Asymptotic Frame Theory for Analog Coding

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Other titles in Foundations and Trends® in Communications and Information Theory

Modeling and Optimization of Latency in Erasure-coded Storage Systems
Vaneet Aggarwal and Tian Lan
ISBN: 978-1-68083-842-8

An Algebraic and Probabilistic Framework for Network Information Theory
S. Sandeep Pradhan, Arun Padakandla and Farhad Shirani
ISBN: 978-1-68083-766-7

Theoretical Foundations of Adversarial Binary Detection
Mauro Barni and Benedetta Tondi
ISBN: 978-1-68083-764-3

Polynomial Methods in Statistical Inference
Yihong Wu and Pengkun Yang
ISBN: 978-1-68083-730-8

Information-Theoretic Foundations of Mismatched Decoding
Jonathan Scarlett, Albert Guillen i Fabregas, Anelia Somekh-Baruch and Alfonso Martinez
ISBN: 978-1-68083-712-4

Coded Computing: Mitigating Fundamental Bottlenecks in Large-Scale Distributed Computing and Machine Learning
Songze Li and Salman Avestimehr
ISBN: 978-1-68083-704-9

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Asymptotic Frame Theory for Analog Coding

Marina Haikin
Amazon, Tel Aviv, Israel
(Previously Tel Aviv University)
mkokotov@gmail.com

Matan Gavish
The Hebrew University of Jerusalem, Israel
gavish@cs.huji.ac.il

Dustin G. Mixon
The Ohio State University, Colombus, OH, USA
mixon.23@osu.edu

Ram Zamir
Tel Aviv University, Israel
zamir@eng.tau.ac.il

now
the essence of knowledge
Boston — Delft

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

M. Haikin, M. Gavish, D.G. Mixon and R. Zamir. *Asymptotic Frame Theory for Analog Coding*. Foundations and Trends® in Communications and Information Theory, vol. 18, no. 4, pp. 526–645, 2021.

ISBN: 978-1-68083-909-8
© 2021 M. Haikin, M. Gavish, D.G. Mixon and R. Zamir

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Foundations and Trends® in Communications and Information Theory

Volume 18, Issue 4, 2021

Editorial Board

Alexander Barg
University of Maryland
USA

Area Editors

Emmanuel Abbe
Princeton University

Arya Mazumdar
UMass Amherst

Olgica Milenkovic
University of Illinois, Urbana-Champaign

Anelia Somekh-Baruch
Bar-Ilan University

Himanshu Tyagi
Indian Institute of Science

Editors

Venkat Anantharam
UC Berkeley

Giuseppe Caire
TU Berlin

Daniel Costello
University of Notre Dame

Albert Guillen i Fabregas
Pompeu Fabra University

Dongning Guo
Northwestern University

Dave Forney
MIT

Te Sun Han
University of Tokyo

Babak Hassibi
Caltech

Michael Honig
Northwestern University

Ioannis Kontoyiannis
Cambridge University

Gerhard Kramer
TU Munich

Amos Lapidoth
ETH Zurich

Muriel Medard
MIT

Neri Merhav
Technion

David Neuhoff
University of Michigan

Alon Orlitsky
UC San Diego

Yury Polyanskiy
MIT

Vincent Poor
Princeton University

Kannan Ramchandran
UC Berkeley

Igal Sason
Technion

Shlomo Shamai
Technion

Amin Shokrollahi
EPF Lausanne

Yossef Steinberg
Technion

Wojciech Szpankowski
Purdue University

David Tse
Stanford University

Antonia Tulino
Bell Labs

Rüdiger Urbanke
EPF Lausanne

Emanuele Viterbo
Monash University

Frans Willems
TU Eindhoven

Raymond Yeung
CUHK

Bin Yu
UC Berkeley

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Editorial Scope

Topics

Foundations and Trends® in Communications and Information Theory publishes survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2021, Volume 18, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328 . Also available as a combined paper and online subscription.

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Contents

1 Introduction **3**

1.1 Notation 10

2 An information-theoretic toy example **11**

2.1 Source coding with erasures 11

2.2 Band-pass interpolation 15

2.3 Non-uniform sampling and signal amplification 19

2.4 From interpolation to frame codes 19

3 Sub-frame performance measures **21**

3.1 The Welch bound 21

3.2 Noise amplification 22

3.3 Spread of eigenvalues 24

3.4 Shannon transform 25

3.5 Statistical RIP and condition number 26

3.6 Sub-frame rectangularity 27

4 Frame theory **28**

4.1 Equiangular tight frames 30

4.2 Other optimal projective codes 37

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

4.3 Frames with well-conditioned subensembles 38

5 Random matrix theory **43**
5.1 Methods of proof in RMT 45

6 Empirical ETF-MANOVA relation **47**
6.1 Distance measures 48
6.2 Frames under study 49
6.3 Empirical observation method 50
6.4 Findings 51

7 Moments of an ETF subset **53**
7.1 Low-order moments: exact analysis 54
7.2 General moments: asymptotic analysis 58
7.3 Concentration 63

8 Sub-frame inequalities **65**

9 Applications **69**
9.1 NOMA/CDMA 69
9.2 DFT codes 71
9.3 Space-time codes (STC) 73
9.4 Over-sampling and Δ-Σ modulation 74
9.5 Multiple-description (MD) source coding 74
9.6 Coded distributed computation (CDC) 76
9.7 Neural networks 78
9.8 Frames in compressed sensing: phase transition 79

10 ETF optimality conjecture **81**
10.1 Erasure Welch bound 81
10.2 Minimum empirical variance and kurtosis 82

11 Conclusion and discussion **85**
11.1 Summary of asymptotic results for ETF subsets 88

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

11.2 Summary of open questions and conjectures 89

Acknowledgements **92**

Appendices **93**

A Entropy-coded dithered quantization **94**

B Information theoretic proofs for sub-frame inequalities **96**

C Variance of $1st$ and $2nd$ moments (proof of Theorem 7.2) **101**

References **104**

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Asymptotic Frame Theory for Analog Coding

Marina Haikin[1], Matan Gavish[2], Dustin G. Mixon[3] and Ram Zamir[4]

[1] *Now at Amazon. Previously at Tel Aviv University where this work was performed; mkokotov@gmail.com*

[2] *Hebrew University of Jerusalem; gavish@cs.huji.ac.il*

[3] *The Ohio State University; mixon.23@osu.edu*

[4] *Tel Aviv University; zamir@eng.tau.ac.il*

ABSTRACT

Over-complete systems of vectors, or in short, frames, play the role of analog codes in many areas of communication and signal processing. To name a few, spreading sequences for code-division multiple access (CDMA), over-complete representations for multiple-description (MD) source coding, space-time codes, sensing matrices for compressed sensing (CS), and more recently, codes for unreliable distributed computation. In this survey paper we observe an information-theoretic random-like behavior of frame subsets. Such sub-frames arise in setups involving erasures (communication), random user activity (multiple access), or sparsity (signal processing), in addition to channel or quantization noise. The goodness of a frame as an analog code is a function of the eigenvalues of a sub-frame, averaged over all sub-frames (e.g., harmonic mean of the eigenvalues relates to least-square estimation error, while geometric mean to the

Marina Haikin, Matan Gavish, Dustin G. Mixon and Ram Zamir (2021), "Asymptotic Frame Theory for Analog Coding", Foundations and Trends® in Communications and Information Theory: Vol. 18, No. 4, pp 526–645. DOI: 10.1561/0100000125.
©2021 M. Haikin, M. Gavish, D.G. Mixon and R. Zamir

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Shannon transform, and condition number to the restricted isometry property).

Within the highly symmetric class of Equiangular Tight Frames (ETF), as well as other "near ETF" families, we show a universal behavior of the empirical eigenvalue distribution (ESD) of a randomly-selected sub-frame: (*i*) the ESD is asymptotically indistinguishable from Wachter's MANOVA distribution; and (*ii*) it exhibits a convergence rate to this limit that is indistinguishable from that of a matrix sequence drawn from MANOVA (Jacobi) ensembles of corresponding dimensions. Some of these results follow from careful statistical analysis of empirical evidence, and some are proved analytically using random matrix theory arguments of independent interest. The goodness measures of the MANOVA limit distribution are better, in a concrete formal sense, than those of the Marchenko–Pastur distribution at the same aspect ratio, implying that deterministic analog codes are better than random (i.i.d.) analog codes. We further give evidence that the ETF (and near ETF) family is in fact superior to any other frame family in terms of its typical sub-frame goodness.

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

1

Introduction

A frame is an "over-complete basis", i.e., a system of vectors that spans the space with more vectors than the space dimension (real or complex). Let us denote the space dimension by m, and the number of vectors by n, where $n > m$. The $m \times n$ frame matrix

$$F = [\mathbf{f}_1 \cdots \mathbf{f}_n] \tag{1.1}$$

is generated by stacking the frame vectors $\mathbf{f}_1, \ldots, \mathbf{f}_n$ as columns, where we restrict attention to unit-norm vectors $\|\mathbf{f}_i\| = 1$. The relative position of frame vectors is determined by their pairwise cross-correlation matrix

$$F^\dagger F = \Big\{ \langle \mathbf{f}_i, \mathbf{f}_j \rangle \Big\}, \tag{1.2}$$

called also Gram or covariance matrix, which is invariant under unitary operation on the frame vectors (e.g., rotation).

This survey proposes an information-theoretic view on the design and analysis of frames with favorable performance. We think of a frame as an "analog code", which can add redundancy [66], [90], [128], [151], remove redundancy [22], or multiplex information directly in the signal space [99], [120], [153]. Multiplication by the frame matrix can expand

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

the dimension $m : n$ hence add redundancy, or reduce the dimension $n : m$ hence compress (with $\mathbf{y} = F^{\dagger}\mathbf{x}$ for the former and $\mathbf{x} = F\mathbf{y}$ for the latter). The aspect ratio n/m is often called the "frame redundancy".

Although information theory tells us that reliable data transmission (adding redundancy) and compression (removing redundancy) can be achieved by *digital* codes, real-world physical-layer communication systems combine *analog* modulation techniques that can be described in terms of frames[1] [16], [17], [22], [67], [94], [158]. We are specifically interested in some old and new applications of frames that involve a combination of *random activity* and *noise*. Performance in these applications is a function of the eigenvalues of (the Gram of) a randomly selected k-subset F_S of the n frame vectors,

$$F_S = [\mathbf{f}_{i_1} \cdots \mathbf{f}_{i_k}] \tag{1.3}$$

where $S = \{i_1, \ldots, i_k\} \subset [n]$, $|S| = k$. (We shall use $[n]$ to denote $\{1, \ldots, n\}$.)

For example, in *non-orthogonal code-division multiple access* (NOMA-CDMA), [127], [140], [145], n users are allocated with m-length spreading sequences using the frame F, but only k out of the n users are active at any given moment, and performance is measured by the *Shannon capacity* of the vector Gaussian channel associated with the $m \times k$ sub-matrix F_S (averaged over the subset S of active users). In transform-based *multiple-description* (MD) source coding, [67], [110], an m-dimensional vector source is expanded into n packets using the frame F, only k packets are received, and performance is measured by the *remote rate-distortion function* associated with F_S (averaged over the subset S of received packets). In coded distributed computation (CDC), [47], [85], [89], the user (master) node expands m sub-computation tasks into n redundant tasks using the frame F, and sends them to n *noisy* computation nodes (where the noise is due to finite precision computation); only k nodes return their answers on time ($n - k$ are stragglers), and

[1] For example, coded-modulation can be thought of as concatenation of an outer digital code with an inner analog code.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

the sub-matrix F_S determines the final average precision (or noise amplification) after the user node decodes the desired computation value. Other "(n, m, k) setups" are listed in Table 1.1.

Table 1.1: (n, m, k) setups featuring analog frame codes

Application	n	m	k	$m \gtrless k$	References
Source with erasures	block-length	bandwidth	important samples	$m > k$	[72], [153]
NOMA-CDMA	users	resources (spread)	active users	$m > k$	[120], [128], [158]
Impulsive channel	block-length	bandwidth	non-erased	$m < k$	[138], [151]
Space-time coding	space (diversity)	time	non-erased antennas	$m < k$	[131], [137], Section 9.3
$\Delta\Sigma$ modulation	over-sampled	original	non-erased	$m < k$	[25], [64]
Multiple descriptions	transmitted	original	received	$m < k$	[67], [110]
Wavelets	coefficients	source	significant coefficients	$m > k$	[81]
Compressed sensing	input	output	sparsity	$m > k$	[24], [40]
Coded computation	workers	computations	non-stragglers	$m < k$	[47], [85], [89]
Neural networks	input	output	features	$m > k$	[7]

In an ideal *noiseless* setup one could choose the frame redundancy n/m equal to the reciprocal of the activity ratio k/n, i.e., $m = k =$ the effective number of users/packets/nodes in the examples above. This is similar to digital erasure correction using *maximum distance separable* (MDS) codes for a channel with k out of n non-erased symbols [13]. However, when *noise* is involved (channel/quantization/computation noise in the setups above), a better trade-off between noise immunity and information rate is obtained by choosing a lower/higher frame redundancy; $k < m < n$ in NOMA-CDMA, or $m < k < n$ in MD and CDC. Thus, m is a design parameter that we can optimize.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

Frame design could be viewed as an attempt to find vectors $\mathbf{f}_1, \ldots, \mathbf{f}_n$ in $\mathbb{R}^m$ or $\mathbb{C}^m$, $n > m$, that are somehow "as orthogonal to each other as possible", either in pairs ($k = 2$) or in larger k-subsets [81], [148]. As we shall see in Chapter 3, the sub-frame performance criteria mentioned above (capacity, rate-distortion function, noise amplification), denoted in general as $\Psi(F_S)$, depend on the spread of the eigenvalues of the Hessian $F_S F_S^\dagger$ or Gram $F_S^\dagger F_S$ matrices[2] of the sub-frame F_S. More mutual orthogonality amounts to a more compact eigenvalue spectrum, and ideal performance occurs when the spectrum shrinks to a delta function, or equivalently, the sub-frame is orthogonal. The redundant nature of the frame, however, implies that most of its subsets are *not* orthogonal. Our target is therefore to find a frame whose *average* performance over all k-subsets

$$\bar{\Psi}(F, k) = \frac{1}{\binom{n}{k}} \sum_{\substack{S \subset [n] \\ |S| = k}} \Psi(F_S) \tag{1.4}$$

is "good"; or in other words, a frame whose *typical* subset has a *compact* eigenvalue spectrum.[3]

We borrow from information theory the probabilistic view of a communication channel, and the notions of typicality and typical-case (rather than worst-case) goodness [34]. The information-theoretic viewpoint leads us to look for frames with a "typically compact" subset spectrum for a given (n, m, k) triplet, and for frame *families* with the best attainable *asymptotic* goodness in the limit as n goes to infinity for fixed (asymptotic) redundancy ratios n/m and k/m. This fresh look on frames turns out to be fruitful, and opens many interesting questions at the intersection of signal processing, random matrix theory, geometry, harmonic analysis and information theory.

Sampling theory suggests the *low-pass frame* (LPF), the frame analog of band-limited interpolation, as a practical candidate for signal

[2] The nonzero eigenvalues of both matrices are the same.

[3] Simple (though only partial) measures for spectrum compactness are the variance and kurtosis; see Section 10.2.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

expansion. This turns out to be a far-from-optimal choice, as we shall see, due to large noise amplification for a typical subset (which corresponds to noisy reconstruction from a non-uniform sampling pattern [83], [98], [122], [144]).

Information theory suggests *random* (i.i.d.) frames as natural candidates for good analog codes. To study the spectrum of these objects, *Random Matrix Theory* (RMT) offers a helpful matrix version of the law of large numbers: the eigenvalue distribution of a typical random matrix tends to *concentrate* towards a fixed distribution in the limit of large dimensions [1], [49]. Indeed, if we choose the elements of the frame matrix F as i.i.d. Gaussian variables, then the subset Gram matrix $F_S^\dagger F_S$ is drawn from a Wishart ensemble, and its spectrum converges almost surely in distribution to the Marchenko–Pastur (MP) distribution with parameter $\beta = k/m$ [93]. We can thus compute the capacity / rate-distortion function / noise amplification (1.4) associated with the Marchenko–Pastur distribution, and obtain some achievable asymptotic performance $\Psi(\mathrm{MP}, \beta)$ for the problems described above.

Is the Marchenko–Pastur distribution - corresponding to random i.i.d. frames - the "most compact" subset spectrum we can hope for? One of the key results of this survey is that better *deterministic* frames do exist. In fact, a certain class of highly symmetric frames obeys asymptotic concentration of the spectrum of a randomly-selected subset to a universal limiting distribution – similarly to the case of a completely random (i.i.d.) matrix. Crucially, this limiting distribution is *more compact* than the MP distribution.

Equiangular tight frames (ETF) are in a sense the most geometrically symmetric family of frames [26], [54], [148]. They have numerous applications in communications and signal analysis, [90], and their study brings together geometry, combinatorics, probability, and harmonic analysis. Interestingly, as we shall see in Chapter 4, one construction of ETFs corresponds to signal expansion with an *irregular* Fourier transform, [153], as opposed to the low-pass frame (LPF) mentioned above.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

A series of recent papers [71]–[75], [92], demonstrated that ETFs, as well as other deterministic tight frames that we term "near ETFs", exhibit an RMT-like behavior familiar from Multivariate ANalysis Of VAriance (MANOVA) [46], [106], [146]. We shall call this phenomenon the *ETF-MANOVA relation*.

Specifically, the work in [74] showed empirically that for any frame within this class of frames, the eigenvalue distribution of a randomly-selected subset appears to be indistinguishable from that of a random matrix taken from the MANOVA (Jacobi) ensemble. The work in [71], [73], [75], [92] further partially proved analytically[4] that as $n \to \infty$, for aspect ratios $m/n \to \gamma$ and $k/m \to \beta$, this eigenvalue distribution converges in distribution almost surely to Wachter's limiting MANOVA distribution parameterized by γ and β [146]. The concluded asymptotic performance (1.4) of the ETF family,

$$\bar{\Psi}(\text{ETF}, \gamma, \beta) = \Psi(\text{MANOVA}, \gamma, \beta), \tag{1.5}$$

is strictly better than $\Psi(\text{MP}, \beta)$, the asymptotic performance of random (i.i.d.) frames, for various performance measures Ψ. Figure 1.1 shows that the gain of MANOVA over MP is ~ 2 dB in noise amplification, and ~ 0.35 bit in capacity. We conjecture that in terms of these performance measures, the MANOVA distribution is, in fact, the *most compact* typical sub-frame spectrum achievable by *any* unit-norm frame.

In this survey paper we propose a common framework for these topics, located at the intersection of information theory and neighboring fields. Chapter 2 addresses primarily the information theory audience, and motivates a passage from digital codes to low-pass interpolation and analog frame codes, through a side-information source coding problem. Chapter 3 formalizes the notion of a performance measure

[4] The proof is complete for the case $\gamma = 1/2$ [92]. For a general $0 < \gamma < 1$, [73] establishes a recursive formula in r for the asymptotic mean rth moment of a randomly-selected ETF subset, for $r = 1, 2, \ldots$. Using a symbolic computer program, we were able to verify that this formula coincides with the first 10 MANOVA moments (above which the complexity explodes). A proof of the identity for a general $r \in \mathbb{N}$ remains a fascinating open problem.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

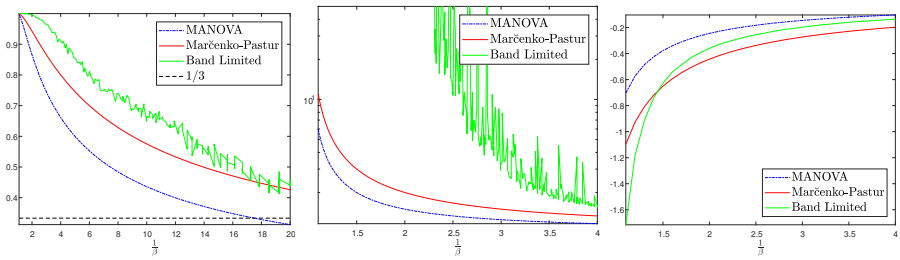


Figure 1.1: Comparison of the asymptotic performance measures $\Psi(\text{MANOVA}, \beta, \gamma)$, $\Psi(\text{Marchenko-Pastur}, \beta)$ and $\Psi(\text{Band Limited}, \beta, \gamma)$, as a function of β, for $\gamma = 1/2$ (corresponding to sub-frame aspect ratio β from a redundancy-2 frame, in the families of ETF, random (i.i.d.), and LPF, respectively). Each figure corresponds to a different sub-frame performance measure Ψ: (Left) restricted isometry property (RIP) associated with compressed sensing (lower is better); (Middle) noise amplification associated with MD and CDC (lower is better); (Right) Shannon transform associated with the capacity of NOMA-CDMA (higher is better). The 1/3 line in the left figure is the sharp RIP bound for sparse signal and low rank matrix recovery [21].

that depends on the eigenvalue spectrum of a sub-frame; e.g., noise amplification amounts to the harmonic mean of the spectrum, while capacity (Shannon transform) amounts to the geometric mean of the spectrum. Chapter 4 gives background from frame theory (in particular, earlier results on the spectral properties of sub-frames motivated by compressed sensing), while Chapter 5 gives the relevant background on random matrix theory.

The two highlights of this survey are (*i*) the *ETF-MANOVA relation*, connecting frame theory with random matrix theory, and (*ii*) the (still mostly open) possibility of *ETF superiority*. The first highlight is divided between two sections: Chapter 6 describes the empirical results of [74] regarding the universal behavior of sub-frames of ETFs and "near ETFs"; and Chapter 7 develops analytically the convergence to the MANOVA limit distribution based on the moment method (see footnote 4 above) [71], [75], [92]. To support the ETF superiority claim, we examine numerically in Chapter 9 some of the applications listed in Table 1.1; and we prove analytically in Chapter 10 the *erasure Welch bound*, [75], which implies that tight frames have the smallest sub-frame spectral

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

variance among all unit-norm frames, and that ETFs have the smallest sub-frame spectral kurtosis among all unit-norm tight frames. In between these two highlights, Chapter 8 proves some sub-frame performance inequalities (in the flavor of the information-theoretic inequalities of [37]), which explain the role of the sub-frame aspect ratio k/m as a design parameter. Finally, Chapter 11 concludes and lists interesting open questions and conjectures that arise in this area.

1.1 Notation

A finite sequence of integers $\{1, \ldots, n\}$ is denoted as $[n]$. Bold letters $\mathbf{f}, \mathbf{x}$ etc. denote column vectors. The $n \times n$ identity matrix is denoted by I_n. Dagger $(\cdot)^\dagger$ denotes transpose or conjugate (Hermitian) transpose, according to the context. The set of all k-subsets of the set $[n]$ is denoted $\binom{[n]}{k}$, or $\{S \subset [n] : \ |S| = k\}$, or simply $\{S : \ |S| = k\}$ when the context is clear. $\mathbb{E}$ denotes expectation. Throughout we try to keep the following glossary:

m	vector space dimension
n	frame size
k	sub-frame size
p	selection probability
γ	frame aspect ratio m/n
β	sub-frame aspect ratio k/m
F	$m \times n$ frame matrix
S	k-subset of $[n]$
F_S	$m \times k$ sub-frame matrix
Ψ	performance measure

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

2

An information-theoretic toy example

To set the stage for the information theory audience, we begin with an "(n, m, k)" source coding problem that originally motivated this study [72], [95].

2.1 Source coding with erasures

Consider the setup shown in Figure 2.1, of a source with "distortion side information" at the encoder [95]. Specifically, the encoder of a (white) Gaussian source $X \sim N(0, \sigma_x^2)$ has access to a statistically independent Bernoulli(p) variable S that reveals whether the source sample X is "important" (to be reconstructed under a squared-error distortion) or not. The distortion measure is thus given by

$$d(x, \hat{x}, s) = \begin{cases} (\hat{x} - x)^2, & \text{if } s = 1 \quad (\text{"important"}) \\ 0, & \text{if } s = 0 \quad (\text{"don't care"}), \end{cases} \tag{2.1}$$

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

and the mean-squared error (MSE), i.e., the average distortion per important sample in encoding the block $X_1, \ldots, X_n$, is given by[1]

$$D = \frac{1}{np} \sum_{t=1}^{n} 1_{\{S_t=1\}} \cdot \mathbb{E}\{(\hat{X}_t - X_t)^2\}. \tag{2.2}$$

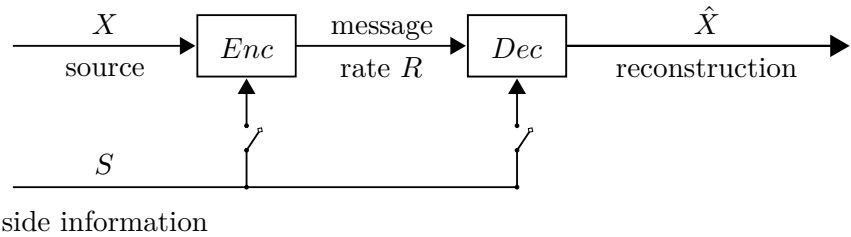


Figure 2.1: Source coding with distortion side information. Our problem of interest is when only the first switch in connected.

If also the decoder had access to the side information, then the best achievable rate-distortion trade off is given by the conditional rate-distortion function of X given S [11], [34]:

$$R_{X|S}(D) = \min_{\hat{X}} I(X; \hat{X}|S) = \frac{p}{2} \cdot \log_2 \left(\frac{\sigma_x^2}{D} \right) \tag{2.3}$$

bits per source sample, where a possible minimizing $p(\hat{x}|x, s)$ in (2.3) is the optimal quadratic-Gaussian test-channel $p^*(\hat{x}|x)$ for $s = 1$, and a degenerate channel (say, $\hat{X} \equiv 0$) for $s = 0$. This simply amounts to ignoring the non-important samples, and communicating the $k \approx np$ out of n important samples at a rate of $(1/2) \log_2(\sigma_x^2/D)$ bits per sample. Interestingly, like in some other favorite side-information problems in information theory, [33], [34], there is no loss for having access to the side information at just one end of the communication channel: the same rate-distortion trade off (2.3) can be achieved even if S is available

[1] For simplicity, the normalization is by the *expected* number of important samples: $np = \mathbb{E}\{\sum_{t=1}^{n} 1_{\{S_t=1\}}\}$.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

only at the encoder.[2] A formal proof is obtained by computing the rate-distortion function of the equivalent source pair (X, S), under the distortion measure (2.1):

$$R_{X,S}(D) = \min_{\hat{X}} I(X, S; \hat{X}) = \frac{p}{2} \cdot \log_2 \left(\frac{\sigma_x^2}{D} \right), \tag{2.4}$$

where the minimizing $p(\hat{x}|x, s)$ now has the optimum marginal $p^*(\hat{x})$ for *both* values of s; i.e., $p(\hat{x}|x, s = 1) = p^*(\hat{x}|x)$ and $p(\hat{x}|x, s = 0) = p^*(\hat{x})$. It follows from the chain rule that the minimizing $\hat{X}$ in (2.4) has *zero* mutual information with the side information S, i.e., $I(S; \hat{X}) = 0$, which is a necessary feature of a zero-loss solution [95].

2.1.1 Complexity of the all-digital solution

The information-theoretic promise comes however at a high price: the complexity of n-dimensional vector quantization. Specifically, to achieve the rate-distortion function (2.4) one needs to draw 2^{nR} codewords of size n at random, and to encode the source vector $X_1, \ldots, X_n$ into a jointly typical codeword: close to the important samples and arbitrary at the non-important samples [34]. Clearly, the encoding complexity is exponential in n. A simple naive solution would be first to encode the side information, i.e., describe the location of the important samples at a rate of $H_b(p) = p \log(1/p) + (1 - p) \log(1/(1 - p))$, e.g., 1 bit per sample for $p = 1/2$; and then to quantize only the important samples using a simple *scalar* quantizer (with just a small loss compared to optimal vector quantization).[3] If however we do not want to pay the rate penalty of ~ 1 bit per sample for encoding the side information, then we cannot replace the complex n-dimensional codebook by simple scalar quantization of the important samples because their location is a

[2] This setup can be seen as a counterpart of the Slepian–Wolf problem, [34], [124], of "statistical side information" at the decoder.

[3] Entropy-coded scalar quantization exceeds the rate-distortion function by only $\approx 1/4$ bit per sample at high resolution conditions [60], [159]

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

priori unknown. Can we simplify the solution without sacrificing too much in rate-distortion performance?

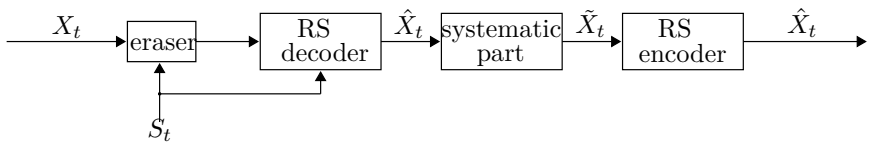


Figure 2.2: Reversed Reed–Solomon coding for a discrete source with erasures known at the encoder. Source $= X_1, \ldots, X_n$, reconstruction $= \hat{X}_1, \ldots, \hat{X}_n$, compressed version $= \tilde{X}_1, \ldots, \tilde{X}_k$.

2.1.2 A simple solution in the discrete case

Martinian et al. suggested an intriguing simple yet optimal solution based on a *reversed Reed-Solomon (RS) code* for a *lossless* variant of this problem [95]. Here, the source X is uniform over a finite alphabet $\mathcal{X}$, the important samples are encoded under a Hamming distortion measure:

$$d(x, \hat{x}, s) = \begin{cases} 1_{\hat{x} \neq x}, & \text{if } s = 1 \quad (\text{“important”}) \\ 0, & \text{if } s = 0 \quad (\text{“don’t care”}), \end{cases} \tag{2.5}$$

and we target a small ($D \to 0$) average distortion. The idea of [95] was to treat the non-important samples as “erasures”, and use an erasure-correction (e.g., RS) code in *reverse*: the encoder applies RS decoding to “correct” the erasures to the codeword $\hat{\mathbf{x}}$ (that matches the source at the non-erased (important) samples $\{t : s_t = 1\}$), and then it transmits the systematic part of $\hat{\mathbf{x}}$; the decoder uses RS encoding to reconstruct $\hat{\mathbf{x}}$ from the systematic part. See Figure 2.2. Clearly, the decoder has no idea which samples are important and which are not, so $I(\mathbf{S}, \hat{\mathbf{X}}) = 0$ like in (2.4). Moreover, assuming that the number of important samples

$$k = pn \tag{2.6}$$

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

is (roughly) constant, the maximum-distance separable (MDS) property of RS codes implies that perfect erasure correction is possible using an (n, k) code.[4] Thus, the size of the systematic part is k, and the coding rate is $R = (k/n) \log_2 |\mathcal{X}|$ bits per sample, as if both the encoder and decoder knew which samples are important and which not.

2.2 Band-pass interpolation

It is tempting to mimic the philosophy of the "reverse" RS-code solution above in the quadratic-Gaussian problem of Figure 2.1. We are inspired by the view of a cyclic code (in particular, an RS code) as consisting of all codewords whose Fourier transform (over the corresponding Galois field) is zero in a certain band of frequencies [13], [14]. Hence, erasure correction can be thought of as "band-limited interpolation", while the optimum ratio k/n corresponds to the "Nyquist sampling rate" [155]. This interpretation suggests a discrete Fourier transform (DFT)-based "analog" coding scheme for real- or complex-valued vector sources, [72], as shown in Figure 2.3. Each codeword is a "low-pass" vector, i.e., an n-dimensional vector whose high DFT coefficients are zero. The encoder looks for a low-pass vector that matches the important samples, and transmits a quantized version (scalar or vector quantization) of its nonzero DFT coefficients. The decoder then reconstructs the source by inverse DFT.

2.2.1 Coding and decoding

To define the scheme precisely, assume for now that as in (2.6) the number of important samples is constant and equal to $k = pn$. Namely, the importance side-information law is "combinatorial" (n choose k) rather than Bernoulli(p). Let m, the codewords bandwidth, $k \le m \le n$, be a design parameter to be optimized later, and let F denote the upper $m \times n$ (lowpass) section from the $n \times n$ DFT matrix:

[4] Implicit in this construction is that the size of the alphabet $\mathcal{X}$ is a prime power, and meets the Galois field requirements of the RS code.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

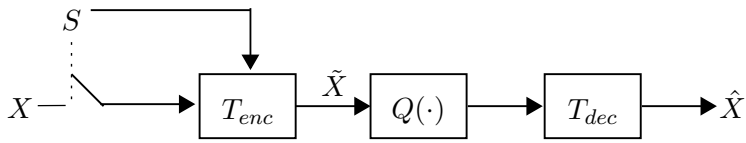


Figure 2.3: Band-limited interpolation coding scheme. Depicted are the source $X_1, \ldots, X_n$; important samples $\{X_t : s_t = 1\}$; encoder $m \times k$ transformation $T_{enc} = F_S(F_S^\dagger F_S)^{-1}$; quantizer input $\tilde{X}_1, \ldots, \tilde{X}_m$; quantizer output $Q(\tilde{X}_1), \ldots, Q(\tilde{X}_m)$; decoder $n \times m$ transformation $T_{dec} = F^\dagger$; and reconstruction $\hat{X}_1, \ldots, \hat{X}_n$. This combination of DFT to handle "erasures" and quantization to introduce "noise" can be viewed as *hybrid analog–digital* coding.

$$F = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 & w & w^2 & \ldots & w^{n-1} \\ 1 & w^2 & w^4 & \ldots & w^{2(n-1)} \\ \vdots & & & & \\ 1 & w^m & w^{2m} & \ldots & w^{m(n-1)} \end{bmatrix}, \tag{2.7}$$

where $w = e^{-j2\pi/n}$ and $j = \sqrt{-1}$, and where the $1/\sqrt{m}$ pre-factor normalizes each column to a unit norm. The vertical axis corresponds to "frequency" and the horizontal axis to "time". The matrix F defines an "m lowpass" codebook, i.e., vectors $\mathbf{x}' \in \mathbb{C}^n$ whose higher $n - m$ DFT coefficients are zero; these vectors live in the row space of F, so $\mathbf{x}' = F^\dagger \tilde{\mathbf{x}}$ for some "frequency vector" $\tilde{\mathbf{x}}$ in $\mathbb{C}^m$. The subset of the time instances of important samples, $S = \{t : s_t = 1\} \subset [n]$, defines the $m \times k$ sub-matrix F_S, (1.3), and a corresponding k sub-vector $\mathbf{x}_S$ of the source vector $\mathbf{x} = (x_1, \ldots, x_n)$. The encoder looks for a codeword $\mathbf{x}'$ that matches the important samples sub-vector $\mathbf{x}_S$, and computes its corresponding frequency vector $\tilde{\mathbf{x}}$:

$$\mathbf{x}_S = (\mathbf{x}')_S = F_S^\dagger \tilde{\mathbf{x}}. \tag{2.8}$$

If $m = k$, then the solution $\tilde{\mathbf{x}} = (F_S^\dagger)^{-1}\mathbf{x}_S$ for (2.8) is unique.[5] For $m > k$ the solution is generally not unique, and for reasons to be clear

[5] Any band of rows of the DFT matrix is "full spark", i.e., any square sub-matrix of a DFT band is invertible [52], [126].

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

soon, a good estimate for $\tilde{\mathbf{x}}$ is the least-squares (LS) "pseudo inverse" solution

$$\tilde{\mathbf{x}}_{\mathrm{LS}} = F_S \cdot (F_S^{\dagger} \cdot F_S)^{-1} \cdot \mathbf{x}_S, \tag{2.9}$$

i.e., the $\tilde{\mathbf{x}}$ satisfying (2.8) with the minimum squared-norm.[6] After estimating $\tilde{\mathbf{x}}$, the encoder transmits a quantized version $Q(\tilde{\mathbf{x}})$ to the decoder, which reconstructs the source via inverse DFT:[7]

$$\hat{\mathbf{x}} = \mathrm{IDFT}[Q(\tilde{\mathbf{x}})|0, \ldots, 0] = F^{\dagger} \cdot Q(\tilde{\mathbf{x}}). \tag{2.10}$$

Indeed, if we neglect the quantization operation $Q(\cdot)$ in (2.10), then by (2.8) the important samples are perfectly reconstructed, i.e., $\hat{x}_t = x_t$, for all t in S.

2.2.2 Rate-distortion performance

The rate-distortion performance of this analog coding scheme depends on the characteristics of the quantization operation $Q(\cdot)$. An efficient analytical tool is the additive-noise "test channel" model of entropy-coded (subtractive) dithered quantization (ECDQ) [159], [161]. For a given pattern S of important samples, the ECDQ coding rate in bits per source sample [72] is given by

$$R_{\mathrm{analog}}(S, D) = \frac{m}{2n} \cdot \log\left(1 + \frac{\frac{1}{m}\mathbb{E}\{\|\tilde{\mathbf{X}}\|^2\}}{D}\right) \tag{2.11}$$

where the expectation $\mathbb{E}\{\cdot\}$ is with respect to randomness of the source $\mathbf{X}$. See Appendix A. As explained earlier, LS estimation (2.9) minimizes the squared norm of $\tilde{\mathbf{X}}$, hence it minimizes the coding rate (2.11) among all possible solutions for (2.8). Using (2.9); recalling the definition of the system parameters: the non-erasure probability p, the aspect ratio

[6] Note that (2.9) is the unique solution $\tilde{\mathbf{x}}$ for (2.8) in the column space of F_S (which is a k-dimensional subspace of $\mathbb{C}^m$), and any other solution can be written as $\tilde{\mathbf{x}}_{\mathrm{LS}} + \mathbf{z}$ for some $\mathbf{z}$ in the orthogonal subspace.

[7] An equivalent solution for the case where n/m is an integer, is to *uniformly* re-sample $\mathbf{x}' = F^{\dagger}\tilde{\mathbf{x}}_{\mathrm{LS}}$, transmit the m quantized uniform samples, and reconstruct by $m : n$ band-limited interpolation.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

of the $m \times n$ matrix F, and the aspect ratio of the $m \times k$ sub-matrix F_S,

$$p = \frac{k}{n}, \quad \gamma = \frac{m}{n} \quad \text{and} \quad \beta = \frac{k}{m} \; ; \tag{2.12}$$

and assuming a white-Gaussian source $\mathbf{X} \sim N(0, \sigma_x^2 I)$, we can rewrite the ECDQ coding rate (2.11) as

$$R_{\text{analog}}(S, D) = \frac{p}{2\beta} \cdot \log \left(1 + \frac{\sigma_x^2}{D} \cdot \beta \cdot \Psi_{\text{MSE}}(F_S) \right) \tag{2.13}$$

where

$$\Psi_{\text{MSE}}(F_S) \stackrel{\Delta}{=} \frac{\text{trace}\left((F_S^\dagger F_S)^{-1} \right)}{k} \tag{2.14}$$

is the *signal amplification factor* of the sub-matrix F_S.[8] Finally, the *operational rate-distortion function* of the system is the average of the ECDQ coding rate over all k-subsets $S \subset [n]$:

$$\bar{R}_{\text{analog}}(D) = \frac{1}{\binom{n}{k}} \sum_{S:|S|=k} R_{\text{analog}}(S, D) \tag{2.15}$$

as in (1.4).

In view of (2.11)-(2.15), the system performance differs from the information-theoretic bound (2.4) in three aspects, which contribute to the loss of the analog coding scheme with respect to the optimum performance (2.3): (i) the "+1" term inside the log; (ii) the design parameter $\beta \leq 1$ that appears both as a pre-log factor and inside the log; and (iii) the signal-amplification factor (2.14). The "+1" term is negligible for high signal-to-distortion ratio $\sigma_x^2 \gg D$, and it can be eliminated (for $\beta = \Psi_{MSE} = 1$) if we replace LS estimation by minimum mean-squared error (MMSE) estimation [72]. As for the role of β, since the function $x \log(1 + a/x)$ is monotonically *increasing* in x, choosing $\beta < 1$ can only be justified if the effect on reducing the signal amplification factor $\Psi_{\text{MSE}}(F_S)$ is stronger for sufficiently many subsets S.

[8] The notation uses "MSE" because (2.14) plays the role of mean-square error (MSE) amplification in other problems; see [72], [83], [97], [122], [144] and Chapter 9.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

2.3 Non-uniform sampling and signal amplification

The signal amplification factor (2.14) plays a key role in understanding frame goodness. As we shall see, it is greater than or equal to 1 for every subset S, and it tends to increase with the subset size k, i.e., with β (see Chapter 8). The equality $\Psi_{MSE}(F_S) = 1$ holds if and only if the column vectors of F_S are orthogonal. Indeed, if $l = n/k$ is an integer and the pattern S happens to fall on a uniform grid (i.e., $t_i = t_1 + (i-1)l$ modulo n, for $i = 1, \ldots, k$), then for $\beta = 1$, the sub-matrix F_S becomes a $k \times k$ DFT matrix, up to a constant phase shift. Thus, F_S is unitary and $F_S^\dagger F_S$ is an identity matrix, $\Psi_{MSE} = 1$, and the coding rate $R_{\text{analog}}(S = \text{uniform}, D)$ coincides with the information-theoretic bound (2.4) at high signal-to-distortion ratio.

A typical pattern S of a Bernoulli process is, however, far from being uniform, and so the signal amplification factor (2.14) is typically strictly greater than one. This phenomena is known in the literature as *noise amplification* in signal recovery from noisy non-uniform samples [83], [98], [122], [144]. It implies that the operational rate-distortion function (2.15) fails to achieve the information-theoretic bound (2.4), as was already observed (for $\beta = 1$) in [95] and [96]. Figure 2.4 shows a histogram of $R_{\text{analog}}(S = \text{random}, D)$ for two values of β. In both cases, the average loss compared to (2.4) is still much higher than the ~ 1 bit penalty of the naive side information coding scheme (Section 2.1.1).

2.4 From interpolation to frame codes

Fortunately, the analog "band-limited interpolation" coding approach can be saved by replacing the low-pass DFT matrix (2.7) by frames which are more robust to the erasure pattern. Analog codes follow a long line of works in the communications literature, that used linear transform-based codes for noise and erasures, under names like "DFT codes" [118], [151], "coding over the reals" [94], "analog BCH/RS codes" [119], [133], and in a broad sense also in *sigma-delta modulation* [28],

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

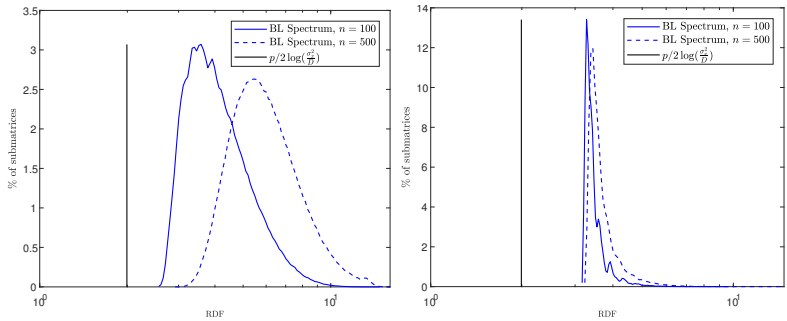


Figure 2.4: Histogram of the operational RDF $R_{\text{analog}}(S, D)$ for random subsets S from a low-pass (BL) frame with two frame sizes ($n = 100$ and 500), for $\sigma_x^2/D = 1000$, $p = 0.4$, and: (left) $\gamma = 1/2$ and $\beta = 0.8$; (right) $\gamma = 2/3$ and $\beta = 0.6$. The vertical line shows Shannon's RDF (2.3). Note that the overall effect of decreasing β below one is to reduce the rate, and that a larger frame size ($n = 500$) is worse than a smaller one ($n = 100$).

code-division multiple access [120], *multiple-description coding* [67], [110], *space-time coding* [131], *array imaging* [83], [129], and *coded distributed computation* [85]. In the context of the current work, the low-pass DFT matrix F is a special case of a *harmonic frame*, i.e., a subset of the rows of the DFT matrix [132], [148]. This viewpoint opens our discussion to general frames in the form of (1.1), and the rich frame theory. Interestingly, in a sharp contrast to the consecutive band of frequencies in the low-pass frame (2.7), the performance of analog coding is optimized by an "irregular" (random-like) selection of the DFT rows,[9] which eventually will lead us to *equiangular tight frames* [153].

[9] Low-pass spectrum can fit (or "explain") well arbitrary sampling values over a *uniform* time grid. Irregular spectrum provides a better (lower squared-norm) fit when the sampling times lie on a *non*-uniform grid [50], [94], [144].

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

3

Sub-frame performance measures

The previous chapter provided an information-theoretic motivation to assess a frame by averaging a specific performance measure (noise amplification) over sub-frames of a fixed size. In this section we show that this, as well as other relevant frame performance measures [140], are functions of the *eigenvalue spectrum* of sub-frames. The sub-frame-average viewpoint (which seems to be new) will motivate our search for "good" frames. We start yet with another well known frame performance measure, the mean-squared pairwise cross correlation, and the notions of the Welch bound and a *tight frame*.

3.1 The Welch bound

The development of code-division multiple access (CDMA) triggered the introduction of the Welch bound (WB) [149] on the average squared-correlation between all distinct pairs of an over-complete set of unit-norm vectors $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ in $\mathbb{R}^m$ or $\mathbb{C}^m$:

$$\frac{1}{(n-1)n} \sum_{i \neq j} |\langle \mathbf{f}_i, \mathbf{f}_j \rangle|^2 \geq \frac{n-m}{(n-1)m}. \tag{3.1}$$

Documents
Document Sharing Hub

Full document details: IEEE Asymptotic Frame Theory for Analog Coding

The Welch bound is achieved with equality if and only if the frame $F = [\mathbf{f}_1 \cdots \mathbf{f}_n]$ is *tight*, i.e., $F \cdot F^\dagger = (n/m) \cdot I_m$ [81], [120], [153]. For example, the low-pass frame (2.7) and the repetition frame $F = [I_m \cdots I_m]$ (or any union of orthonormal bases) are tight. The weakness of the Welch bound (3.1), however, is that it measures the frame as a *whole*. While "the whole is greater than the sum of its parts", in our problems of interest every (or almost every) individual part must be good, so tightness is not sufficient.

3.2 Noise amplification

Let us return to the erasure source coding problem of Section 2.1. At high signal-to-distortion ratio, each term in the operational rate-distortion function (2.15) becomes $(m/2n) \cdot \log(\Psi_{\text{MSE}}(F_S))$, plus some constant (independent of F); this gives rise to a subset-average performance measure of the form (1.4):

$$L_{\text{MSE}}(F,k) = \frac{1}{\binom{n}{k}} \sum_{S:|S|=k} \log(\Psi_{\text{MSE}}(F_S)), \tag{3.2}$$

to which we refer as the "MSE loss" of the frame F. Here, $\Psi_{\text{MSE}}(F_S)$ is the "noise amplification" factor (2.14), and the sum is over all k-subsets S of $[n]$, i.e., over all $m \times k$ sub-matrices F_S of F.

Note that for $k = 2$, the MSE loss (3.2) can be lower bounded using the Welch bound (3.1): $L_{\text{MSE}}(F,2)$ is the average of $\log(1/(1-|\langle \mathbf{f}_i, \mathbf{f}_j \rangle|^2))$ over all distinct pairs of frame vectors, which, by Jensen's inequality and the Welch bound, is greater than or equal to $\log((n-1)m/(m-1)n)$ (which is roughly zero for large m and n), with equality if and only if the frame is tight and all the absolute pairwise cross-correlations are equal, i.e., if and only if F is an equi-angular tight frame.

The MSE performance measure (3.2) implies a frame optimization criterion for any triplet $n \geq m \geq k$:

$$L^*_{\text{MSE}}(n,m,k) = \min_F L_{\text{MSE}}(F,k), \tag{3.3}$$

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

where the minimization is taken over all $m \times n$ unit-norm frames (1.1).[1] Once the minimal loss function (3.3) is known, we can go back and optimize the free bandwidth parameter m in the operational rate distortion function (2.15), for fixed n, k and signal-to-distortion ratio σ_x^2/D.

We note that other problems of interest give rise to the complementary case where $n \geq k \geq m$ (i.e., $\beta = k/m$ is now greater than 1), in which case the MSE performance measure (2.14) is computed with respect to the sub-frame Hessian $F_S F_S^\dagger$ rather than Gram matrix $F_S^\dagger F_S$:

$$\Psi_{\mathrm{MSE}}(F_S) = \frac{\mathrm{trace}\left((F_S F_S^\dagger)^{-1}\right)}{k}.$$

3.2.1 Asymptotic goodness

Information theory tends to look on the asymptotic performance of large block codes, at some fixed redundancy. It is therefore tempting to explore the large dimensional behavior of (3.3), when the aspect ratios $\beta = k/m$ and $\gamma = m/n$ in (2.12) are held fixed:

$$L^*_{\mathrm{MSE}}(\gamma, \beta) \triangleq \liminf_{n\to\infty} L^*_{\mathrm{MSE}}\Big(n, m = \lfloor n\gamma \rfloor, k = \lfloor n\gamma\beta \rfloor\Big), \tag{3.4}$$

and to find a frame family with aspect ratio going to γ that asymptotically achieves it. Alternatively, we can use Bernoulli(p) (rather than combinatorial) selection of the pattern S (as in the original source coding problem of Section 2.1), where the subset size k becomes a random variable with expectation pn. We expect that rare events where k significantly deviates from pn are asymptotically negligible, so

$$L^*_{\mathrm{MSE}}(\gamma, p) = L^*_{\mathrm{MSE}}(\gamma, \beta = p/\gamma). \tag{3.5}$$

Assuming this asymptotic MSE goodness measure is not zero, it reflects a fundamental loss of analog coding with respect to the information theoretic bound (2.3).

[1] The spark of F must be larger than k, otherwise the sum in (3.2) is infinite [52].

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

3.3 Spread of eigenvalues

The performance measures (2.14) and (3.2) are functions of the $k \times k$ sub-frame Gram matrix $F_S^\dagger F_S$. It is useful and insightful to express them in terms of the eigenvalues $\lambda_1, \ldots, \lambda_k$ of $F_S^\dagger F_S$ (i.e., the squared nonzero singular values of the sub-matrix F_S). Recall that the trace of a matrix is equal to the sum of its eigenvalues, so $\lambda_1 + \ldots + \lambda_k = \text{trace}\left((F_S^\dagger F_S)\right) = k$, where the last equality follows because F_S is composed of k unit-norm vectors. Also, the eigenvalues of an inverse matrix are the inverse eigenvalues. Thus, the signal amplification factor (2.14) of a specific sub-matrix can be written as

$$\Psi_{\text{MSE}}(F_S) = (1/k)\left[\lambda_1^{-1} + \ldots + \lambda_k^{-1}\right] = \frac{\left[\frac{\lambda_1 + \ldots + \lambda_k}{k}\right]}{\left[\frac{1/\lambda_1 + \ldots + 1/\lambda_k}{k}\right]^{-1}}, \qquad (3.6)$$

i.e., as the *arithmetic-to-harmonic means ratio* (AHMR) of the eigenvalues of the Gram sub-matrix $F_S^\dagger F_S$. By the *means inequality*, the AHMR is greater than or equal to one, with equality if and only if all the eigenvalues are equal (and in our case, if all the eigenvalues are equal to one). Thus, $\Psi_{\text{MSE}}(F_S)$ is greater than or equal to one, with equality if and only if $F_S^\dagger F_S$ is an identity matrix, i.e., the column vectors of F_S are orthogonal.

Clearly, since $n > m$, not all subsets of $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ can be orthogonal, so $\Psi_{\text{MSE}}(F_S)$ must be greater than one for some patterns S. Figure 3.1 illustrates the spread of the eigenvalues $\lambda_1, \ldots, \lambda_k$ of $F_S F_S^\dagger$ around their mean (i.e., 1) for the low-pass frame (2.7), for three possible patterns: uniform sampling, random-like sampling, and consecutive samples (batch). We observe that for this frame the eigenvalue spread is very sensitive to the pattern S. Thus, a "good" frame should be robust in the sense that the eigenvalue spread of (most of) its subsets is small.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

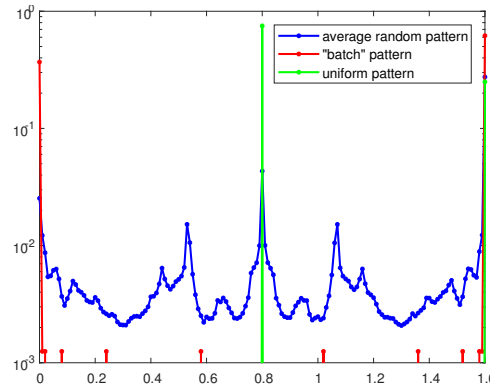


Figure 3.1: Eigenvalues of subsets of the LPF frame (2.7), with $p = 1/2$, $\gamma = 0.625$, $\beta = 0.8$ ($n = 1600$, $m = 1000$, $k = 800$), for several possible patterns S: uniform, random, batch.

3.4 Shannon transform

The Shannon transform of the sub-matrix F_S is the capacity of a multiple-input multiple-output (MIMO) AWGN channel with a transfer matrix F_S [137], [140], [145], [157]. At high signal-to-noise ratio, and assuming $n \geq k \geq m$, the Shannon transform gives rise to a "log-det" *geometric-to-arithmetic means ratio* goodness measure of the $m \times k$ sub-frame F_S:

$$\Psi_{\text{Shannon}}(F_S) = \frac{1}{m} \cdot \log\left(\det[\beta^{-1} \cdot F_S F_S^\dagger]\right) \tag{3.7a}$$

$$= \log\left(\frac{\sqrt[m]{\lambda_1 \cdots \lambda_m}}{\frac{\lambda_1 + \ldots + \lambda_m}{m}}\right). \tag{3.7b}$$

Similarly to (3.6), the argument of the logarithm is smaller than or equal to one, with equality if and only if the eigenvalues of $F_S F_S^\dagger$ are identical. Thus, Ψ_{Shannon} is another measure for the spread of the eigenvalues of F_S. The corresponding average sub-frame performance measure (1.4) is the "Shannon goodness":

$$L_{\text{Shannon}}(F, k) = \frac{1}{\binom{n}{k}} \sum_{S:|S|=k} \Psi_{\text{Shannon}}(F_S) \tag{3.8}$$

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

where the sum is over all k-subsets of $[n]$. As above, we define the associated frame optimization criterion

$$L^*_{\text{Shannon}}(n,m,k) = \max_F L_{\text{Shannon}}(F,k), \tag{3.9}$$

as well as the asymptotic counterparts $L^*_{\text{Shannon}}(\gamma,\beta)$ and $L^*_{\text{Shannon}}(\gamma,p)$ (for Bernoulli(p) selection) as in (3.3)-(3.5).

3.5 Statistical RIP and condition number

We now turn to performance measures which look only on the *edges* of the sub-frame spectrum. In compressed sensing, the Restricted Isometry Property (RIP) is a well-known sufficient condition for sparse recovery [24]. A "sensing" frame F satisfies a δ-RIP of order k if the spectral radius of $F_S^\dagger F_S - I_k$, i.e.,

$$\Psi_{RIP}(F_S) = \max\{\lambda_{\max}(F_S) - 1, 1 - \lambda_{\min}(F_S)\}, \tag{3.10}$$

is uniformly bounded by δ for all k-subsets S of $[n]$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalues of the Gram matrix $F_S^\dagger F_S$. For $\delta < 1/2$, this condition guarantees perfect recovery via ℓ_1-minimization in sensing any $k/2$-sparse vector, while $\delta < 1/3$ allows one to sense any k-sparse vector [21], [23], [24], [56].

The RIP is, however, dominated by the worst subset, which is too restrictive in view of the "typical" approach of information theory. The *statistical RIP* allows some fraction of the subsets to exceed the uniform bound [22]. The fraction of "good" subsets can be written as an average of the indicator of the event $\{\Psi_{RIP}(F_S) \le \delta\}$ over the subset S, and we may wish to maximize it over all $m \times n$ frames F:

$$L^*_{StRIP}(n,m,k,\delta) = \max_F \frac{1}{\binom{n}{k}} \sum_{S:\,|S|=k} 1_{\{\Psi_{RIP}(F_S)\le\delta\}}. \tag{3.11}$$

To complete the picture, we shall also define the closely related notion of sub-frame condition number

$$\Psi_{\text{cond}}(F_S) = \frac{\lambda_{\max}(F_S)}{\lambda_{\min}(F_S)} \tag{3.12}$$

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

and the associated sub-frame average (1.4) (or statistical δ-average as in (3.11)), as well as the corresponding optimum "(n, m, k) frame" as in (3.3) and (3.9).

3.6 Sub-frame rectangularity

In Chapter 8 we show that the sub-frame performance measures above (in particular, $L_{\mathrm{MSE}}(n, m, k)$ and $L_{\mathrm{Shannon}}(n, m, k)$) exhibit a natural monotonic behavior as a function of the sub-frame aspect ratio k/m. The more rectangular is the sub-frame, the better is the average goodness of the frame. See also Figure 2.4.

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

4

Frame theory

In this section we formalize the discussion of frames and introduce a few important examples. Frame theory is the study of overcomplete systems in a Hilbert space. In particular, a sequence $\{\mathbf{f}_i\}_{i\in I}$ of vectors in a Hilbert space H is said to be a *frame* if there exist constants $a, b > 0$ such that

$$a\|\mathbf{x}\|^2 \leq \sum_{i\in I} |\langle \mathbf{x}, \mathbf{f}_i\rangle|^2 \leq b\|\mathbf{x}\|^2 \tag{4.1}$$

for every $\mathbf{x} \in H$. If $a = b$, then $\{\mathbf{f}_i\}_{i\in I}$ is known as a *tight frame*, in which case the frame enjoys a painless reconstruction formula [36]:

$$\mathbf{x} = \frac{1}{a}\sum_{i\in I} \langle \mathbf{x}, \mathbf{f}_i\rangle \mathbf{f}_i.$$

As an example, one may verify that the *low-pass frame* (2.7) obtained from the first m rows of the $n \times n$ discrete Fourier transform matrix is a tight frame for $\mathbb{C}^m$. In general, one may select any m rows from any $n \times n$ unitary matrix to produce a tight frame of n vectors in $\mathbb{C}^m$. As another example, one may combine r orthonormal bases for $\mathbb{C}^m$ to produce a tight frame consisting of $n = rm$ vectors. For instance, the *repetition frame* consists of r copies of the same basis, while the

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

spike-and-sine frame combines the identity basis with the Fourier basis. Note that for *unit-norm tight* frames, the constants in (4.1) satisfy $a = b = n/m$.

Frames were introduced in 1952 by Duffin and Schaeffer [42] in the context of nonharmonic Fourier series. More recently, frames have been used as models for data processing, as information about $\mathbf{x} \in H$ can be recovered from the frame coefficients $\{\langle \mathbf{x}, \mathbf{f}_i \rangle\}_{i \in I}$. Most of the work in this direction has focused on the case in which H has finite dimension, in which case we put $I = [n]$, and we abuse notation by letting F denote both the sequence $\{\mathbf{f}_i\}_{i \in [n]}$ and the matrix whose ith column is $\mathbf{f}_i$. Some frames are particularly well suited for information recovery after corruption in the coefficient domain. For example, of all frames with unit-norm elements, tight frames are the most robust to a single erasure under additive white gaussian noise [66].

Intuitively, a frame is more robust to data corruption when no two frame elements capture similar information. Suppose each frame element has unit norm, and for the sake of illustration, consider $\mathbf{x}$ drawn from a gaussian distribution with identity covariance. Then each frame coefficient $\langle \mathbf{x}, \mathbf{f}_i \rangle$ has standard gaussian distribution, and the covariance between $\langle \mathbf{x}, \mathbf{f}_i \rangle$ and $\langle \mathbf{x}, \mathbf{f}_j \rangle$ is $\langle \mathbf{f}_i, \mathbf{f}_j \rangle$. As such, if we want to minimize the common information between all pairs of frame elements, we are inclined to minimize the *coherence* of F, defined as

$$\nu(F) \triangleq \max_{\substack{i,j \in [n] \\ i \neq j}} |\langle \mathbf{f}_i, \mathbf{f}_j \rangle|.$$

In fact, of all unit norm tight frames, the ones that minimize coherence are most robust to two erasures under additive white gaussian noise [77].

An alternative interpretation is available in the context of coding theory. Recall that given a compact metric space X with metric d, an *optimal code* in (X, d) of size n is a subset $C \subseteq X$ for which the minimum distance

$$\delta(C) \triangleq \min_{\substack{x,y \in C \\ x \neq y}} d(x, y)$$

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding

is maximized. (Here, we ask X to be compact so that an optimal code necessarily exists.) There has been extraordinary success in studying optimal codes in Hamming space and the sphere. In this survey paper, we are concerned with another metric space. Specifically, for each $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $\mathbb{FP}^{m-1}$ denote *projective space*, i.e., the set of 1-dimensional subspaces of $\mathbb{F}^m$. One may define the *chordal distance* over projective space by

$$d(S,T) \triangleq \|P_S - P_T\|_F, \qquad S,T \in \mathbb{FP}^{m-1}.$$

Here, P_S denotes the $m \times m$ orthogonal projection matrix onto S, while $\|\cdot\|_F$ denotes the Frobenius matrix norm. (Notably, antipodal points on the sphere are as far apart as possible in the spherical metric space, but they determine identical points in projective space; despite such differences, both spaces have been treated simultaneously to some extent under the common umbrella of 2-point homogeneous spaces [30], [87], [88].) For unit vectors $\mathbf{x}$ and $\mathbf{y}$, it is straightforward to verify that

$$d(\operatorname{span}\{\mathbf{x}\}, \operatorname{span}\{\mathbf{y}\}) = \sqrt{2(1 - |\langle \mathbf{x}, \mathbf{y}\rangle|^2)}.$$

It follows that optimal codes in projective space correspond to ensembles of vectors that minimize coherence.

In the following subsection, we discuss a particularly nice family of optimal projective codes known as *equiangular tight frames*. Later, we briefly survey other known optimal projective codes before discussing frames with well-conditioned subensembles.

4.1 Equiangular tight frames

This subsection is concerned with a particularly special type of frame. We start with a definition:

Definition 4.1. We say $\{\mathbf{f}_i\}_{i\in[n]}$ in $\mathbb{C}^m$ is an *equiangular tight frame (ETF)* if there exist constants $\nu, r \geq 0$ such that

$$|\langle \mathbf{f}_i, \mathbf{f}_j\rangle| = \begin{cases} 1 & \text{if } i = j \\ \nu & \text{if } i \neq j, \end{cases} \tag{4.2a}$$

Documents
Document Sharing Hub
Full document details: IEEE Asymptotic Frame Theory for Analog Coding