

Κατασκευή προσεγγιστικών equiangular tight frames
και εφαρμογές

Η ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύστασης

του Τμήματος Μηχανικών Η/Υ και Πληροφορικής Εξεταστική
Επιτροπή

από την

Ευαγγελία Τσιλιγιάννη

ως μέρος των Υποχρεώσεων για τη λήψη του

ΔΙΔΑΚΤΟΡΙΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

Ιούλιος 2015

Τριμελής Συμβουλευτική Επιτροπή

1. Λυσίμαχος-Παύλος Κόντης, Αναπληρωτής Καθηγητής του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων (Επιβλέπων).
2. Άγγελος Κατσάγγελος, Professor, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA.
3. Χριστόφορος Νίκου, Αναπληρωτής Καθηγητής του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.

Επταμελής Εξεταστική Επιτροπή

1. Λυσίμαχος-Παύλος Κόντης, Αναπληρωτής Καθηγητής του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
2. Χριστόφορος Νίκου, Αναπληρωτής Καθηγητής του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
3. Άγγελος Κατσάγγελος, Professor, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA.
4. Αριστείδης Λύκας, Καθηγητής του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
5. Κωνσταντίνος Μπλέκας, Επίκουρος Καθηγητής του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
6. Κωνσταντίνος Παρσόπουλος, Επίκουρος Καθηγητής του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
7. Παναγιώτης Τσακαλίδης, Καθηγητής του Τμήματος Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης.

CONTENTS

1	Introduction	1
1.1	Overview	2
1.1.1	Sparse representations	4
1.1.2	Compressed sensing	5
1.1.3	Spreading sequences for s-CDMA	5
1.2	Contributions	6
1.3	Outline	7
2	Frames review	9
2.1	Preliminaries	10
2.2	Finite frames basics	11
2.2.1	Frame operators	12
2.2.2	Tight frames	14
2.2.3	Unit norm tight frames	15
2.2.4	Equiangular tight frames	16
2.3	Connection of frames to graphs	18
2.4	The frame design problem	19
3	Construction of approximately equiangular tight frames	23
3.1	Alternating projections	24
3.1.1	Alternating projections on nonconvex sets	25
3.2	Averaged projections	25
3.2.1	Convergence for averaged projections on prox-regular sets	26
3.3	Construction of incoherent unit norm tight frames	27
3.3.1	Algorithm 1	28
3.3.2	Convergence of Algorithm 1	30
3.3.3	Algorithm 2	32
3.3.4	Convergence of Algorithm 2	33
3.3.5	Experimental results	34
3.4	Construction of nearly equiangular frames	39
3.4.1	Construction of signature matrices	40
3.4.2	Nearly equiangular frames based on signature matrices	44
3.4.3	Nearly equiangular, nearly tight frames based on signature matrices	45

3.5	Comparison of the proposed constructions	47
4	Preconditioning in Sparse and Redundant Representations	51
4.1	The sparse representation problem	52
4.2	Mutual coherence and RIP	53
4.2.1	Mutual Coherence	53
4.2.2	Uniqueness via mutual coherence	54
4.2.3	The Restricted Isometry Property	54
4.2.4	Relation between RIP and mutual coherence	55
4.3	Promoting a sparse solution	56
4.3.1	The ℓ_0 -minimizer	56
4.3.2	Stability of ℓ_0 minimization via the RIP	56
4.3.3	The ℓ_1 -minimizer	57
4.4	The role of the spectral norm	58
4.5	Preconditioning	59
4.5.1	Construction of a preconditioner	61
4.5.2	Preconditioning with incoherent UNTFs	61
4.5.3	Preconditioning with nearly equiangular frames	63
5	Improving Sparse Recovery in Compressed Sensing	67
5.1	Compressed sensing basics	68
5.2	Projection matrices constructions	71
5.2.1	Deterministic projections	72
5.2.2	Structured random projections	73
5.2.3	Optimized projections	74
5.3	Compressed sensing with the proposed frame constructions	75
5.4	Proposed optimized projections	77
5.4.1	Optimized projections using incoherent UNTFs	79
5.5	Preconditioning in compressed sensing	84
6	Spreading sequences for s-CDMA	89
6.1	S-CDMA model	90
6.2	Design of spreading sequences	91
6.2.1	Interuser Interference	91
6.2.2	Welch Bound Equality (WBE) sequences	92
6.2.3	Sum capacity	93
6.3	Optimal spreading sequences for varying number of users	94
6.4	Codebooks from nearly equiangular, nearly tight frames	96
7	Conclusions and future work	99
A	Projections	115

LIST OF FIGURES

3.1	Convergence of Algorithm 1 (alternating projections) for a 60×120 matrix. The mean squared distance between the current iteration and the sets we project on reduces in a linear rate.	30
3.2	Convergence of Algorithm 1 (alternating projections) for a 25×120 matrix. The convergence rate depends on the bound used in eq. (3.4). In (a) we observe a sub-linear convergence rate when the bound equals $1/\sqrt{m}$. In (b) the convergence rate becomes linear as the bound is relaxed to $3/2\sqrt{m}$	31
3.3	Convergence of Algorithm 2 (averaged projections) for a 60×120 matrix. The mean squared distance between the current iteration and the sets we project on reduces in a linear rate.	34
3.4	Convergence of Algorithm 2 (averaged projections) for a 25×120 matrix. The convergence rate depends on the bound used in eq. (3.4). In (a) we observe a sub-linear convergence rate when the bound equals $1/\sqrt{m}$. In (b) the convergence rate becomes linear as the bound is relaxed to $3/2\sqrt{m}$	35
3.5	Mutual coherence (left) and spectral norm (right) as a function of the number of iterations. The experiments involve frames of various dimensions.	36
3.6	Distribution of Gram matrix entries of a 60×120 frame.	37
3.7	Distribution of Gram matrix entries of a 20×120 frame.	37
3.8	The spectrum of the signature matrix of a 64×128 random Gaussian matrix before and after processing the matrix with Algorithms 3 and 4. The black dotted line stands for the spectrum of the signature matrix corresponding to a 64×128 ETF.	42
3.9	Correlation distribution of frame vectors produced with Algorithm 5. μ_{opt} stands for the optimal lowest bound (Welch bound).	45
3.10	Correlation distribution of frame vectors produced with Algorithm 6. μ_{opt} stands for the optimal lowest bound (Welch bound).	45
4.1	Discrepancy between the Gram or pseudo-Gram matrices involved in support estimation and the identity matrix of the same dimensions. The experiments involve $m \times N$ matrices with $m = 64 : 32 : 192$ and $N = 256$	62
4.2	Support recovery rates for sparse representations using OMP for signals with varying support size. The preconditioner's construction was based on the construction of incoherent UNTFs.	63

4.3	Support recovery rates for sparse representations using BP for signals with varying support size. The preconditioner's construction was based on the construction of incoherent UNTFs.	64
4.4	Support recovery rates for sparse representations using Dantzig selector for signals with varying support size. The preconditioner's construction was based on the construction of incoherent UNTFs.	64
4.5	Support recovery rates for sparse representations using Dantzig selector for signals with varying support size. The preconditioner's construction was based on the construction of nearly equiangular, nearly tight frames.	65
5.1	Properties of the effective dictionaries involved in CS reconstruction experiments. In (a) we present mutual coherence as a function of the number of measurements. The bottom brown dash-dotted line represents the lowest possible bound (see eq. (2.25)). In (b) we present spectral norm as a function of the number of measurements. The red dotted line corresponding to our methodology coincide with the lowest possible bound.	80
5.2	Changes in the distribution of the column correlation of a 25×120 frame.	81
5.3	CS performance for random and optimized projection matrices by means of relative MSE in a logarithmic scale. Numerical recovery deploys OMP. In (a) we keep the sparsity level fixed and vary the number of measurements. In (b) we keep the number of measurements fixed and vary the sparsity level. A vanishing graph implies a zero error rate.	82
5.4	CS performance for random and optimized projection matrices by means of relative MSE in a logarithmic scale. Numerical recovery deploys BP. In (a) we keep the sparsity level fixed and vary the number of measurements. In (b) we keep the number of measurements fixed and vary the sparsity level. A vanishing graph implies a zero error rate.	83
5.5	Support recovery rates for OMP and BP, for signals with varying support size acquired with Bernoulli random projections. The signals considered in (a) are sparse under a random Gaussian dictionary. The signals considered in (b) are sparse under a Haar-DCT dictionary.	86
6.1	Standard deviation of the interference term for variable number of active users in an s-CDMA system designed for 128 users.	96

LIST OF TABLES

3.1	Mutual coherence of $m \times N$ frames, with $m = 20 : 20 : 100$ and $N = 120$. . .	38
3.2	Spectral norm of $m \times N$ frames, with $m = 20 : 20 : 100$ and $N = 120$	38
3.3	Spectral norm of $m \times N$ frames with $m = 32 : 16 : 96$ and $N = 128$ obtained with Algorithm 5 and Algorithm 6.	46
3.4	Standard deviation of the Gram matrix entries corresponding to $m \times N$ frames with $m = 32 : 16 : 96$, $N = 128$, obtained with Algorithms 1, 5 and 6.	47
3.5	Mutual coherence of $m \times N$ frames with $m = 32 : 16 : 96$, $N = 128$, obtained with Algorithms 1, 5 and 6.	48
3.6	Average coherence of $m \times N$ frames with $m = 32 : 16 : 96$, $N = 128$, obtained with Algorithms 1, 5 and 6.	49
5.1	Recovery rates for sparse signals of length $N = 120$ obtained with CS, for variable number of measurements, $m = 15 : 5 : 35$, and various types of projection matrices.	76
5.2	Properties of sensing matrices employed in CS experiments. Results involve $m \times N$ matrices with $m \in \{20, 30\}$, $N = 120$	76
5.3	Recovery rates for CS with Bernoulli and optimized projections. When Bernoulli projections are used, recovery involves preconditioning.	87
6.1	Average total squared correlation (TSC) for variable number of active users.	97

LIST OF ALGORITHMS

1	Construction of incoherent UNTFs with Alternating Projections	29
2	Construction of incoherent UNTFs with Averaged Projections	32
3	Signature Matrix Construction I	41
4	Signature Matrix Construction II	43
5	Construction of a nearly equiangular frame	44
6	Construction of a nearly equiangular, nearly tight frame	46
7	OMP: approximately solve ℓ_0 -minimization problem	57

ABSTRACT

Evaggelia V. Tsiligianni, PhD, Computer Science & Engineering Department, University of Ioannina, Greece. July, 2015. Construction of approximately equiangular tight frames and their applications. Thesis Supervisor: Lisimachos P. Kondi.

Frames are considered a natural extension of orthonormal bases to overcomplete spanning systems. In the signal processing community, frames have mainly become popular due to wavelets; however, many other frame families have been employed in numerous applications, including source coding, robust transmission, code division multiple access (CDMA) systems, and coding theory. The most important characteristic of frames is redundancy, which adds more flexibility to signal expansions, facilitating various signal processing tasks.

A finite frame with N vectors in an m -dimensional Hilbert space \mathbb{H}^m is usually identified with the $m \times N$ matrix $F = [f_1 \ f_2 \ \dots \ f_N]$, $m \leq N$, with columns the frame vectors $f_k \in \mathbb{H}^m$, $k = 1, \dots, N$. The most important properties of frames are *mutual coherence* and *spectral norm*. Mutual coherence is a measure of the maximal correlation between the frame vectors and characterizes the degree of similarity between the columns of the matrix F . Spectral norm measures how much a frame can dilate a unit norm coefficient vector. Mutual coherence and spectral norm define particular classes of frames. *Unit norm tight frames (UNTFs)* attain optimal bounds of spectral norm; these frames have unit norm columns and orthogonal rows of equal norm. Unit norm tight frames with small mutual coherence are referred to as *incoherent UNTFs*. The minimum possible mutual coherence is attained by *equiangular tight frames (ETFs)*. The frame vectors of ETFs exhibit identical correlation and these frames are considered closest to orthonormal bases.

ETFs offer erasure-robust transmission in communications and minimize interuser interference when employed as spreading sequences in multiuser communication systems. Due to their incoherence, they are of interest in sparse representations and compressed sensing. However, ETFs do not exist for all frame dimensions and their construction has been proved extremely difficult.

This thesis presents two methods that produce real frames close to ETFs. The proposed constructions are motivated by specific applications, namely, compressed sensing and sparse representations. Concerning sparse or compressible signals, that is, signals with a few significant coefficients, compressed sensing and sparse representations have

experienced a growing interest in the last decade, providing the ability of compact representations that serve various data sources. The mathematical modeling in the heart of these applications involves an underdetermined linear system with more unknowns than equations. Computing its sparsest solution, i.e., the one with the fewest non-vanishing coefficients is tractable with numerical methods. Standard numerical solvers include Orthogonal Matching Pursuit (OMP) and Basis Pursuit (BP).

In sparse and redundant representations, we seek a sparse signal representation with respect to a redundant (overcomplete) dictionary. Performance guarantees for the algorithms deployed to compute the non-vanishing coefficients require that the given dictionary forms an incoherent UNTF. While many incoherent dictionaries are known in the literature, their limited sparsifying ability has promoted the design of learning based dictionaries. Often, learning based dictionaries do not satisfy the necessary properties for numerical computations.

Compressed sensing is a sampling theory that allows signal reconstruction from an incomplete number of measurements. Concerning signals that are sparse or compressible, compressed sensing uses a sensing mechanism implemented by an appropriate matrix, the so-called *projection matrix*. According to theoretical results, the projection matrix must possess a property known as the restricted isometry property (RIP). Constructing RIP matrices is difficult, as evaluation of RIP is combinatorially complex. Random Gaussian or Bernoulli matrices satisfy RIP with high probability. Considering N -dimensional signals with s non-vanishing coefficients, recovery conditions for random matrices require $\mathcal{O}(s \log N)$ measurements. More recent results formulate similar recovery guarantees for projection matrices that form incoherent UNTFs. Thus, a new design strategy involves the construction of projection matrices exhibiting small mutual coherence and spectral norm.

Minimum bounds of mutual coherence and spectral norm are attained by ETFs; therefore, the methods proposed here aim at the construction of frames as close to ETFs as possible. The first method uses results from frame theory and relies on alternating projection ideas. The produced constructions form UNTFs with remarkably small mutual coherence, that is, incoherent UNTFs. The second method relies on recent results showing that there is one-to-one correspondence of ETFs to a special type of graphs. The existence of an ETF is determined by the so-called *signature matrix*. A signature matrix has the form of the adjacency matrix of a graph and its spectrum consists of two distinct eigenvalues. Viewing the construction of a signature matrix as an inverse eigenvalue problem, we develop a numerical algorithm to compute a solution that approximates the signature matrix of an ETF. The second method produces *nearly equiangular, nearly tight frames*, that is, frames with similar column correlation and approximately optimal spectral norm.

The proposed frame constructions are employed as projection matrices in compressed sensing, improving substantially the performance of the deployed algorithms in sparse recovery. Considering that many signals are sparse or compressible under overcomplete dictionaries, incoherent UNTFs are also used for the design of optimized projection ma-

trices with respect to a given representation dictionary. An additional way to employ the proposed frames to solve underdetermined linear systems is the technique of preconditioning. Applying preconditioning to sparse representations, we improve the performance of the algorithms deployed to find the coefficients of the sparse signal. In compressed sensing, preconditioning is used to improve signal recovery when binary matrices are used as projection matrices. Note that binary matrices are considered more suitable for hardware implementation.

Besides compressed sensing and sparse representations, one of the proposed constructions has been employed in the design of near-optimal *codes* or *spreading sequences* in synchronous CDMA systems. Optimal spreading sequences maximize the rate at which the users can transmit and minimize interuser interference. Equal norm tight frames have been proved optimal, if all users in the system are active. When the number of users changes, the only frames that can minimize interuser interference are ETFs. However, only a few ETF constructions are known in the literature. The near optimal codebook presented here has the form of a nearly equiangular, nearly tight frame and minimizes interuser interference even when some users in the system are silent.

ΠΕΡΙΛΗΨΗ

Ευαγγελία Τσιλιγιάννη του Βασιλείου και της Ελένης. PhD, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Ιούλιος, 2015. Κατασκευή προσεγγιστικών equiangular tight frames και εφαρμογές. Επιβλέπων: Λυσίμαχος Παύλος Κόντης.

Τα frames είναι υπερπλήρη συστήματα που παράγουν έναν διανυσματικό χώρο και θεωρούνται επέκταση των ορθοκανονικών βάσεων. Στην επεξεργασία σήματος, τα frames έγιναν γνωστά χάρη στα wavelets. Άλλοι τύποι frames έχουν χρησιμοποιηθεί σε ποικίλες εφαρμογές, όπως είναι η κωδικοποίηση, η εύρωστη μετάδοση και τα συστήματα πολλαπλής προσπέλασης με διαίρεση κώδικα (Code Division Multiple Access–CDMA). Η υπερπληρότητα θεωρείται το πιο σημαντικό χαρακτηριστικό των frames, διότι προσφέρει ευελιξία στην αναπαράσταση ενός σήματος και διευκολύνει την επεξεργασία.

Ένα frame με πεπερασμένο πλήθος διανυσμάτων που παράγει τον m -διάστατο διανυσματικό χώρο \mathbb{H}^m , συνήθως, αναπαριστάται από έναν πίνακα μεγέθους $m \times N$, που έχει ως στήλες τα διανύσματα του frame, δηλαδή, $F = [f_1 \ f_2 \ \dots \ f_N]$, $m \leq N$, $f_k \in \mathbb{H}^m$, $k = 1, \dots, N$. Ως πιο σημαντικές ιδιότητες ενός frame θεωρούνται η *αμοιβαία συνάφεια* (*mutual coherence*) και η *φασματική νόρμα* (*spectral norm*). Η αμοιβαία συνάφεια αποτελεί ένα μέτρο της μέγιστης συσχέτισης των διανυσμάτων του frame και εκφράζει την ομοιότητα μεταξύ των στηλών του πίνακα F . Η φασματική νόρμα αποτελεί μέτρο της μέγιστης δυνατής διαστολής ενός μοναδιαίου διανύσματος, όταν αυτό πολλαπλασιαστεί με το frame. Οι δύο ιδιότητες ορίζουν συγκεκριμένες κατηγορίες frames. Τα *unit norm tight frames* (*UNTFs*) εμφανίζουν τη μικρότερη δυνατή φασματική νόρμα. Τα συγκεκριμένα frames έχουν στήλες μοναδιαίου μέτρου και ορθογώνιες γραμμές ίσου μέτρου. Όταν ένα UNTF εμφανίζει μικρή αμοιβαία συνάφεια, τότε χαρακτηρίζεται ως *incoherent UNTF*. Η ελάχιστη δυνατή αμοιβαία συνάφεια συναντάται στα *equiangular tight frames* (*ETFs*). Τα διανύσματα των ETFs εμφανίζουν ταυτόσημη συσχέτιση και τα frames αυτού του τύπου θεωρούνται ως η καλύτερη προσέγγιση ορθοκανονικών βάσεων.

Τα ETFs έχουν προταθεί για την επίτευξη εύρωστης μετάδοσης σε συστήματα επικοινωνίας, καθώς και για την ελαχιστοποίηση της παρεμβολής μεταξύ των χρηστών σε συστήματα πολλαπλής προσπέλασης. Χάρη στην ελάχιστη αμοιβαία συνάφεια που εμφανίζουν, παρουσιάζουν ενδιαφέρον σε εφαρμογές όπως οι αραιές αναπαραστάσεις (*sparse representations*) και η συμπιεστική δειγματοληψία (*compressed sensing*). Όμως, ETFs δεν υπάρχουν για οποιεσδήποτε διαστάσεις, ενώ η κατασκευή τους έχει αποδειχθεί ιδιαίτερα δύσκολη.

Στην παρούσα διατριβή προτείνονται δύο μέθοδοι για την κατασκευή προσεγγιστικών ETFs. Κίνητρο για τη κατασκευή των προτεινόμενων frames αποτελεί η εφαρμογή τους σε προβλήματα αραιών αναπαραστάσεων και συμπιεστικής δειγματοληψίας. Οι συγκεκριμένες εφαρμογές αφορούν σήματα που μπορούν να παρασταθούν από λίγους μη μηδενικούς συντελεστές, δηλαδή, αραιά ή συμπίεσιμα σήματα, και έχουν γνωρίσει ιδιαίτερη ανάπτυξη την τελευταία δεκαετία, διότι παρέχουν τη δυνατότητα συμπαγών αναπαραστάσεων, χρήσιμων για διάφορους τύπους δεδομένων. Το μαθηματικό μοντέλο που βρίσκεται στην καρδιά των συγκεκριμένων αναπαραστάσεων είναι ένα υπο-ορισμένο γραμμικό σύστημα, με πλήθος εξισώσεων μικρότερο από το πλήθος των αγνώστων. Ο υπολογισμός της αραιότερης λύσης, δηλαδή, της λύσης με το μικρότερο πλήθος μη μηδενικών συντελεστών, είναι εφικτός με τη χρήση κατάλληλων αριθμητικών μεθόδων. Οι πιο γνωστοί αλγόριθμοι είναι ο Orthogonal Matching Pursuit (OMP) και ο Basis Pursuit (BP).

Η αναπαράσταση ενός σήματος με λίγους μη μηδενικούς συντελεστές, συνήθως, επιτυγχάνεται με τη χρήση ενός υπερπλήρους συστήματος αναπαράστασης, που είναι γνωστό ως λεξικό (dictionary). Η αποδοτική λειτουργία των αλγορίθμων που χρησιμοποιούνται για τον υπολογισμό των μη μηδενικών συντελεστών προϋποθέτει την ικανοποίηση συγκεκριμένων συνθηκών. Μια από αυτές απαιτεί το λεξικό να έχει τη μορφή ενός incoherent UNTF. Ωστόσο, γνωστά λεξικά αυτής της μορφής δεν οδηγούν σε ικανοποιητικό επίπεδο αραιότητας. Για το λόγο αυτό πολλά λεξικά έχουν σχεδιαστεί χρησιμοποιώντας τεχνικές εκμάθησης. Συνήθως, όμως, τα λεξικά αυτού του τύπου δεν ικανοποιούν τις συνθήκες που απαιτούν οι αλγόριθμοι υπολογισμού της αραιής αναπαράστασης.

Η θεωρία της συμπιεστικής δειγματοληψίας καθιστά δυνατή την ανάκτηση ενός σήματος από ένα πλήθος ελλιπών μετρήσεων. Η συμπιεστική δειγματοληψία αφορά σήματα που είναι αραιά ή συμπίεσιμα και χρησιμοποιεί έναν μηχανισμό δειγματοληψίας που υλοποιείται με τη βοήθεια κατάλληλου πίνακα, γνωστού ως πίνακα προβολών (*projection matrix*). Σύμφωνα με τη θεωρία, ο πίνακας αυτός πρέπει να έχει την ιδιότητα περιορισμένης ισομετρίας (restricted isometry property–RIP). Η κατασκευή τέτοιων πινάκων είναι ιδιαίτερα δύσκολη, διότι η επαλήθευση της RIP απαιτεί συνδυαστικούς υπολογισμούς. Οι πιο γνωστοί πίνακες που ικανοποιούν τη RIP με μεγάλη πιθανότητα είναι οι τυχαίοι πίνακες Gauss και Bernoulli. Για τους πίνακες αυτούς υπάρχουν θεωρητικά αποτελέσματα που αποδεικνύουν ότι είναι εφικτή η ανάκτηση ενός σήματος μήκους N με s μη μηδενικούς συντελεστές, όταν το πλήθος μετρήσεων είναι της τάξης $\mathcal{O}(s \log N)$. Σύμφωνα πρόσφατα αποτελέσματα, η παραπάνω συνθήκη ανάκτησης ισχύει και όταν ο πίνακας προβολών έχει τη μορφή ενός incoherent UNTF. Συνεπώς, μια νέα στρατηγική κατασκευής πινάκων προβολών περιλαμβάνει την κατασκευή πινάκων με χαμηλή αμοιβαία συνάφεια και μικρή φασματική νόρμα.

Ελάχιστες τιμές τόσο για την αμοιβαία συνάφεια όσο και για τη φασματική νόρμα συναντώνται στα ETFs. Επομένως, οι προτεινόμενες μέθοδοι στοχεύουν στην κατασκευή προσεγγιστικών ETFs. Η πρώτη μέθοδος χρησιμοποιεί αποτελέσματα από τη θεωρία των frames και βασίζεται σε ιδέες που χρησιμοποιούνται στη μέθοδο των εναλλασσόμενων προβολών (alternating projections). Τα frames που παράγει έχουν τη μορφή UNTFs και εμφανίζουν μικρή αμοιβαία συνάφεια, οπότε αποτελούν incoherent UNTFs. Η δεύτερη

μέθοδος βασίζεται σε πρόσφατα αποτελέσματα που αποδεικνύουν την ύπαρξη αμφιμονοσήμαντης αντιστοιχίας μεταξύ ETFs και γράφων συγκεκριμένου τύπου. Η ύπαρξη ενός ETF καθορίζεται από έναν πίνακα, γνωστό ως *πίνακα signature*, που έχει τη μορφή πίνακα γειτονίας γράφου και το φάσμα του αποτελείται από δύο διακριτές ιδιοτιμές. Αντιμετωπίζοντας την κατασκευή του πίνακα signature ως ένα αντίστροφο πρόβλημα ιδιοτιμών (*inverse eigenvalue problem*), προτείνουμε έναν αριθμητικό αλγόριθμο που οδηγεί σε προσεγγιστική λύση. Η δεύτερη μέθοδος παράγει προσεγγιστικά ETFs, με διανύσματα που εμφανίζουν παρόμοια συσχέτιση και σχεδόν βέλτιστη φασματική νόρμα.

Οι προτεινόμενες κατασκευές χρησιμοποιούνται ως πίνακες προβολών για συμπίεστική δειγματοληψία, βελτιώνοντας σημαντικά την απόδοση των σχετικών αλγορίθμων στην ανάκτηση αραιών σημάτων. Επειδή πολλά σήματα έχουν αραιές αναπαραστάσεις ως προς υπερπλήρη λεξικά, χρησιμοποιούμε τα προτεινόμενα *incoherent UNTFs* για την κατασκευή βελτιστοποιημένων πινάκων προβολών σε σχέση με δεδομένο λεξικό. Ένας επιπλέον τρόπος για την αξιοποίηση των προτεινόμενων κατασκευών στην επίλυση υπο-ορισμένων γραμμικών συστημάτων είναι η τεχνική της προρρύθμισης. Εφαρμόζοντας προρρύθμιση σε αραιές αναπαραστάσεις οδηγούμαστε σε καλύτερη απόδοση των αλγορίθμων που χρησιμοποιούνται για τον υπολογισμό των μη μηδενικών συντελεστών. Στη συμπίεστική δειγματοληψία η προρρύθμιση βελτιώνει την ανάκτηση του σήματος, όταν χρησιμοποιούνται δυαδικοί πίνακες προβολών. Σημειώνουμε ότι οι δυαδικοί πίνακες προβολών παρουσιάζουν ευκολότερη πρακτική υλοποίηση.

Εκτός από τις αραιές αναπαραστάσεις και τη συμπίεστική δειγματοληψία, μια από τις προτεινόμενες κατασκευές είναι κατάλληλη για τη σχεδίαση σχεδόν βέλτιστων κωδικών (*codes*) ή ακολουθιών εξάπλωσης (*spreading sequences*) σε συστήματα σύγχρονου CDMA. Είναι γνωστό ότι οι βέλτιστες ακολουθίες έχουν τη μορφή *equal norm tight frames* και οδηγούν σε μεγιστοποίηση του ρυθμού μετάδοσης, ενώ ελαχιστοποιούν την παρεμβολή μεταξύ χρηστών. Ωστόσο, όταν το πλήθος των ενεργών χρηστών είναι μεταβαλλόμενο, τότε οι ακολουθίες είναι βέλτιστες μόνο όταν έχουν τη μορφή ETFs. Δυστυχώς, μόνο λίγες κατασκευές ETFs υπάρχουν στη βιβλιογραφία. Το σύστημα κωδικών που παρουσιάζεται εδώ έχει τη μορφή προσεγγιστικών ETFs και ελαχιστοποιεί την παρεμβολή μεταξύ χρηστών ακόμα και όταν κάποιοι χρήστες είναι ανενεργοί.

CHAPTER 1

INTRODUCTION

1.1 Overview

1.2 Contributions

1.3 Outline

Sometimes the representation of a function or an operator by an overcomplete spanning system is preferable over the use of an orthonormal basis. The reason for this may be that an orthonormal basis with the desired properties does not exist or the deliberate introduction of redundancy. Frames can be regarded as the most natural generalization of the notion of orthonormal bases. Particularly useful in applications are frames in finite dimensional spaces. A finite frame is a spanning set of vectors, which are generally redundant (overcomplete). As frames have more vectors than the dimension of the space, each vector in the space will have infinitely many representations with respect to the frame. While armed with the advantage of redundancy, frames come with the drawback that the frame vectors are linearly dependent.

A finite frame with N vectors in an m -dimensional Hilbert space \mathbb{H}^m is usually identified with the $m \times N$ matrix $F = [f_1 \ f_2 \ \dots \ f_N]$, $m \leq N$, with columns the frame vectors $f_k \in \mathbb{H}^m$, $k = 1, \dots, N$. In many applications there is a need to design frames that are as close to orthonormal bases as possible. Unit norm columns, orthogonal equal norm rows, equal correlation between frame vectors are the desired properties of such frames; the corresponding frame classes are known as unit norm frames, tight frames and equiangular frames, respectively. The most important category of frames includes equiangular unit norm tight frames (ETFs) also known as optimal Grassmannian frames. These frames combine all of the above properties and they also minimize the maximal column correlation $\max_{k \neq \ell} |\langle f_k, f_\ell \rangle|$; therefore, they are considered to be closest to orthonormal bases. Despite their important properties, ETFs do not exist for all frame dimensions and their construction is extremely difficult. Thus, in many applications similar frame constructions are used as substitutes.

This thesis proposes two numerical methods for the construction of frames that are close to ETFs. The obtained frames exhibit small column correlation, a property known as incoherence, and small spectral norm, meaning that they are close to unit norm tight frames. Using these frames in sparse signal recovery in redundant representations and compressed sensing, we substantially improve the performance of the numerical algorithms deployed to find sparse signals. One of the proposed methods yields nearly equiangular frames, which are employed as spreading sequences in synchronous Code Division Multiple Access (s-CDMA) systems, minimizing interuser interference.

1.1 Overview

Let x be a vector of coefficients representing data in a real or complex m -dimensional Hilbert space \mathbb{H}^m . One common approach to data processing is the decomposition of x according to a representation system $\{f_k\}_{k=1}^N$, $N \geq m$, by considering the map

$$x \mapsto (\langle x, f_k \rangle)_{k=1}^N$$

The choice of the representation system is dictated by the treated data and the application of interest. A successful choice enables us to solve a variety of analysis tasks. For example, the sequence $(\langle x, f_k \rangle)_{k=1}^N$ allows compression of x , which is in fact the heart of the new JPEG2000 compression standard when choosing $\{f_k\}_{k=1}^N$ to be a wavelet system.

An accompanying approach is the expansion of the data x by considering sequences $\{c_k\}_{k=1}^N$ satisfying

$$x = \sum_{k=1}^N c_k f_k.$$

It is well known that suitably chosen representation systems allow sparse representations, that is, representations with small number of nonvanishing coefficients.

A representation system that forms an orthonormal basis for \mathbb{H}^m is the standard choice. While orthonormal bases provide unique representations they exhibit important drawbacks. From the decomposition viewpoint, the obtained sequence is far from being robust to erasures. Every single coefficient encapsulates unique information of the data x ; thus, its loss cannot be recovered. From the expansion viewpoint, orthonormal basis rarely yield sparse representations, therefore, they are not suitable for sparsity methodologies like compressed sensing.

These problems can be tackled by allowing the system $\{f_k\}_{k=1}^N$ to be redundant, leading us naturally to the notion of Hilbert frames. Redundancy is a fundamental characteristic of frames and plays a significant role in applications. Due to redundancy frames offer greater design flexibility and can be constructed to fit a particular problem in a manner impossible by a set of linearly independent vectors. For example, in sparse signal representations, a redundant frame can be chosen to fit its content to the data, achieving a high sparsity level that would not be easily obtained using an orthonormal basis. A second major advantage of redundancy is robustness. Frames have the advantage to spread

the information over a wider range of vectors, offering resilience against erasures (losses). Erasures are, for instance, a severe problem in wireless sensor networks when transmission losses occur.

The advantages provided by the frame redundancy come at the cost that the representation may not be unique. Thus, while we have good reasons to trade orthonormal bases for frames, we still want to preserve as many properties of orthonormal bases as possible. To measure the nearness of a frame to an orthonormal basis, we define two important properties. The first is the maximal correlation of the frame vectors defined as the largest absolute normalized inner product between different frame columns

$$\mu(F) = \max_{\substack{1 \leq k, \ell \leq N \\ k \neq \ell}} \frac{|\langle f_k, f_\ell \rangle|}{\|f_k\| \|f_\ell\|}, \quad (1.1)$$

where $\|\cdot\|$ denotes the Euclidean norm. In sparse representations the maximal correlation is referred to as *mutual coherence* [93] and is bounded according to [119]

$$\sqrt{\frac{N-m}{m(N-1)}} \leq \mu(F) \leq 1. \quad (1.2)$$

Frames with small mutual coherence are known as *incoherent*.

An interpretation of the incoherence property from an information theoretic viewpoint is the following. Requiring a matrix F with small mutual coherence, that is, with columns as “independent” as possible, means that the information of a vector x expanded by F is spread in different directions, which makes its recovery easier. As we will see, mutual coherence plays an important role in the existence of a unique solution of underdetermined linear systems as well as in the performance of the algorithms deployed to find sparse solutions.

After the mutual coherence, the spectral norm $\|F\|$ is the most important geometric quantity associated with a frame F . Spectral norm equals the largest eigenvalue of $F^T F$ and measures how much the frame can dilate a unit norm coefficient vector, so it reflects how much the columns of F are “spread out”. A lower bound on the spectral norm of a frame is given by

$$\|F\|^2 \geq \frac{N}{m}. \quad (1.3)$$

When equality holds in this relation, the frame forms a unit norm tight frame (UNTF). Equivalently, the rows of F are mutually orthogonal vectors with equal norms. Minimum bounds of both mutual coherence and spectral norm are achieved by equiangular tight frames. ETFs have unit norm vectors forming equal angles, exhibiting minimal dependency; thus, they are considered to be closest to orthonormal bases. However, the construction of ETFs is extremely difficult, while it has been proved that ETFs do not exist for all frame dimensions.

Mutual coherence and spectral norm define particular classes of frames and play a significant role in applications. Most of the problems employing frames demand certain

desired properties; thus, most frame constructions are application specific. Following this rule, the work presented in this thesis is motivated by the research for a good sensing operator for compressed sensing. The importance of incoherence in sparse signal recovery, both in redundant representations and compressed sensing makes ETFs ideal candidates for these problems [70, 124, 125]. The numerical constructions proposed here produce frames close to ETFs in the sense that the obtained frames exhibit mutual coherence and spectral norm approximating or, sometimes, attaining the minimum bounds. The first method relies on frame theory and constructs incoherent UNTFs. These frames satisfy the theoretical conditions for sparse recovery, and are used in compressed sensing to optimize the measurement process and improve signal reconstruction. The second method is based on results connecting frames to graphs and produces nearly equiangular frames, which are also employed in compressed sensing to improve recovery rates.

Besides compressed sensing, the proposed frames are found useful in a similar problem, namely in sparse reconstruction of redundant representations. The mathematical technique that enables their employment in this problem is referred to as preconditioning. Moreover, based on recent results establishing the important role of equiangularity in designing optimal codes for multiuser communication systems, we employ nearly equiangular frames as spreading sequences in s-CDMA systems to minimize interuser interference.

1.1.1 Sparse representations

In the sparse representations literature, it is common for a basis or frame to be referred to as a dictionary or overcomplete dictionary, respectively, with the dictionary elements being called atoms. A signal expansion under an overcomplete dictionary results in an underdetermined linear system of the form

$$y = Ax, \tag{1.4}$$

where $y \in \mathbb{R}^K$ is the signal of interest, $A \in \mathbb{R}^{K \times N}$, $K < N$, is a redundant dictionary, and $x \in \mathbb{R}^N$ is the vector of the unknown coefficients [58]. Due to the linear dependence between the columns of A , an important issue is the uniqueness of the representation. According to well known results, unique representations can be obtained as long as the involved dictionary is sufficiently incoherent [51]. Having more unknowns than equations, system (1.4) can be solved if we add sparsity priors, requiring x to have only a few nonvanishing coefficients. Conditions that guarantee the performance of sparse reconstruction algorithms [93, 47, 26], besides incoherence, highlight the role of tightness, requiring A to be an incoherent unit norm tight frame [125].

Although constructions of incoherent tight dictionaries appear often in signal processing applications, such dictionaries have a limited ability of sparsifying signals or are suitable only for certain signal types. In this thesis, we propose the use of incoherent unit norm tight frames in the reconstruction of sparse signals, utilizing a technique referred to as preconditioning. Preconditioning is used to transform a system into a form that

is more suitable for numerical solution [6]. Designing a $K \times K$ matrix C such that CA exhibits incoherence and tightness and employing C in (1.4) according to

$$Cy = CAx, \quad \text{or} \quad z = CAx, \quad z = Cy, \quad (1.5)$$

we obtain a system that can be solved more efficiently by the deployed algorithms. An important condition that must be taken into account when designing the preconditioner C is that (1.5) is equivalent to (1.4) if and only if C is invertible.

1.1.2 Compressed sensing

Solving an underdetermined linear system with a sparsity prior has recently received a lot of attention in compressed sensing [49, 25]. Exploiting sparsity, compressed sensing offers simultaneous acquisition and compression of signals, allowing signal reconstruction from an incomplete number of measurements. Considering a sparse signal $x \in \mathbb{R}^N$ under an orthonormal basis or redundant dictionary $A \in \mathbb{R}^{K \times N}$, $K \leq N$, we obtain m linear measurements according to

$$y = PAx, \quad (1.6)$$

using a sensing operator P realized by an $m \times K$, $m \ll K$, matrix. We refer to P as the projection or measurement matrix.

Compressed sensing leads to an underdetermined linear system with m equations and N unknowns, $m \ll N$, and, similarly to the sparse representation problem, relies on numerical methods to find a sparse solution satisfying (1.6). The system matrix is the product of the sensing operator P and the representation dictionary A ; we refer to this product as the effective dictionary. According to theoretical results from sparse representations, the effective dictionary should be an incoherent unit norm tight frame [125].

Successful signal reconstruction in compressed sensing is based on the choice of the projection matrix. Random matrices are considered a universal solution; however, the demand to increase reconstruction accuracy and reduce the necessary number of measurements has led to new theoretical and practical results [54]. A technique used to improve recovery rates in compressed sensing involves the optimization of the projection matrix over the representation dictionary A . Here, we design a projection matrix that yields an effective dictionary having the form of an incoherent unit norm tight frame. Moreover, binary projection matrices that are considered more suitable for hardware implementation may yield recovery rates similar to optimized projections, if the recovery process includes preconditioning.

1.1.3 Spreading sequences for s-CDMA

In synchronous CDMA systems, the users share the entire bandwidth and each user is distinguished from the others by its spreading sequence or code. The capacity region defined as the set of information rates at which users can transmit while retaining reliable

transmission is characterized as a function of the spreading sequences and average input power constraints of the users. Capacity optimal sequences are functions of codebook length as well as the number of users [95, 136].

Suppose that x_1, x_2, \dots, x_N is a set of vectors in \mathbb{R}^m corresponding to N possible users of an s-CDMA system. These vectors form a set of sequences of length m . Optimal spreading sequences have been characterized in [95] to be the Welch Bound Equality (WBE) sequences, that is, equal norm tight frames. WBE sequences minimize the total squared correlation (TSC), that is,

$$TSC = \sum_{i=1}^N \sum_{j=1}^N |\langle x_i, x_j \rangle|^2, \quad (1.7)$$

which results in that the interference experienced by any user is exactly the same. However, WBE sequences do not perform well when the number of users in the cell changes. If the number of the active users is smaller than N , then a code set designed for N users is no longer optimal and new codes should be assigned to all users [76].

The interference experienced by the j -th user in the system depends on the term [95]

$$\sigma(j) = \sqrt{\sum_{i \neq j} |\langle x_i, x_j \rangle|^2}. \quad (1.8)$$

Consider a system with $K < N$ active users. In [76] it was shown that all users experience the same interference, which depends only on K , the current number of active users, if and only if the code set is an equiangular sequence set.

While ETFs constitute an optimal solution for minimizing interuser interference, only a few constructions of ETFs are available. Here, we propose the employment of nearly equiangular frames as spreading sequences and improve interuser interference when the number of users in the system changes.

1.2 Contributions

The main contribution of this thesis is the development of two numerical methods for the construction of frames that are close to ETFs. The first method uses results from frame theory and linear algebra and is based on alternating and averaged projections ideas. The obtained frames are UNTFs with small column correlation, i.e., incoherent UNTFs. The second method uses theoretical results concerning the connection of frames to graphs and employs a heuristic algorithm to produce frames that are nearly equiangular, that is, the frame vectors exhibit similar near optimal correlation. The proposed numerical methods produce frames of any dimensions, which may be employed in various applications requiring ETFs.

Here, we apply the proposed constructions in signal processing applications, namely sparse representations and compressed sensing, and s-CDMA communication systems.

Concerning sparse representations under redundant dictionaries, the proposed frames are utilized in the reconstruction of sparse signals using the technique of preconditioning. Experimental results show that the performance of the deployed numerical solvers is substantially improved. In compressed sensing the proposed frame constructions are used in three ways. First, as projection matrices to acquire sparse signals, attaining high accuracy in signal reconstruction. Second, given the representation dictionary, we construct optimized projection matrices and further improve recovery rates. Third, for the first time, we apply preconditioning in compressed signal acquisition with binary operators. The technique improves the performance of numerical algorithms and is very important for practical compressed sensing applications, because binary matrices have easy hardware implementation. Another application involves the employment of nearly equiangular frames as spreading sequences in s-CDMA systems. Our simulations show that nearly equiangular frames minimize the interuser interference when the number of users in the system changes.

1.3 Outline

This thesis is organized as follows. In Chapter, 2 we review basic results from frame theory and survey important work in frame design. Chapter 3 includes the proposed methods for the construction of frames exhibiting good incoherence and spectral properties. In Chapter 4, we review important results for sparse recovery and use the proposed frames to apply preconditioning of underdetermined linear systems met in sparse representations. In Chapter 5, we address sparse recovery in compressed sensing and explain how the proposed constructions are used to produce optimized projections. We also present reconstruction of sparse signals acquired with Bernoulli projection matrices using preconditioning. The employment of the proposed nearly equiangular frames as spreading sequences in s-CDMA systems is presented in Chapter 6. Finally, Chapter 7 includes conclusions and future research directions.

CHAPTER 2

FRAMES REVIEW

2.1 Preliminaries

2.2 Finite frames basics

2.3 Connection of frames to graphs

2.4 The frame design problem

Introduced by Duffin and Schaeffer [56], frames have been known for over half a century, but they became popular due to wavelets in the late 1980s, when Daubechies, Grossman and Meyer [45, 43] showed their importance for data processing. Generalizing the notion of orthonormal bases, frames are less constrained than bases allowing for redundant (overcomplete) representations, and they are used when more flexibility in choosing a representation is needed.

Traditionally, frames are used in signal and image processing, nonharmonic Fourier series, data compression, and sampling theory [84, 85]. For example, in signal processing, frames are a flexible decomposition tool that facilitates various signal processing tasks, having the ability to capture important signal characteristics and providing numerical stability of reconstruction, resilience to additive noise and resilience to quantization [84, 22]. Finite frames play a central role in the design and analysis of both sparse representations and compressed sensing [124, 125, 9, 8, 27, 41]. Other applications of frames include source coding [43, 69], robust transmission [80, 62], Code Division Multiple Access (CDMA) systems [95, 136, 137, 140], operator theory, coding theory [110], quantum theory and quantum computing [60].

Frame theory might be regarded as partly belonging to applied harmonic analysis, functional analysis, and operator theory, as well as numerical linear algebra and matrix theory. Certain frame categories such as Grassmannian frames have connections to Grassmannian packings, spherical codes and graph theory [119]. Therefore, frame theory and its applications have experienced a growing interest among mathematicians, engineers,

computer scientists, and others. New theoretical insights and novel applications are continually arising due to the fact that the underlying principles of frame theory are basic ideas which are fundamental to a wide canon of areas of research.

2.1 Preliminaries

In this section we present basic definitions and results which we will need later.

Given a positive integer m , we denote by \mathbb{H}^m the real or complex finite Hilbert space of dimension m . This is either \mathbb{R}^m or \mathbb{C}^m . By $\langle \cdot, \cdot \rangle$ we denote the inner product and by $\| \cdot \|$ the corresponding norm. For $x = (x_1, x_2, \dots, x_m)$ and $y = (y_1, y_2, \dots, y_m)$, the inner product is defined as

$$\langle x, y \rangle = \sum_{k=1}^m x_k y_k^*. \quad (2.1)$$

Two vectors $x, y \in \mathbb{H}^m$ are called orthogonal if $\langle x, y \rangle = 0$. The norm is defined as

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{k=1}^m |x_k|^2}. \quad (2.2)$$

A vector $x \in \mathbb{H}^m$ is called normalized if $\|x\| = 1$.

Definition 2.1.1. A system $\{e_k\}_{k=1}^m$ of vectors in \mathbb{H}^m is called:

- i. Linearly independent, if for any scalars $\{a_k\}_{k=1}^m$ and provided that $e_k \neq 0$ for all $k = 1, 2, \dots, m$,

$$\sum_{k=1}^m a_k e_k = 0 \Rightarrow a_k = 0, \text{ for all } k = 1, 2, \dots, m. \quad (2.3)$$

- ii. Complete (or spanning set) if $\text{span}\{e_k\}_{k=1}^m = \mathbb{H}^m$.

- iii. Orthogonal if for all $k \neq \ell$, the vectors e_k and e_ℓ are orthogonal.

- iv. Orthonormal if it is orthogonal and each e_k is normalized.

- v. An orthonormal basis for \mathbb{H}^m if it is complete and orthonormal.

Proposition 2.1.1 (Parseval's identity). *If $\{e_k\}_{k=1}^m$ is an orthonormal basis for \mathbb{H}^m , then for every $x \in \mathbb{H}^m$, we have*

$$\|x\|^2 = \sum_{k=1}^m |\langle e_k, x \rangle|^2. \quad (2.4)$$

It follows that

Corollary 2.1.1. If $\{e_k\}_{k=1}^m$ is an orthonormal basis for \mathbb{H}^m , then for every $x \in \mathbb{H}^m$, we have

$$x = \sum_{k=1}^m \langle e_k, x \rangle e_k \quad \text{for all } x \in \mathbb{H}^m. \quad (2.5)$$

Projections

Definition 2.1.2 (Orthogonal projection). An operator $P : \mathbb{H} \rightarrow \mathbb{H}$ is called a projection, if $P^2 = P$. It is an orthogonal projection if P is also self-adjoint.

For any subspace $W \subset \mathbb{H}^m$, there is an orthogonal projection of \mathbb{H} onto W called the nearest point projection. One way to define it is to pick any orthonormal basis $\{e_k\}_{k=1}^n$, $n \leq m$, and define

$$Px = \sum_{k=1}^n \langle e_k, x \rangle e_k. \quad (2.6)$$

Theorem 2.1.3. Let P be an orthogonal projection onto a subspace W . Then

$$\|x - Px\| \leq \|x - y\| \quad \text{for all } y \in W. \quad (2.7)$$

Analysis and synthesis

Suppose x is a vector of coefficients representing data in \mathbb{H}^m . Considering a general basis F , the following equation expresses the *analysis* or *decomposition* of x under F

$$X = F^* x, \quad (2.8)$$

where $*$ denotes the Hermitian matrix. We can go back to x by

$$x = (F^*)^{-1} X, \quad (2.9)$$

which expresses the *synthesis* or *reconstruction*. If F is an orthonormal basis then $F^* = F^{-1}$, thus, $x = FX$.

2.2 Finite frames basics

Considering a real or complex m -dimensional Hilbert space \mathbb{H}^m , a sequence of $N \geq m$ vectors $\{f_k\}_{k=1}^N$, $f_k \in \mathbb{H}^m$, is a finite frame F , if there are positive constants α, β such that

$$\alpha \|x\|^2 \leq \sum_{k=1}^N |\langle f_k, x \rangle|^2 \leq \beta \|x\|^2, \quad \text{for all } x \in \mathbb{H}^m. \quad (2.10)$$

We refer to α, β as the lower and upper frame bounds, respectively.

The following notions are related to a frame $\{f_k\}_{k=1}^N$.

- (a) The ratio $\rho = N/m$ is referred to as the redundancy of the frame and is a “measure of overcompleteness” of the frame.
- (b) When $\alpha = \beta$, we say that the frame is α -tight, while when $\alpha = \beta = 1$ the frame is called Parseval.

- (c) A frame is called uniform or equal norm, when $\|f_k\| = C$, $C > 0$, for all $k \in \{1, \dots, N\}$, and unit norm, when $\|f_k\| = 1$ for all $k \in \{1, \dots, N\}$.
- (d) For a unit norm frame, the absolute value of the inner product between two frame vectors equals the cosine of the acute angle between the lines spanned by the two vectors. If there is a constant $c > 0$ for which $|\langle f_k, f_\ell \rangle| = c$, $k \neq \ell$, then the frame is called equiangular.
- (e) Any orthonormal basis is a frame with frame bounds $\alpha = \beta = 1$.

A simple example of frames is the so-called Mercedes Benz frame, the smallest redundant family in \mathbb{H}^2 with $N = 3$ vectors. It can be chosen to be a unit norm tight frame if we just select three equally spaced points on the unit circle (i.e., each 120 degrees apart). The vectors to these points from the origin is our unit norm tight frame.

We usually identify the $m \times N$ matrix $F = [f_1 \ f_2 \ \dots \ f_N]$ with columns the frame vectors $f_k \in \mathbb{H}^m$, with the frame itself. The frame bounds are then the lower and upper bounds of the quantity

$$\frac{\|F^*x\|^2}{\|x\|^2} = \frac{\langle F^*x, F^*x \rangle}{\|x\|^2} = \frac{\langle x, FF^*x \rangle}{\|x\|^2}, \quad x \neq 0. \quad (2.11)$$

These bounds are attained at the smallest and largest eigenvalues of FF^* , respectively. We also note that the frame elements span \mathbb{H}^m when $\alpha > 0$; thus, any frame of N elements in m dimensions must satisfy $N \geq m$.

2.2.1 Frame operators

The analysis, synthesis, and frame operators determine the operation of a frame when analysing and reconstructing a signal. The analysis operator—as the name suggests—analyzes a signal in terms of the frame by computing its frame coefficients.

Definition 2.2.1 (Analysis operator). Let $\{f_k\}_{k=1}^N$ be a sequence of vectors in \mathbb{H}^m . Then the associated analysis operator $T : \mathbb{H}^m \rightarrow \mathbb{H}^N$ is defined by

$$Tx := (\langle x, f_k \rangle)_{k=1}^N, \quad x \in \mathbb{H}^m. \quad (2.12)$$

Definition 2.2.2 (Synthesis operator). Let $\{f_k\}_{k=1}^N$ be a sequence of vectors in \mathbb{H}^m with associated analysis operator T . Then, the adjoint operator T^* is called the synthesis operator.

The frame operator might be considered the most important operator associated with a frame. Although it is “merely” the concatenation of the analysis and synthesis operator, it encodes crucial properties of the frame as we will see in the sequel. Moreover, it is also fundamental for the reconstruction of the signal from frame coefficients.

Definition 2.2.3 (Frame operator). Let $\{f_k\}_{k=1}^N$ be a sequence of vectors in \mathbb{H}^m with associated analysis operator T . Then the associated frame operator $S : \mathbb{H}^m \rightarrow \mathbb{H}^m$ is defined by

$$Sx := T^*Tx = \sum_{k=1}^N \langle f_k, x \rangle f_k, \quad x \in \mathbb{H}^m. \quad (2.13)$$

When $\{f_k\}_{k=1}^N$ is an orthonormal basis then $Sx = x$. The matrix representation of the frame operator $S = T^*T$ is the positive semidefinite Hermitian matrix FF^* . The most fundamental property of the frame operator is its invertibility which is crucial for the reconstruction formula.

Allowing the mapping $x \mapsto (\langle x, f_k \rangle)$ to capture the energy of any $x \in \mathbb{H}^m$, reconstruction of x is enabled with the help of some dual frame. In particular, for every frame $F = \{f_k\}_{k=1}^N$ for \mathbb{H}^m , there exists at least one dual frame $\Psi = \{\psi_k\}_{k=1}^N$ such that

$$x = \sum_{k=1}^N \langle f_k, x \rangle \psi_k, \quad \text{for all } x \in \mathbb{H}^m. \quad (2.14)$$

Any orthogonal basis is a frame with frame bounds $\alpha = \beta = 1$ and corresponds to a dual frame $\Psi = F$. The most often-used dual frame is the *canonical* dual frame, namely, the pseudoinverse $\tilde{F} = (FF^*)^{-1}F$. Computing a canonical dual involves the inversion of FF^* . As such when designing a frame it is important to retain control over the eigenvalues $\{\lambda_i\}_{i=1}^m$ of FF^* .

Of particular interest is also the operator generated by first applying the synthesis and then the analysis operator.

Definition 2.2.4 (Grammian operator). Let $\{f_k\}_{k=1}^N$ be a sequence of vectors in \mathbb{H}^m with associated analysis operator T . Then the Grammian operator $R : \mathbb{H}^N \rightarrow \mathbb{H}^N$ is defined by

$$R(a_k)_{k=1}^N = TT^*(a_k)_{k=1}^N = \left(\sum_{\ell=1}^N a_\ell \langle f_k, f_\ell \rangle \right)_{k=1}^N = \sum_{\ell=1}^N (a_\ell \langle f_k, f_\ell \rangle)_{k=1}^N. \quad (2.15)$$

The matrix representation of the Grammian of a frame is called the Gram matrix; this is the $N \times N$ matrix $R = F^*F$ given by

$$\begin{bmatrix} \|f_1\|^2 & \langle f_2, f_1 \rangle & \cdots & \langle f_N, f_1 \rangle \\ \langle f_1, f_2 \rangle & \|f_2\|^2 & \cdots & \langle f_N, f_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle f_1, f_N \rangle & \langle f_2, f_N \rangle & \cdots & \|f_N\|^2 \end{bmatrix}. \quad (2.16)$$

If the frame is unit norm then the entries of the Gram matrix are exactly the cosines of the angles between the frame vectors. The following are fundamental properties of the Gram matrix.

- i. F is an $m \times N$ frame, if and only if the Gram matrix is a self-adjoint projection with rank m .

- ii. F is an $m \times N$ Parseval frame, if and only if the Gram matrix is an orthogonal projection with rank m .
- iii. An operator U on \mathbb{H}^m is unitary, if and only if the Gram matrix of $\{Uf_k\}_{k=1}^N$ coincides with R .
- iv. The nonzero eigenvalues $\{\lambda\}_{i=1}^m$ of F^*F and FF^* are the same; thus

$$\sum_{i=1}^m \lambda_i = \text{trace}(F^*F) = \text{trace}(FF^*). \quad (2.17)$$

Frames $F = \{f_k\}_{k=1}^N$ and $G = \{g_k\}_{k=1}^N$ are unitarily equivalent, if there exists a unitary transformation $U : \mathbb{H}^m \rightarrow \mathbb{H}^m$ with $F = UG := \{Uf_k\}$, $k \in \{1, \dots, N\}$. Therefore, a frame is determined by its Gram matrix up to unitary equivalence.

2.2.2 Tight frames

Let $F = \{f_k\}_{k=1}^N$ be a finite redundant frame in \mathbb{H}^m . If (2.10) holds with $\alpha = \beta$, we have

$$x = \frac{1}{\alpha} \sum_{k=1}^N \langle f_k, x \rangle f_k, \quad \text{for all } x \in \mathbb{H}^m, \quad (2.18)$$

thus obtaining an α -tight frame. In this case, the rows of $\alpha^{-1/2}F$ form an orthogonal family, each with norm $\sqrt{\alpha}$. For an α -tight frame the following property

$$FF^* = \alpha I_m, \quad (2.19)$$

where I_m is the $m \times m$ identity matrix, follows immediately.

Constructing a tight frame is straightforward; we take an orthonormal basis and select the desired number of rows. For example, $m \times N$ harmonic tight frames are obtained by deleting $(N - m)$ rows of an $N \times N$ DFT matrix.

While (2.18) resembles the expansion formula in the case of an orthonormal basis, a tight frame does not constitute an orthonormal basis in general. Because of the linear dependence which exists among frame vectors, the expansion is no longer unique. The expansion is unique in the sense that it minimizes the norm of the expansion among all valid expansions. Because of (2.19), the canonical dual frame $\tilde{F} = (FF^*)^{-1}F$ coincides with the frame itself. Thus, tight frames provide perfect reconstruction. For this reason tight frames are desirable in redundant signal representations.

Considering the spectral properties of an α -tight frame, the following proposition summarizes well-known results.

Proposition 2.2.1 (Spectral properties of tight frames). *A frame is α -tight if and only if one of the following conditions holds:*

- (a) *The nonzero eigenvalues of the Gram matrix equal α .*

(b) The nonzero singular values of F equal $\sqrt{\alpha}$.

(c) The spectral norm of F equals $\sqrt{\alpha}$.

Even though the construction of a tight frame is trivial, we cannot easily design a tight frame with equal-norm columns; such frames exist for certain frame bounds α . For an equal norm α -tight frame with column norms $\|f_k\| = C$, $k = 1, \dots, N$, there holds

$$\text{trace}(F^*F) = \sum_{k=1}^N \|f_k\|^2 = NC^2. \quad (2.20)$$

For the m nonzero eigenvalues of the frame operator there holds

$$\text{trace}(FF^*) = \sum_{i=1}^m \lambda_i = m\alpha. \quad (2.21)$$

Thus, the frame bound is given by

$$\alpha = \frac{N}{m}C^2. \quad (2.22)$$

2.2.3 Unit norm tight frames

Finite frames that are both tight and normalized are called unit norm tight frames (UNTFs) (the term finite normalized tight frames (FNTF) is also used) and possess a significant structure. An intuitive characterization of UNTFs is presented in [69] where the authors demonstrate that if one randomly chooses unit vectors according to a uniform distribution on a sphere, the resulting Bessel sequence is asymptotically a UNTF. A UNTF can be thought of as a sequence that retains the decomposition properties of orthonormal bases while relaxing the need to be a basis. For example a concatenation of α orthonormal bases form an α -UNTF. These expansions gain redundancy and stability at the expense of not having a unique representation.

There is only one choice for the frame bound of a UNTF of N vectors for \mathbb{H}^m , which is given by the following theorem.

Theorem 2.2.5 ([14]). *If $\{f_k\}_{k=1}^N$ is a finite unit norm α -tight frame for an m -dimensional Hilbert space \mathbb{H}^m , then $\alpha = N/m$.*

Therefore, a UNTF in a finite dimensional space is an $m \times N$ matrix such that

- (a) The rows are orthogonal.
- (b) Each row has norm $\sqrt{N/m}$.
- (c) Each column has norm 1.

Another question of interest is whether UNTFs of a given N exist for a Hilbert space \mathbb{H}^m . This question is answered by the following theorem.

Theorem 2.2.6 (Existence of UNTFs [68]). *Given any m, N , with $N \geq m$, there exists a UNTF for \mathbb{H}^m of N vectors.*

Similar to Proposition 2.2.1 the spectral properties of a UNTF are given by the following proposition.

Proposition 2.2.2 (Spectral properties of UNTFs). *A frame is unit norm tight if and only if one of the following conditions holds:*

- (a) *The nonzero eigenvalues of the Gram matrix equal N/m .*
- (b) *The nonzero singular values of F equal $\sqrt{N/m}$.*
- (c) *The spectral norm of F equals $\sqrt{N/m}$.*

The value of the spectral norm of a UNTF is the lowest possible bound for $m \times N$ frames [38]. The spectral norm of an arbitrary frame is often used as a measure of how close a given frame is to a UNTF.

Unit norm tight frames are also known as Welch Bound Equality (WBE) sequences [143]. A quarter century ago Welch published a collection of lower bounds on the maximum magnitude of the inner products of a set of unit norm complex valued vectors and used these results to deduce lower bounds on the maximum magnitudes of correlation functions for sets of periodic sequences. UNTFs were found to meet the lower bounds on the mean square (RMS) magnitude, a quantity that is also known as total squared correlation (TSC). Due to this important property UNTFs are considered optimal spreading sequences for s-CDMA systems [95, 136, 137, 140]. Moreover, their robustness against additive noise and erasures allows for stable reconstruction in communications [69, 68, 31, 80].

2.2.4 Equiangular tight frames

When a unit norm tight frame has vectors forming equal angles we obtain an equiangular tight frame. ETFs exhibit equal column correlation, which is also the smallest possible [119]; thus, they are maximally incoherent equiangular frames. ETFs are arguably the most important class of finite-dimensional frames, and they are the natural choice when one tries to combine the advantages of orthonormal bases with the concept of redundancy provided by frames.

The maximal correlation between different normalized frame vectors is defined as

$$\mu(F) = \max_{\substack{1 \leq k, \ell \leq N \\ k \neq \ell}} |\langle f_k, f_\ell \rangle|, \quad (2.23)$$

and is related to a class of frames known as Grassmannian frames. A Grassmannian frame minimizes the maximal correlation between frame elements among all unit norm frames with the same redundancy.

Definition 2.2.7 (Grassmannian frames [119]). A sequence of vectors $F = \{f_k\}_{k=1}^N$ in \mathbb{H}^m is called a Grassmannian frame, if it is a solution to

$$\min \mu(F), \quad (2.24)$$

where the minimum is taken over all unit norm frames F in \mathbb{H}^m .

The minimum in (2.24) depends on the frame dimensions m, N . The following theorem derives bounds on $\mu(F)$.

Theorem 2.2.8 (Minimum maximal correlation [119]). *Let $F = \{f_k\}_{k=1}^N$ be a frame in \mathbb{H}^m . Then*

$$\mu(F) \geq \sqrt{\frac{N-m}{m(N-1)}}. \quad (2.25)$$

Equality holds, if and only if F is an equiangular tight frame. Furthermore,

$$\begin{aligned} \text{if } \mathbb{H} = \mathbb{R} \quad & \text{equality in (2.25) can only hold if } N \leq \frac{m(m+1)}{2}, \\ \text{if } \mathbb{H} = \mathbb{C} \quad & \text{equality in (2.25) can only hold if } N \leq m^2. \end{aligned} \quad (2.26)$$

In [119] it was shown that the bound in (2.25) is attained by Grassmannian frames that also form unit norm tight frames. These frames are referred to as optimal Grassmannian frames and coincide with equiangular tight frames. As unit norm tight frames with dimensions m, N exist for a specific tightness parameter ($\alpha = N/m$), an optimal Grassmannian frame is an equiangular N/m -tight frame. Therefore, an equiangular tight frame $F = \{f_k\}_{k=1}^N$ in \mathbb{H}^m satisfies the following conditions:

$$\|f_k\| = 1 \quad \text{for } k = 1, \dots, N, \quad (2.27)$$

$$|\langle f_k, f_\ell \rangle| = \sqrt{\frac{N-m}{m(N-1)}} \quad \text{for } k \neq \ell, \quad (2.28)$$

$$\frac{N}{m} \sum_{k=1}^N \langle x, f_k \rangle f_k = x \quad \text{for all } x \in \mathbb{H}^m. \quad (2.29)$$

The lowest bound on the minimal achievable correlation for equiangular frames is also known as *Welch bound* [143], and optimal Grassmannian frames or ETFs are also referred to as *Maximal Welch Bound Equality sequences (MWBE)*.

Equiangular tight frames were introduced by van Lint and Seidel in the setting of discrete geometry [134]. ETFs are particularly interesting and useful. In signal processing, ETFs meet the Welch bound for optimal codes [76]. As spreading sequences in multiuser communication systems the tightness condition allows equiangular tight frames to achieve maximal capacity of a Gaussian channel and their equiangularity allows them to satisfy an interference invariance property [76]. In sparse representations and compressed sensing they are of interest due to their incoherence. ETFs have also been proposed for robust transmission [80, 62].

Despite their important properties and their numerous practical applications, there is no explicit way of constructing ETFs. This problem is connected with other important problems such as packings in Grassmannian spaces and antipodal spherical codes. It has also connections to graph theory, equiangular line sets and coding theory. The techniques reported in [126, 119, 144, 141, 64] construct a few of existent frames. A technique proposed in [119] relies on the connection of frames to graphs and will be discussed next.

2.3 Connection of frames to graphs

Graphs with a lot of structure and symmetry play a central role in graph theory. Different kinds of matrices are used to represent a graph, such as the Laplace matrix or adjacency matrices [21]. What structural properties can be derived from the eigenvalues depends on the specific matrix that is used. The Seidel adjacency matrix Q of a graph Γ is given by

$$Q = \begin{cases} -1 & \text{if the vertices } x, y \in \Gamma \text{ are adjacent,} \\ 1 & \text{if the vertices } x, y \in \Gamma \text{ are nonadjacent,} \\ 0 & \text{if } x = y. \end{cases} \quad (2.30)$$

If Q has only a few distinct eigenvalues, then the graph is strongly regular.

Studies concerning the connection of frames with graphs have shown that the existence of an ETF in a real Hilbert space depends on the existence of a matrix Q with zero diagonal and ± 1 's off-diagonal entries. This matrix corresponds to the adjacency matrix of a special type of strongly regular graphs [119]. From [119, 80] we quote the following definition.

Definition 2.3.1. Given an $m \times N$ equiangular tight frame $F = [f_1 \ f_2 \ \dots \ f_N]$, the Gram matrix can be written in the form

$$R = I + cQ, \quad (2.31)$$

where I is the $N \times N$ identity matrix and c is the Welch bound given by (2.25). The $N \times N$ matrix Q is called the *signature matrix* of the frame F .

The main results about signature matrices are summarized in the following theorem.

Theorem 2.3.2 ([80]). *Let Q be a self-adjoint $N \times N$ matrix, with $q_{i,i} = 0$ for all i and $|q_{i,j}| = 1$ for all $i \neq j$. Then the following are equivalent:*

- i. Q is the signature matrix of an $m \times N$ ETF.*
- ii. $Q^2 = (N - 1)I + \nu Q$ for some necessarily real number ν .*
- iii. Q has exactly two distinct eigenvalues, denoted as $\lambda_1 < \lambda_2$.*

When any of the above conditions hold, the parameters $m, N, \nu, \lambda_1, \lambda_2$ satisfy certain relations [17], implying that for many values of m, N ETFs do not exist. It can be shown that [141]

$$\begin{aligned}\lambda_1 &= -\sqrt{\frac{m(N-1)}{N-m}}, \text{ with multiplicity } N-m, \\ \lambda_2 &= \sqrt{\frac{(N-m)(N-1)}{m}}, \text{ with multiplicity } m.\end{aligned}\tag{2.32}$$

According to [80], there are finitely many possible $N \times N$ signature matrices and finitely many real equiangular frames of N vectors. For more details about the connection between graphs and frames the reader is referred to [119, 80, 17, 141, 18].

Based on the construction of *conference matrices* proposed in [86] and relying on the above results, the authors of [119] proposed the construction of ETFs of size $m \times 2m$. Conference matrices are $N \times N$ matrices with zeros along the diagonal and ± 1 of diagonal entries, satisfying $CC^T = (N-1)I_N$, and play an important role in graph theory [86], [115]. Conference matrices exist for $N = p^\alpha + 1$, where p is an odd prime number and $\alpha \in \mathbb{N}$, and can be constructed explicitly [67, 102]. According to [119], if C^{2m} is a symmetric conference matrix, then there exist $2m$ vectors in \mathbb{R}^m such that the bound (2.25) holds with equality for $N = 2m$. In this case the bound becomes $c = 1/\sqrt{2m-1}$ and the Gram matrix is obtained according to (2.31), having off-diagonal entries equal to $\pm 1/\sqrt{2m-1}$.

2.4 The frame design problem

When designing a frame, the design specifications arise from the application of interest. As a result, there exist a large number of construction methods, as diverse as the applications requiring a frame. Usually, the constructions that come to address specific requirements are difficult to generalize to solve different types of frame design problems. On the other hand, more general constructions coming from the frame community often impose certain restrictions on frame dimensions.

We have seen that properties such as unit normness, tightness and equiangularity define certain classes of frames and play a significant role in applications. Therefore, when design specifications are set they include

- (a) prescribed vector norms,
- (b) prescribed spectral properties,
- (c) correlation constraints such as equiangularity or incoherence.

Considering the construction of a tight frame, it is easy to obtain such a frame by selecting the desired number of rows from an $N \times N$ orthonormal basis. However, most applications require that the vectors comprising the frame have some additional structure. For example, tight frames with prescribed norms, or most required UNTFs, are difficult

to construct, as row orthogonality opposes column unit normness. The design difficulties become stronger when trying to address the main drawback of frames, that is, the correlation between the frame elements. Tightness implies certain restrictions on singular values and singular vectors which combat either column normalization or the requirement for constant inner products between columns [126].

According to [126] finite-dimensional frame design is an algebraic problem. Frame design aims at producing a structured matrix with certain spectral properties, a problem that may require the use of discrete and combinatorial mathematics. Sarwate’s survey paper [112] about tight frames includes constructions of unit-norm frames with methods that have employed algebraic techniques. The last few years, some essentially algebraic algorithms have been proposed that can construct tight frames with nonconstant vector norms [136, 32, 127]. The frames proposed in [136] and [127] were designed with the s-CDMA application in mind, while [32] comes from the frame community.

Algebraic and combinatoric tools are not always effective. In these situations, numerical methods can help to produce constructions with properties that approximate the desired theoretical specifications. Moreover, numerical methods can help researchers develop the insight necessary for completing an algebraic construction. However, the literature does not offer many numerical approaches to frame design.

Regarding the construction of UNTFs, most algorithms provide frames to be used as spreading sequences in s-CDMA systems. This application prompted a long series of papers [132, 109, 133, 3] that describe iterative methods for constructing tight frames with prescribed column norms. Besides spectral and structural properties frames designed for s-CDMA systems may also apply restrictions on the employed alphabet. It is not clear how one could generalize these methods to solve different types of frame design problems.

More recent methods providing general UNTF constructions modify a given frame so that the result is a tight frame. Three techniques are known to belong to this category. In [19], the authors start from a tight frame and approach a UNTF by solving a differential equation. In [29], the authors start from a unit norm frame and increase the degree of tightness using a gradient-descent-based algorithm. Relative primeness of m and N is a condition assumed by both techniques, though in [29] in a weaker sense. The work of [30] comes from the frame community. “Spectral tetris” presented in [30] has the drawback that it often generates multiple copies of the same frame vector.

Regarding the construction of equiangular tight frames, it is known that these frames exist for certain frame dimensions [121] and most existing constructions [126, 119, 144, 107, 141, 64] impose additional restrictions. A survey on known ETFs can be found in [63]. As we have already mentioned, this problem is connected with other important problems such as equiangular line sets and it has been addressed for over 60 years. The problem of constructing any number (especially, the maximal number) of equiangular lines in \mathbb{R}^m is one of the most elementary and at the same time one of the most difficult problems in mathematics. After sixty years of research, we do not know the answer for all dimensions $m \leq 20$ in either the real or complex case.

Recently, the construction of equiangular tight frames has gained the interest of the sparse modelling community, as ETFs are maximally incoherent. Due to new theoretical results in sparse representations and compressed sensing, there is a growing interest for incoherent unit norm tight frames. The few numerical methods that are available in the literature [57, 145, 82] focus on incoherence rather than on spectral properties. Clearly, this is an open research topic.

CHAPTER 3

CONSTRUCTION OF APPROXIMATELY EQUIANGULAR TIGHT FRAMES

3.1 Alternating projections

3.2 Averaged projections

3.3 Construction of incoherent unit norm tight frames

3.4 Construction of nearly equiangular frames

3.5 Comparison of the proposed constructions

The research presented in this thesis is motivated by recent theoretical and practical results formulated in sparse representations and compressed sensing, which highlight the important role of incoherent unit norm tight frames in sparse recovery. Considering that optimal values of incoherence and tightness are observed in equiangular tight frames (ETFs), the frame community aims at perfect ETF constructions. Here, we focus on the improvement of practical applications and propose two methods for the construction of real frames as close to ETFs as possible. We perceive nearness to ETFs by means of mutual coherence and spectral norm and design frames with unit norm vectors that exhibit small mutual coherence and are almost or exactly tight.

The first of the methods developed here is inspired by an algorithm for designing incoherent matrices for compressed sensing. In [57], Elad argued that an optimized projection matrix would be a matrix that reduces the mutual coherence of the effective dictionary involved in sparse recovery and proposed a heuristic algorithm for its construction. Most of the existing work for optimized projections relies on [57] and aims at reducing the mutual coherence. The method developed here introduces, for the first time, the tightness parameter in incoherent matrix design, producing unit norm tight frames with remarkably low incoherence levels.

The second method developed in this thesis is based on the following observation. Studying the properties of the proposed incoherent tight frames, we noticed that these frames have “signature” matrices with eigenvalues approximating the spectrum of a signature matrix of an ETF. Recall that the signature matrix of a real ETF is a matrix with zero diagonal entries, ± 1 off-diagonal entries, and spectrum consisting of two distinct eigenvalues, and defines ETFs up to unitary equivalence. Here, we develop an algorithm for the construction of a matrix satisfying the structural constraints and approximating the spectral constraints of a signature matrix of an ETF. Employing this matrix as a “signature” matrix, we produce frames that are close to ETFs. The most significant property of these frames is that they are nearly equiangular, meaning that the frame vectors form similar angles that are close to the optimal value. This property makes these frames suitable for use in s-CDMA systems as spreading sequences.

Considering the design difficulties when constructing ETFs, the constraints implied by existing constructions and the restrictions coming of frame theory regarding the frame dimensions, the most important characteristic of the proposed algorithms is probably that they can produce frames of any size. Thus, they can provide solutions in many signal processing problems as well as in other applications requiring ETFs.

Both methods proposed here utilize ideas from alternating and averaged projections. However, introducing the tightness parameter in frame design, we actually focus on matrices with certain spectral requirements. Projecting onto spectral sets, that is, sets of matrices defined via properties of their eigenvalues, is an important obstacle the proposed algorithms must face. The spectral sets are not convex, therefore, they do not admit unique projections. We start with a short presentation of alternating and averaged projections, and discuss how these problems could be addressed.

3.1 Alternating projections

Alternating projections [139] is a very simple algorithm for computing a point in the intersection of some convex sets, using a sequence of projections onto the sets. Suppose S and W are closed convex sets in \mathbb{R}^N , and let \mathcal{P}_S and \mathcal{P}_W denote the projection on S and W , respectively. The algorithm starts with any $x_0 \in S$, and then alternately projects onto S and W :

$$y_k = \mathcal{P}_W(x_k), \quad x_{k+1} = \mathcal{P}_S(y_k), \quad k = 0, 1, 2, \dots \quad (3.1)$$

This generates a sequence of points $x_k \in S$ and $y_k \in W$. If S and W are not disjoint, then the sequences x_k and y_k both converge to a point $x \in S \cap W$ [37]. Alternating projections computes a point in the intersection of the sets, provided they intersect. The algorithm does not necessarily produce a point in $x \in S \cap W$ in a finite number of steps, but the sequence x_k (which lies in S) satisfies $\text{dist}(x_k, W) \rightarrow 0$, and likewise for y_k .

Alternating projections is also useful when the sets do not intersect. In this case the following holds. Assume the distance between S and W is attained (i.e., there exist points

in S and W whose distance is $\text{dist}(S, W)$). Then $x_k \rightarrow x^* \in S$, and $y_k \rightarrow y^* \in W$, where $\|x^* - y^*\| = \text{dist}(S, W)$. In other words, alternating projections yields a pair of points in S and W that have minimum distance.

There are many variations and extensions of the basic alternating projections algorithm. For example, we can find a point in the intersection of $k > 2$ convex sets, by projecting onto S_1 , then onto S_2, \dots , and finally onto S_k , and then repeating the cycle of k projections. This is called the sequential or cyclic projection algorithm, instead of alternating projection.

Alternating projections is very popular because of its simplicity and intuitive appeal (see survey article [12]). The method can be slow, but it can be useful when we have some efficient method, such as an analytical formula, for carrying out the projections. Convergence of alternating projections on convex sets has been well studied; however, only a few recent extensions of alternating projections consider the case of nonconvex sets [88], [87].

3.1.1 Alternating projections on nonconvex sets

Iterated projection algorithms and analogous heuristics have been successfully applied in many nonconvex problems, in areas such as inverse eigenvalue problems [35, 36], information theory [126], image processing [142, 13], and more. While alternating projections is quite popular in practice, theoretical understanding is still poor. An important subproblem one must solve in the nonconvex case is that the projection mapping can no longer be single-valued and may be hard to compute. However, the projection problem for some nonconvex sets is relatively easy and computationally inexpensive [88]. Convergence results on nonconvex alternating projection algorithms have been uncommon, and have either focused on a very special case [36], or have been much weaker than for the convex case [42, 126].

The only general convergence study is the work of [88, 87]. In [88] the authors study alternating projections on manifolds and prove local convergence at a linear rate. A more recent publication [87] considers alternating projections on two nonconvex sets, one of which is assumed to be suitably “regular”; the term refers to convex sets, smooth manifolds or feasible regions satisfying the Mangasarian-Fromovitz constraint qualification. The authors show that the method converges locally to a point in the intersection at a linear rate. The convergence of alternating projections on more than two sets, some of which are nonconvex, is still an open problem.

3.2 Averaged projections

Averaged projections is a simple variation of alternating projections. At every step of averaged projections, we project the current iteration onto every set and average the results to obtain the value for the next iteration. We start with $x_0 \in S$ and $y_0 \in W$, we

form the average, $z_0 = (x_0 + y_0)/2$, and set $x_1 = \mathcal{P}_S(z_0)$ and $y_1 = \mathcal{P}_W(z_0)$. Then, we repeat

$$z_k = (x_k + y_k)/2, \quad x_{k+1} = \mathcal{P}_S(z_k), \quad y_{k+1} = \mathcal{P}_W(z_k), \quad k = 1, 2, \dots \quad (3.2)$$

Global convergence of this method in the case of two closed convex sets was proved in [5].

Similar to alternating projections, the method of averaged projections might appear hard to implement on concrete nonconvex problems. The only work analysing convergence of averaged projections for nonconvex sets is the work of [87]. According to [87], studying the convergence of iterative algorithms for nonconvex minimization problems must be equipped with a local theory.

Local linear convergence requires good geometric properties such as convexity, smoothness, or “prox-regularity”. *Prox-regular* sets is a large class of sets that admit unique projections locally. It is known [88] that convex sets and smooth manifolds (see Appendix A) belong to this category. Considering averaged projections on several prox-regular sets, the authors of [87] assert that the method converges locally at a linear rate to a point in the intersection as long as the intersection satisfies some properties.

3.2.1 Convergence for averaged projections on prox-regular sets

The crucial idea behind the convergence analysis presented in [87] is the notion of *strongly regular intersection*. The main result in [87] states that when several prox-regular sets have *strongly regular intersection* at some point, the method converges locally at a linear rate to a point in the intersection. Strongly regular intersection is important to prevent the algorithm from projecting near a *locally extremal point*. The notion of a locally extremal point in the intersection of some sets is the following: if we restrict to a neighborhood of such a point and then translate the sets by small distances, their intersection may render empty. Therefore, not choosing a locally extremal point as initial point in a projections algorithm is a critical hypothesis for convergence.

In order to make clear that strong regularity implies local extremality, we present here the relevant definitions for the case of two sets. For more details the reader is referred to [87].

Definition 3.2.1 (Locally extremal point [87]). Denoting by \mathbb{E} the Euclidean space, consider two sets $H, G \subset \mathbb{E}$. A point $\bar{x} \in H \cap G$ is locally extremal for this pair of sets, if there exists a positive ρ and a sequence of vectors $z_r \rightarrow 0$ in \mathbb{E} such that

$$(H + z_r) \cap G \cap B_\rho(\bar{x}) = \emptyset, \text{ for all } r = 1, 2, \dots,$$

where $B_\rho(\bar{x})$ is the closed ball of radius ρ centered at \bar{x} . Clearly \bar{x} is not locally extremal, if and only if

$$0 \in \text{int}(((H - \bar{x}) \cap \rho B) - (G - \bar{x}) \cap \rho B)), \text{ for all } \rho > 0,$$

where B is the closed unit ball in \mathbb{E} .

Definition 3.2.2 (Strongly regular intersection). Two sets $H, G \subset \mathbb{E}$ have strongly regular intersection at a point $\bar{x} \in H \cap G$ if there exists a constant $\alpha > 0$ such that

$$\alpha \rho B \subset ((H - x) \cap \rho B) - ((G - z) \cap \rho B)$$

for all $x \in H$ near \bar{x} and $z \in G$ near \bar{x} .

By considering the case $x = z = \bar{x}$, we see that strongly regular intersection at a point \bar{x} implies that \bar{x} is not locally extremal. Conversely, finding a point in the intersection of the involved sets that is not locally extremal, implies that the sets have strongly regular intersection at this point.

Now, we can summarize the results of [87] regarding averaged projections.

Theorem 3.2.3. *Consider prox-regular sets $H_1, H_2, \dots, H_L \subset \mathbb{E}$ having strongly regular intersection at a point $\bar{x} \in \cap H_i$, and any constant $k > \text{cond}(H_1, H_2, \dots, H_L | \bar{x})$. Then, starting from any point near \bar{x} , one iteration of the method of averaged projections reduces the mean squared distance*

$$D = \frac{1}{2L} \sum_{i=1}^L d_{H_i}^2$$

by a factor of at least $1 - \frac{1}{k^2 L}$.

The *condition modulus* $\text{cond}(H_1, H_2, \dots, H_L | \bar{x})$ is a positive constant that quantifies strong regularity [87]. The distance d_{H_i} between the current iteration x and the set H_i we project on is defined as $d_{H_i} = \inf\{\|x - X\|_{\mathcal{F}} : X \in H_i\}$, with $\|\cdot\|_{\mathcal{F}}$ denoting the Frobenius norm.

3.3 Construction of incoherent unit norm tight frames

When aiming at minimization of the correlation of a matrix, a common strategy is to work with the Gram matrix. Recall that given an $m \times N$ matrix F , with columns $F = [f_1 \ f_2 \ \dots \ f_N]$, the Gram matrix is the $N \times N$ matrix $G = F^T F$, with the (i, j) entry of G being the correlation between the i -th and the j -th column of F , that is, $g_{ij} = \langle f_i, f_j \rangle$. Reducing column correlation of F is equivalent to applying a “shrinkage” operation on the off-diagonal entries of the Gram matrix. The first method we propose here for the construction of incoherent UNTFs is inspired by the work presented in [57]. In [57], Elad proposed an algorithm for the construction of incoherent matrices, which were used to obtain optimized projection matrices for compressed sensing. In compressed sensing, F stands for the effective dictionary employed in sparse recovery, which comes of the product of the projection matrix P and the sparsifying dictionary D , $F = PD$. In order to minimize the correlation between the columns of F , Elad proposed the following operation

$$\hat{g}_{ij} = \begin{cases} \gamma g_{ij}, & |g_{ij}| \geq t, \\ \gamma t \cdot \text{sgn}(g_{ij}), & t > |g_{ij}| \geq \gamma t, \\ g_{ij}, & \gamma t > |g_{ij}|, \end{cases} \quad (3.3)$$

where γ and t are appropriate scalars. Indeed, the $m \times N$ matrix obtained by the “shrunk” Gram exhibits improved mutual coherence, resulting in higher reconstruction accuracy when used in compressed sensing. Notice that, having computed an incoherent matrix F , optimized projections P are obtained by solving the least squares problem $\min_P \|F - PD\|_2^2$.

Considering the important role of incoherence in sparse signal recovery, many authors have argued that ETFs are ideal candidates for these problems as these frames exhibit the lowest possible mutual coherence. However, very few results concern the employment of ETFs in compressed sensing and the main reason for this are the difficulties in their construction.

The method presented here aims at the construction of frames that are as close to ETFs as possible. The proposed construction strategy is based on the observation that ETFs not only exhibit minimal mutual coherence, but N/m -tightness as well. Thus, we proposed in [128] the following design methodology: Suppose we compute a matrix with small mutual coherence. Then, the problem of approximating an ETF reduces to finding a UNTF that is nearest to the computed incoherent matrix, in Frobenius norm. Computing a UNTF that is nearest to a given matrix, is a matrix nearness problem, which can be solved algebraically by employing the following algebraic theorem.

Theorem 3.3.1 (Nearest tight frame [126, 81]). *Given a matrix $F \in \mathbb{R}^{m \times N}$, $N \geq m$, suppose F has singular value decomposition (SVD) $U\Sigma V^T$. With respect to the Frobenius norm, a nearest α -tight frame F' to F is given by $\sqrt{\alpha} \cdot UV^T$. Assume, in addition, that F has full row-rank. Then $\sqrt{\alpha} \cdot UV^T$ is the unique α -tight frame closest to F . Moreover, one may compute UV^T using the formula $(FF^T)^{-1/2}F$.*

The proposed design methodology is alternating between tightness and incoherence. The algorithm presented in [128] is a preliminary result of our work and utilizes the “shrinkage” operation proposed by Elad to improve incoherence, and Theorem 3.3.1 to improve tightness. Changing the “shrinkage” operation according to

$$\hat{g}_{ij} = \begin{cases} \operatorname{sgn}(g_{ij}) \cdot (1/\sqrt{m}), & \text{if } 1/\sqrt{m} < |g_{ij}| < 1, \\ g_{ij}, & \text{otherwise,} \end{cases} \quad (3.4)$$

we obtain the algorithms presented in [129], which provide a better formulation and a clearer insight of the process described in [128]. The presented construction strategy is implemented utilizing alternating and averaged projections.

3.3.1 Algorithm 1

The first algorithm starts from an arbitrary $m \times N$ matrix that has full rank and sequentially applies (3.4) and Theorem 3.3.1. The “shrinkage” process reduces the matrix mutual coherence, while Theorem 3.3.1 finds an N/m -tight frame that is nearest to the incoherent matrix. The selected bound $1/\sqrt{m}$ is approximately equal to the lowest possible bound (see eq. (2.25)) for large values of N . Other choices of the bound might perform

better depending on the frame dimensions. Combined with Theorem 3.3.1, the proposed Gram matrix processing yields highly incoherent UNTFs.

Algorithm 1 Construction of incoherent UNTFs with Alternating Projections

Input: $m \times N$ frame F_0 , iterations $ITER$

Output: $m \times N$ incoherent UNTF F_{q+1}

for $q := 1$ to $ITER$ **do**

$\hat{F}_q = \text{normc}(F_q)$ // column normalization

$G_q = \hat{F}_q^T \hat{F}_q$ // obtain the Gram matrix

for $i := 1$ to N **do**

for $j := 1$ to N **do**

$\hat{g}_{ij} = g_{ij}$

if $i \neq j$ **then**

if $|g_{ij}| > 1/\sqrt{d}$ **then**

$\hat{g}_{ij} = \text{sgn}(g_{ij})(1/\sqrt{d})$ // apply (3.4) to bound off-diagonal entries

$[U, \Sigma, V] = \text{svd}(\tilde{G}_q)$

$\Sigma = \Sigma(1:m, 1:m)$

$U = U(1:m, 1:m)$

$V = V(1:m, 1:m)$

$\check{G} = U\Sigma V$ // Reduce the rank of \tilde{G}_q to m

$\check{G} = \text{diag}(1./\text{sqrt}(\text{diag}(\check{G}))) \cdot \check{G} \cdot \text{diag}(1./\text{sqrt}(\text{diag}(\check{G})))$ // normalize the Gram matrix

$[U, \Sigma, V] = \text{svd}(\check{G})$ // $U = V$

$S_q = \text{sqrt}(\Sigma)V^T$ // Obtain $S_q \in \mathbb{R}^{m \times N}$ such that $S_q^T S_q = \check{G}_q$

$S'_q = \sqrt{N/m} \cdot (S_q S_q^T)^{-1/2} S_q$ // Find the nearest N/m -tight frame

$F_{q+1} = S'_q$

The algorithm we propose is iterative. We employ as initial matrix F_0 a tight frame nearest to a random Gaussian matrix. In the q -th iteration, the process that reduces the mutual coherence involves “shrinkage” operations on the Gram matrix G_q ; thus, a column normalization step precedes the main steps of our method. After applying (3.4), the modified Gram matrix \tilde{G}_q may have rank larger than m . We obtain the nearest m -rank Gram matrix using SVD. Decomposing the new Gram matrix \check{G}_q , we obtain the incoherent matrix S_q such that $S_q^T S_q = \check{G}_q$. Next, Theorem 3.3.1 is applied to S_q to obtain an incoherent tight frame. Therefore, the q -th iteration of Algorithm 1 involves the following:

1. Obtain the matrix \hat{F}_q , after column normalization of F_q .
2. Calculate the Gram matrix $\hat{G}_q = \hat{F}_q^T \hat{F}_q$ and apply (3.4) to bound the absolute values of the off-diagonal entries, producing \tilde{G}_q .
3. Apply SVD to \tilde{G}_q to force the matrix rank to be equal to m , obtaining \check{G}_q .
4. A matrix $S_q \in \mathbb{R}^{m \times N}$ is obtained such that $S_q^T S_q = \check{G}_q$.

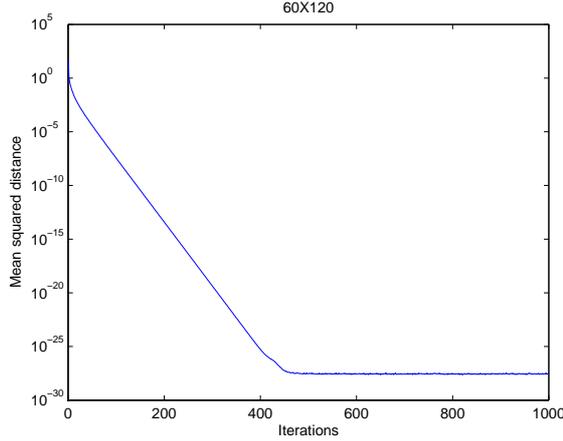


Figure 3.1: Convergence of Algorithm 1 (alternating projections) for a 60×120 matrix. The mean squared distance between the current iteration and the sets we project on reduces in a linear rate.

5. Find S'_q , the nearest N/m -tight frame to S_q , according to $S'_q = \sqrt{N/m} \cdot (S_q S_q^T)^{-1/2} S_q$.
Set $F_{q+1} = S_q$.

3.3.2 Convergence of Algorithm 1

The proposed algorithm is actually an alternating projections algorithm. More particularly, the proposed algorithm projects onto the following sets:

1. The set Y of $N \times N$ Gram matrices of $m \times N$ unit norm frames,

$$Y = \{G \in \mathbb{R}^{N \times N} : G = G^*, g_{ii} = 1, i = 1, \dots, N\}.$$

2. The set Z of $N \times N$ symmetric matrices with bounded off-diagonal entries,

$$Z = \{G \in \mathbb{R}^{N \times N} : G = G^*, |g_{ij}| \leq 1/\sqrt{m}, i \neq j, i, j = 1, \dots, N\}.$$

3. The set W of rank- m , $N \times N$ symmetric matrices,

$$W = \{G \in \mathbb{R}^{N \times N} : G = G^*, \text{rank}(G) = m\}.$$

4. The set S of $N \times N$ Gram matrices of $m \times N$ α -tight frames,

$$S = \{G \in \mathbb{R}^{N \times N} : G = G^*, \text{ with only } m \text{ nonzero eigenvalues, all equal to } \alpha\}.$$

As we have already mentioned, alternating projections has been well studied for closed convex sets. However, from the above sets only Y and Z are convex, whereas W and S are smooth manifolds [88]. Therefore, our discussion regarding convergence of Algorithm

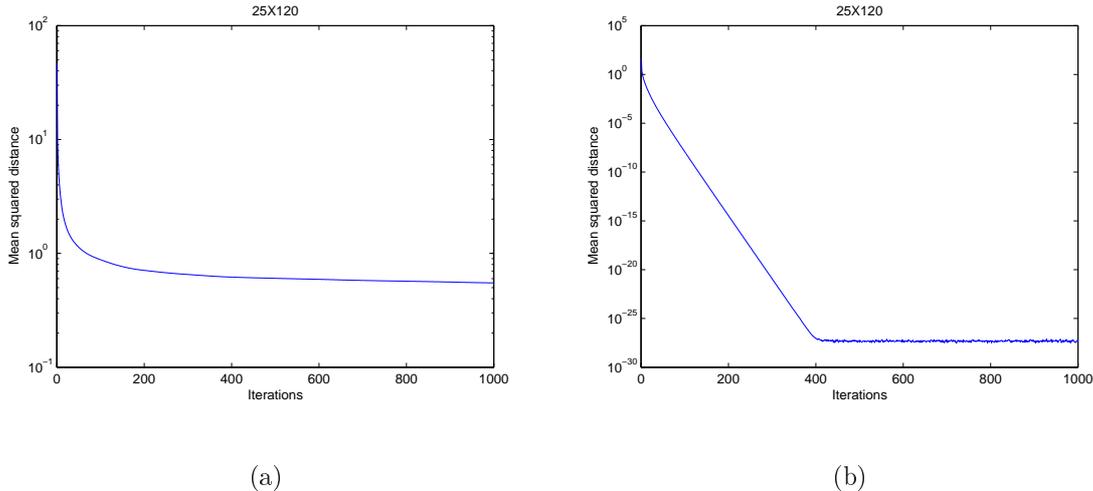


Figure 3.2: Convergence of Algorithm 1 (alternating projections) for a 25×120 matrix. The convergence rate depends on the bound used in eq. (3.4). In (a) we observe a sub-linear convergence rate when the bound equals $1/\sqrt{m}$. In (b) the convergence rate becomes linear as the bound is relaxed to $3/2\sqrt{m}$.

1 is mainly based on numerical results. To illustrate convergence, we need to define the mean squared distance of the current iteration from the sets involved in the projections, that is

$$D(q) = \frac{1}{8}(d^2(G_q, Y) + d^2(G_q, Z) + d^2(G_q, W) + d^2(G_q, S)),$$

where the distance $d(G_q, H)$ between the current iteration G_q and the set H we project on is defined as $d(G_q, H) = d_H = \inf\{\|G_q - X\|_{\mathcal{F}} : X \in H\}$.

In figures 3.1 and 3.2 we display $\log_{10} D(q)$ when Algorithm 1 is applied to a 60×120 and a 25×120 matrix, respectively. Figure 3.1 shows that the proposed algorithm converges at a linear rate, constructing a frame that belongs to the intersection of the involved sets. The zeroing of the mean squared distance implies that the produced frame is indeed an incoherent UNTF. When the frame redundancy increases, the numerical results become a little different. Figure 3.2(a) shows that the convergence rate for a 25×120 frame is sub-linear and the produced frame does not belong to the intersection of the involved sets. Considering the increased difficulties of constructing incoherent frames of high redundancy, this result is not surprising; it is possible that either the intersection is empty or it has properties that bring on difficulties to the proposed algorithm. Experiments performed with a relaxed incoherence level, which is determined by the bound $1/\sqrt{m}$ in eq. (3.4) confirm our conjecture. A relaxed bound yields a broader set Z and increases the probability that the intersection has good properties. Figure 3.2(b) illustrates convergence of Algorithm 1 when the bound $1/\sqrt{m}$ in eq. (3.4) is replaced by $3/2\sqrt{m}$. We can see that the convergence rate becomes linear and the produced matrix belongs to the intersection of the involved sets.

3.3.3 Algorithm 2

Algorithm 2 Construction of incoherent UNTFs with Averaged Projections

Input: $N \times N$ initial Gram matrix G_0 , iterations $ITER$

Output: $m \times N$ incoherent UNTF F_{out}

for $q := 1$ to $ITER$ **do**

$\mathcal{P}_Y(G_q) = \text{diag}(1./\text{sqrt}(\text{diag}(G_q))) \cdot G_q \cdot \text{diag}(1./\text{sqrt}(\text{diag}(G_q)))$ // Normalize the Gram matrix

for $i := 1$ to N **do**

for $j := 1$ to N **do**

$\hat{g}_{ij} = g_{ij}$

if $i \neq j$ **then**

if $|g_{ij}| > 1/\sqrt{d}$ **then**

$\hat{g}_{ij} = \text{sgn}(g_{ij})(1/\sqrt{d})$ // Apply (3.4) to bound the Gram entries

$\mathcal{P}_Z(G_q) = \{\hat{g}_{ij}\}$

$[U, \Sigma, V] = \text{svd}(G_q)$

$\Sigma = \Sigma(1:m, 1:m)$

$U = U(1:m, 1:m)$

$V = V(1:m, 1:m)$

$\mathcal{P}_W(G_q) = U\Sigma V$ // Reduce the rank of G_q to m

$\alpha = N/m$

$\Sigma' = \text{diag}\{\underbrace{\alpha \ \alpha \ \dots \ \alpha}_{N-m}\}$

$\mathcal{P}_S(G_q) = U\Sigma'U^T$ // Symmetric matrix with m eigenvalues all equal to N/m

$G_{q+1} = \frac{1}{4}(\mathcal{P}_Y(G_q) + \mathcal{P}_Z(G_q) + \mathcal{P}_W(G_q) + \mathcal{P}_S(G_q))$ // Apply (3.5)

$[U, \Sigma, V] = \text{svd}(G_{q+1})$ // $U = V$

$\Sigma = \Sigma(1:m, 1:m)$

$U = U(1:m, 1:m)$

$V = V(1:m, 1:m)$

$F_{\text{out}} = \text{sqrt}(\Sigma)V^T$

Considering the difficulties in studying alternating projections on nonconvex sets, we propose here a similar algorithm for the construction of incoherent UNTFs that relies on averaged projections. Suppose G_0 is the initial Gram matrix. We consider the following projections: $\mathcal{P}_Y(G_0)$ the projection onto the set of $N \times N$ symmetric matrices with unit diagonal, $\mathcal{P}_Z(G_0)$ the projection onto the set of $N \times N$ symmetric matrices with bounded off-diagonal entries, $\mathcal{P}_W(G_0)$ the projection onto the set of rank- m $N \times N$ symmetric matrices, $\mathcal{P}_S(G_0)$ the projection onto the set of $N \times N$ symmetric matrices with m nonzero eigenvalues equal to N/m . If G_q is the Gram matrix calculated in the q -th iteration, then a modified version of Algorithm 1 would consider as input in the $(q + 1)$ -th iteration the

average

$$G_{q+1} = \frac{1}{4}(\mathcal{P}_Y(G_q) + \mathcal{P}_Z(G_q) + \mathcal{P}_W(G_q) + \mathcal{P}_S(G_q)). \quad (3.5)$$

The projection $\mathcal{P}_S(G_q)$ can be calculated using Theorem A.2 given in the Appendix A. If $G_q = U\Sigma U^T$ is the eigenvalue decomposition of the symmetric matrix obtained in the q -th iteration, then $\mathcal{P}_S(G_q) = U\Sigma'U^T$ with Σ' being a diagonal matrix with m entries equal to N/m and the rest zero.

Again we start from a random Gaussian matrix and apply Theorem 3.3.1 to obtain a nearest tight frame F_0 ; then we calculate the Gram matrix $G_0 = F_0^T F_0$. In the q -th iteration we execute the following steps:

1. Normalize the Gram matrix to obtain a symmetric matrix with unit diagonal. This is the projection $\mathcal{P}_Y(G_q)$.
2. Apply (3.4) on G_q to bound the absolute values of the off-diagonal entries, producing $\mathcal{P}_Z(G_q)$.
3. Apply SVD to G_q to force the matrix rank to be equal to m , obtaining $\mathcal{P}_W(G_q)$.
4. If $G_q = U\Sigma U^T$ then $\mathcal{P}_S(G_q) = U\Sigma'U^T$ with Σ' being a diagonal matrix with m entries all equal to N/m and the rest zero.
5. Calculate the average Gram matrix G_{q+1} according to (3.5).

3.3.4 Convergence of Algorithm 2

The convergence of averaged projections algorithm is straightforward, considering the results presented in 3.2.1. The sets Y, Z, W and S involved in Algorithm 2 are prox-regular: Y, Z are convex and W, S are smooth manifolds. Their intersection is very likely to be strongly regular; the fact that our initial matrix is a random Gaussian matrix minimizes the probability of choosing an initial point that is near to a locally extremal point. Though we cannot guarantee strong regularity for the above sets, randomness seems to prevent us from irregular solutions. Therefore, we expect that the averaged projections algorithm converges linearly to a point in the intersection of the above sets.

Let us see what experimental results show. Figures 3.3 and 3.4 present mean squared distance for the averaged projections algorithm. Indeed, in Fig. 3.3 the results for a matrix of redundancy equal to 2 confirm a linear convergence rate and are in agreement with our theoretical expectations. Moreover, the zero mean squared distance implies that the obtained frame belongs to the intersection of the involved sets, that is, it forms an incoherent UNTF. The results are a little different for a matrix with higher redundancy. As we can see in Fig. 3.4(b), the rate of convergence becomes sub-linear, indicating that the intersection of the involved sets is either empty or does not have the desired properties. Relaxing the imposed incoherence level, i.e., using a larger bound than $1/\sqrt{m}$ in eq. (3.4), we obtain a broader set Z , increasing the probability that the intersection of the involved

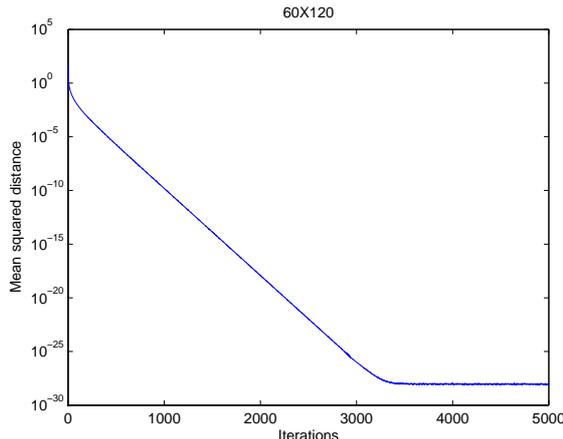


Figure 3.3: Convergence of Algorithm 2 (averaged projections) for a 60×120 matrix. The mean squared distance between the current iteration and the sets we project on reduces in a linear rate.

sets satisfies the sufficient conditions formulated in Theorem 3.2.3. The experiments performed with the new set Z yield a linear convergence rate (Fig. 3.4(b)), confirming our conjecture.

Comparing the convergence of the two proposed algorithms, an important note is that the presented experiments show that the results of the proposed averaged projections algorithm are similar to the alternating projections. Of course, there is a significant difference regarding the slope of the convergence curve; alternating projections is faster than averaged projections. However, the shapes of the curves are identical in all examples employed in our experiments. Therefore, even though the theoretical justification of the proposed alternating projections needs further investigation, the experimental results encourage its use for the proposed constructions. In the next subsection, we present some experiments demonstrating the properties of the obtained frames, showing that both algorithms give similar results.

Before proceeding to more experiments and applications, we would like to clarify a point concerning the incoherence level constraint. One might wonder what is the effect of the imposed incoherence level on the proposed construction. Do we obtain frames with similar properties, regardless of the bound used in eq. (3.4)? The answer is that the frame properties are similar but not identical. Depending on the frame redundancy, there is a lower incoherence bound that should not be exceeded; otherwise, the smaller the incoherence bound we impose, the worse the incoherence level we finally obtain. Thus, the selected bound needs fine tuning. However, the proposed bound $1/\sqrt{m}$ works well for the constructions considered in this thesis.

3.3.5 Experimental results

In order to test the performance of the proposed algorithms, this section includes experimental results that demonstrate the properties of the obtained constructions. The

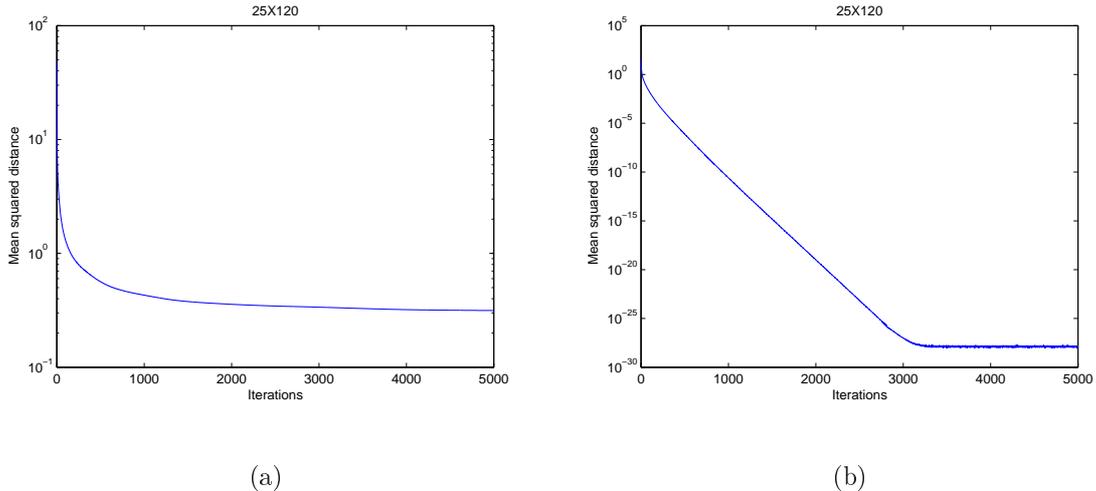
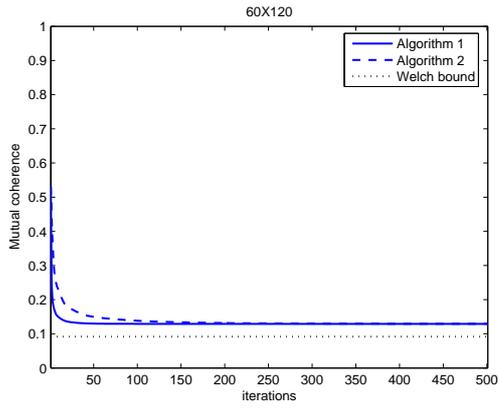


Figure 3.4: Convergence of Algorithm 2 (averaged projections) for a 25×120 matrix. The convergence rate depends on the bound used in eq. (3.4). In (a) we observe a sub-linear convergence rate when the bound equals $1/\sqrt{m}$. In (b) the convergence rate becomes linear as the bound is relaxed to $3/2\sqrt{m}$.

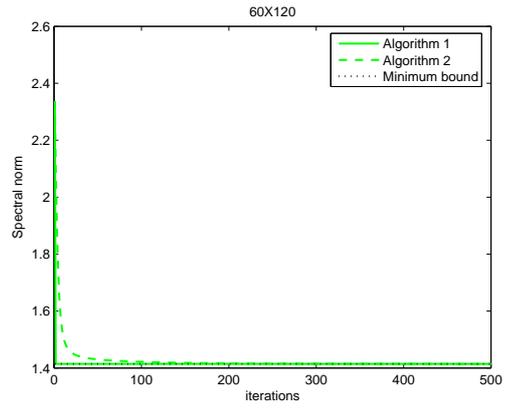
results concern mainly the mutual coherence that expresses the similarity between frame elements and the spectral norm that expresses how close is a frame to a UNTF.

We begin with Fig. 3.5 that illustrates three snapshots of execution including 500 iterations, depicting the achieved mutual coherence and spectral norm at every iteration. The examples involve frames of size 60×120 , 40×120 and 20×120 . The obtained results confirm our convergence discussion, showing that alternating projections algorithm is faster than averaged projections. However, both algorithms finally converge to similar values regarding mutual coherence and spectral norm. The attained results for the spectral norm coincide with the target values, while for the mutual coherence they are close to the minimum bound. Regarding spectral norm, the results for alternating projections are impressive showing that the algorithm meets the minimum bound after only a few iterations; both algorithms finally attain to produce UNTFs. The most important observation concerning the proposed algorithms is that their performance depends on the frame dimensions, or, more accurately, on the frame redundancy ($\rho = N/m$ for an $m \times N$ frame). The lower the frame redundancy, the smaller the distance between the properties of the obtained frames and the target values. This behaviour is more obvious regarding the mutual coherence, but it also affects the spectral norm for large values of redundancy and is in agreement with the convergence discussion of the previous paragraph. Average results presented next confirm these observations.

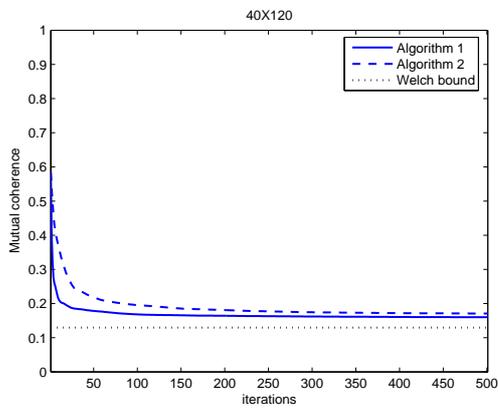
Tables 3.1 and 3.2 include average values of mutual coherence and spectral norm, respectively, for $m \times 120$ frames, with $m = 20 : 20 : 100$. The first column concerns random Gaussian matrices, the second column frames obtained with alternating projections (Algorithm 1) and the third column frames obtained with averaged projections (Algorithm



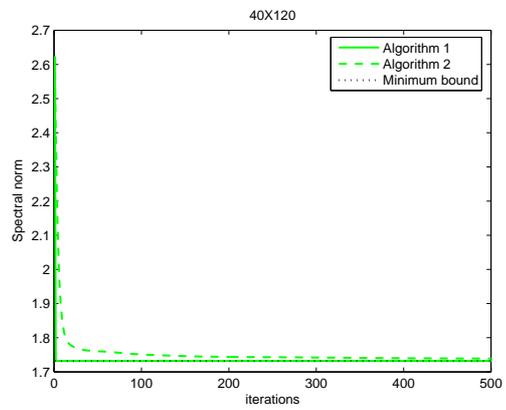
(a) 60×120 frame



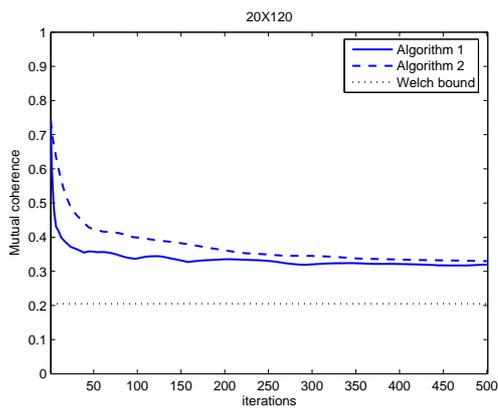
(b) 60×120 frame



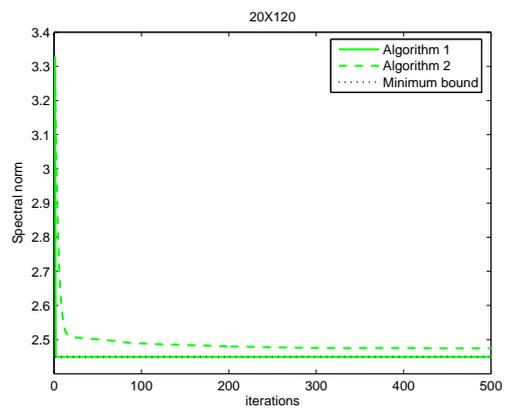
(c) 40×120 frame



(d) 40×120 frame



(e) 20×120 frame



(f) 20×120 frame

Figure 3.5: Mutual coherence (left) and spectral norm (right) as a function of the number of iterations. The experiments involve frames of various dimensions.

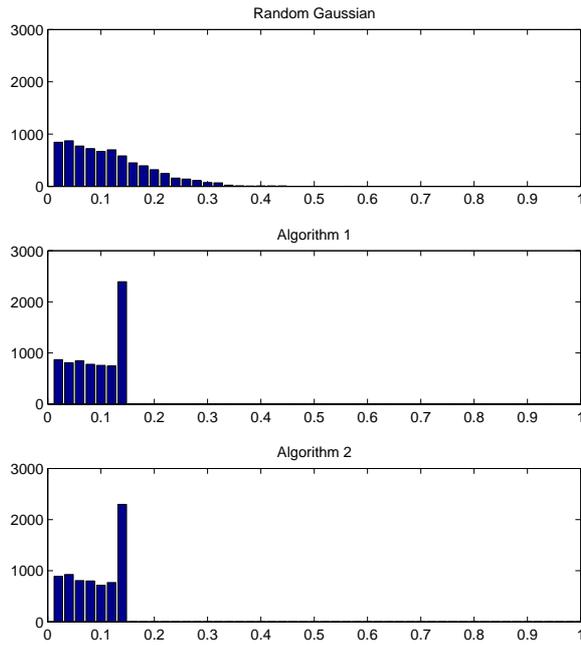


Figure 3.6: Distribution of Gram matrix entries of a 60×120 frame.

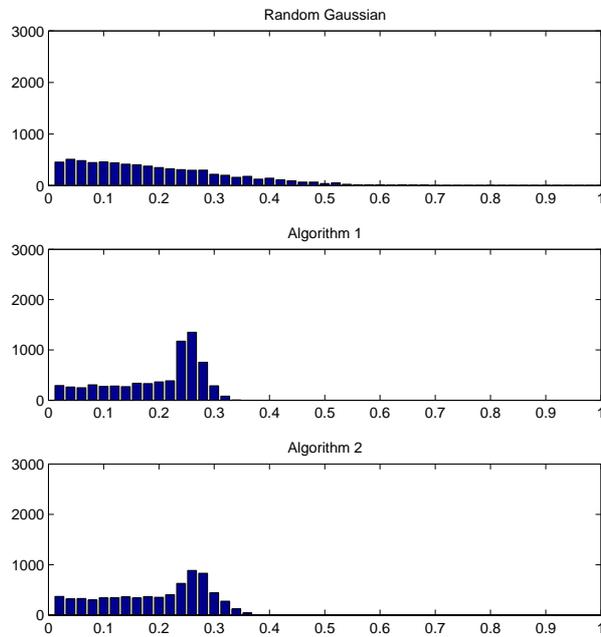


Figure 3.7: Distribution of Gram matrix entries of a 20×120 frame.

Table 3.1: Mutual coherence of $m \times N$ frames, with $m = 20 : 20 : 100$ and $N = 120$.

m	Gaussian	Algorithm 1	Algorithm 2	Optimal
20	0.756	0.339	0.393	0.205
40	0.577	0.169	0.199	0.130
60	0.472	0.129	0.138	0.092
80	0.428	0.112	0.114	0.065
100	0.384	0.100	0.103	0.041

Table 3.2: Spectral norm of $m \times N$ frames, with $m = 20 : 20 : 100$ and $N = 120$.

m	Gaussian	Algorithm 1	Algorithm 2	Optimal
20	3.281	2.450	2.483	2.450
40	2.637	1.732	1.752	1.732
60	2.333	1.414	1.421	1.414
80	2.158	1.225	1.227	1.225
100	2.044	1.095	1.097	1.095

2). The execution of algorithms involves 100 iterations. It is clear that both algorithms yield similar constructions, that is, they produce highly incoherent UNTFs. A small discrepancy between the results of the proposed algorithms can be erased if we increase the number of iterations so that the slow averaged projections algorithm catches up alternating projections. For medium and low redundancy the obtained values for mutual coherence approximate the lowest possible bound.

A better insight into the obtained constructions can be attained by demonstrating the distribution of the off-diagonal entries (absolute values) of the corresponding Gram matrix. Figures 3.6 and 3.7 present results for a 60×120 and a 20×120 frame, obtained after 100 iterations of the proposed algorithms. Compared to the original random Gaussian matrix, most correlation values of the incoherent UNTFs are concentrated near the optimal minimum bound, showing that the obtained constructions are close to ETFs.

Before concluding, we would like to note that the proposed frames are compared with other techniques that produce incoherent matrices for compressed sensing in Chapter 5, where they are used for sensing sparse signals. The reason for choosing not to make a comparison with existing methods at this point is that no other method for designing general purpose incoherent UNTFs has been proposed in the literature. As a final remark, we would like to emphasize that the algorithms proposed here utilizing alternating and averaged projections can yield incoherent UNTFs of any dimensions, providing an efficient tool for the construction of frames that are close to ETFs even if ETFs with the given dimensions do not actually exist.

3.4 Construction of nearly equiangular frames

The second technique we present for the construction of frames that are close to ETFs is based on ideas coming from graph theory. Summarizing the results presented in Section 2.3, an ETF can be defined up to unitary equivalence by its so-called signature matrix. Considering real equiangular frames, the corresponding signature matrix is a symmetric matrix with zero diagonal and ± 1 's off-diagonal entries, and it can be thought of as the adjacency matrix of a graph. The most important property of a signature matrix is its spectrum, consisting of exactly two eigenvalues λ_1, λ_2 , with multiplicity $N - m$ and m , respectively, given by

$$\lambda_1 = -\sqrt{\frac{m(N-1)}{N-m}}, \quad \lambda_2 = \sqrt{\frac{(N-m)(N-1)}{m}}. \quad (3.6)$$

Therefore, the problem of designing an ETF can be reduced to an inverse eigenvalue problem, that is, the construction of a matrix with specific structure and spectrum consisting of two distinct eigenvalues.

Many signature matrices that correspond to ETFs are known and constructions of ETFs based on signature matrices have been proposed in [119]. These techniques impose certain restrictions on frame dimensions. In this thesis, we consider frames of arbitrary dimensions and construct a symmetric matrix with spectrum that approximates the spectrum of the corresponding signature matrix. The obtained matrix is then used for the construction of frames that are close to ETFs. The produced frames are almost tight, with frame vectors forming angles that approximate the optimal value.

Inverse eigenvalue problems (IEPs) concern the construction of a matrix from prescribed spectral data. A large category of IEPs includes structured inverse eigenvalue problems (SIEPs), where given a set \mathcal{N} of specially structured matrices and a set of scalars $\{\lambda_i\}_{i=1}^N$, $\lambda_i \in \mathbb{R}$, corresponding to the desired spectrum, we want to find a matrix $X \in \mathcal{N}$ such that $\sigma(X) = \{\lambda_i\}_{i=1}^N$, where spectrum $\sigma(X)$ [40].

The signature matrix of an ETF is a symmetric matrix with zero diagonal, ± 1 's off-diagonal entries, and spectrum containing the eigenvalues given by (3.6). The problem we need to solve to find a signature matrix is a SIEP formulated as follows.

Signature Matrix Inverse Eigenvalue Problem (SMIEP). *Considering a set of two real numbers, λ_1, λ_2 , given by (3.6), find a symmetric $N \times N$ matrix with zero diagonal, ± 1 's off-diagonal entries, and spectrum*

$$\sigma = \underbrace{\{\lambda_1, \dots, \lambda_1\}}_{N-m}, \underbrace{\{\lambda_2, \dots, \lambda_2\}}_m, \quad m < N. \quad (3.7)$$

SIEPs are difficult to solve and most of the existing algorithms have been designed for problems of special type [40, 101]. The numerical method proposed here for SMIEP does not always produce an exact solution. However, it can produce an approximate solution satisfying structural constraints and approximating spectral constraints. Although such

a matrix is not the signature matrix of an ETF, it can be used to obtain a frame that is close to an ETF as we will see in the sequel.

The work presented here is based on the observation that real frames that are close to ETFs (e.g., incoherent frames proposed in [129]) have “signature” matrices with eigenvalues that approximate the spectrum of a signature matrix corresponding to an ETF. Before proceeding, we need to explain what we call a “signature” matrix of an arbitrary real frame. Suppose we are given an ETF with dimensions m, N . From equation (2.31) we see that we can derive the $N \times N$ signature matrix from the corresponding Gram matrix by keeping the signs of the off-diagonal entries and zeroing the diagonal. In the same manner, we can obtain an $N \times N$ symmetric matrix with ± 1 's off-diagonal entries and zero diagonal from the Gram matrix of an arbitrary $m \times N$ frame. Therefore, we are led to the following definition.

Definition 3.4.1 (Signature matrix of an arbitrary frame). The *signature matrix* Q of an arbitrary $m \times N$ real frame $F = [f_1 \ f_2 \ \dots \ f_N]$ is the $N \times N$ matrix with entries derived from the corresponding Gram matrix, $R = F^T F$, according to

$$q_{ij} = \begin{cases} \text{sgn}(r_{ij}), & i \neq j, \\ 0, & i = j, \end{cases} \quad (3.8)$$

where r_{ij} is the (i, j) entry of R . Obviously, the eigenvalues of an arbitrary signature matrix do not satisfy (3.6).

Now we can explain the main idea of the work presented here. Let us make the assumption that an ETF with arbitrary dimensions m, N exists, and use (3.6), (3.7), to calculate the spectrum of the corresponding signature matrix. If we construct a matrix with spectrum close to (3.7), satisfying the structure of a signature matrix, then, using (2.31), we obtain an $m \times N$ frame with good spectral properties and frame vectors forming angles near the optimal value. We refer to this frame as nearly equiangular.

3.4.1 Construction of signature matrices

A special case of SIEP is the symmetric nonnegative inverse eigenvalue problem (SNIEP), that is, finding a symmetric matrix with nonnegative entries and prescribed spectrum. A numerical method for the solution of SNIEP was presented in [101], where the authors utilize alternating projection ideas and propose an algorithm in which, first, the eigenvalue decomposition is used to impose the desired spectrum, and, subsequently, every negative entry of the obtained matrix is set to zero to obtain a nonnegative matrix.

Inspired by the work of [101], we propose here an algorithm that imposes structural and spectral constraints on a randomly generated symmetric matrix to find a solution to SMIEP. Starting from an initial matrix Q_0 with the prescribed structure, and using an iterative process consisting of two steps, in the k -th iteration we do the following:

Step 1. Compute the eigenvalue decomposition $Q_{k-1} = P\Lambda P^{-1}$, where Λ is a diagonal matrix containing the eigenvalues of Q_{k-1} and P is the matrix of the corresponding

eigenvectors. Then, produce a matrix with the desired spectrum σ according to $\tilde{Q}_k = P\Sigma P^{-1}$, where $\Sigma := \text{diag}(\sigma)$ is the diagonal matrix with entries the desired eigenvalues.

Step 2. Obtain a matrix Q_k with the desired structure that is close to \tilde{Q}_k , by keeping the signs of the off-diagonal entries of \tilde{Q}_k and set the diagonal to zero,

$$q_{ij} = \begin{cases} \text{sgn}(\tilde{q}_{ij}), & i \neq j, \\ 0, & i = j. \end{cases} \quad (3.9)$$

Step 1 replaces the eigenvalues of the given matrix with the requested ones; thus, it yields a matrix with the desired spectrum, impairing the matrix structure. This step actually uses Theorem A.2 given in the Appendix and projects on the spectral set of matrices with spectrum σ . Step 2 yields a matrix exhibiting the requested structure, impairing the matrix spectrum. The above steps bring up Algorithm 3. Note that, due to small numerical inaccuracy, \tilde{Q}_k from Step 1 may not be perfectly symmetric; thus, we perform the following operation: $\tilde{Q}_k := 0.5 \cdot (\tilde{Q}_k^T + \tilde{Q}_k)$.

Algorithm 3 Signature Matrix Construction I

Input: initial $N \times N$ signature matrix Q_0 , spectrum σ , iterations $ITER$

Output: $N \times N$ symmetric matrix Q_k , with zero diagonal, ± 1 's off-diagonal entries and spectrum approximate to σ

$\Sigma := \text{diag}(\sigma)$

for $k := 1$ to $ITER$ **do**

$[P, \Lambda] := \text{EigenDecomp}(Q_{k-1})$ // $Q_{k-1} = P\Lambda P^{-1}$

$Q_k := P\Sigma P^{-1}$ // apply desired spectrum

$Q_k := 0.5 \cdot (Q_k^T + Q_k)$

for every entry of Q_k , q_{ij} , **do**

if $i == j$ **then**

$q_{ij} := 0$ // diagonal entries

else

$q_{ij} := \text{sgn}(q_{ij})$ // off-diagonal entries

$k := k + 1$

Studying the convergence of the proposed algorithm is not a trivial task. Well known results from alternating projections cannot be applied here because convexity conditions for the employed sets are not satisfied, and in case the corresponding ETF does not exist, SMIEP is not solvable. Therefore, our results will be basically experimental. First, we use Algorithm 3 to compute signature matrices of ETFs that are known to exist. Our experiments have shown that the algorithm can produce the signature matrices of ETFs with dimensions $m \times (m + 1)$ in a few iterations. When the algorithm is used to construct ETFs of other dimensions, e.g., 5×10 , 6×16 , it may need a few trials (with different starting matrices) to find the corresponding signature matrices. A possible explanation

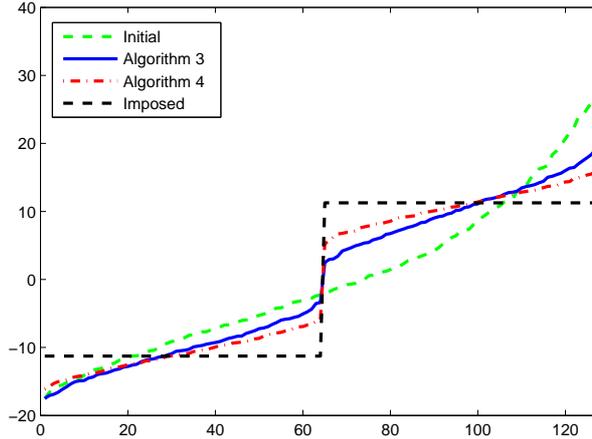


Figure 3.8: The spectrum of the signature matrix of a 64×128 random Gaussian matrix before and after processing the matrix with Algorithms 3 and 4. The black dotted line stands for the spectrum of the signature matrix corresponding to a 64×128 ETF.

for this is that the algorithm may converge locally, thus, finding a solution depends on the starting matrix. As an example we cite the signature matrix of a 6×16 ETF.

$$\begin{array}{cccccccccccccccc}
 0 & +1 & +1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & +1 & -1 \\
 +1 & 0 & +1 & +1 & +1 & -1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 \\
 +1 & +1 & 0 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & +1 & -1 \\
 -1 & +1 & -1 & 0 & -1 & -1 & -1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 \\
 +1 & +1 & +1 & -1 & 0 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & -1 & +1 \\
 -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\
 +1 & -1 & -1 & -1 & +1 & -1 & 0 & +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 \\
 +1 & +1 & -1 & +1 & -1 & -1 & +1 & 0 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & +1 \\
 -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & 0 & +1 & -1 & +1 & -1 & +1 & +1 & +1 \\
 +1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 & 0 & -1 & -1 & -1 & +1 & +1 & -1 \\
 +1 & +1 & -1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 & 0 & +1 & -1 & -1 & +1 & +1 \\
 -1 & +1 & +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 & 0 & -1 & +1 & +1 & +1 \\
 +1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & -1 & -1 & -1 \\
 -1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & 0 & -1 & +1 \\
 +1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 & 0 & -1 \\
 -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & 0
 \end{array}$$

Considering that these frame constructions are already known, the most important result of Algorithm 3 concerns finding the signature matrices of nearly equiangular frames of arbitrary dimensions. Testing the algorithm with signature matrices of frames of various dimensions has shown that after a few iterations we obtain a matrix with the requested structure and significantly improved spectrum that approximates (3.7); therefore, Algorithm 3 yields an approximate solution to SMIEP. Figure 3.8 demonstrates results con-

cerning the spectrum of a signature matrix before and after applying Algorithm 3. The initial signature matrix was obtained by a random Gaussian 64×128 matrix.

Algorithm 4 Signature Matrix Construction II

Input: initial $N \times N$ signature matrix Q_0 , spectrum σ , iterations $ITER$

Output: $N \times N$ symmetric matrix Q_k , with zero diagonal, ± 1 's off-diagonal entries and spectrum approximate to σ

```

 $\Sigma := \text{diag}(\sigma)$ 
for  $k := 1$  to  $ITER$  do
   $[P, \Lambda] := \text{EigenDecomp}(Q_{k-1})$  //  $Q_{k-1} = P\Lambda P^{-1}$ 
   $Q_k := P\Sigma P^{-1}$ 
   $Q_k := 0.5 \cdot (Q_k^T + Q_k)$ 
  for every entry of  $Q_k$ ,  $q_{ij}$ , do
    if  $i == j$  then
      if  $|q_{ij}| < t$  then
         $q_{ij} := 0$  // diagonal entries
      else
        if  $|1 - |q_{ij}|| < t$  then
           $q_{ij} := \text{sgn}(q_{ij})$  // off-diagonal entries
     $k := k + 1$ 
for every off-diagonal entry do
   $q_{ij} := \text{sgn}(q_{ij})$ 
for every diagonal entry do
   $q_{ii} := 0$ 

```

Our experiments with Algorithm 3 have shown that, even though the proposed processing improves the signature matrix spectrum substantially, it becomes ineffective after a few iterations. To further improve our results, we propose to modify the second step as follows. Before changing the value of a matrix entry according to (3.9), we examine its distance from 1 (off-diagonal) or 0 (diagonal). To avoid a significant spectrum impairment, if this distance exceeds a threshold t , we keep the entry unchanged, that is

$$q_{ij} = \begin{cases} \text{sgn}(\tilde{q}_{ij}), & \text{if } |1 - |\tilde{q}_{ij}|| < t, & i \neq j, \\ 0, & \text{if } |\tilde{q}_{ij}| < t, & i = j, \\ \tilde{q}_{ij}, & \text{otherwise.} \end{cases} \quad (3.10)$$

This way the k -th iteration does not produce a matrix having the appropriate entries, but structure is improved gradually. After a number of iterations is reached, we apply (3.9) to finally produce a matrix with the desired structure. Thus, we are led to Algorithm 4. Experimental results showing the improvement achieved with Algorithm 4 are presented in Fig. 3.8.

3.4.2 Nearly equiangular frames based on signature matrices

The signature matrix obtained by Algorithm 4 will be used next to construct a nearly equiangular frame. First, we construct the Gram matrix R according to (2.31). A symmetric $N \times N$ matrix obtained by (2.31) corresponds to an $m \times N$ frame, if it is of rank m . Thus, a rank reduction step follows. Using singular value decomposition (SVD), we keep the m largest eigenvalues and set the rest to zero. The matrix produced after rank reduction may not have ones in the diagonal; therefore, a normalization step follows to ensure that the Gram matrix corresponds to a unit norm frame. Finally, using SVD, we obtain an $m \times N$ frame, which is unit norm, almost tight, with the frame vectors forming angles near the optimal value. The above steps bring up Algorithm 5. Recall that the frame obtained this way is unique up to unitary equivalence.

Algorithm 5 Construction of a nearly equiangular frame

Input: $m \times N$ frame F_0

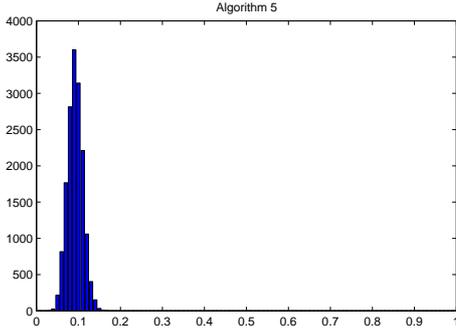
Output: $m \times N$ frame F_{out} , nearly equiangular

```

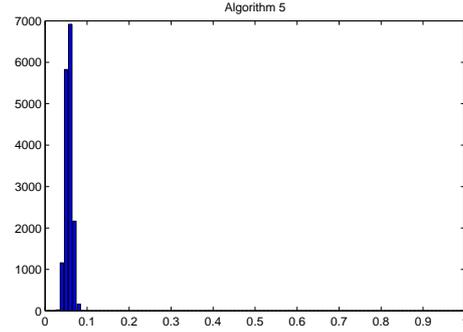
 $R_0 = F_0^T F_0$  // Obtain the initial Gram matrix
// Obtain  $Q_0$  according to (3.8)
 $Q_0 = \text{sgn}(R_0)$ 
 $Q_0(i, i) = 0$ , for all  $i$ 
// Use Algorithm 4 to obtain a signature matrix  $\tilde{Q}$ 
 $\tilde{Q} = \text{Algorithm2}(Q_0)$ 
// Obtain the Gram matrix from (2.31)
 $\tilde{R} = I + c\tilde{Q}$ 
// Reduce the rank of  $\tilde{R}$  to  $m$ 
 $[U, S, V] = \text{svd}(\tilde{R})$ 
 $S = S(1:m, 1:m)$ 
 $U = U(1:m, 1:m)$ 
 $V = V(1:m, 1:m)$ 
 $\check{R} = USV$ 
// Normalize the Gram matrix  $\check{R}$ 
 $\check{R} = \text{diag}(1./\text{sqrt}(\text{diag}(\check{R}))) \cdot \check{R} \cdot \text{diag}(1./\text{sqrt}(\text{diag}(\check{R})))$ 
// Obtain  $F_{\text{out}}$ 
 $[U, S, V] = \text{svd}(\check{R})$  //  $U = V$ 
 $F_{\text{out}} = \text{sqrt}(S)V^T$ 

```

Some results of the produced frames are presented in Fig. 3.9. Figure 3.9(a) demonstrates the frame vectors' correlation for a 64×128 frame, showing that the angles formed by the frame vectors have values around the optimal value of an ETF. Figure 3.9(b) demonstrates the frame vectors' correlation for a 96×128 frame, showing more impressive results for frames of low redundancy.

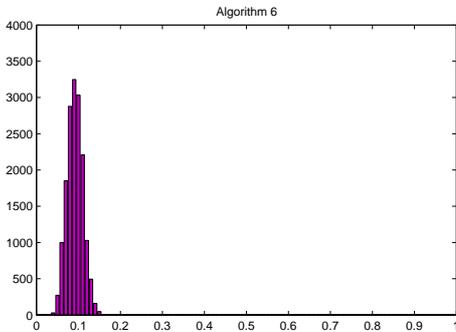


(a) A 64×128 frame ($\mu_{\text{opt}} = 0.0887$).

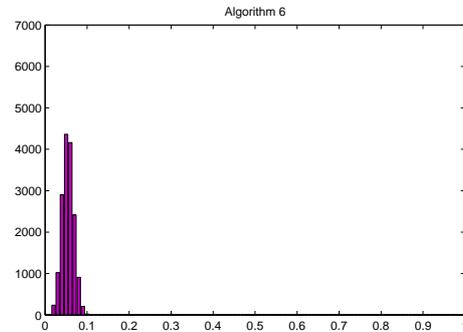


(b) A 96×128 frame ($\mu_{\text{opt}} = 0.0512$).

Figure 3.9: Correlation distribution of frame vectors produced with Algorithm 5. μ_{opt} stands for the optimal lowest bound (Welch bound).



(a) A 64×128 frame ($\mu_{\text{opt}} = 0.0887$).



(b) A 96×128 frame ($\mu_{\text{opt}} = 0.0512$).

Figure 3.10: Correlation distribution of frame vectors produced with Algorithm 6. μ_{opt} stands for the optimal lowest bound (Welch bound).

3.4.3 Nearly equiangular, nearly tight frames based on signature matrices

Algorithm 5 produces frames of any dimensions with the frame vectors forming angles near the optimal value. Even though the obtained frames exhibit good spectral properties, they are not exactly tight, a characteristic that is important for many applications. One way to improve tightness is Theorem 3.3.1 that finds a nearest α -tight frame to a given frame F according to $\sqrt{\alpha}(FF^T)^{-1/2}F$.

Having produced a nearly equiangular $m \times N$ frame with Algorithm 5, we apply Theorem 3.3.1 with $\alpha = N/m$. As tightness opposes unit-normness, we must carry out a few iterations, alternating between these two properties according to Algorithm 6 to obtain a nearly equiangular, nearly tight unit norm frame.

Algorithm 6 Construction of a nearly equiangular, nearly tight frame

Input: $m \times N$ frame F_0 **Output:** $m \times N$ frame F_{out} , nearly equiangular, nearly tight// Compute Q_0 the signature matrix of F_0 .// Obtain a nearly equiangular frame F_1 with Algorithm 5.**for** $k := 1$ to ITER **do** $F_{k+1} := \sqrt{N/m}(F_k F_k^T)^{-1/2} F_k$ // impose tightness $F_{k+1} := \text{normc}(F_{k+1})$ // normalize columns $k := k + 1$

Table 3.3: Spectral norm of $m \times N$ frames with $m = 32 : 16 : 96$ and $N = 128$ obtained with Algorithm 5 and Algorithm 6.

m	Spectral norm		
	Algorithm 5	Algorithm 6	Optimal
32	2.074	2.015	2.000
48	1.716	1.655	1.633
64	1.499	1.440	1.414
80	1.351	1.288	1.265
96	1.250	1.171	1.155

A metric to evaluate how close the obtained frame is to a unit norm tight frame is the spectral norm. Recall that the spectral norm of a unit norm tight frame equals the lowest possible bound $\sqrt{N/m}$. To see the improvement of tightness achieved by Algorithm 6 we construct frames of various dimensions and compute their spectral norm. The results presented in Table 3.3 are averaged over 500 frame samples and concern $m \times N$ frames with $m = 32 : 16 : 96$ and $N = 128$. While Algorithm 6 improves the spectral norm of the obtained frames, it also affects the frame vectors' correlation. We can see that there is a trade-off between equiangularity and tightness, also observed in Figures 3.9, 3.10. Figure 3.10 demonstrates results of the frame vectors' correlation for a 64×128 and a 96×128 frame produced by Algorithm 6. Comparing Fig. 3.10 to Fig. 3.9, we observe a slight deterioration of correlation's distribution, as a price of the improvement of tightness. Therefore, the choice between Algorithm 5 and Algorithm 6 for the construction of nearly equiangular frames, depends on the specific requirements of the related application.

More results regarding the properties of the proposed frames based on signature matrices are presented in the next section, where we provide a comparison with incoherent UNTFs obtained with alternating and average projections.

Table 3.4: Standard deviation of the Gram matrix entries corresponding to $m \times N$ frames with $m = 32 : 16 : 96$, $N = 128$, obtained with Algorithms 1, 5 and 6.

m	Standard deviation			
	Gaussian	Algorithm 1	Algorithm 5	Algorithm 6
32	0.1051	0.0582	0.0439	0.0412
48	0.0862	0.0471	0.0266	0.0257
64	0.0749	0.0422	0.0176	0.0189
80	0.0670	0.0364	0.0119	0.0153
96	0.0612	0.0292	0.0076	0.0136

3.5 Comparison of the proposed constructions

In order to provide a thorough comparison of the proposed frame constructions, we present here numerical results showing the achieved levels of equiangularity and incoherence of the obtained frames and discuss already presented results regarding the spectral norm. The experiments include frames constructed with the signature matrix based Algorithms 5 and 6, and Algorithm 1 that utilizes alternating projections. Averaged projections algorithm yields results very similar to alternating projections, as we have already seen in Tables 3.1 and 3.2, while it is more time consuming. Thus, results for Algorithm 2 are not demonstrated here. The next Tables include average values of the mutual coherence and the average coherence. To evaluate the equiangularity, we study the distribution of the Gram matrix entries. The experiments consider frames of variable redundancy, including constructions of size $m \times 128$, with $m = 32 : 16 : 96$. The obtained measurements are averaged over 500 realizations.

All measurements presented next are related to frame vectors' correlation. However, trying give an answer to the question how close are the proposed frames to ETFs, let us first discuss the obtained values for the spectral norm. Results for the spectral norm of nearly equiangular constructions have been presented in the previous section in Table 3.3. Comparison with incoherent UNTFs regarding the spectral norm is straightforward, as already presented results (see Table 3.2) show that incoherent UNTFs meet the minimum bound, which is also demonstrated in Table 3.3. Table 3.3 shows that equiangular constructions approximate the minimum bound without actually reaching it; therefore, incoherent UNTFs obtained with Algorithm 1 are preferable when tightness is important. However, the spectral norm of frames obtained with Algorithm 6 is very close to the optimal value; thus, Algorithm 6 is expected to provide reliable solutions when the application raises a need for tightness and equiangularity concurrently.

The most important property of the frames produced with the signature matrix based method is that they comprise nearly equiangular vectors. One way to observe equiangularity is to study the distribution of the Gram matrix entries. Besides Figures 3.9, 3.10 depicting the distribution of sample constructions, we present average measurements of

Table 3.5: Mutual coherence of $m \times N$ frames with $m = 32 : 16 : 96$, $N = 128$, obtained with Algorithms 1, 5 and 6.

m	Mutual coherence				
	Gaussian	Algorithm 1	Algorithm 5	Algorithm 6	Optimal
32	0.637	0.220	0.320	0.310	0.154
48	0.534	0.148	0.218	0.214	0.115
64	0.472	0.125	0.158	0.162	0.089
80	0.427	0.112	0.116	0.128	0.069
96	0.392	0.102	0.082	0.103	0.051

the standard deviation of the Gram matrix entries (absolute values). The results presented in Table 3.4 concern nearly equiangular frames constructed with Algorithms 5, 6, and incoherent UNTFs constructed with Algorithm 1. Clearly, Table 3.4 shows that signature matrix based frames are more equiangular compared to incoherent UNTFs, exhibiting smaller values of standard deviation. More impressive results are observed for low redundancy frames, showing the important role redundancy plays in the efficiency of the algorithms, a remark we also have made for Algorithms 1, 2.

Results for the mutual coherence are presented in Table 3.5 and show that smaller values are obtained for incoherent UNTFs. The difference between nearly equiangular frames and incoherent UNTFs is larger when the frame redundancy is high and becomes insignificant for less redundant frames. Truly, for 96×128 frames Algorithm 6 achieves the best results regarding column correlation, that is the smallest mutual coherence and standard deviation. Results obtained for 80×128 frames are also remarkable. Comparing the algorithms based on signature matrices, the observation made in Figures 3.9, 3.10 regarding equiangularity also holds for the mutual coherence; more tight frames are less equiangular and less incoherent. Although the differences in mutual coherence are minor and one would expect that they could hardly affect the applications of interest, the variation of the spectral norm we observed in Table 3.3 may affect the efficiency of the employed frames in applications.

A measure that accounts for all inner products between the columns of a given matrix, and not only for the largest one is average coherence. Given a unit norm matrix $A = [a_1 \ a_2 \ \dots \ a_N]$, average coherence is defined in [55] as

$$\mu_g(A) = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N |\langle a_i, a_j \rangle|^2. \quad (3.11)$$

Results for average coherence of the proposed constructions are presented in Table 3.6 and they are remarkable. All methods yield frames with identical average coherence. It seems that no matter what operations are made on a frame, the resulting construction attains some kind of equilibrium, expressed by the same average coherence. This observation may

Table 3.6: Average coherence of $m \times N$ frames with $m = 32 : 16 : 96$, $N = 128$, obtained with Algorithms 1, 5 and 6.

m	Average coherence				
	Gaussian	Algorithm 1	Algorithm 5	Algorithm 6	Optimal
32	0.042	0.038	0.038	0.038	0.154
48	0.035	0.029	0.029	0.029	0.115
64	0.030	0.023	0.023	0.022	0.089
80	0.028	0.017	0.017	0.017	0.069
96	0.026	0.013	0.013	0.013	0.051

be the key to explain the similar performance in sparse recovery observed for the proposed constructions, as we will see in the next chapters. Definitely, exploring the reasons for which these frames exhibit identical average coherence is a subject for further research.

Concluding the presentation of the developed frame constructions, we would like to make the following remarks. First, both methods proposed here yield frames exhibiting high incoherence levels. As we will see in the next chapters, the proposed frames are appropriate for sparse representations and compressed sensing, improving the performance of sparse recovery algorithms and offering accurate signal reconstruction. Second, regarding the spectral norm, Algorithms 1, 2 yield the best results, producing incoherent UNTFs with spectral norm coinciding with the minimum achievable bound. Signature matrix based constructions are less tight; however, they have a simpler implementation, thus, they are preferable if tightness requirements are loose. Finally, when equiangularity is the main requirement, then the best constructions are obtained with the algorithms based on signature matrices. The frames obtained with Algorithm 6 bridge the distance between tightness and equiangularity, and may be used as spreading sequences in s-CDMA systems, where both properties are required.

CHAPTER 4

PRECONDITIONING IN SPARSE AND REDUNDANT REPRESENTATIONS

4.1 The sparse representation problem

4.2 Mutual coherence and RIP

4.3 Promoting a sparse solution

4.4 The role of the spectral norm

4.5 Preconditioning

Sparse signal representations consist of a linear combination of a small number of elementary signals called atoms. Often, the atoms are chosen from a redundant (overcomplete) dictionary, that is, a collection of atoms with cardinality exceeding the dimension of the signal space. Thus, any signal can be represented by more than one combinations of different atoms [58].

Sparse representations are motivated by the fact that many natural signals are compressible, that is, they can be well approximated by a few large and many small coefficients. Sparseness is one of the reasons for the extensive use of popular transforms such as the Discrete Fourier Transform or the wavelet transform. The aim of these transforms is often to reveal certain structures of a signal and to represent these structures in a compact and sparse form. Sparsity has improved the performance of many signal processing applications such as compression, feature extraction, pattern classification, and noise reduction [58].

The generation of sparse representations with a redundant dictionary is non-trivial. Indeed, the general problem of finding a representation with the smallest number of atoms from an arbitrary dictionary has been shown to be NP-hard. This has led to considerable effort being put into the development of many sub-optimal schemes. A key

contribution to sparse representation problems is considered the work of [34] where the authors proposed a pursuit technique for evaluating sparsity. In general, algorithms for sparse representations form two classes: algorithms that iteratively build up the signal approximation one coefficient at a time, e.g., Matching Pursuit [93], Orthogonal Matching Pursuit [47], and algorithms that process all the coefficients simultaneously, e.g., Basis Pursuit [34]. Even though there exist a range of empirical evidence for the performance of methods built on sparse representation, many fundamental theoretical questions remain to be addressed. The development of novel fast sparse reconstruction algorithms, the theoretical and practical performance of such algorithms, the design and learning of good dictionaries are open research topics in the field [58].

In the heart of sparse representations lies an underdetermined linear system with more unknowns than equations. Uniqueness conditions for the existence of a sparse solution and performance guarantees for the algorithms deployed to find it require that the involved system matrix exhibits incoherence and good spectral properties [125]. While many incoherent tight dictionaries are known, often they cannot provide sufficiently sparse representations or they are not suitable for certain families of signals.

In this chapter, first, we survey well-known results providing the conditions for the existence of unique sparse representations and highlighting the constraints imposed for successful numerical computation. Based on these results, we consider an underdetermined linear system with sparse solutions and apply a mathematical technique referred to as preconditioning that yields a system matrix with good incoherence and spectral properties. While existing work in preconditioning concerns greedy algorithms, the technique presented here can be employed with any standard numerical solver. Our simulations show that the proposed preconditioning substantially improves the recovery rates in sparse representations.

4.1 The sparse representation problem

The weakness of orthogonal transforms to provide highly sparse representations has promoted the development of overcomplete dictionaries. Overcomplete or redundant dictionaries can provide compact representations with a few non-vanishing coefficients. Consider a finite-length real-valued signal x of length m , which we view as an $m \times 1$ column vector in \mathbb{R}^m . Let $\theta \in \mathbb{R}^N$ be a representation of x under an overcomplete dictionary $A \in \mathbb{R}^{m \times N}$, $m < N$,

$$x = A\theta. \tag{4.1}$$

Clearly x and θ are equivalent representations of the same signal, with x in the time domain and θ in the A domain. Assume that $\|\theta\|_0 = s$, where $\|\cdot\|_0$ is the ℓ_0 quasi-norm counting the nonzero coefficients of the treated signal. A sparse representation consists of a linear combination of s columns of A , with $s \ll N$. We refer to s as the *sparsity level* of θ . The set of indices corresponding to the non-vanishing coefficients is referred to as

the *support* of θ .

The sparse representation problem requires the computation of the vector θ , given only the dictionary A and the treated signal x . System (4.1) is underdetermined with fewer equations than unknowns, making the solution ill-posed in general. To avoid the trivial case of having no solution, we assume that the matrix A is of full rank. Thus, the system has infinitely many solutions and if one desires to narrow the choice to one well-defined solution, additional criteria are needed. Therefore, when considering systems of the form (4.1), the following plausible questions are posed:

- (a) When can uniqueness of a sparse solution be claimed?
- (b) Can the solution be reliably and efficiently computed in practice?
- (c) What performance guarantees can be given for various approximate and practical solvers?

Theoretical guarantees for a unique and stable solution satisfying (4.1) set bounds on the maximum sparsity level of the representation and impose certain constraints on the system matrix A .

4.2 Mutual coherence and RIP

Necessary and sufficient conditions ensuring that a signal can have a unique sparse representation under an overcomplete dictionary are phrased in terms of the mutual coherence and the restricted isometry property (RIP). These properties express a measure of the linear dependence between the columns of A , and are used to set restrictions on the maximal sparsity allowed for a unique representation. Note that the results presented here can incorporate sparse representations affected by additive noise, i.e., $x = A\theta + \eta$, $\|\eta\| \leq \epsilon$, with slight modifications.

4.2.1 Mutual Coherence

One of the most important properties related to the geometry of the dictionary A is the maximal column correlation, also known as mutual coherence. Recall that the mutual coherence $\mu(A)$ is a simple numerical way to characterize the degree of similarity between the columns of the matrix A and is defined as the largest absolute normalized inner product between different frame columns [93],

$$\mu(A) = \max_{\substack{1 \leq i, j \leq N \\ i \neq j}} \frac{|\langle a_i, a_j \rangle|}{\|a_i\| \|a_j\|}. \quad (4.2)$$

Mutual coherence is bounded as $0 \leq \mu(A) \leq 1$, with $\mu(A) = 0$ if A is orthogonal (see also Chapter 2). If $A \in \mathbb{R}^{m \times N}$, $m \leq N$, then $\mu(A)$ satisfies

$$\sqrt{\frac{N-m}{m(N-1)}} \leq \mu(A) \leq 1, \quad (4.3)$$

where the lower bound is the well-known Welch bound. Matrices with small mutual coherence are known as *incoherent*. Requiring a matrix A with small mutual coherence, that is, with columns as “independent” as possible, means that the information of x represented by A is spread in different directions, which makes its recovery easier. Mutual coherence plays an important role in the existence of a unique solution of system (4.1) as well as in the performance of the algorithms deployed to find sparse solutions.

4.2.2 Uniqueness via mutual coherence

Mutual coherence can provide a condition that gives an answer to the crucial question regarding the existence of a unique solution of (4.1). The following result was derived in [51].

Theorem 4.2.1 (Incoherence and sparsity [51]). *If the linear system of equations in (4.1) has a solution that satisfies the condition*

$$\|\theta\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mu(A)} \right), \quad (4.4)$$

then this solution is the sparsest one.

Consequently, if a solution satisfies (4.4), then this is the *unique sparsest solution*.

Combining Theorem 4.2.1 with the lower bound of mutual coherence, we can provide an upper bound of sparsity related to the lower dimension m of the matrix A . When $N \geq 2m$, it follows that $\mu(A) \geq (2m-1)^{-1/2}$. Thus, the maximum sparsity level ensuring a unique sparse representation is $\mathcal{O}(\sqrt{m})$. This bound is referred to as *square root bottleneck*.

We must note here that Theorem 4.2.1 is a pessimistic result, and often sparse signal recovery is possible for larger values of $\mathcal{O}(\sqrt{m})$. However, in order to shatter the square root bottleneck, probabilistic analysis is needed as we will see later.

4.2.3 The Restricted Isometry Property

The restricted isometry property (RIP) is a different way to measure the similarity of columns of a matrix and is used to study the uniqueness of the solution and the stability of system (4.1), while it provides conditions for robust recovery in the presence of noise.

Definition 4.2.2. An $m \times N$ matrix A has the *Restricted Isometry Property (RIP)* of order s with $s = 1, 2, \dots$, if there exists a constant $\delta_s \in [0, 1)$ such that

$$(1 - \delta_s)\|\theta\|^2 \leq \|A\theta\|^2 \leq (1 + \delta_s)\|\theta\|^2, \quad \text{for all } \theta \in \mathbb{R}^N. \quad (4.5)$$

We refer to δ_s as the *isometry constant*.

This concept was introduced in [28]. A matrix A obeys the RIP of order s , if δ_s is not too close to one. When this property holds, it implies that the Euclidean norm of θ is approximately preserved, after projecting it on the rows of A . Obviously, if matrix

If A were orthogonal then $\delta_s = 0$. Since we are dealing with non-square matrices this is not possible, thus, we can loosely say that when a matrix obeys RIP of order s , then all subsets of s columns are nearly orthogonal. Clearly, the closer δ_s is to zero, the closer to orthogonal all subsets of s columns of A are.

If \mathcal{S} is a set of columns of the dictionary A , with $|\mathcal{S}| = s$, the following expression of the isometry constant is an immediate consequence of the definition:

$$\delta_s^{\min} = \max_{\mathcal{S} \subseteq \{1, \dots, N\}, |\mathcal{S}|=s} \|A_{\mathcal{S}}^T A_{\mathcal{S}} - I_s\|, \quad (4.6)$$

where I_s is the $s \times s$ identity matrix.

It is interesting to note that the RIP is also related to the condition number of the Gram matrix. In [28, 11], it is pointed out that if A_r denotes the matrix that results by considering r arbitrary columns of A , then the RIP in (4.5) is equivalent to requiring the respective Gram, $A_r^T A_r$, $r \leq s$, to have its eigenvalues within the interval $[1 - \delta_s, 1 + \delta_s]$.

4.2.4 Relation between RIP and mutual coherence

The properties presented so far show that a central issue in sparse representations is the linear independence of vectors involved in the sparse representation. Mutual coherence and RIP try to capture the geometry of the dictionary A and help us to identify well-conditioned subsets of vectors. The size of well-conditioned subdictionaries determines the maximum sparsity level allowed to have a sparse representation under a given dictionary. If \mathcal{S} is a subset of columns of the dictionary A , with $|\mathcal{S}| = s$, then the subdictionary $A_{\mathcal{S}}$ is well-conditioned if $\|A_{\mathcal{S}}^T A_{\mathcal{S}}\| \leq c$, where c is a small constant. The following result connects the mutual coherence with the isometry constant.

Theorem 4.2.3 (Relation between RIP and mutual coherence [50]). *Let A be a dictionary with coherence $\mu = \mu(A)$, and $A_{\mathcal{S}}$ be an arbitrary s -column submatrix of A . Then*

$$\delta_s^{\min} = \|A_{\mathcal{S}}^T A_{\mathcal{S}} - I_s\| \leq (s - 1)\mu, \quad (4.7)$$

where I_s is the $s \times s$ identity matrix. In particular, every collection of s columns is linearly independent when $(s - 1)\mu < 1$.

We must note here that, while mutual coherence of a given matrix can be easily extracted, evaluating the RIP property is NP-hard. However, even though working with the mutual coherence is simpler than working with the complex RIP, the analysis from the point of view of the mutual coherence leads to pessimistic results regarding the maximal sparsity; recovering s components from a sparse signal requires s to be of order $\mathcal{O}(\sqrt{m})$ at most. From a theoretical perspective, the RIP property provides the ability of a probabilistic analysis of sparse recovery, improving substantially the results obtained with deterministic analysis.

4.3 Promoting a sparse solution

When seeking solutions that satisfy system (4.1) the first obstacle we need to surpass is the fact that the system may have infinitely many solutions. Additional criteria to narrow this choice are set through regularization. Thus, we define the general optimization problem

$$\min_{\theta} J(\theta) \text{ subject to } x = A\theta, \quad (4.8)$$

where $J(\theta)$ is a function that imposes sparsity constraints on θ .

4.3.1 The ℓ_0 -minimizer

One way to promote a sparse solution is the ℓ_0 quasi-norm. Choosing $J(\theta) \equiv \|\theta\|_0$, we are led to the following ℓ_0 -minimization problem,

$$\min_{\theta \in \mathbb{R}^N} \|\theta\|_0 \text{ subject to } x = A\theta. \quad (4.9)$$

The discrete and discontinuous nature of the ℓ_0 norm poses many conceptual challenges regarding the solution of (4.9). Problem (4.9) is NP-hard, requiring combinatorial search. The main techniques proposed for its solution include greedy algorithms. Greedy algorithms iteratively approximate the coefficients and the support of the sparse signals. They generate a sequence of locally optimal choices in hope of determining a globally optimal solution, thus, they have the advantage of being very fast and easy to implement. Orthogonal Matching Pursuit (OMP) [47] and its variants (CoSaMP [98], StOMP [53], regularized OMP [99]) belong to this category.

OMP was introduced in [47] as an improved successor of Matching Pursuit (MP) [93]. OMP starts from $\theta^{(0)} = 0$ and it iteratively constructs a k -term approximant $\theta^{(k)}$ by maintaining a set of active atoms. At each stage, it expands that set by one additional atom.

A result that provides performance guarantees for OMP is presented in [124].

Theorem 4.3.1 (Performance guarantee for OMP [124]). *Let A be an $m \times N$ matrix and $\theta \in \mathbb{R}^N$ be a solution of the ℓ_0 minimization problem (4.9) satisfying*

$$\|\theta\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mu(A)} \right).$$

Then OMP with error threshold $\epsilon = 0$ recovers θ .

4.3.2 Stability of ℓ_0 minimization via the RIP

Another fundamental question regarding problem (4.9) concerns the stability of the solution. Considering a slight discrepancy between $A\theta$ and x , which can be interpreted as the presence of noise, we define an error tolerant version of (4.9), with error tolerance $\epsilon > 0$

$$\min_{\theta \in \mathbb{R}^N} \|\theta\|_0 \text{ subject to } \|A\theta - x\| \leq \epsilon. \quad (4.10)$$

Algorithm 7 OMP: approximately solve ℓ_0 -minimization problem

Input: $m \times N$ matrix A , m -dimensional signal x , error threshold t

Output: N -dimensional signal θ

```

 $k = 0$ 
 $\theta^{(0)} = 0$ 
 $r^{(0)} = x$  // initial residual
 $s^{(0)} = \emptyset$  // initial solution support
 $\Phi^{(0)} = [\text{empty matrix}]$  // matrix of chosen atoms
repeat
   $k = k + 1$ 
   $Z = |A^T r^{(k)}|$ 
   $p = \operatorname{argmax}_{1, \dots, m} |Z|$  // find new support entry
   $s^{(k)} = \operatorname{sort}([s^{(k-1)}, p])$  // new support
   $\Phi^{(k)} = A_{s^{(k)}}$  // matrix of chosen atoms
   $\theta^{(k)} = \operatorname{argmin}_{\theta} \|x - \Phi^{(k)}\theta\|_2^2$  s.t.  $\operatorname{support}(\theta) = s^{(k)}$  // new solution estimation
   $r^{(k)} = x - A\theta^{(k)}$  // new residual
until  $r^{(k)} < t$ 

```

Stability issues require that both (4.9) and (4.10) must always give results of the same sparsity. A stability condition involving RIP is given in [58].

Theorem 4.3.2 (Stability of ℓ_0 minimization [58]). *Assume that $\hat{\theta}$ is a candidate solution of (4.10), with $2s_0$ non-vanishing coefficients, satisfying the inequality $\|A\hat{\theta} - x\| \leq \epsilon$. Let us assume that the matrix A satisfies the RIP property for $2s_0$, with $\delta_{2s_0} < 1$. If x_0, \hat{x}_0 are the solutions of (4.9) and (4.10), respectively, then*

$$\|x_0 - \hat{x}_0\| \leq \frac{4\epsilon^2}{1 - \delta_{2s_0}}. \quad (4.11)$$

4.3.3 The ℓ_1 -minimizer

As problem (4.9) is intractable, another approach towards its solution is smoothing the penalty function and replace ℓ_0 -norm with ℓ_1 -norm,

$$\min_{\theta \in \mathbb{R}^N} \|\theta\|_1 \quad \text{subject to} \quad x = A\theta. \quad (4.12)$$

This way we obtain a convex program with computational complexity polynomial in the signal length.

Transforming a computationally intractable problem into a tractable one does not necessarily mean that the solution of (4.9) is similar to the solution of (4.12). A result established in [51, 70] states that if system (4.1) has a solution that satisfies (4.4), then this is the *unique solution* of both ℓ_0 - or ℓ_1 -minimization. A uniqueness condition via the RIP property that also guarantees exact sparse recovery via ℓ_1 -minimization is presented next.

Theorem 4.3.3 (Exact recovery based on RIP [28]). *Suppose that the matrix A in problem (4.12) satisfies RIP of order s , with RIP-constant δ_s . Let θ_s denote the truncated version of θ obtained if we keep its s largest components and set the rest equal to zero.*

I. *If $\delta_{2s} < 1$ and θ is an s -sparse solution of $x = A\theta$, then it is unique.*

II. *If $\delta_{2s} < \sqrt{2} - 1$, then the solution to the ℓ_1 minimizer of (4.12), denoted by $\hat{\theta}$, satisfies the following two conditions*

$$\|\theta - \hat{\theta}\|_1 \leq C_0 \|\theta - \theta_s\|_1, \quad (4.13)$$

and

$$\|\theta - \hat{\theta}\| \leq C_0 s^{-\frac{1}{2}} \|\theta - \theta_s\|_1, \quad (4.14)$$

for some constant C_0 .

This theorem states that if the true vector is a sparse one, i.e., $\theta = \theta_s$, then the ℓ_1 -minimizer recovers the (unique) exact value. On the other hand, if the true vector is not a sparse one, then the minimizer results in a solution whose accuracy is dictated by a procedure that knew in advance the locations of the s largest components of θ . Note that this is a deterministic result; it is always true and not with high probability. Recently, the sufficient condition has been improved to $\delta_{2s} < 0.4931$ [97].

Well-known algorithms deployed to solve (4.12) include Matching Pursuit (MP) [93], Basis Pursuit (BP) [34], iterative thresholding [44], and Dantzig selector [26]. While these solvers require fewer measurements compared to greedy algorithms, they are computationally more complex.

4.4 The role of the spectral norm

Considering the global geometry of an overcomplete dictionary, after the mutual coherence, the most important geometric property is the spectral norm. Spectral norm is a measure of how close is a matrix to a tight frame. Recall that an $m \times N$ frame Φ with $\|\Phi\|^2 = \frac{N}{m}$ is a unit norm tight frame, meaning that the columns of Φ have unit norm and the rows are orthogonal. Results concerning the use of tight frames in sparse representations can be found in [24, 52, 124, 125, 8]. The latest theoretical results that justify the employment of incoherent tight frames in sparse recovery are probabilistic, leading to optimistic bounds on the maximal sparsity for sparse recovery.

Deterministic analysis of sparse representations has shown that, given an overcomplete dictionary, the maximal sparsity depends on the size (number of columns) of well-conditioned subdictionaries. Instead of considering arbitrary sets of columns, the authors of [125] focused on *random* subdictionaries and shattered the square root bottleneck using tools from Banach space probability. The theoretical results presented in [125] necessitate that the system matrix forms an incoherent unit norm tight frame.

Theorem 4.4.1 (Incoherent UNTFs and sparse recovery [125]). *Let A be an $m \times N$ incoherent unit norm tight frame, and θ a sparse representation of an m -dimensional signal x under dictionary A , that is, $x = A\theta$. If θ has $s \leq cm/\log N$ nonzero entries drawn at random (c is some positive constant), then it is the unique solution for ℓ_0 - and ℓ_1 -minimization problems with probability at least 99.44%.*

An $m \times N$ dictionary is characterized as incoherent if its mutual coherence does not exceed $1/\sqrt{m}$.

Theorem 4.4.1 states that the maximum sparsity level is allowed to approach the dimension m of the original time-domain signal. If the dictionary is not a UNTF, then similar results are given as a function of the spectral norm.

Employing spectral norm, mutual coherence and average coherence, the authors of [8] allow for similar sparsity levels providing near-optimal probabilistic guarantees in the performance of a fast greedy algorithm called one-step thresholding (OST). In [8] the average coherence of a unit norm matrix A is defined as

$$\nu(A) = \frac{1}{N-1} \max_i \left| \sum_{i \neq j} \langle a_i, a_j \rangle \right| \quad (4.15)$$

and is a measure of how well the frame elements are distributed in the unit hypersphere. The main result of [8] follows.

Theorem 4.4.2 ([8]). *Let A be an $m \times N$ matrix, with mutual coherence μ and average coherence ν . Suppose $\|A\|^2 = \frac{N}{m}$, $\mu \leq \frac{1}{164 \log N}$ and $\nu \leq \frac{\mu}{\sqrt{m}}$. Then, there exists a constant c such that sorted one-step thresholding fails with probability $\mathbb{P}\{\hat{\theta} \neq \theta\} \leq \frac{6}{N}$, provided that $N \geq 128$ and $m \geq cs \log N$.*

Spectral norm and mutual coherence are also used to provide tighter bounds on the maximal sparsity in case of convex optimization methods in [24], under the additional assumption that the sparse signals have independent nonzero entries with zero median.

4.5 Preconditioning

In linear algebra and numerical analysis, preconditioning is a process that conditions a given problem into a form that is more suitable for numerical treatment [6]. Given a linear system $x = A\theta$, a preconditioner C^{-1} of the matrix A is a matrix such that CA has a smaller condition number than A . Considering an underdetermined linear system with sparse solutions, the aim of the proposed technique is to transform (4.1) into a form that satisfies performance guarantees for the algorithms deployed for its solution. According to theoretical results presented in the previous section, a preconditioner of A should result in a matrix CA that forms an incoherent UNTF.

Although constructions of incoherent tight dictionaries appear often in signal processing applications, such dictionaries have a limited ability of sparsifying signals or are

suitable only for certain signal types. Learning based dictionaries, that have been proposed as an alternative, contain atoms generated from instances belonging to a particular signal family. Every signal in the family can then be represented as a linear combination of a few atoms from the dictionary. As the design of the dictionary is dictated by the characteristics of the treated signals, the obtained dictionary may not satisfy incoherence and/or tightness. Thus, one way to employ incoherent UNTFs in sparse representations is preconditioning.

Let A be an arbitrary $m \times N$ matrix, not satisfying the necessary conditions for sparse recovery. Suppose there exists an $m \times m$ matrix C such that the product CA exhibits good incoherence and spectral properties. Multiplying both sides of (4.1) by C , we obtain

$$Cx = CA\theta \quad \text{or} \quad z = CA\theta, \quad (4.16)$$

where $z = Cx$. Requiring C to be invertible, implies that system (4.1) is equivalent to (4.16). Therefore, solving the following minimization problem

$$\hat{\theta} = \arg \min_{\theta} \|\theta\|_0 \quad \text{subject to} \quad z = CA\theta, \quad (4.17)$$

we obtain a solution that satisfies also (4.9).

Problem (4.17) involves the *effective system matrix* $F = CA$; thus, the efficiency of the numerical algorithms deployed to solve it depends on the properties of F . The question that naturally arises is how can we construct an invertible $m \times m$ matrix C such that the effective matrix F has good incoherence and spectral properties?

The technique of preconditioning in sparse representations was introduced in [113, 114]. The weakness of many overcomplete dictionaries to satisfy incoherence properties, motivated the authors of [114] to propose a modification of thresholding and OMP, such that in the estimation of the unknown support, a matrix different from the original representation dictionary is employed. More particularly, in greedy algorithms like OMP, the estimation of the unknown support depends on the inner products $A^T x = A^T A\theta$. If A were an orthonormal basis, then $A^T A = I_N$, where I_N is the $N \times N$ identity matrix, and the product $A^T x$ would recover the unknown support. Similarly, when employing overcomplete dictionaries, successful recovery is achieved if the Gram matrix has small off-diagonal entries. The authors of [113, 114] introduced a new step, namely, the sensing step, for the estimation of the support of the unknown signal, which employs another dictionary Ψ incoherent to A . The key concept of a frame's coherence is extended to pairs of frames according to the following definition:

Definition 4.5.1 (Mutual coherence of pairs of frames). Given two frames $\Psi = [\psi_1 \ \psi_2 \ \dots \ \psi_N]$, and $A = [a_1 \ a_2 \ \dots \ a_N]$, the mutual coherence between Ψ and A is defined as the maximum absolute normalized inner product between the columns of the given dictionaries

$$\mu(\Psi, A) = \max_{1 \leq i, j \leq N} |\langle \psi_i, a_j \rangle|. \quad (4.18)$$

The sensing step involves the product $\Psi^T A$ that yields a pseudo-Gram matrix with small off-diagonal entries, due to incoherence between Ψ and A . Thus, support estimation described by $\Psi^T x = \Psi^T A \theta$ yields higher recovery rates. Regarding thresholding, an explicit formula for calculating the optimal matrix for support estimation is given in [113].

The method we proposed in [130] considers the recovery of signals that are sparse under overcomplete dictionaries and does not depend on the deployed sparse recovery algorithm. Using the ideas presented in Chapter 3, the proposed preconditioning concerns underdetermined linear systems encountered in sparse representations and aims at the construction of an effective system matrix with good incoherence and spectral properties.

4.5.1 Construction of a preconditioner

As the construction of a preconditioner suitable for system (4.1) aims at the construction of a system matrix with small mutual coherence and small spectral norm, we expect that both of the methods presented in Chapter 3 with slight modifications can be used to obtain a preconditioner. The proposed technique for the construction of a preconditioner involves the following basic steps:

1. Select the initial preconditioner C_{init} to be an $m \times m$ random Gaussian matrix. Set $F_0 = C_{\text{init}} A$.
2. Apply an algorithm that uses F_0 as input to produce a frame \tilde{F} with small mutual coherence and small spectral norm.
3. Obtain the $m \times m$ matrix C solving the minimization problem $\min_C \|CA - \tilde{F}\|$.

The efficiency of the above process depends on the solution of the least squares problem $\min_C \|CA - \tilde{F}\|$, which must yield a preconditioner C such that CA is as close as possible to \tilde{F} . A few iterations between step 2 and step 3 may be necessary to attain a good solution. Moreover, the obtained preconditioner must be invertible, in order to ensure equivalence between the initial and the preconditioned system. Thus, it is important to select an invertible initial matrix C . We are based on [110] and select the initial preconditioner to be a random Gaussian matrix, because a square random matrix will almost never be singular.

Next we discuss the details of every implementation and present experimental results.

4.5.2 Preconditioning with incoherent UNTFs

The first methodology we propose to obtain a preconditioner involves the construction of an incoherent UNTF based on the algorithms proposed in section 3.3. As both algorithms presented there yield similar constructions, we will use Algorithm 1 that converges faster. To construct a preconditioner using Algorithm 1 we employ an iterative process, with the q -th iteration involving the following steps:

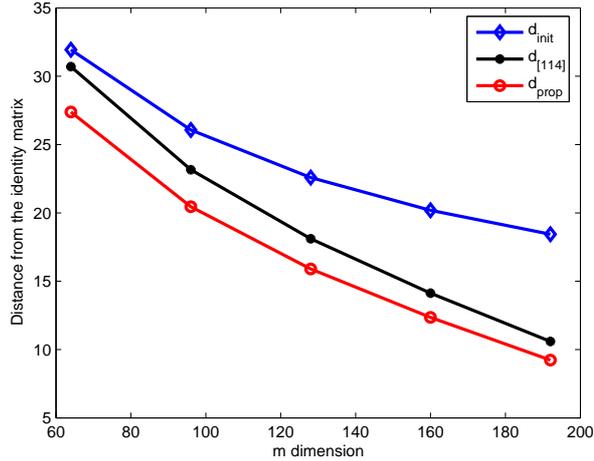


Figure 4.1: Discrepancy between the Gram or pseudo-Gram matrices involved in support estimation and the identity matrix of the same dimensions. The experiments involve $m \times N$ matrices with $m = 64 : 32 : 192$ and $N = 256$.

1. Apply Algorithm 1 on F_q to produce an incoherent UNTF \tilde{F}_q .
2. Find the $m \times m$ matrix C_q by solving the minimization problem $\min_C \|CA - \tilde{F}_q\|$.
3. Set $F_{q+1} = C_q A$.

Indeed, the above process produces an $m \times m$ matrix C_q that yields an effective system matrix $F_{q+1} = C_q A$ forming an incoherent UNTF. As we have already mentioned, the obtained preconditioner can be used in the solution of system (4.1), if and only if it is an invertible matrix; thus, C_q must be invertible. According to our analysis in Chapter 3, there is strong evidence that the algorithm converges locally, meaning that F_q is close to F_0 . Hopefully, the output matrix C_q will be close to the initial matrix C_{init} . Having selected an invertible initial matrix, the probability that the obtained matrix is singular is very low. Experimental results confirm our intuition.

Experimental Results

To test the proposed technique in computing a solution of (4.1), we produce sparse synthetic signals θ of length 256 under overcomplete random Gaussian dictionaries A of size 128×256 , obtaining a signal $x = A\theta$ of length 128. Following the above process, we compute a preconditioner C of size 128×128 and apply it to x to obtain $z = Cx$. Given the efficient matrix $F = CA$ and the signal z , OMP, BP and Dantzig selector are used to compute a sparse solution satisfying $z = CA\theta$. The algorithms are also used to compute a solution given A and x .

Before displaying results concerning the computation of sparse signals, we would like to estimate the appropriateness of the dictionaries involved in signal recovery when greedy algorithms are used. For this reason we compute the discrepancies d_{init} and d_{prop} between the corresponding Gram matrix involved in the sensing step and the identity matrix, that

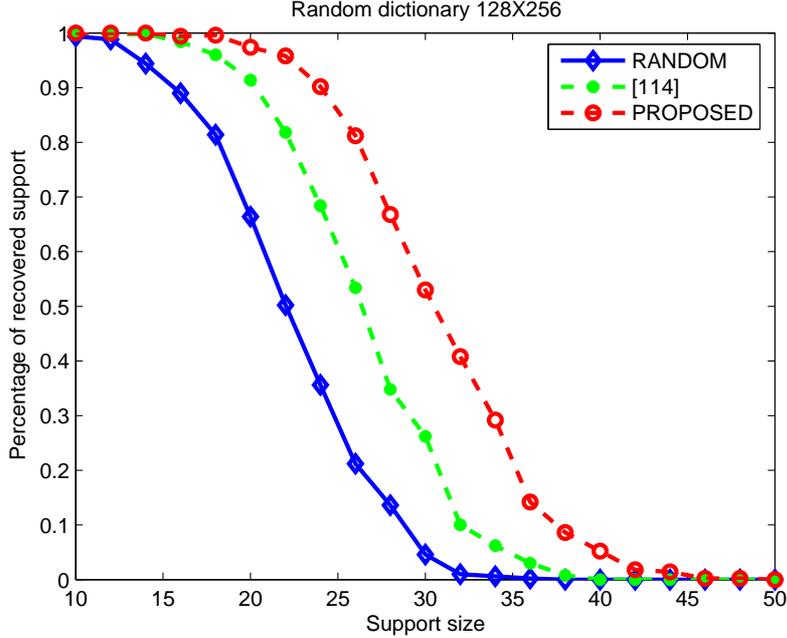


Figure 4.2: Support recovery rates for sparse representations using OMP for signals with varying support size. The preconditioner’s construction was based on the construction of incoherent UNTFs.

is, $d_{\text{init}} = \|A^T A - I_N\|_{\mathcal{F}}$ for the initial dictionary and $d_{\text{prop}} = \|F^T F - I_N\|_{\mathcal{F}}$ for the proposed preconditioning, where \mathcal{F} denotes the Frobenius norm. As our experiments involve a comparison with [114], which employs a matrix Ψ in the sensing step, we also compute the pseudo-Gram matrix $\Psi^T A$ and the distance $d_{[114]} = \|\Psi^T A - I_N\|_{\mathcal{F}}$. Results averaged over 500 experiments are presented in Figure 4.1, involving varying matrix dimensions. The results are best with the proposed construction, indicating improved performance in numerical recovery.

The performance of the deployed algorithms is quantified by computing the percentage of fully recovered support, referred to as recovery rate. Results for signals computed with OMP are demonstrated in Fig. 4.2, including the method proposed in [114]. The results are averaged over 500 experiments and concern signals with varying support size. Clearly, the recovery rates for OMP show that the proposed technique improves algorithm’s performance and surpasses the results in [114]. Similarly, recovery rates for BP and Dantzig selector in Figures 4.3 and 4.4, respectively, confirm that the proposed preconditioning transforms the original system in a manner that is more suitable for finding sparse solutions. The method of [114] concerns only greedy algorithms and is not applicable here.

4.5.3 Preconditioning with nearly equiangular frames

The second methodology we propose to obtain a preconditioner involves the construction of a nearly equiangular, nearly tight frame based on Algorithm 6. Similarly to the previous methodology, the initialization part involves C_{init} to be an $m \times m$ random Gaussian matrix,

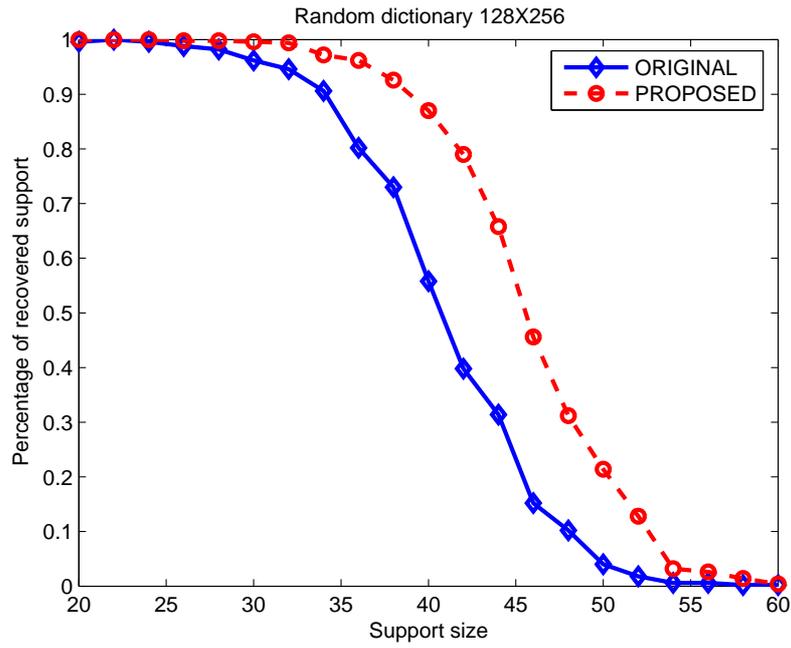


Figure 4.3: Support recovery rates for sparse representations using BP for signals with varying support size. The preconditioner's construction was based on the construction of incoherent UNTFs.

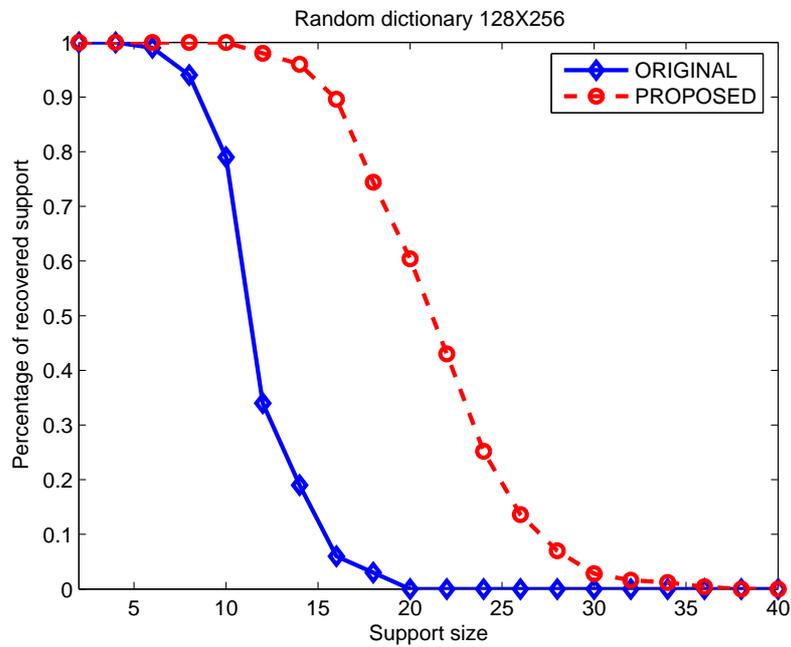


Figure 4.4: Support recovery rates for sparse representations using Dantzig selector for signals with varying support size. The preconditioner's construction was based on the construction of incoherent UNTFs.

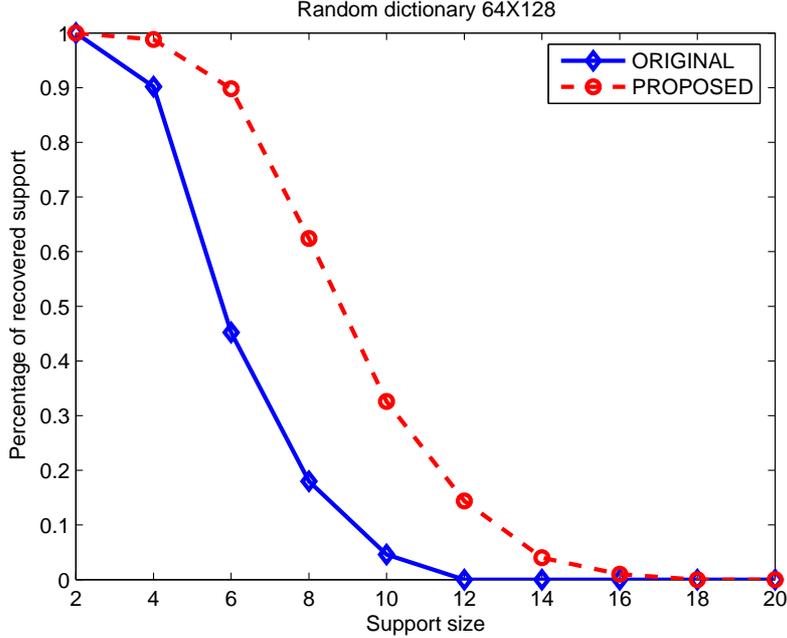


Figure 4.5: Support recovery rates for sparse representations using Dantzig selector for signals with varying support size. The preconditioner’s construction was based on the construction of nearly equiangular, nearly tight frames.

setting $F_0 = C_{\text{init}}A$. After obtaining a nearly equiangular, nearly tight frame, \tilde{F} , we compute the preconditioner solving $\min_C \|CA - \tilde{F}\|$. The process is iterative, with the q -th iteration involving the following:

1. Apply Algorithm 6 on F_q to produce a nearly equiangular, nearly tight frame \tilde{F}_q .
2. Obtain the $m \times m$ matrix C_q solving the minimization problem $\min_C \|CA - \tilde{F}_q\|$.
3. Set $F_{q+1} = C_qA$.

In contrast to the methodology presented in the previous section, experimental results show that the above iterative process does not seem to converge to an optimal solution. Thus, we perform a few iterations, in every iteration we keep the obtained solution and finally choose the preconditioner with the smallest mutual coherence. Regarding the invertibility of the obtained preconditioner, we do not really have any theoretical evidence that the produced matrix is invertible, but the experimental results show that the proposed methodology does not yield singular matrices.

Experimental Results

To test the proposed preconditioning technique, we produce sparse synthetic signals θ of length 128 under overcomplete random Gaussian dictionaries A of size 64×128 . Thus, the preconditioner C is of size 64×64 . Compared to the original random matrix A , the effective dictionary CA exhibits improved mutual coherence and spectral norm. Similarly

to the experiments presented in the previous section, we perform sparse signal recovery using OMP, BP and Dantzig selector. The results are averaged over 500 experiments and concern signals with varying support size. While OMP and BP do not seem to improve their performance substantially, recovery rates for the Dantzig selector are better with the proposed preconditioning and they are demonstrated in Figure 4.5.

CHAPTER 5

IMPROVING SPARSE RECOVERY IN COMPRESSED SENSING

-
- 5.1 Compressed sensing basics
 - 5.2 Projection matrices constructions
 - 5.3 Compressed sensing with the proposed frame constructions
 - 5.4 Proposed optimized projections
 - 5.5 Preconditioning in compressed sensing
-

Compressed sensing or compressive sampling (CS) is a novel theory [49, 25] that merges compression and acquisition, exploiting sparsity to recover signals that have been sampled at a drastically smaller rate than the conventional Shannon/Nyquist theorem imposes. Based on recent mathematical results, CS has enabled signal reconstruction from much fewer data samples, relying on the observation that many natural signals are sparse or compressible, i.e., they can be represented by a few significant coefficients. Recovering a signal from incomplete measurements can be done with computationally efficient methods.

The results of CS have an important impact on numerous signal processing applications including the efficient processing and analysis of high-dimensional data such as audio [71], image [89, 100], video [7], and bioinformatic data [131, 90]. CS has been applied to accelerate the sensing process in medical imaging [15, 66, 72] and to limit the number of sensors in Wireless Visual Sensor Networks (WVSNs) [104]. Other application specific architectures that have been developed include radar analysis [120, 103] and astronomical imaging [16]. Besides signal processing, to date CS theory is extensively utilized by experts to address problems in various fields such as biology [61, 116], medicine [94] and seismology [79].

The standard way to obtain a compressed representation of a signal involves that one computes the coefficients in an appropriate basis and then keeps only the largest coefficients. When complete information on the signal is available, this is certainly a valid strategy. However, when the signal has to be acquired first with a somewhat costly, difficult, or time-consuming measurement process, this seems to be a waste of resources: First, one spends huge efforts to collect complete information on the signal and then one throws away most of the coefficients of the signal to obtain its compressed version. Compressed sensing is an emerging theory that condenses the signal directly into a compressed representation, allowing signal recovery from a number of measurements that is much smaller than the signal length.

Recovering sparse signals from incomplete measurements leads to the ℓ_0 and ℓ_1 minimization problems formulated in sparse representations. In the context of CS, recovery guarantees concern the sensing matrix, i.e., the matrix implementing the sensing mechanism, and involve the restricted isometry property (RIP). At present, a comprehensive CS theory seems established [65] except for a few deep questions such as the improvement of the sensing mechanism and the efficiency of sparse recovery.

In early CS applications, the sensing process was implemented using random matrices. It is known that an $m \times N$ random Gaussian or Bernoulli matrix satisfies RIP with high probability and it can be used to recover an s -sparse signal, provided that the number of measurements m is $\mathcal{O}(s \log(N/s))$ [11]. Recent research aims either at the reduction of the number of measurements or at the improvement in recovery performance. While CS theory concerns non-adaptive measurements, recent work includes optimally designed sensing matrices with respect to a given sparsifying dictionary. Other parameters affecting the design of the sensing mechanism involve the hardware implementation and constraints imposed by the specific application. From this viewpoint significant work is related to matrices that are not completely random and often exhibit considerable structure.

After reviewing basic results from CS theory, we discuss three approaches improving signal recovery in CS. The first includes the employment of the proposed frame constructions as sensing matrices. The second includes the construction of optimized sensing matrices with respect to a given sparsifying dictionary. A third approach considers binary sensing matrices that are more suitable for hardware implementation and improves signal recovery using preconditioning.

5.1 Compressed sensing basics

In signal processing, the conventional Shannon/Nyquist theorem asserts that a signal must be sampled at a rate at least twice its highest frequency in order to be represented without error. Similarly, the fundamental theorem of linear algebra suggests that the number of collected samples (measurements) of a discrete finite-dimensional signal should be at least as large as its length in order to ensure reconstruction. Recovering sparse signals from incomplete measurements relies on recent results that concern the solution

of underdetermined linear systems with numerical methods [93].

Consider a finite-length real-valued signal θ of length N , which we view as an $N \times 1$ column vector in \mathbb{R}^N . CS yields a compressed representation of the treated signal using a sensing mechanism that is realized by an $m \times N$, $m \ll N$, matrix P , which is known as *sensing or projection or measurement matrix*. The linear measurement process is described by

$$y = P\theta, \quad (5.1)$$

where $y \in \mathbb{R}^m$ is the $m \times 1$ vector containing the obtained measurements. Note that the measurement process is non-adaptive, that is, P does not depend in any way on the signal.

Unique identification of a signal from a few measurements is feasible, if we restrict the class of signals we aim to recover. In CS, we assume that θ is a sparse signal, that is, $\|\theta\|_0 = s$, where $\|\cdot\|_0$ is the ℓ_0 quasi-norm counting the non-vanishing coefficients of the treated signal; s is the sparsity level of θ and D is referred to as the sparsifying dictionary. The set of indices corresponding to the non-vanishing coefficients is referred to as the *support* of θ . For signals that are not exactly sparse but compressible, we keep the s most significant coefficients.

System (5.1) is underdetermined with fewer equations than unknowns. A sparse vector satisfying (5.1) can be obtained as the solution of the ℓ_0 -minimization problem

$$\min_{\theta \in \mathbb{R}^N} \|\theta\|_0 \quad \text{subject to} \quad y = P\theta, \quad (5.2)$$

or, alternatively, as the solution of the ℓ_1 -minimization problem

$$\min_{\theta \in \mathbb{R}^N} \|\theta\|_1 \quad \text{subject to} \quad y = P\theta. \quad (5.3)$$

The above minimization problems can be solved efficiently as long as P exhibits certain properties. Results from sparse representations require either that P forms an incoherent unit norm tight frame (Theorem 4.4.1) or that it satisfies the restricted isometry property (Theorem 4.3.3). In this case, well-known algorithms such as OMP [47] and BP [34] can compute the solution of the ℓ_0 - and ℓ_1 -minimization problems. Random Gaussian or random Bernoulli matrices have been proved to exhibit good RIP properties and have been employed in various CS applications.

The theoretical guarantees for sparse recovery in the context of CS are mainly expressed in terms of a sufficient number of measurements. Projection matrices obeying RIP of order s can recover an s -sparse signal, provided that the number of measurements, $m < N$, is of order $\mathcal{O}(s \ln(N/s))$, that is

$$m \geq c_0 s \ln(N/s), \quad (5.4)$$

where c_0 is some constant, which depends on the isometry constant δ_s [11]. Note that m is larger than the sparsity level by an amount controlled by the inequality (5.4). Apparently, the higher the value of s , for which the RIP property of a projection matrix holds true, the larger the range of sparse signals that can be observed.

A recent result that has been formulated in [46] gives a direct expression of the constant involved in (5.4).

Theorem 5.1.1. *Let P be an $m \times N$ matrix that satisfies the RIP of order $2s$ with constant $\delta_s \in (0, \frac{1}{2}]$. Then*

$$m \geq c_1 s \log(N/s), \quad (5.5)$$

where c_1 is some constant, $c_1 = 1/2 \log(\sqrt{24} + 1) \approx 28$.

The restriction to $\delta_s \in (0, \frac{1}{2}]$ is arbitrary and is made merely for convenience. Minor modifications to the argument establish bounds for $\delta \leq \delta_{\max}$ for any $\delta_{\max} < 1$.

The seminal work of [25] where CS theory was first established concerned sparse signal representations under orthonormal bases. Consider a finite-length real-valued signal x of length N , which we view as a $N \times 1$ column vector in \mathbb{R}^N . Let $\theta \in \mathbb{R}^N$ be a sparse representation of x under an orthonormal basis $D \in \mathbb{R}^{N \times N}$,

$$x = D\theta. \quad (5.6)$$

Then, compressed sensing is described by

$$y = PD\theta. \quad (5.7)$$

Setting $F = PD$, system (5.7) can be written in the form

$$y = F\theta, \quad (5.8)$$

with $F \in \mathbb{R}^{m \times N}$ referred to as the *effective dictionary*.

Rephrasing the results formulated in sparse representations to apply to CS, we obtain recovery conditions for the effective dictionary $F = PD$. However, designing an efficient process to recover a signal from incomplete measurements requires theoretical guarantees that concern the sensing mechanism, i.e., the projection matrix P . It has been shown that the above theoretical results that hold for naturally sparse signals also hold for signals that are sparse under orthonormal bases. Requiring P to be a random Gaussian matrix, then the product PD is also an independent identically distributed Gaussian matrix regardless of the choice of the orthonormal sparsifying basis D . Random Gaussian matrices are universal in the sense that PD has the RIP with high probability, therefore, the conditions for sparse recovery for ℓ_0 - and ℓ_1 - minimization problems are satisfied.

Conditions that guarantee recovery of signals that are sparse under redundant dictionaries were established in [27]. In this case, the projection matrix must satisfy a modified RIP property referred to as D -RIP.

Definition 5.1.2 (D-RIP [27]). Let Σ_s be the union of all subspaces spanned by all subsets of s columns of D . A projection matrix, P , obeys the restricted isometry property adapted to D , (D -RIP), with δ_s , if

$$(1 - \delta_s)\|\theta\|^2 \leq \|P\theta\|^2 \leq (1 + \delta_s)\|\theta\|^2, \quad \text{for all } \theta \in \Sigma_s. \quad (5.9)$$

The union of all subspaces, Σ_s , contains all signals x that are s -sparse with respect to the dictionary D . This is the difference with the RIP definition given in section (4.5). All random matrices discussed earlier can be shown to satisfy D -RIP, with overwhelming probability, provided that the number of measurements, m , is at least of order $c_2 s \ln(N/s)$.

5.2 Projection matrices constructions

Compressed sensing was introduced utilizing random projection matrices. The entries of an $m \times N$ random Bernoulli matrix take the value $+\frac{1}{\sqrt{m}}$ or $-\frac{1}{\sqrt{m}}$ with equal probability, while the entries of a Gaussian matrix are independent and follow a normal distribution with expectation 0 and variance $1/m$. With high probability such random matrices satisfy the restricted isometry property with a (near) optimal order in s ; therefore, they allow sparse recovery.

Theorem 5.2.1 (Recovery condition for Gaussian and Bernoulli random matrices [11]). *Let $P \in \mathbb{R}^{m \times N}$ be a Gaussian or Bernoulli random matrix. Let $\epsilon, \delta \in (0, 1)$ and assume*

$$m \geq C\delta^{-2}(s \ln(N/s) + \ln(\epsilon^{-1})) \quad (5.10)$$

for a universal constant $C > 0$. Then with probability at least $1 - \epsilon$ the restricted isometry constant of P satisfies $\delta_s \leq \delta$.

The above Theorem, a simple proof of which can be found in [11], states that all s -sparse vectors θ can be recovered from $y = P\theta$, provided that the number of measurements satisfies $m \geq C\delta^{-2}(s \ln(N/s) + \ln(\epsilon^{-1}))$. Note that setting $C' = C\delta^{-2}$ and choosing $\epsilon = \exp(-cm)$ with $c = 1/(2C')$, we obtain the recovery condition $m \geq 2C's \ln(N/s)$ that we have seen in Theorem 5.1.1.

While random matrices satisfy RIP with high probability, the absence of structure in these matrices leads to infeasible real-world applications. When multiplying arbitrary matrices with signal vectors of high dimension, the lack of any fast matrix multiplication algorithm results in high computational cost. Even storing an unstructured matrix may be difficult. Thus, large scale problems are not practicable with Gaussian or Bernoulli matrices.

Another important issue when considering random matrices is that the fully random matrix approach is sometimes impractical to build in hardware. Applications often do not allow the use of “completely” random matrices, but put certain physical constraints on the measurement process and limit the amount of randomness that can be used. Hardware architectures that have been implemented to enable random measurements in practical settings include the random demodulator [122], random filtering [123] the modulated wideband converter [96], random convolution [108] and the compressive multiplexer [117]. These architectures typically use a reduced amount of randomness and are modeled via matrices that have significantly more structure than a fully random matrix.

The physics of the sensing mechanism and the capabilities of sensing devices may also limit the types of CS matrices that can be implemented in a specific application. Clearly, one reason for proposing new constructions of projection matrices is to address practical limitations appearing in the applications. A research direction towards the solution of such problems involves structured matrices. Important work in construction of structured matrices includes deterministic matrices [48, 83, 4, 78, 20] and structured random matrices [75, 105, 106].

Besides the difficulties in hardware implementation, research on projection matrices is also motivated by the improvement of recovery conditions. New theoretical and practical results concern matrices that are more efficient than random Gaussian or Bernoulli projections. Therefore, another research direction investigates the construction of matrices that lead to fewer necessary measurements or improve the performance of the algorithms deployed in sparse recovery. An interesting approach involves optimized projections [57, 145, 82].

5.2.1 Deterministic projections

From a computational and an application oriented viewpoint it is desirable to have measurement matrices with structure. One class of such matrices includes deterministic matrices. Deterministic constructions [48, 83, 4, 78, 20] may provide the convenience to verify RIP without checking up all s -column submatrices. However, the main drawback of deterministic matrices is that they satisfy poor recovery conditions.

Known deterministic matrices with optimal or near optimal mutual coherence are equiangular tight frames [119] and the Gabor frames generated from the Alltop sequence [78], which are of size $m \times m^2$. Considering a deterministic matrix with mutual coherence $1/\sqrt{m}$, the sparsity level must be of the order of \sqrt{m} (square root bottleneck), or, equivalently, the maximum number of measurements m that must be obtained to ensure a unique solution is $\mathcal{O}(s^2)$. The aforementioned constructions restrict the number of measurements needed to recover an s -sparse signal to $\mathcal{O}(s^2 \log N)$. A construction that managed to go beyond the square root bottleneck [20] provided only a slight improvement.

It is also possible to deterministically construct matrices of size $m \times N$ that satisfy the RIP of order s , but such constructions also require m to be relatively large [48, 83, 20]. For example, the construction in [48] requires $m = \mathcal{O}(s^2 \log N)$ while the construction in [83] requires $m = \mathcal{O}(sN^\beta)$ for some constant β . In many practical settings, this result would lead to an unacceptably large requirement on m . A more optimistic result concerning a specific deterministic construction can be found in [10]; the authors conjecture that ETFs corresponding to Paley graphs of prime order [107] are RIP in a manner similar to random matrices.

5.2.2 Structured random projections

Since it is hard to prove good recovery conditions for deterministic matrices as outlined above, many structured constructions allow some randomness to come into play. This leads to structured random matrices. These matrices are of great interest for computationally efficient sparse recovery, even though they do not precisely attain recovery condition (5.10). The best recovery bounds have the form $\mathcal{O}(Cs \log^\alpha(N/\varepsilon))$, $\alpha > 1$, where $\varepsilon \in (0, 1)$ corresponds to the probability of failure [105]. The important linear scaling of m in s up to log-factors is retained.

An important type of structured random matrices is based on randomly sampled functions [105]. Let $\mathcal{D} \subset \mathbb{R}^d$. Consider a function of the form

$$f(t) = \sum_{k=1}^N x_k \psi_k(t), \quad t \in \mathcal{D}, \quad (5.11)$$

where $x_1, \dots, x_N \in \mathbb{C}$. Let $t_1, \dots, t_N \in \mathcal{D}$ be some points and suppose we are given the sample values

$$y_\ell = f(t_\ell) = \sum_{k=1}^N x_k \psi_k(t_\ell), \quad \ell = 1, \dots, m. \quad (5.12)$$

The corresponding measurement matrix has entries $P_{\ell,k} = (\psi_k(t_\ell))$, $\ell = 1, \dots, m$, $k = 1, \dots, N$. Assuming that the sampling points t_ℓ are selected independently at random, $P_{\ell,k}$ becomes a structured random matrix. So the structure is determined by the function system ψ_k , while the randomness comes from the sampling locations. Sufficient conditions for sparse recovery for CS matrices of the above form require $\mathcal{O}(Cs \ln^2(6N/\varepsilon))$ measurements [105].

The random partial Fourier matrices, which consist of randomly chosen rows of the discrete Fourier matrix can be viewed as a special case of this setup and was studied already in the very first papers on compressed sensing [25]. For these matrices the recovery condition requires $\mathcal{O}(Cs \log(N/\varepsilon))$ measurements. A fast application of a partial Fourier matrix can be computed using the fast Fourier transform (FFT) algorithm.

Another type of structured matrices are partial random circulant and Toeplitz matrices [75, 105, 106]; they were first inspired by applications in communications. A circulant matrix U is a square matrix where the entries in each diagonal are all equal, and where the first entry of the second and subsequent rows is equal to the last entry of the previous row. Since this matrix is square, we perform random subsampling of the rows to obtain a CS matrix $P = RU$, with R being an $m \times N$ subsampling matrix, i.e., a submatrix of the identity $N \times N$ matrix. Circulant and Toeplitz matrices can be applied efficiently using FFT, and they greatly reduce the computational and storage complexity in large-dimensional problems. The recovery guarantees for these matrices require m to be of order $\mathcal{O}(s^{1.5} \log^{1.5} N)$ [54].

5.2.3 Optimized projections

Given a sparse signal θ under a dictionary D , the main criterion when designing a projection matrix P is to enable unique identification of θ from its measurements $y = PD\theta$. While the aforementioned matrix constructions concern non-adaptive projection matrices, designing a projection matrix with respect to a given sparsifying dictionary leads to optimized projections.

A major obstacle in the construction of projection matrices is that verifying RIP is combinatorially complex; we must examine $\binom{N}{s}$ possible combinations of s nonzero entries in the N -length vector θ . Thus, existing optimization techniques concern incoherence. Incoherence is often not satisfied by arbitrary representation dictionaries. As the choice of the sparsifying dictionary is dictated by the nature of the signals we want to measure, one way to improve the structure of the effective dictionary $F = PD$ is the optimization of the projection matrix P . Projections' optimization was first proposed by Elad [57] and involved the improvement of the mutual coherence.

Optimized projections proposed in [57] are based on a “shrinkage” process on the Gram matrix. Suppose we want to obtain CS measurements of a signal that is sparse under a dictionary D . Using a random Gaussian projection matrix P , the sensing mechanism involves the effective dictionary $F = PD$. Let $G = F^T F$ be the corresponding Gram matrix. To improve the mutual coherence, the optimization process “shrinks” the values of the off-diagonal elements of the Gram matrix in order to reduce the correlation between the columns of F . Entries in G with magnitude above a threshold t are “shrunk” by a factor γ . Entries with magnitude below t but above γt are “shrunk” by a smaller amount. Let g_{ij} be the (i, j) entry of the initial Gram matrix. The new Gram matrix elements, \hat{g}_{ij} , are obtained according to

$$\hat{g}_{ij} = \begin{cases} \gamma g_{ij}, & |g_{ij}| \geq t, \\ \gamma t \cdot \text{sgn}(g_{ij}), & t > |g_{ij}| \geq \gamma t, \\ g_{ij}, & \gamma t > |g_{ij}|. \end{cases} \quad (5.13)$$

The “shrinkage” process is applied iteratively. The new Gram matrix yields an effective dictionary \hat{F} with improved mutual coherence. The optimized projection matrix is obtained solving the least squares problem $\min_P \|PD - \hat{F}\|$.

Elad’s technique provoked several algorithms for projections’ optimization each of them employed a different “shrinkage” process on the off-diagonal entries of the Gram matrix [145, 82]. In [145] the authors modify the Gram matrix according to

1. $\hat{g}_{ij} = \begin{cases} 1, & i = j, \\ g_{ij}, & |g_{ij}| < \mu_G, \\ \text{sgn}(g_{ij}) \cdot \mu_G, & \text{otherwise,} \end{cases}$
2. $G_{p+1} = \lambda G_p + (1 - \lambda)G_{p-1}, \quad 0 < \lambda < 1,$

where μ_G is the lowest possible achievable correlation (eq. (2.25)) and G_p is the Gram matrix in the p -th iteration. Similarly, in [82] the proposed “shrinkage” operation is given by the following formula,

$$\hat{g}_{ij} = \text{sgn}(g_{ij})(|g_{ij}| - 0.5 \cdot g_{ij}^2).$$

A similar approach is presented in [55]. Here, the authors’ goal is to produce a Gram matrix that is as close as possible to the identity matrix, introducing the minimization problem

$$\min_F \|F^T F - I\|_{\mathcal{F}}, \quad (5.14)$$

where $\|\cdot\|_{\mathcal{F}}$ denotes the Frobenius norm and I the $N \times N$ identity matrix. Their solution, based on SVD, can work for either the single optimization of the projection matrix given the dictionary or the joint design and optimization of the dictionary and the projection matrix, from a set of training images. In the latter case the authors combine their method with K-SVD [2]. If the dictionary learning process is omitted, the projection matrix optimization is very fast, in contrast to most existing methods that lead to iterative algorithms. Problem (5.14) is also treated in [1], where a solution based on gradient descent is proposed.

5.3 Compressed sensing with the proposed frame constructions

In compressed sensing, we may consider either naturally sparse signals or signals that are sparse with respect to a representation dictionary D . For naturally sparse signals, we employ a projection matrix with the desired properties and take measurements according to $y = P\theta$. If the treated signals are sparse under a representation dictionary D , then the sensing process is described by $y = PD\theta$. In this case, we may consider the product $F = PD$ and optimize F over P such that the projection matrix yields an effective dictionary satisfying the desired properties. In this section we directly employ the proposed frame constructions as projection matrices. The latter consideration involving optimized projections is presented in the next section.

Considering the high incoherence level and the small spectral norm of the frame constructions proposed in Chapter 3, it is of interest to investigate their performance in recovering sparse signals obtained with compressed sensing and compare them with random Gaussian matrices. Therefore, the experiments presented here involve projection matrices of the form of random Gaussian matrices, incoherent UNTFs, nearly equiangular frames and nearly equiangular, nearly tight frames. For the construction of an incoherent UNTF we employ Algorithm 1, while for the construction of nearly equiangular frames we employ Algorithm 5 and Algorithm 6.

Our simulations involve synthetic sparse signals θ of length $N = 120$, with $s = 4$ nonzero coefficients. Considering a projection matrix P of size $m \times N$, with $m = 15 : 5 : 35$ and $N = 120$, we obtain measurements according to $y = P\theta$. The obtained measurements

Table 5.1: Recovery rates for sparse signals of length $N = 120$ obtained with CS, for variable number of measurements, $m = 15 : 5 : 35$, and various types of projection matrices.

m	MSE			
	Gaussian	Alg. 1	Alg. 5	Alg. 6
15	0.01000	0.00821	0.00837	0.00825
20	0.00506	0.00287	0.00300	0.00287
25	0.00180	0.00056	0.00059	0.00059
30	0.00038	$5.650 \cdot 10^{-5}$	$5.609 \cdot 10^{-5}$	$6.887 \cdot 10^{-5}$
35	$9.115 \cdot 10^{-5}$	$3.300 \cdot 10^{-6}$	$2.768 \cdot 10^{-6}$	$5.071 \cdot 10^{-6}$

Table 5.2: Properties of sensing matrices employed in CS experiments. Results involve $m \times N$ matrices with $m \in \{20, 30\}$, $N = 120$.

		Mutual coh.		Average coh.		Spectral norm	
		20	30	20	30	20	30
Type	m						
	Gaussian	0.751	0.647	0.050	0.033	3.290	2.876
	Alg. 1	0.354	0.237	0.042	0.025	2.449	2.000
	Alg. 5	0.463	0.332	0.042	0.025	2.512	2.075
	Alg. 6	0.445	0.319	0.042	0.025	2.459	2.015

are used to find the “unknown” sparse signal, using OMP. For every value of m , we perform 10000 experiments. The quality of the recovered signal is measured computing the Mean Squared Error (MSE). The results demonstrated in Table 5.1 include average values. According to Table 5.1, all proposed frames outperform random Gaussian matrices, improving reconstruction accuracy substantially. In agreement with the established theory, the results depend on the number of acquired measurements, with all types of the proposed frames attaining similar quality of reconstruction for given m .

In order to associate the obtained results for sparse recovery with the properties of the employed projection matrices, we also present results concerning mutual coherence, average coherence and spectral norm. Table 5.2 includes average values over 10000 realizations for projection matrices with dimensions 20×120 and 30×120 . According to these results, the superiority of the proposed frames against random Gaussian matrices is plausible, considering mainly the attained incoherence level. Compared to random Gaussian matrices, nearly equiangular frames produced with Algorithm 5 and Algorithm 6 exhibit reduced mutual coherence by a factor 40 – 50%, while the improvement for incoherent UNTFs produced with Algorithm 1 is higher than 50%.

The next important observations concern a comparison between the proposed frame

constructions. Incoherent UNTFs obtained with Algorithm 1 attain optimal values of spectral norm and the smallest values of mutual coherence, especially when the frames are of high redundancy. According to theoretical results presented in previous sections (see Theorem 4.4.1), one could expect that these frames would yield the highest reconstruction accuracy. However, this is not confirmed by the demonstrated results, which show that the attained reconstruction accuracy is not analogous to the improvement of the aforementioned properties of the employed projection matrices. These results are not that surprising, if we take into account that many authors have argued that mutual coherence may not express well the effectiveness of a matrix in sparse signal recovery [57, 8, 9]. Clearly, other properties of the projection matrix such as average coherence seem to influence the effectiveness of the employed matrix as well. Recall that the notion of average coherence was introduced in [8, 9], where the authors studied its relation to mutual coherence and provided probabilistic guarantees for sparse recovery. While incoherent UNTFs exhibit the smallest mutual coherence and spectral norm values, the values of average coherence are identical for all matrix constructions except from random Gaussian matrices. We conclude that the results obtained in Tables 5.1 and 5.2 indicate that the effectiveness of a matrix involved in sparse recovery seems to depend on all aforementioned properties, with average coherence playing a rather important role.

Concluding, we would like to make a comment concerning the computational cost of Algorithm 5 and Algorithm 6. While Algorithm 6 produces frames with better spectral norm, the achieved improvement slightly affects the reconstruction performance of OMP. Taking into account the additional computational cost introduced by Algorithm 6 and the fact that the matrices employed in CS are practically of large dimensions, we suggest Algorithm 5 as the best choice for the construction of sensing matrices, considering both effectiveness and computational cost. Comparison between Algorithm 5 and Algorithm 1 leads to a similar conclusion, strengthening our preference to Algorithm 5, especially when the application necessitates limitation of resources.

5.4 Proposed optimized projections

Another way to employ the proposed frame constructions in compressed sensing is the method of optimized projections. The method proposed here is based on the alternating and averaged projections algorithms presented in Chapter 3 that produce incoherent UNTFs. As we will explain in the sequel, nearly equiangular frames can be employed to obtain optimized projections in a similar way, with the restriction that the treated signals are sparse under an orthonormal basis.

Despite the existence of theoretical results that highlight the important role of spectral norm, none of the existing methods for the optimization of the projection matrix aims at the construction of effective dictionaries that form tight frames. Tightness was firstly introduced in the optimization of the projection matrix in our preliminary work [128]. Nevertheless, our initial concern when we proposed the algorithm in [128] involved

minimizing the mutual coherence rather than attaining tightness. Based on the observation that the best incoherence levels are obtained by ETFs, which, besides small mutual coherence, also exhibit minimal spectral norm, the algorithm proposed [128] is our first attempt to produce frames close to ETFs.

In optimized projections, we consider the product of the projection matrix and the representation dictionary, that is, $F = PD$, and optimize F over P . The method developed in [128] involves the following operations on the effective dictionary: First, we apply the “shrinkage” process proposed in [57] (see eq. (5.13)) and obtain an effective dictionary with better mutual coherence. Then, we improve the spectral norm of the obtained dictionary finding the nearest (N/m) -tight frame according to Theorem 3.3.1. A third step involves computing the optimized projection matrix solving the minimization problem $\min_P \|PD - F\|$. Aiming at the improvement of this algorithm, we were led to the construction of incoherent UNTFs proposed in [129].

The appropriateness of the projection matrices proposed in [128] is confirmed by the results established in Theorem 4.4.1 [125]. Rephrasing Theorem 4.4.1 to apply to CS, we consider CS measurements of a sparse signal $\theta \in \mathbb{R}^N$ under a dictionary $D \in \mathbb{R}^{K \times N}$, $K \leq N$, according to $y = PD\theta$, where $P \in \mathbb{R}^{m \times K}$, $m \ll K$, is the projection matrix. Theorem 4.4.1 states that θ can be recovered with high probability from $\mathcal{O}(s \log N)$ measurements as long as the effective dictionary $F = PD$ forms an incoherent UNTF. Consequently, an optimization of F over P involves the computation of a projection matrix P such that F is as close to an incoherent UNTF as possible.

Considering the existing optimization techniques for projection matrices and the results established in Theorem 4.4.1, the main steps of an algorithm that leads to optimized projections may be the following:

1. Initialize projections with a random Gaussian matrix P_{init} and compute the initial effective dictionary $F = P_{\text{init}}D$.
2. Apply an algorithm that modifies F to obtain a frame \tilde{F} exhibiting small mutual coherence and spectral norm.
3. Obtain P_{opt} solving $\min_P \|PD - \tilde{F}\|$.

Step 2 can be realized using one of the algorithms presented in Chapter 3. However, the third step of the above process involves the solution of a least squares problem. The obtained solution depends on the sparse representation dictionary D and the computed frame \tilde{F} . When the sparse representation dictionary is redundant, that is, $K < N$, minimization of $\|PD - \tilde{F}\|$ yields an approximate solution P_{opt} . Experiments have shown that if the obtained frame \tilde{F} is an incoherent UNTF constructed with Algorithm 1 or Algorithm 2, then the optimized P_{opt} yields an effective dictionary $F_{\text{opt}} = P_{\text{opt}}D$ that is close to \tilde{F} ; indeed, F_{opt} forms an incoherent UNTF. On the other hand, if the obtained frame \tilde{F} is a nearly equiangular frame, then the optimized P_{opt} yields an effective dictionary that is far from \tilde{F} . In this case, F_{opt} does not exhibit the properties of a nearly equiangular frame.

A nearly equiangular frame could be used to produce optimized projections for signals that are sparse under orthonormal bases, that is, when $K = N$. Then, minimization of $\|PD - \tilde{F}\|$ results in a projection matrix P_{opt} satisfying $P_{\text{opt}}D = \tilde{F}$. In order to present a general solution concerning sparse signals under redundant representation dictionaries and orthonormal bases as well, the optimization method that follows employs algorithms yielding incoherent UNTFs.

5.4.1 Optimized projections using incoherent UNTFs

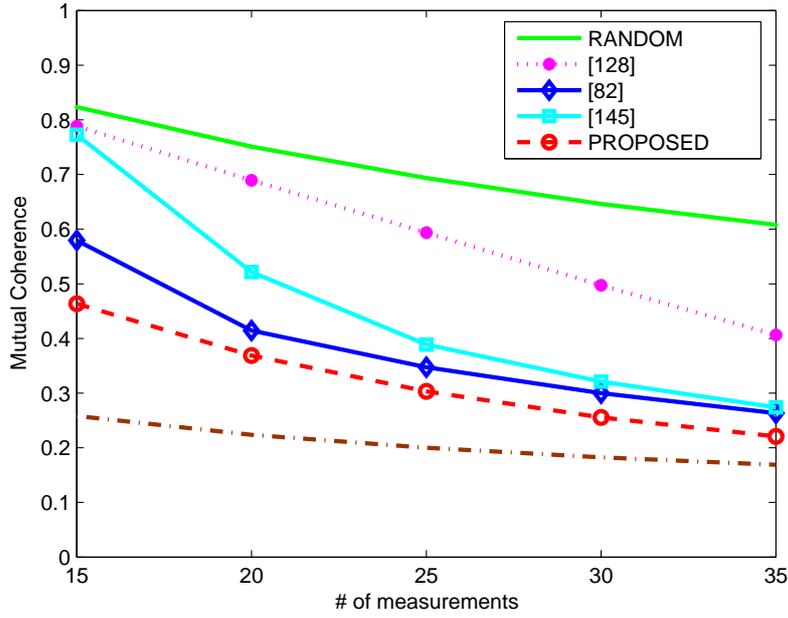
In Chapter 3 we presented two algorithms for constructing incoherent UNTFs. Both algorithms yield similar constructions; therefore, we have decided to employ only one of them in the experiments presented here. We choose the proposed alternating projections (Alg. 1), as it exhibits higher convergence speed. Algorithm 1 is slightly modified to incorporate the optimization step producing the optimized projection matrix. The method yields effective dictionaries with small mutual coherence and small spectral norm.

In our experiments, the proposed optimized projections are compared to our preliminary work [128] and existing constructions presented in [145] and [82]. Although our experiments included the methods of [57], [55] and [1] as well, we only report results with the methods of [145] and [82] since they seem to perform better.

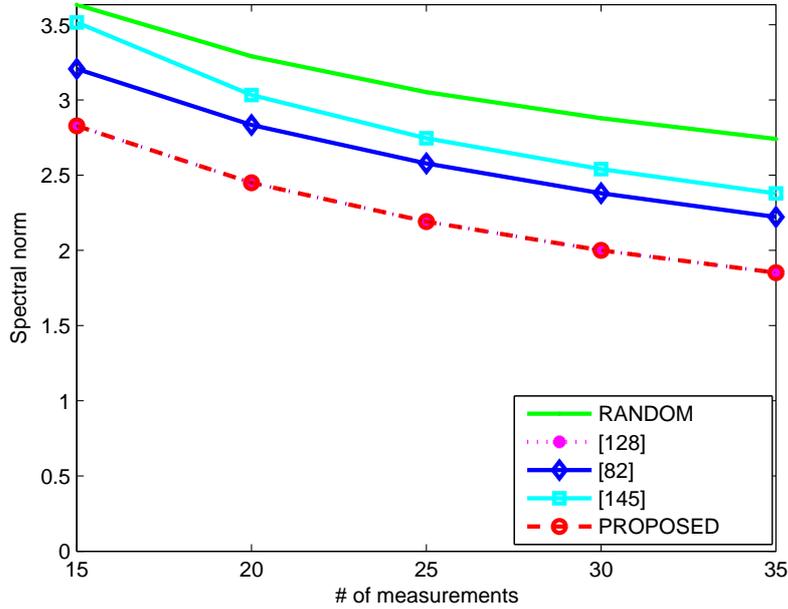
The properties of the effective dictionary

Before proceeding to reconstruction performance of algorithms employed in CS, let us present some results that demonstrate the properties of the obtained incoherent dictionary constructions. The reconstruction experiments that follow involve varying number of measurements, thus, we present here results for $m \times N$ dictionaries with $m = 15 : 5 : 35$ and $N = 120$. For every value of m , we carry out 10000 experiments, in which we construct incoherent matrices with all the methods involved in our CS simulations; All algorithms are executed performing 50 iterations. The properties we are interested in include mutual coherence and spectral norm.

Average results for the mutual coherence are presented in Fig. 5.1(a). We can see that the proposed method leads to a significant reduction of the mutual coherence of the initial matrix by a factor depending on redundancy ($\rho = N/m$). Achieved mutual coherence becomes closer to the lowest possible bound when redundancy decreases (the brown dash-dotted line, in Fig. 5.1(a) stands for the lowest possible bound (see eq. (2.25)). This is a very significant improvement compared to the results of our work in [128] and the other methods presented here. The fact that the proposed method performs well even for very redundant frames is an important advantage over the other competing methods. In Fig. 5.1(b) we demonstrate the spectral norm of the frames under testing, answering the question “how close are the obtained constructions to UNTFs?”. The measurements corresponding to the proposed incoherent UNTFs and our preliminary construction [128] coincide with the lowest bound N/m , confirming that the proposed methodology leads to



(a)



(b)

Figure 5.1: Properties of the effective dictionaries involved in CS reconstruction experiments. In (a) we present mutual coherence as a function of the number of measurements. The bottom brown dash-dotted line represents the lowest possible bound (see eq. (2.25)). In (b) we present spectral norm as a function of the number of measurements. The red dotted line corresponding to our methodology coincide with the lowest possible bound.

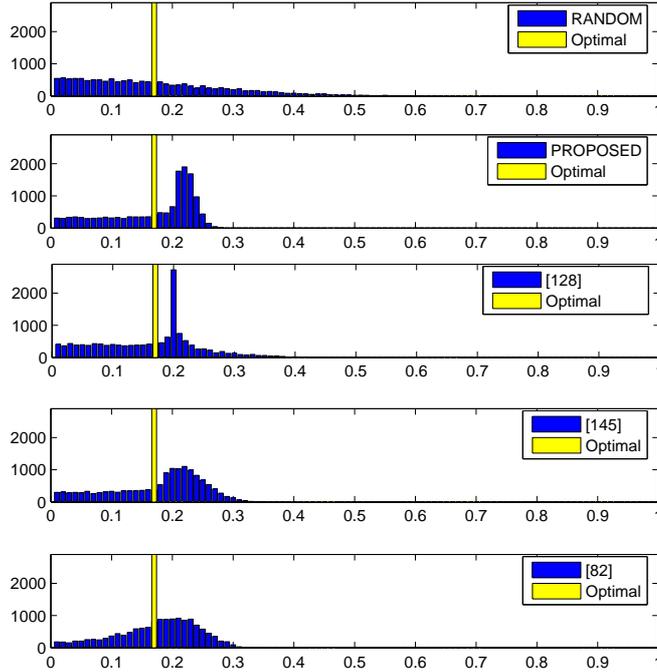


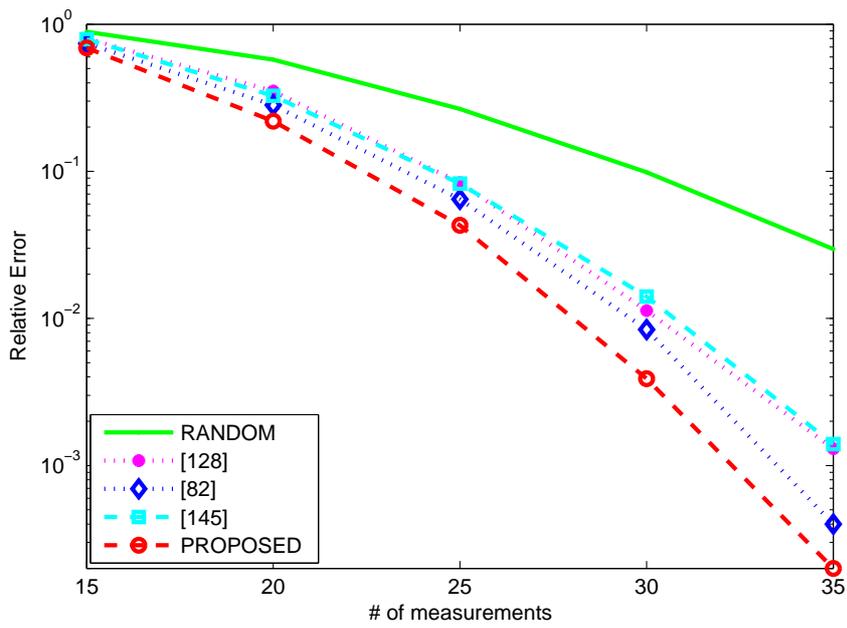
Figure 5.2: Changes in the distribution of the column correlation of a 25×120 frame.

UNTFs.

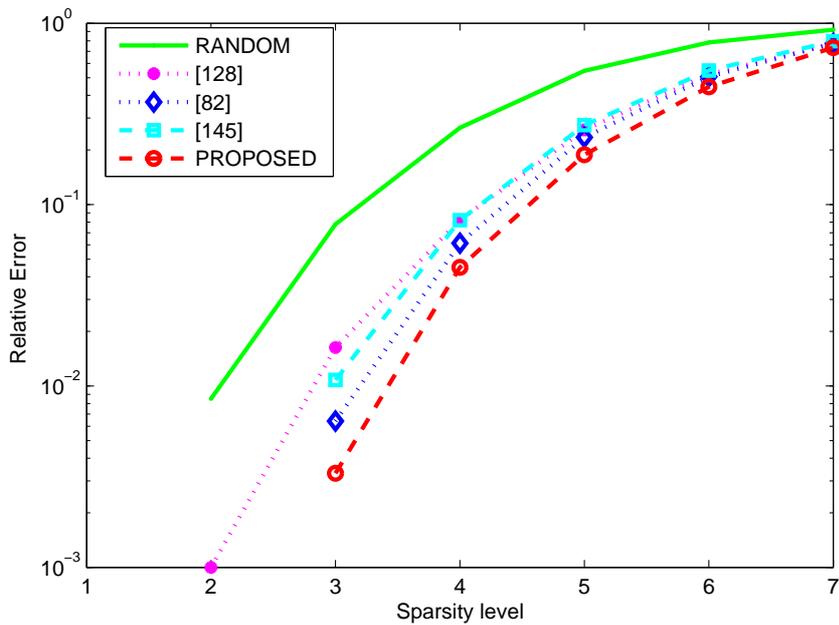
Another way to evaluate the obtained incoherent dictionaries is to consider the distribution of the inner products between distinct columns. Figure 5.2 illustrates a representative example of a 25×120 matrix. The histogram depicts the distribution of the absolute values of the corresponding Gram matrix entries. The results concern the initial matrix and all matrices produced by the employed iterative algorithms, after 50 iterations. The yellow bar rises at the critical interval that includes the minimal achievable correlation, corresponding to the distribution of an optimal Grassmannian frame (the bar's actual height is constrained for clear demonstration of the methods under testing). The proposed method exhibits a significant concentration near the critical interval, combined with a short tail after it, showing that the number of the Gram entries that are closer to the ideal Welch bound is larger than in any other method presented here. Such a result is in agreement with the small mutual coherence values depicted in Fig. 5.1(a).

CS performance

Let us now continue with CS simulations. For each experiment, we generate an s -sparse vector $\theta \in \mathbb{R}^N$ of length N , which constitutes a sparse representation of the K -length synthetic signal $x = D\theta$, $x \in \mathbb{R}^K$, $K \leq N$. We choose the dictionary $D \in \mathbb{R}^{K \times N}$ to be a random Gaussian matrix. Experiments with DCT dictionaries lead to similar results. The locations of the nonzero coefficients in the sparse vector are chosen at random. Besides

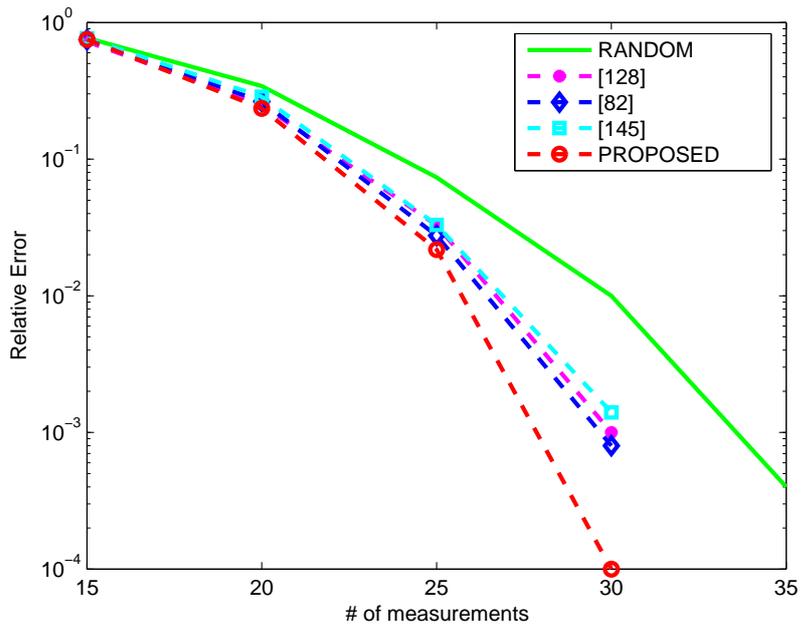


(a)

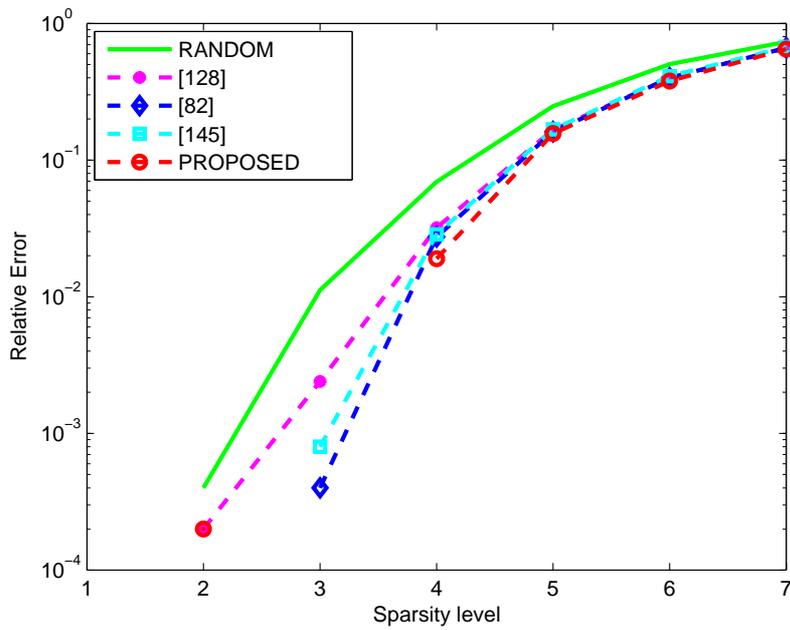


(b)

Figure 5.3: CS performance for random and optimized projection matrices by means of relative MSE in a logarithmic scale. Numerical recovery deploys OMP. In (a) we keep the sparsity level fixed and vary the number of measurements. In (b) we keep the number of measurements fixed and vary the sparsity level. A vanishing graph implies a zero error rate.



(a)



(b)

Figure 5.4: CS performance for random and optimized projection matrices by means of relative MSE in a logarithmic scale. Numerical recovery deploys BP. In (a) we keep the sparsity level fixed and vary the number of measurements. In (b) we keep the number of measurements fixed and vary the sparsity level. A vanishing graph implies a zero error rate.

the effectiveness of the projection matrix P , the reconstruction results also depend on the number of measurements m and the sparsity level of the representation s . Thus, our experiments include varying values of these two parameters. For a specified number of measurements $m \ll K$, we create a random projection matrix $P \in \mathbb{R}^{m \times K}$. After the optimization process, we obtain m projections of the original signal according to $y = PD\theta$. We reconstruct the original sparse signal with OMP and BP.

In all experiments presented here, the synthetic signals are of length $K = 80$ and the respective sparse representations, under the dictionary D , of length $N = 120$. The execution of the optimization algorithm included up to 50 iterations. Two sets of experiments have been considered; the first one includes varying values of the number of measurements m and the second one includes varying values of the sparsity level s of the treated signals. For every value of the aforementioned parameters we perform 10000 experiments and calculate the relative error rate; if the mean squared error of a reconstruction exceeds a threshold of order $\mathcal{O}(10^{-4})$, the reconstruction is considered to be a failure.

Figure 5.3 demonstrates results for OMP. Figure 5.3 (a) presents the relative errors as a function of the number of measurements m , for a fixed sparsity level ($s = 4$) of the treated signal. Figure 5.3 (b) presents the relative errors for a fixed number of measurements ($m = 25$) and varying values of the sparsity level of the signal. It is clear that the projections matrix obtained with the proposed algorithm leads to better reconstruction results compared to random matrices and to matrices produced by the other methods. The observed results are due to the improvement in the effective dictionary properties. Similar results for BP are demonstrated in Figure 5.4.

An important observation regarding CS performance, we have also made in the previous section, is that although we achieved a high quality of reconstruction, the fact that for some values of measurements (e.g., 15) this improvement is not of the same order as the improvement in the mutual coherence, indicates that additional properties should be taken into consideration to decide about the appropriateness of the effective dictionary. This has been pointed out by other authors [57, 55] as well and should be explored both theoretically and experimentally.

5.5 Preconditioning in compressed sensing

Often choosing the projection matrix in a CS application is dictated by specific constraints depending on the application. A major obstacle in most applications is the design of acquisition hardware. Binary random matrices are considered the best option for practical implementation [92, 91]. However, the recovery rates they yield are similar to the ones achieved with random Gaussian matrices at best [11, 91] while certain types of binary projections work well only when combined with specific representation dictionaries [54].

Motivated by the improved performance of sparse recovery algorithms in sparse representations when preconditioning is applied, for the first time to the best of our knowledge, we propose the use of preconditioning in compressed sensing [130]. When sparse signals

are acquired with binary projections, preconditioning can improve the incoherence of the effective dictionary leading to higher accuracy in sparse recovery.

The goal of preconditioning is to transform the linear system describing the measurement process, $y = PD\theta$, $P \in \mathbb{R}^{m \times K}$, $D \in \mathbb{R}^{K \times N}$, into a form that is more suitable for numerical treatment. Employing a preconditioner $C \in \mathbb{R}^{m \times m}$ we obtain the system

$$Cy = CPD\theta \quad \text{or} \quad z = F\theta, \quad (5.15)$$

where $F = CPD$ is the new system matrix. Computing an appropriate preconditioner C is equivalent to constructing a matrix F exhibiting small mutual coherence and small spectral norm. Moreover, the preconditioner C must be an invertible matrix such that $Cy = CPD\theta$ and $y = PD\theta$ are equivalent.

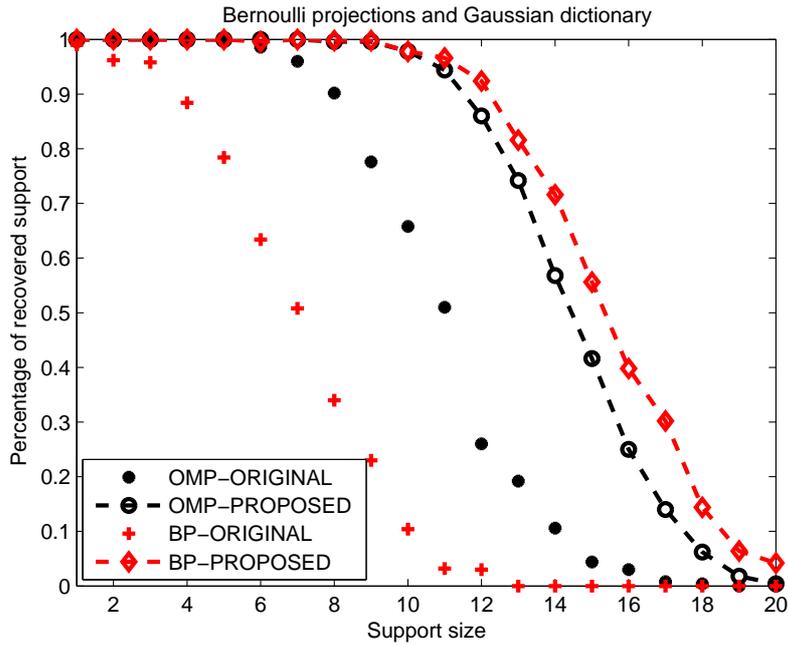
The method developed here is similar to the one proposed in sparse representations. Initializing the preconditioner C with a random Gaussian matrix, the effective dictionary is modified such as the new system matrix $F = CPD$ forms an incoherent UNTF.

1. Initialize preconditioner with a random Gaussian matrix $C = C_{\text{init}}$.
2. Compute the new system matrix $F = CPD$.
3. Modify F such that it forms an incoherent UNTF \tilde{F} .
4. Compute a suitable preconditioner C solving $\min_C \|CPD - \tilde{F}\|$.

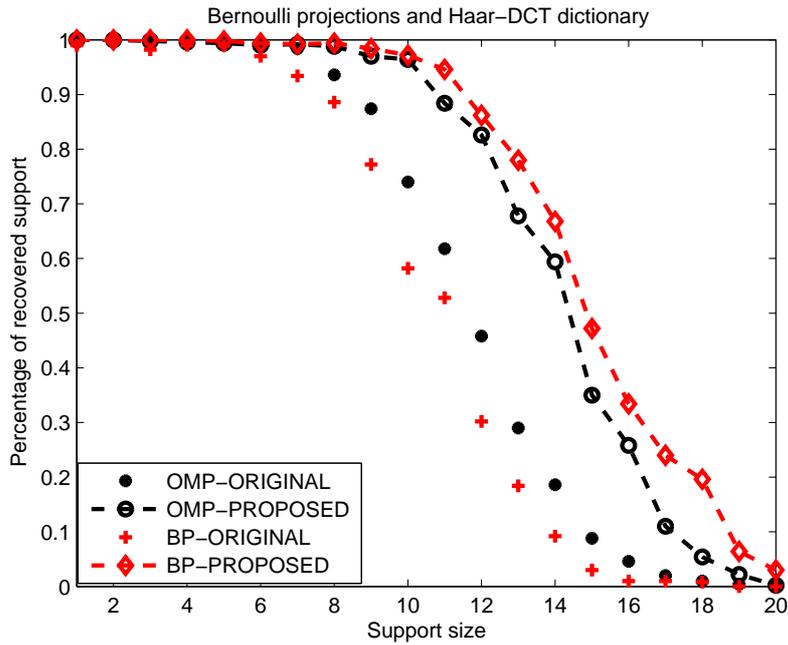
To produce an incoherent UNTF, we employ Algorithm 1 proposed in Chapter 3. Algorithm 1 is modified to incorporate the last step described in the above process such as a preconditioner C is computed in every iteration. We cannot guarantee that the above algorithm yields an invertible matrix C . However, according to our analysis in [129], there is strong evidence that the algorithm converges locally, meaning that the output matrix C is close to the initial matrix C_{init} . Having selected an invertible initial matrix, the probability that the obtained matrix is singular is very small.

Experimental results

In our experiments we consider a practical problem, assuming that the sensing mechanism is implemented by a binary random matrix obtained from a Bernoulli $(0, 1)$ distribution. The first group of experiments involves sparse representation dictionaries D realized by random Gaussian matrices of size 128×256 , while the second group of experiments involves overcomplete Haar-DCT dictionaries of size 128×255 . Assuming sparse signal under the concerned representation dictionary, we construct synthetic signals θ of length $N = 256$ or $N = 255$ depending on the employed dictionary, with varying sparsity level. Signal acquisition is performed according to $y = PD\theta$, where P is a 64×128 random projection matrix with entries $0, 1$. Recovery of the ‘‘unknown’’ θ is performed using OMP and BP. preconditioning is initialized by a 64×64 random Gaussian matrix and is obtained following the steps described above. The performance of the deployed



(a)



(b)

Figure 5.5: Support recovery rates for OMP and BP, for signals with varying support size acquired with Bernoulli random projections. The signals considered in (a) are sparse under a random Gaussian dictionary. The signals considered in (b) are sparse under a Haar-DCT dictionary.

Table 5.3: Recovery rates for CS with Bernoulli and optimized projections. When Bernoulli projections are used, recovery involves preconditioning.

Support Size	OMP		BP	
	Bernoulli-Prec.	Optimized	Bernoulli-Prec.	Optimized
4	1.000	1.000	1.000	1.000
8	0.996	1.000	0.998	1.000
12	0.860	0.870	0.924	0.928
16	0.250	0.248	0.398	0.380
20	0.004	0.006	0.042	0.054

algorithms is quantified by computing the percentage of fully recovered support, referred to as recovery rate.

For the first group of experiments concerning sparse synthetic signals under random Gaussian dictionaries, recovery rates for OMP and BP are presented in Fig. 5.5(a). Averaged over 500 realizations, the results show that preconditioning yields significant improvement in the performance of OMP, and particularly of BP, implying that the proposed technique can be applied successfully in CS. For the second group of experiments concerning sparse synthetic signals under Haar-DCT dictionaries, the recovery rates obtained for OMP and BP are presented in Fig. 5.5(b), confirming that preconditioning can substantially improve the performance of the deployed algorithms.

For further evaluation of the proposed technique, we compare the above results with optimized projections. We consider the first group of experiments, concerning sparse signals under 128×256 random Gaussian dictionaries, and acquire these signals with optimized projection matrices obtained with the method described in the previous section. Table 5.3 demonstrates recovery rates for OMP and BP. The results are similar for both methods, showing that the performance of the deployed algorithms when used with Bernoulli projections and preconditioning is comparable to optimized projections. Considering that Bernoulli matrices are more convenient for hardware implementation, this is an important result for practical compressed signal acquisition.

CHAPTER 6

SPREADING SEQUENCES FOR S-CDMA

6.1 S-CDMA model

6.2 Design of spreading sequences

6.3 Optimal spreading sequences for varying number of users

6.4 Codebooks from nearly equiangular, nearly tight frames

Code Division Multiple Access (CDMA) is an important multiple access technique in wireless networks and other common channel communication systems where a number of users transmit their data using the same physical channel. To distinguish each user from the other, every user is assigned a *code*, also known as *spreading sequence*, which he uses to spread its information on the common channel through modulation. In symbol-synchronous CDMA (s-CDMA) systems, all users are in exact synchronism relative to the receiver, that is, their data symbols are alligned in time. The receiver demodulates the transmitted message upon observing the sum of the transmitted signals embedded in noise.

Our main concern in such systems is to achieve reliable and fair communication using maximum sum rate. The set of information rates at which the users can transmit while retaining reliable transmission is known as *capacity region*. The information theoretic capacity region of Gaussian multiple access channels was addressed in [135] where it was characterized as a function of spreading sequences and average input power constraints of the users. It was suggested in [135] that the choice of the *spreading sequence set* or the *codebook* is left open to the designer of the CDMA system; the spreading sequences could be optimized given the constraints of the problem.

Optimal spreading sequences maximize the *sum capacity*, which is defined as the maximum sum of achievable rates of all users per unit processing gain and the maximum is taken over all choices of spreading sequences. According to results from [111], [136], optimal codebooks are fundamentally a function of the number of active users and the

number of chips. In [111] the authors proved that the spreading sequences that maximize the sum capacity are the ones that minimize the interuser interference. These codebooks form equal norm tight frames, which are also referred to as Welch Bound Equality (WBE) sequences [95].

While WBE sequence sets are of considerable interest in CDMA communication systems, we must note that the properties of a WBE sequence set do not always apply to subsets, meaning that a codebook designed for a specific number of users is no longer optimal, if some users are silent [112, 132]. Therefore, it is of interest to find sequence sets that perform well even when subsets of the available codes are active. This problem was addressed in [77, 76], where the authors constructed codebooks from equiangular tight frames (ETFs) and proved that such codebooks are less sensitive to changes in the number of active users. However, the codebooks proposed in [76], based on conference matrices (see section 2.4), are restricted to certain dimensions.

In this Chapter, first, we briefly review well-known results regarding the design of spreading sequences and characterize optimal spreading sequences for s-CDMA systems. Then, we employ as spreading sequences the proposed nearly equiangular, nearly tight frames and study their performance.

6.1 S-CDMA model

Consider a discrete time symbol synchronous CDMA system with K independent users and processing gain L . The K users want to transmit their information symbols $B^{(k)}$, $k = 1, \dots, K$. Each user is assigned an individual real spreading sequence $s^{(k)}$ of length L , that is, $s^{(k)} = [s_1^{(k)}, s_2^{(k)}, \dots, s_L^{(k)}]$, where L is known as the *spreading factor* of the spread-spectrum system. Each spreading sequence $s^{(k)}$ is assumed to have energy L , i.e.,

$$\langle s^{(k)}, s^{(k)} \rangle = L. \quad (6.1)$$

The users encode their information into real ± 1 valued symbols $B^{(k)}$, which are assumed to be independent Gaussian random variables, with $E[|B^{(k)}|^2] = 1$. In the i -th symbol interval, the users spread their real-valued encoded symbols $B_i^{(k)}$, $k = 1, \dots, K$, by the spreading sequences $s^{(k)}$ and then transmit the L -dimensional symbols

$$B_i^{(k)} s^{(k)} = [B_i^{(k)} s_1^{(k)}, B_i^{(k)} s_2^{(k)}, \dots, B_i^{(k)} s_L^{(k)}].$$

In this manner, the k -th user creates the sequence

$$\dots, B_{-1}^{(k)} s^{(k)}, B_0^{(k)} s^{(k)}, B_1^{(k)} s^{(k)}, \dots$$

Transmitting over a Gaussian multiple access channel and assuming perfect synchronization, the receiver during the i -th symbol period observes the i -th data symbol

$$r_i = w \sum_{k=1}^K B_i^{(k)} s^{(k)} + n_i, \quad (6.2)$$

where w is the received power, assumed the same for all users, n_i is a zero mean Gaussian random vector with correlation matrix $E[NN^T] = n^2 I_L$, and I_L denotes the $L \times L$ identity matrix.

6.2 Design of spreading sequences

Optimal spreading sequences maximize the sum capacity and lead to minimum interuser interference experienced by each user. In [95] Massey and Mittelholzer first identified that spreading sequence sets that minimize interuser interference exhibit minimum total squared correlation. The sequence sets having this property were identified as WBE sequences [95]. Considering the problem of maximizing the capacity of s-CDMA systems, it was shown in [111] that sum capacity is maximized precisely by the same WBE sequences.

6.2.1 Interuser Interference

The observed sequence $r_i = [r_{i1}, r_{i2}, \dots, r_{iL}]$ at the receiver is correlated with the spreading sequence $s^{(k)}$ to produce the detection statistic $S_i^{(k)}$ for the user k ,

$$S_i^{(k)} = \langle r_i, s^{(k)} \rangle = \sum_{j=1}^L r_{ij} s_j^{(k)}.$$

Assuming that $\langle s^{(k)}, s^{(k)} \rangle = L$, the data symbol detection statistic for the user k becomes

$$S_i^{(k)} = wLB_i^{(k)} + w \sum_{\substack{\ell=1 \\ \ell \neq k}}^K B_i^{(\ell)} \langle s^{(k)}, s^{(\ell)} \rangle + \eta_i^{(k)}, \quad (6.3)$$

where $\eta_i^{(k)} = \langle n_i, s^{(k)} \rangle$. The sum

$$\xi_i^{(k)} = w \sum_{\substack{\ell=1 \\ \ell \neq k}}^K B_i^{(\ell)} \langle s^{(k)}, s^{(\ell)} \rangle \quad (6.4)$$

represents the *interuser interference* experienced by the user k . Because the data symbols of the K users are themselves statistically independent and each has mean 0 and variance 1, the interuser interference given by the sum (6.4) has mean 0 and variance

$$\sigma^2(k) = \sum_{\substack{\ell=1 \\ \ell \neq k}}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2. \quad (6.5)$$

The term $\sigma^2(k)$ is also referred to as *interference power*. Equation (6.5) can also be written in the form

$$\sigma^2(k) = \sum_{\ell=1}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2 - L^2. \quad (6.6)$$

The interference caused by the spreading sequences has an effect on the quality of transmission, reducing the signal-to-noise plus interference (SINR) ratio

$$\text{SINR}(k) = \frac{1}{\frac{n^2}{w^2} + \sum_{\ell=1, \ell \neq k}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2}, \quad k = 1, \dots, K. \quad (6.7)$$

Increasing interference results in performance degradation of the s-CDMA system.

Therefore, the sequence design problem for s-CDMA can be formulated as follows:

Problem 6.2.1 (Minimize worst interuser interference). *Choose sequences $s^{(1)}, s^{(2)}, \dots, s^{(K)}$ of length L to minimize*

$$\sigma_{wc}^2 = \max_k \sigma^2(k) = \max_k \sum_{\ell=1}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2 - L^2, \quad (6.8)$$

where σ_{wc} stands for the worst interuser interference.

The optimally solution to problem 6.2.1 will result from a solution, when it exists, to the following problem:

Problem 6.2.2 (Minimize Total Squared Correlation). *Choose sequences $s^{(1)}, s^{(2)}, \dots, s^{(K)}$ of length L to minimize*

$$\sigma_{TOT}^2 = \sum_{k=1}^K \sum_{\ell=1}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2 - KL^2. \quad (6.9)$$

It is easy to show that the necessary and sufficient condition for no interuser interference is

$$\langle s^{(k)}, s^{(\ell)} \rangle = 0, \quad \text{for all } k \neq \ell. \quad (6.10)$$

However, this holds only when $K \leq L$, since there can be at most L orthogonal non-zero sequences of length L .

6.2.2 Welch Bound Equality (WBE) sequences

While orthogonal sequences eliminate interuser interference, it has been shown that non-orthogonal codes are sum capacity optimal. A quarter-century ago, Welch [143] published a collection of lower bounds on the maximum magnitude of the inner products of a set of vectors. One of the main results of [143] concerns lower bounds for the $2m$ -th power of the sum of the inner products between pairs of vectors

$$\sum_{k=1}^K \sum_{\ell=1}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^{2m} \geq \frac{K^2 L^{2m}}{\binom{L+m-1}{m}}. \quad (6.11)$$

Setting $m = 1$ in (6.11), we obtain the Welch bound on the total squared correlation. In [95], Massey and Mittelholzer provided a simple derivation of this bound and first stated the condition for equality.

Theorem 6.2.1 (Bound Total Squared Correlation [95]). *If $s^{(1)}, s^{(2)}, \dots, s^{(K)}$ are sequences in \mathbb{C}^L and all have the same energy L , i.e.,*

$$\|s^{(k)}\|^2 = \langle s^{(k)}, s^{(k)} \rangle = L, \quad k = 1, \dots, K, \quad (6.12)$$

then

$$\sum_{k=1}^K \sum_{\ell=1}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2 \geq K^2 L, \quad (6.13)$$

with equality if and only if the rows $r^{(1)}, r^{(2)}, \dots, r^{(L)}$ of the $L \times K$ array whose columns are $s^{(1)}, s^{(2)}, \dots, s^{(K)}$ are orthogonal and all rows have the same energy, i.e.,

$$\|r^{(\ell)}\|^2 = K, \quad \ell = 1, \dots, L. \quad (6.14)$$

The sequences satisfying (6.13) with equality are known as *Welch Bound Equality (WBE) sequences* [95]. When equality holds the sequences are also characterized as *uniformly good* [95] in the sense that

$$\sum_{\ell=1}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2 = KL, \quad k = 1, \dots, K. \quad (6.15)$$

Recall that the sum in equation (6.15) expresses the variance $\sigma^2(k)$ of the interuser interference (see (6.5)). Therefore, WBE sequences designed for K users when employed as spreading sequences in s-CDMA yield the same interference for every user. From (6.15), (6.7) we see that the SINR is also constant and depends only on K and L .

6.2.3 Sum capacity

Sum capacity is an important measure of overall information capacity of a multiple access channel. It was shown in [135] that the sum capacity is a function of users' spreading sequences and received powers. Sum capacity optimal spreading sequences have been characterized for Gaussian channels [111], [112], fading channels with white noise [136], fading channels with colored noise [137], [3], and with different receivers [138], [73].

Let S be the $L \times K$ matrix with the users' spreading sequences as its columns, $S = [s^{(1)} \ s^{(2)} \ \dots \ s^{(K)}]$, and $W = \text{diag}\{w_1, w_2, \dots, w_K\}$ be the $K \times K$ diagonal matrix of users' received powers. Considering a multiple access channel with zero mean Gaussian noise with correlation matrix $E[NN^T] = n^2 I_L$, the maximum capacity was derived to be [135]

$$C_{\text{sum}} = \frac{1}{2} \log[\det(I_L + n^{-2} S W S^T)]. \quad (6.16)$$

When the received powers of the users are the same, $w_k = w$ for all k , (6.16) reduces to

$$C_{\text{sum}} = \frac{1}{2} \log[\det(I_L + \frac{w}{n^2} S S^T)]. \quad (6.17)$$

A necessary and sufficient condition to attain (6.17) is [111]

$$S^T S = I_K, \quad \text{when } K \leq L, \quad (6.18)$$

$$SS^T = \frac{K}{L}I_L, \quad \text{when } K \geq L, \quad (6.19)$$

where I_K , I_L are the $K \times K$ and $N \times N$ identity matrices, respectively. Therefore, a spreading sequence set should form a set of orthogonal sequences, if the number of users is equal or less than the processing gain, and a unit norm tight frame (UNTF), otherwise.

In [3] it was shown that a spreading sequence set satisfying (6.17) exhibits also minimum total squared correlation (WBE sequences). Therefore, the problem of maximizing the capacity of an s-CDMA system is equivalent to minimizing interuser interference. In [140] it was shown that WBE sequence sets defined in [95] are precisely equal norm tight frames.

In Chapter 2 we have seen that equal norm tight frames and unit norm tight frames exist for any frame dimensions; thus, maximum sum capacity and minimum interuser interference can be always achieved. Constructions of WBE sequences have been described in [95, 112, 127].

6.3 Optimal spreading sequences for varying number of users

Considering that optimal codebooks are a function of the number of active users, practical application of WBE sequences raises the need of reassignment as the number of active users changes. While a WBE sequence designed for K users is capacity optimal and has a nice interference invariance property, the sequence subset ceases to satisfy Welch's bound with equality if any $M < K$ signatures are removed. Therefore, whenever a user leaves or a new user arrives, the subset of remaining sequences will no longer be optimal [112, 132, 77, 76].

Theorem 6.3.1 ([76]). *Let $S = [s^{(1)}, s^{(2)}, \dots, s^{(K)}]$ be a set of WBE sequences of length L and assume $K > L$. If we remove any $M < L$ sequences from or add any $M < L$ equal norm sequences to this set, then the resulting set does not satisfy the Welch's bound with equality.*

Employing a subset of spreading sequences that are not optimal leads to the undesirable property that users would see different amount of interference as a function of their sequence assignment, which can result in capacity or bit error probability degradations. Thus, a system that fully exploits WBE sequences would need (i) a set of spreading sequences for every possible K and (ii) would need to reassign all sequences every time a user arrived or departed from the system.

To mitigate the problems caused by the loss of the WBE property, in [77, 76] the authors studied equiangular frames and employed them as spreading sequences. Perhaps the most interesting property of equiangular sequence sets is that the total interference power for every sequence is only a function of the current number of active sequences and the original dimensionality of the codebook.

Let \mathcal{K} denote the set of integers $1, 2, \dots, K$, and \mathcal{A} a subset of \mathcal{K} that indexes the active sequences. For an arbitrary active sequence k , the interference power is

$$\sigma^2(k) = \sum_{\substack{\ell \in \mathcal{A} \\ \ell \neq k}} |\langle s^{(k)}, s^{(\ell)} \rangle|^2 = (|\mathcal{A}| - 1)c^2, \quad k = 1, \dots, K, \quad (6.20)$$

where c is the equiangular constant, $c = |\langle s^{(k)}, s^{(\ell)} \rangle|$ for any $\ell \neq k$. Note that (6.20) is independent of k and depends only on the number of the active users given by the cardinality of \mathcal{A} . This byproduct of the equiangular property is stated in the following theorem.

Theorem 6.3.2 (Interference Invariance [77]). *The total interference power for any equiangular sequence set is identical for all sequences and depends only on the total number of active sequences.*

A consequence of this theorem is that ETFs are the best of all equiangular sequences since they achieve the lowest bound on the maximum correlation with equality (and thus have the smallest possible c). Considering that for $L \times K$ ETFs there holds

$$\frac{|\langle s^{(k)}, s^{(\ell)} \rangle|}{\|s^{(k)}\| \|s^{(\ell)}\|} = \sqrt{\frac{K-L}{L(K-1)}}, \quad k \neq \ell, \quad (6.21)$$

we obtain

$$c = |\langle s^{(k)}, s^{(\ell)} \rangle| = \frac{K-L}{L(K-1)}L^2, \quad k \neq \ell, \quad (6.22)$$

where we assumed that $\|s^{(k)}\| = L$, for all k . Therefore, for sequence sets obtained by ETFs, the interference power experienced by the k -th user is

$$\sigma^2(k) = \sum_{\substack{\ell \in \mathcal{A} \\ \ell \neq k}} |\langle s^{(k)}, s^{(\ell)} \rangle|^2 = (|\mathcal{A}| - 1) \frac{K-L}{L(K-1)}L^2, \quad (6.23)$$

which is the same for $k = 1, 2, \dots, K$.

It is clear that ETFs are a subclass of WBE sequences since

$$\begin{aligned} \sum_{k=1}^K \sum_{\ell=1}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2 &= \sum_{k=1}^K \left(\sum_{\substack{\ell=1 \\ \ell \neq k}}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2 + L^2 \right) \\ &= K \left((K-1) \frac{K-L}{L(K-1)}L^2 + L^2 \right) \\ &= K \left(\frac{K-L}{L}L^2 + L^2 \right) \\ &= K^2L. \end{aligned} \quad (6.24)$$

Welch's bound was originally stated as a lower bound on the maximum value of $|\langle s^{(k)}, s^{(\ell)} \rangle|$ for $k \neq \ell$ (see eq. (3.4.1)), also referred to as *maximum* Welch bound. Recall that ETFs satisfy the maximum Welch bound with equality and constitute a very important subclass of WBE sequences, also known as maximal WBE (MWBE) sequences [112].

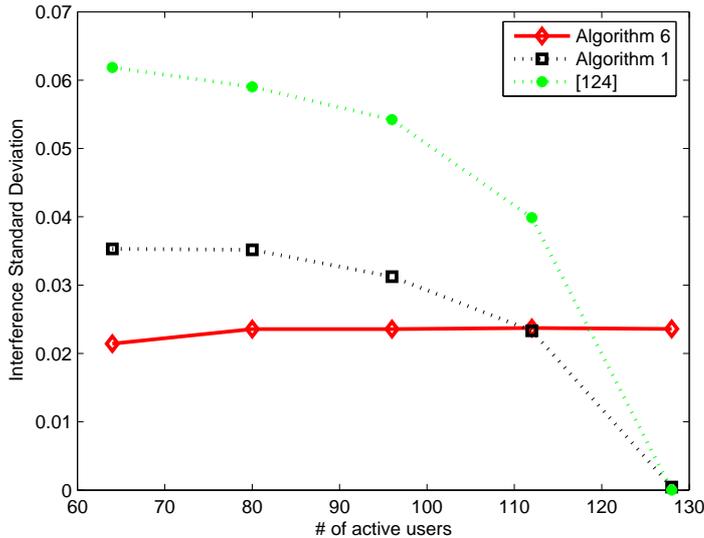


Figure 6.1: Standard deviation of the interference term for variable number of active users in an s-CDMA system designed for 128 users.

6.4 Codebooks from nearly equiangular, nearly tight frames

In Chapter 3 we have seen that Algorithm 6 may produce nearly equiangular, nearly tight frames of any dimensions. As these frames exhibit approximately minimal spectral norm, they are very close to UNTFs; thus, we expect that the frame vectors approximately minimize the total squared correlation as well. Moreover, as the frame vectors exhibit similar correlation, it is expected that the proposed frames lead to similar values of $\sigma^2(k) = \sum_{\ell=1}^K |\langle s^{(k)}, s^{(\ell)} \rangle|^2$, $\ell \neq k$, for all k . As we have discussed, this term is related to the interuser interference and results in variation in the quality of the signal received by the users. Therefore, we propose the employment of nearly equiangular, nearly tight frames as spreading sequences in s-CDMA. We also employ as spreading sequences the proposed incoherent UNTFs produced by Algorithm 1. These frames belong to WBE sequences and are expected to minimize TSC and maximize sum capacity. The proposed frames are compared to a UNTF construction presented in [124] for application to s-CDMA.

Our simulations consider an s-CDMA system with varying activity, that is, the number of users in the system changes, resulting in different subsets of active users. The system is designed for at most $N = 128$ users. The code set includes codes with length 64, thus, it forms a 64×128 frame. For every subset of K active users, the system randomly chooses K frame vectors as codes. The considered subsets of users are of varying size. In every situation, we examine the interference term $\sigma^2(k)$. As a measure of how close we are to the target that all users experience the same interference, we compute the standard deviation of $\sigma^2(k)$. The results are averaged over a series of random trials and are demonstrated in Figure 6.1. Clearly, the obtained results show that nearly equiangular, nearly tight frames, outperform UNTFs (WBE sequences), when the system works with a load up to 85% its total load, exhibiting similar interuser interference for all considered scenarios of activity,

Table 6.1: Average total squared correlation (TSC) for variable number of active users.

# of active users	TSC		
	Algorithm 6	Algorithm 1	[124]
64	95.720	95.742	95.868
80	129.812	129.773	129.706
96	167.769	167.763	167.821
112	209.932	209.876	209.805
128	256.094	256.000	256.000

regardless of the number of active users in the system. On the contrary, concerning UNTF constructions, we see that the smaller the number of active users the higher the variance in interuser interference. However, UNTFs exhibit optimal performance for $K = N$, when $\sigma^2(k)$ is identical for all k , leading to the same interference for every user (see (6.15)).

In Table 6.1, we present average values of the total squared correlation (TSC) observed in the above scenarios of active subsets of users. Both UNTF constructions attain the minimum bound as expected, while nearly equiangular, nearly tight frames exhibit a small discrepancy. As discussed in section 6.2.3, frames that minimize TSC result in optimal sum capacity. Computing the sum capacity corresponding to each frame from (6.17), the observed discrepancy becomes even smaller. We conclude that the proposed nearly equiangular, nearly tight frames satisfy the condition for near optimal sum capacity. Considering that we may produce such frames of any dimensions, the proposed construction offers flexibility when designing codes for an s-CDMA system and provides spreading sequences that lead to near optimal performance.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

In this thesis, we relied on well-known results from frame theory and proposed novel frame constructions that attain small mutual coherence and spectral norm, approximating the corresponding optimal bounds. The proposed frames are successfully employed in sparse representations, compressed sensing, and communications. More particularly, the numerical methods presented here yield three types of frames, namely incoherent UNTFs, nearly equiangular frames and nearly equiangular, nearly tight frames. All proposed frames exhibit remarkable performance, when used to acquire sparse signals in compressed sensing, improving the recovery rates of the deployed algorithms. Incoherent UNTFs are suitable for designing optimized projection matrices for compressed sensing and efficient preconditioners for underdetermined linear systems with sparse solutions that are met in sparse representations and compressed sensing. Nearly equiangular, nearly tight frames approximate UNTFs, which are considered optimal spreading sequences for s-CDMA systems. Exhibiting the additional advantage of approximate equiangularity, they can be employed as spreading sequences in multi-access systems with varying number of users, as they minimize interuser interference.

The mathematical tools used to develop the proposed constructions involve optimization techniques that concern projections onto non-convex sets and numerical methods for the solution of inverse eigenvalue problems. Most theoretical results in these fields have been established over the past decades, yet, important questions such as the projections onto non-convex sets have not been completely answered. It is obvious that any progress in these fields may offer a better insight of the developed techniques and contribute to the improvement of the efficiency of the proposed algorithms.

Theoretical study of the new frame constructions regarding their feasibility in practical problems is an important working direction. While there exist several recovery guarantees for incoherent frames and incoherent UNTFs, we have almost no result for frames that approximate ETFs. It would be of great significance, if the proposed nearly equiangular, nearly tight frames could be accompanied by theoretical results justifying their remarkable performance in simulations. A deep investigation could provide performance bounds deciding the appropriateness of the proposed frames in sparse recovery or their feasibil-

ity to minimize interuser interference, when used as spreading sequences in multi-access systems.

Towards this direction, a quantitative characterization of approximate equiangularity seems useful. The latest work of [23] introduces the concept of ϵ -equiangularity. Considering an $m \times N$ frame $\Phi = \{\varphi_i\}_{i=1}^N$, $\varphi_i \in \mathbb{R}^m$, and denoting by $\mu_{m,N}$ the Welch bound (see (4.3)), the frame is defined as ϵ -equiangular if

$$(1 - \epsilon)\mu_{m,N} \leq |\langle \varphi_i, \varphi_j \rangle| \leq (1 + \epsilon)\mu_{m,N}, \quad (7.1)$$

for any two distinct columns φ_i, φ_j of Φ . Regarding the recovery ability of ϵ -equiangular frames, the authors of [23] rephrase the square root bound on sparsity, $s \leq \sqrt{m}$, to the plausible bound $s \leq \frac{\sqrt{m}}{2(1+\epsilon)}$. Definitely, a further investigation of the advantages and limitations of such frames is of great interest. A theoretical study could also consider ϵ -tight frames with spectral norm that slightly exceeds the minimum bound N/m ($\|\Phi\|^2 < \frac{N}{m} + \epsilon$) and ϵ -unit norm frames with columns of norm close to 1.

In compressed sensing, we have seen that practical problems impose certain restrictions on the design of projections matrices, arising from physical constraints in the related applications. Binary matrices are considered best candidates for hardware implementation. A similar constraint in multi-access systems is that the alphabet of the employed codes may also be restricted. From this perspective, it is a challenge to develop methods that produce frames with specific alphabet, e.g., binary entries, also exhibiting good incoherence and spectral properties. Concerning practical compressed sensing applications, it is important that projection matrices also possess some structure. Recall that structured frames facilitate the design of the acquisition hardware and offer fast and reliable signal reconstruction, improving the performance of sensing devices. Incorporating the above parameters in frame design, while also retaining incoherence and tightness, is a challenge.

The goal of this thesis was the construction of frames that exhibit good incoherence and spectral properties. As equiangular tight frames form a class of frames satisfying optimal bounds regarding incoherence and spectral norm, future research is inevitably connected with new developments in frame theory and, more particularly, new results in the design of ETFs and UNTFs. Of course, the construction of ETFs is an extremely difficult problem—open for over half a century, and is connected with other important problems and conjectures in frame theory that have been stated in [39, 74, 33]. However, from the perspective of an engineer, besides perfect ETF and UNTF constructions, we are also interested in approximate constructions as the ones proposed in this thesis. Any new theoretical foundations contributing to a better understanding of ETFs or UNTFs may provoke the development of new techniques, producing frames that are useful in practical applications.

Sparse representations and compressed sensing have experienced a considerable growth during the past decade. Still important theoretical and practical questions remain open [59, 118]. The work presented here is a typical paradigm of how research in these fields evolves. In the recent years, much of the progress has been inspired from results in other research areas such as frame theory, graph theory, applied harmonic analysis, and

information theory. On the other hand, compressed sensing and sparse representations have played an important role to the evolution of advanced probability theory and, in particular, random matrix theory, convex optimization, and applied harmonic analysis. Furthermore, diffusion of sparse recovery and compressed sensing ideas in areas such as radar analysis, medical imaging, distributed signal processing, and data quantization has also provoked important progress in various practical applications. Clearly, the progress in sparse representations and compressed sensing is a result of interdisciplinary collaborations motivated by one sensible reason: some important problems simply cannot be solved otherwise! An interesting side of this collaborative culture is the way we are thinking about the development of hardware and software when designing sensors and other devices. While, in the past, we addressed these problems separately, it seems that future developments require an interdisciplinary approach, where hardware and algorithms are treated in a truly intergrated manner [118].

BIBLIOGRAPHY

- [1] V. Abolghasemi, D. Jarchi, and S. Sanei. A robust approach for optimization of the measurement matrix in compressed sensing. In *2nd Int. Wor. on Cogn. Inf. Proc. (CIP)*, pages 388–392, 2010.
- [2] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322, 2006.
- [3] P. Anigstein and V. Anantharam. Ensuring convergence of the MMSE iteration for interference avoidance to the global optimum. *IEEE Trans. Inf. Th.*, 49(4):873–885, 2003.
- [4] L. Applebaum, S.D. Howard, S. Searle, and R. Calderbank. Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery. *App. Comp. Harm. Anal.*, 26(2):283–290, 2009.
- [5] A. Auslender. *Méthodes Numériques pour la Résolution des Problèmes d’Optimisation avec Contraintes*. Phd thesis, Faculté des Sciences, Grenoble, 1969.
- [6] O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, New York, 1994.
- [7] Y. Baig, E.M. Lai, and A. Punchihewa. Distributed Video Coding Based on Compressed Sensing. In *ICMEW Workshops*, pages 325–330, 2012.
- [8] W.U. Bajwa, R. Calderbank, and S. Jafarpour. Why Gabor frames? two fundamental measures of coherence and their role in model selection. *J. Commun. Netw.*, 12(4):289–307, 2010.
- [9] W.U. Bajwa, R. Calderbank, and D.G. Mixon. Two are better than one: Fundamental parameters of frame coherence. *App. Comp. Harm. Anal.*, 33(1):58–78, 2012.
- [10] A.S. Bandeira, M. Fickus, D.G. Mixon, and P. Wong. The road to deterministic matrices with the restricted isometry property. arXiv:1202.1234v2 [math.FA], 2012.

- [11] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.
- [12] H.H. Bauschke and J.M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Rev.*, 38(3):367–426, 1996.
- [13] H.H. Bauschke, P.L. Combettes, and D.R. Luke. Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization. *J. Opt. Soc. Am.*, 19(7):1334–1345, 2002.
- [14] J.J. Benedetto and M. Fickus. Finite normalized tight frames. *Adv. Comput. Math.*, 18:357–385, 2003.
- [15] K.T. Block, M. Uecker, and J. Frahm. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magn. Res. Med.*, 57(6):1086–1098, 2007.
- [16] J. Bobin, J.L. Starck, and R. Ottensamer. Compressed sensing in astronomy. *IEEE J. Sel. Top. Signal Process.*, 2:718–726, 2008.
- [17] B. G. Bodmann and V. I. Paulsen. Frames, graphs and erasures. *Linear Algebra Appl.*, 404(15):118–146, 2005.
- [18] B. G. Bodmann, V. I. Paulsen, and M. Tomforde. Equiangular tight frames from complex Seidel matrices containing cube roots of unity. *Linear Algebra Appl.*, 430(1):396–417, 2009.
- [19] B.G. Bodmann and P.G. Casazza. The road to equal-norm Parseval frames. *J. Funct. Anal.*, 258(2):397–420, 2010.
- [20] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova. Explicit constructions of RIP matrices and related problems. *Duke Math. J.*, 159(1):145–185, 2011.
- [21] A.E. Brouwer, A.M. Cohen, and A. Neumaier. *Distance-Regular Graphs*. Springer, Berlin, 1989.
- [22] A.M. Bruckstein, D.L. Donoho, and M. Elad. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Rev.*, 51(1):34–81, 2009.
- [23] D. Bryant and P. Ó Catháin. An asymptotic existence result on compressed sensing. arXiv:1403.2807v2 [math.FA], 2015.
- [24] E. Candès and Y. Plan. Near-ideal model selection by ℓ_1 minimization. *Ann. Statist.*, 37:2145–2177, 2009.

- [25] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Th.*, 52(2):489–509, 2006.
- [26] E. Candès and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Stat.*, 35(6):2313–2351, 2007.
- [27] E.J. Candès, Y.C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *App. Comp. Harm. Anal.*, 31(1):59–73, 2010.
- [28] E.J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Th.*, 51(12):4203–4215, 2005.
- [29] P.G. Casazza, M. Fickus, and D.G. Mixon. Auto-tuning unit norm frames. *App. Comp. Harm. Anal.*, 32(1):1–15, 2012.
- [30] P.G. Casazza, M. Fickus, D.G. Mixon, Y. Wang, and Z. Zhou. Constructing tight fusion frames. *App. Comp. Harm. Anal.*, 30:175–187, 2011.
- [31] P.G. Casazza and J. Kovačević. Equal-norm tight frames with erasures. *Adv. Comput. Math.*, 18:387–430, 2003.
- [32] P.G. Casazza and M.T. Leon. Existence and construction of finite tight frames. unpublished manuscript, 2002.
- [33] P.G. Casazza, D. Redmond, and J.C. Tremain. Real equiangular frames. In *Proc. Conf. Inf. Sci. Syst.*, pages 715–720, 2008.
- [34] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1999.
- [35] X. Chen and M.T. Chu. Constructing a Hermitian matrix from its diagonal entries and eigenvalues. *SIAM J. Matrix Anal. & Appl.*, 16:207–217, 1995.
- [36] X. Chen and M.T. Chu. On the least squares solution of inverse eigenvalue problems. *SIAM J. Numer. Anal.*, 33:2417–2430, 1996.
- [37] W. Cheney and A. Goldstein. Proximity maps for convex sets. In *Proceedings of the AMS*, volume 10, pages 448–450, 1959.
- [38] O. Christensen. *An Introduction to Frames and Riesz Bases*. Boston, MA:Birkhäuser, 2003.
- [39] O. Christensen. Six problems in frame theory. arXiv:1308.5065v1 [math.FA], 2013.
- [40] M. T. Chu and G. H. Golub. *Inverse Eigenvalue Problems: Theory, Algorithms, and Applications*. Oxford University Press, Oxford, 2005.

- [41] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k -term approximation. *J. Amer. Math. Soc.*, 22:211–231, 2009.
- [42] P.L. Combettes and H.J. Trussell. Method of successive projections for finding a common point of sets in metric spaces. *J. Opt. Th. Appl.*, 57(3):487–507, 1990.
- [43] I. Daubechies. *Ten Lectures on Wavelets*. Philadelphia, PA:SIAM, 1992.
- [44] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.
- [45] I. Daubechies, A. Grossman, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27:1271–1283, 1986.
- [46] M. Davenport. *Random observations on random observations: Sparse signal acquisition and processing*. PhD thesis, Rice University, 2010.
- [47] G. Davis, S. Mallat, and M. Avellaneda. Greedy adaptive approximation. *Constr. Approx.*, 13:57–58, 1997.
- [48] R.A. DeVore. Deterministic constructions of compressed sensing matrices. *J. Complexity*, 23(4-6):918–925, 2007.
- [49] D.L. Donoho. Compressed sensing. *IEEE Trans. Inf. Th.*, 52(4):1289–1306, 2006.
- [50] D.L. Donoho and M. Elad. Maximal sparsity representation via ℓ^1 -minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202, 2003.
- [51] D.L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 -minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202, 2003.
- [52] D.L. Donoho, M. Elad, and V.N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Th.*, 52(1):6–18, 2006.
- [53] D.L. Donoho, Y. Tsaig, I. Drori, and J.L. Starck. Sparse solution of underdetermined linear equations by stagewise Orthogonal Matching Pursuit (StOMP). *IEEE Trans. Inf. Th.*, 58(2):1094–1121, 2012.
- [54] M.F. Duarte and Y.C. Eldar. Structured compressed sensing: From theory to applications. *IEEE Trans. Signal Process.*, 59(9):4053–4085, 2011.
- [55] J.M. Duarte-Carvajalino and G. Sapiro. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Trans. Image Process.*, 18(7):1395–1408, 2009.

- [56] R.J. Duffin and A.C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Am. Math. Soc.*, 72(2):341–366, 1952.
- [57] M. Elad. Optimized projections for compressed sensing. *IEEE Trans. Signal Process.*, 55(12):5695–5702, 2007.
- [58] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, New York, 2010.
- [59] M. Elad. Sparse and Redundant Representation Modeling—What Next? *IEEE Signal Process. Lett.*, 19(12):922–928, 2012.
- [60] Y.C. Eldar and G.D. Forney Jr. Optimal tight frames and quantum measurement. *IEEE Trans. Inf. Th.*, 48(3):599–610, 2002.
- [61] S. Erickson and C. Sabatti. Empirical bayes estimation of a sparse vector of gene expression changes. *Stat. Appl. Genet. Mol. Biology*, 4(1):22, 2005.
- [62] M. Fickus and D.G. Mixon. Numerically erasure-robust frames. *Linear Algebra Appl.*, 437(6):1394–1407, 2012.
- [63] M. Fickus and D.G. Mixon. Tables of the existence of equiangular tight frames. arXiv:1504.00253v1 [math.FA], 2012.
- [64] M. Fickus, D.G. Mixon, and J.C. Tremain. Steiner equiangular tight frames. *Linear Algebra Appl.*, 436(5):1014–1027, 2012.
- [65] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Berlin, Germany: Springer, 2013.
- [66] U. Gamper, P. Boesiger, and S. Kozerke. Compressed sensing in dynamic MRI. *Magn. Res. Med.*, 59(2):365–373, 2008.
- [67] J.M. Goethals and J.J. Seidel. Orthogonal matrices with zero diagonal. *Canad. J. Math.*, 19:1001–1010, 1967.
- [68] V.K. Goyal, J. Kovačević, and J.A. Kelner. Quantized frame expansions with erasures. *App. Comp. Harm. Anal.*, 10(3):203–233, 2001.
- [69] V.K. Goyal, M. Vetterli, and N.T. Thao. Quantized overcomplete expansions in \mathbb{R}^N : analysis, synthesis, and algorithms. *IEEE Trans. Inf. Th.*, 44(1):16–31, 1998.
- [70] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inf. Th.*, 49(12):3320–3325, 2003.
- [71] A. Griffin and P. Tsakalides. Compressed sensing of audio signals using multiple sensors. In *Proc. of 16th European Signal Processing Conference*, pages 1–5, 2008.

- [72] M.A. Griswold, P.M. Jakob, R.M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Res. Med.*, 47(6):1202–1210, 2002.
- [73] T. Guess. Optimal sequences for CDMA with decision-feedback receivers. *IEEE Trans. Inf. Th.*, 49(4):886–900, 2003.
- [74] J. Haas. Open problems in frame theory/Phase retrieval. www.math.uh.edu/~johnhaas/, 2013.
- [75] J.D. Haupt, W.U. Bajwa, G. Raz, and R. Nowak. Toeplitz compressed sensing matrices with applications to sparse channel estimation. *IEEE Trans. Inf. Th.*, 56(11):5862–5875, 2010.
- [76] R. W. Heath, T. Strohmer, and A. J. Paulraj. On quasi-orthogonal signatures for CDMA systems. *IEEE Trans. Inf. Th.*, 52(3):1217–1226, 2006.
- [77] R.W. Heath, T. Strohmer, and A.J. Paulraj. Grassmannian signatures for CDMA systems. In *IEEE Global Telecommunications Conference, GLOBECOM '03*, volume 3, pages 1553–1557, 2003.
- [78] M.A. Herman and T. Strohmer. High-Resolution Radar via Compressed Sensing. *IEEE Trans. Signal Process.*, 57(6):2275–2284, 2009.
- [79] F.J. Herrmann, M.P. Friedlander, and O. Yilmaz. Fighting the Curse of Dimensionality: Compressive Sensing in Exploration Seismology. *IEEE Signal Proc. Mag.*, 29(3):88–100, 2012.
- [80] R.B. Holmes and V.I. Paulsen. Optimal frames for erasures. *Linear Algebra Appl.*, 377:31–51, 2004.
- [81] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [82] H. Huang and A. Makur. Optimized measurement matrix for compressive sensing. In *Intl. Conf. Sampling Th. App., Singapore*, 2011.
- [83] P. Indyk. Explicit constructions for compressed sensing of sparse signals. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 30–33, 2008.
- [84] J. Kovačević and A. Chebira. Life beyond bases: The advent of frames (Part I). *IEEE Signal Proc. Mag.*, 24(4):86–104, 2007.
- [85] J. Kovačević and A. Chebira. Life beyond bases: The advent of frames (Part II). *IEEE Signal Proc. Mag.*, 24(5):115–125, 2007.
- [86] P.W.H. Lemmens and J.J. Seidel. Equi-isoclinic subspaces of Euclidean spaces. In *Proc. Nederl. Akad. Wetensch. Series A 76*, pages 98–107. Indag. Math. 35, 1973.

- [87] A.S. Lewis, D.R. Luke, and J. Malick. Local convergence for alternating and averaged nonconvex projections. <http://arxiv.org/pdf/0709.0109.pdf>, 2013.
- [88] A.S. Lewis and J. Malick. Alternating projections on manifolds. *Mathematics of Operations Research*, 33(1):216–234, 2008.
- [89] Q.S. Lian, Y.Y. Gao, L.Li, and P.P. Hao. Image compressed sensing based on DT-CWT. In *Proc. of Int. Conf. on Audio, Language and Image Processing (ICALIP)*, pages 1573–1578, 2008.
- [90] J. Liu and Q. Sun. Mass spectrum data processing based on compressed sensing recognition and sparse difference recovery. In *Proc. of 9th Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD)*, pages 1413–1417, 2012.
- [91] W. Lu, W. Li, K. Kpalma, and J. Ronsin. Near-optimal binary compressed sensing matrix, 2013. arXiv:1304.4071.
- [92] W.L. Lu, K. Kpalma, and J. Ronsin. Sparse binary matrices of LDPC codes for Compressed Sensing. In *Data Compression Conference (CCC)*, 2012.
- [93] S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993.
- [94] M. Mangia, M. Paleari, P.Ariano, R. Rovatti, and G. Setti. Compressed sensing based on rakeness for surface ElectroMyoGraphy. In *Proc. of IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, pages 204–207, 2014.
- [95] J. L. Massey and T. Mittelholzer. Welch’s Bound and Sequence Sets for Code-Division Multiple-Access Systems. In *Sequences II: Methods in Communication, Security and Computer Sciences*, pages 63–78. Berlin:Springer, Heidelberg, 1993.
- [96] M. Mishali and Y.C. Eldar. From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals. *IEEE J. Select. Top. Signal Processing*, 4(2):375–391, 2010.
- [97] Q. Mo and S. Li. New bounds on the restricted isometry constant δ_{2K} . *App. Comp. Harm. Anal.*, 31(3):460–468, 2011.
- [98] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Technical report, California Institute of Technology, Pasadena, 2008.
- [99] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. <http://arxiv.org/abs//0712.1360v1>, 2007.

- [100] S. Olanigan and L. Cao. Multi-scale image compressed sensing with optimized transmission. In *IEEE SiPS Workshop*, pages 59–64, 2013.
- [101] R. Orsi. Numerical methods for solving inverse eigenvalue problems for nonnegative matrices. *SIAM J. Matrix Anal. Appl.*, 28:190–212, 2006.
- [102] R.E.A.C. Paley. On orthogonal matrices. *J. Math. Phys.* , pages 311–320, 1993.
- [103] L.C. Potter, E.Ertin, J.T. Parker, and M. Cetin. Sparsity and compressed sensing in radar imaging. In *Proc. IEEE*, volume 98, pages 1006–1020, 2010.
- [104] S. Pudlewski, A. Prasanna, and T. Melodia. Compressed-Sensing-Enabled Video Streaming for Wireless Multimedia Sensor Networks. *IEEE Trans. Mobile Comput.*, 11(6):1060–1072, 2012.
- [105] H. Rauhut. Compressive sensing and structured random matrices. Theoretical Foundations and Numerical Methods for Sparse Recovery, H. Rauhut, Ed., vol. 9, Radon Series on Computational and Applied Mathematics. De Gruyter, 2010.
- [106] H. Rauhut, J.K. Romberg, and J.A. Tropp. Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmon. Anal.*, 32(2):242–254, 2012.
- [107] J. M. Renes. Equiangular tight frames from Paley tournaments. *Linear Algebra Appl.*, 426:326–330, 2007.
- [108] J. Romberg. Compressive sensing by random convolution. *SIAM J. Imag. Sci.*, 2(4):1098–1128, 2009.
- [109] C. Rose, S. Ulukus, and R.D. Yates. Wireless systems and interference avoidance. *IEEE Trans. Wireless Commun.*, 1(3):415–428, 2002.
- [110] M. Rudelson and R. Vershynin. The Littlewood-Offord Problem and invertibility of random matrices. <http://arxiv.org/abs/math/0703503v2>, 2008.
- [111] M. Rupf and J.L. Massey. Optimum sequence multisets for synchronous code-division multiple-access channels. *IEEE Trans. Inf. Th.*, 40(4):1261–1266, 1994.
- [112] D.V. Sarwate. Meeting the Welch Bound with Equality. In *Sequences and their applications (Singapore, 1998)*. Springer, London, 1999.
- [113] K. Schnass and P. Vandergheynst. Average performance analysis for thresholding. *IEEE Signal Process. Lett.*, 14(11):828–831, 2007.
- [114] K. Schnass and P. Vandergheynst. Dictionary preconditioning for greedy algorithms. *IEEE Trans. Signal Proc.*, 56(5):1994–2002, 2008.

- [115] J.J. Seidel. A survey of two-graphs. In *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo I*, in: *Attidei Convegni Lincei, Vol. 17*, pages 481–511. Accad. Naz. Lincei, Rome, 1976.
- [116] M. Sheikh, O. Milenkovic, and R. Baraniuk. Designing compressive sensing DNA microarrays. In *2nd IEEE CAMSAP Workshop, St. Thomas, VI, USA*, pages 141–144, 2007.
- [117] J.P. Slavinsky, J. Laska, M. Davenport, and R. Baraniuk. The compressive multiplexer for multi-channel compressive sensing. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP), Prague, Czech Republic*, 2011.
- [118] T. Strohmer. Measure what should be Measured: Progress and Challenges in Compressive Sensing. *IEEE Signal Process. Lett.*, 19(12):887–893, 2012.
- [119] T. Strohmer and R.W. Heath. Grassmannian frames with applications to coding and communication. *App. Comp. Harm. Anal.*, 14(3):257–275, 2003.
- [120] T. Strohmer and M. Hermann. Compressed Sensing Radar. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, pages 1509–1512, 2008.
- [121] M. A. Sustik, J. A. Tropp, I. S. Dhillon, and R. W. Heath. On the existence of equiangular tight frames. *Linear Algebra Appl.*, 426:619–635, 2007.
- [122] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk. Beyond Nyquist: Efficient sampling of sparse, bandlimited signals. *IEEE Trans. Inf. Th.*, 56(1):520–544, 2008.
- [123] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk. Random filters for compressive sampling and reconstruction. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2006.
- [124] J.A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inf. Th.*, 50(10):2231–2242, 2004.
- [125] J.A. Tropp. On the conditioning of random subdictionaries. *App. Comp. Harm. Anal.*, 25:1–24, 2008.
- [126] J.A. Tropp, I.S. Dhillon, R.W. Heath, and T. Strohmer. Designing structured tight frames via an alternating projection method. *IEEE Trans. Inf. Th.*, 51(1):188–209, 2005.
- [127] J.A. Tropp, I.S. Dhillon, and R.W. Heath Jr. Finite-step algorithms for constructing optimal CDMA signature sequences. *IEEE Trans. Inf. Th.*, 50(11):2916–2921, 2004.
- [128] E. Tsiligiani, L.P. Kondi, and A.K. Katsaggelos. Use of tight frames for optimized compressed sensing. In *Proc. of 20th Eur. Signal Process. Conf. (EUSIPCO), Bucharest*, 2012.

- [129] E. Tsiligianni, L.P. Kondi, and A.K. Katsaggelos. Construction of Incoherent Unit Norm Tight Frames with Application to Compressed Sensing. *IEEE Trans. Inf. Th.*, 60(4):2319–2330, 2014.
- [130] E. Tsiligianni, L.P. Kondi, and A.K. Katsaggelos. Preconditioning for Underdetermined Linear Systems with Sparse Solutions. *IEEE Signal Process. Lett.*, 22(9):1239–1243, 2015.
- [131] B. Uddin, M.E. Celebi, H. Kingravi, and G. Schaefer. Accurate genomic signal recovery using compressed sensing. In *Proc. of 21st Int. Conf. Pattern Recogn. (ICPR)*, pages 3144–3147, 2012.
- [132] S. Ulukus and R.D. Yates. Iterative construction of optimum signature sequence sets in synchronous CDMA systems. *IEEE Trans. Inf. Th.*, 47(5):1989–1998, 2001.
- [133] S. Ulukus and A. Yener. Iterative joint optimization of CDMA signature sequence and receiver filters. In *Proc. 2002 Conf. Inf. Sciences and Systems (CISS)*. Princeton, NJ, 2002.
- [134] J.H. van Lint and J.J. Seidel. Equiangular point sets in elliptic geometry. In *Proc. Nederl. Akad. Wetensch. Series A 69*, pages 335–348, 1966.
- [135] S. Verdú. Capacity Region of Gaussian CDMA channels: The Symbol-Synchronous case. In *24th Allerton Conf. on Commun. Control and Comput.*, pages 1025–1034, 1986.
- [136] P. Viswanath and V. Anantharam. Optimal sequences and sum capacity of synchronous CDMA systems. *IEEE Trans. Inf. Th.*, 45(6):1984–1991, 1999.
- [137] P. Viswanath and V. Anantharam. Optimal sequences for CDMA under colored noise: a Schur-saddle function property. *IEEE Trans. Inf. Th.*, 48(6):1295–1318, 2002.
- [138] P. Viswanath, V. Anantharam, and D. Tse. Optimal sequences, power control and capacity of synchronous CDMA systems with linear multiuser receivers. In *Information Theory Workshop*, pages 134–135, 1998.
- [139] J. von Neumann. *Functional Operators, volume II*. Princeton University Press, Princeton, NJ, 1950. Reprint of mimeographed lecture notes first distributed in 1933.
- [140] S. Waldron. Generalized Welch bound equality sequences are tight frames. *IEEE Trans. Inf. Th.*, 49(9):2307–2309, 2003.
- [141] S. Waldron. On the construction of equiangular frames from graphs. *Linear Algebra Appl.*, 431(11):2228–2242, 2009.

- [142] C.A. Weber and J.P. Allebach. Reconstruction of frequency-offset fourier data by alternating projection on constraint sets. In *Proc. of 24th Allerton Conf. on Commun. Control and Comput., Urbana-Champaign, IL*, volume 10, pages 194–201, 1986.
- [143] L.R. Welch. Lower bounds on the maximum cross correlation of signals. *IEEE Trans. Inf. Th.*, IT-20:397–399, 1974.
- [144] P. Xia, S. Zhou, and G.B. Giannakis. Achieving the Welch bound with difference sets. *IEEE Trans. Inf. Th.*, 51(5):1900–1907, 2005.
- [145] J. Xu, Y. Pi, and Z. Cao. Optimized projection matrix for compressive sensing. *EURASIP J. Adv. Signal Process.*, pages 43:1–43:8, 2010.

APPENDIX A

PROJECTIONS

Projection onto smooth manifolds

According to [88], a smooth manifold \mathbb{E} is, loosely speaking, a set consisting locally of the solutions of some smooth equations. More precisely, we say that a set $\mathcal{M} \subset \mathbb{E}$ is a C^k -manifold (of codimension d) around a point $x \in \mathcal{M}$, if there exists an open set $U \subset \mathbb{E}$ containing x such that

$$\mathcal{M} \cap U = \{x \in U : F(x) = 0\},$$

where $F : U \rightarrow \mathbb{R}^d$ is a C^k function with surjective derivative throughout U .

Fixed rank matrices is an example of a smooth manifold. Let $\mathbb{E} = M_{m,N}(\mathbb{R})$ be the space of $m \times N$ matrices with the classical inner product $\langle A, B \rangle = \text{trace}(A^T B)$. Routine calculations show that the set of matrices with fixed rank r ,

$$\mathcal{R}_r = \{X \in M_{m,N}(\mathbb{R}) : \text{rank}(X) = r\},$$

is a smooth manifold around any matrix $A \in \mathcal{R}_r$. Using the singular value decomposition $A = UDV^T$ (the two matrices $U = [u_1, u_2, \dots, u_n]$ and $V = [v_1, v_2, \dots, v_m]$ being orthogonal, and the diagonal entries in the diagonal matrix D being written in decreasing order), the tangent space at A to \mathcal{R}_r is

$$T_{\mathcal{R}_r}(A) = \{H \in M_{m,N}(\mathbb{R}) : u_i^T V_j = 0, \quad \text{for all } r < i \leq N, \quad r < j \leq m\}.$$

The following result states that smooth manifolds admit unique projections locally.

Theorem A.1 (Projection onto a manifold [88]). *Let $\mathcal{M} \subset \mathbb{E}$ be a manifold of class C^k (with $k \geq 2$) around a point $\bar{x} \in \mathcal{M}$. Then the projection $P_{\mathcal{M}}$ is well-defined around \bar{x} .*

Projection onto fixed rank matrices is an example of projection onto manifolds and can be computed with the truncated singular value decomposition. If $X \in M_{m,N}(\mathbb{R})$ with $X = U\Sigma V^T$, then the nearest matrix with rank no more than r is

$$\hat{X} = \sum_{i=1}^r \sigma_i u_i v_i^T,$$

where σ_i are the r first singular values of Σ .

Projection onto spectral sets

According to [101], projections onto spectral sets of matrices, that is, sets of matrices defined via properties of their eigenvalues, can be handled using spectral decomposition. Let \mathcal{Q}^N be the space of real symmetric $N \times N$ matrices, equipped with the trace inner product. \mathcal{Q}^N is an Euclidean space. A subset T is spectral if, for every matrix $X \in T$ and every U in the group O^N of orthogonal matrices, we have $U^T X U \in T$. The eigenvalue map $\lambda : \mathcal{Q}^N \rightarrow \mathbb{R}^N$ maps any symmetric matrix X to its eigenvalues arranged in nonincreasing order, $\lambda_1(X) \geq \lambda_2(X) \geq \dots \geq \lambda_N(X)$. It is easy to see that any spectral set can be written in the form $\lambda^{-1}(K) = \{X : \lambda(X) \in K\}$, for some set $K \subset \mathbb{R}^N$, and that we can further restrict K to be permutation-invariant: for every vector $x \in K$ and every P in the group P^N of permutation matrices, we have $Px \in K$. The following result is established in [88].

Theorem A.2 (Spectral projection [88]). *If the point x in the permutation-invariant set $K \subset \mathbb{R}^N$, is a nearest point to the point $y \in \mathbb{R}^N$, then for any orthogonal matrix U , the matrix $U^T \text{diag}(x)U$ is a nearest matrix in the spectral set $\lambda^{-1}(K)$ to the matrix $U^T \text{diag}(y)U$.*

A good example is the set of matrices of some fixed rank. More results regarding projections onto spectral sets can be found in [88].

INDEX

- adjacency matrix, 18
- alternating projections, 25, 30, 34, 35, 38, 41, 47
- analysis operator, 12
- atom, 4, 51, 56, 60
- average coherence, 48, 49, 59, 76, 77
- averaged projections, 25–27, 32–35, 38
- Basis Pursuit (BP), 58, 62–64, 66, 69, 83–87
- Code Division Multiple Access (CDMA), 2, 4–7, 9, 24, 49, 89–94, 96, 97
- conference matrix, 19
- Dantzig selector, 58, 62–66
- dictionary, 4, 5, 7, 51–53, 55, 58–60
- effective dictionary, 5, 27, 65, 70, 74, 75, 78, 79, 84, 85
- extremal point, 26, 33
- frame, 1, 2, 9, 11
 - canonical dual, 13
 - equal norm tight frame, 6, 15, 90, 94
 - equiangular, 1–4, 6, 12, 17
 - equiangular tight frame (ETF), 1–4, 6, 16–21, 23, 24, 28, 38–42, 44, 47, 72, 78, 90, 95, 99, 100
 - finite, 1, 11
 - Grassmannian, 9, 16, 17
 - incoherent, 3–6, 21, 23, 24, 28, 29, 31–33, 38, 40, 47, 48, 52, 54, 58, 59, 61, 69, 75, 76, 78, 79, 81, 85, 96, 99
 - nearly equiangular, 2, 4, 6, 7, 40, 42, 44–48, 63, 65, 75–79, 90, 96, 97, 99
 - tight, 1–6, 11, 12, 14, 15, 23, 28, 45, 49, 52, 58, 63, 65, 75, 77, 78, 90, 96, 97, 100
 - unit norm tight frame (UNTF), 2–5, 15–17, 21, 23, 28, 29, 31–33, 35, 38, 46–48, 58, 59, 61, 69, 75–79, 85, 94, 96, 97, 99
- frame operator, 13
- Gram matrix, 13, 14, 16, 18, 19, 27, 29, 32, 33, 37, 38, 40, 44, 47, 48, 55, 60, 62, 74, 75, 81
- interuser interference, 2, 4, 6, 7, 90–94, 96, 97, 99, 100
- inverse eigenvalue problem (IEP), 39
- isometry constant (RIP-constant), 54, 55, 58, 69, 71
- Maximal Welch Bound Equality sequences (MWBEs), 17, 95
- mutual coherence, 3, 23, 28, 29, 35, 38, 47, 48, 53–55, 58–61, 65, 72, 74, 76–80, 84, 85, 99
- optimized projections, 7, 72, 74, 75, 77–79, 87
- Orthogonal Matching Pursuit (OMP), 52, 56, 60, 62, 63, 66, 69, 76, 77, 84–87
- preconditioning, 4, 5, 7, 52, 59–61, 63, 65, 66, 68, 84, 85, 87
- projection, 11, 115, 116
- projection matrix, 5, 23, 27, 69, 70, 74, 75, 77–79, 84, 85
- prox-regular set, 26, 27, 33

redundancy, 11, 47, 48
Restricted Isometry Property (RIP), 53–55,
57, 58, 68–72, 74

signature matrix, 18, 24, 39–44, 47, 48
smooth manifold, 25, 26, 30, 33, 115
sparsity, 54–59
spectral norm, 2–4, 15, 16, 23, 35, 46–49,
58, 59, 61, 65, 75–80, 85, 96, 100
spreading sequence, 2, 4–7, 24, 49, 89–94,
96, 97, 99, 100
strongly regular intersection, 27
sum capacity, 89–94, 96, 97
synthesis operator, 12

Total Squared Correlation (TSC), 6, 96, 97

underdetermined linear system, 3–5, 7, 52,
53, 59, 61, 69

Welch bound, 17, 18, 54, 81, 92, 95, 100
Welch Bound Equality sequences (WBEs),
6, 16, 90, 92–96

AUTHOR'S PUBLICATIONS

1. E. Tsiligianni, L.P. Kondi, and A.K. Katsaggelos, Construction of Nearly Equiangular Frames and their Applications, submitted.
2. E. Tsiligianni, L.P. Kondi, and A.K. Katsaggelos, Preconditioning for Underdetermined Linear Systems with Sparse Solutions, *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1239–1243, 2015.
3. E. Tsiligianni, L.P. Kondi, and A.K. Katsaggelos, Construction of Incoherent Unit Norm Tight Frames with Application to Compressed Sensing, *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2319–2330, 2014.
4. E. Tsiligianni, L.P. Kondi, and A.K. Katsaggelos, Use of tight frames for optimized compressed sensing, in *Proc. of 20th European Signal Processing Conference (EU-SIPCO)*, Bucharest, 2012.

SHORT VITA

Evaggelia Tsiligianni was born in Ioannina, Greece in 1976. She received the Diploma degree in Electrical and Computer Engineering from National Technical University of Athens, Athens, Greece, in 2001, and the M.S. degree in Computer Science from Computer Science Department, University of Ioannina, Ioannina, Greece, in 2010. Since 2010 she has been a Ph.D. student in Computer Science & Engineering Department, University of Ioannina, Ioannina, Greece. She worked in the Institute for Language and Speech Processing (ILSP/Athena R.C.), Athens, Greece, as a member of the Voice and Sound Technology Department for five years (1999-2004). She has been working as a teacher in secondary education for the last twelve years.