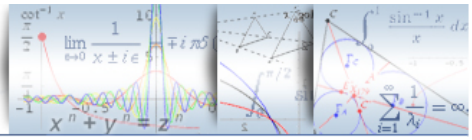
Computation of Sparse Low Degree Interpolating Polynomials and their Application to Derivative-Free Optimization

Afonso José Sousa Bandeira

DEPARTAMENTO DE MATEMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA



Computation of Sparse Low Degree Interpolating Polynomials and their Application to Derivative-Free Optimization

Afonso José Sousa Bandeira

Dissertação para a obtenção do Grau de **Mestre em Matemática**

Área de Especialização em **Análise Aplicada e Matemática Computacional**

Júri

Presidente:	Professor Joaquim João Júdice
Orientador:	Professor Luís Nunes Vicente
Vogal:	Professor Adérito Araújo

Data: Julho de 2010

Resumo

Os métodos de regiões de confiança baseados em interpolação polinomial constituem uma classe de algoritmos importante em Optimização Sem Derivadas. Estes métodos dependem de aproximações locais da função objectivo, feitas por modelos de interpolação polinomial quadrática, construídos, frequentemente, com menos pontos do que elementos das respectivas bases.

Em muitas aplicações práticas, a contribuição das variáveis do problema para a função é tal que a correlação entre vários pares de variáveis se revela nula ou quase nula, o que implica, no caso suave, uma estrutura de esparsidade na matriz Hessiana.

Por outro lado, a teoria sobre recuperação esparsa desenvolvida recentemente na área de *Compressed Sensing* caracteriza condições sob as quais um vector esparso pode ser recuperado, de modo preciso, usando um número reduzido de leituras aleatórias. Tal recuperação é alcançada minimizando a norma ℓ_1 de um vector sujeito às restrições impostas pelas leituras.

Nesta dissertação, sugerimos uma abordagem para a construção de modelos de interpolação quadrática no caso indeterminado, através da minimização da norma ℓ_1 das entradas da matriz Hessiana do modelo. Mostramos que tal procedimento 'recupera' modelos precisos quando a matriz Hessiana da função é esparsa, recorrendo a amostragem aleatória e utilizando consideravelmente menos pontos de amostragem do que elementos nas respectivas bases.

Este resultado serviu de motivação para o desenvolvimento de um método de regiões de confiança baseado em interpolação polinomial, com características mais práticas, nos quais se utilizam modelos quadráticos (determinísticos) de norma ℓ_1 mínima. Os resultados numéricos relatados na tese mostram um desempenho promissor desta abordagem, tanto no caso geral como no esparso.

Palavras Chave: Optimização sem derivadas, métodos de regiões de confiança baseados em interpolação polinomial, amostragem aleatória, recuperação esparsa, compressed sensing, minimização ℓ_1.

Abstract

Interpolation-based trust-region methods are an important class of algorithms for Derivative-Free Optimization which rely on locally approximating an objective function by quadratic polynomial interpolation models, frequently built from less points than bases components.

Often, in practical applications, the contribution of the problem variables to the function is such that several pairwise correlations between variables is zero or negligible, implying, in the smooth case, a sparse structure in the Hessian matrix.

On the other hand, the sparse recovery theory developed recently in the field of Compressed Sensing characterizes conditions under which a sparse vector can be accurately recovered from few random measurements. Such a recovery is achieved by minimizing the ℓ_1-norm of a vector subject to the measurements constraints.

We suggest an approach for building quadratic polynomial interpolation models in the underdetermined case by minimizing the ℓ_1-norm of the entries of the Hessian model. We show that this procedure recovers accurate models when the function Hessian is sparse, using random sampling and considerably less sample points than bases components.

Motivated by this result, we developed a practical interpolation-based trust-region method using deterministic minimum ℓ_1-norm quadratic models and showed that it exhibits a promising numerical performance both in general and in the sparse case.

Keywords: Derivative-free optimization, interpolation-based trust-region methods, random sampling, sparse recovery, compressed sensing, ℓ_1-minimization.

Agradecimentos

I am deeply indebted to my advisor, Professor Luís Nunes Vicente, for the interesting research topic, his scientific guidance and endless availability, and for having taught me how to properly convey and write mathematics.

I also express my gratitude to Prof. Katya Scheinberg, for an instructive collaboration, and to Dr. Rachel Ward, for interesting discussions on the topic of this thesis.

I would also like to thank Prof. Daniel Abreu for having introduced me to mathematical research while working on Applied Harmonic Analysis, during the FCT Program BII, in the academic year of 2008/2009.

Finally, I would like to thank my Parents, for their constant support.

Contents

1 Introduction **1**

2 Use of models in DFO trust-region methods **5**
2.1 Fully linear and fully quadratic models 5
2.2 Quadratic polynomial interpolation models 7
2.2.1 Determined models 8
2.2.2 Underdetermined interpolating models 8

3 Compressed Sensing **11**
3.1 General concepts and properties 11
3.2 RIP and NSP for partial sparse recovery 17

4 Recovery of Sparse Hessians **21**
4.1 Sparse recovery using orthonormal expansions 21
4.2 Sparse recovery using polynomial orthonormal expansions 23
4.2.1 Orthogonal expansions on hypercubes 24
4.2.2 Orthogonal expansions on Euclidian balls 27
4.3 Recovery of functions with sparse Hessian using quadratic models . . 28

5 A practical interpolation-based trust-region method **35**
5.1 Interpolation-based trust-region algorithms for DFO 35
5.2 A practical interpolation-based trust-region method 37
5.3 Numerical results 39

6 Conclusion **45**

Chapter 1

Introduction

Optimization is an important field of Applied Mathematics supported by a comprehensive theory and offering a wide range of applications, recently enriched by the developments in emerging areas such as Machine Learning and Compressed Sensing. A single-objective optimization problem basically consists of minimizing or maximizing a function (usually referred to as the objective function) in a set or region (typically called the feasible region). The mathematical or numerical difficulties of optimization problems are related to the structure of the objective function and the functions defining the feasible region. The least difficult cases occur when these functions are smooth and lead to convexity and when one can access their first and second order derivatives. However, in many real-world applications, the objective function is calculated by some costly black-box simulation which does not provide information about its derivatives. Although one could estimate the derivatives, e.g., by finite differences, such a process often renders too expensive and can produce misleading results in the presence of noise. An alternative is to consider methods that do not require derivative information, and such methods are the subject of study in Derivative-Free Optimization (DFO).

One important class of methods in DFO are interpolation-based trust-region methods. At each iteration, these methods build models of the objective function that locally approximate it in some *trust region* centered at the current iterate point. The model is then minimized in the trust region, and the corresponding minimizer is, hopefully, a better candidate for being a minimizer of the objective function in the trust region, and thus possibly taken as the next iterate. The minimization of the model in the trust region should, however, be relatively easy. The simplest class of models one can think of are the linear ones but they do not capture the curvature of the objective function and thus slow down the convergence of the method. A natural and convenient non-linear class of models are the quadratic ones, being often efficiently used.

In order to compute models of the objective function, one usually uses interpolation techniques. In a general scenario, if one wants to build a *good* quadratic model, then one needs to interpolate using a sample set of cardinality approximately equal to the square of the dimension of the domain of the objective function, which may turn out to be too costly if the objective function is expensive to evaluate. One alternative is to consider underdetermined models, using a sample set of much smaller size than the one needed for determined interpolation. The main idea for building underdetermined quadratic interpolation models is to minimize a norm of the Hessian of the model subject to the interpolating conditions. The use of the vector ℓ_1-norm in this procedure is the main subject of this thesis.

In most applications, the objective function is defined in a high-dimensional space and very often pairs of variables have no correlation, meaning that the corresponding second order derivatives are zero. Thus, one frequently deals with objective functions for which the Hessian exhibits some level or pattern of sparsity. The main idea of our work is to implicitly and automatically take advantage of the sparsity of the Hessian to build accurate underdetermined models by minimizing the ℓ_1-norm of the Hessian model coefficients.

In fact, our work is inspired from the sparse solution recovery theory developed recently in the field of Compressed Sensing, where one characterizes conditions under which a sparse signal can be accurately recovered from very few random measurements. Such type of recovery is achieved by minimizing the ℓ_1-norm of the unknown signal and can be accomplished in polynomial time.

The contribution of this thesis is twofold. First, we show that it is possible to compute fully quadratic models (i.e., models with the same accuracy as second order Taylor models) for functions with sparse Hessians using much less points than the number required for determined quadratic interpolation, when minimizing the ℓ_1-norm of the Hessian model coefficients to build the underdetermined quadratic models [2]. Essentially, letting n be the dimension of the domain of the function, we will show that one can build fully quadratic models based on random sampling by this approach using only $\mathcal{O}(n(\log n)^4)$ sample points (instead of the $\mathcal{O}(n^2)$ required for the determined case) when the number of non-zero elements in the Hessian of the function is $\mathcal{O}(n)$.

Second, we introduce a practical interpolation-based trust-region DFO algorithm exhibiting a competitive numerical performance when compared to the state-of-the-

art DFO software. We have tested our algorithm using both minimum Frobenius and ℓ_1-norm quadratic underdetermined models, showing the ability of the ℓ_1-approach to improve the results of the Frobenius one in the presence of some form of sparsity in the Hessian of the objective function [2].

This thesis is organized as follows. In Chapter 2, we introduce background material about interpolation models. We give a brief introduction to Compressed Sensing in Chapter 3, introducing also a new concept related to *partial* sparse recovery (in Section 3.2). In Chapter 4, we obtain the main result mentioned above for sparse construction of models for functions with sparse Hessians. The proof of this result is based on sparse bounded orthogonal expansions which are briefly described in the beginning of this chapter. In Chapter 5, we introduce our practical interpolation-based trust-region method and present numerical results for the two underdetermined quadratic model variants, defined by minimum Frobenius and ℓ_1-norm minimization. Finally, in Chapter 6 we draw some conclusions and discuss possible avenues for future research.

The thesis makes extensive use of vector, matrix, and functional norms. We have used ℓ_p and $\|\cdot\|_p$ for vector and matrix norms, whenever the arguments are, respectively, vectors or matrices. The notation $B_p(x;\Delta)$ will represent a closed ball in $\mathbb{R}^n$, centered at x and of radius Δ, in the ℓ_p-norm, i.e., $B_p(x;\Delta) = \{y \in \mathbb{R}^n : \|y - x\|_p \leq \Delta\}$. For norms of functions on normed spaces L, we have used $\|\cdot\|_L$.

Chapter 2

Use of models in DFO trust-region methods

2.1. Fully linear and fully quadratic models

One of the main techniques used in Derivative-Free Optimization (DFO) consists of modeling an objective function $f : D \subset \mathbb{R}^n \to \mathbb{R}$, replacing it locally by models that are 'simple' enough to be optimized and sufficiently 'complex' to well approximate f. If one was given the derivatives of the function, then one could use Taylor approximations as polynomial models for f. However, in DFO one has no access to derivatives or accurate derivatives, and so broader classes of models must be consider. As the simplest kind of Taylor approximations in Optimization with derivatives are linear approximations, in DFO the simplest class of models are the so-called fully linear models. Its definition requires f to be smooth up to the first order.

Assumption 2.1.1 *Assume that f is continuously differentiable with Lipschitz continuous gradient (on an open set containing D).*

One does not restrict fully linear models to linear functions, but instead consider models that approximate f *as well* as the linear Taylor approximations. The following definition reduces essentially to the first part of [13, Definition 6.1] and is stated using balls in an arbitrary ℓ_p-norm, with $p \in (0, +\infty]$.

Definition 2.1.1 *Let a function $f : D \to \mathbb{R}$ satisfying Assumption 2.1.1 be given. A set of model functions $\mathcal{M} = \{m : \mathbb{R}^n \to \mathbb{R},\, m \in C^1\}$ is called a fully linear class of models if the following holds:*

There exist positive constants κ_{ef}, κ_{eg}, and ν_1^m, such that for any $x_0 \in D$ and $\Delta \in (0, \Delta_{max}]$ there exists a model function m in $\mathcal{M}$, with Lipschitz continuous gradient and corresponding Lipschitz constant bounded by ν_1^m, and such that

- *the error between the gradient of the model and the gradient of the function*

satisfies

$$\|\nabla f(u) - \nabla m(u)\|_2 \;\leq\; \kappa_{eg}\,\Delta, \quad \forall u \in B_p(x_0;\Delta),$$

- *and the error between the model and the function satisfies*

$$|f(u) - m(u)| \;\leq\; \kappa_{ef}\,\Delta^2, \quad \forall u \in B_p(x_0;\Delta).$$

Such a model m is called fully linear on $B_p(x_0;\Delta)$.

Linear models such as linear Taylor approximations do not necessarily capture the curvature information of the function they are approximating. To achieve better practical local convergence rates in general it is essential to consider nonlinear models. Quadratic models can be considered the simplest nonlinear models and are widely used and studied in Optimization.

Analogously to the linear case, one can consider a wider class of models not necessarily quadratic. For this purpose, the function f has to exhibit smoothness up to the second order.

Assumption 2.1.2 *Assume that f is twice differentiable with Lipschitz continuous Hessian (on an open set containing D).*

Below we state the first part of the definition of fully quadratic models given in [13, Definition 6.2], again using balls in an ℓ_p-norm, with arbitrary $p \in (0, +\infty]$.

Definition 2.1.2 *Let a function $f : D \to \mathbb{R}$ satisfying Assumption 2.1.2 be given. A set of model functions $\mathcal{M} = \{m : \mathbb{R}^n \to \mathbb{R},\, m \in C^2\}$ is called a fully quadratic class of models if the following holds:*

There exist positive constants κ_{ef}, κ_{eg}, κ_{eh}, and ν_2^m, such that for any $x_0 \in D$ and $\Delta \in (0, \Delta_{max}]$ there exists a model function m in $\mathcal{M}$, with Lipschitz continuous Hessian and corresponding Lipschitz constant bounded by ν_2^m, and such that

- *the error between the Hessian of the model and the Hessian of the function satisfies*

$$\|\nabla^2 f(u) - \nabla^2 m(u)\|_2 \;\leq\; \kappa_{eh}\,\Delta, \quad \forall u \in B_p(x_0;\Delta),$$

- *the error between the gradient of the model and the gradient of the function satisfies*

$$\|\nabla f(u) - \nabla m(u)\|_2 \;\leq\; \kappa_{eg}\,\Delta^2, \quad \forall u \in B_p(x_0;\Delta),$$

- *and the error between the model and the function satisfies*

$$|f(u) - m(u)| \leq \kappa_{ef}\,\Delta^3, \quad \forall u \in B_p(x_0;\Delta).$$

Such a model m is called fully quadratic on $B_p(x_0;\Delta)$.

The definition of fully quadratic model, introduced in [13, Section 6.1], requires further that a finite algorithm exists for building such models. However, a less strict Definition 2.1.2 will be adequate to us since, in Chapter 4, we will use random sample sets to build fully quadratic models, and for that reason we are not interested in introducing such an additional requirement. We will return to this subject in Chapter 5 when covering interpolation-based trust-region methods.

The next proposition justifies the fact that the fully quadratic models are generalizations of Taylor approximations of order 2 (the proof is simple and omitted).

Proposition 2.1.1 *Let f satisfy Assumption 2.1.2. Let T be the Taylor approximation of order 2 of f centered at x_0. Then, for any $\Delta \in (0, \Delta_{\max}]$, T is a fully quadratic model for f on $B_p(x_0;\Delta)$ with $\nu_2^m = 0$ and some positive constants κ'_{ef}, κ'_{eg}, and κ'_{eh}.*

2.2. Quadratic polynomial interpolation models

Now, a natural question that arises is how one can find a fully quadratic model for f on some $B_p(x_0;\Delta)$. In DFO such task is usually achieved by interpolation, in particular by quadratic polynomial interpolation. In order to present the techniques for quadratic polynomial interpolation used in DFO, we need to first introduce some basic facts about quadratic bases.

Let $\mathcal{P}_n^2$ be the space of polynomials of degree less than or equal to 2 in $\mathbb{R}^n$. The dimension of this space is $(n+1)(n+2)/2$. A basis ϕ for $\mathcal{P}_n^2$ will be denoted by $\phi = \{\phi_\iota\}$ with ι belonging to a set Υ of indices of cardinality $(n+1)(n+2)/2$. The most natural basis is the one consisting of the monomials, or the *canonical basis*. This basis appears naturally in Taylor models and is given by

$$\bar{\phi} = \left\{1, u_1, ..., u_n, \frac{1}{2}u_1^2, ..., \frac{1}{2}u_n^2, u_1u_2, ..., u_{n-1}u_n\right\}. \tag{2.1}$$

We say that the quadratic function q interpolates f at a given point w if $q(w) = f(w)$. Assume that we are given a set $W = \{w^1, ..., w^k\} \subset \mathbb{R}^n$ of interpolation points.

A quadratic function q that interpolates f at the points in W, written as

$$q(u) = \sum_{\iota \in \Upsilon} \alpha_\iota \phi_\iota(u),$$

must satisfy the following k interpolation conditions $\sum_{\iota \in \Upsilon} \alpha_\iota \phi_\iota(w^i) = f(w^i)$, $i = 1, ..., k$. These conditions form a linear system,

$$M(\phi, W)\alpha = f(W), \tag{2.2}$$

where $M(\phi, W)$ is the interpolation matrix and $f(W)_i = f(w^i),\ i = 1, ..., k$.

2.2.1. Determined models

A sample set W is poised for (determined) quadratic interpolation if the corresponding interpolation matrix $M(\phi, W)$ is non-singular, guaranteeing that there exists one and only one quadratic polynomial q such that $q(W) = f(W)$. It is not hard to prove that this definition of poisedness does not depend on f nor on the basis ϕ (see [13, Chapter 3]). Roughly speaking, W is well poised for (determined) quadratic interpolation if the condition number of a scaled version of $M(\bar{\phi}, W)$ is relatively small (see [13, Section 3.4]). The next theorem, rigorously stated and proved in [13, Chapter 3 and 6], justifies the use of interpolation for building quadratic models.

Theorem 2.2.1 *If $W \subset B_2(x_0; \Delta)$ is a well poised sample set for quadratic interpolation, then the quadratic function q that interpolates f in W is a fully quadratic model for f on $B_2(x_0; \Delta)$.*

For a sample set W to be poised for quadratic interpolation, it has to be of size $(n+1)(n+2)/2$, since $M(\phi, W)$ needs to be non-singular. However, building such a sample set costs $(n+1)(n+2)/2$ evaluations of the function f and that is often too expensive in 'real-world' scenarios. One way of lowering this cost is by considering smaller sample sets, which makes the linear system in (2.2) underdetermined.

2.2.2. Underdetermined interpolating models

To properly introduce the underdetermined case, we need to recall first from [13, Section 2.3] the notions of poisedness for linear interpolation and regression. For this purpose, let us split the basis $\bar{\phi}$ in (2.1) into its linear and quadratic components: $\bar{\phi}_L = \{1, u_1, ..., u_n\}$ and $\bar{\phi}_Q = \bar{\phi} \setminus \bar{\phi}_L$. A sample set W with $|W| = n + 1$ is said to be poised for linear interpolation if $M(\bar{\phi}_L, W)$ is non-singular. Also, a sample set W with $|W| > n + 1$ is said to be poised for linear regression when $M(\bar{\phi}_L, W)$

is full column rank. Roughly speaking, well poisedness, in both cases, means that one has a relatively small condition number of a scaled version of $M(\bar{\phi}_L, W)$, see [13, Section 4.4].

Let us now consider a sample set W, with $|W| \in [n+2, (n+1)(n+2)/2]$, for quadratic underdetermined interpolation. Since, often, such sample sets are poised for linear regression, one could compute linear regression models, possibly leading to fully linear models (see [13, Theorem 2.13]). However, the next theorem (see [13, Theorem 5.4] for a rigorous statement and proof) will guarantee that underdetermined quadratic interpolation will work, at least, as well as linear regression in this respect.

Theorem 2.2.2 *Let q be a quadratic function that interpolates f in W, where $W \subset B_2(x_0; \Delta)$ is a sample set well poised for linear regression. Then, q is a fully linear model (see Definition 2.1.1) for f on $B_2(x_0; \Delta)$, and the error constants κ_{ef} and κ_{eg} are $\mathcal{O}(1 + \|\nabla^2 q\|_2)$.*

Since the system (2.2) is now possibly underdetermined, one needs a criterion for choosing the *best solution* amongst the possible ones. Theorem 2.2.2 suggests that one should build underdetermined quadratic models with 'small' model Hessians. Keeping this in mind, we will describe two underdetermined models that consist in considering the *best solution* of (2.2) as a solution with a 'smallest' model Hessian.

One such underdetermined interpolation model consists of minimizing the Frobenius norm of the Hessian model subject to (2.2). Recalling the split of the basis $\bar{\phi}$ into the linear and the quadratic parts, one can write the interpolation model as

$$q(u) = \alpha_L^T \bar{\phi}_L(u) + \alpha_Q^T \bar{\phi}_Q(u),$$

where α_L and α_Q are the appropriate parts of the coefficient vector α. The minimum Frobenius norm solution [13, Section 5.3] can now be defined as the solution to the following optimization problem

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|\alpha_Q\|_2^2 \\ \text{s.t.} \quad & M(\bar{\phi}_L, W)\alpha_L + M(\bar{\phi}_Q, W)\alpha_Q = f(W). \end{aligned} \tag{2.3}$$

(If $|W| = (n+1)(n+2)/2$, this reduces to determined quadratic interpolation.) Note that (2.3) is a convex quadratic program and thus equivalent to its first order necessary conditions, which can be written in the form (λ being the corresponding

multipliers)

$$\begin{bmatrix} M(\bar{\phi}_Q, W)M(\bar{\phi}_Q, W)^T & M(\bar{\phi}_L, W) \\ M(\bar{\phi}_L, W)^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \alpha_L \end{bmatrix} = \begin{bmatrix} f(W) \\ 0 \end{bmatrix} \tag{2.4}$$

and $\alpha_Q = M(\bar{\phi}_Q, W)^T\lambda$. One says that a set W is poised for minimum Frobenius norm interpolation if problem (2.4) has a unique solution or, equivalently, if the matrix in (2.4) is non-singular. Note also that, due to the choice of $\bar{\phi}$, minimizing $\|\alpha_Q\|_2$ is equivalent to minimizing the Frobenius norm of the upper or lower triangular parts of the Hessian of the model.

From [13, Section 5.3], if a sample set W is (well) poised for minimum Frobenius norm interpolation, one automatically has that W is (well) poised for linear regression. Thus, from Theorem 2.2.2, if W if well poised for minimum Frobenius norm interpolation, then the minimum Frobenius norm interpolating model can be fully linear with error constants κ_{ef} and κ_{eg} of the order of $\mathcal{O}(1 + \|\nabla^2 q\|_2)$ — so, by minimizing the Frobenius norm of the (upper/lower triangular part of the) Hessian of the model, one is also lowering the error constants.

Powell [27] has rather considered the minimization of the Frobenius norm of the difference between the model of the Hessian and a reference matrix. Conn, Scheinberg, and Vicente [13, Section 5.3] have studied version (2.3) in detail.

The second approach to build underdetermined quadratic models, which will be studied in Chapter 4 and is the main object of study of this thesis, consists in considering the solution to the following optimization problem

$$\begin{aligned} \min \quad & \|\alpha_Q\|_1 \\ \text{s.t.} \quad & M(\bar{\phi}_L, W)\alpha_L + M(\bar{\phi}_Q, W)\alpha_Q = f(W), \end{aligned} \tag{2.5}$$

where $\alpha_L, \alpha_Q, \bar{\phi}_L, \bar{\phi}_Q$ are defined as in (2.3). As we will show later, in Chapter 4, this approach is appealing when there is no correlation between some of the variables of the objective function f, meaning that the Hessian of f has several zero elements in its non-diagonal part. In such functions, with sparse Hessian, we will be able to recover, with high probability, fully quadratic models with much less than $(n+1)(n+2)/2$ random points. One other advantage of this approach is that minimizing the ℓ_1-norm subject to linear constraints is also tractable, since (2.5) is equivalent to a linear program (see, e.g., Section 5.3). Obviously, by minimizing the ℓ_1-norm of the entries of the Hessian model one is also lowering its ℓ_2-norm and therefore there is also a direct connection between posing (2.5) and aiming at Theorem 2.2.2.

Chapter 3

Compressed Sensing

3.1. General concepts and properties

In Discrete Signal Processing one is often interested in an inverse problem consisting of recovering a signal (i.e., a vector) x from its discrete Fourier transform $\hat{x}$ (also a vector). If all the components of $\hat{x}$ are available, then this can be easily accomplished by the inverse Fourier transform. However, if only a portion of $\hat{x}$ is known, this problem becomes an underdetermined inverse one. The usual technique to solve this kind of problems is to determine the corresponding *best solution* z such that the corresponding $\hat{z}$ coincides with the known part of $\hat{x}$, whatever *best solution* means in each context, and hope that it is close to, or even coincides with, x. When one has no additional information on x, the classical approach is to find the least squares solution z, in other words the solution with least ℓ_2-norm[1]. However, in many real applications the signals in question are known to be sparse, in the sense of having many zero components (a fact that seems to be heavily explored in modern data compression techniques such as JPEG2000 [33]), and it turns out that the ℓ_2-norm is not appropriate to recover sparse signals. Since we are trying to find sparse solutions to an inverse problem, the naive approach would be to consider the sparsest solution instead of the one with minimal energy, but this problem turns out to be highly combinatorial and NP-Hard [14, 24] and thus a cheaper alternative is desirable.

As the discrete Fourier transform is a linear operator, it can be represented by an $N \times N$ complex matrix F, yielding then $\hat{x} = Fx$. If only a subset $\Omega \subset [N]$ (here $[N]$ denotes the set $\{1, ..., N\}$) of the entries of $\hat{x}$ are known, then all vectors z such that $F_{[\Omega\cdot]}z = \hat{x}_{[\Omega\cdot]}$ are possible solutions to the inverse problem (here $F_{[\Omega\cdot]}$ is the submatrix of F formed by its rows in Ω; $\hat{x}_{[\Omega\cdot]}$ is the subvector of $\hat{x}$ formed by the indices in Ω). Since minimizing the number of non-zero components, i.e., its

[1] The ℓ_2-norm of a signal is often referred to as the energy of the signal.

ℓ_0-*norm*[2],

$$\min \|z\|_0 \quad \text{s. t.} \quad F_{[\Omega\cdot]} z = \hat{x}_{[\Omega\cdot]}, \tag{3.1}$$

is NP-Hard, one must consider a tractable approximation to this problem. The approach that we will consider here is generally referred to as a *convex relaxation*, since one substitutes the non-convex ℓ_0-*norm* by a convex underestimating function that is close to it. Recent results suggest that ℓ_1-norm works well in practice (see [7] for a survey on some of this material). In fact, the ℓ_1-norm is the convex relaxation of the function $g(u) = \|u\|_0$ restricted to $B_\infty(0;1)$ (see [22]). Formally, this strategy consists of solving the optimization problem

$$\min \|z\|_1 \quad \text{s. t.} \quad F_{[\Omega\cdot]} z = \hat{x}_{[\Omega\cdot]}. \tag{3.2}$$

This optimization problem is much easier to solve than (3.1). In fact, (3.2) is equivalent to a linear program (see, e.g., Section 5.3) and, therefore, it can be solved efficiently with state of the art Linear Programming software based on the simplex method or on interior-point methods [25].

One can now consider a much broader setting where the available information about the sparse vector to be recovered is not necessarily of the form $\hat{x}_{[\Omega\cdot]} = F_{[\Omega\cdot]} x$ but, more generally, of the form $y = Ax$, where $x \in \mathbb{R}^N$, $y \in \mathbb{R}^k$, and A is a $k \times N$ matrix (here considered real) with far fewer rows than columns ($k \ll N$). A measure of sparsity is now in order and we will say that a vector x is $s-$sparse if $\|x\|_0 \leq s$. One is interested in *measurements* matrices A such that, for every $s-$sparse vector x, the information given by $y = Ax$ is enough to recover x and, moreover, that such recovery can be accomplished by solving the problem

$$\min \|z\|_1 \quad \text{s. t.} \quad Az = y. \tag{3.3}$$

The next definition will provide an alternative characterization for such matrices. Given $v \in \mathbb{R}^N$ and $S \in [N]$, $v_S \in \mathbb{R}^N$ is a vector defined by $(v_S)_i = v_i$, $i \in S$ and $(v_S)_i = 0$, $i \notin S$. Also, $[N]^{(s)}$ is the set of subsets of $[N]$ with size s.

Definition 3.1.1 (Null Space Property) *The matrix $A \in \mathbb{R}^{k \times N}$ is said to satisfy the Null Space Property (NSP) of order s if, for every $v \in \mathcal{N}(A) \setminus \{0\}$ and for every $S \in [N]^{(s)}$, one has*

$$\|v_S\|_1 < \frac{1}{2}\|v\|_1. \tag{3.4}$$

[2]The ℓ_0-*norm* is defined by $\|u\|_0 = |\{i : u_i \neq 0\}|$ but, strictly speaking, is not a norm.

The term Null Space Property was introduced in [8]. However, we note that the characterization mentioned above and formalized in the following theorem had already been implicitly used in [18].

Theorem 3.1.1 *The matrix A satisfies the Null Space Property of order s if and only if, for every $s-$sparse vector x, problem (3.3) with $y = Ax$ has an unique solution and it is given by x.*

Proof. Let us assume first that every $s-$sparse vector x is the unique minimizer of $\|z\|_1$ subject to $Az = Ax$. Then, in particular, for any $v \in \mathcal{N}(A) \setminus \{0\}$ and for any $S \in [N]^{(s)}$, the $s-$sparse vector v_S is the unique minimizer of $\|z\|_1$ subject to $Az = Av_S$. As $-v_{S^c}$ is a solution of $Az = Av_S$, where $S^c = [N] \setminus S$, one must have $\|v_{S^c}\|_1 > \|v_S\|_1$, and then

$$2\|v_S\|_1 < \|v_{S^c}\|_1 + \|v_S\|_1 = \|v\|_1,$$

and thus (3.4) holds.

To show the other implication, let us now assume that A satisfies the NSP of order s. Then, given an $s-$sparse vector x and a vector z not equal to x and satisfying $Ax = Az$, consider $v = x - z \in \mathcal{N}(A) \setminus \{0\}$ and $S = \text{supp}(x) = \{i : x_i \neq 0\}$ the support of x. One has that

$$\begin{aligned} \|x\|_1 &\leq \|x - z_S\|_1 + \|z_S\|_1 \\ &= \|v_S\|_1 + \|z_S\|_1 \\ &< \|v_{S^c}\|_1 + \|z_S\|_1 \\ &= \| - z_{S^c}\|_1 + \|z_S\|_1 \\ &= \|z\|_1, \end{aligned}$$

(the strict inequality coming from (3.4)), guaranteeing that x is the unique solution of (3.3) with $y = Ax$. ■

The NSP is difficult to be directly verified. On the other hand, the *Restricted Isometry Property* (RIP), introduced in [6] under a different term, has become very popular. The RIP is considerable more useful, although it provides only sufficient conditions for every $s-$sparse vector x to be the unique solution of (3.3) when $y = Ax$. An intuitive reason for being more useful, in opposition to the NSP which exhibits an algebraic and combinatorial nature, is that the RIP is connected to arguments from

Analysis and Operator Theory. The recovery results that we will use later in Chapter 4 in our sparse Hessian recovery setting are obtained by proving that the underlying matrices satisfy the RIP. We present now the definition of *Restricted Isometry Property Constant*, known in the literature as the Restricted Isometry Property.

Definition 3.1.2 (Restricted Isometry Property) *One says that $\delta_s > 0$ is the Restricted Isometry Property Constant, or* RIP *constant, of order s of the matrix $A \in \mathbb{R}^{k\times N}$ if δ_s is the smallest positive real such that:*

$$(1-\delta_s)\,\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_s)\,\|x\|_2^2 \tag{3.5}$$

for every $s-$sparse vector x.

The following theorem (see, e.g., [29]) provides a useful sufficient condition for successful recovery by (3.3) with $y = Ax$.

Theorem 3.1.2 *Let $A \in \mathbb{R}^{k\times N}$. If*

$$2\delta_{2s} + \delta_s < 1, \tag{3.6}$$

where δ_s and δ_{2s} are the RIP constants of order, respectively, s and $2s$ of A, then, for every $s-$sparse vector x, problem (3.3) with $y = Ax$ has an unique solution and it is given by x.

The following proposition will be needed to prove the theorem.

Proposition 3.1.1 *Let A be a real $k \times N$ matrix with RIP constant δ_r and u and v be two vectors with disjoint support such that $|\operatorname{supp}(u)| + |\operatorname{supp}(v)| = r$. Then,*

$$|\langle Au, Av\rangle| \leq \delta_r \|u\|_2 \|v\|_2.$$

Proof. First we note that the inequality (3.5) is equivalent to

$$\left|\|Ax\|_2^2 - \|x\|_2^2\right| \leq \delta_r \|x\|_2^2.$$

If $\operatorname{supp}(x) = S \in [N]^{(r)}$, then

$$\|Ax\|_2^2 - \|x\|_2^2 = \left\|A_{[\cdot S]}x_{[S\cdot]}\right\|_2^2 - \left\|x_{[S\cdot]}\right\|_2^2 = \left\langle \left(\left(A_{[\cdot S]}\right)^T A_{[\cdot S]} - I\right) x_{[S\cdot]}, x_{[S\cdot]} \right\rangle,$$

where $A_{[\cdot S]}$ denotes the submatrix of A formed by the columns in S. Then,

$$\left|\|Ax\|_2^2 - \|x\|_2^2\right| = \left|\left\langle \left(\left(A_{[\cdot S]}\right)^T A_{[\cdot S]} - I\right) x_{[S\cdot]}, x_{[S\cdot]} \right\rangle\right| \leq \left\|A_{[\cdot S]}^T A_{[\cdot S]} - I\right\|_2 \|x\|_2^2$$

and the above inequality is satisfied as equality for some $x \in \mathbb{R}^N$ supported on S, given the definition of the ℓ_2-norm of the real and symmetric matrix $\left(A_{[\cdot S]}\right)^T A_{[\cdot S]} - I$. Thus, the RIP constant of order r can be given as

$$\delta_r = \max_{S \in [N]^{(r)}} \left\| A_{[\cdot S]}^T A_{[\cdot S]} - I \right\|_2 . \tag{3.7}$$

Now, let S_1 and S_2 be, respectively, the support of u and v. From the assumptions of the proposition, $|S_1|+|S_2| = r$. It is easy to see from the disjunction of the supports of u and v that

$$\begin{aligned} \langle Au, Av \rangle &= \left(u_{[(S_1 \cup S_2)\cdot]}\right)^T \left(A_{[\cdot (S_1 \cup S_2)]}\right)^T A_{[\cdot (S_1 \cup S_2)]} v_{[(S_1 \cup S_2)\cdot]} \\ &= \left(u_{[(S_1 \cup S_2)\cdot]}\right)^T \left(\left(A_{[\cdot (S_1 \cup S_2)]}\right)^T A_{[\cdot (S_1 \cup S_2)]} - I \right) v_{[(S_1 \cup S_2)\cdot]}. \end{aligned}$$

Thus,

$$\begin{aligned} \langle Au, Av \rangle &\leq \left\| A_{[\cdot (S_1 \cup S_2)]}^T A_{[\cdot (S_1 \cup S_2)]} - I \right\|_2 \|u\|_2 \|v\|_2 \\ &\leq \delta_r \|u\|_2 \|v\|_2. \end{aligned}$$

■

An interesting fact that can be deduced from (3.7) is that the RIP requires, in particular, that all submatrices of A, obtained considering r columns of A, are well-conditioned, in the sense that the condition number of $\left(A_{[\cdot S]}\right)^T A_{[\cdot S]}$ is bounded by $(1+\delta_r)/(1-\delta_r)$.

Theorem 3.1.2 can now be proved.

Proof. (of Theorem 3.1.2) The strategy is to prove that $2\delta_{2s} + \delta_s < 1$ implies that A satisfies the NSP of order s. Let $v \in \mathcal{N}(A) \setminus \{0\}$ be given and consider the index set S_0 of the s largest entries of v in absolute value. Recursively, consider also the index sets S_i, with $i \geq 1$, given by the indices corresponding to the largest s entries in absolute value of v in $[N] \setminus (S_0 \cup \cdots \cup S_{i-1})$, until this set is empty[3]. We have that $A(v_{S_0}) = -A(v_{S_1} + v_{S_2} + \cdots)$. From Definition 3.1.2

$$\begin{aligned} \|v_{S_0}\|_2^2 &\leq \frac{1}{1-\delta_s} \|A(v_{S_0})\|_2^2 \\ &= \frac{1}{1-\delta_s} \langle A(v_{S_0}), A(-v_{S_1}) + A(-v_{S_2}) + \cdots \rangle \\ &= \frac{1}{1-\delta_s} \sum_{i \geq 1} \langle A(v_{S_0}), A(-v_{S_i}) \rangle. \end{aligned}$$

From Proposition 3.1.1, with $r = 2s$, one has $\langle A(v_{S_0}), A(-v_{S_i}) \rangle \leq \delta_{2s} \|v_{S_0}\|_2 \|v_{S_i}\|_2$. Substituting this latter inequality in the inequality above and dividing both members

[3]The last index set to be chosen will, in general, have less than s elements.

by $\|v_{S_0}\|_2$, one obtains

$$\|v_{S_0}\|_2 \leq \frac{\delta_{2s}}{1-\delta_s} \sum_{i \geq 1} \|v_{S_i}\|_2. \tag{3.8}$$

In addition, one also has that

$$|v_j| \leq \frac{1}{s} \sum_{l \in S_{i-1}} |v_l| \text{ for all } j \in S_i, \tag{3.9}$$

because the absolute values of the nonzero entries of v_{S_i} do not exceed the ones of $v_{S_{i-1}}$, $i \geq 1$. Hence, from (3.9),

$$\|v_{S_i}\|_2 = \left(\sum_{j \in S_i} |v_j|^2 \right)^{\frac{1}{2}} \leq \left(\frac{s}{s^2} \left(\sum_{l \in S_{i-1}} |v_l| \right)^2 \right)^{\frac{1}{2}} \leq \frac{1}{\sqrt{s}} \|v_{S_{i-1}}\|_1,$$

and, using (3.8),

$$\|v_{S_0}\|_2 \leq \frac{\delta_{2s}}{1-\delta_s} \sum_{i \geq 1} \|v_{S_i}\|_2 \leq \frac{\delta_{2s}}{1-\delta_s} \sum_{i \geq 1} \frac{1}{\sqrt{s}} \|v_{S_{i-1}}\|_1 \leq \frac{1}{\sqrt{s}} \frac{\delta_{2s}}{1-\delta_s} \|v\|_1.$$

Since $\|v_{S_0}\|_1 \leq \sqrt{s}\|v_{S_0}\|_2$, one has, from (3.6),

$$\|v_{S_0}\|_1 \leq \frac{\delta_{2s}}{1-\delta_s} \|v\|_1 < \frac{1}{2} \|v\|_1. \tag{3.10}$$

The proof can now be completed since the indices in S_0 correspond to the largest entries of v in absolute value. In fact from (3.10) we then have

$$\|v_S\|_1 \leq \|v_{S_0}\|_1 < \frac{1}{2} \|v\|_1,$$

for any $S \in [N]^{(s)}$. Thus A satisfies the NSP of order s and the proof is completed by appealing to Theorem 3.1.1. ■

Note that, since $\delta_{2s} < 1/3$ trivially implies (together with $\delta_s \leq \delta_{2s}$) the inequality (3.6), the assumptions in Theorem 3.1.2 are often written in the literature as $\delta_{2s} < 1/3$. However, we chose to present, in Theorem 3.1.2, a slightly stronger version.

Although the RIP provides useful sufficient conditions for sparse recovery, it is a difficult and still open problem to find deterministic matrices which satisfy such property when the underlying system is highly underdetermined (see [32] for an interesting discussion on this topic). Intuitively, the difficulty arises from the fact that it is not sufficient to estimate the entries of the matrices in absolute values, since cancellations by the different signs play a vital role. It turns out that random matrices provide a better ground for this analysis because one can use central limit

type theorems and other concentration of measure results that take into account possible cancelations.

We now state one of the classic results in Compressed Sensing. There are many proofs of this result now available (see, e.g., [3] for a particularly interesting one).

Theorem 3.1.3 *Consider the setting of Definition 3.1.2 and $\varepsilon \in (0,1)$. Let $A \in \mathbb{R}^{k\times N}$ be now a normalized Gaussian or Bernoulli random matrix, whose entries are, respectively Gaussian or Bernoulli jointly independent random variables with mean 0 and variance $1/k$, and assume that*

$$k \geq C\delta^{-2}\left(s\log\left(\frac{N}{s}\right) + \log\left(\varepsilon^{-1}\right)\right), \tag{3.11}$$

for some $\delta > 0$ and some universal[4] constant $C > 0$. Then, with probability at least $1-\varepsilon$, the RIP constant of A satisfies $\delta_s \leq \delta$.

The sparse Hessian recovery result that we will derive in Section 4.3 uses a different characterization of RIP involving bounded orthonormal expansions, which will be introduced in Section 4.1. As we will see then, such characterization is a natural generalization of the Discrete Fourier Transform setting mentioned in the beginning of this chapter.

3.2. RIP and NSP for partial sparse recovery

One natural question arising from Chapter 2 (see problem (2.5)) is the existence of a partial recovery result in the sense that sparsity is of interest for only a subset of the vector components. Formally, one has $x = (w,z)^T$, where $w \in \mathbb{R}^{N-r}$ is $(s-r)-$sparse and $z \in \mathbb{R}^r$. A natural generalization of problem (3.3) to this setting of partial sparse recovery is given by

$$\min \|w\|_1 \quad \text{s.t.} \quad A\begin{pmatrix} w \\ z \end{pmatrix} = y. \tag{3.12}$$

We can also define a similar null space property as described in the next definition.

Definition 3.2.1 (Null Space Property for Partial Sparse Recovery) *Let us write $A = (A_1\ A_2)$ where A_1 has the first $N-r$ columns of A and A_2 the last r. We say that A satisfies the Null Space Property (NSP) of order $s-r$ for Partial Sparse*

[4] By universal constant we mean a constant independent of all the problem data, as defined in the literature of Compressed Sensing [4, 5, 6, 17, 29].

Recovery of size $N-r$ with $r \leq s$ if, for every $v \in \mathbb{R}^{N-r} \setminus \{0\}$ such that $A_1 v \in \mathcal{R}(A_2)$ and every $S \in [N-r]^{(s-r)}$, we have

$$\|v_S\|_1 < \frac{1}{2}\|v\|_1. \tag{3.13}$$

Note that when $r = 0$ the NSP for Partial Sparse Recovery reduces to the NSP of Definition 3.1.1. We will be able to reconstruct the part of a vector for which sparsity is of interest if and only if the matrix satisfies the corresponding null space property.

Theorem 3.2.1 *The matrix A satisfies the Null Space Property of order $s-r$ for Partial Sparse Recovery of size $N-r$ if and only if for every $x = (x_1, x_2)^T$ such that $x_1 \in \mathbb{R}^{N-r}$ is $(s-r)-$sparse and $x_2 \in \mathbb{R}^r$, all solutions $(\tilde{x}_1, \tilde{x}_2)^T$ of problem (3.12) with $A(w,z)^T = A(x_1,x_2)^T$ satisfy $\tilde{x}_1 = x_1$.*

Proof. The proof follows the one of Theorem 3.1.1 with appropriate modifications. Let us assume first that for every vector $(x_1, x_2)^T \in \mathbb{R}^N$, where x_1 is an $(s-r)-$sparse vector and $x_2 \in \mathbb{R}^r$, every minimizer $(\tilde{x}_1, \tilde{x}_2)^T$ of $\|w\|_1$ subject to $A(w,z)^T = A(x_1,x_2)^T$ satisfies $\tilde{x}_1 = x_1$. Define A_1 and A_2 as in Definition 3.2.1. Then, in particular, for any $u_1 \in \mathbb{R}^r$ and any $v \neq 0$ such that $A_1 v \in \mathcal{R}(A_2)$ and for any $S \in [N-r]^{(r-s)}$, every minimizer $(\tilde{x}_1, \tilde{x}_2)^T$ of $\|w\|_1$ subject to $A(w,z)^T = A(v_S, u_1)^T$ satisfies $\tilde{x}_1 = v_S$. Since $A_1 v \in \mathcal{R}(A_2)$, there exists u_2 such that $(-v_{S^c}, u_2)^T$ is a solution of $A(w,z)^T = A(v_S,u_1)^T$. As $-v_{S^c} \neq v_S$, $(-v_{S^c}, u_2)^T$ is not a minimizer of $\|w\|_1$ subject to $A(w,z)^T = A(v_S,u_1)^T$, so $\|v_{S^c}\|_1 > \|v_S\|_1$ and, as in the proof of Theorem 3.1.1,

$$\|v_S\|_1 < \frac{1}{2}\|v\|_1.$$

On the other hand, let us assume that A satisfies the NSP of order $s-r$ for Partial Sparse Recovery of size $N-r$ (Definition 3.2.1). Then, given a vector $(x_1,x_2)^T \in \mathbb{R}^N$, with x_1 an $(s-r)-$sparse vector and $x_2 \in \mathbb{R}^r$ and a given vector $(w,z)^T \in \mathbb{R}^N$ with $w \neq x_1$ and satisfying $A(x_1,x_2)^T = A(w,z)^T$, we consider $(v,u)^T = ((x_1-w),(x_2-z))^T \in \mathcal{N}(A)$ which, together with $w \neq x_1$, implies

$A_1 v \in \mathcal{R}(A_2) \setminus \{0\}$. Thus, setting $S = \text{supp}(x)$, one has that

$$\begin{aligned} \|x_1\|_1 &\leq \|x_1 - w_S\|_1 + \|w_S\|_1 \\ &= \|v_S\|_1 + \|w_S\|_1 \\ &< \|v_{S^c}\|_1 + \|w_S\|_1 \\ &= \| - w_{S^c}\|_1 + \|w_S\|_1 \\ &= \|w\|_1, \end{aligned}$$

(the strict inequality coming from (3.13)), guaranteeing that all solutions $(\tilde{x}_1, \tilde{x}_2)$ of problem (3.12) with $A(w,z)^T = A(x_1,x_2)^T$ satisfy $\tilde{x}_1 = x_1$. ■

Note that Theorem 3.2.1 guarantees that, if A satisfies the NSP of order $s-r$ for Partial Sparse Recovery of size $N-r$, given a vector $x = (x_1, x_2)^T$ such that $x_1 \in \mathbb{R}^{N-r}$ is $(s-r)-$sparse and $x_2 \in \mathbb{R}^r$, one is able to recover x_1 solving problem (3.12). This automatically implies that if A_2 (as defined in Definition 3.2.1) is full column rank, then the entire vector is recovered, since then x_2 is uniquely determined by $A_2 x_2 = A_1 x_1 - y$, where y is given by $y = Ax$. Such an assumption on A_2 is reasonable in the applications we will consider in Chapter 4, where r is typically much smaller than N.

We can also define an analogous extension of the RIP to the partial sparse recovery setting but, for that purpose, we will first consider an alternative way to study partial sparsity.

For this purpose, let $A = (A_1 \ A_2)$ as considered above, with the additional reasonable assumption that A_2 has full column rank. Let

$$\mathcal{P} = I - A_2 \left(A_2^T A_2\right)^{-1} A_2^T \tag{3.14}$$

be the matrix representing the projection from $\mathbb{R}^N$ onto $\mathcal{R}\left(A_2\right)^{\perp}$. Then, the problem of recovering $(x_1, x_2)^T$, where x_1 is an $(s-r)-$sparse vector satisfying $A_1 x_1 + A_2 x_2 = y$, can be stated as the problem of recovering an $(s-r)-$sparse vector x_1 satisfying $\left(\mathcal{P}A_1\right) x_1 = \mathcal{P}y$ and then recovering x_2 satisfying $A_2 x_2 = A_1 x_1 - y$. The latter task results in the solution of a linear system of unique solution given that A_2 has full column rank and $(\mathcal{P}A_1)x_1 = \mathcal{P}y$. Note that the former task resumes to the classical setting of Compressed Sensing. These considerations motivate the following definition of RIP for partial sparse recovery.

Definition 3.2.2 (Partial RIP) *We say that $\delta_{s-r}^r > 0$ is the Partial Restricted Isometry Property Constant of order $s-r$ for recovery of size $N-r$ of the matrix*

$A = (A_1\ A_2) \in \mathbb{R}^{k\times N}$ with $r \leq s$ if δ^r_{s-r} is the RIP constant of order $s-r$ (see Definition 3.1.2) of the matrix $\mathcal{P}A_1$, where $\mathcal{P}$ is given by (3.14).

Again, when $r = 0$ the Partial RIP reduces to the RIP of Definition 3.1.2. We also note that, given a matrix $A = (A_1\ A_2) \in \mathbb{R}^{k\times N}$ with Partial RIP constants of order $s-r$ and $2(s-r)$, for recovery of size $N-r$, satisfying $2\delta^r_{2(s-r)} + \delta^r_{s-r} < 1$, then, by Theorem 3.1.2, we have that $\mathcal{P}A_1$ satisfies the NSP of order $s-r$. Thus, given $x = (x_1, x_2)^T$ such that $x_1 \in \mathbb{R}^{N-r}$ is $(s-r)-$sparse and $x_2 \in \mathbb{R}^r$, x_1 can be recovered by minimizing the ℓ_1-norm of z subject to $(\mathcal{P}A_1)z = \mathcal{P}Ax$ and, recalling that A_2 is full-column rank, x_2 is uniquely determined by $A_2x_2 = Ax - A_1x_1$. (In particular, this implies that A satisfies the NSP of order $s-r$ for Partial Sparse Recover of size $N-r$.)

It is still under investigation the advantages of partial recovery (i.e., of the solution of problem (3.12)) over the recovery given by (3.3), in the general setting of Compressed Sensing. In particular it is not known if results similar to Theorem 3.1.3 are possible to obtain, and if so, how would the bound (3.11) be affected — it is clear that Theorem 3.1.3 can be directly applied, ignoring the partial sparsity structure, but one would hope that such result could be improved using the extra information on the structure of the sparsity. Numerical results in the setting of sparse Hessian approximation of Chapter 4 indicate that partial recovery is worth considering.

Chapter 4

Recovery of Sparse Hessians

4.1. Sparse recovery using orthonormal expansions

Let us start by recalling the discrete Fourier transform setting, discussed in the beginning of Chapter 3. One can consider now the vector $\hat{x} \in \mathbb{C}^N$ as a function $\hat{x} : [N] \to \mathbb{C}$ given by

$$\hat{x}(k) = \sum_{j=1}^{N} x_j F_j(k),$$

where $F_j(k) = (1\sqrt{N})e^{2\pi ijk/N}$ is the (j,k) entry of the discrete Fourier transform matrix.

The problem of recovering x from partial information about $\hat{x}$ (of the form $\hat{x}_{[\Omega\cdot]}$ with $\Omega \subset [N]$) can now be stated as recovering the function $\hat{x}$ knowing its values in a subset $\Omega \subset [N]$, and can be accomplished by minimizing the ℓ_1-norm of the vector x of the *expansion coefficients*. Keeping in mind this point of view, one natural generalization is to consider a function $g : \mathcal{D} \to \mathbb{R}$ belonging to a finite dimensional functional space (with a known basis $\phi = \{\phi_1, ..., \phi_N\}$ of functions defined in $\mathcal{D}$), so that g can be written as

$$g = \sum_{j=1}^{N} \alpha_j \phi_j,$$

for some expansion coefficients $\alpha_1, ..., \alpha_N$. We are interested in the problem of recovering g from its values in some finite subset $W = \{w^1, ..., w^k\} \subset \mathcal{D}$ with $k \leq N$, with the additional assumption that g is $s-$sparse, meaning that the expansion coefficients vector α is $s-$sparse. Moreover, one would expect that this task could be accomplished by minimizing the ℓ_1-norm of α, subject to the interpolation conditions,

$$M(\phi, W)\alpha = g(W),$$

with $M(\phi, W)$ the interpolation matrix of elements $[M(\phi, W)]_{ij} = \phi_j(w^i)$ (see (2.2)) and $g(W)$ the vector of components $g(w^i)$, $i = 1, ..., k$, $j = 1, ..., N$.

The purpose of this section is to describe such form of recovery for a certain type of bases. Although the results of this section hold also for complex valued functions,

we will restrict ourselves to the real case, because the functions we are interested in Derivative-Free Optimization are real valued. We consider a probability measure μ defined in $\mathcal{D}$ (having in mind that $\mathcal{D} \subset \mathbb{R}^n$). The basis ϕ will be required to satisfy the following orthogonality property [29].

Definition 4.1.1 (K-bounded orthonormal basis) *A basis $\phi = \{\phi_1, ..., \phi_N\}$ is said to be an orthonormal basis satisfying the K-boundedness condition (in the domain $\mathcal{D}$ for the measure μ) if*

$$\int_{\mathcal{D}} \phi_i(u)\phi_j(u)d\mu(u) = \delta_{ij}$$

and $\|\phi_j\|_{L^\infty(\mathcal{D})} \leq K$, for all $i, j \in [N]$.

As in Compressed Sensing, randomness will play a key role to provide recovery with less points than basis coefficients as it is described in the following theorem, whose proof [29] we omit for the sake of brevity.

Theorem 4.1.1 (Rauhut [29]) *Let $M(\phi, W) \in \mathbb{R}^{k\times N}$ be the interpolation matrix associated with an orthonormal basis satisfying the K-boundedness condition. Assume that the sample set $W = \{w^1, ..., w^k\} \subset \mathcal{D}$ is chosen randomly where each point is drawn independently according to the probability measure μ. Further assume that*

$$\frac{k}{\log k} \geq c_1 K^2 s(\log s)^2 \log N \tag{4.1}$$

$$k \geq c_2 K^2 s \log\left(\frac{1}{\varepsilon}\right), \tag{4.2}$$

where $c_1, c_2 > 0$ are universal constants and $s \in [N]$. Then, with probability at least $1-\varepsilon$, $\varepsilon \in (0, 1)$, every $s-$sparse vector z is the unique solution to the ℓ_1-minimization problem (3.3), with $A = M(\phi, W)$ and $y = g(W)$, where $g(u) = \sum_{j=1}^N z_j\phi_j(u)$.

This result is also true for complex valued matrices [29] but we are only interested in the real case. The theorem is proved by showing that $M(\phi, W)$ satisfies the RIP (Definition 3.1.2). It is worth noting that an optimal result is obtained if one sets $\varepsilon = e^{-\frac{k}{c_2K^2s}}$ in the sense that (4.2) is satisfied as equality. Also, from (4.1) and using $\log s \geq 1$, we obtain $k \geq (\log k)c_1K^2s(\log s)^2 \log N$, so $1 - e^{-\frac{k}{c_2K^2s}} \geq 1 - N^{-\gamma \log k}$, for the universal constant $\gamma = c_1/c_2$. Thus, ε can be set such that the probability of success $1 - \varepsilon$ satisfies

$$1 - \varepsilon \geq 1 - N^{-\gamma \log k}, \tag{4.3}$$

showing that this probability of success grows with N and k.

As it will be seen later in this chapter, we will not be able to use $y = g(W)$ but a noisy version $y = g(W) + \epsilon$, with a known bound on the size of ϵ. In order to extend this approach to handle noise, some modifications of problem (3.3) are needed. Since we are trying to find a sparse solution z to the problem $Az - y = \epsilon$ with ϵ unknown, it is natural to consider, instead of the formulation (3.3), the following optimization problem:

$$\min \|z\|_1 \quad \text{s. t.} \quad \|Az - y\|_2 \leq \eta, \tag{4.4}$$

where η is a positive number. We now present a recovery result, analogous to Theorem 4.1.1, based on the formulation (4.4) and thus appropriate to the noisy case. The proof is available in [29].

Theorem 4.1.2 (Rauhut [29]) *Under the same assumptions of Theorem 4.1.1, with probability at least $1 - \varepsilon$, $\varepsilon \in (0, 1)$, the following holds for every $s-$sparse vector z:*

Let noisy samples $y = M(\phi, W)z + \epsilon$ with

$$\|\epsilon\|_2 \leq \eta$$

be given, for any positive η, and let z^* be the solution of the ℓ_1-minimization problem (4.4) with $A = M(\phi, W)$ and $y = g(W)$, where $g(u) = \sum_{j=1}^{N} z_j \phi_j(u)$. Then,

$$\|z - z^*\|_2 \leq \frac{d}{\sqrt{k}} \eta$$

for some universal constant $d > 0$.

4.2. Sparse recovery using polynomial orthonormal expansions

As mentioned in Chapter 2, we will be interested in recovering first and second order information about an objective function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ in the form of a *local* quadratic model near a point x_0. Therefore we will be interested in the space of quadratic functions *defined* in $B(x_0; \Delta)$. The purpose of this section is to build orthonormal bases for the space of quadratic functions in $B(x_0; \Delta)$, appropriate for the application of Theorem 4.1.2. Since the sparsity of f will appear on its Hessian, one needs an orthonormal basis such that this sparsity is carried to the basis. Thus, such a basis should include multiples of the polynomials $(u_i - x_0)(u_j - x_0)$ which, in turn, should not *appear* in other elements of the basis (later we will set $x_0 = 0$ without

lost of generality). The orthonormal basis should also satisfy the K-boundedness condition for a K not depending on the dimension of the domain of the function, since otherwise the results from Theorems 4.1.1 and 4.1.2 would be much weaker. (Recently, progress has been made to deal with this problem when K grows with the dimension, where the main idea is to pre-condition the interpolation matrix, see [30]). Moreover, the orthonormal basis should preferably be *structured* in a way that a fast basis change algorithm is available to change it to $\bar{\phi}$ in (2.1), to accommodate the case where one is interested in performing this operation relatively fast.

We will start by building such an orthonormal basis for the ball in the infinity norm centered at the origin, $\mathcal{D} = B_\infty(0;\Delta) = [-\Delta,\Delta]^n$, considering there the uniform measure μ normalized as a probability measure (i.e., $\mu\left([-\Delta,\Delta]^n\right) = 1$).

4.2.1. Orthogonal expansions on hypercubes

Let μ be the uniform probability measure on $B_\infty(0;\Delta)$. Note that due to the features of $B_\infty(0;\Delta) = [-\Delta,\Delta]^n$, one has

$$\int_{[-\Delta,\Delta]^n} g(u_i)h(u_1,...,u_{i-1},u_{i+1},...,u_n)du = \tag{4.5}$$

$$= \int_{-\Delta}^{\Delta} g(u_i)du_i \int_{[-\Delta,\Delta]^{n-1}} h(u_1,...,u_{i-1},u_{i+1},...,u_n)du_1\cdots du_{i-1}du_{i+1}\cdots du_n,$$

for appropriate integrable functions g and h.

We want to find an orthonormal basis, with respect to μ, of the second degree polynomials on $B_\infty(0;\Delta)$ that contains multiples of the polynomials $\{u_iu_j\}_{i\neq j}$. Note that we are considering first the non-diagonal part of the Hessian since it has *almost all* the entries, and, as we will see later, this approach will not jeopardize the possible sparsity in the diagonal. We thus start our process of finding the desirable basis by considering the $n(n-1)/2$ polynomials $\{k_2u_iu_j\}_{i\neq j}$, where k_2 is a normalizing constant. Now, note that from (4.5), for different indices i,j,l,

$$\int_{B_\infty(0;\Delta)} u_iu_ju_ld\mu = \int_{B_\infty(0;\Delta)} u_iu_jd\mu = \int_{B_\infty(0;\Delta)} u_iu_j^2d\mu = 0.$$

As a result, we can add to the set $\{k_2u_iu_j\}_{i\neq j}$ the polynomials $\{k_1u_i\}_{1\leq i\leq n}$ and k_0, where k_1 and k_0 are normalizing constants, forming a set of $n(n-1)/2+(n+1)$ orthogonal polynomials.

It remains to find n quadratic polynomials, which will be written in the form $k_3(u_i^2-\alpha_1u_i-\alpha_0)$. We will choose the constants α_0 and α_1 such that these polynomials are orthogonal to the remaining ones. As we need orthogonality with respect

to a multiple of the polynomial u_i,

$$\int_{B_\infty(0;\Delta)} u_i(u_i^2 - \alpha_1 u_i - \alpha_0)d\mu = 0,$$

we must have $\alpha_1 = 0$. Then, orthogonality with respect to a multiple of the polynomial 1 means

$$\int_{B_\infty(0;\Delta)} u_i^2 - \alpha_0 d\mu = 0.$$

Thus,

$$\alpha_0 = \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} u^2 du = \frac{1}{2\Delta}\left(\frac{2}{3}\Delta^3\right) = \frac{1}{3}\Delta^2.$$

Finally, let us calculate the normalization constants. From

$$\int_{B_\infty(0;\Delta)} k_0^2 d\mu = 1$$

we set $k_0 = 1$. From the equivalent statements

$$\begin{aligned}
\int_{B_\infty(0;\Delta)} (k_1 u_i)^2 \, d\mu &= 1, \\
\frac{k_1^2}{|[-\Delta,\Delta]^n|}\int_{B_\infty(0;\Delta)} u_i^2 du &= 1, \\
\frac{k_1^2}{(2\Delta)^n}\int_{-\Delta}^{\Delta} u^2 du \int_{[-\Delta,\Delta]^{n-1}} 1 du &= 1, \\
k_1^2 \int_{-\Delta}^{\Delta} u^2 \frac{du}{2\Delta} &= 1,
\end{aligned}$$

we obtain $k_1 = \sqrt{3}/\Delta$. From the equivalent statements

$$\begin{aligned}
\int_{B_\infty(0;\Delta)} (k_2 u_i u_j)^2 \, d\mu &= 1, \\
k_2^2 \left(\int_{-\Delta}^{\Delta} u^2 \frac{du}{2\Delta}\right)^2 &= 1,
\end{aligned}$$

we conclude that $k_2 = 3/\Delta^2$. And from the equivalent statements

$$\begin{aligned}
\int_{B_\infty(0;\Delta)} \left(k_3\left(u_i^2 - \frac{1}{3}\Delta^2\right)\right)^2 d\mu &= 1, \\
k_3^2 \int_{-\Delta}^{\Delta} \left(u^2 - \frac{1}{3}\Delta^2\right)^2 \frac{1}{2\Delta} du &= 1,
\end{aligned}$$

we obtain

$$k_3 = \frac{3\sqrt{5}}{2}\frac{1}{\Delta^2}$$

We have thus arrived to the following orthonormal basis.

Definition 4.2.1 *We define the basis ψ as the following $(n+1)(n+2)/2$ polynomials:*

$$\begin{cases} \psi_0(u) &= 1 \\ \psi_{1,i}(u) &= \frac{\sqrt{3}}{\Delta}u_i \\ \psi_{2,ij}(u) &= \frac{3}{\Delta^2}u_iu_j \qquad (\textit{for } i \neq j) \\ \psi_{2,i}(u) &= \frac{3\sqrt{5}}{2}\frac{1}{\Delta^2}u_i^2 - \frac{\sqrt{5}}{2}, \end{cases} \tag{4.6}$$

for $i = 1, ..., n$ and $j = 1, ..., n$. We will now slightly abuse of notation and represent any of the 'indices' $(0), (1, i)$, $(2, ij)$ or $(2, i)$ by the letter ι. We will also consider ψ_L the subset of ψ consisting of the polynomials with degree 0 or 1 and ψ_Q the ones with degree 2, as we did in Chapter 2 for $\bar{\phi}$. The basis ψ satisfies the assumptions of Theorems 4.1.1 and 4.1.2, as stated in the following theorem.

Theorem 4.2.1 *The basis ψ (see Definition 4.2.1) is orthonormal and satisfies the K-boundedness condition (see Definition 4.1.1) in $\mathcal{D} = B_\infty(0; \Delta)$ for the uniform probability measure with $K = 3$.*

Proof. From the above derivation and (4.5) one can easily show that ψ is orthonormal in $\mathcal{D} = B_\infty(0; \Delta)$ with respect to the uniform probability measure. So, the only thing left to prove is the boundedness condition with $K = 3$. In fact, it is easy to check that

$$\begin{cases} \|\psi_0\|_{L^\infty(B_\infty(0;\Delta))} &= 1 &\leq 3 \\ \|\psi_{1,i}\|_{L^\infty(B_\infty(0;\Delta))} &= \sqrt{3} &\leq 3 \\ \|\psi_{2,ij}\|_{L^\infty(B_\infty(0;\Delta))} &= 3 &\leq 3 \\ \|\psi_{2,i}\|_{L^\infty(B_\infty(0;\Delta))} &= \sqrt{5} &\leq 3. \end{cases} \tag{4.7}$$

■

We will later be interested in quadratic functions $g = \sum_\iota \alpha_\iota \psi_\iota$ (see Definition 4.2.1) which are $s-$sparse in the coefficients that correspond to the polynomials in ψ_Q, meaning that only s coefficients with respect to this polynomials are non-zero, with s a number between 1 and $n(n+1)/2$. In such cases, the correspondent vector α of coefficients is $(s+n+1)-$sparse. We now state a corollary of Theorem 4.1.2 for sparse recovery in the orthonormal basis ψ which will then be used in the next section. Note that we will write the probability of success in the form $1 - n^{-\gamma \log k}$ which can be derived from (4.3) using $N = \mathcal{O}(n^2)$ and a simple modification of the universal constant γ.

Corollary 4.2.1 *Let $M(\psi, W) \in \mathbb{R}^{k \times N}$ be the matrix of entries $[M(\psi, W)]_{ij} = \psi_j(w^i)$, $i = 1, ..., k$, $j = 1, ..., N$, with $N = (n+1)(n+2)/2$.*

Assume that the sample set $W = \{w^1, ..., w^k\} \subset B_\infty(0; \Delta)$ *is chosen randomly where each point is drawn independently according to the probability uniform measure* μ *in* $B_\infty(0; \Delta)$. *Further assume that*

$$\frac{k}{\log k} \geq 9c\,(s+n+1)\,(\log\,(s+n+1))^2 \log\left(\frac{(n+1)(n+2)}{2}\right),$$

for some universal constant $c > 0$ *and* $s \in \{1, ..., n(n+1)/2\}$. *Then, with probability at least* $1 - n^{-\gamma \log k}$, *for some universal constant* $\gamma > 0$, *the following holds for every vector* z, *having at most* $s + n + 1$ *non-zero expansion coefficients in the basis* ψ*:*

Let noisy samples $y = M(\psi, W)z + \epsilon$ *with*

$$\|\epsilon\|_2 \leq \eta$$

be given (for any positive η*) and let* z^* *be the solution of the* ℓ_1*-minimization problem (4.4) with* $A = M(\psi, W)$. *Then,*

$$\|z - z^*\|_2 \leq \frac{d}{\sqrt{k}}\,\eta$$

for some universal constant $d > 0$.

4.2.2. Orthogonal expansions on Euclidian balls

It would also be natural to consider the ball $B(0; \Delta)$ in the classical ℓ_2-norm, but as we will explain now it does not work as well as using an hypercube, i.e., the ℓ_∞-norm. The first step one would have to take is to select a probability measure. A possible choice would be the uniform measure, especially since the radial measure brings difficulties due to the singularity at the origin. However, one problem with the uniform measure in the ℓ_2-ball is that we no longer have a formula like (4.5), which was heavily used for achieving the orthogonality conditions on the hypercube. It would be possible to construct an orthogonal basis, using, e.g., Gram-Schmidt, but it would most likely lead to some of the following 3 problems:

- The fact that the Hessian is sparse might not necessarily imply sparsity of the expansion coefficients in that basis.
- The basis may be too *unstructured* to allow a fast basis change algorithm in case we are interested in changing to another basis like $\bar{\phi}$ in (2.1).
- The constant K in the K-boundedness condition might grow with n.

Due to these difficulties we will restrict ourselves to the hypercube, $B_\infty(0; \Delta)$.

4.3. Recovery of functions with sparse Hessian using quadratic models

We are interested in recovering first and second order information of a twice continuously differentiable objective function $f : D \to \mathbb{R}$ near a point x_0, with sparse Hessian at the point x_0. First of all, we will need to formalize the sparsity assumption to be made on f. From the motivation given in Chapter 2, sparsity is only considered in the Hessian.

Assumption 4.3.1 (Hessian sparsity) *Assume that $f : D \to \mathbb{R}$ satisfies Assumption 2.1.2 and furthermore that for every $x_0 \in D$ the Hessian $\nabla^2 f(x_0)$ of f at x_0 has at most s non-zero entries, where s is a number between 1 and $n(n+1)/2$. If this is the case, then f is said to satisfy Hessian sparsity of order s.*

We are essentially targeting at functions where the corresponding Hessian matrices are sparse in the non-diagonal parts but our results are general enough to also consider sparsity in the diagonal Hessian components.

The information recovered will come in the form of quadratic models of f near x_0 that are fully quadratic models of f (see Definition 2.1.2), thus yielding the same accuracy of second order Taylor models. Given the result in Section 4.2, we will consider the norm $p = \infty$ in Definition 2.1.2, thus considering there balls of the form $B_\infty(x_0; \Delta)$.

To find a quadratic interpolating model in $B_\infty(x_0; \Delta)$ with the guarantee of being fully quadratic, the number of points required is $(n+1)(n+2)/2$, but evaluating the function at such a number of points can be too expensive. One is typically interested, in Derivative-Free Optimization, in constructing quadratic models with much less sample points. In particular, when f is known to have a sparse Hessian, this information could be used in our favor to significantly reduce the cardinality of the sample set.

Clearly, when f satisfies Hessian sparsity of order s, the second order Taylor expansion of f at x_0 is a quadratic function T with sparse Hessian (and a fully quadratic model of f), thus writing such T as a linear combination of the canonical basis $\bar{\phi}$ for the quadratic functions in $\mathbb{R}^n$ (see (2.1)) yields a sparse representation in $\bar{\phi}$. Derivatives are not available for recovery in Derivative-Free Optimization, but this argument still suggests to use the canonical basis $\bar{\phi}$. However, according to the derivations of Sections 4.1 and 4.2, the basis $\bar{\phi}$ seems not to be a suitable one

for theoretical sparse recovery because it is not orthogonal in an appropriate form. Alternatively, we will consider the orthogonal basis ψ of Definition 4.2.1.

Although the basis ψ is different from the canonical one, it will serve us well as it keeps many properties of the latter one, and can be obtained from it through a few simple transformations. In particular, the sparsity in the Hessian of a quadratic model q will be carried over to sparsity in the representation of q in ψ, since, due to the particular structure of ψ, the expansion coefficients in ψ_Q will be multiples of the ones in $\bar{\phi}_Q$, thus guaranteeing that if the coefficients in the latter are $s-$sparse, so are the ones in the former.

We are now able to use the material developed in Section 4.2 to guarantee the construction, for each x_0 and Δ, and with high probability, of a fully quadratic model of f in $B_\infty(x_0;\Delta)$ using a random sample set of only $\mathcal{O}(n(\log n)^4)$ points, instead of the classical $\mathcal{O}(n^2)$ points. In fact, to find such a fully quadratic model $q(u) = \sum_\iota \alpha_\iota^q \psi_\iota(u)$ (if $x_0 = 0$), one can use problem (4.4) with $z = \alpha^q$, $A = M(\psi, W)$, and $y = f(W)$, written now in the form

$$\begin{aligned} \min \quad & \|\alpha^q\|_1 \\ \text{s.t.} \quad & \|M(\psi, W)\alpha^q - f(W)\|_2 \leq \eta \end{aligned} \tag{4.8}$$

where η is some appropriate positive quantity. Corollary 4.2.1 can then be used to ensure, see (4.15) below, that only $\mathcal{O}(n(\log n)^4)$ points are necessary for recovery around $x_0 = 0$, when the number s of non-zero components of the Hessian of f at $x_0 = 0$ is of the order of n.

Note that we are in fact considering 'noisy' measurements, due to the limitations of only being able to evaluate the function f and to recover quadratics. We will say that a function q^* is the solution to the minimization problem (4.8) if $q^*(u) = \sum_\iota \alpha^* \psi_\iota(u)$, where α^* is the minimizer of (4.8).

Since we are interested in finding a fully quadratic model q for f around x_0, we could consider a translation to reduce the problem to the previous one, writing instead

$$\begin{aligned} \min \quad & \|\alpha^q\|_1 \\ \text{s.t.} \quad & \|M(\psi, W - x_0)\alpha^q - f(W - x_0)\|_2 \leq \eta \end{aligned} \tag{4.9}$$

Thus, without loss of generality, we can consider $x_0 = 0$ and work with the ℓ_1-minimization problem (4.8).

We are finally ready to present our main result.

Theorem 4.3.1 *Let $f : D \to \mathbb{R}$ satisfy Hessian sparsity of order s (see Assumption 4.3.1). For any $\Delta \in (0, \Delta_{\max}]$ and given k random points, $W = \{w^1, ..., w^k\}$, chosen jointly independent with respect to the uniform measure on $B_\infty(0; \infty)$, with*

$$\frac{k}{\log k} \geq 9c\,(s+n+1)\log^2(s+n+1)\log N, \tag{4.10}$$

for some universal constant $c > 0$, then, with probability larger than $1 - n^{-\gamma \log k}$, for some universal constant $\gamma > 0$, the solution q^* to the ℓ_1-minimization problem (4.8), for some η depending on Δ^3 and on the Lipschitz constants of f, is a fully quadratic model of f (see Definition 2.1.2) on $B_\infty(0; \Delta)$ with $\nu_2^m = 0$ and constants κ_{ef}, κ_{eg}, and κ_{eh} not depending on Δ.

Corollary 4.2.1 will provide an L^2 estimate of the difference between the quadratic that one is trying to recover and the one recovered. As one wants L^∞ estimations between the quadratics values, their gradients and their Hessians to satisfy the requirements of fully quadratic models, one needs first to bound these norms by the L^2 norm in the space of quadratic functions in $B_\infty(0; \Delta)$. So, the next lemma will be needed before proving the theorem.

Lemma 4.3.1 *Let q be a quadratic function. Then*

$$|q(u)| \leq \left(3\sqrt{\frac{(n+1)(n+2)}{2}}\right)\|q\|_{L^2(B_\infty(0;\Delta),\mu)} \tag{4.11}$$

$$\|\nabla q(u)\|_2 \leq \left(3\sqrt{5}\sqrt{n+1}\sqrt{n}\right)\frac{1}{\Delta}\|q\|_{L^2(B_\infty(0;\Delta),\mu)} \tag{4.12}$$

$$\left\|\nabla^2 q(u)\right\|_2 \leq \left(3\sqrt{5}\,n\right)\frac{1}{\Delta^2}\|q\|_{L^2(B_\infty(0;\Delta),\mu)} \tag{4.13}$$

for all $u \in B_\infty(0; \Delta)$, where $\|q\|_{L^2(B_\infty(0;\Delta),\mu)} = \left(\int_{B_\infty(0;\Delta)} |q(u)|^2 \frac{du}{(2\Delta)^n}\right)^{1/2}$.

Proof. Recall the orthonormal basis ψ given in (4.6). Let α be a vector in $\mathbb{R}^{(n+1)(n+2)/2}$ such that

$$q(u) = \sum_\iota \alpha_\iota \psi_\iota(u)$$

(see the notation introduced in Definition 4.2.1). Since ψ is orthonormal (with respect to μ), we have that $\|\alpha\|_2 = \|q\|_{L^2(B_\infty(0;\Delta),\mu)}$.

Hence, from (4.7),

$$\|q\|_{L^\infty(B_\infty(0;\Delta))} \leq \sum_\iota |\alpha_\iota| \|\psi_\iota\|_{L^\infty(B_\infty(0;\Delta))} \leq 3\|\alpha\|_1 \leq 3\sqrt{\frac{(n+1)(n+2)}{2}}\|\alpha\|_2.$$

So, $\|q\|_{L^\infty(B_\infty(0;\Delta))} \leq 3\sqrt{(n+1)(n+2)/2}\,\|q\|_{L^2(B_\infty(0;\Delta),\mu)}$, which establishes (4.11).

Also, from (4.6),

$$\begin{aligned}
\left\|\frac{\partial q}{\partial u_i}\right\|_{L^\infty(B_\infty(0;\Delta))} &\leq \sum_\iota |\alpha_\iota| \left\|\frac{\partial \psi_\iota}{\partial u_i}\right\|_{L^\infty(B_\infty(0;\Delta))} \\
&= |\alpha_{1,i}| \left\|\frac{\sqrt{3}}{\Delta}\right\|_{L^\infty(B_\infty(0;\Delta))} + \sum_{j\in[n]\setminus\{i\}} |\alpha_{2,ij}| \left\|\frac{3}{\Delta^2}u_j\right\|_{L^\infty(B_\infty(0;\Delta))} \\
&\qquad + |\alpha_{2,i}| \left\|\frac{3\sqrt{5}}{\Delta^2}u_i\right\|_{L^\infty(B_\infty(0;\Delta))} \\
&= \frac{\sqrt{3}}{\Delta}|\alpha_{1,i}| + \sum_{j\in[n]\setminus\{i\}} \frac{3}{\Delta}|\alpha_{2,ij}| + \frac{3\sqrt{5}}{\Delta}|\alpha_{2,i}| \\
&\leq \frac{3\sqrt{5}}{\Delta} \sum_{\iota\in G_i} |\alpha_\iota|,
\end{aligned}$$

where G_i is the set of indexes $(1,i), (2,i)$, and $(2,ij)$ for every $j \in [n] \setminus \{i\}$, with $|G_i| = n+1$. Then, by the known relations between the norms ℓ_1 and ℓ_2,

$$\begin{aligned}
\frac{3\sqrt{5}}{\Delta} \sum_{\iota\in G_i} |\alpha_\iota| &\leq \frac{3\sqrt{5}}{\Delta}\sqrt{n+1}\sqrt{\sum_{\iota\in G_i} |\alpha_\iota|^2} \\
&\leq \frac{3\sqrt{5}}{\Delta}\sqrt{n+1}\|\alpha\|_2.
\end{aligned}$$

Since the gradient has n components one has

$$\|\nabla q(u)\|_2 \leq \sqrt{n}\left(3\sqrt{5}\sqrt{n+1}\right)\frac{1}{\Delta}\|q\|_{L^2(B_\infty(0;\Delta),\mu)}$$

for all $u \in B_\infty(0;\Delta)$, showing (4.12).

For the estimation of the Hessian, we need to separate the diagonal from the non-diagonal part. For the non-diagonal part, with $i \neq j$,

$$\begin{array}{lllll}
\left\|\frac{\partial^2 q}{\partial u_i \partial u_j}\right\|_{L^\infty(B_\infty(0;\Delta))} & \leq & \sum_\iota |\alpha_\iota| \left\|\frac{\partial^2 \psi_\iota}{\partial u_i \partial u_j}\right\|_{L^\infty(B_\infty(0;\Delta))} & = & |\alpha_{2,ij}| \left\|\frac{3}{\Delta^2}\right\|_{L^\infty(B_\infty(0;\Delta))} \\
& = & |\alpha_{2,ij}|\frac{3}{\Delta^2} & \leq & \frac{3}{\Delta^2}\|\alpha\|_2.
\end{array}$$

For the diagonal part, with $i \in [n]$,

$$\begin{array}{lllll}
\left\|\frac{\partial^2 q}{\partial u_i^2}\right\|_{L^\infty(B_\infty(0;\Delta))} & \leq & \sum_\iota |\alpha_\iota| \left\|\frac{\partial^2 \psi_\iota}{\partial u_i^2}\right\|_{L^\infty(B_\infty(0;\Delta))} & = & |\alpha_{2,i}| \left\|\frac{3\sqrt{5}}{\Delta^2}\right\|_{L^\infty(B_\infty(0;\Delta))} \\
& = & |\alpha_{2,i}|\frac{3\sqrt{5}}{\Delta^2} & \leq & \frac{3\sqrt{5}}{\Delta^2}\|\alpha\|_2.
\end{array}$$

Since the Hessian has n^2 components one has

$$\left\|\nabla^2 q(u)\right\|_2 \leq \left\|\nabla^2 q(u)\right\|_F \leq n\left(3\sqrt{5}\right)\left(\frac{1}{\Delta}\right)^2 \|q\|_{L^2(B_\infty(0;\Delta),\mu)}$$

for all $u \in B_\infty(0;\Delta)$, which proves (4.13). ■

Remark 4.3.1 *Although the estimates in Lemma 4.3.1 are possibly not sharp, the dependency of the error bounds on n cannot be eliminated. In fact, the function*

$$\chi(u) = \sum_{i,j\in[n],\, i\neq j} \sqrt{\frac{2}{n(n-1)}}\frac{3}{\Delta^2}u_i u_j$$

satisfies $\|\chi\|_{L^2(B_\infty(0;\Delta),\mu)} = 1$ *and* $\chi(\Delta,...,\Delta) = \frac{3}{\sqrt{2}}\sqrt{n(n-1)}$.

Proof. (of Theorem 4.3.1) Let T be the second order Taylor model of f centered at 0,

$$T(u) = f(0) + \nabla f(0)^T u + \frac{1}{2}u^T\nabla^2 f(0)u.$$

Since f satisfies Assumption 4.3.1 we know, by Proposition 2.1.1, that, for any $\Delta \in (0, \Delta_{\max}]$, T is a fully quadratic model for f on $B_\infty(0;\Delta)$ with $\nu_2^m = 0$ and some constants κ'_{ef}, κ'_{eg}, and κ'_{eh}. Moreover, $\nabla^2 T(0) = \nabla^2 f(0)$, and so T is a quadratic function in $B_\infty(0;\Delta)$ whose Hessian has at most s non-zero entries. Let $W = \{w^1,...,w^k\}$ be a random sample set where each point is drawn independently according to the uniform probability measure in $B_\infty(0;\Delta)$ such that (4.10) is satisfied. The polynomial T satisfies the assumptions of Corollary 4.2.1 and, for the purpose of the proof, is the quadratic that will be approximately recovered. Now, since T is a fully quadratic model, we have $|f(w^i) - T(w^i)| \leq \kappa'_{ef}\Delta^3$. Therefore

$$\|f(W) - T(W)\|_2 \leq \sqrt{k}\,\kappa'_{ef}\Delta^3. \tag{4.14}$$

Note that one can only recover T approximately given that the values of $T(W) \simeq f(W)$ are 'noisy'.

Then, by Corollary 4.2.1, with probability larger than $1 - n^{-\gamma\log k}$, for a universal constant $\gamma > 0$, the solution q^* to the ℓ_1-minimization problem (4.8) with $\eta = \sqrt{k}\,\kappa'_{ef}\Delta^3$ satisfies

$$\|\alpha^* - \alpha^T\|_2 \leq d\,\kappa'_{ef}\Delta^3,$$

where α^* and α^T are the coefficients of q^* and T in the basis ψ given by (4.6), respectively. Since ψ is an orthonormal basis in $L^2(B_\infty(0;\Delta),\mu)$,

$$\|q^* - T\|_{L^2(B_\infty(0;\Delta),\mu)} = \|\alpha^* - \alpha^T\|_2 \leq d\,\kappa'_{ef}\Delta^3.$$

So, by Lemma 4.3.1,

$$\begin{aligned}
|q^*(u) - T(u)| &\leq d\left(3\sqrt{\frac{(n+1)(n+2)}{2}}\right)\kappa'_{ef}\Delta^3,\\
\|\nabla q^*(u) - \nabla T(u)\|_2 &\leq d\left(3\sqrt{5}\sqrt{n+1}\sqrt{n}\right)\kappa'_{ef}\Delta^2,\\
\left\|\nabla^2 q^*(u) - \nabla^2 T(u)\right\|_2 &\leq d\left(3\sqrt{5}\,n\right)\kappa'_{ef}\Delta,
\end{aligned}$$

for all $u \in B_\infty(0;\Delta)$. Therefore, using (4.14), more specifically $|f(w^i) - T(w^i)| \leq \kappa'_{ef}\Delta^3$, one has

$$\begin{aligned} |q^*(u) - f(u)| &\leq \left(d\left(3\sqrt{\frac{(n+1)(n+2)}{2}}\right)\kappa'_{ef} + \kappa'_{ef}\right)\Delta^3, \\ \|\nabla q^*(u) - \nabla f(u)\|_2 &\leq \left(d\left(3\sqrt{5}\sqrt{n+1}\sqrt{n}\right)\kappa'_{ef} + \kappa'_{eg}\right)\Delta^2, \\ \left\|\nabla^2 q^*(u) - \nabla^2 f(u)\right\|_2 &\leq \left(d\left(3\sqrt{5}\,n\right)\kappa'_{ef} + \kappa'_{eh}\right)\Delta, \end{aligned}$$

for all $u \in B_\infty(0;\Delta)$.

Since q^* is a quadratic function, its Hessian is Lipschitz continuous with Lipschitz constant 0, so one has that $\nu_2^m = 0$. Hence q^* is a fully quadratic model of f on $B_\infty(0;\Delta)$ as we wanted to prove. ■

Note that the result of Theorem 4.3.1 is obtained when the number k of sampling points satisfies (see (4.10) and recall that $N = \mathcal{O}(n^2)$)

$$\frac{k}{\log k} = \mathcal{O}(n(\log n)^2 \log n)$$

when $s = \mathcal{O}(n)$, i.e., when the number of non-zero elements of the Hessian of f at x_0, given by s, is of the order of n. Since $k < (n+1)(n+2)/2$, one obtains

$$k = \mathcal{O}\left(n(\log n)^4\right). \tag{4.15}$$

Another interesting fact about Theorem 4.3.1 is that it is established under no conditions on the sparsity pattern of the Hessian.

Since the sparsity of f is only in the Hessian and keeping in mind the formulation (2.5) and the discussion in Section 3.2, it is desirable to remove from the objective function in (4.8) or (4.9) the coefficients corresponding to ψ_L ($\psi_{1,i}$ for $i \in [n]$ and ψ_0). The version corresponding to (4.8) would then become

$$\begin{array}{ll} \min & \left\|\alpha_Q^q\right\|_1 \\ \text{s.t.} & \left\|M(\psi_L, W)\alpha_L^q + M(\psi_Q, W)\alpha_Q^q - f(W)\right\|_2 \leq \eta. \end{array} \tag{4.16}$$

Problem (4.8) and (4.16) can be solved in polynomial time (with Second-Order Cone Programming software [1]), however they are not linear programs because of their constraints. As one knows that the second order Taylor model T satisfies $\|T(W) - f(W)\|_\infty \leq \eta/\sqrt{k}$ (where $\eta = \sqrt{k}\,\kappa'_{ef}$), because T is fully quadratic for f, one could consider

$$\begin{array}{ll} \min & \left\|\alpha_Q^q\right\|_1 \\ \text{s.t.} & \left\|M(\psi_L, W)\alpha_L^q + M(\psi_Q, W)\alpha_Q^q - f(W)\right\|_\infty \leq \frac{1}{\sqrt{k}}\eta, \end{array}$$

instead of (4.16), which is equivalent to a linear program. In practice, as we have seen in Chapter 2, one imposes the interpolation constrains exactly which corresponds to setting $\eta = 0$ in the above formulations.

Finally, independently of the form of the ℓ_1-recovery problem used, of the slight discrepancy between the bases $\bar{\phi}$ and ψ, and of setting η to zero or not, one must understand that the result of Theorem 4.3.1 cannot strictly validate a practical setting in Derivative-Free Optimization (DFO) but rather provide motivation and insight on the use of ℓ_1-minimization to build underdetermined quadratic models for functions with sparse Hessians. In fact, not only most of the sampling is done deterministically in DFO (as will see in the next chapter) but, also, the constants in the bound (4.10) (and thus in (4.15)) render impractical. In fact, the best known upper bound (see [29]) for the universal constant c appearing in (4.10) is $c < 17190$, making (4.10) only applicable if n is much greater than the values for which DFO problems are tractable (which is of the order of a few dozens). However, such bound is, most likely, not sharp, and, in fact, similar universal constants appearing in the setting of Compressed Sensing are known to be much smaller.

Chapter 5

A practical interpolation-based trust-region method

5.1. Interpolation-based trust-region algorithms for DFO

Trust-region methods are a well known class of algorithms for the numerical solution of nonlinear programming problems [10, 25]. In this section we will give a brief summary of these methods when applied to the unconstrained minimization of a smooth function $f : D = \mathbb{R}^n \to \mathbb{R}$,

$$\min_{x \in \mathbb{R}^n} \; f(x), \tag{5.1}$$

without using the derivatives of the objective function f.

At each iteration k, these methods build a model $m(x_k + s)$ of the objective function around the current iterate x_k and assess how well such model approximates the function in a *trust region* of the form $B_p(x_k; \Delta_k)$, typically with $p = 2$, where Δ_k is the so-called trust-region radius. For this purpose, one has to first determine a step s_k from the solution of the trust-region subproblem

$$\min_{s \in B_2(0;\Delta_k)} \; m_k(x_k + s). \tag{5.2}$$

Then, one compares the actual reduction in the objective function ($ared_k = f(x_k) - f(x_k + s_k)$) to the predicted reduction in the model ($pred_k = m_k(x_k) - m_k(x_k + s_k)$). If the comparison is good ($\rho_k = ared_k/pred_k \geq \eta_1 \in (0, 1)$), then one takes the step and (possibly) increases the trust-region radius (successful iterations). If the comparison is bad ($\rho_k < \eta_0 \in [0, \eta_1]$), then one rejects the step and decreases the trust-region radius (unsuccessful iterations). New iterates might be only accepted based on a sufficient decrease condition of the form $\rho_k \geq \eta_1$, in which case one sets $\eta_0 = \eta_1 \in (0, 1)$. In the setting of Derivative-Free Optimization (DFO) one is interested in accepting new iterates yielding a weaker decrease (so that function evaluations are not unnecessarily wasted), such as a simple decrease $\rho_k \geq \eta_0 = 0$.

Then, one needs to consider an intermediate case (acceptable iterations) of the form $\eta_1 > \rho_k \geq \eta_0$, where the step is accepted and the trust-region radius is decreased.

The global convergence properties of these methods are strongly dependent from the requirement that, as the trust region becomes smaller, the model becomes more accurate, implying in particular that the trust-region radius is bounded away from zero, away from stationarity. Taylor based-models, when derivatives are known, naturally satisfy this requirement. However, in the DFO setting, some provision has to be taken in the model and sample set management to ensure global convergence.

One provision (which have been shown recently by Scheinberg and Toint [31] to be in fact necessary) is the inclusion of the so-called criticality (or stationarity) step, originally introduced by Conn, Scheinberg, and Toint [11]. Essentially, this step (see [13, Section 10.3]) ensures that when a measure of model stationarity is small, the model and the trust region must be changed so that the model becomes fully linear/quadratic in a trust region where the radius is of the order of the measure of model stationarity — and thus guaranteeing that model stationarity happens due to true function stationarity rather than due to badly poised sample sets.

The other provision is the inclusion of model-improvement iterations (see [13, Section 10.3]) when $\eta_1 > \rho_k$ and one cannot certify that the model is fully linear/quadratic. In this case, the trust-region radius is not reduced and one calls a model-improving algorithm to improve the well poisedness of the sample set. Model-improving iterations ensure that acceptable and unsuccessful iterations, where the trust-region radius is decreased, do not occur without model accuracy.

Considering these adaptations, Conn, Scheinberg, and Vicente [12] proved global convergence for first or second order stationary points depending on the use of fully linear or fully quadratic models, making the theory also valid under the more difficult case where acceptance of new iterates is based on simple decrease ($\eta_0 = 0$). It is also shown in [12] that (due to the inclusion of the criticality step) the trust-region radius converges to zero. Scheinberg and Toint [31] have recently shown global convergence to first order stationary points for their *self-correcting geometry* approach which replaces model-improving iterations by an appropriate update of the sample set using only the new trust-region iterates.

5.2. A practical interpolation-based trust-region method

We now introduce a more practical algorithm following some of the basic ideas of the approach introduced by Fasano, Morales, and Nocedal [19], which have also inspired the authors in [31]. The main idea in [19] is to ignore poisedness as much as possible, updating the sample set in successful iterations by including in it the new trust-region iterate and removing from it an appropriate point. However, unlike [19], we discard the sample point farthest away from the new iterate (rather than the sample point farthest away from the current iterate).

In our approach we allow the algorithm to start with less points than those needed to build a determined quadratic model. Whenever there are less points than $p_{\max} = (n+1)(n+2)/2$, we use minimum Frobenius or ℓ_1 norm interpolation to build our models. This poses additional issues to those considered in [19], where $p_{\max}$ points are always used. For instance, until the cardinality of the sample set reaches $p_{\max}$, we never discard points from the sample set and always add new trial points independently of whether or not they are accepted as new iterates, in an attempt to be as greedy as possible when taking advantage of function evaluations.

Another difference from [19] is that we discard points that are too far from the current iterate when the trust-region radius becomes small (this is a kind of weak criticality condition), hoping that the next iterations will refill the sample set resulting in a similar effect as a criticality step. Thus, the cardinality of our sample set might fall below $p_{\min} = n+1$, the number required to build fully linear models in general. In such situations, we never reduce the trust-region radius.

Algorithm 5.2.1 (A practical DFO trust-region algorithm)
Step 0: Initialization.

Initial values. *Select values for the constants* $\epsilon_g (= 10^{-5}) > 0$, $\delta (= 10^{-5}) > 0$, $0 < \eta_0 (= 10^{-3}) < \eta_1 (= 0.25) < 1$, $\eta_2 (= 0.75) > \eta_1$, *and* $0 < \gamma_1 (= 0.5) < 1 < \gamma_2 (= 2)$. *Set* $p_{\min} = n+1$ *and* $p_{\max} = (n+1)(n+2)/2$. *Set the initial trust radius* $\Delta_0 (= 1) > 0$.

Initial sample set. *Let the starting point* x_0 *be given. Select as an initial sample set* $Y_0 = \{x_0, x_0 \pm \Delta_0 e_i, i = 1, \ldots, n\}$, *where the* e_i*'s are the columns of the identity matrix or order* n.

Function evaluations. *Evaluate the objective function at all* $y \in Y_0$.

Set $k = 0$.

Step 1: Model building.

Form a quadratic model $m_k(x_k + s)$ *of the objective function from* Y_k. *If* $|Y_k| = p_{\max}$, *use determined quadratic interpolation. If* $|Y_k| < p_{\max}$, *use minimum Frobenius* ($p = 2$), *or minimum* ℓ_1 ($p = 1$), *norm quadratic interpolation, by solving the problem*

$$\begin{aligned} \min \quad & \tfrac{1}{p}\|\alpha_Q\|_p^p \\ \text{s.t.} \quad & M(\bar{\phi}_L, Y_k)\alpha_L + M(\bar{\phi}_Q, Y_k)\alpha_Q = f(Y_k), \end{aligned} \tag{5.3}$$

where α_Q *and* α_L *are, respectively, the coefficients of order 2 and order less than 2 of the model.*

Step 2: Stopping criteria.

Stop if $\|g_k\| \leq \epsilon_g$ *or* $\Delta_k \leq \delta$.

Step 3: Step calculation.

Compute a step s_k *by solving (approximately) the trust-region subproblem (5.2).*

Step 4: Function evaluation.

Evaluate the objective function at $x_k + s_k$.

Step 5: Selection of the next iterate and trust radius update.

If $\rho_k < \eta_0$, *reject the trial step, set* $x_{k+1} = x_k$, *and reduce the trust-region radius, if* $|Y_k| \geq p_{\min}$, *by setting* $\Delta_k = \gamma_1 \Delta_k$ (**unsuccessful iteration**).

If $\rho_k \geq \eta_0$, *accept the trial step* $x_{k+1} = x_k + s_k$ (**successful and acceptable iterations**).
(Possibly decrease trust-region radius, $\Delta_k = \gamma_1 \Delta_k$, *if* $\rho_k < \eta_1$ *and* $|Y_k| \geq p_{\min}$.*)*

Increase the trust-region radius, $\Delta_{k+1} = \gamma_2 \Delta_k$, *if* $\rho_k > \eta_2$.

Step 6: Update the sample set.

If $|Y_k| = p_{\max}$, *set* $y_k^{out} \in argmax\|y - x_{k+1}\|_2$ *(break ties arbitrarily).*

If the iteration was successful:

If $|Y_k| = p_{\max}$, $Y_{k+1} = Y_k \cup \{x_{k+1}\} \setminus \{y_k^{out}\}$.

If $|Y_k| < p_{\max}$, $Y_{k+1} = Y_k \cup \{x_{k+1}\}$.

If the iteration was unsuccessful:

If $|Y_k| = p_{\max}$, $Y_{k+1} = Y_k \cup \{x_k + s_k\} \setminus \{y_k^{out}\}$ *if* $\|(x_k + s_k) - x_k\|_2 \leq \|y_k^{out} - x_k\|_2$.

If $|Y_k| < p_{\max}$, $Y_{k+1} = Y_k \cup \{x_k + s_k\}$.

Step 7: Model improvement.

When $\Delta_{k+1} < 10^{-3}$, *discard from* Y_{k+1} *all the points outside* $B(x_{k+1}; r\Delta_{k+1})$, *where* r *is chosen as the smallest number in* $\{100, 200, 400, 800, ...\}$ *for which at least three sample points from* Y_{k+1} *are contained in* $B(x_{k+1}; r\Delta_{k+1})$.

Increment k *by* 1 *and return to Step 1.*

5.3. Numerical results

In this section we will describe some of the numerical experiments which have been conducted to test the performance of Algorithm 5.2.1 implemented in MATLAB. We were particularly interested in testing two variants of Algorithm 5.2.1 defined by the norm used to compute the model in (5.3). The first variant makes use of the ℓ_2-norm and leads to minimum Frobenius norm models. As we have seen in Chapter 2, the solution of (5.3) with $p = 2$ is equivalent to the solution of a linear system of the form (2.4) with $W = Y_k$. We solved this system using SVD, regularizing extremely small singular values after the decomposition and before performing the backward solves, in an attempt to remediate extreme ill conditioning caused by nearly poised sample sets. The second approach consisted in using $p = 1$, leading to minimum ℓ_1-norm models and attempting to recover sparsity in the Hessian of the objective function. To solve problem (5.3) with $p = 1$ we first converted it to an equivalent linear program of the form

$$\begin{aligned} \min \quad & \sum_{i=1}^{n(n+1)/2} \left(\alpha_Q^+\right)_i + \left(\alpha_Q^-\right)_i \\ \text{s.t.} \quad & M(\bar{\phi}_L, Y_k)\alpha_L + M(\bar{\phi}_Q, Y_k)\left(\alpha_Q^+ - \alpha_Q^-\right) = f(Y_k) \\ & \alpha_Q^+, \alpha_Q^- \geq 0, \end{aligned} \tag{5.4}$$

where α_Q^- and α_Q^+ correspond, respectively, to the negative and the positive part of α_Q (both problems are equivalent because the optimal solution of (5.4) will satisfy,

for every i, $(\alpha_Q^-)_i = 0$ or $(\alpha_Q^+)_i = 0$). In both cases, $p = 1, 2$, we first scaled the corresponding problems by shifting the sample set to the origin (i.e., translating all the sample points such that the current iterate coincides with the origin) and then scaling the points so that they lie in $B_2(0; 1)$ with at least one scaled point at the border of this ball. This procedure, suggested in [13, Section 6.3], leads to an improvement of the numerical results, especially in the minimum Frobenius norm case.

The trust-region subproblems (5.2) have been solved using the routine `trust.m` from the MATLAB Optimization Toolbox which corresponds essentially to the algorithm of Moré and Sorensen [23]. To solve the linear programs (5.4) we have used the routine `linprog.m` from the same MATLAB toolbox. In turn, `linprog.m` uses in most of the instances considered in our optimization runs the interior-point solver `lipsol.m` developed by Zhang [34]. Also, we have chosen to keep the trust-region radius constant when $\rho_k < \eta_1$ and $|Y_k| \geq p_{\min}$.

In a first set of experiments, we considered the test set of unconstrained problems from the CUTEr collection [20] used in the paper [21], which in turn has also been the one used before in the paper [19]. We kept the problem dimensions from [21] but remove all problems considered there with less than 5 variables. This procedure resulted in the test set described in Table 5.1. Most of this problems exhibit some form of sparsity in the Hessian of the objective function like, for instance, a banded format.

In order to present the numerical results for all problems and all methods (and variants) considered, we have used the so-called performance profiles, as suggested in [15]. Performance profiles are, essentially, plots of a cumulative distribution functions $\rho(\tau)$ representing a performance ratio for the different solvers. Let $\mathcal{S}$ be the set of solvers and $\mathcal{P}$ the set of problems. Let $t_{p,s}$ denote the performance of the solver $s \in \mathcal{S}$ on the problem $p \in \mathcal{P}$ — lower values of $t_{p,s}$ indicate better performance. This performance ratio, $\rho(\tau)$, is defined by first setting $r_{p,s} = t_{p,s}/\min\{t_{p,s} : s \in \mathcal{S}\}$, for $p \in \mathcal{P}$ and $s \in \mathcal{S}$. Then, one defines $\rho_s(\tau) = (1/|\mathcal{P}|)|\{p \in \mathcal{P} : r_{p,s} \leq \tau\}|$. Thus, the value of $\rho_s(1)$ is the probability of the solver s having a better performance over the remaining ones. If we are only interested in determining which solver is the most efficient (in the sense of winning the most), we compare the values of $\rho_s(1)$ for all the solvers. At the other end, solvers with the largest values of $\rho_s(\tau)$ for large τ are the ones who solved the largest number of problems in $\mathcal{P}$. As we are particularly inter-

problem	n	DFO-TR Frob ($acc = 6$)	DFO-TR l1 ($acc = 6$)
ARGLINB	10	57	59
ARGLINC	8	56	57
ARWHEAD	15	195	143
BDQRTIC	10	276	257
BIGGS6	6	485	483
BROWNAL	10	437	454
CHNROSNB	15	993	1004
CRAGGLVY	10	548	392
DIXMAANC	15	330	515
DIXMAANG	15	395	451
DIXMAANI	15	429	361
DIXMAANK	15	727	527
DIXON3DQ	10	–	–
DQDRTIC	10	25	25
FREUROTH	10	249	252
GENHUMPS	5	1449	979
HILBERTA	10	8	8
MANCINO	10	106	73
MOREBV	10	111	105
OSBORNEB	11	1363	1023
PALMER1C	8	–	–
PALMER3C	8	56	53
PALMER5C	6	29	29
PALMER8C	8	60	55
POWER	10	466	428
VARDIM	10	502	314

Table 5.1: The test set used in the first set of experiments and the corresponding dimensions (first two columns). The last two columns report the total number of function evaluations required by Algorithm 5.2.1 to achieve an accuracy of 10^{-6} on the objective function value (versions `DFO-TR Frob` and `DFO-TR l1`). In both cases no success was achieved for two of the problems.

ested in considering a wide range of values for τ, we plot the performance profiles in a log-scale (now, the value at 0 represents the probability of winning over the other solvers).

In the specific case of our experiments, we took the *best* objective function value from [21] (obtained by applying a derivative-based Non-Linear Programming solver), to declare whether a problem was successfully solved or not up to a certain accuracy 10^{-acc}. The number $t_{p,s}$ is then the number of function evaluations needed to achieve an objective function value within an absolute error of 10^{-acc} of the best objective function value; otherwise a failure occurs and the value of $r_{p,s}$ used to build the profiles is set to a significantly large number (see [15]). Other measures of performance could be used for $t_{p,s}$ but the number of function evaluations is the most popular in DFO, being also the most appropriate for expensive objective functions. In Figure 5.1, we plot performance profiles for the two variants of Algorithm 5.2.1

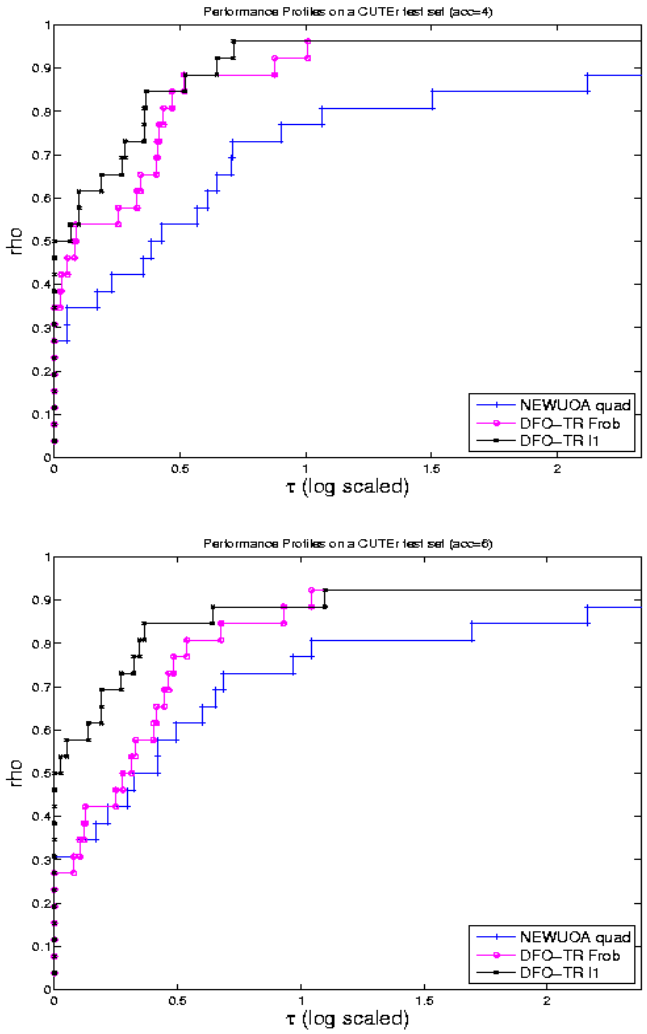


Figure 5.1: Performance profiles comparing Algorithm 5.2.1 (minimum Frobenius and ℓ_1 norm versions) and NEWUOA, on the test set of Table 5.1, for two levels of accuracy (10^{-4} above and 10^{-6} below).

mentioned above and for the state-of-the-art solver NEWUOA [26, 28]. Following [16], and in order to provide a fair comparison, solvers are run first with their own default stopping criterion and if convergence can not be declared another run is repeated with tighter tolerances. In the case of Algorithm 5.2.1, this procedure led to $\epsilon_g = \delta = 10^{-7}$ and a maximum number of 15000 function evaluations, and for NEWUOA we used the data prepared for [21] also for a maximum number of 15000 function evaluations.

Note that NEWUOA requires an interpolation of fixed cardinality in the interval $[2n+1, (n+1)(n+2)/2]$ throughout the entire optimization procedure. We plotted

the extreme possibilities, $2n+1$ and $(n+1)(n+2)/2$, and are reporting results only with the latter one (`NEWUOA quad` in the plots) since it was the one which gave the best results. The two variants of Algorithm 5.2.1, are referred to as `DFO-TR Frob` (minimum Frobenius norm models) and `DFO-TR l1` (minimum ℓ_1-norm models). Two levels of accuracy (10^{-4} and 10^{-6}) are considered in Figure 5.1. One can observe that `DFO-TR l1` is the most efficient version ($\tau = 0$ in the log scale) and basically as robust as the `DFO-TR Frob` version (large values of τ), and that both versions of the Algorithm 5.2.1 seem to outperform `NEWUOA quad` in efficiency and robustness.

problem	n	type of sparsity
ARWHEAD	20	sparse
BDQRTIC	20	banded
BDVALUE	22	banded
BRYDN3D	20	banded
CHNROSNB	20	banded
CRAGGLVY	22	banded
DQDRTIC	20	banded
EXTROSNB	20	sparse
GENHUMPS	20	sparse
LIARWHD	20	sparse
MOREBV	20	banded
POWELLSG	20	sparse
SCHMVETT	20	banded
SROSENBR	20	banded
WOODS	20	sparse

Table 5.2: The test set used in the second set of experiments. For each problem we include the number of variables and the type of sparsity, as described in [9].

In a second set of experiments we ran Algorithm 5.2.1 for the two variants (minimum Frobenius and ℓ_1 norm models) on the test set of CUTEr unconstrained problems used in the paper [9]. These problems are known to have a significant amount of sparsity in the Hessian (this information as well as the dimensions selected is described in Table 5.2). The algorithm has been run now with $\epsilon_g = \delta = 10^{-5}$ and a maximum number of 5000 function evaluations. In Table 5.3, we report the number of objective function evaluations taken as well as the final objective function value obtained. In terms of function evaluations, one can observe that `DFO-TR l1` has approximately 8/9 wins when compared to the `DFO-TR Frob` version, suggesting that the former is more efficient than the latter in the presence of Hessian sparsity. Another interesting aspect of the `DFO-TR l1` version is some apparent ability to produce final model gradients of smaller size.

problem	DFO-TR Frob/l1	# f eval	f val	model ∇ norm
ARWHEAD	Frob	338	3.044e-07	3.627e-03
ARWHEAD	l1	218	9.168e-11	7.651e-07
BDQRTIC	Frob	794	5.832e+01	5.419e+05
BDQRTIC	l1	528	5.832e+01	6.770e-02
BDVALUE	Frob	45	0.000e+00	0.000e+00
BDVALUE	l1	45	0.000e+00	1.297e-22
BRYDN3D	Frob	41	0.000e+00	0.000e+00
BRYDN3D	l1	41	0.000e+00	0.000e+00
CHNROSNB	Frob	2772	3.660e-03	2.025e+03
CHNROSNB	l1	2438	2.888e-03	1.505e-01
CRAGGLVY	Frob	1673	5.911e+00	1.693e+05
CRAGGLVY	l1	958	5.910e+00	8.422e-01
DQDRTIC	Frob	72	8.709e-11	6.300e+05
DQDRTIC	l1	45	8.693e-13	1.926e-06
EXTROSNB	Frob	1068	6.465e-02	3.886e+02
EXTROSNB	l1	2070	1.003e-02	6.750e-02
GENHUMPS	Frob	5000	4.534e+05	7.166e+02
GENHUMPS	l1	5000	3.454e+05	3.883e+02
LIARWHD	Frob	905	1.112e-12	9.716e-06
LIARWHD	l1	744	4.445e-08	2.008e-02
MOREBV	Frob	539	1.856e-04	2.456e-03
MOREBV	l1	522	1.441e-04	3.226e-03
POWELLSG	Frob	1493	1.616e-03	2.717e+01
POWELLSG	l1	5000	1.733e-04	2.103e-01
SCHMVETT	Frob	506	-5.400e+01	1.016e-02
SCHMVETT	l1	434	-5.400e+01	7.561e-03
SROSENBR	Frob	456	2.157e-03	4.857e-02
SROSENBR	l1	297	1.168e-02	3.144e-01
WOODS	Frob	5000	1.902e-01	8.296e-01
WOODS	l1	5000	1.165e+01	1.118e+01

Table 5.3: Results obtained by `DFO-TR Frob` and `DFO-TR l1` on the problems of Table 5.2 (number of evaluations of the objective function, final value of the objective function, and the norm of the final model gradient).

Chapter 6

Conclusion

Since Compressed Sensing emerged, it has been deeply connected to Optimization, using it as a fundamental tool (in particular, to solve ℓ_1-minimization problems). In this thesis, however, we have shown that Compressed Sensing methodology can also serve as a powerful tool for Optimization, in particular for Derivative-Free Optimization (DFO). Namely, we were interested in knowing if it was possible to construct fully quadratic models (essentially models with an accuracy as good as second order Taylor models; see Definition 2.1.2) of a function with sparse Hessian using underdetermined quadratic interpolation on a sample set with much less than $\mathcal{O}(n^2)$ points. We were able to provide, in Theorem 4.3.1, a positive answer to such a question, by considering a random setting and proving that it is possible to construct such models with only $\mathcal{O}(n(\log n)^4)$ points when the number of non-zero components of the Hessian is $\mathcal{O}(n)$. The corresponding quadratic interpolation models were built by minimizing the ℓ_1-norm of the entries of the Hessian model. Our approach was then experimented on a more realistic, deterministic setting, by using these minimum ℓ_1-norm quadratic models in a practical interpolation-based trust-region method (see Algorithm 5.2.1). Our algorithm was able to outperform state-of-the-art DFO methods as shown in the numerical experiments reported in Section 5.3.

One possible way of solving the ℓ_1-minimization problem (2.5) in the context of interpolation-based trust-region methods is to rewrite it as a linear program. This approach was used to numerically test Algorithm 5.2.1 when solving problems (5.3) for $p = 1$. For problems of up to $n = 20, 30$ variables, this way of solving the ℓ_1-minimization problems has produced excellent results in terms of the derivative-free solution of the original minimization problems (5.1) and is still doable in terms of the overall CPU time.

However, for larger values of n, the repeated solution of the linear programs (5.4) becomes significantly heavier. Besides the obvious increase in the dimension in (5.4) and in the number of trust-region iterations, one also has to deal with ill conditioning

due to badly poised sample sets, and it is unclear how to properly regularize. Moreover, it is not simple to warmstart these linear programs. In fact, the number of rows in (5.4) changes frequently making it difficult to warmstart simplex-based methods, and, on the other hand, the difficulties in warmstarting interior-point methods are well known. An alternative is to attempt to approximately solve problem (2.5) by solving $\min \|M(\bar{\phi}, W)\alpha - f(W)\|_2 + \tau\|\alpha_Q\|_1$ for appropriate values of $\tau > 0$. We did some preliminary numerical testing along this avenue but did not succeed in over performing the linear programming approach in any respect. However, it is out of the scope of this thesis a deeper study of the numerical solution of the ℓ_1-minimization problem (2.5) in the context of interpolation-based trust-region methods.

Finally, we would like to stress that building accurate quadratic models for functions with sparse Hessians from function samples could be of interest outside the field of Optimization. The techniques and theory developed in Chapter 4 could also be applicable in other settings of Approximation Theory and Numerical Analysis.

Bibliography

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program.*, 95:3–51, 2003.

[2] A. Bandeira, K. Scheinberg, and L. N. Vicente. Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization. In preparation, 2010.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28:253–263, 2008.

[4] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52:489–509, 2005.

[5] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

[6] E. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52:5406–5425, 2006.

[7] E. J. Candès. Compressive sampling. *Proceedings of the International Congress of Mathematicians Madrid 2006*, Vol. III, 2006.

[8] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22:211–231, 2009.

[9] B. Colson and Ph. L. Toint. Optimizing partially separable functions without derivatives. *Optim. Methods Softw.*, 20:493–508, 2005.

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

[11] A. R. Conn, K. Scheinberg, and Ph. L. Toint. On the convergence of derivative-free methods for unconstrained optimization. In M. D. Buhmann and A. Iserles, editors, *Approximation Theory and Optimization, Tributes to M. J. D. Powell*, pages 83–108. Cambridge University Press, Cambridge, 1997.

[12] A. R. Conn, K. Scheinberg, and L. N. Vicente. Global convergence of general derivative-free trust-region algorithms to first and second order critical points. *SIAM J. Optim.*, 20:387–415, 2009.

[13] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.

[14] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constr. Approx.*, 13:57–98, 1997.

[15] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91:201–213, 2002.

[16] E. D. Dolan, J. J. Moré, and T. S. Munson. Optimality measures for performance profiles. *SIAM J. Optim.*, 16:891–909, 2006.

[17] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.

[18] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inform. Theory*, 47:2845–2862, 2001.

[19] G. Fasano, J. L. Morales, and J. Nocedal. On the geometry phase in model-based algorithms for derivative-free optimization. *Optim. Methods Softw.*, 24:145–154, 2009.

[20] N. I. M. Gould, D. Orban, and Ph. L. Toint. CUTEr (and SifDec), a constrained and unconstrained testing environment, revisited. *ACM Trans. Math. Software*, 29:373–394, 2003.

[21] S. Gratton, Ph. L. Toint, and A. Tröltzsch. Numerical experience with an active-set trust-region method for derivative-free nonlinear bound-constrained optimization. 2010.

[22] H. Y. Le. Convexifying the counting function on $\mathbb{R}^p$ for convexifying the rank function on $\mathcal{M}_{m,n}$. 2010.

[23] J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM J. Sci. Comput.*, 4:553–572, 1983.

[24] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24:227–234, 1995.

[25] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, Berlin, second edition, 2006.

[26] M. J. D. Powell. On trust region methods for unconstrained minimization without derivatives. *Math. Program.*, 97:605–623, 2003.

[27] M. J. D. Powell. Least Frobenius norm updating of quadratic models that satisfy interpolation conditions. *Math. Program.*, 100:183–215, 2004.

[28] M. J. D. Powell. The NEWUOA software for unconstrained optimization without derivatives. In *Nonconvex Optim. Appl.*, volume 83, pages 255–297. Springer-Verlag, Berlin, 2006.

[29] H. Rauhut. *Compressed Sensing and Structured Random Matrices*. In M. Fornasier, editor, Theoretical Foundations and Numerical Methods for Sparse Recovery. Radon Series Comp. Appl. Math. deGruyter 9, 2010.

[30] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1-minimization. *arXiv:10003.0251 [math.NA] (preprint, 2010).*

[31] K. Scheinberg and Ph. L. Toint. Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization. Technical Report 09/06, Dept. of Mathematics, FUNDP, Namur, 2009.

[32] T. Tao. Open question: Deterministic UUP matrices: `http://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices`, 2007.

[33] D. Taubman and M. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. The International Series in Engineering and Computer Science. Kluwer Academic Publishers, Massachusetts, 2002.

[34] Y. Zhang. Solving large-scale linear programs by interior-point methods under the MATLAB environment. *Optim. Methods Softw.*, 10:1–31, 1998.