PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

A GAUSS-NEWTON TRUST-REGION ALGORITHM FOR UNDERSAMPLED MAGNETIC RESONANCE PHASE CONTRAST VELOCITY RECONSTRUCTION

FELIPE CORTÉS SAAVEDRA

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
SERGIO URIBE

Santiago de Chile, May 2018

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

A GAUSS-NEWTON TRUST-REGION ALGORITHM FOR UNDERSAMPLED MAGNETIC RESONANCE PHASE CONTRAST VELOCITY RECONSTRUCTION

FELIPE CORTÉS SAAVEDRA

Members of the Committee:
SERGIO URIBE
CARLOS SING-LONG
MATÍAS COURDURIER
AMADOR GUZMÁN

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Santiago de Chile, May 2018

to those who seek the beauty in truth

ACKNOWLEDGEMENTS

First and foremost I want to thank my advisors Sergio Uribe and Carlos Sing-Long for their priceless advice and dedication in this work. Without them I would've never been introduced to the endless and fascinating world of biomedical imaging research.

I want to specially thank Carlos for introducing me into the beautiful topics of optimization research and numerical algorithms, mostly new and unexplored for me, and for giving me most valuable insight and intuition on them. One of the key attributes I have always looked up on teachers is the ability to communicate intuition and familiarity over complex topics and ideas that only a deep understanding can provide. Carlos not only has shown this ability but also a heartwarming disposition to share his knowledge to whoever wants to hear.

To all my friends and colleagues at the Biomedical Imaging Center Cristóbal, Pame, Julio, Hernán, Joaquín, Francisco, Mariana, Felipe, Nestor, Cristián, Gabiel, Ali and Carlos: I want to express my deepest respect and admiration. Each one of you made all those long days (and nights) worth it. I look forward into working with you and seeing you succeed in each of your future endeavours.

I want to also thank my very best friends Jorge, Andrés, Jimmy, Nacho and Gustavo who gave me much needed distraction and overall fun outside my studies, as well as being selflessly supportive and understanding in my sometimes prolonged absences.

Last, but definitely not least, I want to thank my family for all their support; my brother and sister for always being by my side, and my parents, who have always supported me and my projects with overwhelming love and care. You are the safe net that allow me to feel free to follow my dreams.

This work was partially funded by CONICYT - PIA - Anillo ACT1416

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

LIST OF FIGURES vii

ABSTRACT ix

RESUMEN x

1. Introduction 1

1.1. Magnetic Resonance Imaging 1

1.1.1. Spin Excitation and Signal Generation 1

1.1.2. Image Reconstruction 3

1.2. Phase Contrast 4

1.2.1. 4D Flow 7

1.3. Compressed Sensing 7

1.4. Numerical Optimization 9

1.4.1. Convex Optimization 9

1.4.2. Trust Region Problem 10

2. Gauss-Newton Trust Region algorithm for undersampled MR flow images 12

2.1. Introduction 12

2.2. Theory 14

2.2.1. Phase contrast 15

2.2.2. Data acquisition 16

2.2.3. Existing reconstruction methods 18

2.2.4. Divergence-free velocity fields 19

2.2.5. Recovery procedure 21

2.3. Methods 23

2.3.1. Optimization algorithm 23

2.3.2. Implementation 24

2.3.3. Initial Point 26
2.3.4. Sensitivity Estimation 27
2.3.5. Numerical Phantom 27
2.3.6. *In Vivo* Experiments 28
2.3.7. Error Metrics and Divergence 28
2.3.8. Numerical Experiments 29
2.4. Results 29
2.4.1. Convergence 29
2.4.2. Effect of the initial point 30
2.4.3. Sensitivity Estimation 31
2.4.4. Reconstruction quality 31
2.5. Discussion 36
3. Conclusions 38
REFERENCES 40
APPENDICES 46
A. Differential for the measurement map 47

LIST OF FIGURES

2.1 Magnitude images for different encoding directions of one of the in-vivo data sets, which are noticeably similar and are assumed to be constant in the reconstruction procedure. A low valued noise-like standard deviation of the normalized magnitudes can be also appreciated which further confirms the constant magnitude assumption. 19

2.2 Comparison between reconstructed aortas using a coordinate descent algorithm (2.2b) and the proposed algorithm (2.2c) after the same number of iterations for each one. Notice that the number of iterations of the proposed algorithms accounts for both the convex sub problems iteration and the major trust region problems iterations (70 and 60 in this case). All images were normalized and share the color scale. 25

2.3 Numerical Flow Phantom 28

2.4 Comparison between different initialization points on the reconstruction algorithm for one of the in-vivo datasets and an acceleration factor of 4. The same number of iterations and regularization parameters were used for each reconstruction. All images were normalized and share the color scale. 30

2.5 Comparison between assuming known coil sensitivities and using them as optimization variables. Notice the inaccuracies shown in green circles due to non-convergent areas. All images were normalized and share the color scale. . 31

2.6 Error for different acceleration factors on the phantom data 32

2.7 Comparison of the proposed Gauss-Newton Trust-Region method (GN-TR), Compressed Sensing (CS) and Minimum Energy (ME) reconstructions 33

2.8 Divergence of the resulting velocity fields for different reconstruction algorithms. 34

2.9 Reconstructed velocities of the phantom data for different acceleration rates. Magnitude of the velocity fields are shown 34

2.10 Comparison between a fully sampled and different undersampled reconstruction algorithms in a foot-head velocity of an aorta using 25% of the acquired data. 35

ABSTRACT

4d flow imaging is a Magnetic Resonance (MR) technique that enables the acquisition of a three-dimensional velocity field. The amount of measurements needed renders the technique sensitive to patient movement and precludes its wide adoption due to increased examination times. To accelerate the acquisition process, parallel imaging techniques are currently used and undersampled reconstruction techniques, inspired by Compressed Sensing, have been recently proposed. For the latter, flow incompressibility has been suggested as a method to find a sparsifying domain for the velocity. However, since the acquisition process in 4d flow is non-linear in the variables of interest, magnitude and phase, some questions remain open about how to effectively solve the resulting non-convex optimization problem.

In this work, we address the second point by presenting a novel approach for reconstructing flow images from undersampled phase contrast data that directly handles parallel acquisitions. The approach is based on a Gauss-Newton trust region method that is able to update the optimization variables simultaneously at each iteration and decreases the objective value at each iteration. In addition, we propose to use Total Variation (TV) as a regularizer for the phase. We show this regularizer is able to control the size of the divergence of the velocity and reduces the fluctuations on the recovered phase due to noise. Our results show the proposed method has several advantages: it converges faster, allows for simultaneous recovery and coil sensitivity estimation with minimal increase in computational burden, produces lower reconstruction errors, and is able to control the divergence of the recovered velocity field.

Keywords: Magnetic Resonance Imaging, Phase Contrast, Compressed Sensing, Non-Convex Optimization.

RESUMEN

4d flow es una técnica de adquisición de imágenes por Resonancia Magnética que permite obtener mapas tridimensionales de velocidades en el tiempo. Dada la cantidad de datos requeridos en este tipo de adquisiciones y el movimiento natural de los pacientes durante el examen, se generan distorsiones que reducen la calidad de los resultados. Esto provoca a su vez que la tecnología no esté ampliamente adoptada en el ámbito clínico. Los métodos más utilizados en la actualidad para acelerar los exámenes consisten en técnicas de adquisiciones de imágenes paralelas. Métodos de adquisición submuestreada de Resonancia Magnética han tomado relevancia en los últimos años, inspirados en la técnica de *Compressed Sensing*. Para este último, la incompresibilidad del flujo escaneado ha sido propuesto como una métrica de raleza de la velocidad, pero dado que el proceso de adquisición en *4d flow* es no-lineal en las variables de interés (magnitud y fase compleja), no es posible aplicar directamente los métodos de reconstrucción tradicionales.

En este trabajo, proponemos un método para reconstruir imágenes de flujo a partir de datos submuestrados basado en un algoritmo de Gauss-Newton con región de confianza que permite actualizar de forma conjunta todas las variables del problema en cada iteración. Además, proponemos usar Variación Total como métrica de regularización para la fase. Mostramos que este regularizador es capaz de controlar la magnitud de la divergencia del campo de velocidad resultante, reduciendo además grandes cambios en la fase producto de ruido. Los resultados muestran que el método propuesto presenta una mejor convergencia que otros similares, con bajos errores de reconstrucción y controla la divergencia del campo de velocidades reconstruido.

Palabras Claves: Resonancia Magnética, Contraste de Fase, *Compressed Sensing*, Optimización No-Convexa.

1. INTRODUCTION

1.1. Magnetic Resonance Imaging

Magnetic Resonance (MR) Imaging is one of the most important biomedical exams providing a high range and flexibility over the types of images it can generate while being virtually risk-free when compared to other imaging technology such as X-rays or computed tomography images.

Images generated by a MR scanner show the density of spins rotating at a specific frequency known as the Larmor Frequency which depends on intrinsic parameters of the scanner and the type of precessing atoms.

The image formation process has three distinct stages which will be briefly explained in the following sections:

(i) Spin excitation
(ii) Signal generation
(iii) Image reconstruction

1.1.1. Spin Excitation and Signal Generation

The basic theory of MRI has its origins on the magnetic nature of moving charges applied to the spinning nuclei of hydrogen atoms. While these usually rotate freely in normal conditions, they become aligned when under a constant magnetic field, though out of phase with respect to each other. Naturally, when a magnetic field is applied in the transversal direction of the previous one, rotating at the angular frequency of the spinning protons, their phase become synchronized. Thich is usually known as the in-phase state.

As moving charges induce a variable magnetic field, one can detect these variations using receiver coils and recover the spin density of the system as this value is proportional to the amplitude of these variations. This is the basis of the nuclear magnetic resonance

phenomenon, discovered independently by Purcell and Bloch in the 1940s (Purcell, Torrey, & Pound, 1946; Bloch, 1946) and was the precursor for image formation using the technique.

This does not allow us yet to form an image of the spin density, as no spatial information is being encoded on the received signal. A solution for this was proposed by Lauterbur in 1973 (Lauterbur, 1973) using linear magnetic field gradients, that can cause the spinning particles to acquire a linear de-phase following the direction of such gradients. When precisely calibrated, it can effectively encode spatial information from the excited spins and therefore enable image formation from the acquired signal. For a more detailed explanation see (Liang & Lauterbur, 1999).

The MR signal equation (1.1) follows from solving the Bloch equation (Bloch, 1946) that models the magnetization of particles under a magnetic field

$$S(t) = \int \rho(\vec{x}) e^{-i\gamma \int_0^t \vec{x}\cdot\vec{G}(\xi)d\xi} d\vec{x} \tag{1.1}$$

where $\rho(\vec{x})$ is the magnetization field, $\vec{G}(t)$ is the time-dependant gradient field, and γ is the gyromagnetic constant which depends only on the scanned medium.

One can also rewrite (1.1) with a spatio-temporal reparametrization of S into $\hat{S}$ with the following change of variable

$$\vec{k}(t) = \frac{\gamma}{2\pi} \int_0^t \vec{G}(\xi) d\xi \tag{1.2}$$

Which yields the standard k-space equation for MR signal generation:

$$\hat{S}(\vec{k}) = \int \rho(\vec{x}) e^{-i2\pi\vec{x}\vec{k}} d\vec{x} = \mathcal{F}\{\rho(\vec{x})\}(\vec{k}) \tag{1.3}$$

Where $\mathcal{F}$ corresponds to the Fourier transform operator.

It is worth noting that by changing the gradient shape and duration, different trajectories are transversed in the k-space while performing the data acquisition, thus allowing a great amount of versatility when trying to optimize for good reconstruction procedures.

1.1.2. Image Reconstruction

Eq. (1.3) lets us write the basic reconstruction scheme for magnetic resonance imaging using the inverse Fourier transform:

$$\rho(\vec{x}) = \mathcal{F}^{-1}\left\{\hat{S}(\vec{k})\right\}(\vec{x}) \tag{1.4}$$

As the Fourier transform is orthogonal, it can be easily shown that the solution of (1.4) is equivalent to the solution of the following minimization problem:

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} \frac{1}{2} \|y - \mathcal{F}\rho\|_2^2 \tag{1.5}$$

where y is the original k-space data. When y is assumed to have additive random noise the reconstruction procedure could yield to suboptimal point. To avoid overfitting when reconstructing the noisy image, the reconstruction procedure in (1.5) can be extended to a regularized minimization:

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}}\, R(\rho) \tag{1.6}$$

$$s.t.\ \frac{1}{2} \|y - \mathcal{F}\rho\|_2^2 \leq \epsilon \tag{1.7}$$

where R is a regularizer for the signal ρ. One common choice is the ℓ_2 norm, which leads into the well known Tikhonov regularized linear least squares for which a closed-form solution exists. While it is a computationally efficient problem to solve, it is not directly

clear *why* one would like to minimize the ℓ_2 norm of an image, so several other regularizers are usually used, that both enable a fast and efficient computation of the solution, and enforce properties expected on reconstructed images.

One large family of such regularizers, are the ones that enforce *sparsity* of the reconstructed signal in some known transformation space. This is closely related to the compressibility of a signal, where highly compressible signals have a low dimensionality representation (when compared to the dimensionality of the signal itself) and thus can be expressed in a sparse way. A more detailed discussion can be found in (Donoho, 2006a; Candès, 2014).

The wavelet transform (Jaffard, Meyer, & Ryan, 2001) is a frequently used space for both signal compression and as a regularizer for several image processing algorithms. It is a well known property that the wavelet transform of *natural* looking images[1] have a fast decaying wavelet transform and a rather sparse one too. This leads to an image reconstruction scheme using $R(\rho) = \|\Psi\rho\|_{\ell_1}$, where ψ is the wavelet transform operator, as the regularization function, which penalizes non-zero components of $\Psi\rho$. Now, although the wavelet transform is linear, and the ℓ_1 norm is convex, the reconstruction procedure is not as direct as the Tikhonov regularized one (which is both convex and differentiable over the domain) as it is not differentiable in the origin. However, several optimization algorithms exist to efficiently solve this issue, such as the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) (Beck & Teboulle, 2009).

1.2. Phase Contrast

Equation (1.1) makes several assumptions over both the scanned volume and the whole acquision scheme regarding the magentic field homogeneity, noise level, motion induced distortions, and several others. One relevant assumption that is often not met is the static

[1]Although there is no precise definition of what a *natural* looking image is, it could be argued that it is not needed as it comes from collective common sense, and that medical images fall comfortably into this category.

nature of the magnetization field $\rho(\vec{x})$. Being the signal equation time-dependent, having a moving object produces distortions that are not easily accounted for in the reconstruction stage and is currently an area of high research interest. Nevertheless, the fact that moving particles produce a known distortion can be used in favour of more refined data when choosing specific acquisition sequences.

We can extend (1.1) so that ρ is now a function of both space and time such that each point in space has a given constant velocity. This can be defined as

$$\rho(\vec{x}, t) = \rho_0(\vec{x} + t\vec{v}(\vec{x})) \tag{1.8}$$

By directly replacing the temporal magnetization (1.8) into the signal equation (1.1) we get something equivalent to

$$S(t) = \int \rho(\vec{x}) \exp\left\{-i\gamma \int_0^t \vec{G}(\xi) \cdot (\vec{x} + \vec{v}(x)\xi) d\xi\right\} d\vec{x} \tag{1.9}$$

$$= \int \rho(\vec{x}) \exp\left\{-i\gamma \left(\vec{x} \cdot \int_0^t \vec{G}(\xi) d\xi + \vec{v}(x) \cdot \int_0^t \xi\vec{G}) d\xi\right)\right\} d\vec{x} \tag{1.10}$$

Now, if $\vec{G}(t)$ is chosen such that its first moment is constant, (i.e. $\int \xi\vec{G}(\xi)d\xi = \vec{C}$), for example in bipolar gradients, then Eq. (1.9) can be further reduced to

$$\begin{aligned} S(t) &= \int \rho(\vec{x}) \exp\left\{-i\gamma\vec{v}(x) \cdot \vec{C}\right\} \exp\left\{-i\gamma\vec{x} \cdot \int_0^t \vec{G}(\xi) d\xi\right\} d\vec{x} \\ &= \int \rho(\vec{x}) \exp\left\{-i\gamma\vec{v}(x) \cdot \vec{C}\right\} \exp\left\{-i2\pi\vec{x} \cdot \vec{k}\right\} d\vec{x} \\ &= \mathcal{F}\left\{\rho(\vec{x}) \exp\{-i\gamma\vec{v}(x) \cdot \vec{C}\}\right\}(\vec{k}) \end{aligned}$$

So it turns out that the velocity information of the scanned volume is encoded as a complex phase shift of the reconstructed signal. It is worth noting that $\vec{C}$ is usually chosen

to be constant in all directions and equal to $\frac{\pi}{2\gamma v_{enc}}$ for a constant v_{enc}, such that $\gamma\vec{v}\cdot\vec{C} \in [-\pi/2, \pi/2]$ when $v \in [-v_{enc}, v_{enc}]$. If this constraint is not met, then aliasing artifacts appear in the reconstructed velocity. The v_{enc} parameter is controlled by changing the shape and duration of the bipolar gradients and trades off between high velocity to noise ratio and velocity aliasing.

Because the magnetic field is not completely homogeneous, the magnetization ρ is usually not real and has a complex phase component that would affect the reconstructed velocity. To solve this, Moran developed the phase contrast technique (Moran, 1982) where two acquisitions are made changing the sign of the bipolar gradient, which changes the sign of the constant v_{enc}.

This produces two reconstructed complex signals z_1 and z_2 that encode the velocities with opposing directions with an additive unknown field map ϕ_0

$$z_1 = \rho e^{i(\phi_0 + \frac{\pi}{2}\frac{v}{v_{enc}})}$$
$$z_2 = \rho e^{i(\phi_0 - \frac{\pi}{2}\frac{v}{v_{enc}})}$$

Therefore, the velocity can be reconstructed by subtracting the complex phase of the reconstructed signals.

$$v = \frac{v_{enc}}{\pi}(\angle z_1 - \angle z_2) \tag{1.11}$$

For a more detailed explanation of the technique see (Markl, 2005).

1.2.1. 4D Flow

The 4D flow technique (Markl, Frydrychowicz, Kozerke, Hope, & Wieben, 2012) consists in acquiring a temporal three dimensional velocity field of a volume by performing a phase contrast acquisition for every encoding direction. For this, four aquisitions are done for each spatio-temporal point in order to reconstruct the velocities using Eq. (1.11).

1.3. Compressed Sensing

In the last decade, the magnetic resonance research area took interest in one specific method of signal reconstruction based on the ideas of Donoho and Candès for undersampled reconstruction techniques (Donoho, 2006a; Candès, Romberg, & Tao, 2006). In their work they proved that arbitrary signals with certain degree of sparsity in a known projection space can be reconstructed with a subset of the original information provided some properties over this subset. When applied to magnetic resonance acquisition, the theoretical undersampling factor that can achieve almost perfect reconstruction is well beyond the Nyquist limit that classic signal analysis provides.

The technique gives a formal and precise analysis of the solution of problems where it is known to belong to a sparse subset of the domain. As an intuitive example we present the following canonical problem known as the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996).

$$\underset{x}{\operatorname{argmin}} \frac{1}{2} \|Ax - b\|_2 + \lambda \|x\|_1 \tag{1.12}$$

Notice that while sparsity should, in the most strict way, be penalized by the ℓ_0 norm, it was proven in (Donoho, 2006b; Candes, Romberg, & Tao, 2005), that both solutions are equivalent with high probability.

Magnetic Resonance images are known to be sparse in a variety of spaces, the identity being a common one. Others include finite differences, the wavelet domain and total variation.

As such, we can rewrite the reconstruction scheme in (1.4) to include a ℓ_1 penalization as

$$\hat{\rho} := \underset{\rho}{\operatorname{argmin}} \frac{1}{2} \|y - \mathcal{F}_\Omega \rho\|_2 + \lambda \|\Psi \rho\|_1 \tag{1.13}$$

Notice that now the transform used in the data consistency operator is an undersampled Fourier transform defined as $\mathcal{F}_\Omega = U_\Omega \mathcal{F}$, where U_Ω is a diagonal matrix with 1's on the indices of points that are sampled and 0's otherwise.

It is worth noting several properties of (1.13) that allow for efficient and precise reconstruction such as

Convexity: Having a convex objective function allows one to use several convex optimization algorithms with known convergence rate and error estimates.

Undersampled Reconstruction: By having regularizers that enforce known properties of the reconstructed signal, it is possible to used highly undersampled data, which in turn would allow for faster acquisitions.

Sparse Representation: The ℓ_1 penalization allows one to use several sparsifying transforms known to be well suited for magnetic resonance reconstruction. This can effectively reduce the noise of the reconstructed image and improve the reconstruction quality.

As stated in Section 1.1.2, several algorithms have been implemented to efficiently solve such problems. While initially least-squared minimizing routines were used (such as the conjugate gradient method (Hestenes & Stiefel, 1952)), better performance and convergence can be achieved with specially crafted algorithms for convex optimization with

non-smooth prox-capable (Parikh & Boyd, 2014) terms (such as the ℓ_1 norm). These algorithms include the well known Nesterov's proximal point method (Nesterov, 1983) and Beck and Teboulle's fast version of Nesterov's method FISTA (Beck & Teboulle, 2009).

1.4. Numerical Optimization

While most of equations and problems presented in this document and in the literature are proposed with continuous signals and transforms, they are usually implemented in a discretized way resulting in numerical algorithms that require specific data manipulation and fine tuning to make use of the desired properties of the problems.

For this specific formulation, several optimization techniques and algorithms are used, which are presented in the following sections.

1.4.1. Convex Optimization

A function $f : \mathbb{R}^N \to \mathbb{R}$ is said to be convex if the following property holds for every $x, y \in \mathbb{R}^N$ and every $0 \leq t \leq 1$

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y)$$

It can be easily proven that if f is continuous over a compact set $\Omega \in \mathbb{R}^N$ then f reaches a minimum $x^\star \in \Omega$ and, if f is strictly convex, that minimum is unique. Furthermore, if f is differentiable, then $x^\star$ is the limit of a sequence x_k that only uses local information of f in x_k.

For example $x_k := x_{k-1} - \delta \nabla f(x_{k-1})$ corresponds to the well known steepest descent algorithm, which if f is convex and differentiable, can get arbitrarily close to $x^\star$, i.e $\lim_{k\to\infty} x_k = x^\star$.

Several other algorithms have similar properties and are able to use different assumptions over the objective function to further improve convergence and overall performance.

Proximal algorithms (Parikh & Boyd, 2014), for example, assume functions of the form $f(x) = g(x) + h(x)$ where g, h are convex functions, g has continuous derivative over its domain, and h is non-differentiable function for which its prox $prox_h(x, \lambda)$ defined by

$$prox_h(x, \lambda) = \underset{y}{\text{argmin}}\, h(y) + \frac{1}{2\lambda} \|x - y\|_2^2$$

can be efficiently calculated. These functions include the ℓ_1 norm, for which its prox operator corresponds to a soft threshold around λ and indicator functions where their prox correspond to orthogonal projections over the respective set (Parikh & Boyd, 2014).

One family of such algorithms, and perhaps the most well known, is the FISTA (Beck & Teboulle, 2009) that allows a prox-capable term to be included in the objective function, thus functions such as ℓ_1 can be effectively used in minimization problems.

1.4.2. Trust Region Problem

A known approach to non-convex optimization that uses the better understood structure of convex functions consists on building approximate convex local models around a sequence of iterates, which are found by finding minimums of these local models.

Many decisions are to be made when designing a trust-region algorithm, as the local model construction, size and shape of the trust-region and point acceptance criteria. All of these, while crucial to the performance of the algorithm, are very well documented in the literature (Conn, Gould, & Toint, 2000) and several known algorithms exist to find such parameters (Nocedal & Wright, 2006).

If F is a non-convex function, the trust-region procedure can be defined as

$$x_{k+1} := \underset{\substack{x \\ T(x_k, x) \leq \rho_k}}{\text{argmin}}\ \hat{F}_{x_k}(x) \tag{1.14}$$

where $\hat{F}_x$ is some convex local approximation of F around x_k and $T(x_k, x), \rho_k$ define the trust-region shape and size around x_k. For example, a common implementation is to use a first order approximation of F, i.e.

$$\hat{F}_{x_k}(x) = F(x_k) + \langle \nabla F(x_k), (x - x_k) \rangle \tag{1.15}$$

and the ℓ_2 norm as the trust-region shape, i.e.

$$T_{x_k}(x) = \|x - x_k\|_2 . \tag{1.16}$$

The size of the trust-region can be updated at each iteration to ensure both a sufficiently good approximation of the true objective function, and a step size that allows a good convergence rate.

2. GAUSS-NEWTON TRUST REGION ALGORITHM FOR UNDERSAMPLED MR FLOW IMAGES

2.1. Introduction

4d flow imaging (Markl et al., 2012; Uribe et al., 2009) is a Magnetic Resonance (MR) technique that enables the acquisition of a three-dimensional velocity field. This is achieved by encoding the components of the velocity along three orthogonal directions in the complex phase of the obtained image. Unfortunately, the amount of measurements needed to recover the velocity field is substantial, increasing the data acquisition time. This renders the technique sensitive to patient movement and precludes its wide adoption. Consequently, the problem of accelerating 4d flow data acquisition has garnered significant attention in recent years.

One important breakthrough in the area of accelerated acquisition was the development of parallel imaging (Sodickson & Manning, 1997; Pruessmann, Weiger, Scheidegger, & Boesiger, 1999). In this technique, multiple reception coils with known sensitivities are used, increasing the effective measurements from separate overlapping spatial volumes, allowing a reduction in acquisition times. Since its introduction, several methods have been proposed to increase image quality in parallel imaging (Griswold et al., 2002; Lustig & Pauly, 2010) and today it is a widely used strategy to accelerate the acquisition of 4d flow data.

Another acceleration technique developed in the past decade relies on exact recovery from undersampled data. By enforcing known structural properties over the reconstructed image, it can be hypothesized that only a fraction of data is necessary to recover the whole image. This idea is rigorously formulated by the theory of Compressed Sensing (Candès et al., 2006; Donoho, 2006a) which was used for the first time for undersampled MR data reconstruction in the work of Lustig *et al* (Lustig, Donoho, & Pauly, 2007). They show that reconstruction from highly undersampled data—far below the Nyquist rate—can be achieved without compromising image quality. Compressed Sensing relies on the fact that

the data acquisition process is *linear* and one of its features is that the reconstruction involves solving a *convex* optimization problem.

In contrast with parallel imaging, Compressed Sensing is not readily applicable to accelerate 4d flow as the data acquisition process is non-linear—the velocity is encoded in the complex phase, and the resulting reconstruction problem is non-convex. In addition, it is not *a priori* evident which structure should be enforced in the complex phase. To overcome this, several approaches have been proposed in the literature.

One of them uses Compressed Sensing separately for the undersampled data acquired for each encoding direction, a method we call *frame-by-frame Compressed Sensing*. Then, the magnitude and complex phase can be obtained from the reconstructed complex images associated to each encoding direction (Kwak et al., 2013). The drawback of this approach is that it does not exploit the fact that the magnitude should be the *same* across encoding directions. This information can be enforced in order to improve reconstruction accuracy or increase undersampling rates.

Another approach consists in reconstructing the magnitude and phase data (Zhao, Noll, Nielsen, & Fessler, 2012) by explicitly considering them as variables in a non-convex optimization problem. Along this line, it has also been proposed to enforce structure separately for the the magnitude and phase data, such as sparsity in the wavelet domain and flow incompressibility (Santelli et al., 2016) respectively—enforcing flow incompressibility can be thought as applying a sparsifying transform: the divergence must be zero. Although the techniques show promising results, the approach to solve the resulting non-convex problem commonly uses (block) coordinate descent—that is, to perform alternating minimization on one variable, magnitude or phase, while leaving the other fixed. However, this algorithm has some drawbacks, as it does not exploit the effect that *both* the magnitude and phase information have on data consistency, potentially slowing down convergence and leading to undesirable local minima. These drawbacks become more noticeable when attempting to simultaneously recover the magnitude, phase, and coil sensitivities. In addition, the methods proposed to control the divergence of the recovered field, for instance using

divergence-free wavelets (Deriaz & Perrier, 2006) or finite element methods (Mura et al., 2016), increase the computational burden of solving the optimization problem.

In this work we focus on a different strategy to solve the non-convex optimization problem that results from 4d flow acquisition using a trust-region method (Nocedal & Wright, 2006; Conn et al., 2000). In particular, magnitude data, complex phase data, and coil sensitivities are treated as variables, and they are jointly optimized by the solver. Our results show our method converges faster than coordinate descent methods, and incurs in significantly smaller errors than recovering each encoding direction separately. In addition, instead of using divergence-based regularizers, we show that using a Total Variation (TV) regularizer has a similar effect without the computational burden associated to divergence-based regularizers.

2.2. Theory

In this section we discuss the theoretical framework underlying our method. Although our methods are readily applicable both to 2D or 3D applications, we will focus on presenting the 2D case to simplify the exposition; nevertheless we show results for 2D and 3D cases. Before proceeding we introduce the notation that will be used throughout. We will use standard typeface to denote scalars, boldface to denote vectors or matrices, and italics to denote operators. The symbol $\mathcal{F}$ denotes the (unitary) Fourier transform, and $\mathcal{F}_\Omega$ denotes the restriction of the Fourier transform to those frequencies belonging to Ω; in each case, the domain of the transform will be implied by the context. The symbol $\odot$ will denote the Hadamard product of vectors or matrices, i.e., $\boldsymbol{x} \odot \boldsymbol{y}$ is the vector resulting from computing the entry-wise products of the components of $\boldsymbol{x}$ and $\boldsymbol{y}$. We denote as $\mathbf{exp}\{i\boldsymbol{x}\}$ the vector with k-th entry given by e^{ix_k}, and as $|\boldsymbol{x}|$ the vector with k-th entry given by $|x_k|$. All these operations are defined analogously for matrices. In some cases, we will use functional notation to indicate the value of a matrix at any given pixel or voxel. Therefore, $\boldsymbol{Z}(x)$ represents the entry of the array $\boldsymbol{Z}$ associated to the position x, and $\boldsymbol{v}(x)$ represents the vector of the array $\boldsymbol{v}$ associated to the position x.

2.2.1. Phase contrast

In a phase contrast acquisition one measures the number of spins that were excited by a bipolar gradient and moved out of a voxel along a prescribed direction—the so called *encoding direction*. The motion of the spins induce a complex phase shift proportional to their speed along the encoding direction.

When a full acquisition is performed for each encoding direction, we recover n_p complex images. If the acquisition process is assumed to be instantaneous, distortionless, and has infinite precision, the n_p magnitude images are the same for any encoding direction, and the information about the velocity field is encoded in the changes across the n_p complex phase images. In particular, the velocity field can be typically recovered by taking linear combinations of the phase images (Pelc, Bernstein, Shimakawa, & Glover, 1991).

This leads to the following mathematical model. Let $\boldsymbol{m}$ and $\boldsymbol{\Phi}_p$ be the $n_x \times n_y$ real matrices representing the magnitude image and the phase image for the p-th acquisition respectively. The corresponding complex image is then given by $\boldsymbol{z}_p = \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\}$. If parallel acquisition is performed with n_c coils, we let $\boldsymbol{S}_c$ be the $n_x \times n_y$ real matrix representing the sensitivity of the c-th coil—the justification for assuming real-valued sensitivities will be discussed below. In this case, the c-th coil acquires data corresponding to the complex image $\boldsymbol{Z}_{p,c} = \boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\}$ for the p-th encoding direction. This clearly reduces to a single-coil acquisition if $n_c = 1$ and $\boldsymbol{S}_1$ is a matrix with all its entries equal to one. The model becomes

$$\boldsymbol{Z}_{p,c} = \boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\}\,. \tag{2.1}$$

To simplify notation, throughout our work we will denote $\boldsymbol{Z}$ the collection $\{\boldsymbol{Z}_{p,c}\}$, and let $\boldsymbol{\Phi}$ and $\boldsymbol{S}$ be defined similarly.

For known complex coil sensitivities, we can reduce the problem to the above by writing $\boldsymbol{S}_c = |\boldsymbol{S}_c| \odot \mathbf{exp}\{i\boldsymbol{\Theta}_c\}$ and observing that

$$\boldsymbol{Z}_{p,c} = |\boldsymbol{S}_c| \odot \boldsymbol{m} \odot \mathbf{exp}\{i(\boldsymbol{\Phi}_p + \boldsymbol{\Theta}_c)\}\,.$$

Therefore, the effect is a coil-dependent translation of the phase $\boldsymbol{\Phi}_p$ which, as we will see, can easily handled by the method we will develop for (2.1). If the sensitivities are unknown, even under the simplifying assumptions we have made, the model defined by (2.1) is *invariant* to some transformations of $\boldsymbol{S}$, $\boldsymbol{m}$ and $\boldsymbol{\Phi}$ and, as a consequence, it is not *identifiable*. In other words, it is not possible to uniquely characterize $(\boldsymbol{S}, \boldsymbol{\Phi}, \boldsymbol{m})$ from $\boldsymbol{Z}$. This can be readily seen from the trivial case $n_p = n_c = n_x = n_y = 1$ for which

$$z = sm\,\mathrm{e}\{i\phi\} = sm\,\mathrm{e}\{i(\phi + 2k\pi)\} = (-s)m\,\mathrm{e}\{i(\phi + k\pi)\} = s(-m)\,\mathrm{e}\{i(\phi + k\pi)\}$$

for any integer k. The list above is not exhaustive. Nevertheless, it illustrates the degrees of freedom we have to account for when attempting to find the tuple $(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi})$ from $\boldsymbol{Z}$. In addition, this trivial example helps us illustrate the effect of assuming unknown *complex* sensitivities. In this case there is a degree of freedom in the solution corresponding to an exchange between the phase in the coil sensitivity with the phase in the image and *vice versa*, implying that changes in the phase of the coil sensitivity can be misinterpreted as changes in velocity and *vice versa*. Accounting for these effects is fundamental when assessing the source of potential artifacts in the recovered tuple $(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi})$.

To mitigate this problem, we can remove the degrees of freedom by explicitly imposing that $\boldsymbol{S}, \boldsymbol{m} \geq 0$. In this work, we will not impose this restriction, but we will assume throughout that the tuple $(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi})$ is real.

2.2.2. Data acquisition

In a full acquisition we obtain the entries of the $n_x \times n_y$ complex matrix

$$\boldsymbol{Y}^0_{p,c} = \mathcal{F}(\boldsymbol{Z}_{p,c}) = \mathcal{F}(\boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\}\,),$$

for each one of the n_c coils, and for each one of the n_p encoding directions. Let $\Omega^{(p)}$ be the undersampling pattern for the p-th encoding direction, and let m_p the number of samples acquired for the p-th encoding direction, i.e., $m_p = |\Omega^{(p)}|$. Then the data acquired at the

c-th coil for the p-th encoding direction corresponds to the entries of the $m_p \times 1$ vector

$$\boldsymbol{Y}^0_{p,c} = \mathcal{F}_{\Omega^{(p)}}(\boldsymbol{Z}_{p,c}) = \mathcal{F}_{\Omega^{(p)}}(\boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\}\,). \tag{2.2}$$

In general, the acquired measurements will be corrupted by noise which can be assumed to be additive and Gaussian. The measured data becomes

$$\boldsymbol{Y}_{p,c} \sim \mathcal{N}(\boldsymbol{Y}^0_{p,c}, \sigma^2 \boldsymbol{I}_{m_p}) \tag{2.3}$$

where σ^2 is the noise variance, assumed to be known and independent for each coil and encoding direction. This leads us to define the non-linear map $\mathcal{T}$ by the identities

$$\begin{aligned} \boldsymbol{Y} &= \mathcal{T}(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) \qquad\qquad \Longleftrightarrow \\ \boldsymbol{Y}_{p,c} &= \mathcal{T}_{p,c}(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) \\ &:= \mathcal{F}_{\Omega^{(p)}}(\boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\}\,). \end{aligned}$$

where $\boldsymbol{Y}$ is defined analogously as $\boldsymbol{Z}$. This map is smooth, and although the map is non-linear, it is linear on its first and second variables, but not jointly. This fact will play a key role in the implementation of our recovery procedure. For future reference, we write the explicit expressions for its differential. The proof can be found in Appendix A.

PROPOSITION 2.2.1. *The differential of $\mathcal{T}$ at $(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi})$ is characterized by the expressions*

$$\begin{aligned} d\mathcal{T}_{p,c}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\delta\boldsymbol{S}, \delta\boldsymbol{m}, \delta\boldsymbol{\Phi}) = \\ &\mathcal{F}_{\Omega^{(p)}}(\delta\boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\}\,) + \\ &\mathcal{F}_{\Omega^{(p)}}(\boldsymbol{S}_c \odot \delta\boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\}\,) + \\ &i\mathcal{F}_{\Omega^{(p)}}(\boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}_p\} \odot \delta\boldsymbol{\Phi}_p) \end{aligned}$$

for $p = 1, \ldots, n_p$ *and* $c = 1, \ldots, n_c$*. Similarly, its adjoint is given by*

$$d\mathcal{T}^*_{p,c}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\boldsymbol{y}) = \begin{bmatrix} \mathbf{exp}\{-i\boldsymbol{\Phi}_p\} \odot \boldsymbol{m} \odot \mathcal{F}^*_{\Omega(p)}(\boldsymbol{y}) \\ \boldsymbol{S}_c \odot \mathbf{exp}\{-i\boldsymbol{\Phi}_p\} \odot \mathcal{F}^*_{\Omega(p)}(\boldsymbol{y}) \\ -i\boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{-i\boldsymbol{\Phi}_p\} \odot \mathcal{F}^*_{\Omega(p)}(\boldsymbol{y}) \end{bmatrix}$$

for $p = 1, \ldots, n_p$ *and* $c = 1, \ldots, n_c$*.*

2.2.3. Existing reconstruction methods

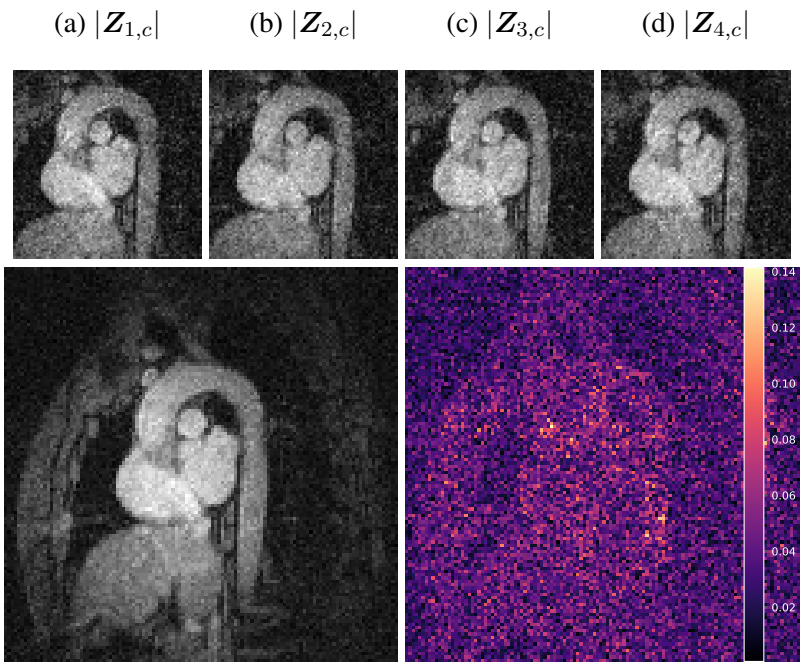
To avoid dealing with (2.1), one of the first approaches proposed in the literature was to consider $\boldsymbol{Z}_{p,c} = \boldsymbol{S}_c \odot \boldsymbol{X}_p$ where $\boldsymbol{X}_p$ is now a $n_x \times n_y$ *complex* array. Typically $\boldsymbol{Z}_{p,c}$ has a sparse representation if $\boldsymbol{X}_p$ does, as the coil sensitivities are smooth functions of the spatial variable. Inspired by Compressed Sensing (Candès et al., 2006), one might recover each $\boldsymbol{Z}_{p,c}$ by solving the convex problem

$$\min_{\boldsymbol{Z}} \ \frac{1}{2\sigma^2} \left\| \boldsymbol{Y}_{p,c} - \mathcal{F}_{\Omega(p)}(\boldsymbol{Z}) \right\|_2^2 + \lambda \left\| \Psi(\boldsymbol{Z}) \right\|_1, \tag{2.4}$$

i.e., the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996), where Ψ is a suitable sparsifying transform and λ is a user-selected regularization parameter. If we let $\boldsymbol{Z}^\star_{p,c}$ be the optimal solution to the above, we can attempt to recover $\boldsymbol{S}$, $\boldsymbol{m}$ and $\boldsymbol{\Phi}$ from the relations

$$|\boldsymbol{Z}^\star_{p,c}| = |\boldsymbol{S}_c| \odot \boldsymbol{m} \quad \text{and} \quad \boldsymbol{Z}^\star_{p,c} = |\boldsymbol{Z}^\star_{p,c}| \odot \mathbf{exp}\{i(\boldsymbol{\Phi}_p + \boldsymbol{\Theta}_c)\}. \tag{2.5}$$

From these relations there are several ways in which one could estimate the magnitudes and phases—applying linear least-squares being the most common (Bydder, Larkman, & Hajnal, 2002). We call this approach *frame-by-frame Compressed Sensing*. Observe (2.5) implies $|\boldsymbol{Z}^\star_{p,c}| = |\boldsymbol{Z}^\star_{p',c}|$ for $p \neq p'$. This condition can be checked directly, and we have found it *does* hold for real data (see Fig. 2.1). This suggests it is reasonable to exploit this redundancy in the data, by considering $\boldsymbol{S}$, $\boldsymbol{m}$ and $\boldsymbol{\Phi}$ independently in order to improve recovery of the velocity field or increase undersampling rates.



(a) $|\boldsymbol{Z}_{1,c}|$ (b) $|\boldsymbol{Z}_{2,c}|$ (c) $|\boldsymbol{Z}_{3,c}|$ (d) $|\boldsymbol{Z}_{4,c}|$

(e) Average image (f) Standard Deviation

Figure 2.1. Magnitude images for different encoding directions of one of the in-vivo data sets, which are noticeably similar and are assumed to be constant in the reconstruction procedure. A low valued noise-like standard deviation of the normalized magnitudes can be also appreciated which further confirms the constant magnitude assumption.

2.2.4. Divergence-free velocity fields

Aside from exploiting the conditions in (2.5) it has also been proposed to exploit the physical constraints on the velocity field. In particular, it has been proposed that blood can be modeled as an incompressible fluid, from where it follows its divergence should be zero. Let x denote the position of a pixel in the image, e.g., a pair (i, j) of indices, and let

$$\boldsymbol{\Phi}(x) = \begin{bmatrix} \boldsymbol{\Phi}_1(x) & \dots & \boldsymbol{\Phi}_{n_p}(x) \end{bmatrix}^t$$

be the vector field that associates to x the vector of n_p phases. The velocity at x is characterized by

$$\boldsymbol{v}(x) = \boldsymbol{B}\boldsymbol{\Phi}(x)$$

where $\boldsymbol{B}$ is a $2 \times n_p$ velocity encoding matrix. The divergence of the velocity field is given by

$$\operatorname{div} \boldsymbol{v} = \operatorname{trace}(\boldsymbol{B} D\boldsymbol{\Phi})$$

where $D\boldsymbol{\Phi}$ is the differential of the field. For a given $\boldsymbol{w}$ we can write

$$\begin{aligned}
\langle \boldsymbol{w},\, \operatorname{div} \boldsymbol{v} \rangle &= \sum\nolimits_x w(x) \operatorname{div} \boldsymbol{v}(x) \\
&= \sum\nolimits_x \operatorname{trace}(w(x) \boldsymbol{B} D\boldsymbol{\Phi}(x)) \\
&=: \sum\nolimits_x \langle w(x) \boldsymbol{B}^t,\, D\boldsymbol{\Phi}(x) \rangle \\
&=: \langle w\boldsymbol{B}^t,\, D\boldsymbol{\Phi} \rangle,
\end{aligned}$$

where $w\boldsymbol{B}^t$ is the field that associates to each pixel x the $n_p \times 2$ matrix $w(x)\boldsymbol{B}^t$. For such fields we can define the ℓ_p/ℓ_q-norm for $1 \leq p, q < \infty$ as

$$\|\boldsymbol{A}\|_{p,q} = \sum\nolimits_x \left(\sum\nolimits_i \left(\sum\nolimits_j A_{ij}(x)^p \right)^{q/p} \right)^{1/q},$$

for $1 \leq p < \infty$ and $q = \infty$ as

$$\|\boldsymbol{A}\|_{p,\infty} = \sum\nolimits_x \max\nolimits_i \left(\sum\nolimits_j A_{ij}(x)^p \right)^{1/p}.$$

Using Hölder's inequality and by the definition of the Total Variation seminorm we see that

$$\begin{aligned}
\langle \boldsymbol{w},\, \operatorname{div} \boldsymbol{v} \rangle &\leq \|w\boldsymbol{B}^t\|_{2,1}^* \|D\boldsymbol{\Phi}\|_{2,1} \\
&= \|w\boldsymbol{B}^t\|_{2,1}^* \sum\nolimits_p \|\boldsymbol{\Phi}_p\|_{\mathrm{TV}},
\end{aligned}$$

where $\|\cdot\|_{2,1}^*$ denotes the dual norm. In this case, it is well-known that $\|\cdot\|_{2,1}^* = \|\cdot\|_{2,\infty}$ (Hardy, Littlewood, & Pólya, 1952). The above indicates that any norm on the divergence is essentially controlled by the TV seminorm of the phases. To be concrete, to control the ℓ_p-norm of the divergence, it suffices to consider

$$\|\operatorname{div} \boldsymbol{v}\|_p \leq \underbrace{\sup_{\|w\|_q \leq 1} \|w\boldsymbol{B}^t\|_{2,\infty}}_{:= C_p(B)} \sum\nolimits_p \|\boldsymbol{\Phi}_p\|_{\mathrm{TV}} \tag{2.6}$$

where $1/p + 1/q = 1$. The constant involved can be easily computed for popular choices of values for p. In fact, for the simple 4-point encoding scheme (Pelc et al., 1991), the encoding matrix becomes

$$\boldsymbol{B} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

and $C_p(B) = \sqrt{2}(n_x n_y)^{1/p}$ whereas for balanced 4-point encoding we have

$$\boldsymbol{B} = \frac{1}{2} \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix}$$

and $C_p(B) = (\sqrt{3}/2)(n_x n_y)^{1/p}$. It follows that the smallest constant is obtained for simple 4-point encoding and $p = \infty$. In this case

$$\| \operatorname{div} \boldsymbol{v} \|_\infty \leq \sqrt{2} \sum\nolimits_p \| \boldsymbol{\Phi}_p \|_{\mathrm{TV}}.$$

2.2.5. Recovery procedure

One of the main tools to address the recovery problem is to use variational methods, i.e., to find a minimizer of

$$\min_{(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi})} F(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) \tag{2.7}$$

for some suitable objective function F. Typically, the objective can be decomposed as the sum of a data consistency function, and a regularizer, which promotes structural properties in the solution. This is clearly the case in (2.4). Given the noise model (2.3) we use the so called loss function

$$\begin{aligned} L(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) &= \frac{1}{2\sigma^2} \| \boldsymbol{Y} - \mathcal{T}(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) \|_2^2 \\ &:= \frac{1}{2\sigma^2} \sum\nolimits_{p,c} \| Y_{p,c} - \mathcal{T}_{p,c}(\boldsymbol{S}_c, \boldsymbol{m}, \boldsymbol{\Phi}_p) \|_2^2 . \end{aligned} \tag{2.8}$$

to enforce data consistency.

To motivate our choice of regularizer, we consider the following structural properties we promote in the solution:

– *Sparse representation of $\boldsymbol{m}$:* There is extensive literature providing empirical evidence for the sparsity of medical images in the wavelet domain (Lustig et al., 2007). For this reason, we consider the regularizer

$$R_m(\boldsymbol{m}) = \|\Psi(\boldsymbol{m})\|_1 \tag{2.9}$$

where Ψ is the Daubechies-4 wavelet transform of $\boldsymbol{m}$.

– *Smoothness of the coil sensitivities:* As coil sensitivity maps are an approximation of the receptive field of an antenna that follows the Biot-Savart law with a quadratic decay over an homogenous field, we assume them to have a certain degree of smoothness. We promote this by the following regularizer

$$R_s(\boldsymbol{S}) = \frac{1}{2}\sum\nolimits_c \|\nabla \boldsymbol{S}_c\|_{2,2}^2\,, \tag{2.10}$$

i.e., the sum of the ℓ_2-norm squared of the gradient of the coil sensitivities at each point.

– *Phase:* As mentioned before, we will consider using the TV seminorm to regularize the phase for each encoding direction. Consequently,

$$R_\phi(\boldsymbol{\Phi}) = \sum\nolimits_p \|\boldsymbol{\Phi}_p\|_{\mathrm{TV}}. \tag{2.11}$$

Therefore, we propose the regularizer

$$R(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) = \lambda_m R_m(\boldsymbol{m}) + \lambda_\phi R_\phi(\boldsymbol{\Phi}) + \lambda_s R_s(\boldsymbol{S}), \tag{2.12}$$

where the parameters $\lambda_m, \lambda_s, \lambda_\phi \geq 0$ are selected by the user (see the discussion in Section 2.3.2). The objective function then becomes

$$F(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) = L(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) + R(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}), \tag{2.13}$$

where only the data consistency term is non-convex, as the regularizers are convex functions. Consequently, we propose the following recovery procedure

$$(\boldsymbol{S}^\star, \boldsymbol{m}^\star, \boldsymbol{\Phi}^\star) \in \underset{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}{\operatorname{argmin}} \; F(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi})$$

where we have used the inclusion symbol instead of an equality sign to emphasize the potential existence of multiple minimizers (see the discussion in Section 2.2.1). Furthermore, since the objective (2.13) is non-convex, it is not clear whether we can find *any* global minimizer in practice, although this is not a necessary condition for an effective reconstruction. We defer the discussion of these issues to Section 2.3.3 below.

2.3. Methods

2.3.1. Optimization algorithm

A popular approach to solve the non-convex optimization problem (2.7) is to use a method akin to coordinate descent (Sun et al., 2017; Santelli et al., 2016). In this approach, one generates iterates by optimizing over one of the variables $\boldsymbol{S}$, $\boldsymbol{m}$ or $\boldsymbol{\Phi}$ while leaving the remaining two fixed. This technique is particularly attractive when minimizing over $\boldsymbol{S}$ or $\boldsymbol{m}$ as the resulting problem is convex. Unfortunately, this approach can be inefficient, as we only change one variable at the time and it is unclear with how much precision one should optimize over $\boldsymbol{S}$ and $\boldsymbol{m}$. As a consequence, convergence can be slow.

Although non-convex, the objective in (2.13) has structural properties that can be exploited to solve (2.7) efficiently. Concretely, the objective is the sum of convex functions composed with smooth non-linear operators; the regularizer in (2.13) is convex whereas the data consistency term is the composition of a strongly convex function with a non-linear smooth map. Solving optimization problems with non-convex objectives of this form has attracted significant attention in recent years.

We propose to use a Gauss-Newton trust region method (Lewis & Wright, 2016; Burke & Ferris, 1995) to solve (2.7). In this method, we generate a sequence of iterates $\{(\boldsymbol{S}^{(n)},$

$\boldsymbol{m}^{(n)}, \boldsymbol{\Phi}^{(n)})\}$ and radii $\{\rho^{(n)}\}$ as follows: at each iterate $(\boldsymbol{S}^{(n)}, \boldsymbol{m}^{(n)}, \boldsymbol{\Phi}^{(n)})$ we consider the convex model (2.15) and we let the candidate for next iterate be the optimal solution to the *convex* optimization problem

$$\begin{aligned} \underset{(\delta\boldsymbol{S},\delta\boldsymbol{m},\delta\boldsymbol{\Phi})}{\text{minimize}} \quad & M_n(\delta\boldsymbol{S},\delta\boldsymbol{m},\delta\boldsymbol{\Phi}) \\ \text{subject to} \quad & \|(\delta\boldsymbol{S},\delta\boldsymbol{m},\delta\boldsymbol{\Phi})\|_2 \leq \rho^{(n)}. \end{aligned} \tag{2.14}$$

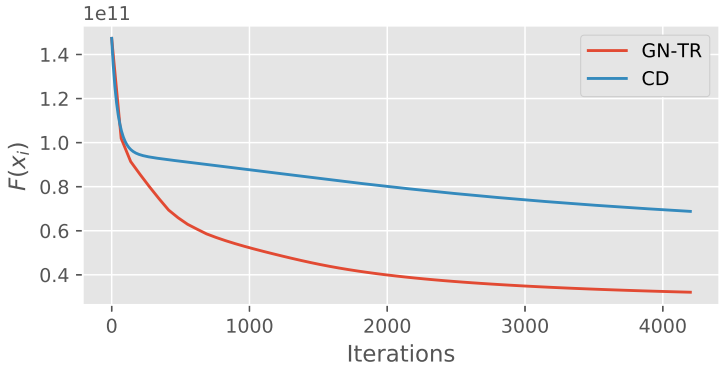
Strategies for accepting or rejecting this new iterate and finding the next trust-region radius $\rho^{(n+1)}$ are well-documented in the literature on trust-region solvers (Nocedal & Wright, 2006). In addition, the iterate obtained by solving (2.14) can be used to obtain a descent direction for the objective in (2.13) and linesearch (Nocedal & Wright, 2006) can be performed to improve the convergence of the method.

The advantage of this approach is that we can use accelerated proximal methods to solve the trust-region problem (2.14), ensuring it can be done efficiently. Furthermore, in using (2.15) one expects to capture *curvature information* about the objective, from where it follows that the iterates generated by this method should outperform those generated by coordinate descent (see Fig. 2.2)

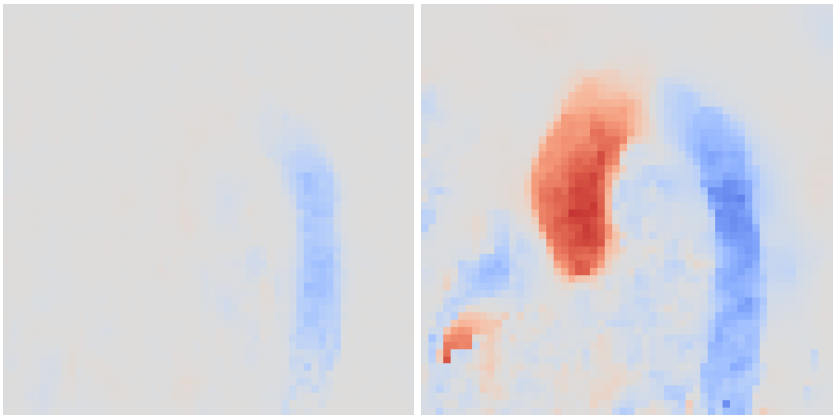
2.3.2. Implementation

Although the convex functions do not need to be smooth, to simplify the implementation we used the Moreau (or Moreau-Yosida) envelope (Moreau, 1965) of the ℓ_1-norm in (2.9) and the ℓ_1/ℓ_2-norm in (2.11) with smoothing parameter 10^{-3} instead of the explicit non-smooth functions. This allowed the trust region problem (2.14) to be solved

$$\begin{aligned} M_n(\delta\boldsymbol{S},\delta\boldsymbol{m},\delta\boldsymbol{\Phi}) := \frac{1}{2\sigma^2}\left\|\boldsymbol{Y} - \mathcal{T}(\boldsymbol{S}^{(n)},\boldsymbol{m}^{(n)},\boldsymbol{\Phi}^{(n)}) - d\mathcal{T}|_{(\boldsymbol{S}^{(n)},\boldsymbol{m}^{(n)},\boldsymbol{\Phi}^{(n)})}(\delta\boldsymbol{S},\delta\boldsymbol{m},\delta\boldsymbol{\Phi})\right\|_2^2 \\ + R(\delta\boldsymbol{S}+\boldsymbol{S}^{(n)},\delta\boldsymbol{m}+\boldsymbol{m}^{(n)},\delta\boldsymbol{\Phi}+\boldsymbol{\Phi}^{(n)}) \end{aligned} \tag{2.15}$$



(a) Objective function value for both algorithm's iterations



(b) Coordinate Descent Algorithm (c) Gauss-Newton trust region Algorithm

Figure 2.2. Comparison between reconstructed aortas using a coordinate descent algorithm (2.2b) and the proposed algorithm (2.2c) after the same number of iterations for each one. Notice that the number of iterations of the proposed algorithms accounts for both the convex sub problems iteration and the major trust region problems iterations (70 and 60 in this case). All images were normalized and share the color scale.

using FISTA (Beck & Teboulle, 2009). The maximum number of iterations for this algorithm was 100, and the Lipschitz constant for the trust region problem was estimated using the backtracking method outlined in (Beck & Teboulle, 2009).

The strategy for choosing the radii and accepting or rejecting the solution to the trust region problem (2.14) can be found in Algorithm 4.1 in (Nocedal & Wright, 2006). The

initial radius was 10 with a minimum and maximum value of 10^{-2} and 10^{3} respectively. In addition, when the solution was accepted, linesearch was performed using the difference between the new and old iterates as the search direction. Since the Moreau envelopes are differentiable, we were able to use Algorithm 3.5 in (Nocedal & Wright, 2006) to perform linesearch over the iterates.

The regularization parameters used in the final implementation were $\lambda_m = 10^3;\ \lambda_\phi = 2 \cdot 10^5;\ \lambda_S = 3 \cdot 10^6$

As mentioned earlier, the constraints $\boldsymbol{S}, \boldsymbol{m} \geq 0$ are not enforced. This leads to potential artifacts where the reconstructed phase shifts in factors of π at the points where the magnitude or the coil sensitivities are negative. Nevertheless, this artifacts are easily corrected by unwrapping the reconstructed phase images and by the fact that the velocities depend on the differences between the phases and not their absolute values (Herráez, Burton, Lalor, & Gdeisat, 2002).

The algorithm was implemented in the Python programming language using the numpy library (Walt, Colbert, & Varoquaux, 2011), with some parts written in C using the FFTW library (Frigo & Johnson, 2005) to improve its performance.

2.3.3. Initial Point

As stated in Section 2.2.5, the initial point for the numerical solver plays a key role in the recovered variables. In fact, from Proposition 2.2.1 it is clear that selecting all variables equal to zero yields a stationary point to (2.7). For this reason, several initial points were tested.

When both the sensitivities and an estimate $\boldsymbol{Z}^{\star}_{p,c}$ of the complex image for the c-th coil and p-th direction are known, we can estimate the magnitude and phase as follows. First,

the magnitude estimate is

$$\boldsymbol{m}^{(0)} = \frac{1}{n_c n_p} \sum_{c=1}^{n_c} \sum_{p=1}^{n_p} |\boldsymbol{Z}^{\star}_{p,c}| \quad \text{and} \quad \boldsymbol{\Phi}^{(0)}_{p} = \angle \sum_{c=1}^{n_c} \boldsymbol{Z}^{\star}_{p,c},$$

where $\angle$ denotes the complex phase of its argument. The methods used to obtain the estimates $\boldsymbol{Z}^{\star}_{p,c}$ were the following.

(i) The *zero-filling* or *minimum-energy* estimate assumes the unobserved data is equal to zero and applies an inverse Fourier transform, i.e., $\boldsymbol{Z}^{\star}_{p,c} = \mathcal{F}^{*}_{\Omega^{(p)}}(\boldsymbol{y}_{p,c})$.
(ii) The LASSO estimate is obtained by the procedure described in Section 2.2.3. In other words, $\boldsymbol{Z}^{\star}_{p,c}$ is the optimal solution to the LASSO problem in (2.4) with data $\boldsymbol{y}_{p,c}$.
(iii) The *zero-velocity* estimate assumes there is no movement in the image, and that the magnitude is a constant equal to one. In other words, $\boldsymbol{\Phi}^{(0)}_{p} = \mathbf{0}$ and $\boldsymbol{m}^{(0)} = \mathbf{1}$.

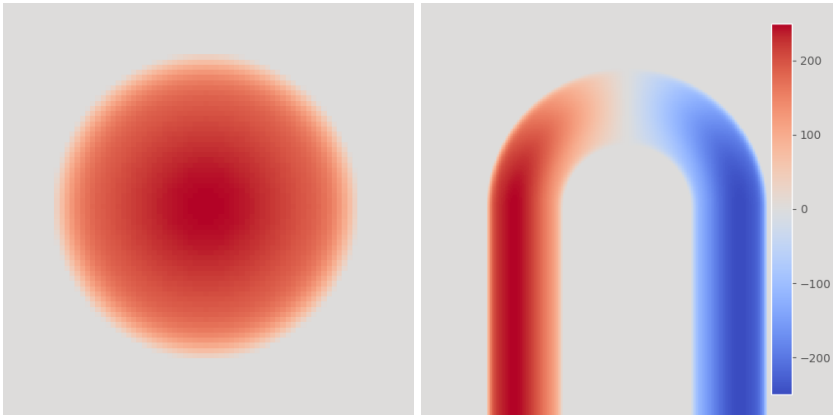
For each one of these initialization strategies, when the coil sensitivities were being simultaneously estimated, they were assumed to be constant and equal to one.

2.3.4. Sensitivity Estimation

While coil sensitivities could be estimated and treated as known for the proposed approach, this did not cause any significant improvement on the resulting image, and slowed down convergence. To test this, coil sensitivities were estimated using the method in (Uecker et al., 2014) with the implementation in (Uecker & Lustig, n.d.).

2.3.5. Numerical Phantom

A three-dimensional numerical phantom was developed to test the algorithm performance with known ground truth data. It is defined as a bent pipe with laminar flow from a fixed velocity profile (see Figs. 2.3a and 2.3b). A Womersley flow model (Womersley, 1955) was used as the profile for an approximated cardiac flow.



(a) Womersley profile (b) Vertical velocity

Figure 2.3. Numerical Flow Phantom

A maximum velocity value of $250cm/s$ was set and the maximum velocity frame was used for the experiments.

2.3.6. *In Vivo* Experiments

Fully-sampled data from five healthy subjects were acquired on a 1.5T MR Philips Achieva Scanner (Best, The Netherlands) using a 4D flow scan in a single sagittal slice of the aorta with parameters FOV $= 230 \times 23$ mm^2, res $= 2.5 \times 2.5$ mm^2 TR/TE $= 5/3$ ms, $v_{enc} = 200$ cm/s, with a five-channel cardiac coil.

2.3.7. Error Metrics and Divergence

Two different error metrics were used to evaluate the performance of the proposed method relative to the fully sampled reconstructed velocity field. These aim to quantify the errors in magnitude, and in direction between the recovered velocity field $\boldsymbol{v}$ and fully-sampled velocity field $\boldsymbol{v}_0$ over a given region of interest (ROI) denoted by $\mathcal{R}$.

(i) The *mean directional error* quantifies the difference between the directions of the recovered and fully-sampled velocity fields over a ROI. It is defined as

$$\text{MDE} = \frac{1}{|\mathcal{R}|} \sum_{x \in \mathcal{R}} \left(1 - \frac{|\boldsymbol{v}(x) \cdot \boldsymbol{v}_0(x)|}{\|\boldsymbol{v}(x)\|_2 \|\boldsymbol{v}_0(x)\|_2} \right). \tag{2.16}$$

(ii) The *normalized root mean square error* quantifies the difference between the magnitudes of the recovered and fully-sampled velocity fields over a ROI. It is also used to quantify the error between the reconstructed and fully-sampled magnitude images. It is defined as

$$\text{nRMSE}_v = \sqrt{\frac{\sum_{x \in \mathcal{R}} (\|\boldsymbol{v}(x)\|_2 - \|\boldsymbol{v}_0(x)\|_2)^2}{\sum_{x \in \mathcal{R}} \|\boldsymbol{v}_0(x)\|_2^2}}. \tag{2.17}$$

The ROI was chosen to be within the boundary of the vessels in the scanned volume. This choice reduces noise artifacts due to complex phase noise on low magnitude areas.

Additionally the divergence of the resulting velocities for the phantom data was estimated using central finite differences to test the performance of the different algorithms.

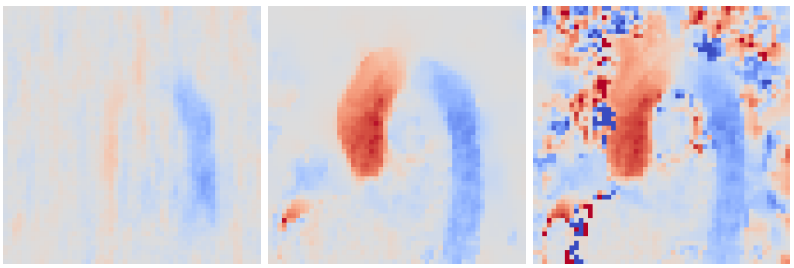
2.3.8. Numerical Experiments

For both phantom and *in vivo* data, the error incurred by the proposed method was compared both against a frame-by-frame Compressed Sensing reconstruction with the ℓ_1-norm regularization of the wavelet coefficients of the complex images, and a minimum ℓ_2-norm reconstruction. In both cases, the undersampling patterns for different acceleration factors were randomly generated using a variable density distribution that fully samples the region around the center of the *k*-space.

2.4. Results

2.4.1. Convergence

In Fig. 2.2 we show the result of solving the reconstruction problem using coordinate descent and using the proposed approach for one of the in-vivo data sets. In Fig. 2.2a we



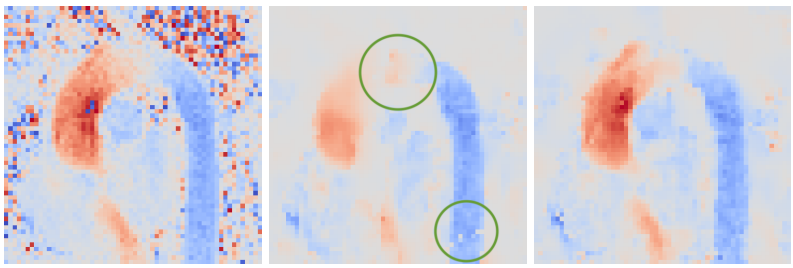
(a) Zero-velocity (b) Minimum-energy (c) LASSO estimate

Figure 2.4. Comparison between different initialization points on the reconstruction algorithm for one of the in-vivo datasets and an acceleration factor of 4. The same number of iterations and regularization parameters were used for each reconstruction. All images were normalized and share the color scale.

observe that over 4000 iterations, our approach achieves an objective value that is about 50% lower than that achieved by coordinate descent. In Figs. 2.2b and 2.2c we see the effect of the convergence in the recovered image. Our approach is able to recover most of the qualitative features of the phase variable (see Fig. 2.2c), whereas the result obtained by coordinate descent does not resolve the main features of the velocity image (see Fig. 2.2b).

2.4.2. Effect of the initial point

Empirical evidence obtained through several experiments suggests that the reconstruction obtained using the *minimum-energy* initial point (see Section 2.3.3) incurs the least error when compared to the LASSO estimate, which generates substantial noise in low magnitude points, and the *zero-velocity* estimate that takes longer to converge. This initial point has the additional advantage of having a low computational cost. Fig. 2.4 shows the resulting recovered phase for the same number of iterations for different initial points. The high variance in the performance regarding the initial point is most likely a consequence of the lack of convexity of the model.



(a) Ground Truth (b) Sensitivities as known values (c) Sensitivities as variables

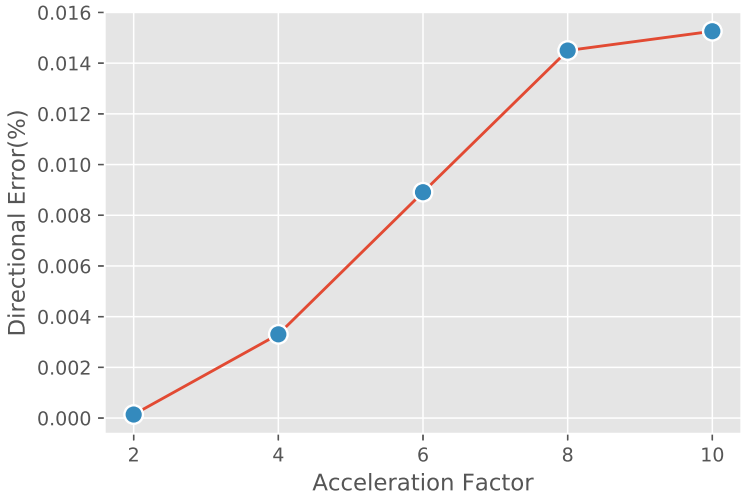
Figure 2.5. Comparison between assuming known coil sensitivities and using them as optimization variables. Notice the inaccuracies shown in green circles due to non-convergent areas. All images were normalized and share the color scale.

2.4.3. Sensitivity Estimation

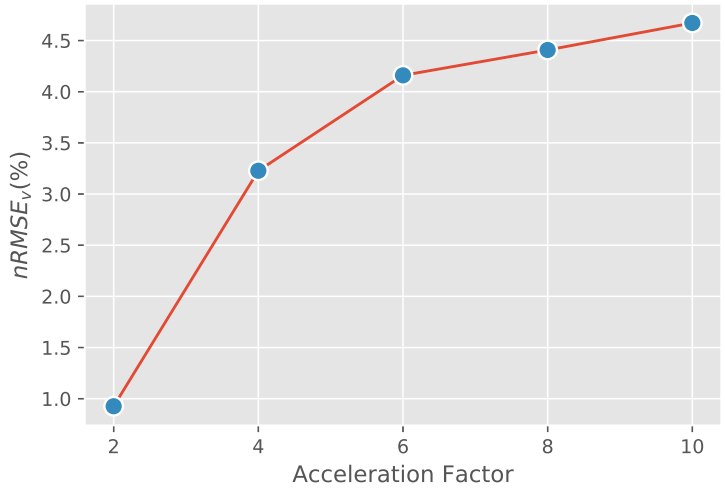
Using the sensitivity as a known value did not improve the quality of the reconstructed velocity field and reduced the convergence of the algorithm. This could be explained due to inaccuracies in the estimated coil sensitivities and phase inhomogeneties of the scanner's magnetic field. Fig. 2.5 shows an example of the resulting velocity after the same number of iterations using known and variable sensitivities. It can be seen in the highlighted areas that the resulting velocity field with estimated sensitivities presents inaccuracies that usually appear in non-convergent points in the optimization process.

2.4.4. Reconstruction quality

The errors incurred by the proposed approach for increasingly large acceleration factors over the phantom data are shown in Fig. 2.6. Magnitude errors stay below 4.5% for all acceleration factors up to 6, and directional errors are mostly negligible for the tested data. Fig. 2.7 compares the proposed method with the *frame-by-frame* Compressed Sensing and the *minimum-energy* reconstruction. It can be seen that the proposed method performs substantially better than the other two methods in both metrics. One interesting fact is that while the Compressed Sensing reconstruction shows lower errors in the velocity magnitude when compared to the *minimum-energy* reconstruction, it performs poorly in the directional



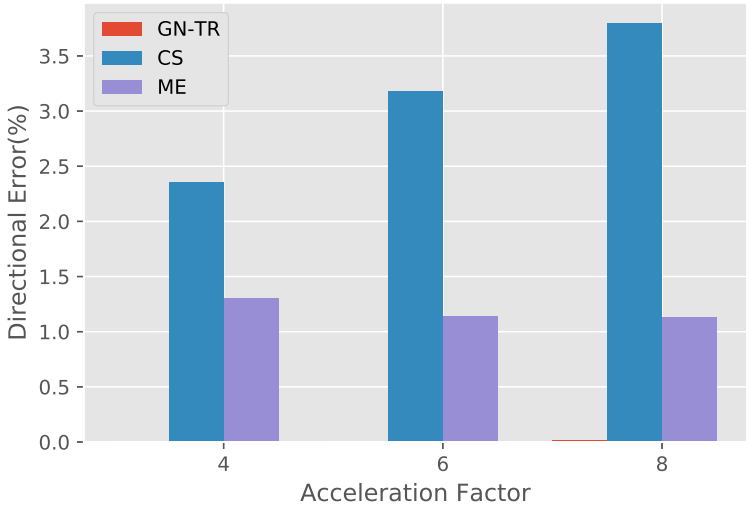
(a) Directional Error



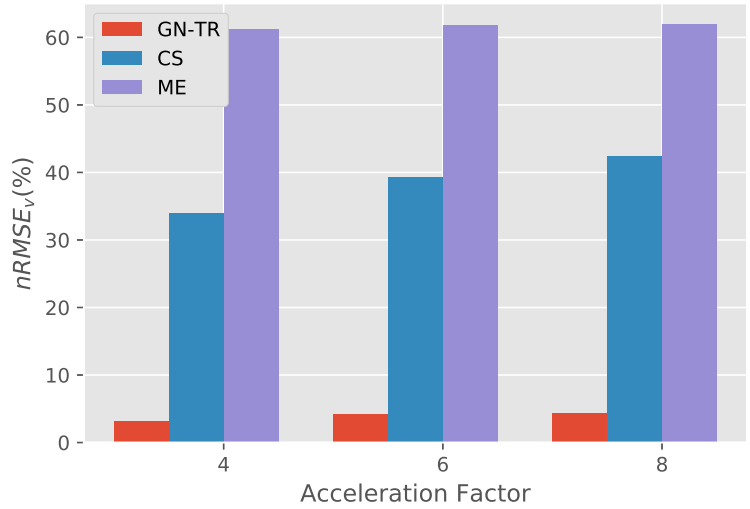
(b) nRMSEv

Figure 2.6. Error for different acceleration factors on the phantom data

error metrics when compared to the other methods. This result is consistent with the fact that a *frame-by-frame* reconstruction does not use information about the whole system and may incur in errors when using the differences between the reconstructed phases. In other words, in the Compressed Sensing procedure, no further properties are enforced over the reconstructed phases other than data consistency.



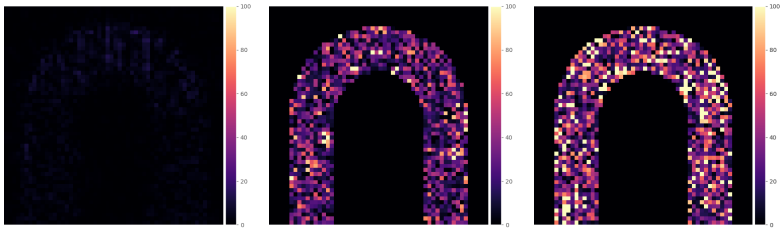
(a) Directional Error



(b) nRMSEv

Figure 2.7. Comparison of the proposed Gauss-Newton Trust-Region method (GN-TR), Compressed Sensing (CS) and Minimum Energy (ME) reconstructions

In Fig. 2.8a we see that the divergence of the velocity field reconstructed with the proposed approach is relatively small. On the other hand, in Fig. 2.8b we see that using frame-by-frame Compressed Sensing results in significantly larger values, by a factor of 10, for



(a) GN-TR (b) CS (c) ME

Figure 2.8. Divergence of the resulting velocity fields for different reconstruction algorithms.

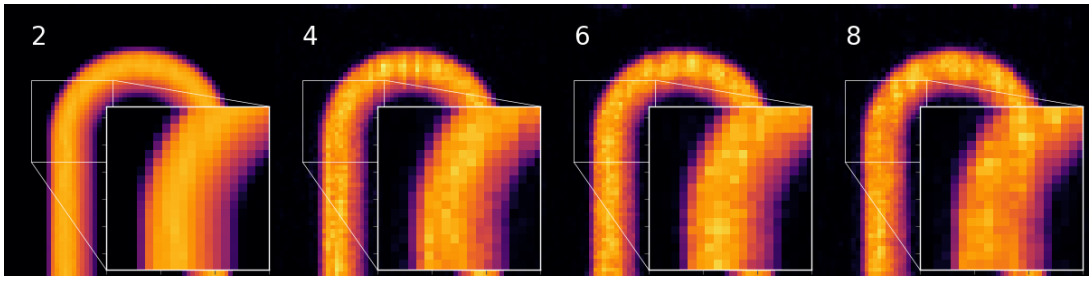


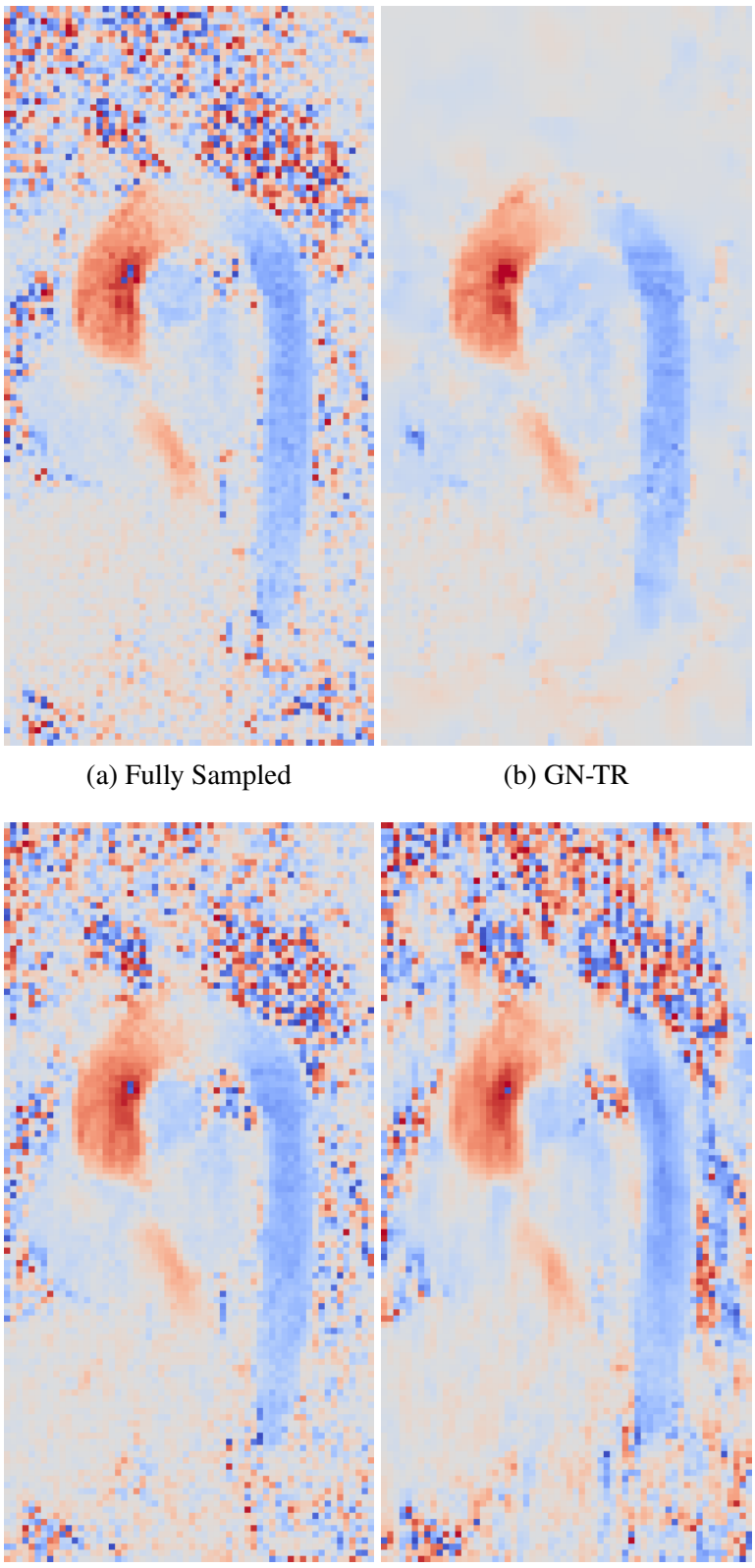
Figure 2.9. Reconstructed velocities of the phantom data for different acceleration rates. Magnitude of the velocity fields are shown

the divergence of the reconstructed velocity field. This shows that using the total-variation seminorm effectively controls the size of the divergence in the reconstructed velocity field, while producing smaller reconstruction errors, as discussed in Section 2.4.4.

Fig. 2.9 compares the reconstructed velocity magnitude for several acceleration rates on the phantom data. While the quality of the reconstructed velocity fields gets degraded for higher acceleration rates, the error measurements are kept low as shown in Fig. 2.6b.

Fig. 2.10 shows the foot-head velocity of an undersampled reconstruction for one of the *in vivo* data sets. It can be appreciated that structural information is correctly preserved in the undersampled reconstruction.

While errors generally remain low in the region of interest, areas in the boundary of vessels or with significant distortions due to motion do present higher velocity and magnitude errors.



(a) Fully Sampled (b) GN-TR

(c) CS (d) ME

Figure 2.10. Comparison between a fully sampled and different undersampled reconstruction algorithms in a foot-head velocity of an aorta using 25% of the acquired data.

2.5. Discussion

We have propose a novel method for the acceleration of 4D flow images, which has distinctive advantages as we discussed next:

(i) *A Gauss-Newton Trust Region method:* The 4d flow data acquisition process leads to a non-convex optimization problem that nonetheless has structure that can be exploited, namely, the objective is a sum of *convex functions* composed with *smooth non-linear maps*. This kind of problem is amenable to a Gauss-Newton trust region method, which considers at each *major* iteration a convex model which has to be minimized over a convex set. This allows us to use standard convex solvers to solve the trust region problem, while being able to leverage convergence results from trust region algorithms even when the objective might be non-smooth. In addition, our results show this method converges in practice faster than coordinate descent methods, and has the advantage of being a *descent method*, i.e., at each major iteration the algorithm decreases the objective value, ensuring the stationary point one obtains is the closest one to the initial point. This contrasts with the use of proximal methods directly on the non-convex problem, which generate sequences of iterates that may not decrease the objective value at each iteration.

(ii) *Joint magnitude, phase and coil sensitivity reconstruction:* Our approach has the advantage of modifying all the optimization variables at the same time without increasing the computational cost over coordinate descent methods. This allows us to easily perform joint reconstruction of the magnitude and phase, along with coil sensitivity estimation. Our results show that estimating the coil sensitivities while reconstructing the magnitude and phase images achieves a lower reconstruction error than simply using sensitivities estimated by standard method

(iii) *Total-variation and divergence regularization:* As mentioned in the introduction, flow incompressibility has been proposed as a structural property that can be promoted on the recovered velocity field. In practice, this reduces to penalizing the divergence of the recovered velocity field, which can be thought as a sparsifying transform for the phase. Instead of using the divergence directly, we propose to simply use the total-variation norm of the

phase. The reason for this is three-fold: (i) the total-variation norm controls the size of the divergence; (ii) computing the total-variation is computationally less demanding than computing the divergence using methods proposed in the literature (Deriaz & Perrier, 2006); (iii) the total-variation norm effectively removes the spurious fluctuations in the phase observed in practice. Our results show that the total-variation is an effective regularizer for 4d flow recovery problems.

Despite the advantages discussed above, our methodology also suffers from some limitations. First is the computational cost for a large 4D flow data sets. The number of variables in the problem, which depends directly on the size of the volume being reconstructed, is the most significant factor in the algorithm's execution speed, ranging from a few minutes for small two dimensional problems to up to 4 hours for the largest phantom experiments. We observe that if the volume has N pixels, with n_p encoding directions and n_c coils, the number of variables involved is of the order of $N(n_p + n_c + 1)$. Further research can be done into the formulation of the optimization problem by introducing auxiliary variables to build a convex relaxation of the original problem. This could benefit from convex optimization theory to improve the convergence of the algorithms. Second, we only test our algorithm in phantoms and in five volunteers, future work will apply our method to more volunteers or patients, in this work our purpose was to introduce the theory and its implementation.

3. CONCLUSIONS

Experimental results with the proposed method show it is possible to achieve 6-fold acceleration without incurring in significant error in the recovered velocity. The reconstruction technique shows that it is possible to jointly recover the optimization variables without compromising the reconstruction quality or the convergence of the algorithm.

The presented work provides a generalized method for non-linear magnetic resonance imaging reconstruction that allows one to think that further improvements can be done to improve the quality of undersampled data reconstruction for dynamic MR images.

Future research topics include building a convex relaxation formulation of the reconstruction model, where the required problem is indirectly solved by using additional variables that map the non-convex model into a higher dimensional convex manifold. This would allow a finer analysis on the convergence of the algorithms while maintaining, or potentially improving, the reconstruction quality.

Other regularization functions could be considered to improve the quality of the reconstruction that enforce known properties of the scanned region. One of such regularizers that have been previously used is some norm of the divergence of the velocity field, which assuming a non-compressible fluid, should be zero. While this could be considered a more reliable physically model, preliminary tests using a finite differences approximation for the divergence showed high numerical inestabilities that outweighted the benefits of the physical correctness.

The use of total variation over the reconstructed complex phases turned out to be a very convenient measure to control the divergence of the velocity field. This could use a deeper analysis and testing to find well defined bounds relative to divergence that could shed some light into new acquisition sequences to improve MR flow imaging.

The software itself that was developed to implement the recovery procedure and perform the documented experiments turned out to be a well featured optimization framework

capable of working with several convez optimization algorithms and some non-convex problems as well. It provides a great amount of flexibility to further extend the capabilities of the problems it can solve, while retaining a compact and robust core focused on efficiency. Further work is planned on improving the efficiency and ease of use of the software for future applications.

REFERENCES

Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, *2*(1), 183–202.

Bloch, F. (1946, October). Nuclear Induction. *Physical Review*, *70*(7-8), 460–474. Retrieved 2018-01-17, from `https://link.aps.org/doi/10.1103/PhysRev.70.460` doi: 10.1103/PhysRev.70.460

Burke, J. V., & Ferris, M. C. (1995, December). A Gauss—Newton method for convex composite optimization. *Mathematical Programming*, *71*(2), 179–194. Retrieved 2018-01-17, from `https://link.springer.com/article/10.1007/BF01585997` doi: 10.1007/BF01585997

Bydder, M., Larkman, D. J., & Hajnal, J. V. (2002, March). Combination of signals from array coils using image-based estimation of coil sensitivity profiles. *Magnetic Resonance in Medicine*, *47*(3), 539–548.

Candes, E., Romberg, J., & Tao, T. (2005, March). Stable Signal Recovery from Incomplete and Inaccurate Measurements. *arXiv:math/0503066*. Retrieved 2018-04-03, from `http://arxiv.org/abs/math/0503066` (arXiv: math/0503066)

Candès, E. J. (2014). Mathematics of sparsity (and a few other things). In *Proceedings of the International Congress of Mathematicians, Seoul, South Korea* (Vol. 123).

Candès, E. J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, *52*(2), 489–509.

Conn, A., Gould, N., & Toint, P. (2000). *Trust Region Methods*. Society for Industrial and Applied Mathematics. Retrieved 2018-01-08, from http://epubs.siam.org/doi/book/10.1137/1.9780898719857 doi: 10.1137/1.9780898719857

Deriaz, E., & Perrier, V. (2006). Divergence-free and curl-free wavelets in two dimensions and three dimensions: application to turbulent flows. *Journal of Turbulence*(7), N3.

Donoho, D. L. (2006a). Compressed sensing. *IEEE Transactions on information theory*, *52*(4), 1289–1306.

Donoho, D. L. (2006b, June). For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, *59*(6), 797–829. Retrieved 2018-01-18, from http://onlinelibrary.wiley.com/doi/10.1002/cpa.20132/abstract doi: 10.1002/cpa.20132

Frigo, M., & Johnson, S. G. (2005). The design and implementation of FFTW3. *Proceedings of the IEEE*, *93*(2), 216–231.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., … Haase, A. (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic resonance in medicine*, *47*(6), 1202–1210.

Hardy, G. H., Littlewood, J. E., & Pólya, G. (1952). *Inequalities*. Cambridge University Press.

Herráez, M. A., Burton, D. R., Lalor, M. J., & Gdeisat, M. A. (2002, December). Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Applied Optics*, *41*(35), 7437–7444. Retrieved 2018-01-17,

from https://www.osapublishing.org/abstract.cfm?uri=ao-41-35-7437 doi: 10.1364/AO.41.007437

Hestenes, M., & Stiefel, E. (1952, December). Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, *49*(6), 409–436.

Jaffard, S., Meyer, Y., & Ryan, R. (2001). *Wavelets*. Society for Industrial and Applied Mathematics. Retrieved 2018-01-17, from http://epubs.siam.org/doi/book/10.1137/1.9780898718119 doi: 10.1137/1.9780898718119

Kwak, Y., Nam, S., Akçakaya, M., Basha, T. A., Goddu, B., Manning, W. J., … Nezafat, R. (2013, September). Accelerated Aortic Flow Assessment with Compressed Sensing With and Without Use of the Sparsity of the Complex Difference Image. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *70*(3), 851–858. Retrieved 2017-12-23, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3604074/ doi: 10.1002/mrm.24514

Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*(242), 190–191.

Lewis, A. S., & Wright, S. J. (2016, July). A proximal method for composite minimization. *Mathematical Programming*, *158*(1-2), 501–546. Retrieved 2018-01-17, from https://link.springer.com/article/10.1007/s10107-015-0943-9 doi: 10.1007/s10107-015-0943-9

Liang, Z.-P., & Lauterbur, P. C. (1999). *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective* (1edition ed.). Bellingham, Wash. : New York: Wiley-IEEE Press.

Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic resonance in medicine*, *58*(6), 1182–1195.

Lustig, M., & Pauly, J. M. (2010, August). SPIRiT: Iterative Self-consistent Parallel Imaging Reconstruction from Arbitrary k-Space. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *64*(2), 457–471. Retrieved 2017-12-23, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2925465/` doi: 10.1002/mrm.22428

Markl, M. (2005). Velocity encoding and flow imaging. *University Hospital Freiburg, Dept of Diagnostic Radiology, Germany*.

Markl, M., Frydrychowicz, A., Kozerke, S., Hope, M., & Wieben, O. (2012). 4d flow MRI. *Journal of Magnetic Resonance Imaging*, *36*(5), 1015–1036.

Moran, P. R. (1982). A flow velocity zeugmatographic interlace for NMR imaging in humans. *Magnetic Resonance Imaging*, *1*(4), 197–203.

Moreau, J. (1965). Proximité et dualité dans un espace hilbertien. *Bulletin de la Société mathématique de France*, *79*, 273–299. Retrieved 2017-12-23, from `http://www.numdam.org/item?id=BSMF_1965__93__273_0` doi: 10.24033/bsmf.1625

Mura, J., Pino, A. M., Sotelo, J., Valverde, I., Tejos, C., Andia, M. E., … Uribe, S. (2016). Enhancing the Velocity Data From 4d Flow MR Images by Reducing its Divergence. *IEEE transactions on medical imaging*, *35*(10), 2353–2364.

Nesterov, Y. (1983). A method of solving a convex programming problem with convergence rate O (1/k2). In *Soviet Mathematics Doklady* (Vol. 27, pp. 372–376).

Nocedal, J., & Wright, S. (2006). *Numerical Optimization*. Springer New York.

Parikh, N., & Boyd, S. (2014, January). Proximal Algorithms. *Found. Trends Optim.*, *1*(3), 127–239. Retrieved 2018-01-18, from http://dx.doi.org/10.1561/2400000003 doi: 10.1561/2400000003

Pelc, N. J., Bernstein, M. A., Shimakawa, A., & Glover, G. H. (1991). Encoding strategies for three-direction phase-contrast MR imaging of flow. *Journal of Magnetic Resonance Imaging*, *1*(4), 405–413.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: sensitivity encoding for fast MRI. *Magnetic resonance in medicine*, *42*(5), 952–962.

Purcell, E. M., Torrey, H. C., & Pound, R. V. (1946, January). Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, *69*(1-2), 37–38. Retrieved 2018-01-18, from https://link.aps.org/doi/10.1103/PhysRev.69.37 doi: 10.1103/PhysRev.69.37

Santelli, C., Loecher, M., Busch, J., Wieben, O., Schaeffter, T., & Kozerke, S. (2016). Accelerating 4d flow MRI by exploiting vector field divergence regularization. *Magnetic resonance in medicine*, *75*(1), 115–125.

Sodickson, D. K., & Manning, W. J. (1997). Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, *38*(4), 591–603.

Sun, A., Zhao, B., Ma, K., Zhou, Z., He, L., Li, R., & Yuan, C. (2017). Accelerated phase contrast flow imaging with direct complex difference reconstruction. *Magnetic resonance in medicine*, *77*(3), 1036–1048.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Uecker, M., Lai, P., Murphy, M. J., Virtue, P., Elad, M., Pauly, J. M., … Lustig, M. (2014, March). ESPIRiT–an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, *71*(3), 990–1001. doi: 10.1002/mrm.24751

Uecker, M., & Lustig, M. (n.d.). *BART Toolbox for Computational Magnetic Resonance Imaging*. Retrieved from `https://mrirecon.github.io/bart/`

Uribe, S., Beerbaum, P., Sørensen, T. S., Rasmusson, A., Razavi, R., & Schaeffter, T. (2009, October). Four-dimensional (4d) flow of the whole heart and great vessels using real-time respiratory self-gating. *Magnetic Resonance in Medicine*, *62*(4), 984–992. doi: 10.1002/mrm.22090

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, *13*(2), 22–30.

Womersley, J. R. (1955). Method for the calculation of velocity, rate of flow and viscous drag in arteries when the pressure gradient is known. *The Journal of physiology*, *127*(3), 553–563.

Zhao, F., Noll, D. C., Nielsen, J.-F., & Fessler, J. A. (2012). Separate magnitude and phase regularization via compressed sensing. *IEEE transactions on medical imaging*, *31*(9), 1713–1723.

APPENDICES

A. DIFFERENTIAL FOR THE MEASUREMENT MAP

As stated in Section 2.2.2, the components of the data consistency map $\mathcal{T}$ are non-linear and defined as

$$\mathcal{T}_{p,c}(\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) = \mathcal{F}_{\Omega^{(p)}}(\boldsymbol{S} \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}\}\,).$$

Observe their form is independent of the coil. To find the differential for each of them, it will be useful to write it as follows

$$\begin{aligned} d\mathcal{T}_{p,c}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\delta\boldsymbol{S}, \delta\boldsymbol{m}, \delta\boldsymbol{\Phi}) =& D_{\boldsymbol{S}}\mathcal{T}_{p,c}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\delta\boldsymbol{S}) + \\ & D_{\boldsymbol{m}}\mathcal{T}_{p,c}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\delta\boldsymbol{m}) + \\ & D_{\boldsymbol{\Phi}}\mathcal{T}_{p,c}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\delta\boldsymbol{\Phi}). \end{aligned}$$

Recall that, by definition, a differentiable function f at x satisfies

$$f(x + \delta x) = f(x) + Df|_x(\delta x) + O(\|\delta x\|^2)$$

where $Df|_x$ is its differential at x. From the definition of $\mathcal{T}_{p,c}$ we can compute

$$\begin{aligned} &\mathcal{T}_{p,c}(\boldsymbol{S} + \delta\boldsymbol{S}, \boldsymbol{m}, \boldsymbol{\Phi}) \\ &= \mathcal{F}_{\Omega^{(p)}}((\boldsymbol{S} + \delta\boldsymbol{S}) \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}\}\,) \\ &= \underbrace{\mathcal{F}_{\Omega^{(p)}}(\boldsymbol{S} \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}\}\,)}_{\mathcal{T}(\boldsymbol{S}_c,\boldsymbol{m},\boldsymbol{\Phi})} + \underbrace{\mathcal{F}_{\Omega^{(p)}}(\delta\boldsymbol{S}_c \odot \boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}\}\,)}_{D_{\boldsymbol{S}}\mathcal{T}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\delta\boldsymbol{S}_c)} \end{aligned}$$

As magnitude and coil sensitivities are both linear over the operator, it is easy to see that

$$D_{\boldsymbol{m}}\mathcal{T}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\delta\boldsymbol{m}) = \mathcal{F}_{\Omega}(\boldsymbol{S} \odot \delta\boldsymbol{m} \odot \mathbf{exp}\{i\boldsymbol{\Phi}\}\,).$$

The phase operator comes from a similar analysis

$$
\begin{aligned}
&\mathcal{T}(\boldsymbol{S}+\boldsymbol{S}, \boldsymbol{m}, \delta\boldsymbol{\Phi}) \\
&= \mathcal{F}_\Omega(\boldsymbol{S} \odot \boldsymbol{m} \odot \textbf{exp}\{i(\boldsymbol{\Phi}+\delta\boldsymbol{\Phi})\}\,) \\
&= \mathcal{F}_\Omega(\boldsymbol{S} \odot \boldsymbol{m} \odot \textbf{exp}\{i\boldsymbol{\Phi}\} \odot \textbf{exp}\{i\delta\boldsymbol{\Phi}\}\,) \\
&= \mathcal{F}_\Omega(\boldsymbol{S} \odot \boldsymbol{m} \odot \textbf{exp}\{i\boldsymbol{\Phi}\} \odot (1 + i\delta\boldsymbol{\Phi} + O(\|i\delta\boldsymbol{\Phi}\|^2)) \\
&= O(\|i\delta\boldsymbol{\Phi}\|^2) + \\
&\underbrace{\mathcal{F}_\Omega(\boldsymbol{S} \odot \boldsymbol{m} \odot \textbf{exp}\{i\boldsymbol{\Phi}\}\,)}_{\mathcal{T}(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})} + \underbrace{\mathcal{F}_\Omega(\boldsymbol{S} \odot \boldsymbol{m} \odot \textbf{exp}\{i\boldsymbol{\Phi}\} \odot i\delta\boldsymbol{\Phi})}_{D_{\boldsymbol{\Phi}}\mathcal{T}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\delta\boldsymbol{\Phi})}
\end{aligned}
$$

Now, using the fact that for every linear operator A, its adoint operator must satisfy $\langle y,\, Ax\rangle = \langle A^* y,\, x\rangle$ for every x, y and properties of the inner product, the adjoint operator for $d\mathcal{T}$ can be found as

$$
\begin{aligned}
\langle \boldsymbol{y},\, D_{\boldsymbol{S}}\mathcal{T}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{S})}(\delta\boldsymbol{S})\rangle &= \langle \boldsymbol{y},\, \mathcal{F}_\Omega(\delta\boldsymbol{S} \odot \boldsymbol{m} \odot \textbf{exp}\{i\boldsymbol{\Phi}\}\,)\rangle \\
&= \langle \mathcal{F}_\Omega^*(\boldsymbol{y}),\, \delta\boldsymbol{S} \odot \boldsymbol{m} \odot \textbf{exp}\{i\boldsymbol{\Phi}\}\,\rangle \\
&= \langle \mathcal{F}_\Omega^*(\boldsymbol{y}) \odot (\boldsymbol{m} \odot \textbf{exp}\{i\boldsymbol{\Phi}\}\,)^*,\, \delta\boldsymbol{S}\rangle \\
&= \langle \underbrace{\mathcal{F}_\Omega^*(\boldsymbol{y}) \odot \boldsymbol{m} \odot \textbf{exp}\{-i\boldsymbol{\Phi}\}}_{D_{\boldsymbol{S}}\mathcal{T}^*|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\boldsymbol{y})},\, \delta\boldsymbol{S}\rangle
\end{aligned}
$$

The adjoint of the magnitude and phase differential operators can be found in essentially the same way:

$$\begin{aligned}
\langle \boldsymbol{y},\, D_{\boldsymbol{m}}\mathcal{T}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{S})}(\delta\boldsymbol{m})\rangle &= \langle \boldsymbol{y},\, \mathcal{F}_\Omega(\boldsymbol{S}\odot\delta\boldsymbol{m}\odot\ \mathbf{exp}\{i\boldsymbol{\Phi}\}\,)\rangle \\
&= \langle \mathcal{F}_\Omega^*(\boldsymbol{y}),\, \boldsymbol{S}\odot\delta\boldsymbol{m}\odot\ \mathbf{exp}\{i\boldsymbol{\Phi}\}\,\rangle \\
&= \langle \mathcal{F}_\Omega^*(\boldsymbol{y})\odot(\boldsymbol{S}\odot\ \mathbf{exp}\{i\boldsymbol{\Phi}\}\,)^*,\, \delta\boldsymbol{m}\rangle \\
&= \langle \underbrace{\mathcal{F}_\Omega^*(\boldsymbol{y})\odot\boldsymbol{S}\odot\ \mathbf{exp}\{-i\boldsymbol{\Phi}\}}_{D_{\boldsymbol{m}}\mathcal{T}^*|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\boldsymbol{y})},\, \delta\boldsymbol{m}\rangle
\end{aligned}$$

and

$$\begin{aligned}
\langle \boldsymbol{y},\, D_{\boldsymbol{\Phi}}\mathcal{T}|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{S})}(\delta\boldsymbol{\Phi})\rangle &= \langle \boldsymbol{y},\, i\mathcal{F}_\Omega(\boldsymbol{S}\odot\boldsymbol{m}\odot\ \mathbf{exp}\{i\boldsymbol{\Phi}\odot\delta\boldsymbol{\Phi}\}\,)\rangle \\
&= \langle -i\mathcal{F}_\Omega^*(\boldsymbol{y}),\, \boldsymbol{S}\odot\boldsymbol{m}\odot\ \mathbf{exp}\{i\boldsymbol{\Phi}\}\ \odot\delta\boldsymbol{\Phi}\rangle \\
&= \langle -i\mathcal{F}_\Omega^*(\boldsymbol{y})\odot(\boldsymbol{S}\odot\boldsymbol{m}\odot\ \mathbf{exp}\{i\boldsymbol{\Phi}\}\,)^*,\, \delta\boldsymbol{\Phi}\rangle \\
&= \langle \underbrace{-i\mathcal{F}_\Omega^*(\boldsymbol{y})\odot\boldsymbol{S}\odot\boldsymbol{m}\odot\ \mathbf{exp}\{-i\boldsymbol{\Phi}\}}_{D_{\boldsymbol{\Phi}}\mathcal{T}^*|_{(\boldsymbol{S},\boldsymbol{m},\boldsymbol{\Phi})}(\boldsymbol{y})},\, \delta\boldsymbol{\Phi}\rangle
\end{aligned}$$