SENSITIVITY AND STABILITY ANALYSIS FOR INVERSE OPTIMIZATION
WITH APPLICATIONS IN INTENSITY-MODULATED RADIATION
THERAPY

by

Ali Goli

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto

© Copyright 2015 by Ali Goli
Abstract

Sensitivity and Stability Analysis for Inverse Optimization with Applications in Intensity-Modulated Radiation Therapy

Ali Goli

Master of Applied Science

Graduate Department of Mechanical and Industrial Engineering

University of Toronto

2015

Inverse optimization aims to use observed data to understand the dynamics of a system by estimating the parameters of the underlying optimization problem. In this dissertation, we first develop an inverse model and investigate its properties for the cases where single and multiple observed data points are available. Next, we study the sensitivity and stability of the inverse model with respect to changes in the model parameters and perturbations in data. In particular, we illustrate the sensitivity of the inverse model with respect to the type of penalty functions used in the objective. Furthermore, we show that perturbations in the observed data can dramatically alter the outcome of the inverse model. We subsequently generalize our model to improve its stability with respect to perturbations in the input. Finally, we perform extensive simulations on synthetic and Intensity-Modulated Radiation Therapy (IMRT) data to validate the effectiveness of the generalized inverse model.
Dedication

I would like to dedicate this work to my beloved wife, Raheleh, for all her love and support.
Acknowledgements

Foremost, I would like to thank my advisor, Prof. Timothy C.Y. Chan, for his endless support, patience, and motivation. I would like to also thank my thesis committee members, Prof. Roy H. Kwon and Prof. Chi-Guhn lee, for their interest and helpful comments.

I thank my fellow labmates for the useful discussions and for all the fun we have had in the last two years. I also want to thank my friends who supported me throughout my M.A.Sc. studies.

I am very grateful to my wife for her patience, love, support and faith in me. Last but not the least, I also like to acknowledge and thank my parents for the sacrifices they made to ensure that I receive a good education.
## Contents

1 **Introduction** 1

1.1 Inverse Optimization .............................. 2

1.2 Motivating Application: IMRT Treatment Planning ............... 3

1.3 Summary of Contributions ............................ 4

1.4 Thesis Outline ...................................... 5

2 **Inverse Optimization Background** 7

2.1 Notation and Preliminaries ............................ 7

2.2 Inverse Continuous Optimization ........................ 8

2.3 Inverse Integer and Mixed Integer Problems ...................... 11

2.4 Connections to Regression and Stability ....................... 12

3 **Intensity-Modulated Radiation Therapy for Prostate Cancer** 14

3.1 Radiotherapy and Prostate Cancer .......................... 14

3.2 Clinical Criteria and a Model for Forward Planning ................ 16

3.3 Classical Parameter Tuning Techniques ........................ 19

3.4 Inverse Optimization and The Treatment Planning Process .............. 20

3.5 Structure of Data and Preprocessing .......................... 21

4 **The Choice of The Penalty Function in Inverse Optimization** 23

4.1 Generalized Single Point Inverse Linear Optimization .............. 24
4.2 Generalized Multi Point Inverse Optimization ........................................... 29
  4.2.1 A Generalized Multi Point Inverse Linear Model ................................. 29
  4.2.2 Behaviour of the Inverse Model with Respect to the Penalty Function ........ 30
4.3 Conclusions ................................................................................................ 42

5 On the Stability of Inverse Optimization ......................................................... 44
  5.1 A Motivational Example ............................................................................ 45
  5.2 Regularization In Regression ...................................................................... 47
    5.2.1 A Robust Regression Model .................................................................. 48
    5.2.2 Lasso Regularization .......................................................................... 48
    5.2.3 Tikhonov Regularization ...................................................................... 49
  5.3 Regularized Inverse Optimization ................................................................. 49
    5.3.1 A Robust Inverse Optimization Model ................................................ 49
    5.3.2 $\ell_1$-Regularized Inverse Optimization .............................................. 50
    5.3.3 $\ell_2$-Regularized Inverse Optimization .............................................. 52
  5.4 Singular Data Sets ...................................................................................... 52
  5.5 Measuring the Stability of Data Sets ............................................................ 57
  5.6 Price of Stability ........................................................................................ 60
  5.7 Computational Results .............................................................................. 61
    5.7.1 Generated Data .................................................................................. 61
    5.7.2 IMRT Treatment Planning .................................................................... 63
  5.8 Conclusions ............................................................................................... 75

6 Conclusions .................................................................................................... 77

A Full-Dimensionality of The Feasible Polyhedron ............................................. 80

Bibliography ...................................................................................................... 85

vi
List of Tables

3.1 A subset of the considered objectives for IMRT treatment planning for prostate cancer at the Princess Margaret Cancer Centre in Toronto, Canada. 18

4.1 Imputed cost vectors for the given solutions (4.9) and constraints (4.8) for different values of \( p \) in formulation (4.7). 33

5.1 Produced weights for patient #1 using different regularization parameters. 68
5.2 Produced weights for patient #2 using different regularization parameters. 68
5.3 Produced weights for patient #3 using different regularization parameters. 68
5.4 Clinical dose criteria for the clinical plan and inversely optimized plans for patient #1. Cell gray levels correspond to violation in the criteria, i.e., darker colour corresponds to more violation. 72
5.5 Log-likelihood of the bladder weights produced by each inverse model using \( OV_{Blad/Rect} \) feature at different expansion levels. 75
List of Figures

3.1 Locations of the OARs and the tumour in prostate cancer . . . . . . . . 15
3.2 Dose distribution of a clinical plan for prostate cancer therapy . . . . . . 16
3.3 DVH curves corresponding to the clinical plan presented in Figure 3.2 . . 17

4.1 The feasible region defined by the inequalities in (4.8) along with the observed solutions given in (4.9). The red arrows indicate the range of cost vectors produced by $\text{GMIO}_p$ for different values of $p$. . . . . . . . . . . 33

5.1 Solution instability of minimizing the sum of the duality gaps ($\text{GMIO}_1$) for problem (5.1). (a) A small perturbation in a single point inverse problem altered the imputed cost vector from $(0.6173, 0.3827)$ to $(-0.1775, 0.8225)$ (shown by red arrows). (b) A small perturbation in a datum in a multi point inverse problem altered the optimal cost vector from $(0.6173, 0.3827)$ to $(-0.1775, 0.8225)$ (shown by red arrows). . . . . . . . . . . . . . . 46

5.2 The instability of minimizing the sum of squares of the duality gaps ($\text{GMIO}_2$). The imputed cost vector changed from $(0.6173, 0.3827)$ to $(-0.1775, 0.8225)$ by a small perturbation in only one data point. . . . . . . . . . . . . . . 47
5.3 The 0.5-stability heat map for (a) non-regularized, (b) \( \ell_1 \)-regularized, and (c) \( \ell_2 \)-regularized inverse problems corresponding to example (5.1). The red areas indicate more instability, i.e., a small perturbation of the original data may change the normalized inversely optimized cost vector by at least 0.5. Note that the dark red areas in (a) are exactly aligned with the locus of centroid of singular data sets for \( \text{GMIO}_1 \), which are a subset of all the lines that have an equal distance from at least two different constraints.

5.4 (a) The stability measure (\( \bar{\rho} \)) for \( \ell_1 \)-regularized inverse problem with parameter \( \lambda \). (b) The price of stability (\( \bar{\Pi} \)) for \( \ell_1 \)-regularized inverse problem with parameter \( \lambda \).

5.5 (a) The stability measure (\( \bar{\rho} \)) for \( \ell_2 \)-regularized inverse problem with parameter \( \lambda \). (b) The price of stability (\( \bar{\Pi} \)) for \( \ell_2 \)-regularized inverse problem with parameter \( \lambda \).

5.6 An illustration of the OVH with schematic view of PTV and bladder, the portion of the green zone separated by the dotted lines illustrates \( OVH_{\text{Blad}}^{0.5} \) and \( OVH_{\text{Blad}}^{1.00} \).

5.7 IMRT treatment planning process using inverse optimization.

5.8 Comparison of DVHs of the non-regularized inverse plan and the clinical plan for patient #1.

5.9 Comparison of DVHs of the \( \ell_1 \)-regularized inverse plan with \( \lambda = 1 \) and the clinical plan for patient #1.

5.10 Comparison of DVHs of the \( \ell_2 \)-regularized inverse plan with \( \lambda = 1 \) and the clinical plan for patient #1.

5.11 Comparison of median error between the clinical plan and non-regularized, Lasso (\( \lambda = 1 \)), and Tikhonov (\( \lambda = 1 \)) inversely optimized plans for the clinical acceptability criteria.
5.12 Comparison of the mismatch between the fitted logistic function and bladder weight produced by non-regularized, $\ell_1$-regularized, $\ell_2$-regularized inverse models for $\lambda = 1$. 

\hspace{1cm} 74
Chapter 1

Introduction

Data-driven inverse optimization uses observed data to estimate parameters of an underlying optimization model. In this dissertation, we pose the question of whether inverse optimization models provide reliable results, and if they do not, how we can develop more reliable inverse models. In particular, this dissertation investigates the following topics:

• Sensitivity of inversely optimized estimates with respect to model parameters,

• Sensitivity of inverse estimates with respect to perturbations in input data,

• Development and performance validation of new models that are less sensitive to perturbations in data, i.e., models that are more stable.

To the best of our knowledge, this dissertation is the first work that studies the sensitivity and stability of inverse models with respect to model parameters and observed data. The contributions are not limited to the theory; we also study the implications of instability by performing simulations on synthetic and Intensity-Modulated Radiation Therapy (IMRT) data.
1.1 Inverse Optimization

The goal of inverse optimization is to impute the parameters of an optimization model by using observed data. Inverse models have been widely used to estimate unknown model parameters in practical problems. Geophysical scientists were the first to introduce inverse problems for estimating seismic wave model parameters, using observed earthquake data [1]. Since then, inverse models have found a wide variety of applications in different fields including healthcare [2, 3], auctions [4], and production planning [5].

In the classical inverse optimization framework, it is tacitly assumed that the given data point is exactly optimal, and the inverse problem aims at estimating the parameters of an optimization model (e.g., constraints or cost function) to make the given data point optimal. Therefore, in classical inverse linear [6] and convex [7] models, optimality constraints like strong duality and KKT conditions are enforced. However, in practical applications, data is subject to error and perturbations and the given data points might not be optimal or even feasible. Modern inverse models [3, 8–11] account for the errors and uncertainty in data and intend to optimize an optimality criterion such as the duality gap or KKT residuals.

In both modern and classical inverse models there exist fixed model parameters whose values are chosen arbitrarily, without mathematical justification. For instance, in some inverse models an inversely optimized cost vector is found by minimally adjusting an a priori given cost vector with respect to some distance function. However, this distance function or norm is chosen arbitrarily, ranging from $\ell_0$ (Hamming distance), weighted $\ell_1$, $\ell_2$, to $\ell_\infty$-distance functions. In this dissertation, we elucidate the profound effect of these fixed model parameters on the behaviour of inverse models.

We also study the stability of inverse models, which we define as the sensitivity of the model with respect to perturbations in the observed data. In particular, we show that even in the case where the model parameters are given, perturbations in observed data may considerably alter the inversely optimized estimates, which we refer to as instability.
We use the connections between inverse optimization and regression and show that previously developed inverse models that mirror LS and LAD regression suffer from similar instability issues. We mitigate the instability issues by developing novel inverse models that were inspired by the classical remedies used for solving similar instability issues in regression. In particular, we introduce two regularized inverse optimization models mirroring Tikhonov [12] and Lasso [13] regularized regression that improve the stability of the inverse solution. We also propose different metrics for quantifying the extent to which our models have been successful in stabilizing the state-of-the-art inverse models. Finally, we perform extensive simulations both on synthetic and IMRT data sets to validate the effectiveness of the proposed methods.

1.2 Motivating Application: IMRT Treatment Planning

The goal in IMRT treatment planning is to optimize the intensities of a set of radiation beams to deliver high doses of x-ray radiation to the tumour, while sparing other organs in the proximity of the tumour. The IMRT treatment planning problem is usually modelled as a multi-objective optimization problem with a weighted objective function, which we refer to as the “forward problem”. The objective function weights in the forward problem are determined through an iterative and time consuming process by a treatment planner and an oncologist. In particular, at each iteration the planner tunes the weights based on feedback from an oncologist and re-solves the forward problem to produce a treatment plan. These iterations will continue until the quality of the reached treatment plan is validated by the oncologist.

In [3], Chan et al. introduced an inverse optimization model that takes a previously validated clinical plan as input and determines a set of inversely optimized weights for the forward problem that can produce a treatment plan similar to the clinical plan. In
later publications Lee et al. [14] and Boutilier et al. [15] showed that the set of inversely optimized weights are closely connected with anatomical features of the patients. Particularly, they proposed a logistic regression model to estimate the inversely optimized weights using the Overlap Volume Histogram (OVH) [16]. Therefore, one can think of the IMRT inverse learning process as a two-phase algorithm. The first phase maps a clinical plan to a set of inversely optimized weights. The second phase links these weights to anatomical features using a regression model. Consequently, for a new patient, anatomical features are employed to predict optimal inversely optimized weights; these weights are then fed into the forward problem to generate a treatment plan for the new patient.

Our initial tests showed that for each patient there exists a range of objective weights that yield acceptable clinical plans. However, inverse models return only a single set of weights for each patient. The regularized models developed in this dissertation will return weights that can be estimated more efficiently and also produce treatment plans that satisfy higher clinical standards. Particularly, our simulations show that by using Lasso regularized inverse models, the inversely optimized plans will have a higher quality, and inversely optimized weights will fit better to the logistic regression model used for predicting them.

1.3 Summary of Contributions

Our contributions are as follows:

1. We provide an in-depth analysis of the behaviour of a class of inverse optimization models and show that perturbations in the model parameters or the data can affect the model considerably. In particular, the model may exhibit instability in certain scenarios.

2. We use the connections of inverse optimization and regression to fix the instability issue by using Lasso and Tikhonov regularization for inverse optimization. We also
show that using regularized inverse models improves the quality of the delivered inversely optimized IMRT treatment plans.

3. We develop metrics for quantifying the trade-off between the stability and efficiency of inverse models. We introduce “predictability” as a new desired feature for inversely optimized weights in IMRT treatment planning and evaluate the effectiveness of our regularized inverse IMRT models by comparing their fitness or predictability with respect to the prediction model.

1.4 Thesis Outline

This dissertation is organized as follows. In Chapter 2, we explain the differences between modern and classical inverse optimization models. We also categorize inverse problems into two groups and examine the main assumptions of each group. Furthermore, we illustrate the connections between inverse optimization and regression, with a focus on stability.

Chapter 3 is focused on radiotherapy for prostate cancer. We present an optimization model used for IMRT treatment planning. We also describe the structure of the data and the preprocessing performed on the data for our simulations.

In Chapter 4, we first develop an inverse optimization model by using a novel normalization technique. Next, we show that our model yields non-trivial solutions and has some other desirable properties. Finally, we illustrate the sensitivity of the model behaviour to the choice of penalty functions in the inverse optimization model.

In Chapter 5, we illustrate that inverse models may exhibit instability in certain scenarios. Next, we apply regularization techniques to mitigate the instability issue and study the connections between our regularized models and a robust inverse optimization model. We also propose metrics for measuring the effect of regularized models on the outcome of the inverse model. Lastly, we perform simulations to elucidate the effect of
regularization methods on the inversely optimized estimates.

Chapter 6 concludes and presents some of the future research directions.
Chapter 2

Inverse Optimization Background

Inverse optimization has gained a lot of attention in the operations research and management science community. Various theories and frameworks for different types of inverse optimization including linear [3, 6], conic [7], and integer programs [17] have been developed and their applications in a wide variety of areas including auctions [4], aggregate production planning [5], portfolio optimization [18], radiation therapy [3], and freight assignment [19] are explored.

This chapter provides a review of different approaches in the inverse optimization literature. Section 2.2 investigates the key assumptions behind inverse continuous models and illustrates the differences between traditional and modern inverse optimization. Section 2.3 is devoted to inverse integer and mixed integer optimization problems that cannot be transformed to a continuous problem with integral solutions (e.g., knapsack problem). Finally, Section 2.4 elucidates the connections of inverse optimization and regression and presents a review of the literature related to the stability of regression models.
2.1 Notation and Preliminaries

Let $\text{FOP}(c) = \min_{x \in X} f(x, c)$ be an optimization problem with a parametric cost function, which we refer to as the forward problem. Given an observed data point $\hat{x}$, a classical inverse optimization problem, which we refer to as $\text{IOP}(\hat{x})$, returns a value $\hat{c} \in C^{opt}$, where $C^{opt} = \{c \mid \hat{x} \in \arg\min_{x \in X} f(x, c)\}$. Therefore, the output of an inverse model is a parameter $\hat{c}$, which yields a cost function $f(x, \hat{c})$ that makes $\hat{x}$ optimal. Based on the structure of the feasible region of the forward problem ($X$), we classify the inverse models into two major categories: inverse continuous and inverse discrete optimization.

2.2 Inverse Continuous Optimization

In inverse optimization, it is assumed that an optimal solution $\hat{x}$ to the forward optimization problem with unknown parameter is observed, and the goal is to estimate the unknown parameter $c$ that makes the observed solution “minimally suboptimal”. That is, if the set of parameter values for which the observed solution is optimal is nonempty, then the inverse model aims to find at least one element of this set; if the observed solution is not optimal for any parameter choice, then the model should find parameter values so that some notion of suboptimality (e.g., the duality gap) is minimized.

Geophysical scientists were the first to introduce inverse problems for estimating seismic wave model parameters, using observed earthquake data [1]. The traditional approaches to inverse optimization mostly revolved around inverse combinatorial optimization and inverse problems in networks and graphs including inverse shortest path and minimum spanning tree [1, 20–22]. For a survey on inverse combinatorial optimization see [18]. Due to the intrinsic characteristics of these problem, classical inverse models tacitly assume that the observed solution is optimal for some parameter values [1, 6, 7, 18, 20–22], or equivalently, it is assumed that the set $C^{opt}$, defined in Section 2.1, is nonempty.
Consequently, traditional inverse models are usually formulated as following:

\[
\begin{align*}
\text{minimize} & \quad d(c, \hat{c}) \\
\text{subject to} & \quad c \in C^{opt},
\end{align*}
\]

where \( d : \mathbb{R}^{K} \times \mathbb{R}^{K} \to \mathbb{R} \) is a distance function. Problem (2.1) determines an optimal value for the unknown parameter \( c \) that makes \( \hat{x} \) optimal and needs minimal adjustment with respect to a given candidate \( \hat{c} \). Formulation (2.1), has been explored for different types of distance functions including \( \ell_0 \) or Hamming distance [23], weighted \( \ell_1 \) distance [6], \( \ell_2 \) distance [7], and \( \ell_{\infty} \) distance [22]. Also the feasible region \( C^{opt} \) is either characterized by enforcing strong duality and dual feasibility [6, 21, 24] or forcing the KKT residuals to be zero in the case of nonlinear forward problems [25–27].

It turns out that in most of the nonlinear inverse problems with a single observation, \( C^{opt} \) is nonempty and the literature centres around developing efficient solution methods and formulations rather than guaranteeing existence of a meaningful optimal solution [7, 25, 26]. However, for cases where the cost function \( f(x, c) \) is constrained to be linear and the feasible region is full-dimensional, the inverse problem (2.1) becomes infeasible (\( C^{opt} \) is empty) for any given \( \hat{x} \) in the interior of the feasible polyhedron [3]. Also for cases where multiple observations are available it might not feasible to make all the observations simultaneously optimal for the same \( c \). In order to accommodate these scenarios, modern inverse optimization techniques were introduced.

The modern, particularly data-driven, inverse methods explicitly consider the impact of noise or other uncertainties that contaminate the observed solution, requiring some notion of error to be minimized such as KKT residuals [8], and duality gap [3, 5, 9, 10]. Let \( A \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^{m}, \) and \( \hat{x} \in \mathbb{R}^{n} \). Below we present a general convex forward problem...
similar to the one studied in [8]:

\[ \text{FOP}(c) : \quad \begin{array}{c}
\text{minimize} \quad f(x, c) \\
\text{subject to} \quad g_i(x) \leq 0, \quad i \in \{1, \ldots, k\}, \\
Ax = b,
\end{array} \quad (2.2) \]

where \( f(x, c) \) and \( g_i(x) \) are differentiable convex functions in \( x \). The inverse optimization problem for a given data point \( \hat{x} \) is formulated as follows:

\[ \text{IOP}(\hat{x}) : \quad \begin{array}{c}
\text{minimize} \quad \Phi(\xi^{(1)}, \xi^{(2)}) \\
\text{subject to} \quad \nabla_x f(\hat{x}, c) + \sum_{i=1}^{k} \mu_i \nabla g_i(\hat{x}) + A^t \lambda = \xi^{(1)}, \\
\mu_i g_i(\hat{x}) = \xi^{(2)}_i, \quad i \in \{1, \ldots, k\}, \\
\mu_i \geq 0, \quad i \in \{1, \ldots, k\},
\end{array} \quad (2.3) \]

where \( \xi^{(1)} \) and \( \xi^{(2)} \) are the stationary and complementary slackness residual vectors, respectively. \( \Phi : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^+ \cup \{0\} \) is a general convex penalty function that is zero if and only if \( \xi^{(1)} = 0 \) and \( \xi^{(2)} = 0 \). In [8], Keshavarz et al. assume that \( f(x, c) = \sum_{i=1}^{p} c_i f_i(x) \) and the inverse model then determines a linear combination of basis functions \( f_i \) that make \( \hat{x} \) minimally suboptimal. The authors also added the \( c_1 = 1 \) constraint to formulation (2.3), which acts as a normalization constraint to avoid trivial solutions.

In [3, 5, 10], the authors focus on a class of linear forward problems and they considered the duality gap to be a measure of optimality. Therefore, given a forward problem
**Chapter 2. Inverse Optimization Background**

**FOP** \((c) = \min_{Ax \geq b} c'x\), the inverse problem is formulated as follows:

\[
\text{IOP}(\hat{x}) : \quad \text{minimize } \epsilon_{c,y} \\
\text{subject to } A'y = c, \\
\epsilon = \frac{c'\hat{x}}{b'y}, \\
y \geq 0.
\] (2.4)

However, note that the relative duality gap constraint \(\epsilon = \frac{c'\hat{x}}{b'y}\) leads to non-convexity of formulation (2.4). In [3, 11], the authors used \(b'y = 1\) as a normalization constraint, which transformed formulation (2.4) to the following equivalent linear optimization problem:

\[
\text{IOP}(\hat{x}) : \quad \text{minimize } c'\hat{x} \\
\text{subject to } A'y = c, \\
b'y = 1, \\
y \geq 0.
\] (2.5)

Chan et al.[3] explored problem (2.5) for a multi-objective forward problem **FOP** \((c) = \min_{Ax \geq b} \sum_{i=1}^{p} \alpha_i c'_i x\), where \(p\) is the number of objectives. Consequently, in the inverse problem, \(c\) is constrained to be a linear combination of \(c_i\)'s. They also illustrated the connections between the dual of the inverse linear multi-objective model with Benson’s method [28] and Pareto surface approximation techniques [29, 30].

In [5, 10], Troutt et al. proposed a nonlinear inverse problem that jointly estimates the cost function and the constraints using a set of observed solutions \(\hat{x}^{(1)}, \ldots, \hat{x}^{(q)}\). The proposed method consists of a master and a sub-problem. The master problem determines the set of constraints \(Ax \geq b\) and the sub-problem returns an optimal cost vector for the set of constraints generated by the master problem. The optimality measure for determining the cost vector is assumed to be \(\sum_{i=1}^{q} \frac{c'_i x_i}{c^* x'}\) or \(\max_{i \in \{1, \ldots, q\}} \frac{c'_i x_i}{c^* x'}\), where \(x^*\) is an
optimal solution to the forward problem \( \min_{Ax \geq b} c^t x \). Although this method seems to be more general than frameworks introduced in [3, 8], its high computational cost makes it less attractive for practical large-scale problems.

### 2.3 Inverse Integer and Mixed Integer Problems

All of the models studied in Section 2.2 hinged on the existence of a set of necessary and sufficient conditions for optimality in convex continuous problems. However, the optimality criteria in linear and convex optimization (e.g., KKT conditions and strong duality) may not hold in the general integer and mixed integer framework. Therefore, characterizing the set \( C_{opt} \) is not as straightforward as it is in the continuous case.

In [17], Schaefer uses superadditive integer duality [31] to find a polyhedral representation for \( C_{opt} \) of an inverse integer program. The author also proves that the inverse optimization problem corresponding to the integer program is \( \mathcal{NP} \)-hard. In [32], Ronald et al. studied an inverse binary knapsack problem under \( \ell_1 \) and \( \ell_\infty \) norms. They also analyzed the computational complexity of the problem in both cases and proposed a bilevel integer linear program for solving the inverse model.

There has been very little research on inverse optimization for mixed integer linear programs (MILP). In [33], Wang proposed a cutting plane algorithm for solving inverse MILPs under \( \ell_1 \) and \( \ell_\infty \) norm. An improved version of the algorithm was published in a later paper [34]. In a recent paper [35], Bulut and Ralphs showed that the inverse mixed integer linear programs are \( \text{coNP} \)-complete. They also proposed a cutting plane method for solving the inverse MILP problem.

### 2.4 Connections to Regression and Stability

The notion of minimizing error or residuals in some optimality criterion as a means to determine parameters values “consistent” with observed data is not unique to inverse
optimization. In fact, this is exactly the goal of regression. Given this conceptual relationship between inverse optimization and regression, it is natural to consider whether certain characteristics of regression, which has been well-studied, are present in inverse optimization. For example, goodness-of-fit is an important measure that quantifies the quality of a regression model’s fit to the data – a recent paper developed a goodness-of-fit measure for inverse linear optimization that has the well-known characteristics of the coefficient of determination in linear regression [11].

Another important characteristic of regression is stability. It is well-known that regression can suffer from instability issues. For example, a small perturbation in one datum can dramatically alter the fitted line in least squares (LS), least absolute deviation (LAD), and Least Median of Squares (LMS) regression models [36–38]. The concept of a singularity set has been used as a quantitative measure of stability [39]. However, regularization techniques have been shown to improve the stability of regression models [40]. Stability has been shown to be an important concept in machine learning [40–42] and optimization [43, 44].
Chapter 3

Intensity-Modulated Radiation Therapy for Prostate Cancer

In this chapter we provide an overview of clinical criteria, proposed solutions, and data structure in Intensity-Modulated Radiation Therapy (IMRT) for prostate cancer. In Section 3.1, we discuss the clinical criteria considered in IMRT treatment planning and present a mathematical optimization model for IMRT treatment planning problem. In Section 3.3, we present an overview of the proposed treatment planning techniques for IMRT. We describe the process of inverse-optimization-based treatment planning in Section 3.4. Finally, we describe the structure of our data and the preprocessing phases in Section 3.5.

3.1 Radiotherapy and Prostate Cancer

Prostate cancer is the most common type of cancer among men [45]. In 2014 alone, 233,000 (27% of total cancer diagnoses among men) cases of prostate cancer were diagnosed in North America [45]. Radiation therapy is one of the most common treatment methods in cancer therapy [46]. IMRT is a form of radiotherapy that delivers radiation beams with different intensities from different angles. In IMRT, the goal is to expose
the cancerous cells to high doses of radiation. Although IMRT has been shown to be an effective approach for destroying the tumour in prostate cancer, the exposure of the organs-at-risk (OARs) in the proximity of the tumour, like bladder and rectum, can lead to post-treatment complications [47]. Therefore, it is crucial to control the amount of radiation that is delivered to the OARs. Figure 3.1 depicts the locations of the OARs in prostate cancer via a Computerized Tomography (CT) scan. Accordingly, the Clinical Target Volume (CTV) refers to the part that contains the cancerous cells. The Planning Target Volume (PTV) is a larger set that encompasses CTV and accounts for tumour movements or other uncertainties. The PTV Inner Ring (PIR) is the set difference of PTV and CTV (i.e., $PIR = PTV \setminus CTV$). The PTV Outer Ring (POR) is a ring around PTV that is used to improve the conformity of the delivered dosage to the tumour. The main OARs in prostate cancer are rectum, bladder, and the femoral heads.

A clinical dose distribution is depicted in Figure 3.2. The radiation exposure of each organ or target is monitored by the Dose Volume Histogram (DVH). Figure 3.3 depicts the...
Figure 3.2: Dose distribution of a clinical plan for prostate cancer therapy

DVH curves for different OARs/targets for the dose distribution presented in Figure 3.2. In Figure 3.3, the $x$ axis is the dose and is measured in Gray ($Gy$). For a value of $xGy$, the DVH curve of an organ/target indicates the portion of that organ/target ($0 \leq y \leq 1$) that is being exposed to at least $xGy$ of radiation. Consequently, for a high quality clinical treatment plan the DVH curves of OARs damp very fast, while the the DVH curves for targets must have a high value till a certain threshold and plunge afterwards.

### 3.2 Clinical Criteria and a Model for Forward Planning

A clinical plan is deemed acceptable if a set of clinical constraints are achieved or the violation of these criteria is minimized. Four types of clinical constraints for each organ/target are considered in this dissertation: min dose, max dose-volume, min dose-volume, and max dose. The value of a min dose constrains the minimum amount of radiation in each
Figure 3.3: DVH curves corresponding to the clinical plan presented in Figure 3.2. Point inside a target. The max dose constraint is also defined similarly. For instance, a min (max) dose objective of $\theta Gy$ for CTV requires all the points in CTV to receive at least (most) $\theta Gy$ of radiation. The max dose-volume requirement is satisfied if no more than a certain portion of an OAR ($p$), receives an specified amount of radiation $\theta$. Consequently, the max dose-volume requirements impose constraints on the Value-at-Risk (VaR). For instance, a max $59.15 Gy - 29\%$ constraint for the bladder indicates that at most 29% of the bladder’s volume can receive 59.15 Gy or more radiation. The min dose-volume is also defined similar to the max dose-volume criterion. However, min dose-volume objectives are usually used for targets as they constrain the minimum amount of radiation delivered to the tissue. A subset of objectives considered for prostate cancer therapy in the Princess Margaret Cancer Centre in Toronto is presented in Table 3.1.
<table>
<thead>
<tr>
<th>#</th>
<th>Organ/Target</th>
<th>Type</th>
<th>Dose (Gy)</th>
<th>Volume (Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CTV</td>
<td>Min dose</td>
<td>78.65</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CTV</td>
<td>Max dose</td>
<td>81.90</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>PIR</td>
<td>Min dose</td>
<td>74.75</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>PIR</td>
<td>Max dose-volume</td>
<td>78.00</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>POR</td>
<td>Max dose</td>
<td>74.10</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Bladder</td>
<td>Max dose-volume</td>
<td>59.15</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>Rectum</td>
<td>Max dose-volume</td>
<td>59.15</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>Left Femur</td>
<td>Max dose-volume</td>
<td>52.00</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>Right Femur</td>
<td>Max dose-volume</td>
<td>52.00</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3.1: A subset of the considered objectives for IMRT treatment planning for prostate cancer at the Princess Margaret Cancer Centre in Toronto, Canada.

In practice, it might not be feasible to meet all the clinical criteria. Therefore, in order to produce an acceptable clinical plan a multi-objective optimization problem is solved, in which some of the clinical limits are imposed as hard constraints and others are penalized with different weights in the objective function. Also minimizing the max dose-volume objectives (VaR type objectives) in the forward problem requires binary variables [48]. Consequently, in order to avoid integer variables associated with the VaR-based objectives, we use a corresponding Conditional Value-at-Risk (CVaR) objective that penalizes every unit of radiation above the specified threshold, which we refer to as the relaxed max dose-volume penalty function. The relaxed max dose-volume penalty function acts as a proxy for minimizing the actual max dose-volume penalty function. In this thesis we rely on the following relaxed multi-objective forward formulation introduced by Chan et al. [3]:
minimize \( \sum_{i \in I} \alpha_i \sum_{v \in O_i} \frac{1}{|O_i|} \max \left\{ \sum_{b \in B} D_{v,\omega_b} \omega_b - \theta_i \right\} + \sum_{j \in J} \alpha_j \max_{v \in O_j} \left\{ \sum_{b \in B} D_{v,\omega_b} \omega_b \right\} \)

subject to \( \sum_{b \in B} D_{v,b,\omega_b} \leq u_v, \quad \forall v \in \mathcal{U}, \)

\( \sum_{b \in B} D_{v,b,\omega_b} \geq l_v, \quad \forall v \in \mathcal{L}, \)

\( \frac{\beta}{|B|} \sum_{b \in B} \omega_b \leq \omega_i \leq \frac{\beta}{|B|} \sum_{b \in B} \omega_b, \quad \forall i \in \mathcal{B}, \)

\( \omega_b \geq 0, \quad \forall b \in \mathcal{B}. \)

(3.1)

In formulation (3.1), \( \omega_b \) is the intensity of the \( b^{th} \) beamlet. \( D \) is the dose-influence matrix and the \( D_{v,b} \) indicates the increase in the amount of radiation (Gy) in the \( v^{th} \) voxel for a unit increase in the intensity of the \( b^{th} \) beamlet. The relaxed max dose-volume objectives and the max dose objectives are indexed by \( I \) and \( J \), respectively. In the first and second sets of constraints a portion of the min and max objectives are enforced as hard constraints. Finally, the third constraint limits the variations of the beamlets to promote the conformity of the induced dose distribution.

### 3.3 Classical Parameter Tuning Techniques

In formulation (3.1), the set of objectives \( I \) and \( J \) and their corresponding weights \( \alpha_i \) and \( \alpha_j \) are tuned for each patient. In practice, the parameter tuning procedure is an iterative and time-consuming process based on trial-and-error. In order to reach an acceptable clinical plan, a treatment planner initializes the tunable parameters in the forward problem and the produced plan is then verified by an oncologist. In case that the plan fails to meet the clinical standards, the oncologist sends feedback to the treatment planner to retune the parameters. In some cases, these iterations may take up to several days to complete and can potentially prolong the treatment process.
There are two major approaches for expediting the treatment planning process. In the first approach the treatment planner provides a set of importance factors after each iteration, which are used to create scoring functions for an optimization problem. These scoring functions are used to formulate and solve an optimization problem, which guides the planner to find an optimal set of weights [49–51]. The second approach determines a set of Pareto optimal treatment plans rather than a single plan. The final prescribed treatment is selected from the Pareto optimal plans based on clinical preferences and expertise [52–54].

3.4 Inverse Optimization and The Treatment Planning Process

In [3], Chan et al. introduced an inverse optimization model that imputes a set of weights for the forward problem (3.1) that can reproduce a plan similar to the clinical plan. Furthermore, they also demonstrated that by limiting the forward problem to have a sparse number of objectives the corresponding inversely optimized plan remains close to the clinical treatment plan. The proposed inverse optimization framework does not provide a solution for IMRT treatment planning per se, but it maps a clinical plan to a small size vector of weights (e.g., with 4-5 elements for prostate cancer). In a later publication [14] Lee et al. showed that the set of inversely optimized weights can be predicted efficiently using anatomical features. The performance of some classical learning algorithms for predicting the inverse weights was also studied by Boutilier et al. [15].
3.5 Structure of Data and Preprocessing

In this thesis we used an extended version of the data set used in [55]. The data set contains the IMRT clinical treatment data for 336 patients who received a treatment at Princess Margaret Cancer Centre in Toronto. The treatment data of 21 patients who had undergone special surgeries or had anatomical anomalies were excluded from our data set. The treatments were delivered using seven 6 MV step-and-shoot intensity modulated x-ray fields with fixed angles at 40°, 80°, 110°, 250°, 280°, 310°, and 335°. The treatment data was saved as DICOM formatted files using Philips Pinnacle. The DICOM files were processed and analyzed using CERR (Computational Environment for Radiotherapy Research) [56].

The patient’s geometry was discretized to voxels based on a contoured CT scan with a resolution of $1\text{mm} \times 1\text{mm} \times 2\text{mm}$. On average the CTV, PTV, bladder, rectum, left femur, and right femur contained 28,776, 62,973, 10,872, 97,604, 46,161, and 46,111 voxels, respectively. To decrease the computational cost of the experiments, the voxels for the PTV (largest structure) were down-sampled by a factor of 10 and the voxels for other organs/targets were downsampled by a factor of 4. The influence matrix ($D$) was also determined using CERR.

The treatment data does not contain the intensity of each beamlet. In fact, the treatment data for each patient only includes the dose of radiation at each voxel. We assume that the treatment plans were produced by solving a problem similar to formulation (3.1). However, for some patients it turns out that there does not exist a feasible solution to formulation (3.1) that can reproduce the exact same clinical plan. Therefore, for any given clinical plan we determine the closest plan that lies within the feasible region of
formulation (3.1) by solving the following minimum mean-square error problem:

\[
\begin{align*}
\text{minimize} & \quad \sum_{v \in V} \left( \sum_{b \in B} D_{v,b} \omega_b - x_v \right)^2 \\
\text{subject to} & \quad \sum_{b \in B} D_{v,b} \omega_b \leq u_v, \quad \forall v \in U, \\
& \quad \sum_{b \in B} D_{v,b} \omega_b \geq l_v, \quad \forall v \in L, \\
& \quad \frac{\beta}{|B|} \sum_{b \in B} \omega_b \leq \omega_i \leq \frac{\beta}{|B|} \sum_{b \in B} \omega_b, \quad \forall i \in B, \\
& \quad \omega_b \geq 0, \quad \forall b \in B,
\end{align*}
\]

(3.2)

where \(x_v\) is the dose at the \(v^{th}\) voxel in the clinical treatment plan and \(D_{v,b}\) is the influence of beamlet \(b\) on voxel \(v\), which was computed using CERR.
Chapter 4

The Choice of The Penalty Function in Inverse Optimization

In this chapter we first present a classical single objective inverse linear model with a non-convex normalization constraint. Then we relax the inverse formulation by proposing a linear normalization method and we show that for a single data point the new formulation for a single data point produces the same optimal solutions as the original non-convex problem. Next we extend our model to the multi data point case with the objective of minimizing an arbitrary penalty function. The value of the proposed linear normalization method is that it remains tractable in the multi data point or multi objective scenarios where closed-form solutions are not available. Furthermore, in the single point scenario the closed-form solutions for the non-convex model remain optimal for the proposed model.

We focus on a class of additive separable penalty functions, where the penalty function is the sum of a set of loss functions for data points. Particularly, we consider a class of loss functions $L(x) = x^p$ for $0 < p < \infty$ and prove that the generalized multi point model will yield meaningful non-trivial solutions for this family. Finally, we investigate the behavior of our models under three decision maker attitudes: loss-averse ($p > 1$),
loss-neutral \((p = 1)\), and loss-seeking \((p < 1)\). We demonstrate that the choice of \(p\) in the multi point inverse problem can profoundly affect the outcome of the multi point inverse models.

## 4.1 Generalized Single Point Inverse Linear Optimization

Consider the classical Single Objective Forward Optimization (SOFO) problem:

\[
\text{SOFO}(c) : \quad \text{minimize} \quad c'x \\
\text{subject to} \quad Ax \geq b, \tag{4.1}
\]

where \(x \in \mathbb{R}^n\), \(c \in \mathbb{R}^n\), \(b \in \mathbb{R}^m\), and \(A \in \mathbb{R}^{m \times n}\). Note that one can multiply both sides of the constraints in (4.1) by a positive diagonal function. Therefore, without loss of generality we assume that rows of \(A\) are normalized such that for all \(i \in \{1,2,\ldots,m\}\), \(\|a_i\|_1 = 1\) by multiplying \(\text{diag}(\frac{1}{\|a_1\|_1}, \ldots, \frac{1}{\|a_m\|_1})\) to both sides of the constraints in (4.1).

Let us start with the following Single point Inverse Optimization (SIO) problem that was proposed in [11]:

\[
\text{SIO}(x^0) : \quad \text{minimize} \quad 0 \\
\text{subject to} \quad A'y = c, \tag{4.2}
\]

\[
y \geq 0,
\]

\[
b'y = c'x^0, \\
\|c\|_1 = 1,
\]

where the first and second constraints ensure dual feasibility, the third constraint enforces strong duality and the last one is used for normalization. We assume that \(x^0\) belongs to the feasible region of problem (4.1). In this dissertation we assume that the feasible polyhedron in the forward problem induced by \(A\) and \(b\) is full-dimensional, since in
the non-full-dimensional cases formulation (4.2) may yield trivial optimal solutions (see Appendix A). For cases where the feasible polyhedron is non-full-dimensional we use the full-dimensional counterpart of the polyhedron in the formulations. The existence of the full-dimensional counterpart and other issues related to the full-dimensionality of the feasible polyhedron are discussed in Appendix A. Furthermore, we assume that the full-dimensional representation does not contain any redundant constraint.

Therefore, we examine the properties of inverse linear models with the following assumptions:

1. The feasible region for the forward problem (4.1) is always full-dimensional or its full-dimensional counterpart is considered in the formulation.

2. \( \mathbf{A} \) and \( \mathbf{b} \) are normalized such that \( \|\mathbf{a}_i\|_1 = 1 \) for all \( i \in \{1, \ldots, m\} \).

3. There does not exist any redundant constraint in the full-dimensional representation, i.e., for all \( i \in \{1, \ldots, m\} \), with removing only the \( i^{th} \) row (\( \mathbf{a}_i^\prime \mathbf{x} \geq \mathbf{b}_i \) constraint) the feasible region will be affected.

4. The given solution \( \mathbf{x}^0 \) is a feasible solution to the forward problem (4.1).

In [11], Chan et al. proposed the following Relaxed Single Point Inverse Optimization (RSIO) problem that minimizes the absolute duality gap:

\[
\text{RSIO : } \min_{\mathbf{y}, \mathbf{c}} \quad \xi^\prime \mathbf{y} \\
\text{subject to } \mathbf{y} \geq 0, \quad A^\prime \mathbf{y} = \mathbf{c}, \\
\|\mathbf{c}\|_1 = 1,
\]

(4.3)

where \( \xi \in \mathbb{R}^m \) is the complementary slackness vector and its \( i^{th} \) element is equal to \( \xi_i = \mathbf{a}_i^\prime \mathbf{x}^0 - \mathbf{b}_i \). To see the equivalence with the formulations proposed by Chan et al. in
The main weakness of model (4.3) is the normalization constraint that makes the problem non-convex. Despite this non-convexity the authors derive closed-form solutions for formulation (4.3). We propose the following Generalized Single point Inverse Optimization (GSIO) problem (4.3) that provides the same closed-form solutions in single point cases, and in addition, is also tractable for generalized multi-point cases whenever closed-form solutions do not exist:

\[ \text{GSIO : } \minimize_{y,c} \xi' y \]

subject to \[ y \geq 0, \]
\[ A'y = c, \]
\[ y'e \geq 1, \]

where \( e \) is the vector of all ones. Theorem 4.1.1 illustrates the connection between problems (4.4) and (4.3), and also provides closed-form solution for formulations (4.3) and (4.4). In particular, it shows that there exists an optimal cost vector for formulation (4.4) such that \( y_i^* = 1, y_j^* = 0 \) for \( j \neq i \) and consequently \( c^* = A'y^* = a_i \). Therefore, there exists at least an optimal cost vector \( c^* \) for the single point inverse problem (4.4) that is perpendicular to one of the constraints.

**Theorem 4.1.1.** The following statements are true:

1. Any feasible point for problem (4.3) is also feasible for problem (4.4).

2. There exists an optimal solution for formulations (4.3) and (4.4) which is attained for \( y_i^* = 1, \) and \( y_j^* = 0 \) for all \( j \neq i, \) where \( i = \arg \min \{ \xi_i \} \).
Proof. 1. Let \((y, c)\) be a feasible solution for problem (4.3); we show that it is also feasible for (4.4):

\[
\| \sum_{i=1}^{m} y_i a_i \|_1 = \| c \|_1 = 1 \Rightarrow 1 = \| \sum_{i=1}^{m} y_i a_i \|_1 (1) \leq \sum_{i=1}^{m} y_i \| a_i \|_1 (2) = \sum_{i=1}^{m} y_i
\]

where (1) follows from the triangle inequality and (2) is because \(\| a_i \|_1 = 1\) for all \(i \in \{1, \ldots, m\}\). Thus, \((y, c)\) is feasible for (4.4) and the proof is complete.

2. Since the feasible region of problem (4.3) is a subset of the feasible region of problem (4.4) and the objectives are the same, then any optimal solution for (4.4) is an optimal solution for (4.3), provided that it is feasible for (4.3).

Formulation (4.4) is an LP and has a finite cost, since \(x^0\) is feasible (i.e., \(Ax^0 \geq b \iff \xi \geq 0\)). Also note that \(c\) appears only in \(c = A'y\) constraint and does not affect the feasible region. In fact, one can disregard the \(c = A'y\) constraint and solve the optimization problem to find the optimal vector \(y^*\) and then determine \(c^* = A'y^*\). Consequently, there exists at least one optimal solution which is attained at a vertex of the polyhedron induced by \(y \geq 0\) and \(y'e = 1\) that has the following form:

\[
y^*_i = 1 \quad \text{and} \quad y^*_j = 0 \quad \forall j \neq i, \quad \text{where} \quad i = \arg \min_t \{\xi_t\}.
\]

Note that the latter optimal solution is also a feasible solution to (4.3), since \(\|a_i\|_1 = 1\) and \(\|c^*\|_1 = \|A'y^*\|_1 = \|a_i\|_1 = 1\).

\(\square\)

Although Theorem 4.1.1 proves that there exists an optimal solution \((y^*, c^*)\) for problem (4.4) that is also optimal for problem (4.3), it does not guarantee that problem (4.4)
does not return a trivial solution (i.e., $\|c^*\|_1 = 0$). The following theorem proves that the set of observed solutions that can yield a trivial solution have measure zero.

**Theorem 4.1.2.** Except for a measure zero set, if $x^0 \in \{x \mid Ax \geq b\}$, then $\|c\| > 0$ for all the optimal solutions $(y^*, c^*)$ of GSIO($x^0$).

**Proof.** First note that if the solution to $i = \arg \min_{t} \{\xi_t\}$ is unique then $c^*$ is unique and is equal to $a_i$ as $a_i \neq 0$ for all $i \in \{1, \ldots, m\}$. Therefore, for $c^* = 0$ the solution to $i = \arg \min_{t} \{\xi_t\}$ cannot be unique. Thus, there exist $i$ and $j$ such that:

$$\xi_i = \xi_j \leq \xi_t \quad \forall t \in \{1, \ldots, m\}.$$ 

However, $\xi_i = \xi_j$ implies that:

$$a_i x^0 - b_i = a_j x^0 - b_j \Rightarrow (a_i - a_j) x^0 = b_i - b_j.$$ 

Therefore, for $c^* = 0$, the observed solution must lie on a $(n-1)$-dimensional hyperplane defined by at least two constraints. Hence, for each pair $(i, j)$ of constraints there is at most one hyperplane $H_{i,j}$ that can potentially yield a trivial solution. Consequently, the set of all solutions $x^0$ that may return a trivial solution is a subset of $\bigcup_{i,j} H_{i,j}$, which is the union of at most $\binom{m}{2}$ lower $(n-1)$-dimensional hyperplanes in $\mathbb{R}^n$. Therefore, the set of observed solutions that may produce a trivial cost vector is a subset of the union of a countable number of measure zero sets and is therefore measure zero.

Theorem 4.1.1 clarifies the connection of our generalized formulation (4.4) and formulation (4.3), and opens the door for a broader analysis of multi point inverse linear optimization.
4.2 Generalized Multi Point Inverse Optimization

In [8], Keshavarz et al. proposed a formulation for imputing a convex objective function by using \( Q \) observations. In order to find a convex objective function that best suits the observations, a penalty function of the form \( \sum_{q=1}^{Q} \Phi(\xi_{\text{stat}}^{(q)}, \xi_{\text{comp}}^{(q)}) \) is minimized, where \( \Phi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow [0, \infty] \) is an arbitrary convex penalty function, and \( \xi_{\text{stat}}^{(q)} \) and \( \xi_{\text{comp}}^{(q)} \) are the stationary and complementary slackness residuals for the \( q^{th} \) data point. Although their formulation is general and can be applied to a wide variety of problems, the choice of \( \Phi \) is not carefully studied. In a similar framework for linear problems we show that the choice of penalty function can profoundly affect the outcome of the optimization.

For the case where only a single observation is available, relative or absolute duality gaps are minimized as measures of optimality. We propose a generalized multi point inverse formulation by considering a regret function and we study the behaviour of the imputed cost vector for different choices of the regret function. Special cases of our generalized model includes optimization problems that aim at minimizing the sum of duality gaps or the sum of squares of duality gaps as measures of optimality.

4.2.1 A Generalized Multi Point Inverse Linear Model

In this section we will extend our methodology to the case where multiple feasible solutions \( x^{(1)}, \ldots, x^{(Q)} \) to the forward problem (4.1) are available. Consider the following Generalized Multi Point Inverse Optimization (GMIO) problem:

\[
\begin{align*}
\text{GMIO}_\Phi : \quad & \text{minimize} \quad \Phi(\epsilon^{(1)}, \ldots, \epsilon^{(Q)}) \\
\text{subject to} \quad & \epsilon^{(q)} = y^{(q)'r}, \quad \forall q \in \{1, \ldots, Q\}, \\
& A^{(q)'y^{(q)}} = c, \quad \forall q \in \{1, \ldots, Q\}, \quad (4.5) \\
& y^{(q)'e} \geq 1, \quad \forall q \in \{1, \ldots, Q\}, \\
& y^{(q)} \geq 0, \quad \forall q \in \{1, \ldots, Q\},
\end{align*}
\]
where $\mathbf{A}^{(q)}$ and $\mathbf{b}^{(q)}$ define the feasible polyhedron for the $q^{th}$ data point, and $\xi^{(q)} = \mathbf{A}^{(q)}\mathbf{x}^{(q)} - \mathbf{b}^{(q)}$ is the complementary slackness vector for the $q^{th}$ data point. Finally, $\Phi : \{\mathbb{R}_+ \cup \{0\}\}^Q \rightarrow \mathbb{R}_+ \cup \{0\}$ is a penalty function which is increasing in each component, with $\Phi(0) = 0$. For the sake of simplicity, we assume that $\Phi$ is an additive separable function. Thus, $\Phi$ has a representation as follows:

$$\Phi(\epsilon^{(1)}, \ldots, \epsilon^{(Q)}) = \sum_{q=1}^{Q} \mathcal{L}_q(\epsilon^{(q)})$$

where $\mathcal{L}_q : \mathbb{R}_+ \cup \{0\} \rightarrow \mathbb{R}_+ \cup \{0\}$ is an arbitrary loss function. In this chapter, we analyze the case where data points share the same loss function (i.e., $\mathcal{L} = \mathcal{L}_q$). Hereinafter, we also assume that $\mathbf{A} = \mathbf{A}_q$, $\mathbf{b} = \mathbf{b}_q$, and $\xi^{(q)} = \mathbf{A}\mathbf{x}^{(q)} - \mathbf{b}$.

### 4.2.2 Behaviour of the Inverse Model with Respect to the Penalty Function

In this section we investigate the properties of formulation (4.5) for a family of functions $\mathcal{L}(x) = x^p$ where $0 < p < \infty$. We demonstrate that the outcome of the optimization may dramatically change by varying the value of $p$. We will investigate the behaviour of the following formulation when $0 < p < \infty$:

$$\begin{align*}
\text{minimize} & \quad \sum_{q=1}^{Q} \left( y^{(q)'\xi^{(q)}} \right)^p \\
\text{subject to} & \quad \mathbf{A}'y^{(q)} = \mathbf{c}, \quad \forall q \in \{1, \ldots, Q\}, \\
& \quad y^{(q)'\mathbf{e}} \geq 1, \quad \forall q \in \{1, \ldots, Q\}, \\
& \quad y^{(q)} \geq 0, \quad \forall q \in \{1, \ldots, Q\}.
\end{align*}$$

(4.6)

**Proposition 4.2.1.** Formulation (4.6) has an optimal solution where $y^{(i)*} = y^{(j)*}$ for all $i, j \in \{1, \ldots, Q\}$ (i.e., $y^{(q)*} = y^*$) and $\mathbf{c}^* = \mathbf{A}'y^*$. 
Proof. Note that $y^{(q)} = e$, and $c = A'e$ yield a feasible solution for formulation (4.6) and the objective value is bounded below by zero as $\xi^{(q)} \geq 0$ and $y^{(k)} \geq 0$, therefore, formulation (4.6) has an optimal solution. Let $(y^{(1)*}, \ldots, y^{(q)*}, c^*)$ be an optimal solution for problem (4.6), note that:

$$0 \leq y^{(q)*}'\xi^{(q)} = y^{(q)*}'(Ax^{(q)} - b) = c^*'x^{(q)} - b'y^{(q)*}.$$ 

Let $j = \arg\max_q \{b'y^{(q)*}\}$, since $A'y^{(j)*} = c^*$ and $y^{(j)*} \geq 0$, $y^{(j)*}$ is dual feasible for all the given solutions $x^{(1)}, \ldots, x^{(Q)}$. Therefore, feasibility of $x^{(q)}$ for the forward problem and weak duality concludes that:

$$y^{(q)*}'\xi^{(q)} = c^*'x^{(q)} - b'y^{(q)*} \geq c^*'x^{(q)} - b'y^{(j)*} = y^{(j)*}'\xi^{(q)}, \quad \forall q \in \{1, \ldots, Q\},$$

where the inequality above holds as a result of $j = \arg\max_q \{b'y^{(q)*}\}$.

Consequently, by setting $y^{(q)}$ to be equal to $y^{(j)*}$ for each $q$, problem (4.6) remains feasible and the optimal cost does not increase. Therefore, there exists an optimal solution for formulation (4.6), where $y^{(q)*} = y^*$ for all $q$ and $c^* = A'y^*$.

Theorem 4.2.1 states that an optimal solution for the following formulation is also optimal for formulation (4.6):

$$\text{GMIO}_p : \begin{array}{ll}
\text{minimize} & \sum_{q=1}^{Q} (y'\xi^{(q)})^p \\
\text{subject to} & A'y = c, \\
& y'e \geq 1, \\
& y \geq 0.
\end{array}$$

One can think of the multi point problem as a process where a cost vector $c$ is chosen such that the overall loss is minimized. Therefore, the value of $p$ determines how individuals
or data points value their loss. Particularly, data points act as individuals who compete for the cost vector and impose a cost on the system based on their level of satisfaction for a chosen $c$. Taking the absolute duality gap to be the real value of loss for each individual, individuals can be loss averse ($p > 1$), loss neutral ($p = 1$), or loss seeking ($0 < p < 1$).

### 4.2.2.1 A Motivational Example

To illustrate the effect of the loss function, we provide a numerical example where the data points and the constraints are fixed and the cost vector is imputed using formulation (4.7) for different values of $p$. Consider the feasible region determined by the following constraints:

$$
0.7692x_1 + 0.2308x_2 \geq 0.0978,
0.9141x_1 - 0.0859x_2 \geq -0.0847,
0.3989x_1 - 0.6011x_2 \geq -0.7231,
-0.6833x_1 - 0.3167x_2 \geq -0.6318,
-0.9842x_1 - 0.0158x_2 \geq -0.4868,
-0.8079x_1 + 0.1921x_2 \geq -0.3010,
-0.4494x_1 + 0.5506x_2 \geq -0.1303. \tag{4.8}
$$

Suppose that the following 10 data points are observed:

$$(0.1801, 0.0062), (0.2439, 0.0744), (0.3367, 0.1054), (0.0931, 0.2047),$$
$$(0.1162, 0.0899), (0.1624, 0.4527), (0.0844, -0.0403), (0.1672, 0.1302),$$
$$(0.2054, 0.1054), (0.2202, 0.0124). \tag{4.9}
$$

Figure 4.1 depicts the feasible region and data points given above.

Table 4.1 provides the optimal cost vectors for different values of $p$ that were determined by solving problem (4.7). As it is illustrated in Table 4.1, formulation (4.7) may
Figure 4.1: The feasible region defined by the inequalities in (4.8) along with the observed solutions given in (4.9). The red arrows indicate the range of cost vectors produced by GMIO\textsubscript{p} for different values of \(p\).

<table>
<thead>
<tr>
<th>Value of (p)</th>
<th>Estimated cost vector ((c^*))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p \to 0)</td>
<td>((0.7692, 0.2308))</td>
</tr>
<tr>
<td>0.3</td>
<td>((0.7692, 0.2308))</td>
</tr>
<tr>
<td>0.5</td>
<td>((-0.4494, 0.5506))</td>
</tr>
<tr>
<td>1</td>
<td>((-0.4494, 0.5506))</td>
</tr>
<tr>
<td>2</td>
<td>((0.46386, 0.53614))</td>
</tr>
<tr>
<td>5</td>
<td>((0.59663, 0.40337))</td>
</tr>
<tr>
<td>10</td>
<td>((0.60956, 0.39044))</td>
</tr>
<tr>
<td>20</td>
<td>((0.66586, 0.33414))</td>
</tr>
<tr>
<td>(p \to \infty)</td>
<td>((-0.44439, 0.55561))</td>
</tr>
</tbody>
</table>

Table 4.1: Imputed cost vectors for the given solutions (4.9) and constraints (4.8) for different values of \(p\) in formulation (4.7).

return a considerably different cost vector depending on the value of \(p\). In the following sections we provide an insight on how the inverse model behaves for different ranges of \(p\).

4.2.2.2 Loss Neutral Attitudes

In the case where decision maker is loss neutral (i.e., \(p = 1\)), formulation (4.7) is equivalent to minimizing the sum of the duality gaps. Theorem 4.2.1 proves that in the loss neutral case problem (4.7) is equivalent to a single point problem. Also corollary 4.2.1.1 shows
that in the loss neutral scenario formulation (4.7) always produces non-trivial solutions except for a measure zero set.

**Theorem 4.2.1.** There exists an optimal solution for formulation (4.7), which is obtained at \( y^*_i = 1 \) and \( y^*_j = 0 \) for all \( j \neq i \), where \( i = \arg \min_t \left\{ \sum_{q=1}^{Q} \xi_t^{(q)} \right\} \), and \( \xi_t^{(q)} \) denotes the \( t \)th component of the complementary slackness vector for the \( q \)th data point. Furthermore, the problem is equivalent to solving a single point problem that finds the optimal cost vector for the centroid of the data set.

**Proof.** \( GMIO_1 \) is an LP and has a bounded optimal value (since \( \xi^{(q)} \geq 0 \)), therefore, at least one of its optimal solutions is attained at a vertex of the feasible region, which has the following form: \( y_i = 1 \) and \( y_j = 0 \) for all \( j \neq i \). Note that the objective can be simplified as:

\[
\sum_{q=1}^{Q} y'_q \xi^{(q)} = y' \sum_{q=1}^{Q} \xi^{(q)}.
\]

Thus, there exists an optimal basic feasible solution where \( y^*_i = 1 \) and \( y^*_j = 0 \) for all \( j \neq i \), where \( i = \arg \min_t \left\{ \sum_{q=1}^{Q} \xi_t^{(q)} \right\} \). Thus, we have:

\[
c^* = A'y^* \Rightarrow \|c^*\|_1 = \|a_i\|_1 = 1
\]

In addition, minimizing \( y' \sum_{q=1}^{Q} \xi^{(q)} \) is equivalent to minimizing \( y' \frac{\sum_{q=1}^{Q} \xi^{(q)}}{Q} \), also define:

\[
\bar{\xi} \triangleq \frac{\sum_{q=1}^{Q} \xi^{(q)}}{Q} = \frac{\sum_{q=1}^{Q} (Ax^{(q)} - b)}{Q} = A \frac{\sum_{q=1}^{Q} x^{(q)}}{Q} - b
\]

\[
\Rightarrow \bar{\xi} = A\bar{x} - b,
\]

where \( \bar{x} = \frac{\sum_{q=1}^{Q} x^{(q)}}{Q} \) is the centroid of data points. Hence, formulation \( GMIO_1 \) is equiv-
alent to the following single point problem:

\[
\text{GSIO}(\bar{x}) : \quad \text{minimize} \quad y' \bar{\xi}
\]

subject to \( y \geq 0, \) \( A'y = c, \) \( y'e \geq 1, \) (4.10)

which concludes the proof.

\[\square\]

Corollary 4.2.1.1. Except for a measure zero set in \( \mathbb{R}^{n \times Q} \) if \((x^{(1)}, \ldots, x^{(Q)}) \in \{x \mid Ax \geq b\}^{Q}\), for all the optimal solutions \((y^*, c^*)\) of \(\text{GMIO}_1\) we have \(\|c^*\|_1 > 0\).

Proof. Let \( S \subset \mathbb{R}^{n \times Q} \) be the set of of all data sets \((x^{(1)}, \ldots, x^{(Q)})\) for which \(\text{GMIO}_1\) has a trivial solution (i.e., \(\|c^*\|_1 = 0\)). Let \(x^{(1)}, \ldots, x^{(Q)}\) be a given data set for which \(\text{GMIO}_p\) yields a trivial solution, as a result of Theorem (4.2.1), the equivalent single point inverse problem \(\text{GSIO}(\bar{x})\) has a trivial solution, where \(\bar{x} = \frac{\sum_{q=1}^{Q} x^{(q)}}{Q}\). Also let \(T\) denote the set of all \(x^0 \in \mathbb{R}^n\) for which \(\text{GSIO}(x^0)\) yields a trivial solution.

For a data set \(x^{(1)}, \ldots, x^{(Q)}\) to yield a trivial solution, the following should hold:

\[\bar{x} \in T \Rightarrow \left( \frac{\sum_{q=1}^{Q-1} x^{(q)}}{Q} + \frac{x^{(Q)}}{Q} \right) \in T \Rightarrow x^{(Q)} \in E^S_{x^{(1)}, \ldots, x^{(Q-1)}},\]

where the set \(E^S_{x^{(1)}, \ldots, x^{(Q-1)}}\) is a translated and scaled version of \(T\) defined as:

\[E^S_{x^{(1)}, \ldots, x^{(Q-1)}} = \{z \mid z = Qx - \sum_{q=1}^{Q-1} x^{(q)}, x \in T\}.\]

Note that \(E^S_{x^{(1)}, \ldots, x^{(Q-1)}}\) is also measure zero, since \(T\) is measure zero. Also note that \(E^S_{x^{(1)}, \ldots, x^{(Q-1)}}\) is a “slice” of \(S\) for fixed \(x^{(1)}, \ldots, x^{(Q-1)}\). For any \((x^{(1)}, \ldots, x^{(Q-1)}) \in \mathbb{R}^{n \times (Q-1)}\) the slice \(E^S_{x^{(1)}, \ldots, x^{(Q-1)}}\) is measure zero in \(\mathbb{R}^n\), therefore, by using Cavalierie’s principle [57] the set \(S\) is measure zero in \(\mathbb{R}^{n \times Q}\). \[\square\]
Note that the extreme points of the \( y \) polyhedron (i.e., \( y_i^* = 1, y_j^* = 0 \) for \( j \neq i \)) are in \( 1 - 1 \) correspondence with the cost vectors perpendicular to constraints of the feasible region defined by \( Ax \geq b \). Consequently, in Table 4.1 the optimal cost vector for the cost neutral case is attained at an extreme point of the \( y \) polyhedron (i.e., \( y \geq 0, \text{ and } y'e = 1 \)) as it is perpendicular to the facet generated by \(-0.4494x_1 + 0.5506x_2 \geq -0.1303\).

### 4.2.2.3 Loss Averse Attitude

For the loss averse case, where the loss functions are convex (i.e., \( p > 1 \)), we again prove that formulation (4.7) returns non-trivial solutions except for a measure zero set. We also investigate the behaviour of the model in the extreme case where \( p \to \infty \), and show that the model behaves similarly to a model that minimizes the maximum duality gap.

In the loss averse case, however, the solution is not necessarily attained at an extreme point of the feasible polyhedron of the inverse problem (i.e., \( y \) polyhedron). Bear in mind that the extreme points of the inverse polyhedron are corresponding to cost vectors that are perpendicular to the constraints of the forward problem. In the loss averse case, as also illustrated in Table 4.1 the optimal cost vector for \( p = 5 \) is equal to \((0.59663, 0.40337)\), which is not perpendicular to any facet of the feasible polyhedron for the forward problem (i.e., extreme point of the inverse polyhedron) and the solutions which are perpendicular to the facets all yield a higher cost. Therefore, unlike the loss neutral case the multi point problem cannot be equivalent to a single point inverse problem, where there always exists an optimal solution which is perpendicular to one of the facets.

**Theorem 4.2.2.** In the case where \( p > 1 \) formulation (4.7) does not yield a trivial optimal solution except for a measure zero set.

*Proof.* Let \( y^* \) be an optimal solution for (4.7) such that \( c = A'y^* = 0 \). Since \( y^* \) is an optimal solution for a convex problem, there exist dual variables \( \nu \) and \( \lambda \) that satisfies
KKT conditions:

Stationarity \( \sum_{q=1}^{Q} (y^*(\xi^{(q)})^{p-1} + \nu^* \cdot e - \lambda^* = 0. \)

Complementary Slackness \( \nu^*(y^* \cdot e - 1) = 0, \)
\( \lambda_i^* y_i^* = 0, \quad i \in \{1, \ldots, m\}. \)

Dual Feasibility \( \lambda^* \geq 0, \quad \nu^* \leq 0, \)

Primal Feasibility \( y^* e \geq 1, \quad y^* \geq 0. \)

Trivial Solution \( \Rightarrow \ c^* = A' y^* = 0. \)

From the definition of \( \xi^{(q)} \) we have:

\( y^* \xi^{(q)} = y^* (A x^{(q)} - b) = (A' y^*) x^{(q)} - b' y^* \quad \forall q \in \{1, \ldots, Q\}, \) (4.12)

where relation (1) is followed by the fact that for a trivial solution \( A' y^* = 0. \) Replacing (4.12) into (4.11) yields:

Stationarity \( (b' y^*)^{p-1} \sum_{q=1}^{Q} \xi^{(q)} + \nu^* \cdot e - \lambda^* = 0. \)

Complementary Slackness \( \nu^*(y^* \cdot e - 1) = 0, \)
\( \lambda_i^* y_i^* = 0, \quad i \in \{1, \ldots, m\}. \)

Dual Feasibility \( \lambda^* \geq 0, \quad \nu^* \leq 0, \)

Primal Feasibility \( y^* e \geq 1, \quad y^* \geq 0. \)

(4.13)
Pick \( \tilde{\nu} \) and \( \tilde{\lambda} \) such that:

\[
\nu^* = (-b'y^*)^{p-1}\tilde{\nu} \quad \& \quad \lambda^* = (-b'y^*)^{p-1}\tilde{\lambda}
\]

Rewriting (4.13) in terms of \( \tilde{\nu} \) and \( \tilde{\lambda} \) yields:

Stationarity \((-b'y^*)^{p-1}\left(\sum_{q=1}^{Q} \xi^{(q)} + \tilde{\nu}e - \tilde{\lambda}\right) = 0\),

Complementary Slackness \((-b'y^*)^{p-1}\tilde{\nu}(y^*e - 1) = 0\),

\[(-b'y^*)^{p-1}\tilde{\lambda}_iy_i^* = 0, \quad i \in \{1, \ldots, m\}, \quad (4.14)\]

Dual Feasibility \((-b'y^*)^{p-1}\tilde{\lambda} \geq 0\),

\[(-b'y^*)^{p-1}\tilde{\nu} \leq 0,\]

Primal Feasibility \(y^*e \geq 1 \quad \& \quad y^* \geq 0\).

Since \(-b'y^* = y^*\xi^{(q)}\) is positive, it can be factored out without affecting the inequalities. Therefore, \(y^*, \tilde{\nu}, \) and \(\tilde{\lambda}\) satisfy KKT conditions for GMIO\(1\). Consequently, for a set of observed solutions \((x^{(1)}, \ldots, x^{(q)})\) if GMIO\(p\) yields a trivial solution, GMIO\(1\) also has a trivial solution. However, GMIO\(1\) returns a trivial solution for only a measure zero set in \(\mathbb{R}^{n\times Q}\), which concludes the proof.

The following theorem demonstrates the connection between minimizing the maximum regret and loss aversion.

**Theorem 4.2.3.** For any sequence of optimal solutions \(\{y^*_n\}_{n=1}^{\infty}\) of GMIO\(_n\) that have nonzero optimal values, there exists a subsequence \(\{y^*_n\}_{n=1}^{\infty}\) that converges to \(y^*\), where
y* is an optimal solution for the following problem:

\[ \begin{align*}
\text{GMIO}_\infty : & \quad \text{minimize}_{y, \epsilon} \| \epsilon \|_\infty \\
& \text{subject to} \quad \epsilon_q = y' \xi^{(q)}, \quad \forall q \in \{1, \ldots, Q\}, \\
& \quad y'.e \geq 1, \\
& \quad y \geq 0.
\end{align*} \] (4.15)

**Lemma 4.2.4.** Define \( f_n \) as follows:

\[ f_n(\epsilon) \triangleq \sqrt[n]{\epsilon_1^n + \cdots + \epsilon_Q^n} = \| \epsilon \|_n \]

and also let \( f(\epsilon) = \| \epsilon \|_\infty \). Suppose that \( \epsilon \in C \), where \( C \) is a bounded set, then \( f_n \) converges uniformly to \( f \) on \( C \).

**Proof.** Since \( C \) is bounded, there exist \( M > 0 \) such that:

\[ \epsilon_i < M \quad \forall \epsilon \in C, \forall i \in \{1, \ldots, Q\}, \]

thus, we have:

\[ \| \epsilon \|_\infty \leq \sqrt[n]{\epsilon_1^n + \cdots + \epsilon_Q^n} \leq \sqrt[n]{Q} \| \epsilon \|_\infty \leq \| \epsilon \|_\infty + (\sqrt[n]{Q} - 1)M, \]

furthermore, for each \( \delta > 0 \) there exists \( N_\delta \) such that \( (\sqrt[n]{Q} - 1)M < \delta \) for all \( n > N_\delta \), which yields:

\[ f(\epsilon) \leq f_n(\epsilon) \leq f(\epsilon) + \delta \quad \forall \epsilon \in C, \forall n > N_\delta, \]

which completes the proof. \( \Box \)

**Proof of Theorem 4.2.3.** Note that the optimal solution of problem (4.7) is bounded below by zero, since \( \xi^{(q)} \geq 0 \). For the case where the optimal value of formulation (4.7) is
positive, any optimal solution $y^*$ satisfies $y^* e \geq 1$. Therefore, all the $y_n^*$'s satisfy the following constraints $y_n^* e = 1$ and $y^* \geq 0$, which define a compact set. Since the sequence \{\$y_n^*$\} is defined on a compact set, it has a subsequence \{\$y_{i_n}^*$\} which converges to some $y^*$. Let $\epsilon^{*n}$ be the vector of duality gaps corresponding to $y_{i_n}^*$, where $\epsilon^{*n}_q = y_{i_n}^* \xi^{(q)}$, note that $y_{i_n}^* \to y^*$ yields $\epsilon^{*n} \to \epsilon$.

Furthermore, $f_n$ converges uniformly to $f$ which is equivalent to:
\[
\forall \delta > 0 \quad \exists N_1 \text{ s.t. } |f_n(\epsilon^{*n}) - f(\epsilon^{*n})| \leq \frac{\delta}{3}.
\]

In addition, since $\epsilon^{*n} \to \epsilon$ and $f$ is continuous, we have:
\[
\forall \delta > 0 \quad \exists N_2 \text{ s.t. } |f(\epsilon^{*n}) - f(\epsilon^*)| \leq \frac{\delta}{3}.
\]

Also let $\tilde{\epsilon}$ be an optimal set of vector of duality gaps for GMIO$_\infty$, then we have:
\[
\forall \delta > 0 \quad \exists N_3 \text{ s.t. } |f_n(\tilde{\epsilon}) - f(\tilde{\epsilon})| \leq \frac{\delta}{3}.
\]

Let $N = \max(N_1, N_2, N_3)$, for $n > N$ we have:
\[
|f_n(\epsilon^{*n}) - f(\epsilon^*)| \leq |f_n(\epsilon^{*n}) - f(\epsilon^{*n})| + |f(\epsilon^{*n}) - f(\epsilon^*)| \leq \frac{2\delta}{3},
\]
\[
f_n(\epsilon^{*n}) \leq f_n(\tilde{\epsilon}) \quad \& \quad f(\tilde{\epsilon}) \leq f(\epsilon^*) \quad \text{(optimality $\epsilon^{*n}$ and $\tilde{\epsilon}$ for $f_n$ and $f$)}
\]
\[
\Rightarrow f(\tilde{\epsilon}) \leq f(\epsilon^*) \leq f_n(\epsilon^{*n}) + \frac{2\delta}{3} \leq f_n(\tilde{\epsilon}) + \frac{2\delta}{3} \leq f(\tilde{\epsilon}) + \delta
\]
\[
\Rightarrow \forall \delta > 0 \quad f(\tilde{\epsilon}) \leq f(\epsilon^*) \leq f(\tilde{\epsilon}) + \delta \quad \Rightarrow f(\tilde{\epsilon}) = f(\epsilon^*).
\]

Thus, $\epsilon^*$ has the same cost as $\tilde{\epsilon}$. Consequently, the corresponding dual vector $y^*$ is also optimal for GMIO$_\infty$ which concludes the proof.

In Theorem 4.2.3 the convergence holds only for a subsequence as the sequence
\(\{y^*_n\}_{n=1}^\infty\) may not converge, consider the following forward problem:

\[
\begin{align*}
\text{minimize} & \quad cx \\
\text{subject to} & \quad -1 \leq x \leq 1.
\end{align*}
\tag{4.16}
\]

Note that there are only two normalized cost vectors \((c = -1 \text{ and } 1)\) for problem (4.16). Let \(a_n = \frac{(-1)^n}{n}\) be a sequence of observations, the optimal sequence of cost vectors returned by \(\text{GMIO}_n\) for \(a_n\) is \((-1)^{n+1}\) which is not convergent.

### 4.2.2.4 Loss Seeking Attitudes

For the class of concave functions where \(0 < p < 1\), the individuals are known to be loss seeking. In this scenario we also consider the extreme case for being loss seeking (i.e., \(p \to 0\)). The following theorem indicates the connection between being loss seeking and minimizing the geometric mean of the losses (duality gaps).

**Theorem 4.2.5.** The optimal solution for formulation (4.7) is obtained for \(y^*_i = 1\) and \(y^*_j = 0\) for all \(j \neq i\), where \(i = \arg \min_t \left\{ \sum_{q=1}^Q \left( y^* \xi_{t}^{(q)} \right)^p \right\} \). Also for the case where \(p \to 0\), it is equivalent to minimizing the product of duality gaps and the optimal solution is obtained for \(y^*_i = 1\) and \(y^*_j = 0\) for all \(j \neq i\), where \(i = \arg \min_t \left\{ \prod_{q=1}^Q |a_t x^{(q)} - b_t| \right\} \).

**Proof.** For \(0 < p < 1\) formulation (4.7) is a concave minimization problem on a polyhedron and has at least one optimal solution that is attained at a vertex of the feasible polyhedron [58]. The set of vertices of the inverse problem polyhedron (i.e., feasible polyhedron for \(\text{GMIO}_p\)) has the following structure: \(y_i = 1\) and \(y_j = 0\) for all \(j \neq i\). Let \(\xi_{t}^{(q)} = a_t^i x^{(q)} - b_t\), then for each \(0 < p < 1\) the objective is to find \(i\) such that:

\[
i = \arg \min_t \left\{ \sum_{q=1}^Q \left( \xi_{t}^{(q)} \right)^p \right\}, \quad p \to 0,
\]
writing the Taylor series of \( \sum_{q=1}^{Q} \left( \xi_t^{(q)} \right)^p \) yields:

\[
\sum_{q=1}^{Q} \left( \xi_t^{(q)} \right)^p = \sum_{q=1}^{Q} e^{p \log \xi_t^{(q)}} \approx \sum_{q=1}^{Q} \left[ 1 + p \log \xi_t^{(q)} + O(p^2) \right], \quad p \to 0.
\]

Therefore, for \( p \to 0 \) the optimal solution is found by solving the following equation:

\[
i = \arg \min_t \left\{ \sum_{q=1}^{Q} 1 + p \log \xi_t^{(q)} \right\} = \arg \min_t \left\{ \sum_{q=1}^{Q} \log \xi_t^{(q)} \right\} = \arg \min_t \prod_{q=1}^{Q} \xi_t^{(q)}.
\]

\[
\square
\]

It is worth mentioning that in the loss seeking case there exists an optimal solution that is attained at a vertex of the inverse problem feasible polyhedron, but the problem is not necessarily equivalent to a single point problem. This is due to the fact that a single data point \( \bar{x} \) may not exists whose \( t \)th complementary residual is equal to \( \bar{\xi}_t = \sum_{q=1}^{Q} \left( \xi_t^{(q)} \right)^p \). The reason why this was the case for \( p = 1 \) only is that all the constraints were linear and it was guaranteed that for the centroid \( \bar{x} \) the \( \bar{\xi}_t = \sum_{q=1}^{Q} \left( \xi_t^{(q)} \right) \) holds.

### 4.3 Conclusions

In this chapter we proposed a novel normalization technique for the inverse linear problem, which has the desired properties of the proposed model in [11] and is also convex. Furthermore, we demonstrated that the proposed single point model will return meaningful non-trivial solutions. We generalized the developed single point model to accommodate multiple data points by minimizing the \( p \)th power mean of the duality gaps. We showed that different values of \( p \) can result in dramatically different outcomes. Particularly, for \( p = 1 \) we demonstrated the equivalence of the multi point problem and a single point problem, and proved that there exists an optimal cost vector which is perpendicular to one of the facets of the forward problem. In the case where \( p < 1 \) we also showed
that there exists an optimal cost vector which is perpendicular to one of the facets of the forward polyhedron but the problem is not necessarily analogous to a single point inverse problem. For the \( p > 1 \) scenario we again proved that our formulations return non-trivial solutions and our results showed that the solution is not necessarily perpendicular to a facet, and therefore it is not equivalent to a single point inverse problem. We also investigated the behaviour of the problem in the extreme cases where \( p \to 0 \) and \( p \to \infty \).

The results presented in this section suggest that the outcome of inverse problem can considerably vary even within a small family of additive separable penalty functions. Therefore, it is crucial to fine-tune the parameters of inverse models for different applications. In fact this also a challenge in the signal processing problems and in order to overcome the issues around the choice of the norm (or equivalently the parameter \( p \)) the optimal value for \( p \) is determined by maximizing a likelihood function [59, 60]. Although it is possible to show that our framework has connections to a Maximum Likelihood Estimation (MLE) problem where \( p \) is the shape parameter of a generalized Gaussian distribution, even in the simple scenarios the likelihood function for the inverse problem becomes non-convex [61] and requires numerical approximations of the size of polytopes. Therefore, iterative heuristic methods like genetic algorithms are employed for solving them which make them intractable for large-scale applications.

In conclusion, we proposed a tractable normalization method for the inverse linear problem and we showed that in the transition from single point to multi point inverse optimization the choice of the penalty function plays a crucial role. The models introduced in this section form the basis for the stability analysis provided in Chapter 5.
Chapter 5

On the Stability of Inverse Optimization

Inverse optimization and structural estimation models have been widely used to estimate parameters of a system by observing its output. Since in practical applications data is imperfect and is subject to error and perturbations, it is crucial to investigate the stability of inversely optimized parameters. Presence of instability in inverse optimization models devalues inversely optimized estimates, since small perturbations may dramatically alter the estimated parameters. However, there has not been any research on the stability of these parameters with respect to perturbations or errors in the observed data.

In this chapter we first illustrate the instability of the inverse linear problem (4.7) for $p = 1, 2$ in Section 5.1. In Sections 5.2 and 5.3, we explore the similarities between regression and inverse optimization, which we then exploit to develop stable inverse optimization models using Lasso and Tikhonov regularization [12, 13]. Furthermore, we establish connections between the regularized inverse optimization and a robust inverse linear problem in Section 5.3. In Section 5.4, we adapt the notion of singularity sets from regression to inverse optimization to study the instability of our inverse models. Next we develop a metric for quantifying the instability of data sets in regularized and non-
regularized formulations in Section 5.5. We also propose a metric to quantify the trade-off between stability and efficiency of regularized inverse formulations in Section 5.6. Finally, we perform different simulations on synthetic and IMRT data to illustrate the impact of regularization on inverse optimization models in Section 5.7.

### 5.1 A Motivational Example

In this section we will provide some examples to illustrate the instability of both the single point and the multi point formulations (4.4) and (4.7). We present examples for unstable data sets in $p = 1, 2$ case for formulations (4.4) and (4.7). In fact, there exists a data set for other $p$’s that may exhibit instability. However, in this section our focus is on the $p = 1, 2$ case.

Figure 5.1 illustrates the instability of the GMIO$_1$ formulation corresponding to the following forward problem:

\[
\begin{align*}
\text{minimize} \quad & c_1 x_1 + c_2 x_2 \\
\text{subject to} \quad & 0.6173 x_1 + 0.3827 x_2 \geq 0.19, \\
& 0.8372 x_1 - 0.1628 x_2 \geq -0.038, \\
& 0.2866 x_1 - 0.7134 x_2 \geq -0.562, \\
& -0.4388 x_1 - 0.5612 x_2 \geq -0.774, \\
& -0.7980 x_1 + 0.2020 x_2 \geq -0.5207, \\
& -0.1775 x_1 + 0.8225 x_2 \geq 0.0072.
\end{align*}
\]

Equation (5.1)

Figure 5.1(a) depicts the single point scenario, where the observed solution is perturbed from $(0.3349, 0.1829)$ to $(0.3250, 0.1882)$ and the inversely optimized cost vector produced by formulation (4.4) changes from $(0.6173, 0.3827)$ to $(-0.1775, 0.8225)$.

In the multi point scenario presented (see Figure 5.1(b)), suppose that the following
data points (solutions) are given:

\[(0.2395, 0.2359), (0.3250, 0.1210), (0.3547, 0.1069), (0.3002, 0.1740),\]
\[(0.4054, 0.1246), (0.2470, 0.1723), (0.2742, 0.1281), (0.3596, 0.1652),\]

by perturbing only the first data point from (0.2395, 0.2359) to (0.2581, 0.2323), the inversely optimized cost vector produced by \textbf{GMIO}$_1$ changes from (0.6173, 0.3827) to (−0.1775, 0.8225).

\[(5.2)\]

Indeed the \(p = 1\) case is not the only scenario for which formulation (4.7) may exhibit instability. For instance, an unstable data set for problem (5.1) in the \(p = 2\) case is depicted in Figure 5.2, where the following solutions are observed:

\[(0.3658, 0.2694), (0.3014, 0.1316), (0.2779, 0.0910), (0.3472, 0.2111),\]
\[(0.3138, 0.1528), (0.3299, 0.1811), (0.2903, 0.1104), (0.3596, 0.2341),\]

and only the first solution is perturbed from (0.3658, 0.2694) to (0.3794, 0.2641). However, the inversely optimized cost vector changes from (0.6173, 0.3827) to (−0.1775, 0.8225).
We note that the notion of minimizing error or residuals in some optimality criterion as a means to determine parameters values “consistent” with observed data is not unique to inverse optimization. In fact, this is exactly the aim of regression models. Interestingly, regression also suffers from similar instability issues as inverse optimization \[38\] and regularization techniques has been shown to be effective in improving the stability of regression models \[40\]. The next section provides an overview on the regularization techniques used in regression and their connections with robust optimization.

### 5.2 Regularization In Regression

In practice both Least Absolute Deviations (LAD) and Least Squares (LS) regression models exhibit instability and the classical remedy to this issue is regularization. Among the regularization techniques the two much explored techniques are Lasso and Tikhonov regularization, which minimize the $\ell_1$ and $\ell_2$ norm of the predictor $\beta$, respectively.

In this section we first present a robust regression model. Next we present Tikhonov and Lasso regularized regression models and we describe how each of these models are
related to the robust regression model.

### 5.2.1 A Robust Regression Model

For given inputs $X$ and outputs $y$, the LAD and LS regression models return a predictor $\beta$ that minimizes the absolute deviations $\|y - X\beta\|_1$ and the squares of the deviations $\|y - X\beta\|_2$, respectively. However, in practice the observed data is subject to error. To account for errors and uncertainties in the observations, one might want to consider the following generalized robust regression problem that minimizes the $p$-norm of deviations:

$$\min_{\beta \in \mathbb{R}^n} \max_{U \in U} \|y - (X + U)\beta\|_p,$$

(5.4)

where $U$ is an uncertainty set around $X$.

Formulation (5.4) finds a robust predictor $\beta$ that better withstands errors and disruptions in the observations. In a paper by Xu et al. [62], the connections between the above robust regression model and the regularized regression models are studied. We will present their findings in the following.

### 5.2.2 Lasso Regularization

Lasso or $\ell_1$ regularization is well-known for its sparsity properties and has been extensively studied in the fields of signal processing, compressed sensing, and machine learning problems [63–65]. Lasso regularized regression minimizes the following objective:

$$\min_{\beta \in \mathbb{R}^n} \|y - X\beta\|_2 + \lambda\|\beta\|_1.$$

(5.5)

The generalized Lasso regularized regression for $\ell_p$ norm is defined as follows:

$$\min_{\beta \in \mathbb{R}^n} \|y - X\beta\|_p + \sum_{i=1}^n c_i |\beta_i|.$$

(5.6)
In [62], Xu et al. show that for $p \geq 1$, problem (5.6) is equivalent to formulation (5.4) for $U = \{(u_1, \ldots, u_n) \mid \|u_i\|_p \leq c_i, i = 1, \ldots, n\}$, where $u_i$’s are the columns of the uncertain matrix $U$.

### 5.2.3 Tikhonov Regularization

Tikhonov or $\ell_2$-regularized regression is one of the classical techniques for improving smoothness and stability of the predictor in regression. The Tikhonov regression minimizes the following objective:

$$\min_{\beta \in \mathbb{R}^n} \|y - X\beta\|_p + \lambda\|\beta\|_2.$$  \hspace{1cm} (5.7)

Similar to the Lasso regularized regression, in [62], authors prove that for $p = 2$ formulation (5.7) is equivalent to (5.4) if $U = \{U \mid \|U\|_F \leq \lambda\}$, where $\|U\|_F$ is the Frobenius norm of matrix $U$.

### 5.3 Regularized Inverse Optimization

Our aim is to exploit the similarities between regression and inverse optimization to address the instability issue in inverse optimization. Therefore, following a similar process as in regression, we first develop a robust inverse optimization model. Next we use the idea of Lasso and Tikhonov regularization from regression to solve the instability of the inverse linear optimization models. We also establish connections between our robust and regularized inverse optimization models.

#### 5.3.1 A Robust Inverse Optimization Model

Inspired by the robust regression problem (5.4), we develop a robust inverse optimization model that accounts for the uncertainties in the observed solutions. Suppose that
\(x_1, \ldots, x_q\) are given observed solutions for forward problem \(\min \{c'x \text{ s.t. } Ax \geq b\}\).

Consider the following robust inverse model, which is developed based on the GMIO\(_p\) model (4.7):

\[
\min_{y, c} \max_{U \in U} \| (X + U)'c - B'y \|_p \\
\text{subject to } A'y = c, \quad y'e = 1, \quad y \geq 0,
\]

where \(X = \begin{bmatrix} x_1 & \cdots & x_Q \end{bmatrix}_{n \times Q}\), \(B = \begin{bmatrix} b & \cdots & b \end{bmatrix}_{m \times Q}\), and \(U\) is a set that represents the uncertainty in the observed data points.

We used the similarities between regression and inverse optimization to develop a robust inverse optimization model. Our aim is to answer the question of whether there exist regularized inverse models that are connected with our robust inverse model, and if so, measure their effectiveness in solving the instability issue in inverse optimization.

### 5.3.2 \(\ell_1\)-Regularized Inverse Optimization

The Lasso or \(\ell_1\)-regularized inverse optimization problem aims at improving the stability of formulation (4.7) by adding an \(\ell_1\) regularizer term to the cost function. The \(\ell_1\)-Regularized Multi point Inverse Optimization (\(\ell_1\)-RMIO) problem is formulated as follows:
\[ \ell_1\text{-RMIO}(p, \lambda) : \quad \text{minimize} \quad \|\epsilon\|_p + \lambda \|c\|_1 \]

subject to
\[ A'y = c, \]
\[ y'\epsilon = 1, \]
\[ \epsilon_q = c'\hat{x}_q - b'y, \]
\[ y \geq 0, \]

where \( \lambda \) is the regularization parameter. The following theorem establishes a connection between formulation (5.9) and robust model (5.8).

**Theorem 5.3.1.** The \( \ell_1 \)-regularized inverse problem (5.9) is equivalent to the robust inverse problem (5.8), if

\[ \mathcal{U} = \left\{ U_{n \times Q} | U' = \begin{bmatrix} | & \ldots & | \\ u_1 & \cdots & u_n \end{bmatrix}, \|u_i\|_p \leq \lambda, \forall i \in \{1, \ldots, n\} \right\}, \]

\[ \text{max}_{U \in \mathcal{U}} \| (X + U)'c - B'y \|_p \leq \text{max}_{U \in \mathcal{U}} \|X'c - B'y\|_p + \|U'c\|_p \]

where the triangle inequality gives (a) and \( \|u_i\|_p \leq \lambda_i \) yields (b). For a fixed \( c \) by setting the following value for \( u_i \)'s the equality is attained:

\[ u_i \triangleq \begin{cases} \lambda_i \text{sgn}(c_i) \frac{X'c - B'y}{\|X'c - B'y\|_p} & \text{if} \quad X' \neq B'y \\ \lambda_i \text{sgn}(c_i)e & \text{o.w.} \end{cases} \]

and the proof is complete.
5.3.3 $\ell_2$-Regularized Inverse Optimization

The $\ell_2$-Regularized Multi Point Inverse Optimization ($\ell_2$-RMIO) model minimizes $\|\epsilon\|_p + \lambda\|c\|_2$ with respect to the constraints in problem (5.9). The following theorem sheds light on the connection between $\ell_2$-RMIO and the robust inverse optimization problem.

**Theorem 5.3.2.** For $p = 2$, the $\ell_2$-regularized inverse problem is equivalent to the robust inverse problem (5.8) for $p = 2$ with $\mathcal{U} = \{U_{n \times Q} \mid \|U\|_F \leq \lambda\}$.

**Proof.** We have:

$$\|\max_{U \in \mathcal{U}} (X + U)'c - B'y\|_2 \leq \max_{U \in \mathcal{U}} \|X'c - B'y\|_2 + \|U'c\|_2$$

$$\leq \|X'c - B'y\|_2 + \lambda\|c\|_2 = \|\epsilon\|_2 + \lambda\|c\|_2,$$

where $(a)$ follows from the triangle inequality and relation $(b)$ is as a result of $\|U\|_F \leq \lambda$.

Also for a given cost vector $c$, one can pick a $U$ with all its rows aligned with $c'X - y'B$, for which the above equalities are attained. Specifically, let $u_i = \frac{c_i}{\|c\|_2} \frac{X'c - B'y}{\|X'c - B'y\|_2}$ then we have:

$$\|U'c\|_2 = \lambda \left\| \sum_{i=1}^{n} c_i^2 \frac{\|X'c - B'y\|_2}{\|c\|_2} \right\| = \lambda\|c\|_2,$$

which shows that the equalities can be attained and the proof is complete. 

5.4 Singular Data Sets

In this section we first propose a formal definition of instability. Next we adapt the notion of singular data sets in [38, 39] for regression to inverse optimization. Then we show that singularity sets are closely related to instability and we characterize the singularity sets of (4.6) in the $p = 1$ case.

Hereinafter, a data set of $Q$ data points in $\mathbb{R}^n$ is denoted by a matrix $X_{n \times Q}$, whose columns represent the data points in the data set. Furthermore, the convergence of
Chapter 5. On the Stability of Inverse Optimization

Data sets is equivalent to convergence of their corresponding matrices under 1-norm (i.e., $X_k \to X$ if and only if $\|X_k - X\|_1 \to 0$).

**Definition 5.4.1.** A data set $X$ is unstable for GMIO$_p$ model if there exist two sequences of data sets $X^k_1$ and $X^k_2$, where $\lim_{k \to \infty} X^k_1 = \lim_{k \to \infty} X^k_2 = X$, and the sequence of optimal solutions to GMIO$_p$ for $X^k_1$ and $X^k_2$ converge to $s^*_1 = (c^*_1, y^*_1)$ and $s^*_2 = (c^*_2, y^*_2)$, where $s^*_1 \neq s^*_2$.

Intuitively, unstable data sets for inverse optimization are analogous to the discontinuity points for functions, where the right and left limits are not equal. Therefore, similar to the discontinuity points in functions, in the neighbourhood of the unstable data sets, the outcome of the inverse problem may abruptly change.

Although definition 5.4.1 illustrates the roles of unstable data sets in inverse optimization, it is not straightforward to use this definition for investigating the instability in the inverse optimization models. The following definition helps us to characterize and measure the instability in our inverse models.

**Definition 5.4.2.** A data set $X$ is singular for GMIO$_p$ if there exist at least two pairs of vectors $s_1 = (c_1, y_1)$ and $s_2 = (c_2, y_2)$, where $s_1$ and $s_2$ are both optimal solutions to GMIO$_p$ and also $s_1 \neq s_2$.

The next theorem characterizes the singular data sets for formulation (4.7) for $p = 1$.

**Theorem 5.4.1.** A data set $X$ is singular for formulation GMIO$_1$ if and only if $\arg \min_{t \in \{1, \ldots, m\}} \{ \xi_t \}$ is not unique, where $\xi_t$ is the $t^{th}$ component of $\xi = \sum_{q=1}^Q Ax_q - b$. 

**Proof.** If the data set $X$ is singular for GMIO$_1$, then the optimal solution of GMIO$_1$ is bounded and not unique for $X$. Therefore, there exist at least two extreme points of the feasible region of (4.7) that are optimal. Note that the extreme points of the feasible region are of form $y_i = 1$ and $y_j = 0$ for all $j \neq i$. Since there are at least two optimal
Chapter 5. On the Stability of Inverse Optimization

solutions that are attained at the extreme points, there exist \( l \) and \( s \) such that:

\[
\sum_{q=1}^{Q} a_t x_q - b_t = \sum_{q=1}^{Q} a_s x_q - b_s \leq \sum_{q=1}^{Q} a_j x_q - b_j \quad \forall j \in \{1, \ldots, m\}.
\]

Therefore, \( t \) and \( s \) are both optimal solutions to \( \arg \min_{t \in \{1, \ldots, m\}} \bar{\xi}_t \).

Conversely, if \( \arg \min_{t \in \{1, \ldots, m\}} \bar{\xi}_t \) is not unique, there exist \( s \) and \( l \) such that:

\[
\sum_{q=1}^{Q} a_t x_q - b_t = \sum_{q=1}^{Q} a_s x_q - b_s \leq \sum_{q=1}^{Q} a_j x_q - b_j \quad \forall j \in \{1, \ldots, m\}. \tag{5.10}
\]

Also note that there exists an optimal solution to (4.7) that is attained at an extreme point of form \( y_i = 1 \) and \( y_j = 0 \) for all \( j \neq i \). Thus, we have:

\[
\sum_{q=1}^{Q} a_i x_q - b_i \leq \sum_{q=1}^{Q} a_j x_q - b_j \quad \forall j \in \{1, \ldots, m\}. \tag{5.11}
\]

Without loss of generality assume that \( a_s \neq a_i \) note that:

\[
\sum_{q=1}^{Q} a_i x_q - b_i = \sum_{q=1}^{Q} a_s x_q - b_s \leq \sum_{q=1}^{Q} a_j x_q - b_j \quad \forall j \in \{1, \ldots, m\}.
\]

Since \( a_i \) is optimal and the latter equation shows that its cost is equal to the cost for \( a_s \), \( \text{GMIO}_1 \) has multiple optimal solutions and the proof is complete.

The following theorem illustrates the connection between the singularity sets and the instability of \( \text{GMIO}_1 \).

**Theorem 5.4.2.** A data set \( X \) is unstable for \( \text{GMIO}_1 \) if and only if it is also singular.

**Lemma 5.4.3.** Let \( a'_i x \geq b_i \) be a non-redundant constraint for the full-dimensional polyhedron induced by \( Ax \geq b \), there exists \( \bar{x} \) such that \( A\bar{x} \geq b \) and

\[
a_i \bar{x} = b_i, \quad \text{and} \quad a_j \bar{x} > b_j, \quad \forall j \neq i.
\]
Proof. Since \( a'_ix \geq b_i \) is a non-redundant constraint there exists \( x^i \) such that:

\[
a'_ix^i < b_i \quad \text{and} \quad a'_jx^i \geq b_j, \quad \forall j \neq i. \tag{5.12}
\]

Also since the feasible polyhedron is full dimensional there exists \( \tilde{x} \) such that:

\[
a'_j\tilde{x} > b_j, \quad \forall j \in \{1, \ldots, m\}. \tag{5.13}
\]

Let \( f(\lambda) = a'_i(\lambda x^i + (1 - \lambda)\tilde{x}) - b_i \), note that \( \lambda x^i + (1 - \lambda)\tilde{x} \) is a feasible point for \( 0 \leq \lambda \leq 1 \) and we have:

\[
(5.12) \Rightarrow f(1) < 0, \quad \text{and} \quad (5.13) \Rightarrow f(0) > 0.
\]

Therefore, using the intermediate value theorem there exists \( 0 < \lambda^* < 1 \), such that \( f(\lambda^*) = 0 \). Let \( \bar{x} = \lambda^*x^i + (1 - \lambda^*)\tilde{x} \), from (5.12), and (5.13) we have:

\[
a'_j(\lambda^*x^i + (1 - \lambda^*)\tilde{x}) - b_j > 0, \quad \forall j \neq i,
\]

\[
f(\lambda^*) = 0 \quad \Rightarrow \quad a'_i(\lambda^*x^i + (1 - \lambda^*)\tilde{x}) - b_i = 0,
\]

which concludes the proof. \( \blacksquare \)

Proof of Theorem 5.4.2. Suppose by way of contradiction that \( \mathbf{X} \) is unstable for GMIO\(_1\) and it is not singular. Therefore, there exist two sequences of data sets \( \mathbf{X}^k_1 \) and \( \mathbf{X}^k_2 \), where \( \lim_{k \to \infty} \mathbf{X}^k_1 = \lim_{k \to \infty} \mathbf{X}^k_2 = \mathbf{X} \), and the sequence of optimal solutions \( \mathbf{s}^{*k}_1 = (\mathbf{c}^{*k}_1, \mathbf{y}^{*k}_1) \) and \( \mathbf{s}^{*k}_2 = (\mathbf{c}^{*k}_2, \mathbf{y}^{*k}_2) \) of GMIO\(_1\) for \( \mathbf{X}^k_1 \) and \( \mathbf{X}^k_2 \) converge to \( \mathbf{s}^*_1 = (\mathbf{c}^*_1, \mathbf{y}^*_1) \) and \( \mathbf{s}^*_2 = (\mathbf{c}^*_2, \mathbf{y}^*_2) \), where \( \mathbf{s}^*_1 \neq \mathbf{s}^*_2 \). Let \( \xi^*_1 \) and \( \xi^*_2 \) be the complementary slackness vectors corresponding to the centroid of data sets \( \mathbf{X}^*_1 \) and \( \mathbf{X}^*_2 \). Since the data sets converge to \( \mathbf{X} \) and complementary slackness residuals of the centroids are linear functions of the data sets, the complementary slackness residuals also converge to their counterpart \( \xi \) for \( \mathbf{X} \).
Since $X$ is not singular, the optimal solution to $\text{GMIO}_1$ is unique for $X$ then either $s^*_1$ or $s^*_2$ is not an optimal solution for the data set $X$. Without loss of generality assume that $s^*_1$ is not an optimal solution of $\text{GMIO}_1$ for $X$. Therefore, there exists an $\epsilon > 0$ and an integer $N_1 > 0$ such that:

$$y^{*k'}_1 \xi \geq y^{*\ell'}_1 \xi + \epsilon, \quad \forall k > N_1,$$

(5.14)

also since $\bar{\xi}_1^k \rightarrow \bar{\xi}$ there exists an integer $N_2 > N_1$ such that:

$$\|\bar{\xi}_1^{k} - \bar{\xi}\|_2 \leq \frac{\epsilon}{2}, \quad \forall k > N_2,$$

(5.15)

Note that $\|y^{*k'}_1\|_2 \leq \|y^{*k'}_1\|_1 = 1$ and $y^{*k'}_1$ is an optimal solution for $X^k_1$. Thus,

$$y^{*k'}_1 \xi \geq y^{*k'}_1 \xi \Rightarrow y^{*k'}_1 (\xi - \bar{\xi} + \bar{\xi}) \geq y^{*k'}_1 \xi$$

$$\Rightarrow \|y^{*k'}_1\|_2 \|\bar{\xi} - \bar{\xi}\|_2 + y^{*k'}_1 \xi \geq y^{*k'}_1 \xi$$

(5.16)

$$\Rightarrow \frac{\epsilon}{2} + y^{*k'}_1 \xi \geq y^{*k'}_1 \xi \Rightarrow \epsilon + y^{*k'}_1 \xi > y^{*k'}_1 \xi \quad \forall k > N_2.$$

However, the latter result is in contradiction with (5.14). Hence, $X$ is singular and the proof is complete.

Conversely, assume that $X$ is singular, we show that it is also unstable. Let $\bar{\xi}$ be the complementary slackness vector corresponding to $X$. In Theorem 5.4.1 we proved that arg min $\{\bar{\xi}_t\}$ is not unique. Therefore, $i, j \in \{1, \ldots, m\}$ exist such that:

$$\bar{\xi}_i = \bar{\xi}_j \leq \bar{\xi}_k, \quad \forall k \in \{1, \ldots, m\}.$$

As a consequence of Lemma 5.4.3, there exist data point $x^i$ and $x^j$ such that:

$$a_i x^i = b_i, \quad \text{and} \quad a_k x^i > b_k, \quad \forall k \neq i,$$

$$a_j x^j = b_j, \quad \text{and} \quad a_k x^j > b_k, \quad \forall k \neq j.$$
Let \( Y_{(n \times Q)} \) and \( Z_{(n \times Q)} \) denote two data sets produced by \( Q \) copies of \( x^i \) and \( x^j \), respectively. Also Note that \( a_i \) and \( a_j \) are unique optimal solutions of \( \text{GMIO}_1 \) for \( Y \) and \( Z \), respectively. Define two sequences of data sets \( X_1^k \) and \( X_2^k \) as follows:

\[
X_1^k = \frac{k - 1}{k} \mathbf{X} + \frac{1}{k} Y, \quad k \in \mathbb{N},
\]
\[
X_2^k = \frac{k - 1}{k} \mathbf{X} + \frac{1}{k} Z, \quad k \in \mathbb{N}.
\]

Note that the sequence of \( \text{GMIO}_1 \) optimal solutions for \( X_1^k \) and \( X_2^k \) are unique and different from each other (corresponding to \( a_i \) and \( a_j \)), and both of the sequences converge to \( \mathbf{X} \) as \( k \to \infty \). Therefore, \( \mathbf{X} \) is unstable and proof is complete.

Therefore, non-singularity of a data set guarantees its stability, and vice versa. Consequently, if a data set is not close to a singular set it will exhibit more stability.

From Theorem 5.4.1, we know that the locus of centroid of singular data sets is a subset of all hyperplanes that have equal distance from at least two different constraints. Similar to corollary 4.2.1.1, the set of singular data sets is measure zero in \( \mathbb{R}^{n \times Q} \). Note that although the set of all singular sets is small, the data sets that are close to singular sets may also exhibit instability. In the next section we introduce a measure for inverse problems to quantify the extent to which a data set exhibits instability.

5.5 Measuring the Stability of Data Sets

In this section we provide a metric for measuring the stability of \( \text{RMIO} \) and \( \text{GMIO} \) problems. We also show that our metric is closely related to the singular data sets. This metric is then used to compare the performance of the regularized inverse models in Section 5.7.

Let \( \mathbf{A} \mathbf{x} \geq \mathbf{b} \) be the set of constraints for the forward problem. Let \( \mathbf{c}^*(\mathbf{X}, \mathbf{A}, \mathbf{b}, \lambda) \) denote an inversely optimized cost vector for a dataset \( \mathbf{X} \) produced by the regularized inverse problem with regularization parameter \( \lambda \). Also assume that the optimal cost
vectors are normalized such that \( \|c^*(X, A, b, \lambda)\|_1 = 1 \). For \( 0 \leq \delta \leq 2 \), the \( \delta \)-stability of the regularized inverse model with parameter \( \lambda \) for a given data set \( \hat{X} \) is measured as follows:

\[
\rho(\hat{X}, A, b, \delta, \lambda) = \min_X \|X - \hat{X}\|_1 \quad \text{subject to} \quad \|c^*(X, A, b, \lambda) - c^*(\hat{X}, A, b, \lambda)\|_1 \geq \delta, \tag{5.17}
\]

where \( \hat{X} \) and \( X \) are matrices whose columns represent the original and perturbed data sets, respectively. The optimal objective value in (5.17) indicates the smallest perturbation in the original data set \( \hat{X} \) that can alter the imputed cost vector for at least \( \delta \).

Theorem 5.5.1 clarifies the connection between the singular data sets and the stability measure defined above for \( \text{GMIO}_1 \).

**Theorem 5.5.1.** The optimal solution \( X^* \) to (5.17) is a singular data set \( \rho(\hat{X}, A, b, \delta, 0) \) is always equal to \( \ell_1 \) distance of \( \hat{X} \) from a singular data set.

**Proof.** In case where the optimal data set \( X^* \) in formulation (5.17) is singular the problem is solved. By way of contradiction assume that the optimal data set \( X^* \) is not singular. Therefore, the optimal solution of \( \text{GMIO}_1 \) for \( X^* \) is unique and is attained in an extreme point of the inverse polyhedron (\( y \)-polyhedron), which is corresponding to one of the constraints in the forward problem, say \( a_i \) for some \( i \in \{1, \ldots, m\} \). Thus, there exist \( \delta > 0 \) such that:

\[
\hat{\xi}_i + \delta < \xi_j \quad \forall j \in \{1, \ldots, m\},
\]

where \( \hat{\xi} = \frac{1}{Q} \sum_{q=1}^{Q} (A\hat{x}_q - b) \) and \( \xi = \frac{1}{Q} \sum_{q=1}^{Q} (Ax_q - b) \) are the complementary slackness vectors corresponding to the centroids of \( \hat{X} \) and \( X^* \). Let \( \bar{X} \) be a data set defined as follows:

\[
\bar{X} = \eta X^* + (1 - \eta)\hat{X},
\]
where \( \eta \) is chosen such that \( 0 \leq \eta \leq 1 \) and \( \eta \geq \max_{j \in \{1, \ldots, m\}} \frac{\xi_j}{\xi_j + \delta} \). Also let \( \tilde{\xi} \) be the complementary slackness vector corresponding to \( \hat{X} \). Note that \( a_i \) is an optimal solution of \( \text{GMIO}_1 \) for \( \hat{X} \), since it \( \text{GMIO}_1 \) has at least one optimal solution which is attained at an extreme point and we also have:

\[
\tilde{\xi}_i = \eta \xi_i + (1 - \eta) \xi_i \leq \eta \xi_j - \eta \delta + (1 - \eta) \xi_i \leq \eta \xi_j - \frac{\xi_j}{\xi_j + \delta} \delta + (1 - \frac{\xi_j}{\xi_j + \delta}) \xi_i = \eta \xi_j \leq \eta \xi_j - \eta \delta + (1 - \xi_j) \xi_i = \eta \xi_j + (1 - \eta) \xi_j \Rightarrow \tilde{\xi}_i \leq \tilde{\xi}_j \quad \forall j \in \{1, \ldots, m\}.
\]

The latter relation shows that the cost value of the extreme point corresponding to the \( i^{th} \) constraint is smaller or equal than any other extreme point.

But for \( \tilde{X} \) we also have:

\[
\| \tilde{X} - \hat{X} \|_1 = \| \eta X^* + (1 - \eta) \tilde{X} - \hat{X} \|_1 = \eta \| \tilde{X} - \hat{X} \|_1 \leq \| X^* - \hat{X} \|_1.
\]

Therefore, \( \tilde{X} \) satisfy the constraint for (5.17) and also yields a lower cost than \( X^* \) which is in contradiction with \( X^* \) being optimal and the proof is complete.

The next theorem shows that for \( \text{GMIO}_1 \) in the extreme case where a data set \( \hat{X} \) is singular, the stability measure \( \rho(\hat{X}, A, b, \delta, \lambda) \) is zero (its minimum) for a sufficiently small \( \delta \).

**Theorem 5.5.2.** Suppose that \( \hat{X} \) is a singular set for \( \text{GMIO}_1 \) and let \( \delta = \min_{i \neq j} \| a_i - a_j \|_1 \), then \( \rho(\hat{X}, A, b, \delta, 0) = 0 \).

**Proof.** In case where \( \hat{X} \) is singular theorem 5.4.1 states that there exist at least two optimal cost vectors corresponding to two constraints, say, \( a_s \) and \( a_t \). Since \( \delta \leq \| a_s - a_t \|_1 \) and they are both optimal for \( \text{GMIO}_1 \), \( X = \hat{X} \) is a feasible solution for (5.17) and the optimal value of cost is zero.

The stability measure in (5.17) is defined for a fixed data set, one can define a general
stability measure for a regularized model with parameter $\lambda$ as follows:

$$\bar{\rho}(A, b, \delta, \lambda) = \mathbb{E}_{\hat{x}} [\rho(\hat{x}, A, b, \delta, \lambda)],$$

where the expectation is taken uniformly over all feasible single data points $\hat{x}$.

### 5.6 Price of Stability

The GMIO$_1$ problem yields a cost vector that attains the minimum sum of duality gaps. However, RMIO also minimizes a regularizer as a side objective and may not return a cost vector that minimizes the sum of duality gaps. Considering the sum of duality gap to be a measure of efficiency, there is a trade-off between the stability and the efficiency of the optimal solutions of an inverse model. The price of stability quantifies the bias that results from regularization or any other method used for stabilization. Let $v^*(X, A, b)$ be the optimal value of GMIO$_1$ for data set $X$. Therefore, $v^*(X, A, b)$ represents the minimum of the sum of duality gaps for data set $X$. let $v^*(X, A, b, \lambda, \ell_i)$ denote the sum of duality gaps for the solution produced by using $\ell_i$-regularization with parameter $\lambda$ for the same data set. The price of stability is defined as follows:

$$\Pi(X, A, b, \lambda, \ell_i) = \frac{v^*(X, A, b, \lambda, \ell_i)}{v^*(X, A, b)}.$$  \hspace{1cm} (5.18)

To derive a metric for a problem which is independent of the data set, one can take the expectation of the metric defined in (5.18) over single data points in the feasible region as follows:

$$\bar{\Pi}(A, b, \lambda) = \mathbb{E}_{\hat{x}} \frac{v^*(\hat{x}, A, b, \lambda, \ell_i)}{v^*(\hat{x}, A, b)},$$  \hspace{1cm} (5.19)

where $\hat{x}$ is assumed to have a uniform distribution over the feasible polyhedron $A\hat{x} \geq b$. 
5.7 Computational Results

In this section we discuss the performance of regularized inverse models for the problem introduced in (5.1) and also for the IMRT data. We illustrate that in low dimensional problems $\ell_2$-regularized (Tikhonov) inverse models outperforms $\ell_1$-regularized (Lasso) inverse models. However, the $\ell_1$-regularized method surpasses the $\ell_2$-regularized inverse models for the IMRT data.

5.7.1 Generated Data

Monte Carlo simulations were used to evaluate the measures defined in (5.17) and (5.19) for problem (5.9). To find $\bar{\rho}$ and $\bar{\Pi}$, a hit-and-run algorithm [66] was used to uniformly sample from the points inside the feasible polyhedron.

Figure 5.3 depicts the $0.5$-stability of the non-regularized, $\ell_1$-regularized , and $\ell_2$-regularized inverse problems with $\lambda = 0.2$ associated with (5.1). Figure 5.3 shows that the $\ell_1$-regularized inverse models do not yield satisfactory results for low dimensional problems as the unstable region for this model has expanded in comparison to the non-regularized model. In fact, the intrinsic sparsity property of $\ell_1$-regularization degrades its performance in low dimensional problems. This is due to the fact that in low dimensional problems there are lower degrees of freedom and “sparsifying” the solution by employing the $\ell_1$-regularizer may indeed degrade the performance of our models. However, as we will see in the next section, in higher dimensional problems the sparsity property improves the performance of the inverse models. In contrast, $\ell_2$-regularized inverse models shrink the unstable areas and yield a better performance for lower dimensional problems.
Figure 5.3: The 0.5-stability heat map for (a) non-regularized, (b) $\ell_1$-regularized, and (c) $\ell_2$-regularized inverse problems corresponding to example (5.1). The red areas indicate more instability, i.e., a small perturbation of the original data may change the normalized inversely optimized cost vector by at least 0.5. Note that the dark red areas in (a) are exactly aligned with the locus of centroid of singular data sets for GMIO$_1$, which are a subset of all the lines that have an equal distance from at least two different constraints.

Figures 5.4 and 5.5 show the trade-off between stability and the efficiency of the solution. As it is illustrated in Figure 5.4 the $\ell_1$-regularized model exhibits more instability than the non-regularized model ($\lambda = 0$) and also returns inefficient solutions (high price of stability). However, Figure 5.5 shows that the $\ell_2$-regularized model exhibits more stability (the dark areas are shrunk) and the price of stability is also lower than that of $\ell_1$-regularized model.
5.7.2 IMRT Treatment Planning

In this section we study the impact of regularization on inversely optimized weights for IMRT treatment planning. We first introduce the Overlap Volume Histogram (OVH) which is a feature for predicting inversely optimized weights [16]. Then we present a logistic regression model developed in [15] for predicting inverse weights in IMRT. We show
that regularized inverse models can better manifest the connections between anatomical features and inversely optimized weights. Consequently, the regularized weights yield a better fit to the models employed for predicting the weights from anatomical features and thus provide better prediction performance.

5.7.2.1 Overlap Volume Histogram

The Overlap Volume Histogram (OVH) has been widely used as a measure that captures spatial information of a patient’s anatomy. OVH has been used to examine the properties of delivered plans and automating the IMRT treatment planning process [14, 15, 67–69]. The OVH captures the spatial correlation between PTV and an OAR. Particularly, $OVH_{Blad}^d$ is defined as follows:

$$OVH_{Blad}^d = \frac{|\{v \mid dist(v, PTV) \leq d, v \in Blad\}|}{|\{v \mid v \in Blad\}|},$$  \hspace{1cm} (5.20)

where $Blad$ denotes the set of voxels in the bladder, $dist(x, S) = \min_{s \in S} \|x - s\|_2$, and $|$ denotes the cardinality of a set. Figure 5.6 illustrates $OVH_{Blad}^{0.5}$ and $OVH_{Blad}^1$ as defined in (5.20).
In this chapter we will use the following metric:

\[
OV_{Blad/Rect}^d = \log \left( \frac{OVH_{Blad}^d}{OVH_{Rect}^d} \right).
\]  

(5.21)

Intuitively, \(OV_{Blad/Rect}^d\) represents the relative distance of bladder and rectum for the tumour’s point of view. For the cases where the bladder is relatively closer to the PTV the value of \(OV_{Blad/Rect}^d\) is positive, however, in the cases where the rectum is located closer to the PTV the value of \(OV_{Blad/Rect}^d\) is negative.

5.7.2.2 An Inverse Paradigm for IMRT Treatment Planning

The key assumption in the inverse-optimization-based treatment planning is that historical clinical plans have been produced by solving a multi-objective problem similar
Chapter 5. On the Stability of Inverse Optimization

66

to (3.1), which can be written in the following compact format:

$$\text{MFO}(\alpha) : \minimize_x \alpha' C x$$

subject to $A x \geq b,$

$$\text{(5.22)}$$

where $x$ is a vector that consists of the intensity of beamlets $\omega$ and a set of auxiliary variables added for linearizing the maximums in the objective function of (3.1).

The aim of inverse optimization is to impute the objective weights in (3.1) so that the optimal solution best represents a clinical plan. The Regularized Inverse Optimization (RIO) problem determines a set of weights $\alpha$ that best fits $\hat{x}$, can be written as follows:

$$\text{RIO}(\hat{x}) : \minimize_{\alpha, y} (\alpha' C \hat{x} - b' y)^p + \lambda \|\alpha' C\|_p^p$$

subject to $A' y = \alpha' C,$

$$y' e = 1,$$

$$y \geq 0,$$

$$\text{(5.23)}$$

where the first and the last constraints ensure dual feasibility and the second constraint is a normalization constraint for avoiding trivial solutions. Also the first term in the objective function intends to minimize the $p^{th}$ power of the duality gap and the second term is $\ell_p$-regularizer, where $p$ and $\lambda$ are the regularization parameter (e.g., $p = 1$ yields Lasso regularizations).

The imputed $\alpha$ for the patients along with the $OV_{Blad/Rect}$ features are then fed into a logistic regression model. To simplify the logistic model we used only the $OV_{\theta Blad/Rect}$ feature for a fixed expansion level $\theta$. Particularly, the logistic regression model solves the following log-likelihood maximization problem for fitting a logistic function to the inversely optimized weights for the bladder:
LR: \[
\text{maximize } \sum_{q=1}^{Q} \left(-\log(1 + \exp^{\beta_0 + \beta_1 OV_q + \beta_2 (OV_q)^2 + \beta_3 (OV_q)^{0.5}}) + \right)
\]
\[
\sum_{q=1}^{Q} \alpha_q^{Blad} (\beta_0 + \beta_1 OV_q + \beta_2 (OV_q)^2 + \beta_3 (OV_q)^{0.5}),
\]

(5.24)

where \(OV_q\) is equal to \(OV_{Blad/Rect}^\theta\) for the \(q^{th}\) patient at \(\theta\) expansion.

Consequently, for a de novo patient the \(OV_{Blad/Rect}^\theta\) is calculated and the inverse weights are predicted by using the trained logistic regression function. The estimated inverse weights are then plugged into the forward problem (3.1). Next, the forward problem is solved to produce the treatment plan. Figure 5.7 illustrates the inverse-optimization-based treatment planning process.

Figure 5.7: IMRT treatment planning process using inverse optimization.

5.7.2.3 Performance of the Inversely Optimized Plans

The treatment data for 315 patients were used to perform the simulations presented in this section. The inverse problem (5.23) was solved with fixed regularization parameter \(\lambda = 1, p = 1\) (Lasso) and \(p = 2\) (Tikhonov) to produce inversely optimized weights
for a five objectives: three relaxed max dose-volume penalty functions at 50 Gy for the rectum, bladder, and PTV ring and two max dose penalty functions for the right and left femoral heads. The weights produced for the first three patients are presented in Tables 5.1 to 5.3. Note that in contrast to \( \ell_2 \)-regularized weights, the \( \ell_1 \)-regularized weights are considerably different from the non-regularized weights, which highlights the effect of sparsity induced by the \( \ell_1 \)-regularizer in high dimensional applications inverse problems like IMRT.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Non-regularized (( \lambda = 0 ))</th>
<th>Lasso with ( \lambda = 1 )</th>
<th>Tikhonov with ( \lambda = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>0.4459</td>
<td>0.2167</td>
<td>0.4517</td>
</tr>
<tr>
<td>Rectum</td>
<td>0.4708</td>
<td>0.6991</td>
<td>0.4534</td>
</tr>
<tr>
<td>L.Fem</td>
<td>0.0042</td>
<td>0.0053</td>
<td>0.0034</td>
</tr>
<tr>
<td>R.Fem</td>
<td>0.0028</td>
<td>0.0022</td>
<td>0.0025</td>
</tr>
<tr>
<td>PTV ring</td>
<td>0.0763</td>
<td>0.0767</td>
<td>0.0890</td>
</tr>
</tbody>
</table>

Table 5.1: Produced weights for patient #1 using different regularization parameters.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Non-regularized (( \lambda = 0 ))</th>
<th>Lasso with ( \lambda = 1 )</th>
<th>Tikhonov with ( \lambda = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>0.4251</td>
<td>0.1498</td>
<td>0.4346</td>
</tr>
<tr>
<td>Rectum</td>
<td>0.4919</td>
<td>0.7650</td>
<td>0.4672</td>
</tr>
<tr>
<td>L.Fem</td>
<td>0.0015</td>
<td>0.0019</td>
<td>0.0013</td>
</tr>
<tr>
<td>R.Fem</td>
<td>0.0014</td>
<td>0.0021</td>
<td>0.0013</td>
</tr>
<tr>
<td>PTV ring</td>
<td>0.0801</td>
<td>0.0812</td>
<td>0.0956</td>
</tr>
</tbody>
</table>

Table 5.2: Produced weights for patient #2 using different regularization parameters.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Non-regularized (( \lambda = 0 ))</th>
<th>Lasso with ( \lambda = 1 )</th>
<th>Tikhonov with ( \lambda = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>0.9480</td>
<td>0.9372</td>
<td>0.9418</td>
</tr>
<tr>
<td>Rectum</td>
<td>0.0087</td>
<td>0.089</td>
<td>0.0136</td>
</tr>
<tr>
<td>L.Fem</td>
<td>0.0077</td>
<td>0.0075</td>
<td>0.0069</td>
</tr>
<tr>
<td>R.Fem</td>
<td>0.0054</td>
<td>0.0061</td>
<td>0.0054</td>
</tr>
<tr>
<td>PTV ring</td>
<td>0.0302</td>
<td>0.0402</td>
<td>0.0323</td>
</tr>
</tbody>
</table>

Table 5.3: Produced weights for patient #3 using different regularization parameters.

The DVHs for non-regularized, \( \ell_1 \)-regularized, and \( \ell_2 \)-regularized inverse plans are compared with the clinical plan DVH for patient #1 in Figures 5.8-5.10. Note that Lasso regularized model has better dose sparing for OARs. In particular, it is easy to see the differences between left femoral head DVHs for different methods.
Figure 5.8: Comparison of DVHs of the non-regularized inverse plan and the clinical plan for patient #1.
Figure 5.9: Comparison of DVHs of the $\ell_1$-regularized inverse plan with $\lambda = 1$ and the clinical plan for patient #1.
Figure 5.10: Comparison of DVHs of the $\ell_2$-regularized inverse plan with $\lambda = 1$ and the clinical plan for patient #1.
Table 5.4 compares the clinical dose criteria of inversely optimized plans with the historical clinical plan for patient #1. The PTV $V_{74.1} \geq 99\%$ (min dose-volume) criterion is not met by any plan including the clinical plan, however, the violation is negligible for all the plans. Also the rectum $V_{70.0} \leq 30$ (max dose-volume) is not met by any plan but $\ell_1$-regularized inverse plan has the least violation among the inversely optimized plans. Finally, only the $\ell_2$-regularized inverse plan does not satisfy the left femoral head max dose-volume constraint. Overall the $\ell_1$-regularized inverse plan outperforms the other inversely optimized plans for patient #1. The performance of the three methods for all 315 patients are compared in Figure 5.11.

<table>
<thead>
<tr>
<th>OAR/Target</th>
<th>Clinical target (%)</th>
<th>Clinical (%)</th>
<th>Non-regularized (%)</th>
<th>Tikhonov (%)</th>
<th>Lasso (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTV</td>
<td>$V_{78.0} \geq 99$</td>
<td>100</td>
<td>99.66</td>
<td>99.64</td>
<td>99.62</td>
</tr>
<tr>
<td>PTV</td>
<td>$V_{74.1} \geq 99$</td>
<td>$98.74$</td>
<td>$98.89$</td>
<td>$98.94$</td>
<td>$98.90$</td>
</tr>
<tr>
<td>PTV</td>
<td>$V_{81.9} \leq 100$</td>
<td>100</td>
<td>99.68</td>
<td>99.67</td>
<td>99.65</td>
</tr>
<tr>
<td>Bladder</td>
<td>$V_{70.0} \leq 30$</td>
<td>25.77</td>
<td>26.30</td>
<td>26.23</td>
<td>26.80</td>
</tr>
<tr>
<td>Bladder</td>
<td>$V_{54.3} \leq 50$</td>
<td>35.73</td>
<td>33.37</td>
<td>33.38</td>
<td>34.32</td>
</tr>
<tr>
<td>Rectum</td>
<td>$V_{70.0} \leq 30$</td>
<td>30.13</td>
<td>31.23</td>
<td>31.28</td>
<td>30.62</td>
</tr>
<tr>
<td>Rectum</td>
<td>$V_{54.3} \leq 50$</td>
<td>44.40</td>
<td>44.97</td>
<td>45.04</td>
<td>44.30</td>
</tr>
<tr>
<td>L.Fem</td>
<td>$V_{52.0} \leq 5$</td>
<td>2.36</td>
<td>0.35</td>
<td>6.19</td>
<td>0</td>
</tr>
<tr>
<td>R.Fem</td>
<td>$V_{52.0} \leq 5$</td>
<td>1.86</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 5.4: Clinical dose criteria for the clinical plan and inversely optimized plans for patient #1. Cell gray levels correspond to violation in the criteria, i.e., darker colour corresponds to more violation.
Figure 5.11: Comparison of median error between the clinical plan and non-regularized, Lasso ($\lambda = 1$), and Tikhonov ($\lambda = 1$) inversely optimized plans for the clinical acceptability criteria.

Figure 5.11 shows that the $\ell_1$-regularized inversely optimized weights represents the clinical plan better than the $\ell_2$-regularized and non-regularized inversely optimized weights. Another desired feature of inversely optimized weights in IMRT is “predictability”, which we investigate in the next section.

5.7.2.4 Predictability of Inversely Optimized Weights

As illustrated in Figure 5.7 for a de novo patient the geometrical features are used to estimate the inversely optimized weights. Consequently, it is also important for the inversely optimized weights to be estimated accurately from the patient geometrical features. In this section we investigate how well inversely optimized weights are represented
using $OV_{Blad/Rect}$ features. Figure 5.12 depicts the relation between inversely optimized weights and $OV_{Blad/Rect}^{0.5}$ for a data set of 315 patients.

![Graphs showing the relation between inversely optimized weights and $OV_{Blad/Rect}^{0.5}$ for different regularization methods.](image)

Figure 5.12: Comparison of the mismatch between the fitted logistic function and bladder weight produced by non-regularized, $\ell_1$-regularized, $\ell_2$-regularized inverse models for $\lambda = 1$.

The parameters of the logistic functions depicted in Figure 5.12 are estimated by solving problem (5.24). The corresponding log-likelihood values for different $OV_{Blad/Rect}$ expansion levels are presented in Table 5.5. Note that $\ell_1$-regularized inverse weights yield a higher likelihood value (i.e., a better fit) to the logistic function than $\ell_2$-regularized and non-regularized weights for all expansion levels. Therefore, the predicted weight in the IMRT treatment planning is less likely to have large errors for the $\ell_1$-regularized weights.
5.8 Conclusions

An inverse optimization problem is unstable if small perturbations in the input (data points) can dramatically alter the output (imputed cost vector). Instability in inverse optimization problems devalues inversely optimized parameters, since small noise levels in the observed data can drastically affect the imputed parameters and therefore yields a different perception of the system. In this chapter we illustrated the instability of the inverse linear optimization models and related it to the similar issues that are also present in regression.

We used the analogy between inverse optimization and regression and employed similar remedies as in regression to stabilize inverse models. Particularly, we used Lasso and Tikhonov regularization techniques and established connections between a robust inverse model and our regularized inverse models.

We also developed metrics for measuring the stability and efficiency of the proposed regularized methods. Our simulations show that Tikhonov outperforms Lasso regularization method for low dimensional inverse problems. However, in our high dimensional IMRT treatment planning problem the Lasso regularized models yield higher quality treatment plans. Furthermore, Lasso regularized weights yield a higher likelihood (lower estimation error) and fit better to our logistic regression models. Overall, we showed the effectiveness of the proposed regularized inverse models and examined the performance
of each technique in high and low dimensional scenarios.
Chapter 6

Conclusions

The classical inverse linear optimization framework cannot accommodate infeasible data points or data points that lie in the relative interior of the feasible polyhedron. Particularly, for a given data point the classical methods return a cost vector that makes the observed data point optimal and consequently if the observed data point is not optimal for any cost vector the inverse problem becomes infeasible. The modern data-driven approaches fix these issues by relaxing some of the hard constraints in the classical models, however, if the classical models are merely relaxed the models may return trivial solutions. Therefore, it is crucial to filter out trivial solutions by adding appropriate normalization constraints to the models. As our first contribution in this dissertation, we proposed a novel normalization technique and proved that our normalization technique forces the problem to generate non-trivial solutions.

In modern data-driven approaches usually multiple observations are used to estimate the cost vector. We analyzed the outcome of a class of inverse models with certain penalty function structures. We investigated the behaviour of multi point inverse models for $p^{th}$ power mean penalty functions and demonstrated that the results may dramatically change by altering $p$. Our results illustrate that the parameters of inverse models may drastically affect the outcome of the inverse problem and therefore it is essential to tune
these parameters for each application.

Aside from the model parameters we also studied the effect of perturbations in the input (observed data) on the output (imputed cost vector) in inverse optimization. We first exemplified the instability of inverse models by considering examples where inverse optimization problem exhibited instability. Then we proposed a formal definition for unstable data sets and a metric to measure instability of data sets. We used the connections between inverse optimization and regression to fix the instability issues by employing Lasso and Tikhonov regularization techniques. We performed simulations on a randomly generated problem and also on IMRT data and our results show that regularization methods can improve the stability of the studied inverse models. Furthermore, our tests verify that $\ell_1$-regularized (Lasso) models outperform $\ell_2$-regularized (Tikhonov) models in our high dimensional IMRT data due to its intrinsic sparsity features. However, in our low-dimensional randomly generated example the $\ell_2$-regularized model is more effective than the $\ell_1$-regularized model.

In IMRT treatment planning process, aside from the quality of plans produced by inversely optimized weights, it is important to be able to efficiently predict inverse weights using geometrical features. Our simulations show that not only the $\ell_1$-regularized weights slightly improve the treatment quality they also fit better to our logistic model. Therefore, the prediction error for $\ell_1$-regularized weights for a de novo patient has less error. We believe that the predictability of inverse weights becomes even more important for more complex radiation treatment planning problems such as head and neck, where the cancer site has a more complicated structure and predicting inverse weights from geometrical features becomes more challenging.

Our tests show that for each patient there exists a range of weights that can produce acceptable treatment plans. Also the range of acceptable weights may differ for different patients. For instance, for some patients a large range of weights produce acceptable treatment plans and they are insensitive to perturbations in the weights of the forward
problem. On the other hand, there exist patients where small perturbations in the weights dramatically alters the treatment plan or equivalently the range of acceptable weights for these patients is smaller. A future research direction is to develop sensitivity metrics to detect sensitive patients and anatomical properties that may lead to sensitivity. These metrics can be used to improve the treatment planning efficiency and quality.

Our foremost contributions in this dissertation was to highlight the sensitivity and stability of inverse models with respect to parameters and observed data. We believe that this study paves the way for a deeper and broader sensitivity and stability analysis of the state-of-the-art inverse models. In particular, one can further exploit the connections of inverse optimization with regression to develop metrics for measuring robustness and reliability of inverse models in the presence of noise and outliers.
Appendix A

Full-Dimensionality of The Feasible Polyhedron

In this appendix we show that in cases where the feasible region for the forward problem is non-full-dimensional the inverse problem may yield trivial solutions. Also in theorem A.0.1 we prove that any polyhedron can be represented as the intersection of an affine subspace and a full-dimensional polyhedron.

**Definition A.0.1.** A non-empty polyhedron $P = \{ x \mid x \in \mathbb{R}^n, Ax \geq b \}$ is full-dimensional if there exists a feasible solution $x \in P$, such that $a_i'x > b_i$ for all $i \in \{1, \ldots, m\}$.

**Proposition A.0.1.** Let $P = \{ x \mid x \in \mathbb{R}^n, Ax \geq b \}$ be a non-full-dimensional polyhedron and Let $x^0 \in P$ be an arbitrary feasible solution. There exists a trivial solution $\tilde{c}$ for problem (4.2) that makes the entire polyhedron optimal.

**Proof.** First, we show that there exists $j \in \{1, \ldots, m\}$ such that $a_j'x = b_j$ for all $x \in P$. Assume the contrary, that is, for each $i \in \{1, \ldots, m\}$ there exists an $x_i$, where $x_i \in P$ and $a_i'x_i > b_i$. Let $\bar{x} = \frac{\sum_{i=1}^{m} x_i}{m}$, note that $\bar{x} \in P$ and for all $i \in \{1, \ldots, m\}$ we have:

$$a_i'\bar{x} \geq \frac{(m-1)b_i + a_i'x_i}{m} > \frac{(m-1)b_i + b_i}{m} > b_i,$$
which is in contradiction with $\mathbf{P}$ being not full-dimensional. Therefore, there exists some $j \in \{1, \ldots, m\}$ such that $a_j' \mathbf{x} = b_j$ for all $\mathbf{x} \in \mathbf{P}$. Let $\tilde{\mathbf{c}} = a_j$, note that the following optimization problem is feasible and its optimal value is attained for all the feasible solutions and is equal to $b_j$:

$$\begin{align*}
\text{minimize} & \quad \tilde{\mathbf{c}}' \mathbf{x} \\
\text{subject to} & \quad \mathbf{A} \mathbf{x} \geq \mathbf{b},
\end{align*}$$

(A.1)

therefore, using strong duality, the dual is also feasible and yields the same optimal value:

$$\begin{align*}
\text{minimize} & \quad \mathbf{b}' \mathbf{y} \\
\text{subject to} & \quad \mathbf{A}' \mathbf{y} = \tilde{\mathbf{c}}, \quad \mathbf{y} \geq \mathbf{0}.
\end{align*}$$

(A.2)

Let $\mathbf{y}^*$ be the optimal solution to the latter problem. Note that $\|\tilde{\mathbf{c}}\| = \|a_j\| = 1$ and $\tilde{\mathbf{c}}' \mathbf{x} = a_j' \mathbf{x} = b_j$ for all $\mathbf{x} \in \mathbf{P}$, therefore, the pair $(\mathbf{y}^*, \tilde{\mathbf{c}})$ is an optimal solution for (4.2) regardless of the given solution $\mathbf{x}^0$.

Theorem A.0.1. For any polyhedron $\mathbf{P} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \mathbf{A}_{m \times n} \mathbf{x} \geq \mathbf{b}\}$, there exist an affine subspace $\mathbf{S} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \mathbf{C}_{l \times n} \mathbf{x} = \mathbf{d}\}$ and a full-dimensional polyhedron $\mathbf{Q} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \tilde{\mathbf{A}}_{k \times n} \mathbf{x} \geq \tilde{\mathbf{b}}\}$ such that $\mathbf{P} = \mathbf{S} \cap \mathbf{Q}$, $\mathbf{C} \tilde{\mathbf{A}}' = 0$, and $k + l \leq m$.

Proof. We prove it by induction on $m$. For $m = 1$, $\mathbf{A} \mathbf{x} \geq \mathbf{b}$ is a half-space and therefore full-dimensional, so $\mathbf{S} = \mathbb{R}^n$ ($\mathbf{C} = 0$), $\mathbf{Q} = \mathbf{P}$ ($\tilde{\mathbf{A}} = \mathbf{A}$), and $\mathbf{C} \tilde{\mathbf{A}}' = 0$. Assume that the statement holds for all positive integers smaller than or equal to some integer $t > 1$, we will prove that the statement also holds for $t + 1$. Let $\mathbf{P}^{(t+1)} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \mathbf{A}_{(t+1) \times n} \mathbf{x} \geq \mathbf{b}\}$ and $\mathbf{P}^{(t)} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \hat{\mathbf{A}}_{t \times n} \mathbf{x} \geq \hat{\mathbf{b}}\}$, where $\hat{\mathbf{A}}_{t \times n}$ and $\hat{\mathbf{b}}$ denote the first $t$ rows of $\mathbf{A}$ and $\mathbf{b}$, respectively. There exists an affine subspace $\mathbf{S}^{(t)} = \{\mathbf{x} \mid \mathbf{C}_{t \times n} \mathbf{x} = \mathbf{d}^{(t)}, \mathbf{x} \in \mathbb{R}^n\}$ and a full-dimensional polyhedron $\mathbf{Q}^{(t)} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \hat{\mathbf{A}}^{(t)}_{r \times n} \mathbf{x} \geq \hat{\mathbf{b}}^{(t)}\}$ such that $\mathbf{P}^{(t)} = \mathbf{S}^{(t)} \cap \mathbf{Q}^{(t)}$ and $\mathbf{C}^{(t)} \hat{\mathbf{A}}^{(t)}$, where $\mathbf{C}^{(t)}$ is an orthonormal basis for $\mathbf{S}^{(t)}$. Note that:
\[ P(t+1) = P(t) \cap H_{t+1} = (S(t) \cap H_{t+1}) \cap Q(t), \]

where \( H_{t+1} = \{ x \mid a'_{t+1}x \geq b_{t+1} \} \) is the half-space induced by the last constraint in \( Ax \geq b \). Define \( a''_t = a'_{t+1} - a'_t C(t) C(t) \) and \( b''_t = b_{t+1} - a'_t C(t) d(t) \). Also let \( H_{t+1}^S \) be as follows:

\[ H_{t+1}^S = \{ x \mid a''_{t+1}x \geq b''_{t+1} \}. \]

Note that we have:

\[ H_{t+1}^S \cap S(t) = H_{t+1} \cap S(t), \]

thus,

\[ P(t+1) = (S(t) \cap H_{t+1}) \cap Q(t) = (Q(t) \cap H_{t+1}^S) \cap S(t). \]

In case that \( Q(t) \cap H_{t+1}^S \) is a full-dimensional polyhedron by concatenating \( a''_{t+1} \) and \( b''_{t+1} \) to \( \tilde{A}(t) \) and \( \tilde{b}(t) \) and constructing \( Q(t+1) = \{ x \mid x \in \mathbb{R}^n, [\tilde{A}(t); a''_{t+1}]x \geq [\tilde{b}(t); b''_{t+1}] \} \) and defining \( S(t+1) = S(t) \), the conditions will hold for the subspace and the full-dimensional polyhedron and the proof is complete.

In the case where \( Q(t) \cap H_{t+1}^S \) is not a full-dimensional polyhedron, we prove that there does not exist \( x \in Q(t) \cap H_{t+1}^S \) such that:

\[ a''_{t+1}x > b''_{t+1}. \]

Assume the contrary. Let \( x \in Q(t) \cap H_{t+1}^S \) and \( \epsilon > 0 \) exist such that:

\[ a''_{t+1}x > b''_{t+1} + \epsilon. \]

Since \( Q(t) \) is full-dimensional there exists \( y \in Q(t) \) such that:

\[ \tilde{A}(t)_{k_t \times n} y > \tilde{b}(t). \]
Choose $0 < \lambda < 1$ such that $\frac{1}{\lambda} - 1 > \frac{b_{t+1}^\perp - a_{t+1}^\perp y}{x}$, it is straightforward to see that $(1 - \lambda)x + \lambda y$ is an interior point for $Q^{(t)} \cap H_{t+1}^\perp$, which is in contradiction with $Q^{(t)} \cap H_{t+1}^\perp$ being not full-dimensional. Therefore, we have:

$$a_{t+1}^\perp x = b_{t+1}^\perp \quad \forall x \in Q^{(t)} \cap H_{t+1}^\perp.$$ 

The latter relation indicates that $a_{t+1}^\perp$ is perpendicular to the entire polyhedron induced by $Q^{(t)} \cap H_{t+1}^\perp$. Without loss of generality assume that $\|a_{t+1}^\perp\| = 1$. Therefore, $a_{t+1}^\perp$ is used to define the subspace which contains the polyhedron. Let $S^{(t+1)} = \{ x \mid \begin{bmatrix} C^{(t)}_{t+1 \times n} ; a_{t+1}^\perp \end{bmatrix} x = [d^{(t)} ; b_{t+1}^\perp] , \quad x \in \mathbb{R}^n \}$, and $C^{(t+1)} = \begin{bmatrix} C^{(t)}_{t+1 \times n} ; a_{t+1}^\perp \end{bmatrix}$. Note that $C^{(t+1)}$ is still orthonormal but $C^{(t+1)}\tilde{A}^{(t+1)} = 0$ does not necessarily hold. Therefore, we need to transform $\tilde{A}^{(t)}$ to $\tilde{A}^{(t+1)}$ (correspondingly $\tilde{b}^{(t)}$ to $\tilde{b}^{(t+1)}$) to transform its rows to become orthogonal to $a_{t+1}^\perp$.

Therefore, each row and element in right hand side of $\tilde{A}^{(t+1)}$ and $\tilde{b}^{(t+1)}$ is calculated as follows:

$$\tilde{a}_i^{(t+1)} = \tilde{a}_i^{(t)} - \tilde{a}_i^{(t)} a_{t+1}^\perp a_{t+1}^\perp,$$

$$\tilde{b}_i^{(t+1)} = \tilde{b}_i^{(t)} - \tilde{a}_i^{(t)} a_{t+1}^\perp b_{t+1}^\perp.$$ 

The polyhedron induced by $\tilde{A}^{(t+1)}$ and $\tilde{b}^{(t+1)}$ is either full-dimensional or not. In case the polyhedron is full-dimensional the proof is complete. In case that the polyhedron is not full dimensional we use the fact that the statement holds for any integer $k \leq t$ and the number of rows of $\tilde{A}^{(t+1)}$ is at most $k_t$, which is smaller than or equal to $t$. Therefore, there exists an equivalent representation with a full-dimensional polyhedron $Q^{(t+1)}$ and a subspace $S^{(t+1)}$ that satisfy the above conditions. Let $S^{(t+1)} = S^{(t+1)} \cap S^{(t)}$. Note that $S^{(t+1)}$ and $Q^{(t+1)}$ satisfy the above conditions, which concludes the proof.

Intuitively, the latter theorem separates the components of the polyhedron that produce non-trivial solutions to the inverse problem ($Q$ polyhedron) from the components that may yield trivial solutions ($S$ subspace). It proves that any polyhedron can be rep-
resented as the intersection of an affine subspace and a full-dimensional polyhedron. In fact, the affine subspace does not contain information about the cost vector, since the vectors in that subspace are all orthogonal to the feasible region and make the entire polyhedron optimal.
Bibliography


[53] H Edwin Romeijn, James F Dempsey, and Jonathan G Li. A unifying framework for


