manner that reduces the number of $A^*A$ evaluations required per iteration.
A. Unconstrained Optimization

Equation (3), as formulated, can be solved using ADMM via Global Consensus [8]. Let $f(x) = (1/2)||Ax - b||_2^2$ and $\rho$ be a prior chosen step size. The consensus iterations are as follows, where $k$ denotes the iteration number:

\[
x^f_{k+1} = \text{prox}_{\rho f}(\tau_k - u^f_k)
\]

\[
x^g_{k+1} = \text{prox}_{\rho g}(\tau_k - u^g_k)
\]

\[
\tau_{k+1} = \left( x^f_{k+1} + x^g_{k+1} \right) / 2
\]

(30)

\[
u^f_{k+1} = u^f_k + x^f_{k+1} - \tau_{k+1}
\]

\[
u^g_{k+1} = u^g_k + x^g_{k+1} - \tau_{k+1}
\]

The proximal operator of $f$ is as follows:

\[
\text{prox}_{\rho f}(v) = \arg\min_x \frac{1}{2} \left[ \left( I / \sqrt{\rho} A \right) x - \left( v / \sqrt{\rho} b \right) \right]^2
\]

(31)

Note that utilizing Conjugate Gradient to solve (31) requires at least two $A^*A$ evaluations: one to calculate the initial residual of the initial estimate ($\tau_k$ in this case) and the other to perform a single iteration. In contrast, the polynomial-preconditioned gradient descent is as follows:

\[
\text{prox}_{\rho f}(v) = \tau_k - P \left[ (I + \rho A^*A)\tau_k - (v + \rho b) \right]
\]

(32)

Here, $P = \rho(I + \rho A^*A)$ is optimized according to (20) or (22). In this case, the later is preferred as, assuming the maximum eigenvalue of $A^*A$ is 1, the spectrum of $I + \rho A^*A$ lies in the interval $[1, 1 + \rho]$.

The advantage of formulating the proximal operator in the form of (32) is that, unlike in Conjugate Gradient, $P$ can be a zero-degree polynomial. This implies the inner loop of (30) requires only a single $A^*A$ evaluation. This is typically sufficient in practice due to the optimal nature of (22).

B. Constrained Optimization

Equation (2) can be equivalently formulated as the following consensus problem:

\[
\arg\min_{x_1, x_2} g(x_1) + \mathcal{I}_{B_b}(x_2) + \mathcal{I}_{B_A}(x_1, x_2)
\]

(33)

Here, convex sets $B_b$ and $B_A$ are defined as:

\[
B_b = \{ x : ||x - b||_2 \leq \epsilon \}
\]

\[
B_A = \{ (x_1, x_2) : Ax_1 = x_2 \}
\]

(34)

$\mathcal{I}_B$ is an indicator function that is zero if the input is a member of $B$, and is infinity otherwise.

The proximal operator of $\mathcal{I}_{B_A}$ involves solving a least squares problem that is very similar to (31). Consequently, the argument made to justify (32) over Conjugate Gradient to reduce the number of $A^*A$ evaluations apply here as well.

C. Multiple Regularizers

Given the above use of Global Consensus to solve (2) and (3), the use of the polynomials in ADMM extends easily to problems involving $N$ regularizations with easy-to-evaluate proximal operators to reduce the number of $A^*A$ evaluations.

IV. Experiments

In the spirit of reproducible research, the data and code used to perform the following experiments can be found at:

https://github.com/sidward/ppcs

All reconstructions were implemented in the Python programming language using SigPy [27]. The polynomial optimizations (20) and (22) were performed using SymPy [28], with the latter leveraging an excellent Chebyshev polynomial package available at https://github.com/mlazaric/Chebyshev.

To verify the efficacy of the preconditioner, four varied MRI reconstructions were studied using the unconstrained formulations (3) and (10). A single application was additionally studied to demonstrate the benefits of the polynomial preconditioner for the constrained formulation (2) and for using Global Consensus to solve an objective with multiple regularizations.

All experiments were performed on an AMD (R) EPYC 7502P 32-Core CPU and an NVIDIA(R) GeForce RTX 2080 Ti GPU. For all experiments, the corresponding measurement matrix $A$ was normalized to have a unitary induced $l_2$-norm, and the measurement vector $b$ was normalized to have unitary $l_2$-norm.

A. Unconstrained Optimization

For all the following unconstrained experiments, the degree of the polynomial ($d$) was a hyper-parameter that was tuned for the application of interest. Similarly, $\lambda$ of (3) and $\lambda_p$ of (10) were tuned for the best qualitative reconstruction performance. The minimum eigenvalue was assumed to be zero for the respective $A^*A$, and hence (20) was used for polynomial design. All reconstruction utilized FISTA [10] with a Nesterov momentum term of $(k - 1)/(k + 4)$, where $k$ is the iteration number, as described in [29]. The iterates with the preconditioner are denoted “Poly. Precond. FISTA" and the non-preconditioned iterates are denoted “FISTA". All seconds reported were measured in real time. The non-preconditioned experiments were allowed to run until the percentage $l_2$-difference between consecutive iterates was less than 0.2%. The preconditioned experiments were allowed to run for as long as the non-preconditioned experiments. Videos demonstrating the convergence results can be found in the code repository: https://github.com/sidward/ppcs.

1) Brain: Figure 1 depicts the convergence of iterates of FISTA with and without an optimized polynomial preconditioner on a publicly available Cartesian compressed sensing brain dataset that was retrospectively under-sampled using the variable density sampling mask depicted in Figure 1. The polynomial degree $d$ was set to 2.

The data were obtained from a fully-sampled, high-resolution acquisition of a human brain on a 1.5T scanner (GE, Waukesha, WI) using an eight-channel coil. The data
were obtained using an inversion-recovery prepared 3D RF-spoiled gradient-echo sequence with the following parameters: TR/TE=12.2/5.2 ms, TI=450 ms, FA=20°, BW=15 kHz, and a matrix size of $256 \times 180 \times 230$ with 1-mm isotropic resolution. The 3D dataset was Fourier transformed along the readout direction and a slice along the readout dimension was taken. This slice was retrospectively under-sampled using the variable-density random sampling mask depicted in Figure 1 and used to perform the experiment.

The unconstrained reconstruction formulation (3) for this experiment is as follows:

$$x^* = \arg \min_x \frac{1}{2} \| M F S W^* x - b \|_2^2 + \lambda \| x \|_1$$  \hspace{1cm} (35)$$

Here, $W$ is the forward Daubechies-4 Wavelet transform, $S$ is the SENSE model of the parallel-imaging acquisition [30] estimated using [31], [32], $F$ is the 2D-Fourier transform and $M$ is the sampling mask depicted in Figure 1. The regularization values for (3) and (10) were set to $\lambda = 5 \times 10^{-5}$ and $\lambda_p = 3 \times 10^{-4}$ respectively.

2) Cardiac: Figure 2 depicts the convergence of iterates of FISTA with and without an optimized polynomial preconditioner on a publicly available spiral cardiac dataset (used in [16]) that was acquired using the trajectory depicted in Figure 2. The polynomial degree $d$ was set to 4.

As described in [16], the data were acquired using a variable density spiral trajectory on a 1.5T scanner (GE Healthcare, Waukesha, WI) using an 8-channel cardiac coil and the HeartVistaRTHawk platform (HeartVista, Los Altos, CA) with a TR of 25.8 ms. The trajectory consisted of 3 interleaves and 3996 readout points, resulting in an effective under-sampling factor of eight. The acquisition field-of-view was $32 \text{ cm} \times 32 \text{ cm}$ with a matrix size of $320 \times 320$.

The unconstrained reconstruction formulation (3) for this experiment is as follows:

$$x^* = \arg \min_x \frac{1}{2} \| F S W^* x - b \|_2^2 + \lambda \| x \|_1$$  \hspace{1cm} (36)$$

Here, $W$ is the forward Daubechies-4 Wavelet transform, $S$ is the SENSE model of the parallel-imaging acquisition [30] estimated using [31], and $F$ is the non-uniform Fourier transform. The regularization values for (3) and (10) were set to $\lambda = 2 \times 10^{-5}$ and $\lambda_p = 4 \times 10^{-4}$ respectively.

This experiment utilized the Toeplitz structure of $F^* F$ for faster evaluation for both (3) and (10), which is not possible when utilizing [16].

3) Radial: Figure 3 depicts the convergence of iterates of FISTA with and without an optimized polynomial preconditioner on a publicly available radial brain dataset from the ISMRM reproducibility challenge which was acquired using the trajectory depicted in Figure 3. The polynomial degree $d$ was set to 5.

https://github.com/mikgroup/sigpy-mri-tutorial
This experiment used the same formulation as the cardiac experiment, with \( \lambda = 1 \times 10^{-6} \) and \( \lambda_p = 2 \times 10^{-5} \).

4) MRF: Figure 4 depicts the convergence of iterates of FISTA with and without an optimized polynomial preconditioner on data acquired using an optimized interleaved spiral-MRF sequence [33] that was acquired on a 3T scanner (GE Healthcare, Waukesha, WI) with IRB approval and informed consent obtained. The polynomial degree \( d \) was set to 9.

The data were obtained using a 3D TGAS-SPI-MRF acquisition with a total of 48 groups that acquired to achieve adequate 3D k-space encoding at each temporal data point [33]. Each acquisition group contains an adiabatic inversion preparation with TI=15ms and 500 variable flip-angle acquisitions with TR/TE=12.5ms/1.75ms and a 1.2s wait time for signal recovery to improve the signal-to-noise ratio, resulting in a net acquisition per group of 7.45s and total acquisition time of approximately 6 minutes. Additionally, a water-exciting rectangular pulse with duration of 2.38ms was used to depress the fat signal [34]. A variable density spiral trajectory with a 16-fold in-plane under-sampling rate at the center of k-space and a 32-fold under-sampling rate at the edge of k-space was used to achieve an encoding at 1-mm isotropic resolution with a field-of-view of \( 220 \times 220 \times 220 \)mm\(^3\) with a readout duration of 6.7 ms. Retrospective under-sampling was performed to simulate a 2 minute acquisition.

The unconstrained reconstruction formulation (3) for this experiment is as follows [21], [33], [35]–[38]:

\[
x^* = \arg \min_x \frac{1}{2} \| \mathcal{F} \Phi x - b \|^2_2 + \lambda \sum_r \| R_r(x) \|_s \\
\]

Here, \( \sum_r \| R_r(x) \|_s \) depicts the locally-low rank constraint [21], \( S \) is the SENSE model of the parallel-imaging acquisition [30] estimated using [39] and \( \mathcal{F} \) is the 3D non-uniform Fourier transform. Note that in this case, the reconstructed \( x \) consists of multiple “coefficient” images such that \( \Phi x \) recovers the temporal evolution of the underlying signal. Please see [33] for more information. The regularization values for (3) and (10) were set to \( \lambda = 2.5 \times 10^{-6} \) and \( \lambda_p = 8.75 \times 10^{-5} \) respectively.

Note that the generalize-ability of the polynomial preconditioner allows it to be used directly in a subspace-reconstruction [21], [33], [35]–[38] without needing to explicitly account for \( \Phi \) when designing the preconditioner. By utilizing \( A^* A \) to construct the preconditioner, the method inherently takes into account information from \( \Phi, S \) and \( \mathcal{F} \).

B. Constrained Optimization

Figure 5 depicts how the reduced number of \( A^* A \) evaluations as discussed in III results in faster optimization of (2) when applied to the Cardiac dataset. Here, \( A = \mathcal{F} \mathcal{S} W^* \) as in IV-A.2, \( g(x) = \| x \|_1 \) and \( \epsilon = 0.1525 \). For the ADMM iterations, \( \rho \) was set to \( 1 \times 10^{-4} \).

C. Multiple Regularizations

Figure 6 depicts how the reduced number of \( A^* A \) evaluations as discussed in III results in faster optimization of Global Consensus with multiple regularizations when applied to the Cardiac dataset.

The objective function solved is:

\[
x^* = \arg \min_x \frac{1}{2} \| \mathcal{F} \mathcal{S} x - b \|^2_2 + \lambda_1 \| W(x) \|_1 + \lambda_2 \| G(x) \|_2^2 \tag{38}
\]

Here, \( A = \mathcal{F} \mathcal{S} \), where \( \mathcal{F} \) and \( S \) are defined as in IV-A.2. Similarly, \( W \) is the same Wavelet transform used in IV-A.2. \( G \) denotes the Finite-Difference operator. The regularizations \( \lambda_1 \) and \( \lambda_2 \) were set to \( 1 \times 10^{-5} \) and \( 5 \times 10^{-6} \) respectively. For the ADMM iterations, \( \rho \) was set to 1.

V. Discussion

The polynomial preconditioner is seen to improve the conditioning of the unconstrained formulation (3), resulting in faster convergence compared to non-preconditioned FISTA. None of the preconditioned unconstrained reconstructions utilized any dual variables or an application-specific preconditioner array, resulting in comparable computational memory requirements to FISTA. The improvement in the rate of convergence for (10) does depend on the application of interest (based on the underlying \( A \)), the degree of the polynomial and value of \( \lambda_p \).

For small \( \lambda_p \), and high degree \( d \), (10) can be solved in 1-2 iterations depending on the application.

With appropriate tuning of \( \lambda_p \) in (10), the reconstructed images with the preconditioner are similar to the solutions of (3) with no noticeable noise coloring. This is as evidenced by Figures 1, 2 and 3, and is hypothesized to be a consequence of (11), where (3) and (10) have identical solutions when \( \lambda = \lambda_p = 0 \).
Fig. 4. Comparing convergence over time between non-preconditioned and polynomial-preconditioned FISTA for Experiment IV-A.4. Each column denotes a “coefficient” image, and all the coefficient images form $x$ in (37). Each coefficient image is scaled by the specified values for visualization. Acquisition time $\approx 2$ minutes.
For the Constrained Optimization and Multiple Regularizations experiments, the use of polynomials to justify a single $A^*A$ evaluation in the inner loop of ADMM was demonstrated to be sufficient. The ADMM parameter $\rho$ was empirically chosen and kept constant to compare the relative performance between using polynomials and Conjugate Gradient for the inner loop. That being said, an exhaustive tuning of $\rho$ or utilizing variable step-sizes might lead to faster convergence.

This work does not explore incorporating a weighting-prior into the spectral cost ($w$ in (21)). It is expected that a reasonable prior estimate of the spectrum of an operator $A$ derived for a specific application will significantly improve the rate of convergence for that application.

This work also does not explore the combination of density compensation with a polynomial-preconditioner, which is possible due to the generalize-ability of the preconditioner.

A limitation of the polynomial preconditioner is with respect to numerical stability. In principle, it is possible to utilize a polynomial $p$ of a high degree $d$. However, in practice, evaluating powers of $A^*A$ can accumulate numerical errors. Therefore, tuning the degree $d$ for the application of interest is required.

The error bound presented in (28) is applicable to any linear operator $P$ such that $P$ is injective on null$(A)^\perp$. Thus, deriving a $P$ to minimize the error bound in (28) is a promising avenue for application-specific preconditioner design. For example, (28) can be used to upper-bound the error of the circulant preconditioner in [15].

VI. Conclusion

This work proposed and validated a polynomial-based preconditioner that accelerates the convergence of iterative proximal methods like FISTA for compressed-sensing linear inverse problems. By leveraging polynomials, the preconditioner only requires matrix-free implementations of the forward model and is highly generalize-able, allowing the method to be easily applied various applications of interest.

VII. Acknowledgements

The authors would like to thank Dr. Sophie Schauman for testing the reproducibility code and Dr. Martin Uecker for providing additional references.

References

[1] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013.
[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
[3] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
[4] K. T. Block, M. Uecker, and J. Frahm, “Undersampled Radial MRI with Multiple Coils. Iterative Image Reconstruction Using a Total Variation Constraint,” *Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1086–1098, 2007.
[5] B. Liu, K. King, M. Steckner, J. Xie, J. Sheng, and L. Ying, “Regularized Sensitivity Encoding (SENSE) Reconstruction Using Bregman Iterations,” *Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 145–152, 2009.
[6] J. A. Fessler, “Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms,” *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 33–40, 2020.
[7] R. Tibshirani, “Regression shrinkage and selection via the LASSO,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
[8] N. Parikh and S. Boyd, “Proximal algorithms,” *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
[9] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed Sensing MRI,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.
[10] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, “Advances in sensitivity encoding with arbitrary k-space trajectories,” Magnetic Resonance in Medicine, vol. 46, no. 4, pp. 638–651, 2001.

K. Koolstra, J. van Gemert, P. Börnert, A. Webb, and R. Remis, “Accelerating compressed sensing in parallel imaging reconstructions using an efficient circulant preconditioner for cartesian trajectories,” Magnetic Resonance in Medicine, vol. 81, no. 1, pp. 670–685, 2019.

S. Ramani and J. A. Fessler, “Parallel MR Image Reconstruction Using Augmented Lagrangian Methods,” IEEE Transactions on Medical Imaging, vol. 30, no. 3, pp. 694–706, 2011.

D. S. Weller, S. Ramani, and J. A. Fessler, “Augmented Lagrangian with Variable Splitting for Faster Non-Cartesian $L_1$-SPIRIT MR Image Reconstruction,” IEEE Transactions on Medical Imaging, vol. 39, no. 5, pp. 1646–1654, 2020.

M. J. Muckley, D. C. Noll, and J. A. Fessler, “Fast, Iterative Sub-Sampled Spiral Reconstruction via Circulant Majorizers,” Proceedings of the International Society of Magnetic Resonance in Medicine, 2016.

F. Ong, M. Uecker, and M. Lustig, “Accelerating Non-Cartesian MRI Reconstruction Convergence Using k-Space Preconditioning,” IEEE Transactions on Medical Imaging, vol. 39, no. 5, pp. 1646–1654, 2020.

A. Chambolle and T. Pock, “A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging,” Journal of Mathematical Imaging and Vision, vol. 40, no. 1, pp. 120–145, 2011.

F. Wang, Z. Dong, T. G. Reese, B. Bilgic, M. Katherine Manhard, J. Chen, J. R. Polimeni, L. L. Wald, and K. Setsompop, “Echo Planar Time-Resolved Imaging (EPTI).” Magnetic Resonance in Medicine, vol. 81, no. 6, pp. 3599–3615, 2019.

D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold, “Magnetic Resonance Fingerprinting,” Nature, vol. 495, no. 7440, pp. 187–192, 2013.

A. Chambolle and C. Dossal, “On the Convergence of the Iterates of the ‘Fast Iterative Shrinkage/Thresholding Algorithm’,” SIAM Journal on Numerical Analysis, vol. 52, no. 2, pp. 340–362, 2015.

A. Van der Sluis and H. A. van der Vorst, “The Rate of Convergence of Conjugate Gradients,” Numerische Mathematik, vol. 48, no. 5, pp. 543–560, 1986.

F. Ong and M. Lustig, “SigPy: A Python Package for High Performance Iterative Reconstruction,” Proceedings of the International Society of Magnetic Resonance in Medicine, Montréal, QC, vol. 4819, 2019.

A. Meurer, C. P. Smith, M. Paprocki, O. Cerlić, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, v. Roučka, A. Saboo, I. Fernando, Š. Kulal, R. Cimrman, and A. Scopatz, “SymPy: Symbolic Computing in Python,” PeerJ Computer Science, vol. 3, p. e103, Jan. 2017.

A. Chambolle and C. Dossal, “On the Convergence of the Iterates of the ‘Fast Iterative Shrinkage/Thresholding Algorithm’,” Journal of Optimization Theory and Applications, vol. 166, no. 3, pp. 968–982, 2015.

K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “SENSE: Sensitivity Encoding for Fast MRI.” Magnetic Resonance in Medicine, vol. 42, no. 3, pp. 952–962, 1999.

M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanaewala, and M. Lustig, “ESPRIT—An Eigenvalue Approach to Autocalibrating Parallel MRI: Where SENSE Meets GRAPPA,” Magnetic Resonance in Medicine, vol. 71, no. 3, pp. 990–1001, 2014.

S. Iyer, F. Ong, K. Setsompop, M. Doneva, and M. Lustig, “SURE-based Automatic Parameter Selection for ESPRIT Calibration,” Magnetic Resonance in Medicine, vol. 84, no. 6, pp. 3423–3437, 2020.

X. Cao, C. Liao, S. Iyer, Z. Wang, Z. Zhou, E. Dai, G. Liberman, Z. Dong, T. Gong, H. He, J. Zhong, B. Bilgic, and K. Setsompop, “Optimized multi-axis spiral projection MR fingerprinting with subspace reconstruction for rapid whole-brain high-isotropic-resolution quantitative imaging,” Magnetic Resonance in Medicine, to be published.

Z. Dong, T. Gong, H. He, J. Zhong, B. Bilgic, and K. Setsompop, “Resonance in Medicine,” vol. 46, no. 3, pp. 1441–1455, 2020.

O. Norbeck, T. Sprenger, E. Avventi, H. Rydén, A. Kiis, J. Berglund, and S. Skare, “Optimizing 3D EPI for Rapid $T_t$-Weighted Imaging,” Magnetic Resonance in Medicine, vol. 64, no. 3, pp. 1090–1094, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.25486

F. H. Petzschner, I. P. Ponce, M. Blaimer, P. M. Jakob, and F. A. Breuer, “Fast MR Parameter Mapping using k-t Principal Component Analysis,” Magnetic Resonance in Medicine, vol. 66, no. 3, pp. 706–716, 2011.

M. Mani, M. Jacob, V. Magnotta, and J. Zhong, “Fast iterative algorithm for the reconstruction of multishot non-cartesian diffusion data,” Magnetic Resonance in Medicine, vol. 46, no. 4, pp. 638–651, 2001.

B. Zhao, K. Setsompop, E. Adalsteinsson, B. Gagoski, H. Ye, D. Ma, Y. Jiang, P. Ellen Grant, M. A. Griswold, and L. L. Wald, “Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling,” Magnetic Resonance in Medicine, vol. 79, no. 2, pp. 933–942, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.26701

L. Ying and J. Sheng, “Joint Image Reconstruction and Sensitivity Estimation in SENSE (JSENSE),” Magnetic Resonance in Medicine, vol. 57, no. 6, pp. 1196–1202, 2007.