Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Raya Horesh	Date

NUMERICAL OPTIMIZATION FOR TRANSPORT AND REGISTRATION PROBLEMS

By

Raya Horesh
Doctor of Philosophy

Mathematics and Computer Science

Eldad Haber, Ph.D.
Advisor

James G. Nagy, Ph.D.
Committee Member

Alessandro Veneziani, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

NUMERICAL OPTIMIZATION FOR TRANSPORT AND REGISTRATION PROBLEMS

by

Raya Horesh
Bachelor of Science

Advisor: Eldad Haber, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy
in Mathematics and Computer Science
2010

Abstract

NUMERICAL OPTIMIZATION FOR
TRANSPORT AND REGISTRATION PROBLEMS
By Raya Horesh

In this thesis we develop numerical methods for the solution of large-scale PDE-based constrained optimization problems. Overall three studies are presented; the first two are application driven, addressing volume preserving image registration and optimal mass transport problems. The third study is more generic and embarks at the development of a new inexact sequential quadratic programming framework. Image registration aims at finding a plausible transformation which aligns images taken at different times, different view-points or by different modalities. This problem is ill-posed and therefore, regularization is required. In that study, elastic regularizer is considered along with volume preserving constraint. A numerical framework based on augmented Lagrangian along with geometrical multigrid preconditioner was devised. The proposed algorithm was tested with real data. The optimal mass transport seeks for an optimal way to move a pile of soil from one site to another using minimal energy, while preserving the overall mass. In that study, a fluid dynamics formulation was considered. This formulation introduces an artificial time stepping, which on the one hand transforms the non-convex problem to a convex one, but on the other hand increases the dimensionality of the problem. A Schur complement and algebraic multigrid formed a preconditioner within a sequential quadratic programming scheme. Results for both three and four dimensional problems were presented. Inside each step of nonlinear optimization, solution for an ill-conditioned, indefinite linear system, known as a KKT system is required. As problem size increases, linear iterative solvers become the bottleneck of the optimization scheme. In the third study, a new approach for inexact step computation is proposed. The general idea is to reduce the number of linear iteration while still maintaining convergence of the overall scheme. This is done, by the embedment of a filter inside a linear solver.

NUMERICAL OPTIMIZATION FOR TRANSPORT AND REGISTRATION PROBLEMS

by

Raya Horesh
Bachelor of Science

Advisor: Eldad Haber, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy
in Mathematics and Computer Science
2010

Acknowledgments

I would like to express my gratitude to my advisor Prof. Eldad Haber for his support and guidance throughout my graduate school experience. His enthusiasm for research and confidence in his students makes him a truly special mentor. Thank you for giving me such a great opportunity. I extend my gratitude to my committee members Prof. James Nagy and Prof. Alessandro Veneziani for sharing their knowledge via their classes and constructive criticisms throughout the course of my research. In addition, I would like to deeply thank Andrew Conn for his sorrow review and his insightful remarks and suggestions.

To my parents, brother and all the immediate and extended family that always had me in their thoughts, regardless of the distance that separated us. Thank you for all your love and support. I know it has been difficult to understand what it is that I do, or what this process is like. I cherish your support.

And of course, no words could capture my appreciation to my husband and friend, Lior, who inspires me everyday to be the best person I can be. Thanks for always listening when I need to talk and for the encouraging words that made such a difference in my journey here. This thesis could not be possible without you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Optimal transport and registration problems are part of a large group of problems, namely inverse problems. Those problems arise in many important applications, including medical imaging, geophysics and economics. The basic idea in inverse problems is to recover the model from some knowledge of the geometry of the problem and from a given measurement data, which in many cases is noisy. In this thesis we will develop numerical methods that are applicable for real life applications, in particular such that involve partial differential equations (PDEs). Often, after discretization, these problems result in large amount of data and unknowns that need to be processed. Thus, special numerical treatment and algorithms need to be developed and implemented for large-scale inverse problems. These algorithms need to be chosen carefully and take into account stable discretizations of the PDEs involved [1]. Moreover, a discretized problem of such a nature is usually ill-conditioned, and requires a suitable and often problem dependent preconditioner. Specifically, we will consider optimal transport and registration problems, where the unknowns to be recovered are the transformation in two or three dimensions

[1] Stability in the context of discretization is further discussed in Chapter 3, Section 2.2 and in Chapter 4, Section 3.

and the density distribution of the images.

1 Image Registration - Problem Statement

The goal of image registration is to *align* two related images by finding a *feasible* and *plausible* transformation. Given two images, $T, R \in \Omega$, we want to find a transformation $u : \mathbb{R}^d \to \mathbb{R}^d$ that transforms T such that R and $T(u)$ are similar on Ω, where d is the dimension of the image. In this setting, the image R which we call the reference image is fixed and image T which we refer to as the template image is deformed. We will first define some basic concepts in image registration.

1.1 An Image

Throughout this thesis, we will consider d-dimensional images, I, which are functions with compact support on domain Ω that maps any spatial point $x \in \mathbb{R}^d$ to a scalar value $I(x) \in \mathbb{R}$. Specifically, if we consider the two dimensional case ($d = 2$), then each such point is called a *pixel* and in the three dimensional case ($d = 3$), it is called *voxel*.

1.2 Distance Measure

A distance measure, sometimes referred to as the image similarity, quantifies the degree of similarity between two or more images. This measure strongly depends on the modalities of the images to be registered. For images taken from the same device, the sum of square differences (SSD) is commonly used (a formal definition is given in Chapter 3, Section 2.1) . Other, more sophisticated approaches, like mutual information or normalized gradient fields, are used in the case of multi-modality registration [80]. We should notice that we require the distance measure to be symmetric, i.e. the distance

from R to T should be the same as the distance from T to R. In general, there is more than one transformation that minimizes the distance, that is, the problem is *ill-posed*. A *well-posed* problem, as was defined by Jacques Hadamard in [56], should have the following three properties: 1. A solution exists, 2. The solution is unique and 3. The solution depends continuously on the data. A problem that does not satisfy one or more of those properties is considered as an ill-posed problem. Inverse problems are usually ill-posed. We should notice, that even if the problem is well-posed, it can still be ill-conditioned. That is, a small error in the data can result in much larger errors in the solution.

1.3 Feasible Transformation

The definition of a *feasible* or *reasonable* transformation is application dependent. As mentioned before, image registration is an ill-posed problem, and therefore, more than one solution exists. In order to choose the most suitable transformation for the specific application at hand we need to reduce our solution space. This can be done by introducing constraints to our optimization problem. There are mainly two ways to introduce a priori information or assumptions regarding our problem. The first is done by introducing a regularization term, by adding an additional term to the objective functional, which is considered as a soft constraint and the second way is by introducing a hard constraint to the problem. There are many popular choices for choosing the regularization term. For example, in some applications, transformations should not result in a folding of the grid or transformations should be volume preserving, see for example [54, 92]. In general, we can divide methods for the registration problem into two approaches, parametric and non-parametric methods. In parametric registration, the registration transformation is given by an explicit model that depends on some parameters. In Section 1.3.1 and

1.3.2 we will introduce common parametric and non-parametric registration methods.

1.3.1 Parametric Registration

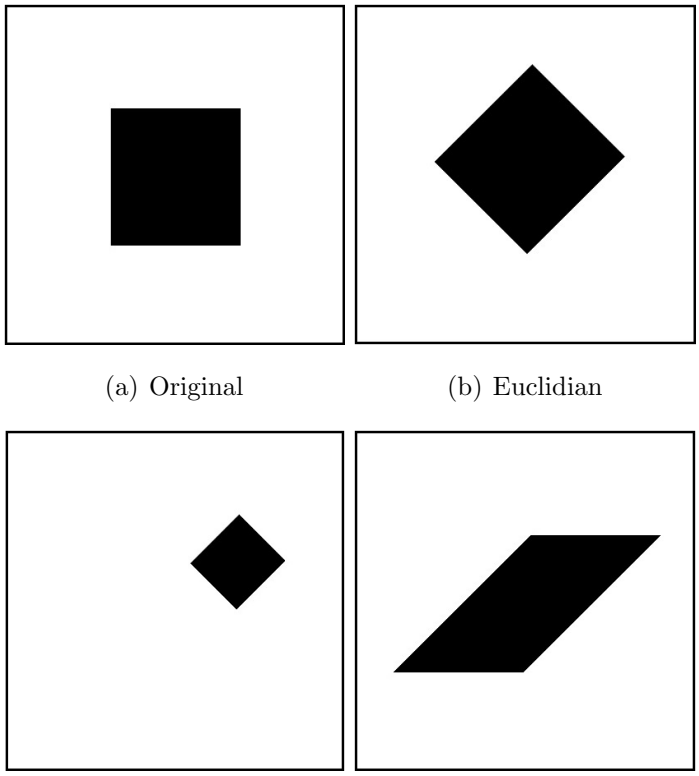
In parametric registration, the problem is restricted to finding real values for a small number of parameters. We can represent the problem mathematically as finding the transformation $u : \mathbb{R}^d \rightarrow \mathbb{R}^d$, which can be expressed as following

$$u(x) = Qx + b$$

where $Q \in \mathbb{R}^{d \times d}$ is an orthogonal matrix and $b \in \mathbb{R}^d$. The common examples are rigid, similarity and affine transformations [93, 110]. Rigid or Euclidian transformations preserves angles and length of line segments and allow only translations and rotations, see Figure 1.1(b). For two dimensional images, a Euclidian transformation has three parameters (θ, t_i, t_j), where θ is the rotation angle and t_i, t_j are translations in the i and j direction, correspondingly. In a similarity transformation, besides the translations and rotations, a uniform isotropic scaling is also allowed. In this case, the objects can become smaller or bigger while presenting their original shape, see Figure 1.1(c). An affine transformations is a more general but still a linear parametric method, which allows for rotation, translation, scaling and shearing. Affine transformation preserves parallelism between lines, see Figure 1.1(d).

1.3.2 Non-parametric Registration

Parametric transformations are considered as global mapping models, since we are interested in parameters that are valid for the entire image. However, global transformations cannot properly handle images which are deformed locally. Least square approaches average the local geometrical deformations over the entire image, which is sometimes not desirable, and moreover, it does



(a) Original (b) Euclidian

(c) Similarity (d) Affine

Figure 1.1: Different types of parametric registrations.

not adequately take into account noise in the images. Information regarding possible models of local geometrical distortion should be incorporated into the registration framework. In non-parametric registration, the solution space is reduced by introducing a regularization term S. The regularized non-parametric registration problem is defined as follows: given two images R, T, a distance measure D and a regularization functional, find a displacement $u : \mathbb{R}^d \to \mathbb{R}^d$ such that it minimizes

$$D(R, T(u)) + \alpha S(u), \ \alpha > 0. \tag{1.1}$$

The parameter α is a regularization parameter that weights the similarity of the images versus the smoothness of the transformation. Choosing small α will give more weight to the distance term while the smoothing term will have little influence on the solution. On the other hand, large α will result in much smoother solutions. There are many algorithms and methods proposed in the literature for choosing a suitable regularization parameter. For example methods such as the discrepancy principle [81], generalized cross-validation (GCV) [49] and L-curve [59, 72] have all been used with varying degrees of success [78]. Practically, none of the aforementioned methods have consistently provided successful results when utilized for the image registration problem. Choosing a regularization parameter for this problem is indeed a non-trivial task. In practice, ad-hoc methods are commonly exercised for this purpose.

Choosing the appropriate regularization functional is application dependent. Many smoothers were suggested throughout the years, such as the diffusive smoother [39], the curvature smoother [40], which were both proposed by Fischer and Modersitzki, and the elastic smoother. In this thesis, the commonly used elastic registration will be considered.

In elastic registration, which was introduced by Bajcsy et al. [7], the images are viewed as pieces of a rubber sheet, on which external forces (registration)

are stretching the image and internal forces defined by stiffness or smoothness (and referred to as Lamé constants) are forcing to bring them into alignment with the minimal amount of bending and stretching. The registration is achieved by locating the minimum energy state in an iterative fashion. The elastic regularization is given by the elastic potential function

$$S(u) := \frac{1}{2} \int_{\Omega} \theta \ (\nabla u : \nabla u) + (\theta + \tau)(\text{div } u)^2 \ dx. \tag{1.2}$$

Here $(A : B) := \sum_{i,j} A_{i,j} B_{i,j}$ denotes a matrix inner product, and θ and τ are the Lamé constants.

The solution of (1.1) is still not unique, and moreover, the transformation might not be feasible or desirable for some applications. For example, one may desire to have parts of the image rigid [75, 76, 69, 79, 52] or maintain a correspondence between so-called landmarks [67, 41, 62]. Nevertheless, an often fundamental requirement in almost all medical imaging applications is that volumes do not shrink below some application dependent threshold. Vanishing volumes are generally not physical in medical imaging and are thus to be avoided. In [55], a method that bounds volume changes has been proposed. This is achieved by constraining shrinkage/expansion of each individual cell in a discretized formulation of the problem. The approach leads to a large scale and highly non-linear inequality constrained optimization problem, where the number of constraints are at the same order as the number of unknowns.

Similar to [55], in this thesis we constrain the feasibility of a transformation by

$$\kappa_m(x) \leq C_{\text{vol}}[u](x) \leq \kappa_M(x) \quad \text{for all} \quad x \in \Omega_\kappa, \quad C_{\text{vol}}[u] := \det(I + \nabla u), \tag{1.3}$$

where the functions κ_m and κ_M and the subset $\Omega_\kappa \subset \Omega$ are supposed to be user supplied and appropriately chosen for a particular application.

In summary, the desired displacement u is a minimizer of the constrained optimization problem

$$\min\ D[u] + \alpha\ S[u] \quad \text{subject to} \quad \kappa_m \leq C_{\text{vol}}[u] \leq \kappa_M. \tag{1.4}$$

We address this problem in more details in Chapter 3.

2 Optimal Mass Transport

2.1 Monge-Kantorovich Problem

The original optimal transport problem concerns with finding the best way to move a pile of soil to an excavation, with the least amount of work.

Let Ω_0 and Ω_T be two diffeomorphic connected subdomains of $\mathbb{R}^d$, and let T, R be Borel measures on Ω_0 and Ω_T, each with a strictly positive density function $T(x) \geq \rho^0_{\text{low}} > 0$ and $R(x) \geq \rho^\tau_{\text{low}} > 0$, respectively and assume that

$$\int_{\Omega_0} T(r)dr = \int_{\Omega_1} R(r)dr = 1.$$

The problem is concerned with finding a mapping that minimizes the transformation cost function, which represents the physical work required for mass transport

$$M(u) = \int_\Omega T(x)|u(x)|^2\, dx \tag{1.5}$$

subject to

$$\int_{u^{-1}(\Omega)} T(x) = \int_\Omega R(x)dx. \tag{1.6}$$

Equation (1.6) is a weak formulation of the Jacobian equation, where the determinant term is added due to coordinates change

$$\det(I_d + \nabla u)T(x + u(x)) = R(x). \tag{1.7}$$

In equation (1.5) the integrand gives the traveled distance squared, weighted by the amount of the transported mass. We can view the cost function (1.5)

as the total amount of effort required to move a pile of mass from one location to another. The L^2 cost has been extensively used in many different applications such as mathematical finance, shape recognition in image processing, computer vision, signal processing and functional analysis (see book [105] and references within).

We can now write our constrained optimization problem for the general L^p cost function, where $1 \leq p \leq 2$ as

$$\begin{aligned} \min \quad & M(u) := \int_{\Omega} T(x)|u(x)|^p \, dx && \text{(1.8a)} \\ \text{s.t.} \quad & c(u) = \det(I_d + \nabla u)T(x + u(x)) - R(x) = 0, && \text{(1.8b)} \end{aligned}$$

where u is a $C^{1,\alpha}$ diffeomorphism from $\Omega_0 \to \Omega_1$. The constraint $c(u) = 0$ (the Jacobian equation) is often referred to as the mass preserving (MP) property.

Even with a simple, quadratic distance function, this is a highly non-linear equality constrained optimization problem. There is extensive analysis as for the existence, uniqueness, and properties of the solution (see for example [1, 32, 105] and the references therein). However, as of today, there are only a small number of papers that deal with finding the solution of the problem, and an even smaller number of papers that deal with efficient *numerical* solutions of the problem [27, 22, 9, 2, 87, 28]. Generally speaking, numerical methods for the solution of the problem can be divided into three approaches.

1. The first approach, for the case $p = 2$, utilizes the property that the transport map can be written as a gradient of convex potential function Ψ, i.e.

   $$u(x) = \nabla\Psi(x)$$

   substituting this into equation (1.7) results in the Monge-Ampère equation

   $$(I_d + H\Psi(x))T(\nabla\Psi(x)) = R(x),$$

where $H\Psi(x)$ is the Hessian of the matrix Ψ. This is a second order non-linear partial differential equation (PDE).

2. The second approach attempts to tackle the constrained optimization problem (4.1) directly. This direct numerical approach raises many numerical challenges, see for example [100].

3. A third approach for the solution of the problem was proposed in the seminal paper of Benamou and Brenier (BB) [9], which suggested a fluid mechanic formulation of the problem.

In this thesis we concentrate on this formulation of the problem, introduce it in Section 2.2 and discuss the numerical solution in Chapter 4.

2.2 Fluid Dynamics Framework for the Monge-Kantorovich Problem

In this section we summarize the fluid dynamics formulation of the L^2 Monge-Kantorovich framework, as was presented in [9]. Let us affix the time interval $[0, \tau]$, and consider all possible sufficiently smooth time-dependent density and velocity field $\rho(x,t) > 0$, $v(x,t)$ $x \in \mathbb{R}^d$, subject to the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \tag{1.9}$$

for $0 \leq t \leq \tau$ and the initial and final conditions

$$\rho(x,0) = T(x), \ \rho(x,\tau) = R(x).$$

Then, we can rewrite the constrained problem as

$$\begin{aligned} \min \quad & \frac{1}{\tau} \int_\Omega \int_0^\tau \rho(x,t)|v(x,t)|^2 \, dxdt & \text{(1.10a)} \\ \text{s.t.} \quad & \frac{\partial \rho}{\partial t} + \nabla(\rho v) = 0 & \text{(1.10b)} \\ & \rho(x,0) = T(x), \ \rho(x,\tau) = R(x). & \text{(1.10c)} \end{aligned}$$

These types of problems are referred to as a PDE constrained optimization problems and their solution was recently addressed in the literature (see for example, [37, 11, 14] and references within).

While methods from PDE optimization can be used here, there are some obvious limitations. The main one is that PDE optimization problems typically assume that it is possible to solve the PDE without the optimization assuming that the control is known. However, for problem (1.10) this is not the case. Note that there are no boundary conditions (BC) for ρ in the PDE and thus one cannot solve uniquely for ρ even if v is known.

3 Inexact Sequential Quadratic Programming

Sequential Quadratic Programming (SQP) is one of the most popular and robust methods for non-linear constrained optimization problems of the form

$$\begin{aligned} \min \quad & f(x) && \text{(1.11a)} \\ \text{s.t.} \quad & c(x) = 0. && \text{(1.11b)} \end{aligned}$$

where $x \in \mathbb{R}^d$, and both the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint $c : \mathbb{R}^n \to \mathbb{R}^m$ are sufficiently smooth. For simplicity, we consider equality constrained optimization problems, although the method is suitable also for problems with inequality constraints.

The underlying principle of the method is based on solution of a series of subproblems designed to minimize a quadratic model of the objective function, subject to a set of constraints. This approach can be viewed as a natural extension of the Newton method for constrained optimization problems.

The Lagrangian of the system can be written as

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^\top c(x)$$

From here, 1^{st} order necessary optimality conditions for stationarity are given by

$$\begin{aligned} \mathcal{L}_x &= g(x) + B(x)^\top \lambda = 0 \\ \mathcal{L}_\lambda &= c(x) = 0, \end{aligned}$$

where $B(x) := \nabla c(x)$ is the Jacobian of the constraints, $g(x) := \nabla f(x)$ is the gradient of the objective function, and λ is a Lagrange multipliers vector.

Using the Newton method on the Karush-Kuhn-Tucker (KKT) conditions yields a linear system whose solution provides a Newton search direction

$$\begin{pmatrix} H_k & B_k^\top \\ B_k & 0 \end{pmatrix} \begin{pmatrix} d_k \\ \delta_k \end{pmatrix} = - \begin{pmatrix} g_k + B_k^\top \lambda_k \\ c_k \end{pmatrix}, \tag{1.12}$$

where H_k is the Hessian $\nabla^2_{xx}\mathcal{L}(x_k, \lambda_k)$ or its symmetric approximation. The new iterate is updated sequentially as follows

$$\begin{aligned} x_{k+1} &= x_k + \alpha d_k \\ \lambda_{k+1} &= \lambda_k + \alpha \delta_k \end{aligned}$$

where α is obtained through a line search procedure.

Thus, at each SQP iteration, we need to solve the linear system (1.12). The KKT system is symmetric and indefinite. Typically for PDE - based problems this system is large-scale and sparse. In cases where the dimensionality and or density of the problem prohibits the use of direct linear solvers for the aforementioned system, some Krylov subspace methods may guarantee a solution within a bounded number of iterations (equal to the dimension of the system in exact arithmetic). Nevertheless, for large linear systems, such an upper limit on the number of iterations is often impractical.

Until recently, available theoretical work prompted for strict satisfaction of the above conditions (1.12), that is, search directions were obtained through

an exhaustively accurate solution of the linearized system. Due to the computational load associated with such solutions, and the underlying intuition that exact solutions of the linearized systems may be somewhat redundant, while the current estimate of the solution is far from the true one, practitioners were driven to consider compromising with inexact computation of the search directions. A family of associated methods are the truncated Newton scheme [83]. Yet, only very recent work by Byrd et al. [19, 20] has retrospectively substantiated this claim and provided an explicit convergence proof.

The fundamental question that arises for an inexact SQP approach is the extent to which inexactness can be allowed? Or more formally and practically, how can one form inexact search directions that will guarantee convergence of the overall optimization scheme?

Byrd et al. work introduced one measure for validating this requirement. Their approach was based on sufficient reduction in a model of an exact penalty function. Inspired by the notion of employing filter methods as an alternative to a merit function, as firstly introduced by Leyffer and Fletcher, in this thesis, we propose the use of a filter-method based measure serving as a decision-maker in the course of an inexact search direction computation.

Filter methods avoid using a merit function which combines the objective and constraint violation into single function. Instead, it considers (1.11) as a bi-objective optimization problem that minimizes $f(x)$ and $h(x) := \|c(x)\|$ for some norm. The concept of domination is borrowed from multi-objective optimization, which defines that x_k dominates point x_l if and only if $f(x_k) \leq f(x_l)$ and $h(x_k) \leq h(h_l)$. The filter is defined as a list of pairs $(h(x_l), f(x_l))$ such that no pair dominates another pair. An illustration of a typical filter is presented in Figure 1.2, where the shaded area represents the region which is dominated by the filter entries. A detailed explanation about the filter is

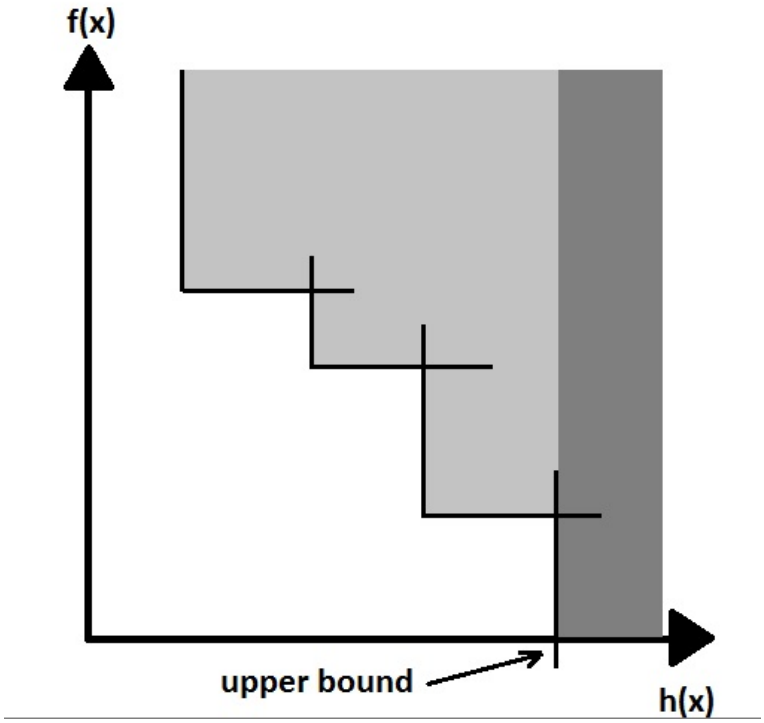


Figure 1.2: Filter method - shaded area represent area that is dominated by the filter entries.

given in Chapter 5.

4 Contributions of this Thesis

The overall purpose of this thesis is to develop efficient numerical optimization methods for large scale PDE-based constrained optimization problems. In particular, the thesis focuses on the challenging problems of registration and optimal transport. More specifically, this thesis presents important contributions in each of the following areas:

1. **Volume Preserving Image Registration**
   A PDE constrained optimization formulation for image registration is considered in this thesis. This formulation involves a highly non-linear PDE which requires special care when discretized. The discrete problem is large-scale and requires special numerical tools. In this work, we solve the problem using the Augmented Lagrangian method, which results in a symmetric positive definite linear system that needs to be solved at each iteration. For that purpose, we use a preconditioned conjugate gradient approach with a geometric multi-grid preconditioner. We develop a new relaxation scheme, which is based on Vanka's method [103] and black-white relaxation, which is generalized to three dimensional domains.

2. **Optimal Mass Transport**
   We consider the fluid mechanic formulation of the Monge-Kantorovich problem, proposed by Benamou and Brenier, as presented in equation (1.10). We reformulated the problem by introducing time-space divergence, where we treat the time as a spatial dimension. The new formulation resembles non-linear flow in porous media. Such problems are well posed as long as the objective function is convex, even without

boundary conditions on the density distribution. We developed robust and stable discretizations on a staggered grid along with an efficient optimization algorithm and an efficient solver for the linear system. We use a Newton-type scheme for the solution of the constrained optimization problem. In particular, we consider a sequential quadratic programming (SQP) method for the solution of the problem, with an inexact step computation of the linearized system using the GMRES method. The resulting KKT system is ill-conditioned and the use of suitable preconditioner is required. We consider a preconditioner based on a Schur complement. The Schur complement in this case is a Laplacian in space-time with possible jumping coefficients, that depend on the density distribution. For the solution of the Schur complement part, we consider the preconditioned conjugate gradient method, with algebraic multi-grid as a preconditoner, which takes into account non-smooth coefficients.

3. **Inexact Filter Based SQP Method**
   Inexact SQP methods are very important for large-scale optimization problems, where the arising linear systems need to be solved using iterative linear algebra algorithms, such as generalized minimal residual (GMRES). In this thesis, we prescribe and characterize a stopping criterion for the iterative linear solver, based on a filter method, such that the computed step makes sufficient progress towards the solution of the linearized problem. To guarantee convergence of the original problem, we incorporate a line search strategy by applying a filter over the non-linear problem. We extend the filter to three dimensions, where the third dimension is checking the dual feasibility term. This work has been embedded successfully within the optimization framework of the optimal mass transport problem mentioned above. The proposed framework, only requires mild traditional assumptions regarding the

problem. Thus, it is applicable to a broad range of constrained optimization problems enabling great versatility and universality of use.

Chapter 2

Background for Relevant Numerical Tools

1 Preface

In this chapter a glossary of relevant numerical tools and terms is introduced. The review begins by a description of some of the fundamental principles of non-linear constrained optimization. In this section, first and second order necessary and sufficient conditions for optimality are elucidated. Further, three popular numerical optimization algorithms that facilitates these conditions are described: the Sequential Quadratic Programming (SQP) which is utilized for solving the optimal mass transport problem (Chapter 4), the Log Barrier approach, which was used in some studies for addressing the aforementioned problem, and the Augmented Lagrangian (AL), which serves as the method of choice for volume preserving image registration (Chapter 3). We later briefly review globalization by means of a line search. This topic, as well as SQP, shall be critically revisited in the inexact SQP chapter (5). The solution of these problems mandates an intermediate solution of a linear system of equations. We shall therefore proceed with a com-

prehensive description of stationary and non-stationary methods as well as multi-grid approaches for the solution of such systems. Fixed-point methods are typically incorporated as preconditioners or smoothers within multi-grid formulations. The latter use is introduced in Chapter 3 within a geometric multi-grid framework for solution of the linear system and in Chapter 4 as part of an algebraic multi-grid preconditioner. We spend some time in describing Krylov subspace linear solvers, and in particular, the Generalized Minimal Residual (GMRES) solver [96, 97]. A modification of this algorithm for incorporation of a filter, as a termination condition will be presented in the inexact SQP chapter 5.

2 Constrained Optimization Problem - Definition

Constrained optimization is defined as minimization (or maximization) of a function subject to a set of constraints acting on some variables. In this chapter, we adopt the common notation used in the finite dimensional constrained optimization literature. Let x be a vector of variables, f be an objective function and c be a vector of constraints. First, we should notice that $\min f(x) = -\max(-f(x))$, thus, the two problems are equivalent and we therefore only consider the minimization problem in this thesis.

A general optimization problem can be written as

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & c_i(x) = 0, \ i \in \mathcal{E} \\ & c_i(x) \geq 0, \ i \in \mathcal{I} \end{aligned} \tag{2.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function, $c : \mathbb{R}^n \to \mathbb{R}^m$ are a set of constraints, that can either be equality ($\mathcal{E}$) or inequality ($\mathcal{I}$) or both. Any point

x that satisfies the constraint is said to be feasible, or otherwise, infeasible.

3 Notation

We first introduce important notation that will be used throughout this thesis. Let $x \in \mathbb{R}^n$ be a column vector with x_i representing its i-th component. We further assume that the function $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

Definition 1 *The gradient of the objective function is defined as*

$$g(x) = \nabla_x f(x)$$

where the i-th component of the vector of the first derivatives $\nabla_x f(x)$ is $\partial f(x)/\partial x_i$. We consider the gradient to be a column vector.

Definition 2 *The Hessian of the objective function is defined as*

$$H(x) = \nabla_{xx} f(x)$$

where the $[i, j]$-th component of the Hessian, is the second derivative

$$\partial^2 f(x)/\partial x_i \partial x_j$$

. One should notice that the Hessian is a symmetric matrix.

Definition 3 *The Jacobian matrix of the constraints is defined as*

$$B(x) = (\nabla_x c(x))^\top$$

Definition 4 *The active set $\mathcal{A}(x)$ at any feasible point x is defined as*

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} | c_i(x) = 0\}$$

Definition 5 *(LICQ) Given the point x and the active set $A(x)$, we say that the linear independence constraint qualification (LICQ) holds if the set of active constraint gradients $\{\nabla c_i(x),\ i \in \mathcal{A}(x)\}$ is linearly independent.*

4 Optimality Conditions for Equality Constrained Optimization

We can now define the Lagrangian function for the equality constrained optimization problem (2.1) with $\mathcal{I} = \emptyset$.

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^\top c(x),$$

where λ is a vector of Lagrange multipliers.

Necessary optimality conditions for equality constrained minimization problem are given below.

Theorem 1 *First Order Necessary Conditions (Karush-Kuhn-Tucker conditions or KKT conditions). If x^* is a local minimizer of $f(x)$ subject to $c(x) = 0$ and f and c are continuously differentiable, then as long as LICQ holds, there exists a vector of Lagrange multipliers λ^*, such that*

$$\begin{aligned} c(x^*) &= 0 \text{ (primal feasibility)} \\ g(x^*) + B^\top(x^*)\lambda^* &= 0 \text{ (dual feasibility)} \end{aligned} \tag{2.2}$$

Theorem 2 *Second Order Necessary Conditions. If x^* is a local minimizer of $f(x)$ subject to $c(x) = 0$ and $\nabla^2 f$ and $\nabla^2 c$ are continuous, then as long as LICQ holds, there exists a vector of Lagrange multipliers λ^*, such that $c(x^*) = 0$, $g(x^*) + B^\top(x^*)\lambda^* = 0$ and*

$$s^\top H(x^*, \lambda^*)s \geq 0, \ \forall s \in \mathcal{N},$$

where

$$\mathcal{N} = \{s \in \mathbb{R}^n | B(x^*)s = 0\}.$$

Theorem 3 *Second Order Sufficient Conditions. If $\nabla^2 f$ and $\nabla^2 c$ are continuous and there exists a vector of Lagrange multipliers λ^*, such that $c(x^*) = 0$, $g(x^*) + B^\top(x^*)\lambda^* = 0$ and*

$$s^\top H(x^*, \lambda^*)s > 0, \ \forall s \in \mathcal{N}$$

where

$$\mathcal{N} = \{s \in \mathbb{R}^n | B(x^*)s = 0\},$$

then x^* is a strict local minimizer of $f(x)$ subject to $c(x) = 0$.

5 Optimality Conditions for Inequality Constrained Optimization

Suppose x^* is a minimizer of the inequality constrained optimization problem (2.1) with $\mathcal{E} = \emptyset$.

Theorem 4 *First Order Necessary Conditions (KKT conditions). If x^* is a local minimizer of $f(x)$ subject to $c(x) \geq 0$ and f and c are continuously differentiable, then as long as LICQ holds, there exists a vector of Lagrange multipliers λ^*, such that*

$$\begin{aligned} c(x^*) &\geq 0 \ \textit{(primal feasibility)} \\ g(x^*) + B^\top(x^*)\lambda^* &= 0 \ \textit{(dual feasibility)} \\ c(x^*) \odot \lambda^* &= 0 \ \textit{(complementary slackness)} \\ \lambda^* &\geq 0 \end{aligned} \tag{2.3}$$

where $\odot$ denotes the Hadamard (or Schur) product.
These conditions are often referred to as the KKT conditions.

Theorem 5 *Second Order Necessary Conditions. If x^* is a local minimizer of $f(x)$ subject to $c(x) \geq 0$ and $\nabla^2 f$ and $\nabla^2 c$ are continuous, then as long as LICQ holds, there exists a vector of Lagrange multipliers λ^*, such that primal and dual feasibility and slackness hold, together with*

$$s^\top H(x^*, \lambda^*)s \geq 0, \ \forall s \in \mathcal{N}_+$$

where

$$s \in \mathcal{N}_+ \Leftrightarrow \begin{cases} \nabla c_i(x^*)^\top s = 0 & \textit{if } c_i(x^*) = 0 \ \& \ \lambda_i^* > 0 \\ \nabla c_i(x^*)^\top s \geq 0 & \textit{if } c_i(x^*) = 0 \ \& \ \lambda_i^* = 0 \end{cases}$$

Theorem 6 *Second Order Sufficient Conditions. If* $\nabla^2 f$ *and* $\nabla^2 c$ *are continuous and there exists a vector of Lagrange multipliers* λ^**, such that* $c(x^*) = 0$*,* $g(x^*) + B^\top(x^*)\lambda^* = 0$ *and*

$$s^\top H(x^*, \lambda^*)s > 0, \ \forall s \in \mathcal{N}_+$$

where

$$s \in \mathcal{N}_+ \Leftrightarrow \begin{cases} \nabla c_i(x^*)^\top s = 0 & \text{if } c_i(x^*) = 0 \ \& \ \lambda_i^* > 0 \\ \nabla c_i(x^*)^\top s \geq 0 & \text{if } c_i(x^*) = 0 \ \& \ \lambda_i^* = 0 \end{cases}$$

then x^* *is a strict local minimizer of* $f(x)$ *subject to* $c(x) \geq 0$*.*

The interested reader may refer to [86] and reference therein.

6 Sequential Quadratic Programming

The basic idea of sequential quadratic programming (SQP) is analogous to Newton's method for unconstrained optimization, where we apply a Newton step to the KKT optimality conditions. In the Quadratic Programming (QP) method, we introduce a quadratic model for the objective function and a linear model of the constraints. In the SQP method, a QP is solved at each iteration. We consider an equality constrained minimization problem

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ \text{subject to} & c(x) = 0. \end{array} \tag{2.4}$$

We can now write the first order KKT conditions for the equality constrained problem (2.4) as a system of $n + m$ equations for $n + m$ unknowns x and λ.

$$F(x, \lambda) = \begin{bmatrix} g(x) + B(x)^\top \lambda \\ c(x) \end{bmatrix} = 0. \tag{2.5}$$

The Jacobian of the system (2.5) with respect to x and λ is then

$$\begin{bmatrix} \nabla^2_{xx} \mathcal{L}(x, \lambda) & B(x)^\top \\ B(x) & 0 \end{bmatrix}. \tag{2.6}$$

To obtain the Newton step, we need to solve the following linear system

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & B(x)^\top \\ B(x) & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g(x) + B(x)^\top \lambda \\ c(x) \end{bmatrix} \tag{2.7}$$

and update the iterate (x_k, λ_k)

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} + \begin{bmatrix} d_k \\ \delta_k \end{bmatrix}. \tag{2.8}$$

If the KKT matrix (2.6) is non-singular, the Newton iteration is well defined. This system is not singular if the following assumptions hold

Assumptions 1

- *(A1) The constraint Jacobian matrix $B(x)$ has full row rank (LICQ).*
- *(A2) The matrix $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ is bounded and second order sufficient, i.e. $s^\top \nabla^2_{xx}\mathcal{L}(x,\lambda)s \geq k\|s\|^2$, $\forall s \neq 0$ with $B(x)s = 0$*

Under these assumptions, Newton's iteration is quadratically convergent, given that the starting point is sufficiently close to x^*.

7 Line Search

Line search methods decide how far we need to traverse along a descent direction d_k and δ_k. The iteration update is then given by

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} + \alpha_k \begin{bmatrix} d_k \\ \delta_k \end{bmatrix}. \tag{2.9}$$

where α_k is a positive scalar, namely the step length. The optimal choice of α will be such that minimizes the merit function $\phi(x_k + \alpha d_k)$ (for unconstrained

case, that is the objective function itself). However, in general this procedure is prohibitively expensive computationally, since it requires too many function evaluations. In practice, an inexact line search is often preformed, such that sufficient reduction in the function's value is obtained. This is typically done by trying out a sequence of values for α and accepting the value for which certain conditions are satisfied. Different line search termination criteria can be considered. The simplest one requires that $\phi(x_k + \alpha d_k) < \phi(x_k)$, but this requirement is not sufficient to guarantee convergence to x^*. Another popular choice is the sufficient decrease or the Armijo condition, which requires that

$$\phi(x_k + \alpha d_k) \leq \phi(x_k) + c_1 \alpha \nabla \phi_k^\top d_k, \quad 0 < c_1 < 1 \tag{2.10}$$

where in practice, $c_1 = 10^{-4}$. Nevertheless, this condition can be satisfied for any small value of α, which may lead to slow convergence. For that purpose, it is advisable to introduce another condition regarding the curvature

$$\nabla \phi(x_k + \alpha d_k)^\top \geq c_2 \nabla \phi_k^\top d_k, \; c_1 < c_2 < 1. \tag{2.11}$$

These two conditions ((2.10) and (2.11)) collectively are known as the Wolfe conditions. Other conditions also exist, but they will not be discussed in this thesis. A common choice of a merit function is the following l_1 function

$$\phi(x_k, \lambda_k) = f(x_k) + \gamma \|c(x_k)\|_1$$

where γ is typically chosen to be $1/\|\lambda_k\|_\infty$. An alternative to the line search procedure for globalization is the family of trust region methods [26]. The underlying principal is to define a region in which a linear or a quadratic model is regarded as trustworthy. A search point is accepted whenever a reasonable match between the predicted reduction and the actual reduction is satisfied. This measure also steers the magnitude of the trusted region: increased for a good match, sustained for a reasonable match and decreased for a poor one.

8 Log-Barrier Method

The log-barrier was first introduced by Frisch [48] in 1955. This methods aims at solving an inequality constrained optimization problem, in the form

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & c(x) \geq 0. \end{aligned} \tag{2.12}$$

The log-barrier function is defined as

$$\mathcal{P}(x, \mu) = f(x) - \mu \sum_{i \in \mathcal{I}} \log c_i(x), \ \mu > 0 \tag{2.13}$$

Under certain conditions, the minimizer of (2.13) approaches x^* as μ approaches 0. In principal, this problem can be solved by means of unconstrained optimization. This method has several drawbacks. One important drawback of this method versus the SQP method is that the function (2.13) become increasingly non-linear as μ is reduced. In addition, the initial starting point has to be feasible, which is a non-trivial requirement in many cases. Moreover, this method results in an ill-conditioned Hessian, as μ approaches 0.

9 Augmented Lagrangian Method

The augmented Lagrangian method is categorized as a penalty method, as a penalty term is added to the Lagrangian function. Let us consider an equality constrained optimization problem, as in (2.4). The augmented Lagrangian is then defined as follows

$$\mathcal{L}(x, \lambda, \mu) = f(x) - \lambda^\top c(x) + \frac{\mu}{2} \|c(x)\|^2 \tag{2.14}$$

The augmented Lagrangian differs from the standard Lagrangian by the rightmost squared term. The KKT condition mandate that $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$

and $c(x^*) = 0$, therefore, the augmented Lagrangian and the standard Lagrangian coincide at the minimum for any μ. At some non-critical point, the optimality conditions are

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = g(x) - B(x)^\top \lambda + \mu B(x)^\top c(x) = 0.$$

From here, at the optimal solution, we obtain that $\lambda^* = \lambda - \mu c(x)$. This formula suggests a way to update the Lagrange multiplier at each iteration

$$\lambda_i^{k+1} = \lambda_i^k - \mu^k c_i(x^k), \tag{2.15}$$

where k denotes the iteration index and i denotes a component index of a vector. Minimization of the augmented Lagrangian results in an addition of a stabilization term to the $(1, 1)$ block of the KKT system, where $\nabla^2 f(x)$ is replaced by $\nabla^2 f(x) + \mu B(x)^\top B(x)$.

It is possible to extend the augmented Lagrangian method to inequality constrained optimization by introducing slack parameters. We can rewrite the inequality constraint $c_i(x) \geq 0,\ i \in \mathcal{I}$ as

$$c_i(x) - s_i = 0,\ s_i \geq 0.$$

It is now possible to recast the inequality constrained optimization problem as an equality constrained problem plus a set of non-negativity constraints.

$$\begin{aligned} \min_{x,s} \quad & f(x) \\ \text{subject to} \quad & c_i(x) = 0,\ i \in \mathcal{E} \\ & c_i(x) - s_i = 0,\ i \in \mathcal{I} \\ & s \geq 0 \end{aligned} \tag{2.16}$$

This optimization problem can be solved efficiently by using the gradient projection method, for a fixed λ and μ. This approach was successfully utilized in the LANCELOT package [25].

10 Stationary Linear Solvers

Stationary iterative methods aim at solving the linear system of equations

$$Ax = b, \tag{2.17}$$

where the update iteration can be expressed in the following form

$$x^k = Bx^{k-1} + c.$$

Note that both B and c do not depend on the iteration count k. In this section we discuss two stationary methods: Jacobi and Gauss-Seidel.

10.1 Jacobi

If we decompose our matrix A into three matrices: D, $-L$ and $-U$, where D is the diagonal, $-U$ is the strictly upper triangular part, and $-L$ is the strictly lower triangular part of A, then we can write

$$(D - U - L)x = b$$

which can be used to define a fixed point iteration of the form

$$x^k = D^{-1}(U + L)x^{k-1} + D^{-1}b.$$

We should notice, that since D is a diagonal matrix, its inverse can be computed easily assuming positive positive definiteness of matrix A. The method will always converge if the matrix A is strictly diagonally dominant. A second convergence condition is when the spectral radius of the iteration matrix $\rho(D^{-1}(U + L)) < 1$.

10.2 Gauss - Seidel

Considering again the linear system in (2.17), as before we can decompose the matrix A into three matrices D, $-L$ and $-U$ as before. The Gauss-Seidel

iteration is therefore as following

$$x^k = (D - L)^{-1}(Ux^{k-1} + b).$$

Though it can be applied to any matrix with non-zero elements on the diagonals, convergence is only guaranteed if the matrix is either diagonally dominant, or symmetric and positive definite. For many cases the convergence of the Gauss-Seidel method is faster than the Jacobi method. Others stationary methods also exist, such as successive over relaxation (SOR), symmetric successive over relaxation (SSOR) and weighted Jacobi. For further reading please refer to [104, 57, 109].

11 Non-stationary Iterative Methods - Krylov Subspace Methods

In this section, we introduce some of Krylov subspace methods [96] for the solution of the linear system of equations (2.17). Let x_0 be the initial solution guess, we can define the initial residual as

$$r_0 = b - Ax_0.$$

When considering Krylov subspace methods, the kth iterate satisfies

$$x_k \in x_0 + \mathcal{K}_k(A, r_0),\ k = 1, 2, \ldots$$

where

$$\mathcal{K}_k(A, r_0) := \text{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\}$$

is kth Krylov subspace, spanned by an A-orthogonal basis. Two vectors v_i, v_j are defined to be A-orthogonal if $v_i^\top Av_j = 0,\ i \neq j$. The Krylov subspace family of solvers are among the most effective and efficient for large and sparse systems of equations. These methods are highly economical with

respect to memory. Most of these methods employ the Lanczos method for orthogonalization, which for symmetric matrices can generate subsequent search directions, using only a three-term vector occurrence. Moreover, an explicit formulation of the matrix A is not compulsory, as long as some mechanism that provides the matrix-vector product exists.

11.1 Conjugate Directions

Solving a linear system of equations (2.17), where $A \in \mathbb{R}^{n \times n}$ is symmetric positive definite (SPD), is equivalent to minimization of a convex quadratic form function

$$f(x) = \frac{1}{2} x^\top A x - b^\top x + c \tag{2.18}$$

since at the minimum, we obtain

$$\frac{\partial f(x)}{\partial x} = Ax - b = 0 \Longleftrightarrow Ax = b.$$

For linear systems, the gradient of the minimized function is equal to the residual error $r(x)$

$$r(x) = Ax - b = \nabla f(x).$$

The general iterative step is given by

$$x_{k+1} = x_k + \alpha_k p_k \tag{2.19}$$

By substitution of (2.19) into (2.18) we obtain the minimized function at the $\{k+1\}$-th iteration

$$f(x_{k+1}) = \frac{1}{2}(x_k + \alpha_k p_k)^\top A (x_k + \alpha_k p_k) - b^\top (x_k + \alpha_k p_k) + c$$

The one-dimensional step-size α_k, a minimizer of this quadratic function along $x_k + \alpha_k p_k$, can be found by differentiation with respect to α_k

$$\frac{\partial f(x_{k+1})}{\partial \alpha_k} = (Ax_k - b)^\top p_k + \alpha_k p_k^\top A p_k = r_k^\top p_k + \alpha_k p_k^\top A p_k = 0$$

therefore, we obtain

$$\alpha_k = -\frac{r_k^\top p_k}{p_k^\top A p_k}.$$

A set of A-conjugate vectors are linearly independent, hence, once conjugate direction $(p_0, p_1, \ldots, p_n)$ has been selected as a search direction at each iteration, f would be minimized within, at most n steps [86]. This important property can be proved by expressing the error term as a linear combination of the search directions

$$e_0 = x_0 - x = \sum_{i=0}^{n-1} \beta_i p_i$$

where the β_i values can be determined by pre-multiplication of the error term by $p_k^\top A$

$$p_k^\top A e_0 = \sum_{i=0}^{n-1} \beta_i p_k^\top A p_i = \beta_k p_k^\top A p_k.$$

This leads to

$$\beta_k = \frac{p_k^\top A e_0}{p_k^\top A p_k} = \frac{p_k^\top A(e_0 + \sum_{i=0}^{k-1} \alpha_i p_i)}{p_k^\top A p_k} = \frac{p_k^\top A e_k}{p_k^\top A p_k}.$$

Since $A e_k = r_k$, it is follows that

$$\beta_k = \frac{p_k^\top r_k}{p_k^\top A p_k} = -\alpha_k.$$

From here, it is apparent that the process of constructing the solution x component-wise, is equivalent to the process of eliminating the error term component by component

$$e_k = e_0 + \sum_{i=0}^{k-1} \alpha_i p_i = \sum_{i=0}^{n-1} \beta_i p_i - \sum_{i=0}^{k-1} \beta_i p_i = \sum_{i=k}^{n-1} \beta_i p_i$$

After n iterations, all the components are deducted, and $e_n = 0$. The main flaw of this method is that construction of any new search direction via a

conjugate Gram-Schmidt process requires retainment of all previous search directions (as is the case for Arnoldi methods, defined on page 34). Furthermore, this approach requires $\mathcal{O}(n^3)$ operations for generation of the full set.

11.2 Conjugate Gradient (CG)

The underlying idea of Conjugate Gradients is almost similar to the Conjugate Directions method, only that now, the search directions are constructed by conjugation of the residuals. The use of the residual is, in some sense, inspired from the Steepest Descent method, and motivated by the fact that the residual is orthogonal to the previous search direction. Therefore, the first search direction is chosen to be the steepest descent direction $-\nabla f(x_k) = -r_k$.

Since the construction of any new search direction does not require explicit elimination of conjugated components from previous search directions, the typical runtime complexity per iteration of this algorithm is reduced to $O(nnz(A))$, where $nnz(A)$ stands for the number of non-zero entries in the matrix A. The convergence is determined by the spectrum of the eigenvalues, with the following relation for the relative error

$$\frac{\|e_k\|_A}{\|e_o\|_A} \leq \min_{q_k} \max_{1 \leq i \leq n} |q_k(\lambda_i)|$$

where q_k is polynomial of degree k satisfying $q_k(0) = 1$ and $\lambda_i, 1 \leq i \leq n$ stands for the set of eigenvalues of $A \in \mathbb{R}^{n \times n}$. The number of required iterations for reduction of the error norm by a factor of ϵ is given by

$$k \leq \left\lceil \frac{\sqrt{\kappa}}{2} \ln \frac{1}{\epsilon} \right\rceil$$

where κ is the condition number. It is therefore evident that the performance of the conjugate gradient algorithm can be improved by application of a

suitable preconditioner, further discussion on that regard is given in Chapters 3 and 4.

11.3 Generalized Minimal Residual Method (GMRES)

The Generalised Minimal Residual method (GMRES) [97] is an extension of the Minimal Residual method (MinRes), which is only applicable to symmetric systems, to unsymmetric systems. At iteration k an approximation x_k of the Krylov subspace is generated by the vector c

$$\mathcal{K}_k(A, c) = \text{span}\{c, Ac, \ldots, A^{k-1}c\}$$

where typically $c = b$. This method solves the least square problem

$$\min_{z \in \mathcal{K}_k(A,b)} \|b - Az\|$$

by constructing an orthonormal basis $V = \{v_1, v_2, \ldots, v_k\}$ for $\mathcal{K}_k(A, c)$ using Arnoldi's method. Starting with $v_1 = b/\|b\|$ the basis for $\mathcal{K}_{k+1}(A, b)$ is constructed recusivly

$$v_{k+1} = \frac{\hat{v_{k+1}}}{\|\hat{v_{k+1}}\|},$$

where

$$v_{k+1} = Av_k + \sum_{i=1}^{k} (v_i^\top A v_k) v_i.$$

This then leads to the following decomposition

$$AV_k = V_{k+1} L_k,$$

where $L_k \in \mathbb{R}^{(k+1)\times k}$ is an upper Hessenberg matrix. The vector z can be represented now as $z = V_k w$ for some w, and therefore

$$Az = V_{k+1} L_k w$$

and

$$b = \|b\| V_{k+1} e_1$$

where e_1 denotes the first column of the identity matrix. The least squares problem is then reduced to

$$\min_{w_k} \|(\|b\|)e_1 - L_k w_k\|$$

$$z_k = V_k w_k$$

It can be observed that in order to generate the next basis vector, A is only accessed in the form of Av_k. The major drawback of the GMRES method is that the computational load and storage required per iteration rises linearly with the iteration count. The customary way to overcome this limitation is by restarting the iteration periodically. After a predefined number of iterations $k_{restart}$, the accumulated data is cleared and the intermediate results are used as the initial data for the next $k_{restart}$ iterations. This procedure is repeated until convergence is achieved. The difficulty is in choosing an appropriate value for $k_{restart}$. If $k_{restart}$ is too small, GMRES may converge slowly, or fail to converge entirely. A value of $k_{restart}$ that is larger than necessary involves excessive work and storage. Unfortunately, there are no general, definite rules for the choice of $k_{restart}$.

Algorithm 1: Arnoldi Method

Input: A, q_1 such that $\|q_1\| = 1$

Output: $q_1, q_2, q_3, \ldots$

for $j = 1, 2, \ldots$ **do**

- $\hat{q}_{k+1} \leftarrow Aq_k$
- **for** $i = 1, 2, \ldots, j$ **do**
  - $h_{ij} \leftarrow \hat{q}_{j+1}^{\top} q_i$
  - $\hat{q}_{j+1} \leftarrow \hat{q}_{j+1} - h_{ij} q_i$
- $h_{j+1,j} \leftarrow \|\hat{q}_{j+1}\|$
- $q_{j+1} \leftarrow \hat{q}_{j+1}/h_{j+1,j}$

Algorithm 2: GMRES Algorithm

Input: A, b, x_0

Output: x_k

set $r_0 \leftarrow b - Ax_0$

set $q_1 \leftarrow r_0/\|r_0\|$

for $j = 1, 2, \ldots$ **do**

  Compute q_{k+1} and $h_{i,k}, i = 1, 2, \ldots, k+1$ using Arnoldi method

  Solve the least square problem $\min_y \|\beta e_1 - A_{k+1,k}y\|$ to find y_k, where $\beta = \|r_0\|$

  Set $x_k \leftarrow x_0 + Q_k y_k$ where $Q_k \in \mathbb{R}^{n\times k}$ has q_i as columns

The above algorithm will be revisited and modified in Chapter 5 to accommodate inexact step computation within the SQP framework. Despite the popularity of the GMRES solver for solution of large-scale non-symmetric linear systems, its reliance on long recurrences, often shadows its overall appeal. As an alternative, a set of short recurrence Krylov space solvers exists. To name some, restarted GMRES (discussed earlier in this section), bi-conjugated gradient (Bi-CG) and stabilized bi-conjugated gradients (Bi-CGSTAB). These methods will not be discussed here, the reader should refer for further reading [102]. Studies show that each of these algorithms may outperform the others for a given system. Therefore, there are no definitive guidelines for favoring one upon the other. Nevertheless, typically much greater improvement can be achieved by an appropriate choice of a preconditioner.

12 Multi-grid

Multi-grid methods are designed to solve systems with optimal time complexity. Specifically, the number of operations needed to solve a problem scales linearly with the size of the problem. Although the multi-grid methodol-

ogy may be applied to linear and non-linear equations, in this thesis we restrict our attention to the linear case $Ax = b$. The fundamental principle of multi-grid is the relationship between relaxation and coarse-grid correction. When applied to a linear system, a stationary iterative method (e.g. Jacobi or Gauss-Seidel iteration) rapidly reduces high-frequency or oscillatory errors in the approximate solution. After several applications of the relaxation procedure the approximate solution consists primarily of low frequency or slow-to-converge error components that are not effectively reduced by further relaxation. However, when these components are restricted to a coarser grid they become oscillatory and are damped by relaxation applied to the coarser problem. The approximate solution on the coarse grid is then interpolated back to the finer grid to update the fine-grid approximate solution. This procedure is applied recursively to form a hierarchy of grids until the coarsest problem can be solved exactly. The processes of relaxation and coarse-grid correction comprise the multi-grid cycle. The efficiency of the multi-grid cycle depends on the effectiveness of relaxation on oscillatory error modes and the accuracy with which slow-to-converge error components are represented on coarser grids. For an algorithmic example of a V-cycle geometric multi-grid, see algorithm 3 below.

Algorithm 3: Multi-grid V-cycle algorithm

multi-grid-Vcycle(A, b, x_0)
if $l = N$ **then**
  return Solve(A_l, b_l)
else
  $x_l \leftarrow \text{PreSmooth}(A_l, b_l, x_l)$
  $r_l \leftarrow b_l - A_l x_l$
  $b_{l+1} \leftarrow \text{Restrict}(r_l)$
  $A_{l+1} \leftarrow \text{Restrict}(A_l)$
  $x_{l+1} \leftarrow 0$
  $x_{l+1} \leftarrow \text{multi-grid-Vcycle}(A_{l+1}, b_{l+1}, x_{l+1})$
  $x_l \leftarrow x_l + \text{Interpolate}(x_{l+1})$
  $x_l \leftarrow \text{PostSmooth}(A_l, b_l, x_l)$
  return x_l

Multi-grid methods can be classified into two types. The geometric multi-grid method [101] uses a hierarchy of explicitly built meshes. Coarser grids are determined by a predefined geometric simplification. Restriction and interpolation operators are defined in a manner consistent with the coarsening of the domain. Often, the coarse level equations correspond to a coarser discretizations of the initial (continuous) problem. The second group of methods avoid the geometrical complexities associated with the concurrent handling of a hierarchy of meshes and builds the multi-grid procedure purely on algebraic concepts. Unlike the geometrical multi-grid, where the hierarchy of meshes and the prolongation operators are defined from rediscretization, the algebraic multi-grid (AMG) [95] approach tries to construct the hierarchy of coarse spaces and the prolongation operators only from matrix data, making assumptions on the underlying differential equation and its discretization. While using multi-grid as a solver over classical elliptic PDE problems is possible, it is most effective as a preconditioner for an iterative solver.

Chapter 3

Volume Preserving Image Registration

1 Introduction

Image registration is commonly used in image guided medical applications such as treatment planing, radiation therapy, surgery planing, and many more; see, e.g. [58, 108, 78, 35, 98, 77, 34, 73, 43, 42] and references therein. The goal of image registration is to *align* two related images by finding a *feasible* and *plausible* transformation. To achieve this goal, mathematical definitions of the components *aligned*, *feasible*, and *plausible* are required. Since image registration is an ill-posed and under-determined problem [31, 78], these components are the building blocks of any registration approach. By choosing them carefully it is possible to dramatically reduce the inherent non-uniqueness in the problem and obtain reliable results.

Alignment is often quantified by the distance between two images. For images taken from the same device, the sum of square differences (SSD) is commonly used [18]. Other, more sophisticated approaches, like mutual information or normalized gradient fields, are used in the case of multi-modality

registration; see [80] and references within. In this study, the SSD is chosen as a template for a general distance measure but could be substituted by any other suitable approach. The definition of *feasible* and *plausible* or *reasonable* are mostly application dependent. Reasonable transformations typically have small norm and the choice of an appropriate norm distinguishes between different registration techniques. Choosing a criterion for a reasonable transformation leads to a particular regularization technique. For example, diffusion [39], elastic [6], and curvature [40] regularizations have all been used with varying degrees of success [78]. In this study, the commonly used elastic registration serves as a template for regularization but could be replaced by any other of the aforementioned regularization operators. Choosing the set of feasible transformations is also problem dependent. For example, one may desire to have parts of the image rigid [75, 76, 69, 79, 52] or maintain a correspondence between so-called landmarks [67, 41, 62]. Nevertheless, an often fundamental requirement in almost all medical imaging applications is that volumes do not shrink below some application dependent threshold. Vanishing volumes imply shocks, which are generally non-physical in medical imaging and therefore should be avoided. In [55], a method that bounds volume changes has been proposed. That was achieved by constraining shrinkage/expansion of each individual cell in a discretized formulation of the problem. The approach led to a large scale highly non-linear inequality constrained optimization problem, where the number of constraints was at the same order as the number of unknowns. As the number of unknowns for practical applications is of the order of millions and even billions, an efficient numerical scheme becomes mandatory.

The goals of this chapter are as follows. First, the intent is to explore and develop an efficient optimization technique for the solution of the problem. In [55], a primal log-barrier (Section 8 in Chapter 2) method has been implemented. Such a method is known to have reliable convergence properties

but it tends to converge rather slowly [85, 15]. The proposed new approach is based on an augmented Lagrangian (AL) method; see Section 9, Chapter 2 and [47] and references within for background on the AL method. Second, a multi-grid method for the linear systems obtained at each iteration of the AL iteration is developed. This combination of a robust optimization technique with advanced linear algebra tools leads to a highly effective algorithm that can tackle large-scale problems, as our numerical results will show.

The chapter is organized as follows. Section 2 presents the mathematical description of the continuous problem and its associated discretization. Section 3 briefly reviews the optimization procedure, the augmented Lagrangian method, and discusses the adaptation of the algorithm to constrained image registration. Section 4 presents a multi-grid method for the solution of the linear system obtained at each step of the optimization course. Section 5 demonstrates the properties of the new approach to clinical data. Finally, Section 6 summarizes and suggests future research directions.

This study has been published in the Journal of Numerical Linear Algebra with Applications [53].

2 Mathematical Formulation and Discretization

This section states the registration problem as a continuous constrained optimization problem and discusses its discretization.

2.1 Continuous Formulation

Let $\Omega = (0,1)^3 \subset \mathbb{R}^3$ be a domain and $T, R : \Omega \to \mathbb{R}$ denote the template and reference image, respectively. The goal is to compute a displacement

$u : \mathbb{R}^3 \to \mathbb{R}^3$, such that ideally $T(x+u(x)) = R(x)$ for all $x = [x^1, x^2, x^3] \in \Omega$. More precisely, the distance is measured by the L_2-norm (or sum of squared differences):

$$\mathcal{D}[u] := \tfrac{1}{2} \, \|T[\mathrm{id} + u] - R\|^2_{L_2(\Omega)} := \frac{1}{2} \int_\Omega (T(x + u(x)) - R(x))^2 \, dx. \quad (3.1)$$

As outlined in the introduction, $\mathcal{D}$ does not have a unique minimizer and the elastic potential is used for regularization,

$$\mathcal{S}[u] := \frac{1}{2} \int_\Omega \theta \, (\nabla u : \nabla u) + (\theta + \tau)(\mathrm{div} \, u)^2 \, dx, \quad (3.2)$$

where $(A : B) := \sum_{i,j} A_{i,j} B_{i,j}$ denotes a matrix inner product, and θ and τ are the Lamé constants; see [80] for details.

Similar to [55], feasibility of a transformation is constrained by

$$\kappa_m(x) \leq \mathcal{C}_{\mathrm{vol}}[u](x) \leq \kappa_M(x) \quad \text{for all} \quad x \in \Omega_\kappa, \quad \mathcal{C}_{\mathrm{vol}}[u] := \det(I + \nabla u), \quad (3.3)$$

where the functions κ_m and κ_M and the subset $\Omega_\kappa \subset \Omega$ are provided by the user and are customarily prescribed for a particular application.

In summary, the desired displacement u is a minimizer of the constrained optimization problem

$$\min \mathcal{D}[u] + \alpha \, \mathcal{S}[u] \quad \text{subject to} \quad \kappa_m \leq \mathcal{C}_{\mathrm{vol}}[u] \leq \kappa_M, \quad (3.4)$$

where α is a regularization parameter, appropriately chosen to balance the data-fit and the regularity of the displacement.

Problem (3.4) is a non-trivial constrained optimization problem. The difficulties arise from two sources. First, the constraint is highly non-linear and involves products of derivatives. For comparison, the well-known Navier-Stokes equations [36] involve only terms of type $u^j \; \partial_i u^j$, whereas Problem (3.4) involves complex terms like $\partial_i u^1 \; \partial_j u^2 \; \partial_k u^3$, where $i, j, k \in \{1, 2, 3\}$. Second, the size of the discrete 3D optimization problem (3.4) can become

large. For example, for high resolution images, a proper discretization may result in millions of unknowns and constraints. As a consequence, no standard optimization routine can be used and specialized routines are required.

2.2 Discretization

Similar to other optimization problems, which stem from continuous space, a decision whether to discretize the continuous problem first and then to optimize or to compute the optimality conditions in functional space and only then to discretize the problem has to be made. This option is often refered to as *discretize then optimize* versus *optimize then discretize*. Discretize first then optimize is usually preferable whenever possible, as it leads to simpler algorithmic treatment of the optimization problem; see [37] for a further discussion.

The optimization problem (3.4) is discretized using a finite volume approach. The domain Ω is partitioned into $n_1 \times n_2 \times n_3$ cells of size $h_1 \times h_2 \times h_3$ of volume $h := h_1 h_2 h_3$. This results in a nodal grid of $n := (n_1+1)(n_2+1)(n_3+1)$ nodal nodes; see Figure 3.1. Full integer vectors $i = (i_1, i_2, i_3)$, $i_j = 0, ..., n_j$, are used for the nodal points and half integers are used for cell centered points. The displacement is discretized on a nodal grid, where with a lexicographical ordering of the grid points, $x_i := [x^1_{i_1}, x^2_{i_2}, x^3_{i_3}]$, $x^{h,k} := [x^k_1, ..., x^k_n] \in \mathbb{R}^n$ and $u^{h,k} \approx [u^k(x_1), ..., u^k(x_n)] \in \mathbb{R}^n$

$$x^h := [x^{h,1}, x^{h,2}, x^{h,3}] \in \mathbb{R}^{3n}, \quad \text{and} \quad u^h := [u^{h,1}, u^{h,2}, u^{h,3}] \in \mathbb{R}^{3n}.$$

Similar to other registration algorithms, the images T and R are represented at cell-centered knots; see [80] for details. Thus, an interpolation operator $\mathcal{I}^n_c$ is used to represent u on cell-centers. A discretization of $\mathcal{D}$ (3.1) based on this averaging and a midpoint quadrature rule is given by

$$D^h(u^h) := \tfrac{h}{2} \, \|T(\mathcal{I}^n_c(x^h + u^h)) - R(\mathcal{I}^n_c x^h)\|^2. \tag{3.5}$$

The integral of the squared norm of the gradient of the displacement over a voxel $\Omega_{i+\frac{1}{2}}$ is

$$\int_{\Omega_{i+\frac{1}{2}}} (\nabla u : \nabla u)\ dx = \sum_{j,k=1}^{3} \int_{\Omega_{i+\frac{1}{2}}} (\partial_j u^k)^2\ dx.$$

Note that $\int_{\Omega_{i+\frac{1}{2}}} (\partial_j u^k)^2\ dx = h\ (\partial_j u^k(x_{i+\frac{1}{2}}))^2 + \mathcal{O}(h^2)$. Let $e_j \in \mathbb{R}^3$ denote the jth unit vector. The cell-centered value is approximated by the average of four finite difference approximations on the edges; see also Figure 3.1:

$$\partial_j u^k(x_{i+\frac{1}{2}}) \approx [\partial_j^h u^{h,k}]_{i+\frac{1}{2}} := \frac{1}{4} \sum\nolimits_{p_j = i_j,\ p_\ell \in \{i_\ell, i_\ell+1\} \text{ for } \ell \neq j} \quad [\partial_j^h u^{h,k}]_{p+\frac{1}{2}e_j},$$

where the second order accurate discrete partial derivative operator ∂_j^h is defined by

$$[\partial_j^h u^{h,k}]_{p+\frac{1}{2}e_j} := \tfrac{1}{h_j}\ (u^{h,k}_{p+e_j} - u^{h,k}_p).$$

The integral over the domain Ω is obtained by summing the averaged squares of the finite difference approximations of the derivatives on the edges. Note that the alternative, i.e. summing the squares of the averaged derivative approximation yields a non-trivial nullspace and consequently may lead to instability. In order to find a quadratic form for the discretization, the following identity is used,

$$a^\top\ [d_i^2]_{i=1}^n = a^\top \mathrm{diag}(d)\ d = \mathrm{diag}([a_i d_i]_{i=1}^n)\ d = d^\top \mathrm{diag}(a)\ d = \|d\|_a^2.$$

Introducing the edge-to-cell-center averaging operators $\mathcal{A}_c^{e,j}$,
$a := [(\mathcal{A}_c^{e,1})^\top e, (\mathcal{A}_c^{e,2})^\top e, (\mathcal{A}_c^{e,3})^\top e]$, then with e being the vector with all one entries, the Kronecker-product $\otimes$ [16], and the discrete gradient $(\nabla^h)^\top := I_3 \otimes [(\partial_1^h)^\top, (\partial_2^h)^\top, (\partial_3^h)]^\top$, it thus holds that

$$\int_\Omega (\nabla u : \nabla u)\ dx = h\ \|\nabla^h u^h\|_a^2 + \mathcal{O}(h^2).$$

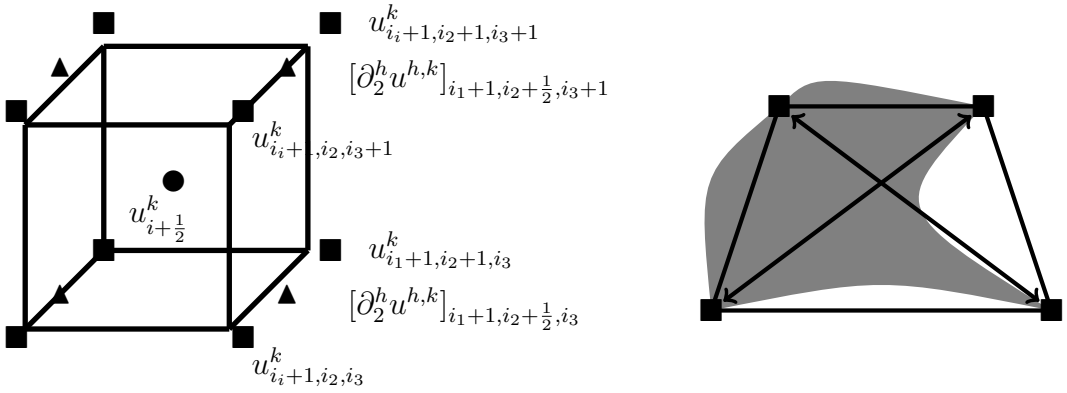


Figure 3.1: Left: Voxel $\Omega_{i+\frac{1}{2}}$ with cell-center (•) and nodal (■) representation of u^k and approximation of $\partial_2 u^k$ on edges (▲); Right: 2D representation, nodal points (■), deformed volume (gray), its linear approximation (black polygon), and the diagonals used to compute its approximated volume (arrows).

The divergence of u is approximated using the same concepts,

$$\begin{aligned} \text{div } u(x_{i+\frac{1}{2}}) &= \sum_{j=1}^{3} \partial_j u^j(x_{i+\frac{1}{2}}) \approx \sum_{j=1}^{3} [\partial_j^h u^{h,j}]_{i+\frac{1}{2}}, \\ [\text{div } u(x_{i+\frac{1}{2}})]_i &\approx \text{div}^h u^h := [\mathcal{A}_c^{e,1}\partial_1^h, \mathcal{A}_c^{e,2}\partial_2^h, \mathcal{A}_c^{e,3}\partial_3^h]\ u^h. \end{aligned}$$

A discretized version of the regularizer is thus given by

$$S^h(u^h) := \frac{1}{2}\ (u^h)^\top A_{\text{reg}} u^h := \frac{h}{2}\Big(\theta \|\nabla^h u^h\|_a^2 + (\theta+\tau)\|\text{div}^h u^h\|^2\Big). \tag{3.6}$$

It remains to discretize the constraint. Note that

$$\int_{\Omega_{i+\frac{1}{2}}} \mathcal{C}_{\text{vol}}[u](x)\ dx = h\ \mathcal{C}_{\text{vol}}[u](x_{i+\frac{1}{2}}) + \mathcal{O}(h^2).$$

With $C^{h,\text{vol}}_{i+\frac{1}{2}} := h\ \mathcal{C}_{\text{vol}}[u](x_{i+\frac{1}{2}})$ and the cell-centered discretized derivatives $d_j^k := [\mathcal{A}_c^{e,j}\partial_j^h u^{h,k}]_{i+\frac{1}{2}}$, it holds that

$$C^{h,\text{vol}}_{i+\frac{1}{2}} = h \begin{vmatrix} 1+d_1^1 & d_2^1 & d_3^1 \\ d_1^2 & 1+d_2^2 & d_3^2 \\ d_1^3 & d_2^3 & 1+d_3^3 \end{vmatrix}.$$

Note that a first order approximation to the determinant is $\mathcal{C}_{\text{vol}}[\varepsilon u] \approx 1 + \varepsilon \operatorname{div} u$. Thus, it is expected that at least in the first iteration the Jacobian of the constraint behaves like the divergence operator. This interpretation plays a key role in the stability analysis of the discretization presented in the section addressing local Fourier analysis.

3 Optimization

This section describes an optimization approach for the discretized problem (3.4),

$$\min J(u) := D(u) + \alpha\, S(u) \quad \text{subject to} \quad C(u) \geq 0, \tag{3.7}$$

where for ease of presentation the superscript $(\cdot)^h$ has been dropped. Moreover, only one of the bound constraints is addressed as the handling of the upper bound is along the same lines. For ease of presentation, the constraint has been rephrased: $C(u) := C^{h,\text{vol}}(u) - \kappa_m$.

The goal is to derive a Newton-type scheme. Since the size of this problem is rather large, employment of inexact solvers for Newton-systems is inevitable. While inexact Newton type methods are well established for unconstrained optimization [86, 70], their adaptation to constrained optimization is in its infancy [19]. Therefore, methods involving unconstrained optimization at their core are used in this study.

Two such methods as introduced in Chapter 2 are the primal log barrier method (LB) (Section 8), which was used in [55], and the augmented Lagrangian method (AL) (Section 9), which is practiced in this chapter. The AL has some well-known advantages over the LB: it tends to converge faster and to avoid some of the ill-conditioning behavior that is inherited in LB [86]; see also [47] for a recent algorithmic development of AL. As shown in the next sections, the AL method generates better numerical systems to be solved at

each iteration and thus allows for more efficient calculation.

3.1 Augmented Lagrangian Method

By introducing a non-negative slack vector s, problem (3.7) is transformed to

$$\min J(u) := D(u) + \alpha\, S(u) \quad \text{subject to} \quad C(u) - s = 0 \text{ and } s \geq 0, \quad (3.8)$$

which exhibits simpler bound constraints. This enables to handle the inequality constraints by AL and the slack variable by using an active set method as described below. Following (2.14), the augmented Lagrangian is given by

$$L(u, s, p; \mu) := J(u) - p^\top (C(u) - s) + \frac{\mu}{2} \|C(u) - s\|^2.$$

The AL algorithm divides the problem into a two step decoupled iteration. In the first step the Lagrange multiplier p and the penalty μ are fixed and the following is solved (approximately)

$$\min L(u, s; p, \mu) \quad \text{subject to} \quad s \geq 0 \text{ and } p, \mu \text{ given.}$$

The second step is a point-wise update of p, as in (2.15), and μ:

$$p \leftarrow p - \mu(C(u) - s) \quad \text{and} \quad \mu \leftarrow \gamma\mu, \quad (3.9)$$

where $\gamma > 1$ is a moderate constant, e.g. $\gamma = 1.1$. The main advantage of the AL method over a LB method is that the parameter μ is substantially smaller. As shown next, this has a positive effect on the solution of the linear system solved at each iteration. Obviously, the computationally challenging part is the minimization on the Lagrangian, which is described in the next section.

3.2 Minimization of the Augmented Lagrangian

Various methods can be used for the minimization of the augmented Lagrangian. Since the slack variables have simple bound constraints, a modified version of the projected Newton method as suggested in [74] is used. The method is an active set method and involves two steps. The first step takes a projected steepest descent direction to identify the active set. In the second step, an inexact Newton method is used for the non-active set.

Given the initial guesses for u and s, the first stage of a projected Newton scheme updates the variables by a projected gradient

$$u \leftarrow u - t\ \nabla_u L \quad \text{and} \quad s \leftarrow \mathcal{P}(s - t\ \nabla_s L), \tag{3.10}$$

where t is chosen such that the AL decreases and $\mathcal{P}$ is a projection to the non-negative cone,

$$\begin{aligned} \mathcal{P}(x) &= [\max\{x_i, 0\}]_{i=1}^n, \\ \nabla_u L &= \nabla J + \nabla C^\top(\mu(C - s) - p), \qquad (3.11) \\ \nabla_s L &= p - \mu(C - s). \qquad (3.12) \end{aligned}$$

Let $\mathcal{I}$ denote the set of inactive constraints and $\mathcal{A}$ denote the set of active constraints. The next stage is to set $s_\mathcal{A} := 0$ and to take an inexact Newton-like step for an update of $s_\mathcal{I}$ and u. In particular, a Gauss-Newton approximation to the Hessian of J is used

$$\nabla^2 J \approx A := \nabla^2_{\text{GN}} D + \alpha \nabla^2 S,$$

where $\nabla^2_{\text{GN}} D$ denotes a Gauss-Newton type approximation to the Hessian of the distance measure [55]. Note that only the second derivative of the data fitting component is approximated but the regularization component is exact. The Gauss-Newton system for the augmented Lagrangian thus reads

$$\begin{pmatrix} A + \mu \nabla C^\top \nabla C & -\mu \nabla C^\top P^\top \\ -\mu P \nabla C & \mu I \end{pmatrix} \begin{pmatrix} \delta u \\ \delta s_\mathcal{I} \end{pmatrix} = \begin{pmatrix} \nabla_u L \\ P \nabla_s L \end{pmatrix},$$

where P is a sub-matrix of the identity matrix such that $s_{\mathcal{I}} = Ps$.

To invert the system $\delta s_{\mathcal{I}}$ is eliminated,

$$\delta s_{\mathcal{I}} = \tfrac{1}{\mu}\, P\nabla_s L + P\nabla C\delta u, \tag{3.13}$$

obtaining the system

$$H(\mu)\ \delta u = g, \tag{3.14}$$

with

$$\begin{aligned} H(\mu) &:= A + \mu\nabla C^\top (I - PP^\top)\nabla C, \\ g &:= \nabla_u L + \nabla C^\top P^\top P\nabla_s L. \end{aligned}$$

Note that $I - PP^\top$ is a diagonal matrix with zero diagonal entries for the inactive indices $\mathcal{I}$ and ones for the active indices $\mathcal{A}$. The matrix A is symmetric positive semi-definite with a small null space due to the boundary conditions, the matrix $\nabla C^\top (I - PP^\top)\nabla C$ is also symmetric positive semi-definite. However, our experience shows that for all practical cases, the resulted matrix $H(\mu)$ is symmetric positive definite.

Algorithm 4 summarizes these steps.

Algorithm 4: AL for constrained image registration: $u \leftarrow$ `ALIR`.

for $\ell = 1, 2, \ldots$ **do**

- Compute the gradients $\nabla_u L$ (cf. (3.11)) and $\nabla_s L$ (cf. (3.12)).
- Update u and s using (3.10).
- Determine the inactive set $\mathcal{I}$ and the active set $\mathcal{A}$, such that $s_{\mathcal{A}} = 0$.
- Solve the system (3.14) for δu and compute $\delta s_{\mathcal{I}}$ using (3.13).
- Update $u \leftarrow u + \delta u$ and $s \leftarrow s + \delta s$.
- Update the Lagrange multipliers p and the penalty μ using (3.9).

Obviously, the bulk of the work in Algorithm 4 is done in the fourth step, where the symmetric positive definite system (3.14) has to be solved. The next section presents a multi-grid method for the solution of this system.

4 Multi-grid Solver

Each iteration of the AL needs to solve the linear system (3.14). Here, a conjugate gradient (CG) algorithm with a V-cycle multi-grid [101] as preconditioner is used. This combination of a multi-grid preconditioned CG has several attractive attributes, including linear complexity and robustness.

For classical multi-grid methods to be efficient, the linear system is required to have special properties. In particular and as discussed in the next section, h-ellipticity is required. Multi-grid methods have three main components which are described subsequently: a *smoother*, *restriction* and *prolongation* operators, and a *coarse grid solver*; see also [101] for details.

4.1 Smoother

For a generic elliptic equation, iterative procedures such as Gauss-Seidel or Jacobi are typically used. However, for systems of PDE's, the coupling of the different components has to be taken into consideration [101].

Since system (3.14) is a tightly coupled system of PDE's for $[\delta u^1, \delta u^2, \delta u^3]$, experiments with two different smoothers are performed. First, a simple block Jacobi smoother coupling the unknowns at each node is used. Thus, a cycle through all nodes on the computational grid is performed, where on each node an associated 3-by-3 system is solved. Second, a Vanka type smoother [103] is used. Here, a cell $\Omega_{i+\frac{1}{2}}$ and the eight associated nodes are considered. This yields systems of 24 unknowns to be solved for each of the cells. A classical Vanka smoother proceeds in a Gauss-Seidel fashion, i.e. using the updated values for the computations on a next cell. This sequential nature of the Vanka smoother is a drawback for parallel algorithms. Thus, we suggest a new variant of the Vanka smoother. The idea is to use an eight color labeling of the cells, where cells of one color are not direct neighbors; see Figure 3.2. A reordering of the unknowns according to these colors yields a linear system

with uncoupled blocks consisting of 24 unknowns each; an exemplarily plot of the non-zero elements of the original and reordered Hessian is shown in Figure 3.2. The reordered sparse Hessian system is solved using a block-Jacobi iteration. This reordering leads to locally strongly coupled blocks of unknowns which are relaxed simultaneously. While the Vanka smoother is roughly 8 times more expensive than the nodal block Jacobi smoother, it takes more of the coupling into consideration and is hence a more efficient smoother.

4.2 Prolongation and Restriction Operators

Other components of a multi-grid scheme include the prolongation and restriction operators. Here, a full weighted operator [101] is used for restriction and its adjoint for prolongation.

4.3 Coarse Grid Solver

Similar to many other approaches [10], the coarse grid discretization may not result in a small linear system. For example, even if the coarsest mesh consists of only 16^3 cells, the number of unknowns is already about $15,000$. Thus, an exact solution of the coarse problem may be too costly. Therefore, a conjugate gradient method [63] with symmetric Gauss-Seidel preconditioning is used to compute an approximated solution on the coarsest mesh. However, it is important to choose a small tolerance for the approximate solution on the coarsest mesh; here $\text{tol} = 10^{-9}$ is chosen.

4.4 Discretization Properties

This subsection provides a local Fourier analysis (LFA) and an h-ellipticity analysis of the system (3.14). For ease of presentation, the analysis is for

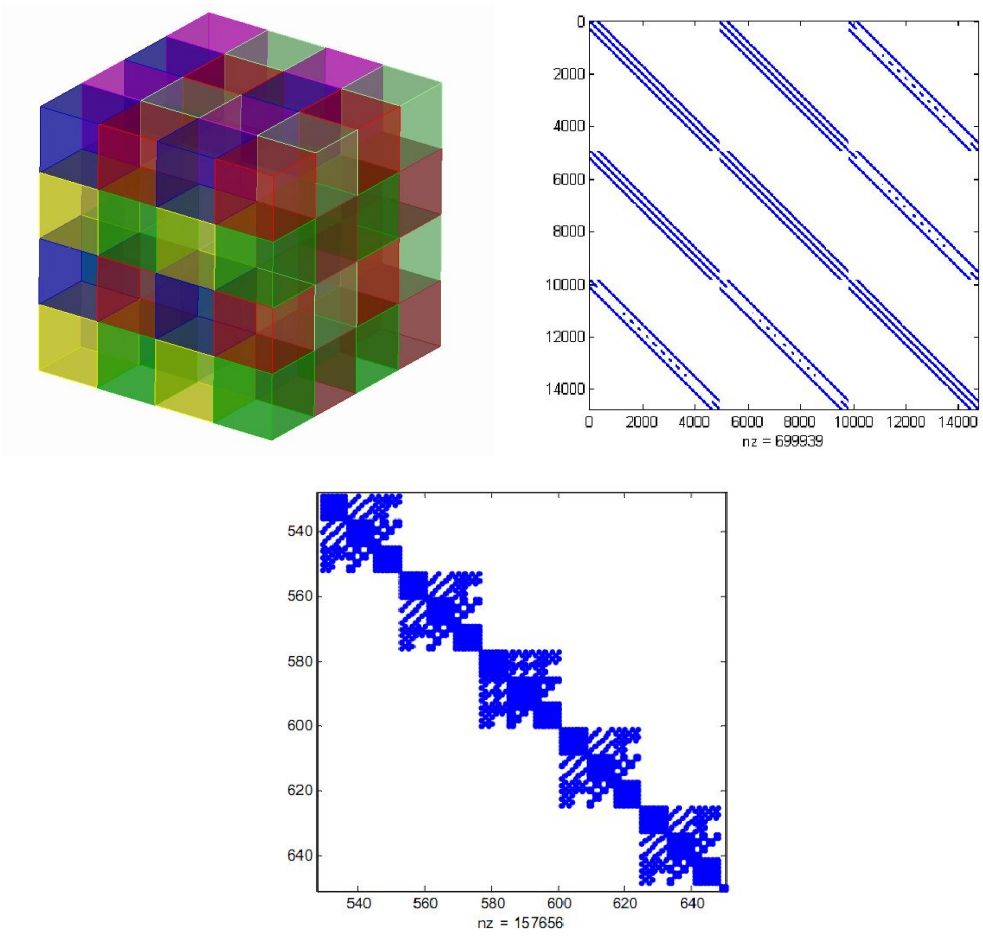


Figure 3.2: Top-Left: the eight colors of a mesh. Top-Right: Original Hessian. Bottom: Zoom-in of reordered block diagonal Hessian.

spatial dimension of two, assuming $h := h_1 = h_2$. To simplify the LFA further, it is assumed that $I - PP^\top = I$ and $\nabla^2_{\text{GN}} D = 0$, which is a worst case scenario, as adding a semi-positive definite matrix improves the spectrum. The discrete operators are locally linearized and by freezing coefficients of $\nabla C^\top (I - PP^\top) \nabla C$, the operator can be considered as having constant coefficients. Finally, the operators are assumed to be defined on an infinite grid.

Introducing the shortcuts $\nu := h\alpha\theta/2$ and $\xi := h\alpha(\theta + \eta)/2$, the discrete operator under consideration is

$$H_0 = \nu \boldsymbol{\Delta}^h + \xi \text{div}^{h^\top} \text{div}^h + \mu \nabla C^\top \nabla C,$$

with

$$\begin{aligned} \boldsymbol{\Delta}^h &\sim \begin{pmatrix} \Delta & 0 \\ 0 & \Delta \end{pmatrix}, \qquad \text{div}^{h^\top} \text{div}^h \sim \begin{pmatrix} \partial_1\partial_1 & \partial_1\partial_2 \\ \partial_2\partial_1 & \partial_2\partial_2 \end{pmatrix} \\ \nabla C^\top \nabla C &\sim \begin{pmatrix} (\alpha_1\partial_1 + \alpha_2\partial_2)^2 & (\alpha_1\partial_1 + \alpha_2\partial_2)(\alpha_3\partial_1 + \alpha_4\partial_2) \\ (\alpha_3\partial_1 + \alpha_4\partial_2)(\alpha_1\partial_1 + \alpha_2\partial_2) & (\alpha_3\partial_1 + \alpha_4\partial_2)^2 \end{pmatrix}, \end{aligned}$$

where α_j are constants arising from the linearization.

The stencils for the discrete operators are

$$\Delta \sim \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} /h^2, \quad \partial_1\partial_1 = (\partial_2\partial_2)^\top \sim \begin{pmatrix} -2 & 4 & -2 \\ -4 & 8 & -4 \\ -2 & 4 & -2 \end{pmatrix} /h^2,$$

$$\partial_1\partial_2 \sim \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} /h^2.$$

The fundamental quantities in LFA are the grid functions

$$\varphi(\sigma, x) = e^{i\sigma^\top x/h} = e^{i\sigma_1 x_1/h} e^{i\sigma_2 x_2/h},$$

where i denotes the imaginary unit, x a spatial position, $\sigma \in \Sigma^2$ a frequency, and $\Sigma := [-\pi, \pi)$ the continuous frequency interval, which is partitioned into

low and high frequencies:

$$\Sigma^{\text{low}} := [-\tfrac{\pi}{2}, \tfrac{\pi}{2}) \quad \text{and} \quad \Sigma^{\text{high}} := \Sigma \backslash \Sigma^{\text{low}}.$$

Appling the discretized operators to φ yields the following symbols

$$\begin{aligned}
\mathcal{F}[\Delta^h] &= 2 - \cos\sigma_1 - \cos\sigma_2, \\
\mathcal{F}[\partial_1^h \partial_1^h] &= (2 - 2\cos\sigma_1)(2 + 2\cos\sigma_2), \\
\mathcal{F}[\partial_2^h \partial_2^h] &= (2 + 2\cos\sigma_1)(2 - 2\cos\sigma_2), \\
\mathcal{F}[\partial_1^h \partial_2^h] &= 4\sin\sigma_1 \sin\sigma_2.
\end{aligned}$$

These intermediates are used to compute the h-ellipticity of H_0, which provides a qualitative criterion for the existence of local smoothers for a given discrete operator and is thus directly related to the definition of the smoothing factor; see [101] for details. The h-ellipticity of a discrete operator H is defined via the Fourier-symbol $\widehat{H} := \mathcal{F}[H]$ of H:

$$E_h(H) := \frac{\min\{|\det \widehat{H}(\sigma)| : \sigma \in \Sigma^{\text{high}}\}}{\max\{|\det \widehat{H}(\sigma)| : \sigma \in \Sigma\}}.$$

The h-ellipticity values for H_0 are summarized in Table 3.1, where all coefficients and the Lamé constants are set to one, such that all components have equal weight. The penalty parameter μ varies between h^{2k}, $k = 1, 0, -1, -2$.

Table 3.1: Numerical h-ellipticity values for different grid sizes.

grid size	$\mu = h^2$	$\mu = 1$	$\mu = 1/h^2$	$\mu = 1/h^4$
8×8	0.26	0.27	0.05	$7.8 \cdot 10^{-4}$
16×16	0.26	0.27	0.05	$2.0 \cdot 10^{-4}$
32×32	0.26	0.26	0.05	$4.9 \cdot 10^{-5}$
64×64	0.26	0.26	0.05	$1.2 \cdot 10^{-5}$

Table 3.1 clearly shows that as long as μ maintains moderate values the discretization is highly h-elliptic and reasonable convergence rates for the multi-grid solver are to be expected. However, for large values of μ the h-ellipticity is close to zero and any multi-grid method is expected to be inefficient. One reason for this problem is related to the null-space of the linearized constraint, which has a high frequency checkerboard pattern. At the limit $\mu \to \infty$, for which the matrix $\nabla C^\top \nabla C$ dominates the Hessian, the h-ellipticity of the system is expected to deteriorate. This comes as no surprise: similar discretizations of the Navier-Stokes problem suffer from the very same short-comings; see [101].

To summarize, just as in similar applications, the Hessian is sufficiently h-elliptic for moderate choices of μ. The use of an AL technique is highly beneficial from a linear algebra perspective, since μ can be chosen as a moderate constant. Choosing $\mu = \mathcal{O}(1)$ yields an efficient multi-grid convergence rate. In contrast, in the log-barrier method very similar linear systems are to be solved, where now $\mu \to \infty$ is required and thus, convergence problems may be anticipated.

5 Numerical Results

In this section we test the performance of our algorithm. To begin with, two different components are tested. The first test assesses the performance and behavior of the AL method for a model problem but discretized using different mesh sizes and different multi-grid solvers. The second test quantifies the multi-grid solver and its efficiency for different mesh sizes. The model problem was chosen to be the registration of two 3D SPECT images of a swine's heart at different phases of the cardiac cycle. The first image is taken at the systolic stage (hereafter the template image) whereas the second image

is taken at the diastolic stage (reference image), see Figure 3.3. The data was provided by Tracy Faber from the department of Radiology at Emory University, as part of a controled experiment aimed to evaluate the heart's motion during the cardiac cycle. This problem is particularly challenging since the motion from systole to diastole is highly non-linear.

For this application, the extent to which the volume of the heart in the systolic phase is allowed to shrink compared to its diastolic volume is $\kappa_m = 0.5$. This value is based on the physiological properties of the cardiac muscle [38]. As there is a clear trend in volumetric change (shrinkage from the diastolic to the systolic stage), the upper bound κ_M is made inactive and is thus not specified. As mentioned earlier, this bound can be treated in similar manner as the lower bound κ_m, when justified by physiological or physical considerations.

Optimal performance of an algorithm is measured by means of mesh independence. This implies that if each iteration is of linear complexity with respect to the number of unknowns, then the overall algorithm is optimal, i.e. linear with respect to the problem size.

In the first example, the finest grid is $96 \times 96 \times 80$ and the grid sizes vary from $24 \times 24 \times 20$ via $48 \times 48 \times 40$ to $96 \times 96 \times 80$. The initial value for the penalty parameter is $\mu = \frac{2}{h_1 h_2 h_3}$, and it is increased by a factor of 1.1 at each AL iteration. Figure 3.3 presents the transformed template for grid dimensions of $48 \times 48 \times 40$ and the difference image. The transformed template visually resembles the reference image to a good extent.

A quantitative evaluation of the difference between the reference and the template compared to the reference and the transformed template are displayed in Figure 3.4 left and right respectively. Table 3.2 records the number of AL iterations required for convergence over different grid sizes. As can be seen from this table, the AL algorithm is almost optimal: although the

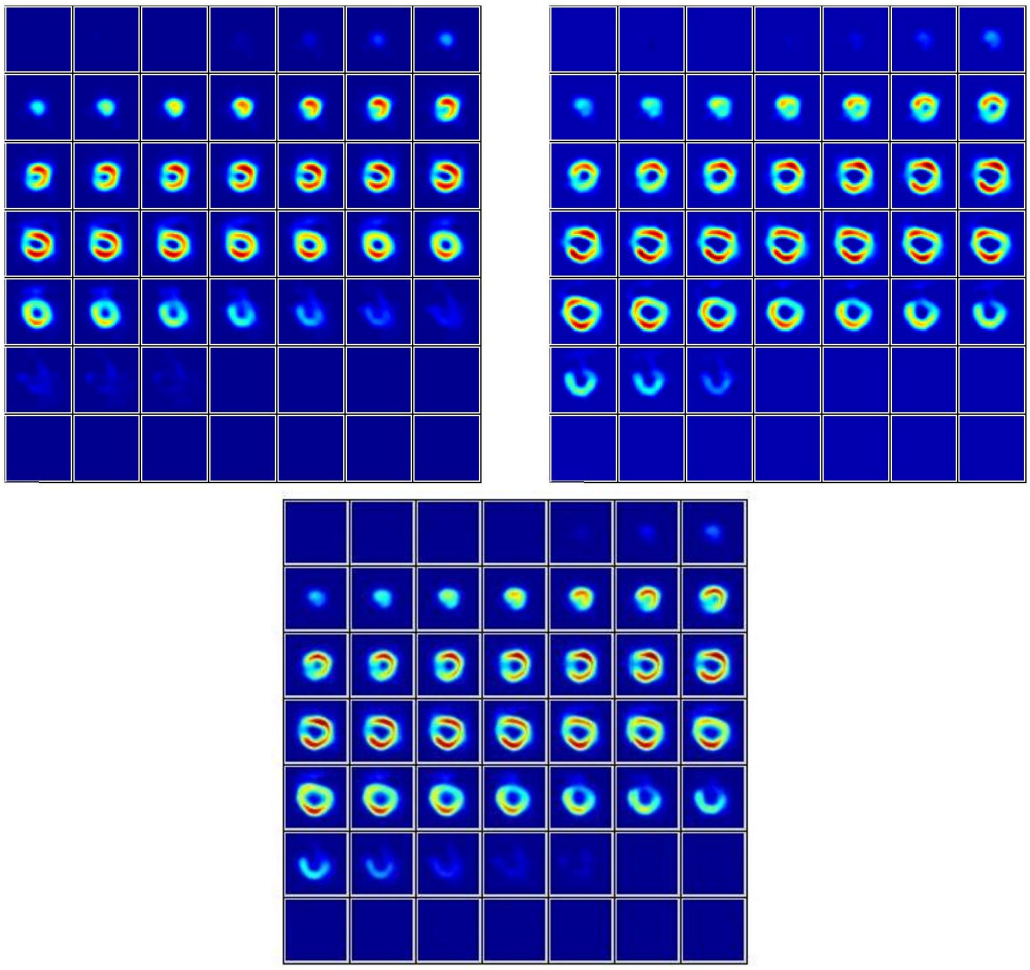


Figure 3.3: 2D slices of 3D SPECT images of the heart, during systolic and diastolic phases. Top left: template image during systolic phase, top right: reference image during the diastolic phase, bottom: Transformed template image.

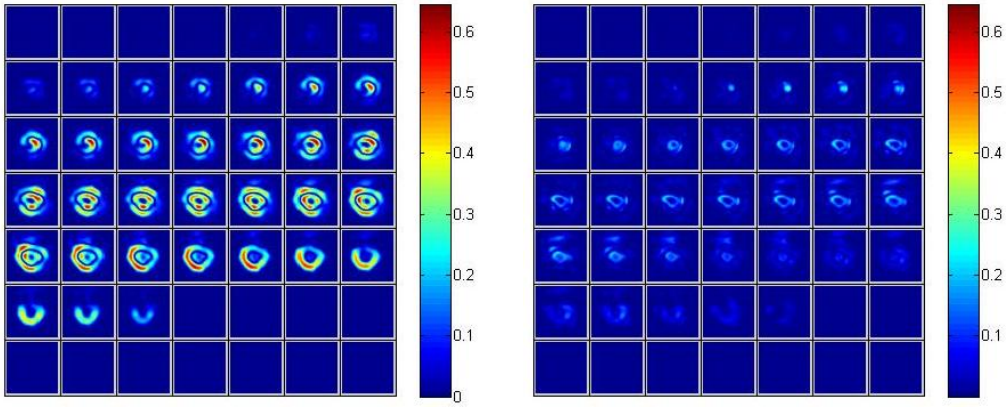


Figure 3.4: Left: difference between the template and the reference image; Right: difference between the transformed template and the reference image

number of variables increase by a factor of 8 between different grid sizes, only a few number of non-linear iterations are needed to achieve convergence.

Table 3.2: Number of AL iterations required for convergence for different grid sizes.

Grid Size	AL iterations
$24 \times 24 \times 20$	10
$48 \times 48 \times 40$	11
$96 \times 96 \times 80$	14

Another interesting feature of the proposed algorithm is the automatic identification of the active set, the region in which equality for the constraints holds. Note that if the active set is empty, the unconstrained problem is solved. Figure 3.5 visualizes the Jacobian of the transformation. As it turns out, there are quite a few voxels for which the lower bound of the

volume is attained. These voxels correspond to the active set at the end of the optimization process. The number of active constraints during the optimization process is recorded in Table 3.3. Though the active set is rather small, it is concentrated in the region of interest, i.e. where most of the deformation occurs. Such an observation can typically be inspected from an image of the Lagrange multipliers.

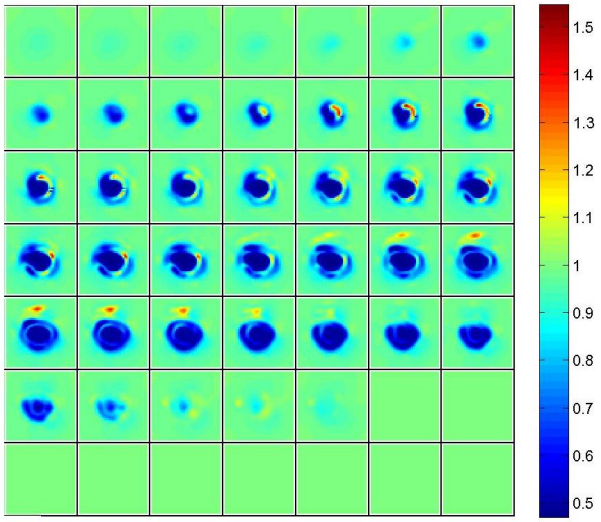


Figure 3.5: Volume changes of the discrete transformation (Jacobian).

Finally, the deformed grid is presented in Figure 3.6 and provides an independent insight into properties of the transformation. Similar to the Jacobian as shown in Figure 3.5, a general impression regarding the changes each volume is undergoing is visible. In addition, these images can provide a valuable tool to confirm the regularity of a transformation and assure that no folding or twisting occurs. Note that for most or even all medical applications, folding or twisting of tissue is considered to be unacceptable.

Table 3.3: The size of the active set (in voxel and percentage) at each AL iteration, for grids of different dimensions, using an eight colored Vanka-type smoother.

AL iteration	Grid $24 \times 24 \times 20$ (out of 11520)	Grid $48 \times 48 \times 40$ (out of 92160)	Grid $96 \times 96 \times 80$ (out of 737280)
1	166 (1.4%)	3696 (4.0%)	29701 (4.0%)
2	185 (1.6%)	2377 (2.6%)	20359 (2.8%)
3	245 (2.1%)	2782 (3.0%)	23681 (3.2%)
4	296 (2.6%)	2456 (2.7%)	21860 (3.0%)
5	320 (2.8%)	2831 (3.1%)	24255 (3.3%)
6	272 (2.4%)	2465 (2.7%)	20268 (2.7%)
7	312 (2.7%)	2793 (3.0%)	22575 (3.1%)
8	290 (2.5%)	2644 (2.9%)	23539 (3.2%)
9	315 (2.7%)	2767 (3.0%)	23038 (3.1%)
10	286 (2.5%)	2832 (3.1%)	23824 (3.2%)
11		2732 (3.0%)	23840 (3.2%)
12			23593 (3.2%)
13			23876 (3.2%)
14			23881 (3.2%)

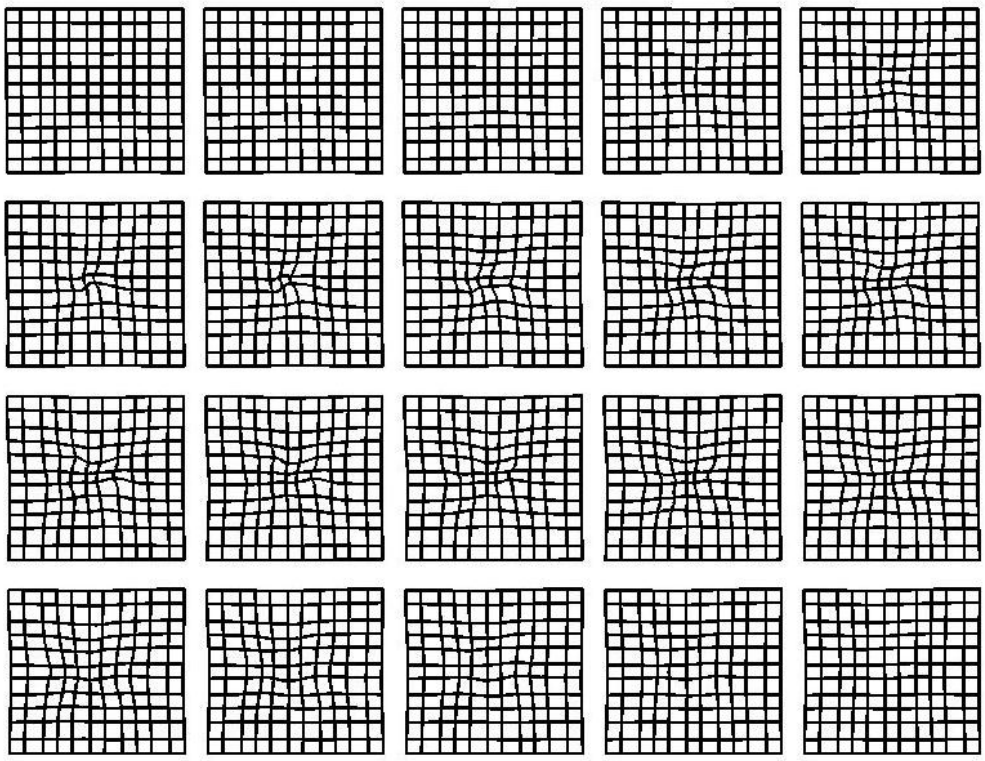


Figure 3.6: 2D slices of the transformation grid (ordered: left to right, top to bottom), with $2h$ spacing

Although the number of AL iterations is almost mesh independent, the overall algorithm is only optimal if the linear solver has linear complexity. Thus, the convergence rate of the multi-grid preconditioner is of particular importance and thus, it is tested during different phases of the optimization process using two different smoothers: a block nodal Jacobi and a Vanka smoother. The performance of PCG solver with the multi-grid preconditioner and the different smoothers is presented in Table 3.4. The performance is compared for the first iteration (no active set), an iteration in the midst of the optimization process, and the last iteration of the process. It is important to note that the emphasis was given to the algorithmic framework and its scalability, rather than optimizing code efficiency. Thus, iteration count, which should be platform and coding language independent, where compared.

Table 3.4: Comparison of numbers of PCG iterations required for converge to the tolerance of 10^{-3} at 1^{st}, 6^{th} and last iteration of the AL algorithm: different grid sizes, V-cycle multi-grid preconditioner with different smoothers.

	eight colored Vanka-type smoother		
Grid Size	1^{st} AL iteration	6^{th} AL iteration	Last AL iteration
$24 \times 24 \times 20$	4	7	9
$48 \times 48 \times 40$	4	8	10
$96 \times 96 \times 80$	5	11	16
	nodal Jacobi smoother		
Grid Size	1^{st} AL iteration	6^{th} AL iteration	Last AL iteration
$24 \times 24 \times 20$	6	22	25
$48 \times 48 \times 40$	6	24	33
$96 \times 96 \times 80$	7	25	40

Table 3.4 shows that, as expected, classic mesh independence behavior is

observable for the first iteration, while the mesh independency is mildly deteriorated for the final iterations. This latter observation can be explained as follows. The contribution from the active set in the Hessian has the form $\nabla C^\top W \nabla C$, where W is a diagonal matrix with Boolean diagonal entries. In the above simplified Fourier analysis, freezing coefficients are considered. Analogously to many other multi-grid methods, convergence deteriorates when the coefficient matrix W is not smooth or varies substantially [101]. Methods such as operator induced prolongation and restriction are needed in order to avoid these phenomena and to obtain a perfectly scalable method. The proposed scheme avoids the use of such prolongations as these techniques are inheritably more expansive and the deterioration of the proposed scheme is only mild. In Chapter 4, we consider the case where the coefficient matrix is not smooth. In this case, algebraic multi-grid method is considered.

6 Summary

A novel and efficient framework for 3D volume constrained image registration is proposed. The new framework is based on an augmented Lagrangian formulation of the inverse problem utilizing a multi-grid preconditioned conjugate gradients solver for the reduced system. The proposed formulation included two smoothing strategies. Both strategies rely on local coupling of the unknowns. These strategies are particularly favorable for parallel implementation, as they entail a block diagonal structure. One may consider three levels of parallelization of the aforementioned Vanka smoother: each color designates a block diagonal matrix of the size (grid size (n) / number of colors (8)), and thereby accommodate the running of 8 parallel matrix inversions. Since within each color-based block diagonal sub matrix, one may further consider an inner division of block diagonal matrices of the size of 24×24, each of these smaller matrices can be processed in parallel. Lastly,

for large-parallel systems the combination of the two options can be sought.

Comparisons of the number of outer iterations for two smoothers (eight colored Vanka and nodal Jacobi) show similar mesh independent behavior. However, in terms of inner iterations, the eight colored Vanka method turns out to be substantially more effective. These results are theoretically plausible, because in the eight colored Vanka smoother the inner nodes are relaxed eight times, compared with a single relaxation in the nodal Jacobi approach.

Various numerical experiments with SPECT data have been conducted. As the results for other tests give similar numerical results, these are not included in the chapter. From an optimization perspective, the augmented Lagrangian formulation weighs the constraints moderately within the first few iterations by maintaining the augmentation parameter μ relatively small. This leaves greater freedom to align the template to the reference image. As the algorithm advances in terms of iterations, μ increases, which of course puts more focus on the the volume constraints.

From a medical imaging perspective, the transformed template image closely resembles the reference, indicating a good correspondence. Moreover, by satisfying the volume change constraints on the transformation a meaningful solution is generated automatically.

Chapter 4

Mass Preserving Image Registration

1 Introduction

The optimal mass transport problem is of cardinal importance in science and engineering. Applications include econometrics, fluid dynamics, automatic control, transportation, statistical physics, shape optimization, expert systems, and meteorology [91, 105]. The problem was first formulated by the civil engineer Gaspar Monge in 1781, and is concerned with finding an optimal way, in the sense of minimal transportation cost, of moving a pile of soil from one site to another. Much later the problem was extensively analyzed by Leonid Kantorovich [68], and is now known as the Monge–Kantorovich problem.

There are several formulations of the problem [1, 91, 105] of varying degrees of generality. We recall here the formulation of the Monge–Kantorovich problem for smooth densities and domains in a Euclidean space. For more general measures, see [1]. Let Ω_0 and Ω_1 be two diffeomorphic connected subdomains in $\mathbb{R}^d$, and let T and R be Borel measures [90] on Ω_0 and Ω_1, each

with a strictly positive density function $T(x) \geq T^0_{\text{low}} > 0$ and $R_1 \geq R^1_{\text{low}} > 0$, respectively. Assume

$$\int_{\Omega_0} T(x)dx = \int_{\Omega_1} R(x)dx,$$

so that the same total mass is associated with Ω_0 and Ω_1.

Under some mild assumptions, the Monge-Kantorovich problem may be expressed as follows:

$$\min \qquad M(u) := \frac{1}{p}\int_{\Omega} T(x)|u(x)|^p \, dx \tag{4.1a}$$

$$\text{s.t.} \qquad c(u) = \det(I_d + \nabla u)T(x + u(x)) - R(x) = 0, \tag{4.1b}$$

where u is a $C^{1,\alpha}$ diffeomorphism from $\Omega_0 \to \Omega_1$. The constraint $c(u) = 0$ (the Jacobian equation) is often referred to as the mass preserving (MP) property. Here, we consider the classical case of $p = 2$ as well as $1 \leq p \leq 2$ and attempt to address the limiting case of $p = 1$. Note, the constraint in this problem bares a great resemblance to the one introduced in the volume preserving image registration chapter (Chapter 3).

Even with a simple, quadratic distance function, the problem (4.1) is regarded as a highly non-linear equality constrained optimization problem. Extensive analysis as to the existence, uniqueness, and properties of the solution is available (see for example [1, 32, 105] and the references therein). However, while a large body of literature deals with the analysis of the problem, surprisingly a relatively small number of papers are concerned with finding solutions for the problem, and even a smaller number of publications deal with devising efficient *numerical* solutions for this challenging problem [2, 9, 22, 27, 28, 100, 87].

Generally speaking, numerical methods for the solution of the problem can be divided into three approaches. In the first approach, for the case $p = 2$, one prescribes the transformation u to honor the property $u = \nabla\Psi$ where Ψ is a concave function and thereby solves the Monge-Ampère equation [29,

88]. The second approach attempts to tackle the constrained optimization problem head-on, for example see [100].

A third approach for the solution of the problem, which serves as a starting point for this study, was proposed in the seminal paper of Benamou and Brenier [9]. Their research reconstructs an optimal path from T to R by solving a convex optimization problem with a linear space-time transport partial differential equation as a constraint. Their approach is particularly useful whenever a transportation path is needed. Its disadvantage is that it increases the dimensionality of the problem by recasting the problem as a space-time control problem. In the original work of Benamou and Brenier naïve discretization and an augmented Lagrangian method were practiced for the solution of the problem. While reproducing the results of the paper we have observed some stability issues as well as deterioration of the algorithm for large-scale problems. Nonetheless, the Benamou and Brenier method replaces a generally non-convex problem with a convex one and further avoids some other intricacies such as boundary conditions and difficulties when the density contrasts are large. In fact, for the latter case all algorithms known to us have failed when the density contrast exceeded a value of 50.

The goal of this study is to develop a numerical solution framework, based upon the Benamou-Brenier fluid dynamics formulation. We require this algorithm to be fast and efficient, capable of resolving problems comprising large density contrast, and lastly, we aim at the design of a generic formulation, for which alternative cost functionals can be readily incorporated, such as in [94].

The chapter is structured as follows. In Section 2 we reformulate the problem by following the Benamou and Brenier derivation and present its extension to more general L_p functionals. In Section 3 we introduce conservative discretization of the problem. In Section 4 we layout the optimization algorithm and the linear solver used to solve the optimization problem and the

linear systems that arise. In Section 5 we conduct a few numerical experiments and finally, in Section 6 we summarize the study.

2 Problem Reformulation

Let us consider the time interval $t \in [0, 1]$. The idea of Benamou-Brenier was to replace the non-convex problem (4.1) with a convex PDE constrained optimization problem

$$\min \quad \int_{t=0}^{1} \int_{\Omega} \rho(x,t)|v(x,t)|^2 \, dx \, dt \tag{4.2a}$$

$$\text{s.t.} \quad \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0 \tag{4.2b}$$

$$\rho(x,0) = T(x) \quad \rho(x,1) = R(x) \tag{4.2c}$$

Here, $v(x,t)$ is the velocity and $\rho(x,t) > 0$ is the density. This formulation essentially describes the transport evolution of the mass density distribution T to the distribution R, where minimal energy is invested, while the mass obeys a mass-preserving PDE. The above problem falls into the PDE constrained optimization problems category. Generic solutions for this family of problems were recently presented in the literature (see for example, [11, 14, 37] and references within). In the context of PDE optimization problems the density, ρ, is the state and the velocity, v, is the control. One then minimizes the objective function subject to fulfilment of the PDE constraints. This formulation is similar to the one solved in [14].

While conventional PDE optimization methods can be used here, there are some obvious limitations to their use. The main limitation is related to the underlying assumption that given the controls, v, the PDE can be solved independently from the optimization problem. For the problem described in equation (4.2) this is not the case. Note that there are no boundary conditions (BC) for ρ in the PDE and thus one cannot solve for ρ even if v is

known. Although it is possible to fabricate boundary conditions (and this has been done), such boundary conditions can bring disastrous effects upon the reconstruction. For example, the use of some inflow and outflow BC strongly biases v, similarly, periodic BCs are non-physical. As we see next, BCs for ρ are actually not needed in order to solve the problem. In fact, resolving the boundary conditions is an integral part of the optimization problem. To see that, we continue and follow the Benamou-Brenier formulation by setting the momentum $m = \rho v$ and replacing equation (4.2) with a convex optimization problem

$$\begin{aligned} \min \quad & f(m,\rho) = \int_{t=0}^{1} \int_{\Omega} \frac{|m(x,t)|^2}{\rho(x,t)}\, dx\, dt && \text{(4.3a)} \\ \text{s.t.} \quad & c(m,\rho) = \frac{\partial \rho}{\partial t} + \operatorname{div}\, m = 0 && \text{(4.3b)} \\ & \rho(x,0) = T \quad \rho(x,1) = R. && \text{(4.3c)} \end{aligned}$$

It is interesting to note that the above problem has the same structure as of the mixed form of non-linear flow in porous media [84]. This becomes apparent by setting $w = (m,\rho)^\top = (m_1, \ldots, m_d, \rho)^\top$, where d is the dimension of the problem, and rewriting the problem as

$$\begin{aligned} \min \quad & f(w) \\ \text{s.t.} \quad & \nabla^{\text{st}} \cdot w = 0 \\ & w(x,0) = T(x),\ w(x,1) = R(x) \end{aligned}$$

where $\nabla^{\text{st}}\cdot$ is a space-time divergence. If we treat time as a spatial dimension then the MK problem and non-linear flow in porous media have the exact same general variational form (see [8, 84]). Furthermore, it is well-known that such problems are well-posed as long as f is smooth and convex even in the absence of boundary conditions for $w_{d+1} = \rho$. As we shall see later, the lack of BC for ρ leads to BC for the Lagrange multipliers. To illustrate

this notion for the simplest case, let $f(w) = \int_\Omega |w|^2\, dx$. For this objective, a classical result yields the mixed form of the Poisson equation [17]. Thus, we see that the optimization problem at hand is nothing but a non-linear Poisson-like equation.

This similarity to the mixed form of flow in porous media motivates our research here. A large body of literature and highly efficient algorithms have been developed for such problems (e.g. [66, 82] and many others). It is merely sensible on our behalf to put this techniques into a good use. Understanding this similarity also helps in developing robust and stable discretization, in choosing efficient optimization algorithms, as well as indicating efficient solvers for the linear systems. We address each of these components separately in the following sections.

The above formulation allows us to easily consider alternative convex functionals. For example, we may consider the function $|v|^p\rho(x)$, which lends itself to the energy functional

$$\frac{|v(x,t)|^p}{\rho(x,t)} = \frac{|m(x,t)|^p}{\rho(x,t)^{p-1}}. \tag{4.4}$$

This functional is strictly convex for $1 < p \leq 2$. Interestingly, for the case $p = 1$ we obtain that the energy is reduced to $|m|$. It is possible to verify that the optimization problem in this case is not well-posed [32]. This indeed corresponds to ill-posedness of the original MK problem for the L_1 case. Nonetheless, in our numerical tests we shall investigate the use of a sequence of p's that converges to $p = 1$ as a possible approach to computing the L_1 solution, assuming it exists.

3 Discretization

It is important to note that the discretization of optimization problems consisting of differential constraints may not be straightforward. In fact, insta-

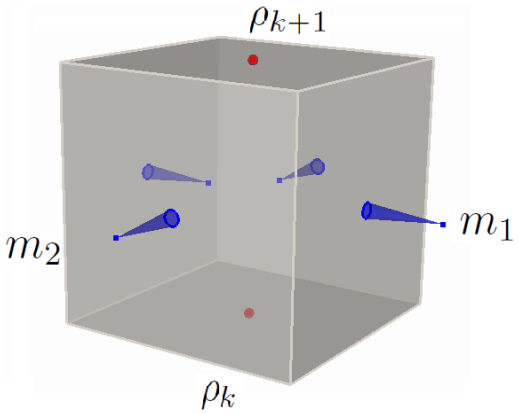


Figure 4.1: A box and the staggered grid in space-time. ρ is discretized at the beginning and end of each time interval and m is discretized on a staggered grid in space

bilities may arise if naïve discretization techniques are used. This topic has been studied extensively for the Stokes' problem and for flow in porous media (see for example [17, 101] and references within). To obtain stable discretizations using finite difference or finite volume, h-ellipticity criterion needs to be met. For finite element discretization the Ladysenskaja, Babuska, Brezzi (LBB) conditions need to be fulfilled [17]. Thus, special attention must be given to maintain stability in such problems.

Here we propose to use a finite volume approach in space-time. For simplicity of the discussion we describe the discretization in 2D and time. The extension to 3D is straightforward and will be experimented and evaluated in Section 5. We discretize space-time using $n_1 \times n_2 \times n_3$ rectangular cells. For simplicity we assume uniform spacing h, yet, this can be easily generalized. Consider the (i, j, k) box in space time, Ω_{ijk}. By integrating the constraint over Ω_{ijk} and using the Gauss theorem along with the midpoint integration

rule we obtain

$$h^{-3}\int_{\Omega_{ijk}} \frac{\partial \rho}{\partial t} + \mathrm{div} m \, dx\, dt = h^{-3}\int_{S_{ijk}} (\frac{\partial \rho}{\partial t} + \mathrm{div} m)\cdot n \, dS =$$
$$h^{-1}\Big((m_1^{i+\frac{1}{2},j,k} - m_1^{i-\frac{1}{2},j,k}) + (m_2^{i,j+\frac{1}{2},k} - m_2^{i,j-\frac{1}{2},k}) + (\rho^{i,j,k+\frac{1}{2}} - \rho^{i,j,k-\frac{1}{2}}) \Big)$$
$$+ \mathcal{O}(h^2).$$

This suggests a staggered grid discretization for the variables m_1, m_2 and ρ. Indeed, it is well known that a natural compact discretization of the divergence is staggered for problems such as fluid dynamics, mixed formulation of the Poisson equation and Maxwell's equations. Thus, we discretize m_1 at the points $(i+\frac{1}{2}, j, k)$, m_2 at $(i, j+\frac{1}{2}, k)$ and ρ at $(i, j, k+\frac{1}{2})$ with $i, j, k = 0, \dots n$. It is important to highlight that the use of a non-staggered grid for the discretization of the divergence may incur instabilities that manifest themselves due to the null space of the resulting system [101]. Also, note that the staggered discretization does not require boundary conditions for either m or ρ. This is an innate property of the problem at hand.

Finally, discretizing T and R and defining the vector

$$q = h^{-1}\left(T^\top, \underbrace{0, \dots, 0}_{n_1 n_2 (n_3 - 1)}, R^\top\right)^\top$$

we obtain a compact discretization of the constraint

$$D_1 m_1 + D_2 m_2 + D_3 \rho = q \tag{4.5}$$

where D_j are derivative matrices in x, y and t directions.

In order to discretize the objective function we use a combination of the midpoint and the trapezoidal methods. It is straightforward to verify that for sufficiently smooth m and ρ we have

$$\int_{\Omega_{ijk}} \frac{|m|^2}{\rho}\, dx\, dt = h^3 \frac{(m_1^{i+\frac{1}{2},j,k})^2 + (m_1^{i-\frac{1}{2},j,k})^2 + (m_2^{i,j+\frac{1}{2},k})^2 + (m_2^{i,j-\frac{1}{2},k})^2}{4\rho^{i,j,k-\frac{1}{2}}} +$$

$$h^3 \frac{(m_1^{i+\frac{1}{2},j,k})^2 + (m_1^{i-\frac{1}{2},j,k})^2 + (m_2^{i,j+\frac{1}{2},k})^2 + (m_2^{i,j-\frac{1}{2},k})^2}{4\rho^{i,j,k+\frac{1}{2}}} + \mathcal{O}(h^2)$$

We can write the objective in a matrix form as

$$\int_\Omega \frac{|m|^2}{\rho}\, dx\, dt = h^3 \left\{ \frac{A_1(m_1)^2 + A_2(m_2)^2}{A_t^+ \rho} + \frac{A_1(m_1)^2 + A_2(m_2)^2}{A_t^- \rho} \right\} + \mathcal{O}(h^2) \tag{4.6}$$

where A_1, A_2 are averaging matrices for the x and y directions and $A_t^\pm$ chooses the appropriate ρ on each side of the space-time cube. In our notation, the expressions m^2 and $1/\rho$ are applied point-wise. This expression can also be written as

$$\int_\Omega \frac{|m|^2}{\rho}\, dx\, dt = h^3 (A_s m^2) A_t(\frac{1}{\rho}) + \mathcal{O}(h^2) \tag{4.7}$$

where $A_s = [A_1, A_2]$ and A_t are averaging matrices in space and time with respect to the vectors $m = [m_1, m_2]$ and ρ.

Note that since we use a staggered grid, averaging is required for evaluation of the objective function, which is defined point-wise. Averaging can introduce instabilities into the discretization unless performed appropriately. Here we *first square then average* and *first divide then average*. These are crucial steps in obtaining stable discretization. For example, this guarantees that $f(m, \rho) \to \infty$ as $\rho \to 0$. This property is particularly important for the avoidance from obtaining non-physical negative densities.

4 Optimization

In this section we shortly describe the constrained optimization framework for the discrete problem

$$\begin{aligned} \min \quad & f(m, \rho) = h^3 (A_s m^2) A_t \left(\frac{1}{\rho}\right) \\ \text{s.t.} \quad & C(m, \rho) := D \left[m^\top; \rho^\top\right]^\top - q = 0 \end{aligned}$$

We use a Newton-type scheme for the solution of this constrained optimization problem. Such methods enjoy the advantage of being mesh independent [106], that is, the number of iterations needed for convergence are independent of the mesh. When such methods are combined with an appropriate multi-grid solver for the linear system one can obtain linear complexity convergence.

Since we consider large-scale 3D and 4D space-time problems, the use of inexact Newton solvers is inevitable. In particular, we consider the sequential quadratic programming (SQP) method for the solution of the problem, with inexact step computation of the linearized system. We can write the Lagrangian of the problem

$$\mathcal{L}(m,\rho) = (A_s m^2)A_t\left(\frac{1}{\rho}\right) + \lambda^\top (D\left[m^\top;\rho^\top\right]^\top - q)$$

where λ is a vector of Lagrange multipliers. From here it follows that the first order necessary conditions are

$$\nabla_m \mathcal{L} = 2MA_s^\top A_t\left(\frac{1}{\rho}\right) + (D_1, D_2)^\top \lambda = 0 \tag{4.8a}$$

$$\nabla_\rho \mathcal{L} = -KA_t^\top A_s(m^2) + D_3^\top \lambda = 0 \tag{4.8b}$$

$$\nabla_\lambda \mathcal{L} = D\left[m^\top;\rho^\top\right]^\top - q = 0 \tag{4.8c}$$

where $M := \mathrm{diag}(m)$, $K := \mathrm{diag}(\frac{1}{\rho^2})$.

In principle, this non-linear system of equations for m, ρ and λ can be solved by using Newton's method. However, since the mixed second derivative of the objective function introduces negative terms in the Hessian, a Gauss-Newton approximation to the Hessian of the objective is therefore sought [51].

The saddle point system which is solved inexactly at each SQP iteration can be written as

$$\begin{pmatrix} \hat{A} & D^\top \\ D & 0 \end{pmatrix}\begin{pmatrix} \delta w \\ \delta\lambda \end{pmatrix} = -\begin{pmatrix} \nabla_w \mathcal{L} \\ \nabla_\lambda \mathcal{L} \end{pmatrix}, \tag{4.9}$$

where $\hat{A} := \widehat{\nabla^2 \mathcal{L}}$ is a symmetric positive definite approximation of the Hessian of the objective function

$$\hat{A} = \begin{pmatrix} 2\text{diag}(A_s^\top A_t(\frac{1}{\rho})) & 0 \\ 0 & 2\text{diag}(A_t^\top A_s(m^2))\text{diag}(\frac{1}{\rho^3}) \end{pmatrix}.$$

The saddle point system is symmetric and indefinite [10]. It is also ill-conditioned, large and sparse. We shall explore two solution approaches. In the first approach, we solve the above system inexactly using preconditioned GMRES [97] to obtain an update for the variables

$$w \leftarrow w + \alpha \delta w \tag{4.10}$$

$$\lambda \leftarrow \lambda + \alpha \delta \lambda \tag{4.11}$$

where α is determined by a line search.

Inexactness in the solution of the linear system can be incorporated into the SQP framework using a similar method to the one proposed in [19]. We determine our stopping criteria for GMRES based on a so-called filter [46], and terminate the linear iterations whenever the step is accepted by the filter. Thus, the GMRES solver has two stopping criteria and *both* have to be fulfilled in order to stop the iteration. First, the prescribed tolerance is achieved and second, the solution is accepted by the filter. Only if both criteria are fulfilled is the iteration stopped; for further details see Chapter 5.

Since the saddle point system (4.9) is ill-conditioned, the use of a suitable preconditioner is essential. We consider a preconditioner based on the Schur complement (see [10] and references within)

$$\begin{pmatrix} \hat{A} & D^\top \\ 0 & S \end{pmatrix} \begin{pmatrix} \delta w \\ \delta \lambda \end{pmatrix} = - \begin{pmatrix} \nabla_w \mathcal{L} \\ \nabla_\lambda \mathcal{L} + D\hat{A}^{-1} \nabla_w \mathcal{L} \end{pmatrix},$$

where $S = -D^\top \hat{A}^{-1} D$ is the Schur complement of the system. Using this preconditioner, preconditioned GMRES iterations converge within three iterations in exact arithmetic [10]. The difficulty lies in inverting the Schur complement.

We should notice that since $\hat{A}$ is a diagonal matrix, its inverse can be trivially computed. Moreover, the Schur complement is a generalized Laplacian in space-time of the form $\nabla^{\text{st}} \cdot \sigma(\nabla^{\text{st}} \cdot)^*$ with coefficients $\sigma = \text{diag}(\hat{A}^{-1})$ which depends only on ρ and m^2. Therefore, the Schur complement is suitable for multi-grid methods. Since we consider models of high density contrast, we should pursue a multi-grid algorithm that can account for jumping coefficients. In our algorithm we applied a single V-cycle of a Ruge-Stüben style algebraic multi-grid algorithm [95] for the solution of the Schur complement part of the preconditioner.

While the above approach grants optimal complexity in theory it may be slow in practice. This is due to the large setup time and memory requirements of algebraic multi-grid methods. A simpler approach that is not optimal but can converge quickly for smaller mesh sizes is to solve the system (4.9) by elimination. We first eliminate δw, and obtain a symmetric positive definite system for $\delta\lambda$

$$-S\delta\lambda = \nabla_\lambda \mathcal{L} - D\hat{A}^{-1}\nabla_w \mathcal{L} \tag{4.12}$$

which is solved by the conjugated gradients (CG) method with a Symmetric Gauss-Seidel preconditioner. We then compute the update for w by

$$\delta w = -\hat{A}^{-1}(D^\top \delta\lambda + \nabla_w \mathcal{L})$$

and update the variables as in (4.11).

In numerical experiments we have found out that the "break even" point between the approaches is when the mesh size is roughly 64^3.

5 Numerical Experiments

In our numerical experiments we would like to examine two aspects of the algorithm. Firstly, we would like to test the algorithm for the L_2 distance. Here, we consider the two approaches elucidated above: a combination of GMRES and multi-grid within the Schur complement preconditioner versus the solution of the reduced order system by elimination and preconditioned CG. Secondly, we would like to test the behavior of the algorithm when the norm of the MK functional, p, is gradually reduced from two to one. Note that for the L_1 norm we have that

$$f(m,\rho) = \int_\Omega \int_0^1 |m| \, dt \, dx$$

and thus the problem is not well-defined [32]. We therefore observe at the limit $p \to 1$ of

$$f(m,\rho) = \int_\Omega \int_0^1 \frac{|m|^p}{\rho^{p-1}} \, dt \, dx$$

and examine the behavior of the algorithm as we gradually reduce the norm p to one.

As a set of model problems we considered the 2D images displayed below in Figure 4.2. For the initial density distribution T an image of four circle quarters, one at each corner was deployed. The final distribution R was represented by an image of a circle positioned in the center, where we verified that the two images have the same total volume, see Figure 4.2.

We first consider grid size of 42×42 in space and 42 discretization steps in time, with density contrast of 10. We shall later increase the grid fineness to examine the performance of our multi-grid scheme. In Figure 4.3 the solution for density distributions with contrast 10 and for $p = 2$ is displayed; for each time step excluding the initial and final stage, from left to right, top to bottom.

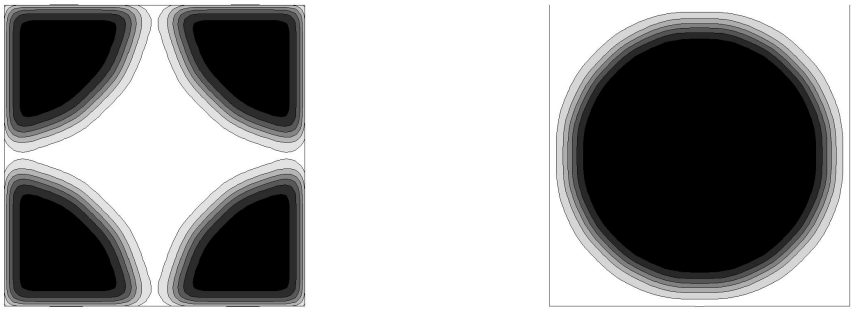


Figure 4.2: Left: initial density distribution image T, right: final density distribution image R

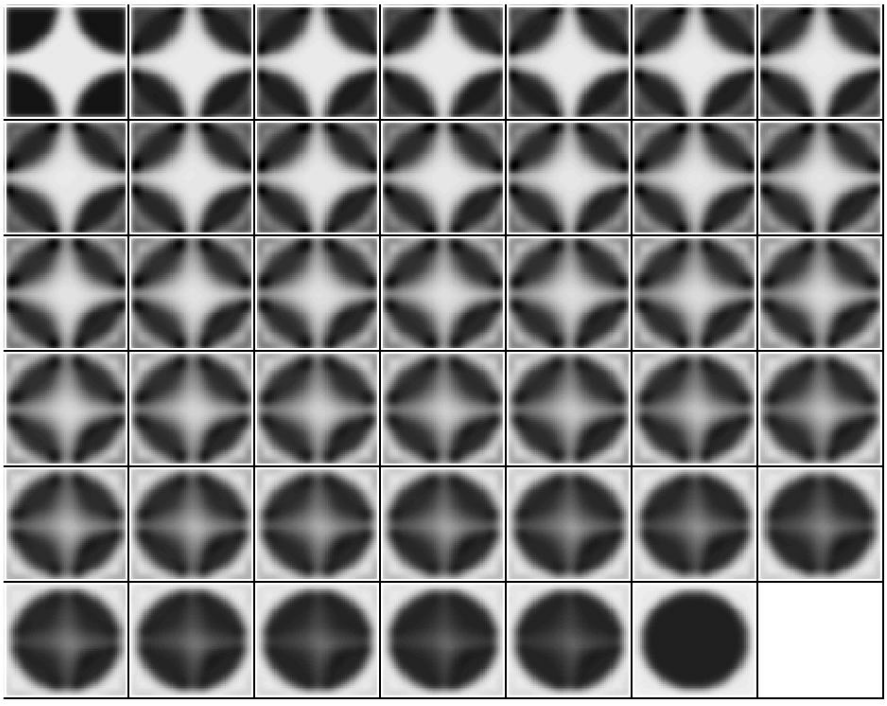


Figure 4.3: Solution for density distribution for $p = 2$

Next we to check the optimality of our algorithm with respect to mesh independence behavior. The overall algorithm is optimal if its iteration count is independent with respect to the problem size, that is, if each problem is of linear complexity with respect to the number of unknowns. For this test, we considered a grid hierarchy; starting from a coarse grid of $16 \times 16 \times 10$ (in space and time) through a grid of $32 \times 32 \times 20$ to a grid of $64 \times 64 \times 40$. We set the tolerance for the outer (SQP) iteration to 10^{-4}, and as mentioned before, each step was computed inexactly up to a relative residual error tolerance of 10^{-1}.

In Table 4.1 we record the number of SQP iterations required for convergence on different grid dimensions to the desired tolerance for $p = 2$. As can be deduced from the table, the SQP algorithm is almost optimal: although the number of variables increased by a factor of 2^3 between the different grid sizes, the number of SQP iterations required for convergence was altered only mildly. In order to verify that the overall algorithm is also optimal, we need

Grid Size	SQP iterations
$16 \times 16 \times 10$	16
$32 \times 32 \times 20$	13
$64 \times 64 \times 40$	14

Table 4.1: Number of SQP iterations required for convergence on different grid sizes, for $p = 2$ and density contrast 10.

to check that the linear solver is of linear complexity. Thus, the convergence rate of the multi-grid preconditioned linear solver is of particular interest. The performance of GMRES with a multi-grid preconditioner is presented in Table 4.2. In this table the inner iteration count is compared for the first iteration of SQP, an iteration in the midst of the optimization process,

and the last iteration of the process. Table 4.2 shows a small increment in

Grid Size	1^{st}	Midst	Last
$16 \times 16 \times 10$	2	3	2
$32 \times 32 \times 20$	2	5	3
$64 \times 64 \times 40$	2	6	7

Table 4.2: Comparison of numbers of preconditioned GMRES iterations required for convergence to the tolerance of 10^{-1} at 1^{st}, midst and last iteration of the SQP algorithm.

the number of iterations when the mesh is refined. Although the number of iterations is increasing, the increment is still negligible in comparison to the increase in mesh size. It is interesting to note that the increase in the number of iterations of the linear solver is not due to a deterioration of the multi-grid solver. The reason for the increase is that the GMRES solution for a tolerance of 10^{-1} was determined to be inappropriate for descent by the filter. Thus more iterations in the GMRES solver were needed to fulfill the filter criteria.

In a third experiment we test our algorithm for a higher density contrast. For that purpose, we consider the 2D images as before, but this time with a density contrast of 100. Again, our setup involves an hierarchy of models of increasing grid sizes, starting from 16×16 in space and 8 in time and up to 64×64 in space and 32 time steps. As before, we set the tolerance for the outer (SQP) iteration to 10^{-4}, and each step is computed inexactly up to a relative residual error tolerance of 10^{-1}. In this case, we take the multi-level approach, where we interpolate the solution from the coarser grid to finer one, such that the initial guess on the finer grid is closer to the optimal

solution. In Table 4.3 we record the number of SQP iterations required for convergence on different grid sizes to the desired tolerance for $p = 2$.

Grid Size	SQP iterations
$16 \times 16 \times 8$	28
$32 \times 32 \times 16$	12
$64 \times 64 \times 32$	14

Table 4.3: Number of SQP iterations required for convergence on different grid sizes, for $p = 2$ and density contrast 100.

For this problem, the number of GMRES iterations was consistently two throughout the optimization process, for all grid sizes. As can be seen, the algorithm performs well for a large density contrast. No other algorithm known to us is capable of solving this problem efficiently.

In a fourth experiment, we test our algorithm on 4D problems (3D in space plus time). Since we do not have an effective 4D algebraic multi-grid code the linear system is solved by elimination as described in Section 4. Again, we consider an hierarchy of grids, starting from $16 \times 16 \times 16$ in space and 8 time steps, through $32 \times 32 \times 32$ with 16 times step up to a grid of size $64 \times 64 \times 64$ in space and 32 time steps. In Table 4.4 we present the number of SQP iterations required for convergence on different grid dimensions to the desired tolerance for $p = 2$. In Table 4.5 we summarize the number of preconditioned CG iterations required for convergence to a tolerance of 10^{-4} at different stages of the SQP iterations. The table demonstrates the mesh independence property of our algorithm even in 4D settings. Since we use symmetric Gauss-Seidel as a preconditioner our linear solver is not

Grid Size	SQP iterations
$16 \times 16 \times 16 \times 8$	10
$32 \times 32 \times 32 \times 16$	11
$64 \times 64 \times 64 \times 32$	10

Table 4.4: Number of SQP iterations required for convergence on different grid sizes, for $4D$ problem, with $p = 2$ and density contrast 10.

mesh independent. Nonetheless, we are still able to obtain solutions of 4D problems with a moderate computational effort.

Grid Size	1^{st}	Midst	Last
$16 \times 16 \times 16 \times 8$	21	38	32
$32 \times 32 \times 32 \times 16$	39	86	60
$64 \times 64 \times 64 \times 32$	71	187	137

Table 4.5: Comparison of numbers of preconditioned CG iterations required for convergence to a tolerance of 10^{-4} at the 1^{st}, midst and last iteration of the SQP algorithm.

Finally, we test the performance of our algorithm when gradual reduction of the norm p from two to one is imposed. We shall address a 2D problem in space and time, with grid size 64×64 with 32 time steps, and density contrast 10. We elect to solve the reduced system using preconditioned CG. Simple continuation is used to obtain a solution with $p \to 1$. We start with $p = 2$ and execute a single SQP iteration. We then reduce p and repeat the process, starting from the previous solution. We use 100 continuation steps to get from $p = 2$ to $p = 1.01$. In Figure 4.4 the convergence of the primal (equation (4.8c)) and dual feasibility (equations (4.8a) and (4.8b))

are presented as a function of the norm p. As expected, the norms of the primal and dual feasibility are reduced as we progress with the continuation and decrease the p. Although we have used a simple continuation, this was sufficient in order to obtain an approximation to the L_1 problem. More efficient continuation methods for the problem are not considered in this study.

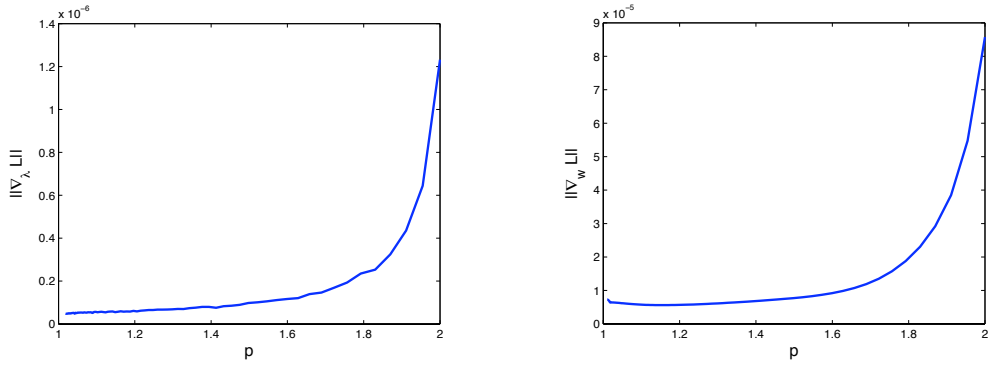


Figure 4.4: Left: Norm of primal feasibility as a function of p, right: Norm of dual feasibility as a function of p

6 Conclusions and Summary

A novel numerical framework for the 3D and 4D optimal mass transport problems using the fluid mechanics formulation is proposed. The new framework treats the time dependency of the PDE as yet another spatial dimension in a staggered grid discretization. This formulation does not require explicit prescription of boundary conditions for the density distribution. Instead, these boundary conditions are introduced implicitly as part of our optimization scheme. In order to account for large density contrast in the models, we have incorporated an algebraic multi-grid approach to choose a preconditioner for

the solution of the linearized system. This approach proved to be optimal with respect to the mesh size. We have also considered a distance function for the general norm p, where $1 \leq p \leq 2$. We started our iterations at $p = 2$ and slowly relaxed the norm until $p = 1.01$, while conducting only a few iterations for each p.

From an imaging perspective, the transformation of the initial density distribution closely resembles the final distribution. This indicates a good correspondence and overall good performance of the proposed algorithm. We believe that the proposed algorithm can readily be disseminated for a large range of applications of various objectives and high density contrasts.

Chapter 5

Filter-Based Inexact SQP Method

1 Background and Motivation

Sequential Quadratic Programming (see Section 6 in Chapter 2) is regarded among the most successful and effective constrained optimization methods practiced to date. Among its virtues, probably the most profound, is its Newton-like inherited quadratic convergence rate, for exact arithmetic and small residuals. Moreover, the method is relatively easy to implement (whenever 2^{nd} order derivatives are available) and typically performs robustly even when its mild set of prerequisite assumptions are slightly violated.

As asymptotically effective the method is, in its original form it may often converge slowly (time-wise), especially for large-scale problems. The main computational bottle-neck emerges from the redundant computational effort devoted to an exhaustively accurate solution of the intermediate linearized KKT system (SQP sub-problem) within each SQP iteration. For medium to large-scale problems, this operation by-far dominates all other computational operations, due to the frequent ill-conditioning and dimensions of the KKT

systems. It is important to note that this attribute has also been inherited from its unconstrained counterpart approach, the Newton method. Inspired by solution approaches of the latter for large-scale problems (e.g. truncated Newton, inexact Newton [83, 30]), accuracy-controlled iterative linear solvers have been recently incorporated in the form of *inexact* SQP. In particular, Krylov-subspace linear solvers such as GMRES [97] are commonly in use.

For the sake of completeness, it is important to highlight at this point, that inexactness can be introduced in different levels. First, the function or the constraints can be represented inexactly; this approach is often known as approximated physics (in the context of PDE based problems), and bears high level of relevance when evaluation of the exact function (or constraint) is computationally intensive [3]. Second, derivatives may be given approximately, for example quasi-Newton methods such as Gauss-Newton or BFGS [44, 86]. Third, the solution of the sub-problem can be computed inexactly. Lastly, further degree of inexactness can be introduced in the globalization process. This is often performed by an inexact line-search or trust region [99, 26] approach.

While approximated physics is obviously problem dependent, and both quasi-Newton-like optimization approaches as well as inexact line search methods have been extensively studied, the notion of the inexact solution of the sub-problem has been under-researched. Moreover, due to the aforementioned recognized computational burden associated with the solution of the sub-problem, in this chapter we shall concentrate in that aspect of inexactness.

The elegant resolution of this problem by accuracy-controlled iterative linear solvers offers a trade-off between accuracy of the linearized systems and computational effort invested. Yet, it also introduces a new problem: while mild degradation of solution accuracy may yield computational savings, severe degradation is likely to lead to breakdown of the algorithm and con-

vergence failure. Thus, a central research question would be to devise a mechanism for the definition of the linear solution tolerance, such that convergence of the inexact SQP algorithm is guaranteed. While for inexact Newton methods, definition of such a criteria is concerned merely with retaining optimality, for the constrained optimization case, one needs to account also for maintaining feasibility.

A variety of methods for inexact step computation for constrained optimization have been proposed recently. Jäger and Sachs [65] described an inexact reduced SQP method in a Hilbert space. Lalee, Nocedal, and Plantenga [71], Byrd, Hribar, and Nocedal [21], and Heinkenschloss and Vicente [61] proposed composite step trust-region approaches where the step was computed as an approximate solution to an SQP sub-problem. Similarly, Walther [107] provides a composite step method that allows incomplete constraint Jacobian information. The approach used by Byrd, Curtis and Nocedal [19] approach had some features in common with the algorithms of Biros and Ghattas [12, 13] Haber and Ascher [50], and Prudencio, Byrd and Cai [89]. They followed a full-space SQP method and perform a line-search to promote convergence. The importance of their work can mainly be attributed to the fact that they were the first to present conditions that guarantee global convergence of inexact SQP steps.

Inspired by the pioneering research of Byrd, Curtis and Nocedal, where full-space SQP formulation was pursued, we decided to follow the same track. However, while in their approach a stopping criteria for the iterative linear solver was based on reduction in a merit function, our approach is based on the filter-method [46], such that the computed step makes sufficient progress towards the solution on the linearized problem. To guarantee convergence of the original problem, we then introduce a line-search by using a filter over the non-linear problem.

In the next section an introduction to the filter method is brought. We

then proceed by introducing the inexact SQP formulation, and further we outline a convergence proof for the proposed formulation. In the next section, numerical results are presented, and lastly we summarize this study.

Other than the numerical studies presented in this chapter, the proposed inexact SQP approach was incorporated in Chapter 4 for the solution of the optimal mass transport problem.

2 Filter Method

The filter method was first introduced by Fletcher in a plenary talk at the SIAM Optimization Conference in Victoria, 1996; a more comprehensive description of the method can be found in [45]. The method is an alternative for the common use of a penalty (merit) function in non-linear programming, which can exhibit slow convergence for large penalty parameters (denoted by α in Section 7). The goal was to develop a method which will serve as a global convergence safeguard while only mildly intervene with Newton's method. Instead of combining the objective and the constraint violation into one function, the filter method views the constrained optimization problem (2.1), as a bi-objective optimization problem that minimizes $f(x)$ and $h(x) = \|c(x)\|_1$. The second objective is more transparent (and perhaps more important), since we must guarantee that the solution satisfies $h(x^*) = 0$. A filter is defined as a list of pairs $(h(x_l), f(x_l))$ such that no pair dominates another pair, where a point x_k is defined to dominate a point x_l if and only if $f(x_k) \leq f(x_l)$ and $h(x_k) \leq h(x_l)$.

For example, a general outline of a trust-region SQP [26] is as follows [45]. At iteration $k = 0$, we initialize the filter $\mathcal{F}_k = (U, -1)$, where U is an initial upper bound on the constraint violation. We proceed by accepting only steps that are not dominated by the current filter. For an acceptable point, we set $x_{k+1} = x_k + s$, and possibly increase the trust-region radius and update the

filter by adding the previous point and removing any dominated entries. If, on the other hand, the step is dominated by the current filter, then we reject the step, i.e. set $x_{k+1} = x_k$, reduce the trust-region radius, and proceed to the next iteration.

In order to ensure convergence several refinements are required:

1. Filter Envelope - To avoid convergence to non-stationary infeasible limit points where $h_* > 0$, we add an envelope around the current filter. Now, a new iterate is acceptable if, $\forall\ l \in \mathcal{F}_k$,

$$h_{k+1} \leq \beta h_l \quad \text{or} \quad f_{k+1} \leq f_l - \gamma h_{k+1},$$

   where $0 < \beta, \gamma < 1$ are constants. This sloping envelope [23, 24] provides slightly stronger convergence properties. If an infinite number of points are added to the filter, and $f(x)$ is bounded below, then the limit point must be feasible, see Lemma 1 in [23]. In Figure 5.1 a typical filter is illustrated, where the shaded area shows the region dominated by the filter entries and the dashed line represents the filter's envelope.

2. Sufficient Reduction - The filter alone cannot ensure convergence to stationary points. For an instance, if the sequence satisfies $h_{k+1} \leq \beta h_k$, then the iterates could converge to an arbitrary feasible point. Therefore, if the constraint violation becomes small, a sufficient reduction condition similar to unconstrained optimization is then enforced. Let us denote the predicted reduction by $\Delta q_k := -\Delta f_k^\top s - \frac{1}{2} s^\top H_k s$ and introduce the following switching condition:

$$\begin{gathered} \text{if } (\ \Delta q_k > 0) \text{ then} \\ \text{check } f_k - f_{k+1} \geq \sigma \Delta q_k, \end{gathered}$$

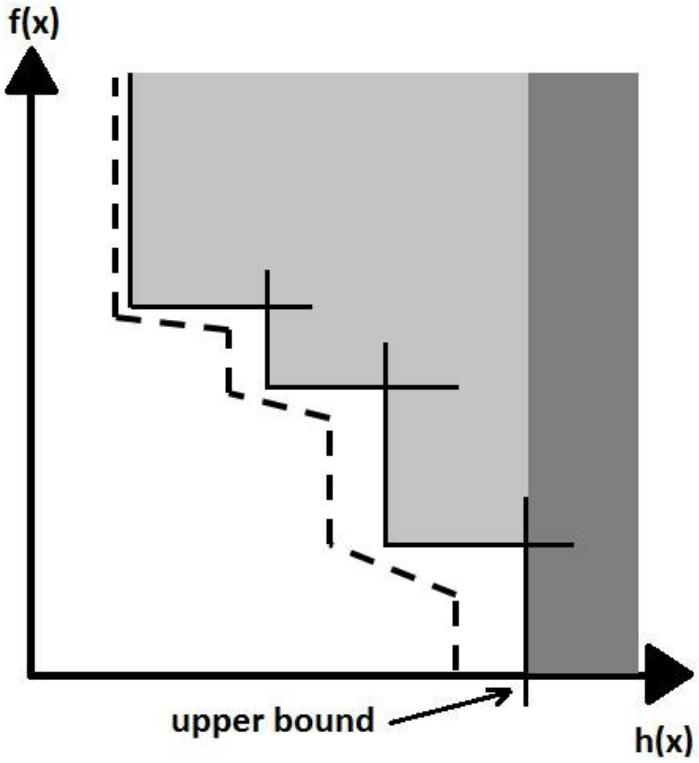


Figure 5.1: Filter method - shaded area represent area that is dominated by the filter entries. All pairs $(f(x), h(x))$ that are below and to the left of the dashed line are acceptable to the filter.

where $\sigma \in (0, 1)$ is a constant.

3. Feasibility Restoration - By reducing the trust-region radius, the QP may become inconsistent. The inconsistency indicates that the current point is too far from the feasible set to make meaningful progress to optimality or that the search direction subproblem is not determined properly or at least ideally. Hence an SQP-like algorithm that minimizes the constraint violation $h(x)$ can be invoked. The restoration phase is terminated once an acceptable point has been found.

3 Constrained Optimization Framework

We shall focus now on the following equality constrained optimization problem

$$\min_x \ f(x) \ \text{ s.t. } c(x) = 0, \tag{5.1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is convex and the constraints $c : \mathbb{R}^n \to \mathbb{R}^m$ are sufficiently smooth.

The Lagrangian of the system can be written as

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^\top c(x).$$

Accordingly, necessary conditions for an optimal solution (1^{st} order conditions - KKT conditions) are brought by

$$\begin{aligned} \mathcal{L}_x &= g(x) + B(x)^\top \lambda = 0 \\ \mathcal{L}_\lambda &= c(x) = 0, \end{aligned}$$

where $B(x) := \nabla c(x)$ is the Jacobian of the constraints, $g(x) := \nabla f(x)$ is the gradient of the objective function, and λ is the Lagrange multipliers vector.

Using Newton's method, a KKT system can be constructed. Such a system needs to be solved at each SQP iteration to obtain a Newton step

$$\begin{pmatrix} H_k & B_k^\top \\ B_k & 0 \end{pmatrix} \begin{pmatrix} d_k \\ \delta_k \end{pmatrix} = - \begin{pmatrix} g_k + B_k^\top \lambda_k \\ c_k \end{pmatrix}, \tag{5.2}$$

where H_k is the Hessian $\nabla^2_{xx}\mathcal{L}(x_k, \lambda_k)$ or its symmetric approximation. The new iterate is updated as follows

$$\begin{aligned} x_{k+1} &= x_k + \alpha d_k \\ \lambda_{k+1} &= \lambda_k + \alpha \delta_k \end{aligned}$$

where α is computed using a line-search.

4 Inexact SQP Framework

A sequential quadratic programming algorithm is used for the solution of the equality constraint problem (5.1). At each SQP iteration, the linear system (5.2) needs to be solved. The symmetric KKT system is indefinite and sparse for most partial differential equations based problems. The solution of this system is obtained by preconditioned Krylov subspace methods (e.g preconditioned GMRES [97]), which guarantee that the solution is found within a bounded number of iterations, as the dimension of the system, in exact arithmetic. However, for very large linear systems, this upper bound on the number of iterations is often not practical. In this work, we propose to solve the linear system inexactly, similarly to [19]. However, rather than relying on a reduction of a merit function as a stopping criterion to the linear iterations, we embed a criteria based on the filter method.

Therefore, at each SQP iteration we solve the perturbed system

$$\begin{pmatrix} H_k & B_k^\top \\ B_k & 0 \end{pmatrix} \begin{pmatrix} d_k \\ \delta_k \end{pmatrix} = - \begin{pmatrix} g_k + B_k^\top \lambda_k \\ c_k \end{pmatrix} + \begin{pmatrix} \rho_k \\ r_k \end{pmatrix}. \tag{5.3}$$

Here, we shall consider a three-dimensional filter based upon primal feasibility, dual feasibility and the objective function. The axes are defined as follows: the x-axis is $h_k(x) := \|c(x_k)\|_2$, the y-axis is $p_k(x, \lambda) := \|g(x_k) + B(x_k)^\top \lambda_k\|_2$ and the z-axis is $f(x_k)$, where $f(x_k)$ and $g(x_k)$ are the objective function and its derivative (gradient) at iterate k, and $c(x_k)$ and $B(x_k)$ are the constraints and its derivative (Jacobian) at iterate k.

The GMRES step is accepted by the filter only if the new point x_{k+1} is not dominated by the current filter. By introducing a filter envelope [23, 24], the new iterate is acceptable by the filter if, $\forall\ l \in \mathcal{F}_k$,

$$\begin{aligned} h_{k+1} &\leq \beta h_l \\ \text{or } p_{k+1} &\leq \beta p_l \\ \text{or } f_{k+1} &\leq f_l - \gamma(h_{k+1} + p_{k+1}), \end{aligned} \tag{5.4}$$

where $0 < \beta, \gamma < 1$ are some constants.

Algorithm 5: Inexact SQP with filter

Input: $(x_0, \lambda_0), \tau, \bar{\alpha}, k \leftarrow 0, \mathcal{F}_0$, $0 < \hat{\beta} < \beta$, $0 < \sigma < 1$

Output: (x_k, λ_k)

if $h_0 > 0$ *and* $p_0 > 0$ **then**
  $\mathcal{F}_0 \leftarrow (x_0, \lambda_0)$
$k \leftarrow 0$
// SQP
while *not converged* **do**
  form the KKT matrix and the rhs
  $j \leftarrow 0$, converged $\leftarrow$ FALSE
  // GMRES
  while *not converged* **do**
    compute update step (d_{kj}, δ_{kj})
    if (ρ_{kj}, r_{kj}) *satisfies* $\|\rho_{kj}\|_2 \leq \hat{\beta} p_l$ *OR* $\|r_{kj}\|_2 \leq \hat{\beta} h_l$ *for* $\forall l \in \mathcal{F}_k$ **then**
      accept the step (d_{kj}, δ_{kj})
      converged $\leftarrow$ TRUE
  $d_k \leftarrow d_{kj}$, $\delta_k \leftarrow \delta_{kj}$
  converged $\leftarrow$ FALSE
  $j \leftarrow 0$, $\alpha_j \leftarrow 1$
  // Line Search
  while *not converged AND* $\alpha_j \geq \bar{\alpha}$ **do**
    if $(x_k, \lambda_k) + \alpha_j (d_k, \delta_k)$ *is acceptable by the filter* **then**
      if $\Delta q_k(1) > 0$ **then**
        if $f_k - f_{k+1} \geq \sigma \Delta q_k(\alpha)$ **then**
          accept the step $(x_{k+1}, \lambda_{k+1}) \leftarrow (x_k, \lambda_k) + \alpha_j (d_k, \delta_k)$
          converged $\leftarrow$ TRUE
      else
        accept the step $(x_{k+1}, \lambda_{k+1}) \leftarrow (x_k, \lambda_k) + \alpha_j (d_k, \delta_k)$
        converged $\leftarrow$ TRUE
    $\alpha_{j+1} \leftarrow \alpha_j / 2$
  if $h_{k+1} > 0$ *AND* $p_{k+1} > 0$ **then**
    $\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k \cup (x_{k+1}, \lambda_{k+1})$

where $\Delta q_k(\alpha) = -\alpha g_k^\top s - \frac{\alpha^2}{2} s^\top H_k s$.

5 Numerical Examples

In this section we shall demonstrate the utility of the proposed approach to the solution of two PDE constrained optimization inversion problems. The first of which concerns with the famous generalized Laplace PDE, whereas the latter is the Monge-Kantorovich optimal mass transport problem which was introduced in Chapter 4. In our numerical tests we shall conduct a comparison between the linear solver iteration count of the proposed methodology versus the conventional SQP algorithm.

5.1 Generalized Laplace Inverse Problem

We would like to solve the following optimization problem [60] with partial differential equations as a constraints

$$\begin{aligned} \min_{y,u} \quad & \tfrac{1}{2}\|y-d\|^2 + \alpha R(u) \\ \text{s.t.} \quad & A(u)y - q = 0 \end{aligned}$$

Here, finite difference discretization is practiced, where u is discretized on cell centers and y is on the nodes. For simplicity of presentation, we address the one dimensional case, and thus formulate the problem as

$$\begin{aligned} \min_{y,u} \quad & \tfrac{1}{2}\|y-d\|^2 + \tfrac{\alpha}{2}\|Lu\|^2 \\ \text{s.t.} \quad & A(u)y - q = 0 \end{aligned} \tag{5.5}$$

where L is the Laplacian matrix, and the following constraint is imposed

$$\frac{d}{dx} e^u \frac{d}{dx} y - q = 0.$$

This problem appears in many applications, among the popular once are electrical resistivity tomography and hydrology, see [64] and reference therein. Then we can write the discretized matrix A as

$$A(u) = D^\top \text{diag}(e^u) D$$

where D is discretized gradient operator (with Neumann boundary conditions).

By employing an inexact Newton-type method, we can write the Lagrangian

$$\mathcal{L}(y, u, \lambda) = \frac{1}{2}\|y - d\|^2 + \frac{\alpha}{2}\|Lu\|^2 + \lambda^\top (A(u)y - q).$$

The necessary optimality conditions (1^{st} order conditions - KKT conditions) are given by

$$\begin{aligned} \mathcal{L}_y &= y - d + A(u)^\top \lambda = 0 \\ \mathcal{L}_u &= \alpha L^\top L u + G^\top \lambda = 0 \\ \mathcal{L}_\lambda &= A(u)y - q = 0 \end{aligned}$$

where $G := D\text{diag}(Dy)\text{diag}(e^u)$

Using Newton's method, a KKT system can be formed. Specifically, at each Newton step it is necessary to solve the linear system

$$\begin{pmatrix} I & 0 & A^\top \\ 0 & \alpha L^\top L & G^\top \\ A & G & 0 \end{pmatrix} \begin{pmatrix} \delta y \\ \delta u \\ \delta \lambda \end{pmatrix} = - \begin{pmatrix} \mathcal{L}_y \\ \mathcal{L}_u \\ \mathcal{L}_\lambda \end{pmatrix} \tag{5.6}$$

For our numerical tests a 2D problem was addressed. The dimensions of the grids were 32×32, 64×64 and 128×128, and the solution tolerance for the outer iteration was prescribed to be 10^{-4}. The solution tolerances for the inner problem were 10^{-8} and 10^{-1} for the exact and inexact SQP ($\hat{\beta}$ in Algorithm 5) respectively. Convergence properties of the SQP method using regular GMRES linear solver mandates exact step computation. Therefore,

in our comparison, we benchmark the performance of such conventional SQP method, that is a small tolerance is prescribed for the GMRES linear solver, and its performance with those of the proposed filter-based inexact SQP method.

In Table 5.1 a comparison of the number of linear iterations for exact and filter-based inexact SQP method is provided.

Table 5.1: Comparison of the number of linear iterations for exact and the filter-based inexact SQP method for 2D generalized Laplace inverse problem

Grid Size	Exact SQP (iters)	Inexact SQP (iters)
32×32	54	29
64×64	40	24
128×128	48	27

The results above display a significant reduction in the overall number of iterations. It is important to note that an arbitrary reduction of the linear solver tolerance within the conventional SQP scheme to 10^{-1} results in a line-search break. This behavior is of course indeed in agreement with the theory, as convergence is only guaranteed for conventional SQP when an exact step is computed. That is in agreement with the theory, which requires exact step calculation.

5.2 Optimal Mass Transport

In this test we have incorporated the proposed filter-based inexact SQP solution for the Monge-Kantorovich optimal mass transport problem addressed in Chapter 4. The constraints in this problem are governed by time-dependent mass preservation PDEs in the fluid dynamics formulation. The only modification upon the formulation presented in Section 4, is that here, we solve

equation (4.1) using the proposed filter-based inexact SQP formulation. For this numerical test, the dimensions of the grids were $16 \times 16 \times 8$, $32 \times 32 \times 16$ and $64 \times 64 \times 32$ in space and time, density contrast of 10, and solution tolerance for the outer iteration prescribed to be 10^{-4}. As before, in Table 5.2 we have compared the performance of exact SQP with an inner tolerance of 10^{-8} with the filter-based inexact SQP method and an inner tolerance of 10^{-1}

Table 5.2: Comparison of the number of linear iterations for the exact and the filter-based inexact SQP method for the 3D Monge-Kantorovich problem.

Grid Size	Exact SQP (iters)	Inexact SQP (iters)
$16 \times 16 \times 8$	33	22
$32 \times 32 \times 16$	61	49

For these problems, similar reduction in the overall number of iterations can be observed.

5.3 Conclusions

In this section a new inexact SQP method was proposed. The method relies on the filter method to determine termination of the inner linear solution process and thereby provides superior convergence speed upon the use of conventional SQP.

One of the advantages of this new approach is that it can readily replace any present SQP algorithm, and be used as a plug in. This virtue enables the methods great dissemination for a broad range of problems.

As highlighted in the introduction (Section 1), addressing inexactness in the solution of the inner linear system concurs only one of a multitude of options for imposition of inexactness for non-linear optimization. There is

still quite a significant breadth in exploring other potential directions related to different aspects of inexactness, as well as to investigate the interplay between them.

The author would like to acknowledge at this point Sven Leyffer, Andrew Conn, Andreas Wächter and Frank Curtis for their invaluable input and advice in the course of the development and evaluation of the proposed algorithm.

Chapter 6

Summary and Future Work

In this thesis, numerical techniques for large-scale PDE constrained optimization problems were developed. Specifically, the volume preserving image registration and the optimal mass transport problems were considered.

The first study is aimed at the incorporation of volume constraints for 3D image registration. Such constraints play an important role in medical applications, where the associated volume represents bodily tissues, and thereby is subject to physical limitations regarding its ability to shrink or expand. Since the constraint involves a determinant of a differential term, it forms a highly non-linear inequality. This characteristic poses a great numerical challenge that requires special care during the discretization process as well as a tailored optimization scheme. Stable discretization was obtained by discretization over a staggered grid, which effectively eliminates the need for long distance differences. By introducing slack parameters the problem was decomposed into an equality constrained problem along with simple bound constraints. In order to achieve fast convergence, an augmented Lagrangian formulation was favored. The resulting KKT system was solved by elimination and therefore entailed a symmetric positive definite system which was solved by a preconditioned conjugated gradients algorithm. Local Fourier

analysis of the system established the appropriateness of multi-grid schemes as optimal, mesh- independent preconditioners for the problem under hand. Thus, V - cycle geometric multi-grid incorporating a custom-made Vanka-type smoother was developed. The algorithm displayed visually viable registration solutions which were validated to be feasible. Convergence results reassessed the expected mesh-independent behavior. It is important to note, that these results were obtained for real data in realistic settings. As the role of image registration in medical diagnosis is growing rapidly, we hope that the proposed algorithm will serve practitioners in their routine diagnosis work. This study was published in the Numerical Linear Algebra with Applications journal, 2009 and presented at the SIAM Imaging Science conference, 2010 in Chicago, the Copper Mountain Multi-grid conference, 2009 and as a seminar at the IBM TJ Watson Research Center, 2009.

In the second study the optimal mass transport problem was addressed. This important problem has applications in econometrics, fluid dynamics, automatic control, transportation, statistical physics, shape optimization, expert systems, and meteorology. In this problem the objective was minimization of an energy term, whereas the space-time transport PDEs served as constraints. The formulation we considered introduce additional artificial time dimension that increases the dimensionality of the problem on the one hand. On the other hand, it successfully replaces a generally, non-convex problem by a convex one, as well as avoids some other complexities such as boundary conditions and numerical difficulties when densities are small. Since the time dependency of the PDE was treated as a spatial dimension, the resulting formulation resembles flow in porous media. Finite volume discretization formulations require special care in order to preserve numerical stability of the solution. Thus, three and four dimensional space-time staggered grid discretizations were employed. The discretized PDE constrained optimization was solved by a sequential quadratic programming

algorithm. Two approaches were considered for the solution of the resulting KKT system. In the first approach, the system was solved inexactly using preconditioned GMRES, where an algebraic multi-grid preconditioner was facilitated. In the second approach the system was solved by elimination, resulting in a reduced system which was symmetric positive definite. This system was then solved by preconditioned conjugated gradients with a Gauss-Seidel preconditioner. Numerical results seemed plausible and consistent, even for problems with large density contrasts. From the convergence stand point, again, mesh-independent convergence was achieved. This study was presented at the Monge-Kantorovich Optimal Transport - Theory and Applications conference, 2009, the Copper Mountain conference on Iterative Methods, 2010, the SIAM Imaging Science conference, 2010 and in the SIAM Annual Meeting 2010.

Lastly, the 3rd study addressed large-scale PDE constrained optimization from a broader and more general perspective. The common computational bottleneck for both aforementioned studies, as well as for most PDE constrained optimization problems lies in the exhaustively accurate solution of the linear system, within the overall non-linear solution scheme. Thus, the overall idea for this study was to prescribe an appropriate stopping criterion for the iterative linear solver, such that an inexact search direction is computed, while still maintaining convergence of the overall algorithm. Our stopping criterion is based on the filter method, which simultaneously controls reduction in the objective as well as constraint violation. The proposed framework has been applied effectively for several test problems, including the abovementioned optimal mass transport problem. The algorithm requires only mild modification of currently employed Krylov linear solvers, and therefore can be readily applied to a wide range of applications. This research work was presented in the Copper Mountain conference on Iterative Methods, 2010, the SIAM Imaging Science conference, 2010 and in the SIAM

Annual Meeting 2010.

The fields of numerical and computational methods for simulation-based optimization and imaging receives growing amount of interest. There are many unanswered questions as well as important applications that need to be addressed. For an instance, incorporation of compressive sensing ideas for PDE constrained optimization. Recent advances in the field of compressive sensing have not yet been sufficiently disseminated for observation operators that are PDE based. The utilization of concepts such as sparsity and spatio-temporal compressive sampling in the PDE context should be further explored. Another example is large-scale inequality constrained optimization. While an extensive body of research has been concentrated in equality constrained optimization, its inequality counterpart is still under-researched, despite its utility for a broad range of practical applications.

Bibliography

[1] L. Ambrosio. Lecture Notes on Optimal Transport Problems, *2000 Lecture Notes in Mathematics "Mathematical aspects of evolving interfaces" CIME Series, Madeira (PT)* **1812**, P. Colli and J.F. Rodrigues Eds., 1–52, 2003.

[2] S. Angenent, S. Haker, and A. Tannenbaum. Minimizing Flows for the Monge-Kantorovich Problem *SIAM J. Math. Anal*, 35:61–97, 2003.

[3] A. C. Antoulas, D. C. Sorensen and S. Gugercin, A survey of model reduction methods for large-scale systems. *Contemporary Mathematics*, 280:193–219, 2001.

[4] U. M. Ascher, H. Huang, and K. Doel. Artificial time integration, 2006.

[5] U. M. Ascher. Numerical Methods for Evolutionary Differential Equations. *Society for Industrial & Applied Mathematics*, 2008.

[6] R. Bajcsy and S. Kovačič. Toward an individualized brain atlas elastic matching. Technical Report MS-CIS-86-71 Grasp Lap 76, Dept. of Computer and Information Science, Moore School, University of Philadelphia, 1986.

[7] R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Comput. Vision Graph. Image Process.* 46:1–21, 1989.

[8] J. Bear, "Dynamics of Fluids in Porous Media,"" *Dover Publications, New York,* (1972).

[9] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem *Numer. Math,* 375–393, 2000.

[10] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, **14**:1–137 (2005),

[11] L. T. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. v. B. Waanders, Real-Time PDE-Constrained Optimization *Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA, 2007.

[12] G. Biros and O. Ghattas Parallel Lagrange-Newton-Krylov-Schur methods for PDEconstrained optimization. Part I: the Krylov-Schur solver. *SIAM Journal on Scientific Computing*, 27:687–713, 2005.

[13] G. Biros and O. Ghattas Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization. Part II: the Lagrange-Newton solver, and its application to optimal control of steady viscous flows. *SIAM Journal on Scientific Computing* 27:714-739, 2005.

[14] A. Borzi and K. Kunisch. A globalization strategy for the multigrid solution of elliptic optimal control problems. *Optimization Methods and Software*, 21:445–459, 2006.

[15] J.P. Boyd. *Chebyshev & Fourier Spectral Methods.* Springer-Verlag, 1989.

[16] J. W. Brewer. Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems*, 25:772–780, 1978.

[17] F. Brezzi and M. Fortin, "Mixed and Hybrid Finite Element Methods," *Springer-Verlag*, (1991)

[18] L.G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24:325–376, 1992.

[19] R. H. Byrd, F. E. Curtis, and J. Nocedal. An Inexact SQP Method for Equality Constrained Optimization. *SIAM J. on Optimization* 19:351–369, 2008.

[20] R. H. Byrd, F. E. Curtis, and J. Nocedal. An inexact Newton method for nonconvex equality constrained optimization. *Math. Program.* 122:273–299, 2009.

[21] R. H. Byrd, M. E. Hribar, and J. Nocedal, An interior point algorithm for large scale nonlinear programming. *SIAM Journal on Optimization*, 9:877–900, 1999.

[22] R. Chartr, K. Vixie, B. Wohlberg, and E. Bollt. A gradient descent solution to the Monge-Kantorovich problem, 2005.

[23] C.M. Chin. A global convergence theory of a filter line search method for nonlinear programming. Technical report, Department of Electrical Engineering, University of Malaya, Kuala Lumpur, Malaysia, August 2002.

[24] C.M. Chin and R. Fletcher. On the global convergence of an SLP-filter algorithm that takes EQP steps. *Mathematical Programming*, 96(1):161177, 2003.

[25] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Numerical experiments with the LANCELOT package (Release A) for large-scale nonlinear optimization. *Math. Program.*, 73:73–110, 1996.

[26] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust-region Methods*, *MPS-SIAM Series on Optimization* (SIAM, Philadelphia, 2000.

[27] M. J. P. Cullen and R. J. Purser. An extended Lagrangian theory of semi-geostrophic frontogenesis. *J. Atmos. Sci*, 1477–1497, 1984.

[28] R. G. E. J. Dean. Numerical methods for fully nonlinear elliptic equations of the Monge-Ampère type. *Comput. Methods Appl. Mech. Eng.*, 195:1344–1386, 2006.

[29] G. L. Delzanno, L. Chacn, J. M. Finn, Y. Chung and G. Lapenta "An optimal robust equidistribution method for two-dimensional grid adaptation based on Monge-Kantorovich optimization," *Journal of Computational Physics archive*, **227(23)**, (2008), pp. 9841–9864

[30] R.S. Dembo and T. Steihaug. Truncated-Newton Algorithms for Large-Scale Unconstrained Optimization. *Math. Prog.*, 26:190–212, 1983.

[31] M. Droske and M. Rumpf. A variational approach to non-rigid morphological registration. *SIAM Appl. Math.*, 64(2):668–687, 2004.

[32] L. C. Evans. Partial differential equations and Monge-Kantorovich mass transfer survey paper. *Current Developments in Mathematics, 1997, International Press*, 1999.

[33] W. Gangbo and R. J. Mccann. The geometry of optimal transportation. *Acta Math*, 1996.

[34] C. Glasbey. A review of image warping methods. *Journal of Applied Statistics*, 25:155–171, 1998.

[35] A. A. Goshtasby. *2-D and 3-D Image Registration.* Wiley Press, New York, 2005.

[36] P. M. Gresho and R. L. Sani. Incompressible Flow and the Finite Element Method. John Wiley & Sons, 2000.

[37] M. D. Gunzburger. Perspectives in Flow Control and Optimization. *Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA, 2002.

[38] A. C. Guyton and J. E. Hall. Textbook of Medical Physiology *W.B. Saunder Company*, 1986.

[39] B. Fischer and J. Modersitzki. Fast diffusion registration. *AMS Contemporary Mathematics, Inverse Problems, Image Analysis, and Medical Imaging*, 313:117–129, 2002.

[40] B. Fischer and J. Modersitzki. Curvature based image registration. *J. of Mathematical Imaging and Vision*, 18(1):81–85, 2003.

[41] B. Fischer and J. Modersitzki. Combining landmark and intensity driven registrations. *PAMM*, 3:32–35, 2003.

[42] B. Fischer and J. Modersitzki. Ill-posed medicine – an introduction to image registration. *Inverse Problems*, 24:1–19, 2008.

[43] J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer Jr. Image registration, handbook of medical imaging. *Volume 2: Medical Image Processing and Analysis, SPIE*, pages 447–513, 2000.

[44] R. Fletcher Practical methods of optimization (2nd ed.) *New York: John Wiley and Sons*, 1987

[45] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91:239-270, 2002.

[46] R. Fletcher, S. Leyffer and Ph. L. Toint. A Brief History of Filter Methods. *SIAG Optimization Views-and-News*, 18(1):2–12, 2007.

[47] M. P. Friedlander and M. A. Saunders. A globally convergent linearly constrained Lagrangian methods for nonlinear optimization. *SIAM J. Optimization*, 15:863–897, 2005.

[48] K. R. Frisch. The logarithmic potential method of convex programming. Memorandum of May 13, University Institute of Economics, Oslo, Norway, 1955.

[49] G. H. Golub and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[50] E. Haber and U. M. Ascher Preconditioned all-at-once methods for large, sparse parameter estimation problems. *Inverse Problems*, 17:1847–1864, 2001.

[51] E. Haber and U. Ascher and D. Oldenburg, "On Optimization Techniques for Solving Nonlinear Inverse Problems," *Inverse Problems,***16** (2000) pp. 1263–1280 Volume = 16,

[52] E. Haber, S. Heldmann, and J. Modersitzki. A framework for image-based constrained registration with an application to local rigidity. *LAA*, pages 1–16, 2009. In Press.

[53] E. Haber, R. Horesh and J. Modersitzki. Numerical optimization for constrained image registration. *Numerical Linear Algebra with Applications, Special Issue: Copper Mountain Conference on Multigrid Methods 2009*, 17:343–359, 2010.

[54] E. Haber and J. Modersitzki, Numerical methods for volume preserving image registration. In *Inverse Problems, Institute of Physics Publishing*, pp. 1621–1638 (2004).

[55] E. Haber and J. Modersitzki. Image registration with guaranteed displacement regularity. *Int. J. on Comp. Vision*, 71(3):361 – 372, 2006.

[56] J. Hadamard. Lectures on Cauchy's Problem in Linear Differential Equations. *Yale University Press*, New Haven, 1923.

[57] L. A. Hageman and D. M. Young. Applied Iterative Methods, 1981.

[58] J. Hajnal, D. Hawkes, and D. Hill. *Medical Image Registration*. CRC Press, 2001.

[59] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34:561–580, 1992.

[60] L. Hanson E. Haber. Model problems in pde-constrained optimization. Technical report, Emory University, 2007.

[61] M. Heinkenschloss and L. N. Vicente, Analysis of inexact trust-region SQP algorithms. *SIAM Journal on Optimization*, 12:283–302, 2001.

[62] P. Hellier and C. Barillot. Coupling dense and landmark-based approaches for non rigid registration. *IEEE Transactions on Medical Imaging*, 22:217–227, 2003.

[63] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.*, 49:409–436, 1952.

[64] Edited by D. S. Holder Electrical Impedance Tomography – Methods, History and Applications Taylor and Francis, 2004.

[65] H. Jäger and E. W. Sachs Global convergence of inexact reduced SQP methods. *Optimization Methods and Software*, 7: 83-110, 1996.

[66] E. W. Jenkins, C. T. Kelley, C. T. Miller, and C. E. Kees, "An Aggregation-based Domain Decomposition Preconditioner for Groundwater Flow" *SIAM J. Sci. Comp.,* **(23)** (2001), pp. 430-441

[67] H. J. Johnson and G. E. Christensen. Consistent landmark and intensity-based image registration. *IEEE TMI*, 21(5):450–461, 2002.

[68] L. Kantorovich. On a problem of monge. *Uspekhi Mat. Nauk.*, 3:225–226, 1948.

[69] S. L. Keeling and W. Ring. Medical image registration and interpolation by optical flow with maximal rigidity. *J. of Mathematical Imaging and Vision*, 23(1):47–65, 2005.

[70] C.T. Kelley. *Iterartive Methods for Optimization.* SIAM, Philadelphia, 1999.

[71] M. Lalee, J. Nocedal, and T. D. Plantenga, On the implementation of an algorithm for large-scale equality constrained optimization. *SIAM Journal on Optimization*, 8: 682–706, 1998.

[72] C. L. Lawson and R. J. Hanson. Solving Least Squares Problems Classics in Applied Mathematics, new edition ed. Society for Industrial Mathematics, 1987.

[73] H.Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32:129–149, 1999.

[74] C.J Lin and J. More. Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, 9:1100–1127, 1999.

[75] J. A. Little, D.L.G Hill, and D.J. Hawkes. Deformations incorporating rigid structures. *Computer Vision and Image Understanding*, 66(2):223–232, 1997.

[76] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens. Nonrigid image registration using free-form deformations with a local rigidity constraint. In *MICCAI, volume 3216 of LNCS*, pages 639–646, 2004.

[77] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[78] J. Modersitzki. *Numerical Methods for Image Registration.* Oxford University Press, New York, 2004.

[79] J. Modersitzki. FLIRT with rigidity – image registration with a local non-rigidity penalty. *IJCV*, pages 153–163, 2008. DOI: 10.1007/s11263-007-0079-3.

[80] J. Modersitzki. *FAIR: Flexible Algorithms for Image Registration.* SIAM, Philadelphia, 2009. (To appear).

[81] V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417 (1966).

[82] J. D. Moulton, J. E. Dendy, and J. M. Hyman, "The black box multigrid numerical homogenization algorithm," *J. Comput. Phys.,* **141**:1–29 (1998).

[83] S. G. Nash A Survey of Truncated-Newton Methods. *Journal of Computational and Applied Mathematics.*, 124:45–59, 2000.

[84] R. A. Nicolaides, "Existence, Uniqueness and Approximation for Generalized Saddle Point Problems," *SIAM Journal on Numerical Analysis,* **18**:349–357 (1982).

[85] P.E. Nikravesh and I.S. Chung. Application of euler parameters to the dynamic analysis of three-dimensional constrained mechanical systems. *J. Mechanical Design*, 104:785–791, 1982.

[86] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, New York, 1999.

[87] A. Oberman. Wide stencil finite difference schemes for the elliptic Monge-Ampere equation and functions of the eigenvalues of the Hessian. *Discrete and Continuous Dynamical Systems series B (DCDS B)*, 10:221–238, 2008.

[88] V. Oliker and L. Prussner, On the numerical solution of the equation $z_{xx}^2 + z_{yy}^2 - z_{xy}^2 = f$ and its discretizations, *Numerische Mathematik* **54** (1988), pp. 271-293.

[89] E. E. Prudencio, R. Byrd, and X. C. Cai, Parallel full space SQP Lagrange-Newton-Krylov-Schwarz algorithms for PDE-constrained optimization problems. *SIAM Journal on Scientific Computing*, 27:1305–1328, 2006.

[90] J.D. Pryce, Basic methods of functional analysis. *Hutchinson University Library. Hutchinson.* p. 217, 1973.

[91] S. Rachev and L. Rüschendorf, *Mass Transportation Problems*, Vol. I, Probab. Appl., Springer-Verlag, New York, 1998.

[92] T. Rohlfing and C. R. Maurer, Jr. Intensity-Based Non-rigid Registration Using Adaptive Multilevel Free-Form Deformation with an Incompressibility Constraint *MICCAI 2001: Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, 111–119, 2001.

[93] L. Romero and F. Calderon. Chap. [10] A Tutorial on Parametric Image Registration *Scene Reconstruction Pose Estimation and Tracking*, R. Stolkin, ed., (I-Tech Education and Publishing, p. 530, 2007.

[94] J. Rubinstein and G. Wolansky, "Intensity control with a free-form lens," *J. Opt. Soc. Am. A*, **24** (2007), pp. 463–469

[95] J. W. Ruge and K. Stuben. Algebraic multigrid. In *Multigrid methods*, volume 3 of Frontiers Appl. Math. SIAM, Philadelphia, PA, 73–130, 1987.

[96] Y. Saad. Iterative Methods for Sparse Linear Systems. *Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA, 2003.

[97] Y.Saad and M. H. Schultz. Gmres: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7(3):856–869, July 1986.

[98] O. Scherzer. *Mathematical Models for Registration and Applications to Medical Imaging*. Springer, New York, 2006.

[99] Z. J. Shi and J. Shen. New Inexact Line Search Method for Unconstrained Optimization. *Journal of optimization theory and applications*, 127(2):425-446, November 2005.

[100] A. Tannenbaum. E. Haber, T. Rehman, to appear, *SIAM J. on Scientific Computing* (unpublished).

[101] U. Trottenberg, C. Oosterlee, and A. Schuller. *Multigrid*. Academic Press, 2001.

[102] H. A. van der Vorst, Bi-CGSTAB: A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Sys-

tems. *SIAM Journal on Scientific and Statistical Computing*, 13:631–644, 1992.

[103] S.P. Vanka. Block-implicit multigrid solution of navier-stokes equations in primitive variables. *J. Comp. Phys.*, 65:138–158, 1986.

[104] R. S. Varga. Matrix Iterative Analysis, 2000.

[105] C. Villani. Topics in optimal transportation. In *of Graduate Studies in Mathematics, American Mathematical Society*, 2003.

[106] S. Volkwein, Mesh-Independence of Lagrange-SQP Methods with Lipschitz-Continuous Lagrange Multiplier Updates, *Optimization Methods and Software*, **17**, (2002) , pp. 77-111.

[107] A. Walther, A first-order convergence analysis of trust-region methods with inexact Jacobians. Tech. Report MATH-WR-01-2005, Institute of Scientific Computing, Technische Universitat Dresden, Dresden, Germany, 2005.

[108] T. S. Yoo. *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis*. AK Peters Ltd, 2004.

[109] D. M. Young. Iterative Solution of Large Linear Systems, 1971.

[110] B. Zitova and J. Flusser. Image registration methods: a survey. Image and Vision Computing, 21:977–1000, 2003.

Index