

# Gauge duality and low-rank spectral optimization

by

Ives José de Albuquerque Macêdo Júnior

B.Sc., Universidade Federal de Pernambuco, 2005

M.Sc., Associação Instituto Nacional de Matemática Pura e Aplicada, 2007

D.Sc., Associação Instituto Nacional de Matemática Pura e Aplicada, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2015

© Ives José de Albuquerque Macêdo Júnior 2015

# Abstract

The emergence of compressed sensing and its impact on various applications in signal processing and machine learning has sparked an interest in generalizing its concepts and techniques to inverse problems that involve quadratic measurements. Important recent developments borrow ideas from matrix lifting techniques in combinatorial optimization and result in convex optimization problems characterized by solutions with very low rank, and by linear operators that are best treated with matrix-free approaches. Typical applications give rise to enormous optimization problems that challenge even the very best workhorse algorithms and numerical solvers for semidefinite programming.

The work presented in this thesis focuses on the class of low-rank spectral optimization problems and its connection with a theoretical duality framework for gauge functions introduced in a seminal paper by Freund (1987). Through this connection, we formulate a related eigenvalue optimization problem more amenable to the design of specialized algorithms that scale well and can be used in practical applications.

We begin by exploring the theory of gauge duality focusing on a slightly specialized structure often encountered in the motivating inverse problems. These developments are still made in a rather abstract form, thus allowing for its application to different problem classes.

What follows is a connection of this framework with two important classes of spectral optimization problems commonly found in the literature: trace minimization in the cone of positive semidefinite matrices and affine nuclear norm minimization. This leads us to a convex eigenvalue optimization problem with rather simple constraints, and involving a number of variables equal to the number of measurements, thus with dimension far smaller than the primal.

The last part of this thesis exploits a sense of strong duality between the primal-dual pair of gauge problems to characterize their solutions and to devise a method for retrieving a primal minimizer from a dual one. This allows us to design and implement a proof of concept solver which compares favorably against solvers designed specifically for the PhaseLift formulation of the celebrated phase recovery problem and a scenario of blind image deconvolution.

# Preface

The work presented herein is the product of a joint work and active collaboration between myself and my supervisor, Professor Michael Friedlander. The material that appears in Chapter 2 and parts of Chapter 3 are the result of additional collaboration with Dr. Ting Kei Pong while he was a postdoc at UBC. The developments that result from this joint work and collaboration have been published in the SIAM Journal on Optimization; see Friedlander, Macêdo, and Pong (2014).

Some parts of Chapter 3 as well as the developments and results appearing in Chapters 4 and 5 have been submitted to the SIAM Journal on Scientific Computing and made publicly available in preprint form; see Friedlander and Macêdo (2015).

The proof of concept implementation described in Chapter 4 was developed in MATLAB jointly by myself and Professor Michael Friedlander, while the specialized solvers against which we compare our results in Chapter 5 were kindly provided by some of their authors.

# Table of Contents

Abstract . . . . .	ii
Preface . . . . .	iii
Table of Contents . . . . .	iv
List of Tables . . . . .	vi
List of Figures . . . . .	vii
Acknowledgments . . . . .	viii
Dedication . . . . .	x
<b>1 Introduction . . . . .</b>	<b>1</b>
1.1 From sparse to low-rank signal recovery . . . . .	2
1.1.1 Phase retrieval via matrix lifting . . . . .	8
1.1.2 Blind deconvolution and biconvex compressive sensing . . . . .	14
1.2 Norm-minimization and gauge duality . . . . .	17
1.3 Thesis overview and contributions . . . . .	25
<b>2 Gauge optimization and duality . . . . .</b>	<b>27</b>
2.1 Preliminaries . . . . .	29
2.1.1 Polar sets . . . . .	30
2.1.2 Gauge functions . . . . .	31
2.1.3 Antipolar sets . . . . .	35
2.2 Antipolar calculus . . . . .	36
2.2.1 Linear transformations . . . . .	37
2.2.2 Unions and intersections . . . . .	40
2.3 Duality derivations . . . . .	41
2.3.1 The gauge dual . . . . .	43
2.3.2 Lagrange duality . . . . .	45

*Table of Contents*

---

2.4	Strong duality . . . . .	46
2.5	Variational properties of the gauge value function . . . . .	50
2.6	Extensions . . . . .	52
2.6.1	Composition and conic side constraints . . . . .	52
2.6.2	Nonnegative conic optimization . . . . .	54
<b>3</b>	<b>Spectral gauge optimization and duality . . . . .</b>	<b>56</b>
3.1	Semidefinite nonnegative conic optimization . . . . .	56
3.1.1	Trace minimization in the PSD cone . . . . .	57
3.2	Nuclear-norm minimization . . . . .	60
3.2.1	Reduction to PSD trace minimization . . . . .	62
3.3	Spectral lower models for convex minimization . . . . .	63
3.3.1	Spectral lower models for the dual gauge objective . . . . .	64
3.3.2	Proximal-bundle subproblems . . . . .	67
<b>4</b>	<b>Low-rank spectral optimization . . . . .</b>	<b>71</b>
4.1	Derivation of the approach . . . . .	73
4.1.1	Recovering a primal solution . . . . .	73
4.1.2	Primal recovery subproblem . . . . .	75
4.2	The implementation of a proof of concept solver . . . . .	77
4.2.1	Dual descent . . . . .	77
4.2.2	Primal recovery . . . . .	80
4.2.3	Primal-dual refinement . . . . .	80
4.3	Extensions . . . . .	81
4.3.1	Weighted trace minimization . . . . .	82
4.3.2	Weighted affine nuclear-norm minimization . . . . .	83
<b>5</b>	<b>Numerical experiments . . . . .</b>	<b>86</b>
5.1	Phase recovery . . . . .	86
5.1.1	Random Gaussian signals . . . . .	87
5.1.2	Random problems with noise . . . . .	88
5.1.3	Two-dimensional signal . . . . .	89
5.2	Blind deconvolution . . . . .	91
<b>6</b>	<b>Conclusions . . . . .</b>	<b>93</b>
	<b>Bibliography . . . . .</b>	<b>96</b>

# List of Tables

Table 2.1	The main rules of the antipolar calculus . . . . .	37
Table 5.1	Phase retrieval comparisons for random complex Gaussian signals of size $n = 128$ measured using random complex Gaussian masks . . . . .	88
Table 5.2	Phase retrieval comparisons for problems with noise . . . . .	89
Table 5.3	Phase retrieval comparisons on a 2-dimensional image. . . . .	90
Table 5.4	Blind deconvolution comparisons. . . . .	92

# List of Figures

Figure 1.1	Unit balls of the 1-, 2- and max-norms in $\mathbb{R}^2$ . . . . .	3
Figure 1.2	Unit balls of the 1-, 2- and max-norms in $\mathbb{R}^3$ . . . . .	3
Figure 1.3	Unit balls of the nuclear- and operator-norms for real symmetric 2-matrices . . . . .	5
Figure 1.4	Unit ball of the nuclear-norm for 2-by-2 real symmetric matrices . . . . .	6
Figure 1.5	Unit ball of the operator-norm for 2-by-2 real symmetric matrices . . . . .	6
Figure 1.6	Visualizing the structural content encoded in the Fourier phase . . . . .	9
Figure 2.1	Visual depiction of the counter-example described in Example 2.2.1 . . . . .	42
Figure 3.1	Visualizing the geometry of the gauge dual feasible set	59
Figure 3.2	Generalized unit balls of the trace-gauge and its polar for real symmetric 2-by-2 matrices . . . . .	60
Figure 3.3	Generalized unit ball of the trace-gauge for real symmetric 2-matrices . . . . .	61
Figure 3.4	Generalized unit ball of the polar to the trace-gauge for real symmetric 2-matrices . . . . .	61
Figure 5.1	Image used for large phase retrieval experiment ( $\approx 2\text{MP}$ )	90
Figure 5.2	Images used for the blind deconvolution experiments .	92

# Acknowledgments

First and foremost, I would like to express my utmost respect, admiration and gratitude to my supervisor, Professor Michael Friedlander. During all these years working together, he has provided me insightful guidance, enthusiastic encouragement and tireless support. His work ethic and attitude towards collaboration has given me an example to aim for.

I would like to further express my gratitude to my colleagues Ting Kei Pong, Nathan Krislock, and Ernie Esser, who has left us way too young and this World a bit less friendly and jovial. Our numerous and insightful discussions and collaboration have had a great impact on me and broadened my perspectives. I am thankful and privileged to have crossed paths with you.

Thank you to professors Nick Harvey and Philip Loewen for very kindly accepting to be part of my Supervisory Committee and providing insightful advice during early and later stages of this research. To Joyce Poon, for making my life so much easier by taking care of and providing guidance through so many administrative matters.

I am also grateful to professors Eldad Haber and Yaniv Plan to generously accepting to be part of my committee as University Examiners and for their thoughtful feedback on this research. I am humbled and thankful to Professor Robert Freund for his acceptance to serve as my External Examiner and for his presence and comments during my Final Doctoral Exam.

Without the financial support provided by awards from the University of British Columbia and its Department of Computer Science, by means of their Four-Year Doctoral Fellowship (4YF) and International Partial Tuition Scholarship, I would have not been able to even start this endeavour. For that, I am very grateful.

I am thankful to my colleagues Gabriel Goh and Julie Nutini for many insightful discussions and great times during our years as students under Michael and to this day. My times in the Scientific Computing Lab were filled with great interactions and good-hearted discussions with them, Farbod Roosta-Khorasani, Michael Wathen, Essex Edwards and, in earlier times, Ewout van den Berg, who convinced me a second doctorate might make sense if working with Michael Friedlander.

## *Acknowledgments*

---

My years at UBC provided me with the opportunity to meet some truly amazing people. I am deeply grateful to the numerous friends I made while living in St. John's College and while part of UBC's Salsa-Rueda and Latin Dance clubs. You guys kept me sane throughout most of this crazy time.

A special thank you to my very dear friends Cristina, Emilio, Clarissa, Julio, Mario and Patricia. You guys made my visits to Calgary so very much warm, no matter how cold it was outside. With you, I would feel like I was back home.

To 錡雨, I am grateful for your support, companionship, patience and love. We endured together much pain, frustration and anxiety that is often involved in the later stages of a doctorate. Without you by my side, this would have been a much steeper journey.

At last, I am grateful and forever indebted to the unconditional love and neverending emotional support from my parents, Rosélis and Ives, and my sisters, Catarine and Jéssica. You are my constants.

*To*  
*My Parents, Rosélie and Ives,*  
*et*  
*My Sisters, Catarina and Jéssica.*

# Chapter 1

## Introduction

Recovering a signal from partial, and possibly corrupted, measurements is a fundamental and ubiquitous problem in science and engineering. In many applications, the nature of the measurement acquisition process limits the type and number of observations available for recovery. This limited availability of data in relation to the unknown signal's ambient space often leads to highly ill-posed inverse problems.

Common approaches to resolve this ambiguity often lead to optimization problems whose solution one wishes to (perhaps approximately) reproduce the measurements while also exhibiting some application-dependent extremal property expected from a good estimate. Classical examples of these approaches are Tikhonov/ $l_2$ -norm regularization and the problem of obtaining a minimum-norm solution to an underdetermined linear system.

The past decade has seen the emergence and fast growth of sparse recovery and compressed sensing, where one expects the signal to possess some low-dimensional structure, typically represented by the number of coefficients necessary for its representation as a linear combination of elements from some basis or larger dictionary. This approach has received much interest for its theoretical recovery guarantees and the development of effective computational methods for solving the challenging optimization problems that are required.

Motivated by the successful application of these developments, there has been a growing body of work that extends concepts and techniques from sparse recovery to problems with different senses of low-dimensionality and more general measurements. Recent examples include extensions of sparse recovery results to low-rank matrices—in which the rank of a matrix is the counterpart to the number of nonzeros in a vector—and problems where the measurements are quadratic, which are then seen as linear measurements from a “lifted” matrix space in which the unknown signal is rank-1.

For the latter examples, there have been suitable theoretical recovery results, but the computational cost associated to the lifting technique has limited the scale of the problems and the practical use of these approaches. The main purpose of this thesis is to study the low-rank spectral optimization problems that arise from these techniques and propose numerical methods able

to handle large-scale problems that arise in practice.

In the following sections we contextualize and introduce the problems that motivate this work along with existing approaches for their solution. We then describe the class of gauge optimization problems, which neatly captures these and a wide variety of other formulations that arise from inverse problems, and provides the basis for our subsequent developments.

## 1.1 From sparse to low-rank signal recovery

In a number of applications of signal processing, it has been observed that digital signals admit compact approximate representations using techniques such as transform coding, i.e., they can be well represented by just a few elements from a suitable basis or dictionary of functions and their coefficients.

This simple observation forms the basis for many compression schemes, such as the ubiquitous JPEG (Taubman and Marcellin, 2001), and naturally leads to the ambitious endeavour of searching for an optimally sparse signal that satisfies a set of measurements.

More formally, let  $s \in \mathbb{R}^d$  be a vector representation of our digital signal,  $\Phi \in \mathbb{R}^{d \times n}$  a basis or overcomplete dictionary of elementary signals (i.e.,  $d \leq n$  and  $\text{rank } \Phi = d$ ),  $\Psi \in \mathbb{R}^{m \times d}$  a model for a linear measurement process and  $b \in \mathbb{R}^m$  a vector of possibly corrupted observations under this process. This quest for a sparsest representation of  $s$  with respect to  $\Phi$ , while approximating the measurements  $\|b - \Psi s\|_2 \leq \epsilon$ , can be posed as the optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|x\|_0 \quad \text{subject to} \quad \|b - Ax\|_2 \leq \epsilon,$$

where  $s = \Phi x$ ,  $A = \Psi \Phi \in \mathbb{R}^{m \times n}$  and  $\|x\|_0 := \#\{i = 1, \dots, n \mid x_i \neq 0\}$ . Although easy to state, this problem is extremely difficult to solve in general, in fact it is known to be NP-hard; c.f. Natarajan (1995).

While studying exact sparse overcomplete representations of signals, Chen, Donoho, and Saunders (2001) propose a convex relaxation that substitutes the function  $\|\cdot\|_0$  by the 1-norm of vectors. This approach leads to a convex minimization problem of the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|x\|_1 \quad \text{subject to} \quad \|b - Ax\|_2 \leq \epsilon,$$

where  $\|x\|_1 := \sum_{i=1}^n |x_i|$ , which we refer to as the *basis pursuit denoise* (BPDN) problem. This convex optimization problem can be solved in practice using a number of techniques; e.g., see van den Berg and Friedlander (2011).

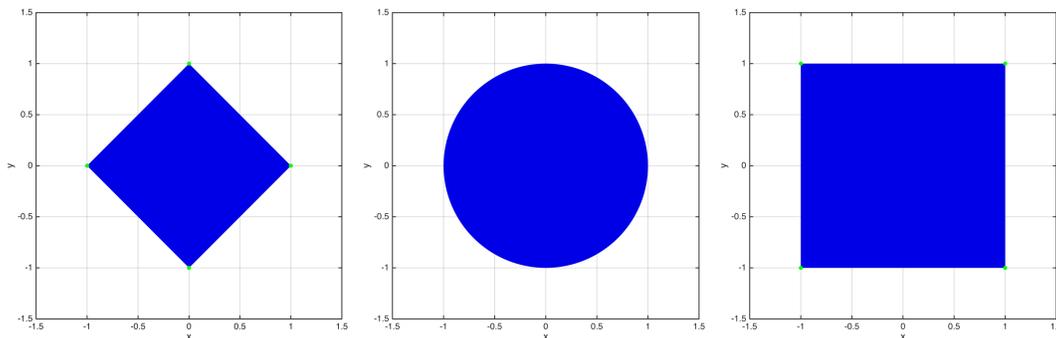


Figure 1.1: Unit balls of the 1-, 2- and max-norms in  $\mathbb{R}^2$ .

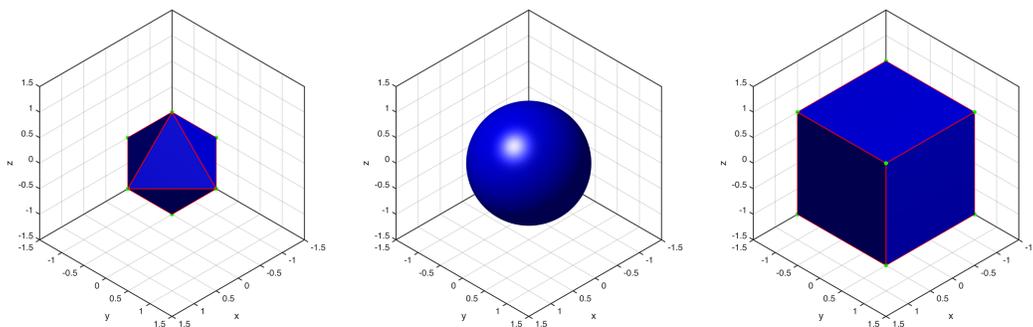


Figure 1.2: Unit balls of the 1-, 2- and max-norms in  $\mathbb{R}^3$ .

By looking at the unit balls of different  $p$ -norms (Figures 1.1 and 1.2), one can intuitively observe that a minimum-norm point on an affine subspace in general position is most likely to be sparse if the norm used is the 1-norm.

The practical observation that *basis pursuit* ( $\epsilon = 0$ ) solutions often lead to an exact recovery of sparse signals led to further study of the relationships between the original and the convexified problems. This convex optimization approach to sparse recovery gained tremendous momentum with theoretical results providing a number of conditions under which these two problems are in fact equivalent, or at least bound the error for the BPDN solution—e.g., see Candès and Tao (2005); Donoho (2006); Bruckstein, Donoho, and Elad (2009).

Motivated by the practical success and available theoretical tools and techniques developed for compressed sensing based on the 1-norm heuristic, there has been an effort to extend this framework to other classes of problems in which different notions of complexity play roles analogous to that of the *cardinality* (i.e., number of nonzeros) in sparse recovery.

Most notable is the generalization to matrices in which the rank function serves as the matrix counterpart to the number of nonzeros in a vector. Applications in which a low-rank matrix is sought appear in many disciplines and examples include: factor analysis, where one looks for low-degree statistical models; the computation of low-order controllers or realizations of linear systems in control theory; and finding low-dimensional Euclidean embeddings. In such scenarios, one seeks a lowest-rank matrix satisfying a number of linear measurements (possibly within a bound on the misfit), leading to the problem

$$\underset{X \in \mathbb{R}^{n_1 \times n_2}}{\text{minimize}} \quad \|X\|_0 \quad \text{subject to} \quad \|b - \mathcal{A}X\|_2 \leq \epsilon, \quad (1.1)$$

where  $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$  is a linear map,  $\|X\|_0 := \|\sigma(X)\|_0 = \text{rank } X$  and  $\sigma : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}_+^{n_{\min}}$  maps a matrix to the vector of its singular values in nonincreasing order ( $n_{\min} := \min\{n_1, n_2\}$ ), i.e.,  $\sigma_1(X) \geq \dots \geq \sigma_{n_{\min}}(X) \geq 0$ . This is indeed a generalization: a vector can be encoded as a diagonal matrix and the rank function then gives the vector's cardinality.

By observing the characterization of the rank function in terms of the cardinality of the vector of singular values, one can naturally devise a convex relaxation by taking the 1-norm of that vector. This is in fact the convex heuristic for rank minimization advocated and proved tightest by Fazel, Hindi, and Boyd (2001); see also Fazel (2002). This approach leads to convex optimization problems of the form

$$\underset{X \in \mathbb{R}^{n_1 \times n_2}}{\text{minimize}} \quad \|X\|_1 \quad \text{subject to} \quad \|b - \mathcal{A}X\|_2 \leq \epsilon, \quad (1.2)$$

where  $\|X\|_1 := \|\sigma(X)\|_1 = \sum_{i=1}^{n_{\min}} \sigma_i(X)$  is the *nuclear norm* of a matrix (also known as the *trace norm*, Ky Fan  $n$ -norm and Schatten 1-norm).

**Remark 1.1.1.** We shall adopt the Schatten  $p$ -norm notation throughout, i.e.,  $\|X\|_p := \|\sigma(X)\|_p$  with  $1 \leq p \leq \infty$ , which closely parallels the notation of the vector case. A peculiar consequence of this choice of notation is that  $\|X\|_2$  will in fact denote the *Frobenius norm* of  $X$ , as opposed to its *operator norm* denoted by  $\|X\|_\infty = \|\sigma(X)\|_\infty = \sigma_1(X)$ .

Figure 1.3 depicts the unit balls of the nuclear and operator norms in a parameterization of the space of real symmetric 2-by-2 matrices (in this parameterization, the Frobenius norm has the same geometry as that of the 2-norm in  $\mathbb{R}^3$ , depicted in Figure 1.2). The brighter kinks and edges depict the extreme points of these convex bodies. In the case of the nuclear norm, the brighter points along the circumferences forming the edges of the cylinder correspond to rank-1 matrices, which intuitively justifies the minimization of

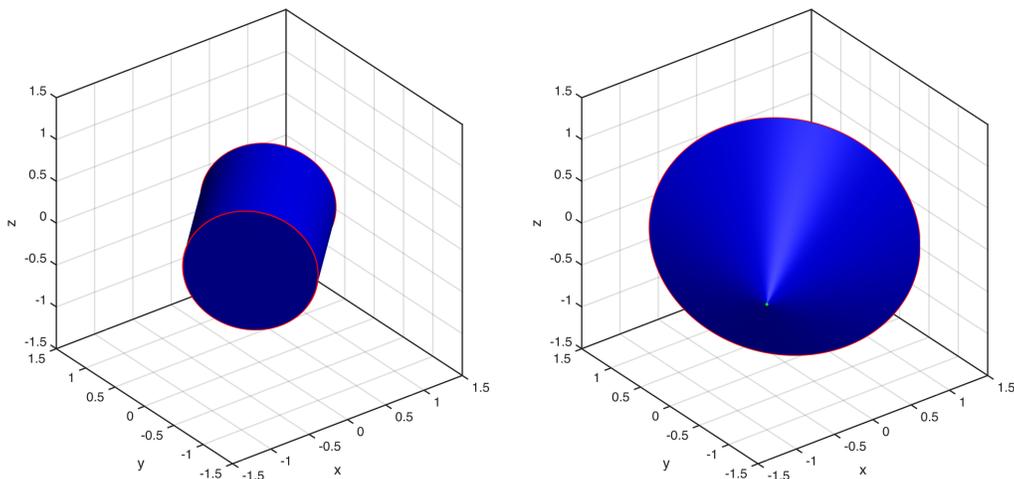


Figure 1.3: Unit balls of the nuclear- and operator-norms for real symmetric 2-matrices. Depictions above correspond to the parameterization  $\left[ x, \frac{z}{\sqrt{2}}; \frac{z}{\sqrt{2}}, y \right]$ . Figures 1.4 and 1.5 provide different viewing positions.

this norm in the search for low-rank solutions. Figures 1.4 and 1.5 provide different viewpoints to better visualize the geometry of these sets.

Early work by Recht, Fazel, and Parrilo (2010) exploited this connection between the 1-norm relaxation from sparse recovery and the nuclear-norm heuristic from rank minimization to extend concepts and techniques from compressed sensing and provide the first sufficient guarantees for the recovery of low-rank matrices. Central to their work is a suitable generalization of the *restricted isometry property* (RIP) to operators on low-rank matrices, introduced first for sparse vectors by Candès and Tao (2005).

**Definition 1.1.1** (Restricted Isometry Property). Let  $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$  be a linear map and  $n_{\min} := \min\{n_1, n_2\}$ . For every  $1 \leq r \leq n_{\min}$ , define the  $r$ -restricted isometry constant to be the smallest number  $\delta_r(\mathcal{A})$  such that

$$(1 - \delta_r) \|X\|_2 \leq \|\mathcal{A}X\|_2 \leq (1 + \delta_r) \|X\|_2$$

holds for all matrices  $X \in \mathbb{R}^{n_1 \times n_2}$  with  $\text{rank } X \leq r$ .

With this RIP concept,  $X_0 \in \mathbb{R}^{n_1 \times n_2}$  a fixed matrix of rank at most  $r$  and  $b := \mathcal{A}X_0$ , the following result provides a condition for injectivity of  $\mathcal{A}$  on the set of matrices with rank bounded by  $r$ .

1.1. From sparse to low-rank signal recovery

---

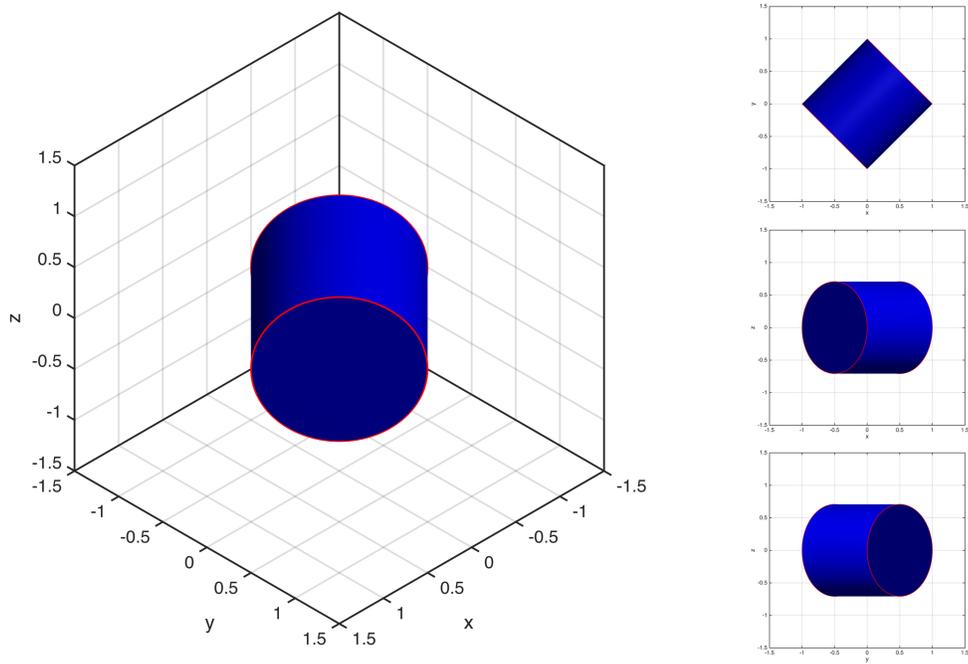


Figure 1.4: Unit ball of the nuclear-norm for 2-by-2 real symmetric matrices.

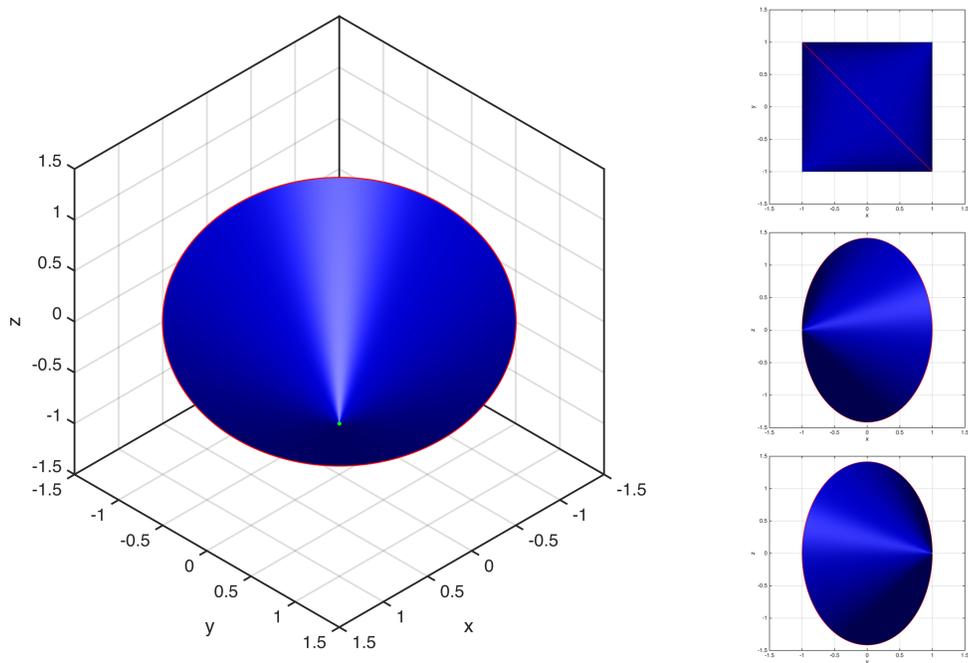


Figure 1.5: Unit ball of the operator-norm for 2-by-2 real symmetric matrices.

**Theorem 1.1.1** (Recht et al. (2010)). Suppose that  $\delta_{2r} < 1$  for some integer  $r \geq 1$ . Then  $X_0$  is the only matrix of rank at most  $r$  satisfying  $\mathcal{A}X = b$ .

The following result then provides a connection between rank minimization and its convex relaxation. Let  $X_*$  be a solution of the nuclear-norm minimization problem (1.2) with  $\epsilon = 0$  and the remaining data as above.

**Theorem 1.1.2** (Recht et al. (2010)). Suppose that  $r \geq 1$  is such that  $\delta_{5r} < 1/10$ . Then  $X_* = X_0$ .

Recht et al. (2010) then exploited these sufficient conditions to provide a family of random measurement maps  $\mathcal{A}$  for which RIP is satisfied with high probability whenever  $m$  is sufficiently large, however still asymptotically much smaller than  $n_1 n_2$ .

A typical example of such constructions with probabilistic guarantees is the case when the entries of  $\mathcal{A}$  are assumed i.i.d. Gaussian random variables with zero mean and variance  $1/m$ . The result below is but a particular case of a theorem of Recht et al. (2010).

**Theorem 1.1.3** (Recht et al. (2010, Theorem 4.2)). Fix  $0 < \delta < 1$ . If  $\mathcal{A}$  is a random variable as above, then, for every  $1 \leq r \leq n_{\min}$ , there exist positive constants  $c_0$  and  $c_1$  depending only on  $\delta$  such that, with probability at least  $1 - \exp(-c_1 m)$ ,  $\delta_r(\mathcal{A}) \leq \delta$  whenever  $m \geq c_0 r(n_1 + n_2) \log(n_1 n_2)$ .

Because the number of degrees of freedom of  $n_1 \times n_2$  real matrices with rank at most  $r$  is  $r(n_1 + n_2 - r)$ , this result ensures that the sufficient recoverability conditions above are very likely satisfied without having to perform a number of measurements anywhere near the dimension of the ambient space ( $= n_1 n_2$ ) with a qualitatively modest increase on the underlying dimensionality of the problem.

Although these results generalize in a natural manner their counterparts in sparse recovery, the measurement process induced by such families of operators—in which one essentially measures random projections of the unknown matrix and accesses information about all of its entries—does not commonly arise in practice.

Later recovery results studied the problem of *matrix completion* (also known as *collaborative filtering* or *the Netflix problem*), in which one seeks to retrieve a low-rank matrix by observing a relatively small subset of its entries. Motivated by this problem, Candès and Recht (2009) introduced similar probabilistic results under a more practical measurement scenario. Avoiding a

number of technical definitions, their results essentially state that “most  $n$ -by- $n$  matrices” of rank  $r$  can be recovered exactly, with high probability, via nuclear-norm minimization from  $m$  entries observed uniformly at random as long as  $m \geq Crn^{1+c} \log n$ , for constants  $C$  and  $1/5 \leq c \leq 1/4$ .

These works sparked an interest in providing more refined conditions under which recoverability of low-rank matrices can be attained via nuclear norm minimization under a variety of linear measurement models, these include Recht (2011); Recht, Xu, and Hassibi (2011); Eldar, Needell, and Plan (2012).

The salient characteristics in these results relevant for the design of structured optimization methods are that the number of measurements is considerably smaller than the dimension of the ambient matrix space, as well as the typical rank expected in the solution is also low, while the linear measurement operator admits fast evaluation on factored low-rank matrices as does the product of its adjoint (and transpose) to a vector, i.e.,  $(\mathcal{A}^*y)v$  and  $(\mathcal{A}^*y)^T u$ .

In the following subsections we describe and contextualize two important problems in scientific imaging and signal processing to which these techniques for low-rank matrix completion have been extended. Interesting characteristics shared by these two problems are that they involve quadratic measurements in their nonconvex formulations, typically involve large dimensions already in this original form, and are expected to lead to rank-1 solutions of their relaxations.

### 1.1.1 Phase retrieval via matrix lifting

In a number of imaging science applications, the recovery of a complex signal from magnitude observations of its Fourier transform is a fundamental problem. In such applications, physical properties of the measurement system typically prevent gathering phase information or at least without incurring large errors (Millane, 2006).

Since much of the structural content of a signal is encoded in its phase information, the problem of *phase retrieval* appears naturally and must be tackled. In Figure 1.6 we illustrate this property by computing the Fourier transform of two real images, swapping their phases, and visualizing the results from their inverse Fourier transforms—independently across color channels.

Assuming that magnitude measurements are available, one can think of a model in which the signal is first probed by a linear process prior to observing its magnitudes. In such a scenario, we seek to recover a complex signal  $x \in \mathbb{C}^n$  from measurements of the form  $|Ax|$ , where  $|\cdot|$  denotes the componentwise absolute-value map and  $A \in \mathbb{C}^{m \times n}$  models the initial probing process, in which one would naturally expect  $m \geq n$ . An example is the case when  $A = F$  the *discrete Fourier transform* (DFT) matrix.



(a) Signal  $x_1$



(b) Signal  $x_2$



(c)  $F^{-1}\{|Fx_1|(Fx_2)/|Fx_2|\}$



(d)  $F^{-1}\{|Fx_2|(Fx_1)/|Fx_1|\}$

Figure 1.6: Visualizing the structural content encoded in the Fourier phase.

This is a scenario in which the most popular algorithms operate, the alternating (nonconvex) projection methods introduced in the seminal works of Gerchberg and Saxton (1972) and Fienup (1978, 1982). The main difficulties associated with these approaches lie on the need for a careful exploitation of prior knowledge about the signal—e.g., support, nonnegativity, real-valuedness etc.—, sensible parameter selection and availability of a good initial iterate.

Motivated by these difficulties, structured multiple illumination techniques in diffraction imaging, and prior connections to rank minimization and matrix completion problems by Chai, Moscoso, and Papanicolaou (2011), Candès, Eldar, Strohmer, and Voroninski (2013a) propose and study a more applicable measurement model and a convex relaxation for the phase recovery problem.

We start with a description of the convex relaxation by noticing that

$$|Ax|_i^2 = (Ax)_i \overline{(Ax)_i} = [(Ax)(Ax)^*]_{ii} = [A(xx^*)A^*]_{ii},$$

leads us to a problem of the form

$$\text{find } x \in \mathbb{C}^n \quad \text{such that} \quad \text{diag}[A(xx^*)A^*] = b,$$

where  $b \in \mathbb{R}^m$  denotes the input (squared) magnitude measurements.

Observing the intrinsic global-phase ambiguity in the measurement process, in which  $\exp(i\theta)x$  evaluates to the same measurements as  $x$ , we can identify these  $x$  with a rank-1 positive semidefinite Hermitian matrix  $X$  via  $X = xx^*$ . Therefore, if it does admit a solution, this nonconvex feasibility problem can be recast equivalently as one of rank minimization with the form

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \text{rank } X \quad \text{subject to} \quad \mathcal{A}X = b \text{ and } X \succeq 0,$$

where  $X \succeq 0$  denotes that the matrix  $X$  is Hermitian and positive semidefinite.

Employing the nuclear-norm convex relaxation for rank minimization, we arrive at the problem

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \|X\|_1 \quad \text{subject to} \quad \mathcal{A}X = b \text{ and } X \succeq 0,$$

which can then be simplified by noticing that  $\sigma(X) = \lambda(X)$  implies  $\|X\|_1 = \text{trace } X$  for all  $X \succeq 0$ , where  $\lambda : \mathcal{H}^n \rightarrow \mathbb{R}^n$  denotes the ordered eigenvalue map, i.e.,  $\lambda_1(X) \geq \dots \geq \lambda_n(X)$ . This results in the trace minimization SDP

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \text{trace } X \quad \text{subject to} \quad \mathcal{A}X = b \text{ and } X \succeq 0.$$

This rationale was employed by Chai et al. (2011) in their study of a class of array imaging problems in which only magnitude measurements can

be obtained by current sensors. They exploit the connection with low-rank recovery to provide conditions on the imaging and scatterer configurations under which the rank minimization problem admits a unique solution. Making use of then recent low-rank recovery results, this was accomplished by proving sufficient conditions for  $\delta_2(\mathcal{A}) < 1$  and invoking Theorem 1.1.1 above.

Candès et al. (2013a) introduce a measurement model inspired on a number of structured illumination techniques in scientific imaging. In this model, an illumination of the signal  $x$  is modulated by a *coded diffraction pattern* (CDP), modeled as a diagonal matrix  $C = \text{Diag}(c) \in \mathbb{C}^{n \times n}$ , and the resulting signal  $Cx$  undergoes a measurement of the squared magnitude of its Fourier transform, leading to measurements of the form  $|FCx|^2$ . This procedure is performed  $L \geq 1$  times with different CDPs, which can be deterministic or drawn from a random ensemble, e.g., Bernoulli entries modeling a binary mask.

In our notation, this model leads to  $m = nL$  measurements of the form

$$A := \begin{pmatrix} FC_1 \\ \vdots \\ FC_L \end{pmatrix}, \quad \text{where } C_k := \text{Diag}(c_k), \quad k = 1, \dots, L.$$

It also give rise to very large optimization problems involving linear maps only feasibly treated as matrix-free operators.

In that work, Candès et al. (2013a) provide a deterministic construction of CDPs for which, if the trace minimization problem admits a rank-1 solution, it matches the measured signal up to a global phase. The main shortcomings were lack of guarantees that the convex problem would have a rank-1 minimizer and whether the approach would be stable to noisy measurements. These issues were further studied by Candès, Strohmer, and Voroninski (2013b) in the case where the rows of  $A$  are Gaussian or uniformly drawn from the sphere of radius  $\sqrt{n}$  and the measurements have bounded errors. The convex problem proposed is the one we call here (PhaseLift) (although this originally assumes  $\epsilon = 0$ )

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \text{trace } X \quad \text{subject to} \quad \|b - AX\|_2 \leq \epsilon \text{ and } X \succeq 0. \quad (\text{PhaseLift})$$

The authors show that, for an arbitrary signal and  $\epsilon = 0$ , if  $m \geq c_0 n \log n$  measurements are taken from the distributions above, this problem has a unique solution and it has rank-1 with probability at least  $1 - 3 \exp(-\gamma m/n)$ , for positive constants  $c_0, \gamma$ .

For uniform measurements of the sphere, they show that, under the same conditions, if the measurements are contaminated by additive noise bounded on the 2-norm by  $\epsilon > 0$ , the best rank-1 approximation  $\hat{X}$  to the solution

of (PhaseLift) will be within a distance of  $C_0\epsilon$  to the measured signal in the Frobenius norm, with nearly the same probability as the noiseless case.

These results were further refined by Candès and Li (2014) removing the logarithmic term, the dependence on the signal—i.e., draw  $A$  and the probabilistic bounds hold for all signals—and proving that the feasible set of the  $\epsilon = 0$  problem is in fact a singleton, a result also shown by Demanet and Hand (2014) but with a less tight lower bound on the number of measurements.

Much more recently, similar results have been proved for the semidefinite relaxation under the CDP model in which the  $c_k$  are drawn for certain random ensembles. Candès, Li, and Soltanolkotabi (2015b) show that  $L \geq c_0 \log^4 n$  ensure that the feasible set of the  $\epsilon = 0$  problem is a singleton with probability at least  $1 - 1/n$ . Their result being further refined by Gross, Krahmer, and Kueng (2014) decreasing the exponent on the logarithm from 4 down to 2.

Although the matrix lifting and convexification techniques relax the quadratic measurements to linear operators on matrices, they lead to incredibly challenging convex optimization problems due to the quadratic increase in the dimensionality of the primal space. This calls for customized computational methods that exploit the structure of these operators and the (nearly) low-rank properties of the minimizers.

In the following, we describe two main approaches designed for these convex relaxations of the phase retrieval problem and contrast their results with those we obtain in Chapter 5.

### Convex prox-gradient methods

In their original paper proposing the PhaseLift formulation, Candès et al. (2013a) propose a computational strategy based on PSD constrained least squares (or alternative misfit) regularized by the trace. For a given regularization parameter  $\rho > 0$ , their basic approach aims to solve the problem

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad f(X) := \left\{ \frac{1}{2} \|b - \mathcal{A}X\|_2^2 + \rho \text{trace } X \right\} \quad \text{subject to} \quad X \succeq 0.$$

To solve this problem, the authors leverage the excellent MATLAB package **TFOCS**, introduced by Becker, Candès, and Grant (2011). **TFOCS** provides a wrapper for a number of related first-order algorithms to create customized solvers for a variety of convex models. The main building block used to solve the problem above is the implementation of a prox-gradient operator, i.e., a routine to solve

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad f(X_k) + \langle \nabla f(X_k), X - X_k \rangle + \frac{1}{2\alpha_k} \|X - X_k\|_2^2 \quad \text{subject to} \quad X \succeq 0.$$

By elementary manipulations, it is easy to see that this amounts to computing a projection onto the PSD cone, i.e.,  $\mathcal{P}_{\succeq 0}(X_k - \alpha_k \nabla f(X_k))$ . This operator can be evaluated by computing the positive eigenvalues and eigenvectors of a matrix. Formally, if  $X = Q \text{Diag}(\lambda(X))Q^*$  is an eigendecomposition of  $X$ , we have that  $\mathcal{P}_{\succeq 0}(X) = Q \text{Diag}([\lambda(X)]_+)Q^*$ , where  $([x]_+)_i := \max\{0, x_i\}$  denotes the positive-part operator. Assuming  $X_k$  is available as a low-rank factorization, the structure of  $\nabla f(X_k) = -\mathcal{A}^*(b - \mathcal{A}X_k) + \rho I$  allows for efficient products between the matrix to be projected and a given vector, thus amenable to exploit Lanczos-based eigensolvers to compute the positive rightmost eigenpairs.

A downside of this approach is that there is no *a priori* knowledge nor control on how many positive eigenvalues may appear throughout the iterations. Aware of this issue, the authors heuristically fix the maximum number of computed eigenpairs and mention that the convergence guarantees provided by the convex optimization algorithms implemented in TFOCS are lost. Nevertheless, good empirical results have been reported in their paper, where this number is taken between 10 and 20.

For it being one of the first computational methods able to solve the PhaseLift formulation in instances just too large for off-the-shelf SDP solvers, and whose code was kindly provided to us by Prof. Thomas Strohmer, we contrast the results obtained with our proof-of-concept solver to those of their implementation later in Chapter 5.

It is noteworthy that Candès et al. (2013a) also propose an iteratively reweighted method in which sequences of problems similar to the one above are solved and achieve better low-rank recovery results with fewer measurements. Although we have not performed experiments with reweighted formulations, we show in §4.3 how they can be dealt with and fit within the approach we propose.

### Gradient descent on a nonconvex formulation

During the later stages of our research, Candès, Li, and Soltanolkotabi (2015a) proposed an approach for the phase recovery problem in which one employs a simple gradient descent method on a nonlinear least squares (or alternative misfit function) formulation. Their approach aims to solve the problem

$$\underset{x \in \mathbb{C}^n}{\text{minimize}} \quad f(x) := \frac{1}{4m \|x_0\|_2^2} \|b - \mathcal{A}(xx^*)\|_2^2.$$

Using a predefined sequence of steplengths  $(\alpha_k)$ , this leads to the sequence

$$x_{k+1} := x_k + \frac{\alpha_k}{\|x_0\|_2^2} \left\{ \frac{1}{m} [\mathcal{A}^*(b - \mathcal{A}(x_k x_k^*))] x_k \right\}$$

and has been reported to provide excellent numerical solutions.

The key idea behind the effectiveness of this method, and difference from classical nonconvex approaches, lies in a judicious choice of its first iterate  $x_0$  :

$$x_0 \in \arg \max_{\|x\|_2^2 = n \frac{\sum_{i=1}^m b_i}{\|A\|_2^2}} \left\langle \left( \frac{1}{m} \mathcal{A}^* b \right) x, x \right\rangle.$$

This way, the first iterate is computed by a suitable rescaling of an eigenvector of  $\mathcal{A}^* b$  associated to its rightmost eigenvalue  $\lambda_1(\mathcal{A}^* b)$ , which can be computed via matrix-free Lanczos-based methods, thus taking advantage of fast operators and allowing the solution of problems far larger than the prox-based methods used by Becker et al. (2011).

Naturally, this nonconvex approach with spectral initialization requires a certain number of measurements to be taken for the initial iterate to be close enough to the basin of attraction of a solution. The remarkable achievement of this work is that Candès et al. (2015a) provide guarantees under which, if a large enough number of measurements in the coded diffraction model are taken ( $m \geq Cn \log^d n$ , for constant  $C$  and  $2 \leq d \leq 4$ ), the algorithm will converge to an exact solution with high probability.

Similar provably effective nonconvex approaches with spectral initialization have been proposed for Gaussian measurements  $A$  that induce operators of the form  $\mathcal{A}X = \text{diag}(AXA^*)$ —see Netrapalli, Jain, and Sanghavi (2013)—and the much more recent works by White, Sanghavi, and Ward (2015) and Chen and Candès (2015), the latter of which came to our attention at the time of writing this thesis.

In Chapter 5, we contrast results obtained by our solver with those provided by an implementation of this approach made available by its authors.

### 1.1.2 Blind deconvolution and biconvex compressive sensing

Motivated by applications in image processing and wireless communications, Ahmed, Recht, and Romberg (2014) study a restricted version of the blind deconvolution problem under the framework of low-rank matrix recovery.

Succinctly, the blind (circular) deconvolution problem can be stated as: given measurements of the form  $b = f_1 * f_2 \in \mathbb{C}^m$ , where  $* : \mathbb{C}^m \times \mathbb{C}^m \rightarrow \mathbb{C}^m$  denotes the discrete circular convolution operator, recover the generating signals  $f_1, f_2 \in \mathbb{C}^m$  up to scaling (since  $cf_1$  and  $c^{-1}f_2$  evaluate to the same measurements for all  $c \neq 0$ ).

This problem has been approached from a number of perspectives and algorithms have been proposed for it in different scenarios by exploiting various application-domain specific priors, e.g., see Levin, Weiss, Durand, and Freeman (2011) for uses and solution schemes in natural image processing.

Ahmed et al. (2014) analyze an idealized version of this problem in which the signals are spread out in known subspaces, i.e., given full column-rank matrices  $B_1 \in \mathbb{C}^{m \times n_1}$  and  $B_2 \in \mathbb{C}^{m \times n_2}$ , then  $f_1 = B_1 x_1$  and  $f_2 = B_2 x_2$  for some (dense) coefficient vectors  $x_1 \in \mathbb{C}^{n_1}$  and  $x_2 \in \mathbb{C}^{n_2}$ . This leads to the problem

$$\text{find } x_1 \in \mathbb{C}^{n_1} \text{ and } x_2 \in \mathbb{C}^{n_2} \quad \text{such that} \quad (B_1 x_1) * (B_2 x_2) = b.$$

Invoking the convolution theorem, and denoting by  $F \in \mathbb{C}^{m \times m}$  the unitary DFT matrix ( $F^{-1} = F^*$ ), we observe the following chain of equalities

$$\begin{aligned} (B_1 x_1) * (B_2 x_2) &= F^* \text{diag}[(F B_1 x_1)(F B_2 x_2)^T] \\ &= F^* \text{diag}[(F B_1)(x_1 \overline{x_2}^*)(\overline{F B_2})^*]. \end{aligned}$$

Defining  $A_1 := F B_1 \in \mathbb{C}^{m \times n_1}$  and  $A_2 := \overline{F B_2} \in \mathbb{C}^{m \times n_2}$ , we can employ a matrix lifting technique similar to the one used for the phase recovery problem and identify the coefficient vectors  $x_1$  and  $x_2$  with the matrix  $X = x_1 \overline{x_2}^*$ . Thus reformulating the problem above as one of rank minimization

$$\underset{X \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad \text{rank } X \quad \text{subject to} \quad F^* \text{diag}(A_1 X A_2^*) = b.$$

A convex relaxation follows immediately from our previous discussions by substituting the rank function by its convex surrogate  $\|\cdot\|_1$ , leading to a nuclear-norm minimization problem of the form

$$\underset{X \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad \|X\|_1 \quad \text{subject to} \quad \|b - \mathcal{A}X\|_2 \leq \epsilon, \quad (\text{NNM})$$

where we use  $\mathcal{A}X := F^* \text{diag}(A_1 X A_2^*)$  and have introduced bounded-misfit constraints to allow for errors, as described by Ahmed et al. (2014).

Ahmed et al. (2014) then provide technical conditions on these bases and the spread of  $f_1$  in Fourier domain under which exact recovery is achieved in the noiseless case with  $\epsilon = 0$ .

In order to experimentally illustrate the recovery ability of this approach in a scenario of two-dimensional image deblurring, they design experiments where the image support of a blurring kernel is known (but not its values) and the image to be recovered lies in a known subspace represented as a subset

of the Haar wavelet basis (or is estimated from the Haar coefficients of the blurred image). Their numerical results are very encouraging and we compare them with our results on one of their experiments in Chapter 5. This choice was made essentially to compare the optimization approaches, as we make no claims of contributions regarding recovery ability once the data for the convex relaxations has been defined.

It is noteworthy that the matrix lifting rationale used above can be employed for more general bilinear (or sesquilinear) measurements. An approach in this direction was presented recently by Ling and Strohmer (2015), who propose an alternative scheme to achieve both low-rank and sparse recovery (where  $x_1$  and  $x_2$  are expected to be sparse) by using a matrix lifting to linearize the measurement operators, and the *elementwise* 1-norm on the lifted matrix to encourage sparsity. Although they provide recovery guarantees for their measurement models, there is still a lack of scalable computational approaches for their *SparseLift* formulation.

In the following we provide some details on the method used by Ahmed et al. (2014) to solve (NNM).

### Augmented Lagrangian for low-rank SDPs

Ahmed et al. (2014) use the nonlinear programming (NLP) algorithm of Burer and Monteiro (2003) to solve nuclear-norm minimization problems whose size would be too large for off-the-shelf SDP packages.

Their approach first reformulates (NNM), with  $\epsilon = 0$ , into a fully equivalent SDP as proposed by Fazel (2002). This leads to a problem of the form

$$\begin{aligned} & \underset{\substack{X \in \mathbb{C}^{n_1 \times n_2}, \\ U \in \mathcal{H}^{n_1}, \\ V \in \mathcal{H}^{n_2}}}{\text{minimize}} & \quad \frac{1}{2} \text{trace} \begin{pmatrix} U & X \\ X^* & V \end{pmatrix} \\ & \text{subject to} \\ & \quad \mathcal{A}X = b \quad \text{and} \quad \begin{pmatrix} U & X \\ X^* & V \end{pmatrix} \succeq 0. \end{aligned}$$

Next, this method represents  $X$  in a low-rank factored form  $X = ZZ^*$ , with  $Z \in \mathbb{C}^{(n_1+n_2) \times r}$  and  $r \geq 1$  chosen *a priori* to be small. This way, the SDP above can be reformulated as the nonconvex constrained NLP

$$\underset{Z \in \mathbb{C}^{(n_1+n_2) \times r}}{\text{minimize}} \quad \frac{1}{2} \|Z\|_2^2 \quad \text{subject to} \quad \hat{\mathcal{A}}(ZZ^*) = b,$$

where we define  $\hat{\mathcal{A}} : \mathcal{H}^{n_1+n_2} \rightarrow \mathbb{C}^m$  as the extension of the measurement operator  $\mathcal{A}$  to Hermitian matrices that acts on the top-right off-diagonal block.

At this point the classical augmented Lagrangian method is applied to the resulting problem. This leads to a sequence of subproblems of the form

$$Z_{k+1} \in \arg \min_{Z \in \mathbb{C}^{(n_1+n_2) \times r}} \frac{1}{2} \|Z\|_2^2 + \langle y_k, b - \hat{\mathcal{A}}(Z_k Z_k^*) \rangle + \frac{\rho_k}{2} \|b - \hat{\mathcal{A}}(Z_k Z_k^*)\|_2^2,$$

in which we omit the details related to rules for updating the dual variables  $y_k \in \mathbb{C}^m$  and regularization parameters  $\rho_k > 0$ . The numerical solution of these subproblems (to first-order stationarity) is then carried out using a standard limited-memory quasi-Newton method via the `minFunc` package by Schmidt (2005), and thus being able to take full advantage of matrix-free FFT routines.

Such an approach relies on the effectiveness of the subproblem solver to provide good solutions and that the *a priori* choice of the rank  $r$  is large enough to capture the expected solution, which seems reasonable in the scenarios where low-rank recovery is likely. For problems with positive  $\epsilon$ , the authors mention they stop the iterations as soon as the misfit falls below that value, not necessarily solving (NNM) but providing a feasible point.

Chapter 5 compares the solution computed by this method, from an implementation made available online by the authors, to that output from our solver in an instance of image deblurring.

## 1.2 Norm-minimization and gauge duality

Chandrasekaran, Recht, Parrilo, and Willsky (2012) generalize results of sparse and low-rank recovery under linear measurements to more abstract notions of low-complexity. They observe that the constructions of convex relaxations for the cardinality function of vectors and for the rank of matrices can be suitably generalized by the notion of *atomic norms* defined below.

**Definition 1.2.1** (Atomic norm). Let  $\mathcal{T} \subset \mathbb{R}^n$  be compact (with elements called *atoms*), such that all of its elements are extreme points of  $\text{conv } \mathcal{T}$  and  $0 \in \text{conv } \mathcal{T}$ . The *gauge of  $\mathcal{T}$* , denoted as  $\|\cdot\|_{\mathcal{T}} : \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ , is defined by

$$\|x\|_{\mathcal{T}} := \inf\{t > 0 \mid x \in t \text{conv } \mathcal{T}\}. \quad (1.3)$$

If  $\mathcal{T}$  is centrally symmetric about the origin (i.e.,  $a \in \mathcal{T}$  implies  $-a \in \mathcal{T}$ ), then  $\|\cdot\|_{\mathcal{T}}$  is a norm and called the *atomic norm* induced by  $\mathcal{T}$ .

With this concept, the set  $\mathcal{S} := \{\pm e_1, \dots, \pm e_n\}$  (where  $e_k \in \mathbb{R}^n$  is the  $k$ th-element of the canonical basis for  $\mathbb{R}^n$ ) will induce the atomic norm  $\|\cdot\|_{\mathcal{S}} =$

1.2. Norm-minimization and gauge duality

---

$\|\cdot\|_1 : \mathbb{R}^n \rightarrow \mathbb{R}_+$ . Under this framework, the atoms of  $\mathcal{S}$  are taken to be the sparsest unit 2-norm elements in  $\mathbb{R}^n$ , i.e. the elements of the canonical basis and their negated counterparts. Similarly, the nuclear norm is induced by taking the gauge of all rank-1 matrices with unit operator-norm—i.e.,

$$\mathcal{L} := \{xy^T \mid x \in \mathbb{R}^{n_1}, y \in \mathbb{R}^{n_2}, 1 = \|x\|_2 = \|y\|_2\}$$

induces  $\|\cdot\|_{\mathcal{L}} = \|\cdot\|_1 : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}_+$ .

In their approach, the notion of low-complexity is embedded in the choice of the collection of atoms  $\mathcal{T}$ , which would be dictated by the problem at hand. The studied recovery procedure is given by the convex optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|x\|_{\mathcal{T}} \quad \text{subject to} \quad \|b - Ax\|_2 \leq \epsilon, \tag{1.4}$$

where  $0 \leq \epsilon < \|b\|_2$  prescribes the admissible measurement misfit away from the vector of observations  $b \in \mathbb{R}^m$ .

Assuming measurement maps  $A \in \mathbb{R}^{m \times n}$  whose entries are drawn i.i.d. from the standard Gaussian distribution, Chandrasekaran et al. (2012) provide probabilistic recoverability and stability bounds similar to those discussed in §1.1 for a range of different “low-complexity” models generalized by the concept of atomic norm and the solution of the convex optimization problem (1.4). For sparse and low-rank recovery, they provide much tighter lower bounds on the number of measurements sufficient for the effectiveness of the relaxation.

Candès and Recht (2013) also leveraged this concept of atomic norms and their dual-norms to provide a unified treatment of recoverability bounds from noiseless linear measurements for both (block) sparse vectors and low-rank matrices.

As mentioned by Chandrasekaran et al. (2012), most of their developments are applicable even in some cases where the collection of atoms  $\mathcal{T}$  does not induce a norm—e.g., sparse nonnegative vectors and low-rank positive semidefinite matrices. In such scenarios, convex problems of the form (1.4) are particular cases of a general class of *gauge optimization* studied by Freund (1987).

In his seminal paper, Freund (1987) develops a duality framework around a class of optimization problems that generalizes the minimization of a  $p$ -norm, convex QPs, and linear programs with nonnegative optimal values.

This class of problems can be succinctly described with the formulation

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad x \in \mathcal{C}, \tag{1.5}$$

where  $\mathcal{X}$  is a finite-dimensional real inner-product space,  $\kappa : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a *gauge* function—i.e., a nonnegative convex function that is positively homogeneous and satisfies  $\kappa(0) = 0$ —and  $\mathcal{C} \subseteq \mathcal{X}$  is a closed convex set.

**Remark 1.2.1.** It must be noted that we will often assume  $\kappa$  to be closed (i.e., its epigraph  $\text{epi } \kappa := \{(x, t) \in \mathcal{X} \times \mathbb{R} \mid \kappa(x) \leq t\}$  is a closed set) and that  $\mathcal{C}$  does not contain the origin, otherwise  $x = 0$  would be a solution of (1.5).

As can be readily inferred from definitions, the class of gauge functions subsumes all norms and seminorms. Generalizing the notion of dual norms, we have that of a *polar* gauge  $\kappa^\circ : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  defined by

$$\kappa^\circ(y) := \inf\{t > 0 \mid t\kappa(x) \geq \langle x, y \rangle, \forall x \in \mathcal{X}\}, \quad (1.6)$$

giving rise to a natural generalization of the Cauchy-Schwartz inequality

$$\langle x, y \rangle \leq \kappa(x)\kappa^\circ(y), \quad \forall x, y \in \mathcal{X}, \quad (1.7)$$

which holds whenever the product makes sense, i.e.,  $\{\kappa(x), \kappa^\circ(y)\} \neq \{0, +\infty\}$ .

Although motivated by problems in which  $\mathcal{C}$  is polyhedral, Freund (1987) proposed an approach to handle general nonlinear constraints based on a concept of the *antipolar* set to  $\mathcal{C}$ , denoted by  $\mathcal{C}' \subset \mathcal{X}$  and defined as

$$\mathcal{C}' := \{y \in \mathcal{X} \mid \langle x, y \rangle \geq 1, \forall x \in \mathcal{C}\}. \quad (1.8)$$

From its definition, it can be readily observed that this set is closed and convex and does not contain the origin (however it might be empty).

An interesting consequence of the concepts of polar gauges and antipolar sets is that they lead to a certain duality relationship represented by

$$x \in \mathcal{C} \text{ and } y \in \mathcal{C}' \implies 1 \leq \kappa(x)\kappa^\circ(y), \quad (1.9)$$

naturally leading to a gauge minimization problem of the form

$$\underset{y \in \mathcal{X}}{\text{minimize}} \quad \kappa^\circ(y) \quad \text{subject to} \quad y \in \mathcal{C}', \quad (1.10)$$

which is the corresponding *dual* to (1.5) proposed by Freund (1987).

In his study of duality relationships between the gauge optimization problems (1.5) and (1.10), Freund (1987) provides a general weak duality result for that primal-dual pair. In his theorem, compiled below, the *domain* of a function  $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ , denoted by  $\text{dom } f$ , is defined as

$$\text{dom } f := \{x \in \mathcal{X} \mid f(x) < +\infty\},$$

i.e., it corresponds to the set of points at which the function is finite-valued.

**Theorem 1.2.1** (Freund (1987, Theorem 1A (Weak Duality))).

Let  $p^* \in \mathbb{R}_+ \cup \{+\infty\}$  and  $d^* \in \mathbb{R}_+ \cup \{+\infty\}$  be optimal values for (1.5) and (1.10), respectively. Then

- (i) If  $x \in \mathcal{C} \cap \text{dom } \kappa$  and  $y \in \mathcal{C}' \cap \text{dom } \kappa^\circ$ , then  $\kappa(x)\kappa^\circ(y) \geq 1$ , and hence  $p^*d^* \geq 1$ .
- (ii) If  $p^* = 0$ , then (1.10) is *essentially infeasible*, i.e.,  $d^* = +\infty$ .
- (iii) If  $d^* = 0$ , then (1.5) is *essentially infeasible*, i.e.,  $p^* = +\infty$ .
- (iv) If  $x^* \in \mathcal{C} \cap \text{dom } \kappa$  and  $y^* \in \mathcal{C}' \cap \text{dom } \kappa^\circ$  and  $\kappa(x^*)\kappa^\circ(y^*) = 1$ , then  $x^*$  and  $y^*$  are optimal solutions of (1.5) and (1.10), respectively.

This previous result is a direct consequence of (1.9) and it illustrates the multiplicative notion of duality in gauge optimization, which is different from the usual additive one from Lagrangian duality in general convex optimization.

For problems in which  $\mathcal{C}$  is polyhedral, Freund (1987) introduces a qualification condition for strong duality to hold.

**Definition 1.2.2** (Projection property and qualification). A convex set  $S \subseteq \mathcal{X}$  satisfies the *projection property* if, for all linear maps  $A : \mathcal{X} \rightarrow \mathbb{R}^m$ , we have that  $AS = \{Ax \mid x \in S\}$  is closed.

A gauge function  $\kappa : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  satisfies the *projection qualification* if both  $\{x \in \mathcal{X} \mid \kappa(x) \leq 1\}$  and  $\{y \in \mathcal{X} \mid \kappa^\circ(y) \leq 1\}$  satisfy the projection property.

With these concepts, the following strong duality result can be proved.

**Theorem 1.2.2** (Freund (1987, Theorem 2)). Assume that  $\kappa$  satisfies the projection qualification and  $\mathcal{C}$  is polyhedral. Let  $p^* \in \mathbb{R}_+ \cup \{+\infty\}$  and  $d^* \in \mathbb{R}_+ \cup \{+\infty\}$  be optimal values for (1.5) and (1.10), respectively. Then

- (i) If  $\mathcal{C} \cap \text{dom } \kappa \neq \emptyset$  and  $\mathcal{C}' \cap \text{dom } \kappa^\circ \neq \emptyset$ , then  $p^*d^* = 1$ ; and the optimal values of (1.5) and (1.10) are achieved for some  $x^*$  and  $y^*$ .
- (ii)  $\mathcal{C} \cap \text{dom } \kappa = \emptyset$  (i.e.,  $p^* = +\infty$ ) if and only if  $d^* = 0$ ; and  $d^* = 0$  is achieved for some  $y^* \in \mathcal{C}'$  if  $\mathcal{C} = \emptyset$ .
- (iii)  $\mathcal{C}' \cap \text{dom } \kappa^\circ = \emptyset$  (i.e.,  $d^* = +\infty$ ) if and only if  $p^* = 0$ ; and  $p^* = 0$  is achieved for some  $x^* \in \mathcal{C}$  if  $\mathcal{C}' = \emptyset$ .

Freund (1987) provides alternative strong duality results under qualifications similar to the classical Slater conditions, i.e., when the intersections in Theorem 1.2.2 (i) involve relative interiors. Later in §2.4 we provide more refined sufficient conditions for strong gauge duality in an abstract setting.

These results were used by Freund (1987) to show that the gauge dual for the minimization of strictly convex  $p$ -norms provides tighter lower bounds when compared to the corresponding Lagrange dual. Other authors have exploited these for applications ranging from proving robust stability of linear dynamical systems to approximate Farkas Lemmas in linear programming and tight inequalities in probability; see Teboulle and Kogan (1994); Todd and Ye (1998); Bertsimas and Popescu (2005).

Our motivation is the natural gauge optimization form of the convex relaxations to inverse problems in low-complexity recovery. For these problems, the convex relaxation typically has the form

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad \rho(b - Ax) \leq \epsilon, \quad (1.11)$$

where  $A : \mathcal{X} \rightarrow \mathbb{R}^m$  is a linear map and  $\rho : \mathbb{R}^m \rightarrow \mathbb{R}$  is also a gauge, modeling the measurement misfit and typically a  $p$ -norm.

As we shall see later in Chapter 2, the corresponding dual gauge problem has the form

$$\underset{y \in \mathbb{R}^m}{\text{minimize}} \quad \kappa^\circ(A^*y) \quad \text{subject to} \quad \langle b, y \rangle - \epsilon\rho(y) \geq 1, \quad (1.12)$$

which can be contrasted to the Lagrange dual problem

$$\underset{y \in \mathbb{R}^m}{\text{maximize}} \quad \langle b, y \rangle - \epsilon\rho(y) \quad \text{subject to} \quad \kappa^\circ(A^*y) \leq 1. \quad (1.13)$$

In the applications that motivate our work,  $\rho$  is typically rather simple while  $\kappa$  has a more complicated structure. Our approach described later in Chapter 4 leverages the fact that  $m$  is expected to be much smaller than  $\dim \mathcal{X}$  to solve the dual problem (1.12), whose feasible set is much simpler and is thus more amenable to computational approaches for nonsmooth convex optimization than its Lagrangian counterpart (1.13).

**Examples.** The following self-contained examples, some condensed from our previous and later discussions, illustrate the versatility of the gauge optimization formulation in modeling a variety of problems, and also provide a motivation for the abstract treatment that we give to gauge duality in the following Chapter 2.

**Example 1.2.1** (Norms and minimum-length solutions). Norms are special cases of gauge functions that are finite everywhere, symmetric, and zero only at the origin. (Semi-norms drop the last requirement, and allow the function to be zero at other points.) Let  $\kappa(x) = \|x\|$  be any norm, and  $\mathcal{C} = \{x \mid Ax = b\}$  describe the solutions to an underdetermined linear system. Then (1.5) yields a minimum-length solution to the linear system  $Ax = b$ . This problem can be modeled as an instance of (1.11) by letting  $\rho$  be any gauge function for which  $\rho^{-1}(0) = \{0\}$  and setting  $\epsilon = 0$ . The polar  $\kappa^\circ = \|\cdot\|_*$  is the norm dual to  $\|\cdot\|$ , and  $\mathcal{C}' = \{A^*y \mid \langle b, y \rangle \geq 1\}$ ; cf. Corollary 2.3.2. The corresponding gauge dual (1.10) is then

$$\underset{y \in \mathcal{Y}}{\text{minimize}} \quad \|A^*y\|_* \quad \text{subject to} \quad \langle b, y \rangle \geq 1.$$

**Example 1.2.2** (Sparse optimization and atomic norms). In his thesis, van den Berg (2009) describes a framework for sparse optimization based on the formulation where  $\kappa$  is a gauge, and the function  $\rho$  is differentiable away from the origin. The nonnegative regularization parameter  $\epsilon$  influences the degree to which the linear model  $Ax$  fits the observations  $b$ . This formulation is specialized by van den Berg and Friedlander (2011) to the particular case in which  $\rho$  is the 2-norm. In that case,  $\mathcal{C} = \{x \mid \|Ax - b\|_2 \leq \epsilon\}$  and

$$\mathcal{C}' = \{A^*y \mid \langle b, y \rangle - \epsilon \|y\|_2 \geq 1\};$$

cf. Corollary 2.3.1. Teuber, Steidl, and Chan (2013) consider a related case where the misfit between the model and the observations is measured by the Kullback-Leibler divergence.

Chandrasekaran et al. (2012) describe how to construct regularizers that generalize the notion of sparsity in linear inverse problems. In particular, they define the gauge

$$\|x\|_{\mathcal{T}} := \inf \{t \geq 0 \mid x \in t \text{conv } \mathcal{T}\} \tag{1.14}$$

over the convex hull of the set of canonical atoms given by the set  $\mathcal{T}$ . If  $0 \in \text{int conv } \mathcal{T}$  and  $\mathcal{T}$  is bounded and symmetric, i.e.,  $\mathcal{T} = -\mathcal{T}$ , then the definition (1.14) yields a norm. For example, if  $\mathcal{T}$  consists of the set of unit  $n$ -vectors that contain a single nonzero element, then (1.14) is the 1-norm; if  $\mathcal{T}$  consists of the set of rank-1 matrices with unit spectral norm, then (1.14) is the Schatten 1-norm. The polar  $\kappa^\circ(y) = \sup \{\langle y, a \rangle \mid a \in \text{conv}(\{0\} \cup \mathcal{T})\}$  is the support function of the closure of  $\text{conv}(\{0\} \cup \mathcal{T})$ . Jaggi (2013) catalogs various sets of atoms that yield commonly used gauges in machine learning.

**Example 1.2.3** (Conic gauge optimization). In this example we demonstrate that it is possible to cast any convex conic optimization problem in the gauge framework. Let  $\mathcal{K}$  be a closed convex cone, and let  $\mathcal{K}^*$  denote its dual. Consider the primal-dual pair of feasible conic problems:

$$\underset{x}{\text{minimize}} \quad \langle c, x \rangle \quad \text{subject to} \quad Ax = b, \quad x \in \mathcal{K}, \quad (1.15a)$$

$$\underset{y}{\text{maximize}} \quad \langle b, y \rangle \quad \text{subject to} \quad c - A^*y \in \mathcal{K}^*. \quad (1.15b)$$

Suppose that  $\hat{y}$  is a dual-feasible point, and define  $\hat{c} = c - A^*\hat{y}$ . Because  $\hat{c} \in \mathcal{K}^*$ , it follows that  $\langle \hat{c}, x \rangle \geq 0$  for all  $x \in \mathcal{K}$ . In particular, the primal problem can be equivalently formulated as a gauge optimization problem by defining

$$\kappa(x) = \langle \hat{c}, x \rangle + \delta_{\mathcal{K}}(x) \quad \text{and} \quad \mathcal{C} = \{x \mid Ax = b\}, \quad (1.16)$$

where  $\delta_{\mathcal{K}}$  is the indicator function on the set  $\mathcal{K}$ . (More generally, it is evident that any function of the form  $\gamma + \delta_{\mathcal{K}}$  is a gauge if  $\gamma$  is a gauge.) This formulation is a generalization of the nonnegative linear program discussed by Freund, and we refer to it as conic gauge optimization. The generalization captures some important problem classes, such as trace minimization of positive semidefinite (PSD) matrices, which arises in the phase-retrieval problem (c.f. §§1.1.1). This is an example where  $c \in \mathcal{K}^*$ , in which case the dual-feasible point  $\hat{y} = 0$  is trivially available for the gauge reformulation; cf. §3.1.

**Example 1.2.4** (Semidefinite programming relaxation for MAX-CUT).

A concrete example of conic gauge programming, in the simple case where  $c \in \mathcal{K}^*$ , is the semidefinite programming relaxation of the max-cut problem studied by Goemans and Williamson (1995). Let  $G = (V, E)$  be an undirected graph, and  $D = \text{diag}((d_v)_{v \in V})$ , where  $d_v$  denotes the degree of vertex  $v \in V$ . The max-cut problem can be written as

$$\underset{x}{\text{maximize}} \quad \frac{1}{4} \langle D - A, xx^T \rangle \quad \text{subject to} \quad x \in \{-1, 1\}^V,$$

where  $A$  denotes the adjacency matrix associated with  $G$ . The semidefinite programming relaxation for this problem is derived by lifting  $xx^T$  into a PSD matrix, as we have done for phase retrieval in §1.1.1:

$$\underset{X}{\text{maximize}} \quad \frac{1}{4} \langle D - A, X \rangle \quad \text{subject to} \quad \text{diag } X = e, \quad X \succeq 0,$$

where  $e$  denotes the vector of all ones. The constraint  $\text{diag } X = e$  implies that  $\langle D, X \rangle = \sum_{v \in V} d_v = 2|E|$  is constant. Thus, the optimal value is equal to

$$|E| - \frac{1}{4} \cdot \min_X \{ \langle D + A, X \rangle \mid \text{diag } X = e, X \succeq 0 \}, \quad (1.17)$$

and the solution can be obtained by solving this latter problem. Note that  $D + A$  is PSD because it has nonnegative diagonals and is diagonally dominant. (In fact, it is possible to reduce the problem in linear time to one where  $D + A$  is positive definite by identifying its bipartite connected components.) Because the dual of the cone of PSD matrices is itself, and the trace inner product between PSD matrices is nonnegative, (1.17) falls into the class of conic gauge problems defined by (1.15a).

**Example 1.2.5** (Submodular functions). Let  $V = \{1, \dots, n\}$ , and consider the set-function  $f : 2^V \rightarrow \mathbb{R}$ , where  $f(\emptyset) = 0$ . The Lovász (1983) extension  $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{R}$  of  $f$  is given by

$$\hat{f}(x) = \sum_{k=1}^n x_{j_k} [f(\{j_1, \dots, j_k\}) - f(\{j_1, \dots, j_{k-1}\})],$$

where  $x_{j_1} \geq x_{j_2} \geq \dots \geq x_{j_n}$  are the sorted elements of  $x$ . Clearly, the extension is positively homogeneous and vanishes at the origin. As shown by Lovász, the extension is convex if and only if  $f$  is *submodular*, i.e.,

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad \text{for all } A, B \subset V;$$

see also (Bach, 2013, Proposition 2.3). If  $f$  is additionally non-decreasing, i.e.,

$$A, B \subset V \text{ and } A \subset B \implies f(A) \leq f(B),$$

then the extension is nonnegative over  $\mathbb{R}_+^n$ . Thus, when  $f$  is a submodular and non-decreasing set function, that function plus the indicator on the nonnegative orthant, i.e.,  $\hat{f} + \delta_{\mathbb{R}_+^n}$ , is a gauge. Bach (2013) surveys the properties of submodular functions and their application in machine learning; see Proposition 3.7 therein.

## 1.3 Thesis overview and contributions

The main body of this thesis is comprised of four chapters following the present introduction. In Chapter 2 we present the theoretical framework of gauge optimization and extend its duality theory with a number of results useful for modeling and deriving duals with focus given to problems involving gauge-constrained linear measurements, an abstraction of the types of measurements frequently encountered in the resulting convex relaxations of nonlinear inverse problems that motivated this research.

Chapter 3 specializes this abstract gauge duality framework to problems involving matrix variables. It connects the spectral optimization problems of trace minimization in the PSD cone and nuclear-norm minimization to a concrete primal-dual pair of constrained gauge minimization problems. The main result is the formulation of a dual convex eigenvalue optimization problem, whose feasible set is defined by a rather simple constraint and is amenable to matrix-free numerical methods, a requirement for scalability to large problems.

Chapter 4 completes the abstract to concrete flow of this work with an analysis of the strong duality and optimality conditions of the primal PSD trace minimization and dual constrained eigenvalue minimization problems. These conditions induce an approach for recovering low-rank primal solutions from dual minimizers and are exploited to design a computational method to solve the primal-dual gauge pair. A description of the main algorithmic components of our proof-of-concept solver is presented and extensions of our analysis to the reweighted formulations by Candès et al. (2013a) and Mohan and Fazel (2010) conclude that chapter.

The numerical experiments in Chapter 5 contrast the results of our implementation to those obtained by solvers designed for each of the two main problems we study. For the PhaseLift formulation of phase recovery, our results are compared against those of TFOCS (Becker et al., 2011; Candès et al., 2013b) and Wirtinger Flow (Candès et al., 2015a) on a number of random problems and a large two-dimensional image, used to validate the ability of our approach to solve problems of a practical scale. An instance of blind deconvolution for motion deblurring of a two-dimensional image is used to compare our solver's results to those obtained by the augmented Lagrangian solver proposed in (Ahmed et al., 2014) and assess its feasibility in solving nuclear norm minimization problems of such scale.

We conclude this work in Chapter 6 with a discussion of our developments along with our perspective of interesting avenues for further investigation.

## **Reproducible research**

The data files and MATLAB scripts used to generate the tables and figures from all numerical experiments presented in Chapter 5 can be obtained at:

<http://www.cs.ubc.ca/labs/scl/low-rank-opt>

# Chapter 2

## Gauge optimization and duality

Gauge functions significantly generalize the notion of a norm, and gauge optimization, as defined by Freund (1987), seeks the element of a convex set that is minimal with respect to a gauge function. This conceptually simple problem can be used to model a remarkable array of useful problems that arise in a range of fields (see examples in §1.2). The gauge structure of these problems allows for a special kind of duality framework which we explore and specialize to a particular form that exposes some useful properties of the gauge optimization framework (such as the variational properties of its value function), and yet maintains most of the generality of the abstract form shown in §1.2.

As observed in the TFOCS-based method discussed in Chapter 1, one approach to solving linear inverse problems is to optimize a regularization function over the set of admissible deviations between the observations and the forward model. Although regularization functions come in a wide range of forms depending on the particular application, they often share some common properties. In this chapter we describe and study the class of gauge optimization problems, which neatly captures a wide variety of regularization formulations that arise from the convex relaxations described in Chapter 1 and others from fields such as machine learning and inverse problems. We explore the duality and variational properties particular to this family of problems.

All of the problems that we consider can be expressed as

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad x \in \mathcal{C}, \quad (\text{P})$$

where  $\mathcal{X}$  is a finite-dimensional Euclidean space,  $\mathcal{C} \subseteq \mathcal{X}$  is a closed convex set, and  $\kappa : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a *gauge* function, i.e., a nonnegative, positively homogeneous convex function that vanishes at the origin. (We assume that  $0 \notin \mathcal{C}$ , since otherwise the origin is trivially a solution of the problem.) This class of problems admits a duality relationship that is different from Lagrange duality, and is founded on the gauge structure of its objective. Indeed, Freund (1987) defines the dual gauge counterpart to be

$$\underset{y \in \mathcal{X}}{\text{minimize}} \quad \kappa^\circ(y) \quad \text{subject to} \quad y \in \mathcal{C}', \quad (\text{D})$$

where the set

$$\mathcal{C}' := \{ y \mid \langle y, x \rangle \geq 1 \text{ for all } x \in \mathcal{C} \} \quad (2.1)$$

is the antipolar of  $\mathcal{C}$  (in contrast to the better-known polar of a convex set), and the polar  $\kappa^\circ$  (also a gauge) is the function that satisfies the Cauchy-Schwartz-like inequality most tightly:

$$\langle x, y \rangle \leq \kappa(x) \kappa^\circ(y), \quad \forall x \in \text{dom } \kappa, \forall y \in \text{dom } \kappa^\circ; \quad (2.2)$$

see (2.5) for the precise definition. It follows directly from this inequality and the definition of  $\mathcal{C}'$  that all primal-dual feasible pairs  $(x, y)$  satisfy the weak-duality relationship

$$1 \leq \kappa(x) \kappa^\circ(y), \quad \forall x \in \mathcal{C} \cap \text{dom } \kappa, \forall y \in \mathcal{C}' \cap \text{dom } \kappa^\circ. \quad (2.3)$$

This duality relationship stands in contrast to the more usual Lagrange framework, where the primal and dual objective values bound each other in an additive sense.

**A roadmap of our developments.** Freund's analysis of gauge duality is mainly concerned with specialized linear and quadratic problems that fit into the gauge framework, and with the pair of abstract problems (P) and (D). Our treatment in this chapter considers the particular formulation of (P) given by

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad \rho(b - Ax) \leq \epsilon, \quad (\text{P}_\rho)$$

where  $\rho$  is also a gauge. Typical applications might use  $\rho$  to measure the mismatch between the model  $Ax$  and the measurements  $b$ , and in that case, it is natural to assume that  $\rho$  vanishes only at the origin, so that the constraint reduces to  $Ax = b$  when  $\epsilon = 0$ . This formulation is only very slightly less general than (P) because any closed convex set can be represented as  $\{x \mid \rho(b - x) \leq 1\}$  for some vector  $b$  and gauge  $\rho$ ; cf. §2.1.2. However, it is sufficiently concrete that it allows us to develop a calculus for computing gauge duals for a wide range of existing problems. (Conic side constraints and a linear map in the objective can be easily accommodated; this is covered in §2.6.)

The special structure of the functions in the gauge program  $(\text{P}_\rho)$  leads to a duality framework that is analogous to the classical Lagrange-duality framework. The gauge program dual to  $(\text{P}_\rho)$  is given by

## 2.1. Preliminaries

---

$$\underset{y \in \mathcal{X}}{\text{minimize}} \quad \kappa^\circ(A^*y) \quad \text{subject to} \quad \langle y, b \rangle - \epsilon \rho^\circ(y) \geq 1, \quad (\text{D}_\rho)$$

which bears a striking similarity to the Lagrange dual problem

$$\underset{y \in \mathcal{X}}{\text{maximize}} \quad \langle y, b \rangle - \epsilon \rho^\circ(y) \quad \text{subject to} \quad \kappa^\circ(A^*y) \leq 1. \quad (\text{D}_\ell)$$

Note that the objective and constraints between the two duals play different roles. (These two duals are derived in §2.3 under suitable assumptions.) A significant practical difference between these two formulations is when  $\rho$  is a simple Euclidean norm and  $\kappa$  is a more complicated function (such as the one described in Example 1.2.2). The result is that the Lagrange dual optimizes a “simple” objective function over a potentially “complicated” constraint; in contrast, the situation is reversed in the gauge optimization formulation.

We develop in §2.2 an antipolar calculus for computing the antipolars of sets such as  $\{x \mid \rho(b - Ax) \leq \epsilon\}$ , which corresponds to the constraint in our canonical formulation  $(\text{P}_\rho)$ . This calculus is applied in §2.3 to derive the gauge dual  $(\text{D}_\rho)$ .

The formal properties of the polar and antipolar operations are described in §§2.1–2.2. In §2.4 we develop conditions sufficient for strong duality, i.e., for there to exist a primal-dual pair that satisfies (2.3) with equality. Our derivation parts with the “ray-like” assumption used by Freund, and in certain cases further relaxes the required assumptions by leveraging connections with established results from convex analysis and Fenchel duality.

## 2.1 Preliminaries

In this section we review known facts about polar sets, gauges and their polars, and introduce results that are useful for our subsequent analysis. We mainly follow Rockafellar (1970) and use a number of results from his monograph: see §14 in that text for a discussion of polarity operations on convex sets, and §15 for a discussion of gauge functions and their corresponding polarity operations. A few results from other books are partially compiled and credited when needed for convenient reference.

We use the following notation throughout. For a closed convex set  $\mathcal{D}$ , let  $\text{ri } \mathcal{D}$  and  $\text{cl } \mathcal{D}$  denote, respectively, the relative interior and the closure of  $\mathcal{D}$ . The indicator function of the set  $\mathcal{D}$  is denoted by  $\delta_{\mathcal{D}}$  and defined as  $\delta_{\mathcal{D}}(x) = 0$ , if  $x \in \mathcal{D}$ , and  $\delta_{\mathcal{D}}(x) = +\infty$ , otherwise. The recession cone of  $\mathcal{D}$  is defined below; see Auslender and Teboulle (2003, Definition 2.1.2).

**Definition 2.1.1** (Recession cone). Let  $\mathcal{D}$  be a nonempty set in  $\mathcal{X}$ . Then the *recession* (or *asymptotic*) *cone* of the set  $\mathcal{D}$ , denoted by  $\mathcal{D}_\infty$ , is the set of vectors  $d \in \mathcal{X}$  that are limits in direction of the sequences  $\{x_k\} \subset \mathcal{D}$ , namely

$$\mathcal{D}_\infty := \left\{ d \in \mathcal{X} \mid \exists t_k \rightarrow +\infty, \exists x_k \in \mathcal{D} \text{ with } \lim_{k \rightarrow +\infty} \frac{x_k}{t_k} = d \right\}.$$

For a gauge  $\kappa : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ , its domain is denoted by  $\text{dom } \kappa = \{x \mid \kappa(x) < +\infty\}$ , and its epigraph is denoted by  $\text{epi } \kappa = \{(x, \mu) \mid \kappa(x) \leq \mu\}$ . A function is called closed if its epigraph is closed, which is equivalent to the function being lower semi-continuous; c.f. Rockafellar (1970, Theorem 7.1). Let  $\text{cl } \kappa$  denote the gauge whose epigraph is  $\text{cl epi } \kappa$ , which is the largest lower semi-continuous function smaller than  $\kappa$ ; see Rockafellar (1970, p. 52). Finally, for any  $x \in \text{dom } \kappa$ , the subdifferential of  $\kappa$  at  $x$  is denoted  $\partial\kappa(x) = \{y \mid \kappa(u) \geq \kappa(x) + \langle y, u - x \rangle, \forall u\}$ .

We make the following blanket assumptions throughout. The set  $\mathcal{C}$  is a nonempty closed convex set that does not contain the origin; the set  $\mathcal{D}$  is a nonempty convex set that may or may not contain the origin, depending on the context. The gauge function  $\rho : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ , used in  $(P_\rho)$ , is closed; when  $\epsilon = 0$ , we additionally assume that  $\rho^{-1}(0) = \{0\}$  essentially to ensure that  $\{x \mid \rho(b - Ax) \leq 0\} = \{x \mid Ax = b\}$ , as would be expected in the case of a zero tolerance on the mismatch between  $Ax$  and the observations  $b$ .

### 2.1.1 Polar sets

The *polar* of a nonempty closed convex set  $\mathcal{D}$  is defined as

$$\mathcal{D}^\circ := \{y \mid \langle x, y \rangle \leq 1, \forall x \in \mathcal{D}\},$$

which is necessarily closed convex, and contains the origin.

The *bipolar theorem* states that if  $\mathcal{D}$  is closed, then it contains the origin if and only if  $\mathcal{D} = \mathcal{D}^{\circ\circ}$ ; see Rockafellar (1970, Theorem 14.5).

When  $\mathcal{D} = \mathcal{K}$  is a closed convex cone, the polar is equivalently given by

$$\mathcal{K}^\circ := \{y \mid \langle x, y \rangle \leq 0, \forall x \in \mathcal{K}\}.$$

The positive polar cone (also known as the *dual cone*) of  $\mathcal{D}$  is given by

$$\mathcal{D}^* := \{y \mid \langle x, y \rangle \geq 0, \forall x \in \mathcal{D}\}.$$

The polar and positive polar are related via the closure of the conic hull, i.e.,

$$\mathcal{D}^* = (\text{cl cone } \mathcal{D})^* = -(\text{cl cone } \mathcal{D})^\circ, \quad \text{where} \quad \text{cone } \mathcal{D} := \bigcup_{\lambda \geq 0} \lambda \mathcal{D}.$$

### 2.1.2 Gauge functions

All gauges can be represented in the form of a *Minkowski function*  $\gamma_{\mathcal{D}}$  of some nonempty convex set  $\mathcal{D}$ , i.e.,

$$\kappa(x) = \gamma_{\mathcal{D}}(x) := \inf \{ \lambda \geq 0 \mid x \in \lambda \mathcal{D} \}. \quad (2.4)$$

In particular, one can always choose  $\mathcal{D} = \{x \mid \kappa(x) \leq 1\}$ , and the above representation holds. The *polar of the gauge*  $\kappa$  is defined by

$$\kappa^\circ(y) := \inf \{ \mu > 0 \mid \langle x, y \rangle \leq \mu \kappa(x), \forall x \}, \quad (2.5)$$

which explains the inequality (2.2). Because  $\kappa$  is a proper convex function, one can also define its *convex conjugate*:

$$\kappa^*(y) := \sup_x \{ \langle x, y \rangle - \kappa(x) \}. \quad (2.6)$$

It is well known that  $\kappa^*$  is a proper closed convex function (Rockafellar, 1970, Theorem 12.2). The next fact is used in the proof of the proposition following it and compiled from Auslender and Teboulle (2003, Proposition 3.1.3) for convenient reference.

**Proposition 2.1.1.** Let  $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  be proper closed and convex. Then  $0 \in \text{int dom } f^*$  if and only if the optimal set  $\{x \mid f(x) = \inf f\}$  is nonempty and compact.

The next result collects properties relating the polar and conjugate of a gauge.

**Proposition 2.1.2.** For the gauge  $\kappa : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ , it holds that

- (i)  $\kappa^\circ$  is a closed gauge function;
- (ii)  $\kappa^{\circ\circ} = \text{cl } \kappa = \kappa^{**}$ ;
- (iii)  $\kappa^\circ(y) = \sup_x \{ \langle x, y \rangle \mid \kappa(x) \leq 1 \}$  for all  $y$ ;
- (iv)  $\kappa^*(y) = \delta_{\kappa^\circ(\cdot) \leq 1}(y)$  for all  $y$ ;
- (v)  $\text{dom } \kappa^\circ = \mathcal{X}$  if  $\kappa$  is closed and  $\kappa^{-1}(0) = \{0\}$ ;
- (vi)  $\text{epi } \kappa^\circ = \{ (y, \lambda) \mid (y, -\lambda) \in (\text{epi } \kappa)^\circ \}$ .

*Proof.* The first two items are proved in Theorems 15.1 and 12.2 of Rockafellar (1970). Item (iii) follows directly from the definition (2.5) of the polar gauge. To prove item (iv), we note that if  $g(t) = t$ ,  $t \in \mathbb{R}$ , then the so-called monotone conjugate  $g^+$  is

$$g^+(s) = \sup_{t \geq 0} \{ st - t \} = \delta_{[0,1]}(s),$$

where  $s \geq 0$ . Now, apply Rockafellar (1970, Theorem 15.3) with  $g(t) = t$ , and  $\kappa^{**}$  in place of  $f$  in that theorem to obtain that  $\kappa^{***}(y) = \delta_{[0,1]}(\kappa^{**\circ}(y))$ . The conclusion in item (iv) now follows by noting that  $\kappa^{***} = \kappa^*$  and  $\kappa^{**\circ} = \kappa^{\circ\circ} = \kappa^\circ$ . To prove item (v), note that the assumptions together with Proposition 2.1.1 show that  $0 \in \text{int dom } \kappa^*$ . This together with item (iv) and the positive homogeneity of  $\kappa^\circ$  shows that  $\text{dom } \kappa^\circ = \mathcal{X}$ . Finally, item (vi) is stated on Rockafellar (1970, p. 137) and can also be verified directly from the definition.  $\square$

In many interesting applications, the objective in (P) is the composition  $\kappa \circ A$ , where  $\kappa$  is a gauge and  $A$  is a linear map. Clearly,  $\kappa \circ A$  is also a gauge. The next result gives the polar of this composition.

**Proposition 2.1.3.** Let  $A$  be a linear map. Suppose that either

- (i)  $\text{epi } \kappa$  is polyhedral; or
- (ii)  $\text{ri dom } \kappa \cap \text{range } A \neq \emptyset$ .

Then

$$(\kappa \circ A)^\circ(y) = \inf_u \{ \kappa^\circ(u) \mid A^*u = y \}.$$

Moreover, the infimum is attained when the value is finite.

*Proof.* Since  $\kappa \circ A$  is a gauge, we have from Proposition 2.1.2(iii) that

$$(\kappa \circ A)^\circ(y) = \sup_x \{ \langle y, x \rangle \mid \kappa(Ax) \leq 1 \} = - \inf_x \{ \langle -y, x \rangle + \delta_{\mathcal{D}}(Ax) \},$$

where  $\mathcal{D} = \{ x \mid \kappa(x) \leq 1 \}$ . Since  $\kappa$  is positively homogeneous, we have  $\text{dom } \kappa = \bigcup_{\lambda \geq 0} \lambda \mathcal{D}$ . Hence,  $\text{ri dom } \kappa = \bigcup_{\lambda > 0} \lambda \text{ri } \mathcal{D}$  from Rockafellar (1970, p. 50). Thus, assumption (ii) implies that  $\text{ri } \mathcal{D} \cap \text{range } A \neq \emptyset$ . On the other hand, assumption (i) implies that  $\mathcal{D}$  is polyhedral; and  $\mathcal{D} \cap \text{range } A \neq \emptyset$  because they both contain the origin. Use these conclusions and apply Rockafellar (1970, Corollary 31.2.1) (see also Rockafellar's remark right after that

corollary for the case when  $\mathcal{D}$  is polyhedral) to conclude that

$$\begin{aligned} (\kappa \circ A)^\circ(y) &= -\sup_u \{ -(\langle -y, \cdot \rangle)^*(-A^*u) - (\delta_{\mathcal{D}})^*(u) \} \\ &= -\sup_u \{ -\kappa^\circ(u) \mid A^*u = y \}, \end{aligned}$$

where the second equality follows from the definition of conjugate functions and Proposition 2.1.2(iii). Moreover, from that same corollary, the supremum is attained when finite. (Note that Rockafeller's statement of that corollary is formulated for the difference between convex and concave function, and must be appropriately adapted to our case.) This completes the proof.  $\square$

Suppose that a gauge is given as the Minkowski function of a nonempty convex set that may not necessarily contain the origin. The following proposition summarizes some properties concerning this representation.

**Proposition 2.1.4.** Suppose that  $\mathcal{D}$  is a nonempty convex set. Then

- (i)  $(\gamma_{\mathcal{D}})^\circ = \gamma_{\mathcal{D}^\circ}$ ;
- (ii)  $\gamma_{\mathcal{D}} = \gamma_{\text{conv}(\{0\} \cup \mathcal{D})}$ ;
- (iii) If  $\text{conv}(\{0\} \cup \mathcal{D})$  is closed, then  $\gamma_{\mathcal{D}}$  is closed;
- (iv) If  $\kappa = \gamma_{\mathcal{D}}$ ,  $\mathcal{D}$  is closed, and  $0 \in \mathcal{D}$ , then  $\mathcal{D}$  is the unique closed convex set containing the origin such that  $\kappa = \gamma_{\mathcal{D}}$ ; indeed,  $\mathcal{D} = \{x \mid \kappa(x) \leq 1\}$ .

*Proof.* Item (i) is proved in Rockafellar (1970, Theorem 15.1). Item (ii) follows directly from the definition. To prove (iii), we first notice from item (ii) that we may assume without loss of generality that  $\mathcal{D}$  contains the origin. Notice also that  $\gamma_{\mathcal{D}}$  is closed if and only if  $\gamma_{\mathcal{D}} = \gamma_{\mathcal{D}}^{**}$ . Moreover,  $\gamma_{\mathcal{D}}^{**} = \gamma_{\mathcal{D}^{\circ\circ}} = \gamma_{\text{cl}\mathcal{D}}$ , where the first equality follows from Proposition 2.1.2(ii) and item (i), while the second equality follows from the bipolar theorem. Thus,  $\gamma_{\mathcal{D}}$  is closed if and only if  $\gamma_{\mathcal{D}} = \gamma_{\text{cl}\mathcal{D}}$ . The latter holds when  $\mathcal{D} = \text{cl}\mathcal{D}$ . Finally, the conclusion in item (iv) was stated on Rockafellar (1970, p. 128); indeed, the relation  $\mathcal{D} = \{x \mid \kappa(x) \leq 1\}$  can be verified directly from definition.  $\square$

From Proposition 2.1.2(iv) and Proposition 2.1.4(iv), it is not hard to prove the following formula on the polar of the sum of two gauges of independent variables.

**Proposition 2.1.5.** Let  $\kappa_1$  and  $\kappa_2$  be gauges. Then  $\kappa(x_1, x_2) := \kappa_1(x_1) + \kappa_2(x_2)$  is a gauge, and its polar is given by

$$\kappa^\circ(y_1, y_2) = \max \{ \kappa_1^\circ(y_1), \kappa_2^\circ(y_2) \}.$$

*Proof.* It is clear that  $\kappa$  is a gauge. Moreover,

$$\kappa^*(y_1, y_2) = \kappa_1^*(y_1) + \kappa_2^*(y_2) = \delta_{\mathcal{D}_1 \times \mathcal{D}_2}(y_1, y_2),$$

where  $\mathcal{D}_i = \{x \mid \kappa_i^\circ(x) \leq 1\}$  for  $i = 1, 2$ ; the first equality follows from the definition of the convex conjugate and the fact that  $y_1$  and  $y_2$  are decoupled, and the second equality follows from Proposition 2.1.2(iv). This together with Proposition 2.1.4(iv) implies that

$$\begin{aligned} \kappa^\circ(y_1, y_2) &= \inf \{ \lambda \geq 0 \mid y_1 \in \lambda \mathcal{D}_1, y_2 \in \lambda \mathcal{D}_2 \} \\ &= \max \{ \inf \{ \lambda \geq 0 \mid y_1 \in \lambda \mathcal{D}_1 \}, \inf \{ \lambda \geq 0 \mid y_2 \in \lambda \mathcal{D}_2 \} \} \\ &= \max \{ \gamma_{\mathcal{D}_1}(y_1), \gamma_{\mathcal{D}_2}(y_2) \} = \max \{ \kappa_1^\circ(y_1), \kappa_2^\circ(y_2) \}. \end{aligned}$$

This completes the proof.  $\square$

The following corollary is immediate from Propositions 2.1.3 and 2.1.5.

**Corollary 2.1.1.** Let  $\kappa_1$  and  $\kappa_2$  be gauges. Suppose that either

- (i)  $\text{epi } \kappa_1$  and  $\text{epi } \kappa_2$  are polyhedral; or
- (ii)  $\text{ri dom } \kappa_1 \cap \text{ri dom } \kappa_2 \neq \emptyset$ .

Then

$$(\kappa_1 + \kappa_2)^\circ(y) = \inf_{u_1, u_2} \{ \max \{ \kappa_1^\circ(u_1), \kappa_2^\circ(u_2) \} \mid u_1 + u_2 = y \}. \quad (2.7)$$

Moreover, the infimum is attained when finite.

*Proof.* Apply Proposition 2.1.3 with  $Ax = (x, x)$  and the gauge  $\kappa_1(x_1) + \kappa_2(x_2)$ , whose polar is given by Proposition 2.1.5.  $\square$

The *support function* for a nonempty convex set  $\mathcal{D}$  is defined as

$$\sigma_{\mathcal{D}}(y) = \sup_{x \in \mathcal{D}} \langle x, y \rangle.$$

It is straightforward to check that if  $\mathcal{D}$  contains the origin, then the support function is a (closed) gauge function. Indeed, we have the following relationship between support and Minkowski functions (Rockafellar, 1970, Corollary 15.1.2).

**Proposition 2.1.6.** Let  $\mathcal{D}$  be a closed convex set that contains the origin. Then  $\gamma_{\mathcal{D}}^{\circ} = \sigma_{\mathcal{D}}$  and  $\sigma_{\mathcal{D}}^{\circ} = \gamma_{\mathcal{D}}$ .

The following result relates the domain of the support function and the recession cone via polarity; see Auslender and Teboulle (2003, Theorem 2.2.1).

**Theorem 2.1.1.** If  $\mathcal{D}$  is nonempty and convex, then  $(\text{dom } \sigma_{\mathcal{D}})^{\circ} = \mathcal{D}_{\infty}$ .

### 2.1.3 Antipolar sets

The antipolar  $\mathcal{C}'$ , defined by (2.1), is nonempty as a consequence of the separation theorem. Freund's 1987 derivations are largely based on the following definition of a ray-like set. (As Freund mentions, the terms *antipolar* and *ray-like* are not universally used.)

**Definition 2.1.2.** A set  $\mathcal{D}$  is *ray-like* if for any  $x, y \in \mathcal{D}$ ,

$$x + \alpha y \in \mathcal{D} \quad \text{for all } \alpha \geq 0.$$

Note that the antipolar  $\mathcal{C}'$  of a (not necessarily ray-like) set  $\mathcal{C}$  must be ray-like.

The following result is analogous to the bipolar theorem for antipolar operations; see McLinden (1978, p. 176) and Freund (1987, Lemma 3).

**Theorem 2.1.2** (Bi-antipolar theorem).  $\mathcal{C} = \mathcal{C}''$  if and only if  $\mathcal{C}$  is ray-like.

The following proposition, stated by McLinden (1978, p. 176), follows from the bi-antipolar theorem.

**Proposition 2.1.7.**  $\mathcal{C}'' = \bigcup_{\lambda \geq 1} \lambda \mathcal{C}$ .

The next fact is used in the proof of the lemma following it; it has been drawn from Auslender and Teboulle (2003, Proposition 2.1.5) and compiled here for convenient reference.

**Proposition 2.1.8.** Let  $\mathcal{D}$  be a nonempty convex set in  $\mathcal{X}$ . Then the recession cone  $\mathcal{D}_\infty$  is a closed convex cone. Moreover, defining

$$\begin{aligned} D(x) &:= \{d \in \mathcal{X} \mid x + td \in \text{cl } \mathcal{D}, \forall t > 0\} \forall x \in \mathcal{D}, \\ E &:= \{d \in \mathcal{X} \mid \exists x \in \mathcal{D} \text{ such that } x + td \in \text{cl } \mathcal{D}, \forall t > 0\}, \\ F &:= \{d \in \mathcal{X} \mid d + \text{cl } \mathcal{D} \subset \text{cl } \mathcal{D}\}, \end{aligned}$$

we have that  $D(x)$  is in fact independent of  $x$ , now denoted simply by  $D$ , and  $\mathcal{D}_\infty = D = E = F$ .

The next lemma relates the positive polar of a convex set, its antipolar and the recession cone of its antipolar.

**Lemma 2.1.1.**  $\text{cl cone}(\mathcal{C}') = \mathcal{C}^* = (\mathcal{C}')_\infty$ .

*Proof.* It is evident that  $\text{cl cone}(\mathcal{C}') \subseteq \mathcal{C}^*$ . To show the converse inclusion, take any  $x \in \mathcal{C}^*$  and fix an  $x_0 \in \mathcal{C}'$ . Then for any  $\tau > 0$ , we have

$$\langle c, x + \tau x_0 \rangle \geq \tau \langle c, x_0 \rangle \geq \tau \quad \text{for all } c \in \mathcal{C},$$

which shows that  $x + \tau x_0 \in \text{cone } \mathcal{C}'$ . Taking the limit as  $\tau$  goes to 0 shows that  $x \in \text{cl cone}(\mathcal{C}')$ . This proves the first equality.

Next we show the second equality, and begin with the observation that  $\mathcal{C}^* \subseteq (\mathcal{C}')_\infty$ . Conversely, suppose that  $x \in (\mathcal{C}')_\infty$  and fix any  $x_0 \in \mathcal{C}'$ . Then, by Proposition 2.1.8,  $x_0 + \tau x \in \mathcal{C}'$  for all  $\tau > 0$ . Hence, for any  $c \in \mathcal{C}$ ,

$$\frac{1}{\tau} \langle c, x_0 \rangle + \langle c, x \rangle = \frac{1}{\tau} \langle c, x_0 + \tau x \rangle \geq \frac{1}{\tau}.$$

Since this is true for all  $\tau > 0$ , we must have  $\langle c, x \rangle \geq 0$ . Since  $c \in \mathcal{C}$  is arbitrary, we conclude that  $x \in \mathcal{C}^*$ .  $\square$

## 2.2 Antipolar calculus

In general, it may not always be easy to obtain an explicit formula for the Minkowski function of a given closed convex set  $\mathcal{D}$ . Hence, we derive some elements of an antipolar calculus that allows us to express the antipolar of a more complicated set in terms of the antipolars of its constituents. These rules are useful for writing down the explicit gauge duals of problems such as  $(P_\rho)$ . Table 2.1 summarizes the main elements of the calculus.

As a first step, the following formula gives an expression for the antipolar of a set defined via a gauge. The formula follows directly from the definition of polar functions.

Table 2.1: The main rules of the antipolar calculus; the required assumptions are made explicit in the specific references.

Result	Reference
$(AC)' = (A^*)^{-1}C'$	Proposition 2.2.2
$(A^{-1}C)' = \text{cl}(A^*C')$	Propositions 2.2.3 and 2.2.4
$(C_1 \cup C_2)' = C'_1 \cap C'_2$	Proposition 2.2.5
$(C_1 \cap C_2)' = \text{cl conv}(C'_1 \cup C'_2)$	Proposition 2.2.6

**Proposition 2.2.1.** Let  $C = \{x \mid \rho(b-x) \leq \epsilon\}$  with  $0 < \epsilon < \rho(b)$ . Then  

$$C' = \{y \mid \langle b, y \rangle - \epsilon\rho^\circ(y) \geq 1\}.$$

*Proof.* Note that  $y \in C'$  is equivalent to  $\langle x, y \rangle \geq 1$  for all  $x \in C$ . Thus, for all  $x$  such that  $\rho(b-x) \leq \epsilon$ ,

$$\langle x-b, y \rangle \geq 1 - \langle b, y \rangle \iff \langle b-x, y \rangle \leq \langle b, y \rangle - 1.$$

From Proposition 2.1.2(iii), this is further equivalent to  $\epsilon\rho^\circ(y) \leq \langle b, y \rangle - 1$ .  $\square$

Proposition 2.2.1 is very general since any closed convex set  $\mathcal{D}$  containing the origin can be represented in the form of  $\{x \mid \rho(x) \leq 1\}$ , where  $\rho(x) = \inf \{\lambda \geq 0 \mid x \in \lambda\mathcal{D}\}$ ; cf. (2.4). For conic constraints in particular, one obtains the following corollary by setting  $\rho(x) = \delta_{-\mathcal{K}}(x)$ .

**Corollary 2.2.1.** Let  $C = \{x \mid x \in b + \mathcal{K}\}$  for some closed convex cone  $\mathcal{K}$  and a vector  $b \notin -\mathcal{K}$ . Then

$$C' = \{y \in \mathcal{K}^* \mid \langle b, y \rangle \geq 1\}.$$

Note that Proposition 2.2.1 excludes the potentially important case  $\epsilon = 0$ ; however, Corollary 2.2.1 can instead be applied by defining  $\mathcal{K} = \rho^{-1}(0) = \{0\}$ .

### 2.2.1 Linear transformations

We now consider the antipolar of the image of  $C$  under a linear map  $A$ .

**Proposition 2.2.2.** It holds that

$$(AC)' = (A^*)^{-1}C'.$$

Furthermore, if  $\text{cl}(AC)$  does not contain the origin, then both sets above are nonempty.

*Proof.* Note that  $y \in (AC)'$  is equivalent to

$$\langle y, Ac \rangle = \langle A^*y, c \rangle \geq 1 \quad \text{for all } c \in \mathcal{C}.$$

The last relation is equivalent to  $A^*y \in \mathcal{C}'$ . Hence,  $(AC)' = (A^*)^{-1}\mathcal{C}'$ . Furthermore, the assumption that  $\text{cl}(AC)$  does not contain the origin, together with an argument using supporting hyperplanes, implies  $(AC)'$  is nonempty. This completes the proof.  $\square$

We have the following result concerning the pre-image of  $\mathcal{C}$ .

**Proposition 2.2.3.** Suppose that  $A^{-1}\mathcal{C} \neq \emptyset$ . Then

$$(A^{-1}\mathcal{C})' = \text{cl}(A^*\mathcal{C}'),$$

and both sets are nonempty.

*Proof.* Recall from our blanket assumption (cf. §2.1) that  $\mathcal{C}$  is a closed convex set not containing the origin. It follows that  $\text{cl}(A^*\mathcal{C}')$  is nonempty. Moreover,  $A^{-1}\mathcal{C}$  is also a closed convex set that does not contain the origin. Hence,  $(A^{-1}\mathcal{C})'$  is also nonempty.

We next show that  $\text{cl}(A^*\mathcal{C}')$  does not contain the origin. Suppose that  $y \in A^*\mathcal{C}'$  so that  $y = A^*u$  for some  $u \in \mathcal{C}'$ . Then for any  $x \in A^{-1}\mathcal{C}$ , we have  $Ax \in \mathcal{C}$  and thus

$$\langle x, y \rangle = \langle x, A^*u \rangle = \langle Ax, u \rangle \geq 1,$$

which shows that  $y \in (A^{-1}\mathcal{C})'$ . Thus, we have  $A^*\mathcal{C}' \subseteq (A^{-1}\mathcal{C})'$  and consequently that  $\text{cl}(A^*\mathcal{C}') \subseteq (A^{-1}\mathcal{C})'$ . Since the set  $A^{-1}\mathcal{C}$  is nonempty,  $(A^{-1}\mathcal{C})'$  does not contain the origin. Hence, it follows that  $\text{cl}(A^*\mathcal{C}')$  also does not contain the origin.

Now apply Proposition 2.2.2 with  $A^*$  in place of  $A$ , and  $\mathcal{C}'$  in place of  $\mathcal{C}$ , to obtain

$$(A^*\mathcal{C}')' = A^{-1}\mathcal{C}''.$$

Taking the antipolar on both sides of the above relation, we arrive at

$$(A^*\mathcal{C}')'' = (A^{-1}\mathcal{C}'')'. \tag{2.8}$$

Since  $\mathcal{C}'$  is ray-like, it follows that  $\text{cl}(A^*\mathcal{C}')$  is also ray-like. Since  $\text{cl}(A^*\mathcal{C}')$  does not contain the origin, we conclude from the bi-antipolar theorem that  $(A^*\mathcal{C}')'' = \text{cl}(A^*\mathcal{C}')$ . Moreover, we have

$$(A^{-1}\mathcal{C}'')' = \left( \bigcup_{\lambda \geq 1} \lambda A^{-1}\mathcal{C} \right)' = (A^{-1}\mathcal{C})',$$

where the first equality follows from Proposition 2.1.7, and the second equality can be verified directly from definition. The conclusion now follows from the above discussion and (2.8).  $\square$

The next result is used in the proof of the proposition following it and we compile it here for convenient reference; see (Berman, 1973, Lemma 3.1).

**Theorem 2.2.1** (R.A. Abrams). Let  $K \subset \mathcal{X}$  and  $A$  a linear map from  $\mathcal{X}$ . Then  $AK$  is closed if and only if  $K + \ker A$  is closed.

We now have the following further consequence.

**Proposition 2.2.4.** Suppose that  $A^{-1}\mathcal{C} \neq \emptyset$ , and either  $\mathcal{C}$  is polyhedral or  $\text{ri } \mathcal{C} \cap \text{range } A \neq \emptyset$ . Then  $(A^{-1}\mathcal{C})'$  is nonempty and

$$(A^{-1}\mathcal{C})' = A^*\mathcal{C}'.$$

*Proof.* We will show that  $A^*\mathcal{C}'$  is closed under the assumption of this proposition. Then the conclusion follows immediately from Proposition 2.2.3.

Abrams's Theorem 2.2.1 asserts that  $A^*\mathcal{C}'$  is closed if and only if  $\mathcal{C}' + \ker A^*$  is closed. We will thus establish the closedness of the latter set.

Suppose that  $\mathcal{C}$  is a polyhedral. Then it is routine to show that  $\mathcal{C}'$  is also a polyhedral and thus  $\mathcal{C}' + \ker A^*$  is closed. Hence, the conclusion of the corollary holds under this assumption.

Finally, suppose that  $\text{ri } \mathcal{C} \cap \text{range } A \neq \emptyset$ . From Theorem 2.1.1 and the bipolar theorem mentioned in §2.1.1, we have  $\text{cl dom } \sigma_{\mathcal{C}'} = [(\mathcal{C}')_{\infty}]^{\circ}$ , where  $(\mathcal{C}')_{\infty}$  is the recession cone of  $\mathcal{C}'$ , which turns out to be just  $\mathcal{C}^*$  by Lemma 2.1.1. From this and the bipolar theorem, we see further that

$$\text{cl dom } \sigma_{\mathcal{C}'} = [\mathcal{C}^*]^{\circ} = [(\text{cl cone } \mathcal{C})^*]^{\circ} = -\text{cl cone } \mathcal{C},$$

and hence  $\text{ri dom } \sigma_{\mathcal{C}'} = -\text{ri cone } \mathcal{C}$ , thanks to Rockafellar (1970, Theorem 6.3). Furthermore, the assumption that  $\text{ri } \mathcal{C} \cap \text{range } A \neq \emptyset$  is equivalent to  $\text{ri cone } \mathcal{C} \cap \text{range } A \neq \emptyset$ , since  $\text{ri cone } \mathcal{C} = \bigcup_{\lambda > 0} \lambda \text{ri } \mathcal{C}$ ; see Rockafellar (1970, p. 50). Thus, the assumption  $\text{ri } \mathcal{C} \cap \text{range } A \neq \emptyset$  together with Rockafellar (1970, Theorem 23.8) imply that

$$\mathcal{C}' + \ker A^* = \partial \sigma_{\mathcal{C}'}(0) + \partial \delta_{\text{range } A}(0) = \partial(\sigma_{\mathcal{C}'} + \delta_{\text{range } A})(0).$$

In particular,  $\mathcal{C}' + \ker A^*$  is closed.  $\square$

## 2.2.2 Unions and intersections

Other important set operations are union and intersection, which we discuss here. Ruys and Weddepohl (1979, Appendix A.1) outline additional rules.

**Proposition 2.2.5.** Let  $\mathcal{C}_1$  and  $\mathcal{C}_2$  be nonempty closed convex sets. Then

$$(\mathcal{C}_1 \cup \mathcal{C}_2)' = \mathcal{C}'_1 \cap \mathcal{C}'_2.$$

If  $0 \notin \text{cl conv}(\mathcal{C}_1 \cup \mathcal{C}_2)$ , then the sets above are nonempty.

*Proof.* By definition, every  $y \in (\mathcal{C}_1 \cup \mathcal{C}_2)'$  obeys  $\langle y, x \rangle \geq 1$  for all  $x \in (\mathcal{C}_1 \cup \mathcal{C}_2) \supseteq \mathcal{C}_1$ , so  $y \in \mathcal{C}'_1$ . Likewise,  $y \in \mathcal{C}'_2$ , and thus  $y \in \mathcal{C}'_1 \cap \mathcal{C}'_2$ . The converse is equally direct. Moreover, if we assume further that  $0 \notin \text{cl conv}(\mathcal{C}_1 \cup \mathcal{C}_2)$ , then  $(\mathcal{C}_1 \cup \mathcal{C}_2)' = [\text{cl conv}(\mathcal{C}_1 \cup \mathcal{C}_2)]'$  is nonempty. This completes the proof.  $\square$

We now consider the antipolar of intersections. Note that it is necessary to assume that both  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are ray-like, which was missing from Ruys and Weddepohl (1979, Property A.5). (The necessity of this assumption is demonstrated by Example 2.2.1, which follows the proposition.)

**Proposition 2.2.6.** Let  $\mathcal{C}_1$  and  $\mathcal{C}_2$  be nonempty ray-like closed convex sets not containing the origin. Suppose further that  $\mathcal{C}_1 \cap \mathcal{C}_2 \neq \emptyset$ . Then

$$(\mathcal{C}_1 \cap \mathcal{C}_2)' = \text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2),$$

and both sets are nonempty.

*Proof.* From the fact that both  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are closed convex sets not containing the origin, it follows that  $\mathcal{C}'_1$  and  $\mathcal{C}'_2$  are nonempty and hence  $\text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2) \neq \emptyset$ . Moreover, because  $\mathcal{C}_1 \cap \mathcal{C}_2$  does not contain the origin,  $(\mathcal{C}_1 \cap \mathcal{C}_2)'$  is also nonempty.

We first show that  $\text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2)$  does not contain the origin. To this end, let  $y \in \mathcal{C}'_1 \cup \mathcal{C}'_2$ . For any  $x \in \mathcal{C}_1 \cap \mathcal{C}_2$ , we have  $\langle y, x \rangle \geq 1$ , which shows that  $\mathcal{C}'_1 \cup \mathcal{C}'_2 \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2)'$ , and hence  $\text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2) \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2)'$ . Since  $\mathcal{C}_1 \cap \mathcal{C}_2$  is nonempty,  $(\mathcal{C}_1 \cap \mathcal{C}_2)'$  does not contain the origin. Consequently,  $\text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2)$  does not contain the origin, as claimed.

Now apply Proposition 2.2.5, with  $\mathcal{C}'_1$  in place of  $\mathcal{C}_1$  and  $\mathcal{C}'_2$  in place of  $\mathcal{C}_2$ , to obtain

$$(\mathcal{C}'_1 \cup \mathcal{C}'_2)' = \mathcal{C}''_1 \cap \mathcal{C}''_2 = \mathcal{C}_1 \cap \mathcal{C}_2.$$

Take the antipolar of both sides to obtain

$$(\mathcal{C}_1 \cap \mathcal{C}_2)' = (\mathcal{C}'_1 \cup \mathcal{C}'_2)'' = [\text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2)]'' = \text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2),$$

where the second equality follows from the definition of antipolar, and the third equality follows from the observation that  $\text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2)$  is a nonempty ray-like closed convex set not containing the origin. This completes the proof.  $\square$

The following counter-example shows that the requirement that  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are ray-like cannot be removed from Proposition 2.2.6. Refer to Figure 2.1 for a visual depiction of its construction.

**Example 2.2.1** (Set intersection and the ray-like property).

Consider the sets

$$\mathcal{C}_1 = \{(x_1, x_2) \mid 1 - x_1 \leq x_2 \leq x_1 - 1\} \quad \text{and} \quad \mathcal{C}_2 = \{(x_1, x_2) \mid x_1 = 1\}.$$

Define  $H_1 = \{(x_1, x_2) \mid x_1 + x_2 \geq 1\}$  and  $H_2 = \{(x_1, x_2) \mid x_1 - x_2 \geq 1\}$  so that  $\mathcal{C}_1 = H_1 \cap H_2$ . Clearly the set  $\mathcal{C}_2$  is not ray-like, while the sets  $\mathcal{C}_1$ ,  $H_1$ , and  $H_2$  are. Moreover, all four sets do not contain the origin. Furthermore,  $\mathcal{C}_1 \cap \mathcal{C}_2$  is the singleton  $\{(1, 0)\}$ , and hence a direct computation shows that  $(\mathcal{C}_1 \cap \mathcal{C}_2)' = \{(y_1, y_2) \mid y_1 \geq 1\}$ .

Next, it follows directly from the antipolar definition that

$$\mathcal{C}'_2 = \{(y_1, 0) \mid y_1 \geq 1\}.$$

Also note that  $H_1 = L_1^{-1}I$ , where  $L_1(x_1, x_2) = x_1 + x_2$  and  $I = \{u \mid u \geq 1\}$ . Thus, by Proposition 2.2.4,  $H'_1 = \{(y_1, y_1) \mid y_1 \geq 1\}$ .

Similarly,  $H'_2 = \{(y_1, -y_1) \mid y_1 \geq 1\}$ . Because  $H_1$  and  $H_2$  are ray-like, it follows from Proposition 2.2.6 that

$$\mathcal{C}'_1 = (H_1 \cap H_2)' = \text{cl conv}(H'_1 \cup H'_2),$$

which contains  $\mathcal{C}'_2$ . Thus,

$$\text{cl conv}(\mathcal{C}'_1 \cup \mathcal{C}'_2) = \mathcal{C}'_1 \subsetneq \{(y_1, y_2) \mid y_1 \geq 1\} = (\mathcal{C}_1 \cap \mathcal{C}_2)'.$$

## 2.3 Duality derivations

We derive in this section the gauge and Lagrange duals of the primal problem  $(P_\rho)$ . Let

$$\mathcal{C} = \{x \mid \rho(b - Ax) \leq \epsilon\} \tag{2.9}$$

denote the constraint set, where  $\rho$  is a closed gauge and  $0 \leq \epsilon < \rho(b)$ . We also consider the associated set

### 2.3. Duality derivations

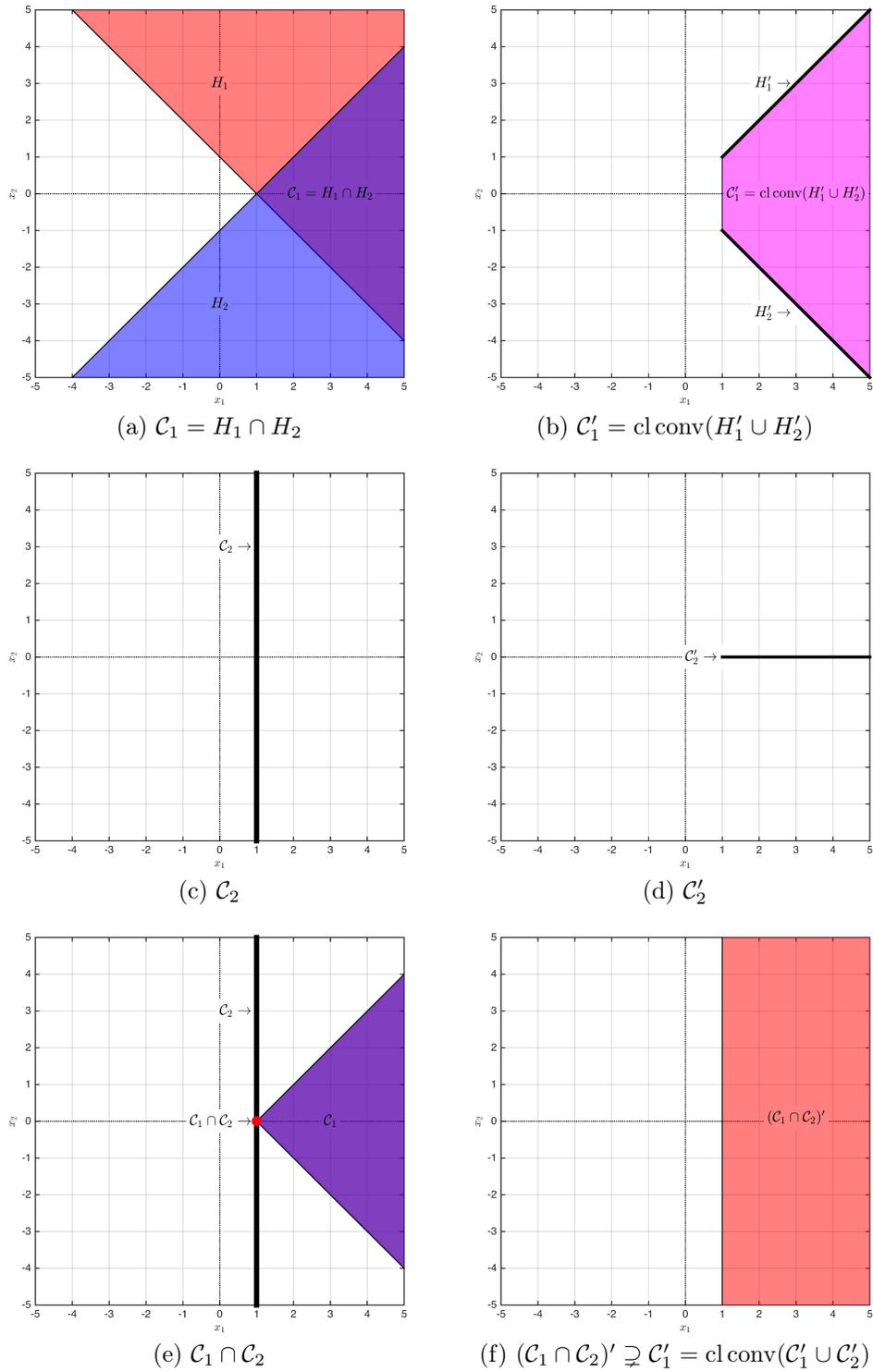


Figure 2.1: Visual depiction of the counter-example described in Example 2.2.1

$$\mathcal{C}_0 = \{ u \mid \rho(b - u) \leq \epsilon \}, \quad (2.10)$$

and note that  $\mathcal{C} = A^{-1}\mathcal{C}_0$ . Recall from our blanket assumption in §2.1 that when  $\epsilon = 0$ , we only consider closed gauges  $\rho$  with  $\rho^{-1}(0) = \{0\}$ .

### 2.3.1 The gauge dual

We consider two approaches for deriving the gauge dual of  $(P_\rho)$ . The first uses explicitly the abstract definition of the gauge dual (D). The second approach redefines the objective function to also contain an indicator for the nonlinear gauge  $\rho$  where  $\mathcal{C}$  is an affine set. This alternative approach is instructive, because it illustrates the modeling choices that are available when working with gauge functions.

#### First approach

The following combines Proposition 2.2.4 with Proposition 2.2.1, and gives an explicit expression for the antipolar of  $\mathcal{C}$  when  $\epsilon > 0$ .

**Corollary 2.3.1.** Suppose that  $\mathcal{C}$  is given by (2.9), where  $0 < \epsilon < \rho(b)$ , and  $\mathcal{C}_0$  is given by (2.10). If  $\mathcal{C}_0$  is polyhedral, or  $\text{ri } \mathcal{C}_0 \cap \text{range } A \neq \emptyset$ , then

$$\mathcal{C}' = \{ A^*y \mid \langle b, y \rangle - \epsilon \rho^\circ(y) \geq 1 \}.$$

As an aside, we present the following result, which follows from Corollary 2.2.1 and Proposition 2.2.4, concerning a general closed convex cone  $\mathcal{K}$ .

**Corollary 2.3.2.** Suppose that  $\mathcal{C} = \{ x \mid Ax - b \in \mathcal{K} \}$  for some closed convex cone  $\mathcal{K}$  and  $b \notin -\mathcal{K}$ . If  $\mathcal{K}$  is polyhedral, or  $(b + \text{ri } \mathcal{K}) \cap \text{range } A \neq \emptyset$ , then

$$\mathcal{C}' = \{ A^*y \mid \langle b, y \rangle \geq 1, y \in \mathcal{K}^* \}.$$

These results can be used to obtain an explicit representation of the gauge dual problem. We rely on the antipolar calculus developed in §2.2. Assume

$$\mathcal{C}_0 \text{ is polyhedral, or } \text{ri } \mathcal{C}_0 \cap \text{range } A \neq \emptyset. \quad (2.11)$$

Consider separately the cases  $\epsilon > 0$  and  $\epsilon = 0$ .

**Case 1 ( $\epsilon > 0$ ):** Apply Corollary 2.3.1 to derive the antipolar set

$$\mathcal{C}' = \{ A^*y \mid \langle b, y \rangle - \epsilon \rho^\circ(y) \geq 1 \}. \quad (2.12)$$

**Case 2** ( $\epsilon = 0$ ): Here we use the blanket assumption (see §2.1) that  $\rho^{-1}(0) = \{0\}$ , and in that case,  $\mathcal{C} = \{x \mid Ax = b\}$ . Apply Corollary 2.3.2 with  $\mathcal{K} = \{0\}$  to obtain

$$\mathcal{C}' = \{A^*y \mid \langle b, y \rangle \geq 1\}. \quad (2.13)$$

Since  $\rho^{-1}(0) = \{0\}$  and  $\rho$  is closed, we conclude from Proposition 2.1.2(v) that  $\text{dom } \rho^\circ = \mathcal{X}$ . Hence, (2.13) can be seen as a special case of (2.12) with  $\epsilon = 0$ .

These two cases can be combined, and we see that when (2.11) holds, the gauge dual problem (D) for  $(P_\rho)$  can be expressed as  $(D_\rho)$ . If the assumptions (2.11) are not satisfied, then in view of Proposition 2.2.3, it still holds that (D) is equivalent to

$$\underset{u, y}{\text{minimize}} \quad \kappa^\circ(u) \quad \text{subject to} \quad u \in \text{cl} \{A^*y \mid \langle y, b \rangle - \epsilon \rho^\circ(y) \geq 1\}.$$

This optimal value can in general be less than or equal to that of  $(D_\rho)$ .

### Second approach

This approach does not rely on assumptions (2.11). Define the function  $\xi(x, r, \tau) := \kappa(x) + \delta_{\text{epi } \rho}(r, \tau)$ , which is a gauge because  $\text{epi } \rho$  is a cone. Then  $(P_\rho)$  can be equivalently reformulated as

$$\underset{x, r, \tau}{\text{minimize}} \quad \xi(x, r, \tau) \quad \text{subject to} \quad Ax + r = b, \quad \tau = \epsilon. \quad (2.14)$$

Invoke Proposition 2.1.5 to obtain

$$\begin{aligned} \xi^\circ(z, y, \alpha) &= \max \{ \kappa^\circ(z), (\delta_{\text{epi } \rho})^\circ(y, \alpha) \} \\ &\stackrel{(i)}{=} \max \{ \kappa^\circ(z), \delta_{(\text{epi } \rho)^\circ}(y, \alpha) \} \\ &\stackrel{(ii)}{=} \kappa^\circ(z) + \delta_{(\text{epi } \rho)^\circ}(y, \alpha) \stackrel{(iii)}{=} \kappa^\circ(z) + \delta_{\text{epi } (\rho^\circ)}(y, -\alpha), \end{aligned}$$

where (i) follows from Proposition 2.1.4(i), (ii) follows from the definition of indicator function, and (iii) follows from Proposition 2.1.2(vi). As Freund (1987, §2) shows for gauge programs with linear constraints, the gauge dual is given by

$$\underset{y, \alpha}{\text{minimize}} \quad \xi^\circ(A^*y, y, \alpha) \quad \text{subject to} \quad \langle y, b \rangle + \epsilon \alpha \geq 1,$$

which can be rewritten as

$$\underset{y, \alpha}{\text{minimize}} \quad \kappa^\circ(A^*y) \quad \text{subject to} \quad \langle y, b \rangle + \epsilon \alpha \geq 1, \quad \rho^\circ(y) \leq -\alpha.$$

(The gauge dual for problems with linear constraints also follows directly from Corollary 2.3.2 with  $\mathcal{K} = \{0\}$ .) Further simplification leads to the gauge dual program  $(D_\rho)$ .

Note that the transformation used to derive (2.14) is very flexible. For example, if  $(P_\rho)$  contained the additional conic constraint  $x \in \mathcal{K}$ , then  $\xi$  could be defined to contain an additional term given by the indicator of  $\mathcal{K}$ .

Even though this approach does not require the assumptions (2.11) used in §2.3.1, and thus appears to apply more generally, it is important to keep in mind that we have yet to impose conditions that imply strong duality. In fact, as we show in §2.4, the assumptions required there imply (2.11).

### 2.3.2 Lagrange duality

Our derivation of the Lagrange dual problem  $(D_\ell)$  is standard, and we include it here as a counterpoint to the corresponding gauge dual derivation. We begin by reformulating  $(P_\rho)$  by introducing an artificial variable  $r$ , and deriving the dual of the equivalent problem

$$\underset{x,r}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad Ax + r = b, \quad \rho(r) \leq \epsilon. \quad (2.15)$$

Define the Lagrangian function

$$L(x, r, y) = \kappa(x) + \langle y, b - Ax - r \rangle.$$

The Lagrange dual problem is given by

$$\underset{y}{\text{maximize}} \quad \inf_{x, \rho(r) \leq \epsilon} L(x, r, y).$$

Consider the (concave) dual function

$$\begin{aligned} \ell(y) &= \inf_{x, \rho(r) \leq \epsilon} L(x, r, y) \\ &= \inf_{x, \rho(r) \leq \epsilon} \left\{ \langle y, b \rangle - \langle y, r \rangle - (\langle A^*y, x \rangle - \kappa(x)) \right\} \\ &= \langle y, b \rangle - \sup_{\rho(r) \leq \epsilon} \langle y, r \rangle - \sup_x \left\{ \langle A^*y, x \rangle - \kappa(x) \right\} \\ &= \langle y, b \rangle - \epsilon \rho^\circ(y) - \delta_{\kappa^\circ(\cdot) \leq 1}(A^*y), \end{aligned}$$

where the first conjugate on the right-hand side follows from Proposition 2.1.2(iii) when  $\epsilon > 0$ , and when  $\epsilon = 0$ , it is a direct consequence of the assumption that  $\rho^{-1}(0) = \{0\}$  so that  $\text{dom } \rho^\circ = \mathcal{X}$  from Proposition 2.1.2(v); the last conjugate

follows from Proposition 2.1.2(iv). The Lagrange dual problem is obtained by maximizing  $\ell$ , leading to  $(D_\ell)$ .

Strictly speaking, the Lagrangian primal-dual pair of problems that we have derived is given by (2.15) and  $(D_\ell)$ , but it is easy to see that  $(P_\rho)$  is equivalent to (2.15) in the sense that the respective optimal values are the same, and that solutions to one problem readily lead to solutions for the other. Thus, without loss of generality, we refer to  $(D_\ell)$  as the Lagrange dual to the primal problem  $(P_\rho)$ .

## 2.4 Strong duality

Freund's 1987 analysis of the gauge dual pair is mainly based on the classical separation theorem. It relies on the ray-like property of the constraint set  $\mathcal{C}$ . Our study of the gauge dual pairs allows us to relax the ray-like assumption. By establishing connections with the Fenchel duality framework, we can develop strong duality conditions that are analogous to those required for Lagrange duality theory.

The Fenchel dual (Rockafellar, 1970, §31) of  $(P)$  is given by

$$\underset{y}{\text{maximize}} \quad -\sigma_{\mathcal{C}}(-y) \quad \text{subject to} \quad \kappa^\circ(y) \leq 1, \quad (2.16)$$

where we use  $(\delta_{\mathcal{C}})^* = \sigma_{\mathcal{C}}$  and Proposition 2.1.2(iv) to obtain  $\kappa^* = \delta_{[\kappa^\circ \leq 1]}$ . Let  $v_p$ ,  $v_g$ , and  $v_f$ , respectively, denote the optimal values of  $(P)$ ,  $(D)$  and (2.16). The following result relates their optimal values and dual solutions.

**Theorem 2.4.1** (Weak duality). Suppose that  $\text{dom } \kappa^\circ \cap \mathcal{C}' \neq \emptyset$ . Then

$$v_p \geq v_f = 1/v_g > 0.$$

Furthermore,

- (i) if  $y^*$  solves (2.16), then  $y^* \in \text{cone } \mathcal{C}'$  and  $y^*/v_f$  solves  $(D)$ ;
- (ii) if  $y^*$  solves  $(D)$  and  $v_g > 0$ , then  $v_f y^*$  solves (2.16).

*Proof.* The fact that  $v_p \geq v_f$  follows from standard Fenchel duality theory. We now show that  $v_f = 1/v_g$ .

Because  $\text{dom } \kappa^\circ \cap \mathcal{C}' \neq \emptyset$ , there exists  $y_0$  such that  $\kappa^\circ(y_0) \leq 1$  and  $y_0 \in \tau \mathcal{C}'$  for some  $\tau > 0$ . In particular, because  $-\sigma_{\mathcal{C}}(-y) = \inf_{c \in \mathcal{C}} \langle c, y \rangle$  for all  $y$ , it follows from the definition of  $v_f$  that

$$v_f = \sup_y \{ \inf_{c \in \mathcal{C}} \langle c, y \rangle \mid \kappa^\circ(y) \leq 1 \} \geq \inf_{c \in \mathcal{C}} \langle c, y_0 \rangle \geq \tau > 0. \quad (2.17)$$

## 2.4. Strong duality

---

Hence

$$v_f = \sup_{y, \lambda} \{ \lambda \mid \kappa^\circ(y) \leq 1, -\sigma_{\mathcal{C}}(-y) \geq \lambda, \lambda > 0 \}. \quad (2.18)$$

From this, we have further that

$$\begin{aligned} v_f &= \sup_{y, \lambda} \{ \lambda \mid \kappa^\circ(y/\lambda) \leq 1/\lambda, -\sigma_{\mathcal{C}}(-y/\lambda) \geq 1, 1/\lambda > 0 \} \\ &= \sup_{y, \mu} \{ 1/\mu \mid \kappa^\circ(\mu y) \leq \mu, -\sigma_{\mathcal{C}}(-\mu y) \geq 1, \mu > 0 \}. \end{aligned}$$

Inverting both sides of this equation gives

$$\begin{aligned} 1/v_f &= \inf_{y, \mu} \{ \mu \mid \kappa^\circ(\mu y) \leq \mu, -\sigma_{\mathcal{C}}(-\mu y) \geq 1, \mu > 0 \} \\ &= \inf_{w, \mu} \{ \mu \mid \kappa^\circ(w) \leq \mu, -\sigma_{\mathcal{C}}(-w) \geq 1, \mu > 0 \} \\ &\stackrel{(i)}{=} \inf_{w, \mu} \{ \mu \mid \kappa^\circ(w) \leq \mu, w \in \mathcal{C}', \mu > 0 \} \\ &= \inf_{w, \mu} \{ \mu \mid \kappa^\circ(w) \leq \mu, w \in \mathcal{C}' \} \\ &= \inf_w \{ \kappa^\circ(w) \mid w \in \mathcal{C}' \} = v_g, \end{aligned} \quad (2.19)$$

where equality (i) follows from the definition of  $\mathcal{C}'$ . This proves  $v_f = 1/v_g$ .

We now prove item (i). Assume that  $y^*$  solves (2.16). Then  $v_f$  is nonzero (by (2.17)) and finite, and so is  $v_g = 1/v_f$ . Then  $y^* \in \text{cone } \mathcal{C}'$  because  $-\sigma_{\mathcal{C}}(-y^*) = \inf_{c \in \mathcal{C}} \langle c, y^* \rangle = v_f > 0$ , and we see from (2.19) that  $y^*/v_f$  solves (D). We now prove item (ii). Note that if  $y^*$  solves (D) and  $v_g > 0$ , then  $\kappa^\circ(y^*) > 0$ . One can then observe similarly from (2.19) that  $y^*/v_g = v_f y^*$  solves (2.16). This completes the proof.  $\square$

Fenchel duality theory allows us to use Theorem 2.4.1 to obtain several sufficient conditions that guarantee strong duality, i.e.,  $v_p v_g = 1$ , and the attainment of the gauge dual problem (D). For example, applying Rockafellar (1970, Theorem 31.1) yields the following corollary.

**Corollary 2.4.1** (Strong duality I). Suppose that  $\text{dom } \kappa^\circ \cap \mathcal{C}' \neq \emptyset$  and  $\text{ri dom } \kappa \cap \text{ri } \mathcal{C} \neq \emptyset$ . Then  $v_p v_g = 1$  and the gauge dual (D) attains its optimal value.

*Proof.* From  $\text{ri dom } \kappa \cap \text{ri } \mathcal{C} \neq \emptyset$  and Rockafellar (1970, Theorem 31.1), we see that  $v_p = v_f$  and  $v_f$  is attained. The conclusion of the corollary now follows immediately from Theorem 2.4.1.  $\square$

## 2.4. Strong duality

---

We would also like to guarantee *primal attainment*. Note that the gauge dual of the gauge dual problem (D) (i.e., the bidual of (P)) is given by

$$\underset{x}{\text{minimize}} \quad \kappa^{\circ\circ}(x) \quad \text{subject to} \quad x \in \mathcal{C}'', \quad (2.20)$$

which is not the same as (P) unless  $\mathcal{C}$  is ray-like and  $\kappa$  is closed; see Theorem 2.1.2 and Proposition 2.1.2(ii). However, we show in the next proposition that (2.20) and (P) always have the same optimal value when  $\kappa$  is closed (even if  $\mathcal{C}$  is not ray-like), and that if the optimal value is attained in one problem, it is also attained in the other.

**Proposition 2.4.1.** Suppose that  $\kappa$  is closed. Then the optimal values of (P) and (2.20) are the same. Moreover, if the optimal value is attained in one problem, it is also attained in the other.

*Proof.* From Proposition 2.1.7, we see that (2.20) is equivalent to

$$\underset{\lambda, x}{\text{minimize}} \quad \lambda \kappa(x) \quad \text{subject to} \quad x \in \mathcal{C}, \lambda \geq 1,$$

which clearly gives the same optimal value as (P). This proves the first conclusion. The second conclusion now also follows immediately.  $\square$

Hence, we obtain the following corollary, which generalizes Freund (1987, Theorem 2A) by dropping the ray-like assumption on  $\mathcal{C}$ .

**Corollary 2.4.2** (Strong duality II). Suppose that  $\kappa$  is closed, and that  $\text{ri dom } \kappa \cap \text{ri } \mathcal{C} \neq \emptyset$  and  $\text{ri dom } \kappa^{\circ} \cap \text{ri } \mathcal{C}' \neq \emptyset$ . Then  $v_p v_g = 1$  and both values are attained.

*Proof.* The conclusion follows from Corollary 2.4.1, Proposition 2.4.1, the fact that  $\kappa = \kappa^{\circ\circ}$  for closed gauge functions, and the observation that  $\text{ri dom } \kappa \cap \text{ri } \mathcal{C} \neq \emptyset$  if and only if  $\text{ri dom } \kappa \cap \text{ri } \mathcal{C}'' \neq \emptyset$ , since  $\text{ri } \mathcal{C}'' = \bigcup_{\lambda > 1} \lambda \text{ri } \mathcal{C}$  (Rockafellar, 1970, p. 50) and  $\text{dom } \kappa$  is a cone.  $\square$

Before closing this section, we specialize Theorem 2.4.1 to study the relationship between the Lagrange ( $D_\ell$ ) and gauge ( $D_p$ ) duals. Let  $v_\ell$  denote the optimal value of ( $D_\ell$ ). We use the fact that, for any  $y$ ,

$$-\sigma_{\mathcal{C}}(-y) = \inf_{c \in \mathcal{C}} \langle c, y \rangle \begin{cases} > 0 & \text{if } y \in \text{cone } \mathcal{C}' \setminus \{0\}, \\ \leq 0 & \text{otherwise,} \end{cases} \quad (2.21)$$

which is directly verifiable using the definition of  $\mathcal{C}'$ .

## 2.4. Strong duality

**Corollary 2.4.3.** Suppose that  $\mathcal{C}$  is given by (2.9), where  $0 \leq \epsilon < \rho(b)$ , assumption (2.11) holds, and  $\text{dom } \kappa^\circ \cap \mathcal{C}' \neq \emptyset$ . Then  $v_l = v_f > 0$ . Moreover,

- (i) if  $y^*$  solves  $(D_\ell)$ , then  $y^*/v_l$  solves  $(D_\rho)$ ;
- (ii) if  $y^*$  solves  $(D_\rho)$  and  $v_g > 0$ , then  $v_l y^*$  solves  $(D_\ell)$ .

*Proof.* From (2.21), for any  $y \in \text{cone } \mathcal{C}' \setminus \{0\}$ , we have  $-\sigma_{\mathcal{C}}(-y) = \inf_{c \in \mathcal{C}} \langle c, y \rangle > 0$  and is hence finite. Note that  $\inf_{c \in \mathcal{C}} \langle c, y \rangle = \inf_{c, r} \{ \langle c, y \rangle \mid Ac + r = b, \rho(r) \leq \epsilon \}$ . Use this reformulation and proceed as in §2.3.2 to obtain the dual function

$$\begin{aligned} \ell(u) &= \inf_{c, \rho(r) \leq \epsilon} \{ \langle u, b \rangle - \langle u, r \rangle - (\langle A^*u, c \rangle - \langle c, y \rangle) \} \\ &= \langle b, u \rangle - \sup_{\rho(r) \leq \epsilon} \langle u, r \rangle - \sup_c \{ \langle A^*u - y, c \rangle \} \\ &= \langle b, u \rangle - \epsilon \rho^\circ(u) - \delta_{A^*u=y}(u). \end{aligned}$$

The dual problem to  $\inf_{c \in \mathcal{C}} \langle c, y \rangle$  is given by maximizing  $\ell$  over  $u$ . Because of assumption (2.11) and the finiteness of  $-\sigma_{\mathcal{C}}(-y)$ ,

$$\inf_{c \in \mathcal{C}} \langle c, y \rangle = \sup_{y=A^*u} \{ \langle b, u \rangle - \epsilon \rho^\circ(u) \}, \quad (2.22)$$

and the supremum is attained, which is a consequence of Rockafellar (1970, Corollary 28.2.2 and Theorem 28.4). On the other hand, for any  $y \notin \text{cone } \mathcal{C}' \setminus \{0\}$ , we have from weak duality and (2.21) that

$$\sup_{y=A^*u} \{ \langle b, u \rangle - \epsilon \rho^\circ(u) \} \leq \inf_{c \in \mathcal{C}} \langle c, y \rangle \leq 0. \quad (2.23)$$

Since  $\text{dom } \kappa^\circ \cap \mathcal{C}' \neq \emptyset$ , we can substitute (2.22) into (2.18) and obtain

$$\begin{aligned} 0 < v_f &= \sup \{ \lambda \mid \kappa^\circ(y) \leq 1, -\sigma_{\mathcal{C}}(-y) \geq \lambda > 0 \} \\ &= \sup \{ \langle b, u \rangle - \epsilon \rho^\circ(u) \mid \kappa^\circ(A^*u) \leq 1, A^*u \in \text{cone } \mathcal{C}' \setminus \{0\} \} \\ &= \sup \{ \langle b, u \rangle - \epsilon \rho^\circ(u) \mid \kappa^\circ(A^*u) \leq 1 \} = v_l, \end{aligned}$$

where the last equality follows from (2.22), (2.23), and the positivity of  $v_f$ . This completes the first part of the proof. In particular, the Fenchel dual problem (2.16) has the same optimal value as the Lagrange dual problem  $(D_\ell)$ , and  $y^* = A^*u^*$  solves (2.16) if and only if  $u^*$  solves  $(D_\ell)$ . Moreover, since assumption (2.11) holds, §2.3.1 shows that  $(D)$  is equivalent to  $(D_\rho)$ . The conclusion now follows from these and Theorem 2.4.1.  $\square$

We next state a strong duality result concerning the primal-dual gauge pair  $(P_\rho)$  and  $(D_\rho)$ .

**Corollary 2.4.4.** Suppose that  $\mathcal{C}$  and  $\mathcal{C}_0$  are given by (2.9) and (2.10), where  $0 \leq \epsilon < \rho(b)$ . Suppose also that  $\kappa$  is closed,

$$\text{ri dom } \kappa \cap A^{-1} \text{ri } \mathcal{C}_0 \neq \emptyset, \quad \text{and} \quad \text{ri dom } \kappa^\circ \cap A^* \text{ri } \mathcal{C}'_0 \neq \emptyset. \quad (2.24)$$

Then the optimal values of  $(P_\rho)$  and  $(D_\rho)$  are attained, and their product is equal to 1.

*Proof.* Since  $A^{-1} \text{ri } \mathcal{C}_0 \neq \emptyset$ ,  $A$  satisfies the assumption in (2.11). Then §2.3.1 shows that (D) is equivalent to  $(D_\rho)$ . Moreover, from Rockafellar (1970, Theorem 6.6, Theorem 6.7), we see that  $\text{ri } \mathcal{C} = A^{-1} \text{ri } \mathcal{C}_0$  and  $\text{ri } \mathcal{C}' = A^* \text{ri } \mathcal{C}'_0$ . The conclusion now follows from Corollary 2.4.2.  $\square$

This last result also holds if  $\mathcal{C}_0$  were polyhedral; in that case, the assumptions (2.24) could be replaced with  $\text{ri dom } \kappa \cap \mathcal{C} \neq \emptyset$  and  $\text{ri dom } \kappa^\circ \cap \mathcal{C}' \neq \emptyset$ .

## 2.5 Variational properties of the gauge value function

Thus far, our analysis has focused on the relationship between the optimal values of the primal-dual pair  $(P_\rho)$  and  $(D_\rho)$ . As with Lagrange duality, however, there is also a fruitful view of dual solutions as providing sensitivity information on the primal optimal value. Here we provide a corresponding variational analysis of the gauge optimal-value function with respect to perturbations in  $b$  and  $\epsilon$ .

Sensitivity information is captured in the subdifferential of the value function

$$v(h, k) = \inf_x f(x, h, k), \quad (2.25)$$

with

$$f(x, h, k) = \kappa(x) + \delta_{\text{epi } \rho}(b + h - Ax, \epsilon + k). \quad (2.26)$$

Following the discussion in Aravkin, Burke, and Friedlander (2013, Section 4), we start by computing the conjugate of  $f$ , which can be done as follows:

$$\begin{aligned} f^*(z, y, \tau) &= \sup_{x, h, k} \{ \langle z, x \rangle + \langle y, h \rangle + \tau k - \kappa(x) - \delta_{\text{epi } \rho}(b + h - Ax, \epsilon + k) \} \\ &= \sup_{x, w, \mu} \{ \langle z + A^*y, x \rangle - \kappa(x) + \langle y, w \rangle + \tau \mu - \delta_{\text{epi } \rho}(w, \mu) \} - \langle b, y \rangle - \tau \epsilon \\ &= \kappa^*(z + A^*y) + \delta_{\text{epi } \rho}^*(y, \tau) - \langle b, y \rangle - \tau \epsilon. \end{aligned}$$

2.5. Variational properties of the gauge value function

---

Use Proposition 2.1.2(iv) and the definition of support function and convex conjugate to further transform this as

$$\begin{aligned}
 f^*(z, y, \tau) + \langle b, y \rangle + \tau\epsilon &= \delta_{\kappa^\circ(\cdot) \leq 1}(z + A^*y) + \sigma_{\text{epi } \rho}(y, \tau) \\
 &\stackrel{(i)}{=} \delta_{\kappa^\circ(\cdot) \leq 1}(z + A^*y) + \delta_{(\text{epi } \rho)^\circ}(y, \tau) \\
 &\stackrel{(ii)}{=} \delta_{\kappa^\circ(\cdot) \leq 1}(z + A^*y) + \delta_{\text{epi}(\rho^\circ)}(y, -\tau) \\
 &= \delta_{\kappa^\circ(\cdot) \leq 1}(z + A^*y) + \delta_{\rho^\circ(\cdot) \leq \cdot}(y, -\tau),
 \end{aligned}$$

where equality (i) follows from Proposition 2.1.6 and Proposition 2.1.4(i), and equality (ii) follows from Proposition 2.1.2(vi). Combining this with the definition of the value function  $v(h, k)$ ,

$$\begin{aligned}
 v^*(y, \tau) &= \sup_{h, k} \{ \langle y, h \rangle + \tau k - v(h, k) \} \\
 &= \sup_{x, h, k} \{ \langle y, h \rangle + \tau k - f(x, h, k) \} \\
 &= f^*(0, y, \tau) = -\langle b, y \rangle - \epsilon\tau + \delta_{\kappa^\circ(\cdot) \leq 1}(A^*y) + \delta_{\rho^\circ(\cdot) \leq \cdot}(y, -\tau).
 \end{aligned} \tag{2.27}$$

In view of Rockafellar and Wets (1998, Theorem 11.39), under a suitable constraint qualification, the set of subgradients of  $v$  is nonempty and given by

$$\begin{aligned}
 \partial v(0, 0) &= \operatorname{argmax}_{y, \tau} \{ -f^*(0, y, \tau) \} \\
 &= \operatorname{argmax}_{y, \tau} \{ \langle b, y \rangle + \epsilon\tau \mid \kappa^\circ(A^*y) \leq 1, \rho^\circ(y) \leq -\tau \} \\
 &= \left\{ (y, -\rho^\circ(y)) \mid y \in \operatorname{argmax}_y \{ \langle b, y \rangle - \epsilon\rho^\circ(y) \mid \kappa^\circ(A^*y) \leq 1 \} \right\},
 \end{aligned} \tag{2.28}$$

in terms of the solution set of  $(D_\ell)$  and the corresponding function value of  $\rho^\circ(y)$ . We state formally this result, which is a consequence of the above discussion and Corollary 2.4.3.

**Proposition 2.5.1.** For fixed  $(b, \epsilon)$ , define  $v$  as in (2.25) and  $f$  as in (2.26). Then

$$\operatorname{dom} f(\cdot, 0, 0) \neq \emptyset \iff 0 \in A \operatorname{dom} \kappa - [\rho(b - \cdot) \leq \epsilon],$$

and hence

$$(0, 0) \in \operatorname{int} \operatorname{dom} v \iff 0 \in \operatorname{int}(A \operatorname{dom} \kappa - [\rho(b - \cdot) < \epsilon])$$

If  $(0, 0) \in \text{int dom } v$  and  $v(0, 0) > 0$ , then  $\partial v(0, 0) \neq \emptyset$  with

$$\begin{aligned} \partial v(0, 0) &= \left\{ (y, -\rho^\circ(y)) \mid y \in \underset{y}{\operatorname{argmax}} \{ \langle b, y \rangle - \epsilon \rho^\circ(y) \mid \kappa^\circ(A^*y) \leq 1 \} \right\} \\ &= \left\{ v(0, 0) \cdot (y, -\rho^\circ(y)) \mid y \in \underset{y}{\operatorname{argmin}} \{ \kappa^\circ(A^*y) \mid \langle b, y \rangle - \epsilon \rho^\circ(y) \geq 1 \} \right\}. \end{aligned}$$

*Proof.* It is routine to verify the properties of the domain of  $f(\cdot, 0, 0)$  and the interior of the domain of  $v$ . Suppose that  $(0, 0) \in \text{int dom } v$ . Then the value function is continuous at  $(0, 0)$  and hence  $\partial v(0, 0) \neq \emptyset$ . The first expression of  $\partial v(0, 0)$  follows directly from Rockafellar and Wets (1998, Theorem 11.39) and the discussions preceding this proposition.

We next derive the second expression of  $\partial v(0, 0)$ . Since  $(0, 0) \in \text{int dom } v$  implies  $0 \in \text{int}(A \text{ dom } \kappa - [\rho(b - \cdot) < \epsilon])$ , the linear map  $A$  satisfies assumption 2.11. Moreover, as another consequence of Rockafellar and Wets (1998, Theorem 11.39),  $(0, 0) \in \text{int dom } v$  also implies that  $v(0, 0) = \sup_{y, \tau} \{-f^*(0, y, \tau)\}$ , which is just the optimal value of the Lagrange dual problem  $(D_\ell)$ . Furthermore,  $v(0, 0)$  being finite and nonzero together with the definition of  $(D_\ell)$  and (2.12) implies that  $\text{dom } \kappa^\circ \cap \mathcal{C}' \neq \emptyset$ . The second expression of  $\partial v(0, 0)$  now follows from these three observations and Corollary 2.4.3.  $\square$

## 2.6 Extensions

The following examples illustrate how to extend the canonical formulation  $(P_\rho)$  to accommodate related problems. It also provides an illustration of the techniques that can be used to pose problems in gauge form and how to derive their corresponding gauge duals.

### 2.6.1 Composition and conic side constraints

A useful generalization of  $(P_\rho)$  is to allow the gauge objective to be composed with a linear map, and for the addition of conic side constraints. The composite objective can be used to capture, for example, problems such as weighted basis pursuit (e.g., Candés, Wakin, and Boyd (2008); Friedlander, Mansour, Saab, and Yilmaz (2012)), or together with the conic constraint, problems such as nonnegative total variation (Krishnan, Lin, and Yip, 2007).

The following result generalizes the canonical primal-dual gauge pair  $(P_\rho)$  and  $(D_\rho)$ .

**Proposition 2.6.1.** Let  $D$  be a linear map and  $\mathcal{K}$  be a convex cone. The following pair of problems constitute a primal-dual gauge pair:

$$\underset{x}{\text{minimize}} \quad \kappa(Dx) \quad \text{subject to} \quad \rho(b - Ax) \leq \epsilon, \quad x \in \mathcal{K}, \quad (2.29a)$$

$$\underset{y, z}{\text{minimize}} \quad \kappa^\circ(z) \quad \text{subject to} \quad \langle y, b \rangle - \epsilon \rho^\circ(y) \geq 1, \quad D^*z - A^*y \in \mathcal{K}^*. \quad (2.29b)$$

*Proof.* Reformulate (2.29a) as a gauge optimization problem by introducing additional variables, and lifting both the cone  $\mathcal{K}$  and the epigraph  $\text{epi } \rho$  into the objective by means of their indicator functions: use the function  $f(x, s, r, \tau) := \delta_{\mathcal{K}}(x) + \kappa(s) + \delta_{\text{epi } \rho}(r, \tau)$  to define the equivalent gauge optimization problem

$$\underset{x, s, r, \tau}{\text{minimize}} \quad f(x, s, r, \tau) \quad \text{subject to} \quad Dx = s, \quad Ax + r = b, \quad \tau = \epsilon.$$

As with §2.3.1, observe that  $f$  is a sum of gauges on disjoint variables. Thus, we invoke Proposition 2.1.5 to deduce the polar of the above objective:

$$\begin{aligned} f^\circ(u, z, y, \alpha) &= \max \{ \delta_{\mathcal{K}^\circ}^\circ(u), \kappa^\circ(z), \delta_{\text{epi } \rho}^\circ(y, \alpha) \} \\ &\stackrel{(i)}{=} \max \{ \delta_{\mathcal{K}^\circ}(u), \kappa^\circ(z), \delta_{(\text{epi } \rho)^\circ}(y, \alpha) \} \\ &\stackrel{(ii)}{=} \max \{ \delta_{\mathcal{K}^*}(-u), \kappa^\circ(z), \delta_{\text{epi}(\rho^\circ)}(y, -\alpha) \} \\ &\stackrel{(iii)}{=} \delta_{\mathcal{K}^*}(-u) + \kappa^\circ(z) + \delta_{\text{epi}(\rho^\circ)}(y, -\alpha), \end{aligned}$$

where (i) follows from Proposition 2.1.4(i), (ii) follows from Proposition 2.1.2(vi), and (iii) follows from the definition of indicator function. Moreover, use Corollary 2.3.2 to derive the antipolar of the linear constraint set

$$\mathcal{C} = \{ (x, s, r, \tau) \mid Dx = s, Ax + r = b, \tau = \epsilon \},$$

which leads us to

$$\mathcal{C}' = \{ (-D^*z + A^*y, z, y, \alpha) \mid \langle b, y \rangle + \epsilon \alpha \geq 1 \}.$$

From the above discussion, we obtain the following gauge program

$$\underset{y, z, \alpha}{\text{minimize}} \quad \delta_{\mathcal{K}^*}(D^*z - A^*y) + \kappa^\circ(z) + \delta_{\text{epi}(\rho^\circ)}(y, -\alpha) \quad \text{subject to} \quad \langle b, y \rangle + \epsilon \alpha \geq 1.$$

Bringing the indicator functions down to the constraints leads to

$$\underset{y, z, \alpha}{\text{minimize}} \quad \kappa^\circ(z) \quad \text{subject to} \quad \langle y, b \rangle + \epsilon \alpha \geq 1, \quad \rho^\circ(y) \leq -\alpha, \quad D^*z - A^*y \in \mathcal{K}^*;$$

further simplification by eliminating  $\alpha$  yields the gauge dual problem (2.29b).  $\square$

## 2.6.2 Nonnegative conic optimization

Conic optimization subsumes a large class of convex optimization problems that ranges from linear, to second-order, to semidefinite programming, among others. Example 1.2.3 describes how a general conic optimization problem can be reformulated as an equivalent gauge problem; see (1.16).

We can easily accommodate a generalization of (1.16) by embedding it within the formulation defined by (1.15a), and define

$$\underset{x}{\text{minimize}} \quad \langle c, x \rangle + \delta_{\mathcal{K}}(x) \quad \text{subject to} \quad \rho(b - Ax) \leq \epsilon, \quad (2.30)$$

with  $c \in \mathcal{K}^*$ , as the conic gauge optimization problem. The following result describes its gauge dual.

**Proposition 2.6.2.** Suppose that  $\mathcal{K} \subset \mathcal{X}$  is a convex cone and  $c \in \mathcal{K}^*$ . Then the gauge

$$\kappa(x) = \langle c, x \rangle + \delta_{\mathcal{K}}(x)$$

has the polar

$$\kappa^\circ(u) = \inf \{ \alpha \geq 0 \mid \alpha c \in \mathcal{K}^* + u \}, \quad (2.31)$$

with  $\text{dom } \kappa^\circ = \text{span}\{c\} - \mathcal{K}^*$ . If  $\mathcal{K}$  is closed and  $c \in \text{int } \mathcal{K}^*$ , then  $\kappa$  has compact level sets, and  $\text{dom } \kappa^\circ = \mathcal{X}$ .

*Proof.* From Proposition 2.1.2, we have that

$$\begin{aligned} \kappa^\circ(u) &= \sup \{ \langle u, x \rangle \mid \kappa(x) \leq 1 \} \\ &= \sup \{ \langle u, x \rangle \mid \langle c, x \rangle \leq 1 \text{ and } x \in \mathcal{K} \} \\ &= \inf \{ \alpha \geq 0 \mid \alpha c - u \in \mathcal{K}^* \}, \end{aligned} \quad (2.32)$$

where the strong (Lagrangian) duality relationship in the last equality stems from the following argument. First consider the case where  $u \in \text{dom } \kappa^\circ$ . Because the maximization problem in (2.32) satisfies Slater's condition, equality follows from Rockafellar (1970, Corollary 28.2.2 and Theorem 28.4). Next, consider the case where  $u \notin \text{dom } \kappa^\circ$ , where  $\kappa^\circ(u) = +\infty$ . The last equality then follows from weak duality. For the domain, note that the minimization problem is feasible if and only if  $u \in \text{span}\{c\} - \mathcal{K}^*$ ; hence  $\text{dom } \kappa^\circ = \text{span}\{c\} - \mathcal{K}^*$ .

To prove compactness of the level sets of  $\kappa$  when  $\mathcal{K}$  is closed and  $c \in \text{int } \mathcal{K}^*$ , define  $\gamma := \inf_x \{ \langle c, x \rangle \mid \|x\| = 1, x \in \mathcal{K} \}$  and observe that compactness of the feasible set in this minimization implies that the infimum is attained and that  $\gamma > 0$ . Thus, for any  $x \in \mathcal{K} \setminus \{0\}$ ,  $\langle c, x \rangle \geq \gamma \|x\| > 0$  and, consequently, that  $\{x \in \mathcal{X} \mid \kappa(x) \leq \alpha\} = \{x \in \mathcal{K} \mid \langle c, x \rangle \leq \alpha\} \subset \{x \in \mathcal{X} \mid \|x\| \leq \alpha/\gamma\}$ .

This guarantees that the level sets of  $\kappa$  are bounded, which establishes their compactness. From this and Proposition 2.1.2(iii), we see that  $\kappa^\circ(u)$  is finite for any  $u \in \mathcal{X}$ .  $\square$

**Remark 2.6.1.** Note that even though the polar gauge in (2.31) is closed, it is not necessarily the case that it has a closed domain. For example, let  $\mathcal{K}$  be the cone of PSD 2-by-2 matrices, and define

$$c = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad u_n = \begin{pmatrix} 0 & 1 \\ 1 & -\frac{1}{n} \end{pmatrix},$$

for each  $n = 1, 2, \dots$ . Use the expression (2.31) to obtain that  $\kappa^\circ(u_n) = n$ . Hence,  $u_n \in \text{dom } \kappa^\circ$ , but  $\lim_{n \rightarrow +\infty} u_n \notin \text{dom } \kappa^\circ$ .

This is an example of a more general result described by Ramana, Tunçel, and Wolkowicz (1997, Lemma 2.2), which shows that the cone of PSD matrices is *devious* (i.e., for every nontrivial proper face  $F$  of  $\mathcal{K}$ ,  $\text{span } F + \mathcal{K}$  is not closed). The concept of a devious cone seems to be intimately related to the closedness of the domain of polar gauges such as (2.31) because  $\text{span}\{c\} - \mathcal{K}^* = -(\text{span } F + \mathcal{K}^*)$ , where  $F \subseteq \mathcal{K}^*$  is the smallest face of  $\mathcal{K}^*$  that contains  $c$ ; see Tunçel and Wolkowicz (2012, Proposition 3.2).

With that in mind, it is interesting to derive a representation for the closure of the domain of (2.31). It follows from Rockafellar (1970, Corollary 16.4.2) that  $\text{cl dom } \kappa^\circ = \text{cl}(\text{span}\{c\} - \mathcal{K}^*) = (\{c\}^\perp \cap \text{cl } \mathcal{K})^\circ$ .

# Chapter 3

## Spectral gauge optimization and duality

The gauge duality framework developed in Chapter 2, albeit inspired by convex formulations arising from practical inverse problems, was still approached in a rather abstract form.

In this chapter we connect this general class of problems with those convex relaxations for phase recovery and blind deconvolution presented in §1.1. Our goal is to derive their corresponding dual problems and to show that they admit a unified description suitable for the design of numerical methods for their minimization.

Due to its expressiveness and direct relationship to the PhaseLift formulation, we begin by specializing the nonnegative conic gauge optimization class from §2.6.2 to matrix problems in which the abstract cone is instantiated to be that of positive semidefinite Hermitian matrices. This specialization along with further analysis lead us to a constrained eigenvalue minimization problem derived from the dual gauge problem, which will serve as the basis for our approach described in Chapter 4.

This is followed by an analogous derivation of a dual gauge problem associated to the nuclear norm relaxation described in §1.1.2, which is then further reduced to a problem of the same form as that for PhaseLift—thus allowing for a unified approach to both problems.

We conclude this chapter with a construction of a family of lower approximations to this spectral gauge dual objective and a discussion of their possible application within the framework of proximal bundle methods for nonsmooth convex minimization.

### 3.1 Semidefinite nonnegative conic optimization

For our purposes, an instance of semidefinite nonnegative conic optimization is a problem of the form

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \langle C, X \rangle \quad \text{subject to} \quad \rho(b - \mathcal{A}X) \leq \epsilon \text{ and } X \succeq 0, \quad (3.1)$$

### 3.1. Semidefinite nonnegative conic optimization

---

where  $C \succeq 0$ ,  $\rho : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  is a closed gauge function such that  $\rho^{-1}(0) = \{0\}$ ,  $b \in \mathbb{R}^m \setminus \{0\}$ ,  $\epsilon < \rho(b)$ , and  $\mathcal{A} : \mathcal{H}^n \rightarrow \mathbb{R}^m$  is a linear map defined by matrices  $A_k \in \mathcal{H}^n$  in such a way that

$$(\mathcal{A}X)_k = \langle X, A_k \rangle := \text{trace}(A_k^* X) = \text{trace}(A_k X)$$

for each  $k = 1, \dots, m$  and all  $X \in \mathcal{H}^n$ .

We begin with the specialization of the general nonnegative conic gauge optimization presented in §2.6.2 to problems of the form (3.1). In that abstract formulation, we take  $\mathcal{X} = \mathcal{H}^n$  to be the real vector space of  $n$ -by- $n$  Hermitian matrices,  $\mathcal{K} \subset \mathcal{X}$  to be the (self-dual) cone of positive semidefinite matrices,  $c = C$  and  $A = \mathcal{A}$ , this specialization translates (2.30) to the problem

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \kappa(X) := \langle C, X \rangle + \delta_{\succeq 0}(X) \quad \text{subject to} \quad \rho(b - \mathcal{A}X) \leq \epsilon, \quad (3.2)$$

where the PSD constraint from (3.1) is moved to the objective function turning it into the gauge  $\kappa$ , a technique used in §2.6.2 for more general convex cones.

Invoking Proposition 2.6.2, we have that the gauge polar to  $\kappa$  is given by

$$\begin{aligned} \kappa^\circ(U) &= \inf_{\alpha \in \mathbb{R}} \{ \alpha \geq 0 \mid \alpha C \succeq U \} \\ &= \max\{0, \inf\{ \alpha \in \mathbb{R} \mid \alpha C \succeq U \}\} \\ &= \max\{0, \lambda_1(U, C)\} \\ &= [\lambda_1(U, C)]_+, \end{aligned}$$

where  $\lambda_1(U, C)$  denotes the rightmost generalized eigenvalue—i.e., largest real *not* in absolute value—corresponding to the generalized eigenvalue problem  $Ux = \lambda Cx$  (which might be  $+\infty$  as in the example given in Remark 2.6.1).

With this characterization, the abstract gauge dual  $(D_\rho)$  derived in §2.3.1 specializes to the spectral optimization problem

$$\underset{y \in \mathbb{R}^m}{\text{minimize}} \quad [\lambda_1(\mathcal{A}^* y, C)]_+ \quad \text{subject to} \quad \langle b, y \rangle - \epsilon \rho^\circ(y) \geq 1, \quad (3.3)$$

which is feasible for  $0 \leq \epsilon < \rho(b)$ , but not necessarily finite due to assuming only semidefiniteness of  $C$ ; c.f. Remark 2.6.1.

In the following, we further specialize and analyze this primal-dual pair to its most commonly encountered form from convex relaxations.

#### 3.1.1 Trace minimization in the PSD cone

The lifted formulation of phase retrieval is an example of the semidefinite conic gauge optimization problem (3.2) where  $C$  is the identity matrix and  $\rho : \mathbb{C}^m \rightarrow \mathbb{R}$  is some norm  $\|\cdot\|$  (typically the 2-norm).

### 3.1. Semidefinite nonnegative conic optimization

For these cases, (3.2) reduces to the problem of minimizing the trace of a PSD matrix that satisfies a prescribed bound on the (norm-measured) linear measurements misfit, while the dual consists of a constrained eigenvalue minimization problem. This neatly specializes the abstract primal-dual gauge pair

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad \rho(b - Ax) \leq \epsilon, \quad (3.4a)$$

$$\underset{y \in \mathcal{Y}}{\text{minimize}} \quad \kappa^\circ(A^*y) \quad \text{subject to} \quad \langle y, b \rangle - \epsilon \rho^\circ(y) \geq 1, \quad (3.4b)$$

to the spectral gauge pair that forms the core of our approach

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \text{trace } X + \delta(X \mid \cdot \succeq 0) \quad \text{subject to} \quad \|b - \mathcal{A}X\| \leq \epsilon, \quad (3.5a)$$

$$\underset{y \in \mathbb{R}^m}{\text{minimize}} \quad [\lambda_1(\mathcal{A}^*y)]_+ \quad \text{subject to} \quad \langle y, b \rangle - \epsilon \|y\|_* \geq 1, \quad (3.5b)$$

where  $\|\cdot\|_*$  denotes the norm dual to  $\|\cdot\|$  and  $\mathcal{A}^*y = \sum_{k=1}^m A_k y_k$ .

Figure 3.1 illustrates the geometry of the gauge dual feasible set in (3.5b) for different norms and values of  $\epsilon$  in the case where  $m = 2$  and  $b = (0, 1) \in \mathbb{R}^2$ . The generalized “unit balls” of the gauge trace  $X + \delta(X \mid \cdot \succeq 0)$  and its corresponding polar  $[\lambda_1(U)]_+$  are depicted in Figure 3.2 for 2-by-2 real symmetric matrices parameterized as  $\begin{bmatrix} x, & \frac{z}{\sqrt{2}}; & \frac{z}{\sqrt{2}}, & y \end{bmatrix}$ , where  $(x, y, z)$  correspond to the coordinate axes in  $\mathbb{R}^3$ . Figures 3.3 and 3.4 provide alternative viewing positions.

Assuming the original problem with the cone constraints is feasible and  $b \neq 0$ , we can further simplify the dual objective and safely eliminate the positive-part operator: since in this case  $\kappa(X)$  is obviously strictly positive for all nonzero  $X$ , and is additionally finite over the feasible set of the original cone-constrained problem, it follows from weak gauge duality (cf. Theorem 2.4.1) that  $\kappa^\circ(\mathcal{A}^*y)$  is positive for all dual feasible points. In other words,

$$0 < [\lambda_1(\mathcal{A}^*y)]_+ = \lambda_1(\mathcal{A}^*y) \quad (3.6)$$

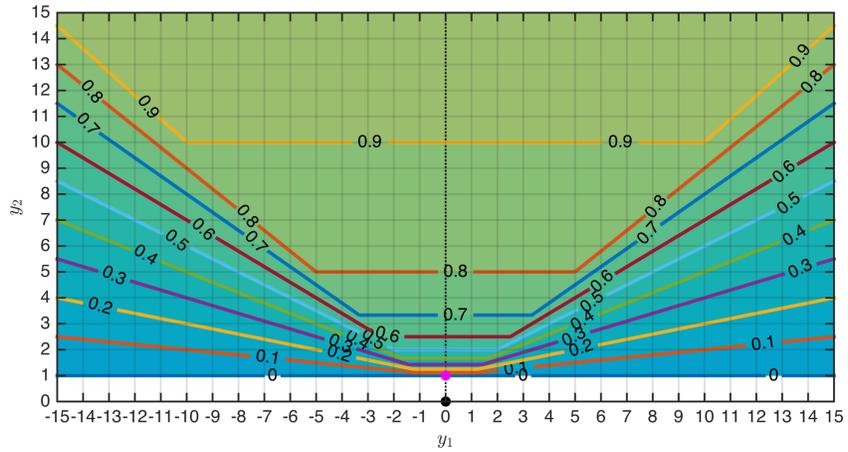
for all dual feasible points  $y$ , leading to the equivalent spectral problem

$$\underset{y \in \mathbb{R}^m}{\text{minimize}} \quad \lambda_1(\mathcal{A}^*y) \quad \text{subject to} \quad \langle b, y \rangle - \epsilon \|y\|_* \geq 1, \quad (3.7)$$

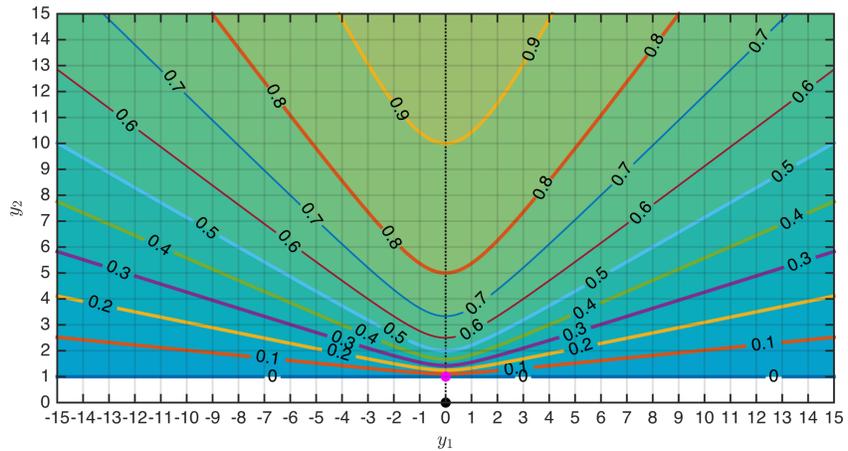
which will play a central role in our computational approach described in Chapter 4.

In practice, we need to be prepared to detect infeasibility. The failure of condition (3.6) in fact furnishes a certificate of infeasibility for the original cone-constrained problem (3.1): if  $\lambda_1(\mathcal{A}^*y) \leq 0$  for some dual-feasible vector  $y$ , it follows from weak gauge duality that  $\kappa(X)$  is necessarily infinite over the feasible set of (3.5a)—i.e.,  $X \not\succeq 0$  for all  $X$  feasible in problem (3.5a).

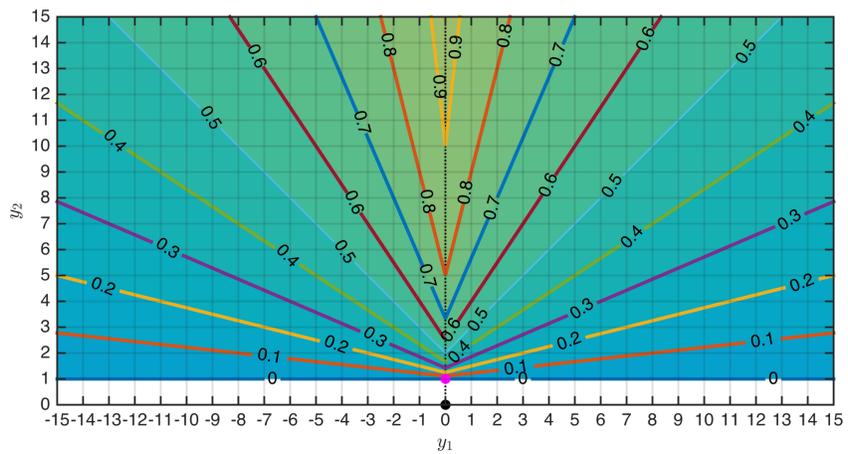
### 3.1. Semidefinite nonnegative conic optimization



(a)  $\|\cdot\| = \|\cdot\|_1$  and  $\|\cdot\|_* = \|\cdot\|_\infty$



(b)  $\|\cdot\| = \|\cdot\|_2$  and  $\|\cdot\|_* = \|\cdot\|_2$



(c)  $\|\cdot\| = \|\cdot\|_\infty$  and  $\|\cdot\|_* = \|\cdot\|_1$

Figure 3.1: Visualizing the dual feasible set for  $b = (0, 1)$  and  $\epsilon \in [0 : .1 : .9]$ .

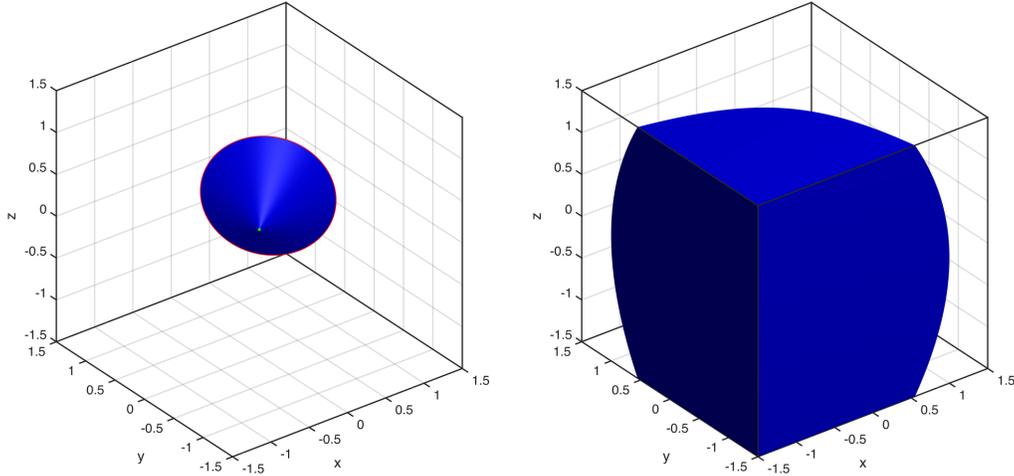


Figure 3.2: Generalized unit balls of the trace-gauge and its polar for real symmetric 2-by-2 matrices (intersected with  $[-1.5, 1.5]^3$  for illustration purposes). Depictions above correspond to the parameterization  $\left[x, \frac{z}{\sqrt{2}}; \frac{z}{\sqrt{2}}, y\right]$ . Figures 3.3 and 3.4 provide different viewing positions.

## 3.2 Nuclear-norm minimization

Recall the nuclear-norm minimization problem described in Chapter 1,

$$\underset{X \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad \|X\|_1 = \sum_i \sigma_i(X) \quad \text{subject to} \quad \|b - \mathcal{A}X\| \leq \epsilon,$$

where the measurement operator  $\mathcal{A} : \mathbb{C}^{n_1 \times n_2} \rightarrow \mathbb{C}^m$  is defined by matrices  $A_k \in \mathbb{C}^{n_1 \times n_2}$  as  $(\mathcal{A}X)_k := \langle X, A_k \rangle := \text{trace}(A_k^* X)$ ,  $k = 1, \dots, m$ , and  $\epsilon < \|b\|$ .

The dual to this gauge minimization problem admits a far simpler treatment if we exploit the norm structure of the objective. By observing that its polar is the corresponding dual norm (i.e., the operator norm  $\|\cdot\|_\infty = \sigma_1(\cdot)$ ) and identifying  $\mathcal{Y} = \mathbb{C}^m$  as a real  $2m$ -dimensional vector space with the inner-product  $\Re \langle \cdot, \cdot \rangle$ , we have the following primal-dual gauge pair

$$\underset{X \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad \|X\|_1 \quad \text{subject to} \quad \|b - \mathcal{A}X\| \leq \epsilon, \quad (3.8a)$$

$$\underset{y \in \mathbb{C}^m}{\text{minimize}} \quad \|\mathcal{A}^* y\|_\infty \quad \text{subject to} \quad \Re \langle y, b \rangle - \epsilon \|y\|_* \geq 1, \quad (3.8b)$$

where  $\mathcal{A}^* y = \sum_{k=1}^m A_k y_k \in \mathbb{C}^{n_1 \times n_2}$  is the adjoint of  $\mathcal{A}$  evaluated at  $y$ .

### 3.2. Nuclear-norm minimization

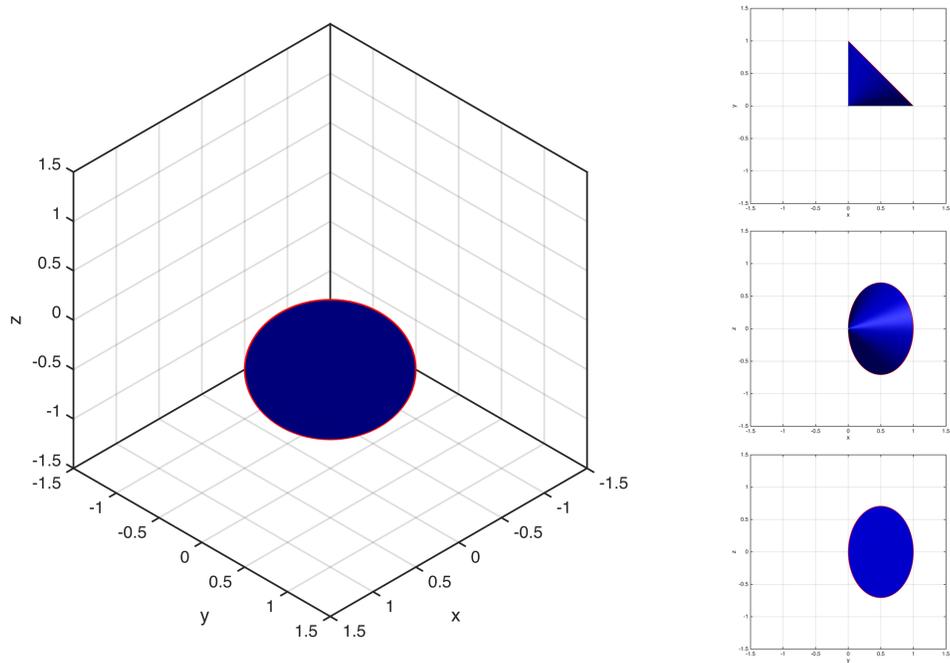


Figure 3.3: Generalized unit ball of the trace-gauge for symmetric 2-matrices.

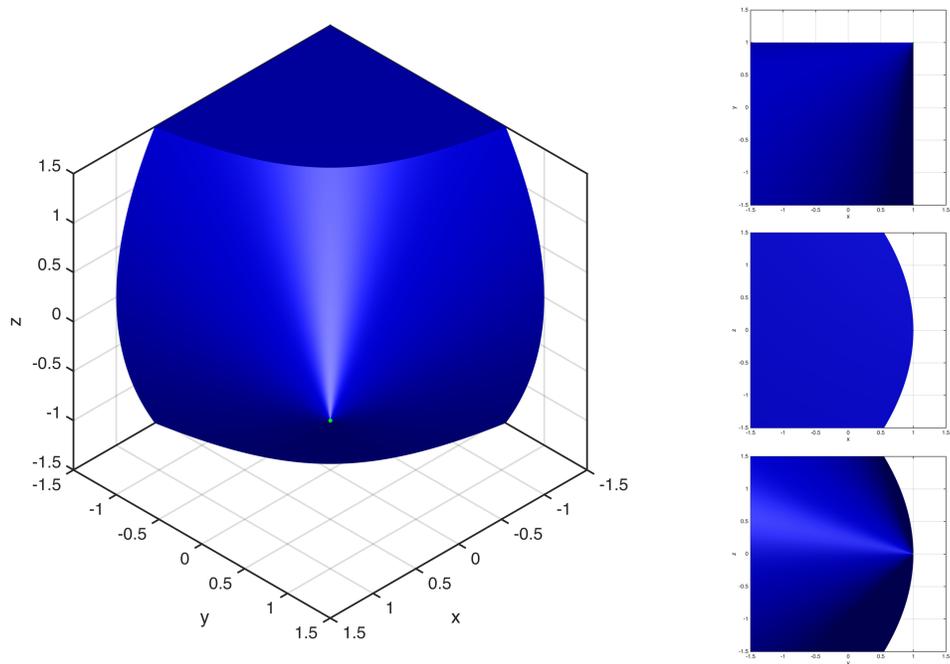


Figure 3.4: Generalized unit ball of the polar to the trace-gauge for real symmetric 2-matrices. Set is unbounded, so it is illustrated here within a box.

### 3.2.1 Reduction to PSD trace minimization

It will be convenient, for both theoretical and algorithmic developments of the approach we discuss in Chapter 4, to embed the nuclear-norm minimization problem (3.8a) within the symmetric SDP formulation (3.5a). The results are no less general, and it will allow us to solve both problems with what is essentially a single algorithmic approach and software implementation.

The reduction to the Hermitian trace-minimization problem (3.5a) leverages the SDP formulation of the nuclear norm as introduced by Fazel (2002) and used by Ahmed et al. (2014) for their algorithmic approach, as we discussed in §1.1.2. It embeds the problem in a larger Hermitian space and separates the complex measurements into their real and imaginary parts, leading to a problem of the form

$$\begin{aligned}
 & \underset{\substack{U \in \mathcal{H}^{n_1}, V \in \mathcal{H}^{n_2} \\ X \in \mathbb{C}^{n_1 \times n_2} \\ r_1, r_2 \in \mathbb{R}^m}}{\text{minimize}} & \left\langle \frac{1}{2} \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}, \begin{pmatrix} U & X \\ X^* & V \end{pmatrix} \right\rangle \\
 & \text{subject to} & \left\langle \frac{1}{2} \begin{pmatrix} 0 & A_k \\ A_k^* & 0 \end{pmatrix}, \begin{pmatrix} U & X \\ X^* & V \end{pmatrix} \right\rangle + r_{1k} = \Re b_k, \\
 & & \left\langle \frac{i}{2} \begin{pmatrix} 0 & A_k \\ -A_k^* & 0 \end{pmatrix}, \begin{pmatrix} U & X \\ X^* & V \end{pmatrix} \right\rangle + r_{2k} = \Im b_k, \\
 & & \|r_1 + ir_2\| \leq \epsilon, \quad \begin{pmatrix} U & X \\ X^* & V \end{pmatrix} \succeq 0, \quad k = 1, \dots, m,
 \end{aligned} \tag{3.9}$$

where the residual vectors  $r_1, r_2 \in \mathbb{R}^m$  are introduced simply to allow a compact presentation.

The gauge dual obtained from this problem when instantiated in (3.3) will have the exact same form as (3.8b) once the dual variables  $y_1, y_2 \in \mathbb{R}^m$  are “compactified” back into a complex vector  $y = y_1 + iy_2 \in \mathbb{C}^m$  and the objective is simplified by observing the following equalities

$$\begin{aligned}
 \left[ \lambda_1 \left( \begin{pmatrix} 0 & \frac{1}{2}(\mathcal{A}^*y) \\ \frac{1}{2}(\mathcal{A}^*y)^* & 0 \end{pmatrix}, \frac{1}{2}I \right) \right]_+ &= \left[ \lambda_1 \left( \begin{pmatrix} 0 & \mathcal{A}^*y \\ (\mathcal{A}^*y)^* & 0 \end{pmatrix} \right) \right]_+ \\
 &= [\|\mathcal{A}^*y\|_\infty]_+ \\
 &= \|\mathcal{A}^*y\|_\infty,
 \end{aligned}$$

where we use the fact that the eigenvalues of the matrix  $[0, Z; Z^*, 0]$  are  $\pm\sigma_k(Z)$  for each  $k = 1, \dots, \min\{n_1, n_2\}$  and  $|n_1 - n_2|$  zeros.

Based on this reduction, from now on we focus entirely on the PSD trace minimization formulation (3.5a), its gauge dual (3.5b) and the corresponding constrained eigenvalue optimization problem (3.7).

### 3.3 Spectral lower models for convex minimization

In the next chapter we describe a computational approach for problem (3.5b) that is based on a projected subgradient method (PSGM) to solve that convex eigenvalue minimization problem. Although simple and promising (c.f., experiments described in Chapter 5), that approach inherits some difficulties from PSGMs—e.g., lack of natural stopping criteria—for which workarounds need to be devised.

Bundle methods for nonsmooth convex minimization provide natural mechanisms to address such issues, however they require good (lower) approximations for the objective function. The main goal of this section is to describe the construction of a family of such lower models that seems suitable for the spectral objectives appearing in the dual gauge problems from §3.1 and §3.2.

Although promising, these approximations have yet to be thoroughly exploited in the design of computational alternatives to the PSGM approach. To support the first steps towards such an endeavour, the following subsections describe the construction of these lower models and briefly outline their use within the context of proximal bundle methods.

**An overview.** Many classical methods for nonsmooth convex minimization have been designed for problems of the form

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad x \in F, \quad (\text{CM})$$

where  $f : \mathcal{X} \rightarrow \mathbb{R}$  is a convex objective function and  $F \subseteq \mathcal{X}$  is the set of feasible points, a closed and convex subset of a finite-dimensional real Euclidean space  $\mathcal{X}$ . Due to the simplicity of the dual constraints in (3.5b) and the nonsmoothness of the objective, we focus on casting that dual gauge problem in this framework and specialize the numerical methods to that class of problems.

In order to compute solutions for problems in the general formulation (CM), many methods assume that it is possible to query the objective function (seen as a “black-box”) for its value and subgradients at given points ( $F$  might allow for the solution of certain quadratic problems associated with it as an additional constraint, e.g., projection onto  $F$ ). A common structure in these methods is that such functional values and subgradients are used to construct suitable lower models for the original objective function (e.g., bundle methods (Bonnans, Gilbert, Lemaréchal, and Sagastizábal, 2006, Chapter 10)).

Following this reasoning, we proceed by presenting a family of lower models for the dual gauge objective in (3.5b) and will assume that the constraints in

that problem are simple enough to allow for the operations required in those methods to be performed without significant overhead (since those constraints can be transformed into a particularly simple second-order cone constraint, we believe this is not unreasonable).

### 3.3.1 Spectral lower models for the dual gauge objective

Our approach to construct lower models for the dual gauge objective in (3.5b) is based on forming outer-approximations for its epigraph in a manner equivalent to that used by Helmberg and Rendl (2000).

Whenever  $U \in \mathbb{C}^{n \times k}$  is an arbitrary matrix with  $C$ -orthonormal columns (i.e.,  $U^*CU = I$ ), we have that

$$\begin{aligned}
 \text{epi} [\lambda_1(\mathcal{A}^* \cdot, C)]_+ &= \left\{ (y, t) \in \mathcal{Y} \times \mathbb{R}_+ \mid t \geq \lambda_1(\mathcal{A}^*y, C) \right\} \\
 &= \left\{ (y, t) \in \mathcal{Y} \times \mathbb{R}_+ \mid tC \succcurlyeq \mathcal{A}^*y \right\} \\
 &\subseteq \left\{ (y, t) \in \mathcal{Y} \times \mathbb{R}_+ \mid tU^*CU \succcurlyeq U^*(\mathcal{A}^*y)U \right\} \\
 &= \left\{ (y, t) \in \mathcal{Y} \times \mathbb{R}_+ \mid tI \succcurlyeq U^*(\mathcal{A}^*y)U \right\} \\
 &= \left\{ (y, t) \in \mathcal{Y} \times \mathbb{R}_+ \mid t \geq \lambda_1(U^*(\mathcal{A}^*y)U) \right\} \\
 &= \text{epi} [\lambda_1(U^*(\mathcal{A}^* \cdot)U)]_+ \\
 &= \text{epi} [\lambda_1(\mathcal{A}_U^* \cdot)]_+,
 \end{aligned}$$

where we define the reduced measurement operator  $\mathcal{A}_U : \mathcal{H}^k \rightarrow \mathbb{C}^m$  as  $\mathcal{A}_U(Z) := \mathcal{A}(UZU^*)$ , which has an adjoint  $\mathcal{A}_U^* : \mathbb{C}^m \rightarrow \mathcal{H}^k$  defined by  $\mathcal{A}_U^*y = U^*(\mathcal{A}^*y)U$ .

The advantage of using such approximations is that, while the original function involved a typically very large linear matrix inequality (LMI) with matrices of order  $n$ , the outer-approximation is described by LMIs of a given order  $k$ . As long as we can control this order  $k$ , we have the possibility to deal with much reduced problems (an observation also made in Helmberg and Rendl (2000) and exploited via proximal bundle methods (Bonnans et al., 2006, Chapter 10)).

#### Relationship with approximate subdifferentials

Key observations for the suitability of such lower models show up when they are built using generalized eigenvectors of the adjoint measurement operator at a given point and related to approximate subdifferentials of the objective at

feasible points. To that end, let us begin with a simple characterization of the subdifferential of the abstract gauge objective in (3.4b) presented in

**Proposition 3.3.1.** Let  $\kappa : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a closed gauge function with compact level-sets and  $A : \mathcal{X} \rightarrow \mathcal{Y}$  be a linear map. Then, for each  $\varepsilon \geq 0$ , we have that

$$g \in \partial_\varepsilon(\kappa^\circ(A^*\cdot))(y) \iff g = Au, \kappa(u) \leq 1 \text{ and } \varepsilon \geq \kappa^\circ(A^*y) - \langle A^*y, u \rangle.$$

*Proof.* Since  $\kappa$  has compact level-sets, we have that  $\kappa^\circ$  is finite-valued everywhere and the first and second of the following chain of equivalences hold by (Hiriart-Urruty and Lemaréchal, 1993, Theorem XI.3.2.1) and (Hiriart-Urruty and Lemaréchal, 1993, Proposition XI.1.2.1), respectively:

$$\begin{aligned} g \in \partial_\varepsilon(\kappa^\circ(A^*\cdot))(y) &\iff g = Au \text{ and } u \in \partial_\varepsilon(\kappa^\circ(\cdot))(A^*y) \\ &\iff g = Au \text{ and } (\kappa^\circ)^*(u) + \kappa^\circ(A^*y) - \langle A^*y, u \rangle \leq \varepsilon \\ &\iff g = Au \text{ and } \delta_{\kappa(\cdot) \leq 1}(u) + \kappa^\circ(A^*y) - \langle A^*y, u \rangle \leq \varepsilon \\ &\iff g = Au, \kappa(u) \leq 1 \text{ and } \kappa^\circ(A^*y) - \langle A^*y, u \rangle \leq \varepsilon. \quad \square \end{aligned}$$

By noticing that, if the primal (3.5a) is feasible, weak gauge duality and Slater's constraints qualification in the Lagrangian dual of (3.5a) give us

$$0 < (\text{trace } X)^{-1} \leq [\lambda_1(\mathcal{A}^*y, C)]_+ = \lambda_1(\mathcal{A}^*y, C),$$

for every (dual) feasible point in (3.5b).

Observing these two results and that the assumption  $C \succ 0$  implies that the primal gauge objective  $\kappa(X) = \langle C, X \rangle + \delta_{\succ 0}(X)$  has compact level sets (c.f. Proposition 2.6.2), we can relate our family of outer-approximations with a class of subdifferential-based lower models for our objective in (3.5b) centered at a given dual feasible point. Let  $\hat{y} \in \mathcal{Y}$  be feasible in (3.5b),  $U \in \mathbb{C}^{n \times k}$  have as columns generalized eigenvectors corresponding to the (decreasingly ordered) rightmost generalized eigenvalues of  $\mathcal{A}^*\hat{y}$  with respect to  $C$  (counting multiplicities) and  $Z \in \mathcal{H}^k$  be such that  $\text{trace } Z = 1$  and  $Z \succcurlyeq 0$ , then

$$\begin{aligned} \langle \hat{y}, \mathcal{A}(UZU^*) \rangle &= \langle U^*(\mathcal{A}^*\hat{y})U, Z \rangle \\ &= \langle \text{Diag}(\lambda_{[k]}(\mathcal{A}^*\hat{y}, C)), Z \rangle \\ &\geq \lambda_k(\mathcal{A}^*\hat{y}, C) \\ &= [\lambda_1(\mathcal{A}^*\hat{y}, C)]_+ - ([\lambda_1(\mathcal{A}^*\hat{y}, C)]_+ - \lambda_k(\mathcal{A}^*\hat{y}, C)) \\ &= [\lambda_1(\mathcal{A}^*\hat{y}, C)]_+ - (\lambda_1(\mathcal{A}^*\hat{y}, C) - \lambda_k(\mathcal{A}^*\hat{y}, C)), \end{aligned}$$

### 3.3. Spectral lower models for convex minimization

---

where the last equality uses the observation above that  $\lambda_1(\mathcal{A}^*\hat{y}, C) > 0$  at dual feasible points  $y$ . By noticing that  $\langle C, UZU^* \rangle = \langle U^*CU, Z \rangle = \langle I, Z \rangle = \text{trace } Z$ , we deduce that

$$\partial_{\hat{\varepsilon}} [\lambda_1(\mathcal{A}^*\cdot, C)]_+(\hat{y}) \supseteq \left\{ \mathcal{A}_U Z \mid \text{trace } Z = 1, Z \succcurlyeq 0 \right\},$$

where  $\hat{\varepsilon} := \lambda_1(\mathcal{A}^*\hat{y}, C) - \lambda_k(\mathcal{A}^*\hat{y}, C)$ . Moreover, it is easy to see that the set on the right-hand side of this inclusion has a nonempty intersection with  $\partial [\lambda_1(\mathcal{A}^*\cdot, C)]_+(\hat{y})$ , a property that will justify the use of such lower models in proximal bundle methods as we will see later.

These considerations shed some light on the structure of our family of lower models as a possibly uncountable supremum of linear lower models given by approximate subgradient inequalities:

$$\begin{aligned} & [\lambda_1(\mathcal{A}^*y, C)]_+ \\ & \geq \left[ \sup_{\substack{\text{trace } Z=1 \\ Z \succcurlyeq 0}} \left\{ \begin{array}{l} [\lambda_1(\mathcal{A}^*\hat{y}, C)]_+ + \langle y - \hat{y}, \mathcal{A}_U Z \rangle \\ - \left( [\lambda_1(\mathcal{A}^*\hat{y}, C)]_+ - \langle \lambda_{[k]}(\mathcal{A}^*\hat{y}, C), \text{diag}(Z) \rangle \right) \end{array} \right\} \right]_+ \\ & = \left[ \sup_{\substack{\text{trace } Z=1 \\ Z \succcurlyeq 0}} \left\{ \begin{array}{l} \lambda_1(\mathcal{A}^*\hat{y}, C) + \langle \mathcal{A}_U^*(y - \hat{y}), Z \rangle \\ - \lambda_1(\mathcal{A}^*\hat{y}, C) + \langle \lambda_{[k]}(\mathcal{A}^*\hat{y}, C), \text{diag}(Z) \rangle \end{array} \right\} \right]_+ \\ & = \left[ \sup_{\substack{\text{trace } Z=1 \\ Z \succcurlyeq 0}} \left\{ \langle \mathcal{A}_U^*(y - \hat{y}), Z \rangle + \langle \lambda_{[k]}(\mathcal{A}^*\hat{y}, C), \text{diag}(Z) \rangle \right\} \right]_+ \\ & = \left[ \sup_{\substack{\text{trace } Z=1 \\ Z \succcurlyeq 0}} \left\{ \langle \mathcal{A}_U^*y, Z \rangle - \langle \mathcal{A}_U^*\hat{y}, Z \rangle + \langle \lambda_{[k]}(\mathcal{A}^*\hat{y}, C), \text{diag}(Z) \rangle \right\} \right]_+ \\ & = \left[ \sup_{\substack{\text{trace } Z=1 \\ Z \succcurlyeq 0}} \left\{ \langle \mathcal{A}_U^*y, Z \rangle - \langle \text{Diag}(\lambda_{[k]}(\mathcal{A}^*\hat{y}, C)), Z \rangle + \langle \lambda_{[k]}(\mathcal{A}^*\hat{y}, C), \text{diag}(Z) \rangle \right\} \right]_+ \\ & = \left[ \sup_{\substack{\text{trace } Z=1 \\ Z \succcurlyeq 0}} \left\{ \langle \mathcal{A}_U^*y, Z \rangle - \langle \lambda_{[k]}(\mathcal{A}^*\hat{y}, C), \text{diag}(Z) \rangle + \langle \lambda_{[k]}(\mathcal{A}^*\hat{y}, C), \text{diag}(Z) \rangle \right\} \right]_+ \\ & = \left[ \sup_{\substack{\text{trace } Z=1 \\ Z \succcurlyeq 0}} \langle \mathcal{A}_U^*y, Z \rangle \right]_+ \\ & = [\lambda_1(\mathcal{A}_U^*y)]_+. \end{aligned}$$

### 3.3.2 Proximal-bundle subproblems

In a seminal paper, Correa and Lemaréchal (1993) introduced general conditions under which the classical proximal bundle method (Bonnans et al., 2006, Chapter 10) converges to a solution of (CM).

The basic idea in those methods consists of approximating  $f$  using lower models  $\check{f}_i$  around the main iterates  $x^k$  and minimizing a stabilization of the form  $\check{f}_i + \frac{1}{2\alpha_k} \|\cdot - x^k\|_2^2 + \delta_F(\cdot)$ . By suitably constructing those lower models to satisfy a few general conditions and updating the main iterates when sufficient descent is achieved, it can be shown that the sequence of main iterates converges to a minimizer of  $f$  in  $F$  when such a point exists.

In the following considerations, we skip many details on the outer proximal bundle iterations and focus our attention on verifying that the family of lower models presented in the previous subsection can be used within the framework presented in (Correa and Lemaréchal, 1993) and on the structure of the proximal bundle's stabilized subproblems when specialized to the dual gauge problem for the generalized PhaseLift formulation (3.5b).

#### Conditions on lower models for use in proximal bundle methods

Bundle methods generate iterates approximating solutions of (CM) by solving problems of the form

$$\hat{x}^k := \arg \min_{x \in \mathcal{X}} \check{f}_k(x) + \frac{1}{2\alpha_s} \|x - x^s\|_2^2 \quad \text{subject to } x \in F,$$

in order to achieve a tentative iterate  $\hat{x}^k \in F$ . If a certain sufficient decrease condition is satisfied, this step is called a *serious step* and the stability center  $x^s$  is updated with the new serious iterate  $x^{s+1} := \hat{x}^k$ . Otherwise, the step is declared a *null step* and the null iterate  $\hat{x}^k$  is only used to update the lower model.

In principle, the lower model can be updated at any time after either a serious or a null step as long as the conditions below, introduced in (Correa and Lemaréchal, 1993, page 269), are satisfied:

$$\begin{aligned} & \check{f}_k(x) \leq f(x), \quad \forall k = 0, \dots \text{ and } \forall x \in F, \\ \max \left\{ \begin{array}{l} \check{f}_k(\hat{x}^k) + \langle \alpha_s^{-1}(x^s - \hat{x}^k), \cdot - \hat{x}^k \rangle, \\ f(\hat{x}^k) + \langle \hat{g}^k, \cdot - \hat{x}^k \rangle \end{array} \right\} & \leq \check{f}_{k+1}, \quad \text{if } \hat{x}^k \in F \text{ was generated} \\ & \hspace{15em} \text{in a null step,} \end{aligned}$$

where  $\hat{g}^k \in \partial f(\hat{x}^k)$  is arbitrary.

This result is remarkable since it presents very meager conditions on the construction of lower models for the objective which will still satisfy the convergence results for a proximal bundle iteration based on them. In fact, the maximum of the two affine functions on the left-hand side of the second condition can be used to construct such lower models since it satisfies the first condition. This can be verified by noticing that the second term in that maximum corresponds to a subgradient linearization of the convex  $f$ , and then by observing that the optimality condition for the stabilized subproblem implies that the first term in the maximum is a subgradient linearization of  $f_k$  in  $F$ —i.e.,  $\alpha_s^{-1}(x^s - \hat{x}^k) \in \partial \check{f}_k(\hat{x}^k) + \mathcal{N}_F(\hat{x}^k)$ —, which is a lower model for  $f$  by assumption.

Naturally, one would not expect good practical convergence behaviour if only these meager assumptions are satisfied (i.e., using arbitrary lower models after serious steps and the maximum of these two linearizations). However, by designing more accurate lower models satisfying these conditions, one might be able to trade-off convergence performance and complexity of the lower models and difficulty in the numerical solution of the subproblems.

It is noteworthy that the spectral bundle method introduced by Helmberg and Rendl (2000) can be roughly seen as implementing lower models constructed taking the maximum of the first affine function above (also known in the bundle literature as the *aggregate* linearization (Bonnans et al., 2006, Chapter 10)) and substituting the second subgradient linearization with a spectral lower model as presented in the previous subsection. For those,  $U$  is constructed in such a way that it contains columns with a few of the rightmost eigenvectors of the corresponding  $\mathcal{A}^* \hat{y}^k$ , as well as re-orthonormalized columns from previously computed eigenvectors.

### Proximal bundle subproblems with spectral lower models

From our discussion above, the design and maintenance of lower models for the dual gauge objective in (3.5b) allow for considerable flexibility. As long as the models lie below the objective at all times and above the aggregate linearization and *some* subgradient linearization at the most recent subproblem solution, we have the freedom to manipulate the model function in order to balance memory usage and subproblem solution costs against the accuracy/tightness of those lower models while still satisfying the assumptions used to prove convergence of the serious iterates to a minimizer.

To present our construction, we first present the structure of the proximal

bundle subproblems when specialized to (3.5b):

$$\hat{y}^k := \underset{y \in \mathbb{C}^m}{\text{minimize}} \quad \check{f}_k(y) + \frac{1}{2\alpha_s} \|y - y^s\|_2^2 \quad \text{subject to} \quad \Re \langle y, b \rangle - \epsilon \|y\|_2 \geq 1,$$

where  $\check{f}_k$  is our current lower model for  $[\lambda_1(\mathcal{A}^*, C)]_+$ .

A simple and natural construction of lower models can be derived from the conditions stated above based on the lower models from the previous subsection. Given a  $C$ -orthonormal matrix  $U_k \in \mathbb{C}^{n \times r_k}$  whose columns comprise at least one generalized eigenvector of  $\mathcal{A}^* \hat{y}^k$  corresponding to  $\lambda_1(\mathcal{A}^* \hat{y}^k)$  and given also arbitrary  $C$ -orthonormal  $\{U_i\}_{i \in \mathcal{I}_k}$ , we have that lower models of the form

$$\check{f}_{k+1} := \max \left\{ 0, \check{f}_k(\hat{y}^k) + \langle \alpha_s^{-1}(y^s - \hat{y}^k), \cdot - \hat{y}^k \rangle, \lambda_1(\mathcal{A}_{U_{k-1}}^* \cdot), \max_{i \in \mathcal{I}_k} \{ \lambda_1(\mathcal{A}_{U_i}^* \cdot) \} \right\},$$

satisfy the conditions presented by Correa and Lemaréchal (1993). This is due to our considerations in the previous subsection regarding the relationship between lower models of the form  $[\lambda_1(\mathcal{A}_U^* \cdot)]_+$  and the (approximate) subdifferentials of the dual gauge objective at dual feasible points.

**Remark 3.3.1.** It is noteworthy that  $U_k$  does not need to contain more than one column, as long as one of its columns is a generalized eigenvector of  $\mathcal{A}^* \hat{y}^k$  corresponding to  $\lambda_1(\mathcal{A}^* \hat{y}^k, C)$ , not all columns need to be generalized eigenvectors and  $\mathcal{I}_k$  could be empty (i.e., no additional  $U_i$  at all). In fact, one can show that this is (by equivalence) the case in the spectral bundle method Helmer and Rendl (2000) where  $U_k$  contains some generalized eigenvectors and  $\hat{y}^k$  as well as  $C$ -reorthonormalized columns of previously computed generalized eigenvectors.

Using lower models like these, we can reformulate bundle subproblems as

$$\begin{aligned} & \underset{(y,t) \in \mathbb{C}^m \times \mathbb{R}}{\text{minimize}} && t + \frac{1}{2\alpha_s} \|y - y^s\|_2^2 \\ & \text{subject to} && \\ & && \Re \langle y, b \rangle - \epsilon \|y\|_2 \geq 1, \\ & && \hat{t}_{k-1} + \langle \alpha_s^{-1}(y^s - \hat{y}^{k-1}), y - \hat{y}^{k-1} \rangle \leq t, \\ & && \mathcal{A}_{U_k}^* y \preceq tI, \\ & && \mathcal{A}_{U_i}^* y \preceq tI, \quad i \in \mathcal{I}_{k-1}, \\ & && t \geq 0, \end{aligned}$$

where  $\hat{t}_{k-1} := \check{f}_k(\hat{y}^k)$ . These problems can be solved by off-the-shelf solvers implementing interior point methods (typically after introducing multiple, but small, semidefinite slacks and second-order cone auxiliary variables).

In this approach, we trade-off the minimization of a linear objective over one huge ( $n \times n$ ) linear matrix inequality (LMI) on  $m$  (possibly complex) variables by the sequential minimization of simple quadratic objectives over controllably-many controllably-small LMIs (still on  $m$  variables).

As mentioned earlier, we would not expect to be able to arbitrarily decrease the complexity of the subproblems and the induced lower models without impacting the number of iterations needed to achieve fixed stopping criteria. However, this family of spectral lower models satisfying the conditions of the framework presented by Correa and Lemaréchal (1993) allows much flexibility for bundle maintenance while retaining convergence guarantees.

Although the family of lower models described in this section allows for a natural exploitation within the framework of bundle methods for nonsmooth convex minimization, the approach and proof of concept solver we describe next make use of a simpler projected subgradient descent method for the dual spectral optimization problem (3.5b). The use of such models for the numerical solution of that problem, along with the primal recovery technique presented next, seems to be an interesting topic for further investigation.

# Chapter 4

## Low-rank spectral optimization

Having established in Chapter 3 the connection between those matrix-lifted convex relaxations of phase recovery and blind deconvolution from Chapter 1 and the gauge duality framework studied in Chapter 2, we are in the position to develop a computational strategy for those problems.

The approach that we propose is designed for the numerical solution of the following problems in the scenarios of low-rank recovery described in §1.1:

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \text{trace } X \quad \text{subject to } \|b - \mathcal{A}X\| \leq \epsilon, \quad X \succeq 0, \quad (4.1a)$$

$$\underset{X \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad \|X\|_1 := \sum_{i=1}^{\min\{n_1, n_2\}} \sigma_i(X) \quad \text{subject to } \|b - \mathcal{A}X\| \leq \epsilon, \quad (4.1b)$$

where the parameter  $\epsilon \in [0, \|b\|)$  controls the admissible deviations between the linear model  $\mathcal{A}X$  and the vector of observations  $b$ . (The particular properties of  $b$  and the linear operators  $\mathcal{A}$  are detailed in §1.1.) Our strategy for both problems is based on first solving a related constrained Hermitian eigenvalue optimization problem over a simple constraint, and then using that solution to recover a solution of the original problem. This eigenvalue problem is highly structured, and because the constraint is easily handled, we are free to apply a projected first-order method with inexpensive per-iteration costs that scales well to very large problems.

The method that we develop applies to the much broader class of conic optimization problems with nonnegative objective values. We pay special attention to the low-rank spectral problems just mentioned because their measurement operators and solution are highly structured and can be exploited both theoretically and computationally.

In the following we summarize our overall scheme and its building blocks are described in the next sections.

**A roadmap of our approach.** Our strategy for these low-rank spectral optimization problems is based on solving the following dual constrained eigenvalue optimization problem derived via gauge duality in Chapter 3:

$$\underset{y \in \mathbb{R}^m}{\text{minimize}} \quad \lambda_1(\mathcal{A}^*y) \quad \text{subject to} \quad \langle b, y \rangle - \epsilon \|y\|_* \geq 1. \quad (4.2)$$

The dimension of the variable  $y$  in the eigenvalue optimization problem corresponds to the number of measurements. In the context of phase retrieval and blind deconvolution, Candès et al. (2015b) and Ahmed et al. (2014) show that the number of measurements needed to recover with high probability the underlying signals is within logarithmic factors of the signal length (see §1.1). The crucial implication is that the dimension of the dual problem grows slowly as compared to the dimension of the primal problem, which grows quadratically on the signal length.

In our implementation, we apply a simple first-order projected subgradient method to solve this convex constrained spectral optimization problem. The dominant cost at each iteration of our algorithm is the computation of right-most eigenpairs of the  $n \times n$  Hermitian linear operator  $\mathcal{A}^*y$ , which are used to construct descent directions for (4.2). The structure of the measurement operators then allows us to use Krylov-based eigensolvers, such as ARPACK (Lehoucq, Sorensen, and Yang, 1998), for obtaining these leading eigenpairs. Primal solution estimates  $X$  are then recovered and retained in a low-rank factored form via relatively small constrained least-squares problems, described next in §4.2.

An analogous approach based on the classical Lagrangian duality would also lead to a dual optimization problem in the same space as our dual eigenvalue problem (c.f. considerations in §§2.3.2 and characterization in §3.1):

$$\underset{y \in \mathbb{R}^m}{\text{maximize}} \quad \langle y, b \rangle - \epsilon \|y\|_* \quad \text{subject to} \quad \mathcal{A}^*y \preceq I. \quad (4.3)$$

Note that the Lagrange dual possesses a rather simple objective and a difficult linear matrix inequality of order  $n$  as a constraint. Precisely the reverse situation holds for the gauge dual (4.2), which has a relatively simple constraint.

It is well known that SDPs with a constant-trace property—i.e.,  $\mathcal{A}X = b$  implies  $\text{trace}(X)$  is constant—have Lagrange dual problems that can be reformulated as unconstrained eigenvalue problems. This approach is used by Helmberg and Rendl (2000) to develop a spectral bundle method. The applications that we consider however, do not necessarily exhibit this property.

## 4.1 Derivation of the approach

There are two key theoretical pieces needed for our approach. The first is the derivation of the convex constrained eigenvalue optimization problem (4.2), which we described in §3.1.1 (see equation (3.7)). The second piece is described next and consists of the derivation of a subproblem that allows recovery of a primal solution  $X$  from a solution of (4.2).

### 4.1.1 Recovering a primal solution

Our derivation of a subproblem for primal recovery proceeds in two stages. The first stage develops necessary and sufficient optimality conditions for the primal-dual gauge pair (3.5a) and (3.5b). The second stage uses these to derive a subproblem that can be used to recover a primal solution from a dual solution.

#### Strong duality and optimality conditions

From Theorem 2.4.1, the weak gauge duality condition

$$1 \leq \kappa(X)\kappa^\circ(\mathcal{A}^*y) \quad (4.4)$$

holds for all primal-dual feasible pairs  $(X, y)$ . The following result asserts that if the pair is optimal, then that inequality must necessarily hold tightly.

**Proposition 4.1.1** (Strong duality). If (4.1a) is feasible and  $0 \leq \epsilon < \|b\|$ , then

$$\left[ \min_{\substack{X \in \mathcal{H}^n \\ \|b - \mathcal{A}X\| \leq \epsilon}} \text{trace } X + \delta(X \mid \cdot \succeq 0) \right] \cdot \left[ \inf_{\substack{y \in \mathbb{R}^m \\ \langle y, b \rangle - \epsilon \|y\|_* \geq 1}} [\lambda_1(\mathcal{A}^*y)]_+ \right] = 1. \quad (4.5)$$

*Proof.* We proceed by reasoning about the Lagrangian-dual pair (4.1a) and (4.3). We then translate these results to the corresponding gauge-dual pair (3.5a) and (3.5b).

The primal problem (4.1a) is feasible by assumption. Because its Lagrange dual problem (4.3) admits strictly feasible points (e.g.,  $y = 0$ ), it follows from Rockafellar (1970, Theorems 28.2 and 28.4) that the primal problem attains its positive minimum value and that there is zero duality gap between the Lagrange-dual pair (a broadly known fact of satisfying dual Slater conditions).

Moreover, because the primal problem (4.1a) attains its positive minimum value for some  $\widehat{X}$ , and there is zero duality gap, there exists a sequence  $\{y_j\}$  such that  $[\lambda_1(\mathcal{A}^*y_j)]_+ \leq 1$  and  $\langle y_j, b \rangle - \epsilon \|y_j\|_* \nearrow \text{trace } \widehat{X}$ . Because  $\text{trace } \widehat{X} > 0$ , we can take a subsequence  $\{y_{j_k}\}$  for which  $\langle y_{j_k}, b \rangle - \epsilon \|y_{j_k}\|_*$  is uniformly bounded above zero. Define the sequence  $\{\widehat{y}_k\}$  by  $\widehat{y}_k := y_{j_k} (\langle y_{j_k}, b \rangle - \epsilon \|y_{j_k}\|_*)^{-1}$ . Then  $\langle \widehat{y}_k, b \rangle - \epsilon \|\widehat{y}_k\|_* = 1$  for all  $k$ , which is a feasible sequence for the gauge dual problem (3.5b). Weak gauge duality (4.4) and the definition of  $\widehat{y}_k$  then implies that

$$(\text{trace } \widehat{X})^{-1} \leq [\lambda_1(\mathcal{A}^*\widehat{y}_k)]_+ \leq (\langle y_{j_k}, b \rangle - \epsilon \|y_{j_k}\|_*)^{-1} \searrow (\text{trace } \widehat{X})^{-1}.$$

Multiply the series of inequalities by  $\text{trace } \widehat{X}$  to obtain (4.5).  $\square$

Note the lack of symmetry in the statement of Proposition 4.1.1: the primal problem is stated with a “min”, but the dual problem is stated with an “inf”. This is because the dual Slater condition—i.e., strict feasibility of the corresponding Lagrange-dual problem (4.3)—allows us to assert that a primal optimal solution necessarily exists. However, we cannot assert in general that a dual optimal solution exists because the corresponding primal feasible set does not necessarily satisfy the Slater condition. We comment further on this theoretical question in the concluding Chapter 6.

The following result characterizes gauge primal-dual optimal pairs. It relies on von Neumann’s trace inequality, which says that: for Hermitian  $A$  and  $B$ ,

$$\langle A, B \rangle \leq \langle \lambda(A), \lambda(B) \rangle,$$

and equality holds if and only if  $A$  and  $B$  admit a *simultaneous ordered eigen-decomposition*, i.e.,  $A = U \text{Diag}[\lambda(A)]U^*$  and  $B = U \text{Diag}[\lambda(B)]U^*$  for some unitary matrix  $U$ ; see Lewis (1996).

**Proposition 4.1.2** (Optimality conditions). If (4.1a) is feasible and  $0 \leq \epsilon < \|b\|$ , then  $(X, y) \in \mathcal{H}^n \times \mathbb{R}^m$  is primal-dual optimal for the gauge dual pair (3.5a) and (3.5b) if and only if the following conditions hold:

1.  $X \succeq 0$  and  $\|b - \mathcal{A}X\| = \epsilon$ ;
2.  $\langle y, b \rangle - \epsilon \|y\|_* = 1$ ;
3.  $\langle y, b - \mathcal{A}X \rangle = \|y\|_* \|b - \mathcal{A}X\|$ ;
4.  $\lambda_i(X) \cdot (\lambda_1(\mathcal{A}^*y) - \lambda_i(\mathcal{A}^*y)) = 0, i = 1, \dots, n$ ;

5.  $X$  and  $\mathcal{A}^*y$  admit a simultaneous ordered eigendecomposition.

*Proof.* By strong duality (Proposition 4.1.1), the pair  $(X, y) \in \mathcal{H}^m \times \mathbb{R}^m$  is primal-dual optimal if and only if they are primal-dual feasible and the product of their corresponding objective values is equal to one. In this case,

$$\begin{aligned}
 1 &= [\text{trace } X + \delta(X \mid \cdot \succeq 0)] \cdot [\lambda_1(\mathcal{A}^*y)]_+ && \text{(strong duality)} \\
 &= \langle e, \lambda(X) \rangle \cdot \lambda_1(\mathcal{A}^*y) \\
 &= \langle \lambda_1(\mathcal{A}^*y) \cdot e, \lambda(X) \rangle \\
 &\geq \langle \lambda(\mathcal{A}^*y), \lambda(X) \rangle && (\lambda_1(\mathcal{A}^*y) \geq \lambda_i(\mathcal{A}^*y) \text{ and } X \succeq 0) \\
 &\geq \langle \mathcal{A}^*y, X \rangle && \text{(von Neumann's trace inequality)} \\
 &= \langle y, \mathcal{A}X \rangle \\
 &= \langle y, b \rangle - \langle y, b - \mathcal{A}X \rangle \\
 &\geq \langle y, b \rangle - \|y\|_* \|b - \mathcal{A}X\| && \text{(Cauchy-Schwartz inequality)} \\
 &\geq \langle y, b \rangle - \epsilon \|y\|_* && \text{(primal feasibility)} \\
 &\geq 1. && \text{(dual feasibility)}
 \end{aligned}$$

Thus all of the above inequalities hold with equality. This proves conditions 1–4. Condition 5 follows from again invoking von Neumann's trace inequality and noting its implication that  $X$  and  $\mathcal{A}^*y$  share a simultaneous ordered eigenvalue decomposition. Sufficiency of those conditions can be verified by simply following the reverse chain of reasoning and again noticing that the inequalities can be replaced by equalities.  $\square$

### 4.1.2 Primal recovery subproblem

The optimality conditions stated in Proposition 4.1.2 furnish the means for deriving a subproblem that can be used to recover a primal solution from a dual solution. The next result establishes an explicit relationship between primal solutions  $X$  and  $\mathcal{A}^*y$  for an arbitrary optimal dual solution  $y$ .

**Corollary 4.1.1.** Suppose that the conditions of Proposition 4.1.2 hold. Let  $y \in \mathbb{R}^m$  be an arbitrary optimal solution for the dual gauge program (3.5b),  $r_1 \in \{1, \dots, n\}$  be the multiplicity of  $\lambda_1(\mathcal{A}^*y)$ , and  $U_1 \in \mathbb{C}^{n \times r_1}$  be the matrix formed by the first  $r_1$  eigenvectors of  $\mathcal{A}^*y$ . Then a matrix  $X \in \mathcal{H}^n$  is a solution for the primal problem (3.5a) if and only if there

#### 4.1. Derivation of the approach

exists an  $r_1 \times r_1$  matrix  $S \succeq 0$  such that

$$X = U_1 S U_1^* \quad \text{and} \quad (b - \mathcal{A}X) \in \epsilon \partial \|\cdot\|_*(y). \quad (4.6)$$

*Proof.* The assumptions imply that the optimal dual value is positive. If  $y \in \mathbb{R}^m$  is an optimal solution to (3.5b), the positive-homogeneity of its objective and constraint, and the positivity of the optimal value, allow us to deduce that the dual constraint must be active, i.e.,  $\langle y, b \rangle - \epsilon \|y\|_* = 1$ . Thus condition 2 of Proposition 4.1.2 holds.

The construction of  $X$  in (4.6) guarantees that it shares a simultaneous ordered eigendecomposition with  $\mathcal{A}^*y$ , and that it has rank of  $r_1$  at most. Thus, conditions 4 and 5 of Proposition 4.1.2 hold.

We now show that conditions 1 and 3 of the proposition hold. The sub-differential  $\partial \|\cdot\|_*$  corresponds to the set of maximizers of the linear function that defines the dual norm; i.e.,  $\|\cdot\|_* := \max_{\|z\| \leq 1} \Re \langle \cdot, z \rangle$ . Then because  $(b - \mathcal{A}X) \in \epsilon \partial \|\cdot\|_*(y)$ , it holds that  $\|b - \mathcal{A}X\| \leq \epsilon$  and  $\epsilon \|y\|_* = \langle y, b - \mathcal{A}X \rangle \leq \|y\|_* \|b - \mathcal{A}X\| \leq \epsilon \|y\|_*$ , implying that  $\|b - \mathcal{A}X\| = \epsilon$  and  $\langle y, b - \mathcal{A}X \rangle = \|y\|_* \|b - \mathcal{A}X\|$ . This way, condition 1 and 3 of Proposition 4.1.2 are also satisfied. Hence, all the conditions of the proposition are satisfied, and the pair  $(X, y) \in \mathcal{H}^n \times \mathbb{R}^m$  is optimal.

Suppose now that  $X \in \mathcal{H}^n$  is optimal for (3.5a). We can invoke Proposition 4.1.2 on the pair  $(X, y) \in \mathcal{H}^n \times \mathbb{R}^m$ . Condition 4 implies that any eigenvector of  $\mathcal{A}^*y$  associated to an eigenvalue  $\lambda_i(\mathcal{A}^*y)$  with  $i > r_1$  is in the nullspace of  $X$ , therefore there is an  $r_1 \times r_1$  matrix  $S \succeq 0$  such that  $X = U_1 S U_1^*$ . Conditions 1 and 3 imply that  $\|b - \mathcal{A}X\| \leq \epsilon$  and  $\langle y, b - \mathcal{A}X \rangle = \epsilon \|y\|_*$ , thus verifying that  $(b - \mathcal{A}X) \in \epsilon \partial \|\cdot\|_*(y)$ , as required.  $\square$

Corollary 4.1.1 thus provides us with a way to recover a solution to our model problem (4.1a) after computing a solution to the gauge dual problem (3.5a). When the residual in (4.1a) is measured in the 2-norm, condition (4.6) simplifies, and implies that the matrix  $S$  that defines  $X = U S U^*$  can be obtained by solving

$$\underset{S \succeq 0}{\text{minimize}} \quad \|\mathcal{A}(U_1 S U_1^*) - b_\epsilon\|_2^2, \quad \text{with } b_\epsilon := b - \epsilon y / \|y\|. \quad (4.7)$$

When the multiplicity  $r_1$  of the eigenvalue  $\lambda_1(\mathcal{A}^*y)$  is much smaller than  $n$ , this optimization problem is relatively inexpensive. In particular, if  $r_1 = 1$ —which may be expected in some matrix-lifted applications such as PhaseLift—the optimization problem is over a scalar  $s$  that can be obtained immediately as

$$s = [\langle \mathcal{A}(u_1 u_1^*), b_\epsilon \rangle]_+ / \|\mathcal{A}(u_1 u_1^*)\|_2^2$$

where  $u_1$  is a rightmost eigenvector of  $\mathcal{A}^*y$ . This approach exploits the complementarity relation on eigenvalues in condition 4 of Proposition 4.1.2 to reduce the dimensionality of the primal solution recovery. Its computational difficulty effectively depends on finding a dual solution  $y$  at which the rightmost eigenvalue has low multiplicity  $r_1$ .

## 4.2 The implementation of a proof of concept solver

The effectiveness of our approach hinges on efficiently solving the constrained eigenvalue optimization problem (4.2) in order to generate solution estimates  $y$  and rightmost eigenvector estimates  $U_1$  of  $\mathcal{A}^*y$  that we can feed to (4.7). The two main properties of this problem that drive our approach are that it has a nonsmooth objective and that projections on the feasible set are inexpensive. Our implementation is based on a basic projected-subgradient descent method, although certainly other choices are available. For example, Nesterov (2009) and Richtárik (2011) propose specialized algorithms for minimizing positively homogeneous functions with affine constraints; some modification of this approach could possibly apply to (4.3). Another possible choice is Helmberg and Rendl’s (2000) spectral bundle method. For simplicity, and because it has proven sufficient for our needs, we use a standard projected subgradient method as described below.

### 4.2.1 Dual descent

The generic projected subgradient method is based on the iteration

$$y_+ = \mathcal{P}(y - \alpha g), \tag{4.8}$$

where  $g$  is a subgradient of the objective at the current iterate  $y$ ,  $\alpha$  is a positive steplength, and the operator  $\mathcal{P} : \mathbb{R}^m \rightarrow \mathbb{R}^m$  gives the Euclidean projection onto the feasible set. For the objective function  $f(y) = \lambda_1(\mathcal{A}^*y)$  of (3.7), the subdifferential has the form

$$\partial f(y) = \{ \mathcal{A}(U_1 T U_1^*) \mid T \succeq 0, \text{ trace } T = 1 \}, \tag{4.9}$$

where  $U_1$  is the  $n \times r_1$  matrix of rightmost eigenvectors of  $\mathcal{A}^*y$  (Overton, 1992, Theorem 3). A Krylov-based eigenvalue solver can be used to evaluate  $f(y)$  and a subgradient  $g \in \partial f(y)$ . Such methods require products of the

form  $(\mathcal{A}^*y)v$  for arbitrary vectors  $v$ . In many cases, these products can be computed without explicitly forming the matrix  $\mathcal{A}^*y$ . In particular, for the applications described in §1.1, these products can be computed entirely using fast operators such as the FFT and the fast *discrete Wavelet transform* (DWT), with essentially  $\mathcal{O}(n \log n)$  arithmetic operations. Similar efficiencies can be used to compute a subgradient  $g$  from the forward map  $\mathcal{A}(U_1 T U_1^*)$ .

For large problems, further efficiencies can be obtained simply by computing a single eigenvector  $u_1$ , i.e., any unit-norm vector in the range of  $U_1$ . In our implementation, we typically request at least *two* rightmost eigenpairs: this gives us an opportunity to detect if the leading eigenpair is isolated. If it is, then the subdifferential contains only a single element, which implies that  $f$  is in fact differentiable at that point.

Any sequence of step lengths  $\{\alpha_k\}$  that satisfies the generic conditions

$$\lim_{k \rightarrow \infty} \alpha_k = 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty$$

is sufficient to guarantee that the value of the objective at  $y_k$  converges to the optimal value (Bertsekas, 2015, Proposition 3.2.6). A typical choice is  $\alpha_k = \mathcal{O}(1/k)$ . Our implementation defaults to a Barzilai-Borwein steplength (Barzilai and Borwein, 1988) with a nonmonotonic linesearch (Zhang and Hager, 2004) if it is detected that a sequence of iterates is differentiable (by observing separation of the leading eigenpair); and otherwise it falls back to a decreasing step size.

The projection operator  $\mathcal{P}$  onto the dual-feasible set (3.7) is inexpensive when the residual is measured in the 2-norm. In particular, if  $\epsilon = 0$ , the dual-feasible set is a halfspace, and the projection can be accomplished in linear time. When  $\epsilon$  is positive, the projection requires computing the roots of a 1-dimensional degree-4 polynomial, which in practice requires little additional expense. In the following, we describe the approach we use for implementing the projection operator.

### Computing the projection onto the gauge dual feasible set

In order to compute the projection onto the gauge dual feasible set, we need to be able to solve the following optimization problem:

$$\underset{y \in \mathbb{C}^m}{\text{minimize}} \quad \frac{1}{2} \|y - \hat{y}\|_2^2 \quad \text{subject to} \quad \Re \langle b, y \rangle - \epsilon \|y\|_2 \geq 1, \quad (4.10)$$

where  $\hat{y} \in \mathbb{C}^m$  is arbitrary and  $\|b\|_2 > \epsilon$ , otherwise the feasible set is empty. The following result allows for a concrete way to compute this projection by

solving for the roots of a quartic polynomial and then picking one satisfying a few simple conditions.

**Proposition 4.2.1.** Given  $b \in \mathbb{C}^m \setminus \{0\}$  and  $\epsilon \in [0, \|b\|_2)$ , there is a unique  $y \in \mathbb{C}^m$  that solves (4.10). Moreover, it is characterized by

$$y = \begin{cases} \hat{y} & \text{if } \Re \langle b, \hat{y} \rangle - \epsilon \|\hat{y}\|_2 \geq 1; \\ (\hat{y} + \lambda b) \left(1 - \frac{\lambda \epsilon}{\|\hat{y} + \lambda b\|_2}\right) & \text{otherwise;} \end{cases}$$

where  $\lambda > 0$  is a positive real root of the degree-4 polynomial

$$p(\lambda) = \|\hat{y} + \lambda b\|_2^2 [\Re \langle b, \hat{y} + \lambda b \rangle + \lambda \epsilon^2 - 1]^2 - \epsilon^2 [\|\hat{y} + \lambda b\|_2^2 + \lambda \Re \langle b, \hat{y} + \lambda b \rangle]^2$$

which satisfies

$$\|\hat{y} + \lambda b\|_2 [\Re \langle b, \hat{y} + \lambda b \rangle + \lambda \epsilon^2 - 1] = \epsilon [\|\hat{y} + \lambda b\|_2^2 + \lambda \Re \langle b, \hat{y} + \lambda b \rangle]$$

and for which  $\|\hat{y} + \lambda b\|_2 - \lambda \epsilon$  is positive and smallest.

*Proof.* Existence and uniqueness follow from the fact that the feasible set is nonempty, closed and convex, and from the strong convexity of the objective. It is clear that if  $\hat{y}$  is feasible, it is also the solution. This way we can focus on the second part of the characterization. Defining  $h(\cdot) := 1 - \Re \langle b, \cdot \rangle + \epsilon \|\cdot\|_2$ , we have that  $\partial h(y) = \left\{ \epsilon \frac{y}{\|y\|_2} - b \right\} \neq 0$ . And the necessary and sufficient first-order optimality conditions tell us we need to find  $\lambda > 0$  such that

$$y - \hat{y} + \lambda \left( \epsilon \frac{y}{\|y\|_2} - b \right) = 0 \quad \text{and} \quad \Re \langle b, y \rangle - \epsilon \|y\|_2 = 1.$$

Manipulating the first condition, we have that

$$\begin{aligned} y \left( 1 + \frac{\lambda \epsilon}{\|y\|_2} \right) &= \hat{y} + \lambda b \quad \implies \quad \|y\|_2 = \|\hat{y} + \lambda b\|_2 - \lambda \epsilon \\ &\implies \quad y = (\hat{y} + \lambda b) \left( 1 - \frac{\lambda \epsilon}{\|\hat{y} + \lambda b\|_2} \right), \end{aligned}$$

which we can then substitute into the second condition, resulting in

$$\|\hat{y} + \lambda b\|_2 [\Re \langle b, \hat{y} + \lambda b \rangle + \lambda \epsilon^2 - 1] = \epsilon [\|\hat{y} + \lambda b\|_2^2 + \lambda \Re \langle b, \hat{y} + \lambda b \rangle].$$

The result follows by squaring both sides and noticing that  $\lambda$  must then satisfy  $p(\lambda) = 0$ .  $\square$

Our implementation is motivated by the approach outlined above as it leads to a simple and inexpensive enough MATLAB implementation. An alternative approach to compute  $\lambda$  can be devised by deriving the Lagrangian dual problem to (4.10) and then employing an efficient iterative method to solve the resulting one-dimensional problem. This might lead to improved robustness and accuracy, but we have not identified a particular need for that in the course of our numerical experimentations.

### 4.2.2 Primal recovery

At each iteration of the descent method (4.8) for the constrained eigenvalue optimization problem (4.2), we compute a corresponding primal estimate

$$X_+ = U_1 S_+ U_1^* \tag{4.11}$$

maintained in factored form. The matrix  $U_1$  has already been computed in the evaluation of the objective and its subgradient; see (4.9). The positive semidefinite matrix  $S_+$  is the solution of the primal-recovery problem (4.7).

A byproduct of the primal-recovery problem is that it provides a suitable stopping criterion for the overall algorithm. Because the iterations  $y_k$  are dual feasible, it follows from Corollary 4.1.1 that if (4.7) has a zero residual, then the dual iterate  $y_k$  and the corresponding primal iterate  $X_k$  are optimal. Thus, we use the size of the residual to determine a stopping test for approximate optimality.

### 4.2.3 Primal-dual refinement

The primal-recovery procedure outlined in §4.2.2 is used as a stopping criterion and to provide primal solutions, it does not directly affect the sequence of dual iterates from (4.8). In our numerical experiments, we find that significant gains can be had by refining the primal estimate (4.11) and feeding it back into the dual sequence. We use the following procedure, which involves two auxiliary subproblems that add relatively little to the overall cost.

Inspired by the Wirtinger Flow algorithm described in §§1.1.1, the first step is to refine the primal estimate obtained via (4.7) by using its solution to determine the starting point  $Z_0 = U_1 S_+^{1/2}$  for the smooth unconstrained non-convex problem

$$\underset{Z \in \mathbb{C}^{n \times r}}{\text{minimize}} \quad h(Z) := \frac{1}{4} \|\mathcal{A}(ZZ^*) - b_\epsilon\|_2^2. \tag{4.12}$$

In effect, we continue to minimize (4.7), where additionally  $U_1$  is allowed to vary. Several options are available for solving this smooth unconstrained problem. Our implementation has the option of using a steepest-descent iteration with a spectral steplength and non-monotone linesearch (Zhang and Hager, 2004), or a limited-memory BFGS method (Nocedal and Wright, 2006, §7.2). The main cost at each iteration is the evaluation of the gradient

$$\nabla h(Z) = \mathcal{A}^*(\mathcal{A}(ZZ^*) - b_\epsilon)Z. \quad (4.13)$$

We thus obtain a candidate improved primal estimate  $\hat{X} = \hat{Z}\hat{Z}^*$ , where  $\hat{Z}$  is a solution of (4.12). When  $\epsilon = 0$ , this non-convex problem coincides with the problem used by Candès et al. (2015a) in their Wirtinger Flow algorithm. They use the initialization  $Z_0 = \gamma u_1$ , where  $u_1$  is a leading eigenvector of  $\mathcal{A}^*b$ , and  $\gamma = n \sum_i b_i / \sum_i \|a_i\|_2^2$ . Our initialization, on the other hand, is based on a solution of the primal-recovery problem (4.7).

The second step of the refinement procedure is to construct a candidate dual estimate  $\hat{y}$  from a solution of the constrained linear-least-squares problem

$$\underset{y \in \mathbb{R}^m}{\text{minimize}} \quad \frac{1}{2} \left\| (\mathcal{A}^*y)\hat{Z} - \hat{\lambda}\hat{Z} \right\|_2^2 \quad \text{subject to} \quad \langle y, b \rangle - \epsilon \|y\|_* \geq 1, \quad (4.14)$$

where  $\hat{\lambda} := 1/\text{trace } \hat{X} \equiv 1/\|\hat{Z}\|_2^2$  is the reciprocal of the primal objective value associated with  $\hat{X}$ . This constrained linear-least-squares problem attempts to construct a vector  $\hat{y}$  such that the columns of  $\hat{Z}$  correspond to eigenvectors of  $\mathcal{A}^*\hat{y}$  associated with  $\hat{\lambda}$ . If  $f(\hat{y}) < f(y_+)$ , then  $\hat{y}$  improves on the current dual iterate  $y_+$  obtained by the descent method (4.8), and we are free to use  $\hat{y}$  in its place. This improved estimate, which is exogenous to the dual descent method, can be considered a “spacer” iterate, as described by Bertsekas (1999, Proposition 1.2.6). Importantly, it does not interfere with the convergence of the underlying descent method. The projected-descent method used to solve the dual problem and generate the dual sequence can also be applied to (4.14), though in this case the objective is guaranteed to be differentiable.

### 4.3 Extensions

The formulations (4.1) that we have considered so far are stated in their simplest form. Reweighted formulations, as introduced by Mohan and Fazel (2010) and Candès et al. (2013a), are also useful, and are accommodated by our approach. In the next subsections we provide alternative derivations of the corresponding gauge duals for these weighted formulations and specialize the primal-from-dual recovery conditions from Corollary 4.1.1.

### 4.3.1 Weighted trace minimization

Candès, Eldar, Strohmer, and Voroninski (2013a) show that an iteratively reweighted sequence of trace minimization problems of the form

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \langle W^{-1}, X \rangle \quad \text{subject to} \quad \|b - \mathcal{A}X\| \leq \epsilon, \quad X \succeq 0, \quad (4.15)$$

where  $W \succ 0$ , can improve the range of signals that can be recovered using the PhaseLift relaxation. Each problem in the sequence uses the previous solution to derive the next weight matrix  $W$ , which is a low-rank update to a small positive multiple of the identity.

Define the maps

$$\mathcal{W}(\cdot) := W^{\frac{1}{2}}(\cdot)W^{\frac{1}{2}} \quad \text{and} \quad \mathcal{A}_W := \mathcal{A} \circ \mathcal{W}.$$

It is evident that  $X \succeq 0$  if and only if  $\mathcal{W}^{-1}(X) \succeq 0$ , and so the weighted trace minimization problem can be stated equivalently as

$$\underset{X \in \mathcal{H}^n}{\text{minimize}} \quad \text{trace } \mathcal{W}^{-1}(X) \quad \text{subject to} \quad \|b - \mathcal{A}_W(\mathcal{W}^{-1}(X))\| \leq \epsilon, \quad \mathcal{W}^{-1}(X) \succeq 0.$$

Because  $\mathcal{W}$  is a bijection, we can optimize over  $\hat{X} := \mathcal{W}^{-1}(X)$  instead of  $X$ :

$$\underset{\hat{X} \in \mathcal{H}^n}{\text{minimize}} \quad \text{trace } \hat{X} \quad \text{subject to} \quad \|b - \mathcal{A}_W \hat{X}\| \leq \epsilon, \quad \hat{X} \succeq 0. \quad (4.16)$$

This clearly falls within the structure of (4.1a), and has the corresponding gauge dual

$$\underset{y \in \mathbb{R}^m}{\text{minimize}} \quad [\lambda_1(\mathcal{A}_W^* y)]_+ \quad \text{subject to} \quad \langle y, b \rangle - \epsilon \|y\|_* \geq 1. \quad (4.17)$$

Observing that  $\lambda_1(\mathcal{A}_W^* y) = \lambda_1(W^{\frac{1}{2}}(\mathcal{A}^* y)W^{\frac{1}{2}}) = \lambda_1(\mathcal{A}^* y, W^{-1})$ , the dual gauge problem corresponding to (4.15) has the form

$$\underset{y \in \mathbb{R}^m}{\text{minimize}} \quad [\lambda_1(\mathcal{A}^* y, W^{-1})]_+ \quad \text{subject to} \quad \langle y, b \rangle - \epsilon \|y\|_* \geq 1. \quad (4.18)$$

This shows that the introduction of a weighting matrix that is not a simple multiple of the identity leads to a dual gauge problem involving the minimization of the rightmost *generalized* eigenvalue of  $\mathcal{A}^* y$  with respect to that weight. Now that we have a formulation for the gauge dual problem, we focus on how a primal solution to the original weighted trace minimization can be computed given a dual minimizer. The following result extends Corollary 4.1.1.

**Corollary 4.3.1.** Suppose that problem (4.15) is feasible and  $0 \leq \epsilon < \|b\|$ . Let  $y \in \mathbb{R}^m$  be an arbitrary optimal solution for the dual gauge (4.18),  $r_1 \in \{1, \dots, n\}$  be the multiplicity of  $\lambda_1(\mathcal{A}^*y, W^{-1})$ , and  $U_1 \in \mathbb{C}^{n \times r_1}$  be the matrix formed by the first  $r_1$  generalized eigenvectors of  $\mathcal{A}^*y$  with respect to  $W^{-1}$ . Then  $X \in \mathcal{H}^n$  is a solution for the primal problem (4.15) if and only if there exists  $S \succeq 0$  such that

$$X = U_1 S U_1^* \quad \text{and} \quad (b - \mathcal{A}X) \in \epsilon \partial \|\cdot\|_*(y).$$

*Proof.* A solution for (4.18) is clearly a solution for (4.17). We may thus invoke Corollary 4.1.1 and assert that  $\hat{X} \in \mathcal{H}^n$  is a solution for (4.16) iff there is  $S \succeq 0$  such that  $\hat{X} = \hat{U}_1 S \hat{U}_1^*$  and  $(b - \mathcal{A}_W \hat{X}) \in \epsilon \partial \|\cdot\|_*(y)$ , where  $\hat{U}_1 \in \mathbb{C}^{n \times r_1}$  is a matrix formed by the first  $r_1$  eigenvectors of  $\mathcal{A}_W^* y = W^{\frac{1}{2}}(\mathcal{A}^* y)W^{\frac{1}{2}}$ . From the structure of  $\mathcal{W}$ , we have that  $X$  is a solution to (4.15) iff  $X = \mathcal{W}(\hat{X})$ . Thus,  $X = W^{\frac{1}{2}} \hat{U}_1 S \hat{U}_1^* W^{\frac{1}{2}} = U_1 S U_1^*$ , where  $U_1 := W^{\frac{1}{2}} \hat{U}_1$  corresponds to the first  $r_1$  generalized eigenvectors of  $\mathcal{A}^* y$  with respect to  $W^{-1}$ .  $\square$

Once again, this provides us with a way to recover a solution to the weighted trace minimization problem by computing a solution to the gauge dual problem (now involving the rightmost generalized eigenvalue) and then solving a problem of potentially much reduced dimensionality.

### 4.3.2 Weighted affine nuclear-norm minimization

We can similarly extend the reweighted extension to the non-symmetric case (4.1b). Let  $W_1 \in \mathcal{H}^{n_1}$  and  $W_2 \in \mathcal{H}^{n_2}$  be invertible. The weighted nuclear-norm minimization problem becomes

$$\underset{X \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad \|W_1^{-1} X W_2^{-*}\|_1 \quad \text{subject to} \quad \|b - \mathcal{A}X\| \leq \epsilon. \quad (4.19)$$

Define the weighted quantities

$$\mathcal{W}(\cdot) = W_1(\cdot)W_2^* : \mathbb{C}^{n_1 \times n_2} \rightarrow \mathbb{C}^{n_1 \times n_2}, \quad \mathcal{A}_W = \mathcal{A} \circ \mathcal{W}, \quad \text{and} \quad \hat{X} := \mathcal{W}^{-1}(X).$$

The weighted problem can then be stated equivalently as

$$\underset{\hat{X} \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad \|\hat{X}\|_1 \quad \text{subject to} \quad \|b - \mathcal{A}_W \hat{X}\| \leq \epsilon,$$

### 4.3. Extensions

---

which, following the approach introduced in Fazel (2002), can be embedded in a symmetric problem:

$$\begin{aligned}
 & \underset{\substack{\widehat{U} \in \mathcal{H}^{n_1} \\ \widehat{V} \in \mathcal{H}^{n_2} \\ \widehat{X} \in \mathbb{C}^{n_1 \times n_2}}}{\text{minimize}} && \left\langle \frac{1}{2}I, \begin{pmatrix} \widehat{U} & \widehat{X} \\ \widehat{X}^* & \widehat{V} \end{pmatrix} \right\rangle \\
 & \text{subject to} && \begin{pmatrix} \widehat{U} & \widehat{X} \\ \widehat{X}^* & \widehat{V} \end{pmatrix} \succeq 0 \text{ and } \|b - \mathcal{A}_W \widehat{X}\| \leq \epsilon.
 \end{aligned} \tag{4.20}$$

Define the measurement operator from  $\mathcal{M} : \mathcal{H}^{n_1+n_2} \rightarrow \mathbb{C}^m$  by the map

$$\begin{pmatrix} \widehat{U} & \widehat{X} \\ \widehat{X}^* & \widehat{V} \end{pmatrix} \mapsto \mathcal{A}_W \widehat{X},$$

and identify  $\mathbb{C}^m$  with  $\mathbb{R}^{2m}$  as a real inner-product space. The adjoint of the measurement operator is then given by

$$\mathcal{M}^* y = \begin{pmatrix} 0 & \mathcal{A}_W^* y \\ (\mathcal{A}_W^* y)^* & 0 \end{pmatrix},$$

where  $\mathcal{A}_W^* y = \frac{1}{2} \sum_{i=1}^m W_1 A_i W_2^* y_i$ . We can now state the gauge dual problem:

$$\underset{y \in \mathbb{C}^m}{\text{minimize}} \quad \left[ \lambda_1(\mathcal{M}^* y, \frac{1}{2}I) \right]_+ \quad \text{subject to} \quad \Re \langle y, b \rangle - \epsilon \|y\|_* \geq 1. \tag{4.21}$$

Observe the identity

$$\begin{aligned}
 \lambda_1(\mathcal{M}^* y, \frac{1}{2}I) &= \lambda_1(2\mathcal{M}^* y) \\
 &= \left[ \lambda_1 \left( \begin{pmatrix} 0 & \sum_{i=1}^m W_1 A_i W_2^* y_i \\ (\sum_{i=1}^m W_1 A_i W_2^* y_i)^* & 0 \end{pmatrix} \right) \right]_+ \\
 &= \left[ \|W_1(\mathcal{A}^* y)W_2^*\|_\infty \right]_+ = \|W_1(\mathcal{A}^* y)W_2^*\|_\infty.
 \end{aligned}$$

We can now deduce the simplified form for the gauge dual problem:

$$\underset{y \in \mathbb{C}^m}{\text{minimize}} \quad \|W_1(\mathcal{A}^* y)W_2^*\|_\infty \quad \text{subject to} \quad \Re \langle y, b \rangle - \epsilon \|y\|_* \geq 1. \tag{4.22}$$

This weighted gauge dual problem can be derived from first principles using the tools from §3.1.1 by observing that the primal problem is already in standard gauge form (c.f. §3.2.1). We present this alternative approach,

however, to make explicit the close connection between the (weighted) nuclear-norm minimization problem and the (weighted) trace-minimization problem described in §4.3.1.

The following result provides a way to characterize solutions of the nuclear norm minimization problem when a solution to the dual gauge problem is available.

**Corollary 4.3.2.** Suppose that problem (4.19) is feasible and  $0 \leq \epsilon < \|b\|$ . Let  $y \in \mathbb{C}^m$  be an arbitrary optimal solution for the dual gauge problem (4.22),  $r_1 \in \{1, \dots, n\}$  be the multiplicity of  $\sigma_1(W_1(\mathcal{A}^*y)W_2^*)$ ,  $U_1 \in \mathbb{C}^{n_1 \times r_1}$  and  $V_1 \in \mathbb{C}^{n_2 \times r_1}$  be the matrices formed by the first  $r_1$  left- and right-singular-vectors of  $W_1(\mathcal{A}^*y)W_2^*$ , respectively. Then  $X \in \mathbb{C}^{n_1 \times n_2}$  is a solution for the primal problem (4.19) if and only if there exists  $S \succeq 0$  such that  $X = (W_1U_1)S(W_2V_1)^*$  and  $(b - \mathcal{A}X) \in \epsilon \partial \|\cdot\|_*(y)$ .

*Proof.* A solution for (4.22) is clearly a solution for (4.21); this way we invoke Corollary 4.1.1 and have that  $(\widehat{U}, \widehat{V}, \widehat{X}) \in \mathcal{H}^{n_1} \times \mathcal{H}^{n_2} \times \mathbb{C}^{n_1 \times n_2}$  induce a solution for (4.20) iff there is  $S \succeq 0$  such that  $\widehat{X} = \widehat{U}_1 S \widehat{V}_1^*$  and  $(b - \mathcal{A}_W \widehat{X}) \in \epsilon \partial \|\cdot\|_*(y)$ , where  $\widehat{U}_1 \in \mathbb{C}^{n_1 \times r_1}$  and  $\widehat{V}_1 \in \mathbb{C}^{n_2 \times r_1}$  are matrices formed by the first  $r_1$  left- and right-singular-vectors of  $\mathcal{A}_W^* y = W_1(\mathcal{A}^*y)W_2^*$ . From the structure of  $\mathcal{W}$ , we have that  $X$  is a solution to (4.19) iff  $X = \mathcal{W}(\widehat{X})$ . This way,  $X = (W_1 \widehat{U}_1) S (W_2 \widehat{V}_1)^*$ .  $\square$

# Chapter 5

## Numerical experiments

This chapter reports on a set of numerical experiments for solving instances of the phase retrieval and blind deconvolution problems described in Chapter 1. The various algorithmic pieces described in §4.2 have been implemented as a MATLAB software package. The implementation uses MATLAB’s `eigs` routine for the eigenvalue computations described in §4.2.1, thus only accessing  $\mathcal{A}^*y$  via its product with a given vector and not requiring to explicitly form a large dense matrix. We implemented a projected gradient-descent method, which is then used for solving (4.8), (4.12), and (4.14).

### 5.1 Phase recovery

We conduct three experiments for phase retrieval via the PhaseLift formulation. The first experiment is for a large collection of small 1-dimensional random signals, and is meant to contrast our approach against a general-purpose convex optimization algorithm and a specialized non-convex approach. The second experiment tests problems where the vector of observations  $b$  is contaminated by noise, hence testing the case where  $\epsilon > 0$ . The third experiment tests the scalability of the approach on a large 2-dimensional natural image.

Our phase retrieval experiments follow the approach outlined in Candès et al. (2015a). The diagonal matrices  $C_k \in \mathbb{C}^{n \times n}$  encode diffraction patterns that correspond to the  $k$ th “mask” ( $k = 1, \dots, L$ ) through which a signal  $x_0 \in \mathbb{C}^n$  is measured. The measurements are given by

$$b = \mathcal{A}(x_0 x_0^*) := \text{diag} \left[ \begin{pmatrix} FC_1 \\ \vdots \\ FC_L \end{pmatrix} (x_0 x_0^*) \begin{pmatrix} FC_1 \\ \vdots \\ FC_L \end{pmatrix}^* \right],$$

where  $F$  is the unitary discrete Fourier transform (DFT). The adjoint of the associated linear map  $\mathcal{A}$  is then

$$\mathcal{A}^*y := \sum_{k=1}^L C_k^* F^* \text{Diag}(y_k) FC_k,$$

where  $y = (y_1, \dots, y_L)$  and  $\text{Diag}(y_k)$  is the diagonal matrix formed from the vector  $y_k \in \mathbb{R}^n$ . The main cost in the evaluation of the forward map  $\mathcal{A}(VV^*)$  involves  $L$  applications of the DFT for each column of  $V$ . Each evaluation of the adjoint map applied to a vector  $v$ —i.e.,  $(\mathcal{A}^*y)v$ —requires  $L$  applications of both the DFT and its inverse. In the experimental results reported below, the columns labeled “nDFT” indicate the total number of DFT evaluations used over the course of a run. The costs of these DFT evaluations are invariant across the different algorithms, and dominate the overall computation.

### 5.1.1 Random Gaussian signals

In this section we consider a set of experiments for different numbers of masks. For each value of  $L = 6, 7, \dots, 12$ , we generate a fixed set of 100 random complex Gaussian vectors  $x_0$  of length  $n = 128$ , and a set of  $L$  random complex Gaussian masks  $C_k$ .

Table 5.1 summarizes the results of applying four different solvers to each set of 100 problems. The solver **GAUGE** is our implementation of the approach described in §4.2; **TFOCS** (Becker et al., 2011) is a first-order conic solver applied to the primal problem (4.1a). The version used here was modified to avoid explicitly forming the matrix  $\mathcal{A}^*y$  (Strohmer, 2013). The algorithm **WFLOW** (Candès et al., 2015a) is a non-convex approach that attempts to recover the original signal directly from the feasibility problem (4.12), with  $\epsilon = 0$ . To make sensible performance comparisons to **WFLOW**, we add to its implementation a stopping test based on the norm of the gradient (4.13); the default algorithm otherwise uses a fixed number of iterations. We also show the results of applying the **GAUGE** code in a “feasibility” mode that exits with the first candidate primal-feasible solution; this is the solver labeled **GAUGE-feas**. This derivative of **GAUGE** is in some respects akin to **WFLOW**, with the main difference that **GAUGE-feas** uses starting points generated by the dual-descent estimates, and generates search directions and step-lengths for the feasibility problem from a spectral gradient algorithm. The columns labeled “xErr” report the median relative error  $\|x_0x_0^* - \widehat{x}\widehat{x}^*\|_F / \|x_0\|_2^2$  of the 100 runs, where  $\widehat{x}$  is the solution returned by the corresponding solver. The columns labeled “%” give the percentage of problems solved to within a relative error of  $10^{-2}$ . This column is excluded for **GAUGE** and **GAUGE-feas** because these solvers obtained the prescribed accuracy for *all* problems in each test set. At least on this set of artificial experiments, the **GAUGE** solver (and its feasibility variant **GAUGE-feas**) appear to be most efficient.

### 5.1. Phase recovery

Table 5.1: Phase retrieval comparisons for random complex Gaussian signals of size  $n = 128$  measured using random complex Gaussian masks. Numbers of the form  $n_{-e}$  are a shorthand for  $n \cdot 10^{-e}$ .

$L$	GAUGE		TFOCS			GAUGE-feas		WFLOW		
	nDFT	xErr	nDFT	xErr	%	nDFT	xErr	nDFT	xErr	%
12	18,330	1.6 <sub>-6</sub>	2,341,800	3.6 <sub>-3</sub>	100	3,528	1.3 <sub>-6</sub>	5,232	1.2 <sub>-5</sub>	100
11	19,256	1.5 <sub>-6</sub>	2,427,546	4.3 <sub>-3</sub>	100	3,344	1.4 <sub>-6</sub>	4,906	1.6 <sub>-5</sub>	100
10	19,045	1.4 <sub>-6</sub>	2,857,650	5.5 <sub>-3</sub>	100	3,120	1.6 <sub>-6</sub>	4,620	2.1 <sub>-5</sub>	100
9	21,933	1.6 <sub>-6</sub>	$1.2 \cdot 10^7$	7.5 <sub>-3</sub>	89	2,889	1.4 <sub>-6</sub>	4,374	2.5 <sub>-5</sub>	100
8	23,144	2.1 <sub>-6</sub>	$1.1 \cdot 10^7$	1.2 <sub>-2</sub>	22	2,688	1.9 <sub>-6</sub>	4,080	3.3 <sub>-5</sub>	100
7	25,781	1.8 <sub>-6</sub>	6,853,245	2.4 <sub>-2</sub>	0	2,492	2.0 <sub>-6</sub>	3,836	5.2 <sub>-5</sub>	95
6	34,689	3.0 <sub>-6</sub>	2,664,126	6.4 <sub>-2</sub>	0	2,424	2.5 <sub>-6</sub>	3,954	9.5 <sub>-5</sub>	62

#### 5.1.2 Random problems with noise

In this set of experiments, we assess the effectiveness of the gauge-based solvers to problems with  $\epsilon > 0$ , which could be useful in recovering signals with noise. For this purpose, it is convenient to generate problem instances with noise and known primal-dual solutions, which we can do by using Corollary 4.1.1. Each instance is generated by first sampling octanary masks  $C_k$ —as described by Candès et al. (2015a)—and real Gaussian  $y \in \mathbb{R}^m$ ; a solution  $x_0 \in \mathbb{C}^n$  is then chosen as a unit-norm rightmost eigenvector of  $\mathcal{A}^*y$ , and the measurements are computed as  $b := \mathcal{A}(x_0x_0^*) + \epsilon y / \|y\|$ , where  $\epsilon$  is chosen as  $\epsilon := \|b - \mathcal{A}(x_0x_0^*)\| = \eta \|b\|$ , for a given noise-level parameter  $0 < \eta < 1$ .

For these experiments, we generate 100 instances with  $n = 128$  for each pairwise combination  $(L, \eta)$  with

$$L \in \{6, 9, 12\} \quad \text{and} \quad \eta \in \{0.1\%, 0.5\%, 1\%, 5\%, 10\%, 50\%\}.$$

Table 5.2 summarizes the results of applying the solvers **GAUGE**, **GAUGE-feas**, and **WFLOW** to these problems. It is not clear that **GAUGE-feas** and **WFLOW** are relevant for this experiment, since they do not directly attempt to solve the norm-constrained PSD trace minimization problem, but for interest we include them in the results. **GAUGE** is generally successful in recovering the rank-1 minimizer for most problems—even for cases with significant noise, though in these cases the overall cost increases considerably.

### 5.1. Phase recovery

Table 5.2: Phase retrieval comparisons for problems with noise, i.e.,  $\epsilon > 0$ . Numbers of the form  $n_{-e}$  are a shorthand for  $n \cdot 10^{-e}$ .

$L$	$\eta$	GAUGE			GAUGE-feas			WFLOW		
		nDFT	xErr	%	nDFT	xErr	%	nDFT	xErr	%
12	0.1%	29,988	2.9 <sub>-3</sub>	100	936	3.4 <sub>-3</sub>	100	14,856	7.1 <sub>-4</sub>	100
9	0.1%	36,292	2.7 <sub>-3</sub>	100	774	2.9 <sub>-3</sub>	100	11,511	7.7 <sub>-4</sub>	100
6	0.1%	50,235	2.9 <sub>-3</sub>	100	612	3.2 <sub>-3</sub>	100	8,922	1.2 <sub>-3</sub>	98
12	0.5%	27,252	4.9 <sub>-3</sub>	100	936	4.6 <sub>-3</sub>	100	14,808	2.5 <sub>-3</sub>	100
9	0.5%	31,032	5.0 <sub>-3</sub>	100	774	4.4 <sub>-3</sub>	100	11,430	3.5 <sub>-3</sub>	100
6	0.5%	38,766	5.2 <sub>-3</sub>	100	606	5.9 <sub>-3</sub>	100	8,790	5.7 <sub>-3</sub>	98
12	1.0%	22,620	8.5 <sub>-3</sub>	97	936	7.2 <sub>-3</sub>	100	14,712	5.1 <sub>-3</sub>	100
9	1.0%	24,813	8.5 <sub>-3</sub>	96	774	7.3 <sub>-3</sub>	100	11,331	7.1 <sub>-3</sub>	99
6	1.0%	$2 \cdot 10^5$	7.1 <sub>-3</sub>	98	600	9.9 <sub>-3</sub>	53	8,634	1.1 <sub>-2</sub>	8
12	5.0%	$9 \cdot 10^5$	4.7 <sub>-3</sub>	90	936	3.3 <sub>-2</sub>	0	14,148	2.6 <sub>-2</sub>	0
9	5.0%	$8 \cdot 10^5$	4.2 <sub>-3</sub>	90	774	3.4 <sub>-2</sub>	0	10,701	3.5 <sub>-2</sub>	0
6	5.0%	$5 \cdot 10^5$	4.8 <sub>-3</sub>	98	600	4.4 <sub>-2</sub>	0	7,728	5.4 <sub>-2</sub>	0
12	10.0%	$1 \cdot 10^6$	4.1 <sub>-3</sub>	89	912	6.7 <sub>-2</sub>	0	13,548	5.3 <sub>-2</sub>	0
9	10.0%	$7 \cdot 10^5$	3.6 <sub>-3</sub>	91	765	6.7 <sub>-2</sub>	0	10,125	7.1 <sub>-2</sub>	0
6	10.0%	$5 \cdot 10^5$	3.2 <sub>-3</sub>	100	588	8.2 <sub>-2</sub>	0	7,098	1.1 <sub>-1</sub>	0
12	50.0%	$5 \cdot 10^5$	2.7 <sub>-3</sub>	94	888	3.7 <sub>-1</sub>	0	11,424	3.3 <sub>-1</sub>	0
9	50.0%	$3 \cdot 10^5$	2.0 <sub>-3</sub>	99	738	3.5 <sub>-1</sub>	0	8,586	4.3 <sub>-1</sub>	0
6	50.0%	$2 \cdot 10^5$	2.4 <sub>-3</sub>	95	588	3.7 <sub>-1</sub>	0	7,176	6.3 <sub>-1</sub>	0

#### 5.1.3 Two-dimensional signal

We conduct a third experiment on a stylized application in order to assess the scalability of the approach to larger problem sizes. In this case the measured signal  $x_0$  is a two-dimensional real-valued image of size  $1600 \times 1350$  pixels, a grayscale version of the image shown in Figure 5.1, which corresponds to  $n = 2.16 \cdot 10^6$ . The dimension of the ambient space of the primal lifted formulation is  $\binom{n+1}{2} > 2.3 \cdot 10^{12}$ , which makes it clear that the resulting SDP is enormous, and must be handled by a specialized solver. We have excluded TFOCS from the list of candidate solvers because it cannot make progress on this example. We generate 10 and 15 octanary masks. Table 5.3 summarizes the results. The column headers carry the same meaning as Table 5.1.



Figure 5.1: Image used for phase retrieval experiment; size  $1600 \times 1350$  pixels (7.5MB).

Table 5.3: Phase retrieval comparisons on a 2-dimensional image.

$L$	GAUGE		GAUGE-feas		WFLOW	
	nDFT	xErr	nDFT	xErr	nDFT	xErr
15	200,835	$2.1 \cdot 10^{-6}$	5,700	$2.1 \cdot 10^{-6}$	8,100	$4.1 \cdot 10^{-6}$
10	195,210	$5.8 \cdot 10^{-7}$	12,280	$9.1 \cdot 10^{-7}$	12,340	$2.1 \cdot 10^{-5}$

## 5.2 Blind deconvolution

In this blind deconvolution experiment, the convolution of two signals  $s_1 \in \mathbb{C}^m$  and  $s_2 \in \mathbb{C}^m$  are measured. Let  $B_1 \in \mathbb{C}^{m \times n_1}$  and  $B_2 \in \mathbb{C}^{m \times n_2}$  be two bases. The circular convolution of the signals can be described by

$$\begin{aligned} b &= s_1 * s_2 = (B_1 x_1) * (B_2 x_2) \\ &= F^{-1} \text{diag} \left( (F B_1 x_1) (F B_2 x_2)^T \right) \\ &= F^{-1} \text{diag} \left( (F B_1) (x_1 \bar{x}_2^*) (\overline{F B_2})^* \right) =: \mathcal{A}(x_1 \bar{x}_2^*), \end{aligned}$$

where  $\mathcal{A}$  is the corresponding asymmetric linear map with the adjoint

$$\mathcal{A}^* y := (F B_1)^* \text{Diag}(F y) (\overline{F B_2}).$$

Because  $F$  is unitary, it is possible to work instead with measurements

$$\hat{b} \equiv F b = \text{diag} \left( (F B_1) (x_1 \bar{x}_2^*) (\overline{F B_2})^* \right)$$

in the Fourier domain. For the experiments that we run, we choose to work with the former real-valued measurements  $b$  because they do not require accounting for the imaginary parts, and thus the number of constraints in (3.9) that would be required otherwise is reduced by half.

We follow the experimental setup outlined by Ahmed et al. (2014) and use the data and code that they provide. In that setup,  $B_1$  is a subset of the Haar wavelet basis, and  $B_2$  is a mask corresponding to a subset of the identity basis. The top row of Figure 5.2 shows the original image, a depiction of the motion blur kernel, and the observed blurred image. The second row of the figure shows the image reconstructed using the augmented Lagrangian code provided by Ahmed et al. (2014), **GAUGE**, and **GAUGE-feas**. Table 5.4 summarizes the results of applying the three solvers. The columns headed “nDFT” and “nDWT” count the number of discrete Fourier and wavelet transforms required by each solver; the columns headed “xErr1” and “xErr2” report the relative errors  $\|x_i - \hat{x}_i\|_2 / \|x_i\|_2$ ,  $i = 1, 2$ , where  $\hat{x}_i$  are the recovered solutions; the column headed “rErr” reports the relative residual error  $\|b - \mathcal{A}(\hat{x}_1 \bar{\hat{x}}_2^*)\|_2 / \|b\|_2$ . Although the non-convex augmented Lagrangian approach yields visibly better recovery of the image, the table reveals that the solutions are numerically similar, and are recovered with far less work.

It is noteworthy that this scenario is extremely limited in the context of blind deconvolution problems. Our aim with its setup is to compare the computational approach described in Chapter 4 against the augmented Lagrangian method used by Ahmed et al. (2014) for the numerical solution of the nuclear norm minimization problem arising from relaxing the lifted formulation presented in §1.1.2.

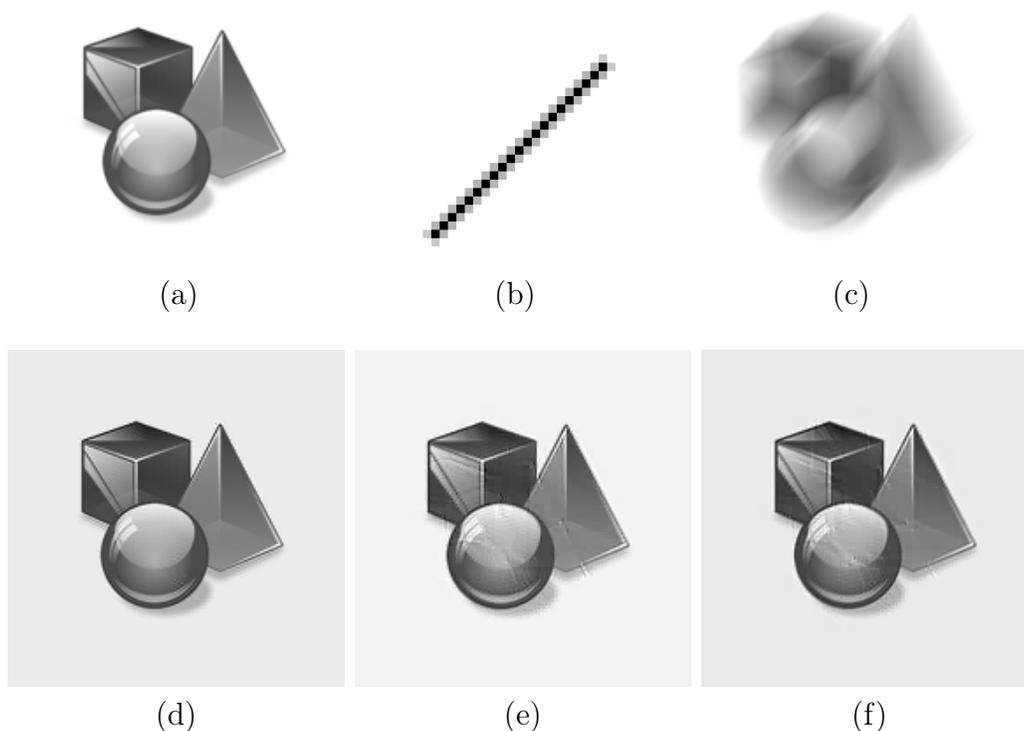


Figure 5.2: Images used for the blind deconvolution experiments: (a) original image; (b) zoom of the motion-blurring kernel; (c) blurred image; (d) image recovered by the augmented Lagrangian approach; (e) image recovered by **GAUGE**; (f) image recovered by **GAUGE-feas**. The shaded background on the recovered images is an artifact of the uniform scaling used to highlight any error between the original and recovered signals.

Table 5.4: Blind deconvolution comparisons.

solver	nDFT	nDWT	xErr1	xErr2	rErr
augmented Lagrangian	92,224	76,872	$7.9 \cdot 10^{-2}$	$5.0 \cdot 10^{-1}$	$1.4 \cdot 10^{-4}$
<b>GAUGE</b>	17,432	8,374	$1.9 \cdot 10^{-2}$	$5.4 \cdot 10^{-1}$	$3.8 \cdot 10^{-4}$
<b>GAUGE-feas</b>	4,128	2,062	$8.4 \cdot 10^{-2}$	$5.5 \cdot 10^{-1}$	$4.0 \cdot 10^{-4}$

# Chapter 6

## Conclusions

The reconstruction of low-complexity signals from given data, under different measurement and noise models comprises an important field of research that generalizes many recent developments in compressed sensing, and has an impact on a broad range of applications. It encompasses different notions of complexity—e.g., sparsity, low rank, atomic norms—, reconstruction algorithms and the theoretical conditions under which these achieve recoverability, either exactly or approximately. Recent approaches to these problems rely on convexifying a suitable, but hard, optimization reformulation and providing conditions under which, for certain measurement and noise models, the solution of the resulting convex optimization problem will very likely yield a solution of the original problem.

This thesis focused on the practical aspect of providing an effective approach to numerically solve the large spectral optimization problems arising from the matrix lifting convex relaxation approach for phase recovery (Candès et al., 2013b) and blind deconvolution (Ahmed et al., 2014). By identifying both problems as particular cases of a class of gauge optimization problems, via suitable reformulations, we were able to abstract some of the structure and contribute results useful to our approach while expanding the theoretical framework of gauge optimization.

In the following paragraphs, we discuss our salient developments while presenting some interesting opportunities for further investigation.

**A theoretical framework based on gauge duality.** Chapter 2 focused on presenting and extending the duality theory of gauge optimization to a particular—but still rather broad—class of problems, and on providing theoretical tools to simplify the manipulations commonly involved in modeling problems that fit this framework. Some results for dealing with antipolar sets were presented in §2.2 which can be invoked in the derivation of dual problems.

The structure particular to gauge optimization allows for an alternative to the usual Lagrange duality (a parallel explored in §2.3), which was later useful for providing an avenue of exploration for modeling and algorithm development. Depending on the particular application, it may prove computationally

convenient or more efficient to use existing algorithms to solve the gauge dual rather than the Lagrange dual problem. For example, a variation of the projected subgradient method was used to exploit the relative simplicity of the gauge dual constraints in the eigenvalue optimization problem (3.7). As with methods that solve the Lagrange dual problem, a procedure would be needed to recover the primal solution. Although this is difficult to do in general, for specific problems it is possible to develop a primal-from-dual recovery procedure via the optimality conditions, an approach undertaken later in Chapter 4.

More generally, an important question left unanswered is if there exists a class of algorithms that can leverage this special structure. It is intriguing the possibility of developing a primal-dual algorithm specialized to the gauge-constrained primal-dual gauge pair.

The sensitivity analysis presented in §2.5 relied on existing results from Lagrange duality. It would be preferable, however, to develop a line of analysis that is self-contained and based entirely on gauge duality theory and some form of “gauge multipliers”. In this regard, if we define the value function as  $\tilde{v}(b, \sigma) = \inf_x \{ \kappa(x) + \delta_{\text{epi } \rho}(b - Ax, \sigma) \}$ , then  $\tilde{v}$  is a gauge. It is conceivable that sensitivity analysis could be carried out based on studying its polar, given by

$$\tilde{v}^\circ(y, \tau) = \inf \{ \mu \geq 0 \mid (y, \tau) \in \mu \mathcal{D} \} = \kappa^\circ(A^*y) + \delta_{\rho^\circ(\cdot) \leq \cdot}(y, -\tau),$$

where  $\mathcal{D} = \{ (y, \tau) \mid \kappa^\circ(A^*y) \leq 1, \rho^\circ(y) \leq -\tau \}$ . This formula follows from Proposition 2.1.2(iv) and a computation of  $\tilde{v}^*$  similar to the one leading to (2.27). This approach would be in contrast to the usual sensitivity analysis, which is based on studying a certain (convex) value function and its conjugate.

Extensions of the standard primal-dual gauge pair are presented in §2.6. They were motivated by subproblems that show up in iteratively reweighted algorithms for trace and nuclear-norm minimization. These are further specialized in Chapter 3 to provide concrete primal-dual pairs for trace minimization in the PSD cone, connecting our spectral optimization problems to the gauge framework.

**An approach to low-rank spectral optimization.** One of the criticisms that have been leveled at relaxation approaches based on matrix lifting is that they lead to problems that are too difficult to be useful in practice due to the quadratic increase in the primal ambient space. This has led to work on non-convex recovery algorithms that may not have as-strong statistical recovery guarantees, but are nonetheless effective in practice; Netrapalli et al. (2013); Candès et al. (2015a); White et al. (2015). A major motivation for

this work was to determine whether it would be possible to develop convex optimization algorithms that are as efficient as non-convex approaches. The numerical experiments on these problems presented in Chapter 5 suggest that the gauge-dual approach may prove effective. Indeed, other convex optimization algorithms may be possible, and clearly the key to their success will be to leverage the special structure of these problems.

A theoretical question not addressed here is to delineate conditions under which dual attainment will hold. In particular, the conclusion (4.5) of Theorem 4.1.1 is asymmetric: it is possible to assert that, if the primal problem is feasible, a primal point exists that attains the primal optimal value while satisfying the constraints (because the Lagrange dual is strictly feasible), but in general one cannot assert that a dual point exists such that it attains the dual optimal value while still being feasible. A related theoretical question is to understand the relationship between the quality of a “near optimal” dual solution, and the quality of the primal estimate obtained from it by the primal recovery procedure.

During the course of our experiments, we noticed that the rightmost eigenvalue of  $\mathcal{A}^*y$  seems to remain fairly well separated from the others across iterations. This appears to contribute to the overall effectiveness of the dual-descent method. Is there a special property of these problems or of the algorithm that encourages this separation? It seems likely that there are solutions  $y$  at which the objective is not differentiable, and in that case, one can wonder if there are algorithmic devices that might be used to avoid such points.

The dual-descent method used to solve the dual subproblem (cf. §4.2.1) is only one possible algorithm among many. Other more specialized methods, such as the spectral bundle method of Helmberg and Rendl (2000), its second-order variant (Helmberg, Overton, and Rendl, 2014), or the stochastic-gradient method of d’Aspremont and Karoui (2014), may prove effective alternatives. Chapter 3 includes the construction of a family of lower models for the dual gauge objective which might be exploited to construct bundle-type methods for its minimization, this seems an interesting avenue for further investigation.

It was convenient to embed the nuclear-norm minimization problem (4.1b) in the SDP formulation (4.1a) because it allows us to use the same solver for both problems. Further efficiencies, however, may be gained by implementing a solver that applied directly to the corresponding gauge dual

$$\underset{y \in \mathbb{C}^m}{\text{minimize}} \quad \|\mathcal{A}^*y\|_\infty \quad \text{subject to} \quad \Re \langle y, b \rangle - \epsilon \|y\|_* \geq 1.$$

This would require an iterative solver for evaluating leading singular values and singular vectors of the non-symmetric operator  $\mathcal{A}^*y$ , such as PROPACK (Larsen, 2001).

# Bibliography

- A. Ahmed, B. Recht, and J. Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, 2014.
- A. Y. Aravkin, J. V. Burke, and M. P. Friedlander. Variational properties of value functions. *SIAM Journal on Optimization*, 23(3):1689–1717, 2013.
- A. Auslender and M. Teboulle. *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2003.
- F. Bach. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends in Machine Learning*, 6(2-3):145–373, 2013.
- J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8:141–148, 1988.
- S. Becker, E. Candès, and M. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3:165–218, 2011.
- A. Berman. *Cones, Matrices and Mathematical Programming*, volume 79 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1973.
- D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, second edition, 1999.
- D. P. Bertsekas. *Convex optimization algorithms*. Athena Scientific, Belmont, MA, 2015.
- D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: a convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005.

- J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical optimization: Theoretical and practical aspects*. Universitext. Springer-Verlag, Berlin, second edition, 2006. ISBN 3-540-35445-X.
- A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.
- S. Burer and R. D. C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2, Series B):329–357, 2003.
- E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, December 2005.
- E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted L1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.
- E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, April 2015a.
- E. J. Candès and X. Li. Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. *Foundations of Computational Mathematics*, 14(5):1017–1026, October 2014.
- E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.
- E. J. Candès and B. Recht. Simple bounds for recovering low-complexity models. *Mathematical Programming*, 141(1-2, Series A):577–589, 2013.
- E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013a.
- E. J. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013b.
- E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, 2015b.

- A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements. *Inverse Problems*, 27(1):015005, 2011.
- V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.
- S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.
- Y. Chen and E. J. Candès. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *CoRR*, abs/1505.05114, 2015. URL <http://arxiv.org/abs/1505.05114>. (Retrieved on December 16, 2015).
- R. Correa and C. Lemaréchal. Convergence of some algorithms for convex minimization. *Mathematical Programming*, 62(2, Series B):261–275, 1993.
- A. d’Aspremont and N. E. Karoui. A stochastic smoothing algorithm for semidefinite programming. *SIAM Journal on Optimization*, 24(3):1138–1177, 2014.
- L. Demanet and P. Hand. Stable optimizationless recovery from phaseless linear measurements. *Journal of Fourier Analysis and Applications*, 20(1):199–221, 2014.
- D. L. Donoho. For most large underdetermined systems of linear equations, the minimal  $l_1$ -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59, 2006.
- Y. C. Eldar, D. Needell, and Y. Plan. Uniqueness conditions for low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 33(2):309–314, 2012.
- M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Elec. Eng. Dept, Stanford University, 2002.
- M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *American Control Conference*, Arlington, 2001.
- J. R. Fienup. Reconstruction of an object from the modulus of its fourier transform. *Optics Letters*, 3(1):27–29, Jul 1978.

- J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21 (15):2758–2769, Aug 1982.
- R. M. Freund. Dual gauge programs, with applications to quadratic programming and the minimum-norm problem. *Mathematical Programming*, 38 (1):47–67, 1987.
- M. P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz. Recovering compressively sampled signals using partial support information. *IEEE Trans. Inform. Theory*, 58(2):1122–1134, 2012.
- M. P. Friedlander and I. Macêdo. Low-rank spectral optimization. *CoRR*, abs/1508.00315, 2015. URL <http://arxiv.org/abs/1508.00315>. (Retrieved on December 16, 2015).
- M. P. Friedlander, I. Macêdo, and T. K. Pong. Gauge optimization and duality. *SIAM Journal on Optimization*, 24(4):1999–2022, 2014. doi: 10.1137/130940785. URL <https://dx.doi.org/10.1137/130940785>. (Retrieved on December 16, 2015).
- R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik*, 35: 237–246, 1972.
- M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, November 1995.
- D. Gross, F. Kraemer, and R. Kueng. Improved recovery guarantees for phase retrieval from coded diffraction patterns. *CoRR*, abs/1402.6286, 2014. URL <http://arxiv.org/abs/1402.6286>. (Retrieved on December 16, 2015).
- C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10(3):673–696, 2000.
- C. Helmberg, M. L. Overton, and F. Rendl. The spectral bundle method with second-order information. *Optimization Methods and Software*, 29(4): 855–876, 2014.
- J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. II: Advanced theory and bundle methods*, volume 306 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

- M. Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 427–435, 2013.
- D. Krishnan, P. Lin, and A. M. Yip. A primal-dual active-set method for non-negativity constrained total variation deblurring problems. *IEEE Transactions on Image Processing*, 16(11):2766–2777, 2007.
- R. M. Larsen. Combining implicit restart and partial reorthogonalization in Lanczos bidiagonalization, 2001. URL <http://sun.stanford.edu/~rmunk/PROPACK/>. (Retrieved on December 16, 2015).
- R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*, volume 6. SIAM, 1998.
- A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2354–2367, Dec 2011.
- A. S. Lewis. Convex analysis on the hermitian matrices. *SIAM Journal on Optimization*, 6(1):164–177, 1996.
- S. Ling and T. Strohmer. Self-calibration and biconvex compressive sensing. *CoRR*, abs/1501.06864, 2015. URL <http://arxiv.org/abs/1501.06864>. (Retrieved on December 16, 2015).
- L. Lovász. Submodular functions and convexity. In A. Bachem, B. Korte, and M. Grötschel, editors, *Mathematical Programming The State of the Art*, pages 235–257. Springer, 1983.
- L. McLinden. Symmetric duality for structured convex programs. *Transactions of the American Mathematical Society*, 245:147–181, 1978.
- R. P. Millane. Recent advances in phase retrieval. *Proceedings of SPIE 6316, Image Reconstruction from Incomplete Data IV, 63160E*, September 2006.
- K. Mohan and M. Fazel. Reweighted nuclear norm minimization with application to system identification. In *American Control Conference (ACC), 2010*, pages 2953–2959, June 2010.
- B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234, 1995.

- Y. Nesterov. Unconstrained convex minimization in relative scale. *Mathematics of Operations Research*, 34(1):180–193, 2009.
- P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems 26*, pages 2796–2804, 2013.
- J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, second edition, 2006.
- M. L. Overton. Large-scale optimization of eigenvalues. *SIAM Journal on Optimization*, 2(1):88–120, 1992.
- M. V. Ramana, L. Tunçel, and H. Wolkowicz. Strong duality for semidefinite programming. *SIAM Journal on Optimization*, 7(3):641–662, 1997.
- B. Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12:3413–3430, 2011.
- B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.
- B. Recht, W. Xu, and B. Hassibi. Null space conditions and thresholds for rank minimization. *Mathematical Programming*, 127(1, Series B):175–202, 2011.
- P. Richtárik. Improved algorithms for convex minimization in relative scale. *SIAM Journal on Optimization*, 21(3):1141–1167, 2011.
- R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
- R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, New York, 1998.
- P. H. Ruys and H. M. Weddepohl. Economic theory and duality. In J. Kriens, editor, *Convex analysis and mathematical economics*, pages 1–72. Springer, Berlin, 1979.
- M. Schmidt. minFunc: unconstrained differentiable multivariate optimization in Matlab, 2005. URL <http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>. (Retrieved on December 16, 2015).

- T. Strohmer. Personal communication, December 2013.
- D. S. Taubman and M. W. Marcellin. *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, Norwell, MA, USA, 2001. ISBN 079237519X.
- M. Teboulle and J. Kogan. Applications of optimization methods to robust stability of linear systems. *Journal of Optimization Theory and Applications*, 81(1):169–192, 1994.
- T. Teuber, G. Steidl, and R. H. Chan. Minimization and parameter estimation for seminorm regularization models with I-divergence constraints. *Inverse Problems*, 29(3), 2013.
- M. J. Todd and Y. Ye. Approximate Farkas lemmas and stopping rules for iterative infeasible-point algorithms for linear programming. *Mathematical Programming*, 81(1, Series A):1–21, 1998.
- L. Tunçel and H. Wolkowicz. Strong duality and minimal representations for cone optimization. *Computational Optimization and Applications*, 53(2): 619–648, 2012.
- E. van den Berg. *Convex optimization for generalized sparse recovery*. PhD thesis, University of British Columbia, December 2009.
- E. van den Berg and M. P. Friedlander. Sparse optimization with least-squares constraints. *SIAM Journal on Optimization*, 21(4):1201–1229, 2011.
- C. D. White, S. Sanghavi, and R. Ward. The local convexity of solving systems of quadratic equations. *CoRR*, abs/1506.07868, 2015. URL <http://arxiv.org/abs/1506.07868>. (Retrieved on December 16, 2015).
- H. Zhang and W. Hager. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM Journal on Optimization*, 14(4):1043–1056, 2004.