On the optimal sampling design for the Graph Total Variation decoder: recovering piecewise-constant graph signals from a few vertex measurements

by
Rodrigo CERQUEIRA GONZALEZ PENA

Lausanne, Switzerland
2019

Quem me dera
um mapa de tesouro
que me leve a um velho baú
cheio de mapas do tesouro

— Paulo Leminski

Acknowledgements

First of all, I am grateful that Pierre Vandergheynst accepted me into his group (LTS2) and had the insight to propose a research topic that would keep me interested, and busy, for years. Thank you in particular for encouraging me to go to any conference I was drawn to, even if I had no work to present there.

At LTS2, my colleagues could not have been more welcoming. Benjamin Ricaud and Xavier Bresson were fundamental in introducing me to research life. Andreas Loukas became the source for discussing theory, and the ideal partner for playing questionable, but heartfelt, Flight of the Conchords covers. Through ups and downs, the Ph.D. students of LTS2 developed a good sense of comradery. I will not disclose, however, how much of this development involved table football. Listing everyone from LTS2, and neighboring groups, who helped me go through my doctoral studies would fill most of this page, but I thank you all for the time we shared.

Most of my Ph.D. was funded by the European Union's FP7-PEOPLE-2013-ITN program under grant agreement n^{o} 607290 SpaRTaN. While I do appreciate the money, I would like to mostly thank the *people* from SpaRTaN and its partner network, MacSeNet. We learned together and we laughed together, and I believe we are all better off for having met one another. Especial thanks go to everyone at the CVSSP at the University of Surrey and at the Fraunhofer IDMT for hosting me.

Much to my family's fault, I consider a life in pursuit of education to be a noble endeavor. Often when surrounded by books I am transported back home. I know that being nine thousand kilometers apart is hard on us, so I appreciate your support. My grandfather, being an immigrant himself, told me upon my moving that I now "belonged to the world". I agree, but I also believe in visiting for some coffee and loud conversations around the table.

Lastly, and most importantly, I would like to thank my wife and confidante extraordinaire, Anna. She was a published author years before me, but never bragged about it. Anna lifted my mood when I had a hard time putting thoughts to paper; she will probably continue doing so long after I forget what I wrote here. This work is dedicated to you.

Lausanne, 30 September 2019 *R. C. G. P.*

Abstract

Compressed Sensing teaches us that measurements can be traded for offline computation if the signal being sensed has a simple enough representation. Proper decoders can exactly recover the high-dimensional signal of interest from a lower-dimensional vector of that signal's observations. In graph domains — like social, similarity, or interaction networks — the relevant signals often have to do with the network's cluster structure. Partitioning a graph into different communities induces a piecewise-constant signal, an object that can be decoded via Graph Total Variation ($\mathcal{G}$-TV) minimization even if it is not fully observed. In fact, assume that such a signal can only be accessed by querying vertices at random. Then, we could sensibly ask: what are the sampling probabilities that minimize the number of queries required for a successful $\mathcal{G}$-TV recovery? This thesis is an attempt to answer this question through the study of the success conditions in $\mathcal{G}$-TV minimization programs. I show that the recovery error in these programs undergoes a phase transition in terms of the number of measurements, with a threshold that explicitly depends on the vertex-sampling probabilities. It suffices to minimize this threshold to obtain an optimal sampling design. Yet, sampling optimally in practice has problems of its own. While numerical experiments reveal that it is important to focus on the places of the graph where the signal varies, implementing the optimal design without actually knowing the signal-to-be-sampled remains an open issue.

Keywords: Total Variation, graph signal processing, community structure, piecewise-constant signal, convex optimization, ℓ_1 minimization, analysis sparsity, representer theorem, minimum restricted eigenvalue, small ball method, inexact dual certificate, golfing scheme

Résumé

L'Acquisition Comprimée nous enseigne qu'il est possible d'échanger des mesures contre du calcul hors-ligne tant que le signal mesuré a une représentation suffisamment simple. Des décodeurs appropriés peuvent récupérer exactement le signal d'intérêt de grande dimension à partir d'un vecteur de plus petite dimension contenant des observations de ce signal. Dans les domaines de graphes — comme les réseaux sociaux, de similarité ou d'interaction —, les signaux pertinents ont souvent à voir avec la structure des clusters du réseau. Le partitionnement d'un graphe en différentes communautés induit un signal constant par morceaux, un objet qui peut être décodé via la minimisation de la Variation Totale sur le Graphe ($\mathscr{G}$-TV) même s'il n'est pas complètement observé. En fait, supposons qu'un tel signal ne soit accessible qu'en interrogeant des nœuds de manière aléatoire. Alors, nous pourrions raisonnablement demander : quelles sont les probabilités d'échantillonnage minimisant le nombre de requêtes nécessaires à une reconstruction $\mathscr{G}$-TV réussie ? Cette thèse tente de répondre à cette question en étudiant les conditions de réussite des programmes de minimisation $\mathscr{G}$-TV. Je montre que l'erreur de reconstruction dans ces programmes subit une transition de phase en termes du nombre de mesures, avec un seuil qui dépend explicitement des probabilités d'échantillonnage des nœuds. Il suffit de minimiser ce seuil pour obtenir un plan d'échantillonnage optimal. Cependant, échantillonner optimalement dans la pratique pose des problèmes particuliers. Bien que les expériences numériques révèlent qu'il est important de se concentrer sur les endroits du graphe où le signal varie, la mise en œuvre du plan optimal sans connaître réellement le signal à échantillonner reste une question ouverte.

Mots clés : Variation Totale, traitement du signal dans les graphes, structure de communauté, signal constant par morceaux, optimisation convexe, minimisation ℓ_1, *analysis sparsity*, théorème du représentant, valeur propre restreinte minimale, *small ball method*, *inexact dual certificate*, *golfing scheme*

Contents

List of Acronyms

i.i.d.	Independent and Identically Distributed
$\mathscr{G}$**-TV**	Graph Total Variation
Ber(π)	Bernoulli Sampling Model
CS	Compressed Sensing
CSWR(π)	Coordinate Sampling with Replacement
FBF	Forward-Backward-Forward
GSP	Graph Signal Processing
KKT	Karush-Kuhn-Tucker
LHS	left-hand side
MAP	Maximum a Posteriori
NNSP	Network Null Space Property
p.s.d.	Positive Semi-Definite
RHS	right-hand side
RIP	Restricted Isometry Property
SBM	Stochastic Block Model
TV	Total Variation
w.h.p.	with high probability
w.r.t.	with respect to

List of Symbols

s	A scalar
$\mathbf{v}$	A vector
$\mathbf{M}$	A matrix
$\mathscr{S}$	A set
M_{ij}	The entry in row i, column j of matrix $\mathbf{M}$
v_i	The i^{th} entry of a vector $\mathbf{v}$
$\nabla f(\mathbf{v})$	The gradient of function f at vector $\mathbf{v}$
$\text{diag}(\mathbf{M})$	The vector formed by extracting the main diagonal of matrix $\mathbf{M}$
$\text{diag}(\mathbf{v})$	The diagonal matrix whose main diagonal is equal to vector $\mathbf{v}$
$\mathbf{I}_n$	The $n \times n$ identity matrix
$\mathbf{M}^+$	The Moore-Penrose pseudoinverse of matrix $\mathbf{M}$
$\Pi_{\mathscr{S}}$	The orthogonal projection operator onto set $\mathscr{S}$
$\mathbf{P}_{\mathscr{S}}$	The orthogonal projection operator onto $\text{span}\{\mathbf{e}_i : i \in \mathscr{S}\}$. Also refered to as "coordinate projection".
$\mathbf{M}^\top$	The transpose of matrix $\mathbf{M}$
$\mathscr{D}(f, \mathbf{v})$	The descent cone of function f at vector $\mathbf{x}$
$\partial f(\mathbf{v})$	The subdifferential set of function f at vector $\mathbf{v}$
$\mathbb{B}_{\mathscr{G}\text{-}\mathbf{TV}}$	The set of unit Graph Total Variation semi-norm vectors in $\mathbb{R}^n$
$\mathbb{B}_q^n$	The set of unit q-norm vectors in $\mathbb{R}^n$, that is, $\{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v}\|_q \leq 1\}$
$\text{bd}(\mathscr{S})$	The boundary of a set $\mathscr{S}$
$\mathbb{S}^{n-1}$	The Euclidean sphere in $\mathbb{R}^n$. Can also be written as $\text{bd}(\mathbb{B}_2^n)$
$\mathscr{S}^n$	The cartesian product $\underbrace{\mathscr{S} \times \cdots \times \mathscr{S}}_{n \text{ times}}$ of a set $\mathscr{S}$
$\text{conv}(\mathscr{S})$	The convex hull of set $\mathscr{S}$
$\mathscr{G}_{k,n}$	The Grassmannian manifold of k-dimensional subspaces in $\mathbb{R}^n$
$\mathscr{S}^{m \times n}$	The set of $m \times n$ matrices with entries that take value in a set $\mathscr{S}$
$[n]$	The set $\{1, 2, \ldots, n\}$ of positive integers going from 1 to n
$\mathbb{R}$	The set of real numbers
$\{\tilde{\mathbf{e}}_i\}_{i=1}^N$	The set of standard basis vectors in $\mathbb{R}^N$
$\{\mathbf{e}_i\}_{i=1}^n$	The set of standard basis vectors in $\mathbb{R}^n$
$\|\mathbf{v}\|_0$	The number of non-zero elements in vector $\mathbf{v}$. Also known as the ℓ_0 "norm"
$\|\mathbf{v}\|_p$	The ℓ_p norm $\left(\sum_i \|\mathbf{v}_i\|^p\right)^{1/p}$ of a vector $\mathbf{v}$, for $p \in (0, \infty)$
$\|\mathbf{v}\|_\infty$	The maximum absolute entry in a vector $\mathbf{v}$. Also known as the ℓ_∞ norm

List of Figures

List of Tables

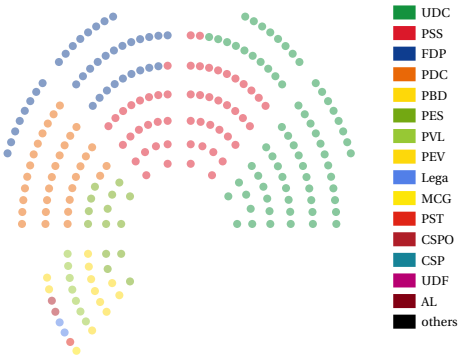
1 Introduction

> Eventually everything connects - people, ideas, objects.
> The quality of the connections is the key to quality per se.
>
> — Charles Eames

How much do the votes of politicians reflect their party affiliations? Stokes writes that "political parties are endemic to democracy" [68, p.245], so wondering about the empirical expression of party beliefs is important to diagnose issues in the democratic process. It helps inform the debate of whether political parties "made modern democracy", contributing to its responsiveness and realization of public goods, or "are an inextricable weed" in democracy's garden, "partial to their own conception of the good" [68, pp.263–264]. In Switzerland, the National Council is one of the main stages of federal politics [1]. Since 1963, it has had a fixed number of seats (200) divided into shares proportional to each canton's percentage of the total population. Every four years the Swiss people elect a new council, the most recent of which has held office from 2015 to 2019. It is the 50th legislature since the foundation of the Swiss federation, in 1848. Figure 1.1 displays a photo of the Council hall, next to a scatter plot of the National Councillors taking part in the 50th legislature, highlighting their party affiliations.

[1] It is one of the two houses that form the Federal Assembly of Switzerland, the other being the Council of States.

(a) The Council hall at the Federal Palace in Bern, Switzerland. © 2006 http://www.parlament.ch



(b) Council members and their party affiliations during the 50th legislature.

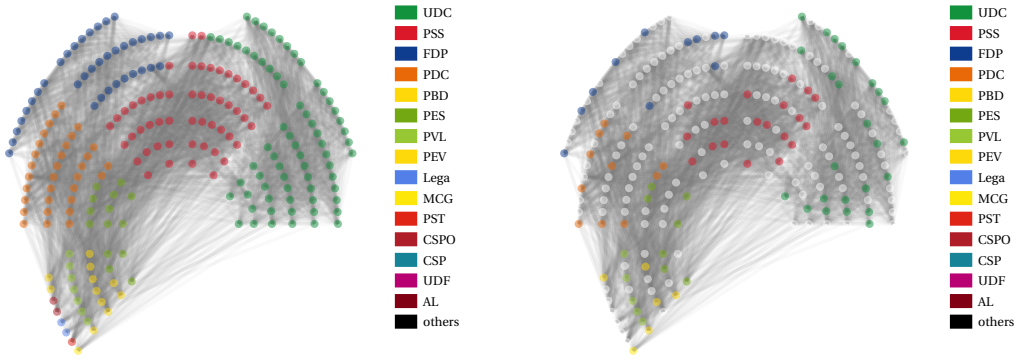
Figure 1.1 – The Swiss National Council. The scatter plot on the right has more "chairs" than the 200 in the actual council hall on the left because I account for councillors that resigned or were replaced throughout the 50th legislature.

The Swiss Parliament has an Open Data policy, allowing — in particular — the consultation of voting results through a database of parliamentary votes [2]. With this data, we can find out how each National Councillor voted in each of the affairs in the 50th legislature, and, comparing the voting patterns, record how similar the councilors are to one another in their voting behaviors. I depict these similarities in the form of a network (or graph) on Figure 1.2a, with edges connecting councilors who voted the most alike during the period 2015–2019. [3] Despite increasing polarization of the Council since the 1990s — tied to the rise of the right-wing UDC party [4] — it is not rare to find in Figure 1.2a councilors from different parties that nonetheless vote alike. Had the Council been completely polarized, vertices of different colors would never connect. Still, party affiliation is not meaningless, as there seem to be more connections within than across parties. But how much of the party division in the Swiss National Council is encoded in the network structure induced by the voting data? Or, more concretely, if we knew the party affiliations of half the councilors, as in Figure 1.2b, could we infer the other half of the labels based on the connectivity information?

[2] See https://www.parlament.ch/en/ratsbetrieb/abstimmungen/abstimmungs-datenbank-nr. I was pointed to this source by D. Debruyn, Y. Morize, N. Orgland, and S. Stettler, who were all EPFL Master's students at the time.

[3] I disclose the precise way in which this graph is constructed only in Chapter 6. For now, it suffices to interpret connected vertices as representing councilors with similar voting decisions.

[4] https://www.swissinfo.ch/eng/political-drift_polarisation-of-swiss-parliament-continues/43752300



(a) The graph with all the party labels.

(b) The graph with half the party labels.

Figure 1.2 – The voting-similarity graph for the Swiss National Council. Each vertex represents a council member, and the edges connect members that had similar voting patterns during the 50th legislature.

These musings are part of the usual pipeline in modern Data Science [76]. An object of interest in the real-world (Swiss National Council) has some available data (vote results) that can be used to infer some property (party labels) of the object in question. The need for such inference may come from difficulties in measuring the desired property or simply out of scientific interest on the predictive powers of the available data. In the contrived example of the missing party labels in Figure 1.2b, we might be just interested to see if the voting patterns reflect the party affiliations. But in graphs such as the Internet or large social networks querying every single node for some property can be very expensive. The missing information in these cases is intrinsically due to the large size of the network. To infer, or recover, the full content from scarce observations, there are two main issues that practitioners concern themselves with. First, which kind of — and how many — measurements are available. Second, which machinery to use to retrieve the missing information. The first point pertains to what I call the *sampling* stage of the pipeline; the second, to the *decoding* stage. In this thesis, I assume the freedom to act in both, choosing a suitable decoder for an important class of signals and subsequently looking for an optimal sampling strategy adapted to this decoder. Optimality here refers to reducing the number of measurements needed from the sampling stage for a successful decoding stage. An added benefit of studying optimal sampling lies in the possibility to quantify how suited the decoder is in retrieving the signals of interest. The more samples the decoder needs, the less suited it is. Take, for example, the Swiss National Council data, and imagine we chose a decoder for missing party labels (like those of Figure 1.2b) based on the voting-similarity connections. This decoder's optimal number of samples could, therefore, function as a numerical proxy to how well the councilors' votes echo their party affiliations.

1.1 The main objects and questions in the thesis

Networks (or graphs) have long been objects of interest in Mathematics and Computer Science, but they have found their way into Signal Processing over the last decade or so. Graph Signal Processing (GSP) is now an established subfield [66, 64, 55], concerned with any quantities whose support can be interpreted as being a graph. I have already shown a graph signal. Each member (vertex) of the Swiss National Council was associated with a party color, so we can understand the mapping "councilor $\mapsto$ color" in Figure 1.2b as a signal living on the voting-similarity graph. On the Internet, a relevant signal is the number of data packets at each router; on a social network, it might be people's likelihood of buying a given product. Even when a graph is not naturally present — as in the Swiss National Council example —, representing the variable dependencies through a network may be a beneficial pre-processing step. To connect a vertex to only a few neighbors is akin to performing statistical selection, restricting a variable's predictors to its most similar peers. Moreover, these few connections implicitly constrain the interactions between vertices to happen only locally, so comparisons of signal values can be more efficiently computed.

One of the central tenets of GSP can be stated as "connected vertices have similar signal values". [5] It is the glue that binds the graph and the signals that it supports, but it gives room to specify what "similar values" means. This thesis deals with a particular set of graph signals deemed *piecewise-constant*. In loose terms, these objects assume constant values over sizeable swathes of the graph but are allowed to vary abruptly between the constant pieces. The councilors' colors in Figure 1.2b is an example of a piecewise-constant signal: each party defines a piece of constant color, but those colors are allowed to vary abruptly between pieces. In general, any classification or segmentation task can be interpreted as producing a piecewise-constant signal on some graph. Indeed, imagine a social network, then assign value 1 to every person who has watched the 2010 movie "The Social Network", and value 0 otherwise. The $\{0,1\}$-valued labels thus defined form a piecewise-constant signal over the social network. Whenever we know the value of a signal at a vertex of the graph, I will say that this vertex has been sampled. Sampled vertices represent the knowledge that an oracle made available to us to help us figure out the full underlying signal. This thesis admits oracles that provide *vertex samples independently at random*, according to a probability distribution set beforehand. They are a model for any stakeholder that has a budget number of samples with which to query the network for a signal of interest. Some vertices may be more important than others, so they could be sampled with higher likelihood; the samples are kept independent to facilitate the mathematical treatment of the process. Intuitively, if a company wants to sell a product, then "influencers" in a social network should be queried for their propensity to advertise the merchandise. Readers will find in Chapter 2 all the formalism that I use in subsequent chapters when speaking of graphs, signals on graphs, and the sampling of these signals.

[5] An analogous statement, "similar variables have similar outputs", is the central assumption of the related field of Semi-Supervised Learning [15].

The decoders in this thesis manifest themselves in terms of *convex programs*. That is, their output is chosen by minimizing some convex loss function. Piecewise-constant signals vary only across few of the edges on their graph support, so it makes sense to pair them up with a recovery function that penalizes edge-differences. [6] This idea leads to the central subjects of the thesis, Graph Total Variation ($\mathscr{G}$-TV) decoders. They are programs that minimize a semi-norm of the form $\|\mathbf{D}\cdot\|_1$, for some difference operator $\mathbf{D}$. Total Variation (TV) minimization is a standard tool for processing "classic" time-series or images that have rare, but sudden changes in value. Think here of a quantized waveform, or a Mondrian painting. I will show how TV and piecewise-constant graph signals are also intimately related. Chapter 3 provides the required background on convex recovery and introduces the $\mathscr{G}$-TV decoders so central to this thesis.

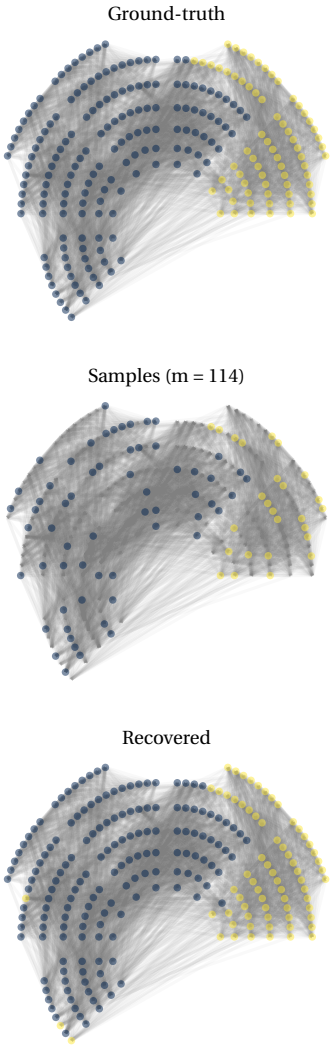


Figure 1.3 – Example of the processing pipeline highlighting the sampling and decoding (recovery) stages. See the paragraph on the left for details.

The performance of the decoder depends on what happens at the sampling stage. To get a better feeling for this, let us revisit the Swiss National Council example, this time with a different graph signal. The top row of Figure 1.3 depicts the indicator function of the Swiss People's Party (UDC in French), the largest party in the 50th legislature, occupying about 30% of the total seats. The signal takes value 1 (yellow) at vertices representing UDC members and 0 (blue) everywhere else. UDC is politically the farthest to the right on the National Council, so the party should be fairly well encoded in the voting patterns represented by the edges of the graph. In the middle row of Figure 1.3, I have sampled the UDC indicator function at 50% of the vertices, uniformly at random. Many instances of both yellow (UDC) and blue (non-UDC) vertices appear in the sample because the party constitutes a considerable share of the Council. The third row of Figure 1.3 shows the decoded signal as output by $\mathscr{G}$-TV interpolation, yet to be introduced in Chapter 3. Most of the vertex labels are recovered correctly, but there are visibly wrong assignments. This means that sampling uniformly at random the labels of *half* the councilors is still not enough to guarantee a perfect recovery using our decoder. Sampling uniformly at random, however, is not the only way to query the vertices under our oracle model. We could sample more often the vertices

[6] In a way, this recovery procedure is an instance of transductive learning [15, Chapter 24] where the sampled vertices form the training set, the unsampled vertices form the testing set, and the loss implicitly defines a search space.

that are poorly connected, or the hubs of the graph, or even mix those two strategies. The possibilities are endless. Rather than think of an ad hoc plan of action, this thesis asks

> What is the vertex sampling design that minimizes the number of measurements required for the success of $\mathscr{G}$-TV decoders?

The best hope one has to find a sampling design that answers the question above is by studying the conditions for successful recovery under $\mathscr{G}$-TV minimization. In the course of this study, we will encounter typical concepts that usually arise when dealing with stochastic objects in high dimensions [77, Ch. 1], such as concentration, universality, and sharp transitions. In particular, I will show that the recovery error drops suddenly to zero (with a high likelihood) at a critical number of measurements that depends on the vertex sampling probabilities. Minimize this threshold and one can find the optimal sampling design.

1.2 Contributions

To prove that a convex program returns the desired output, one must produce a *certificate*, an object whose very existence guarantees the success of the recovery procedure. Chapters 4 and 5 describe two parallel attempts that I made towards producing such guarantees.

In Chapter 4, the certificate manifests itself as a positive lower bound on a minimum gain functional. This view has proven fruitful when Gaussian-like vectors are used to measure the ground-truth signal and has become standard in the literature [14, 72]. However, the nature of vertex sampling gets in the way of the usual small-ball method [50, 38] used to lower bound the minimum gain. As a result, I ultimately fail in this attempt, but I end the chapter pointing towards a possible redemption, reliant on better knowledge of the coordinate structure of a certain "descent cone".

The real win comes in Chapter 5, while seeking a *dual* certificate for the $\mathscr{G}$-TV decoder. There, the Karush-Kuhn-Tucker conditions of the problem motivate a blueprint for an iterative golfing scheme [28] that produces, in the end, the desired recovery guarantee. It is this chapter that the number of measurements implying a successful recovery is upper bounded by an expression that depends on the vertex sampling probabilities. The optimal sampling design then comes directly as a corollary. Interestingly, the sampling probabilities in this design depend on how each vertex perturbs the graph difference operator, restricted to the edges across which the piecewise-constant, signal-to-be-recovered changes in value.

Chapter 6 closes the thesis with a numerical tour to balance out the mostly theoretical discussion up to that point. There I plot — for a variety of graphs and signals of interest —, the phase transition underwent by the recovery error of a $\mathscr{G}$-TV decoder when the number of sampled vertices changes. Experiments show how these phase transitions can be improved

using proper sampling designs, but the question of how to make proper designs that are also *practical* remains open by the end of the book.

1.3 The presentation of the proofs

Most chapters in this thesis contain theorems with somewhat lengthy proofs, whose immediate presentation would disturb the flow of the text. For this reason, I have gathered detailed arguments largely in appendices at the end of the relevant chapters. The proofs themselves are written in a hierarchical structure, based on the guidelines of L. Lamport [43]. For the unfamiliar reader, this means that I build a sequence of discrete, provable claims leading to the desired result. Each of these discrete claims is settled as true either by appealing to established knowledge or by presenting a sequence of discrete sub-claims. In the end, the full proof resembles a nested list. I have avoided too much nesting due to the obvious limitations of the printed format. Still, the hierarchical structure should make it easier to distill the arguments and find exactly where I use each assumption in the statement of their respective theorems.

1.4 Reproducibility and Open Science

This work is licensed under a Creative Commons Attribution 4.0 International License. All the computer code required to reproduce the contents of this thesis is distributed under the MIT License, and hosted at the following repository:

https://github.com/rodrigo-pena/phd-thesis

2 Graphs, signals, and sampling

Graphs are combinatorial objects, but much of graph *signal processing* is a matter of linear algebra. A real-valued signal supported on a network can be embedded on a Euclidean space with dimension equal to the number of vertices. In this space, various difference operators can be defined and used as a basis for Fourier analysis or dynamical process analogs for graph signals [66].

The signals that this thesis cares about are piecewise-constant, a characteristic tantamount to having few variations (or jumps) in value across edges. To measure such variations, we work with an operator $\mathbf{D}$, called the graph gradient matrix, whose induced p-semi-norms $\|\mathbf{D}\cdot\|_p$ yield various ways to quantify the amount of signal variation on the graph. The semi-norms corresponding to $p = 1, 2$ receive special names: Graph Total Variation ($\mathscr{G}$-TV) and Dirichlet form, respectively. The former is one of the main objects of study in the thesis; the latter appears sporadically, as a comparison point to the $\mathscr{G}$-TV.

The act of sampling signal values is also cast in the language of linear algebra. Vertex measurements can be obtained through multiplication with a random matrix $\mathbf{A}$ indexed by a sampling set Ω. Specifying the inclusion probabilities of indices into Ω is called a sampling design. This design is represented numerically by a vector $\boldsymbol{\pi}$, whose influence on the inclusion probabilities I present in two alternative ways, the Bernoulli Sampling Model (Ber($\boldsymbol{\pi}$)) and Coordinate Sampling with Replacement (CSWR($\boldsymbol{\pi}$)). Convenience dictates which one is used when: Ber($\boldsymbol{\pi}$) in Chapter 4; CSWR($\boldsymbol{\pi}$) in Chapter 5. From the perspective of recovery problems, we will see that these two models are essentially equivalent.

This chapter's goal is to explain how the processing and sampling of graph signals can be seen as issues about vectors and matrices. Readers can use it as a reference because much of the notation in the thesis is already established here.

2.1 Graph signal processing

Graphs, or networks, are tuples $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of a vertex set $\mathcal{V}$ and an edge set $\mathcal{E}$. The latter contains ordered pairs $e_{vu} = (v, u)$ that indicate directed connections from some vertex $v \in V$ to another vertex $u \in V$. We will consider a graph to be *undirected* if $(v, u) \in \mathcal{E} \iff (v, u) \in \mathcal{E}$, that is, if vertex v connects to u if and only if u also connects to v. Moreover, a graph has no *self-loops* if $(v, v) \notin \mathcal{E}, \forall\, v \in \mathcal{V}$, and is *connected* if, starting from any vertex, one can visit all the others by following the edges in $\mathcal{E}$. The graphs in the numerical examples of Chapter 6 are all connected, undirected, and without self-loops. Any other reference to graphs in the rest of the text can be assumed to be valid for generic networks.

In machine learning and signal processing, graphs are commonly used to encode "closeness" or "similarity" between the objects represented by the vertices. A common way to quantify those similarities is by attaching non-negative weights to the edges of the graph. The larger the weight, the more similar the corresponding connected vertices. Formally, the weight assignment is done through a function $w : \mathcal{E} \to \mathbb{R}_{\geq 0}$. But an alternative representation of the weights can be derived once we impose an arbitrary ordering $v_1 < v_2 < \cdots < v_n$ to the n vertices in $\mathcal{V}$. With this fixed ordering, we can build a weighted adjacency matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$ with entries given by

$$\forall\, i, j \in [n], \quad W_{ij} = \begin{cases} w((v_i, v_j)) & \text{if } (v_i, v_j) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

On undirected graphs, any pair $(v_i, v_j), (v_j, v_i)$ really represents the same undirected edge. In this case, we make the weight function symmetric, that is, $w((v_i, v_j)) = w((v_j, v_i)), \forall\, v_i, v_j \in \mathcal{V}$. As a consequence, W for an undirected graph is a symmetric matrix. Whenever the weight function satisfies $w \equiv 1$, we say that the graph is unweighted, recovering in W the classic adjacency matrix used in algebraic graph theory.

I reserve the variable n for the number of vertices on a graph. In other words, for any $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we will have $n := |\mathcal{V}|$. Alluding to the imposed order on the vertices of $\mathcal{V}$, we can employ a one-to-one mapping between $\mathcal{V}$ and $[n] := \{1, 2, \ldots, n\}$. This mapping, takes $i \mapsto v_i$, and vice versa, for every $i \in [n]$. Without fear of ambiguity then — and for the sake of presentation —, I interchangeably refer to the vertex set as either $\mathcal{V}$ or $[n]$. Similarity, we keep the letter N for referring to the number of edges, $|\mathcal{E}|$. The capital N also works as a mnemonic device for the fact that the number of edges will almost always be larger than the number of vertices. In fact, for any connected, undirected graph $\mathcal{G}$, the bounds $|\mathcal{V}| - 1 \leq |\mathcal{E}| \leq \frac{|\mathcal{V}|(|\mathcal{V}|-1)}{2}$ apply by a simple counting exercise. The lower bound is reached if $\mathcal{G}$ is a tree, whereas the upper bound holds whenever $\mathcal{G}$ is a complete graph.

Any function $f : \mathcal{V} \to \mathbb{R}^n$ is thought of as a signal on the graph. Intuitively, the naming is justified by imagining each vertex $v \in \mathcal{V}$ as having a real value $f(v)$ living on top of it. We can

then refer to the graph as the support of the signal. This view is borrowed from graphs such as sensor networks, where each vertex (sensor) has a clear signal component (e.g., temperature) attached to it. But we can abstract from this example and refer to, say, the numerical labels of vertices in a clustered network as a graph signal as well.

For processing reasons, it is useful to identify the set of graph signals with the set of vectors in $\mathbb{R}^n$, via the bijection between $\mathcal{V}$ and $[n]$. In practice, this only means that for any $\mathbf{x} \in \mathbb{R}^n$ there exists a graph signal f such that $\mathbf{x} = \big(f(v_i)\big)_{i=1}^n$, for some ordering $v_1 < v_2 < \cdots < v_n$ of the vertices in $\mathcal{V}$. From now on, having this identification in mind, I will only refer to graph signals as vectors in $\mathbb{R}^n$.

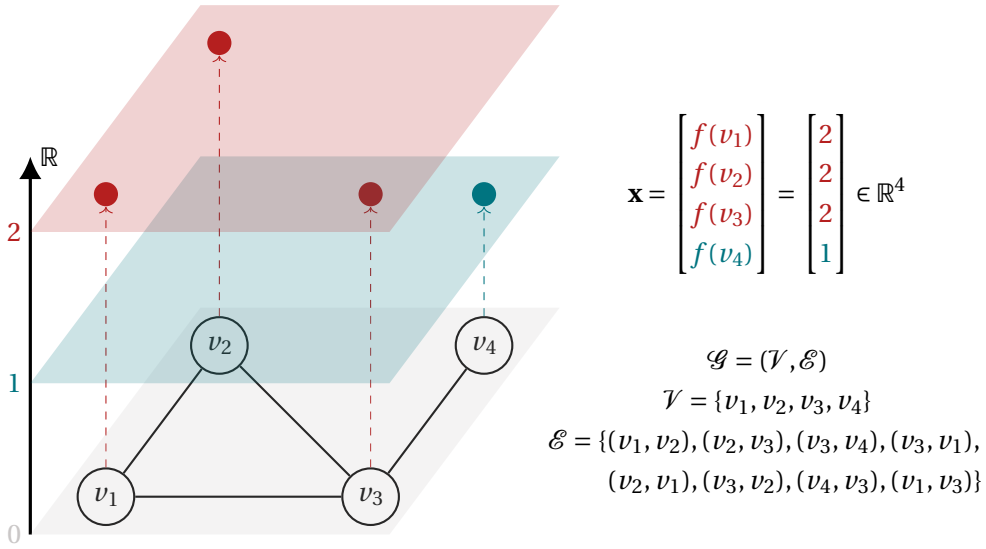


Figure 2.1 – A signal supported on a graph. The graph is a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of vertices and edges, and the signal is a function mapping vertices to the reals (in this case). The vector representation $\mathbf{x}$ of the graph signal gathers the function values at each vertex, once an (arbitrary) vertex ordering has been fixed.

2.1.1 Piecewise-constant graph signals

In Figure 2.1 is what we would intuitively call a piecewise-constant signal: it assigns a constant value of 2 for vertices v_1, v_2, v_3 and another constant value of 1 to vertex v_4. We could even think that this signal encodes the community structure of this graph, with v_1, v_2, v_3 in one community and v_4 in another.

To properly define a signal that is constant by pieces, it helps to explore a notion of "step-functions" on graphs, similarly to how Figure 2.1 distinguishes between a "red" step and a "blue" step. The indicator vector of a subset $\mathcal{W} \subset \mathcal{V}$ is the object with entries

$$\mathbb{1}_{\{\mathcal{W}\}}[i] = \begin{cases} 1 & \text{if } v_i \in \mathcal{W} \\ 0 & \text{otherwise} \end{cases}, \quad \forall i \in [n]. \tag{2.1}$$

The blue component of the signal in Figure 2.1 is the indicator vector of the vertex set $\{v_4\}$, while the red component is twice the indicator vector of $\{v_1, v_2, v_3\}$. Should we then state that any graph signal $\mathbf{x} \in \mathbb{R}^n$ is piecewise-constant if it can be written as a linear combination of indicator vectors? No, because the set of indicators $\{\mathbb{1}_{\{i\}}\}_{i=1}^n \equiv \{\mathbf{e}_i\}_{i=1}^n$ forms the standard basis in $\mathbb{R}^n$, so *any* graph signal lies in its linear span. To earn the adjective of piecewise-constant, a signal must also have relatively few variations, which we can quantify with help from difference operators.

2.1.2 Difference operators on graphs

The weights in the adjacency matrix $\mathbf{W}$ are supposed to represent how "close" vertices are to one another. Hence, it is natural to consider gradient maps $\nabla_{\mathcal{G}} f : \mathcal{E} \to \mathbb{R}^N$ of graph signals $f : \mathcal{V} \to \mathbb{R}^n$ as producing edge differences [1]

$$(\nabla_{\mathcal{G}} f)[e] = \sqrt{\mathbf{W}_{j,i}}\big(f(v_j) - f(v_i)\big), \quad \forall e = (v_i, v_j) \in \mathcal{E}. \tag{2.2}$$

This gradient assigns each $e = (v_i, v_j)$ to a real number quantifying the variation of the signal f from vertex v_i to vertex v_j. If $f(v_j)$ is larger, this variation is positive, indicating that the signal *increases* when going from v_i to v_j.

As we have done for vertices, fix now an (arbitrary) ordering $e_1 < e_2 < \cdots < e_N$ of the edges in $\mathcal{E}$. The gradient map can then be encoded in a weighted, signed, matrix $\mathbf{D} \in \mathbb{R}^{N \times n}$ with entries [2]

$$\mathbf{D}_{kl} = \begin{cases} \sqrt{\mathbf{W}_{ji}} & \text{if } l = i \\ -\sqrt{\mathbf{W}_{ij}} & \text{if } l = j \\ 0 & \text{otherwise} \end{cases}, \quad \forall k \in [N], l \in [n], \text{ with } e_k = (v_i, v_j). \tag{2.3}$$

If $\mathbf{x} = (f(v_i))_{i=1}^n$ is the vector representation of a graph signal f, then the gradient map $\nabla_{\mathcal{G}} f$ can be just as well expressed by the matrix-vector multiplication $\mathbf{Dx}$. For this reason, I will refer to $\mathbf{D}$ as the gradient (or difference) operator associated with graph $\mathcal{G}$.

Going back to "step functions", one can verify that $\mathbf{D}\mathbb{1}_{\mathcal{W}}$ is supported [3] on the boundary $\partial\mathcal{W} \subset \mathcal{E}$ corresponding to the edges between the vertex set $\mathcal{W}$ and its complement $\mathcal{V} \setminus \mathcal{W}$. In graph theory, $\partial\mathcal{W}$ is also known as the cut-set determined by the partition $(\mathcal{W}, \mathcal{V} \setminus \mathcal{W})$ of the vertex set.

We can use any ℓ_p-norm to define a measure of "size" of a cut-set in a way that accounts for

[1] The square root in the expression is standard [66] and has to do with obtaining a clean expression for a related difference operator, the graph Laplacian.

[2] For readers familiar with graph theory, this is a version of the transpose incidence matrix.

[3] The word "support" here refers to the edges corresponding to the non-zero entries of vector $\mathbf{D}\mathbb{1}_{\mathcal{W}}$.

the edge weights:

$$\begin{aligned}\left\|\mathbf{D}\mathbb{1}_{\{\mathscr{W}\}}\right\|_p^p &= \sum_{e=(v_i,v_j)\in\mathscr{E}} \left(\mathbf{W}_{ji}\right)^{p/2} \left|\mathbb{1}_{\{\mathscr{W}\}}[j] - \mathbb{1}_{\{\mathscr{W}\}}[i]\right|^p \\ &= \sum_{e=(i,j)\in\partial\mathscr{W}} \left(\mathbf{W}_{ji}\right)^{p/2}.\end{aligned}$$

The choice of p only influences — a priori — the importance given to the edge weights for the size computation.

More generally, a linear combination of indicator vectors induces several partial cut-sets, one for each constant piece. I call the union of these partial cut-sets the *jump-set*, indicating across which edges the signal changes value. For piecewise-constant graph signals, we should expect the size of the jump set to be small with respect to the total number N of edges in the graph. The jump-set can be identified with an index set in $[N]$ via the action of $\mathbf{D}$.

Definition 2.1.1 (Jump-set)**.** The jump-set of a graph signal $\mathbf{x}\in\mathbb{R}^n$ is the set

$$\mathscr{S} := \operatorname{supp}(\mathbf{Dx}) \tag{2.4}$$

containing the indices of the non-zero entries of the weighted-edge-differences vector $\mathbf{Dx}$.

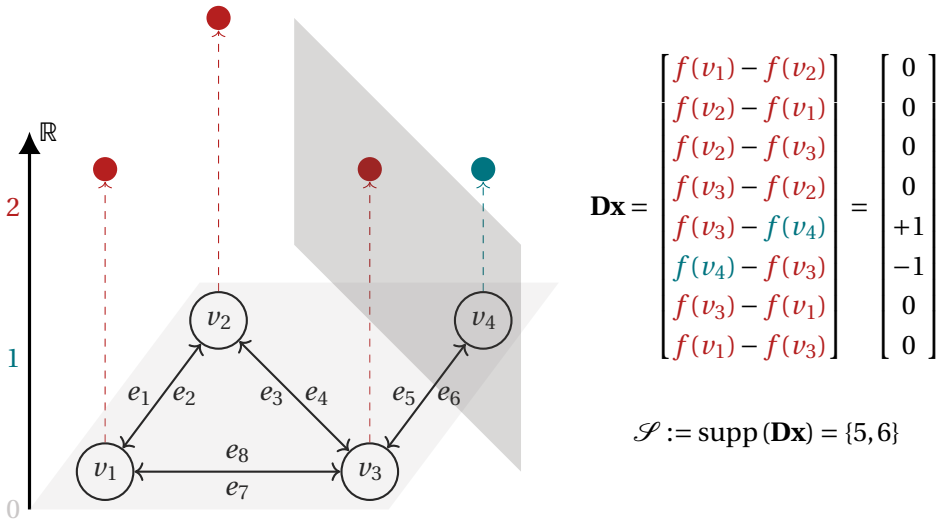


Figure 2.2 – Difference operation on the graph signal from Figure 2.1, assuming unit-weight connections. The piecewise-constant signal varies only across two of the eight directed edges in the graph. The jump-set $\mathscr{S}$ indexes the twin edges e_5 and e_6, whose cut splits the graph into "red" and "blue" communities.

We can also measure the size of the jump-set using ℓ_p norms of $\mathbf{Dx}$. Let $\mathbf{x} = \sum_{l=1}^{L} \alpha_l \mathbb{1}_{\{\mathscr{W}_l\}}$ be a signal with L constant pieces, taking values $\{\alpha_l\}_{l=1}^{L}$ on corresponding disjoint vertex subsets

$\mathcal{W}_1, \ldots, \mathcal{W}_L$. The functional $\|\mathbf{Dx}\|_p^p$ decomposes into L terms, one for each constant piece:

$$\begin{aligned}\|\mathbf{Dx}\|_p^p &= \sum_{e=(v_i,v_j)\in\mathcal{E}} \left(\mathbf{W}_{ji}\right)^{p/2} \left|\mathbf{x}[j]-\mathbf{x}[i]\right|^p \\ &= \sum_{l=1}^{L} \sum_{v_i\in\mathcal{W}_l} \sum_{v_j\in\mathcal{W}_k\neq\mathcal{W}_l} \left(\mathbf{W}_{ji}\right)^{p/2} |\alpha_l-\alpha_k|^p .\end{aligned} \tag{2.5}$$

I give special names to the functionals corresponding to two particular choices of p.

Definition 2.1.2 (Graph Total Variation ($\mathcal{G}$-TV))**.** The Graph Total Variation of a signal $\mathbf{x}\in\mathbb{R}^n$ is given by

$$\|\mathbf{Dx}\|_1 = \sum_{i\in[n]} \sum_{j\in[n]} \sqrt{\mathbf{W}_{ji}} \left|\mathbf{x}[j]-\mathbf{x}[i]\right| \tag{2.6}$$

Definition 2.1.3 (Dirichlet form)**.** The Dirichlet form of a signal $\mathbf{x}\in\mathbb{R}^n$ is given by

$$\|\mathbf{Dx}\|_2^2 = \sum_{i\in[n]} \sum_{j\in[n]} \mathbf{W}_{ji} \left(\mathbf{x}[j]-\mathbf{x}[i]\right)^2 \tag{2.7}$$

Graph Total Variation is a constant presence in this thesis, whereas the Dirichlet form appears in chapters 3 and 6 as a comparison point for results concerning the $\mathcal{G}$-TV semi-norm.

The null space of $\mathbf{D}$ is a particularly important object in the analysis of the $\mathcal{G}$-TV decoders introduced in the next chapter. As usual for difference operators, null($\mathbf{D}$) is non-trivial, that is, it contains a subspace dimension at least one. If the associated graph is *connected*, this subspace is span($\mathbf{1}$), the set of *constant* vectors in $\mathbb{R}^n$. If, however, the graph has *disconnected* parts, then null($\mathbf{D}$) will also contain the vectors that are constant on each of the connected sub-graphs. In the limiting case of a graph without edges, null($\mathbf{D}$) $\equiv \mathbb{R}^n$.

2.2 Sampling

To sample is to request the values $f(v)$ for every vertex v in some query set $\mathcal{Q}\subset\mathcal{V}$. In the vector interpretation of graph signals, this process is the same as measuring a coordinate subset of a point in $\mathbb{R}^n$. To sense coordinates, consider the standard basis $\{\mathbf{e}_i\}_{i=1}^n$ of $\mathbb{R}^n$, where each vector $\mathbf{e}_i$ contains a one at the i^{th} coordinate and zeros otherwise. Using this basis, a coordinate sample $y_i\in\mathbb{R}$ of a graph signal $\mathbf{x}=(f(v_i))_{i=1}^n$ at vertex v_i is nothing but the inner product

$$y_i = \langle \mathbf{e}_i, \mathbf{x}\rangle = x_i.$$

More generally, given a sampling set $\Omega \in [n]$ of cardinality $|\Omega| = m$, we can form a matrix

$$\mathbf{A} := \left(\mathbf{e}_i^\top\right)_{i\in\Omega} \in \mathbb{R}^{m\times n} \tag{2.8}$$

and define a sampling vector $\mathbf{y} = (y_i)_{i=1}^n$ via the linear operation $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Call $\mathbf{A} = \mathbf{A}(\Omega)$ the measurement matrix associated with the coordinate sampling set Ω. Sometimes it will be convenient to "lift" the co-domain of $\mathbf{A}$ to $\mathbb{R}^n$, by inserting zero-valued rows for the coordinates outside the sampling set. The resulting square matrix is $\sum_{i\in\Omega} \mathbf{e}_i\mathbf{e}_i^\top =: \mathbf{P}_\Omega$, the orthogonal projection operator onto the sampling set. With that in mind, I will abuse notation every once in a while and write $\mathbf{A} = \mathbf{P}_\Omega$ as a shorthand.

A *sampling design* is a blueprint for choosing Ω, even if implicitly defined. In general, designs can be either deterministic or probabilistic, but we will consider only the latter [4]. I assign a numeric template to sampling design in the form of a vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$. This vector assigns sampling probabilities to each element of $[n]$, but not necessarily their *inclusion probabilities* into Ω. Nonetheless, a larger π_i implies a more likely sample of vertex v_i. We will assume that each coordinate is sampled *independently* from the others, a convenient constraint for the probabilistic estimates in chapters 4 and 5.

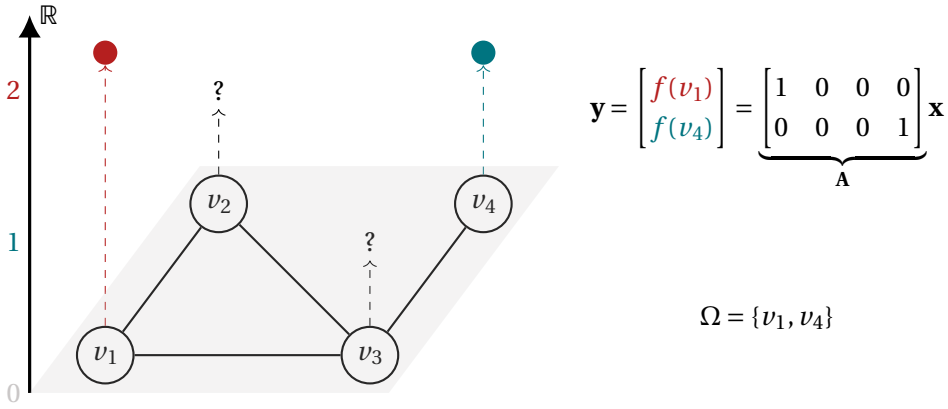


Figure 2.3 – Sampling the graph signal of Figure 2.1. A vertex is more likely to be in the sampling set Ω if its corresponding entry in $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)$ is large. Vector $\mathbf{y}$ gathers the measured (sampled) signal values, a linear operation over the ground-truth $\mathbf{x}$.

There exist parametric possibilities for modeling the probabilities in $\boldsymbol{\pi}$. For an example, Jung [34] takes each π_i from an exponential family, parametrized by the signal value x_i at vertex v_i. We will proceed otherwise, seeing the entries of $\boldsymbol{\pi}$ merely as unspecified real numbers in the interval $[0, 1]$.

[4] Technically, it is possible to produce deterministic sampling designs probabilistic one by setting probability masses to either zero or one.

2.2.1 Bernoulli Sampling Model (Ber(π))

Define random selectors $\{\delta_i\}_{i=1}^n$ by drawing n independent, $\{0,1\}$-valued Bernoulli random variables, each of which according to the probabilities

$$\mathbb{P}(\{\delta_i = 1\}) = \pi_i = 1 - \mathbb{P}(\{\delta_i = 0\}), \forall i \in [n]. \tag{2.9}$$

These selectors induce a sampling set according to the rule $\Omega = \{i \in [n] : \delta_i = 1\}$. Note that each vertex $i \in [n]$ is *included* in the sampling set with probability

$$\mathbb{P}(\{i \in \Omega\}) = \pi_i.$$

An advantage of using Bernoulli selectors that the independent coordinate samples *do not repeat*. There is no redundance in the sampling set, in the sense that the number of unique coordinate observations is equal to the cardinality of Ω. The downside is that this cardinality is itself a *random variable*. Indeed, $|\Omega| = \sum_{i=1}^n \delta_i$, so the number of samples is not determined a priori. Fortunately, the distribution $|\Omega|$ is fairly well concentrated around its expectation $\overline{m} := \mathbb{E}(|\Omega|) = \sum_{i=1}^n \pi_i$. An application of the scalar Bernstein inequality (Lemma A.0.1) yields

$$\frac{\overline{m}}{2} \leq |\Omega| \leq \frac{3\overline{m}}{2}, \tag{2.10}$$

with probability at least $1 - 2\exp\left(-\overline{m}/9\right)$.

The (lifted) measurement matrix induced from the Bernoulli model can be written directly in terms of the selectors, bypassing the sampling set Ω:

$$\mathbf{A} = \sum_{i \in \Omega} \mathbf{e}_i^\top \mathbf{e}_i^\top = \sum_{i=1}^n \delta_i \mathbf{e}_i^\top \mathbf{e}_i^\top, \tag{2.11}$$

this expression will be very useful in Chapter 4.

2.2.2 Coordinate Sampling with Replacement (CSWR(π))

Let ω be a random variable taking values in $[n]$ with probabilities $\mathbb{P}(\{\omega = i\}) = \pi_i$, for each $i \in [n]$. Draw m *i.i.d.* copies, $\omega_1, \omega_2, \ldots, \omega_m$, of ω and define the sampling *multiset* $\Omega = \{\omega_i\}_{i=1}^m$ containing all those copies, including repetitions. This process effectively implements independent coordinate sampling with replacement. In particular, setting $\pi = (1/n, \ldots, 1/n)$, one retrieves uniform random sampling.

Compared to the Bernoulli model, the likelihood that any given vertex $i \in [n]$ is in the sampling

set has a more complicated expression:

$$\begin{aligned}\mathbb{P}(\{i \in \Omega\}) &= \mathbb{P}(\{\omega_1 = i \vee \cdots \vee \omega_m = i\}) \\ &= 1 - \mathbb{P}(\{\omega_1 \neq i \wedge \cdots \wedge \omega_m \neq i\}) \\ &= 1 - \mathbb{P}(\{\omega \neq i\})^m \quad (i.i.d.) \\ &= 1 - (1 - \pi_i)^m.\end{aligned}$$

However, the total number of measurements is now deterministic and equal to m. This property proves to be convenient in Chapter 5.

2.2.3 Reconstruction from samples

Behind every sampling procedure, there is an underlying signal $\mathbf{x}$ which is the real object of interest. Sometimes the measurements only need to be numerous enough to estimate the mean, variance or other simple statistics of $\mathbf{x}$. We, however, want to recover the *full* signal, in the spirit of Compressed Sensing (CS).

It should be expected that our ability to reconstruct a sub-sampled signal depends on how many measurements we have taken. If we sample all of the n coordinates of $\mathbf{x}$ then the problem is trivially solved; if we sample none, there is no hope for recovery. The interesting cases are somewhere in between, especially when the number of measurements m is much smaller than the dimension n of the signal, a setup we informally refer to as $m \ll n$. In terms of linear algebra, recovery of $\mathbf{x}$ from $\mathbf{y} = \mathbf{A}\mathbf{x}$ is an attempt to invert a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ that is rank-deficient. This problem has infinitely many solutions and is therefore ill-posed. In Chapter 3, I present decoders $\mathfrak{D} : \mathbb{R}^m \to \mathbb{R}^n$, based on convex optimization, that remedy this situation. They revert the measurement process, yielding $\mathbf{x} = \mathfrak{D}(\mathbf{y})$, as long as $\mathbf{x}$ belongs to a restricted class of vectors within $\mathbb{R}^n$.

I have previously postulated that the optimal sampling design for a fixed decoder $\mathfrak{D}$ will minimize the number of measurements needed for a successful recovery. Intuitively, such a design should sample more often the coordinates that contribute the most to the recovery confidence, while neglecting those that do not add as much value. In terms of the probabilities in $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$, each entry π_i can be thus seen as an importance measure of coordinate i from the perspective of the decoder. These guidelines become formal once we can precise how each of the sampling probabilities in $\boldsymbol{\pi}$ affects the performance of the decoder.

2.2.4 "Equivalence" between the sampling models

Using the Ber($\boldsymbol{\pi}$) or the CSWR($\boldsymbol{\pi}$) model in our context is essentially a matter of convenience. Indeed, we can borrow an argument by Candès *et al.* [12, Appendix], which I have detailed in Appendix 2.A. The reasoning holds in the context of any decoder whose success probability, $\mathbb{P}(\{\text{Success}\})$, is monotonically increasing with the sample size m. This is the case for the

procedures commonly used in Compressed Sensing.

To begin, $\mathbb{P}(\{\text{Success}\})$ when sampling *without* replacement can only be larger than the corresponding success probability when sampling *with* replacement when we have the same number of measurements in both settings. The Ber($\boldsymbol{\pi}$) model is without replacement, but produces a *random* number of measurements. The tails of this distribution allow the redundant CSWR($\boldsymbol{\pi}$) model to "catch-up", as long as the number of redundant samples in the latter is slightly larger than the average produced by Ber($\boldsymbol{\pi}$). Ultimately the two models, CSWR($\boldsymbol{\pi}$) and Ber($\boldsymbol{\pi}$), need a similar number of measurements to reach the same recovery success probability.

A related discussion comparing the use of sampling with or without replacement can be found in Gross and Nesme [29], grounded on a classical moment domination result by Hoeffding [30].

2.3 Summary

We can reason about signal processing on graphs by defining appropriate vector spaces and linear operators therein that depend on the connections between vertices. Fundamentally, signal variations across edges can be measured a map $\mathbf{x} \mapsto \mathbf{D}\mathbf{x}$. Piecewise-constant graph signals are the ones that have few edge variations, or a "small" jump-set. The size of the jump-set can be measured by ℓ_p semi-norms induced by the graph gradient operator, allowing us to assess the degree to which a signal can be deemed piecewise-constant. Table 2.1 is a record of important Graph Signal Processing objects that are often referenced in the following chapters.

Concept	Notes
Number of vertices	n
Number of edges	N
Jump-set of a signal $\mathbf{x}$	$\mathscr{S} := \operatorname{supp}(\mathbf{D}\mathbf{x})$
null ($\mathbf{D}$)	Contains at least span($\mathbf{1}$)
Graph Total Variation ($\mathscr{G}$-TV)	$\|\mathbf{D}\mathbf{x}\|_1 = \sum_{i,j} \sqrt{\mathbf{W}_{ji}} \left\|\mathbf{x}[j] - \mathbf{x}[i]\right\|$
Dirichlet form	$\|\mathbf{D}\mathbf{x}\|_2^2 = \sum_{i,j} \mathbf{W}_{ji} \left(\mathbf{x}[j] - \mathbf{x}[i]\right)^2$

Table 2.1 – Summary of objects related to the graph gradient operator $\mathbf{D} \in \mathbb{R}^{N \times n}$.

Vertex sampling is also a linear operation, $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$, where the rows of $\mathbf{A}$ are taken from the standard basis of $\mathbb{R}^n$, indexed by a sampling set $\Omega \subset [n]$. A sampling design is a choice of a vector $\boldsymbol{\pi} = (\pi_i, \ldots, \pi_n)$ that determines the inclusion probabilities of elements of $[n]$ into Ω.

I introduced two models for independent random sampling, named Ber($\boldsymbol{\pi}$) and CSWR($\boldsymbol{\pi}$), that are almost equivalent in the context of recovery problems. Their use in chapters 4 and 5, respectively, is dictated by convenience. Table 2.2 is a reference point for some properties of these two sampling models.

Model	Rows of $\mathbf{A}$	Repeated samples	Number of samples ($\lvert\Omega\rvert$)
Ber($\boldsymbol{\pi}$)	independent	no	$m \in \left[\frac{1}{2}\boldsymbol{\pi}^\top\mathbf{1}, \frac{3}{2}\boldsymbol{\pi}^\top\mathbf{1}\right]$ w.h.p.
CSWR($\boldsymbol{\pi}$)	independent	yes	m deterministic

Table 2.2 – Summary of sampling models used to construct the measurement matrix $\mathbf{A}$ from a random (multi-)set of integers $\Omega \subset [n]$.

Appendix 2.A "Equivalence" of sampling models

This appendix adapts reasoning found in [12, Appendix], arguing for a certain equivalence between the sampling models introduced in Section 2.2. For the purposes of the sort of decoders used in this thesis, a recovery guarantee obtained from one sampling model automatically implies a similar recovery guarantee for the other.

Assume we have an algorithm that takes as input samples of some n-dimensional vector and outputs a vector in $\mathbb{R}^n$. Denote by $\texttt{Success}$ the event whereby this algorithm produces the correct output. The notion of success can be arbitrary, as long as its probability never decreases as the number of samples increases. In other words, if the samples are indexed by a set Ω, the quantity $\mathbb{P}(\{\texttt{Success}\})$ is monotonically increasing with increasing $|\Omega|$.

Now, denote by $\mathbb{P}_{\mathrm{Ber}(\boldsymbol{\pi})}(\{\texttt{Success}\})$ and $\mathbb{P}_{\mathrm{CSWR}(\boldsymbol{\pi})}(\{\texttt{Success}\})$ the probabilities of success induced by the distributions of the Bernoulli Sampling Model ($\mathrm{Ber}(\boldsymbol{\pi})$) and the Coordinate Sampling with Replacement ($\mathrm{CSWR}(\boldsymbol{\pi})$), respectively. Recall that $\mathrm{CSWR}(\boldsymbol{\pi})$ always draws m samples, while the number of samples produced by $\mathrm{Ber}(\boldsymbol{\pi})$ is random, with mean denoted by $\overline{m} = \mathbb{E}(|\Omega|)$. I will put the number of samples in evidence by referring to the Bernoulli Sampling Model as $\mathrm{Ber}(\overline{m})$ and to the Coordinate Sampling with Replacement as $\mathrm{CSWR}(m)$.

Lemma 2.A.1. *Let $\overline{m} = \frac{1}{1+\varepsilon} m$, for any $\varepsilon > 0$. Then, a success probability of P under CSWR(m) implies a success probability of $P\left[1 - 2\exp\left(\frac{-3\varepsilon\overline{m}}{8}\right)\right]$ under Ber($\overline{m}$).*

PROOF:

⟨1⟩1. A direct calculation reveals

$$\begin{aligned}
\mathbb{P}_{\mathrm{Ber}(\overline{m})}(\{\texttt{Success}\}) &\geq \sum_{k=m}^{n} \mathbb{P}_{\mathrm{Ber}(\overline{m})}(\{\texttt{Success}||\Omega| = k\})\, \mathbb{P}_{\mathrm{Ber}(\overline{m})}(\{|\Omega| = k\}) \\
&\geq \mathbb{P}_{\mathrm{CSWR}(m)}(\{\texttt{Success}\}) \sum_{k=m}^{n} \mathbb{P}_{\mathrm{Ber}(\overline{m})}(\{|\Omega| = k\}) \\
&= \mathbb{P}_{\mathrm{CSWR}(m)}(\{\texttt{Success}\})\, \mathbb{P}_{\mathrm{Ber}(\overline{m})}(\{|\Omega| \geq m\}) \\
&= \mathbb{P}_{\mathrm{CSWR}(m)}(\{\texttt{Success}\}) \left[1 - \mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\left\{|\Omega| < c\,\overline{m}\right\}\right)\right] \\
&\geq \mathbb{P}_{\mathrm{CSWR}(m)}(\{\texttt{Success}\}) \left[1 - 2\exp\left(-3c\,\overline{m}/8\right)\right]
\end{aligned}$$

PROOF:

⟨2⟩1. Conditioned on its cardinality, the distribution of Ω under $\mathrm{Ber}(\overline{m})$ is equivalent to coordinate sampling *without* replacement. A sample without replacement always implies more distinct elements in Ω than a sample with replacement (for the same number of measurements). Thus, for any $k \geq m$,

$$\mathbb{P}_{\mathrm{Ber}(\overline{m})}(\{\texttt{Success}||\Omega = k|\}) \geq \mathbb{P}_{\mathrm{CSWR}(m)}(\{\texttt{Success}\}).$$

⟨2⟩2. The mononicity assumption also implies, for every $k \leq m$,

$$\mathbb{P}_{\mathrm{CSWR}(k)}(\{\texttt{Success}\}) \leq \mathbb{P}_{\mathrm{CSWR}(m)}(\{\texttt{Success}\}).$$

⟨2⟩3. Finally, the scalar Bernstein inequality (Lemma A.0.1) uncovers the tail bound

$$\mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{|\Omega| < m\}\right) = \mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\left\{|\Omega| < (1+\varepsilon)\,\overline{m}\right\}\right) \leq 2\exp\left(-3\varepsilon\,\overline{m}/8\right).$$

⟨1⟩2. Q.E.D.

Lemma 2.A.2. *There exists some $\varepsilon_1 \in (0,1)$ such that if $m = \frac{1+\varepsilon}{1-\varepsilon_1}\,\overline{m}$, then a success probability of P under Ber($\overline{m}$) implies a corresponding success probability of $P - 2\exp\left(\frac{-3\varepsilon_2\,\overline{m}}{8}\right)$ under CSWR(m), for any $\varepsilon_2 > 0$.*

PROOF:
⟨1⟩1. There exists some ε_1 for which

$$\begin{aligned}
\mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{\texttt{Success}\}\right) = &\sum_{k=1}^{(1-\varepsilon_1)m-1} \mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{\texttt{Success}||\Omega| = k\}\right)\mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{|\Omega| = k\}\right) \\
&+ \sum_{k=(1-\varepsilon_1)m}^{n} \mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{\texttt{Success}||\Omega| = k\}\right)\mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{|\Omega| = k\}\right) \\
&\leq \mathbb{P}_{\mathrm{CSWR}(m)}\left(\{\texttt{Success}\}\right) + \mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{|\Omega| \geq (1-\varepsilon_1)m\}\right) \\
&\leq \mathbb{P}_{\mathrm{CSWR}(m)}\left(\{\texttt{Success}\}\right) + 2\exp\left(\frac{-3\varepsilon_2\,\overline{m}}{8}\right)
\end{aligned}$$

PROOF:
⟨2⟩1. Conditioned on its cardinality, the distribution of Ω under Ber($\overline{m}$) is equivalent to coordinate sampling *without* replacement. For the same number of measurements, a sample without replacement always implies more distinct elements in Ω than a sample with replacement with the same number of measurements. However, there exists some ε such that the number of distinct elements in a sample without replacement of size $(1-\varepsilon)m$ is smaller than the number of distinct elements in a sample with replacement of size m. Take ε_1 to be the infimum among such ε, and the monotonicity in the success likelihood for the algorithm will imply

$$\mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{\texttt{Success}||\Omega = k|\}\right) \leq \mathbb{P}_{\mathrm{CSWR}(m)}\left(\{\texttt{Success}\}\right),$$

for any $k \leq (1-\varepsilon_1)m$.
⟨2⟩2. Once again, the tail bound is given by the scalar Bernstein inequality (Lemma A.0.1) uncovers

$$\mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\{|\Omega| > (1-\varepsilon_1)m\}\right) = \mathbb{P}_{\mathrm{Ber}(\overline{m})}\left(\left\{|\Omega| > (1+\varepsilon_2)\,\overline{m}\right\}\right) \leq 2\exp\left(-3\varepsilon_2\,\overline{m}/8\right).$$

⟨1⟩2. Q.E.D.

3 Recovery via convex programs

We left Chapter 2 with an *ill-posed* problem: retrieve $\mathbf{x}$ by observing $\mathbf{y} := \mathbf{A}\mathbf{x}$. Due to unsampled vertices, the measurement matrix $\mathbf{A}$ is rank-deficient, so it cannot be inverted. In other words, there is no way to distinguish, a priori, a signal $\mathbf{x}$ from all the points in the set $\{\mathbf{z} \in \mathbb{R}^n : \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}\}$.

The culprit for ill-posedness is the size of $\mathbb{R}^n$, the default search space. Linear algebra tells us that $\{\mathbf{z} \in \mathbb{R}^n : \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}\}$ has infinitely many points whenever $\mathbf{A}$ is non-invertible. But in this thesis we only care about *piecewise-constant* signals, so we should not have to look for answers in the whole of $\mathbb{R}^n$. We should aim to find a smaller set $\mathcal{Z} \subset \mathbb{R}^n$ — containing piecewise-constant signals — for which the search space $\{\mathbf{z} \in \mathcal{Z} : \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}\}$ reduces to the singleton $\{\mathbf{x}\}$. That is, restricted to $\mathcal{Z}$, the only vector with samples $\mathbf{y} = \mathbf{A}\mathbf{x}$ is $\mathbf{x}$ itself. There would be no loss of information by representing $\mathbf{x}$ in the compressed form $\mathbf{y}$.

In modern signal processing, the variational principle is a popular way to constrain the search space for inverse problems. First, one picks a function $f : \mathbb{R}^n \to \mathbb{R}$ that quantifies a key property of the signal $\mathbf{x}$ to be recovered. The function here is seen as a "complexity cost", taking small values for signals that look like $\mathbf{x}$ and large values otherwise. Then, one simply looks among the vectors agreeing with the measurements for the one that *minimizes* f.

$$\min_{\mathbf{z} \in \mathbb{R}^n} f(\mathbf{z}) \text{ subject to } \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}. \tag{Pf}$$

We will call optimization programs like (Pf) *interpolation* problems, because the "penalty" f informs how to fill in the missing data, without changing the values of the sampled points.

Interpolation is fine if the measurements are noiseless. Noise demands that we adapt (Pf), but a simple tweak is usually enough. Let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ be the noisy samples and assume that we have (for some $q \geq 1$ and $\eta \geq 0$) the upper bound $\|\mathbf{e}\|_q^q \leq \eta$ on the noise component. Replace the equality constraint in (Pf) by $\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_q^q \leq \eta$, allowing the recovered values at the sampled points to differ from the measurements, but restricting the difference to the noise level. The

resulting *regression* program [1] reads

$$\min_{z\in\mathbb{R}^n} f(\mathbf{z}) \text{ subject to } \|\mathbf{Az}-\mathbf{y}\|_q^q \leq \eta. \tag{Pf-η}$$

Problem (Pf-η) actually generalizes (Pf): it suffices to take $\eta = 0$ to get the latter from the former. Why then bother to define the two versions? This mostly has to do with Chapter 5, whose arguments apply an interpolation problem only. As that chapter contains the main theoretical contributions in this thesis, the numerical experiments in Chapter 6 also consider only noiseless settings. But in Chapter 4 we can work directly with regression, which allows more general conclusions. In any case, the text will make clear whether a statement is valid for a problem like (Pf-η) or solely for (Pf).

Noise considerations aside, we still have to pick a function f adapted to the graph signals that we care for. I proposed in Chapter 2 that a piecewise-constant graph signal $\mathbf{x}$ is characterized by few jumps in value across the edges. Equivalently, the jump-set $\mathscr{S} := \text{supp}(\mathbf{Dx})$ has small cardinality, or is sparse, using the language of Compressed Sensing (CS) [2]. We can write the cardinality of $\mathscr{S}$ as $|\mathscr{S}| = \|\mathbf{Dx}\|_0$, so the most direct proposal for a cost function appropriate to piecewise-constant graph signals would be the map $\mathbf{z} \overset{f}{\mapsto} \|\mathbf{Dz}\|_0$.

However, the ℓ_0 "norm" is not convex, and it pays off to have a convex function in recovery programs. The reasons are both analytical and numerical. First, a convex f turns (Pf-η) and (Pf) into convex problems, where every local minimum is a global minimum. Convexity brings with it a range of theoretical tools to analyze the properties of the solutions to the optimization problems. Second, convex problems have numerical solvers with *convergence* guarantees. These solvers can approximate the global minimum with often relatively few iterations. Furthermore, the lack of sub-optimal local minima in convex problems ultimately decouples the numerical and analytical aspects of their solutions. In contrast, in non-convex problems such as the training of deep linear neural networks, the *trajectories* of the gradient descent solver inform the properties of the solutions found in practice [5]. With decoupled analytical and numerical aspects, we can study convex programs independently of their practical implementation.

In Compressed Sensing, the ℓ_1-norm is the standard "convexification" of the ℓ_0 "norm". This choice brings us to the Graph Total Variation ($\mathscr{G}$-TV) interpolation decoder

$$\min_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{Dz}\|_1 \text{ such that } \mathbf{Ax} = \mathbf{Az}. \tag{P1}$$

Naturally, (P1) also has its noisy alternative,

$$\min_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{Dz}\|_1 \text{ subject to } \|\mathbf{Az}-\mathbf{y}\|_q^q \leq \eta. \tag{P1-η}$$

[1] I name it regression in contrast with the interpolation version, because the penalty f potentially denoises the sampled values, in addition to filling in the missing data.

[2] Some CS researchers would also say that $\mathbf{x}$ is *co-sparse* under the action of $\mathbf{D}$ [54].

which I call the $\mathscr{G}$-TV regression problem. Most of this thesis is dedicated to dwelling on these two especial decoders.

But before that, the next section discusses some base conditions for unique and exact solution in the *general* interpolation program (Pf). In the process, I introduce the descent cone and the subdifferential of $\|\mathbf{D}\cdot\|_1$, sets that play a fundamental role towards linking the number of vertex samples with the success of $\mathscr{G}$-TV minimization.

Following this, I argue why the $\mathscr{G}$-TV semi-norm is a good convex surrogate to $\|\mathbf{D}\cdot\|_0$ as a signature for piecewise-constant graph signals. The reasoning is geometric, relying on the atomic status of $\|\mathbf{D}\cdot\|_1$. I also include a comparison with the Dirichlet form $\|\mathbf{D}\cdot\|_2^2$, using representer theorems to show that $\mathscr{G}$-TV minimization is less dependent on the *form* of the measurement matrix $\mathbf{A}$.

This chapter ends by connecting our recovery programs with the wider Compressed Sensing literature. Indeed, we can see in $\mathscr{G}$-TV minimization an instance of general co-sparse (or analysis-sparse) programs. What is exceptional in our setting is the sampling procedure. In CS, Gaussian-like measurement vectors are the norm; comparatively little can be found on coordinate-sampled, co-sparse models. Much of the difficulty in the subsequent chapters 4 and 5 stems from the lack of tools to deal with our measurement matrix.

3.1 When is the solution of convex interpolation unique?

Due to the minimization principle, the only obstacles to $\mathbf{x}$ being the sole solution of (Pf) are the vectors $\mathbf{z}\in\mathbb{R}^n$ for which $f(\mathbf{z})\le f(\mathbf{x})$. To investigate the impact of these vectors, we can fix ourselves on $\mathbf{x}$ and check how much of $\mathbb{R}^n$ we cover by only moving in the directions that decrease f. The conic hull of these descent directions is called the descent cone.

Definition 3.1.1 (Descent cone)**.** Let $\mathbf{x}$ be a fixed point in $\mathbb{R}^n$. The descent cone of a convex function $f:\mathbb{R}^n\to\mathbb{R}$ at $\mathbf{x}$ is the set

$$\mathscr{D}(f,\mathbf{x}) := \operatorname{cone}\left(\left\{\mathbf{u}\in\mathbb{R}^n : f(\mathbf{x}+\mathbf{u})\le f(\mathbf{x})\right\}\right) \tag{3.1}$$

$$=: \bigcup_{\tau\ge 0}\left\{\tau\mathbf{u}\in\mathbb{R}^n : f(\mathbf{x}+\mathbf{u})\le f(\mathbf{x})\right\}. \tag{3.2}$$

Unique recovery in (Pf) then turns out to be a question about how the descent cone $\mathscr{D}(f,\mathbf{x})$ intersects the null space of the measurement matrix $\mathbf{A}$. The precise statement is given in Theorem 3.1.1.

Theorem 3.1.1 ([14, Prop. 2.1],[35, Thm. 3])**.** *Vector* $\mathbf{x}$ *is the unique solution of problem* (Pf) if and only if *the trivial intersection*[3] $\mathscr{D}(f,\mathbf{x})\cap\operatorname{null}(\mathbf{A})=\{\mathbf{0}\}$ *takes place.*

[3] Both $\mathscr{D}(f,\mathbf{x})$ and $\operatorname{null}(\mathbf{A})$ contain at least $\mathbf{0}$, so "trivial intersection" highlights the situation when these sets agree only at the origin.

The proof is standard but short, so I include it here.

PROOF:

⟨1⟩1. " $\Longleftarrow$ ": Assume that $\mathscr{D}(f,\mathbf{x}) \cap \text{null}(\mathbf{A}) = \{\mathbf{0}\}$, and let $\mathbf{u} \in \mathbb{R}^n$ be such that $f(\mathbf{x}+\mathbf{u}) \leq f(\mathbf{x})$. Vector $\mathbf{x}+\mathbf{u}$ is thus a solution of problem $(\mathrm{P}f)$ if $\mathbf{A}(\mathbf{x}+\mathbf{u}) = \mathbf{A}\mathbf{x}$. It follows that $\mathbf{x}+\mathbf{u} = \mathbf{x}$.

PROOF:

⟨2⟩1. $\mathbf{u} \in \mathscr{D}(f,\mathbf{x})$ by construction.

⟨2⟩2. $\mathbf{A}(\mathbf{x}+\mathbf{u}) = \mathbf{A}\mathbf{x} \implies \mathbf{u} \in \text{null}(\mathbf{A})$

⟨2⟩3. $\mathbf{u} \in \mathscr{D}(f,\mathbf{x}) \cap \text{null}(\mathbf{A}) \implies \mathbf{u} = 0$, by assumption.

□

⟨1⟩2. " $\Longrightarrow$ ": Assume that $\mathbf{x}$ is the unique solution of $(\mathrm{P}f)$, and pick $\mathbf{u} \in \mathbb{R}^n$ such that $\mathbf{A}(\mathbf{x}+\mathbf{u}) = \mathbf{A}\mathbf{x}$. Then $\mathbf{u} \notin \mathscr{D}(f,\mathbf{x})$, unless $\mathbf{u} = \mathbf{0}$.

PROOF:

⟨2⟩1. $\mathbf{u} \in \text{null}(\mathbf{A})$ by construction.

⟨2⟩2. If $\mathbf{u} \neq \mathbf{0}$, then $f(\mathbf{x}+\mathbf{u}) > f(\mathbf{x})$ because $\mathbf{x}$ is the unique feasible minimizer.

⟨2⟩3. $\mathbf{u} \in \mathscr{D}(f,\mathbf{x}) \iff \mathbf{u} \neq \mathbf{0}$

□

⟨1⟩3. Q.E.D.

$\mathbf{x}$ is the unique solution of $(\mathrm{P}f) \iff \mathscr{D}(f,\mathbf{x}) \cap \text{null}(\mathbf{A}) = \{\mathbf{0}\}$.

Theorem 3.1.1 is a geometric result, illustrated in Figure 3.1. The cone $\mathscr{D}(f,\mathbf{x})$ is a deterministic object, the fruit of the signal we want to recover and its implicit modeling through function f. The null space of $\mathbf{A}$ is a random subspace: its exact orientation depends on which vertices are sampled in the graph.

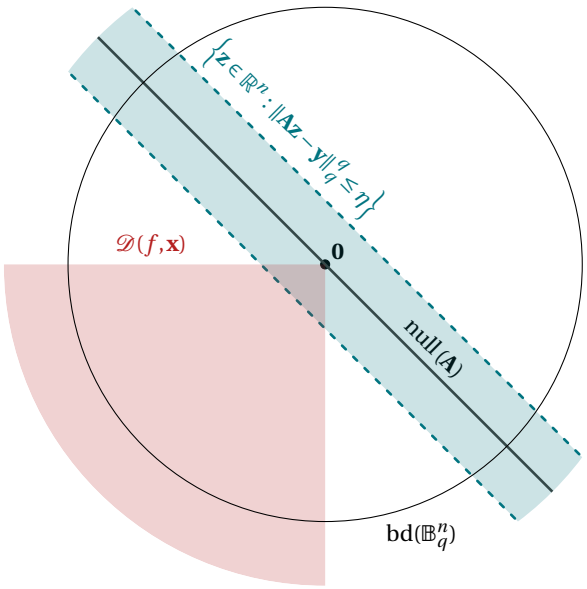


Figure 3.1 – Illustration of the trivial intersection property $\mathscr{D}(f,\mathbf{x}) \cap \text{null}(\mathbf{A}) = \{\mathbf{0}\}$.

Figure 3.1 points at an intuitive idea: the narrower the descent cone, the easier should be the for the trivial intersection to take place. The precise notion of width is discussed in Chapter 4, where we will also see that the same geometric idea can potentially be used to arrive at robust recovery guarantees for the *regression* problem (Pf-η). The reader may suspect that this is possible by examining how the blue slab $\left\{\mathbf{z}\in\mathbb{R}^n : \|\mathbf{A}\mathbf{z}-\mathbf{y}\|_q^q \leq \eta\right\}$ wraps around null$(\mathbf{A})$. Potential regression solutions can distance themselves from $\mathbf{x}$ at most by approximately the noise level.

In the language of information theory, the trivial intersection property is the same as lossless dimension reduction using matrix $\mathbf{A}$. There is no ambiguity in representing $\mathbf{x}$ by $\mathbf{A}\mathbf{x}$, provided we use (Pf) as decoder. Gaussian matrices are known for their dimension reduction properties when applied to finite [31] or sparse [22] sets of vectors. And more recently broader classes of sub-Gaussian matrices have also been shown to behave as such [56]. Can one say the same about our coordinate sampling matrix $\mathbf{A}$? This is essentially the question studied in Chapter 4.

Specializing to Graph Total Variation interpolation, we can also obtain a simpler, *necessary* condition for uniqueness. It involves the null space of $\mathbf{D}$ and is recorded in Proposition 3.1.1.

Proposition 3.1.1. *If* null$(\mathbf{D})\cap$ null$(\mathbf{A})\neq\{\mathbf{0}\}$*, then problem* (P1) *has infinitely many solutions*[4].

The subspace null$(\mathbf{D})$ belongs to $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$, so the trivial intersection property in Theorem 3.1.1 is strictly stronger. Nonetheless, Proposition 3.1.1 is a first glimpse on the compatibility required from the analysis and measurement operators if vector $\mathbf{x}$ is to be properly recovered.

PROOF: Assume that $\mathbf{z}^\star$ is in the solution set of (Pf-η) and let $\mathbf{h}$ be an arbitrary point in the subspace null$(\mathbf{D})\cap$ null$(\mathbf{A})\subset\mathbb{R}^n$. Note that $\mathbf{z}^\star+\mathbf{h}$ is feasible, as $\mathbf{A}(\mathbf{z}^\star+\mathbf{h})-\mathbf{A}\mathbf{x}=\mathbf{A}\mathbf{z}^\star-\mathbf{A}\mathbf{x}=\mathbf{0}$. Furthermore, $\|\mathbf{D}(\mathbf{z}^\star+\mathbf{h})\|_1=\|\mathbf{D}\mathbf{z}^\star\|_1$, so $\mathbf{z}^\star+\mathbf{h}$ is also a solution. Since $\mathbf{h}$ was arbitrary and the subspace to which it belongs is not zero-dimensional, the claim holds. □

Working with the descent cone is not the only way to derive uniqueness results for (Pf). Readers familiar with optimization problems in Calculus know that maxima and minima of a differentiable function can be found where the gradient vanishes. The analogous object in the variational analysis of non-differentiable functions is the subdifferential set.

Definition 3.1.2 (Subdifferential)**.** Let $\mathbf{x}$ be a point in $\mathbb{R}^n$. The subdifferential of a convex function $f:\mathbb{R}^n\to\mathbb{R}$ at $\mathbf{x}$ is the set

$$\partial f(\mathbf{x}) := \left\{\mathbf{u}\in\mathbb{R}^n : f(\mathbf{z})-f(\mathbf{x})\geq\langle\mathbf{u},\mathbf{z}-\mathbf{x}\rangle,\ \ \forall\mathbf{z}\in\mathbb{R}^n\right\}. \tag{3.3}$$

The subdifferential $\partial f(\mathbf{x})$ is *polar* to the closure of $\mathscr{D}(f,\mathbf{x})$ [62, Thm. 23.7], meaning that $\langle\mathbf{v},\mathbf{u}\rangle\leq 0$ whenever $\mathbf{u}\in\overline{\mathscr{D}(f,\mathbf{x})}$ and $\mathbf{v}\in\partial f(\mathbf{x})\}$. Intuitively, both sets share the origin, but "point" towards opposite directions. The trivial intersection property can be equivalently expressed

[4] Actually, readers can examine the proof to convince themselves that the conclusion is valid whenever the interpolation problem (Pf) employs a function of the form $f(\cdot)=g(\mathbf{D}\cdot)$, where $g:\mathbb{R}^N\to\mathbb{R}$ is convex.

in terms of the subdifferential under the guise of Karush-Kuhn-Tucker (KKT) conditions. I explore this optic in Chapter 5.

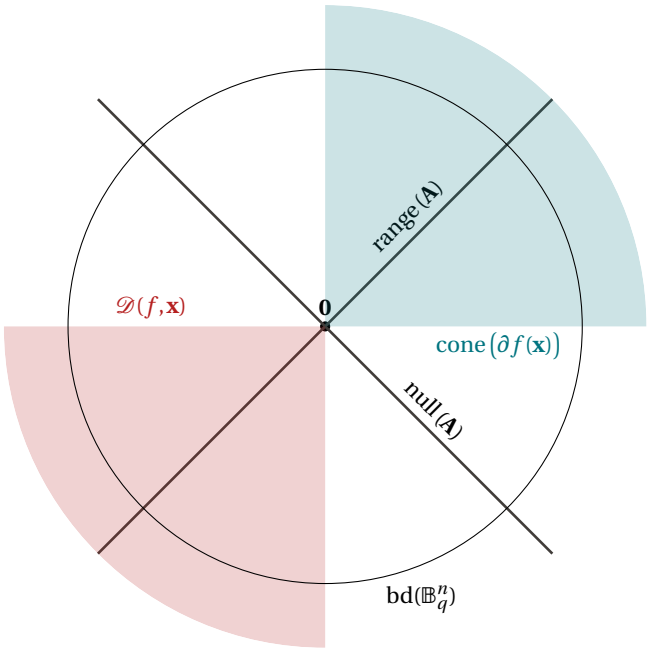


Figure 3.2 – Illustration of the polarity between the descent cone and the subdifferential of a convex function f. Exact recovery happens either if $\mathscr{D}(f,\mathbf{x})$ and null $(\mathbf{A})$ intersect trivially, or equivalently if $\partial f(\mathbf{x})$ and range $(\mathbf{A})$ intersect non-trivially.

The subdifferential of the $\mathscr{G}$-TV semi-norm has simple defining expressions, which I write down in the next proposition. I omit its proof because it is standard in variational analysis, and not very informative to us. Just recall that $\mathbf{P}_{\mathscr{S}} = \sum_{i\in\mathscr{S}} \tilde{\mathbf{e}}_i \tilde{\mathbf{e}}_i^\top$, denoting the coordinate projection indexed by set $\mathscr{S}$.

Proposition 3.1.2. *Let* $\mathscr{S} := \operatorname{supp}(\mathbf{Dx})$ *for some matrix* $\mathbf{D} \in \mathbb{R}^{N\times n}$ *and some vector* $\mathbf{x} \in \mathbb{R}^n$. *A point* $\mathbf{z} \in \mathbb{R}^n$ *belongs to the subdifferential* $\mathbf{z} \in \partial\|\mathbf{D}\cdot\|_1(\mathbf{x})$ *if and only if* $\mathbf{z} = \mathbf{D}^\top\mathbf{w}$ *for some* $\mathbf{w} \in \mathbb{R}^N$ *satisfying simultaneously*

$$\mathbf{P}_{\mathscr{S}}\mathbf{w} = \operatorname{sign}(\mathbf{Dx}), \quad \textit{and} \tag{3.4}$$

$$\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}\|_\infty \leq 1. \tag{3.5}$$

The polar relationship between descent cone and the subdifferential allow us deduce a corresponding expression for (the closure of) $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$. The result, given as Proposition 3.1.3, has an accompanying proof in Appendix 3.A.1.

Proposition 3.1.3. *Let* $\mathscr{S} := \operatorname{supp}(\mathbf{Dx})$. *The topological closure of* $\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})$ *can be written as*

$$\overline{\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})} = \left\{\mathbf{u} \in \mathbb{R}^n : \left\langle \operatorname{sign}(\mathbf{Dx}), \mathbf{Du}\right\rangle \leq -\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Du}\|_1\right\}. \tag{3.6}$$

This expression allows us to connect the trivial intersection in Theorem 3.1.1 with the so-called *null-space property* commonly required in Compressed Sensing [22, Ch. 4]:

$$\|\mathbf{P}_{\mathscr{S}}\mathbf{Du}\|_1 < \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Du}\|_1, \quad \text{for all } \mathbf{u} \in \operatorname{null}(\mathbf{A}) \setminus \{\mathbf{0}\}. \tag{3.7}$$

Note that $\|\mathbf{P}_{\mathscr{S}}\mathbf{Du}\|_1 \geq \left\langle \mathbf{P}_{\mathscr{S}} \operatorname{sign}(\mathbf{Dx}), \mathbf{Du}\right\rangle = \left\langle \operatorname{sign}(\mathbf{Dx}), \mathbf{Du}\right\rangle$, hence (3.7) implies the trivial intersection $\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x}) \cap \operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$, through Proposition 3.1.3.

3.2 Why $\mathscr{G}$-TV minimization for piecewise-constant graph signals?

In science, it is often a good idea to apply Occam's razor: among competing hypothesis, pick the simplest one. In signal processing, sparse linear models are among the simplest. Let $\mathbf{t}_1, \mathbf{t}_2, \cdots \in \mathscr{T}$ be a set of vectors in $\mathbb{R}^n$, and $c_1, c_2, \cdots \in \mathbb{R}$ be constants. A linear model of a signal $\mathbf{x}$, using the *atoms* in $\mathscr{T}$ takes the form

$$\mathbf{x} = \sum_i c_i \mathbf{t}_i. \tag{3.8}$$

The multiplying constants quantify the importance of each atom in describing $\mathbf{x}$. If only a few of the constants is different than zero, we say that the model is sparse. Equivalently, $\mathbf{x}$ is explained by only a few of the atoms in $\mathscr{T}$.

On the one hand, if we want the output of (Pf) to have have a sparse linear expression as (3.8), then the set $\mathbf{x} + \mathscr{D}(f, \mathbf{x})$ gathering the potential solutions should contain the atoms in $\mathscr{T}$. On the other hand, Theorem 3.1.1 indicates that the descent cone should not be too "wide", lest it intersect with the null space of $\mathbf{A}$. The convex hull of $\mathscr{T}$, denoted conv($\mathscr{T}$), is the narrowest convex set containing $\mathscr{T}$, at least from the perspective of the angles at the atoms. Thus, whenever $\mathbf{x}$ is explained by combining few of such atoms, making $\mathscr{D}(f, \mathbf{x})$ the conic hull of $-\mathbf{x} + \operatorname{conv}(\mathscr{T})$ yields a fairly narrow cone. See Figure 3.3 for an illustration of this fact.

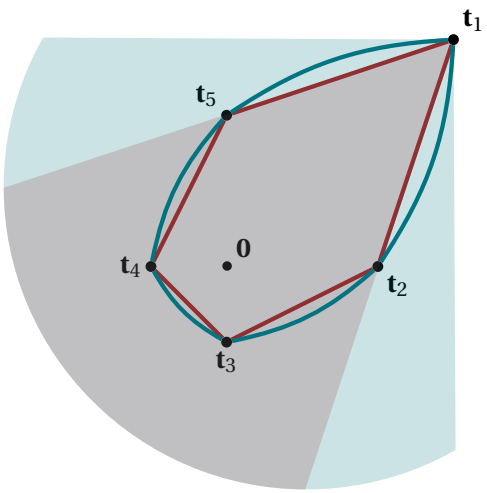


Figure 3.3 – An atomic set, $\mathscr{T} = \{\mathbf{t}_1, \ldots, \mathbf{t}_5\}$, whose convex hull is bounded by the thick red lines. The thick blue lines form the boundary of a larger convex set containing the atoms in $\mathscr{T}$. Suppose that $\mathbf{x} = \mathbf{t}_1$. If f is the gauge of $\operatorname{conv}(\mathscr{T})$, then the descent cone of this function at $\mathbf{x}$ has the same shape as the red cone. If, however, f is the gauge of the larger convex set, then its descent cone at $\mathbf{x}$ is as wide as the blue cone.

The function f for which $\mathscr{D}(f, \mathbf{x}) = \operatorname{cone}(-\mathbf{x} + \operatorname{conv}(\mathscr{T}))$ is the *gauge* of $\operatorname{conv}(\mathscr{T})$ [14]. This gauge is the map $\mathbf{z} \xmapsto{f} \|\mathbf{z}\|_{\mathscr{T}} := \inf\{t > 0 : \mathbf{z} \in t \operatorname{conv}(\mathscr{T})\}$, which becomes a Minkowski *norm* whenever $\mathscr{T}$ is symmetric about the origin. For a familiar example, the ℓ_1 norm in $\mathbb{R}^n$ is the gauge of the convex hull of standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

But what would be appropriate atomic set $\mathscr{T}$ for modelling piecewise-constant signals on graphs? Well, we have already a linear model for those signals in Chapter 2, using indicator vectors. Denote by $\mathscr{V}_1, \mathscr{V}_2, \ldots$ all possible vertex subsets on the graph, and let $\mathbb{1}_{\mathscr{V}_1}, \mathbb{1}_{\mathscr{V}_2}, \ldots$ be the corresponding indicator vectors. A piecewise-constant graph signal with P pieces is then of the form

$$\mathbf{x} = \sum_i c_i \mathbb{1}_{\mathscr{V}_i}, \tag{3.9}$$

where we can always choose only P of the constants $c_1, c_2, \ldots$ to be non-zero. The matching vertex subsets are disjoint and represent each of the constant pieces of the signal. Given (3.9) and the discussion of the previous paragraphs, it makes sense then pick indicator vectors of vertex subsets as our signal atoms. This modeling by a sparse linear combination of indicator vectors also appears in Jung [34] — for example — under the name of "clustering hypothesis".

The $\mathscr{G}$-TV semi-norm induces a fairly narrow descent cone if $\mathbf{x}$ is described by a sparse combination of vertex-subset indicators. Formally, Proposition 3.2.1 shows that such indicators are

the "extreme points" [5] of the $\|\mathbf{D}\cdot\|_1$-ball. As a result the tip of the cone $\mathscr{D}(\|\mathbf{D}\cdot\|1,\mathbf{x})$ resembles the neighborhood of of a hyper-edge on the polyhedron $\mathrm{conv}(\{\mathbb{1}_{\mathcal{V}_1},\mathbb{1}_{\mathcal{V}_2},\ldots\})$. The fewer atoms it takes to describe $\mathbf{x}$, the narrower this tip potentially is. Refer to Appendix 3.A.2 for a proof of the proposition — and a precise definition of "up-to-additive-constant extreme points".

Proposition 3.2.1 (Extreme points of $\mathbb{B}_{\mathcal{G}-TV}$, adapted from [69])**.** *A vector $\mathbf{t}^\star\in\mathbb{R}^n$ is an (up-to-additive-constant) extreme point of $\{\mathbf{x}\in\mathbb{R}^n:\|\mathbf{D}\mathbf{x}\|_1\leq 1\}$* if and only if *the entries of $\mathbf{t}^\star$ take on exactly two distinct values, $v_1\neq v_2\in\mathbb{R}$. In particular, such extreme points can be written as $\mathbf{t}^\star=(v_2-v_1)\cdot\mathbb{1}_{\mathcal{S}_{v_2}(\mathbf{t}^\star)}+v_1\mathbf{1}$, where $\mathcal{S}_{v_2}(\mathbf{t}^\star):=\{i\in\mathbb{R}^n:t_i=v_2\}$.*

3.2.1 Graph Total Variation versus the Dirichlet form

I have just presented the $\mathcal{G}$-TV semi-norm can be interpreted as the atomic gauge induced by indicator vectors of vertex subsets. But from the pure perspective of counting signal jumps, there is not much difference a priori between $\|\mathbf{D}\mathbf{x}\|_1=\sum_{i,j}\sqrt{W_{ij}}|x_i-x_j|$ and the Dirichlet form [6] $\|\mathbf{D}\mathbf{x}\|_2^2=\sum_{i,j}W_{ij}(x_i-x_j)^2$ as signatures for piecewise-constant signals. How different it is to solve

$$\min_{\mathbf{z}\in\mathbb{R}^n}\|\mathbf{D}\mathbf{z}\|_2^2 \text{ subject to } \|\mathbf{A}\mathbf{z}-\mathbf{y}\|_q^q\leq\eta, \tag{3.10}$$

rather than (P1-η)? After all, a decoder similar to (3.10) has been used by Puy *et al.*[60] for recovering sub-sampled, bandlimited [7] graph signals.

The answer I present in this section adapts the representer theorems of Unser *et al.*[74, 75] to compare the *form* of the points in the solution set of both ℓ_1 and ℓ_2 settings. A second, numerical perspective on the difference between (3.10) and (P1-η) can be found in Chapter 6.

To place ourselves within the same context as Unser *et al.*, let us generalize a bit the regression problems (3.10) and (P1-η). Denote by $\mathscr{C}$ a convex, compact set in $\mathbb{R}^m$ and consider the program

$$\mathscr{M}_p=\min_{\mathbf{z}\in\mathbb{R}^n}\|\mathbf{D}\mathbf{z}\|_p^p \text{ such that } \mathbf{A}\mathbf{z}\in\mathscr{C}. \tag{3.11}$$

Both (P1-η) and (P1-η) are particular instances of (3.11), achieved by setting $\mathscr{C}=\{\mathbf{v}\in\mathbb{R}^m:\|\mathbf{v}-\mathbf{A}\mathbf{x}\|_q^q\leq\eta\}$. This set is indeed convex and compact by virtue of being a scaled, translated norm ball.

[5] "Extreme points" is in quotation marks because $\mathbb{B}_{\mathcal{G}-\mathrm{TV}}$ does not actually have proper extreme points, consequence of the non-trivial null space of $\mathbf{D}$. Indeed, suppose $\mathbf{t}$ were an extreme point of $\mathbb{B}_{\mathcal{G}-\mathrm{TV}}$, and use some constant $c\in\mathbb{R}$ to create new points $\mathbf{s}:=\mathbf{t}+c\mathbf{1}$ and $\mathbf{u}:=\mathbf{t}-c\mathbf{1}$. Then, $\|\mathbf{D}\mathbf{s}\|_1=\|\mathbf{D}\mathbf{u}\|_1=\|\mathbf{D}\mathbf{t}\|_1$, because $\mathrm{null}(\mathbf{D})\supset\mathrm{span}(\mathbf{1})$. This implies $\mathbf{s},\mathbf{t}\in\mathbb{B}_{\mathcal{G}-\mathrm{TV}}$. But at the same time, $\mathbf{t}=\frac{1}{2}(\mathbf{s}+\mathbf{u})$, so $\mathbf{t}$ cannot be an extreme point of $\mathbb{B}_{\mathcal{G}-\mathrm{TV}}$. The proposition goes around this problem by defining extreme points up to the null space of $\mathbf{D}$.

[6] The name "Dirichlet form" is inspired by noting that $\frac{1}{2}\mathbf{D}^\top\mathbf{D}$ is usually called the *Laplacian* matrix $\mathbf{L}$ [66]. Hence, $\|\mathbf{D}\mathbf{x}\|_2^2=\mathbf{x}^\top\mathbf{L}\mathbf{x}$ is a Laplacian quadratic form.

[7] Let $\mathbf{L}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ be the singular value decomposition of the graph Laplacian matrix, with eigenvector columns ordered from smallest to largest w.r.t. their respective eigenvalues. A signal is said to be *k-bandlimited* if it lies in the span of the first k eigenvectors in $\mathbf{U}$.

The first representer theorem concerns the ℓ_1 version of the general program (3.11). Its proof, presented in Appendix 3.B.1, is based on [74, Theorem 19].

Theorem 3.2.1. *Consider problem* (3.11) *with* $p = 1$ *and assume that:*

1. $\operatorname{null}(\mathbf{D}) \cap \operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$.
2. *The set* $\mathscr{C} \subset \mathbb{R}^m$ *is convex and compact.*
3. *The pre-image* $\mathbf{A}^{-1}(\mathscr{C}) := \{\mathbf{z} \in \mathbb{R}^n : \mathbf{A}\mathbf{z} \in \mathscr{C}\}$ *is non-empty.*

Then, the extreme points of the solution set $\mathscr{M}_1$ *are of the form*

$$\mathbf{z}^\star = \mathbf{D}^+\mathbf{a}^\star + \mathbf{b}^\star, \tag{3.12}$$

where $\mathbf{a}^\star \in \mathbb{R}^N$ *has at most* m *non-zero coefficients, and* $\mathbf{b}^\star \in \operatorname{null}(\mathbf{D})$.

The graph gradient matrix $\mathbf{D}$, mapping signals to edge differences, embodies our prior information that piecewise-constant graph signals vary little across edges. The columns of the Moore-Penrose pseudo-inverse, $\mathbf{D}^+$, transfers the edge-differences perspective back to the vertex domain. The extreme points of the solution set of (3.11) are effectively explained by a small selection of columns of the pseudo-inverse. Consequently, the signals that we hope to recover have to be, in a sense, "compatible" with the network structure because $\mathbf{D}^+$ depends solely on the graph. For an example of compatible signal, turn to Figure 3.4, where we meet once again the Swiss National Council graph introduced in Chapter 1. The presence of an atom of $\mathbf{D}^+$ so well-aligned with the party split indicates that the graph connections are a good predictor of whether a council member belongs or not to UDC. In other words, the party is a fairly well-defined community in the network.

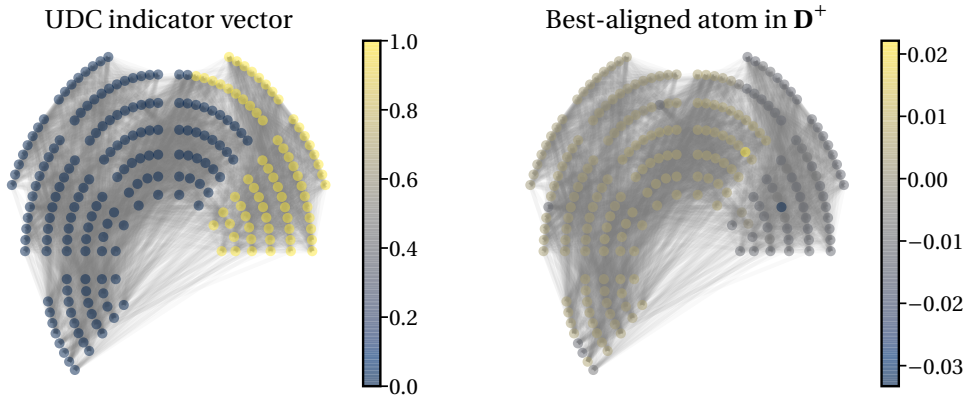


Figure 3.4 – A piecewise-constant signal that is compatible with the graph structure. In the Swiss National Council graph depicted above, council members are linked by similarity in their voting patterns. The left image displays the indicator vector of the right-wing UDC party. Its best match among the columns (atoms) of the matrix $\mathbf{D}^+$, in terms of inner product, is shown on the right. Since $\mathbf{D}^+$ depends on the graph alone, the UDC party is fairly well encoded in the edge distribution and forms an effective community in the network.

Let us turn our attention now to the ℓ_2 version of problem (3.11). Unlike before, the measurement matrix $\mathbf{A}$ makes an appearance on the representer theorem.

Theorem 3.2.2. *Consider problem* (3.11)*, with* $p = 2$*, and use the same set of assumptions as in Theorem 3.2.1. Then,* all *the points of the solution set* $\mathscr{M}_2$ *are of the form*

$$\mathbf{z}^\star = \mathbf{D}^+\mathbf{A}^\top\mathbf{v} + \mathbf{b}^\star, \tag{3.13}$$

where $\mathbf{v} \in \mathbb{R}^m$ *is a fixed vector, and* $\mathbf{b}^\star \in \operatorname{null}(\mathbf{D})$.

Both theorems 3.2.1 and 3.2.2 include an unavoidable term in the null space of $\mathbf{D}$, but the solution set induced by the Dirichlet form depends explicitly on the measurement matrix $\mathbf{A}$. Unser *et al.*[74] use this contrast to argue that ℓ_1 regularization is better at imposing prior information *despite* the form of the measurements. This position will become even more appealing in Chapter 6, where I show that the recovery error of the Dirichlet form decoder is more sensitive to $\mathbf{A}$ than to the graph connections represented in $\mathbf{D}$.

3.3 Relevant recovery results in the literature

Problems with (potentially) redundant dictionaries [8] such as (P1-η) have been studied in Compressed Sensing at least since Rauhut *et al.*[61] and Candès *et al.*[11]. If $\mathbf{D}$ were a tight frame and $\mathbf{A}$ a Gaussian matrix, then the recovery error $\|\mathbf{z}^\star - \mathbf{x}\|_2$ of any solution $\mathbf{z}^\star$ to (P1-η)

[8] The matrix $\mathbf{D}$ in $\|\mathbf{D}\mathbf{x}\|_1$ is often called dictionary or analysis operator because it maps $\mathbf{x}$ to a potentially sparse representation.

would be inversely proportional to $s := \|\mathbf{Dx}\|_0$, with an optimal dependence on the noise level [11, Theorem 1.2]. The argument that Candès *et al.*use to show this relies on **A** satisfying a certain **D**-Restricted Isometry Property. The number $(N-s)$ [9] is sometimes called the cosparsity level of **x** under the analysis operator **D**. Nam *et al.*[54] and Kabanava and Rauhut [36] are good places to find more results about cosparse models in compressed sensing. Still, these papers focus on frames for dictionaries Gaussian matrices measurements. Our graph gradient matrix is not a frame [10], and our measurement vectors are not Gaussian.

In Kabanava *et al.* [37] and Krahmer *et al.* [40] we start to see difference operators being used as the analysis transform, but the measurements are still (sub-)Gaussian-like. Poon [59] also studies Total Variation minimization but in the context of weighted Fourier sampling. All of these works restrict themselves to signals lying on the Euclidean grid, so their difference matrix **D** can be interpreted as the graph gradient of the grid graph. I would like to be able to deal with more general graphs. Kabanava *et al.*[37] attack the Total Variation problem by examining the descent cone of $\|\mathbf{D}\cdot\|_1$ and deriving a direct certificate of recovery, a strategy that we will explore in Chapter 4. Poon [59], on the other hand, employs a golfing scheme, derived from the properties of the subgradient of $\|\mathbf{D}\cdot\|_1$, arriving at a so-called dual certificate. This is the line of work developed in Chapter 5, which is also greatly informed by the work of Boyer *et al.*[10].

One can also refer to Jung *et al.*[33, 32, 34] for a more network-oriented perspective on Total Variation recovery problems. Under their "clustering hypothesis", a piecewise-constant graph signal naturally induces a partition of the vertices into disjoint clusters. The sampled vertices are said to "resolve" the partition if they satisfy the Network Null Space Property defined below.

Definition 3.3.1 (Network Null Space Property (NNSP)[33])**.** The measurement matrix **A** satisfies the NNSP with respect to an edge set $\mathscr{S} \subseteq \mathscr{E}$ if

$$\|\mathbf{P}_{\mathscr{S}}\mathbf{Du}\|_1 \leq \frac{1}{2}\|(\mathbf{I}_n - \mathbf{P}_{\mathscr{S}})\mathbf{Du}\|_1, \tag{3.14}$$

for any $\mathbf{u} \in \operatorname{null}(\mathbf{A}) \setminus \{\mathbf{0}\}$.

Similarly to other null-space properties in compressed sensing, the NNSP is shown to imply exact recovery in the interpolation problem (P1).

Proposition 3.3.1 ([33])**.** *If* **A** *satisfies the NNSP w.r.t.* $\mathscr{S} = \operatorname{supp}(\mathbf{Dx})$*, then* **x** *is the unique solution* (P1)*.*

Jung *et al.* [33] give deterministic conditions upon which the sampled vertices resolve the partition induced by a piecewise signal of interest. These conditions look into the hop distances of sampled vertices to the borders of the partition, but do not otherwise inform how to design an optimal *random* sampling for the recovery problem.

[9] Recall that $\mathbf{D} \in \mathbb{R}^{N\times n}$.

[10] Due to its non-trivial null-space.

In this thesis, I approach optimal sampling design for the decoders (P1) and (P1-η) in a similar manner as Puy *et al.*[60] did. They showed that a type of coherence-weighted sampling can minimize the number of rows that $\mathbf{A}$ needs to represent an approximate isometry map for bandlimited graph signals. The decoder is naturally constructed by penalizing non-bandlimited vectors. In our case, we have decoders (P1) and (P1-η) that were hand-picked for another class of graph signals, piecewise-constant. But the optimal sampling design is still deduced by minimizing the sample complexity of the decoder with respect to the vertex sampling probabilities $\boldsymbol{\pi} = (\pi_i, \ldots, \pi_n)$.

What kind of sample complexity expressions can one expect? Readers familiar with Compressed Sensing might know that $m = \mathcal{O}\left(s \log(n/s)\right)$ random Gaussian measurements are enough to recover, with high probability, any s-sparse vector via ℓ_1 minimization [22, Chapter 9]. The "sparsity-level" in Graph Total Variation ($\mathcal{G}$-TV) problems is $s := \|\mathbf{Dx}\|_0$, which is the number of edges in the jump-set $\mathcal{S} := \operatorname{supp}(\mathbf{Dx})$ of signal $\mathbf{x}$. Can one hope for sample complexity thresholds in (P1) that are proportional to $\|\mathbf{Dx}\|_0$? Surprisingly, such thresholds appear in the work of Lee *et al.* [48]. They even apply for coordinate sampling, and have accompanying optimal designs. Why not just adapt Lee *et al.*'s work to the context of graph signals to answer the main question in this thesis?

The reason is that Lee *et al.* [48]'s bounds can be vacuous when dealing with graphs signals. I will show in Chapter 6 examples where $|\mathcal{S}|$ is large in comparison to the number of vertices, but we can still successfully interpolate the respective signals using relatively few samples. For such large jump-sets, Lee *et al.* [48] would prescribe a number of measurements, m, larger than the total number of vertices, n. Therefore our sample complexity should be more than just a function of the cardinality of the jump-set and potentially depend on the signal-to-be-recovered [11].

This conclusion is not particularly new in the context of analysis-sparse (or co-sparse) models. Recent studies make a good case for considering only non-uniform guarantees whenever the decoder computes the ℓ_1 norm in the image of a non-invertible analysis operator $\mathbf{D}$. For instance, Giryes *et al.* [25, 26] show that no algorithm can accurately recover a co-sparse signal from a number of measurements proportional to the signal's manifold dimension. Meanwhile, Genzel *et al.* [24] construct two examples of analysis operators for which a signal has the same cosparsity level, but lead to completely different recovery thresholds [12]. Once again, those observations are to reinforce the fact that one should avoid naively using sample complexity thresholds proportional to $\|\mathbf{Dx}\|_0$ in problems like (Pf).

[11] Results with this characteristic are often called "non-uniform" in Compressed Sensing. They are named in contrast with results that are valid for a whole class of signals sharing a certain descriptive parameter. For example, any statement containing "... for all s-sparse vectors ..." is a uniform result.

[12] Genzel *et al.* [24] prescribe instead thresholds that depend on the partition of the Gram matrix $\mathbf{DD}^\top$ by the support $\mathcal{S} := \operatorname{supp}(\mathbf{Dx})$. Through $\mathcal{S}$, their guarantee depends on $\mathbf{x}$, hence is non-uniform.

3.4 Summary and final notes

The variational principle is a popular paradigm for recovering signals from subsampled linear measurements. Among all vectors satisfying the observation constraints, the returned solution is the one that minimizes a loss function f. This function is often chosen to be convex for analytical and numerical reasons.

There is also precedent in the literature for setting the search space directly, without resorting to a penalty f. Problems of the type "$\min_{z\in\mathbb{R}^n} \|\mathbf{Az}-\mathbf{y}\|_q^q$ subject to $\mathbf{z}\in\mathscr{K}$", where $\mathscr{K}$ is a convex set, are sometimes called the $\mathscr{K}$-lasso [57]. Despite the difference at first sight, there is a $\rho>0$ for which the $\mathscr{K}$-lasso becomes equivalent to (Pf-η) by setting $\mathscr{K}=\{\mathbf{z}\in\mathbb{R}^n : f(\mathbf{z})\leq\rho\}$ [13]. Some researchers, like Bora *et al.* [7], might even use non-convex constraint sets $\mathscr{K}$. They set out to recover images of human faces from few pixel measurements, having access to a pre-trained Generative Adversarial Network (GAN). This GAN excels in creating natural-looking faces, so Bora *et al.*argue that its range can be used as the recovery search space. It is an implicit way to avoid points in $\mathbb{R}^n$ that do not look like human faces, but it definitely makes the problem harder to study mathematically.

The $\mathscr{G}$-TV semi-norm is a good penalty choice for recovering piecewise-constant signals on graphs. These signals have a sparse representation in terms of the extreme points of the $\mathscr{G}$-TV ball. As a result, $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$ should be fairly narrow when $\mathbf{x}$ is piecewise constant, making it easier for a trivial intersection with null$(\mathbf{A})$ to take place. Ultimately, the trivial intersection is what guarantees the exact recovery of $\mathbf{x}$ from $\mathbf{Ax}$. I have also argued that minimizing $\|\mathbf{D}\cdot\|_1$ is preferable to minimizing $\|\mathbf{D}\cdot\|_2^2$ if we want solutions less dependent on the form of the measurement matrix.

Our main decoders, (P1) and (P1-η), fall under the scope of co-sparse models in Compressed Sensing. But the most relevant work in this field (Lee *et al.* [48]) prescribes sample complexity bounds that are potentially vacuous when dealing with graph signals. In the end, I have to reach new recovery guarantees from scratch, with the next two chapters compiling attempts to do so. Chapter 4 deals with a gain functional certifying that the descent cone $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$ and the subspace null$(\mathbf{A})$ intersect only at $\mathbf{0}$. In parallel, Chapter 5 uses the characteristics of the subdifferential $\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})$ as a blueprint to iteratively build a recovery certificate.

[13] This can be proved by the method of Lagrange multipliers [9, Chapter 5].

Appendix 3.A Geometry of the $\mathscr{G}$-TV semi-norm

3.A.1 Proof of Proposition 3.1.3

Here I adapt [41, Lemma 4.1], where Krahmer and Stöger characterize the descent cone of the matrix spectral norm. But first, let me precise the notion of polarity for cones.

Definition 3.A.1 (Polar cone)**.** Let $\mathscr{K} \subset \mathbb{R}^n$ be a cone. The corresponding polar cone, denoted $\mathscr{K}^\circ$, is the set

$$\mathscr{K}^\circ := \{\mathbf{v} \in \mathbb{R}^n : \langle \mathbf{v}, \mathbf{u} \rangle \leq 0, \forall \mathbf{u} \in \mathscr{K}\}. \tag{3.15}$$

I will say that a vector $\mathbf{v}$ is polar to a set $\mathscr{S}$ if $\langle \mathbf{v}, \mathbf{s} \rangle \leq 0, \forall \mathbf{s} \in \mathscr{S}$. Similarly, $\mathbf{v}$ is polar to another vector $\mathbf{u}$ if $\langle \mathbf{v}, \mathbf{u} \rangle \leq 0$.

PROOF: The desired expressions are a consequence of the subdifferential $\partial \|\mathbf{D}\cdot\|_1(\mathbf{x})$ being polar to $\overline{\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})}$ [62, Thm. 23.7]. Recall from Proposition 3.1.2 that $\mathbf{v} \in \partial \|\mathbf{D}\cdot\|_1(\mathbf{x})$ if and only if $\mathbf{v} = \mathbf{D}^\top \operatorname{sign}(\mathbf{D}\mathbf{x}) + \mathbf{D}^\top(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}$, for some $\mathbf{w} \in \mathbb{R}^N$ satisfying $\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}\|_\infty \leq 1$. We then split the argument in two parts:

⟨1⟩1. " $\Longleftarrow$ ": If $\left\langle \operatorname{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{u} \right\rangle \leq -\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{u}\|_1$ then $\mathbf{u} \in \overline{\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})}$.

PROOF: For any given $\mathbf{v} \in \overline{\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})}^\circ$, directly compute

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{u} \rangle &= \left\langle \mathbf{D}^\top \operatorname{sign}(\mathbf{D}\mathbf{x}) + \mathbf{D}^\top(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}, \mathbf{u} \right\rangle \\
&= \left\langle \operatorname{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{u} \right\rangle + \langle (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}, \mathbf{D}\mathbf{u} \rangle \\
&\leq \left\langle \operatorname{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{u} \right\rangle + \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{u}\|_1 \cdot \underbrace{\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}\|_\infty}_{\leq 1} \quad \texttt{(Hölder inequality)} \\
&\leq 0. \qquad \texttt{(Assumption)}
\end{aligned}$$

Therefore, $\mathbf{u}$ is polar to $\overline{\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})}^\circ$, meaning $\mathbf{u} \in \overline{\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})}$. □

⟨1⟩2. " $\Longrightarrow$ ": If $\mathbf{u} \in \overline{\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})}$ then $\left\langle \operatorname{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{u} \right\rangle \leq -\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{u}\|_1$.

PROOF:

⟨2⟩1. Pick a vector $\mathbf{w} \in \mathbb{B}_\infty^N$ for which $\langle \mathbf{w}, (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{u} \rangle = \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{u}\|_1$. Then $\mathbf{v} = \mathbf{D}^\top \operatorname{sign}(\mathbf{D}\mathbf{x}) + \mathbf{D}^\top(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}$ is a valid subgradient in $\partial \|\mathbf{D}\cdot\|_1(\mathbf{x})$, because

$$\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}\|_\infty \leq \underbrace{\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\|_\infty}_{\leq 1} \cdot \underbrace{\|\mathbf{w}\|_\infty}_{\leq 1} \leq 1.$$

⟨2⟩2. Since $\mathbf{v} \in \partial \|\mathbf{D}\cdot\|_1(\mathbf{x})$ is polar to $\overline{\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})}$, we conclude that

$$\begin{aligned}
0 &\geq \langle \mathbf{u}, \mathbf{v} \rangle \qquad \texttt{(Assumption)} \\
&= \left\langle \mathbf{D}^\top \operatorname{sign}(\mathbf{D}\mathbf{x}) + \mathbf{D}^\top(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{w}, \mathbf{u} \right\rangle \\
&= \left\langle \operatorname{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{u} \right\rangle + \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{u}\|_1. \qquad \texttt{(Choice of } \mathbf{w}\texttt{)}
\end{aligned}$$

□

⟨1⟩3. Q.E.D.

3.A.2 Proof of Proposition 3.2.1

I will need to establish some notation before delving into the argument.

Definition 3.A.2 (Graph cut)**.** The *graph cut* associated with two vertex sets $\mathscr{S}_1, \mathscr{S}_2 \subset [n]$ is the set

$$\operatorname{cut}(\mathscr{S}_1, \mathscr{S}_2) := \left\{ e = \left(v_i, v_j\right) \in \mathscr{E} : (i \in \mathscr{S}_1 \wedge j \in \mathscr{S}_2) \vee (j \in \mathscr{S}_1 \wedge i \in \mathscr{S}_2) \right\}. \tag{3.16}$$

We also define the size of $\operatorname{cut}(\mathscr{S}_1, \mathscr{S}_2)$ to be

$$|\operatorname{cut}(\mathscr{S}_1, \mathscr{S}_2)| := \sum_{i \in \mathscr{S}_1} \sum_{j \in \mathscr{S}_2} \sqrt{W_{ij}}. \tag{3.17}$$

Definition 3.A.3 (Up-to-additive-constant extreme point)**.** We say that $\mathbf{t} \in \mathscr{S}$ is an up-to-additive-constant extreme point of $\mathscr{S}$ when we can write $\mathbf{t} = \beta\mathbf{s} + (1-\beta)\mathbf{u}$, using $\beta \in (0,1)$ and $\mathbf{s}, \mathbf{u} \in \mathscr{S}$, only if $\mathbf{s} - \mathbf{t} = c_s \mathbf{1}$ and $\mathbf{u} - \mathbf{t} = c_u \mathbf{1}$, for some constants $c_s, c_u \in \mathbb{R}$.

Lastly, define also the following *level sets* of $\mathbf{t} \in \mathbb{R}^n$, indexed by $v \in \mathbb{R}$:

$$\mathscr{S}_v(\mathbf{t}) := \{i \in \mathbb{R} : t_i = v\} \tag{3.18}$$

$$\mathscr{S}_{v^+}(\mathbf{t}) := \{i \in \mathbb{R} : t_i > v\} \tag{3.19}$$

$$\mathscr{S}_{v^-}(\mathbf{t}) := \{i \in \mathbb{R} : t_i < v\}. \tag{3.20}$$

I will prove a slightly stronger claim than the one stated in the main text: the up-to-additive-constants extreme points $\mathbf{t}^\star$ of $\mathbb{B}_{\mathscr{G}-\mathrm{TV}}$ are of the form

$$\mathbf{t}^\star = |\operatorname{cut}(\mathscr{S}_{v_1}(\mathbf{t}^\star), \mathscr{S}_{v_2}(\mathbf{t}^\star))| \cdot \mathbb{1}_{\mathscr{S}_{v_2}(\mathbf{t}^\star)} + v_1 \mathbf{1}. \tag{3.21}$$

The argument is adapted from Szlam *et al.* [69].

PROOF: Let $v_1 < v_2 < \cdots < v_d \in \mathbb{R}$ be all the distinct values taken by the coordinates of some $\mathbf{t} \in \mathbb{R}^n$. Trivially, $1 \le d \le n$. I will split the argument into the cases $d = 1$, $d = 2$, and $d \ge 3$, and show that $\mathbf{t}$ is an extreme point of $\mathbb{B}_{\mathscr{G}-\mathrm{TV}}$ if and only if $d = 2$. The precise expression for $\mathbf{t}^\star$ will appear as a consequence of the computations in the proof.

⟨1⟩1. If $d = 1$, $\mathbf{t}$ is not an extreme point of $\mathbb{B}_{\mathscr{G}-\mathrm{TV}}$.
PROOF: Vector $\mathbf{t} = v_1\mathbf{1}$ is constant, so $\|\mathbf{D}\mathbf{t}\|_1 = \sum_{(i,j)} \sqrt{W_{i,j}} |v_1 - v_1| = 0$. Since $\|\mathbf{D}\mathbf{t}\|_1 < 1$, vector $\mathbf{t}$ cannot be an extreme point of $\mathbb{B}_{\mathscr{G}-\mathrm{TV}}$.

⟨1⟩2. If $d \ge 3$ and $\|\mathbf{D}\mathbf{t}\|_1 = 1$, $\mathbf{t}$ is still not an extreme point of $\mathbb{B}_{\mathscr{G}-\mathrm{TV}}$.
PROOF: I will construct perturbations $\mathbf{s} := \mathbf{t} + \varepsilon_p \mathbb{1}_{\mathscr{S}_{v_p}(\mathbf{t})} - \varepsilon_q \mathbb{1}_{\mathscr{S}_{v_q}(\mathbf{t})}$ and $\mathbf{u} := \mathbf{t} - \varepsilon_p \mathbb{1}_{\mathscr{S}_{v_p}(\mathbf{t})} + \varepsilon_q \mathbb{1}_{\mathscr{S}_{v_q}(\mathbf{t})}$, for some small $\varepsilon_p, \varepsilon_q > 0$ and some $v_p, v_q \in \{v_1, \ldots, v_d\}$. With these constructions, I show that $\mathbf{s}, \mathbf{u} \in \mathbb{B}_{\mathscr{G}-\mathrm{TV}}$. Since $\mathbf{t} = \frac{1}{2}\mathbf{s} + \frac{1}{2}\mathbf{u}$, this will imply that $\mathbf{t}$ cannot be an

extreme point of $\mathbb{B}_{\mathscr{G}-\mathrm{TV}}$.

⟨2⟩1. Note first that if we pick any positive $\varepsilon < \min_{k\neq l\in[d]} |v_k - v_l|$ the effect of the perturbation $\varepsilon \mathbb{1}_{\mathscr{S}_v(\mathbf{t})}$ decouples from $\|\mathbf{Dt}\|_1$:

$$\begin{aligned}
\|\mathbf{D}(\mathbf{t}+\varepsilon\mathbb{1}_{\mathscr{S}_v(\mathbf{t})})\|_1 &= \sum_{(i,j)\in\mathscr{E}} \sqrt{W_{ij}}\left|t_i - t_j + \varepsilon\mathbb{1}_{\mathscr{S}_v(\mathbf{t})}(i) - \varepsilon\mathbb{1}_{\mathscr{S}_v(\mathbf{t})}(j)\right| \\
&= 2\sum_{i\in\mathscr{S}_{v^-(\mathbf{t})}}\sum_{j\in\mathscr{S}_v(\mathbf{t})}\sqrt{W_{ij}}\underbrace{|t_i - t_j - \varepsilon|}_{=t_j-t_i+\varepsilon} \\
&\quad + 2\sum_{i\in\mathscr{S}_v(\mathbf{t})}\sum_{j\in\mathscr{S}_v(\mathbf{t})}\sqrt{W_{ij}}|t_i - t_j| \\
&\quad + 2\sum_{i\in\mathscr{S}_v(\mathbf{t})}\sum_{j\in\mathscr{S}_{v^+}(\mathbf{t})}\sqrt{W_{ij}}\underbrace{|t_i - t_j + \varepsilon|}_{=t_j-t_i-\varepsilon} \\
&= \|\mathbf{Dt}\|_1 + \varepsilon\left(\underbrace{\left[\sum_{i\in\mathscr{S}_{v^-}(\mathbf{t})}\sum_{j\in\mathscr{S}_v(\mathbf{t})}\sqrt{W_{ij}}\right]}_{=:|\mathrm{cut}(\mathscr{S}_{v^-}(\mathbf{t}),\mathscr{S}_v(\mathbf{t}))|} - \underbrace{\left[\sum_{i\in\mathscr{S}_v(\mathbf{t})}\sum_{j\in\mathscr{S}_{v^+}(\mathbf{t})}\sqrt{W_{ij}}\right]}_{=:|\mathrm{cut}(\mathscr{S}_{v^+}(\mathbf{t}),\mathscr{S}_v(\mathbf{t}))|}\right).
\end{aligned}$$

⟨2⟩2. Therefore, if we pick $\varepsilon_p, \varepsilon_q < \frac{1}{2}\min_{k\neq l\in[d]} |v_k - v_l|$, we can write the $\mathscr{G}$-TV semi-norm of $\mathbf{s}$ as

$$\begin{aligned}
\|\mathbf{Ds}\|_1 &= \left\|\mathbf{D}\left(\mathbf{t} + \varepsilon_p\mathbb{1}_{\mathscr{S}_{v_p}(\mathbf{t})} - \varepsilon_q\mathbb{1}_{\mathscr{S}_{v_q}(\mathbf{t})}\right)\right\|_1 \\
&= \underbrace{\|\mathbf{Dt}\|_1}_{=1} + \varepsilon_p\left(\left|\mathrm{cut}(\mathscr{S}_{v_p^-(\mathbf{t})},\mathscr{S}_{v_p}(\mathbf{t}))\right| - \left|\mathrm{cut}(\mathscr{S}_{v_p^+(\mathbf{t})},\mathscr{S}_{v_p}(\mathbf{t}))\right|\right) \\
&\qquad - \varepsilon_q\left(\left|\mathrm{cut}(\mathscr{S}_{v_q^-(\mathbf{t})},\mathscr{S}_{v_q}(\mathbf{t}))\right| - \left|\mathrm{cut}(\mathscr{S}_{v_q^+(\mathbf{t})},\mathscr{S}_{v_q}(\mathbf{t}))\right|\right)
\end{aligned}$$

and, similarly for $\mathbf{u}$,

$$\begin{aligned}
\|\mathbf{Du}\|_1 &= \left\|\mathbf{D}\left(\mathbf{t} - \varepsilon_p\mathbb{1}_{\mathscr{S}_{v_p}(\mathbf{t})} + \varepsilon_q\mathbb{1}_{\mathscr{S}_{v_q}(\mathbf{t})}\right)\right\|_1 \\
&= 1 - \varepsilon_p\left(\left|\mathrm{cut}(\mathscr{S}_{v_p^-(\mathbf{t})},\mathscr{S}_{v_p}(\mathbf{t}))\right| - \left|\mathrm{cut}(\mathscr{S}_{v_p^+(\mathbf{t})},\mathscr{S}_{v_p}(\mathbf{t}))\right|\right) \\
&\qquad + \varepsilon_q\left(\left|\mathrm{cut}(\mathscr{S}_{v_q^-(\mathbf{t})},\mathscr{S}_{v_q}(\mathbf{t}))\right| - \left|\mathrm{cut}(\mathscr{S}_{v_q^+(\mathbf{t})},\mathscr{S}_{v_q}(\mathbf{t}))\right|\right)
\end{aligned}$$

⟨2⟩3. For both $\mathbf{s}$ and $\mathbf{u}$ to be on $\mathrm{bd}(\mathbb{B}_{\mathscr{G}-\mathrm{TV}})$, *i.e.*, $\|\mathbf{Ds}\|_1 = \|\mathbf{Du}\|_1 = 1$, it suffices to pick

$\varepsilon_p, \varepsilon_q$ such that

$$\varepsilon_p = \varepsilon_q \left(\frac{\left|\mathrm{cut}\left(\mathcal{S}_{v_q^-(\mathbf{t})}, \mathcal{S}_{v_q}(\mathbf{t})\right)\right| - \left|\mathrm{cut}\left(\mathcal{S}_{v_q^+(\mathbf{t})}, \mathcal{S}_{v_q}(\mathbf{t})\right)\right|}{\left|\mathrm{cut}\left(\mathcal{S}_{v_p^-(\mathbf{t})}, \mathcal{S}_{v_p}(\mathbf{t})\right)\right| - \left|\mathrm{cut}\left(\mathcal{S}_{v_p^+(\mathbf{t})}, \mathcal{S}_{v_p}(\mathbf{t})\right)\right|} \right).$$

⟨2⟩4. We conclude that $\mathbf{t}$ is not an extreme point of $\mathbb{B}_{\mathcal{G}-\mathrm{TV}}$, because $\mathbf{t} = \frac{1}{2}\mathbf{s} + \frac{1}{2}\mathbf{u}$ and $\mathbf{s}, \mathbf{u} \in \mathbb{B}_{\mathcal{G}-\mathrm{TV}}$.

⟨1⟩3. If $d = 2$ and $\|\mathbf{Dt}\|_1 = 1$, $\mathbf{t}$ is an (up-to-additive-constant) extreme point of $\mathbb{B}_{\mathcal{G}-\mathrm{TV}}$.
PROOF: It suffices to show that if $\mathbf{t} = \beta\mathbf{s} + (1-\beta)\mathbf{u}$ for some $\beta \in (0,1)$ and $\mathbf{s}, \mathbf{u} \in \mathrm{bd}(\mathbb{B}_{\mathcal{G}-\mathrm{TV}})$ with $\mathbf{s} \neq \mathbf{u}$, then $\mathbf{s}$ and $\mathbf{u}$ differ from $\mathbf{t}$ only by a constant.

⟨2⟩1. Since $\mathbf{t} = (v_2 - v_1)\mathbb{1}_{\mathcal{S}_{v_2}} + v_1\mathbf{1}$, $\mathbf{Dt}$ is supported on the edges in $\mathrm{cut}(\mathcal{S}_{v_2}(\mathbf{t}), \mathcal{S}_{v_1}(\mathbf{t})) \subset \mathcal{E}$.

⟨2⟩2. Define a new graph $\mathcal{R} = (\mathcal{V}, \mathcal{E}|_{\mathrm{cut}(\mathcal{S}_{v_2}(\mathbf{t}), \mathcal{S}_{v_1}(\mathbf{t}))})$ which is the restricted version of the original graph $\mathcal{G}$ to the edges in $\mathrm{cut}(\mathcal{S}_{v_2}(\mathbf{t}), \mathcal{S}_{v_1}(\mathbf{t}))$. Correspondingly, define a new graph difference operator $\mathbf{D}_{\mathcal{R}}$ by setting to zero the entries in $\mathbf{D}$ related to the edges in the complement of $\mathrm{cut}(\mathcal{S}_{v_2}(\mathbf{t}), \mathcal{S}_{v_1}(\mathbf{t}))$.

⟨2⟩3. By construction, the graph TV semi-norm of $\mathbf{t}$ is conserved in this new graph, *i.e.*, $\|\mathbf{Dt}\|_1 = \|\mathbf{D}_{\mathcal{R}}\mathbf{t}\|_1 = 1$.

⟨2⟩4. I can then use the triangle inequality to get the relation

$$\begin{aligned} 1 &= \|\mathbf{D}_{\mathcal{R}}\mathbf{t}\|_1 \\ &= \|\mathbf{D}_{\mathcal{R}}(\beta\mathbf{s} + (1-\beta)\mathbf{u})\|_1 \\ &\leq \beta\|\mathbf{D}_{\mathcal{R}}\mathbf{s}\|_1 + (1-\beta)\|\mathbf{D}_{\mathcal{R}}\mathbf{v}\|_1. \end{aligned}$$

⟨2⟩5. On the other hand, $\|\mathbf{Dt}\|_1 := \sum_{(i,j)} \sqrt{W_{ij}}|t_i - t_j|$ is monotonically decreasing w.r.t. the edge weights, and $\mathcal{R}$ had fewer edges than $\mathcal{G}$. Therefore, since both $\mathbf{s}$ and $\mathbf{u}$ are assumed to be on $\mathrm{bd}(\mathbb{B}_{\mathcal{G}-\mathrm{TV}})$, we thus verify

$$\begin{aligned} \beta\|\mathbf{D}_{\mathcal{R}}\mathbf{s}\|_1 + (1-\beta)\|\mathbf{D}_{\mathcal{R}}\mathbf{v}\|_1 &\leq \beta\underbrace{\|\mathbf{Ds}\|_1}_{=1} + (1-\beta)\underbrace{\|\mathbf{Dv}\|_1}_{=1} \\ &= \beta + (1-\beta) \\ &= 1. \end{aligned}$$

⟨2⟩6. By the two previous steps, we must have $\|\mathbf{Ds}\|_1 = \|\mathbf{D}_{\mathcal{R}}\mathbf{s}\|_1 = \|\mathbf{D}_{\mathcal{R}}\mathbf{u}\|_1 = \|\mathbf{Du}\|_1 = 1$. Hence, both $\mathbf{Ds}$ and $\mathbf{Du}$ must also be supported on $\mathrm{cut}(\mathcal{S}_{v_2}(\mathbf{t}), \mathcal{S}_{v_1}(\mathbf{t}))$.

⟨2⟩7. Finally, having verified that $\mathbf{Dt}, \mathbf{Ds}, \mathbf{Du}$ share the same support, and that $\|\mathbf{Dt}\|_1 = \|\mathbf{Ds}\|_1 = \|\mathbf{Du}\|_1$, we must conclude that $\mathbf{t}, \mathbf{s}, \mathbf{u}$ can differ only by elements on the null space of $\mathbf{D}$. The latter corresponds to the space of vectors of the form $c\mathbf{1}$, for $c \in \mathbb{R}$, so the claim is proved. □

⟨1⟩4. If $d = 2$ and $\mathbf{t}^\star$ is an (up-to-additive-constants) extreme point of $\mathbb{B}_{\mathcal{G}-\mathrm{TV}}$, then $\mathbf{t}^\star = |\mathrm{cut}(\mathcal{S}_{v_1}(\mathbf{t}^\star), \mathcal{S}_{v_2}(\mathbf{t}^\star))| \cdot \mathbb{1}_{\mathcal{S}_{v_2}(\mathbf{t}^\star)} + v_1\mathbf{1}$.
PROOF: This is just a computation exercise. Since $\mathbf{t}^\star = (v_2 - v_1)\mathbb{1}_{\mathcal{S}_{v_2}} + v_1\mathbf{1}$ and $\|\mathbf{Dt}^\star\|_1 = 1$,

we obtain the identity

$$\begin{aligned}1 &= \|\mathbf{D}\mathbf{t}^\star\|_1 \\ &= \sum_{(i,j)} \sqrt{W_{ij}}|t_i - t_j| \\ &= 2\sum_{i\in\mathscr{S}_{v_2}}\sum_{j\in\mathscr{S}_{v_1}} \sqrt{W_{ij}}(v_2 - v_1) \\ &= 2\left|\mathrm{cut}(\mathscr{S}_{v_2}(\mathbf{t}^\star), \mathscr{S}_{v_1}(\mathbf{t}^\star))\right| \cdot (v_2 - v_1).\end{aligned}$$

Therefore, $(v_2 - v_1) = \frac{1}{2|\mathrm{cut}(\mathscr{S}_{v_2}(\mathbf{t}^\star), \mathscr{S}_{v_1}(\mathbf{t}^\star))|}$. □

⟨1⟩5. Q.E.D.

We have exhausted the options for the number d of distinct coordinates, so $\mathbf{t}$ is an extreme point of $\mathbb{B}_{\mathscr{G}-\mathrm{TV}}$ if and only if $d = 2$. Furthermore, any such extreme point is of the form $2|\,\mathrm{cut}(\mathscr{S}_{v_1}(\mathbf{t}^\star), \mathscr{S}_{v_2}(\mathbf{t}^\star))| \cdot \mathbb{1}_{\mathscr{S}_{v_2}(\mathbf{t}^\star)} + v_1\mathbf{1}$.

Appendix 3.B Representer theorems

Throughout this appendix, we will consider problems of the type

$$\mathscr{M}_p = \min_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{D}\mathbf{z}\|_p^p \text{ such that } \mathbf{A}\mathbf{z} \in \mathscr{C}, \tag{3.11}$$

with $p \geq 1$, and assume that

$$\mathrm{null}(\mathbf{D}) \cap \mathrm{null}(\mathbf{A}) = \{\mathbf{0}\} \tag{A1}$$

$$\mathscr{C} \subset \mathbb{R}^m \text{ is compact and convex} \tag{A2}$$

$$\mathbf{A}^{-1}(\mathscr{C}) := \{\mathbf{z} \in \mathbb{R}^n : \mathbf{A}\mathbf{z} \in \mathscr{C}\} \text{ is non-empty.} \tag{A3}$$

The representer theorems I will prove are based on the work of Unser *et al.* [74], but here I extend their results to operators $\mathbf{D}$ without a right-inverse. [14] The first basic fact that I need to establish is the following characterization of the sets $\mathscr{M}_p$, for $p \geq 1$. It follows the reasoning of Unser *et al.* [74, Lemma 20].

Lemma 3.B.1. *For any $p \geq 1$, $\mathscr{M}_p$ is a non-empty, convex, and compact subset of $\mathbb{R}^n$.*

The proof relies on features of the interactions between linear maps and convex, compact sets, as well as on properties of norms in finite-dimensional spaces. I begin by characterizing the pre-image $\mathbf{A}^{-1}(\mathscr{C})$. Then, I use the Bolzano-Weierstrass theorem to show that $\mathscr{M}_p$ is not empty. Finally, writing $\mathscr{M}_p$ as an intersection of "well-behaved" sets, we can conclude that it must be convex and compact.

[14] A matrix has a right inverse if and only if it represents a surjective linear map. Our graph gradient operator, $\mathbf{D} \in \mathbb{R}^{N\times n}$, is not guaranteed to have a right inverse because $N > n$ in general — hence $\mathbf{D}$ cannot represent a surjective map.

PROOF:

⟨1⟩1. $\mathbf{A}^{-1}(\mathscr{C})$ is closed because $\mathscr{C}$ is closed and the linear map $L : \mathbf{z} \mapsto \mathbf{Az}$ is continuous.

⟨1⟩2. $\mathbf{A}^{-1}(\mathscr{C})$ is also convex, because the pre-image of a convex set $\mathscr{C}$ by a linear map is convex.

⟨1⟩3. $\mathscr{M}_p$ is non-empty.

PROOF:

⟨2⟩1. For notation sake, let $\gamma := \inf_{\mathbf{z} \in \mathbf{A}^{-1}(\mathscr{C})} \|\mathbf{Dz}\|_p^p$

⟨2⟩2. Let $(\mathbf{z}^{(i)})_{i \in \mathbb{N}}$ be a sequence of elements of $\mathbf{A}^{-1}(\mathscr{C})$ inducing a *decreasing* sequence $(\|\mathbf{Dz}^{(i)}\|_p^p)_{i \in \mathbb{N}}$ of norms such that $\lim\inf_{i \to \infty} \|\mathbf{Dz}^{(i)}\|_p^p = \gamma$.

⟨2⟩3. For each $i \in \mathbb{N}$, decompose $\mathbf{z}^{(i)}$ uniquely as $\mathbf{z}^{(i)} = \mathbf{D}^+ \mathbf{a}^{(i)} + \mathbf{b}^{(i)}$, where $\mathbf{a}^{(i)} = \mathbf{Dz}^{(i)}$ is an element of range $\left(\mathbf{D}^\top\right)$ and $\mathbf{b}^{(i)}$ is an element of null $(\mathbf{D})$.

⟨2⟩4. The sequence $(\mathbf{a}^{(i)})_{i \in \mathbb{N}}$ is bounded, because $\|\mathbf{a}^{(i)}\|_p^p = \|\mathbf{Dz}^{(i)}\|_p^p \leq \|\mathbf{Dz}_{i-1}\|_p^p \leq \cdots \leq \|\mathbf{Dz}_1\|_p^p$.

⟨2⟩5. The sequence $(\mathbf{b}^{(i)})_{i \in \mathbb{N}}$ is also bounded.

PROOF: To see this, note that assumption A1implies $\|\mathbf{b}^{(i)}\|_p^p \leq c\|\mathbf{Ab}^{(i)}\|_p^p$ for some positive constant $c > 0$. Then, compute

$$\begin{aligned} \|\mathbf{b}^{(i)}\|_p^p &\leq c\|\mathbf{Ab}^{(i)}\|_p^p \\ &= c\|\mathbf{Az}^{(i)} - \mathbf{AD}^+ \mathbf{a}^{(i)}\|_p^p \\ &\leq c\|\mathbf{Az}^{(i)}\|_2 + c\|\mathbf{AD}^+ \mathbf{a}^{(i)}\|_p^p. \end{aligned}$$

The left term on the RHS, $\|\mathbf{Az}^{(i)}\|_p^p$, is bounded because $\mathbf{Az}^{(i)} \in \mathscr{C}$ and $\mathscr{C}$ is a compact set. The right term is bounded because $\mathbf{a}^{(i)}$ is bounded and $\mathbf{AD}^+$ is a finite dimensional linear operator, hence bounded. □

⟨2⟩6. We can then extract a sub-sequence from $(\mathbf{z}^{(i)})_{i \in \mathbb{N}}$ that converges to some point $\mathbf{z}^{(\infty)} = \mathbf{D}^+ \mathbf{a}^{(\infty)} + \mathbf{b}^{(\infty)}$.

PROOF: Bolzano-Weierstrass theorem, using the boundedness of both $(\mathbf{a}^{(i)})_{i \in \mathbb{N}}$ and $(\mathbf{b}^{(i)})_{i \in \mathbb{N}}$ sequences. □

⟨2⟩7. The converging point $\mathbf{z}^{(\infty)}$ must satisfy $\|\mathbf{Dz}^{(\infty)}\|_p^p \leq \gamma$.

PROOF: Indeed, $\|\mathbf{Dz}^{(\infty)}\|_p^p \leq \|\mathbf{Dz}^{(i)}\|_p^p$, for any $i \in \mathbb{N}$, due to the sequence $(\|\mathbf{Dz}^{(i)}\|_p^p)_{i \in \mathbb{N}}$ being decreasing. Therefore, taking the limit on both sides of this inequality, $\mathbf{z}^{(\infty)}$ must satisfy $\|\mathbf{Dz}^{(\infty)}\|_p^p \leq \inf_{\mathbf{z} \in \mathbf{A}^{-1}(\mathscr{C})} \|\mathbf{Dz}\|_p^p =: \gamma$. □

⟨2⟩8. On the other hand, $\|\mathbf{Dz}^{(\infty)}\|_p^p \geq \inf_{\mathbf{z} \in \mathbf{A}^{-1}(\mathscr{C})} \|\mathbf{Dz}\|_p^p =: \gamma$ because $\mathbf{A}^{-1}(\mathscr{C})$ is closed, containing all limits of sequences of its elements.

⟨2⟩9. $\|\mathbf{Dz}^{(\infty)}\|_p^p = \gamma$, so $\mathscr{M}_p$ contains at least one point, namely $\mathbf{z}^{(\infty)}$.

□

⟨1⟩4. Finally, I can show that $\mathscr{M}_p$ is both convex and compact.

PROOF:

⟨2⟩1. Let $\mathscr{L}_p = \left\{\mathbf{z} \in \mathbb{R}^n : \|\mathbf{Dz}\|_p^p \leq \gamma\right\}$. We can write $\mathscr{M}_p$ as the intersection $\mathscr{M}_p = \mathscr{L}_p \cap \mathbf{A}^{-1}(\mathscr{C}) = \left\{\left[\mathscr{L}_p \cap \text{range}\,(\mathbf{D})\right] \cap \mathbf{A}^{-1}(\mathscr{C})\right\} \oplus \left\{\left[\mathscr{L} \cap \text{null}\,(\mathbf{D})\right] \cap \mathbf{A}^{-1}(\mathscr{C})\right\}$.

⟨2⟩2. $\mathscr{L}_p \cap \text{range}\,(\mathbf{D})$ is convex and compact because it is a norm ball in a subset of $\mathbb{R}^n$.

⟨2⟩3. Hence, $\left[\mathscr{L}_p \cap \operatorname{range}(\mathbf{D})\right] \cap \mathbf{A}^{-1}(\mathscr{C})$ is both convex and compact, by virtue of being the intersection of two convex sets, one closed and the other compact.

⟨2⟩4. For the second term in the direct sum, consider splitting the feasible set as $\mathbf{A}^{-1}(\mathscr{C}) = \left[\mathbf{A}^{-1}(\mathscr{C}) \cap \operatorname{range}\left(\mathbf{A}^\top\right)\right] \oplus \left[\mathbf{A}^{-1}(\mathscr{C}) \cap \operatorname{null}(\mathbf{A})\right]$. The first term is a one-to-one linear mapping from $\mathscr{C}$ to a set in $\mathbb{R}^n$. Therefore, the compactness of $\mathscr{C}$ implies the compactness of $\left[\mathbf{A}^{-1}(\mathscr{C}) \cap \operatorname{range}\left(\mathbf{A}^\top\right)\right]$. Moreover, the latter is a hyperplane slice of a convex set, so it is itself convex. As for the term $\left[\mathbf{A}^{-1}(\mathscr{C}) \cap \operatorname{null}(\mathbf{A})\right]$, it is the empty set if $\mathbf{0} \notin \mathscr{C}$, or equal to $\operatorname{null}(\mathbf{A})$ otherwise.

⟨2⟩5. Since $\|\mathbf{Dz}\|_p^p = 0 \iff \mathbf{z} \in \operatorname{null}(\mathbf{D})$, we have the identity $\mathscr{L}_p \cap \operatorname{null}(\mathbf{D}) = \operatorname{null}(\mathbf{D})$. Thus, by assumption A1,

$$\left[\mathscr{L}_p \cap \operatorname{null}(\mathbf{D})\right] \cap \left[\mathbf{A}^{-1}(\mathscr{C}) \cap \operatorname{null}(\mathbf{A})\right] = \begin{cases} \emptyset & \text{if } \mathbf{0} \notin \mathscr{C} \\ \{\mathbf{0}\} & \text{otherwise.} \end{cases}$$

In any case,

$$\begin{aligned} \left[\mathscr{L}_p \cap \operatorname{null}(\mathbf{D})\right] \cap \mathbf{A}^{-1}(\mathscr{C}) &= \operatorname{null}(\mathbf{D}) \cap \left\{\left[\mathbf{A}^{-1}(\mathscr{C}) \cap \operatorname{range}\left(\mathbf{A}^\top\right)\right] \oplus \left[\mathbf{A}^{-1}(\mathscr{C}) \cap \operatorname{null}(\mathbf{A})\right]\right\} \\ &= \operatorname{null}(\mathbf{D}) \cap \left[\mathbf{A}^{-1}(\mathscr{C}) \cap \operatorname{range}\left(\mathbf{A}^\top\right)\right], \end{aligned}$$

which is a hyperplane slice of a convex and compact set, thus also convex and compact.

⟨2⟩6. At last, we conclude that $\mathscr{M}_p$ is both convex and compact, because those properties are preserved under direct sum.

□

⟨1⟩5. Q.E.D.

3.B.1 Proof of Theorem 3.2.1

Let me restate below an informal reminder of this section's goal.

Claim. The extreme points $\mathbf{z}^\star$ of $\mathscr{M}_1$ are of the form $\mathbf{z}^\star = \mathbf{D}^+\mathbf{a}^\star + \mathbf{b}^\star$, where $\mathbf{a}^\star$ has at most m non-zero coordinates, and $\mathbf{b}^\star \in \operatorname{null}(\mathbf{D})$.

The main tool we will need is the next lemma.

Lemma 3.B.2. *Let* $\mathbf{D}(\mathbf{A}^{-1}(\mathscr{C})) := \{\mathbf{Dz} \in \mathbb{R}^N : \mathbf{Az} \in \mathscr{C} \subset \mathbb{R}^m\}$. *The extreme points of the set*

$$\widetilde{\mathscr{M}_1} := \min_{\mathbf{a} \in \mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))} \|\mathbf{a}\|_1 \tag{3.22}$$

have at most m non-zero coefficients.

The proof can be deduced from [74, Theorem 6], but I give the full argument here for completeness.

PROOF:

⟨1⟩1. First of all, $\widetilde{\mathscr{M}_1}$ is non-empty, convex and compact. This is a consequence of Lemma 3.B.1, with the linear transformation $\mathbf{z} \mapsto \mathbf{D}\mathbf{z}$ mapping $\mathscr{M}_1$ to $\widetilde{\mathscr{M}_1}$.

⟨1⟩2. By the Krein-Milman theorem, $\widetilde{\mathscr{M}_1}$ is then the closed convex hull of its extreme points.

⟨1⟩3. Let $\mathbf{a}^\star$ be one such extreme point, chosen arbitrarily. I will show that $\|\mathbf{a}^\star\|_0 \leq m$, implying the main claim.

PROOF:

⟨2⟩1. For notation sake, let $\gamma := \min_{\mathbf{a} \in \mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))} \|\mathbf{a}\|_1$, so that $\|\mathbf{a}^\star\|_1 = \gamma$.

⟨2⟩2. Prooceed by contradiction, assuming that $\|\mathbf{a}^\star\|_0 \geq m+1$. Without loss of generality, we can say that $\{\mathbf{a}_j^\star\}_{j=1}^{m+1}$ forms a set of non-zero coordinates of $\mathbf{a}^\star$.

⟨2⟩3. Define a new vector $\overline{\mathbf{a}} := \mathbf{a}^\star - \sum_{j=1}^{m+1} \mathbf{a}_j^\star \mathbf{e}_j$, where $\{\mathbf{e}_j\}_{j=1}^N$ forms the standard basis in $\mathbb{R}^N$. Note that by construction $\overline{\mathbf{a}}$ and $\mathbf{a}^\star$ have disjoint supports.

⟨2⟩4. Now, for each $j \in [m+1]$, define $\mathbf{v}_j := \mathbf{A}\mathbf{D}^+(\mathbf{a}_j^\star \mathbf{e}_j) \in \mathbb{R}^m$. Since any collection of $m+1$ vectors in $\mathbb{R}^m$ is linearly dependent, there must exist constants $c_1, c_2, \ldots, c_{m+1}$ for which $\sum_{j=1}^{m+1} c_j \mathbf{v}_j = \mathbf{0}$.

⟨2⟩5. Using these same constants, define a new vector in $\mathbb{R}^N$ through $\mathbf{a}_0 := \sum_{j=1}^{m+1} c_j \mathbf{a}_j^\star \mathbf{e}_j$. We already know that $\mathbf{a}^\star \in \mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))$, but we further remark that the perturbations $\mathbf{a}^\star - \varepsilon \mathbf{a}_0$ and $\mathbf{a}^\star + \varepsilon \mathbf{a}_0$ are also both in $\mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))$, for any $\varepsilon > 0$.

PROOF: $\mathbf{A}\mathbf{D}^+\mathbf{a}_0 = \sum_{j=1}^{m+1} c_j \mathbf{v}_j = \mathbf{0}$ by construction. Hence, $\mathbf{A}\mathbf{D}^+(\mathbf{a}^\star - \varepsilon \mathbf{a}_0) = \mathbf{A}\mathbf{D}^+\mathbf{a}^\star = \mathbf{A}\mathbf{D}^+(\mathbf{a}^\star + \varepsilon \mathbf{a}_0)$ for any $\varepsilon > 0$. □

⟨2⟩6. I now claim that $\sum_{j=1}^{m+1} c_j |\mathbf{a}_j^\star| = 0$, implying $\|\mathbf{a}^\star \pm \varepsilon \mathbf{a}_0\|_1 = \gamma$.

PROOF:

⟨3⟩1. Suppose otherwise that $\sum_{j=1}^{m+1} c_j |\mathbf{a}_j^\star| \neq 0$, and pick $\varepsilon \in \left(\frac{-1}{\max_{j \in [m+1]} |c_j|}, \frac{1}{\max_{j \in [m+1]} |c_j|} \right)$. Then either $\|(\mathbf{a}^\star - \varepsilon \mathbf{a}_0)\|_1 < \gamma$ or $\|(\mathbf{a}^\star + \varepsilon \mathbf{a}_0)\|_1 < \gamma$.

PROOF: This follows by computing

$$\begin{aligned}
\|(\mathbf{a}^\star \pm \varepsilon \mathbf{a}_0)\|_1 &= \left\| \overline{\mathbf{a}} + \sum_{j=1}^{m+1} (1 \pm \varepsilon c_j) \mathbf{a}_j^\star \mathbf{e}_j \right\|_1 \\
&\underset{\text{disjoint support}}{=} \|\overline{\mathbf{a}}\|_1 + \sum_{j=1}^{m+1} |1 \pm \varepsilon c_j| |\mathbf{a}_j^\star| \\
&\underset{\text{choice of } \varepsilon}{=} \|\overline{\mathbf{a}}\|_1 + \sum_{j=1}^{m+1} (1 \pm \varepsilon c_j) |\mathbf{a}_j^\star| \\
&\underset{\text{reordering}}{=} \|\mathbf{a}\|_1 \pm \sum_{j=1}^{m+1} \varepsilon c_j |\mathbf{a}_j^\star| \\
&= \gamma \pm \sum_{j=1}^{m+1} \varepsilon c_j |\mathbf{a}_j^\star|. \square
\end{aligned}$$

⟨3⟩2. But since $\gamma = \min_{\mathbf{a} \in \mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))} \|\mathbf{a}\|_1$ and we have established that both $\mathbf{a}^\star \pm \varepsilon \mathbf{a}_0$ belong to $\mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))$, the conclusion of the last step is absurd. Thus, $\sum_{j=1}^{m+1} c_j |\mathbf{a}_j^\star| = 0$ and $\|\mathbf{a}^\star \pm \varepsilon \mathbf{a}_0\|_1 = \gamma$, as claimed.

□

⟨2⟩7. A direct consequence of $\|\mathbf{a}^\star \pm \varepsilon \mathbf{a}_0\|_1 = \gamma$ is then that $\mathbf{a}^\star \pm \varepsilon \mathbf{a}_0 \in \widetilde{\mathscr{M}_1}$

⟨2⟩8. We have reached our contradiction: we are able to write $\mathbf{a}^\star$ as the convex combination $\mathbf{a}^\star = \frac{1}{2}(\mathbf{a}^\star + \varepsilon \mathbf{a}_0) + \frac{1}{2}(\mathbf{a}^\star - \varepsilon \mathbf{a}_0)$ of two points in $\widetilde{\mathscr{M}_1}$. Thus, $\mathbf{a}^\star$ *cannot* be an extreme point of $\widetilde{\mathscr{M}_1}$.

□

⟨1⟩4. Q.E.D.
Any extreme point of $\widetilde{\mathscr{M}_1} \subset \mathbb{R}^N$ must have at most m non-zero coordinates.

With Lemma 3.B.2 at hand, I am finally ready to prove Theorem 3.2.1.

PROOF OF THEOREM 3.2.1:

⟨1⟩1. The extreme points $\mathbf{z}^\star$ of $\mathscr{M}_1$ satisfy the equation $\mathbf{D}\mathbf{z}^\star = \mathbf{a}^\star$, where $\mathbf{a}^\star$ has at most m non-zero coefficients.
PROOF: Applying the change of variable $\mathbf{D}\mathbf{z} = \mathbf{a}$, call on Lemma 3.B.2 and realize that the extreme points of $\mathscr{M}_1$ are mapped to the extreme points of $\widetilde{\mathscr{M}_1}$ through the linear transformation $\mathbf{z} \mapsto \mathbf{D}\mathbf{z}$. □

⟨1⟩2. We can express any $\mathbf{z} \in \mathbb{R}^n$ as $\mathbf{z} = \mathbf{D}^+\mathbf{D}\mathbf{z} + (\mathbf{I}_n - \mathbf{D}^+\mathbf{D})\mathbf{z}$, by seeing $\mathbb{R}^n$ as a direct sum between range$\left(\mathbf{D}^\top\right)$ and null$(\mathbf{D})$.

⟨1⟩3. Q.E.D.
Any extreme point of $\mathbf{z}^\star$ of $\mathscr{M}_1$ takes the form $\mathbf{z}^\star = \mathbf{D}^+\mathbf{D}\mathbf{z}^\star + \mathbf{b}^\star = \mathbf{D}^+\mathbf{a}^\star + \mathbf{b}^\star$, for some $\mathbf{b}^\star \in$ null$(\mathbf{D})$, and $\mathbf{a}^\star \in \mathbb{R}^N$ satisfying $\|\mathbf{a}^\star\|_0 \leq m$.

3.B.2 Proof of Theorem 3.2.2

This section's goal is the following:

Claim. All the points in $\mathscr{M}_2$ are of the form $\mathbf{z}^\star = \mathbf{D}^+\mathbf{D}\mathbf{A}^\top\mathbf{v} + \mathbf{b}^\star$, where $\mathbf{v} \in \mathbb{R}^m$ is a fixed vector, and $\mathbf{b}^\star \in$ null$(\mathbf{D})$.

This time I adapt [74, Theorems 5, 9, 18] to use as our main tool.

Lemma 3.B.3. *Let* $\mathbf{D}(\mathbf{A}^{-1}(\mathscr{C})) := \{\mathbf{D}\mathbf{z} \in \mathbb{R}^N : \mathbf{A}\mathbf{z} \in \mathscr{C} \subset \mathbb{R}^m\}$. *The set*

$$\widetilde{\mathscr{M}_2} := \min_{\mathbf{a} \in \mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))} \|\mathbf{a}\|_2^2 \tag{3.23}$$

has a single point, $\mathbf{a}^\star$. *Furthermore, this point is of the form* $\mathbf{a}^\star = \mathbf{D}\mathbf{r}^\star$ *for some* $\mathbf{r} \in$ range$\left(\mathbf{A}^\top\right)$.

PROOF:

⟨1⟩1. The set $\mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))$ is convex and closed because it is a linear mapping of a convex and closed set $\mathscr{C}$. It is also non-empty because $\mathbf{A}^{-1}(\mathscr{C})$ is assumed to be non-empty.

⟨1⟩2. The solution set $\widetilde{\mathscr{M}_2}$ contains thus a single point, namely the the orthogonal projection of the origin, $\mathbf{0}$, onto the convex set $\mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))$. Let us call this single point $\mathbf{a}^\star$.

⟨1⟩3. Let $\mathbf{z}^\star$ be a point in $\mathbb{R}^n$ such that $\mathbf{a}^\star = \mathbf{D}\mathbf{z}^\star$. Decompose this point as $\mathbf{z}^\star = \mathbf{r}^\star + \mathbf{n}^\star$, where $\mathbf{r}^\star \in$ range$\left(\mathbf{A}^\top\right)$ and $\mathbf{n}^\star \in$ null$(\mathbf{A})$. Then, we must have $\mathbf{n}^\star = \mathbf{0}$.
PROOF:

⟨2⟩1. By assumption A1, the orthogonal projection operators $\mathbf{D}^+\mathbf{D}$ and $\mathbf{A}^+\mathbf{A}$ commute.

⟨2⟩2. This commutative property leads to the fact $\|\mathbf{Dz}^\star\|_2^2 = \|\mathbf{Dr}^\star\|_2^2 + \|\mathbf{Dn}^\star\|_2^2$. Indeed,

$$\begin{aligned}
\|\mathbf{Dz}^\star\|_2^2 &= \|\mathbf{Dr}^\star\|_2^2 + 2\langle \mathbf{Dn}^\star, \mathbf{Dr}^\star\rangle + \|\mathbf{Dn}^\star\|_2^2 \\
&= \|\mathbf{Dr}^\star\|_2^2 + 2\langle \mathbf{n}^\star, \mathbf{D}^\top\mathbf{Dr}^\star\rangle + \|\mathbf{Dn}^\star\|_2^2 \\
&\underset{\text{for some } \mathbf{z}\in\mathbb{R}^n}{=} \|\mathbf{Dr}^\star\|_2^2 + 2\langle \mathbf{n}^\star, \mathbf{D}^+\mathbf{DA}^+\mathbf{Az}^\star\rangle + \|\mathbf{Dn}^\star\|_2^2 \\
&\underset{\text{commutativity}}{=} \|\mathbf{Dr}^\star\|_2^2 + 2\langle \underbrace{\mathbf{n}^\star}_{\in\text{null}(\mathbf{A})}, \underbrace{\mathbf{A}^+\mathbf{AD}^+\mathbf{Dz}^\star}_{\in\text{range}(\mathbf{A}^\top)}\rangle + \|\mathbf{Dn}^\star\|_2^2 \\
&= \|\mathbf{Dr}^\star\|_2^2 + 0 + \|\mathbf{Dn}^\star\|_2^2
\end{aligned}$$

⟨2⟩3. The term $\mathbf{n}^\star$ must be in null $(\mathbf{D})$.
PROOF: To see this, note that $\mathbf{Az}^\star = \mathbf{Ar}^\star$, so $\mathbf{Ar}^\star$ is also in $\mathbf{D}(\mathbf{A}^{-1}(\mathscr{C}))$. But because $\mathbf{Dz}^\star$ is a norm minimizer, together with the result from the previous step, that $\|\mathbf{Dr}^\star\|_2^2 \geq \|\mathbf{Dz}^\star\|_2^2 = \|\mathbf{Dr}^\star\|_2^2 + \|\mathbf{Dn}^\star\|_2^2 \iff \|\mathbf{Dn}^\star\|_2^2 = 0$. Hence, $\mathbf{n}^\star \in$ null $(\mathbf{D})$. □

⟨2⟩4. In summary, $\mathbf{n}^\star \in$ null $(\mathbf{A}) \cap$ null $(\mathbf{D})$. Calling upon assumption A1 once again, we conclude that $\mathbf{n}^\star = \mathbf{0}$.

□

⟨1⟩4. Q.E.D.
The single point belonging to $\widetilde{\mathscr{M}_2}$ has the form $\mathbf{a}^\star = \mathbf{Dr}^\star$, for some $\mathbf{r}^\star \in \text{range}\left(\mathbf{A}^\top\right)$.

Finally, proving Theorem 3.2.2 is very similar to proving Theorem 3.2.1, but this time I call on Lemma 3.B.3 instead of Lemma 3.B.2.

PROOF OF THEOREM 3.2.2:

⟨1⟩1. Every point $\mathbf{z}^\star \in \mathscr{M}_2$ satisfies the equation $\mathbf{Dz}^\star = \mathbf{DA}^\top\mathbf{v}$, for a fixed vector $\mathbf{v} \in \mathbb{R}^m$.
PROOF: With the change of variable $\mathbf{Dz} = \mathbf{a}$, apply Lemma 3.B.3 and realize that the points of $\mathscr{M}_2$ are all mapped to the single point in $\widetilde{\mathscr{M}_2}$ through the linear transformation $\mathbf{z} \mapsto \mathbf{Dz}$. □

⟨1⟩2. We can express any $\mathbf{z} \in \mathbb{R}^n$ as $\mathbf{z} = \mathbf{D}^+\mathbf{Dz} + (\mathbf{I}_n - \mathbf{D}^+\mathbf{D})\mathbf{z}$.

⟨1⟩3. Q.E.D.
Any point of $\mathbf{z}^\star$ of $\mathscr{M}_2 \subset \mathbb{R}^n$ takes the form $\mathbf{z}^\star = \mathbf{D}^+\mathbf{Dz}^\star + \mathbf{b}^\star = \mathbf{D}^+\mathbf{DA}^\top\mathbf{v} + \mathbf{b}^\star$, for a fixed vector $\mathbf{v} \in \mathbb{R}^m$, and some $\mathbf{b}^\star \in$ null $(\mathbf{D})$.

4 Direct certificates: measurement gain inside the descent cone

In Chapter 3 I hinted at how the geometry of the descent cone influences the correctness of the solutions to the *interpolation* problem (Pf). In this chapter, I will show which sense of narrowness in the descent cones can lead to robust recovery guarantees for *regression* problems (Pf-η), and in particular (P1-η). Recall that the regression setting admits noisy measurements of the type $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ where we have a bound, $\|\mathbf{e}\|_q^q \leq \eta$, on the noise level.

Notions of width appear as a consequence of translating the trivial intersection property from Theorem 3.1.1 into something computable. This quantity is a positive lower bound on the measurement operator's gain restricted to the descent cone. I this lower bound a direct recovery certificate, in contrast to the dual certificates investigated in the next chapter.

Strategies for lower bounding the minimum gain vary according to which random matrix plays the role of measurement operator. It is enlightening at first to imagine what would happen if we had Gaussian measurement vectors. Then, I introduce Mendelson's small-ball method as a potential path towards a direct certificate in our setting. Unfortunately, the "spikiness" of our coordinate-sampling matrices proves to be a burden down this road.

The chapter finishes with an open question, but points towards a way to get a direct certificate for a robust recovery in (P1-η). A way that requires knowing more about the *coordinate structure* of descent cones induced by the $\mathscr{G}$-TV semi-norm. Ultimately, the sample complexity of $\mathscr{G}$-TV decoders only gets a workable expression in Chapter 5.

4.1 A positive gain functional as a recovery certificate

We can quantify the trivial intersection property in terms of any q-norm by noting that

$$\mathscr{D}(f,\mathbf{x}) \cap \operatorname{null}(\mathbf{A}) = \{\mathbf{0}\} \iff \|\mathbf{A}\mathbf{u}\|_q^q > 0, \forall \mathbf{u} \in \mathscr{D}(f,\mathbf{x}) \setminus \{\mathbf{0}\}. \tag{4.1}$$

This motivates the definition of a minimum gain functional as a computable proxy for the property.

Definition 4.1.1 (Minimum q-gain)**.** For any $q \geq 1$, the minimum q-gain of a measurement operator $\mathbf{A}$, restricted to the descent cone $\mathscr{D}(f,\mathbf{x})$ is the quantity

$$\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right) = \inf_{\mathbf{u}\in\mathscr{D}(f,\mathbf{x})\cap \mathrm{bd}(\mathbb{B}_q^n)} \|\mathbf{A}\mathbf{u}\|_q^q. \tag{4.2}$$

The right-hand side of (4.1) now simply reads as $\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right) > 0$. Automatically, by Theorem 3.1.1, we conclude that a positive minimum q-gain yields the uniqueness of $\mathbf{x}$ as a solution of the *interpolation* problem (Pf). In the literature, $\sqrt{\gamma_{\min}^{(2)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)}$ is also known as simply "minimum gain" [14], or "minimum conic singular value" [72]. Rather than being a mere numeric translation of a geometric property, minimum gain functionals also inform on the robustness of convex recovery programs to noise.

A *regression* program of the type

$$\min_{\mathbf{z}\in\mathbb{R}^n} f(\mathbf{D}\mathbf{z}) \text{ subject to } \|\mathbf{A}\mathbf{z}-\mathbf{y}\|_q^q \leq \eta. \tag{Pf-η}$$

recovers $\mathbf{x}$ robustly from $\mathbf{y} = \mathbf{A}\mathbf{x}+\mathbf{e}$ if any solution $\mathbf{z}^\star$ is as close to $\mathbf{x}$ as the noise $\mathbf{e}$ permits. Theorem 4.1.1 shows how the distance between $\mathbf{z}^\star$ and $\mathbf{x}$ is inversely proportional to the minimum gain functional, and directly proportional to the noise level.

Theorem 4.1.1 ([14, Prop. 2.2], [72, Prop. 2.6], [37, Thm. 4])**.** *If* $\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right) > 0$ *and* $\|\mathbf{e}\|_q^q \leq \eta$*, then any solution* $\mathbf{z}^\star$ *of problem* (Pf-η) *satisfies*

$$\|\mathbf{z}^\star-\mathbf{x}\|_q^q \leq \frac{2\eta}{\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)}. \tag{4.3}$$

Therefore, the larger $\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)$, the more robust the corresponding decoder is. Theorem 4.1.1 is a powerful result with a simple proof.

PROOF: A straightforward adaptation of the argument in [37, Theorem 4], connecting two separate inequalities to reach the claim. Let $\mathbf{z}^\star$ be some solution (Pf-η).

⟨1⟩1. $\|\mathbf{z}^\star - \mathbf{x}\|_q^q \leq \frac{\|\mathbf{A}(\mathbf{z}^\star - \mathbf{x})\|_q^q}{\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)}$

PROOF:

⟨2⟩1. $f(\mathbf{z}^\star) \leq f(\mathbf{x})$, because $\mathbf{z}^\star$ is a solution of (Pf-η).

⟨2⟩2. Hence, $\mathbf{z}^\star - \mathbf{x} \in \mathscr{D}(f,\mathbf{x})$.

⟨2⟩3. With $\mathbf{u} := \frac{1}{\|\mathbf{z}^\star - \mathbf{x}\|_q^q} \in \mathscr{D}(f,\mathbf{x}) \cap \mathbb{B}_q^n$, we conclude that $\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right) \cdot \|\mathbf{u}\|_q^q \leq \|\mathbf{A}\mathbf{u}\|_q^q$, by definition of the minimum q-gain.

□

⟨1⟩2. $\|\mathbf{A}(\mathbf{z}^\star - \mathbf{x})\|_q^q \leq 2\eta$.

PROOF: Just use the triangle inequality and the feasibility of both $\mathbf{z}^\star$ and $\mathbf{x}$:

$$\begin{aligned}\|\mathbf{A}(\mathbf{z}^\star - \mathbf{x})\|_q^q &= \|(\mathbf{A}\mathbf{z}^\star - \mathbf{y}) - (\mathbf{A}\mathbf{x} - \mathbf{y})\|_q^q \\ &\leq \|\mathbf{A}\mathbf{z}^\star - \mathbf{y}\|_q^q + \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_q^q \\ &\leq \eta + \eta.\end{aligned}$$

□

⟨1⟩3. Q.E.D.

$$\|\mathbf{z}^\star - \mathbf{x}\|_q^q \leq \frac{\|\mathbf{A}(\mathbf{z}^\star - \mathbf{x})\|_q^q}{\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)} \leq \frac{2\eta}{\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)}.$$

The dependence of $\|\mathbf{z}^\star - \mathbf{x}\|_q^q$ on the noise level is optimal, because all we know about $\mathbf{e}$ is the bound $\|\mathbf{e}\|_q^q \leq \eta$. Furthermore, Theorem 3.1.1 from Chapter 3 becomes a mere corollary, reached by making $\eta \to 0$ in the recovery error bound. Hence there is no loss in considering only regression — and not interpolation — problems in this chapter.

In Compressed Sensing, a metric condition more commonly used than $\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right) > 0$ is the Restricted Isometry Property (RIP), demanding that

$$c\|\mathbf{z}\|_q \leq \|\mathbf{A}\mathbf{z}\|_q \leq C\|\mathbf{z}\|_q, \;\; 0 < c < C, \;\; \forall \mathbf{z} \in \mathscr{D}(f,\mathbf{x}). \tag{4.4}$$

But left-hand side of (4.4) is equivalent to a positive minimum q-gain. The RIP is thus a stronger condition; too strong, in fact. Recent research [19] suggests that the right-hand side of (4.4) gives rise to a *gap* between the optimal and the RIP-certifiable numbers of measurements required in ℓ_1-recovery problems. The reason for the gap seems to be that the inequality $\|\mathbf{A}\mathbf{z}\|_q \leq C\|\mathbf{z}\|_q$ requires of $\mathbf{A}$ either very good concentration properties or a large number of rows [45]. Despite the popularity of the RIP, I will therefore focus only on how our measurement matrices can be made to satisfy $\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right) > 0$.

4.2 Interlude: what is known for Gaussian measurements?

The properties of Gaussian random vectors are among the easiest to characterize theoretically. It is no surprise then that most of the work in convex recovery [14, 72, 78, 57, 49, 58, 79] (in general), and Compressed Sensing [22, 23, 20, 19] (in particular) employ Gaussian measurement ensembles.

Gordon's "escape through a mesh" theorem [27, Cor. 1.2] connects the minimum 2-gain of a Gaussian matrix [1] $\mathbf{A}$ to its number of rows (measurements) through the Gaussian width of the descent cone [14, Cor. 3.3]. This notion of conic width through the lens of Gaussian vectors can be defined as follows.

Definition 4.2.1 (Conic Gaussian width [72, Def. 3.1])**.** Let $\mathcal{K} \subset \mathbb{R}^n$ be a cone and $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ be a standard Gaussian vector. The conic Gaussian width of $\mathcal{K}$ is the quantity

$$w(\mathcal{K}) := \mathbb{E}\left(\sup_{\mathbf{u} \in \mathcal{K} \cap \mathbb{S}^{n-1}} \langle \mathbf{g}, \mathbf{u} \rangle \right). \tag{4.5}$$

The conic Gaussian width is computed for the descent cones of some atomic norms in Chandrasekaran *et al.* [14]. The authors can then use those widths to arrive at the number of Gaussian measurements required for a robust recovery in the respective decoders.

More than that, it is even possible to precisely describe the phase transition undergone by the probability of recovery in convex recovery problems with Gaussian measurements. Such a result, derived from conic integral geometry tools, is presented next.

Theorem 4.2.1 (Phase transition [4, Thm. II])**.** *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *be a random measurement matrix with* i.i.d. *standard Gaussian entries. Define* `Success` *as the event that* $\gamma_{\min}^{(q)}\left(\mathscr{D}(f, \mathbf{x}), \mathbf{A}\right) > 0$ *takes place. Then, for any* $\varepsilon \in (0,1)$

$$\begin{cases} m \leq w(\mathscr{D}(f,\mathbf{x}))^2 - \sqrt{8n\log(4/\varepsilon)} & \implies \quad \mathbb{P}\left(\texttt{Success}\right) \leq \varepsilon \\ m \geq w(\mathscr{D}(f,\mathbf{x}))^2 + \sqrt{8n\log(4/\varepsilon)} + 1 & \implies \quad \mathbb{P}\left(\texttt{Success}\right) \geq 1 - \varepsilon \end{cases} \tag{4.6}$$

In words, the theorem says that if the number of Gaussian measurements crosses a barrier of $\mathcal{O}\left(\sqrt{n}\right)$, centered at $w(\mathscr{D}(f,\mathbf{x}))^2$, then the probability of success suddenly jumps from almost zero to almost one. Sharp transition phenomena like this one are ubiquitous in high-dimensional geometry [21, 56]. Below, I give a brief sketch of the proof of Theorem 4.2.1, because it complements the geometric intuition from Chapter 3.

[1] By "Gaussian matrix" I always mean a random matrix whose entries are independent draws from the standard Gaussian distribution.

PROOF SKETCH: If the entries of $\mathbf{A} \in \mathbb{R}^{m\times n}$ are *i.i.d.* standard Gaussian, then the solutions to the equation $\mathbf{Au} = \mathbf{0}$ lie almost surely on an $(n-m)$-dimensional subspace of $\mathbb{R}^n$ drawn uniformly at random [2]. The probability that a fixed cone, $\mathscr{D}(f,\mathbf{x})$, and a uniformly random subspace, $\text{null}(\mathbf{A})$, intersect only at $\mathbf{0}$ is given by the kinematic formula studied in conic integral geometry [65]. The kinematic formula is expressed as a sum of certain "conic intrinsic volumes", which are shown to concentrate sharply around the square of the conic Gaussian width [4] [3]

It is possible to rank different objectives f, from worst to best, according to how small is $w(\mathscr{D}(f,\mathbf{x}))$. This is one of the main benefits of pinpointing the sample complexity threshold so precisely. The width functional almost singlehandedly determines how many Gaussian measurements are needed for a successful recovery; good objectives f should require a small number of measurements. Lemma 4.2.2 shows a possible upper bound to the Gaussian width of the descent cone induced by the $\mathscr{G}$-TV semi-norm.

Lemma 4.2.2 (Conic Gaussian width of $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$ [37])**.** *Let* $\mathscr{S} := \text{supp}(\mathbf{Dx})$ *be the support of* $\mathbf{x}$ *under the action of the analysis operator* $\mathbf{D} \in \mathbb{R}^{N\times n}$*, and define* $\mathbf{P}_{\mathscr{S}}$ *to be the orthogonal projection operator onto* $\text{span}\{\mathbf{e}_i : i \in \mathscr{S}\}$*. Then, with* $\mathbf{g} \sim \mathscr{N}(\mathbf{0}, \mathbf{I}_n)$*,*

$$w(\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x}))^2 \leq n - \left[\frac{\mathbb{E}\left(\left\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dg}\right\|_1\right)}{\max\limits_{\|\mathbf{z}\|_2\leq 1}\|\mathbf{Dz}\|_1}\right]^2. \tag{4.7}$$

The $\mathscr{G}$-TV is suited to recover a signal $\mathbf{x}$ from Gaussian measurements only if the descent cone $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$ has small Gaussian width. Referring to Theorem 4.2.1 and inequality (4.7), the closer the term

$$\left[\mathbb{E}\left(\left\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dg}\right\|_1\right) \Big/ \max_{\|\mathbf{z}\|_2\leq 1}\|\mathbf{Dz}\|_1\right]^2$$

is to n, the fewer observations are needed for a successful recovery. Among these terms, the only depending on $\mathbf{x}$ (through the jump-set $\mathscr{S}$) is $\mathbb{E}\left(\left\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dg}\right\|_1\right)$. Intuitively, this expectation is made larger the smaller the jump-set is. Therefore, even if measured by Gaussian vectors, piecewise-constant graph signals (due to their small jump-set) seem to be efficiently reconstructed by $\mathscr{G}$-TV decoders.

I should remark that there are currently better estimates for the Gaussian width of $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$. The *sampling-rate function* in Genzel *et al.* [24] gives a tighter upper bound than the one in Lemma 4.2.2. I avoided introducing this function for the sake of presentation, but the reader is invited to check Genzel *et al.*'s paper because it discusses interesting properties of ℓ_1-analysis recovery programs in general.

[2] More precisely, a subspace distributed according to the Haar measure on the Grassmannian manifold $\mathscr{G}_{(n-m),n}$ invariant to the group of rotations $\text{SO}(n)$.

[3] Actually, Amelunxen *et al.*show that the conic intrinsic volumes concentrate around the so-called "statistical dimension" of the descent cone, $\delta(\mathscr{D}(f,\mathbf{x}))$. Nevertheless, this quantity is bounded as $w(\mathscr{D}(f,\mathbf{x}))^2 \leq \delta(\mathscr{D}(f,\mathbf{x})) \leq w(\mathscr{D}(f,\mathbf{x}))^2 + 1$ [4, Prop. 10.2], so the Gaussian width and the statistical dimension are essentially equivalent for the purposes of characterizing the phase transition in Theorem 4.2.1.

4.3 The small-ball method and its shortcomings

If the measurement matrix $\mathbf{A}$ is not Gaussian, which tools from probability theory can still be used to show when the random object $\gamma^{(q)}_{\min}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)$ is positive? The sampling matrices defined in Chapter 2 have *independent* rows, so the minimum q-gain is a bounded function of many independent random variables. By concentration of measure, we could argue that $\gamma^{(q)}_{\min}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)$ is then essentially constant [70], taking values that are almost always close to its mean. If $\mathbb{E}\left(\gamma^{(q)}_{\min}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)\right) > 0$ then the minimum q-gain would also be positive *with high probability*.

The downside of such concentration arguments is that they only work properly if the marginals of $\gamma^{(q)}_{\min}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)$ have well-behaved tails. The small-ball method [50, 38] was developed by Mendelson and others with the explicit goal of "obtaining high-probability, uniform estimates in heavy-tailed situations" [52, p. 7]. The method's name comes from its main assumption, a positive lower bound on the small-ball probability $\inf_{\mathbf{u}\in\mathbb{S}^{n-1}} \mathbb{P}\left(\{|\langle\mathbf{u},\mathbf{a}\rangle| > 0\}\right)$, where $\mathbf{a}$ is distributed as the rows of $\mathbf{A}$ in our context.

Definition 4.3.1 (Small-ball condition)**.** A random vector $\mathbf{v}$ satisfies a small-ball condition with constants $\kappa > 0$ and $\delta \in (0,1)$ if $\mathbb{P}\left(\{|\langle\mathbf{v},\mathbf{u}\rangle| \geq \kappa\|\mathbf{u}\|_2\}\right) \geq \delta$.

The small-ball condition can be linked to identifiability questions about linear functionals [46, 44]. Let X be a random variable distributed according to some probability measure μ. A class of linear functionals $\mathscr{F} = \{\langle\mathbf{v},\cdot\rangle : \mathbf{v} \in \mathscr{S}\}$ is *identifiable* under μ if $\mathbb{P}\left(\{\langle\mathbf{v},X\rangle \neq \langle\mathbf{u},X\rangle\}\right) > 0$ for every $\mathbf{u},\mathbf{v} \in \mathscr{S}$. This is equivalent to assuming $\mathbb{P}\left(\{|\langle\mathbf{v},X\rangle| > 0\}\right) > 0$ [46, 44], which is a small-ball condition. This condition is thus weak in the sense that it simply demands the distribution of random vectors X to be able to distinguish the functions in $\mathscr{F}$ — with some non-zero probability.

The first step towards the small-ball method is to see the minimum q-gain functional as the infimum of a non-negative empirical process induced by the rows of $\mathbf{A}$. Denote those rows by $\{\mathbf{a}_i\}_{i=1}^m$, so as to unpack the q-norm in Definition 4.1.1 as

$$\gamma^{(q)}_{\min}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right) = \inf_{\mathbf{u}\in\mathscr{D}(f,\mathbf{x})\cap\mathbb{B}^n_q} \left(\sum_{i=1}^m |\langle\mathbf{a}_i,\mathbf{u}\rangle|^q\right)^{1/q}. \tag{4.8}$$

Then, a series of non-trivial manipulations of this expression ends up lower bounding the minimum q-gain by the difference of two functionals, one related to a small-ball condition and the other to a new notion of width for the descent cone $\mathscr{D}(f,\mathbf{x})$. The first of these terms is the *marginal tail function*.

Definition 4.3.2 (Marginal tail function [72])**.** The marginal tail function, at level $\xi \geq 0$, of a random vector $\mathbf{v}$ restricted to a set $\mathscr{S}$ is defined as

$$Q_\xi(\mathbf{v},\mathscr{S}) := \inf_{\mathbf{u}\in\mathscr{S}\cap\mathbb{S}^{n-1}} \mathbb{P}\left(\{|\langle\mathbf{v},\mathbf{u}\rangle| \geq \xi\}\right) \tag{4.9}$$

The second functional appearing the the lower bound of $\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)$ is the *mean empirical width*. It has a similar expression to the conic Gaussian width of the previous section, but the expectation is taken with respect to a Rademacher average of the rows of $\mathbf{A}$.

Definition 4.3.3 (Mean empirical width [72])**.** Let $\varepsilon_1,\ldots,\varepsilon_m$ be *i.i.d.* copies of a Rademacher random variable [4]. The mean empirical width of a set $\mathscr{S}$, as measured by m *i.i.d.* copies, $\mathbf{v}_1,\ldots,\mathbf{v}_m$, of a random vector $\mathbf{v}$, is the quantity

$$W_m(\mathbf{v},\mathscr{S}) := \mathbb{E}\left(\sup_{\mathbf{u}\in\mathscr{S}\cap\mathbb{S}^{n-1}}\left\langle \underbrace{\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varepsilon_i\mathbf{v}_i}_{=:\mathbf{h}},\mathbf{u}\right\rangle\right) \tag{4.10}$$

In passing, note that whenever $\mathbf{v}$ has bounded moments the Central Limit Theorem tells us that the distribution of $\mathbf{h}$ tends to $\mathscr{N}(0,\mathbb{E}\left(\mathbf{v}\mathbf{v}^\top\right))$ as $m\to\infty$. If, on top of that, $\mathbf{v}$ is isotropic [5] then $W_m(\mathbf{v},\mathscr{S})$ will approximate the Gaussian width $w(\mathscr{S})$ as the number of *i.i.d.* copies of $\mathbf{v}$ grows. The functional W_m is really then an empirical analogous of the notion of set width induced by standard Gaussian vectors.

The precise expression relating the minimum q-gain to the marginal tail function and the mean empirical width is given in Proposition 4.3.1. Its proof is taken from Tropp [72], but I reproduce it in Appendix 4.A.1 for the interested reader.

Proposition 4.3.1 ([72, Prop. 5.1])**.** *Let the rows of* $\mathbf{A}\in\mathbb{R}^{m\times n}$ *be* i.i.d. *copies of a random vector* $\mathbf{a}$*. Then, for any constants* $\xi,t>0$*, and with probability at least* $1-\exp\left(\frac{-t^2}{2}\right)$*, we have the lower bound*

$$\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right) \geq m^{\frac{2-q}{2q}}\left[\xi\sqrt{m}Q_\xi(\mathbf{a},\mathscr{D}(f,\mathbf{x})) - 2W_m(\mathbf{a},\mathscr{D}(f,\mathbf{x})) - \xi t\right]. \tag{4.11}$$

This lower bound combines with Theorem 4.1.1 to form in the following robust recovery result for regression problems of the type (Pf-η).

Corollary 4.3.0.1. *Let the rows of* $\mathbf{A}\in\mathbb{R}^{m\times n}$ *be* i.i.d. *copies of a random vector* $\mathbf{a}$*. Then, for any constants* $\tau,\xi,t>0$*, any solution* $\mathbf{z}^\star$ *of problem* (Pf-η) *satisfies*

$$\|\mathbf{z}^\star-\mathbf{x}\|_q^q \leq \frac{2\eta}{\tau m^{(2-q)/2q}}, \tag{4.12}$$

with probability at least $1-\exp\left(\frac{-t^2}{2}\right)$*, provided that*

$$m \geq \left(\frac{2W_m(\mathbf{a},\mathscr{D}(f,\mathbf{x})) + \xi t + \tau}{\xi Q_\xi(\mathbf{a},\mathscr{D}(f,\mathbf{x}))}\right)^2. \tag{4.13}$$

[4] That is, a $\{-1,1\}$-valued random variable ε for which $\mathbb{P}(\{\varepsilon=-1\})=\mathbb{P}(\{\varepsilon=1\})=1/2$.

[5] $\mathbb{E}\left(\mathbf{v}\mathbf{v}^\top\right)=\mathbf{I}_n$

PROOF: Set $\tau := \xi\sqrt{m}Q_\xi(\mathbf{a},\mathscr{S}) - 2W_m(\mathbf{a},\mathscr{S}) - \xi t$ as a lower-estimate for $\gamma_{\min}^{(q)}\left(\mathscr{D}(f,\mathbf{x}),\mathbf{A}\right)$. This estimate is positive by condition (4.13). Then, combine Proposition 4.3.1 and Theorem 4.1.1. □

In light of Corollary 4.3.0.1, the small-ball method prescribes the following three steps for uncovering the sample complexity of decoders like (Pf-η) whenever the measurement matrix has independent rows.

1: Bound $Q_\xi(\mathbf{a},\mathscr{D}(f,\mathbf{x}))$ below.
2: Bound $W_m(\mathbf{a},\mathscr{D}(f,\mathbf{x}))$ above.
3: Return m according to Corollary 4.3.0.1.

There are many tools for attacking the non-trivial steps 1 and 2. Unless we already know the constants in the small-ball condition satisfied by $\mathbf{a}$, a Paley-Zygmund inequality [18, Prop. 3.3.1] may lower-bound the marginal tail function. Generic chaining [71, Ch. 2] or cone polarity [72, Prop 7.1], in turn, may majorize the mean empirical width by simpler objects. Indeed, in Appendix 4.A.2 I adapt an argument of Tropp to arrive at the following estimate for the mean empirical width associated with the $\mathscr{G}$-TV regression decoder (P1-η). Notice its similarity to Lemma 4.2.2, even though the rows of $\mathbf{A}$ are not necessarily Gaussian random vectors.

Lemma 4.3.1 (Mean empirical width of $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$)**.** *Let $\mathscr{S} := \operatorname{supp}(\mathbf{Dx})$ be the support of $\mathbf{x}$ under the action of the analysis operator $\mathbf{D}\in\mathbb{R}^{N\times n}$, and define $\mathbf{P}_{\mathscr{S}}$ to be the orthogonal projection operator onto* $\operatorname{span}\{\mathbf{e}_i : i\in\mathscr{S}\}$*. Recall that, given m* i.i.d. *copies of a random vector $\mathbf{a}$, we define $\mathbf{h} := \frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i\mathbf{a}_i$ as their Rademacher average. Then, the following upper bound holds:*

$$W_m(\mathbf{a},\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x}))^2 \le \mathbb{E}\left(\|\mathbf{a}\|_2^2\right) - \left[\frac{\mathbb{E}\left(\|(\mathbf{I}_N-\mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1\right)}{\max\limits_{\|\mathbf{z}\|_2\le 1}\|\mathbf{Dz}\|_1}\right]^2. \tag{4.14}$$

This bound is manageable even when we consider the sampling matrices defined in Chapter 2. We only need an estimate for the second moment of the rows of $\mathbf{A}$, and we can borrow from the literature [37] ways to deal with the rightmost term in (4.14). Our coordinate sampling matrices only become a problem when dealing with the marginal tail function.

Koltchinskii and Mendelson [38] — and later Tropp [72] — state that the marginal tail function reflects the absolute continuity of the distribution of the random vector $\mathbf{v}$. Hence $Q_\xi(\mathbf{v},\cdot)$ may be quite small when the distribution of $\mathbf{v}$ is "spiky", an adjective we can certainly give to the sampling vectors used in this thesis. After all, the rows of $\mathbf{A}$ are drawn among the standard basis vectors in $\mathbb{R}^n$, so the distribution of our sampling vectors is supported on n points only. Meanwhile, the distribution of Gaussian vectors is supported on the whole of $\mathbb{R}^n$.

For a concrete example of how large the marginal tail function can be for absolutely continuous random vectors, consider the following proposition.

Proposition 4.3.2. *If* $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$*, then*

$$\lim_{\xi \to 0} \inf_{\mathbf{u} \in \mathbb{S}^{n-1}} \mathbb{P}\left(\left\{|\langle \mathbf{g}, \mathbf{u} \rangle| \geq \xi\right\}\right) = 1. \tag{4.15}$$

PROOF: It suffices to realize that $\langle \mathbf{g}, \mathbf{u} \rangle \sim \mathcal{N}(\mathbf{0}, \underbrace{\|\mathbf{u}\|_2^2}_{=1})$ and compute

$$\begin{aligned}
\lim_{\xi \to 0} \inf_{\mathbf{u} \in \mathbb{S}^{n-1}} \mathbb{P}\left(\left\{|\langle \mathbf{g}, \mathbf{u} \rangle| \geq \xi\right\}\right) &= \lim_{\xi \to 0} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathbb{1}_{\{|t| \geq \xi\}} e^{-t^2/2} dt \\
&= \lim_{\xi \to 0} \frac{2}{\sqrt{2\pi}} \int_{\xi}^{\infty} e^{-t^2/2} dt \qquad \texttt{(2× Gaussian tail)} \\
&= 1.
\end{aligned}$$

□

Now, contrast Proposition 4.3.2 with the next one concerning random vectors from the Coordinate Sampling with Replacement (CSWR($\boldsymbol{\pi}$)) model.

Proposition 4.3.3. *Let* $\mathbf{a} \in \mathbb{R}^n$ *be distributed as an arbitrary row of a matrix following the CSWR(*$\boldsymbol{\pi}$*) model. Then,*

$$\lim_{\xi \to 0} \inf_{\mathbf{u} \in \mathbb{S}^{n-1}} \mathbb{P}\left(\{|\langle \mathbf{a}, \mathbf{u} \rangle| \geq \xi\}\right) \leq \frac{1}{n}. \tag{4.16}$$

PROOF:

⟨1⟩1. The distribution of vector $\mathbf{a}$ is supported on the set of standard basis vectors $\{\mathbf{e}_i\}_{i=1}^n$. This distribution is parametrized as $\mathbb{P}(\{\mathbf{a} = \mathbf{e}_i\}) = \pi_i, \forall i \in [n]$, where $\sum_{i=1}^n \pi_i = 1$.

⟨1⟩2. Thus, $\mathbb{P}(\{|\langle \mathbf{a}, \mathbf{u} \rangle| \geq \xi\}) = \sum_{i=1}^n \mathbb{1}_{\{|\langle \mathbf{e}_i, \mathbf{u} \rangle| \geq \xi\}} \pi_i$.

⟨1⟩3. We can pick $\mathbf{u} \in \mathbb{S}^{n-1}$ orthogonal to all but one of the standard basis vectors. Pick this non-orthogonal vector as the one associated with the smallest sampling probability. Then,

$$\begin{aligned}
\lim_{\xi \to 0} \inf_{\mathbf{u} \in \mathbb{S}^{n-1}} \mathbb{P}\left(\{|\langle \mathbf{a}, \mathbf{u} \rangle| \geq \xi\}\right) &= \lim_{\xi \to 0} \inf_{\mathbf{u} \in \mathbb{S}^{n-1}} \sum_{i=1}^{n} \mathbb{1}_{\{|\langle \mathbf{e}_i, \mathbf{u} \rangle| \geq \xi\}} \pi_i \\
&= \min_{i \in [n]} \pi_i.
\end{aligned}$$

⟨1⟩4. Q.E.D.

The claim holds by noting that $\left(n \cdot \min_{i \in [n]} \pi_i\right) \leq \sum_{i=1}^n \pi_i = 1$.

Through Corollary 4.3.0.1, a marginal tail function as minuscule as in Proposition 4.3.3 would lead to a *vacuous* sample complexity — unless the mean empirical width were impractically small [6]. That is for most practical convex functions in (Pf-η), the corollary would guarantee

[6] The related Gaussian width for convex cones commonly used in recovery problems varies between $\mathcal{O}(n)$ and $\mathcal{O}(\log n)$ [4, Table 3.1]

robust recovery only if $m > n$, a number of coordinate samples larger than the total number of coordinates in the signal-to-be-recovered.

This is the main roadblock in using the standard small-ball method to arrive at the sample complexity for the decoders in this thesis. Still, the reason why this roadblock was reached might help in future attempts to certify coordinate-sampled convex recovery programs via the minimum q-gain functional. My unsatisfactory estimate for our marginal tail functional was a consequence of being able to pick a vector $\mathbf{u} \in \mathbb{S}^{n-1}$ orthogonal to all but one of the standard basis vectors in $\mathbb{R}^n$. This means, however, that this pick is itself one of the standard basis vectors. But in employing Proposition 4.3.3 I willfully ignored how the descent cone might restrict this choice [7]. If the vectors in $\mathscr{D}(f,\mathbf{x}) \cap \mathbb{S}^{n-1}$ are shown to be "far" from any given coordinate axis [8], we might be able to get a better lower bound on Q_ξ even when the small-ball condition does not strictly hold.

Indeed, take the toy example $\mathscr{D}(f,\mathbf{x}) \equiv \operatorname{span}(\mathbf{1})$. Then, for any $\mathbf{u} \in \mathscr{D}(f,\mathbf{x}) \cap \mathbb{S}^{n-1}$, we have $|\langle \mathbf{a}, \mathbf{u} \rangle| = 1/\sqrt{n}$. Hence, $\lim_{\xi \to 0} \inf_{\mathbf{u} \in \mathscr{D}(f,\mathbf{x}) \cap \mathbb{S}^{n-1}} \mathbb{P}(\{|\langle \mathbf{a}, \mathbf{u} \rangle| \geq \xi\}) = 1$, just as in the Gaussian case. In the next section I will further explore how the coordinate information of the descent cone could be used to bypass the shortcomings we found in the standard small-ball method.

4.4 Exploring the coordinate structure of the descent cone

I will focus in this section on the Bernoulli Sampling Model (Ber($\boldsymbol{\pi}$)) of independent vertex sampling (see Chapter 2). As a reminder, this model uses *i.i.d.* Bernoulli selectors to build a measurement matrix $\mathbf{A} = \sum_{i=1}^n \delta_i \mathbf{e}_i \mathbf{e}_i^\top$. The q-gain of $\mathbf{A}$ for any vector $\mathbf{u} \in \mathbb{R}^n$ is then a sum,

$$\|\mathbf{A}\mathbf{u}\|_q^q = \sum_{i=1}^n \delta_i |\langle \mathbf{e}_i, \mathbf{u} \rangle|^q = \sum_{i=1}^n \delta_i |u_i|^q, \tag{4.17}$$

of independent random variables whose expectation has the form

$$\mathbb{E}\left(\|\mathbf{A}\mathbf{u}\|_q^q\right) = \sum_{i=1}^n \mathbb{E}(\delta_i)\,|u_i|^q = \sum_{i=1}^n \pi_i |u_i|^q. \tag{4.18}$$

A simple application of the Bernstein inequality shows that $\|\mathbf{A}\mathbf{u}\|_q^q$ does not deviate too much from its expectation. The precise estimate — whose proof I put in Appendix 4.A.3 — is given the following lemma.

[7] We can safely ignore the descent cone in marginal tail function for random vectors satisfying a small-ball condition with large constants κ, δ (see Definition 4.3.1). In such cases, the effects of the measurement vectors and the convex objective on the minimum gain functional essentially decouple: Q_ξ is a sort of condition number for the measurements; W_m deals with the geometry of the descent cone.

[8] We could make this statement precise, as Mendelson does, by defining sets with "regular coordinate structure" [53].

Lemma 4.4.1. *Suppose that matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$*, with* $n \geq 2$ *rows comes from the Ber(*$\boldsymbol{\pi}$*) sampling model, and let* $\mathbf{u}$ *be any vector in* $\mathbb{R}^n$*. Set* $\tau := \sum_{i=1}^n \pi_i |u_i|^q$*. If* $\tau \geq \frac{32}{3n}\|\mathbf{u}\|_\infty^q \log\left(\frac{2}{\varepsilon}\right)$*, we then observe*

$$\frac{\tau}{2} \leq \|\mathbf{A}\mathbf{u}\|_q^q \leq \frac{3\tau}{2} \tag{4.19}$$

with probability at least $1-\varepsilon$*.*

The condition $\sum_{i=1}^n \pi_i |u_i|^q \geq \frac{32}{3n}\|\mathbf{u}\|_\infty^q \log\left(\frac{2}{\varepsilon}\right)$ has to do with the geometry of the set to which vector $\mathbf{u}$ belongs. If it holds, then $\mathbf{u}$ belongs — as Mendelson [53] would say — to a set of "regular coordinate structure", because enough coordinates in $\mathbf{u}$ are larger than some constant. Such vectors are in some sense "far" from the coordinate axes (when measuring distances by angle, for example). The set $\operatorname{span}(\mathbf{1})$, evoked in the end of the previous section, is an extreme example of a set with regular coordinate structure. Note in passing that the vertex probabilities, $\pi_1, \ldots, \pi_n$, in the condition, indicating where the sampling design may affect the recovery guarantees.

Still, Lemma 4.4.1 is not the end of the story; a direct certificate for problem (P1-η) is only obtained if we have a lower bound on $\inf_{\mathbf{u}\in\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})\cap\mathbb{S}^{n-1}} \|\mathbf{A}\mathbf{u}\|_q^q$. One possible line of attack towards this goal is an ϵ-net argument. That is, show that all points in $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})\cap\mathbb{S}^{n-1}$ are at most ϵ-far from a finite set $\mathscr{F}$ with regular coordinate structure, and then apply the previous lemma in a union bound over $\mathbf{u}\in\mathscr{F}$ [9].

But even then a question remains open: what finite set $\mathscr{F}$ with a regular coordinate structure forms an ϵ-net for $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})\cap\mathbb{S}^{n-1}$? Or, even more fundamentally, *when* is the intersection $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})\cap\mathbb{S}^{n-1}$ close to such an $\mathscr{F}$? Answers to these questions require a better geometric characterization of the descent cone induced by the $\mathscr{G}$-TV semi-norm for different classes of graphs and signals. I do not have this characterization, but looking for it should be an interesting endeavor.

4.5 Summary and further notes

This chapter delved into the possibility of providing a direct recovery certificate for the Graph Total Variation regression (P1-η) by lower-bounding a minimum q-gain functional.

Coordinate sampling ensembles such as ours are somewhat unusual in compressed sensing, so I chose to show first what would happen if we had Gaussian measurements instead. In this exercise, we saw how the size of the descent cone (via the conic Gaussian width) informs how many Gaussian vectors it takes to encode enough information about the ground-truth signal and ensure a robust recovery. Other notions of width appear in similar settings in the literature, induced by other specific classes of linear measurements. For instance, Sivakumar et al. [67]

[9] This is the strategy employed by Mendelson [53] in the context of sparse recovery problems from subsampled, random convolutions.

define a mean *exponential* width to deal with measurement vectors with sub-exponential tails. In any case, the conclusion is always the same: narrower cones lead to better sample complexities.

Mendelson's small-ball method has been involved in many success stories regarding lower-bounds for non-negative empirical processes like our minimum q-gain functional. But the small-ball condition of our coordinate sampling vectors is very poor. It predicts a marginal tail function that is too small and finally leads to a vacuous sample complexity for practical convex decoders. I should mention that Mendelson has kept building upon the original method. A somewhat recent improvement was replacing the small-ball assumption by a *stable lower bound condition* [51], extending the scope of empirical processes can be dealt with.

By the end of this chapter I can but point towards a direction where the $\mathscr{G}$-TV regression (P1-η) might be given a direct recovery certificate. The path requires a better understanding of the geometry of the descent cone $\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})$, especially in what concerns the coordinate structure of the set $\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x}) \cap \mathbb{S}^{n-1}$. But this characterization I leave as an open problem.

In the next chapter, I will finally show a recovery certificate – even if only for the noiseless, interpolation program (P1). An optimal sampling design will then be revealed as the one that minimizes the number of samples the certificate demands.

Appendix 4.A Proofs

4.A.1 Proof of Proposition 4.3.1

In this section, $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a random matrix whose rows, $\mathbf{a}_1, \dots, \mathbf{a}_m$, are *i.i.d.* copies of a random vector $\mathbf{a}$. The reader should recall Definitions 4.1.1, 4.3.2, and 4.3.3 for the minimum q-gain, the marginal tail function, and the mean empirical width, respectively. The minimum q-gain of $\mathbf{A}$, restricted to a set $\mathscr{S}$, will be seen as a non-negative empirical process induced by the random vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$:

$$\gamma_{\min}^{(q)}(\mathscr{S}, \mathbf{A}) = \inf_{\mathbf{u}\in\mathscr{S}\cap \mathrm{bd}(\mathbb{B}_q^n)} \|\mathbf{A}\mathbf{u}\|_q^q = \inf_{\mathbf{u}\in\mathscr{S}\cap \mathrm{bd}(\mathbb{B}_q^n)} \sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{u}\rangle|^q .$$

I will then show that for any constants $\xi, t > 0$, and with probability larger than $1 - \exp\left(\frac{-t^2}{2}\right)$, the lower bound

$$\gamma_{\min}^{(q)}(\mathscr{S}, \mathbf{A}) \geq m^{\frac{2-q}{2q}} \left[\xi\sqrt{m} Q_\xi(\mathbf{a}, \mathscr{S}) - 2W_m(\mathbf{a}, \mathscr{S}) - \xi t\right]$$

takes place. The argument is taken from Tropp [72, Sec. 2.5.5].

PROOF:

⟨1⟩1. Use, successively, the Lyapunov and Markov inequalities to reach the starting lower bound

$$\left(\frac{1}{m}\sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{u}\rangle|^q\right)^{1/q} \geq \frac{1}{m}\sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq \frac{\xi}{m}\sum_{i=1}^m \mathbb{1}_{\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq \xi\}}.$$

⟨1⟩2. Add and subtract $\mathbb{P}(\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq 2\xi\})$ on the RHS of the inequality in the previous step. Then take the infimum over $\mathscr{S}$ on both sides:

$$\inf_{\mathbf{u}\in\mathscr{S}} \left(\frac{1}{m}\sum_{i=1}^m |\langle \mathbf{a}_i, \mathbf{u}\rangle|^q\right)^{1/q} \geq \xi \overbrace{\inf_{\mathbf{u}\in\mathscr{S}} \mathbb{P}(\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq 2\xi\})}^{=:Q_{2\xi}(\mathbf{a},\mathscr{S})} - \frac{\xi}{m}\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^m \left[\mathbb{P}(\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq 2\xi\}) - \mathbb{1}_{\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq \xi\}}\right]$$

⟨1⟩3. Control the supremum on the new RHS using the bounded differences inequality [8, Sec. 6.1]. This is possible because the summands inside the supremum are independent and bounded in magnitude by one. With probability larger than $1 - \exp\left(-t^2/2\right)$, we then have

$$\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^m \left[\mathbb{P}(\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq 2\xi\}) - \mathbb{1}_{\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq \xi\}}\right] \leq \mathbb{E}\left(\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^m \left[\mathbb{P}(\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq 2\xi\}) - \mathbb{1}_{\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq \xi\}}\right]\right) + t\sqrt{m}.$$

⟨1⟩4. It remains to bound the expected supremum to the right. Let $\varepsilon_1, \dots, \varepsilon_m$ be *i.i.d.* copies of a Rademacher random variable; I claim that the following holds:

$$\mathbb{E}\left(\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^{m} \left[\mathbb{P}\left(\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq 2\xi\}\right) - \mathbb{1}_{\{|\langle \mathbf{a}_i,\mathbf{u}\rangle|\geq\xi\}}\right]\right) \leq \frac{2}{\xi}\mathbb{E}\left(\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^{m} \varepsilon_i \langle \mathbf{a}_i, \mathbf{u}\rangle\right),$$

PROOF:

⟨2⟩1. Define a "soft" indicator function $\psi_\xi : \mathbb{R} \to [0,1]$ (see Figure 4.1) as the map

$$s \mapsto \psi_\xi(s) := \begin{cases} 0, & |s| \leq \xi \\ \frac{|s|-\xi}{\xi}, & \xi < |s| \leq 2\xi \\ 1, & |s| > 2\xi \end{cases}$$

We will need two, easily-verifiable properties of this function. First, it is "sand-

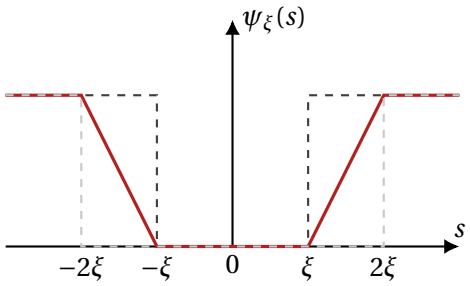


Figure 4.1 – "Soft" indicator function

wiched" by two indicator functions, $\mathbb{1}_{\{|\cdot|\geq 2\xi\}} \leq \psi_\xi(\cdot) \leq \mathbb{1}_{\{|\cdot|\geq\xi\}}$. Second, the product map $s \mapsto \xi\psi_\xi(s)$ is a contraction.

⟨2⟩2. Use the soft indicator function — and its properties — to apply a symmetrization procedure [47, Lemma 6.3], and then the Rademacher comparison principle [47, Theorem 4.12] to the expected supremum. As a result, the claim is proved:

$$\begin{aligned}
&\mathbb{E}\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^{m} \left[\mathbb{P}\left(\{|\langle \mathbf{a}_i, \mathbf{u}\rangle| \geq \xi\}\right) - \mathbb{1}_{\{|\langle \mathbf{a}_i,\mathbf{u}\rangle|\geq\xi\}}\right] \\
&\qquad = \mathbb{E}\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^{m} \left[\mathbb{E}\mathbb{1}_{\{|\langle \mathbf{a}_i,\mathbf{u}\rangle|\geq 2\xi\}} - \mathbb{1}_{\{|\langle \mathbf{a}_i,\mathbf{u}\rangle|\geq\xi\}}\right] \\
&\qquad (\text{``sandwiched''}\ \psi_\xi) \leq \mathbb{E}\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^{m} \left[\mathbb{E}\psi_\xi(\langle \mathbf{a}_i, \mathbf{u}\rangle) - \psi_\xi(\langle \mathbf{a}_i, \mathbf{u}\rangle)\right] \\
&\qquad (\texttt{symmetrization}) \leq 2\mathbb{E}\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^{m} \varepsilon_i \psi_\xi(\langle \mathbf{a}_i, \mathbf{u}\rangle) \\
&\qquad (\texttt{contraction of}\ \xi\psi_\xi) \leq \frac{2}{\xi} \underbrace{\mathbb{E}\sup_{\mathbf{u}\in\mathscr{S}} \sum_{i=1}^{m} \varepsilon_i \langle \mathbf{a}_i, \mathbf{u}\rangle}_{=:\sqrt{m}W_m(\mathbf{a},\mathscr{S})}
\end{aligned}$$

□

⟨1⟩5. Q.E.D.

The desired lower bound on the minimim q-gain functional arises by combining steps ⟨1⟩2, ⟨1⟩3 and ⟨1⟩4:

$$\gamma_{\min}^{(q)}(\mathscr{S},\mathbf{A}) := \inf_{\mathbf{u}\in\mathscr{S}} \left(\frac{1}{m}\sum_{i=1}^{m} |\langle \mathbf{a}_i,\mathbf{u}\rangle|^q\right)^{1/q} \geq m^{\frac{2-q}{2q}} \left[\xi\sqrt{m}Q_\xi(\mathbf{a},\mathscr{S}) - 2W_m(\mathbf{a},\mathscr{S}) - \xi t\right].$$

4.A.2 Proof of Lemma 4.3.1

As a reminder, I use the symbols $\mathscr{S} := \operatorname{supp}(\mathbf{Dx})$ for the jump-set of $\mathbf{x}$, $\mathbf{P}_{\mathscr{S}} := \sum_{i\in\mathscr{S}} \mathbf{e}_i\mathbf{e}_i^\top$ for the corresponding orthogonal projection operator, and $\mathbf{h} := \frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i\mathbf{a}_i$ for the Rademacher average of vectors $\mathbf{a}_1,\ldots,\mathbf{a}_m$. I will prove the bound

$$W_m(\mathbf{a},\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x}))^2 \leq \mathbb{E}\left(\|\mathbf{a}\|_2^2\right) - \left[\frac{\mathbb{E}\left(\|(\mathbf{I}_N-\mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1\right)}{\max\limits_{\|\mathbf{z}\|_2\leq 1}\|\mathbf{D}\mathbf{z}\|_1}\right]^2$$

for the mean empirical width of the $\mathscr{G}$-TV descent cone, as measured by *i.i.d.* copies, $\mathbf{a}_1,\ldots,\mathbf{a}_m$, of a random vector $\mathbf{a}$.

The argument — an adaptation of [37, Lem. 1 & Thm. 3] — works by relating the descent cone $\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x})$ to the subdifferential

$$\partial\|\mathbf{D}\cdot\|_1(\mathbf{x}) := \left\{\mathbf{v}\in\mathbb{R}^n : \langle \mathbf{v},\mathbf{z}-\mathbf{x}\rangle \leq \|\mathbf{D}\mathbf{z}\|_1 - \|\mathbf{D}\mathbf{x}\|_1, \forall \mathbf{z}\in\mathbb{R}^n\right\}. \tag{4.20}$$

In particular, the defining expressions of the subdifferential are called upon (see Proposition 3.1.2).

PROOF:

⟨1⟩1. Tropp [72, Prop. 7.1] gives us the initial upper bound

$$W_m(\mathbf{a},\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x}))^2 \leq \mathbb{E}\left(\inf_{\tau\geq 0}\inf_{\mathbf{v}\in\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})}\|\mathbf{h}-\tau\mathbf{v}\|_2^2\right)$$

⟨1⟩2. We can exchange the expectation and the first infimum:

$$\mathbb{E}\left(\inf_{\tau\geq 0}\inf_{\mathbf{v}\in\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})}\|\mathbf{h}-\tau\mathbf{v}\|_2^2\right) \leq \inf_{\tau\geq 0}\mathbb{E}\left(\inf_{\mathbf{v}\in\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})}\|\mathbf{h}-\tau\mathbf{v}\|_2^2\right)$$

⟨1⟩3. Consider a fixed $\tau\geq 0$. I claim that $\mathbb{E}\left(\inf\limits_{\mathbf{v}\in\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})}\|\mathbf{h}-\tau\mathbf{v}\|_2^2\right) \leq \mathbb{E}\left(\|\mathbf{h}\|_2^2\right) - \left[\frac{\mathbb{E}(\|(\mathbf{I}_N-\mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1)}{\max\limits_{\|\mathbf{z}\|_2\leq 1}\|\mathbf{D}\mathbf{z}\|_1}\right]^2$

PROOF:

⟨2⟩1. Start with the following characterization from subdifferential calculus:

$$\begin{aligned}\mathbf{v} \in \partial\|\mathbf{D}\cdot\|_1(\mathbf{x}) &\iff \mathbf{v} \in \mathbf{D}^\top \partial\|\cdot\|_1(\mathbf{D}\mathbf{x}) \\ &\iff \exists \mathbf{u} \in \mathbb{B}_\infty^N : \mathbf{v} = \mathbf{D}^\top \left[\underbrace{\mathbf{P}_{\mathscr{S}} \operatorname{sign}(\mathbf{D}\mathbf{x})}_{=\operatorname{sign}(\mathbf{D}\mathbf{x})} + (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u} \right]\end{aligned}$$

⟨2⟩2. Note that $\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})$ is a compact set, because it is the linear image of points in $\mathbb{B}_\infty^N$ by the finite-dimensional linear operator $\mathbf{D}^\top$.

⟨2⟩3. Therefore, there exists at least one $\mathbf{v}^\star$ in $\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})$ for which the maximum $\max_{\mathbf{v}\in\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})} \langle \mathbf{h}, \mathbf{v}\rangle =: \langle \mathbf{h}, \mathbf{v}^\star\rangle$ is achieved.

⟨2⟩4. Fix this $\mathbf{v}^\star \in \partial\|\mathbf{D}\cdot\|_1(\mathbf{x})$, and compute

$$\begin{aligned}\inf_{\mathbf{v}\in\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})} \|\mathbf{h} - \tau\mathbf{v}\|_2^2 &\le \|\mathbf{h}\|_2^2 - \tau\langle \mathbf{h}, \mathbf{v}^\star\rangle + \tau^2\|\mathbf{v}^\star\|_2^2 \\ &= \|\mathbf{h}\|_2^2 - 2\tau \max_{\mathbf{v}\in\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})} \langle \mathbf{h}, \mathbf{v}\rangle + \tau^2 \max_{\|\mathbf{z}\|_2\le 1} \langle \mathbf{z}, \mathbf{v}^\star\rangle \\ &= \|\mathbf{h}\|_2^2 - 2\tau\langle \mathbf{h}, \mathbf{D}^\top \operatorname{sign}(\mathbf{D}\mathbf{x})\rangle - 2\tau \max_{\mathbf{u}\in\mathbb{B}_\infty^N} \left\langle \mathbf{h}, \mathbf{D}^\top(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}\right\rangle \\ &\qquad + \tau^2 \max_{\|\mathbf{z}\|_2\le 1} \left\langle \mathbf{z}, \underbrace{\mathbf{D}^\top\left[\operatorname{sign}(\mathbf{D}\mathbf{x}) + (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}^\star\right]}_{(\texttt{For some } \mathbf{u}^\star \in \mathbb{B}_\infty^N)} \right\rangle \\ &= \|\mathbf{h}\|_2^2 - 2\tau\langle \mathbf{h}, \mathbf{D}^\top \operatorname{sign}(\mathbf{D}\mathbf{x})\rangle - 2\tau\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1 \\ &\qquad + \tau^2 \max_{\|\mathbf{z}\|_2\le 1} \left\langle \mathbf{D}\mathbf{z}, \left[\operatorname{sign}(\mathbf{D}\mathbf{x}) + (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}^\star\right]\right\rangle \\ (\texttt{Hölder ineq.}) &\le \|\mathbf{h}\|_2^2 - 2\tau\langle \mathbf{h}, \mathbf{D}^\top \operatorname{sign}(\mathbf{D}\mathbf{x})\rangle - 2\tau\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1 \\ &\qquad + \tau^2 \max_{\|\mathbf{z}\|_2\le 1} \|\mathbf{D}\mathbf{z}\|_1 \cdot \underbrace{\left\|\operatorname{sign}(\mathbf{D}\mathbf{x}) + (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}^\star\right\|_\infty}_{\le 1}\end{aligned}$$

⟨2⟩5. Recall that $\mathbb{E}(\mathbf{h}) = \mathbf{0}$, so taking the expectation on both sides of the previous inequality yields

$$\mathbb{E}\left(\inf_{\mathbf{v}\in\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})} \|\mathbf{h} - \tau\mathbf{v}\|_2^2\right) \le \|\mathbf{h}\|_2^2 - 2\tau\mathbb{E}\left(\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1\right) + \tau^2 \max_{\|\mathbf{z}\|_2\le 1} \|\mathbf{D}\mathbf{z}\|_1.$$

⟨2⟩6. Calculus tells us that $\tau = \frac{\mathbb{E}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1}{\max_{\|\mathbf{z}\|_2\le 1} \|\mathbf{D}\mathbf{z}\|_1^2}$ is the minimizer of the RHS above. Replace this value of τ onto the inequality to reach the claim. □

⟨1⟩4. Finally, check that $\mathbb{E}\left(\|\mathbf{h}\|_2^2\right) = \mathbb{E}\left(\|\mathbf{a}\|_2^2\right)$.

PROOF: By direct calculation,

$$\mathbb{E}\left(\|\mathbf{h}\|_2^2\right) := \mathbb{E}\left(\left\|\frac{1}{\sqrt{m}}\sum_{i=1}^m \varepsilon_i \mathbf{a}_i\right\|_2^2\right)$$

$$
\begin{aligned}
&= \frac{1}{m}\sum_{i,j}\mathbb{E}\left(\varepsilon_i\varepsilon_j\right)\mathbb{E}\left(\langle \mathbf{a}_i,\mathbf{a}_j\rangle\right)\\
&\quad \texttt{(Independence of } \mathbf{a} \texttt{ and } \varepsilon\texttt{)}\\
&= \frac{1}{m}\left[\sum_{i\neq j}\underbrace{\mathbb{E}\left(\varepsilon_i\varepsilon_j\right)}_{=\mathbb{E}(\varepsilon_i)\mathbb{E}(\varepsilon_j)=0}\mathbb{E}\left(\langle \mathbf{a}_i,\mathbf{a}_j\rangle\right)\right.\\
&\qquad \left.+\sum_{i=1}^{m}\underbrace{\mathbb{E}\left(\varepsilon_i^2\right)}_{=1}\mathbb{E}\left(\langle \mathbf{a}_i,\mathbf{a}_i\rangle\right)\right]\\
&\quad \texttt{(Independence and zero mean of } \varepsilon_1,\ldots,\varepsilon_m\texttt{)}\\
&= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}\left(\|\mathbf{a}_i\|_2^2\right)\\
&= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}\left(\|\mathbf{a}\|_2^2\right)\\
&\quad \texttt{(Identical distribution of } \mathbf{a}_1,\ldots,\mathbf{a}_m\sim\mathbf{a}\texttt{)}\\
&= \mathbb{E}\left(\|\mathbf{a}\|_2^2\right)\ \square
\end{aligned}
$$

⟨1⟩5. Q.E.D.

Join the estimates in each step to unveil $W_m(\mathbf{a},\mathscr{D}(\|\mathbf{D}\cdot\|_1,\mathbf{x}))^2 \leq \mathbb{E}\left(\|\mathbf{a}\|_2^2\right) - \left[\frac{\mathbb{E}(\|(\mathbf{I}_N-\mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1)}{\max_{\|\mathbf{z}\|_2\leq 1}\|\mathbf{D}\mathbf{z}\|_1}\right]^2$, the desired upper bound.

4.A.3 Proof of Lemma 4.4.1

For $\mathbf{A}\in\mathbb{R}^{n\times n}$ coming from the Ber($\boldsymbol{\pi}$) sampling model, and $\mathbf{u}\in\mathbb{R}^n$, I claim that

$$\|\mathbf{A}\mathbf{u}\|_q^q \in \left[\frac{\tau}{2},\frac{3\tau}{2}\right]$$

with likelihood larger than $1-\varepsilon$, as long as $\tau := \sum_{i=1}^n \pi_i|u_i|^q \geq \frac{32}{3n}\|\mathbf{u}\|_\infty^q \log\left(\frac{2}{\varepsilon}\right)$.

PROOF: We only need to apply the scalar Bernstein inequality (Appendix A) to the random deviation $\|\mathbf{A}\mathbf{u}\|_q^q - \mathbb{E}\|\mathbf{A}\mathbf{u}\|_q^q$.

⟨1⟩1. Set $X := \|\mathbf{A}\mathbf{u}\|_q^q - \mathbb{E}\left(\|\mathbf{A}\mathbf{u}\|_q^q\right) = \sum_{i=1}^n \underbrace{(\delta_i-\pi_i)|u_i|^q}_{=:X_i}$. The scalar X is a sum of independent, zero-mean random variables X_i, because the Bernoulli selectors $\{\delta_i\}_{i=1}^n$ are independent.

⟨1⟩2. Each term in the sum is upper bounded almost surely as

$$|X_i| = \underbrace{|\delta_i-\pi_i|}_{\leq 1}|u_i|^q \leq \max_{i\in[n]}|u_i|^q = \|\mathbf{u}\|_\infty^q =: B.$$

⟨1⟩3. We can also write an upper bound for their second moments as

$$\mathbb{E}\left(X_i^2\right) = \mathbb{E}\left((\delta_i - \pi_i)^2\right)|u_i|^{2q} = \pi_i(1-\pi_i)|u_i|^{2q} \le \pi_i|u_i|^{2q} \le B\pi_i|u_i|^q.$$

and set $\sigma^2 := \frac{B}{n}\sum_{i=1}^n \pi_i|u_i|^q$.

⟨1⟩4. The scalar Bernstein inequality in Lemma A.0.1, along with the moment estimates above, give us the deviation probability

$$\begin{aligned}\mathbb{P}\left(\left\{\left|\|\mathbf{Au}\|_q^q - \mathbb{E}\left(\|\mathbf{Au}\|_q^q\right)\right| > t\right\}\right) &\le 2\exp\left(-\frac{3n}{8}\min\left\{\frac{t^2}{\sigma^2}, \frac{t}{B}\right\}\right) \\ &= 2\exp\left(-\frac{3n}{8}\frac{t}{\|\mathbf{u}\|_\infty^q}\min\left\{\frac{nt}{\sum_{i=1}^n \pi_i|u_i|^q}, 1\right\}\right).\end{aligned}$$

⟨1⟩5. Picking $t = \frac{1}{2}\sum_{i=1}^n \pi_i|u_i|^q =: \frac{1}{2}\mathbb{E}\|\mathbf{Au}\|_q^q$, we have $\frac{nt}{\sum_{i=1}^n \pi_i|u_i|^q} \ge 1$, as long as $n \ge 2$. With this choice, the deviation inequality becomes

$$\mathbb{P}\left(\left\{\left|\|\mathbf{Au}\|_q^q - \mathbb{E}\left(\|\mathbf{Au}\|_q^q\right)\right| > \frac{1}{2}\sum_{i=1}^n \pi_i|u_i|^q\right\}\right) \le 2\exp\left(-\frac{3n}{32}\frac{\sum_{i=1}^n \pi_i|u_i|^q}{\|\mathbf{u}\|_\infty^q}\right).$$

⟨1⟩6. Q.E.D.

The claim holds by setting the RHS of the last inequality to at most ε.

5 Dual certificates: KKT conditions and the golfing scheme

As we exit our search for a direct recovery certificate — cut short by lack of coordinate information on $\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})$ —, this chapter looks for an alternative in the polar opposite of the descent cone. The subdifferential $\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})$ appears when considering the Karush-Kuhn-Tucker (KKT) conditions for the solutions of

$$\min_{\mathbf{z}\in\mathbb{R}^n}\|\mathbf{Dz}\|_1 \text{ such that } \mathbf{Ax} = \mathbf{Az}. \tag{P1}$$

One of these conditions relies on the existence of a certain *dual* vector $\mathbf{u} \in \mathbb{R}^N$ living in the co-domain of the linear transformation represented by $\mathbf{D}$. The very existence of the dual vector $\mathbf{u}$ can be seen as a recovery certificate for (P1); the hard task is *proving* that such a vector exists. Nevertheless, I show how to use the KKT conditions as a blueprint for an iterative scheme producing approximations of the dual vector that are still valid certificates. This idea gives rise to *a version* of the golfing scheme [28], popular in Compressed Sensing, and whose convergence depends on well-behaving tails in random matrices born of the interaction between $\mathbf{D}$ and $\mathbf{A}$. Although powerful, this scheme(by its very construction) applies only to the noiseless, interpolation problem (P1). [1]

In the end, I can reach a sample complexity threshold for the $\mathscr{G}$-TV interpolation under Coordinate Sampling with Replacement (CSWR($\boldsymbol{\pi}$)). More importantly, this threshold explicitly depends on the sampling probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ of CSWR($\boldsymbol{\pi}$). The corresponding optimal design is then just a corollary of sample complexity result, achieved by minimizing the threshold level with respect to $\boldsymbol{\pi}$. Although simple to state, the optimal sampling design is difficult to evaluate in practice, but some approximations of it are examined in the next chapter.

[1] This restriction is the main downside of the certificates in this section, as compared to the ones that we could have obtained in the previous chapter.

5.1 Lagrange dual problem and the KKT conditions

The method of Lagrange multipliers [9, Ch. 5] gives us a dual perspective on problem (P1). First, consider augmenting its objective function in the following way. Let $\boldsymbol{\nu} = (\nu_1, \dots, \nu_m) \in \mathbb{R}^m$ be a vector with an entry for each of the m implicit equations in $\mathbf{Az} = \mathbf{Ax}$. The numbers $\nu_1, \dots, \nu_m$ will act as the Lagrange multipliers for the equality constraint $\mathbf{Az} - \mathbf{Ax} = \mathbf{0}$. The multipliers augment the objective through the map

$$\mathbf{z}, \boldsymbol{\nu} \mapsto \mathfrak{L}(\mathbf{z}, \boldsymbol{\nu}) := \|\mathbf{Dz}\|_1 + \langle \boldsymbol{\nu}, \mathbf{Az} - \mathbf{Ax} \rangle , \tag{5.1}$$

The function $\mathfrak{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is then deemed the Lagrangian of the problem. Second, use the Lagrangian to define the dual problem [2]

$$\max_{\boldsymbol{\nu} \in \mathbb{R}^m} \min_{\mathbf{z} \in \mathbb{R}^n} \|\mathbf{Dz}\|_1 + \langle \boldsymbol{\nu}, \mathbf{Az} - \mathbf{Ax} \rangle , \tag{P1-dual}$$

whose objective has an optimal value identical to the one of (P1) [3].

The just-defined (P1-dual) is a saddle-point problem, convex in $\mathbf{z}$ and concave [4] in $\boldsymbol{\nu}$. Variational analysis [63, Thm. 8.15] tells us that $(\mathbf{z}^\star, \boldsymbol{\nu}^\star)$ is the corresponding saddle-point (or optimal pair) if the inclusions

$$\mathbf{0} \in \partial_{\boldsymbol{\nu}} \mathfrak{L}(\mathbf{z}^\star, \boldsymbol{\nu}^\star), \text{ and} \tag{5.2}$$

$$\mathbf{0} \in \partial_{\mathbf{z}} \mathfrak{L}(\mathbf{z}^\star, \boldsymbol{\nu}^\star) \tag{5.3}$$

take place [5]. The Lagrangian is differentiable with respect to $\boldsymbol{\nu}$, so we can unpack (5.2) as

$$\begin{aligned} \mathbf{0} \in \partial_{\boldsymbol{\nu}} \mathfrak{L}(\mathbf{z}^\star, \boldsymbol{\nu}^\star) &\iff \mathbf{0} = \nabla_{\boldsymbol{\nu}} \left\{ \|\mathbf{Dz}^\star\|_1 + \left\langle \cdot, \mathbf{Az}^\star - \mathbf{Ax} \right\rangle \right\} (\boldsymbol{\nu}^\star) \\ &\iff \mathbf{0} = \mathbf{Az}^\star - \mathbf{Ax}. \end{aligned}$$

For the second inclusion, we have to deal with the subdifferential of $\|\mathbf{D}\cdot\|_1$ at $\mathbf{z}^\star$. It helps to recall this set's defining expressions from Proposition 3.1.2, which let us write $\mathbf{z} \in \partial \|\mathbf{D}\cdot\|_1(\mathbf{z}^\star)$ $\iff \mathbf{z} = \mathbf{D}^\top \left[\operatorname{sign}(\mathbf{Dx}) + (\mathbf{I}_N + \mathbf{P}_{\mathscr{S}})\, \mathbf{u} \right]$ for some $\mathbf{u} \in \mathbb{R}^N$ satisfying $\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}\|_\infty \leq 1$. As a result, we can read (5.3) as

$$\begin{aligned} \mathbf{0} \in \partial_{\mathbf{z}} \mathfrak{L}(\mathbf{z}^\star, \boldsymbol{\nu}^\star) &\iff \mathbf{0} = \partial \|\mathbf{D}\cdot\|_1(\mathbf{z}^\star) + \nabla_{\mathbf{z}} \left\{ \left\langle \boldsymbol{\nu}^\star, \mathbf{A}\cdot - \mathbf{Ax} \right\rangle \right\} (\mathbf{z}^\star) \\ &\iff \mathbf{0} = \mathbf{D}^\top \left[\operatorname{sign}(\mathbf{Dx}) + (\mathbf{I}_N + \mathbf{P}_{\mathscr{S}})\, \mathbf{u}^\star \right] + \mathbf{A}^\top \boldsymbol{\nu}^\star, \end{aligned}$$

where $\mathbf{u}^\star$ is such that $\mathbf{z}^\star = \mathbf{D}^\top \mathbf{u}^\star$ *and* $\left\| (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}) \mathbf{u}^\star \right\|_\infty \leq 1$.

[2] Taking (P1) as the reference, *primal* problem.

[3] This is guaranteed because the primal problem is convex and Slater's condition is satisfied: there is at least one strictly feasible point $\mathbf{z}$, namely $\mathbf{z} \equiv \mathbf{x}$, belonging to the relative interior of the primal's objective [9, Sec. 5.2.3].

[4] It is actually *linear* in $\boldsymbol{\nu}$, hence both convex *and* concave.

[5] In Calculus, this is analogous to finding the critical points of a differentiable function in the places where the derivative vanishes.

Together, the unpacked saddle-point expressions form the so-called Karush-Kuhn-Tucker (KKT) conditions for the optimality of the primal-dual pair $(\mathbf{z}^\star, \boldsymbol{\nu}^\star)$:

$$\mathbf{A}\mathbf{z}^\star = \mathbf{A}\mathbf{x}$$

and

$$\mathbf{z}^\star = \mathbf{A}^\top \boldsymbol{\nu}^\star = \mathbf{D}^\top \mathbf{u}^\star \;:\; \begin{cases} \mathbf{P}_{\mathscr{S}} \mathbf{u}^\star = \operatorname{sign}(\mathbf{D}\mathbf{x}) \\ \left\| (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}) \mathbf{u}^\star \right\|_\infty \leq 1 \end{cases} .$$

The first of these just restates the interpolation constraint; the second lists the ingredients we will need in the rest of the chapter. Problem (P1) is only succesful if each of its optimal points $\mathbf{z}^\star$ is identical to $\mathbf{x}$. In this case the first of the KKT conditions is trivially satisfied; let us focus on the second one then.

Note that not much is demanded of the optimal dual point $..bmv^\star$: it just needs to be different than zero, lest $\mathbf{z}^\star$ be zero as well. The equation $\mathbf{z}^\star = \mathbf{A}^\top \boldsymbol{\nu}^\star$, in other words, just says that $\mathbf{z}^\star$ should be in the range of $\mathbf{A}^\top$. The only degree of freedom left is vector the $\mathbf{u}^\star$. If there is some such vector simultaneously satisfying

$$\mathbf{D}^\top \mathbf{u}^\star \in \operatorname{range}\left(\mathbf{A}^\top\right), \tag{5.4}$$

$$\mathbf{P}_{\mathscr{S}} \mathbf{u}^\star = \operatorname{sign}(\mathbf{D}\mathbf{x}) \tag{5.5}$$

$$\left\| (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}) \mathbf{u}^\star \right\|_\infty \leq 1, \tag{5.6}$$

and $\mathbf{x} = \mathbf{D}^\top \mathbf{u}^\star$, then it certifies $\mathbf{x}$ as the solution of (P1). We suddenly see in $\mathbf{u}^\star$ the (dual) certificate we had been looking for.

But before we rush, remember the random nature of the sampling matrix $\mathbf{A}$. It turns impractical the search for a vector $\mathbf{u}^\star$ satisfying an equality constraint like $\mathbf{P}_{\mathscr{S}} \mathbf{u}^\star = \operatorname{sign}(\mathbf{D}\mathbf{x})$, *while* having a deterministic image $\mathbf{D}^\top \mathbf{u}^\star$ that lies in a random subspace. Fortunately, I show in the next section how some points $\mathbf{u} \in \mathbb{R}^N$ can be enough of a recovery certificate despite $\mathbf{P}_{\mathscr{S}} \mathbf{u}$ merely approximating $\operatorname{sign}(\mathbf{D}\mathbf{x})$. This relaxation does not come — of course — without a cost, for it demands a more precise control over the interplay between operators $\mathbf{D}$ and $\mathbf{A}$. Still, it pays off, later on, when the defining expressions for these inexact certificates are turned into the blueprint for an effective golfing scheme.

5.2 Inexact dual certificates for $\mathscr{G}$-TV interpolation

Inexact dual certificates are a staple of exact recovery studies in Compressed Sensing, especially when the measurement ensemble is "structured" [6] [3, 10, 13]. Ever since Candès and Plan [13], the scope of such certificates has been incrementally extended. Using the form of (P1) as a template, we can say that inexact dual certificates were initially shown to exist only when the

[6] As opposed to measurement ensembles like the Gaussian which are considered "unstructured"

sparsifying transform $\mathbf{D}$ was the identity. Then, other proofs started admitting programs with tight frames [11] or injective operators [48]. From the perspective of the measurement matrix, the traditionally imposed restraints regarded the covariance structure of $\mathbf{A}$. It was only after Kueng and Gross [42] that non-isotropic [7] measurements could be dealt with.

The lemma that I present next generalizes the progress discussed in the historical account of the previous paragraph. It shows that inexact dual vectors can certify (P1) for a large class of matrices $\mathbf{D}$ and $\mathbf{A}$. In particular, the analysis operator $\mathbf{D}$ need not be injective, as long as its null space intersects trivially with the one of $\mathbf{A}$ [8]. Moreover, the random properties of $\mathbf{A}$ are allowed to be regularized by a free parameter in the form of a matrix $\mathbf{B}$.

Lemma 5.2.1 (Inexact Dual Certificate)**.** *Set* $\mathbf{M} := \left[\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]^\top$*, using some matrix* $\mathbf{B} \in \mathbb{R}^{m\times n}$*, and let* $\mathbf{u}$ *be some vector in* $\mathbb{R}^N$*. The point* $\mathbf{x} \in \mathbb{R}^n$ *is certified by* $\mathbf{u}$ *to be the unique solution of* (P1) *if all of the following hold:*

$$\operatorname{null}(\mathbf{D}) \cap \operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}, \tag{A1}$$

$$\|\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\|_2 \le 1/3, \tag{A2}$$

$$\max_{k\notin\mathscr{S}} \left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^\top \mathbf{e}_{\mathbf{k}}\right\|_2 \le 1, \tag{A3}$$

$$\mathbf{D}^\top \mathbf{u} \in \operatorname{range}\left(\mathbf{A}^\top\right), \tag{A4}$$

$$\|\mathbf{P}_{\mathscr{S}}(\mathbf{u} - \operatorname{sign}(\mathbf{Dx}))\|_2 \le 1/3, \tag{A5}$$

$$\|(\mathbf{I_N} - \mathbf{P}_{\mathscr{S}})\mathbf{u}\|_\infty \le 1/3. \tag{A6}$$

To prove this result, I took inspiration from Boyer *et al.* [10, Appendix A]. The reader can find the full argument in Appendix 5.A.1, but here is the gist of it.

PROOF SKETCH: If a perturbation $\mathbf{z} = \mathbf{x} + \mathbf{h}$ is a solution of P1, and the assumptions above hold, I argue then that $\mathbf{h} \equiv \mathbf{0}$ in order to avoid the contradiction $\|\mathbf{Dz}\|_1 > \|\mathbf{Dx}\|_1$. This line of reasoning follows closely the seminal proof in Candès and Plan [13, Lemma 3.2], but I make the necessary adaptations due to the non-trivial null space of $\mathbf{D}$.

As might be expected from the proof sketch, if we set $\mathbf{D} = \mathbf{I}_n$ and $\mathbf{B} = \mathbf{A}^\top$ then Lemma 5.2.1 reduces to Candès and Plan [13, Lemma 3.2]. The more recent work of Lee *et al.*has a similar-looking statement [48, Lemma 21], that is nonetheless only valid for $\mathbf{D}$ injective and $\mathbf{A}$ such that $\mathbf{A}^\top\mathbf{A}$ is a projection matrix. This can also be seen as a specialization of Lemma 5.2.1, since the injectivity of $\mathbf{D}$ turns Assumption A1 trivial.

By the way, let me remark on some things about the conditions in the lemma. Readers will correctly identify assumptions A4–A6 as consequences of the KKT conditions 5.4–5.6. The absolute constant of $1/3$ in the assumed inequalities is a presentation choice; examine the proof in Appendix 5.A.1 to convince oneself that other numbers could be used — if they are, at

[7] A random matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is isotropic if $\mathbb{E}\left(\mathbf{A}^\top\mathbf{A}\right) = \mathbf{I}_n$.

[8] Recall from Proposition 3.1.1 that $\operatorname{null}(\mathbf{D}) \cap \operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$ is a *necessary* condition for unique solutions in (P1).

least, strictly less than 1. The pair A2 and A3, on the contrary, are artifacts of the proof. They could potentially be replaced if different arguments were devised.

In light of Lemma 5.2.1, I will call any vector $\mathbf{u} \in \mathbb{R}^N$ satisfying A4–A6 an *inexact dual certificate* for $\mathbf{x}$ as *the* solution of P1. The next section shows how to actually produce such vectors using the lemma's assumptions as guidelines.

5.3 The golfing scheme for producing certificates

In the same spirit of the classic golfing scheme [28], we will start with a simple guess $\mathbf{u}^{(0)}$ which is then iteratively updated — via statistically independent adjustments — becoming some vector $\mathbf{u}^{(L)}$ after L steps. The exact number of steps is specified later on, but if $\mathbf{u}^{(L)}$ is to be an inexact dual certificate, we need to make sure that the error vectors

$$\mathbf{w}^{(l)} := \mathbf{P}_{\mathscr{S}}\left(\operatorname{sign}(\mathbf{Dx}) - \mathbf{u}^{(l)}\right), \quad l \in [L], \tag{5.7}$$

end up into the origin-centered Euclidean ball of radius $\frac{1}{3}$. At the same time, we have to guarantee that $(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}^{(L)}$ is within the cube $\frac{1}{3}\mathbb{B}_\infty^N$. The name of the iterative scheme gets is thus a metaphor: we imagine a ball at the initial position $\mathbf{u}^{(0)}$ and push it towards some ideal point $\mathbf{u}^\star$ specified by $\mathbf{P}_{\mathscr{S}}\mathbf{u}^\star = \operatorname{sign}(\mathbf{Dx})$ and $\left\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}^\star\right\|_\infty \leq 1$. Each iteration draws us closer to $\mathbf{u}^\star$, just as each golf shot should bring the ball closer to the hole (see Figure 5.1).

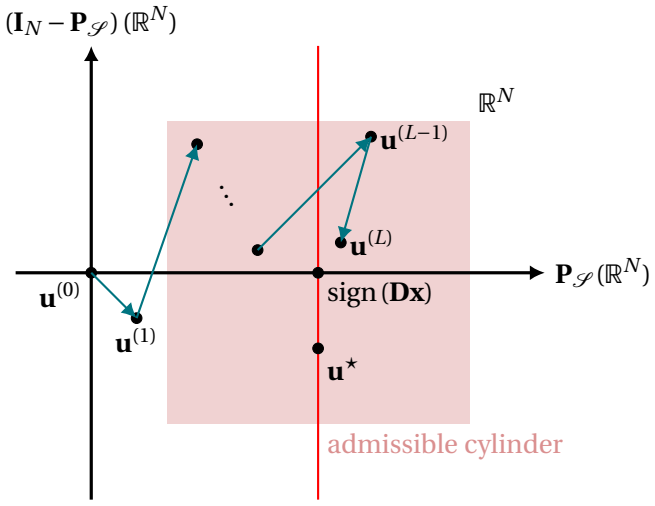


Figure 5.1 – The golfing scheme. Pick an initial guess $\mathbf{u}^{(0)}$ and iteratively push it towards a region where inexact dual certificates are located. The admissible cylinder is given by the set $\left\{\mathbf{u} \in \mathbb{R}^N : \ \mathbf{P}_{\mathscr{S}}\mathbf{u} - \operatorname{sign}(\mathbf{Dx}) \in \frac{1}{3}\mathbb{B}_2^N \ \text{ and } \ (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u} \in \frac{1}{3}\mathbb{B}_\infty^N\right\}$, according to Lemma 5.2.1. The red line segment represents the region of $\mathbb{R}^N$ containing the vectors that satisfy the KKT conditions (5.5) and (5.6) exactly.

A good place for picking the initial guess $\mathbf{u}^{(0)}$ is within the slab

$$\left\{\mathbf{u}\in\mathbb{R}^N : \left\|\left(\mathbf{I}_N-\mathbf{P}_{\mathscr{S}}\right)\mathbf{u}^\star\right\|_\infty \leq \frac{1}{3}\right\}.$$

Then, we could simply update each next iterate in the direction of the error vector $\mathbf{w}^{(l)}$, an assignment like the one below.

1: ...
2: $\mathbf{u}^{(l)} \leftarrow \mathbf{u}^{(l-1)} + \mathbf{w}^{(l)}$
3: ...

If we choose, for example, $\mathbf{u}^{(0)} = \mathbf{0}$ then within a single iteration $\mathbf{u}^{(1)} = \operatorname{sign}(\mathbf{Dx})$ would immediate yield a perfect certificate according to the KKT conditions (5.5) and (5.6). But the procedure above ignored an important detail: a valid dual certificate $\mathbf{u}^{(L)}$ should *also* satisfy $\mathbf{D}^\top\mathbf{u}^{(L)} \in \operatorname{range}\left(\mathbf{A}^\top\right)$.

But the range condition in (A4) is stated in the *primal* domain, because $\mathbf{D}^\top$ maps vectors in $\mathbb{R}^N$ to $\mathbb{R}^n$. To transform it into a condition on the *dual* domain (where $\mathbf{u}^{(0)},\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(L)}$ live), I will employ assumption (A1). The trivial intersection $\operatorname{null}(\mathbf{D})\cap\operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$ implies the one-to-one equivalence [9]

$$\mathbf{D}^\top\mathbf{u}\in\operatorname{range}\left(\mathbf{A}^\top\right) \iff \underbrace{(\mathbf{D}^+)^\top\mathbf{D}^\top}_{=\mathbf{D}\mathbf{D}^+}\mathbf{u}\in\operatorname{range}\left((\mathbf{D}^+)^\top\mathbf{A}^\top\right), \tag{5.8}$$

for any $\mathbf{u}\in\mathbb{R}^N$. Each iterate $\mathbf{u}^{(l)}$ can be expressed as $\mathbf{u}^{(l)} = \mathbf{D}\mathbf{D}^+\mathbf{u}^{(l)} + \left(\mathbf{I}_N-\mathbf{D}\mathbf{D}^+\right)\mathbf{u}^{(l)}$, by means of a direct sum of $\mathbb{R}^N$. Hence, by the equivalence (5.8), we are only required to write the orthogonal projection of $\mathbf{u}^{(l)}$ onto the range of $\mathbf{D}$ as $(\mathbf{D}^+)^\top\mathbf{A}^\top\mathbf{v}$, for some vector $\mathbf{v}\in\mathbb{R}^m$.

We can now improve the prototypical assignment by incorporating this new constraint using an additional sequence of vectors, $\{\mathbf{v}^{(l)}\}_{l\in[L]}$. Instead of simply adding the error vector $\mathbf{w}^{(l)}$, modify the part of it which lies within the range of $\mathbf{D}$ and add the result to the current iterate:

1: ...
2: $\mathbf{u}^{(l)} \leftarrow \mathbf{u}^{(l-1)} + \left(\mathbf{I}_N-\mathbf{D}\mathbf{D}^+\right)\mathbf{w}^{(l)} + (\mathbf{D}^+)^\top\mathbf{A}^\top\mathbf{v}^{(l)}$
3: ...

As long as $\mathbf{u}^{(0)}$ starts in $\left\{\mathbf{u}\in\mathbb{R}^N : \mathbf{D}\mathbf{D}^+\mathbf{u}\in\operatorname{range}\left((\mathbf{D}^+)^\top\mathbf{A}^\top\right)\right\}$, we can rest assured that each subsequent $\mathbf{u}^{(l)}$ will remain in the same set. Fortunately, picking $\mathbf{u}^{(0)} = \mathbf{0}$ will always guarantee this situation. But which form should the vectors $\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(L)}$ take?

It would be convenient if each $\mathbf{v}^{(l)}$ were a function of $\mathbf{w}^{(l)}$, so we would not have to deal with

[9] It suffices to note that $\operatorname{null}(\mathbf{D})\cap\operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$ implies that the orthogonal projection operators defined by $\mathbf{D}$ and $\mathbf{A}$ *commute*.

a separate sequence of vectors. Consider the following parametrization: $\mathbf{v}^{(l)} = \mathbf{B}^\top \mathbf{D}^\top \mathbf{w}^{(l)}$, for some matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$ to be chosen later. This expression — although not intuitive — allows the update directions to be then expressed as linear transformation of the error vector using familiar matrices:

$$\begin{aligned}\left(\mathbf{I}_N - \mathbf{D}\mathbf{D}^+\right)\mathbf{w}^{(l)} + (\mathbf{D}^+)^\top \mathbf{A}^\top \mathbf{v}^{(l)} &= \left[\mathbf{I}_N - \mathbf{D}\mathbf{D}^+ + (\mathbf{D}^+)^\top \mathbf{A}^\top \mathbf{B}^\top \mathbf{D}^\top\right]\mathbf{w}^{(l)} \\ &= \left[\mathbf{I}_N - (\mathbf{D}^+)^\top (\mathbf{I}_n - \mathbf{B}\mathbf{A})^\top \mathbf{D}^\top\right]\mathbf{w}^{(l)} \\ &= \left[\mathbf{I}_N - \underbrace{\left(\mathbf{D}(\mathbf{I}_n - \mathbf{B}\mathbf{A})\mathbf{D}^+\right)^\top}_{=\mathbf{M}}\right]\mathbf{w}^{(l)}.\end{aligned}$$

The matrix $\mathbf{M} \in \mathbb{R}^{N\times N}$ is the same as the one in the statement of Lemma 5.2.1. By retracing our steps, we can now see that multiplying any vector in $\mathbb{R}^N$ by $\mathbf{I}_N - \mathbf{M}$ forces the result to be in the set $\{\mathbf{u} \in \mathbb{R}^N : \mathbf{D}\mathbf{u} \in \text{range}(\mathbf{A}^\top)\}$. But unlike $\mathbf{D}\mathbf{D}^+$ or $\mathbf{A}\mathbf{A}^+$, the matrix $\mathbf{I}_N - \mathbf{M}$ does not represent an orthogonal projection, in general. The need to control certain semi-norms of $\mathbf{M}$ in Lemma 5.2.1 should seem less alien now. With matrix $\mathbf{M}$, our working version of the golfing scheme update reads as

---

1: ...
2: $\mathbf{u}^{(l)} \leftarrow \mathbf{u}^{(l-1)} + (\mathbf{I}_N - \mathbf{M})\mathbf{w}^{(l)}$
3: ...

Keep in mind that we still have an an unspecified matrix $\mathbf{B}$ hidden within the assignment above. Let us address it now. Our measurement operator is a random matrix, so matrix $\mathbf{M}$ — being a function of $\mathbf{A}$ — is also random. If $\mathbf{M}$ is zero-mean, then $\mathbf{u}^{(l)}$ is updated as initially intended at least on average. That is, $\mathbb{E}\left(\mathbf{u}^{(l)} - \mathbf{u}^{(l-1)}\right)$ points straight in the direction of the error vector $\mathbf{w}^{(l)}$. So how can we make $\mathbf{M}$ zero mean? Assume that $\mathbf{A}$ has a full rank covariance matrix [10], then $\mathbb{E}(\mathbf{M}) = \mathbf{0}$ if

$$\mathbf{B} = \mathbb{E}\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top. \tag{5.9}$$

From now on, assume $\mathbf{B}$ as in the expression above. Not only will it have the aforementioned effect on the average direction of the updates, but it will also simplify some probabilistic estimates of $\mathbf{M}$ later on in the chapter. I leave open the question of whether other choices of this conditioning matrix are useful or not.

Regardless, the vector $\mathbf{u}^{(L)}$ at the final iteration is only a proper inexact certificate if we can properly control the semi-norms $\|\mathbf{P}_{\mathscr{S}}(\mathbf{u}^{(L)} - \text{sign}(\mathbf{D}\mathbf{x}))\|_2$ and $\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}^{(L)}\|_\infty$. The first of

[10] This is true of both Ber($\boldsymbol{\pi}$) and CSWR($\boldsymbol{\pi}$) sampling models in this thesis.

these concerns directly the error vector $\mathbf{w}^{(L)}$, which can now be unpacked as [11]

$$\begin{aligned}
\mathbf{w}^{(L)} &:= \mathbf{P}_{\mathscr{S}}(\operatorname{sign}(\mathbf{Dx}) - \mathbf{u}^{(L)}) \\
&= \mathbf{P}_{\mathscr{S}}\left(\operatorname{sign}(\mathbf{Dx}) - \mathbf{u}^{(L-1)} - (\mathbf{I}_N - \mathbf{M})\,\mathbf{w}^{(L-1)}\right) \\
&= \mathbf{P}_{\mathscr{S}}\left(\mathbf{w}^{(L-1)} - \mathbf{w}^{(L-1)} + \mathbf{M}\mathbf{w}^{(L-1)}\right) \\
&= \mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{w}^{(L-1)} \\
&\cdots \\
&= [\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}]^L\,\mathbf{w}^{(0)} \\
&=: [\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}]^L \operatorname{sign}(\mathbf{Dx}).
\end{aligned}$$

If we show $\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}$ to be a contraction then the length of the error vectors decreases geometrically with each update. More precisely, if at each iteration $l \in [L]$ there is a constant $\alpha \in (0,1)$ such that $\left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{w}^{(l)}\right\|_2 < \alpha \left\|\mathbf{w}^{(l)}\right\|_2$, then the norm of the error vector at the end of the of golfing steps is

$$\begin{aligned}
\|\mathbf{P}_{\mathscr{S}}(\mathbf{u}^{(L)} - \operatorname{sign}(\mathbf{Dx}))\|_2 &= \|\mathbf{w}^{(L)}\|_2 \\
&= \left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{w}^{(L-1)}\right\|_2 \\
&\leq \alpha \left\|\mathbf{w}^{(L-1)}\right\|_2 \\
&\cdots \\
&\leq \alpha^L \|\operatorname{sign}(\mathbf{Dx})\|_2 \\
&= \alpha^L \sqrt{|\mathscr{S}|}. \qquad (5.10)
\end{aligned}$$

For the second semi-norm that we need to control, suppose that $\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}$ *as well as* its complement, $(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\mathbf{M}\mathbf{P}_{\mathscr{S}}$, are contractions — now in the infinity norm. In other words, let all steps $l \in [L]$ admit some constant $\beta \in (0,1)$ such that $\left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{w}^{(l)}\right\|_\infty$ and $\left\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{w}^{(l)}\right\|_\infty$ are smaller than $\beta \left\|\mathbf{w}^{(l)}\right\|_\infty$ [12]. Then we could see that the coordinates in the orthogonal complement of the error vectors $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(L+1)}$ do not drift far from zero throughout the process:

$$\begin{aligned}
\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{u}^{(L)}\|_\infty &= \left\|\sum_{i=1}^{L}(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,(\mathbf{I}_N - \mathbf{M})\,\mathbf{P}_{\mathscr{S}}\mathbf{w}^{(l-1)}\right\|_\infty \\
&= \left\|\sum_{i=1}^{L}(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{w}^{(l-1)}\right\|_\infty \\
\texttt{(triangle ineq.)} &\leq \sum_{i=1}^{L}\left\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{w}^{(l-1)}\right\|_\infty \\
&\leq \sum_{i=1}^{L}\beta\left\|\mathbf{w}^{(l-1)}\right\|_\infty
\end{aligned}$$

[11] Keep in mind that $\mathbf{w}^{(L)} = \mathbf{P}_{\mathscr{S}}\mathbf{w}^{(L)}$.

[12] This simultaneous control over the infinity norms of $\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}$ and $\left(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\right)\mathbf{M}\mathbf{P}_{\mathscr{S}}$ is an idea taken from Boyer *et al.* [10] to avoid having $\|\left(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\right)\mathbf{M}\mathbf{P}_{\mathscr{S}}\mathbf{u}^{(L)}\|_\infty$ depend on the size to the support $\mathscr{S} := \operatorname{supp}(\mathbf{Dx})$.

$$
\begin{aligned}
&= \sum_{i=1}^{L} \beta \left\| \mathbf{P}_{\mathscr{S}} \mathbf{M} \mathbf{P}_{\mathscr{S}} \mathbf{w}^{(l-2)} \right\|_{\infty} \\
&\leq \sum_{i=1}^{L} \beta^2 \left\| \mathbf{w}^{(l-2)} \right\|_{\infty} \\
&\cdots \\
&\leq \sum_{i=1}^{L} \beta^l \left\| \mathbf{w}^{(0)} \right\|_{\infty} \\
&= \sum_{i=1}^{L} \beta^l \underbrace{\left\| \operatorname{sign}(\mathbf{D}\mathbf{x}) \right\|_{\infty}}_{=1} \\
&= \frac{1-\beta^L}{1-\beta}.
\end{aligned} \tag{5.11}
$$

If we wish $\mathbf{u}^{(L)}$ to an inexact dual certificate as predicted by Lemma 5.2.1, our new objective is then to find out when we can guarantee contraction constants $\alpha, \beta \in (0,1)$ that satisfy $\alpha^L \sqrt{|\mathscr{S}|} \leq \frac{1}{3}$ and $\frac{1-\beta^L}{1-\beta} \leq \frac{1}{3}$.

Recall that $\mathbf{M}$ is a random matrix, so the desired contraction inequalities have to be probabilistic estimates. At this stage, such estimates are somewhat hampered by the statistical dependence between $\mathbf{M}$ and each error vector $\mathbf{w}^{(l)}$ [13]. The traditional way around this issue is to *force* each iterate $\mathbf{u}^{(l)}$ to be *independent* of each other by using different matrices in the place of $\mathbf{M}$ for each update. This idea goes back to Gross [28] — who introduced the golfing scheme — and is based on selecting L independent matrices, $\mathbf{A}^{(1)} \in \mathbb{R}^{m_1 \times n}, \ldots, \mathbf{A}^{(L)} \in \mathbb{R}^{m_L \times n}$ such that

$$
\bigcup_{l=1}^{L} \operatorname{range}\left(\mathbf{A}^{(l)\top}\right) \subset \operatorname{range}\left(\mathbf{A}^{\top}\right) \quad \text{and} \quad m_1 + m_2 + \cdots + m_L = m. \tag{5.12}
$$

The way I will define each of these smaller matrices is by simply stacking successive rows of our sampling operator. That is, $\mathbf{A}^{(1)}$ consists of the first m_1 rows of $\mathbf{A}$, $\mathbf{A}^{(2)}$ the next m_2 rows, an so on. This strategy works whenever the original matrix has independent rows, like the ones in this thesis [14].

We now use each of the independent chunks of $\mathbf{A}$ at its corresponding iteration. By that I mean

[13] To get a feeling for this, consider the semi-norm $\left\| \left(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\right) \mathbf{M} \mathbf{P}_{\mathscr{S}} \mathbf{w}^{(l)} \right\|_{\infty} := \max_{k \notin \mathscr{S}} \left| \left\langle \tilde{\mathbf{e}}_k, \mathbf{M} \mathbf{P}_{\mathscr{S}} \mathbf{w}^{(l)} \right\rangle \right|$. The moments of the quantities $\left\langle \mathbf{P}_{\mathscr{S}} \mathbf{M}^{\top} \tilde{\mathbf{e}}_k, \mathbf{w}^{(l)} \right\rangle$ inside the maximum cannot be factored into a product with a term depending only $\mathbf{M}$ and another relying only on $\mathbf{w}^{(l)}$. This is because both these objects are functions of the same random matrix $\mathbf{A}$. As a result, it is infeasible to condition on $\mathbf{w}^{(l)}$ in order to obtain a bound of the type $\left\| \left(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\right) \mathbf{M} \mathbf{P}_{\mathscr{S}} \mathbf{w}^{(l)} \right\|_{\infty} \leq \beta \left\| \mathbf{w}^{(l)} \right\|_{\infty}$ without having detailed knowledge on how the distribution of $\mathbf{A}$ affects the action of $\mathbf{M}$ on $\mathbf{w}^{(l)}$

[14] The same idea can work in other contexts. For example, Boyer *et al.*[10] study row-block measurements, so their matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)}$ are built by stacking independent *blocks of rows*.

to define an operator analogous to the matrix $\mathbf{M}$, for each $l \in [L]$, via the expression

$$\mathbf{M}^{(l)} := \left[\mathbf{D}\left(\mathbf{I}_n - \mathbb{E}\left(\mathbf{A}^{(l)\top}\mathbf{A}^{(l)}\right)^{-1}\mathbf{A}^{(l)\top}\mathbf{A}^{(l)}\right)\mathbf{D}^{+}\right]^{\top}. \tag{5.13}$$

Once that is done, we can finally settle on the final version of the golfing scheme, which I record here as Algorithm 1. Notice how $\mathbf{M}^{(l)}$ and $\mathbf{w}^{(l)}$ are now independent because the error vector is only a function of $\mathbf{M}^{(l-1)}, \mathbf{M}^{(l-2)}, \ldots, \mathbf{M}^{(1)}$ but not $\mathbf{M}^{(l)}$.

Algorithm 1 Golfing scheme

1: Set $\mathbf{u}^{(0)} = \mathbf{0}$
2: **for** $l = 1, \ldots, L$ **do**
3: $\quad \mathbf{u}^{(l)} \leftarrow \mathbf{u}^{(l-1)} + \left[\mathbf{I}_N - \mathbf{M}^{(l)}\right]\mathbf{w}^{(l)}$ ▷ with $\mathbf{M}^{(l)}$ as in (5.13), and $\mathbf{w}^{(l)}$ as in (5.7)
4: **end for**
5: **return** $\mathbf{u}^{(L)}$ as a potential certificate

Some readers may rightfully wonder, is the range constraint $\mathbf{D}^{\top}\mathbf{u}^{(L)} \in \operatorname{range}\left(\mathbf{A}^{\top}\right)$ still satisfied with the artificially introduced independent iterates? The answer is yes, because $\mathbf{D}^{\top}\mathbf{u}^{(L)} \in \bigcup_{l=1}^{L} \operatorname{range}\left(\mathbf{A}^{(l)\top}\right)$ and the right-hand side belongs to $\operatorname{range}\left(\mathbf{A}^{\top}\right)$ by construction. Still, notice the trade-off as we exchange a potentially larger certificate search space for the chance to work with independent iterates.

In the end, any iterative scheme proposing inexact dual certificates is only useful if the returned vector, $\mathbf{u}^{(L)}$, has all the qualities laid out by Lemma 5.2.1 — at least with high probability. The following lemma makes this wish precise in what concerns the golfing scheme of Algorithm 1. I include its proof in Appendix 5.A.2, but it essentially just assigns probabilistic figures to the assumptions we have gathered up to this point.

Lemma 5.3.1. *The vector $\mathbf{u}^{(L)}$ produced by Algorithm 1 after $L := 1 + \left\lceil \frac{\log|S|}{2\log 3} \right\rceil$ steps is, with probability larger than $1-\varepsilon$, an inexact dual certificate for $\mathbf{x}$ as the unique solution of P1 if*

$$\operatorname{null}(\mathbf{D}) \cap \operatorname{null}(\mathbf{A}) = \{0\}, \tag{A1}$$

$$\mathbb{P}\left(\left\{\|\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\|_2 > 1/3\right\}\right) \leq \frac{\varepsilon}{3}, \tag{A2}$$

$$\mathbb{P}\left(\left\{\max_{k\notin\mathscr{S}} \left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^{\top}\mathbf{e}_k\right\|_2 > 1\right\}\right) \leq \frac{\varepsilon}{3}, \tag{A3}$$

and, for all $\mathbf{v} \in \mathbb{R}^N$ and $l \in [L]$,

$$\mathbb{P}(\{\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\|_2 > (1/3)\|\mathbf{v}\|_2\}) \leq \varepsilon/3L, \tag{A5-l}$$

$$\mathbb{P}(\{\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\|_\infty > (1/4)\|\mathbf{v}\|_\infty\}) \leq \varepsilon/3L, \tag{A6(a)-l}$$

$$\mathbb{P}(\{\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\|_\infty > (1/4)\|\mathbf{v}\|_\infty\}) \leq \varepsilon/3L. \tag{A6(b)-l}$$

The correctness of the golfing scheme is but a means towards knowing how many samples it takes for $\mathbf{x}$ to be the unique vector decoded by P1. As usual in Compressed Sensing, such sample complexities appear as a consequence of enforcing tail bounds like the ones in Lemma 5.3.1. To be concrete, imagine that assumptions A2 and A3 were to hold as long as $m = m'$, and assumptions A5-l–A6(b)-l if $m_1 = m_2 = \cdots = m_L = m''$ [15] Then we would be allowed to conclude, with high probability, that problem P1 outputs only $\mathbf{x}$, provided that $\mathbf{A}$ has at least $m = \max\{m', L \cdot m''\}$ rows. I will finish this chapter giving precise figures to this argument, as it applies to the Coordinate Sampling with Replacement (CSWR($\boldsymbol{\pi}$)) model.

5.4 An optimal vertex-sampling design for $\mathscr{G}$-TV interpolation

In the CSWR($\boldsymbol{\pi}$) model, the measurement operator is formed by staking m standard basis vectors of $\mathbb{R}^n$, picked independently at random [16]. The picks are determined by *i.i.d.* copies of a random variable ω taking values in $[n]$ with probabilities $\mathbb{P}(\{\omega = k\}) = \pi_k, \forall k \in [n]$. The *i.i.d.* copies, $\omega_1, \ldots, \omega_m$, form a sampling set Ω with which we express the sampling matrix as $\mathbf{A} = \left(\mathbf{e}_{\omega_i}^\top\right)_{\omega_i \in \Omega}$.

Skipping some computations [17], the CSWR($\boldsymbol{\pi}$) model induces the following expression for the matrix $\mathbf{M}$ appearing in Lemma 5.3.1:

$$\mathbf{M} := \left[\mathbf{D}\left(\mathbf{I}_n - \mathbb{E}\left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\mathbf{A}\right)\mathbf{D}^+\right]^\top = \frac{1}{m}\sum_{i=1}^{m}\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\right)\mathbf{D}^+\right]^\top. \tag{5.14}$$

In words, $\mathbf{M}$ is thus a sum of independent perturbations of the orthogonal projection matrix $\mathbf{D}\mathbf{D}^+$ by random, rank-one matrices. Each rank-one matrix is associated with a vertex of the graph via the probabilities $\pi_1, \ldots, \pi_n$. By its very construction, a matrix $\mathbf{M}^{(l)}$ only differs then from $\mathbf{M}$ by restricting the limits of the sum in (5.14) to m_l consecutive rows. I will fix each of the chunks $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)}$ of $\mathbf{A}$ in the golfing scheme to be of the same size[18] (*i.e.*,$m_1 = m_2 = \cdots = m_L$) in order to write

$$\mathbf{M}^{(l)} = \frac{1}{m_1}\sum_{i=(l-1)\cdot m_1+1}^{l\cdot m_1}\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\right)\mathbf{D}^+\right]^\top \tag{5.15}$$

at once, for all $l \in [L]$.

The golfing scheme's ability to output an inexact dual certificate depends on the tails of functions of $\mathbf{M}$ and $\{\mathbf{M}^{(l)}\}_{l\in[L]}$. I will show that these tails are well-behaved if certain moments of the respective matrices are well-behaved. Correspondingly, define the following deterministic

[15] Recall that $m_1, \ldots, m_L$ are the number of rows, respectively, of the independent sub-matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)}$ induced by $\mathbf{A}$.

[16] See Chapter 2.

[17] It suffices to note that $\mathbf{A}^\top\mathbf{A} = \sum_{i=1}^m \mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top$ and $\mathbb{E}\left(\mathbf{A}^\top\mathbf{A}\right) = m\,\mathrm{diag}\,(\boldsymbol{\pi})$.

[18] There is no point in doing otherwise for the CSWR($\boldsymbol{\pi}$) model, because the rows of $\mathbf{A}$ are statistically indistinguishable from each other.

parameters (whose notation I borrowed from Boyer *et al.* [10]).

Definition 5.4.1.

$$\Theta(\mathscr{S},\boldsymbol{\pi}) := \max_{i\in[n]} \left\| \left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_i}\mathbf{e}_i\mathbf{e}_i^\top\right)\mathbf{D}^+ \right]^\top \mathbf{P}_{\mathscr{S}} \right\|_{\infty\to\infty} \tag{5.16}$$

$$= \max_{i\in[n]} \max_{k\in[N]} \left\| \tilde{\mathbf{e}}_k^\top \left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_i}\mathbf{e}_i\mathbf{e}_i^\top\right)\mathbf{D}^+ \right]^\top \mathbf{P}_{\mathscr{S}} \right\|_1 \tag{5.17}$$

$$\Upsilon(\mathscr{S},\boldsymbol{\pi}) := \sup_{\|\mathbf{v}\|_\infty\le 1} \sum_{i=1}^{n} \pi_i \cdot \left\| \left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_i}\mathbf{e}_i\mathbf{e}_i^\top\right)\mathbf{D}^+ \right]^\top \mathbf{P}_{\mathscr{S}}\mathbf{v} \right\|_2^2 \tag{5.18}$$

$$\Gamma(\mathscr{S},\boldsymbol{\pi}) := \max\{\Theta(\mathscr{S},\boldsymbol{\pi}), \Upsilon(\mathscr{S},\boldsymbol{\pi})\}. \tag{5.19}$$

The main theorem of this chapter is the next result, which provides a sample complexity threshold for exact recovery in (P1) under CSWR($\boldsymbol{\pi}$) measurements. To prove it, simply call upon certain versions of the Bernstein inequality, as seen in Appendix 5.A.3. Good Bernstein bounds rely on the good estimation of moments, so I avoided approximations, deferring them to Chapter 6. The elaborate expressions hidden under $\Gamma(\mathscr{S},\boldsymbol{\pi})$ are a consequence of this choice.

Theorem 5.4.1 (Sample complexity of (P1) under CSWR($\boldsymbol{\pi}$) measurements)**.** *Let* $\mathbf{A}\in\mathbb{R}^{m\times n}$ *be the measurement matrix in the CSWR(*$\boldsymbol{\pi}$*) model and* $\mathbf{D}\in\mathbb{R}^{N\times n}$ *be the analysis matrix, denoting* $\mathscr{S} := \operatorname{supp}(\mathbf{Dx})$ *for some* $\mathbf{x}\in\mathbb{R}^n$*. With probability larger than* $1-\varepsilon$*, vector* $\mathbf{x}$ *is the sole output of* (P1) *if* $\operatorname{null}(\mathbf{D})\cap\operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$ *and*

$$m \ge 38\cdot\Gamma(\mathscr{S},\boldsymbol{\pi})\cdot\log(|\mathscr{S}|)\cdot\log\left(\frac{63\cdot N\cdot\log(|\mathscr{S}|)}{\varepsilon}\right). \tag{5.20}$$

As far as this thesis is concerned, the best sampling design for the (P1) decoder is the one that minimizes its sample complexity. This design — according to Theorem 5.4.1 — should therefore minimize $\Gamma(\mathscr{S},\boldsymbol{\pi})$, since this is the only factor in the sample complexity bound that depends on the sampling probabilities $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_n)$. The next corollary just formalizes this statement.

Corollary 5.4.1.1 (Optimal CSWR($\boldsymbol{\pi}$) design)**.** *Let* $\mathbf{D}\in\mathbb{R}^{N\times n}$ *be the analysis matrix for some* $\mathbf{x}\in\mathbb{R}^n$*, yielding the co-support* $\mathscr{S} := \operatorname{supp}(\mathbf{Dx})$*. The CSWR(*$\boldsymbol{\pi}$*) design that minimizes the number of measurements required by Theorem 5.4.1 to exactly recover* $\mathbf{x}$ *from* (P1) *with high probability is*

$$\boldsymbol{\pi} = \begin{cases} \textit{arg}\ \min\limits_{\mathbf{p}\in\mathbb{R}^n} & \Gamma(\mathscr{S},\mathbf{p}) \\ \textit{subject to} & \mathbf{p}\succeq\mathbf{0}\ \textit{ and }\ \langle\mathbf{p},\mathbf{1}\rangle = 1. \end{cases} \tag{5.21}$$

The optimal design — despite being easy to state — is not necessarily straightforward to implement. After all, the objective $\Gamma(\mathscr{S}, \boldsymbol{\pi})$ is the maximum of two rather complicated expressions, $\Theta(\mathscr{S}, \boldsymbol{\pi})$ and $\Upsilon(\mathscr{S}, \boldsymbol{\pi})$. Boyer *et al.* [10], in a similar situation, suggest looking for common upper bounds to $\Theta(\mathscr{S}, \boldsymbol{\pi})$ and $\Upsilon(\mathscr{S}, \boldsymbol{\pi})$, and optimize that instead. But finding appropriate upper bounds for our setting goes beyond the scope of this chapter. It is also important to mention the dependence of $\Gamma(\mathscr{S}, \boldsymbol{\pi})$ on $\mathscr{S} = \operatorname{supp}(\mathbf{Dx})$ [19]: since $\mathbf{x}$ is the hidden "ground-truth" signal, how can one estimate the actions of the projection matrix $\mathbf{P}_{\mathscr{S}}$ without knowing $\mathbf{x}$ a priori? The reader will find some numerical experiments addressing this questions in Chapter 6.

5.5 Summary and final notes

The Karush-Kuhn-Tucker (KKT) conditions of the interpolation problem (P1) reveal that dual certificate vectors arise in the interaction of range ($\mathbf{A}$) and the subdifferential $\partial\|\mathbf{D}\cdot\|_1(\mathbf{x})$. In fact, merely approximating the KKT conditions can be enough to guarantee exact recovery. Using Lemma 5.2.1 as blueprint, I formulated a golfing scheme that produces potential certificates. Experienced readers might spot how Algorithm 1 encompasses other golfing schemes from the literature, derived from particular instances of problem (P1) [20].

When $\mathbf{A}$ comes from the CSWR($\boldsymbol{\pi}$) model, the success of the golfing scheme demands a number of measurements proportional to a term $\Gamma(\mathscr{S}, \boldsymbol{\pi})$. Here lies the explicit connection between the sample complexity of (P1) and the sampling design $\boldsymbol{\pi}$. Finding the optimal design is thus a matter of minimizing $\Gamma(\mathscr{S}, \boldsymbol{\pi})$, a quantity related to moments of random matrices induced by $\mathbf{D}$ and $\mathbf{A}$. The practical aspects of sampling optimally are discussed in the next chapter.

Lastly, a short note about the absence of the regression version (P1-η) in this chapter. Employing the golfing scheme can be suboptimal when dealing with noisy measurements [41]. By this, I mean that the number of measurements predicted by the scheme is knowingly not the best possible for some measurement ensembles. I decided then to restrict this chapter to setting with noiseless samples, hoping that ideas like the ones in Chapter 4 could prove to be effective in the future to study the sample complexity of $\mathscr{G}$-TV regression decoders.

[19] Recall that, in the context of piecewise-constant graph signals analysed via the graph gradient, $\mathscr{S}$ is called the jump-set of the signal.

[20] Problems that have, for example, $\mathbf{D} = \mathbf{I}_n$ and a Gaussian matrix for $\mathbf{A}$.

Appendix 5.A Proofs

5.A.1 Proof of Lemma 5.2.1

As a reminder, I follow the strategy of Candès and Plan [13, Lemma 3.2]: assume that some perturbation $\mathbf{x}+\mathbf{h}$ is a solution of (P1), and then show that assumptions A1 – A6 imply $\mathbf{h}=\mathbf{0}$, lest the contradiction $\|\mathbf{D}(\mathbf{x}+\mathbf{h})\|_1 > \|\mathbf{D}\mathbf{x}\|_1$ take place.

⟨1⟩1. Suppose $\mathbf{x}+\mathbf{h}$ is a solution of (P1). It suffices to consider $\mathbf{h} \in \text{null}(\mathbf{A})$ such that $\mathbf{h} \perp \text{null}(\mathbf{D})$.
PROOF:
⟨2⟩1. From the feasilibilty condition, $\mathbf{A}(\mathbf{x}+\mathbf{h}) = \mathbf{A}\mathbf{x} \implies \mathbf{h}=\mathbf{0}$. So $\mathbf{h}$ has to be in the null space of $\mathbf{A}$ in order for $\mathbf{x}+\mathbf{h}$ to be a solution.
⟨2⟩2. If, on top of that, $\mathbf{h} \in \text{null}(\mathbf{D})$, then assumption A1 implies $\mathbf{h}=\mathbf{0}$ and the uniqueness claim holds trivially.
⟨2⟩3. Therefore, the interesting perturbations are the ones belonging to the intersection $\text{null}(\mathbf{A}) \cap \text{range}\left(\mathbf{D}^\top\right)$.
□

⟨1⟩2. Define $\mathbf{g} := \text{sign}(\mathbf{D}\mathbf{x}) + (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\text{sign}(\mathbf{D}\mathbf{h}) \in \mathbb{R}^N$. Vector $\mathbf{g}$ is a valid subgradient of $\|\cdot\|_1$ at $\mathbf{D}\mathbf{x}$.
PROOF:Indeed, we verify

$$\begin{aligned} \mathbf{P}_{\mathscr{S}}\mathbf{g} &= \text{sign}(\mathbf{D}\mathbf{x}) \\ \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{g}\|_\infty = \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\text{sign}(\mathbf{D}\mathbf{h})\,\|_\infty &= \left\|\text{sign}((\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h})\right\|_\infty \leq 1, \end{aligned}$$

so $\mathbf{g}$ is a valid subgradient, by Proposition 3.1.2. □

⟨1⟩3. Use the expression of $\mathbf{g}$ and the fact that it is a subgradient to lower bound $\|\mathbf{D}(\mathbf{x}+\mathbf{h})\|_1$ as

$$\begin{aligned} \|\mathbf{D}(\mathbf{x}+\mathbf{h})\|_1 &\geq \|\mathbf{D}\mathbf{x}\|_1 + \left\langle \mathbf{g}, \mathbf{D}\mathbf{h}\right\rangle \\ &= \|\mathbf{D}\mathbf{x}\|_1 + \left\langle \text{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{h}\right\rangle + \left\langle (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\text{sign}(\mathbf{D}\mathbf{h}), \mathbf{D}\mathbf{h}\right\rangle \\ &= \|\mathbf{D}\mathbf{x}\|_1 + \left\langle \text{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{h}\right\rangle + \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1 . \end{aligned}$$

We need now to provide estimates to the terms to the right of $\|\mathbf{D}\mathbf{x}\|_1$ in the bound above.

⟨1⟩4. Let $\mathbf{u} \in \mathbb{R}^N$ be any vector satisfying A4. Then, $\langle \mathbf{u}, \mathbf{D}\mathbf{h}\rangle = 0$.
PROOF: Since $\mathbf{h} \in \text{null}(\mathbf{A})$, then $\langle \mathbf{v}, \mathbf{h}\rangle = 0$ for any $\mathbf{v} \in \text{range}\left(\mathbf{A}^\top\right)$. But $\mathbf{D}^\top\mathbf{u} \in \text{range}\left(\mathbf{A}^\top\right)$, so $\langle \mathbf{u}, \mathbf{D}\mathbf{h}\rangle = \langle \mathbf{D}^\top\mathbf{u}, \mathbf{h}\rangle = 0$. □

⟨1⟩5. Add $0 = \langle \mathbf{u}, \mathbf{D}\mathbf{h}\rangle - \langle \mathbf{u}, \mathbf{D}\mathbf{h}\rangle$ to the bound of step ⟨1⟩3, obtaining

$$\begin{aligned} \|\mathbf{D}(\mathbf{x}+\mathbf{h})\|_1 &\geq \|\mathbf{D}\mathbf{x}\|_1 + \underbrace{\langle \mathbf{u}, \mathbf{D}\mathbf{h}\rangle}_{=0} - \left\langle \mathbf{u} - \text{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{h}\right\rangle + \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1 \\ &= \|\mathbf{D}\mathbf{x}\|_1 - \left\langle \mathbf{u} - \text{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{h}\right\rangle + \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1. \end{aligned}$$

Next, we upper bound $\left|\left\langle \mathbf{u} - \text{sign}(\mathbf{D}\mathbf{x}), \mathbf{D}\mathbf{h}\right\rangle\right|$.

⟨1⟩6. Let $\mathbf{u}$ further abide by assumptions A5 and A3. Then,

$$\left|\left\langle \mathbf{u} - \operatorname{sign}(\mathbf{Dx}), \mathbf{Dh} \right\rangle\right| \leq \frac{1}{3}\|\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\|_2 + \frac{1}{3}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1.$$

PROOF:

⟨2⟩1. Split $\mathbf{u} - \operatorname{sign}(\mathbf{Dx})$ into range$(\mathbf{P}_{\mathscr{S}})$ and null$(\mathbf{P}_{\mathscr{S}})$.

⟨2⟩2. Then, using the triangle and Hölder inequalities, along with assumptions A5 and A3, compute

$$\begin{aligned}
\left|\left\langle \mathbf{u} - \operatorname{sign}(\mathbf{Dx}), \mathbf{Dh} \right\rangle\right| &= \left|\left\langle \mathbf{P}_{\mathscr{S}}(\mathbf{u} - \operatorname{sign}(\mathbf{Dx})), \mathbf{Dh} \right\rangle + \left\langle (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})(\mathbf{u} - \operatorname{sign}(\mathbf{Dx})), \mathbf{Dh} \right\rangle\right| \\
&= \left|\left\langle \mathbf{P}_{\mathscr{S}}(\mathbf{u} - \operatorname{sign}(\mathbf{Dx})), \mathbf{P}_{\mathscr{S}}\mathbf{Dh} \right\rangle + \langle (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}, (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh} \rangle\right| \\
&\leq \left|\left\langle \mathbf{P}_{\mathscr{S}}(\mathbf{u} - \operatorname{sign}(\mathbf{Dx})), \mathbf{P}_{\mathscr{S}}\mathbf{Dh} \right\rangle\right| + |\langle (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\mathbf{u}, (\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\,\mathbf{Dh} \rangle| \\
&\leq \|\mathbf{P}_{\mathscr{S}}(\mathbf{u} - \operatorname{sign}(\mathbf{Dx}))\|_2 \cdot \|\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\|_2 \\
&\qquad + \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}\|_\infty \cdot \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1 \\
&\leq \frac{1}{3}\|\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\|_2 + \frac{1}{3}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1.
\end{aligned}$$

□

⟨1⟩7. Pair the result from step ⟨1⟩6 with assumption A3, to read the lower bound from step ⟨1⟩5 as

$$\|\mathbf{D}(\mathbf{x} + \mathbf{h})\|_1 \geq \|\mathbf{Dx}\|_1 - \frac{1}{3}\|\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\|_2 + \frac{2}{3}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1.$$

Continue by controlling $\|\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\|_2$ in terms of $\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1$.

⟨1⟩8. Assumptions A2 and A3 imply

$$\|\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\|_2 < \frac{3}{2}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1.$$

PROOF:

⟨2⟩1. Crucially, assumption A2 implies that the matrix $\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{P}_{\mathscr{S}}$ is *invertible*. Indeed, we can bound the norm of its inverse, using the Neumann series, as

$$\begin{aligned}
\left\| \left[\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{P}_{\mathscr{S}}\right]^{-1} \right\|_2 &= \left\| \sum_{k=0}^{\infty} \left[\mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{P}_{\mathscr{S}}\right]^k \right\|_2 \\
&\leq \sum_{k=0}^{\infty} \left\| \mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{P}_{\mathscr{S}} \right\|_2^k \\
&\leq \sum_{k=0}^{\infty} \frac{1}{3^k} \\
&= \frac{3}{2}.
\end{aligned}$$

⟨2⟩2. We can then write the projection matrix $\mathbf{P}_{\mathscr{S}}$ in the slightly convoluted form

$$\mathbf{P}_{\mathscr{S}} = \left[\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{P}_{\mathscr{S}}\right]^{-1}\left[\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]\mathbf{P}_{\mathscr{S}}.$$

⟨2⟩3. Furthermore, $\mathbf{h} \perp \text{null}(\mathbf{D})$ implies $\mathbf{h} = \mathbf{D}^+\mathbf{Dh}$, while $\mathbf{h} \in \text{null}(\mathbf{A})$ implies $\mathbf{BAh} = 0$ for any matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$. Together, these facts entail the identity $\mathbf{h} = (\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{Dh}$.

⟨2⟩4. Gathering these observations, and using the shorthand $\mathbf{P}_{\mathscr{S}^c} := \mathbf{I}_N - \mathbf{P}_{\mathscr{S}}$, we are able to write

$$\begin{aligned}
\|\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\|_2 &= \Big\| \left[\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{P}_{\mathscr{S}}\right]^{-1} \\
&\qquad \left[\mathbf{P}_{\mathbb{S}} - \mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{P}_{\mathscr{S}}\right]\mathbf{Dh}\Big\|_2 \\
&\le \left\| \left[\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\mathbf{P}_{\mathscr{S}}\right]^{-1} \right\|_2 \times \\
&\qquad \left\|\mathbf{P}_{\mathscr{S}}\left[\mathbf{I}_N - \mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\right\|_2 \\
&\le \frac{3}{2}\left\|\mathbf{P}_{\mathscr{S}}\left[\mathbf{I}_N - \mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\right\|_2 \\
&=: \frac{3}{2}\left\|\mathbf{P}_{\mathscr{S}}\left[\mathbf{I}_N - \mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right](\mathbf{I}_N - \mathbf{P}_{\mathscr{S}^c})\mathbf{Dh}\right\|_2 \\
&= \frac{3}{2}\Bigg\| \mathbf{P}_{\mathscr{S}}\underbrace{\left[\mathbf{I}_N - \mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]\mathbf{Dh}}_{=0} \\
&\qquad -\mathbf{P}_{\mathscr{S}}\left[\mathbf{I}_N - \mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]\mathbf{P}_{\mathscr{S}^c}\mathbf{Dh}\Big\|_2 \\
&= \frac{3}{2}\left\|\mathbf{P}_{\mathscr{S}}\left[\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]\mathbf{P}_{\mathscr{S}^c}\mathbf{Dh}\right\|_2 \\
&\le \frac{3}{2}\sum_{k\in\mathscr{S}^c}\left\|\mathbf{P}_{\mathscr{S}}\left[\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]\tilde{\mathbf{e}}_k\right\|_2 \cdot |\tilde{\mathbf{e}}_k^\top\mathbf{Dh}| \\
&\le \frac{3}{2}\max_{k\in\mathscr{S}^c}\left\|\mathbf{P}_{\mathscr{S}}\left[\mathbf{D}(\mathbf{I}_n - \mathbf{BA})\mathbf{D}^+\right]\tilde{\mathbf{e}}_k\right\|_2 \cdot \|\mathbf{P}_{\mathscr{S}^c}\mathbf{Dh}\|_1 \\
&\le \frac{3}{2}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1.
\end{aligned}$$

□

⟨1⟩9. Steps ⟨1⟩7 and ⟨1⟩8 combine to yield the lower bound

$$\|\mathbf{D}(\mathbf{x}+\mathbf{h})\|_1 \ge \|\mathbf{Dx}\|_1 + \left(\frac{2}{3} - \frac{1}{2}\right)\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1 = \|\mathbf{Dx}\|_1 + \frac{1}{6}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1.$$

⟨1⟩10. But then we must conclude that $(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh} = \mathbf{P}_{\mathscr{S}}\mathbf{Dh} = 0$. In other words, $\mathbf{h} \in \text{null}(\mathbf{D})$.
PROOF:

⟨2⟩1. Vector $\mathbf{x}+\mathbf{h}$ is assumed to be a minimizer of $\|\mathbf{Dz}\|_1$, subject to $\mathbf{Az} = \mathbf{Ax}$. Hence, $\|\mathbf{D}(\mathbf{x}+\mathbf{h})\|_1 \le \|\mathbf{Dx}\|_1$, because $\mathbf{x}$ is trivially feasible.

⟨2⟩2. In order to avoid contradiction in step ⟨1⟩9, we must then have $\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1 = 0$, meaning $(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh} = \mathbf{0}$.

⟨2⟩3. The second assertion, $\mathbf{P}_{\mathscr{S}}\mathbf{Dh} = \mathbf{0}$, ultimately follows from the dominance relation $\|\mathbf{P}_{\mathscr{S}}\mathbf{Dh}\|_2 < \frac{3}{2}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{Dh}\|_1$.

⟨1⟩11. Q.E.D.

The only point $\mathbf{h}$ satisfying both $\mathbf{h} \in \text{null}(\mathbf{D})$ and $\mathbf{h} \perp \text{null}(\mathbf{D})$ is $\mathbf{h} = \mathbf{0}$. Therefore, $\mathbf{x} + \mathbf{h} = \mathbf{x} + \mathbf{0} = \mathbf{x}$ is the only solution of problem (P1).

Remark. The arguably most unnatural step in this proof, ⟨1⟩8⟨2⟩1, was inspired by a comment in Boyer *et al.* [10, Appendix A]. This step is ultimately the reason I could directly adapt the classical proof from Candès and Plan [13]. To wit, Lee *et al.* [48, Lemma 21] arrive at a result that is similar to (but slightly weaker than) Lemma 5.2.1, but their proof — derived from Chen and Chi [16] — requires splitting their argument into two complementary cases. The first supposes $\|\mathbf{P}_{\mathscr{S}}\mathbf{D}\mathbf{h}\|_2 < \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1$, while the second $\|\mathbf{P}_{\mathscr{S}}\mathbf{D}\mathbf{h}\|_2 > \|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1$. Our implicit invertibility assumption in $\|\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\|_2 \leq 1/3$ (a version of which Lee *et al.* [48] *also* require) makes the second case above irrelevant: $\|\mathbf{P}_{\mathscr{S}}\mathbf{D}\mathbf{h}\|_2$ is always dominated by $\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{D}\mathbf{h}\|_1$.

5.A.2 Proof of Lemma 5.3.1

All we need to do is verify that the assumptions of Lemma 5.2.1 hold with high probability. Recall that $\mathbf{w}^{(l)} := \mathbf{P}_{\mathscr{S}}\left(\text{sign}(\mathbf{D}\mathbf{x}) - \mathbf{u}^{(l)}\right)$ is the error vector at each iteration $l \in [L]$, and the updates of the golfing scheme are given by $\mathbf{u}^{(l)} = \mathbf{u}^{(l-1)} + \left[\mathbf{I}_N - \mathbf{M}^{(l)}\right]\mathbf{w}^{(l)}$. The rest of the proof is a computation exercise.

PROOF:

⟨1⟩1. Refer back to expression 5.7. If assumptions (A5-l) hold for each $\mathbf{v} \in \mathbb{R}^N$ and $l \in [L]$, then the choice of $L := 1 + \left\lceil \frac{\log|S|}{2\log 3} \right\rceil$ implies that, with probability at least $1 - \varepsilon/3$,

$$\begin{aligned}\|\mathbf{P}_{\mathscr{S}}(\mathbf{u}^{(L)} - \text{sign}(\mathbf{D}\mathbf{x}))\|_2 &\leq \left(\frac{1}{3}\right)^L \sqrt{|S|} \\ &\leq \frac{1}{3\sqrt{|\mathscr{S}|}}\sqrt{|\mathscr{S}|} \\ &\leq \frac{1}{3}.\end{aligned}$$

⟨1⟩2. Similarly, refferring to inequality (5.11), assume (A6(a)-l) and (A6(b)-l). For any $L \geq 1$, we have then

$$\begin{aligned}\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{u}^{(L)}\|_\infty &\leq \frac{1}{4}\left(\frac{1 - 1/4^L}{1 - 1/4}\right) \\ &= \frac{1}{3}\left(1 - \frac{1}{4^L}\right) \\ &\leq \frac{1}{3},\end{aligned}$$

with probability at least $1 - \varepsilon/3$.

⟨1⟩3. By construction, $\mathbf{u}^{(L)}$ satisfies $\mathbf{D}^\top\mathbf{u}^{(L)} \in \text{range}\left(\mathbf{A}^\top\right)$ with probability 1.

⟨1⟩4. Together with assumptions (A2) and (A3), we verify all the requirements of Lemma 5.2.1 with probability at least $1 - \left(\frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}\right) = 1 - \varepsilon$.

⟨1⟩5. Q.E.D.
With probability larger than $1-\varepsilon$, vector $\mathbf{u}^{(L)}$ is an inexact dual certificate for $\mathbf{x}$, according to Lemma 5.2.1. Therefore, with the same likelihood, vector $\mathbf{x}$ is the unique solution of problem (P1).

5.A.3 Proof of Theorem 5.4.1

PROOF: The argument consists of going through each of the conditions in Lemma 5.3.1, ensuring that the golfing scheme will produce an inexact dual certificate for the uniqueness of $\mathbf{x}$ as the solution of (P1). All the operators we have to deal with are sums of bounded, independent random matrices, allowing us to employ the Bernstein inequalities in Appendix A to derive the necessary tail bounds.

⟨1⟩1. I claim that $\mathbb{P}\left(\left\{\|\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\|_{2\to 2} > 1/3\right\}\right) \leq \frac{\varepsilon}{3}$ if $m \geq 24\cdot\Theta(\mathscr{S},\boldsymbol{\pi})\cdot\log\left(\frac{6|\mathscr{S}|}{\varepsilon}\right)$.

PROOF: Write

$$\mathbf{X} := \mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}} = \frac{1}{m}\sum_{i=1}^{m} \underbrace{\mathbf{P}_{\mathscr{S}}\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\mathbf{P}_{\mathscr{S}}\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}}_{=:\mathbf{X}_i},$$

a sum of independent, zero-mean random matrices $\{\mathbf{X}_i\}_i$.

⟨2⟩1. Bound each $\mathbf{X}_i$ almost surely with $\Theta(\mathscr{S},\boldsymbol{\pi})$:

$$\begin{aligned}
\|\mathbf{X}_i\|_{2\to 2} &\leq \underbrace{\|\mathbf{P}_{\mathscr{S}}\|_{2\to 2}}_{=1}\cdot\left\|\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}\right\|_{2\to 2}\\
&\leq \max_{i\in[n]}\left\|\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_i}\mathbf{e}_i\mathbf{e}_i^\top\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}\right\|_{2\to 2}\\
&\leq \Theta(\mathscr{S},\boldsymbol{\pi}).
\end{aligned}$$

⟨2⟩2. Bound the second moment matrices in the positive definite order:

$$\begin{aligned}
\mathbf{0} \preceq \mathbb{E}\left(\mathbf{X}_i\mathbf{X}_i^\top\right) &= \mathbf{P}_{\mathscr{S}}\mathbf{D}\mathbf{D}^+\underbrace{\mathbb{E}\left(\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}\right)}_{=\mathbf{0}}\\
&\qquad + \mathbb{E}\left(\frac{1}{\pi_{\omega_i}}\left[\mathbf{D}^+\right]^\top\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\mathbf{D}^\top\mathbf{P}_{\mathscr{S}}\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}\right)\\
&\preceq \mathbb{E}\left(\frac{1}{\pi_{\omega_i}}\left[\mathbf{D}^+\right]^\top\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\mathbf{D}^\top\mathbf{P}_{\mathscr{S}}\right)\\
&\qquad \times\max_{i\in[n]}\left\|\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_i}\mathbf{e}_i\mathbf{e}_i^\top\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}\right\|_{2\to 2}\\
&\preceq \Theta(\mathscr{S},\boldsymbol{\pi})\cdot\mathbf{D}\mathbf{D}^+\mathbf{P}_{\mathscr{S}}\\
&\preceq \Theta(\mathscr{S},\boldsymbol{\pi})\cdot\mathbf{I}_N, \text{ (}\mathbf{D}\mathbf{D}^+ \text{ is an orthogonal projector)}
\end{aligned}$$

and, by symmetry, $\mathbb{E}\left(\mathbf{X}_i^\top\mathbf{X}_i\right) \preceq \Theta(\mathscr{S},\boldsymbol{\pi})\cdot\mathbf{I}_N$.

⟨2⟩3. Set the variance parameter $\nu(\mathbf{X}) := \max\left\{\mathbb{E}\left(\mathbf{X}\mathbf{X}^\top\right), \mathbb{E}\left(\mathbf{X}^\top\mathbf{X}\right)\right\} = \frac{1}{m}\Theta(\mathscr{S}, \boldsymbol{\pi})$.

⟨2⟩4. With these moment bounds, the matrix Bernstein inequality in Lemma A.0.3 gives the tail bound

$$\mathbb{P}\left(\left\{\|\mathbf{P}_{\mathscr{S}}\mathbf{M}\mathbf{P}_{\mathscr{S}}\|_{2\to 2} > 1/3\right\}\right) \leq 2|\mathscr{S}| \cdot \exp\left(-\frac{m}{24\Theta(\mathscr{S}, \boldsymbol{\pi})}\right).$$

⟨2⟩5. This probability is less than $\varepsilon/3$ if $m \geq 24 \cdot \Theta(\mathscr{S}, \boldsymbol{\pi}) \cdot \log\left(\frac{6|\mathscr{S}|}{\varepsilon}\right)$.

□

⟨1⟩2. I claim that $\mathbb{P}\left(\left\{\max_{k \notin \mathscr{S}} \left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^\top \mathbf{e}_k\right\|_2 > 1\right\}\right) \leq \frac{\varepsilon}{3}$ if $m \geq 24 \cdot \Theta(\mathscr{S}, \boldsymbol{\pi}) \cdot \log\left(\frac{6(N-|\mathscr{S}|)}{\varepsilon}\right)$.

PROOF: Note the domination relation

$$\max_{k \notin \mathscr{S}} \left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^\top \mathbf{e}_k\right\|_2 \leq \left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^\top \left(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\right)\right\|_{2\to 2} = \|\left(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\right)\mathbf{M}\mathbf{P}_{\mathscr{S}}\|_{2\to 2},$$

implying $\mathbb{P}\left(\left\{\max_{k \notin \mathscr{S}} \left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^\top \mathbf{e}_k\right\|_2 > 1\right\}\right) \leq \mathbb{P}\left(\left\{\|\left(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}}\right)\mathbf{M}\mathbf{P}_{\mathscr{S}}\|_{2\to 2} > 1\right\}\right)$. The right-hand side is less than $\varepsilon/3$ if $m \geq 24 \cdot \Theta(\mathscr{S}, \boldsymbol{\pi}) \cdot \log\left(\frac{6(N-|\mathscr{S}|)}{\varepsilon}\right)$, by precisely the same arguments given in step ⟨1⟩1. □

⟨1⟩3. For each $l \in [L]$, I claim that $\mathbb{P}\left(\left\{\left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\right\|_2 > (1/3)\|\mathbf{v}\|_2\right\}\right) \leq \varepsilon/3L$, as long as the row size satisfies $m_l \geq 8 \cdot \max\{3\Upsilon(\mathscr{S}, \boldsymbol{\pi}), \Theta(\mathscr{S}, \boldsymbol{\pi})\} \cdot \log\left(\frac{6L}{\varepsilon}\right)$.

PROOF: The problem is the same for each $l \in [L]$, so we consider only $l = 1$. Fix $\mathbf{v} \in \mathbb{B}_2^N$ and write

$$\mathbf{P}_{\mathscr{S}}\mathbf{M}^{(1)}\mathbf{P}_{\mathscr{S}}\mathbf{v} = \frac{1}{m_1}\sum_{i=1}^{m_1} \underbrace{\mathbf{P}_{\mathscr{S}}\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top \mathbf{P}_{\mathscr{S}}\right)\mathbf{D}^+\right]^\top \mathbf{P}_{\mathscr{S}}\mathbf{v}}_{=:\mathbf{v}_i},$$

a sum of independent, zero-mean random vectors $\{\mathbf{v}_i\}_i$.

⟨2⟩1. Bound each $\mathbf{v}_i$ almost surely with $\Theta(\mathscr{S}, \boldsymbol{\pi})$:

$$\begin{aligned}\|\mathbf{v}_i\|_2 &\leq \left\|\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\right)\mathbf{D}^+\right]^\top \mathbf{P}_{\mathscr{S}}\right\|_{2\to 2} \cdot \underbrace{\|\mathbf{v}\|_2}_{\leq 1} \cdot \\ &\leq \Theta(\mathscr{S}, \boldsymbol{\pi}).\end{aligned}$$

⟨2⟩2. Bound the second moment as

$$\begin{aligned}\mathbb{E}\left(\|\mathbf{v}_i\|_2^2\right) &= \sum_{i=1}^n \pi_i \left\|\mathbf{P}_{\mathscr{S}}\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top \mathbf{P}_{\mathscr{S}}\right)\mathbf{D}^+\right]^\top \mathbf{P}_{\mathscr{S}}\mathbf{v}\right\|_2^2 \\ &\leq \sum_{i=1}^n \pi_i \left\|\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top \mathbf{P}_{\mathscr{S}}\right)\mathbf{D}^+\right]^\top \mathbf{P}_{\mathscr{S}}\mathbf{v}\right\|_2^2 \\ &\leq \Upsilon(\mathscr{S}, \boldsymbol{\pi}).\end{aligned}$$

⟨2⟩3. Set the variance parameter $\sigma^2 = \frac{1}{m_1^2}\sum_{i=1}^{m_1} \mathbb{E}\left(\|\mathbf{v}_i\|_2^2\right) \leq \frac{1}{m_1}\Upsilon(\mathscr{S}, \boldsymbol{\pi})$.

⟨2⟩4. With these moment bounds, the vector Bernstein inequality in Lemma A.0.2 gives

the tail bound

$$\mathbb{P}\left(\left\{\left\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\right\|_2 > (1/3)\|\mathbf{v}\|_2\right\}\right) \leq 2\exp\left(-\frac{m_1}{8}\min\left\{\frac{1}{3\Upsilon(\mathscr{S},\boldsymbol{\pi})}, \frac{1}{\Theta(\mathscr{S},\boldsymbol{\pi})}\right\}\right).$$

⟨2⟩5. This probability is less than $\varepsilon/3L$ if $m_1 \geq 8\cdot\max\{3\Upsilon(\mathscr{S},\boldsymbol{\pi}),\Theta(\mathscr{S},\boldsymbol{\pi})\}\cdot\log\left(\frac{6L}{\varepsilon}\right)$.

□

⟨1⟩4. For each $l \in [L]$, I claim that both $\mathbb{P}(\{\|\mathbf{P}_{\mathscr{S}}\mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\|_\infty > (1/4)\|\mathbf{v}\|_\infty\}) \leq \varepsilon/3L$ and its complement $\mathbb{P}(\{\|(\mathbf{I}_N - \mathbf{P}_{\mathscr{S}})\mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\|_\infty > (1/4)\|\mathbf{v}\|_\infty\}) \leq \varepsilon/3L$ hold, provided that the number of rows satisfies $m_l \geq 8\cdot\max\{3\Upsilon(\mathscr{S},\boldsymbol{\pi}),\Theta(\mathscr{S},\boldsymbol{\pi})\}\cdot\log\left(\frac{6N}{\varepsilon}\right)$.
PROOF: Once again, the problem is the same for each $l \in [L]$, so we consider only $l = 1$. Fix $k \in [N]$, some $\mathbf{v} \in \mathbb{B}_\infty^N$, and write

$$X := \left\langle \tilde{\mathbf{e}}_k, \mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\right\rangle = \frac{1}{m_1}\sum_{i=1}^{m_1}\underbrace{\left\langle \tilde{\mathbf{e}}_k, \left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\mathbf{P}_{\mathscr{S}}\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}\mathbf{v}\right\rangle}_{=:X_i}$$

a sum of independent, zero-mean random variables $\{X_i\}_i$.

⟨2⟩1. Bound each X_i almost surely with $\Theta(\mathscr{S},\boldsymbol{\pi})$:

$$\begin{aligned}|X_i| &= \left\|\tilde{\mathbf{e}}_k^\top\left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\mathbf{P}_{\mathscr{S}}\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}\right\|_1 \cdot \underbrace{\|\mathbf{v}\|_\infty}_{\leq 1} \\ &\leq \Theta(\mathscr{S},\boldsymbol{\pi}).\end{aligned}$$

⟨2⟩2. Bound the second moment as

$$\begin{aligned}\mathbb{E}\left(|X_i|^2\right) &= \sum_{i=1}^n \pi_i\left|\left\langle \tilde{\mathbf{e}}_k, \left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_{\omega_i}}\mathbf{e}_{\omega_i}\mathbf{e}_{\omega_i}^\top\mathbf{P}_{\mathscr{S}}\right)\mathbf{D}^+\right]^\top\mathbf{P}_{\mathscr{S}}\mathbf{v}\right\rangle\right|^2 \\ &\leq \Upsilon(\mathscr{S},\boldsymbol{\pi}).\end{aligned}$$

⟨2⟩3. Set the variance parameter $\sigma^2 = \frac{1}{m_1^2}\sum_{i=1}^{m_1}\mathbb{E}\left(|X_i|^2\right) \leq \frac{1}{m_1}\Upsilon(\mathscr{S},\boldsymbol{\pi})$.

⟨2⟩4. The scalar Bernstein inequality in Lemma A.0.1 gives the tail bound

$$\mathbb{P}\left(\left\{\left\langle \tilde{\mathbf{e}}_k, \mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\right\rangle > (1/4)\|\mathbf{v}\|_2\right\}\right) \leq 2\exp\left(-\frac{3m_1}{32}\min\left\{\frac{1}{4\Upsilon(\mathscr{S},\boldsymbol{\pi})}, \frac{1}{\Theta(\mathscr{S},\boldsymbol{\pi})}\right\}\right)$$

for each fixed $k \in [N]$.

⟨2⟩5. Taking the union bound over $\mathscr{S}$ and then over $N \setminus [N]$ yields in turn

$$\mathbb{P}\left(\left\{\left\langle \tilde{\mathbf{e}}_k, \mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\right\rangle > (1/4)\|\mathbf{v}\|_2\right\}\right) \leq 2|\mathscr{S}| \times \exp\left(-\frac{3m_1}{32}\min\left\{\frac{1}{4\Upsilon(\mathscr{S},\boldsymbol{\pi})}, \frac{1}{\Theta(\mathscr{S},\boldsymbol{\pi})}\right\}\right)$$

$$\mathbb{P}\left(\left\{\left\langle \tilde{\mathbf{e}}_k, \mathbf{M}^{(l)}\mathbf{P}_{\mathscr{S}}\mathbf{v}\right\rangle > (1/4)\|\mathbf{v}\|_2\right\}\right) \leq 2(N-|\mathscr{S}|) \times \exp\left(-\frac{3m_1}{32}\min\left\{\frac{1}{4\Upsilon(\mathscr{S},\boldsymbol{\pi})}, \frac{1}{\Theta(\mathscr{S},\boldsymbol{\pi})}\right\}\right)$$

⟨2⟩6. Both these probabilities are less than $\varepsilon/3L$ if $m_1 \geq \frac{32}{3}\cdot\max\{4\Upsilon(\mathscr{S},\boldsymbol{\pi}),\Theta(\mathscr{S},\boldsymbol{\pi})\}\cdot\log\left(\frac{6NL}{\varepsilon}\right)$.

⟨1⟩5. We now call upon the definition of $\Gamma(\mathscr{S},\boldsymbol{\pi}) := \max\{4\Upsilon(\mathscr{S},\boldsymbol{\pi}),\Theta(\mathscr{S},\boldsymbol{\pi})\}$. All requirements in Lemma 5.3.1 depending on matrices $\mathbf{M}^{(l)}$ are simultaneously attained if

$$m = \sum_{l=1}^{L} m_l \geq \frac{32L}{3}\cdot\Gamma(\mathscr{S},\boldsymbol{\pi})\cdot\log\left(\frac{6NL}{\varepsilon}\right),$$

whereas the requirements depending on matrix $\mathbf{M}$ are enforced if

$$m \geq 24\cdot\Theta(\mathscr{S},\boldsymbol{\pi})\cdot\log\left(\frac{6N}{\varepsilon}\right).$$

Recalling that $L \geq 2 + \left\lceil\frac{\log|S|}{2\log 3}\right\rceil$, if suffices then to set

$$m \geq \frac{32}{3}\left(2+\frac{\log|\mathscr{S}|}{2\log(3)}\right)\cdot\Gamma(\mathscr{S},\boldsymbol{\pi})\cdot\log(|\mathscr{S}|)\cdot\log\left(\frac{6\cdot N\cdot\left(2+\frac{\log|\mathscr{S}|}{2\log(3)}\right)}{\varepsilon}\right),$$

which can be simplified to $m \geq 38\cdot\Gamma(\mathscr{S},\boldsymbol{\pi})\cdot\log(|\mathscr{S}|)\cdot\log\left(\frac{63\cdot N\cdot\log(|\mathscr{S}|)}{\varepsilon}\right)$ if we assume $|\mathscr{S}| \geq 3$. [21]

⟨1⟩6. Q.E.D.

All the conditions of Lemma 5.3.1 hold simutaneously with probability larger than $1-\varepsilon$. Therefore — with the same likelihood — the golfing scheme certifies $\mathbf{x}$ to be the unique solution of (P1).

[21] We do not lose in doing so, since a co-support $\mathscr{S} = \mathrm{supp}(\mathbf{Dx})$ of 3 is orders of magnitude below what is normally encountered in applications.

6 A numerical tour

The goal of this chapter is to display how the theoretical discussion presented thus far relates to the practice of signal recovery on graphs. First of all, I show how to implement the Graph Total Variation ($\mathscr{G}$-TV) decoders as efficient iterative procedures derived from a proximal splitting technique. They are efficient in the sense that they require only a sequence of sparse matrix-vector multiplications and some inexpensive elementwise operations. Then, I present four datasets with signals that exhibit small jump-sets with respect to their graph support [1]. The first contains draws of random graphs under the Stochastic Block Model (SBM); the corresponding signals are the community-indicator vectors. The SBM is traditionally used to emulate clustered networks, so this data could be seen as a sort of baseline for comparisons. The next two datasets have each a single, fixed graph. The `email-EU-core` network encodes email exchanges across departments in a European research institution; the `swiss-national-council` network [2] connects members of the Swiss Parliament by "voting similarity". The piecewise-constant signals in each of these two are derived from the "natural" communities in their respective graphs. The final dataset, `BSDS300`, relates to an image segmentation task. Each graph in this collection represents a natural image by mapping pixels to vertices and patch color similarity to edges. The segmentation masks of the images can then be interpreted as piecewise-constant graph signals. The second half of the chapter is a series of numerical experiments on our four datasets. For information on how to reproduce them, visit

https://github.com/rodrigo-pena/phd-thesis/blob/master/python/README.md

To supplement a discussion from Chapter 3, I show how the behavior of the interpolation error changes if we minimize the $\mathscr{G}$-TV semi-norm $\|\mathbf{D}\cdot\|_1$ or the Dirichlet form $\|\mathbf{D}\cdot\|_2^2$. Under $\mathscr{G}$-TV minimization, the recovery error goes through a sharp phase transition as the number of measurements increases; under the Dirichlet form, the error decays smoothly. The main set

[1] Recall that small jump-sets are what qualify a graph signal as piecewise-constant.

[2] The same from Chapter 1.

of experiments then investigates the sampling design's effect over the recovery error of $\mathscr{G}$-TV interpolation. For that, I compare uniform random sampling with two designs inspired by the results of Chapter 5. To reduce the number of samples needed for recovery, the experiments show that it is important to know more about the signal's jump-set than just its size.

6.1 Implementing the $\mathscr{G}$-TV decoders with proximal splitting

There is a base algorithm that can be used to solve both [3]

$$\min_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{D}\mathbf{z}\|_1 \text{ such that } \mathbf{P}_\Omega\mathbf{z} = \mathbf{P}_\Omega\mathbf{x} \tag{P1}$$

and

$$\min_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{D}\mathbf{z}\|_1 \text{ subject to } \|\mathbf{P}_\Omega\mathbf{z} - \mathbf{y}\|_q^q \leq \eta. \tag{P1-η}$$

To use it, I need to state each of these problems in the generic unconstrained form

$$\min_{\mathbf{z}\in\mathbb{R}^n} \; f(\mathbf{z}) + g(\mathbf{D}\mathbf{z}) + h(\mathbf{z}) \tag{6.1}$$

using convex functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^N \to \mathbb{R}$, and a function $h : \mathbb{R}^n \to \mathbb{R}$ that is both convex *and* differentiable. I will do that first for problem (P1), with help from a *convex indicator function*.

Definition 6.1.1 (Convex indicator function)**.** The convex indicator function of a set $\mathscr{C}$ is the mapping

$$\mathbf{z} \mapsto \iota_{\mathscr{C}}(\mathbf{z}) = \begin{cases} 0, & \text{if } \mathbf{z} \in \mathscr{C}, \\ +\infty & \text{otherwise.} \end{cases} \tag{6.2}$$

This function is useful whenever one needs to transform a constrained problem like $\min_{\mathbf{z}\in\mathscr{C}} f(\mathbf{z})$ into its unconstrained equivalent $\min_{\mathbf{z}\in\mathbb{R}^n} f(\mathbf{z}) + \iota_{\mathscr{C}}(\mathbf{z})$. The *interpolation* problem has a constraint set $\mathscr{C} := \{\mathbf{z} \in \mathbb{R}^n : \mathbf{P}_\Omega\mathbf{z} = \mathbf{P}_\Omega\mathbf{x}\}$, so the template (6.1) represents (P1) if we set

$$f(\cdot) = \iota_{\{\mathbf{z}\in\mathbb{R}^n : \mathbf{P}_\Omega\mathbf{z} = \mathbf{P}_\Omega\mathbf{x}\}}(\cdot) \tag{6.3}$$

$$g(\cdot) = \|\cdot\|_1 \tag{6.4}$$

$$h(\cdot) \equiv 0. \tag{6.5}$$

For the *regression* problem, I will proceed differently. Let us focus on the case when $q = 2$ in

[3] Here I incur in a slight abuse of notation, which was already foreseen in Chapter 2, Section 2.2. Still (from the implementation's perspective) it is more convenient to work with the square projection operator $\mathbf{P}_\Omega \in \mathbb{R}^{n\times n}$ than with the rectangular sampling matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$. Through zero-padding, one can avoid ever going from $\mathbb{R}^n$ to $\mathbb{R}^m$. Furthermore, identifying $\mathbf{A}$ with $\mathbf{P}_\Omega$ turns matrices like $\mathbf{A}^\top\mathbf{A}$ — that would appear in the algorithms — into $\mathbf{P}_\Omega^\top\mathbf{P}_\Omega = \mathbf{P}_\Omega$, because the later is a projection operator. Overall, the algorithms' statement is cleaner using $\mathbf{P}_\Omega$.

the error estimate, so that $\mathbf{z} \mapsto \left\|\mathbf{P}_\Omega \mathbf{z} - \mathbf{y}\right\|_2^2$ is a convex, differentiable function. Then there is some regularization hyperparameter ρ for which the choice

$$f(\cdot) = 0 \tag{6.6}$$

$$g(\cdot) = \|\cdot\|_1 \tag{6.7}$$

$$h(\cdot) = \frac{\rho}{2}\|\mathbf{P}_\Omega(\cdot) - \mathbf{y}\|_2^2 \tag{6.8}$$

expresses in (6.1) the same problem as (P1-η) [9, Ch. 5].

The Forward-Backward-Forward (FBF) primal-dual procedure in Komodakis and Pesquet [39, Algorithm 6] is a numerical solver for any problem of the form (6.1). It is based on a proximal splitting technique, alternating between (forward) gradient steps and (backward) calls to a proximity operator. Between each forward and backward steps the matrix $\mathbf{D}$ connects the primal and dual spaces. [4] Interested readers can find a more detailed account in Appendix 6.A. Meanwhile, I present Algorithms 2 and 3 as the respective translations of the base FBF procedure for the particular problems (P1) and (P1-η). Therein, $\mathrm{soft}_\gamma(\cdot)$ represents the soft thresholding operation, defined coordinatewise by

$$\forall i,\ \ w_i \mapsto \begin{cases} w_i - \gamma, & \text{if } w_i > \gamma \\ 0, & \text{if } |w_i| \leq \gamma \\ w_i + \gamma, & \text{if } w_i < -\gamma \end{cases}.$$

This function shows up because it is the proximity operator associated with the ℓ_1 norm. The algorithms admit some leeway in specifying the step sizes for each iteration. As long as the sequence $(\gamma_n)_{n\in\mathbb{N}}$ stays within the given intervals, convergence is guaranteed [39].

Algorithm 2 FBF primal-dual iterations for solving (P1)

$\mathbf{z}_0 \leftarrow \mathbf{0} \in \mathbb{R}^n$ ▷ Primal variable
$\mathbf{d}_0 \leftarrow \mathbf{0} \in \mathbb{R}^N$ ▷ Dual variable
repeat
  pick $\gamma_n \in \left(0, \frac{1}{1+\|\mathbf{D}\|_2}\right)$
  $\left(\mathbf{w}_{1,n},\ \mathbf{w}_{2,n}\right) \leftarrow \left(\mathbf{z}_n - \gamma_n \mathbf{D}^\top \mathbf{d}_n,\ \mathbf{d}_n + \gamma_n \mathbf{D}\mathbf{z}_n\right)$ ▷ Forward
  $\left(\mathbf{p}_{1,n},\ \mathbf{p}_{2,n}\right) \leftarrow \left((\mathbf{I}_n - \mathbf{P}_\Omega)\mathbf{w}_{1,n} + \mathbf{P}_\Omega \mathbf{x},\ \mathbf{w}_{2,n} - \mathrm{soft}_{\gamma_n}(\mathbf{w}_{2,n})\right)$ ▷ Backward
  $\left(\mathbf{q}_{1,n},\ \mathbf{q}_{2,n}\right) \leftarrow \left(\mathbf{p}_{1,n} - \gamma_n \mathbf{D}^\top \mathbf{p}_{2,n},\ \mathbf{p}_{2,n} + \gamma_n \mathbf{D}\mathbf{p}_{1,n}\right)$ ▷ Forward
  $\left(\mathbf{z}_{n+1},\ \mathbf{d}_{n+1}\right) \leftarrow \left(\mathbf{z}_n - \mathbf{w}_{1,n} + \mathbf{q}_{1,n},\ \mathbf{d}_n - \mathbf{w}_{2,n} + \mathbf{q}_{2,n}\right)$
until convergence
return $\mathbf{z}_{n+1}$

[4] The primal space in our problems is $\mathbb{R}^n$ (where $\mathbf{x}$ lives), while the dual space is $\mathbb{R}^N$ (the co-domain of $\mathbf{D}$).

Algorithm 3 FBF primal-dual iterations for solving (P1-η)

$\mathbf{z}_0 \leftarrow \mathbf{0} \in \mathbb{R}^n$ ▷ Primal variable
$\mathbf{d}_0 \leftarrow \mathbf{0} \in \mathbb{R}^N$ ▷ Dual variable
repeat
  pick $\gamma_n \in \left(0, \frac{1}{1+\rho+\|\mathbf{D}\|_2}\right)$
  $\left(\mathbf{w}_{1,n},\ \mathbf{w}_{2,n}\right) \leftarrow \left(\mathbf{z}_n - \gamma_n\left[\rho(\mathbf{P}_\Omega\mathbf{z}_n - \mathbf{y}) + \mathbf{D}^\top\mathbf{d}_n\right],\ \mathbf{d}_n + \gamma_n\mathbf{D}\mathbf{z}_n\right)$ ▷ Forward
  $\left(\mathbf{p}_{1,n},\ \mathbf{p}_{2,n}\right) \leftarrow \left(\mathbf{0},\ \mathbf{w}_{2,n} - \mathrm{soft}_{\gamma_n}(\mathbf{w}_{2,n})\right)$ ▷ Backward
  $\left(\mathbf{q}_{1,n},\ \mathbf{q}_{2,n}\right) \leftarrow \left(\mathbf{p}_{1,n} - \gamma_n\left[\rho(\mathbf{P}_\Omega\mathbf{z}_n - \mathbf{y}) + \mathbf{D}^\top\mathbf{p}_{2,n}\right],\ \mathbf{p}_{2,n} + \gamma_n\mathbf{D}\mathbf{p}_{1,n}\right)$ ▷ Forward
  $\left(\mathbf{z}_{n+1},\ \mathbf{d}_{n+1}\right) \leftarrow \left(\mathbf{z}_n - \mathbf{w}_{1,n} + \mathbf{q}_{1,n},\ \mathbf{d}_n - \mathbf{w}_{2,n} + \mathbf{q}_{2,n}\right)$
until convergence
return $\mathbf{z}_{n+1}$

Other algorithms could certainly be used to implement (P1) and (P1-η), but the FBF procedure is good enough for two reasons. The first of these, I have already mentioned: the same general algorithm applies to both problems. The second reason has to do with numerical efficiency. Each step in Algorithms 2 and 3 requires only elementwise operations, and matrix-vector multiplications using the graph gradient operator $\mathbf{D}$. Graphs used in practice are often sparse [5], so that $\mathbf{D}$ has few non-zero entries, making multiplications with it cheap to compute.

6.2 The data

To test the $\mathscr{G}$-TV decoders, we should let them try to recover the sort of signal that motivated their study. Signals with a small jump-set can always be found in the indicator vectors of the communities in a clusterable graph. I present here four datasets, drawn from different domains, yielding graphs with different cluster structures. The clusters are reflected — in varying degrees — on the constant pieces of the assembled signals therein. Each dataset in this section is used in at least one experiment later on in the chapter.

6.2.1 Community indicator vectors in the Stochastic Block Model

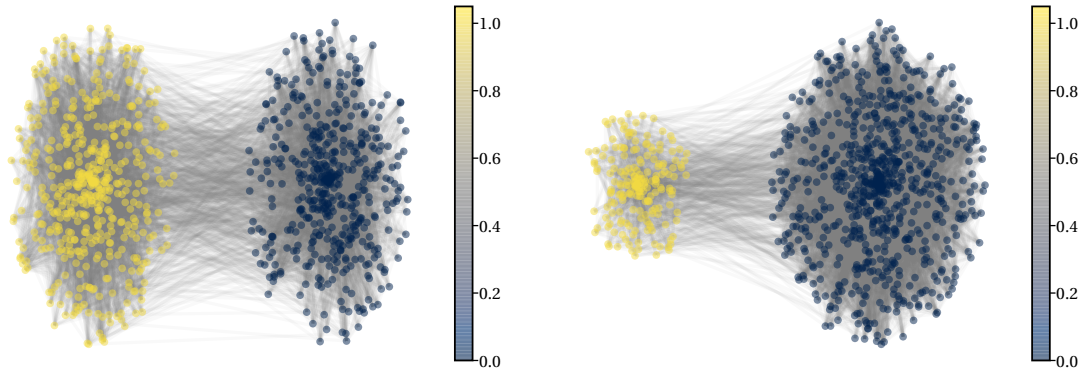
The standard way to simulate graphs with a community structure is via the Stochastic Block Model (SBM)[6]. The SBM describes a *distribution* of random graphs, and one may refer to it using the notation SBM$(n, \mathscr{C}, \mathbf{p}, \mathbf{Q})$ when the drawn graphs have a fixed number n of vertices [1]. The set $\mathscr{C} = \{\mathscr{C}_1, \ldots, \mathscr{C}_k\}$ lists the communities, which are themselves lists of vertices in the graph. The vector $\mathbf{p} = (p_1, \ldots, p_k)$ lists the probabilities with which vertices within each

[5] Sparse graphs are ones that have few edges. Generally, graphs are considered sparse their number of edges is on the order of their number of vertices. Dense graphs can have a number of edges on the order of the square of the number of vertices. As the number of vertices grows, storing dense graphs quickly becomes a problem on most systems.

[6] According to Abbe [1], this is the most commonly used name for it in the Statistics literature. The model is also known as "planted partitions" in Computer Science, or "inhomogeneous random graph" in Mathematics.

community connect to each other. [7] The matrix **Q**, in turn, gathers the probabilities of connection across communities. [8].

In my experiments, I only consider SBM graphs with two communities, having the same internal connection probability p. In other words, $p_1 = p_2 = p$. Since there are only two communities, the matrix **Q** can be reduced to a single probability parameter q. I will use shorthand and refer to this restricted distribution as SBM$(n, 2, p, q)$. If the two communities in the model are the same size, they become statistically indistinguishable, or "symmetric". I will sometimes highlight this setting by referring to it as 2-SSBM while calling the (potentially) unbalanced alternative 2-SBM. Note that I might also write the partition of vertices between communities in evidence as a sum of two terms. For example, the unbalanced 2-SBM of Figure 6.1 has 200 vertices in the first community and 800 in the second, hence the appended "(200 + 800)". Balanced or unbalanced, the SBM experiments always use the indicator vector of the smallest community as the piecewise-constant signal. For matters of graph signal processing, I enforce a unitary edge weight between connected vertices in the SBM graphs. Thus, their gradient matrix, **D**, only has entries valued -1, 0, or 1.



(a) Balanced 2-SSBM(500 + 500)

(b) Unbalanced 2-SBM(200 + 800)

Figure 6.1 – Graphs drawn from the SBM$\left(1000, 2, 4.5\frac{\log(1000)}{1000}, 0.5\frac{\log(1000)}{1000}\right)$ model, varying the relative community sizes. The vertex colors represent the indicator vector of the leftmost community.

My use of community indicator vectors is inspired by one of the basic questions in the study of Stochastic Block Models: can one retrieve the community partitions by looking at the edge structure of graphs drawn from the distribution? Researchers like Abbe *et al.* [2, 1] have used the Maximum a Posteriori (MAP) estimator as a tool to prove when this question can be answered. Theorem 6.2.1 shows this solvability threshold for the parameters of the 2-SSBM.

[7] For example, let v_i, v_j be two vertices in the same community, $\mathscr{C}_1$, and write $v_i \sim v_j$ if these vertices are connected. Then $\mathbb{P}(\{v_i \sim v_j\}) = p_1$.

[8] Take $v_i \in \mathscr{C}_1$ again, but v_j in another community, $\mathscr{C}_2$. Then $\mathbb{P}(\{v_i \sim v_j\}) = Q_{12}$.

Theorem 6.2.1 ([1, Thm. 7.1])**.** *Exact recovery in the symmetric SBM*$\left(n, 2, a\frac{\log n}{n}, b\frac{\log n}{n}\right)$ *is solvable if and only if*

$$\left(\sqrt{a}-\sqrt{b}\right)^2 > 2. \tag{6.9}$$

To retrieve the community partitions, the graphs drawn from the SBM distribution have to be at least *connected* [9]. Abbe [1] points out that $a+b>2$ is the parameter regime for which the 2-SSBM graphs are connected with high probability. The theorem above implies then that $2(\sqrt{ab}-1)$ is the extra factor needed to go from connectivity to exact recovery. But Abbe's result concerns retrieving the community partitions when all we can observe is the edge structure of the SBM graphs. When I sample the community indicator vectors, I effectively call upon an oracle to reveal the true assignment of certain vertices. How does the recovery error of the $\mathscr{G}$-TV interpolation (P1) behaves for SBM distributions with parameter regimes in the neighborhood of the one from Theorem 6.2.1?

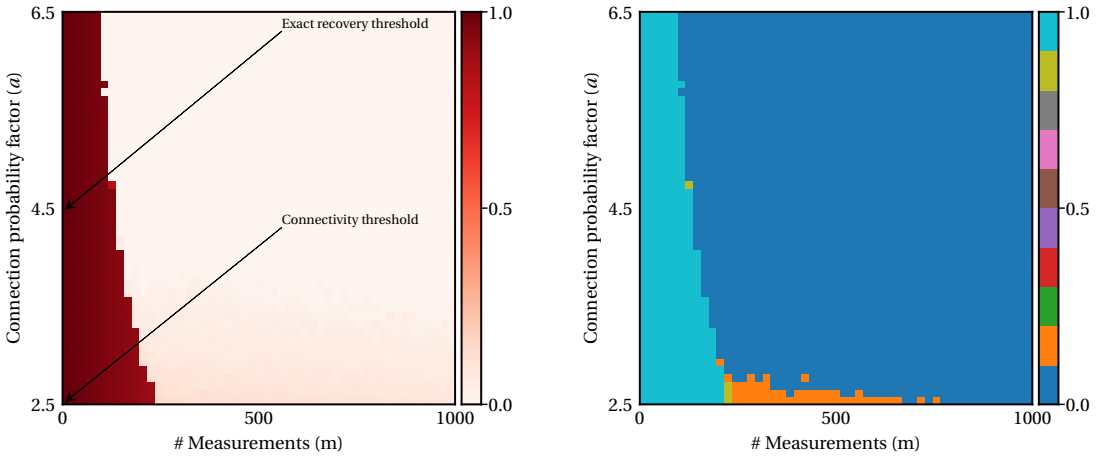
Figure 6.2 shows an initial answer to that question for the 2-SSBM. There, I use uniform random sampling (with replacement) to query the vertex community assignments. The number of measurements varies from zero to the number of vertices [10]. On the vertical axis, I vary the intra-community connection parameter a, starting from the connectivity regime ($a+b>2$) — in the bottom — and passing by the solvability regime from Theorem 6.2.1 — in the middle. The higher up on the vertical axis, the denser each community tends to be in terms of the number of edges. The average number of edges connecting vertices between different communities stays the same throughout. I measure the recovery error as a normalized [11] Euclidean distance between the signal estimated by (P1) and the "ground-truth" indicator vector. The plot to the right of the figure is a quantized version of the one to the left, presented so that we can better distinguish the level sets in the recovery error.

Note the clear phase transition when the number of samples reaches a critical value. The threshold happens earlier the higher up on the plot, where the number of edges within the communities becomes progressively larger (on average) than the size of the indicator vector's jump-set. But even the bottom half of the plot exhibits very small recovery errors. This observation does not contradict Theorem 6.2.1 — of course — but simply attests to the extra information included in the measurements themselves. To allow comparisons with this initial plot, the remaining SBM experiments in this chapter employ the SBM$\left(1000, 2, a\frac{\log(1000)}{1000}, 0.5\frac{\log(1000)}{1000}\right)$ distributions within the same parameter range, *i.e.*,with a taking values in the interval $[2.5, 6.5]$.

[9] Recall from Chapter 2 that a graph is connected if one can visit all the vertices by traveling only on the edges of the graph.

[10] Note however that since the sampling is with replacement, having the number of measurements equal to the number of vertices does not automatically imply zero recovery error because there can be redundant samples.

[11] The normalization factor is the inverse of the Euclidean norm of the ground-truth signal.



(a) Recovery error (normalized Euclidean distance)

(b) Quantized version

Figure 6.2 – Phase transition for the error of (P1) in recovering the community indicator vector of graphs drawn from the symmetric SBM$\left(1000, 2, a\frac{\log(1000)}{1000}, 0.5\frac{\log(1000)}{1000}\right)$ from uniform random samples. Each pixel represents the median recovery error across 25 independent trials.

6.2.2 Department indicator vectors in `email-EU-core`

The `email-EU-core` data [12] represents some email exchanges by people from a large European research institution [80]. The institution is split into 42 departments, and each of the 1005 individuals in the dataset belongs to exactly one such department. An edge of unitary weight connects two people in the network if they exchanged at least one email. As one could expect, communication tends to stay restricted to within departments, so the network clusters are a reflection of the departmental makeup. Thus, the department indicator vectors have small jump sets, as edges across departments are fewer than within. But the number of people in each department is relatively small, so to be fairer with the baseline uniform sampling design I pick as "ground-truth" signal for the `email-EU-core` dataset the indicator vector of the union of the five largest departments. In doing so, the ground-truth takes value 1 at about 38% of the vertices and value 0 at the rest 62%.

6.2.3 Party indicator vectors in `swiss-national-council`

I construct the `swiss-national-council` with data extracted from the 50$^{\text{th}}$ legislature in the Swiss Database of Parliamentary Votes [13]. I associate each council member to a feature vector indicating how they voted in each of 3395 affairs accounted for at the time of writing. If they voted `Yes` in some affair, the corresponding entry in the feature vector receives value 1; if they voted `No` the entry gets value -1. Affairs for which the councilor did not register a vote

[12] http://snap.stanford.edu/data/email-Eu-core.html
[13] https://www.parlament.ch/en/ratsbetrieb/abstimmungen/abstimmungs-datenbank-nr

(whatever the reason) get a value of 0. The idea underneath this numerical translation is to use the feature vectors to compare the voting patterns of different council members. Investigating which metric is best for this task is beyond the scope of this thesis, so I opt for the traditional Euclidean distance.

The graph construction takes place directly at the level of the weighted adjacency matrix. Let $\mathbf{f}_i$ and $\mathbf{f}_j$ the the feature vectors of councillors i and j. These councillors are then connected with edge weight $W_{ij} = \exp\left(-\|\mathbf{f}_i - \mathbf{f}_j\|_2^2/\sigma\right)$, where σ is set as the mean distance between feature vectors in the dataset. [14] Then, to enforce a sparse graph, I keep only the 25 largest edge weights for each council member, setting the rest to 0 (while making sure that the weight matrix remains symmetric). In the end, councilors that voted identically throughout all the affairs get a unitary edge weight, whereas members that voted differently enough are disconnected.

To put a signal on this graph, I went over the Swiss Parliament files listing the party affiliations of all councilors since 1848. [15] Intuition tells us that members of the same party stand to vote more similarly than people from different parties. So the party indicator vectors should strongly correlate with the cluster structure of the graph I have constructed. I decided to use as "ground-truth" for the experiments the indicator vector of the two largest right-wing parties in the 50$^{\text{th}}$ legislature, UDC and FDP. By joining these two parties, the ground-truth signal takes value 1 at around 46% of the vertices and 0 at the other 54%, an almost even split of the National Council that improves the chances of recovery under uniform random sampling. Meanwhile — because the two joined parties are to the right in political spectrum — this concocted ground-truth should still have a sparse jump-set since the members of UDC and FDP have different voting patterns than the mostly centrist and left-wing politicians in the other half of the Council.

6.2.4 Image segmentation masks in `BSDS300`

The Berkeley Segmentation Dataset and Benchmark [16] contains 300 images, split into train (100) and test (200) sets. Each of these images is accompanied by human-generated segmentation masks. What I call the `BSDS300` dataset is a collection of graphs for each of the 300 images, along with vector representations of the segmentation masks — interpreted as graph signals.

The first step in constructing a graph for each image is to identify each pixel in the image with a vertex in the graph. [17] Then, I connect — with unitary edge weights — pixels that are

[14] Exponential kernels such as these are commonly used in Machine Learning and Data Science for similarity computations. The exponential always outputs positive similarity values, and the negative exponent ensures that these values decay quickly towards zero as the compared feature vectors become more distant.

[15] https://www.parlament.ch/en/ratsmitglieder

[16] https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/

[17] To get graphs of manageable size for the experiments, I subsample the original images by a factor of 12 before starting the graph construction. That is, I keep every 12th row and column of the original image, and use the pixels of this lower-resolution version as the vertices of the graph.

adjacent *in the image*. This first set of connections encodes "spatial similarity": neighboring pixels in the image are neighbors in the graph. The next set of connections encodes color information via "patch similarity". To understand it, let $\mathbf{p}_i \in \mathbb{R}^{147}$ be a vectorized, $7 \times 7 \times 3$ patch containing the RGB values in the image neighborhood centered at pixel i. [18] These color patches are used as feature vectors in the same way as voting data was in the construction of the `swiss-national-council` graph. I compute edge weights $W_{ij} = \exp\left(-\|\mathbf{p}_i - \mathbf{p}_j\|_2^2/\sigma\right)$ for each pair of pixels i and j, with σ set as the mean patch distance in the image. For the sake of a sparse graph, I then set most of these weights to zero, except for the 3 largest attached to each pixel. [19] The final step is to add the "spatial" and "patch" weight matrices, whose result represents a hybrid graph that connects pixels either because they are next to each other in the original image grid or because they are the center of similar color patches.

Each segmentation mask in the Berkeley dataset is seen as a piecewise-constant signal for the graph of the image that the mask segments. Step-by-step, I assign an arbitrary integer to each segment in the mask. Then — for example —, if a pixel in the original image belongs to segment number 7, its corresponding vertex in the graph gets a signal value of 7. This process goes for each vertex until we have a piecewise-constant graph signal that represents the segmentation of the original image. This signal should have a corresponding small jump-set because the spatial and color information encoded in the graph structure should correlate with the image segmentation. After all, humans often select as segments continuous objects with homogeneous colors. The edges connecting pixels within the same segment should then be more numerous than the ones across.

6.2.5 Data summary

For future reference, the table on the next page summarizes some key characteristics of the datasets used in this chapter. The number of vertices in the SBM models is fixed, but since they represent distributions of random graphs, I give their corresponding edge counts as expected values. There is a range of values to these expected values because I use SBM models with varying intra-community connection probabilities. The `BSDS300` line is special in the sense that there is a different ground-truth signal for each image in the dataset, hence the range of values in the number of edges in the signal's jump-set. As a final note, the complete, undirected graph on 1000 vertices has 499500 edges, so all the graphs in these four datasets are comparatively edge-sparse.

[18] If i is a border pixel, use some form of padding to obtain the patch.

[19] This is approximately what happens, because the weight matrix has to be symmetric as well in the end.

Dataset	Number of vertices	Number of edges	Signal to be recovered	Edges in the jump-set
`2-SSBM`(500 + 500)	1000	5172 – 12066	Indicator vector of one of the communities	863
`2-SBM`(200 + 800)	1000	6416 – 15796	Indicator vector of the smallest community	553
`email-EU-core`	1005	16064	Indicator vector of the five largest departments	4581
`swiss-national-council`	228	4259	Indicator vector of the two largest right-wing parties	641
`BSDS300`	1107	2977 – 5198	Segmentation mask	102 – 2617

Table 6.1 – Summary of the data used in our numerical tour. The edge counts for the SBM rows are expected values for the parameter regime considered in the experiments.

6.3 Graph Total Variation ($\mathscr{G}$-TV) vs. Dirichlet form

In Chapter 3, I argued for minimizing the $\mathscr{G}$-TV semi-norm $\|\mathbf{D}\mathbf{z}\|_1$ against the the Dirichlet form $\|\mathbf{D}\mathbf{z}\|_2^2$ in the recovery of piecewise-constant signals. Back then, I used representer theorems to show that the solutions of $\mathscr{G}$-TV decoders depend less on the measurement matrix than their Dirichlet form counterparts. Here I will reinforce this argument by plotting how the recovery error in the two settings varies as we increase the number of vertex measurements that we take. To control for the sampling design, the sampled vertices are always chosen uniformly at random (with replacement) throughout this section.

Let us look first at the behavior of error when recovering the community indicator vector on SBM graphs. Figure 6.3 shows the difference between $\mathscr{G}$-TV and Dirichlet form interpolation considering unbalanced 2-SBM(200 + 800) graphs. Similarly to the first SBM plot in Figure 6.2, the vertical axes vary the intra-community connection probability — moving upwards yields denser communities —, and the horizontal axes vary the number of uniform random samples of the signal. On the one hand, there is a sharp phase transition on the $\mathscr{G}$-TV column, reminiscent of the one in Figure 6.2 for the 2–SSBM(500 + 500) dataset. But this time the transition curve is more to the right, a consequence of the size imbalance in the communities of 2–SBM(200 + 800). Under uniform random sampling, one needs to sample more often to get enough information on the smallest community. On the other hand, the recovery error in the Dirichlet form column decreases smoothly as one gathers more and more samples. Even if its error level-sets are almost indifferent to the density of connections within the communities, their smooth decrease is not fast enough to reach the lowest error levels of the $\mathscr{G}$-TV plot. All in all, $\mathscr{G}$-TV interpolation seems to rely more on the contrast between the ground-truth's jump-set and the rest of the edges in the graph; what impacts most the Dirichlet form decoder are the measurement constraints.

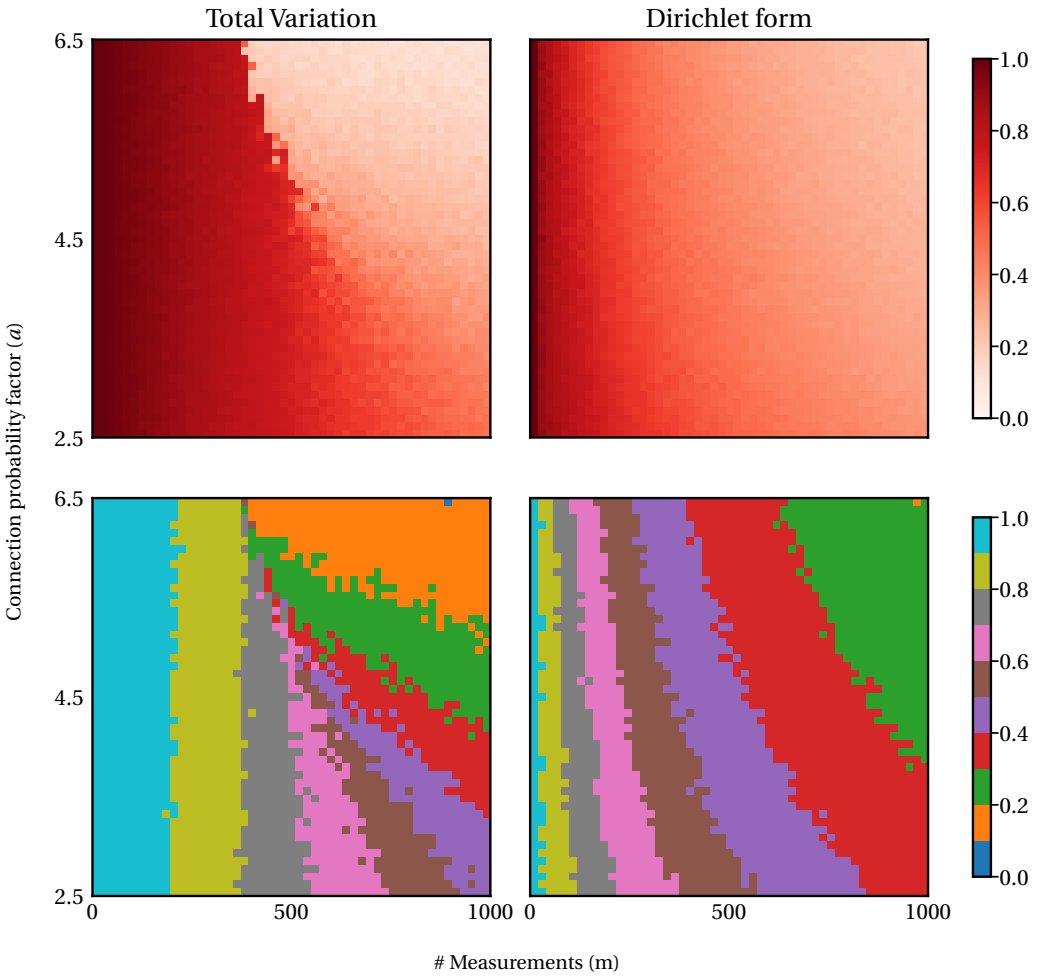


Figure 6.3 – Effect of the decoder's objective on the interpolation error when recovering the indicator vector of the smallest community on 2-SBM(200 + 800) graphs from samples taken uniformly at random. Left column: $\|\mathbf{Dz}\|_1$ (Graph Total Variation semi-norm). Right column: $\|\mathbf{Dz}\|_2^2$ (Dirichlet form). Each pixel on the top row represents the median, across 25 independent trials, of the normalized Euclidean distance from the recovered vector to the ground-truth. The bottom row has quantized versions of the plots from the top row, to better discern the error level-sets.

If we do the same comparison now using the `swiss-national-council` data, some of the same behaviors arise, but with interesting twists. See Figure 6.4. The $\mathcal{G}$-TV recovery error transitions sharply, but now this happens in two stages. My hypothesis for this behavior is that it is an artifact of the way I assembled the ground-truth signal. Although the UDC and FDP parties have a joint voting pattern that distinguishes them from the rest of the Council, these parties themselves do not vote identically in every affair. I suppose that the first drop in the error comes when the decoder approximately accounts for the largest, UDC component of the ground truth; the second drop would come when enough members of FDP are sampled

as belonging to the same signal piece as UDC. The error in Dirichlet form interpolation does not change in stages; it decreases smoothly and is even smaller than the $\mathscr{G}$-TV error in the beginning, just as in the previous experiment. The surprising observation this time is the error curves of the two decoders catching up at some point and proceeding to decrease smoothly at the same rate. The recovery errors to the right of the plot are still considerably large, but this might just be an indication that there are many councilors in the blue part of the ground-truth that vote very similarly to the UDC or FDP members. In other words, the large interpolation error is possibly a consequence of a large jump-set for the ground-truth signal.

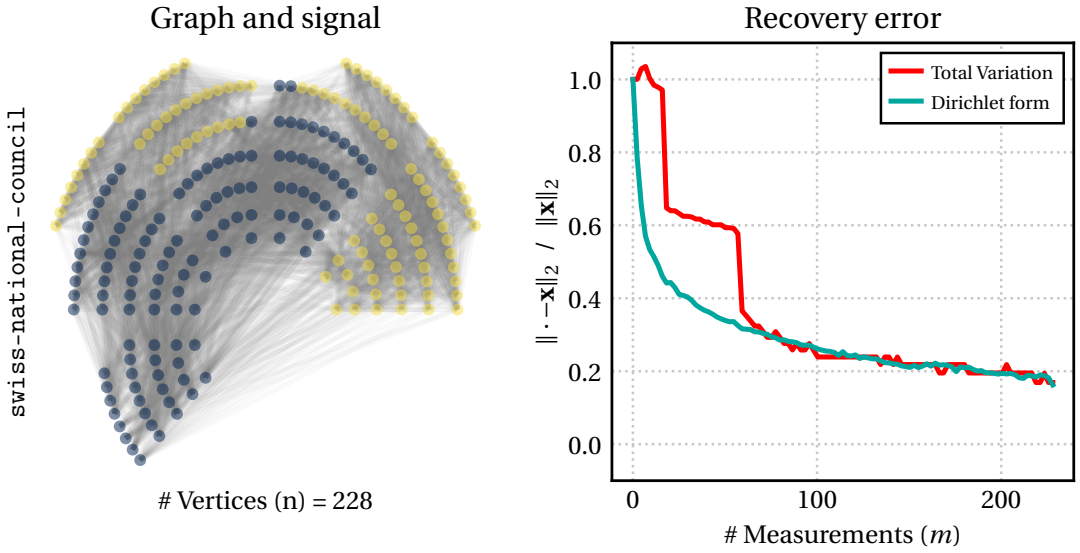


Figure 6.4 – Effect of the decoder's objective on the interpolation error when recovering the indicator vector of the two largest right-wing parties (UDC and FDP) in the `swiss-national-council` graph from samples taken uniformly at random. Red curve: $\|\mathbf{Dz}\|_1$ (Graph Total Variation semi-norm). Blue curve: $\|\mathbf{Dz}\|_2^2$ (Dirichlet form). Each point on the curves represents the median, across 51 independent trials, of the normalized Euclidean distance from the recovered vector to the ground-truth.

6.4 Effects of the sampling design for $\mathscr{G}$-TV interpolation

At the end of Chapter 5, I gave an explicit expression for an optimal vertex-sampling design when working with the $\mathscr{G}$-TV interpolation problem (P1). It prescribes sampling probabilities $\boldsymbol{\pi}$ that minimize $\Gamma(\mathscr{S}, \boldsymbol{\pi})$, a functional that also depends on the jump-set $\mathscr{S} := \operatorname{supp}(\mathbf{Dx})$ of the signal-to-be-recovered. This optimization program is easy to state, but hard to implement, all due to the indirect definition of the objective as $\Gamma(\mathscr{S}, \boldsymbol{\pi}) := \max\{\Theta(\mathscr{S}, \boldsymbol{\pi}), \Upsilon(\mathscr{S}, \boldsymbol{\pi})\}$. Out of the two random matrix moment estimates in the definition of Γ, it is Θ the one with the simplest expression. I will thus ignore Υ in this section and investigate sampling designs that minimize different upper bounds to Θ, comparing their behavior with the baseline uniform random sampling.

To define the designs we will enquire into, recall Definition 5.4.1 and consider the following succession of upper bounds for $\Theta(\mathcal{S}, \boldsymbol{\pi})$:

$$\begin{aligned}
\Theta(\mathcal{S}, \boldsymbol{\pi}) &= \max_{i\in[n]} \max_{k\in[N]} \left\| \tilde{\mathbf{e}}_k^\top \left[\mathbf{D}\left(\mathbf{I}_n - \frac{1}{\pi_i} \mathbf{e}_i \mathbf{e}_i^\top \right) \mathbf{D}^+ \right]^\top \mathbf{P}_{\mathcal{S}} \right\|_1 \\
&\leq \max_{k\in[N]} \left\| \tilde{\mathbf{e}}_k^\top \left[\mathbf{D}\mathbf{D}^+ \right]^\top \mathbf{P}_{\mathcal{S}} \right\|_1 + \max_{i\in[n]} \max_{k\in[N]} \frac{1}{\pi_i} \left\| \tilde{\mathbf{e}}_k^\top \left[\mathbf{D}\mathbf{e}_i \mathbf{e}_i^\top \mathbf{D}^+ \right]^\top \mathbf{P}_{\mathcal{S}} \right\|_1 \\
&\leq \sqrt{N} + \max_{i\in[n]} \max_{k\in[N]} \frac{1}{\pi_i} \left| \tilde{\mathbf{e}}_k^\top (\mathbf{D}^+)^\top \mathbf{e}_i \right| \cdot \left\| \mathbf{e}_i^\top \mathbf{D}^\top \mathbf{P}_{\mathcal{S}} \right\|_1 \\
&= \sqrt{N} + \max_{i\in[n]} \max_{k\in[N]} \frac{\left\| (\mathbf{D}^+)^\top \mathbf{e}_i \right\|_\infty \cdot \| \mathbf{P}_{\mathcal{S}} \mathbf{D}\mathbf{e}_i \|_1}{\pi_i} \qquad (6.10) \\
&\leq \sqrt{N} + \underbrace{\| \mathbf{P}_{\mathcal{S}} \|_{1\to 1}}_{=|\mathcal{S}|} \cdot \max_{i\in[n]} \max_{k\in[N]} \frac{\left\| (\mathbf{D}^+)^\top \mathbf{e}_i \right\|_\infty \cdot \| \mathbf{D}\mathbf{e}_i \|_1}{\pi_i}. \qquad (6.11)
\end{aligned}$$

The sampling design $\boldsymbol{\pi}$ minimizing the bound in (6.10) is the one that satisfies

$$\pi_i = \frac{\left\| (\mathbf{D}^+)^\top \mathbf{e}_i \right\|_\infty \cdot \| \mathbf{P}_{\mathcal{S}} \mathbf{D}\mathbf{e}_i \|_1}{\sum_{j=1}^n \left\| (\mathbf{D}^+)^\top \mathbf{e}_j \right\|_\infty \cdot \left\| \mathbf{P}_{\mathcal{S}} \mathbf{D}\mathbf{e}_j \right\|_1}, \forall i \in [n]. \qquad (6.12)$$

I call it the "*jump-set coherence*" design, because it depends on the jump-set $\mathcal{S}$ — through $\mathbf{P}_{\mathcal{S}}$ — and on the coherence [20] between $\mathbf{D}$ and the standard basis vectors in $\mathbb{R}^n$. [21] Note that this design is not necessarily practical, since it requires — *a priori* — knowing the jump-set of the "ground-truth" graph signal. Another sampling design arises when we minimize the the looser bound in (6.11):

$$\pi_i = \frac{\left\| (\mathbf{D}^+)^\top \mathbf{e}_i \right\|_\infty \cdot \| \mathbf{D}\mathbf{e}_i \|_1}{\sum_{j=1}^n \left\| (\mathbf{D}^+)^\top \mathbf{e}_j \right\|_\infty \cdot \left\| \mathbf{D}\mathbf{e}_j \right\|_1}, \forall i \in [n]. \qquad (6.13)$$

It no longer depends on the jump-set of $\mathbf{x}$, but still has the coherence terms. I name this design "*naive coherence*", because it is a consequence of naively controlling the jump-set $\mathcal{S}$ via its cardinality. Getting rid of the jump-set, though, has practical benefits: as long as we know the graph (represented in $\mathbf{D}$) we can implement the *naive coherence* design. There is no need to consider the inaccessible ground-truth signal. But there are other designs that are simple to implement; uniform random sampling being arguably the simplest. *Naive coherence* sampling only makes sense for our purposes if it implies a successful $\mathcal{G}$-TV interpolation using fewer measurements than under uniform random sampling, and not many more than under the *jump-set coherence* design. I contrast these three sampling strategies — summarized on Table 6.2 — in the set of experiments that follows.

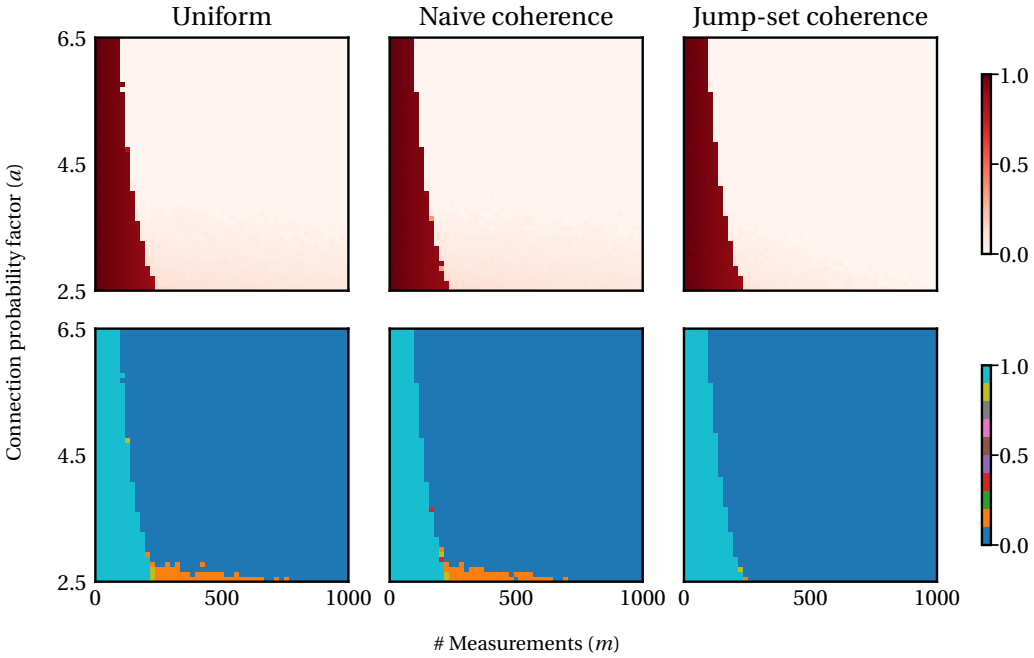
[20] I use "coherence" here in the sense of Compressed Sensing.

[21] Recall that our sampling matrix is a stack of standard basis vectors, so the quantified coherence is really between the analysis and measurement operators.

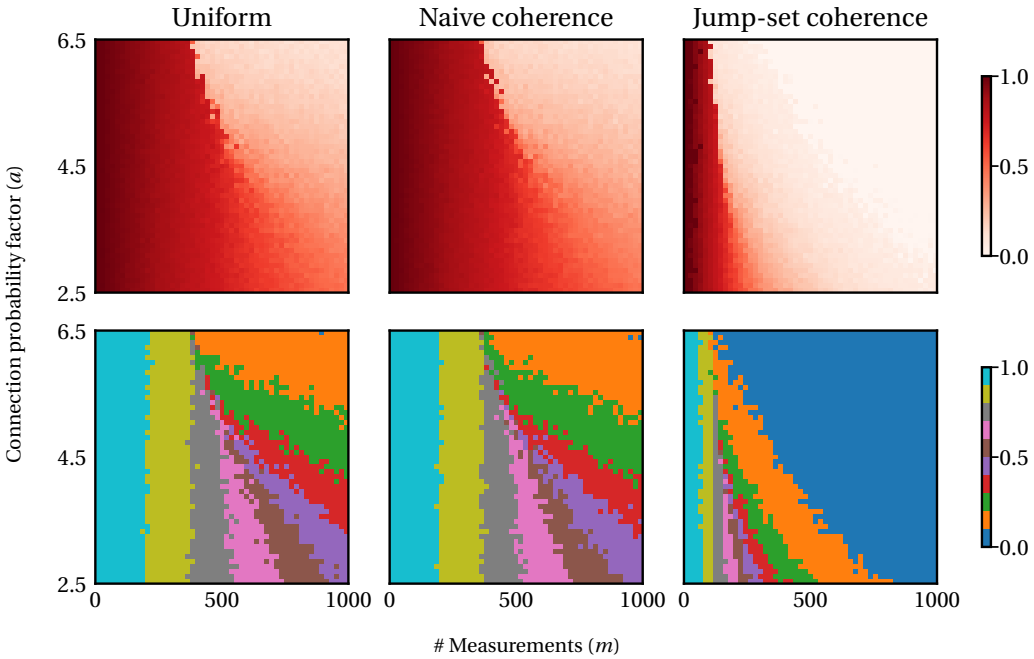
Sampling design	Expression ($\forall i \in [n]$)
Uniform	$\pi_i = 1/n$
Naive coherence	$\pi_i \propto \left\|(\mathbf{D}^+)^\top \mathbf{e}_i\right\|_\infty \cdot \|\mathbf{D}\mathbf{e}_i\|_1$
Jump-set coherence	$\pi_i \propto \left\|(\mathbf{D}^+)^\top \mathbf{e}_i\right\|_\infty \cdot \|\mathbf{P}_{\mathscr{S}}\mathbf{D}\mathbf{e}_i\|_1$

Table 6.2 – Summary of the sampling designs compared in the experiments. The expressions refer to the sampling probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ used in the CSWR($\boldsymbol{\pi}$) model for vertex-sampling of graph signals (see Section 2.2).

Let us begin by revisiting the SBM datasets. Figure 6.5 shows the recovery error of $\mathscr{G}$-TV interpolation under the three sampling designs, contrasting balanced and unbalanced community settings. We have already seen the plots from the first column, but note how similar they are with the ones from the second column. The sample complexity for a correct output in (P1) under *naive coherence sampling* is the same as if the vertices were sampled uniformly at random. The naive control over the ground-truth's jump-set is not enough to change the phase transition profile, a conclusion made all the more convincing once we examine the figure's third column. The change is subtle for the balanced 2-`SSBM`$(500 + 500)$ dataset but very pronounced for the unbalanced case. Recovering the indicator vector unbalanced 2-`SBM`$(200 + 800)$ graphs is the naturally hard setting for the uniform sampling strategy since the sampled vertices have only a one-in-four chance of belonging to the smallest community; in the balanced graphs, every other sample belongs to either community. The *jump-set coherence* design seems to regularize the unbalanced signals, making the phase transition of their recovery look like the one for the symmetric SBM under *uniform random sampling*. The underlying cause of this regularization is found upon examining expression (6.12): only the vertices connected to edges on the ground-truth's jump-set are sampled with probability larger than zero.



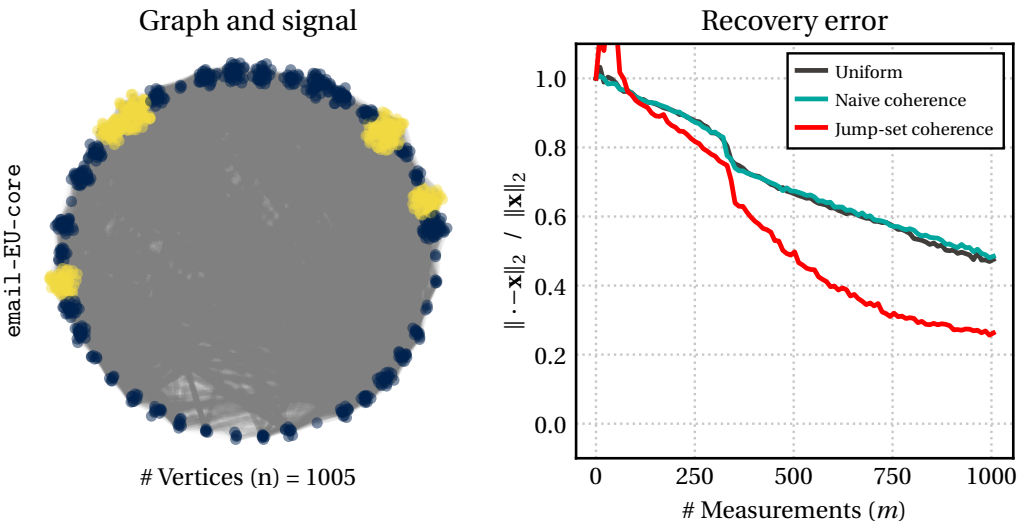
(a) 2-SSBM(500 + 500)



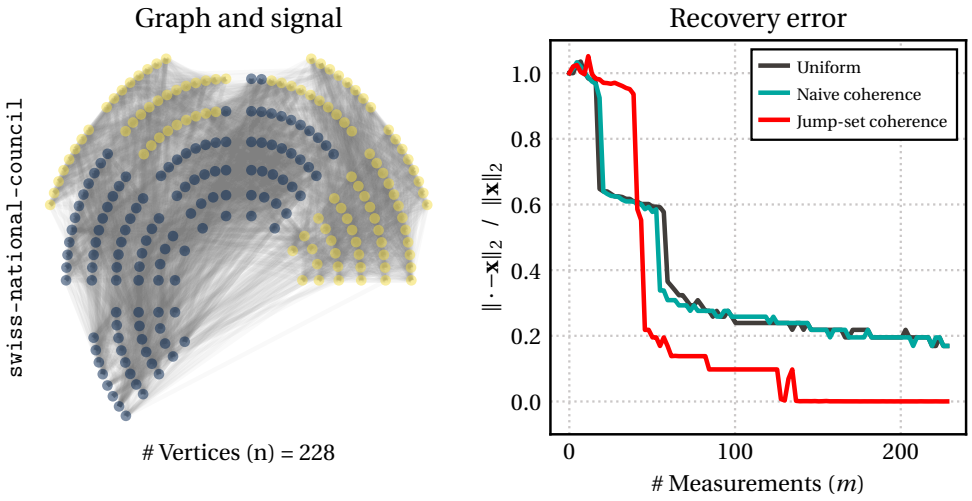
(b) 2-SBM(200 + 800)

Figure 6.5 – Impact of three sampling designs on the recovery error of $\mathscr{G}$-TV interpolation for the community indicator vectors of SBM graphs. Each pixel represents the median error over 25 independent trials. Each plot is paired with its quantized version highlighting some error level-sets.

We can draw similar conclusions for the effects of the three sampling designs using the "real-world" datasets `email-EU-core` and `swiss-national-council`, but we find also some surprises. Figure 6.6 shows that once again the *naive coherence* design behaves just as poorly as *uniform random sampling*; the red curves are more intriguing.



(a) Ground-truth: indicator vector of the 5 largest departments



(b) Ground-truth: indicator vector of the 2 largest right-wing parties

Figure 6.6 – Impact of three sampling designs on the recovery error of $\mathscr{G}$-TV interpolation for the signals of the `email-EU-core` and `swiss-national-council` datasets. Each point in the plot represents the median error over 51 independent trials.

On the plots of Figure 6.6, the *jump-set coherence design* reaches smaller recovery errors than the ones of the two others, but when the measurements are very few this design's error becomes the largest. I think this has to do with how well the ground-truth's jump-set reflects the natural partitions of the graph. When the $\mathscr{G}$-TV decoder has very few measurements, it must rely almost only on the graph structure to do the interpolation. For the SBM datasets this dependence was not detrimental because the signals being recovered were the indicator vectors of the natural communities in the graphs. For the `email-EU-core` and `swiss-national-council` datasets I constructed assembled the ground-truth based only on an intuitive idea of how the respective graphs would cluster. When sampling a few vertices, only at the signal transitions, the *jump-set coherence design* may lack the sample variety of the other designs that influences the $\mathscr{G}$-TV decoder to settle for a less-than-natural partition of the graph. In the end, however, the role of the graph in recovery problems like ours is to *inform* the composition of the signal; very rarely in practice does the graph encode *exactly* a signal of interest. The fact that the *jump-set coherence design* reaches — globally — lower error levels than the other two designs is more important. After all, the error values to the left of the plot are *all* very large. To the right — as the number of samples increases—, the *jump-set coherence design* even results in exact recovery for the `swiss-national-council` data, hinting that the voting patterns of the Swiss National Councillors reflect their political leaning better than the email exchanges in the `email-EU-core` reflect the institution's departmental makeup.

To finish off, let us sweep the `BSDS300` dataset and compare the error curves for the three sampling designs when recovering the segmentation mask graph signals. Figure 6.7 displays these side-by-side, with a gray curve for each image in the dataset.

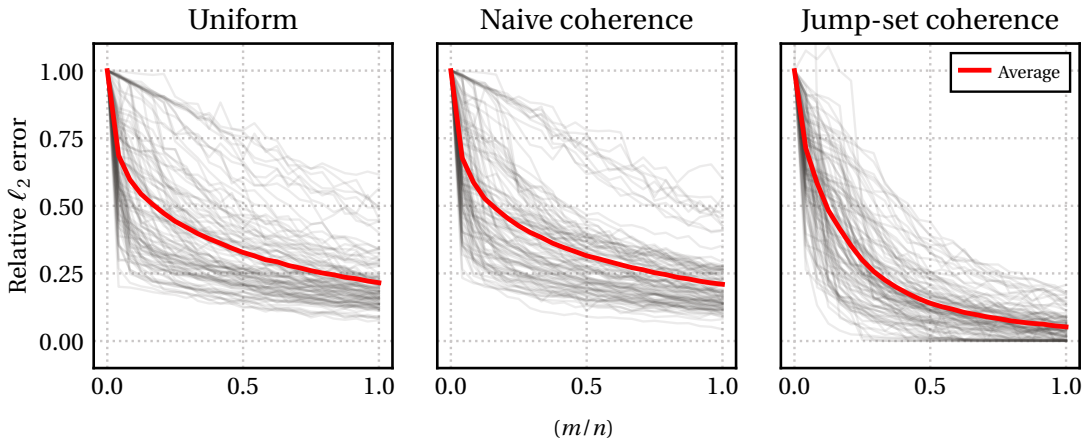


Figure 6.7 – Impact of three sampling designs on the recovery error of $\mathscr{G}$-TV interpolation for the segmentation masks in `BSDS300`. Each grey curve on each of the three plots corresponds to an image in the dataset. Each point in these curve records the median error over 15 independent trials. The red curve traces the point-wise average the grey ones.

Note first that not every image in the dataset admits a sharp phase transition in the recovery

error. For some, the error drops suddenly at a critical number of samples; for others, the error decreases smoothly and remains large even when a lot of pixels are queried. For the first time in our experiments one could argue that there is *some* improvement in using the *naive coherence design* over uniform random sampling, but not enough to justify its use. The *jump-set coherence design* leads to exact recovery for some segmentation masks, but for others, the recovery error remains fairly large despite the use of unfair knowledge about the ground-truth jump-set in the sampling design. Let us examine the images associated with extremes of this behavior to see where this disparity comes from. The left column of Figure 6.8 shows the image whose segmentation mask leads to the smallest recovery error under *jump-set coherence sampling*; the right column shows the image with the largest respective error. The larger error seems to have to do with the presence of several small pieces in the image's segmentation. Even when the sampling design is restricted to query only vertices belonging to the jump-set, it can still miss samples from some of the several small segments.

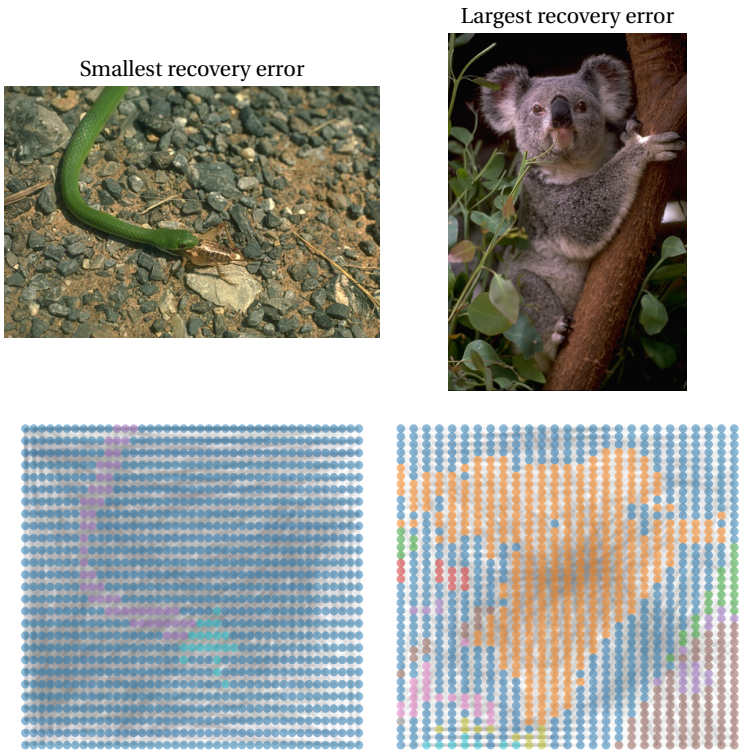


Figure 6.8 – Segmentation mask signals in the BSDS300 dataset that yield the smallest and largest recovery errors when taking $m = n$ samples (with replacement) under the *jump-set coherence design*.

6.5 Summary

The thorough study of a recovery program should not ignore the issues involved in its practical use. Fortunately, the $\mathcal{G}$-TV decoders in (P1) and (P1-η) can be implemented numerically in efficient ways. The one I presented is based on a single primal-dual proximal splitting procedure whose iterations required, in the end, operations that can be made cheaper the sparser the graph is. Upon convergence, the estimated solution can be close or far to the true underlying signal, depending on the number of vertex-samples taken. I have shown that when the ground-truth is the indicator vector of a graph cluster, the recovery error of $\mathcal{G}$-TV interpolation undergoes a sharp phase transition with respect to the number of samples. The threshold happens earlier the denser are the clusters. In contrast, the error under Dirichlet form interpolation when recovering these signals decreases smoothly and slowly, depending more on the actual number of measurements than on the edge structure of the graph.

Putting into practice the optimal sampling design for $\mathcal{G}$-TV interpolation also has issues of its own. To simplify the complicated objective of the optimal design, I proposed for it two progressively looser upper bounds, whose optimizers gave rise to two sampling designs. Both are a form of coherence sampling, familiar in Compressed Sensing, but one of them also uses information from the jump-set of the ground-truth signal. I compared these two designs with the baseline uniform random sampling, concluding that some information about the ground-truth's jump-set must be used to change the phase transition profile of the recovery error in $\mathcal{G}$-TV interpolation; *naive coherence sampling* behaves just as poorly as *uniform random sampling*. The practical issue that remains is how to account for the jump-set in the sampling designs without actually resorting to the — *unknowable* — ground-truth signal?

Appendix 6.A More on primal-dual proximal splitting

In this appendix I sketch, in a little more detail than in the main text, the proximal splitting approach for numerically solving programs like the $\mathscr{G}$-TV decoders (P1) and (P1-η). We may consider here more general optimization problems of the type

$$\min_{\mathbf{z}\in\mathbb{R}^n} \; f(\mathbf{z}) + g(\mathbf{D}\mathbf{z}) + h(\mathbf{z}), \tag{P_{std}}$$

where $f:\mathbb{R}^n\to\mathbb{R}$, $g:\mathbb{R}^N\to\mathbb{R}$ and $h:\mathbb{R}^n\to\mathbb{R}$ are convex functions, the latter of which is also differentiable (with a Lipschitz-continuous gradient).

The proximity operator of a convex function $f:\mathbb{R}^n\to\mathbb{R}$ is the map

$$\mathbf{z}\mapsto \operatorname{prox}_f(\mathbf{z}) = \arg\min_{\mathbf{v}\in\mathbb{R}^n} \; \frac{1}{2}\|\mathbf{z}-\mathbf{v}\|_2^2 + f(\mathbf{v}). \tag{6.14}$$

The problem on the right-hand side admits a unique solution for every $\mathbf{z}\in\mathbb{R}^n$. One can interpret proximity operators as generalizing projections. Indeed, take f to be the convex indicator function of a set $\mathscr{C}$, *i.e.,*, the functions defined by

$$\mathbf{z}\mapsto \iota_{\mathscr{C}}(\mathbf{z}) = \begin{cases} 0, & \text{if } \mathbf{z}\in\mathscr{C}, \\ +\infty & \text{otherwise} \end{cases}.$$

Then, for any $\mathbf{z}\in\mathbb{R}^n$, the vector $\operatorname{prox}_f(\mathbf{z})$ is exactly the orthogonal projection of $\mathbf{z}$ onto $\mathscr{C}$ [17, Table 2, entry i]. Just like orthogonal projections, proximity maps allow us to split the space into complementary halves. Given a convex function f, the so-called Moreau decomposition of any $\mathbf{v}\in\mathbb{R}^n$ is given by

$$\mathbf{v} = \operatorname{prox}_f(\mathbf{v}) + \operatorname{prox}_{f^*}(\mathbf{v}),$$

where f^* is the Fenchel conjugate of f, a function defined as $\mathbf{v}\mapsto \sup_{\mathbf{u}\in\mathbb{R}^n} \langle \mathbf{u},\mathbf{v}\rangle - f(\mathbf{u})$.

But perhaps the two most important properties of prox_f — in what concerns iterative solvers — are its firm non-expansiveness and the fact that its fixed point set matches the set of minimizers of f [17]. The gradient descent map $\mathbf{w}\mapsto\mathbf{w}-\gamma\nabla h(\mathbf{w})$ can also be non-expansive (with a proper choice of step size γ) and, similarly to the proximity map, its fixed point set is equal to the set of minimizers of h. *Proximal splitting* techniques take advantage of these two properties to solve problems of the type $\min_{\mathbf{z}} f(\mathbf{z}) + h(\mathbf{z})$ with alternating calls to the gradient of h (forward step) and proximal operator of f (backward step). In rough terms, the iterations look like the ones below and repeat until $\mathbf{z}$ stops changing noticeably, coming close enough to a minimizer of the sum $f(\cdot)+h(\cdot)$.

$\mathbf{z} \leftarrow \mathbf{z} - \gamma \nabla h(\mathbf{z})$ ▷ Forward step

…

$\mathbf{z} \leftarrow \text{prox}_{\gamma f}(\mathbf{z})$ ▷ Backward step

Adding g to the objective changes things a bit, because the domain of this function is different from that of f and h. *Primal-dual* proximal splitting methods address this situation by working simultaneously with two variables, called the primal $\mathbf{z} \in \mathbb{R}^n$ and the dual $\mathbf{d} \in \mathbb{R}^n$. The direct linear map $\mathbf{z} \mapsto \mathbf{D}\mathbf{z}$ and its transpose, $\mathbf{d} \mapsto \mathbf{D}^\top \mathbf{d}$ connect the primal and dual domains between the forward and backward steps:

$(\mathbf{z}, \mathbf{d}) \leftarrow \left(\mathbf{z} - \gamma \nabla h(\mathbf{z}) - \gamma \mathbf{D}^\top \mathbf{d}, \mathbf{d} + \gamma \mathbf{D}\mathbf{z}\right)$ ▷ Forward step

…

$(\mathbf{z}, \mathbf{d}) \leftarrow \left(\text{prox}_{\gamma f}(\mathbf{z}), \text{prox}_{\gamma g^*}(\mathbf{d})\right)$ ▷ Backward step

The specific primal-dual proximal splitting method used in this chapter comes from Komodakis and Pesquet [39, Algorithm 6]. I reproduce all of its steps in Algorithm 4, but note that at its core the algorithm consists of forward gradient steps, backward proximity steps, and a linear connection between primal and dual spaces via matrix $\mathbf{D}$. The paper of Komodakis and Pesquet lists other numerical solvers for (P_{std}), but the one in Algorithm 4 has more intuitive step size parameters and allows the computation of prox_f and prox_g in parallel.

Algorithm 4 FBF primal-dual iterations for solving (P_{std})

$\mathbf{z}_0 \leftarrow \mathbf{0} \in \mathbb{R}^n$ ▷ Initial primal variable
$\mathbf{d}_0 \leftarrow \mathbf{0} \in \mathbb{R}^N$ ▷ Initial dual variable
repeat
  pick $\gamma_n \in \left(0, \frac{1}{1+\|\mathbf{D}\|_2+\rho}\right)$ ▷ Step size
  $\left(\mathbf{w}_{1,n},\ \mathbf{w}_{2,n}\right) \leftarrow \left(\mathbf{z}_n - \gamma_n \left[\nabla h(\mathbf{z}_n) + \mathbf{D}^\top \mathbf{d}_n\right],\ \mathbf{d}_n + \gamma_n \mathbf{D}\mathbf{z}_n\right)$ ▷ Forward step
  $\left(\mathbf{p}_{1,n},\ \mathbf{p}_{2,n}\right) \leftarrow \left(\text{prox}_{\gamma_n f}\left(\mathbf{w}_{1,n}\right),\ \text{prox}_{\gamma_n g^*}\left(\mathbf{w}_{2,n}\right)\right)$ ▷ Backward step
  $\left(\mathbf{q}_{1,n},\ \mathbf{q}_{2,n}\right) \leftarrow \left(\mathbf{p}_{1,n} - \gamma_n \left[\nabla h(\mathbf{p}_{2,n}) + \mathbf{D}^\top \mathbf{p}_{2,n}\right],\ \mathbf{p}_{2,n} + \gamma_n \mathbf{D}\mathbf{p}_{1,n}\right)$ ▷ Forward step
  $\left(\mathbf{z}_{n+1},\ \mathbf{d}_{n+1}\right) \leftarrow \left(\mathbf{z}_n - \mathbf{w}_{1,n} + \mathbf{q}_{1,n},\ \mathbf{d}_n - \mathbf{w}_{2,n} + \mathbf{q}_{2,n}\right)$ ▷ Update primal/dual variables
until convergence
return $\mathbf{z}_{n+1}$

Algorithms 2 and 3 in the main text are straightforward specializations of Algorithm 4. To see that this is true, first identify g with $\|\cdot\|_1 : \mathbb{R}^N \to \mathbb{R}$. Then the proximal mapping of its Fenchel conjugate is given by

$$\begin{aligned}\mathbf{w} \mapsto &\text{prox}_{\gamma_n g^*}(\mathbf{w}) \\ &= \mathbf{w} - \text{prox}_{\gamma_n g}(\mathbf{w}) \\ &= \mathbf{w} - \text{soft}_{\gamma_n}(\mathbf{w}).\end{aligned}$$

Note that I used the Moreau decomposition, along with the fact that $\operatorname{prox}_{\gamma_n\|\cdot\|_1}$ is the soft thresholding operator $\operatorname{soft}_{\gamma_n}(\mathbf{w})$ [17, Table 2, entry ii]. Next, in the interpolation problem related to Algorithm 2, we have $h \equiv 0$ and so $\nabla h \equiv \mathbf{0}$. The interpolation constraint is expressed via function f with the convex indicator function $\iota_{\{\mathbf{z}\in\mathbb{R}^n:\mathbf{P}_\Omega\mathbf{z}=\mathbf{P}_\Omega\mathbf{x}\}}(\cdot)$. Its proximity map is the same as the orthogonal projection onto set $\{\mathbf{z}\in\mathbb{R}^n : \mathbf{P}_\Omega\mathbf{z} = \mathbf{P}_\Omega\mathbf{x}\}$:

$$\begin{aligned}\mathbf{w} \mapsto &\operatorname{prox}_{\gamma_n f}(\mathbf{w}) \\ &= \arg\min_{\{\mathbf{z}\in\mathbb{R}^n:\mathbf{P}_\Omega\mathbf{z}=\mathbf{P}_\Omega\mathbf{x}\}} \|\mathbf{z}-\mathbf{w}\|_2 \\ &= (\mathbf{I}_n - \mathbf{P}_\Omega)\mathbf{w} + \mathbf{P}_\Omega\mathbf{x}.\end{aligned}$$

For the regression problem, we can express the constraints using a differentiable function, $h(\cdot) = \frac{\rho}{2}\|\mathbf{P}_\Omega(\cdot) - \mathbf{y}\|_2^2$. Its corresponding gradient map is $\mathbf{z} \mapsto \nabla h(\mathbf{z}) = \rho\mathbf{P}_\Omega^\top\mathbf{P}_\Omega\mathbf{z} - \mathbf{y} = \rho\mathbf{P}_\Omega\mathbf{z} - \mathbf{y}$. There is no need for another function in the objective, so we may set $f \equiv 0$, which finishes the specialization of Algorithm 4 into Algorithm 3.

7 Conclusions

> I want to stand as close to the edge as I can without going over. Out on the edge you see all kinds of things you can't see from the center.
>
> — Kurt Vonnegut, Player Piano

To find the missing values, we first connected the dots. Discrete signals are often informed by the similarity of its support points. Groups of friends tend to watch the same movies; politicians that vote alike tend to be in the same party. Those similarities can be concisely represented through a network, and the original signal becomes a graph signal. Still, I have shown at the beginning of this thesis how to express the sampling and analysis of graph signals in the familiar language of linear algebra. The graph gradient matrix, $\mathbf{D}$, arose as the analysis operator for piecewise-constant signals, that is, those that have few edge-variations. Matrix $\mathbf{A}$, representing vertex-sampling, was formed by stacking independent random rows from the standard basis in $\mathbb{R}^n$, according to two alternative sampling models — Ber($\boldsymbol{\pi}$) and CSWR($\boldsymbol{\pi}$).

To figure the whole from its samples, we exploited a signature feature. For piecewise-constant graph signals, a good potential signature is the count of edges across which the signal varies. Among all possible ways to "count" the edge differences, I contrasted the Graph Total Variation ($\mathscr{G}$-TV) and the Dirichlet form. Both theoretically and numerically, I argued that $\mathscr{G}$-TV minimization is less sensitive to the sampling *operator*. In particular, its recovery error drops suddenly close to zero as long as *the number* of measurements is large enough.

To reveal the best sampling design, we minimized the sample complexity. I explored two approaches to study the recovery conditions under $\mathscr{G}$-TV minimization. The "direct" path of Chapter 4 is the more general — dealing automatically with noisy samples — but I was forced to leave it without a punchline. What blocked it was the lack of knowledge on the coordinate

structure of the decent cone $\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})$. The "dual" path of Chapter 5 was more fruitful, revealing the number of measurements required for exact $\mathscr{G}$-TV interpolation as a function of the sampling probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$. What followed was an explicit expression for the optimal vector $\boldsymbol{\pi}$.

To realize the practical issues, we finally took a numerical tour. Implementing the decoders was fairly easy; implementing the optimal sampling design is less so. Nevertheless, we saw how an approximation of the optimal design can visibly change the phase transition profile of the recovery error in $\mathscr{G}$-TV interpolation. But this approximate design has to depend on the ground-truth signal's jump-set, knowledge of which is unobtainable a priori. Indeed — for all the different datasets of Chapter 6 —, the *naive coherence design* behaved just as well as *uniform random sampling*, while the *jump-set coherence* alternative moved the sample complexity threshold towards lower measurement levels.

I leave the readers with two lists that summarize what we have learned in this thesis and what we still have to learn about sampling and subsequent $\mathscr{G}$-TV recovery of piecewise-constant graph signals.

7.1 Takeaways

- **Piecewise-constant signals and the geometry of $\mathbb{B}_{\mathscr{G}-TV}$.** The extreme "points" of the Graph Total Variation ($\mathscr{G}$-TV) ball are essentially indicator vectors of vertex subsets. These indicators are basic elements for a sparse description of piecewise-constant graph functions. On the one hand, the solutions of $\mathscr{G}$-TV minimization will thus tend to be sparse combinations of vertex-subset indicator vectors. On the other hand, the extreme points of the solution set can be *represented* by a selection of columns from the pseudo-inverse $\mathbf{D}^+$. Piecewise-constant graph signals that are "compatible" with the graph structure require relatively few atoms from $\mathbf{D}^+$ in its description. Compatible signals can be potentially recovered by $\mathscr{G}$-TV minimization using only a few vertex samples.

- **Optimal sampling design for $\mathscr{G}$-TV interpolation.** The sampling probabilities that minimize the recovery threshold for $\mathscr{G}$-TV interpolation depend on vertex-induced perturbations of the projection operator $\mathbf{D}\mathbf{D}^+$, restricted to the jump-set $\mathscr{S} := \operatorname{supp}(\mathbf{D}\mathbf{x})$ of the signal $\mathbf{x}$ to be recovered.

- **Importance of the jump-set in the sample complexity threshold.** Attempting to write the sample complexity of $\mathscr{G}$-TV decoders as simply proportional to the jump-set's cardinality might incur in vacuous bounds. This is true, first, because $|\mathscr{S}|$ in piecewise-constant graph signals can be much larger than the signal's dimension and, yet, these signals can still be successfully recovered from relatively few measurements. Second, a sampling design based on this *naive* control over the jump-set induces the same recovery error profile as if the vertices were sampled uniformly at random.

7.2 Open problems

- **A direct certificate for $\mathscr{G}$-TV recovery.** Knowing more about the coordinate structure of the descent cone $\mathscr{D}(\|\mathbf{D}\cdot\|_1, \mathbf{x})$ — whenever $\mathbf{x}$ is a piecewise-constant graph signal — can lead to a high-probability lower bound on the minimum q-gain functional. Such a result would be more powerful than the certificate from Chapter 5, because it would imply robustness guarantees for the noisy regression problem (P1-η). In fact, it would even be useful in more general settings where the measurements undergo some non-linear transformation before becoming available [57].

- **Practical sampling designs.** The optimization program in Corollary 5.4.1.1 is explicit but not practical a priori. I used a simplified version of it in Chapter 6 but even then the sampling design that performed well was the one that required knowing the signal-to-be-sampled before even sampling it. We are thus left with a modeling problem. The optimal design depends on the action of the projection operator $\mathbf{P}_{\mathscr{S}}$, but knowing this operator is tantamount to knowing the jump-set $\mathscr{S}$ of the signal we want to recover. How much is it possible to assume, in practice, about the actions of $\mathbf{P}_{\mathscr{S}}$ without explicitly knowing $\mathscr{S}$?

A Bernstein inequalities

Many of the random objects encounter in this thesis can be written as a sum of independent random variables. The terms of the sum have bounded moments, so the sum itself takes on values close to its expectation with high probability. This concentration phenomenon can be quantified in terms of the Bernstein-type inequalities that I present here.

Lemma A.0.1 (Scalar Bernstein inequality [6, p.117])**.** . *Let $X_1, \dots, X_n$ be independent random variables, taking values in $\mathbb{R}$, and such that, for each $i \in [n]$,*

- $\mathbb{E}(X_i) = 0$,
- $\sigma^2 := \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(X_i^2\right) < \infty$,
- $|X_i| \leq B < \infty$.

Then, $\forall\, t > 0$, the tails of the sum $S_n := X_1 + \cdots + X_n$ behave as

$$\mathbb{P}\left(\{|S_n| \geq tn\}\right) \leq 2\exp\left(\frac{-3n}{8}\min\left\{\frac{t^2}{\sigma^2}, \frac{t}{B}\right\}\right). \tag{A.1}$$

Lemma A.0.2 (Vector Bernstein inequality [47, p.164])**.** *Let $X_1, \dots, X_n$ be independent random vectors in a Banach space equipped with norm $\|\cdot\|$, such that, for each $i \in [n]$,*

- $\mathbb{E}(X_i) = 0$,
- $\sigma^2 := \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\|X_i\|^2\right) < \infty$,
- $\|X_i\| \leq B < \infty$.

Then, $\forall\, t > 0$, the tails of the sum $S_n := X_1 + \cdots + X_n$ behave as

$$\mathbb{P}\left(\{\|S_n\| \geq tn\}\right) \leq 2\exp\left(\frac{-3n}{8}\min\left\{\frac{t^2}{\sigma^2}, \frac{t}{B}\right\}\right). \tag{A.2}$$

Lemma A.0.3 (Matrix Bernstein inequality [73, Thm. 1.6.2])**.** *Let $X_1, \dots, X_n \in \mathbb{C}^{d_1 \times d_2}$ be independent random matrices, and denote by $\|\cdot\|$ the spectral norm (maximal singular value). Assume that, for each $i \in [n]$,*

- $\mathbb{E}(X_i) = 0$,
- $\sigma^2 := \max\left\{\left\|\sum_{i=1}^n \mathbb{E}\left(X_i X_i^*\right)\right\|, \left\|\sum_{i=1}^n \mathbb{E}\left(X_i^* X_i\right)\right\|\right\} < \infty$,
- $\|X_i\| \leq B < \infty$.

Then, $\forall\, t > 0$, the tails of the sum $S_n := X_1 + \cdots + X_n$ behave as

$$\mathbb{P}\left(\{\|S_n\| \geq tn\}\right) \leq (d_1 + d_2) \exp\left(\frac{-3n}{8} \min\left\{\frac{t^2}{\sigma^2}, \frac{t}{B}\right\}\right). \tag{A.3}$$

Remark: The main difference between the matrix Bernstein inequality and the other two instances is the presence of a dimensional factor $(d_1 + d_2)$ which cannot be removed in the general case. As a consequence, there is a limited range of t for which the matrix Bernstein inequality is informative [73, p. 77]. Additionally, Tropp highlights that the definition of the variance term σ^2 through a maximum of two terms reflects the existence of two different squares for a general matrix $\mathbf{M}$, namely $\mathbf{M}^\top \mathbf{M}$ and $\mathbf{M}\mathbf{M}^\top$. A scalar $s \in \mathbb{R}$ has only one square, so only one second moment to consider.

Bibliography

[1] ABBE, E. Community Detection and Stochastic Block Models. *Foundations and Trends® in Communications and Information Theory 14*, 1-2 (2018), 1–162. 00108.

[2] ABBE, E., BANDEIRA, A. S., AND HALL, G. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory 62*, 1 (2015), 471–487.

[3] ADCOCK, B., HANSEN, A. C., POON, C., AND ROMAN, B. Breaking the coherence barrier: A new theory for compressed sensing. *Forum of Mathematics, Sigma 5* (2017). 00131.

[4] AMELUNXEN, D., LOTZ, M., MCCOY, M. B., AND TROPP, J. A. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference. A Journal of the IMA 3* (2014), 224–294.

[5] ARORA, S., COHEN, N., GOLOWICH, N., AND HU, W. A Convergence Analysis of Gradient Descent for Deep Linear Neural Networks. In *International Conference on Learning Representations* (2019). 00023.

[6] ARTSTEIN-AVIDAN, S., GIANNOPOULOS, A., AND MILMAN, V. *Asymptotic Geometric Analysis, Part I*, vol. 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, Rhode Island, June 2015. 00088.

[7] BORA, A., JALAL, A., PRICE, E., AND DIMAKIS, A. G. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (2017), JMLR. org, pp. 537–546. 00127.

[8] BOUCHERON, S., LUGOSI, G., AND MASSART, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence*, 1st ed ed. Oxford University Press, Oxford, 2013. 01181 OCLC: ocn818449985.

[9] BOYD, S., AND VANDENBERGHE, L. *Convex Optimization*. web.stanford.edu, 2009.

[10] BOYER, C., BIGOT, J., AND WEISS, P. Compressed sensing with structured sparsity and structured acquisition. *Applied and Computational Harmonic Analysis 46*, 2 (Mar. 2019), 312–350.

[11] CANDÈS, E. J., ELDAR, Y. C., NEEDELL, D., AND RANDALL, P. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis 31* (2011), 59–73.

[12] CANDÈS, E. J., LI, X., MA, Y., AND WRIGHT, J. Robust Principal Component Analysis? *J. ACM 58*, 3 (June 2011), 11:1–11:37. 04220.

[13] CANDES, E. J., AND PLAN, Y. A Probabilistic and RIPless Theory of Compressed Sensing. *IEEE Transactions on Information Theory 57*, 11 (Nov. 2011), 7235–7254. 00464.

[14] CHANDRASEKARAN, V., RECHT, B., PARRILO, P. A., AND WILLSKY, A. S. The Convex Geometry of Linear Inverse Problems. *Foundations of Computational Mathematics 12* (Oct. 2012), 805–849.

[15] CHAPELLE, O., SCHÖLKOPF, B., AND ZIEN, A. *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 2006.

[16] CHEN, Y., AND CHI, Y. Robust spectral compressed sensing via structured matrix completion. *IEEE Transactions on Information Theory 60*, 10 (2014), 6576–6601. 00142.

[17] COMBETTES, P. L., AND PESQUET, J.-C. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011, pp. 185–212. 01895.

[18] DE LA PEÑA, V. H., AND GINÉ, E. *Decoupling*. Probability and Its Applications. Springer New York, New York, NY, 1999. 00094.

[19] DIRKSEN, S., LECUÉ, G., AND RAUHUT, H. On the Gap Between Restricted Isometry Properties and Sparse Recovery Conditions. *IEEE Transactions on Information Theory 64*, 8 (Aug. 2018), 5478–5487. 00026.

[20] DIRKSEN, S., AND MENDELSON, S. Robust one-bit compressed sensing with non-Gaussian measurements. *arXiv:1805.09409 [cs, math]* (May 2018). 00000.

[21] DONOHO, D., AND TANNER, J. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 367*, 1906 (Nov. 2009), 4273–4293. 00313.

[22] FOUCART, S., AND RAUHUT, H. *A Mathematical Introduction to Compressive Sensing*. Springer Science & Business Media, New York, NY, Aug. 2013.

[23] FOYGEL, R., AND MACKEY, L. Corrupted Sensing: Novel Guarantees for Separating Structured Signals. *IEEE Transactions on Information Theory 60* (2014), 1223–1247. 00067.

[24] GENZEL, M., KUTYNIOK, G., AND MÄRZ, M. L^1-Analysis Minimization and Generalized (Co-)Sparsity: When Does Recovery Succeed? *arxiv.org 1710* (Oct. 2017), arXiv:1710.04952.

[25] GIRYES, R., PLAN, Y., AND VERSHYNIN, R. On the Effective Measure of Dimension in the Analysis Cosparse Model. *IEEE Transactions on Information Theory 61* (2015), 5745–5753.

[26] GIRYES, R., PLAN, Y., AND VERSHYNIN, R. On the effective measure of dimension in total variation minimization. In *2015 International Conference on Sampling Theory and Applications (SampTA)* (May 2015), pp. 593–597. 00001.

[27] GORDON, Y. On Milman's inequality and random subspaces which escape through a mesh in ℝn. vol. 1317. Springer, Berlin, Heidelberg, Berlin, Heidelberg, 1988, pp. 84–106.

[28] GROSS, D. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory 57*, 3 (Mar. 2011), 1548–1566. 00689.

[29] GROSS, D., AND NESME, V. Note on sampling without replacing from a finite collection of matrices. *arXiv:1001.2738 [quant-ph]* (Jan. 2010). 00056.

[30] HOEFFDING, W. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association 58*, 301 (1963), 13–30. 06998.

[31] JOHNSON, W. B., AND LINDENSTRAUSS, J. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics 26* (1984), 189–206. 00000.

[32] JUNG, A. The Network Nullspace Property for Compressed Sensing of Big Data over Networks. *CoRR abs/1302.3921 stat.ML* (2017), arXiv:1705.04379.

[33] JUNG, A., HERO, A. O. I., MARA, A., AND JAHROMI, S. Semi-Supervised Learning via Sparse Label Propagation. *arxiv.org 1612* (Dec. 2016), arXiv:1612.01414.

[34] JUNG, A., TRAN, N., AND MARA, A. When is network lasso accurate? *Frontiers in Applied Mathematics and Statistics 3* (2018), 28. 00018.

[35] KABANAVA, M., AND RAUHUT, H. Analysis l^1-recovery with Frames and Gaussian Measurements. *Acta Applicandae Mathematicae 140* (2014), 173–195.

[36] KABANAVA, M., AND RAUHUT, H. Cosparsity in Compressed Sensing. No. Chapter 11. Springer International Publishing, Cham, 2015, pp. 315–339.

[37] KABANAVA, M., RAUHUT, H., AND ZHANG, H. Robust analysis l^1-recovery from Gaussian measurements and total variation minimization. *European Journal of Applied Mathematics 26*, 6 (2015), 917–929. 00021.

[38] KOLTCHINSKII, V., AND MENDELSON, S. Bounding the smallest singular value of a random matrix without concentration. *International Mathematics Research Notices 2015*, 23 (2015), 12991–13008. 00096.

[39] KOMODAKIS, N., AND PESQUET, J.-C. Playing with Duality. *IEEE Signal Processing Magazine* (Oct. 2015), 31–54.

[40] KRAHMER, F., KRUSCHEL, C., AND SANDBICHLER, M. Total Variation Minimization in Compressed Sensing. vol. 28. Birkhäuser, Cham, Cham, 2017, pp. 333–358.

[41] KRAHMER, F., AND STÖGER, D. On the convex geometry of blind deconvolution and matrix completion. *arXiv:1902.11156 [cs, math]* (Feb. 2019). 00000.

[42] KUENG, R., AND GROSS, D. RIPless compressed sensing from anisotropic measurements. *Special Issue on Sparse Approximate Solution of Linear Systems 441* (Jan. 2014), 110–123. 00032.

[43] LAMPORT, L. How to write a 21st century proof. *Journal of Fixed Point Theory and Applications 11* (2012), 43–63.

[44] LECUÉ, G., AND MENDELSON, S. Regularization and the small-ball method II: Complexity dependent error rates. *The Journal of Machine Learning Research 18*, 1 (2017), 5356–5403. 00020.

[45] LECUÉ, G., AND MENDELSON, S. Sparse recovery under weak moment assumptions. *Journal of the European Mathematical Society 19*, 3 (2017), 881–904. 00052.

[46] LECUÉ, G., AND MENDELSON, S. Regularization and the small-ball method I: Sparse recovery. *The Annals of Statistics 46*, 2 (2018), 611–641. 00048.

[47] LEDOUX, M., AND TALAGRAND, M. *Probability in Banach Spaces: Isoperimetry and Processes*. Classics in Mathematics. Springer, Berlin ; London, 2011. 03091 OCLC: ocn751525992.

[48] LEE, K., LI, Y., JIN, K. H., AND YE, J. C. Unified Theory for Recovery of Sparse Signals in a General Transform Domain. *IEEE Transactions on Information Theory 64*, 8 (Aug. 2018), 5457–5477. 00000.

[49] LIAW, C., MEHRABIAN, A., PLAN, Y., AND VERSHYNIN, R. A simple tool for bounding the deviation of random matrices on geometric sets. vol. 2169. Springer, Cham, 2017, pp. 277–299.

[50] MENDELSON, S. Learning without Concentration. *Journal of the ACM 62* (June 2015), 1–25.

[51] MENDELSON, S. Learning without concentration for general loss functions. *Probability Theory and Related Fields 33* (June 2017), 1497–44.

[52] MENDELSON, S. Approximating the covariance ellipsoid. *arXiv:1804.05402 [cs, stat]* (Apr. 2018). 00000.

[53] MENDELSON, S., RAUHUT, H., AND WARD, R. Improved bounds for sparse recovery from subsampled random convolutions. *The Annals of Applied Probability 28*, 6 (2018), 3491–3527. 00012.

[54] NAM, S., DAVIES, M. E., ELAD, M., AND GRIBONVAL, R. The cosparse analysis model and algorithms. *Applied and Computational ... 34* (2013), 30–56.

[55] ORTEGA, A., FROSSARD, P., KOVAČEVIĆ, J., MOURA, J. M., AND VANDERGHEYNST, P. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE 106*, 5 (2018), 808–828. 00144.

[56] OYMAK, S., AND TROPP, J. A. Universality laws for randomized dimension reduction, with applications. *Information and Inference: A Journal of the IMA 7*, 3 (Sept. 2018), 337–446. 00033.

[57] PLAN, Y., AND VERSHYNIN, R. The generalized lasso with non-linear observations. *IEEE Transactions on Information Theory 62* (Mar. 2016), 1528–1537.

[58] PLAN, Y., VERSHYNIN, R., AND YUDOVINA, E. High-dimensional estimation with geometric constraints. *Information and Inference. A Journal of the IMA 6* (2017), 1–40.

[59] POON, C. On the Role of Total Variation in Compressed Sensing. *SIAM Journal on Imaging Sciences 8*, 1 (Jan. 2015), 682–720. 00019.

[60] PUY, G., TREMBLAY, N., GRIBONVAL, R., AND VANDERGHEYNST, P. Random sampling of bandlimited signals on graphs. *Applied and Computational Harmonic Analysis* (2016). 00052.

[61] RAUHUT, H., SCHNASS, K., AND VANDERGHEYNST, P. Compressed Sensing and Redundant Dictionaries. 1–19.

[62] ROCKAFELLAR, R. T. *Convex Analysis*. Princeton University Press, 1970. 25143.

[63] ROCKAFELLAR, R. T., AND WETS, R. J. B. *Variational Analysis*. Springer Science & Business Media, June 2009.

[64] SANDRYHAILA, A., AND MOURA, J. M. Discrete signal processing on graphs. *IEEE transactions on signal processing 61*, 7 (2013), 1644–1656. 00693.

[65] SCHNEIDER, R., AND WEIL, W. *Stochastic and Integral Geometry*. Springer Science and Business Media, Berlin, Heidelberg, Sept. 2008.

[66] SHUMAN, D. I., NARANG, S. K., FROSSARD, P., ORTEGA, A., AND VANDERGHEYNST, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine 30* (Apr. 2013), 83–98.

[67] SIVAKUMAR, V., BANERJEE, A., AND RAVIKUMAR, P. K. Beyond sub-gaussian measurements: High-dimensional structured estimation with sub-exponential designs. In *Advances in Neural Information Processing Systems* (2015), pp. 2206–2214. 00020.

[68] STOKES, S. C. Political Parties and Democracy. *Annual Review of Political Science 2*, 1 (June 1999), 243–267. 00234.

[69] SZLAM, A., AND BRESSON, X. Total Variation and Cheeger Cuts. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (Haifa, Israel, June 2010), J. Fürnkranz and T. Joachims, Eds., Omnipress, pp. 1039–1046. 00000.

[70] TALAGRAND, M. A new look at independence. *The Annals of Probability 24* (Jan. 1996), 1–34.

[71] TALAGRAND, M. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. No. 3. Folge, volume 60 in Ergebnisse Der Mathematik Und Ihrer Grenzgebiete : A Series of Modern Surveys in Mathematics. Springer, Heidelberg ; New York, 2014. 00139 OCLC: ocn873529598.

[72] TROPP, J. A. Convex recovery of a structured signal from independent random linear measurements. In *Sampling Theory, a Renaissance*. Springer, 2015, pp. 67–101. 00112.

[73] TROPP, J. A. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning 8*, 1-2 (2015), 1–230. 00360.

[74] UNSER, M., FAGEOT, J., AND GUPTA, H. Representer Theorems for Sparsity-Promoting l1 Regularization. *IEEE Transactions on Information Theory 62* (2016), 5167–5180.

[75] UNSER, M., FAGEOT, J., AND WARD, J. P. Splines are Universal Solutions of Linear Inverse Problems with Generalized-TV regularization. *arxiv.org math.FA* (Mar. 2016), arXiv:1603.01427.

[76] VAN DER AALST, W. Data Science in Action. In *Process Mining*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016, pp. 3–23. 00797.

[77] VAN HANDEL, R. Probability in High Dimension:. Tech. rep., Defense Technical Information Center, Fort Belvoir, VA, June 2014. 00035.

[78] VERSHYNIN, R. Estimation in high dimensions: A geometric perspective. In *Sampling Theory, a Renaissance*. Springer, 2015, pp. 3–66. 00105.

[79] VERSHYNIN, R. Lectures in Geometric Functional Analysis. Apr. 2017.

[80] YIN, H., BENSON, A. R., LESKOVEC, J., AND GLEICH, D. F. Local Higher-Order Graph Clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2017), KDD '17, ACM, pp. 555–564. 00105.

Rodrigo Cerqueira Gonzalez Pena

Avenue d'Échallens 112, 1004 Lausanne, Switzerland
Phone: +41 76 542 02 55
Email: rodrigocgp@gmail.com
Website: rodrigo-pena.github.io
GitHub: rodrigo-pena
LinkedIn: rodrigocgpena

Education

École Polytechnique Fédérale de Lausanne (EPFL) *Ph.D. Computer Science* 2019
Lausanne, Switzerland
Working at the Signal Processing Laboratory 2 (LTS2), on machine learning and compressed sensing techniques for signals living on graphs.
- Thesis supervisor: Pierre Vandergheynst

Universidade de Brasília *B.Sc. Electrical Engineering* 2014
Brasília, Brazil
I specialized in Signal Processing, completing a thesis project on saliency-based quality metrics for images and videos.
- Thesis supervisor: Mylène C.Q. Farias

École Nationale Supérieure d'Électronique, Informatique, Télécommunications, Mathématique et Mécanique de Bordeaux (ENSEIRB-MATMECA) *Exchange Student* 2012–2013
Bordeaux, France
Funded by a scholarship awarded by the Brazilian government. At ENSEIRB-MATMECA, I completed the second year of the French Engineering Schools system.

Research Experience

École Polytechnique Fédérale de Lausanne (EPFL) *Ph.D. thesis* 2015–2019
Lausanne, Switzerland
I focused on how to provide recovery guarantees for inverse problems supported on graphs. In particular, I studied the construction of random vertex-sampling designs that allow recovering the underlying network signal by minimizing its graph total-variation. Possible applications include learning class labels of elements of a network.

Fraunhofer-Institut für Digitale Medientechnologie *Visiting Researcher* Mar-Apr 2018
Ilmenau, Germany
Collaborating with specialists in industry applications of machine learning for audio signals. I designed a time-dependent harmonic (in the sense of music theory) similarity measure for song excerpts, building upon work from a past Master's student at the institute. The visit was promoted and funded by the EU's Marie Curie Initial Training Network "SpaRTaN".

University of Surrey *Visiting Researcher* Jan–Mar 2017
Guildford, United Kingdom
Visiting the Centre for Vision, Speech and Signal Processing (CVSSP). The goal was to learn from their expertise in machine learning for audio signals, while trading some of my knowlegde of machine learning on graphs. This visit was also promoted and funded by the EU's Marie Curie Initial Training Network "SpaRTaN".

Laboratoire de l'Intégration du Matériau au Système (IMS) *Intern* Jun 2013
Bordeaux, France
I implemented a robust fuzzy k-means algorithm targeted at clustering image texture patches.

Teaching Experience

École Polytechnique Fédérale de Lausanne (EPFL) *Teaching Assistant* 2015–2018
Lausanne, Switzerland

- *EE-558 A Network Tour of Data Science*

  Fall 2018. Master's-level, project-oriented course. I helped curate, guide, and evaluate the projects of 46 teams of 4 students.
- *MICRO-310 Signals and Systems I*

  Fall 2015–2017. Bachelor's level course. I helped design and organize exercise sessions for 150+ students each semester. Eventually, I prepared and presented mini-lectures at these sessions. I also partially prepared and corrected their final exams.
- *MICRO-311 Signals and Systems II*

  Spring 2017–2018. Bachelor's level course. Similar class size and responsibilities as MICRO-310. The difference was only in content: MICRO-311 focuses on the analysis of discrete-time, shift-invariant systems, whereas MICRO-310 dwells on the continuous-time version.

École Polytechnique Fédérale de Lausanne (EPFL) *Project Supervisor* 2016–2019
Lausanne, Switzerland

- *"Audiovisual Source Separation Using Neural Networks"*

  Master thesis, B. Inan, 2019 (co-supervised with B. Ricaud, H. Peic Tukuljac and researchers from Logitech).
- *"Audio Blind Source Separation for Noise Reduction"*

  Master thesis, V. Pollet, 2019 (co-supervised with B. Ricaud, H. Peic Tukuljac and researchers from Logitech).
- *"Graph Representation of Music Database"*

  Master thesis, H. Parmantier and A. Basille, 2016 (co-supervised with K. Benzi).

Publications

Conference Proceedings

[1] **Pena, Rodrigo**, Bresson, Xavier, and Vandergheynst, Pierre. **2016**. "Source localization on graphs via ℓ_1 recovery and spectral graph theory". In: *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. Bordeaux, France: IEEE. ISBN: 978-1-5090-1929-8. DOI: 10.1109/IVMSPW.2016.7528230.

Conferences/Workshops

Invited

Signal Processing with Adaptive Sparse Structured Representations (SPARS) *Poster Presenter* Jul 2019
Toulouse, France

Applied Machine Learning Days (AMLD) *Speaker / Facilitator* Jan 2019
Lausanne, Switzerland
I co-organized, with Michaël Defferard, a full-day session on learning and processing over networks. We designed interactive Python notebooks for the 50+ attendees, and supplemented the section with lecture slides for explaining theoretical concepts.

graphSIP: Traitement du signal sur graphes - Applications aux nuages de points 3D et en neuroscience *Speaker* Sep 2018
Aussois, France
I presented two, one-hour lectures for researchers interested in applying graph signal processing to their problems. I talked about spectral clustering, Laplacian eigenmaps, and graph learning.

IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP) *Poster Presenter* Jul 2016
Bordeaux, France

Attended

13th Iternational Conference on Sampling Theory and Applications (SampTA) Jul 2019
Bordeaux, France

Conference on Learning Theory (COLT) Jul 2018
Stockholm, Sweden

Concentration of Measure and its Applications (Cargèse International School) May 2018
Cargèse, France

Signal Processing with Adaptive Sparse Structured Representations (SPARS) Jun 2017
Lisbon, Portugal

Awards and Honors

Best student paper award Jul 2016
IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)

Outstanding student in graduating class Sep 2014
Awarded by the professors of the Electrical Engineering department at Universidade de Brasília.

Grants and Fellowships

Marie Curie Initial Training Network (ITN) *Early Stage Researcher Fellow* Jul 2015
European Union's Seventh Framework Programme (FP7- PEOPLE-2013-ITN) grant agreement 607290 SpaRTaN, covering the Ph.D. salary as well as travel and training costs.

CAPES/Brafitec Jul 2012
Brazilian government scholarship funding exchange opportunities at Frech engineering schools.

Service

Reviewer

- IEEE Transactions on Signal Processing
- IEEE GlobalSIP

Skills

Programming

Scientific Computing	Python (numpy, matplotlib, scikit-learn, pandas), Matlab
Version Control	git
Writing	LaTeX, markdown

I am a contributing developer to two Python packages managed by the Signal Processing Laboratory 2 (LTS2) at EPFL:

- `pygsp`, containing useful tools for *graph signal processing*, and
- `pyunlocbox`, implementing optimization solvers based on *proximal-splitting algorithms*.

Languages

English	Fully proficient
French	Fully proficient
Portuguese	Native
Spanish	Intermediate

Communication

As a Marie Curie Initial Training Network fellow, I had access to substantial training on communication skills related to public speaking and science outreach. As a consequence, I have harnessed basic skills on

- design and data visualization,
- effective use of body language,
- on-camera speaking,

Teaching

I have attended two workshops at EPFL's Teaching Support Centre while a Ph.D. student: "Teaching Toolkit" and "Presenting and explaining in class". Together with my teaching assistant experience, these workshops helped me assimilate strategies for

- teaching one-to-one effectively,
- organizing exercise sections,
- structuring a lecture.

Personal Details

Married, 28, Brazilian citizenship.

126