Department of Communications and Networking

Probabilistic Models and Algorithms for Energy-Efficient Large and Dense Wireless Sensor Networks

Giancarlo Pastor Figueroa

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 38/2016

Probabilistic Models and Algorithms for Energy-Efficient Large and Dense Wireless Sensor Networks

Giancarlo Pastor Figueroa

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall S4 of the school on 19 April 2016 at 12.

Aalto University
School of Electrical Engineering
Department of Communications and Networking
Rey Juan Carlos University, Escuela Internacional de Doctorado,
Programa de Doctorado Interuniversitario en Multimedia y Comunicaciones

Supervising professors
Professor Riku Jäntti, Aalto University, Finland
Professor Antonio J. Caamaño Fernández, Rey Juan Carlos University, Spain

Preliminary examiners
Professor Stian N. Anfinsen, UiT The Arctic University of Norway, Norway
Professor Jari Iinatti, University of Oulu, Finland

Opponents
Professor Mikael Johansson, KTH Royal Institute of Technology, Sweden
Professor Baltasar Beferull-Lozano, University of Agder, Norway

Aalto University publication series
DOCTORAL DISSERTATIONS 38/2016

ISBN 978-952-60-6681-3 (printed)
ISBN 978-952-60-6682-0 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-6682-0

Unigrafia Oy
Helsinki 2016

Finland

441 697
Printed matter

Universidad
Rey Juan Carlos

TESIS DOCTORAL

Probabilistic Models and Algorithms for Energy-Efficient Large and Dense Wireless Sensor Networks

Autor:

Giancarlo Pastor Figueroa

Directores:

Antonio J. Caamaño Fernández

Riku Jäntti

Programa de Doctorado Interuniversitario en Multimedia y Comunicaciones

Escuela Internacional de Doctorado

2016

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Giancarlo Pastor Figueroa

Name of the doctoral dissertation
Probabilistic models and algorithms for energy-efficient large and dense wireless sensor networks

Publisher School of Electrical Engineering

Unit Department of Communications and Networking

Series Aalto University publication series DOCTORAL DISSERTATIONS 38/2016

Field of research Communications Engineering

Manuscript submitted 9 October 2015 **Date of the defence** 19 April 2016

Permission to publish granted (date) 14 December 2015 **Language** English

☒ **Monograph** ☐ **Article dissertation** ☐ **Essay dissertation**

Abstract

The concept of sparsity is fundamental to understanding data collection in foreseeable large scenarios. This intuition is based on the fact that future analytics will only perform efficiently over compact information domains. This scenario is even more critical in Wireless Sensor Networks (WSN), since small sensors suffer from major limitations. Hence, Compressive Data Aggregation (CDA) is required to alleviate the work load of the sensors.

This research has proposed a compressed sensing-based protocol which follows a random sensing principle. This principle allows for the characterization of sensor interactions using methods from stochastic geometry. This work will deliver: a) a methodology to jointly analyze the compression and communication aspects of CDA in WSNs and b) a novel protocol that effectively implements compressed sensing with collision avoidance. It is called Stochastic Compressive Data Aggregation (S-CDA).

Although there have been great advances, there is a lack of methodologies to jointly analyze the compression and communication aspects of CDA in WSNs. Consequently, this dissertation seeks to bridge this gap.

Resumen

Los futuros sistemas de análisis de datos funcionarán de forma eficiente sólo sobre dominios compactos de información. Las Redes Inalámbricas de Sensores (*Wireless Sensor Networks*, WSN) necesitan operar en armonía con aquellos dominios. La razón es que los diminutos sensores que conforman una WSN sufren de grandes limitaciones. En este contexto, el concepto de compresibilidad es fundamental para formular modelos y desarrollar soluciones. Aunque hubo grandes progresos en el análisis de recogida de datos en WSNs, hay una falta de metodologías que analicen de forma conjunta los aspectos de compresión y comunicación. La presente tesis busca llenar ese vacío utilizando técnicas de sensado compresivo y geometría estocástica. Concretamente, la tesis tiene como objetivos: primero, explicar, de forma formal, las propiedades que deben de tener las funciones de compresibilidad; segundo, desarrollar métodos para caracterizar, de forma estadística, la potencia de interferencia en redes inalámbricas; y tercero, diseñar un protocolo para la recogida eficiente de datos en WSNs a partir de un modelo probabilístico que considere los aspectos de compresión y comunicación.

Keywords compressed sensing (sensado comprimido), data aggregation (agregación de datos), stochastic geometry (geometría estocástica)

ISBN (printed) 978-952-60-6681-3 **ISBN (pdf)** 978-952-60-6682-0

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2016

Pages 148 **urn** http://urn.fi/URN:ISBN:978-952-60-6682-0

A! Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Giancarlo Pastor Figueroa

Väitöskirjan nimi
Todennäköisyyspohjaisia malleja ja algoritmeja energiatehokkaille ja tiheille langattomille anturiverkoille

Julkaisija Sähkötekniikan korkeakoulu

Yksikkö Tietoliikenne- ja tietoverkkotekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 38/2016

Tutkimusala Tietoliikennetekniikka

Käsikirjoituksen pvm 09.10.2015 **Väitöspäivä** 19.04.2016

Julkaisuluvan myöntämispäivä 14.12.2015 **Kieli** Englanti

☒ **Monografia** ☐ **Artikkeliväitöskirja** ☐ **Esseeväitöskirja**

Tiivistelmä

Laitteiden internetin ja muiden tietojärjestelmien tuottama erittäin laajat tietomassat aiheuttavat haasteita datan keräämiselle ja analytiikalle. Erityisen haasteellista suurien datamäärien kerääminen on langattomissa anturiverkoissa, antureiden laskentakyvyn, muistin, tiedonsiirron ja energian rajoitteiden takia. Näin ollen langattomissa anturiverkoissa tarvitaan uusia tehokkaita ja luotettavia menetelmiä tiedon kokoamiseksi yhteen. Harvuuden käsite mahdollistaa datan keruuprosessin paremman ymmärryksen sekä uusien kompaktiin esitystapaan perustuvien tehokkaiden datan keruu ja analyysimenetelmien kehittämisen. Pakattu datan kokoaminen (*Compressive Data Aggregation*, CDA) mahdollistaa työkuorman vähentämisen datan keräys- tai fuusiopisteessä. Tämä puolestaan pidentää akkujen kestoa ja kasvattaa verkon toiminta-aikaa. Tässä työssä esitetään pakattuun mittaukseen perustuva MAC-protokolla radioresurssien jakamiseksi antureiden välillä huomioiden CDA:n tuomat mahdollisuudet. Protokollan satunnaisuus mahdollistaa sen toiminnan mallintamisen käyttämällä hiljattain esitettyjä stokastisen geometrian menetelmiä. Kehitetty MAC-protokolla muistuttaa kantoaallon tunnistukseen perustuvaa CSMA-protokollaa ja sen toiminnalle on todennäköisyyteen perustuvat rajat.

Erityisesti tässä työssä esitetään a) metodologia langattomien anturiverkkojen datan pakatun kokoamisen ja tietoliikenteen yhtäaikaiseen analyysiin; b) uusi stokastinen datan kokoaminen (S-CDA) MAC-protokolla, joka tehokkaasti yhdistää pakatun mittauksen ja pakettien törmäyksen välttämisen.

Aikaisempi tutkimus on analysoinut CDA:ta epäyhtenäisesti olettaen joko tietoliikenteen tai pakatun mittauksen ideaaliseksi. Tässä työssä tarkastellaan näitä molempia yhdessä.

Avainsanat datan kokoaminen, pakattu mittaus, stokastinen geometria

ISBN (painettu) 978-952-60-6681-3 **ISBN (pdf)** 978-952-60-6682-0

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2016

Sivumäärä 148 **urn** http://urn.fi/URN:ISBN:978-952-60-6682-0

Preface

The research work for this doctoral thesis was carried out *en cotutelle* between the Department of Communications and Networking at Aalto University and the Department of Signal Theory and Communications, Telematics and Computing at Rey Juan Carlos University around the years 2012-2015. The completion and wrap-up phases in 2015 were carried out while visiting VTT Technical Research Centre of Finland and the Department of Mathematics at the University of Texas at Austin.

This work was funded by research grants from the Spanish Ministry of Science and Innovation, the Finnish Cultural Foundation, the European Institute of Innovation and Technology, and my supervisors' own research funds. I would also like to acknowledge the grants from the Spanish Ministry of Economy and Competitiveness, the Research Foundation of Helsinki University of Technology, the Rey Juan Carlos University, and Sirkka and Heikki Luoma.

First of all, I am grateful to my esteemed supervisors, Prof. Antonio J. Caamaño and Prof. Riku Jäntti, for their guidance, support, enthusiasm, and trust that allowed me to encounter my research interests and strengths. My unofficial adviser, Prof. Inmaculada Mora-Jiménez, undoubtedly also contributed to my research.

The pre-examiners of this dissertation, Prof. Stian N. Anfinsen and Prof. Jari Iinatti, are greatly acknowledged for their valuable comments which certainly enabled to improve the readability of the manuscript. I must also express special gratitude to Dr. Duan Ruifeng for reviewing an early version of my manuscript. Furthermore, I would like to thank in advance Prof. Mikael Johansson and Prof. Baltasar Beferull-Lozano for finding time to serve as opponents at the public examination. It is an honor to share this unique experience, hopefully in a friendly manner, with such distinguished scientists.

I am thankful for all the opportunities to gain experiences beyond research: Dr. Jose Costa-Requena, for sharing with me his developed know-how of proposals writing and project management; Dr. Samuli Aalto, for giving me the opportunity to enjoy teaching after a long pause on this duty; and Prof. Jussi Autere, for supervising my business training in the EIT Digital doctoral school. Special thanks go to two mathematicians that have strongly motivated my most recent research: Research Prof. Ilkka Norros and Prof. François Baccelli.

For always handling all practical matters smoothly, I acknowledge the efforts of: Ana I. Ruiz, Elena Párraga and the HR team from Rey Juan Carlos University; Heli Liukko, Sari Kiveliö, Anita Bisi, Viktor Nässi, Kati Voutilainen and the doctoral team from Aalto University; and Katri Sarkio and Soili Adolfsson from EIT Digital.

To the fellows at work. I am pleased to be acquainted with Alejandro Rivera, Mihaela I. Chidean, Estrella Evers, Luis M. López, Eduardo del Arco, Antonio G. Marqués, Jóse L. Rojo, Aamir Mahmood, Alex Grigorevski, David González G., Juha Zidbeck (I am grateful for the hospitality at VTT), Tero Tyrväinen, Pirkko Kuusela, Jorma Kilpi, Jae Oh Woo, Christopher Mollén, Ali Khezeli, and Jianhua Mo, to name a few.

Last but foremost, I would like to thank my dear family. They are truthfully my source of strength and values: my mother Trinidad, the memory of my father German, my siblings, and my new family in Finland. My ultimately warmest thanks go to you Salla, for all the love, happiness and peace you have brought me. Also for your songs that I missed when research has taken me away.

Espoo, February 23, 2016,

Giancarlo Pastor Figueroa

Resumen

Antecedentes

Con el propósito fundamental de comprender la naturaleza y el comportamiento de sistemas reales, y de construir los modelos adecuados y proponer las soluciones correctas, las futuras Tecnologías de la Información y las Comunicaciones (*Information and Communications Technology*) utilizan grandes cantidades de datos de múltiples fuentes. El desarrollo alcanzado en estas tecnologías en las últimas décadas esta permitiendo niveles impresionantes de eficiencia en las tareas masivas de recogida y almacenamiento de datos, así como de su procesamiento. Sin embargo, la continua y creciente demanda de mayores cantidades de datos urge la búsqueda de nuevos paradigmas.

Las Redes Inalámbricas de Sensores (*Wireless Sensor Networks*), aquellas infraestructuras encargadas de la recogida de los datos en las tecnologías de la información y las comunicaciones, son las primeras en enfrentar esta urgente necesidad de nuevos paradigmas. En particular, estas redes deben implementar protocolos capaces de abordar la tarea de recogida de datos con la consiguiente tarea de compresión de información.

Las redes inalámbricas de sensores son un conjunto de dispositivos sensores electrónicos, llamados nodos en los modelos matemáticos, los cuales por lo general sufren de grandes limitaciones de procesamiento, almacenamiento, comunicaciones y energía. Comúnmente, estas redes se despliegan para observar fenómenos físicos de muy variada naturaleza, esenciales en aplicaciones de monitorización del medio ambiente o de agricultura de precisión, entre otras muchas posibles aplicaciones. Así, los datos recogidos son obtenidos de grandes áreas geográficas, durante extensos periodos de tiempo.

El despliegue de las redes inalámbricas de sensores es muchas veces sencillo y rentable debido a los grandes avances en miniaturización de la electrónica y comunicaciones inalámbricas. Sin embargo, el diseño óptimo de las técnicas de recogida y de comunicación en estas redes todavía conlleva grandes retos, especialmente en escenarios de gran escala. En concreto, estas técnicas deben ser eficientes, de baja complejidad, y escalables para garantizar una recogida fiable y observaciones de alta fidelidad, para una red con un mínimo mantenimiento. Aquí, el término escalabilidad se refiere a que el rendimiento de la red no se verá afectado por el número de sus sensores o densidad de la red.

Las nuevas teorías probabilísticas (y estadísticas), tales como el Sensado Compresivo (*Compressed Sensing*) y la Geometría Estocástica (*Stochastic Geometry*), favorecen los modelos derivados de estos grandes escenarios. La presente Tesis adoptará ambas teorías con el propósito de cumplir sus objetivos.

Objetivos

El objetivo general de la presente Tesis es el desarrollo de un protocolo para la recogida eficiente de datos en una gran y densa red inalámbrica de sensores, y un modelo teórico, con énfasis en las etapas de comunicaciones y compresión, de forma que permita abordar el análisis probabilístico del protocolo, creando procedimientos específicos para cada etapa pero al mismo tiempo, siendo aplicables a otros distintos protocolos para este tipo de redes.

En esta Tesis se han abordado principalmente dos áreas específicas de creciente interés teórico y práctico: primero, las Señales Comprimibles (*Sparse Signals*); y segundo, las Distribuciones de Colas Pesadas (*Heavy Tailed Distributions*). Es interesante notar que cada una de estas áreas presenta como denominador común el interés que se tiene en entender el comportamiento de los elementos de menor amplitud, concretamente, en cuantificar la velocidad de su decaimiento. Estos elementos son, en las señales comprimibles, aquellos que pueden considerarse despreciables y que permitirán un mayor nivel de compresión de la señal. En las distribuciones con colas pesadas, estos elementos se encuentran en las colas de las distribuciones, y son los que están asociados a valores remotos que la variable aleatoria en cuestión podría tomar durante una realización.

Las señales comprimibles van a modelar el contenido de información de los datos recogidos por la red inalámbrica de sensores. Mientras que las distribuciones de colas pesadas van a modelar, o aproximar, la distribución que sigue la potencia de interferencia resultante de la interacción entre sensores. Así, ambas áreas van a modelar componentes esenciales en el protocolo propuesto, de acuerdo con el objetivo general. Sin embargo, la naturaleza tan diversa de dichas áreas hace que cada una represente por sí misma un objetivo específico de la presente Tesis.

El primer objetivo específico consiste en profundizar en la definición matemática de Compresibilidad (*Sparsity*) y en la evaluación y diseño de funciones que midan esta propiedad en los datos, son las llamadas Funciones de Compresibilidad (*Sparsity Functions*). Estas funciones tradicionalmente provienen del análisis de Entropía (*Entropy*) o de teorías de análisis económico, y se construyen a partir de normas o pseudo-normas de vectores. En general, estas funciones miden la desviación de una señal con respecto a dos extremos. En el entorno de las señales comprimibles, en un extremo se encuentran aquellas señales con una única componente de información, la que caracteriza por sí sola y de forma completa a una señal, y que por ello consiguen un nivel máximo de compresibilidad; y en el otro extremo se encuentran aquellas señales con todas sus componentes conteniendo una cantidad igual de información relevante para su caracterización, lo que imposibilita conseguir compresión alguna.

Así, un apoyo sustancial en la búsqueda de soluciones comprimibles en muchos problemas de sensado comprimido o de Aprendizaje Máquina (*Machine Learning*), entre otros muchos problemas actuales, ha de venir necesariamente del análisis de compresibilidad y de las funciones utilizadas para medir dicha propiedad. Sin embargo, existen muy pocos precedentes de axiomas que permitan caracterizar la compresibilidad, y existe una gran necesidad de disponer de funciones que sirvan de base para inducir esta propiedad, sobretodo en problemas actuales de optimización, con mayor impacto en grandes dimensiones. Este primer objetivo pretende entonces explicar, de forma formal o mediante una axiomatización, aquellas leyes que deben de respetar las funciones de compresibilidad.

En esta Tesis, además del análisis del concepto de compresibilidad, se analiza la entropía, que es otro concepto fundamental del procesamiento de señales y la teoría de la información. En el caso de la entropía, las

funciones a diseñarse deberán medir la desviación de una distribución de probabilidad con respecto a los siguientes extremos. Por un lado, las distribuciones de probabilidad que concentran toda su masa de probabilidad en una única posible salida, es decir las que constituyen el conjunto de (variables aleatorias) constantes; mientras que en el otro extremo se tiene a la distribución que distribuye toda su masa de probabilidad de forma equivalente sobre todas sus posibles salidas, es decir la distribución de una variable aleatoria uniforme.

El segundo objetivo específico consiste en la adaptación y aplicación de métodos de aproximación de funciones de distribución de colas pesadas. El propósito de estos métodos es la evaluación de la potencia de interferencia agregada en las redes inalámbricas de sensores. Este objetivo deberá permitir incrementar sustancialmente la precisión de aproximaciones actuales de estas distribuciones y específicamente, apoyar el análisis de las métricas de rendimiento de la red.

Este es otro ámbito de naturaleza muy diversa al anterior, pero de indudable interés práctico. El análisis de trazas de interferencia de redes aleatorias, tales como las redes inalámbricas de sensores o las redes celulares, revela que estas trazas siguen distribuciones de colas pesadas. Datos reales recogidos, o aquellos generados a partir de modelos probabilísticos, corroboran esta hipótesis. Sin embargo, las técnicas de aproximación de distribuciones de cola pesada, procedentes de muchos sistemas físicos reales, han tenido, hasta la fecha, un alcance limitado en este ámbito. Esta situación se debe principalmente a la dificultad del manejo de estas distribuciones. Se deberán analizar tanto individual como conjuntamente métodos procedentes de teorías estadísticas clásicas, y utilizar datos extraídos de modelos probabilísticos validados en la literatura científica, concretamente provenientes de modelos de red construidos a partir de conceptos de geometría estocástica. Este objetivo pretende desarrollar aproximaciones para caracterizar, de forma estadística y asintótica, la distribución de colas pesadas de la potencia de interferencia en redes inalámbricas.

El tercer objetivo específico es el más próximo al objetivo general de la Tesis y consiste en el desarrollo de un protocolo que permita la recogida eficiente de datos, a partir de un modelo probabilístico de las interacciones entre sensores, y utilizando los resultados parciales de los objetivos específicos anteriores.

El análisis de grandes cantidades de datos y el desarrollo teórico de

nuevos protocolos de recogida de datos representan hoy, sin duda, un área de investigación muy activa en distintos dominios, y en los últimos años se han propuesto numerosos modelos y protocolos. En concreto, en esta Tesis se analiza el dominio de la frecuencia por ser un espacio en el que muchas señales poseen una representación comprimible. Por consiguiente, esta Tesis contribuye a mejorar el proceso de recogida de datos en aplicaciones reales de muy diversa naturaleza.

Así, las siguientes cuestiones son fundamentales. ¿Cómo desarrollar una caracterización axiomática de compresibilidad? ¿De qué manera se pueden caracterizar las trazas de interferencia cuando su distribución es de colas pesadas? ¿Cuál es la manera natural de implementar el sensado compresivo sobre las redes inalámbricas de sensores? Cada una de estas preguntas de investigación se responderá con un capítulo de la Tesis.

Metodología

La metodología de la Tesis se apoya en dos teorías probabilísticas concebidas para grandes objetos matemáticos. Estas son el sensado compresivo y la geometría estocástica.

El sensado compresivo es una nueva forma de pensar acerca de la recogida masiva de datos. A diferencia de los enfoques tradicionales donde toda la señal es observada, en el sensado compresivo se utiliza un principio aleatorio de observación que opera eficientemente sobre el dominio de la información. Las oportunidades de negocios posibles a partir de la teoría de sensado compresivo están siendo reconocidas recientemente por la industria, lo que se corrobora al encontrar el término "compresivo" en el nombre de muchos nuevos dispositivos y métodos, como es la Resonancia Magnética Compresiva (*Compressive Magnetic Resonance*), entre otros. En este contexto, el concepto de compresibilidad es fundamental para formular adecuadamente los modelos y soluciones. La importancia de este concepto se demostrará en la Tesis.

En las redes inalámbricas de sensores, el sensado compresivo se aplica a la recogida de datos bajo el título amalgamado de Recogida Compresiva de Datos (*Compressive Data Aggregation*). El objetivo de la recogida compresiva de datos es reducir el número de observaciones y aliviar la carga de las comunicaciones. Dada su importancia, su literatura es extensa. Sin embargo, aunque hubo grandes progresos, las investigaciones anteriores han analizado la recogida compresiva de datos en las

redes inalámbricas de sensores desde perspectivas complementarias: por un lado algunas investigaciones asumen comunicaciones perfectas y se centran únicamente en el análisis de compresión; otras por el contrario, analizan las comunicaciones y asumen que los datos poseen perfecta compresibilidad. En consecuencia, no existe un modelo que permita un análisis conjunto de los aspectos de compresión y la comunicación de la recogida compresiva de datos. Esta Tesis pretende llenar este vacío.

Las técnicas de recogida compresiva de datos permiten diseñar protocolos eficientes de recogida de datos que aprenden las relaciones (o correlaciones) existentes entre un conjunto suficientemente representativo de los datos, cada uno de ellos extraídos de vecindades de sensores. En los últimos, años estas técnicas de sensado compresivo han experimentado un avance espectacular, tanto en fundamentos teóricos como en su aplicación a distintas y numerosas tareas de monitorización. Más importante aún, se pronostica que el análisis masivo de datos recogidos funcionará de forma eficiente únicamente sobre dominios compactos de información. En este sentido, la contribución de la Tesis es consistente con otros componentes de las tecnologías de la información y las comunicaciones.

La segunda teoría utilizada en la Tesis es geometría estocástica. Esta teoría estudia grandes estructuras matemáticas de carácter aleatorio. De forma general, estas estructuras consisten en arreglos aleatorios de puntos, los cuales pueden reemplazarse por Conjuntos Cerrados Aleatorios (*Random Closed Sets*), generando arreglos aleatorios de mayor complejidad, como son los procesos de líneas o de planos, entre otros objetos aún más abstractos. Aquí, el Proceso de Poisson de Puntos (*Poisson Point Process*) es el modelo más aceptado para las redes inalámbricas de sensores con despliegue aleatorio. El motivo es que, bajo este modelo, las posiciones de los sensores son independientes y están distribuidos de manera uniforme a lo largo de toda el área de monitorización.

Es importante notar que el proceso de Poisson de puntos posee una gran manejabilidad matemática, lo que permite el análisis probabilístico de protocolos avanzados de Control de Acceso al Medio (*Medium Access Control*). Estos modelos demuestran que, en general, la distribución de la potencia de interferencia en la red será muy asimétrica y lejos de ser Gaussiana. En esencia, la distribución presentará colas pesadas. Por ello, se requieren de nuevos métodos que permitan la estimación precisa de trazas de interferencia para una medición adecuada de su efecto en los protocolos de comunicación.

Resultados

Así, la presente Tesis propone un protocolo para la recogida de datos basado en un principio de sensado aleatorio. Este principio aleatorio permitirá el estudio en simultáneo de los aspectos de compresión y comunicación del protocolo. Básicamente, las interacciones entre sensores se caracterizarán utilizando métodos de geometría estocástica, y el efecto de la potencia de interferencia en la red se evaluarán utilizando expansiones asintóticas y Momentos Logarítmicos (*Log-moments*). Por otro lado, el nivel de compresión alcanzado por el protocolo propuesto se analizará utilizando conceptos de sensado compresivo y de Convexidad Generalizada (*Generalized Convexity*). El protocolo resultante se asemeja a Acceso Múltiple con Escucha de Portadora y Detección de Colisiones (*Carrier Sense Multiple Access with Collision Avoidance*), pero se diferencia en que de forma efectiva implementa una recogida compresiva.

Los tres capítulos principales de la Tesis se resumen a continuación:

En el Capítulo 3 ("Principios Matemáticos de Compresibilidad y Entropía: Axiomas y Funciones Fundamentales"), las propiedades y las funciones de compresibilidad y de entropía son analizadas conjuntamente. Este capítulo propone una caracterización axiomática. Como resultado del análisis axiomático, se deriva una lista con un número mínimo de axiomas para la evaluación de distintas funciones de compresibilidad (y de entropía). Notablemente, esta lista es constructiva y permite el diseño de funciones fundamentales de compresibilidad que ofrecen nuevas formulaciones para el problema de reconstrucción de funciones comprimibles de distribución de probabilidad. Igual de importante, las funciones desarrolladas en este capítulo generalizan funciones clásicas como son, la entropía de Rényi y la entropía de Tsallis, las cuales son generalizaciones de la famosa entropía de Shannon.

En el Capítulo 4 ("Expansiones Asintóticas para Trazas de Interferencia con Distribuciones de Colas Pesadas"), las distribuciones de la potencia de interferencia son analizadas. Estas distribuciones, bajo condiciones generales, exhibirán una cola pesada cuyos momentos no pueden ser definidos a partir de un cierto orden, usualmente un orden muy bajo. Esto restringe la aplicabilidad de los métodos tradicionales de aproximación de distribuciones. Entonces, para llenar ese vacío, este capítulo generaliza la expansión de Edgeworth para funciones de densidad de probabilidad, y la expansión de Cornish-Fisher para cuantiles, ambas basadas en momentos

estadísticos, a expansiones asintóticas análogas basadas en momentos logarítmicos. Debido a los momentos logarítmicos, las expansiones propuestas heredan la capacidad de caracterizar las distribuciones de colas pesadas. Más aún, su implementación es sencilla siguiendo los paralelos con los métodos tradicionales.

En el Capítulo 5 ("Análisis Estocástico de la Recogida Compresiva de Datos"), el cual contiene los resultados principales de la Tesis, se desarrolla un protocolo de recogida eficiente de datos para las redes inalámbricas de sensores basados en sensado compresivo. Se le llama, Recogida Compresiva y Estocástica de Datos (*Stochastic Compressive Data Aggregation*). En el modelo, las posiciones de los sensores siguen un proceso de Poisson de puntos, lo que permitirá el análisis conjunto de los aspectos de compresión y de comunicación del protocolo, utilizando resultados clásicos de sensado compresivo y geometría estocástica, respectivamente. Aquí, las técnicas de geometría estocástica permiten el análisis de la configuración de la red y de los modelos de canal. Luego, el protocolo propuesto se asegurará de la llegada de un número mínimo de paquetes para alcanzar el nivel deseado de fidelidad. Este número de paquetes se describe en términos de la densidad de la red después de aplicar el protocolo. Luego, un cálculo a posteriori de la interferencia agregada en la red comprueba que el efecto de la interferencia es insignificante sobre la densidad efectiva. Aparte, por el lado de la compresión, usando las posiciones de los sensores, se construye una nueva matriz de sensado compresivo. Esta nueva matriz satisface criterios clásicos descritos en la literatura de sensado compresivo.

La interacción entre los capítulos se puede resumir como sigue. Una red inalámbrica de sensores se despliega de forma aleatoria sobre una cierta región geográfica para observar un fenómeno físico cuya representación en el dominio de las frecuencias es comprimible. Esta hipótesis es válida teniendo en cuenta que muchas señales reales son comprimibles, por ejemplo, las fotografías son comúnmente comprimibles en el dominio de las ondículas. Primero, el nivel inicial de compresibilidad del fenómeno físico se aproxima utilizando unas pocas observaciones utilizando las funciones de compresibilidad desarrolladas en el Capítulo 3. Este número de observaciones lo determina el protocolo propuesto en el Capítulo 5 que, teniendo en cuenta la resolución deseada y el nivel de compresibilidad del fenómeno físico, protege el número requerido de paquetes para una apropiada reconstrucción. Este protocolo, proporciona entonces la

densidad efectiva de la red inalámbrica de sensores. A posteriori, las expansiones desarrolladas en el Capítulo 4 se emplean para evaluar la degradación del rendimiento de la red debido a la interferencia, que resulta ser imperceptible. Finalmente, un algoritmo de reconstrucción, adaptado a los requerimientos del problema, utiliza la matriz de sensado compresivo construida con las posiciones de los sensores y las observaciones parciales para reconstruir el fenómeno físico en cuestión.

Conclusiones

Un análisis crítico de los resultados de la Tesis abre la discusión y permite reconocer algunas limitaciones técnicas.

Por el lado de compresión, el dominio de información supuesto es el de frecuencias o de Fourier. Esto significa que el fenómeno físico bajo monitorización debe ser comprimible en el dominio de las frecuencias. Aunque esta es una hipótesis válida para una gran variedad de fenómenos físicos, como la humedad o temperatura, fenómenos con representaciones menos estructuradas, o aquellas de gran variabilidad en el tiempo, requerirán considerar diferentes dominios de información. Sin embargo, aunque el diseño de un protocolo adaptado al dominio del fenómeno físico puede reducir el número de observaciones, su implementación supondrá un mayor número de paquetes de control para primero, determinar el dominio de información, y luego para construir las matrices de sensado compresivo adecuadas.

Por el lado de la comunicación, el modelo de red, o de las posiciones de sus sensores, es el proceso de Poisson de puntos. Esto significa que las posiciones de los sensores desplegados deben ser completamente aleatorias. Nuevamente, aunque esta es una hipótesis válida en las redes inalámbricas de sensores, la naturaleza del fenómeno físico podría exigir diferentes configuraciones de red. Por ejemplo, si se desean capturar mayores detalles de zonas específicas. Los Procesos Cluster de Puntos (*Cluster Point Processes*) serían adecuados en estos escenarios. Sin embargo, estos procesos son mucho menos manejables analíticamente. Adicionalmente, estos procesos generan componentes en frecuencia que no se pueden suprimir dado que poseen información relevante para la reconstrucción. En el modelo aleatorio propuesto, caracterizado por el proceso de Poisson de puntos, estas componentes espurias se pueden tratar como ruido incorrelado lo cual facilita considerablemente la tarea

de reconstrucción.

Es importante mencionar que, aunque el protocolo propuesto se describe en un escenario de un solo salto, los resultados son directamente aplicables a escenarios con múltiples saltos. En este caso, en lugar de enviar los paquetes a un nodo central, los sensores deberán transmitir sus paquetes a otros sensores dentro de su vecindad.

Como última reflexión. Los métodos de monitorización compresiva de datos en las redes inalámbricas de sensores propuesta en esta Tesis permitirán aligerar la carga de trabajo de los sensores de una red inalámbrica de sensores y así conseguir una monitorización prolongada. Con el continuo crecimiento de las tareas de monitorización, y consiguientemente del tamaño de las redes inalámbricas de sensores, y por supuesto el de las tecnologías de la información y las comunicaciones, la necesidad de protocolos y algoritmos que operen eficientemente sobre dominios compactos de información es inminente.

Contents

List of Abbreviations

AIHT	Accelerated Iterative Hard Thresholding
ALOHA	ALOHA Protocol
ApEn	Approximate Entropy
BEP	Bit Error Rate
BPP	Binomial Point Process
CDA	Compressive Data Aggregation
CDF	Cumulative Density Function
CFEC	Cornish-Fisher Expansion using Cumulants
CFEL	Cornish-Fisher Expansion using Log-Cumulants
CSMA	Carrier Sense Multiple Access
DFT	Discrete Fourier Transform
EEC	Edgeworth Expansion using Cumulants
EEL	Edgeworth Expansion using Log-Cumulants
FASTA	Fast Adaptive Shrinkage/Thresholding Algorithm
FdBF	Faá di Bruno Formula
FKS	First Kind Statistics or Classical Statistics
GCFEC	Generalized Cornish-Fisher Expansion using Cumulants
GCFEL	Generalized Cornish-Fisher Expansion by Log-Cumulants
IHT	Iterative Hard Thresholding
IID	Independent and Identically Distributed

IoT	Internet of Things
LSA	Licensed Shared Access
MAC	Medium Access Control
MAP	Medium Access Probability
ME-CDA	Minimum-Energy Compressive Data Aggregation
MIC	Mutual Incoherence Constant
MIP	Mutual Incoherence Property
MoLC	Method of Log-Cumulants
MoLM	Method of Log-Moments
MoM	Method of Moments
NDFT	Non-Uniform Discrete Fourier Transform
NDFT-k	Non-uniform discrete Fourier transform of type k
PCP	Poisson Cluster Process
PDF	Probability Density Function
PGFL	Probability Generating Functional
PMF	Probability Mass Function
PPP	Poisson Point Process
RACS	Random Access Compressed Sensing
RDFT	Random Discrete Fourier Transform
RIC	Restricted Isometry Constant
RIP	Restricted Isometry Property
RSA	Random Sequential Adsorption
RV	Random Variable
S-CDA	Stochastic Compressive Data Aggregation
SampEn	Sample Entropy
SIR	Signal to Interference Ratio

SKS	Second Kind Statistics or Log-statistics
SPP	Strauss Point Process
SSI	(Sequential Spatial) Inhibition Process
TC-CDA	Treelet-based Clustered CDA
TVWS	TV White Spaces
WCS	Wireless Compressive Sensing
WSN	Wireless Sensor Network

List of Symbols

Greek Symbols

γ	Interference threshold
δ_k	Restricted isometry constant of order k
$\delta_\Omega(i)$	Activation indicator of the ith sensor
ϵ	Sensing noise
η	Inverse mean of fading
Θ	Point process (model) of the network
Θ_{on}	Point process of the active network
Θ_Ω	Point process of the transmitting network
$\kappa_{X,k}$	kth cumulant of the distribution of RV X
$\tilde{\kappa}_{X,k}$	kth log-cumulant of the distribution of RV X
λ	Intensity of the point process Θ
λ_{on}	Intensity of the point process Θ_{on}
λ_1	Transmission rate of RACS
λ_Ω	Intensity of the point process Θ_Ω
μ	Mutual incoherence constant
μ_X	Mean of the distribution of RV X
$\tilde{\mu}_X$	Log-mean of the distribution of RV X
ν	Carrier sensing threshold
ξ	Fading and shadowing effects
$\pi(.)$	Permutation operator
π_b	Parameter of Bernoulli RV
Π	Partition set
ρ	Rate of activation
Σ	Transformation to information domain
$\Sigma_b(.)$	Transformation to b-bins-histogram domain
σ_X	Standard deviation of the distribution of RV X

$\tilde{\sigma}_X$	Standard log-deviation of the distribution of RV X
$\varphi_X(.)$	1st characteristic function of the distribution of RV X
$\tilde{\varphi}_X(.)$	1st characteristic log-function of the distribution of RV X
Φ	Measurement matrix
$\psi_X(.)$	2nd characteristic function of the distribution of RV X
$\tilde{\psi}_X(.)$	2nd characteristic log-function of the distribution of RV X
Ψ	Sparsifying matrix
Ω	Compressed sensing matrix

Latin Symbols

b	Number of atoms, associated index $j = 1, \ldots, b$
$\mathcal{B}$	Deployment region
$\mathcal{B}_k(.)$	kth Bell polynomial
D_x^k	kth derivative with respect to the variable x
$E_c(.)$	Exponential integral function
$\mathbb{E}_X$	Expectation with respect to the distribution of RV X
$\mathbb{E}^!$	Expectation with respect to the Palm probability $\mathbb{P}^!$
$f_X(.)$	Probability density function of the RV X
$\mathbf{f}_a$	ath set of coordinates of a grid array, $a = 1, 2$
$f_{\mathcal{N}}(.)$	Density of the standard normal distribution
$F_X(.)$	Cumulative density function of RV X
$\mathcal{F}$	Fourier transform
$G_\Theta(.)$	PGFL of the point process Θ
$\mathcal{G}_\Theta(.)$	Conditional PGFL of the point process Θ
$\mathsf{h}^b(.)$	Entropy function given b atoms
$\mathcal{H}_k(.)$	kth Hermite polynomial
$I_{\mathbf{x}}$	Interference received at the sensor $\mathbf{x}$
k	Sparsity of coefficients $\mathbf{w}$
k_Ω	Estimated sparsity of the sample coefficients $\mathbf{w}_\Omega$
$\mathcal{K}_k$	Partition set
$\ell(.)$	Path-loss attenuation function
$\ell_p(.)$	p-pseudo-norm of vectors, $0 \leq p < 1$
$\ell_p(.)$	p-norm of vectors, $1 \leq p$
$\mathcal{L}(.)$	Logarithmic transformation
$\mathcal{L}^{-1}\mathcal{N}$	Set of nearly-log-normal distributions
$m_{X,k}$	kth moment of the distribution of RV X
$\tilde{m}_{X,k}$	kth log-moment of the distribution of RV X

m	Number of samples
$\mathcal{M}$	Mellin transform
n	Signal dimension, associated index $i = 1, \ldots, n$
$\mathcal{N}$	Set of nearly-normal distributions
$\mathcal{N}_i$	Neighborhood of the ith sensor
p	1st parameter of core function
p_{map}	Medium access probability
p_{suc}	Successful transmission probability
$\mathbf{p} = (p_j)$	Probability mass function, $p_j = \mathbb{P}(V = a_j)$
$\tilde{\mathbf{p}}$	Vector $\mathbf{p}$ after an action
$\mathbb{P}(.)$	Probability function
$\mathbb{P}^!(.)$	Palm probability
q	2nd parameter of core function
r_i	Distance from origin to the ith sensor
$\Re(.)$	Real part of complex number
$\mathsf{s}(.)$	Sparsity function
$\mathcal{S}(.)$	Standardization operation
$\mathbb{S}_+^b$	Probability simplex
t_i	Time mark of the ith sensor
$\mathbf{t}_a$	ath set of coordinates of a random array, $a = 1, 2$
T	Observation window
T_{coh}	Coherence time
u_q	qth quantile of the RV U
U	Reference RV
v_i	Measurement of the ith sensor
$\mathbf{v} = (v_i)$	Signal representation of physical variable, with elements v_i
$\mathbf{v}_\Omega$	Samples of the vector $\mathbf{v}$
$\mathbf{v}^\#$	Recovered signal
V	RV defined on $\mathcal{V}$
$\mathcal{V} = \{a_j\}$	Countable sample space with atoms (events) a_j
$\mathbf{w} = (w_i)$	Coefficients of the vector $\mathbf{v}$, with components w_i
$\mathbf{w}_\Omega$	Coefficients of sample signal $\mathbf{v}_\Omega$
$\tilde{\mathbf{w}}$	Coefficients $\mathbf{w}$ after an action
$\mathbf{w}^\#$	Recovered coefficients
x_q	qth quantile of the RV X
$\mathbf{x}_i$	ith sensor (node)
X	Arbitrary RV
$\mathbf{y}$	Observations or sample vector

1. Introduction

1.1 Background and Research Environment

Modern information technologies aim to measure *everything*, *everywhere*, and *every time*. This massive data acquisition principle provides new possibilities for the mathematical understanding of the nature and behavior of physical systems. Adequate models of these systems make it possible to describe them in an intellectual way, by computing characteristic statistics that touch on the essentials.

Consequently, these modern information technologies have created three major fields of research. While the first of these fields, the Internet of Things (IoT) [1], offers a communication infrastructure, the second, Big Data [2], offers a processing infrastructure. The third field, which offers a sensing capability, is the Wireless Sensor Network (WSN) [3]. In the context of this thesis, it is more appropriately described as a large and dense WSN.

A WSN is a collection of electronic sensor devices, denoted as nodes in mathematical models, which possess sensing, low-power computing, and low-range communication capabilities. Hence, there is still a need for separate communication and processing infrastructures. These devices are deployed to observe and inform about physical variables, which are essential for the implementation of precision agriculture [4] and environmental monitoring [5] applications, among many other usage scenarios.

Although progress in electronics miniaturization and wireless communication allows for cost-effective and easy-to-deploy networks, the realization of WSNs offers intrinsic challenges to the design of acquisition, processing, communication, and networking techniques, especially in large scale scenarios. Specifically, these techniques must be power

efficient, low in complexity, and scalable to ensure reliable and high-fidelity data collection from a network with minimum maintenance [6]. Scalability means that system performance is not affected by the size (number or density) of the network. New probabilistic theories, such as compressed sensing [7] [8], favor these large networks.

Compressed sensing is a new way to thinking about the acquisition of massive data. Instead of the common approach of collecting all the data from a large array of sensors, compressed sensing randomly and efficiently samples the data in its information domain. In this theory, a knowledge of sparsity holds great importance. These opportunities are only just now being recognized by industry, evidenced by the term "compressive" being added to many new devices and methods, such as compressive magnetic resonance imaging [9] and compressive sensing radar [10].

Research on compressed sensing is very active and could be organized in three areas: (1) design and analysis of optimization algorithms for sparse recovery [11] [12] [13]; (2) compressed sensing matrices that preserve the information content [14] [15] [16]; and (3) less generalized, functions that promote sparsity [17] [18] [19]. This thesis focuses on the latter two of these areas.

In WSNs, compressed sensing is applied to data aggregation [20] under the amalgamated title of Compressive Data Aggregation (CDA) [21]. The goal in CDA is to reduce the number of measurements and alleviate the communication load, especially near central sensors that gather the data. Given its prevalence, the literature on CDA has become vast. However, while some research assumes perfect communications and focuses on the compression analysis which proposes practical compressed sensing matrices [22] [23], other research assumes a given compressed sensing matrix and focuses on the communication analysis under imperfect channels [24] [25]. Hence, as there is no model that allows for a joint analysis of the compression and communication aspects of CDA, this thesis seeks to bridge this gap.

1.2 Objectives and Scope

The scope of this thesis is WSN-based monitoring scenarios, and the main objective is to develop efficient acquisition protocol with probabilistic guarantees. The research process of the thesis is based on two probabilis-

Table 1.1. Topics, key references and contributions of the dissertation.

Topic	**References**	**Contribution**
Essentials of Compressed Sensing	[7], [8]	
Algorithms for Sparse Recovery	[11], [12], [13]	
Compressed Sensing Matrices	[14], [15], [16]	Chapter 5
Sparsity Functions	[17], [18], [19]	Chapter 3
Compressive Data Aggregation	[21], [22], [23]	
Stochastic CDA	*Research gap*	Chapter 5
Stochastic Data Aggregation	[20], [24], [25]	
Aggregate Interference	[28], [29], [30]	Chapter 4
Medium Access Control	[31], [32], [33]	Chapter 5
Network Modeling	[34], [35], [36]	
Essentials of Stochastic Geometry	[26], [27]	

tic theories developed for large mathematical objects, compressed sensing and stochastic geometry [26] [27]. To outline the structure of this thesis, Table 1.1 illustrates the specific areas of research as related to this thesis along with a few key references.

Stochastic geometry studies large random structures. In communications, the applied research which is supported on this theory could be organized into three key areas: (1) sensors locations or network modeling via point processes [34] [35] [36]; (2) analysis of medium access control (MAC) protocols [31] [32] [33]; and more generalized, (3) analysis of aggregate network interference [28] [29] [30].

The Poisson Point process (PPP) [37] is the most accepted model for WSNs with random deployment. Under this model, sensors are independent and uniformly distributed throughout the sensing field. The PPP possesses high mathematical tractability which allows for the probabilistic analysis and tuning of advanced MAC protocols. However, sensor interactions produced by these protocols, in conjunction with the geography of the sensing field, organize the sensors such that their new positions follow less tractable point processes. Moreover, under mild conditions [29], the resulting aggregate interference will be strongly skewed and far from normal [38]. Essentially, the new interference pattern will follow a heavy-tailed model. An accurate estimation of interference traces enables the measurement of the interference effect in the given protocols.

Thus, the following research questions are of critical concern:

- *How is it possible to develop an axiomatic characterization of sparsity?*;
- *In what way(s) can heavy-tailed interference traces be characterized?*; and
- *What is the best way to implement the random sensing of compressed sensing in WSNs?*

Each research question is answered with a chapter as described in the next structure.

1.3 Dissertation Structure

The remainder of this thesis is organized into five chapters. Each chapter gives a short introduction to the subject and highlights the contributions in the subject area. An overview of related work precedes the contributions and results. Following this structure, Chapter 2 introduces the system model and basic analytical tools, Chapters 3 and 4 develop the tools for the analysis of the CDA protocol as described in Chapter 5, and Chapter 6 concludes and presents future directions for research:

Ch2 Chapter 2 describes briefly the theories and tools adopted in the thesis. It also introduces the network, channel, and signal models.

Ch 3 The properties of sparsity and entropy are jointly analyzed. Sparsity and entropy are principal theoretical concepts in signal processing and information theory. Previous efforts have contributed to individually understanding sparsity and entropy for specific research perspectives. Accordingly, Chapter 3 proposes a joint axiomatic characterization. This chapter gathers and introduces inherent and first principle criteria such as axioms and attributes that jointly characterize these principal concepts. The proposed set of axioms is constructive and allows for the derivation of *core functions* which offer new formulations to the recovery of sparse probability measures. The sparsity functions developed in this chapter will be utilized to estimate (and track) the level of sparsity in the data aggregation protocol proposed in Chapter 5.

Ch 4 Under mild conditions, the distributions of network interference traces exhibit a heavy tail, are strongly skewed, and far from

normal. Their characterization have gained increasing attention, especially with the emergence of different types of wireless networks. However, the cumulants of heavy-tailed distributions cannot be commonly defined above a certain order. This restricts the applicability of the classical approximation methods. To bridge this gap, Chapter 4 extends the usability of the Edgeworth and Cornish-Fisher expansions, which are based on cumulants, to analogous asymptotic expansions based on logarithmic cumulants or log-cumulants. The proposed expansions inherit the capability of log-cumulants to characterize heavy-tailed distributions [39] and parallel classical expansions; thus, they are readily implemented. Subsequent numerical simulations show the accuracy of the proposed expansions to approximate the tails of interference under different network scenarios. The expansions developed in this chapter will be used to measure the effect of interference in the data aggregation protocol proposed in Chapter 5.

Ch 5 Chapter 5 contains the main results of the thesis. This chapter introduces a data aggregation protocol for WSNs based on compressed sensing. This protocol is called Stochastic Compressive Data Aggregation (S-CDA). In the model, the locations of sensors follow a PPP which allows for the joint analysis of the compression and communication aspects of S-CDA using classic results from compressed sensing and stochastic geometry. On the communication side, a Carrier Sense Multiple Access (CSMA) type protocol is designed and analyzed such that the required number of sensors (in terms of network density) are active and transmit their measurements. The resulting compressed sensing matrix is then discussed using the Restricted Isometry Property (RIP) [7] and Mutual Incoherence Property (MIP) [15] criteria.

Ch 6 Chapter 6 concludes the thesis and discusses further paths of research.

The main results, theorems, and lemmas will be proved in the Appendix Section.

The interaction between chapters can be summarized as follows. In a monitoring scenario, a WSN is randomly deployed to observe a physical phenomenon which is sparsely represented in the frequency domain. This assumption is justified as real signals are commonly sparse under suitable

domains, e.g. most images are sparse in the domain of wavelets. The initial level of sparsity of the sensing field is first computed using the measured data from all the sensors. After this learning step, the level of sparsity is estimated using core sparsity, developed in Chapter 3, by employing only a few measurements or samples. The number of samples here is determined according to the MAC protocol. Given the desired resolution and the current sparsity level, a CSMA-type protocol is tuned and implemented to protect the required number of random packets. This protocol, developed in Chapter 5, provides the initial density of samples. Then, the expansions developed in Chapter 4 are employed to evaluate the network performance degradation due to interference. This evaluation could refine the parameters of the protocol which in turn provides the corrected density. The new density defines the number of samples required for the desired level of fidelity. Finally, a sparse recovery algorithm uses the compressed sensing matrix (which is constructed with the locations of the WSN, the acquired samples, and the estimated level of sparsity) to recover the signal. The complete process is described in Figure 1.1.

1.4 Dissertation Contribution

The main contributions of each chapter are:

1. Chapter 3 introduces a refined and constructive set of axioms for sparsity and entropy which allows for the derivation of new functions;

2. Chapter 4 derives two new asymptotic expansions, and describes a methodology to extend other approximation methods for skewed and heavy-tailed data traces; and

3. Chapter 5 develops a framework for the joint analysis of the compression and communication aspects of data aggregation in randomly deployed WSNs.

Part of the research results have been documented in six peer-reviewed conference papers (references [40], [41], [42], [43], [44] and [45])[1]; the notion of linking sparsity and entropy axioms was originally described

[1]Also, a peer-reviewed journal paper (reference [46]) has been recently accepted for publication.

in [40] in a shorter form. This paper also proposes low-complexity entropy functions which were implemented and evaluated in a real scenario in [42]. In addition, the Edgeworth expansion using log-cumulants was originally published in [43] for the approximation of common interference models. Other interference models, generated by random geometric models, were estimated in [44] using log-cumulants. The resulting model and communication analysis of Chapter 5 includes some aspects which were originally presented in [45] and partly in [41], respectively, but they are both elaborated further here.

However, significant aspects of the thesis, namely, the refinement of axioms (Section 3.2) and formal derivation of functions (Section 3.4) in Chapter 3; approximation of quantiles (Section 4.3.2) in Chapter 4; and compression and communication analysis (Sections 5.2.2 and 5.2.3) in Chapter 5 have not yet been published. This thesis gathers all these obtained research results into the form of a monograph.

Paper [42] was written by Mihaela I. Chidean. The author contributed to Section III, which describes the properties and implementation of low-complexity entropy functions as well as revising and commenting on the rest of the paper. Professor Riku Jäntti motivated the study of Cornish-Fisher expansions using log-cumulants (Section 4.3.2). The author was the main author of the other aforementioned publications ([40], [41], [43], [44] and [45]) and was responsible for the planning and oversight of all experimental work, analysis of the test results, numerical modeling, and development of the theory presented in the thesis. The supervisors of the thesis and any other individuals contributed to the work only through discussions and by providing comments on the related conference papers and on the manuscript of the thesis. The supervisors stimulated the original research ideas and contributed to the research with valuable comments and suggestions.

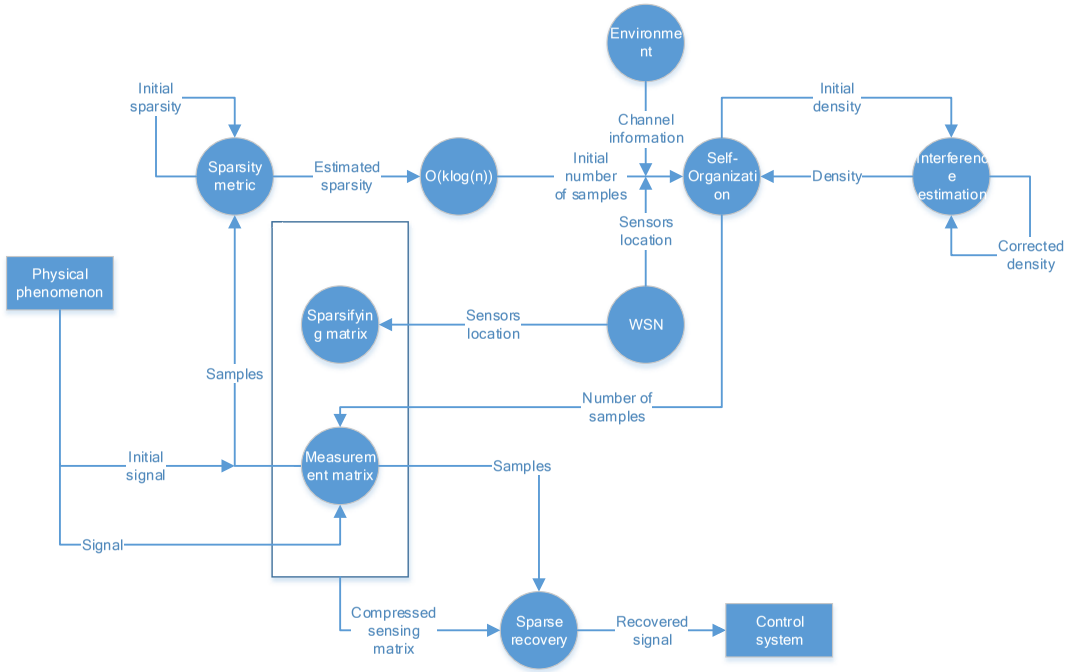


Figure 1.1. S-CDA protocol. A WSN is randomly deployed to observe a phenomenon which is sparse in the frequency domain. The sparsity of the sensing field is computed by employing only a few samples. Given the desired resolution and the current sparsity level, a CSMA-type protocol is tuned and implemented to protect the required number of random packets. Then, asymptotic expansions are employed to evaluate the network performance degradation due to interference. Finally, a sparse recovery algorithm uses the compressed sensing matrix to recover the signal.

2. Models and Tools from Compressed Sensing and Stochastic Geometry

This section describes the system model and key concepts from compressed sensing and stochastic geometry.

2.1 Compressed Sensing

2.1.1 Sparse Recovery

In Compressed Sensing [7], a physical variable $\mathbf{v} \in \mathbb{R}^n$ of sparse coefficients $\mathbf{w} \in \mathbb{C}^n$, with $\mathbf{v} = \Psi\mathbf{w}$, is recovered from samples $\mathbf{y} \in \mathbb{R}^m$, with $\mathbf{y} = \Phi\mathbf{v}$ and $m \ll n$. The matrix Φ models the measurement process, e.g. it is the routing matrix in the data aggregation scenario. On the other hand, the matrix Ψ is typically the Fourier transform and coefficients $\mathbf{w}$ represent the frequency content of $\mathbf{v}$.[1] The theory is based on three assumptions:

1. $\mathbf{v}$ is sparse or compressible in a known domain Ψ (CS1);
2. Φ and Ψ are incoherent (CS2); and
3. m is in the order of $k \log n$ (CS3).

The incoherence means that rows of Φ do not possess sparse representations in terms of columns of Ψ; k represents the sparsity level of $\mathbf{w}$. Under the assumptions, *perfect recovery* is reached with high probability. This probability is of the form $1 - \mathcal{O}(n^{-m})$ ($\mathcal{O}$ denotes order), which states that compressed sensing is a theory for large signals [7].

The recovery is based on searching the coefficients $\mathbf{w}^{\#}$ for the sparsest representation or *minimum complexity*, subject to (s.t.) the measurement

[1] The data aggregation scenario defines the domain of definitions, e.g. the signals are real, and the coefficients are complex.

Φ and sparsifying Ψ processes,

$$\mathbf{w}^{\#} = \arg \begin{cases} \min_{\mathbf{u}\in\mathbb{C}^n} & \|\mathbf{u}\|_1, \\ \text{s.t.} & \Omega\mathbf{u} = \mathbf{y} \end{cases} \tag{2.1}$$
$$\mathbf{v}^{\#} = \Psi\mathbf{w}^{\#},$$

where $\|.\|_1$ is the ℓ_1-norm; $\Omega = \Phi\Psi$ is the compressed sensing matrix; and Φ and Ψ are the measurement and sparsifying matrices, respectively.

In signal processing, sparsity functions describe the efficiency of the basis representation. The ℓ_0-pseudo-norm is known as the canonical sparsity count. It is also known as strict or hard sparsity since it counts the cardinality of the support of a function, e.g. the number of non-zero elements of a vector. In the following, this sparsity count will be denoted as "ℓ_0".[2] Its lack of a useful derivative prompts combinatorial optimization when used as the objective functional in maximizing-sparsity problems.

For (2.1) and related problems, the adoption of the ℓ_1-norm as a *proxy function* enables derivation of efficient algorithms and performance guarantees [8]. "*ℓ_1 is a convex surrogate for ℓ_0 count. It is the best surrogate in the sense that the ℓ_1 ball is the smallest convex body containing all 1-sparse objects of the form $\pm\mathbf{e}_n$*" [47] ($\pm\mathbf{e}_n$ is an element of the canonical basis). This function will be denoted "ℓ_1". "ℓ_1" is not however a sparsity function [17].

Definition 2.1.1 *[Restricted isometry property (RIP)] The Restricted Isometry Constant (RIC) [7] of the order k, denoted as δ_k, is defined as [7]*

$$\delta_k = \min_{\delta}\{\delta : 1-\delta \leq \frac{\|\Omega\mathbf{u}\|_2^2}{\|\mathbf{u}\|_2^2} \leq 1+\delta, \forall k\text{-}\textit{sparse}\ \mathbf{u}\}, \tag{2.2}$$

where $\|.\|_2$ is the ℓ_2-norm.

Essentially, a small RIC indicates that the joint (measurement and sparsifying) process Ω performs as an isometry by preserving the energy of the signal. The RIC is present in many results of compressed sensing. However, its computational complexity is challenging [48], and new trends [49] and recent evidence [50] suggest that the RIP criterion is not used in many practical scenarios. Hence, the exact calculation of the RIC is outside the scope of this work.

[2]The quotation marks "." help to preserve and tries to alert about the misleading notation which treats ℓ_0-pseudo- (ℓ_1-) norm as a sparsity function. In fact, ℓ_0-pseudo- (ℓ_1-) norm measures non-sparsity. This misleading notation led [17] to erroneously treat ℓ_0 (no quotation marks) as a sparsity function.

Definition 2.1.2 *[Mutual incoherence property (MIP)] Let Ω_i be the i-th unit normalized column vector of Ω. As such, the Mutual Incoherence Constant (MIC) is defined as [15]*

$$\mu = \max_{i \neq j} |\langle \Omega_i, \Omega_j \rangle|, \tag{2.3}$$

which is bounded as [16]

$$\sqrt{\frac{n-m}{(n-1)m}} \leq \mu < \frac{1}{2k-1}. \tag{2.4}$$

The MIP is a computationally verifiable condition that ensures sparse recovery.

2.1.2 Signal Model

Let $\mathbf{v} = (v_i)$ be a smooth sensing field, such as temperature, pressure, or humidity maps, whose representation in the frequency domain $\mathbf{w} = (w_i)$, consists of few frequency components f_i,

$$\mathbf{v} = \Re\left(\sum_{i=1}^{k} e^{\sqrt{-1} f_i}\right) \quad \leftrightarrow \quad \mathbf{w} = \sum_{i=1}^{k} A_i f_i, \tag{2.5}$$

where $\Re$ denotes the real part, and $A_i \in \mathbb{R}$ denotes the amplitude of frequency component f_i. Many physical signals have sparse representations in suitable domains, e.g. most images are compressible in the wavelets domain. Furthermore, it is assumed that the sensing field follows a process with coherence time, T_{coh}, such that $\mathbf{v}(t_1) \approx \mathbf{v}(t_2)$ for $|t_1 - t_2| \leq T_{\mathsf{coh}}$. This assumption defines an observation window of the size $T \leq T_{\mathsf{coh}}$, which allows for the dropping of time index from the sensor measurements v_i.

2.2 Stochastic Geometry

2.2.1 Network Modeling

Stochastic geometry is another theory of large objects. In it, one object is the point process Θ which is analyzed using the Probability Generating Functional (PGFL) and its conditional form [51].

Definition 2.2.1 *[PGFL of point process Θ]*

$$G_\Theta w = \mathbb{E}\left[\prod_{\mathbf{x}_i \in \Theta} w(\mathbf{x}_i)\right], \tag{2.6}$$

where the function, $w : \mathbf{x}_i \in \Theta \mapsto [0, 1]$, s.t. $1 - w$ is integrable.

Definition 2.2.2 *[Conditional PGFL of point process Θ]*

$$\mathcal{G}_{\Theta} w = \mathbb{E}^{!}\left[\prod_{\mathbf{x}\in\Theta} w(\mathbf{x})\right], \tag{2.7}$$

where the expectation $\mathbb{E}^{!}$ is with respect to the reduced Palm measure $\mathbb{P}^{!}$.

A tractable point process is the PPP [37]. In a PPP, the sensors $\mathbf{x}_i$ are independent and uniformly distributed in a region, $\mathcal{B}$. For a density of sensors λ, the number of sensors n_1 and n_2, falling within the disjointed regions $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively, are independent and identically distributed (IID) RVs,

$$n_a = \Theta(\mathcal{B}_a) \sim \text{Poisson}(\lambda|\mathcal{B}_a|), \quad a = 1, 2, \tag{2.8}$$

where the function $|.|$ measures the area. Definitions 2.2.1 and 2.2.2 coincide for the PPP [51],

$$G_{\Theta} w = \mathcal{G}_{\Theta} w = \exp(-\lambda \int_{\mathbb{R}^2} (1 - w(\mathbf{u}))d\mathbf{u}). \tag{2.9}$$

A point process closely related to the PPP is the Binomial Point Process (BPP) [34]. In a BPP, a finite number of sensors $\mathbf{x}_i$, namely n, are independent and uniformly distributed in region $\mathcal{B}$. As the number of sensors and the size of the region grow, so that the density λ remains constant, the BPP converges to a PPP.[3] Hence, a BPP can be studied from a PPP conditioned on its number of sensors.

The complexity of modern wireless communication networks prohibits an analysis by the simulation of such complex systems. Instead, a stochastic approach has attracted a considerable amount of attention in recent years since it leads to general results and, in certain cases, to an unambiguous insight into the behavior of networks. Under this approach, the target network is modeled as a random object, where the positions of sensors, denoted as $\mathbf{x}_i$, follow a spatial point process $\Theta = \{\mathbf{x}_i\}$ [52].

2.2.2 Network and Channel Models

Let a WSN which consists of n sensors randomly deployed in a bounded region $\mathcal{B}$ be such that the density of sensors is λ. A widely accepted probabilistic model for this WSN is [51]

$$\text{PPP } \Theta = \{\mathbf{x}_i\} \text{ with intensity } \lambda, \tag{2.10}$$

[3] As in the traditional case for binomial and Poisson RVs.

(conditioned on $i = 1, \ldots, n$), where $\mathbf{x}_i$ denotes the tag and location of the i^{th} sensor. The sink or fusion center of the WSN, denoted as $\mathbf{x}_0$, is the typical (receiver) sensor which is commonly located at the origin, i.e. at position $\mathbf{x}_0 = \mathbf{0}$. For computational purposes (cf. 2.9), the sink could be either included or not in Θ as the underlying process is the PPP.

Let the amplitude variation of a received radio signal be the product of the distance-dependent power attenuation function ℓ (bounded and non-increasing), shadowing or log-normal variation in the local mean power level (large-scale fading), and (small-scale) fading due to movement in the order of a wavelength. The use of this model is common practice in radio communications. The shadowing and fading effects (and transmission power) of the model are gathered in the random variable (RV) ξ. Then, the signal power from $\mathbf{x}_i$ received at the typical sensor is

$$\xi_i \ell(\|\mathbf{x}_i - \mathbf{x}_0\|_2) = \xi_i \ell(r_i), \text{ with } r_i = \|\mathbf{x}_i\|_2. \tag{2.11}$$

Several models exist for path loss including the variants of the Okumura-Hata and Walfisch-Ikegami formulas, to simple power laws where the decay exponent is typically found in the range of 3 to 5 depending on the environment. Likewise, numerous statistical models have been proposed for fading, most notably the Rayleigh, Rice, and Nagakami-m probability laws. Moreover, shadowing has been empirically observed to obey approximately a log-normal distribution in a wide variety of propagation environments [53].

In interference analysis, the target network Θ is the effective set of active sensors (transmitters or interferers). The interference, aggregated with respect to this target network, is measured at a typical (receiver) sensor, denoted as $\mathbf{x}$, which is commonly located at the center of the network. Then, the interference at $\mathbf{x}$ can be described as

$$I_{\mathbf{x}} = \sum_{\mathbf{x}_i \in \Theta \setminus \{\mathbf{x}\}} \xi_i \ell(\|\mathbf{x} - \mathbf{x}_i\|_2). \tag{2.12}$$

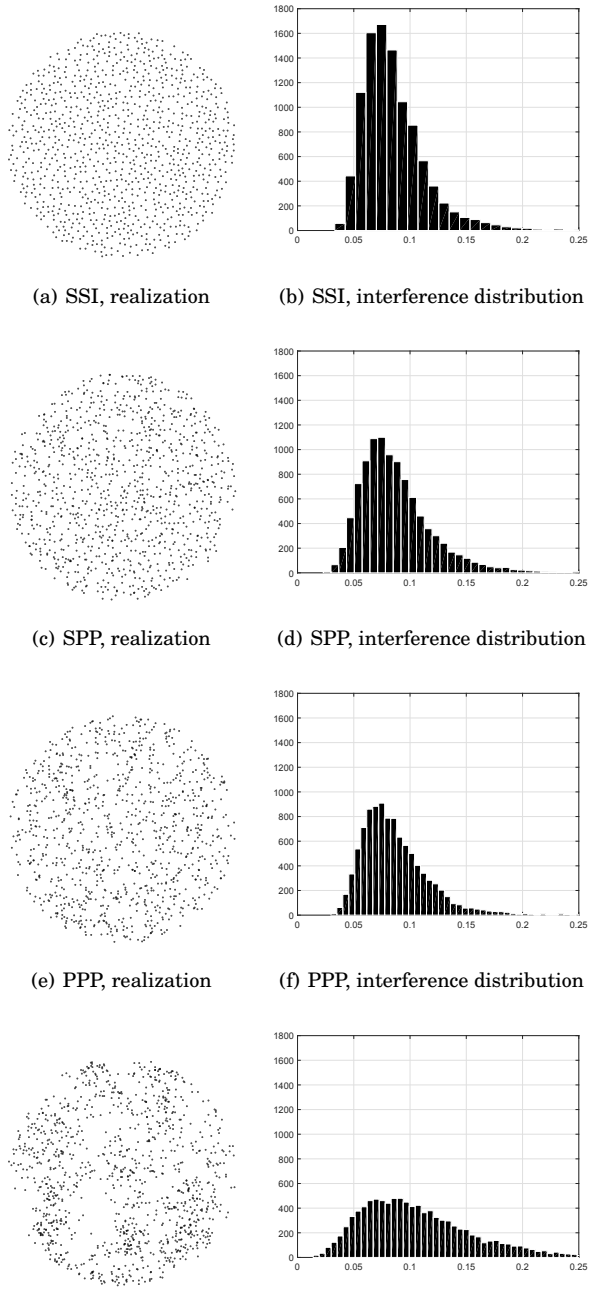
As well, the spatial arrangement of the sensors clearly affects the distribution of interference. On this point, several models have been recently proposed for capturing the spatial randomness of the topology of the interference source. The most widely used model is the PPP, mainly due to its analytical tractability. However, several physical and practical constraints make the placement of the sensors in real-world networks non-uniform and irregular, leading to repulsion (or attraction) between sensors, resulting in more regular (or more clustered) point patterns

than for the PPP. For instance, MAC protocols or the arrangement of the sensors by themselves may induce clustering. Hence, more accurate models than the PPP for actual deployments are the Poisson Cluster Process (PCP), (Sequential Spatial) Inhibition Process (SSI) and Strauss Process (SPP). Nevertheless, the limited analytical tractability of the aforementioned point processes makes it difficult or even impossible to obtain closed-form analytical expressions for network performance [38].

Essentially, the triplet network regularity[4] and density, path loss, and fading-shadowing conditions should determine the tail behavior of the interference. The network density is connected to the exclusion radius around the typical receiver and to the number of interferers. A finite number of (dominant) interferers around this receiver breaks the normality assumption as the central limit theorem does not apply. The unbounded behavior of path loss around the origin (controlled by network density) can lead to heavy tails as well. Moreover, under bounded path loss, the fading and shadowing effects dominate the tail behavior of the interference. However, as common fading models are light-tailed, shadowing will control the heaviness of the tails [54].

Figure 2.1 plots the networks in order of their decreasing regularity (see Section 4.4 for simulation details). The grid configuration is the more regular network (not included in the simulations as it is a very artificial deployment). It is followed by the SSI which includes a mechanism that stipulates a minimum distance d between (two) sensors, which fixes an exclusion radius and reduces the effect of near interferers. The SPP is derived in a similar fashion as the SSI, with a specified fraction of sensors allowed within a distance d of any given sensor. The PPP can be considered in a neutral position in the ordering since it is distributed uniformly at random. Finally, the less regular network is the PCP, whose clusters create a highly varying local density.

[4]Regularity refers to local density invariance.



(a) SSI, realization (b) SSI, interference distribution

(c) SPP, realization (d) SPP, interference distribution

(e) PPP, realization (f) PPP, interference distribution

(g) PCP, realization (h) PCP, interference distribution

Figure 2.1. Aggregate interference in random networks: Poisson point process (PPP), Poisson Cluster process (PCP), (Sequential Spatial) Inhibition process (SSI), and Strauss point process (SPP). Figures 2.1(a), 2.1(c), 2.1(e) and 2.1(g) show network realizations, Figures 2.1(b), 2.1(d), 2.1(f) and 2.1(h) show the sample density of the interference. The channel model considers a distance-dependent (r) attenuation function $\ell(r) = (1 + r^{\alpha})^{-1}$, with $\alpha = 4$, and a log-normal shadowing effect, with $\mu = 1$ and $\sigma = 1.25$. Each network realization is drawn such that the expected number of interferers is approximately 1000.

3. Mathematics of Sparsity and Entropy: Axioms and Core Functions

3.1 Introduction

For a given vector $\mathbf{v}$, the uncertainty of its elements defines the compressibility of its coefficients $\mathbf{w}$ in a given domain. This dependence, between the (elements') uncertainty and (coefficients') compressibility of a vector, suggests a connection between the two families of functions that measure these properties. However, what axioms do these families have in common? This chapter adopts the following definitions with the aim of responding to this question. Let the elements of the vector $\mathbf{v}$ follow a distribution with the probability mass function (PMF) $\mathbf{p}$.

Definition 3.1.1 *[Compressibility or sparsity] The property of concentrating most of the energy into a few components.*

Definition 3.1.2 *[Uncertainty or entropy] The property of not concentrating most of the probability mass into a few atoms.*

Sparsity and entropy functions quantify properties 3.1.1 and 3.1.2, respectively. One approach to evaluate the goodness of these functions is through the validation of criteria (see [17] [55] for sparsity, [56] [57] for entropy). Following this approach, this chapter gathers and introduces inherent and first principle criteria as the axioms and attributes that jointly characterize the sparsity and entropy functions. As expected, it follows that criteria of both families of functions are strictly complementary.

In [58], a set of conditions are identified under which distributions can be ordered by variance and entropy can be used to in a similar way. This suggests that entropy could be extracted from the "shape" of distributions. Intuitively, for a given variance, random processes with similarly sorted PMFs should present similar levels of entropy. Hence,

for a given vector with elements following a univariate distribution, the uncertainty of the underlying process equals the non-compressibility of its PMF. The following toy examples illustrate this relationship. Let RV $V \in \{a_1, a_2\}$, with PMF $\mathbf{p} = (p_1, p_2)$, be such that $p_j = \mathbb{P}(V = a_j)$, $j = 1, 2$.

1. Assume $p_1 > p_2$: if p_1 increases (thus p_2 decreases since $p_1, p_2 \geq 0$ and $\|\mathbf{p}\|_1 = 1$), then a_1 is even more certain to appear, i.e. if the compressibility of $\mathbf{p}$ increases, then the uncertainty of V decreases,

$$\mathsf{sparsity}(\mathbf{p}) \uparrow \equiv \mathsf{entropy}(V) \downarrow . \tag{3.1}$$

2. Assume $\mathbf{p} = (p_1, p_2) = (1, 0)$ (constant RV): if p_2 increases (thus p_1 decreases), then a_1 is not the unique possible outcome, i.e. if the compressibility of $\mathbf{p}$ decreases, then the uncertainty of V increases,

$$\mathsf{sparsity}(\mathbf{p}) \downarrow \equiv \mathsf{entropy}(V) \uparrow . \tag{3.2}$$

The sparsity functions derived in this chapter can help to estimate and track the level of sparsity of large signals such as the information collected from a sensing field.

Entropy functions (related to uncertainty, information, fuzziness, or complexity measures) and sparsity functions (related to compressibility, fairness, or dispersion measures) are not regarded as easily characterizable. Hence, these functions have been thoroughly studied in numerous articles from different perspectives and research areas. The following presents the relevant references which contribute explicitly to the set of axioms and attributes of the sparsity and entropy functions of Section 3.2.

The main method for the axiomatic characterization of the entropy consist of treating inherent or satisfactory properties as axioms for the Shannon's entropy function [59]. This started in 1948 with [59], which established the continuity and monotonicity of entropy functions; [60] added concavity for the fuzziness functions and [61] relaxed it to quasi-concavity for the dispersion functions; [62] added concentration for the information functions; [56] added maximality for the entropy functions; and [57] added symmetry for the information functions.

On the other hand, although sparse data and sparsity have attracted considerable attention in recent years, concentration, scaling, homogeneous growth, and replication were originally applied in 1920 by [63] in a financial setting to measure the inequity of wealth distribution; [61] added bounds for the dispersion functions; [17] [55] added symmetry and

continuity for the sparsity and fairness functions, respectively; and [64] added quasi-convexity for the reward-risk ratio functions.

For sparsity and entropy, previous works have separately established axioms and relationships, and derived generalizations. Thus, the contribution of the chapter is threefold.

1. The main contribution of this chapter is a *refined and constructive set of axioms (and attributes) of the sparsity and entropy functions* (Section 3.2), which allows for the derivation of function generalizations.

2. The derived sparsity functions generalize the Hoyer measure, Gini index and pq-means and *offer novel formulations to the compressed sensing problem* (Section 3.4).

3. The derived entropy functions generalize the Rényi and Tsallis entropies, and *offer simple formulations to the recovery of sparse probability measures* (Section 3.4.2).

3.2 Axioms

This section gathers and introduces inherent and first principle criteria as axioms that jointly characterize the sparsity and entropy functions. From these axioms, attributes will be derived in the next section.

Let V be a discrete and real-valued RV, defined on a countable sample space $\mathcal{V} = \{a_1, \dots, a_b\} \subseteq \mathbb{R}$, and let $\mathbf{v} = (v_i) \in \mathcal{V}^n$ be a vector containing n realizations of RV V. The definition of the vector $\mathbf{v}$ allows two possible and exchangeable interpretations. It is an arbitrary signal as a whole or a random process formed by its elements. Throughout this chapter, both interpretations are adopted, and different operations are applied to the vector $\mathbf{v}$ according to the analysis performed. The analysis of sparsity of the signal will be from the compressibility perspective. Analogously, the analysis of entropy of the process will be from the uncertainty perspective. The analogies between these perspectives, that will help to derive axioms and attributes, are summarized in Table 3.1. The following operations will provide adequate representations for the analysis.

For the compressibility assessment of the signal $\mathbf{v}$ a transformation to the information domain, e.g. frequency, is applied using basis $\Sigma \in \mathbb{R}^{n \times n}$,

$$\Sigma : \mathbb{R}^n \to \mathbb{R}^n_+; \mathbf{v} \mapsto \mathbf{w} = \Sigma \mathbf{v}, \tag{3.3}$$

Table 3.1. Analogies between sparsity functions and entropy functions.

Description	**Sparsity** s	**Entropy** h
Property	compressibility	uncertainty
Argument (element)	coefficients (component)	PMF (atom)
Argument's unit	energy	(probability) mass
High value if few...	... dominant components	... unlikely atoms
Low value if few...	... negligible components	... likely atoms

where $\mathbf{w} = (w_i)$ is the vector of the coefficients of $\mathbf{v}$. Note that $\Sigma = \Psi^{-1}$, cf. equation (2.1). For ease of notation, since compressibility is measured from magnitude, the energy values of the components w_i are assumed to be non-negative, i.e. $w_i \in \mathbb{R}_+$. Essentially, redefine the components (in terms of true components) as $w_i = |w_i|$. Under this model, a sparsity function, which measures the compactness of the signal $\mathbf{v}$ under the basis Σ is

$$s : \mathbb{R}_+^n \to \mathbb{R}; \mathbf{w} = \Sigma\mathbf{v} \mapsto \mathsf{s}(\mathbf{w}). \tag{3.4}$$

For the uncertainty calculation of process $\mathbf{v}$ a b-bins-based ℓ_1-normalized histogram method is applied using operator Σ_b,

$$\Sigma_b : \mathbb{R}^n \to [0,1]^b; \mathbf{v} \mapsto \mathbf{p} = \Sigma_b(\mathbf{v}), \tag{3.5}$$

where $\mathbf{p} = (p_j)$ is the sample PMF of $\mathbf{v}$, with $\|\mathbf{p}\|_1 = 1$ and

$$p_j = \mathbb{P}(V = a_j) \approx \frac{\text{cardinality}(\{i : v_i = a_j\})}{n}. \tag{3.6}$$

Similarly, define $\mathbb{S}_+^b = \{\mathbf{p} \in [0,1]^b : \|\mathbf{p}\|_1 = 1\}$ (probability simplex). Then, for a given variance σ^2, an entropy function which measures the randomness of the process $\mathbf{v}$ from the occurrences of its outcomes using the method Σ_b, is

$$\mathsf{h}^b : \mathbb{S}_+^b \to \mathbb{R}; \mathbf{p} = \Sigma_b(\mathbf{v}) \mapsto \mathsf{h}^b(\mathbf{p}). \tag{3.7}$$

Under this notation, the following enumerates a collection of the axioms and attributes of the sparsity and entropy functions. The axioms and attributes describe the effect of different actions on the argument of functions, e.g. changes to its pattern, ratios, or relative differences (see Table 3.2 for a mathematical description). In the case of entropy analysis, and in order to stay inside the probability simplex, the actions related

to regularity and homogeneous growth require an inner normalization $\|\tilde{\mathbf{p}}\|_1 = 1$, where $\tilde{\mathbf{p}}$ denotes $\mathbf{p}$ after an action, such as the normalization. The histogram model described above makes this normalization step transparent.

The following five are the axioms:

3.2.1 Continuity

S. (*Sparsity*) Slight energy perturbation retains the ratio of dominant and negligible components. [55]

H. (*Entropy*) Slight probability mass perturbation retains the ratio of unlikely and likely atoms. [59]

3.2.2 Symmetry

S. Energy permutation retains the ratio of dominant and negligible components. [61]

H. Probability mass permutation retains the ratio of unlikely and likely atoms. [57]

3.2.3 Concentration (Dalton's 1st Law)

S. Moving energy from negligible to dominant components shortens the set of dominant components. [63]

H. Moving probability mass from unlikely to likely atoms shortens the set of likely atoms. [62]

3.2.4 Scaling (Dalton's Modified 2nd Law)

S. Scaled components contain proportional amounts of energy in a proportional number of components. [63]

H. Scaled occurrences contain proportional amounts of probability mass in a proportional number of atoms (and by the $\|\mathbf{p}\|_1 = 1$ normalization).

3.2.5 Replication (Dalton's 4th Law)

S. Concatenating replicas of (all) components retains the ratio of dominant and negligible components. [63]

H. Adding realizations following the same law retains the ratio of unlikely and likely atoms. This axiom states that entropy is intrinsic to a given process (see Subsection 3.4.3 for details).

3.3 Attributes

3.3.1 Bounds

S. The least compressible coefficients distribute equally their energy between all components, e.g. a white noise. The most compressible coefficients distribute their energy to one single component, e.g. a single "tone". [61]

H. The most uncertain distribution distributes equally its probability mass between all atoms, e.g. a uniform RV. The least uncertain distribution distributes its probability mass to one single atom, e.g. a constant RV. [59]

Lemma 3.3.1 *Axioms 3.2.3 and 3.2.4 imply attribute 3.3.1.*

3.3.2 Quasi-convexity

S. Quasi-convexity prefers extremes to averages. [64]

H. Quasi-concavity encourages diversification. [60]

Lemma 3.3.2 *Axioms 3.2.1, 3.2.2, 3.2.3 and 3.2.4 imply attribute 3.3.2.*

3.3.3 Monotonicity

This is a simple version of axiom 3.2.3.

S. For two signals with a non-zero pair of components, the signal with the largest ratio of highest-energy by lowest-energy is more compressible. [55]

H. For two processes with a non-impossible pair of atoms, the process with the largest ratio of highest-mass by lowest-mass is less uncertain.

Lemma 3.3.3 *Axiom 3.2.3 implies attribute 3.3.3.*

3.3.4 Completeness

S. Adding zero-energy components concentrates the energy into relatively less components. [17]

H. Considering impossible (zero-mass) atoms concentrates the mass into relatively less atoms.

Lemma 3.3.4 *[17] Axioms 3.2.3, 3.2.4, and 3.2.5 imply attribute 3.3.4.*

3.3.5 Regularity

S. A significant concentration of energy in one single component makes the rest negligible. [17]

H. A significant concentration of probability mass in one single atom makes the rest unlikely.

Lemma 3.3.5 *[17] Axioms 3.2.3 and 3.2.4 imply attribute 3.3.5.*

3.3.6 Homogeneous Growth (Dalton's 3rd Law)

S. Increasing relatively the energy of negligible components makes them relatively more dominant. [63]

H. Increasing relatively the probability mass of unlikely atoms makes them relatively more likely.

Lemma 3.3.6 *Axiom 3.2.4 and attributes 3.3.1 and 3.3.2 imply attribute 3.3.6.*

3.3.7 Schur-Convexity

S. Schur-convexity holds.

H. Schur-concavity holds.

Lemma 3.3.7 *Axiom 3.2.2 and attribute 3.3.2 imply attribute 3.3.7.*

Table 3.2. Axioms 3.2.1-3.2.5 and attributes 3.3.1-3.3.8 of sparsity and entropy functions, with $\alpha, \beta > 0$. Operator π denotes permutation.

Name	**Sparsity**	**Entropy**
Continuity 3.2.1	$\tilde{\mathbf{w}} \to \mathbf{w}, \mathsf{s}(\tilde{\mathbf{w}}) \to \mathsf{s}(\mathbf{w})$	$\tilde{\mathbf{p}} \to \mathbf{p}, \mathsf{h}^b(\tilde{\mathbf{p}}) \to \mathsf{h}^b(\mathbf{p})$
Symmetry 3.2.2	$\tilde{\mathbf{w}} = \pi(\mathbf{w}), \mathsf{s}(\tilde{\mathbf{w}}) = \mathsf{s}(\mathbf{w})$	$\tilde{\mathbf{p}} = \pi(\mathbf{p}), \mathsf{h}^b(\tilde{\mathbf{p}}) = \mathsf{h}^b(\mathbf{p})$
Concentration 3.2.3	$\begin{cases} \tilde{\mathbf{w}} = (\tilde{w}_i), \tilde{w}_{i_1} = w_{i_1} + \alpha, \tilde{w}_{i_2} = w_{i_2} - \alpha, w_{i_1} \geq w_{i_2} \\ \tilde{w}_i = w_i, i \neq \{i_1, i_2\}, \mathsf{s}(\tilde{\mathbf{w}}) > \mathsf{s}(\mathbf{w}) \end{cases}$	$\begin{cases} \tilde{\mathbf{p}} = (\tilde{p}_j), \tilde{p}_{j_1} = p_{j_1} + \alpha, \tilde{p}_{j_2} = p_{j_2} - \alpha, p_{j_1} \geq p_{j_2} \\ \tilde{p}_j = p_j, j \neq \{j_1, j_2\}, \mathsf{h}^b(\tilde{\mathbf{p}}) < \mathsf{h}^b(\mathbf{p}) \end{cases}$
Scaling 3.2.4	$\tilde{\mathbf{w}} = \alpha\mathbf{w}, \mathsf{s}(\tilde{\mathbf{w}}) = \mathsf{s}(\mathbf{w})$	$\tilde{\mathbf{p}} = \alpha\mathbf{p}, \mathsf{h}^b(\tilde{\mathbf{p}}) = \mathsf{h}^b(\mathbf{p})$
Replication 3.2.5	$\tilde{\mathbf{w}} = \mathbf{w} \| \ldots \| \mathbf{w}, \mathsf{s}(\tilde{\mathbf{w}}) = \mathsf{s}(\mathbf{w})$	$\tilde{\mathbf{p}} = \mathbf{p} \| \ldots \| \mathbf{p}, \mathsf{h}^b(\tilde{\mathbf{p}}) = \mathsf{h}^b(\mathbf{p})$
Bounds 3.3.1	$\tilde{\mathbf{w}} = \alpha\mathbf{1}_n, \hat{\mathbf{w}} = \mathbf{0}_{n-1} \| \alpha, \mathsf{s}(\tilde{\mathbf{w}}) \leq \mathsf{s}(\hat{\mathbf{w}})$	$\tilde{\mathbf{p}} = \alpha\mathbf{1}_b, \hat{\mathbf{p}} = \mathbf{0}_{b-1} \| \alpha, \mathsf{h}^b(\tilde{\mathbf{p}}) \geq \mathsf{h}^b(\hat{\mathbf{p}})$
Quasi-Convexity 3.3.2	$\alpha < 1, \mathsf{s}(\alpha\mathbf{w} + (1-\alpha)\tilde{\mathbf{w}}) < \max\{\mathsf{s}(\mathbf{w}), \mathsf{s}(\tilde{\mathbf{w}})\}$	$\alpha < 1, \mathsf{h}^b(\alpha\mathbf{p} + (1-\alpha)\tilde{\mathbf{p}}) > \min\{\mathsf{h}^b(\mathbf{p}), \mathsf{h}^b(\tilde{\mathbf{p}})\}$
Monotonicity 3.3.3	$\frac{\alpha}{2} \leq w < \tilde{w} \leq \alpha, \mathsf{s}(\tilde{w}, \alpha - \tilde{w}) > \mathsf{s}(w, \alpha - w)$	$\frac{\alpha}{2} \leq p < \tilde{p} \leq \alpha, \mathsf{h}^b(\tilde{p}, \alpha - \tilde{p}) > \mathsf{h}^b(p, \alpha - p)$
Completeness 3.3.4	$\tilde{\mathbf{w}} = \mathbf{w} \| 0, \mathsf{s}(\tilde{\mathbf{w}}) > \mathsf{s}(\mathbf{w})$	$\tilde{\mathbf{p}} = \mathbf{p} \| 0, \mathsf{h}^b(\tilde{\mathbf{p}}) < \mathsf{h}^b(\mathbf{p})$
Regularity 3.3.5	$\begin{cases} \exists\beta, \tilde{\mathbf{w}} = (\tilde{w}_i), \hat{\mathbf{w}} = (\hat{w}_i), \tilde{w}_i = \hat{w}_i = w_i, i \neq i_1 \\ \tilde{w}_{i_1} = w_{i_1} + \beta, \hat{w}_{i_1} = w_{i_1} + \beta + \alpha, \mathsf{s}(\tilde{\mathbf{w}}) < \mathsf{s}(\hat{\mathbf{w}}) \end{cases}$	$\begin{cases} \exists\beta, \tilde{\mathbf{p}} = (\tilde{p}_j), \hat{\mathbf{p}} = (\hat{p}_j), \tilde{p}_j = \hat{p}_j = p_j, j \neq j_1 \\ \tilde{p}_{j_1} = p_{j_1} + \beta, \hat{p}_{j_1} = p_{j_1} + \beta + \alpha, \mathsf{h}^b(\tilde{\mathbf{p}}) > \mathsf{h}^b(\hat{\mathbf{p}}) \end{cases}$
Homogeneous Growth 3.3.6	$\tilde{\mathbf{w}} = \mathbf{w} + \alpha\mathbf{1}_n, \mathsf{s}(\tilde{\mathbf{w}}) < \mathsf{s}(\mathbf{w})$	$\tilde{\mathbf{p}} = \mathbf{p} + \alpha\mathbf{1}_b, \mathsf{h}^b(\tilde{\mathbf{p}}) > \mathsf{h}^b(\mathbf{p})$
Schur-Convexity 3.3.7	$\mathbf{w} \succ \tilde{\mathbf{w}}, \mathsf{s}(\mathbf{w}) \geq \mathsf{s}(\tilde{\mathbf{w}})$	$\mathbf{p} \succ \tilde{\mathbf{p}}, \mathsf{h}^b(\mathbf{p}) \leq \mathsf{h}^b(\tilde{\mathbf{p}})$
Triangle Inequality 3.3.8	$\mathsf{s}(.) \geq 0, \mathsf{s}(\mathbf{w} + \tilde{\mathbf{w}}) \leq \mathsf{s}(\mathbf{w}) + \mathsf{s}(\tilde{\mathbf{w}})$	$\mathsf{h}^b(.) \leq 0, \mathsf{h}^b(\mathbf{p} + \tilde{\mathbf{p}}) \geq \mathsf{h}^b(\mathbf{p}) + \mathsf{h}^b(\tilde{\mathbf{p}})$

Table 3.3. Sparsity functions.

Function	Description	s_{pq} family
"ℓ_0" or sparsity count	$-\frac{1}{n}\|\mathbf{w}\|_0$	not a member
"ℓ_1" (proxy)	$-\frac{1}{n}\|\mathbf{w}\|_1$	not a sparsity func.
Kurtosis	$n\left(\frac{\|\mathbf{w}\|_4}{\|\mathbf{w}\|_2}\right)^4$	not a member
Gini index	$\frac{2}{1+n}\frac{\|\mathbf{w}\|_{1^*}}{\|\mathbf{w}\|_1}-1$	$p=1^*, q=1$
Hoyer measure	$\frac{1}{\sqrt{n}-1}\left(\sqrt{n}-\frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2}\right)$	$p=1, q=2$
pq-means	$-n^{\frac{1}{q}-\frac{1}{p}}\frac{\|\mathbf{w}\|_p}{\|\mathbf{w}\|_q}$	$p\leq 1, q>1$
$\mathsf{s}_{1\infty}$ or $\max$-sparsity	$-\frac{1}{n}\frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_\infty}$	$p=1, q=\infty$

3.3.8 Triangle Inequality

S. If $s>0$, the triangle inequality holds.

H. If $h<0$, the reverse triangle inequality holds.

Lemma 3.3.8 *If* $\mathsf{s}(.)>0$*, axiom 3.2.4 and attribute 3.3.2 imply attribute 3.3.8.*

3.4 Core Functions

3.4.1 Core Sparsity

Apart from the "ℓ_0" and "ℓ_1", other sparsity functions proposed previously for (2.1) will be discussed here. Among these functions, the kurtosis [65], Gini index [66], Hoyer measure [67], and pq-means $\{p\leq 1, q>1\}$ [17] (denoted pq-means in the following) satisfy most of the six-rules criteria of [17], where concentration, scaling, homogeneous growth, and replication are their axioms, and regularity and completeness are their attributes. [40] extends these criteria and derives $\max$-sparsity. Table 3.3 defines all these functions.

Further inspection of sparsity functions (Table 3.3) reveals a common functional form in their construction. The metrics of inequality of wealth or fairness, i.e. non-sparsity, exhibit the same functional form, e.g. the Theil [68] and Atkinson [69] indexes. This simple observation enables the derivation of a generalization.

Theorem 3.4.1 *[Core sparsity] The function*

$$\mathsf{s}_{pq}(\mathbf{w}) = -n^{\frac{1}{q}-\frac{1}{p}} \frac{\|\mathbf{w}\|_p}{\|\mathbf{w}\|_q}, \quad \{p \leq 1, q \geq 1, p \neq q\}, \tag{3.8}$$

satisfies axioms 3.2.1-3.2.5.

Core sparsity is continuous (except at the origin), symmetric, scale-invariant, and semi-strictly quasi-convex.

The kurtosis is out of the range for p and q in (3.8). This function failed the test using the criteria of [17]; hence, it will not be further analyzed. Nevertheless, the Gini index (weighted mean) and Hoyer measure are positive affine transformations of core sparsity. Positive affine transformations are increasing functions; hence, they preserve the quasi-convexity attribute 3.3.2 by Theorem A.1.2 [70].

Table 3.4 shows the results of the test using the axioms and attributes of Section 3.2. "ℓ_0" measures the strict sparsity; hence, it fails the concentration and regularity, which describe smooth variations. "ℓ_1" clearly fails most of the axioms and thus attributes. Core sparsity will satisfy all axioms, as it is designed to. Hoyer and Gini, which are closely related to core sparsity, also satisfy all axioms except replication, which is satisfied only asymptotically.

More importantly, *core sparsity extends pq-means*. [71] presents pq-means with $\{p \leq 1, q = 2\}$. *"This, (ℓ_p) normalization by the ℓ_2-norm may turn out to be the best sparseness criterion. This, however, has yet to be further investigated.* [71]". And although [17] grants the origin (citation sense) of pq-means to [71], (anonymous) reviewers suggested these functions to [17] as stated in its acknowledgments. Thus, pq-means has neither a clear origin, nor a formal derivation in the literature. Thus, the Appendix derives core sparsity (which extends pq-means) from axioms 3.2.1-3.2.5.

A key factor of core sparsity is its resemblance of "ℓ_0". This resemblance is shown in Figure 3.1. Figures 3.1(d)-3.1(f) illustrate the smoothness, symmetry, and quasi-convexity of the sparsity functions; and Figures 3.1(g)-3.1(l) indicate that the unit balls of their respective functional forms and of "ℓ_0" are the same. A difference between the functions in Figure 3.1 is their behavior at the origin. "ℓ_0" attains its maximum at the origin, which makes sense from a practical perspective since no information is required to reconstruct null signals. However, core sparsity is not defined at the origin due to continuity, e.g. in Figure 3.1,

$$0 = \lim_{\epsilon \to 0} \mathsf{s}_{pq}((\epsilon, \epsilon)) \neq \lim_{\epsilon \to 0} \mathsf{s}_{pq}((\epsilon, 0)) = 1. \tag{3.9}$$

Table 3.4. Test results of sparsity functions. If the function satisfies (asymptotically) the axiom or attribute a "✓" ("→") is shown, otherwise "✗".

	"ℓ_0"	"ℓ_1"	Hoyer	Gini	s_{pq}
Continuity 3.2.1	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12
Symmetry 3.2.2	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12
Concentration 3.2.3	✗ Proof A.1.13	✗ Proof A.1.22	✓ Proof A.1.32	✓ Proof A.1.41	✓ Proof A.1.58
Scaling 3.2.4	✓ Proof A.1.14	✗ Proof A.1.23	✓ Proof A.1.33	✓ Proof A.1.42	✓ Proof A.1.59
Replication 3.2.5	✓ Proof A.1.15	✓ Proof A.1.24	→ Proof A.1.34	→ Proof A.1.43	✓ Proof A.1.60
Bounds 3.3.1	✓ Proof A.1.16	✗ Proof A.1.25	✓ Proof A.1.35	✓ Proof A.1.44	✓ Proof A.1.61
Quasi-convexity 3.3.2	✓ Proof A.1.17	✓ Proof A.1.26	✓ Proof A.1.36	✓ Proof A.1.45	✓ Proof A.1.62
Monotonicity 3.3.3	✓ Proof A.1.18	✗ Proof A.1.27	✓ Proof A.1.37	✓ Proof A.1.46	✓ Proof A.1.63
Completeness 3.3.4	✓ Proof A.1.19	✓ Proof A.1.28	✓ Proof A.1.38	✓ Proof A.1.47	✓ Proof A.1.64
Regularity 3.3.5	✗ Proof A.1.20	✗ Proof A.1.29	✓ Proof A.1.39	✓ Proof A.1.48	✓ Proof A.1.65
Homogeneous Growth 3.3.6	✓ Proof A.1.21	✓ Proof A.1.30	✓ Proof A.1.40	✓ Proof A.1.49	✓ Proof A.1.66
Schur-Convexity 3.3.7	✓ Lemma 3.3.7	✓ Lemma 3.3.7	✓ Lemma 3.3.7	✓ Lemma 3.3.7	✓ Lemma 3.3.7
Triangle Inequality 3.3.8	✓ Lemma 3.3.8	✗ Proof A.1.31	✓ Lemma 3.3.8	✓ Lemma 3.3.8	✓ Lemma 3.3.8

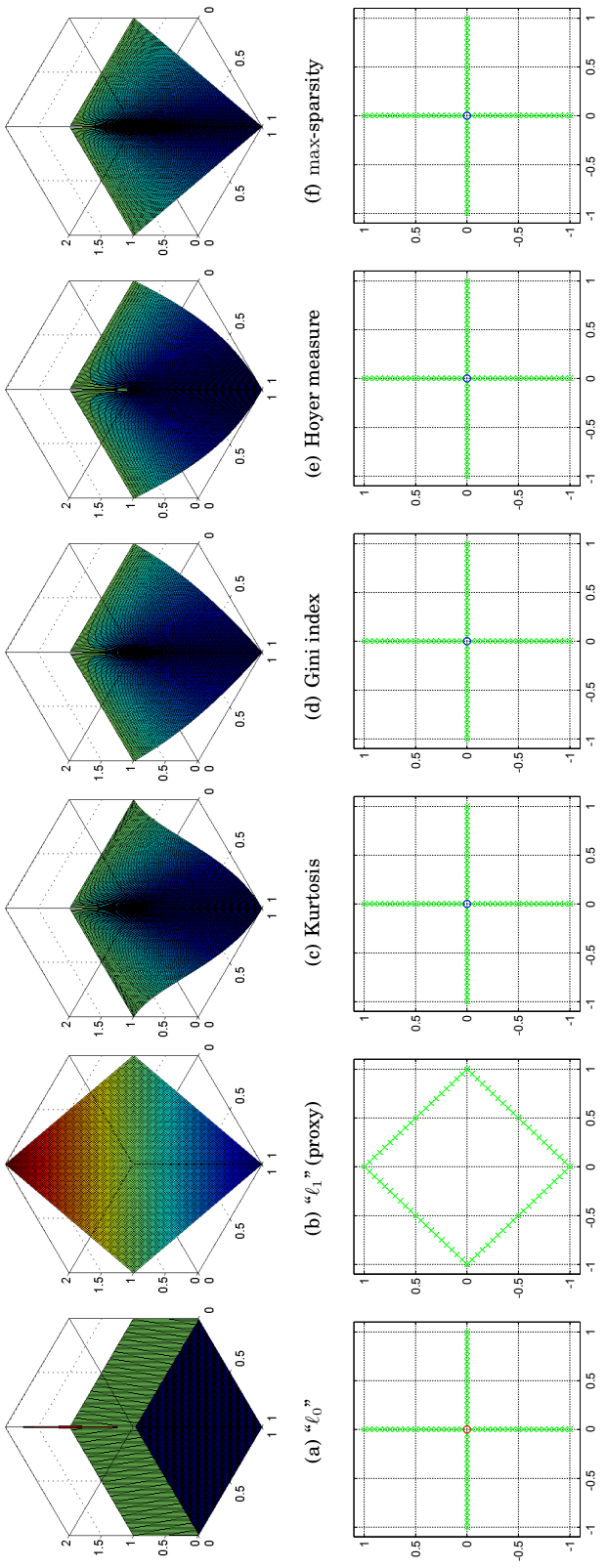


Figure 3.1. Sparsity functions. Figures 3.1(a)-3.1(f): [0,2]-normalized sparsity of $\mathbf{w} \in [0,1]^2$ (*x-axis:* w_1; *y-axis:* w_2; *z-axis:* $s(\mathbf{w})$). Figures 3.1(g)-3.1(l): Unit ball of functional forms in $\mathbf{w} \in [-1,1]^2$ (*x-axis:* w_1; *y-axis:* w_2; *z-axis:* $s(\mathbf{w}) = 1$).

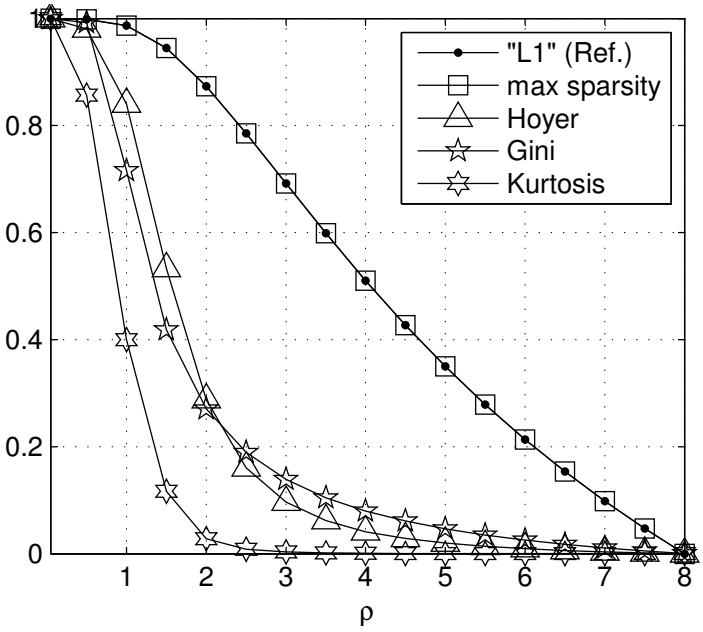
Figure 3.2(a) shows the versatility of $\max$-sparsity (and of normalized "ℓ_1") to assess the sparsity of arbitrarily-sparse coefficients and the localized domain of the kurtosis, Gini index, and Hoyer measure, which are better distinguishing highly sparse coefficients. Notably, the p and q parameters localize the domain of the core sparsity. To which, Figure 3.2(b) presents this localization process for $\mathbf{w}$, following a power law decay, i.e. $\mathbf{w} = (w_i)$, with $w_i = ri^{-\frac{1}{p}}$. This would suggest the (slightly) superior performance, in terms of number of measurements m (the dimension of $\mathbf{y}$), of recovery strategies based on these last functions compared with the common ℓ_1-minimization, as reported in [18] for ℓ_p, with $p \in (0, 1)$; in [72] for the Gini index, whose stochastic algorithm tends to be unstable; in [73] for kurtosis, although the function selects an inappropriate basis; and in [74] for the Hoyer measure in the framework of image regularization.

Thus, *the appropriate choice of parameters p and q would offer a localized-sparsity formulation* of (2.1), since core sparsity strongly encourages coefficients of a given (or estimated) level of sparsity. Furthermore, the core sparsity is differentiable (except at $w_i = 0$, $\forall i$) which is appropriate for gradient-based optimization, and by the quasi-convexity, its solution to (2.1) is attained at an extreme point of the polytope defined by linear (equality) constraints [75], e.g. $\Omega\mathbf{u} = \mathbf{y}$. Then, a simple exact algorithm can solve (2.1) by walking along the extreme points of the feasible region. Still, the potential of the core sparsity in sparse recovery has not been studied.

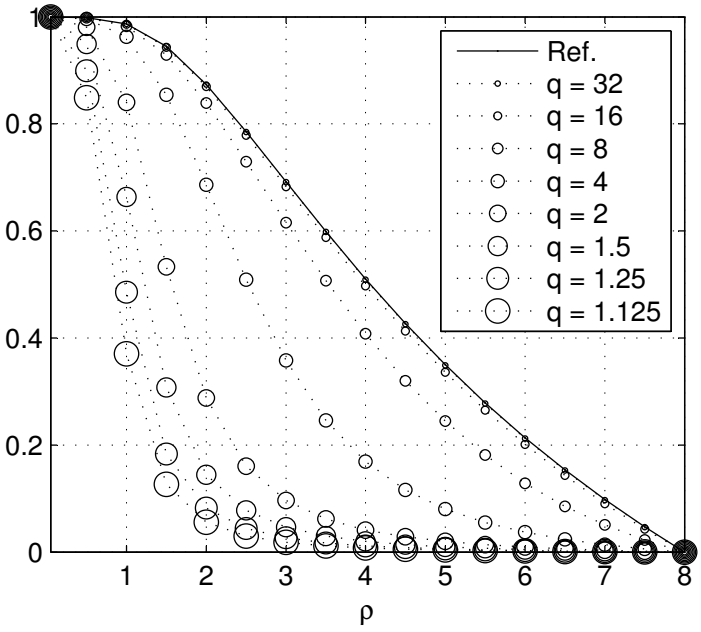
3.4.2 Core Entropy

Entropy is a concept of physics and information theory that characterizes the unpredictability of the state of systems and the events of processes.

In physics, the Tsallis entropy [76] generalizes the Boltzmann-Gibbs entropy. In information theory, the Rényi entropy [77] generalizes the entropy functions via its parameter $p \geq 0$: the Hartley entropy or $\log\max$ entropy [78] ($p = 0$) counts the cardinality of the non-zero probability event; the Shannon entropy ($p \to 1$) uses an expression similar to the Boltzmann-Gibbs entropy to measure the average unpredictability of events; the collision entropy ($p = 2$); the $\log\min$ entropy [79] ($p = \infty$) is the most conservative entropy function since it measures the unpredictability of the most likely event. Hence, the $\log\min$ entropy is never greater than the Shannon entropy (see Figures 3.3 and 3.4).



(a) Sparsity functions (*x-axis:* ρ*; y-axis:* $s(\mathbf{w})$)



(b) Core sparsity s_{1q} (*x-axis:* ρ*; y-axis:* $s(\mathbf{w})$)

Figure 3.2. Sparsity functions. [0,1]-normalized sparsity of coefficients $\mathbf{w} = (w_i)$ of length $n = 1000$, with $w_i = ri^{-\frac{1}{\rho}}$, $r > 1$ and $\rho \in [0.03, 8]$.

Table 3.5. Entropy functions. In the table, $\langle .,. \rangle$ denotes the inner product of vectors.

Function	**Description**	h^b_{pq} **family**
$\log\max$ **entropy**	$\log_2 \|\mathbf{p}\|_0$	$p = 0,\ q = 1$
Shannon entropy	$-\langle \mathbf{p}, \log_2 \mathbf{p} \rangle$	$p = 1,\ q \downarrow 1$
Collision entropy	$-2\log_2 \|\mathbf{p}\|_2$	$p = 1,\ q = 2$
$\log\min$ **entropy**	$-\log_2 \|\mathbf{p}\|_\infty$	$p = 1,\ q = \infty$
Rényi entropy	$\frac{p}{1-p}\log_2 \|\mathbf{p}\|_p$	$p \neq 1,\ q = 1$
Tsallis entropy	$\frac{1}{p-1}(1 - \|\mathbf{p}\|_p^p)$	$p \neq 1,\ q = 1$
$\mathsf{h}^b_{1\infty}$ **or** $\min$ **entropy**	$\frac{1}{\|\mathbf{p}\|_\infty}$	$p = 1,\ q = \infty$

All these functions are defined in Table 3.5, and Table 3.6 indicates the results of the test using the axioms and attributes of Section 3.2. Since the evaluated functions are all derived from the core entropy, they will satisfy all the axioms and thus attributes.

In computer science, entropy characterizes the complexity of sequences. In this context, the entropy functions are called complexity algorithms. Among these algorithms, the Lempel-Ziv complexity [80] characterizes the randomness of a sequence of symbols by measuring its rate of (new) pattern generation. Following a distinct approach, the Approximate Entropy (ApEn) [81] examines finite-length time series for similar epochs, and the Sample Entropy (SampEn) [82] performs similarly to ApEn but without counting self-matches. Both algorithms emerged from the formulation of the Kolmogorov complexity for finite-sample approximations. Although the axioms and attributes of Section 3.2 are not applicable to these algorithms, they will be utilized for comparison.

Although the Rényi entropy and Tsallis entropy are concepts from different fields and are applied to different objects, both functions possess the same functional form that was identified in the sparsity functions (Tables 3.3 and 3.5). The key difference between the Rényi entropy and Tsallis entropy (and the sparsity functions of Section 3.4) is the logarithmic transformation which provides the information units of the Rényi entropy and which, simultaneously preserves the quasi-concavity attribute by Theorem A.1.2 [70]. The same applies to the power function in the Tsallis entropy. Thus, by following Section 3.2 and Section 3.4, and removing all unit transformations, the process allows for the derivation of a generalization.

Table 3.6. Test results of entropy functions. If the function satisfies the axiom or attribute a "✓" is shown, otherwise "✗".

	Shannon	Rényi	Tsallis	min-h	h^b_{pq}
Continuity 3.2.1	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12
Symmetry 3.2.2	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12	✓ Proof A.1.12
Concentration 3.2.3	✓ Proof A.1.68	✓ Proof A.1.77	✓ Proof A.1.84	✓ Proof A.1.50	✓ Proof A.1.58
Scaling 3.2.4	✓ Proof A.1.69	✓ Proof A.1.69	✓ Proof A.1.69	✓ Proof A.1.51	✓ Proof A.1.59
Replication 3.2.5	✓ Proof A.1.70	✓ Proof A.1.70	✓ Proof A.1.70	✓ Proof A.1.52	✓ Proof A.1.60
Bounds 3.3.1	✓ Proof A.1.71	✓ Proof A.1.78	✓ Proof A.1.85	✓ Proof A.1.53	✓ Proof A.1.61
Quasi-Concavity 3.3.2	✓ Proof A.1.72	✓ Proof A.1.79	✓ Proof A.1.86	✓ Proof A.1.36	✓ Proof A.1.62
Monotonicity 3.3.3	✓ Proof A.1.73	✓ Proof A.1.80	✓ Proof A.1.87	✓ Proof A.1.54	✓ Proof A.1.63
Completeness 3.3.4	✓ Proof A.1.74	✓ Proof A.1.81	✓ Proof A.1.88	✓ Proof A.1.55	✓ Proof A.1.64
Regularity 3.3.5	✓ Proof A.1.75	✓ Proof A.1.82	✓ Proof A.1.89	✓ Proof A.1.56	✓ Proof A.1.65
Homogeneous Growth 3.3.6	✓ Proof A.1.76	✓ Proof A.1.83	✓ Proof A.1.90	✓ Proof A.1.57	✓ Proof A.1.66
Schur-Concavity 3.3.7	✓ Lemma 3.3.7	✓ Lemma 3.3.7	✓ Lemma 3.3.7	✓ Lemma 3.3.7	✓ Lemma 3.3.7
Triangle Inequality 3.3.8	✓ Lemma 3.3.8	✓ Lemma 3.3.8	✓ Lemma 3.3.8	✓ Lemma 3.3.8	✓ Lemma 3.3.8

Theorem 3.4.2 *[Core entropy] The function*

$$\mathsf{h}^{b}_{pq}(\mathbf{p}) = \frac{\|\mathbf{p}\|_p}{\|\mathbf{p}\|_q}, \quad \{p \leq 1, q \geq 1, p \neq q\}, \tag{3.10}$$

satisfies axioms 3.2.1-3.2.5.

The core entropy is continuous (except at the origin which corresponds to the empty set), symmetric, and semi-strictly quasi-concave. It follows the quotient-of-weighted-functions form of general entropy functions [83], e.g. the Aczel-Daróczy entropy [84].

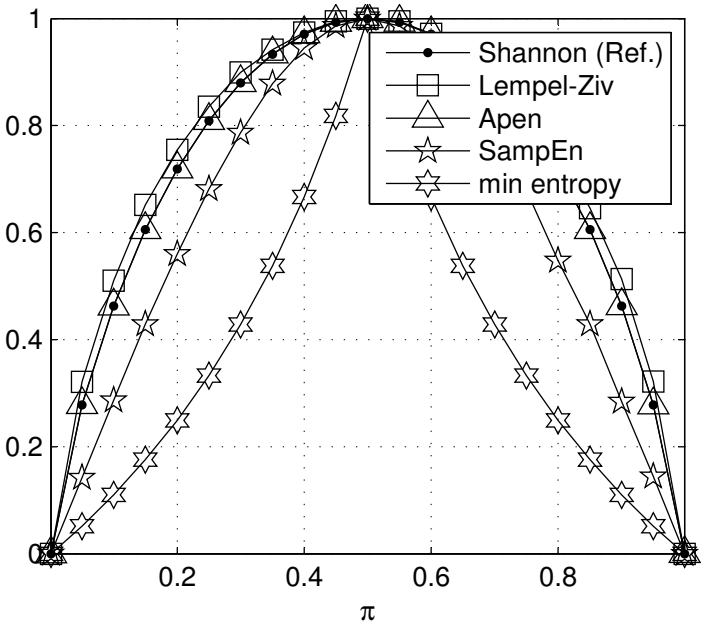
Interestingly, application of the function

$$\mathsf{a}(.) = -\frac{1}{.} \tag{3.11}$$

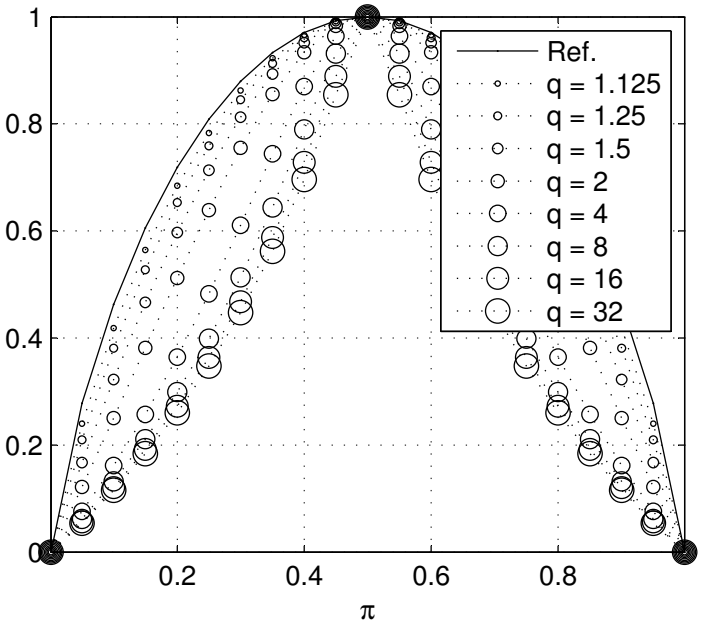
to the core entropy increases the domain of its parameters, p and q, to the set $\{p \geq 1, q \leq 1, p \neq q\}$ and results in a non-positive valued core entropy. Then, the Tsallis entropy ($p \neq 1$, $q = 1$), Rényi entropy ($p \neq 1$, $q = 1$), and their special cases (Shannon, etc.) are transformations (by the logarithm and power functions) of the core entropy. All these previous transformations (including $\mathsf{a}(.)$) are increasing functions; hence, they preserve the (now) quasi-concavity attribute by Theorem A.1.2 [70]. Note that the domain extension of parameters p and q to $\{p \geq 1, q \leq 1, p \neq q\}$ also applies to the core sparsity, which will now be non-negative.

Some features of the core entropy are shown in Figure 3.3 which presents the entropy of a RV following the Bernoulli distribution with the parameter π_b, the canonical example of information theory textbooks [85]. In addition, Figure 3.3(a) indicates the smoothness, monotonicity, symmetry, and quasi-concavity of entropy functions and complexity algorithms while Figure 3.3(b) shows the same features in core entropy. By setting the parameter to $p = 1$, it enables an appreciation of the most interesting case: the core entropy converges to the Shannon entropy as q decreases to 1 (denoted as $q \downarrow 1$). Unsurprisingly, the same tendency is present for $q = 1$ fixed and p increasing to 1 (denoted as $p \uparrow 1$).

For general univariate distributions, [58] identifies the conditions under which entropy defines an order (on distributions). Figure 3.4 illustrates this ordering phenomenon for several distributions. For a better visual comparison, these test distributions do not include the uniform distribution which reaches the maximum in the entropy functions. This is followed by Figure 3.4(a) which shows the ordering obtained using the entropy functions and complexity algorithms.



(a) Entropy functions (*x-axis:* π_b*; y-axis:* $h(\mathbf{p})$)



(b) Core entropy h^b_{1q} (*x-axis:* π_b*; y-axis:* $h(\mathbf{p})$)

Figure 3.3. Entropy functions. [0,1]-normalized entropy of $\mathbf{p} = (\pi_b, 1 - \pi_b)$ of $N = 10000$ samples $v_i \sim \text{Bern}(\pi_b)$ (Bernoulli RV with parameter π_b), with $\pi_b \in [0, 1]$.

Despite its simplicity, the core entropy $\mathsf{h}^b_{1\infty}$ or $\min$ entropy respects the same trend ordering as reported by the other functions. As in Figure 3.3(b), Figure 3.4(b) indicates that the ordering trend drawn by the core entropy, with $p = 1$ fixed and $q \downarrow 1$, converges to the ordering obtained by the Shannon entropy.

3.4.3 On Relative and Absolute Functions

Core sparsity is a relative measure, and core entropy is an absolute measure. The following discusses this subtle difference and its consequences for the way in which both functions follow the axioms (and attributes) of Section 3.2.

For the entropy functions, the replication attribute does not refer to a larger number of events or states. As the number of "replicas" of $\mathbf{p}$ goes to infinity, the *replication attribute states that entropy is intrinsic to a process (system)*. Hence, the entropy functions measure an absolute property as it is the uncertainty of a process (system) of fixed possible events (states).

Analogously, the sparsity functions measure a relative property as it is the efficiency of a basis representation. For example, consider the coefficients $\mathbf{w} = (1, 0)$ and $\tilde{\mathbf{w}} = (1, 1, 0, 0)$, both require 50% of their content for an exact representation, hence both coefficients should present the same level of sparsity.

This subtle difference is characterized by the dimension normalization, $n^{\frac{1}{q}-\frac{1}{p}}$, which is present only in the core sparsity. Furthermore, (discrete) entropy is defined for discrete RVs the countable sample spaces whose cardinality could be infinity, i.e. when $b \to \infty$. Thus, normalization of the core entropy would not be possible in general.

More importantly, replication and completeness attributes should not be confused with the following two common axioms of entropy functions [57].

Lemma 3.4.3 *Core entropy satisfies*

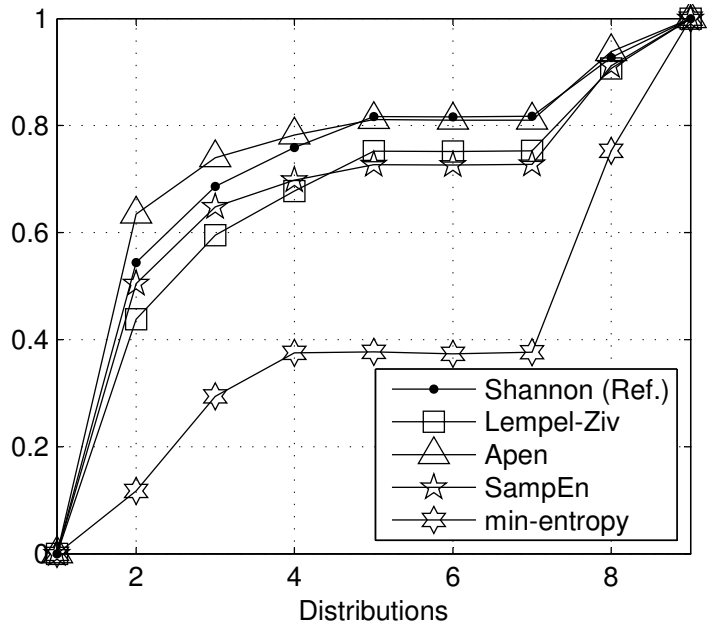
$$\mathsf{h}^b_{pq}(\mathbf{p}) = \mathsf{h}^{b+1}_{pq}(\mathbf{p} \| 0), \tag{3.12}$$

where $.\|.$ *denotes concatenation of vectors.*

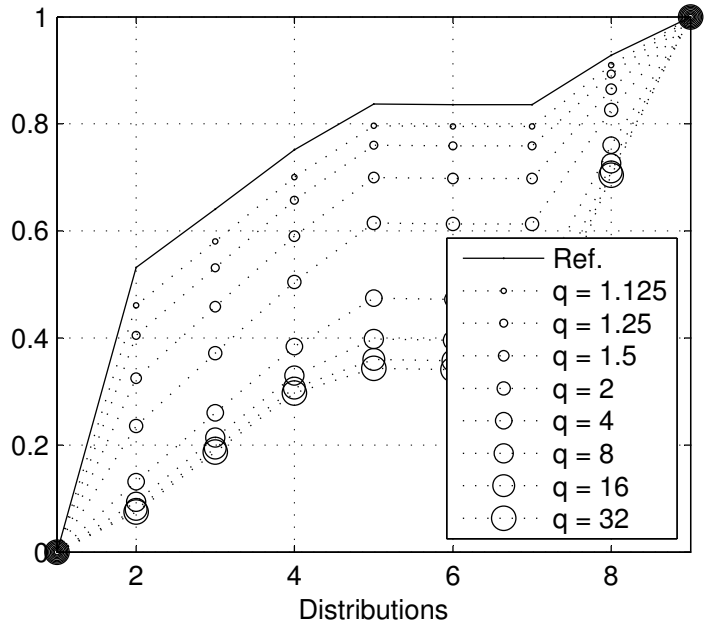
Lemma 3.4.4 *Core entropy satisfies, for* $b_1 < b_2$*,*

$$\mathsf{h}^{b_1}_{pq}(\mathbf{1}_{b_1}) < \mathsf{h}^{b_2}_{pq}(\mathbf{1}_{b_2}), \tag{3.13}$$

where $\mathbf{1}_c$ *is the vector of ones of length* c*.*



(a) Entropy functions (*x-axis: distribution index; y-axis:* $h(\mathbf{p})$)



(b) Core entropy h^b_{1q} (*x-axis: distribution index; y-axis:* $h(\mathbf{p})$)

Figure 3.4. Entropy and ordering of univariate distributions. Unit-variance ($\sigma = 1$) test distributions: (1) Constant (all probability mass in one single event, and deviation $\sigma = 0$), (2) Pareto, (3) Frechet, (4) Log-normal, (5) Exponential, (6) Gamma, (7) Weibull, (8) Gumbel, and (9) Gaussian. Parameters of algorithms (ApEn, SampEn, Lempel-Ziv): alphabet size is 10, $m = 2$ and $r = 0.2\sigma$.

3.4.4 Recovery of Sparse Probability Measures

Consider the minimum (Shannon) entropy problem on the probability simplex with moment constraints [86],

$$\begin{cases} \min_{\mathbf{u}\in\mathbb{S}_+^b} & \exp(\mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{u})) \\ \text{s.t.} & \mathbb{E}_{\mathbf{u}}[V^k] = \mu_k,\ 1 \le k \le m-1, \end{cases} \tag{3.14}$$

where $\mathbb{E}_{\mathbf{u}}$ denotes the expectation with respect to $\mathbf{u}$.

Write the compressed sensing matrix $\Omega = (\omega_{ij})$, with $\omega_{ij} = (a_j)^{i-1}$; and moments $\mathbf{y} = (1, \mu_1, \ldots, \mu_{m-1})$. Since the $\log\max$ entropy upper bounds the Shannon entropy, the solution of (3.14) is upper-bounded [87] by the solution of

$$\begin{cases} \min_{\mathbf{u}\ge\mathbf{0}} & \mathsf{h}^b_{\max}(\mathbf{u}) \\ \text{s.t.} & \Omega\mathbf{u} = \mathbf{y} \end{cases} \equiv \begin{cases} \min_{\mathbf{u}\ge\mathbf{0}} & \|\mathbf{u}\|_0 \\ \text{s.t.} & \Omega\mathbf{u} = \mathbf{y}, \end{cases} \tag{3.15}$$

where $\mathsf{h}^b_{\max}$ denotes the core entropy h^b_{01} or $\max$ entropy. The deterministic measurement matrix Ω is Vandermonde [88], and the linear constraint $\Omega\mathbf{u} = \mathbf{y}$ includes the $\|\mathbf{u}\|_1 = 1$ normalization. Then, under the common hypotheses of compressed sensing (Section 2.1), (3.15) is equivalent to

$$\begin{cases} \min_{\mathbf{u}\ge\mathbf{0}} & \|\mathbf{u}\|_1 \\ \text{s.t.} & \Omega\mathbf{u} = \mathbf{y} \end{cases} \equiv \begin{cases} \min_{\mathbf{u}\ge\mathbf{0}} & 1 \\ \text{s.t.} & \Omega\mathbf{u} = \mathbf{y}. \end{cases} \tag{3.16}$$

However, the formulation (3.16) is useless in this case since it produces a feasibility problem whose solution is not unique if the problem is underdetermined. Nevertheless, if Ω satisfies the common compressed sensing hypotheses, the solution to (3.15) could be approximated by other sparsity functions,

$$\min \mathsf{h}^b_{\min}(\mathbf{p}) \le \min \mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{p}) \le \min \mathsf{h}^b_{\max}(\mathbf{p}), \tag{3.17}$$

where $\mathsf{h}^b_{\max}$ denotes the core entropy $\mathsf{h}^b_{1\infty}$ or $\min$ entropy.[1]

3.5 Summary

A complete characterization of sparsity and entropy is given which offers new functions and tools to efficiently solve sparse recovery problems. As

[1] $\min$ entropy measures the similarity between $\mathbf{p}$ and $\mathbf{1}_b$, the most uncertain distribution, i.e. uniform distribution [40].

such, this chapter proposed a formalism, a joint axiomatic characterization, for sparsity and entropy. The proposed set of axioms is constructive and allows for the derivation of the *core functions*, core sparsity, and core entropy. Both core functions consist of the functional forms of well-known sparsity and entropy functions. Finally, the core functions were applied to the sparse recovery problem where they offered efficient formulations.

Core sparsity will be utilized in Chapter 5 to estimate the effective sparsity of coefficients from the samples. The estimation follows the method developed by [89], which applies the following lemma to estimate each norm in the ratio of core sparsity.

Lemma 3.5.1 *[89] Suppose* $\mathbf{x} \in \mathbb{R}^n$ *is fixed, and* $\mathbf{a} \sim$ *stable*$(\alpha, \gamma)^{\otimes n}$ *with parameters* $\alpha \in (0, 2]$ *and* $\gamma > 0$. *Then, the RV* $\langle \mathbf{x}, \mathbf{a} \rangle$ *is distributed according to stable*$(\alpha, \|\mathbf{x}\|_\alpha)$.

Using this lemma, the estimation of $\|\mathbf{w}\|_\alpha$, with $\alpha = \{p, q\}$, is equivalent to estimate the scale parameter γ of a stable law from an IID sample.

Future work will follow two paths. Concerning entropy, there is a need for a compatibility study of the proposed axioms and specific axioms of entropy, especially those related to relative and conditional entropy should be performed. Some of these specific axioms were already verified here. Concerning sparsity, there is a need to study the core sparsity in a general optimization framework, where the core sparsity can be "tuned" to match the optimal recovery strategy according to the sparsity level.

4. Asymptotic Expansions for Heavy-tailed Distributed Interference Traces

4.1 Introduction

The statistical characterization of interference has gained increasing attention, especially with the emergence of different types of wireless networks, such as ad hoc, sensor, cognitive radio networks, and heterogeneous cellular networks. The statistical interference estimation is of special interest in spectrum-sharing applications where multiple systems need to coexist on the same band without causing excessive interference to each other. Examples of such systems include secondary access to TV White Spaces (TVWS) [90] and Licensed Shared Access (LSA) [91].

Interference is the sum of signals from devices, such as sensors, operating over the same channel. It depends on random parameters, such as the spatial density of sensors and environment fading and shadowing conditions. This randomness is one of the main performance-limiting factors in networks, since the tail of the interference distributions is directly related to the outage probability that communication does not succeed. As the number of sensors increases, there might be a tendency to approximate interference by a normal RV, as individual interferences could be assumed to be independent. However, this approximation is not valid except under specific conditions [29], and physical mechanisms exist that give rise to heavy-tailed interference [92]. In clustered networks, if the distance to the nearest receiver is not appropriately lower bounded, then interference is shown to be heavy-tailed; otherwise, the tail is predominantly dictated by fading [54]. Hence, interference distributions commonly exhibit a heavy tail, are strongly skewed, and far from normal [38].

Most of the existing literature focuses on analytical characterizations of interference. However, only in a few exceptional cases, it is possible

to obtain a closed-form law of the interference. Alternatively, numerical method based on importance sampling is proposed for light-tailed models in [30]. More simple approaches propose to extrapolate the heavy-tailed interference by known distributions as a means to provide tractable expressions for key performance metrics. For networks with repulsion and inhibition, moment and log-moment matching methods are applied in [93] and [44], respectively. Similarly, in [94], a symmetric truncated-stable distribution is adopted since it accommodates the heavy-tailed behavior that exhibits the dominant contribution of a few interferers in the receiver vicinity. Other approaches consider expansions. For instance, the Edgeworth expansion in terms of cumulants is proposed in [28] and [95] for interference when it is close to the normal distribution, for receivers with single and multiple antennas, respectively. However, both papers conclude using a log-normal model since it is more accurate to approximate the aggregated interference. Following the premise that interference follows an skewed and heavy-tailed distribution, this chapter proposes expansions which introduce perturbation terms to fit a log-normal model to the underlying interference.

Heavy-tailed models commonly lack closed-form expressions for the Probability Density Functions (PDF), and either the moments and cumulants cannot be defined starting from a certain order. Consequently, methods need to resort to numerical approximations. Note that the applicability of the method of moments (MoM) [65] is conditioned on the existence of moments up to the number of unknown parameters. The method of simulated quantiles, where simulations of theoretical quantiles are matched with the sample counterparts, is introduced in [96]. Then, a goodness of fit hypothesis test is developed in [97] which, using bootstrap estimates, is able to discriminate between symmetric heavy-tailed models. Using the generalized Taylor series and fractional moments, the distribution is represented in [98]. However, the fractional derivatives and complex numbers setting lead to complicated expressions.

The Second Kind Statistics (SKS) or log-statistics formalism is developed for positive-valued RVs in [39]. The aforementioned paper presents a theorem of strong conditions that verifies the existence of log-moments and log-cumulants of arbitrary order. This new formalism proves considerably easier to use for heavy tail applications than traditional approaches based on classical statistics or First Kind Statistics (FKS), e.g. those using moments and cumulants. For instance, the

parameter estimation of general (heavy-tailed) α-stable models using the method of log-moments (MoLM) [39] or log-cumulants (MoLC) [99] is addressed in [100]. A different approach is utilized in [101], which resorts to a maximum entropy technique and log-moments constraints to recover heavy-tailed densities. The resulting approximate density is the log-normal and constraints introduce perturbation terms to fit the underlying distribution. However, this method requires strong necessary and sufficient conditions for solvability.

This chapter presents a different approach to characterize heavy-tailed models. It is based on expansions. The study of expansions of distributions formed a major part of statistical developments during the early part of the 19th century [102]. It did not last long before an expansion was carried out by considering the normal distribution as the first term of a series, later known as the Gram-Charlier series [103]. From this, the Edgeworth [104] and Cornish-Fisher [105] expansions can be derived. Essentially, in these expansions, moments or cumulants introduce perturbation terms to the normal (distribution) reference to approximate near-normal distributions.

However, interference cannot be always accurately fitted by light-tailed distributions and methods based on normality. To fill this gap, this chapter derives asymptotic expansions from the log-statistics perspective that generalize the Edgeworth and generalized Cornish-Fisher expansions. The proposed expansions are based on a log-normal (distribution) reference, and hence they are designed to approximate near-log-normal distributions. The proposed expansions parallel traditional expansions. Thus, they are readily implemented and inherit the same order of convergence on the log-scale. Hence, they possess better properties as discussed later. In particular, the proposed expansions neatly approximate the tail behavior of different heavy-tailed phenomena. Interestingly, they will also be able to accurately approximate light-tailed or near-normal data.

Network interference is not in most cases normal, and approximation methods should be robust enough to work under both scenarios, normal (symmetric and light-tailed) and non-normal (skewed and heavy-tailed). Hence, this chapter contributes to research on the probabilistic characterization of RVs in three specific ways:

- Traditional asymptotic expansions in terms of cumulants are written in terms of log-cumulants; hence, they are able to approximate skewed and heavy-tailed distributions;

- The proposed expansions automatically tackle also the original scenario under normality; and
- The proposed expansions tightly characterize arbitrary interference; hence, they will be key ingredients to predict performance (metrics) of wireless networks.

4.2 Theoretical Background

This section briefly describes elements and expansions of classical statistics.

4.2.1 Classical Statistics or First Kind Statistics (FKS)

The Fourier transform $\mathcal{F} : f \mapsto \hat{f}$ of the PDF f_X of a real-valued RV X defines the (first) characteristic function,

$$\varphi_X(u) = \hat{f}_X(u) = \mathbb{E}_X[e^{iuX}] = \sum_{k=0}^{\infty} \frac{(iu)^k}{k!} m_{X,k}, \tag{4.1}$$

where $\mathbb{E}_X$ is the expectation with respect to RV X, and k^{th} moments are obtained as

$$m_{X,k} = \mathbb{E}_X[X^k] = (-i)^k D_u^k \varphi_X(u)|_{u=0}. \tag{4.2}$$

Similarly, the second characteristic function ψ_X is defined as

$$\psi_X(u) = \log \varphi_X(u) = \sum_{k=1}^{\infty} \frac{(iu)^k}{k!} \kappa_{X,k}, \tag{4.3}$$

where k^{th} cumulants are obtained as

$$\kappa_{X,k} = (-i)^k D_u^k \psi_X(u)|_{u=0}, \tag{4.4}$$

where D_u^k is the k^{th} order derivative with respect to the variable u.

4.2.2 Classical Asymptotic Expansions

Traditional expansions are based on presented elements of classical statistics. For ease of notation, the following invertible scale transformations are used.

Definition 4.2.1 *[Logarithm $\mathcal{L}$ and standardization $\mathcal{S}$ transformations] The logarithm $\mathcal{L} : X \mapsto \log X$ and standardization $\mathcal{S} : X \mapsto \frac{X-\mu_X}{\sigma_X}$, where the mean $\mu_X = \mathbb{E}_X[X]$ and the variance $\sigma_X^2 = \mathbb{E}_X[(X-\mu_X)^2]$.*

Edgeworth Expansion using Cumulants (EEC)

The original Edgeworth expansion approximates the set of near-normal distributions $\mathcal{N}$ in terms of cumulants. It performs a Fourier inversion of the characteristic function in a Taylor series representation. While performing the inversion, the method capitalizes on a key relationship between the derivatives of the normal distribution and Hermite polynomials.

Let real-valued RV $X \sim \mathcal{N}(\mu_X, \sigma_X)$ and its standardization RV Y given by,

$$Y = \mathcal{S}(X) = \frac{X - \mu_X}{\sigma_X}, \tag{4.5}$$

i.e. $Y \sim \mathcal{N}(0,1) = \{X : X \sim \mathcal{N}, m_{X,1} = 0, m_{X,2} = 1\}$. The density of RV Y is given below:

Theorem 4.2.2 *[Density of $Y \sim \mathcal{N}(0,1)$] [104] Let RV Y be a zero-mean and unit-variance RV, i.e. $\kappa_{Y,1} = 0$ and $\kappa_{Y,2} = 1$. Then,*

$$f_Y(y) = f_{\mathcal{N}}(y) \left(1 + \sum_{s \geq 1} \sum_{\mathcal{K}_s} \omega_{Y,s} \mathcal{H}_{s+2r}(y) \right), \tag{4.6}$$

where $f_{\mathcal{N}}$ is standard normal distribution, set

$$\mathcal{K}_k = \{(k_1, \ldots, k_k; r) : \sum_{j \leq k} j k_j = k, k_j \in \mathbb{Z}_+; r = \sum_{j \leq k} k_j\}, \tag{4.7}$$

$$\omega_{Y,s} = \prod_{m \leq s} \frac{1}{k_m!} s_{Y,m+2}^{k_m}, \tag{4.8}$$

$$s_{Y,k} = \frac{\kappa_{Y,k}}{k!}, \tag{4.9}$$

and $\mathcal{H}_k$ is the k^{th} Hermite polynomial.

From the density of RV $Y \sim \mathcal{N}(0,1)$ in (4.6), the density of arbitrary RV $X \sim \mathcal{N}(\mu_X, \sigma_X)$ is recovered.

Lemma 4.2.3 *[Density of RV $X \sim \mathcal{N}(\mu_X, \sigma_X)$] By simple calculations,*

$$f_X(x) = \frac{1}{\sigma_X} f_Y \left(\frac{x - \mu_X}{\sigma_X} \right). \tag{4.10}$$

Note that Theorem 4.2.2 describes the Edgeworth expansion using cumulants of standardized RV X, i.e. of RV Y, instead of those of RV X. To recover the original formulation of [104], which uses the cumulants of RV X, apply the following technical result.

Lemma 4.2.4 *Let $a > 0$. Then, for $k > 1$,*

$$\kappa_{X+a,k} = \kappa_{X,k}, \tag{4.11}$$

$$\kappa_{aX,k} = a^k \kappa_{X,k}. \tag{4.12}$$

Thus, simply replace $\kappa_{Y,k} = \sigma_X^k \kappa_{X,k}$, $k > 1$ in (4.9).

Generalized Cornish-Fisher Expansion using Cumulants (GCFEC)

The generalized Cornish-Fisher expansion approximates the quantiles x_q of an arbitrary RV X in terms of the quantiles u_q of a reference RV U, and *vice versa*, i.e. it calculates $x_q = x_q(u_q)$ and $u_q = u_q(x_q)$. The method is based on the Fourier inversion of the Gram-Charlier series representation.

Theorem 4.2.5 *[Quantiles of arbitrary RV X] [105] Define the function $\gamma(x) = \frac{F_X(x)-F_U(x)}{f_U(x)}$. Then,*

$$u_q = x_q + \sum_{k\geq 1} \frac{\gamma(x_q)^k}{k!} C_k(x_q), \tag{4.13}$$

$$x_q = u_q - \sum_{k\geq 1} \frac{D_{(k)}(\gamma(u_q))^k}{k!}, \tag{4.14}$$

where the following recurrences hold

$$C_k(x) = \begin{cases} 1, & k = 1, \\ P(x) = -\frac{f'_U(x)}{f_U(x)}, & k = 2, \\ ((k-1)P(x) + D_x^1)C_{k-1}(x), & k > 2, \end{cases} \tag{4.15}$$

$$D_{(k)} = \prod_{j=1}^{k-1} (jP(u) - D_u^1). \tag{4.16}$$

Both the general standardizing (4.13) and inverse (4.14) expansion depend on the function $\gamma(x)$.

Theorem 4.2.6 *[105] Let*

$$\lambda_k = \kappa_{X,k} - \kappa_{U,k}. \tag{4.17}$$

Then,

$$\gamma(x) = -\sum_{\pi\in\Pi} \lambda_\pi P_m(x), \tag{4.18}$$

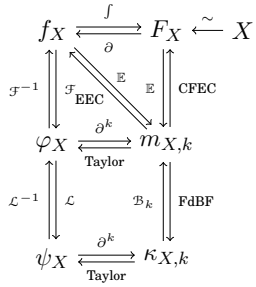
where $\Pi = \{\{(s_i, \rho_i)\}_{i=1}^k : m = \sum_{i=1}^k \rho_i s_i, k = \sum_{k=1}^k \rho_i\}$ is a partition set, coefficients λ_π are

$$\lambda_\pi = \prod_{i=1}^{k} \frac{1}{\rho_i!} \left(\frac{\lambda_{s_i}}{s_i!}\right)^{\rho_i}, \tag{4.19}$$

and polynomials P_m satisfy the recurrence

$$P_m(x) = \begin{cases} 1, & m = 1, \\ P(x) = -\frac{f'_U(x)}{f_U(x)}, & m = 2, \\ (P(x) - D_x^1)P_{m-1}(x), & m > 2. \end{cases} \tag{4.20}$$

Table 4.1. Relations in FKS.



In the original Cornish-Fisher expansion (using cumulants) the reference distribution is the (standard) normal. This leads to a significant amount of simplifications.

Cornish-Fisher Expansion using Cumulants (CFEC)
In the original Cornish-Fisher expansion, RV U is normal distributed such that cumulants of order greater than 2 are zero [106]. Thus, the method employs terms presented as the sums of the products of the Hermite polynomials and possesses a very simple functional form of the polynomial, $P(u) = -\frac{f'_U(u)}{f_U(u)} = u$.

Corollary 4.2.7 *[Quantiles of RV $X \sim \mathcal{N}(\mu_X, \sigma_X)$] [106] Let RV $Y = \mathcal{S}(X)$. Then, the quantiles of RV Y are*

$$y_q = v_q + \kappa_{Y,3} h_1(v_q) + \kappa_{Y,4} h_2(v_q) + \kappa_{Y,3}^2 h_{11}(v_q) + \dots \\ \kappa_{Y,5} h_3(v_q) + \kappa_{Y,3}\kappa_{Y,4} h_{12}(v_q) + \dots, \tag{4.21}$$

where v_q are the quantiles of RV $V = \mathcal{S}(U) \sim \mathcal{N}(0,1)$ and polynomials

$$h_1(v_q) = \frac{1}{6}\mathcal{H}_2(v_q), \tag{4.22}$$

$$h_2(v_q) = \frac{1}{24}\mathcal{H}_3(v_q), \tag{4.23}$$

$$h_{11}(v_q) = -\frac{1}{36}(2\mathcal{H}_3(v_q) + \mathcal{H}_1(v_q)), \tag{4.24}$$

and so on [107]. Then, the quantiles of RV X are recovered as

$$x_q = \mathcal{S}^{-1}(y_q) = \mu_X + \sigma_X y_q. \tag{4.25}$$

Table 4.1 shows the relationships in classical statistics and the role of traditional expansions.

The analytical calculation of characteristic functions of certain RVs on $\mathbb{R}_+$ is sometimes difficult or even impossible. This motivates the introduction of a formalism that makes it possible to perform an effective

analysis of distributions defined on $\mathbb{R}_+$. It is called here log-statistics [39]. Under this formalism, the elements of classical statistics are redefined, namely the characteristic functions yielding moments and cumulants. The remaining of this section briefly describes these elements.

4.2.3 Log-Statistics or Second Kind Statistics (SKS)

Log-statistics follows the same construction as in classical statistics to define characteristic log-functions, by replacing the Fourier transform by the Mellin transform [108]; and to define log-moments and log-cumulants, by the derivation of these functions.

The Mellin transform $\mathcal{M} : f \mapsto \acute{f}$ of the PDF f_X of a non-negative-valued RV X defines the (first) characteristic log-function,

$$\tilde{\varphi}_X(v) = \acute{f}_X(v) = \mathbb{E}_X[X^{v-1}] = \sum_{k=0}^{\infty} \frac{(v-1)^k}{k!} \tilde{m}_{X,k}, \tag{4.26}$$

where k^{th} log-moments are obtained as

$$\tilde{m}_{X,k} = \mathbb{E}_X[(\log X)^k] = D_v^k \tilde{\varphi}(v)_X|_{v=1}. \tag{4.27}$$

Similarly, the second characteristic log-function is defined as

$$\tilde{\psi}_X(v) = \log \tilde{\varphi}_X(v) = \sum_{k=1}^{\infty} \frac{(v-1)^k}{k!} \tilde{\kappa}_{X,k}, \tag{4.28}$$

where k^{th} log-cumulants are obtained as

$$\tilde{\kappa}_{X,k} = D_v^k \tilde{\psi}(v)_X|_{v=1}. \tag{4.29}$$

Essentially, log-statistics, of such similarity to classical statistics, cannot lead to intrinsically new relations and methods. More importantly, methods using log-statistics should have clear advantages over traditional methods (those using elements from classical statistics) to treat heavy-tailed models [39].

4.2.4 On the Log-Statistics Relationship

Since characteristic functions in log-statistics are defined accordingly as in classical statistics, relationships between log-moments and log-cumulants are identical to those existing between the moments and cumulants [39]. According to similar arguments as those presented in [43], the log-moments and log-cumulants are related by the combinatorial formula of Faá di Bruno (FdBF)

$$\tilde{\kappa}_{X,k} = \tilde{m}_{X,k} - \sum_{j=1}^{k-1} \binom{k-1}{j-1} \tilde{\kappa}_{X,j} \tilde{m}_{X,k-j}, \tag{4.30}$$

and reversely through

$$\tilde{m}_{X,k} = \mathcal{B}_k(\tilde{\kappa}_{X,1}, \ldots, \tilde{\kappa}_{X,k}), \tag{4.31}$$

where $\mathcal{B}_k$ is the k^{th} complete Bell polynomial.

4.3 Proposed Asymptotic Expansions

Here, the proposed expansions, which were originally developed within classical statistics, are derived in terms of the elements of log-statistics.

4.3.1 Expansion of Density

Edgeworth Expansion using Log-cumulants (EEL)
Let a non-negative-valued RV $X \sim \mathcal{L}^{-1}\mathcal{N}(\tilde{\mu}_X, \tilde{\sigma}_X)$ and its standardization RV Y given by,

$$Y \sim \mathcal{L}^{-1}\mathcal{N}(0,1) = \{X : X \sim \mathcal{L}^{-1}\mathcal{N}, \tilde{m}_{X,1} = 0, \tilde{m}_{X,2} = 1\}, \tag{4.32}$$

and RV $Z = \mathcal{L}(Y) \sim \mathcal{N}(0,1)$,

$$Z = \mathcal{L}(Y) = \mathcal{SL}(X) = \frac{\log X - \tilde{\mu}_X}{\tilde{\sigma}_X}, \tag{4.33}$$

where log-mean $\tilde{\mu}_X = \mathbb{E}_X[\log X]$ and log-variance $\tilde{\sigma}_X^2 = \mathbb{E}_X[(\log X - \tilde{\mu}_X)^2]$. The density of the standard RV Y is given by the following result.

Theorem 4.3.1 *[Density of RV $Y \sim \mathcal{L}^{-1}\mathcal{N}(0,1)$] Let RV Y be a zero-log-mean and unit-log-variance RV, i.e. $\tilde{\kappa}_{Y,1} = 0$ and $\tilde{\kappa}_{Y,2} = 1$, which defines the RV $Z = \mathcal{L}(Y) \sim \mathcal{N}(0,1)$. Then,*

$$f_Z(z) = f_{\mathcal{N}}(z)\left(1 + \sum_{s\geq 1}\sum_{\mathcal{K}_s} \tilde{\omega}_{Y,s}\mathcal{H}_{s+2r}(z)\right), \tag{4.34}$$

where $f_{\mathcal{N}}$, $\mathcal{K}_k$, and r are defined as in Theorem 4.2.2,

$$\tilde{\omega}_{Y,s} = \prod_{m\leq s}\frac{1}{k_m!}\tilde{s}_{Y,m+2}^{k_m}, \tag{4.35}$$

$$\tilde{s}_{Y,k} = \frac{\tilde{\kappa}_{Y,k}}{k!}, \tag{4.36}$$

and $\mathcal{H}_k$ is the k^{th} Hermite polynomial. Then, the density of RV Y is

$$f_Y(y) = \frac{1}{y}f_Z(\log y). \tag{4.37}$$

From the density of RV $Y \sim \mathcal{L}^{-1}\mathcal{N}(0,1)$ in (4.37), the density of arbitrary RV $X \sim \mathcal{L}^{-1}\mathcal{N}(\tilde{\mu}_X, \tilde{\sigma}_X)$ is recovered by,

Lemma 4.3.2 *[Density of RV $X \sim \mathcal{L}^{-1}\mathcal{N}(\tilde{\mu}_X, \tilde{\sigma}_X)$] By simple calculations,*

$$f_X(x) = \frac{1}{x\tilde{\sigma}_X}\left(\frac{x}{e^{\tilde{\mu}_X}}\right)^{\frac{1}{\tilde{\sigma}_X}} f_Y\left(\left(\frac{x}{e^{\tilde{\mu}_X}}\right)^{\frac{1}{\tilde{\sigma}_X}}\right). \tag{4.38}$$

As in the EEC, in order to formulate the EEL in terms of the log-cumulants of RV X instead of those of standardized RV Y, apply the following technical result.

Lemma 4.3.3 *Let $a > 0$. Then, for $k > 1$,*

$$\tilde{\kappa}_{aX,k} = \tilde{\kappa}_{X,k}, \tag{4.39}$$

$$\tilde{\kappa}_{X^a,k} = a^k \tilde{\kappa}_{X,k}. \tag{4.40}$$

Thus, simply replace $\tilde{\kappa}_{Y,k} = \tilde{\sigma}_X^k \tilde{\kappa}_{X,k}$ in (4.36).

Essentially, the EEC represents near-normal distributions by adjusting the (standard) normal distribution using the cumulants. Analogously, the EEL represents near-log-normal distributions by adjusting, again, the (standard) normal distribution but in the logarithmic scale and using the log-cumulants. Thus, the EEL could be compactly formulated in terms of the (standard) log-normal distribution.

4.3.2 Approximation of Quantiles

Generalized Cornish-Fisher Expansion using Log-cumulants (GCFEL)
The key tool to develop the GCFEL is the Gram-Charlier series. Thus, in order to develop this expansion using log-cumulants, the series should also be first written in terms of log-cumulants.

Lemma 4.3.4 *[Gram-Charlier series using log-cumulants] Let*

$$\tilde{\lambda}_k = \tilde{\kappa}_{X,k} - \tilde{\kappa}_{U,k}, \tag{4.41}$$

and RVs $Y = \mathcal{L}(X)$ and $V = \mathcal{L}(U)$. Then,

$$F_Y(y) = \exp\left(\sum_{k\geq 1} \frac{\tilde{\lambda}_k}{k!}(-D_y^1)^k\right) F_V(y). \tag{4.42}$$

Now, developing F_Y about F_V as a Charlier differential series, and collecting terms of like degree in the resulting series gives

Theorem 4.3.5 *The function*

$$\tilde{\gamma}_X(x) = \gamma_Y(\log x) = -\sum_{\pi\in\Pi} \tilde{\lambda}_\pi P_m(\log x), \tag{4.43}$$

where the partition set Π *and the polynomials* P_m *are defined as in Theorem 4.2.6, and*

$$\tilde{\lambda}_\pi = \prod_{i=1}^{k} \frac{1}{\rho_i!} \left(\frac{\tilde{\lambda}_{s_i}}{s_i!} \right)^{\rho_i}. \tag{4.44}$$

Finally, the Theorems 4.2.5 and 4.3.5 and the exponential transformation of RV Y, $X = \mathcal{L}^{-1}(Y)$, prove the following.

Theorem 4.3.6 *[Quantiles of arbitrary RV* X*] Let* $\gamma_Y(y) = \frac{F_Y(y) - F_V(y)}{f_V(y)}$*, with RV* $Y = \mathcal{L}(X)$ *and* $V = \mathcal{L}(U)$*. Then,*

$$u_q = x_q e^{\sum_{k \geq 1} \frac{\gamma_Y(\log x_q)^k}{k!} C_k(x_q)}, \tag{4.45}$$

$$x_q = u_q e^{-\sum_{k \geq 1} \frac{D_{(k)}(\gamma_Y(\log u_q))^k}{k!}}. \tag{4.46}$$

Note that Theorems 4.2.5 and 4.3.6 are general since both adopt an arbitrary reference distribution for RV U. This means that the GCFEC or GCFEL approximate arbitrary distributions, not necessarily near-normal or near-log-normal, as long as the cumulants or log-cumulants exist, respectively.

It should not be a surprise by now that it is possible and straightforward to write the simplified Cornish-Fisher expansion in terms of the log-cumulants.

Cornish-Fisher Expansion using Log-cumulants (CFEL)

Corollary 4.3.7 *[Quantiles of RV* $X \sim \mathcal{L}^{-1}\mathcal{N}(\tilde{\mu}_X, \tilde{\sigma}_X)$*] Let RVs* $Y = \mathcal{L}(X)$ *and* $Z = \mathcal{S}(Y)$*, i.e.* $Z = \mathcal{S}\mathcal{L}(X)$*. Then, the quantiles of RV* Z *are*

$$\begin{aligned} z_q = w_q + \tilde{\kappa}_{Z,3} h_1(w_q) + \tilde{\kappa}_{Z,4} h_2(w_q) + \tilde{\kappa}_{Z,3}^2 h_{11}(w_q) + \ldots \\ \tilde{\kappa}_{Z,5} h_3(w_q) + \tilde{\kappa}_{Z,3}\tilde{\kappa}_{Z,4} h_{12}(w_q) + \ldots, \end{aligned} \tag{4.47}$$

where $w_q = v_q$ *are the quantiles of RV* $V = \mathcal{S}(U) \sim \mathcal{N}(0,1)$ *and the polynomials* h_m *are defined as in Theorem 4.2.7. Then, the quantiles of RV* Y *are recovered as*

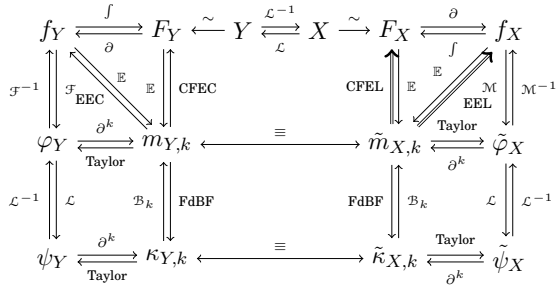
$$y_q = \mathcal{S}^{-1}(z_q) = \mu_y + \sigma_y z_q, \tag{4.48}$$

and those of RV X *as*

$$x_q = \mathcal{L}^{-1}(y_q) = \mathcal{L}^{-1}\mathcal{S}^{-1}(z_q) = e^{\tilde{\mu}_X + \tilde{\sigma}_X z_q}. \tag{4.49}$$

Table 4.2 summarizes these relationships (and highlights the proposed expansions using double lines).

Table 4.2. Relations in SKS. Proposed expansions in double lines.



4.4 Numerical Experiments and Discussion

This section describes numerical experiments and demonstrates an application of the proposed expansions in the performance metrics. Finally, the section discusses further properties and implementation details.

4.4.1 Numerical Experiments

To study the tail behavior of interference in networks that follow different spatial processes (PPP, PCP, SSI, or SPP) [109], the channel model considered is the distance-dependent path loss model $\ell(r) = (1 + r^{\alpha})^{-1}$, with r denotes the distance and $\alpha = 4$, no light-tailed fading (since it does not contribute heavy-tailed interference), and log-normal shadowing with the parameters, $\mu = 1$ and $\sigma = 1.25$, i.e. $e^{\xi_i} \sim \mathcal{N}(\mu, \sigma)$. These parameters are selected so that they will produce heavy tails in the distribution of the interference. The network is drawn such that the expected number of interferers is approximately 1000 in each realization. The number of realizations is 10000. In addition, the expansions proposed in Section 4.3 are applied to characterize the interference aggregated from these networks of varying regularity. In all expansions, four orders are utilized since this should be enough to represent any data as reported in [110].

Figure 4.1 illustrates the approximation results of the EEC and EEL. The EEC clearly diverges trying to approximate the density of the skewed and heavy-tailed interference. Moreover, as it is known to happen for the EEC, the resulting approximation presents negative values. On the other hand, the EEL successfully approximates the interference after a few orders, and it even works with just one single order. The reason is that the EEL is an asymptotic approximation for distributions with a broad spec-

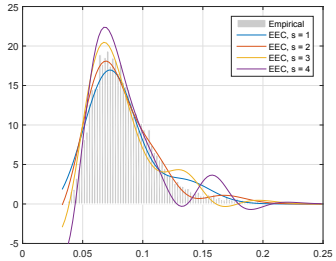
trum of tails, as it will be discussed in Section 4.4.2. Moreover, this result is in accordance with [28] and [95] where the log-normal distribution is proposed to approximate similar interference patterns. More importantly, the EEL barely presents negative values as it (frequently) happens in the traditional EEC. Although imperceptible, negative values appear for 1 and 2 orders in the SSI network. This is due to the rapid growth of density around zero. Concerning the tails, these are tightly approximated and more importantly, the EEL clearly converges. The gamma and log-normal models are commonly proposed to approximate the heavy-tailed behavior of interference. However, these two models fail fitting tests [44]. Another common model is the α-stable. The matching results of this model are shown in Figure 4.1 via the method of Koutrouvelis [111]. However, the method of Koutrouvelis clearly fails to approximate both tails; hence, it presents an incorrect mode.

Performance Metrics

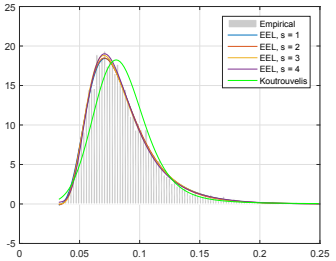
The success probability given by $\mathbb{P}(I_{\mathbf{x}} < \gamma)$, where $I_{\mathbf{x}}$ is the interference received at sensor x and γ is a design parameter, is a useful quantification for performance analysis in cellular systems because performance at the cell edge is usually interference limited [112]. From this, other metrics can be derived, such as the Signal to Interference Ratio (SIR) and Bit Error Probability (BEP). Thus, the proposed expansions are used to approximate this performance metric.

The expression $\mathbb{P}(I_{\mathbf{x}} < \gamma)$ can be written using the cumulative form of the EEL (no integral) or simply by using the CFEL. The approximation results of the CFEC and CFEL are shown in Figure 4.2. The CFEC diverges trying to calculate the quantiles of skewed and heavy-tailed interference (or outage probability). On the other hand, the CFEL successfully calculates the quantiles of the interference after a few orders, and it even works with just one single order as in the EEL case. Moreover, the results from the α-stable model via the Koutrouvelis method are also included for comparison. The method of Koutrouvelis overestimates the outage probability for all networks, which represents a conservative approximation.

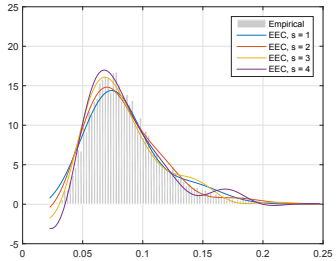
Interestingly, EEL and CFEL expansions show their best performance in different regions of the density. The EEL approximates more neatly the tails of the density in comparison with the CFEL, which approximates more neatly the mode of the density.



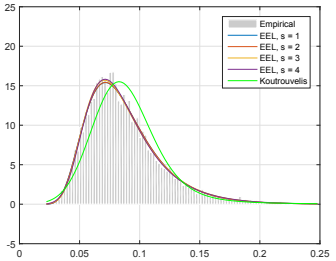
(a) SSI network, EEC approximation



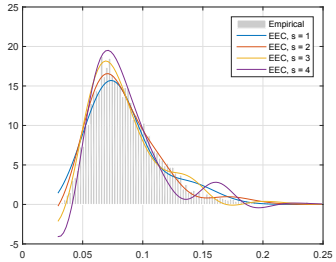
(e) SSI network, EEL approximation



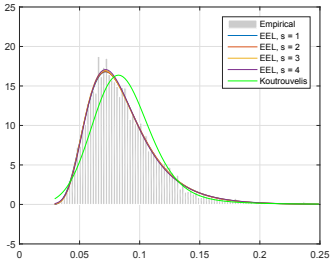
(b) SPP network, EEC approximation



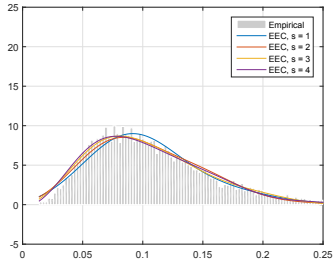
(f) SPP network, EEL approximation



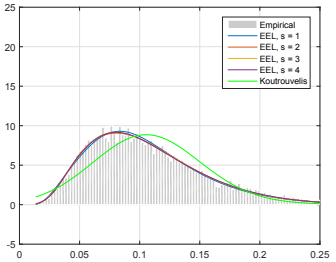
(c) PPP network, EEC approximation



(g) PPP network, EEL approximation

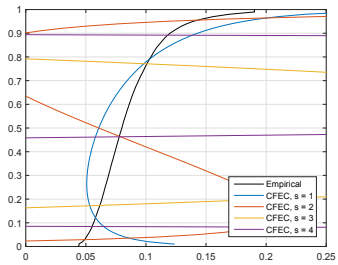


(d) PCP network, EEC approximation

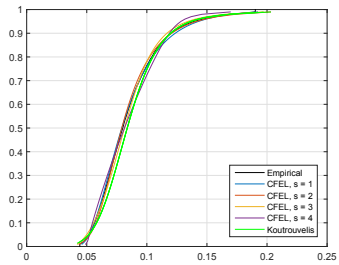


(h) PCP network, EEL approximation

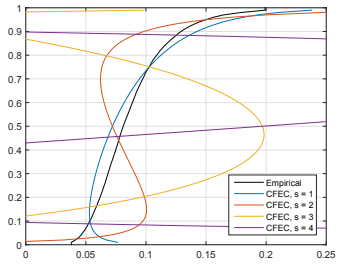
Figure 4.1. Probability density function of interference $I_{\mathbf{x}}$ (*x-axis: t; y-axis:* $f_{I_{\mathbf{x}}}(t)$). Analytical approximation results of the EEC and EEL. The empirical data are collected from the aggregate interference generated by networks of interferers of varying regularity following the channel and network models as defined in Figure 2.1.



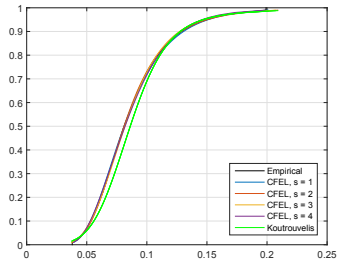
(a) SSI network, CFEC approximation



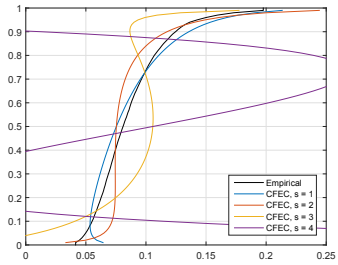
(e) SSI network, CFEL approximation



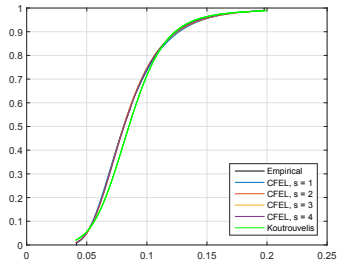
(b) SPP network, CFEC approximation



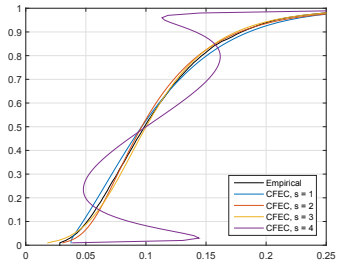
(f) SPP network, CFEL approximation



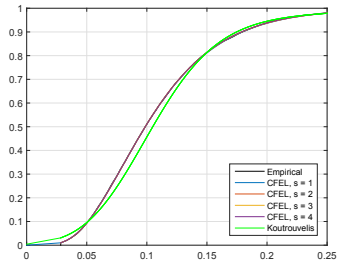
(c) PPP network, CFEC approximation



(g) PPP network, CFEL approximation



(d) PCP network, CFEC approximation



(h) PCP network, CFEL approximation

Figure 4.2. Cumulative density function of interference $I_{\mathbf{x}}$ (*x-axis: t; y-axis:* $F_{I_{\mathbf{x}}}(t)$). Analytical approximation results of the CFEC and CFEL. The empirical data are collected from the aggregate interference in networks of varying regularity following the channel and network models as defined in Figure 2.1.

4.4.2 Extensions to Symmetric and Light-tailed Interference

By the implicit Mellin transform in log-statistics (4.27) and arising log-normal reference, the proposed EEL and (G)CFEL expansions are designed to approximate non-negative-valued RVs following a near-log-normal distribution. However, in practice, these expansions will also be useful for real-valued RVs following a near-normal distribution. For instance, under very specific conditions, the interference may follow a light-tailed model. See for example [44] where the light-tailed interference is generated by the light-tailed fading terms. In the proposed expansions, the heaviness of the tail of the log-normal reference allows for a perturbance (in particular, an attenuation) of these tails as desired to fit a large spectrum of tails. Basically, small values in the Hermite polynomials can generate the light-tails of near-normal distributions. Formally, this could be proved using the decay order of the tails. Furthermore, note that real-valued RVs following a near-normal distribution possess finite variance. Hence, a simple shift (that is adding a large value) can make these RVs non-negative. In the EEL, this shift is straightforward. In the (G)CFEL, it is more delicate, but also feasible. Accordingly, Figure 4.3 illustrates an example of how the EEL accurately approximates normally distributed data with the parameters, $\mu = 0$ and $\sigma = 1$.

Note that the opposite, using traditional expansions, such as EEC or (G)CFEC, to approximate near-log-normal distributions, is not possible, as clearly shown by Figures 4.1 and 4.2. In this case, the shifting step should be replaced by an exponentiation, but this will not be enough in general. The reason is that the normal reference of traditional expansions does not have sufficient tails that can be perturbed to form heavy tails. In fact, this is why traditional expansions commonly diverge when trying to approximate heavy tails. Essentially, the Hermite polynomials compensate for the light tails of normal reference with large values which provoke a divergence of the traditional expansions.

Other Heavy-tailed Scenarios

In the analysis of real systems, the inputs, noise, and outputs are usually characterized in statistical terms. Here, normality, characterized by light tails, is typically assumed for its convenient analytical properties. However, if this assumption does not hold then the analysis may be misleading and there is a significant risk of drawing incorrect conclusions about the system [113]. Hence, it is more accepted nowadays to assume

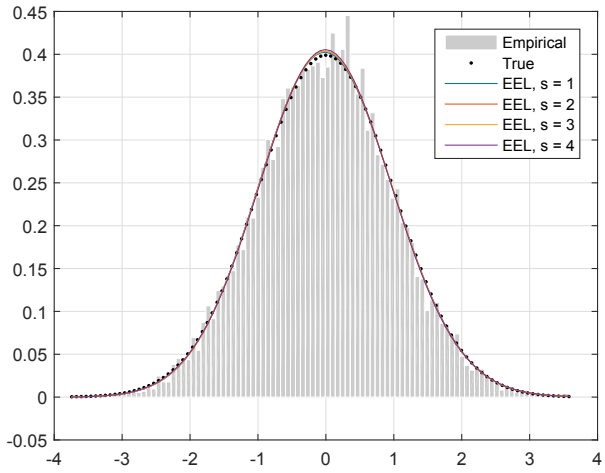


Figure 4.3. Approximation of normally distributed data, parameters $\mu = 0$ and $\sigma = 1$, via the EEL (*x-axis:* x*; y-axis:* $f_X(x)$).

skewed (asymmetric) and heavy-tailed models, i.e. the counterpart of normality. In computer network traffic, heavy tails arise in most data that can be extracted from traffic traces [114]. In cloud-computing infrastructures, a real cluster usage datasets have revealed that the job and task durations and requested resources appear to follow a heavy-tailed distribution [115]. In dynamic spectrum access, if either the primary user busy time or the secondary user message size is heavy-tailed, then the time to complete the transmissions of secondary users will be as well [116]. The proposed expansions derived in this chapter represent new tools that truly and asymptotically will approximate the heavy-tailed behavior of many systems.

4.4.3 Implementation Details

The diagrams in Tables 4.1 and 4.2 summarized the relationships between the elements of classical statistics and log-statistics, respectively. See also [117] for extensions of these relationships to matrix-variates. For instance, in classical statistics, Table 4.1 shows that if the set of moments or cumulants are known, and exist, the PDF can be obtained using the EEC and the Cumulative Density Function (CDF) using the CFEC. For log-statistics, Table 4.2 illustrates that if the set of log-moments or log-cumulants are known, the PDF and CDF can be obtained using the proposed EEL and CFEL, respectively.

Table 4.3. Paths to perform the Edgeworth expansion. Double lines describe the followed path: expressions (4.4, 4.12, 4.9, 4.8, 4.6, 4.10) (FKS) and (4.29, 4.40, 4.36, 4.35, $\equiv$, 4.34, 4.37, 4.38) (SKS). Plain lines describe an alternative simple path. Dotted lines describe other possible paths.

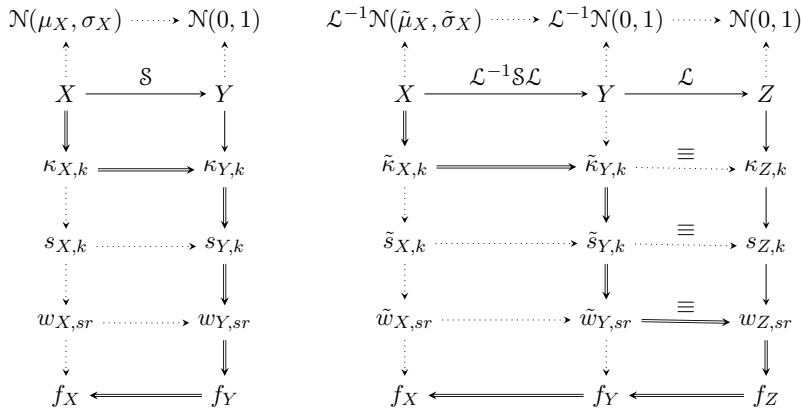
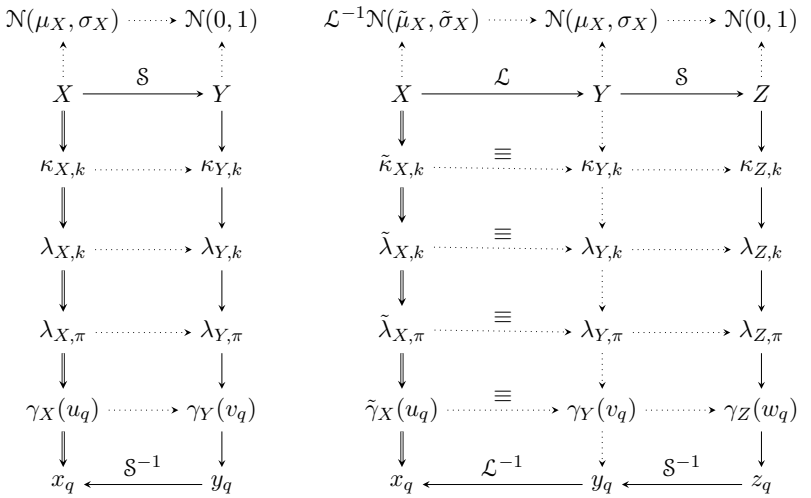


Table 4.2 also highlights a key relationship between elements of both formalisms. Basically, the k^{th} log-moment of a certain RV X equals the k^{th} moment of the RV $\log X$. Using this observation, and the logarithmic and standardization transformations, Table 4.3 presents different paths to implement the EEC and EEL. For both expansions, using cumulants and log-cumulants, the simple path consists of transforming the data to standard normal and then undoing the transformation. For clarification, Table 4.4 shows this for the (G)CFEC and (G)CFEL. Hence, implementation of the proposed expansions is relatively straightforward.

4.5 Summary

There is considerably recent interest in skewed and heavy-tailed distributions, but there is a lack of methods to handle them. [118] demonstrates that although very little can be said in terms of moments about the central part of the distribution, the tail is much more sharply defined by the finite moments. When moments do not exist, log-moments (or log-cumulants) are of great help since they will offer mechanisms, as the proposed expansions, to bound the tail behavior of the underlying distribution. Hence, this chapter extends the usability of the Edgeworth and Cornish-Fisher expansions to analogous asymptotic expansions, denoted as EEL and GCFEL, which are based on elements of a formalism called log-statistics. The proposed expansions possess better convergence attributes

Table 4.4. Paths to perform the Cornish-Fisher expansion. Double lines describe the followed path: expressions (4.4, 4.17, 4.19, 4.18, 4.14) (FKS) and (4.29, 4.41, 4.44, 4.43, 4.46) (SKS). Plain lines describe an alternative simple path. Dotted lines describe other possible paths.



in comparison with the original expansions and inherit the capability of log-statistics to characterize heavy-tailed distributions. Furthermore, the proposed expansions are also able to approximate non-skewed and light-tailed distributions due to the tails of their arising log-normal references. It is important to note that relationships of classical statistics remain valid for log-statistics; hence, the extension of other traditional methods should be straightforward. The chapter illustrates alternative and simple paths to formally perform other extensions. As the main objective of the chapter, the EEL and GCFEL are successfully applied to characterize interference traces drawn using models from stochastic geometry. Under mild conditions, the distribution of these traces exhibits a heavy tail and is strongly skewed [38]. Key performance metrics could be expressed in terms of these expansions.

These expansions will be utilized to assess the effect of interference in the protocol developed in the next Chapter 5.

5. Stochastic Analysis of Compressive Data Aggregation

5.1 Introduction

This chapter introduces S-CDA, a data aggregation protocol for WSNs. S-CDA follows a probabilistic model which helps to develop a framework to analyze simultaneously the communication and compression aspects of CDA. Under this framework, a WSN with randomly deployed sensors is analyzed. Moreover, the previous results will be adopted from the technical Chapters 3 and 4. In particular, the sparsity functions from Chapter 3 will estimate the level of sparsity of the sensing field from the partial samples. Then, the asymptotic expansions from Chapter 4 will be utilized to evaluate the results by computing the effect of the aggregate interference in the network.

The research on CDA is vast. For instance, [24] proposed Wireless Compressive Sensing (WCS) for energy harvesting sensor nodes and derived a measurement matrix for a network suffering from Rayleigh fading; [25] proposed Random Access Compressed Sensing (RACS) also under fading channels and sensors using ALOHA protocol (ALOHA) to access the medium; [22] proposed Minimum-Energy and High-Fidelity CDA (ME-CDA) which employs diffusion wavelets and integer optimization to find the shortest routing path; and [23] proposed Treelet-based Clustered CDA (TC-CDA) and studied the recovery error using treelets as a sparsifying matrix. In these studies, [24] and [25] assume a given sparse representation of the data under imperfect communications and focus on the communication analysis, while [22] and [23] assume perfect communications and focus on the compression proposing useful sparsifying matrices.

S-CDA will be mainly compared against RACS. Note that in RACS,

the key mechanism to reduce packet collisions and collect more packets is actually the interference (and fading effects). However, RACS does not follow the essence of compressed sensing, the efficient random sensing. RACS introduces redundancy in terms of repeated packets from continuous rate λ_1 samples from the entire network. Contrary to WCS and RACS, S-CDA uses the PPP to design and analyze a sparsifying matrix, the Random Discrete Fourier Transform (RDFT), which is defined in (5.5). Contrary to the diffusion wavelets and treelets of [22] and [23], respectively, the RDFT is readily implemented in a simple matrix-form.

The main contributions of this chapter are:

- a probabilistic model for CDA in WSNs;
- a framework to jointly analyze the compression and communication aspects of CDA; and
- the analysis of a new sparsifying matrix suitable for random deployments and other random sensing scenarios.

5.2 Proposed Data Aggregation Protocol

This section describes S-CDA, which is motivated by the remark:

"One can regard the possibility of digital compression as failure of sensor design. If it is possible to compress measured data, one might argue that too many measurements were taken" (David Brady, Duke University) [47].

This is the essence of compressed sensing. In the data aggregation scenario, the remark refers to the failure of the WSN design by an excess of collected measurements. For randomly deployed WSNs, S-CDA is a CDA protocol which follows the efficient random principle of compressed sensing.

5.2.1 Stochastic Compressive Data Aggregation (S-CDA)

S-CDA consists of three mechanisms.

Random Sensing Mechanism (CS2)
At each frame, each sensor $\mathbf{x}_i \in \Theta$ is active with a small probability $\rho \ll 1$ and acquires a measurement v_i. The set of active sensors is

$$\Theta_{\mathsf{on}} = \{\mathbf{x}_i \in \Theta : \delta_\Omega(i) = 1, \delta_\Omega(i) \sim \text{Bernoulli}(\rho)\}, \tag{5.1}$$

where the indicator variable $\delta_\Omega(i)$ equals 1 if the sensor $\mathbf{x}_i$ is active, cf. key condition (CS2) in Chapter 3.

In [7], the Bernoulli model is introduced for selecting the set of (frequency) information content. Then, it was explained how the results from this model can be translated into those for the uniform model. This is the so-called random sensing of compressed sensing.

In S-CDA, the random activation of sensors performs an independent thinning of the PPP Θ, resulting in PPP Θ_{on} with an intensity $\lambda_{\mathsf{on}} = \rho\lambda$ which preserves the randomness of the measurement process.

Free-Collision Medium Access Mechanism (CS3)
At each frame, each active sensor $\mathbf{x}_i \in \Theta_{\mathsf{on}}$ competes with its neighborhood $\mathcal{N}_i$ such that only the sensor that tries earlier, event associated to a random time RV $t_i \sim \text{Uniform}([0,1])$, achieves transmission, denoted $\mathbf{x}_i \in \Theta_\Omega$. The set of retained sensors is,

$$\Theta_\Omega = \{\mathbf{x}_i \in \Theta_{\mathsf{on}} : t_i < t_j, \forall \mathbf{x}_j \in \mathcal{N}_i\}. \tag{5.2}$$

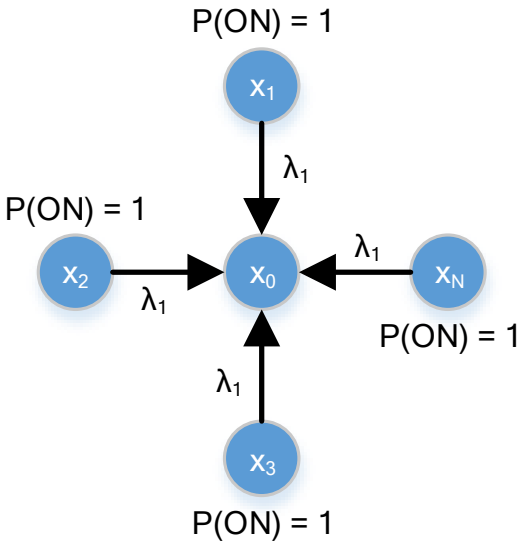
Then, each retained sensor $\mathbf{x}_i$ encodes its measurement v_i along with its location tag $\mathbf{x}_i$ into a packet, which is then modulated and transmitted.

By the end of the observation interval, the gathered data at the sink can be expressed as

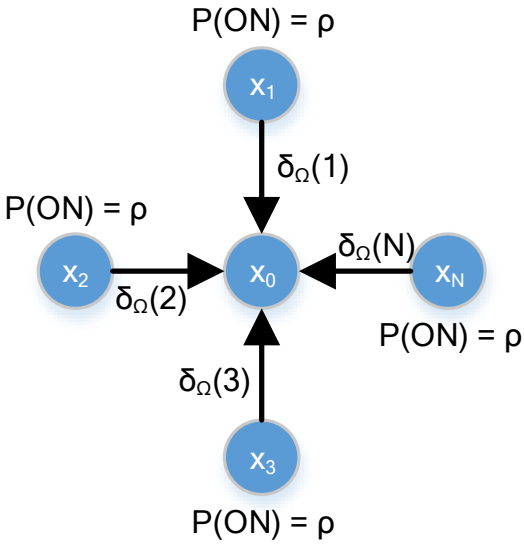
$$\mathbf{y} = \Phi\mathbf{v} + \epsilon, \tag{5.3}$$

where the matrix Φ models the collection of the packets, and ϵ represents the sensing noise which arises due to limitations in the sensing devices. Note that the sink can form the matrix Φ from the sample packets, since they carry the location tags.

In S-CDA, a small portion of the network is active and obtains permission to transmit using a CSMA-type protocol. This MAC protocol allows for the avoidance of collisions and ensures that the required m samples are received at the sink. Contrarily, in RACS [25], the whole network is active, and each sensor acquires and transmits measurements at a continuous rate λ_1 (measurements per second), independently of the other sensors. This rate controls which sensors are active at any particular time in an ALOHA access fashion. Figure 5.1 describes these two models. Contrary to other methods, RACS and S-CDA use the Fourier transform instead of a basis designed from the correlation of distances between sensors, i.e. RACS and S-CDA are not adaptive models.



(a) RACS



(b) S-CDA

Figure 5.1. Description of RACS and S-CDA. In each frame: in RACS, all sensors are active and transmit continuously at rate $\lambda_1 \gg 1$ (ALOHA fashion); in S-CDA, sensors are active with spatial frequency $\rho \ll 1$, and only the sensors with $\delta_\Omega(i) = 1$ are given access to transmit once per frame (CSMA fashion).

Sparse Recovery Mechanism

At the sink, the major frequency content is recovered from the gathered data using an Iterative Hard Thresholding (IHT) algorithm [12]. At iteration $j+1$, the IHT solves

$$\hat{\mathbf{v}}^{j+1} = H_k(\hat{\mathbf{v}}^j + \zeta\Omega^{\mathsf{T}}(\mathbf{y} - \Omega)\mathbf{v}^j) \tag{5.4}$$

where H_k is the hard thresholding operator, $(.)^{\mathsf{T}}$ denotes matrix conjugate, and compressed sensing matrix $\Omega = \Phi\Psi$. Basically, the operator H_k sets all but the k largest (in magnitude) elements in a vector to zero.

For computational analysis, since S-CDA preserves the randomness of the measurement process, Φ can be regarded as m rows of an $n \times n$ identity matrix, picked at random, i.e. each row consists of a single 1 in the position corresponding to the sensor contributing the useful packet. Since the signal $\mathbf{v}$ is assumed to be smooth and the measurements follow a PPP, the RDFT Ψ is the conjugate of the matrix for the Non-uniform Discrete Fourier transform of type 3 (NDFT-3) (with PPP distributed measurements), which is given by

$$\Psi = \frac{1}{n_1} z^{\mathbf{t}_1\mathbf{f}_1^{\mathsf{T}} + \mathbf{t}_2\mathbf{f}_2^{\mathsf{T}}}, \tag{5.5}$$

where $n_1 = \lceil\sqrt{n}\rceil$ (nearest upper integer), $z = e^{\frac{2\pi\sqrt{i}}{n_1}}$, measurement coordinates $\mathbf{t}_a = (\mathbf{x}_i(a))$, $a = 1, 2$, and frequency coordinates,

$$\mathbf{f}_1 = \otimes_{i=0}^{n_1-1} k\mathbf{1}_{n_1}, \tag{5.6}$$

$$\mathbf{f}_2 = \otimes_{i=0}^{n_1-1} [N_1] \quad (\text{with } [n_1] = [0, \ldots, n_1 - 1]) \tag{5.7}$$

where $\otimes$ denotes the concatenation of vectors. Finally, the sink builds a map of the sensing field as

$$\mathbf{v}^{\#} = \Psi\mathbf{w}^{\#}. \tag{5.8}$$

In a nutshell, first, S-CDA implements the random sensing of compressed sensing; then, based on CSMA, S-CDA collects the required number of packets for adequate recovery; and finally, S-CDA recovers the relevant frequency content and builds a map of the sensing field.

5.2.2 Compression Analysis

In RACS [25], the analysis of the data aggregation from equispaced sensors is based on the Discrete Fourier Transform (DFT), also denoted NDFT-1. It was claimed there that this model could be extended to

cases where the sensors are placed randomly, by replacing the DFT by the Non-Uniform Discrete Fourier Transform (NDFT). However, such an extension is not trivial. The major drawback of the NDFT with random measurements or frequencies is that an inversion formula, which should play the role of the sparsifying matrix Ψ generally does not exist [119]. Furthermore, a potential inversion via numerical methods will strongly depend on the position of measurements and allocation of frequencies.

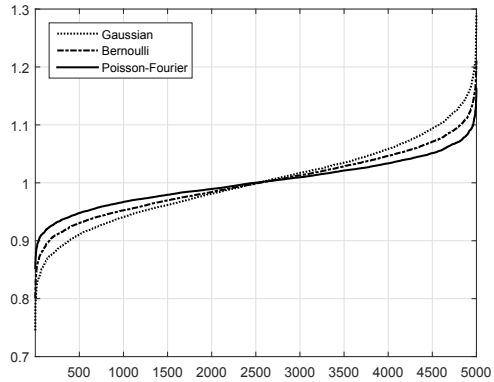
The NDFT represents the frequency content in different ways, according to the pair configuration (measurements, frequencies),

1. NDFT-1 or DFT: (equispaced, equispaced),
2. NDFT-2: (equispaced, irregular),
3. NDFT-3: (irregular, equispaced), and
4. NDFT-4: (irregular, irregular).

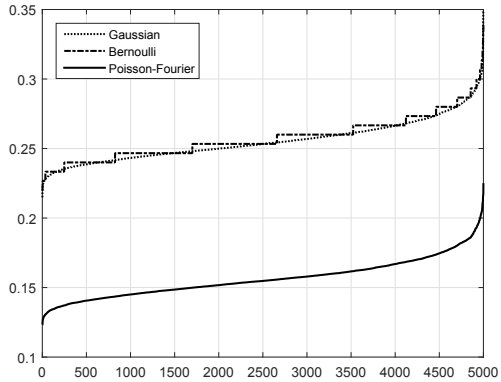
The frequency content from equispaced measurements in the NDFT-1 and NDFT-2 is periodic. In contrast, the frequency content from random measurements in the NDFT-3 and NDFT-4 is not periodic and exhibits random-like unwanted frequency components or *smeared aliasing*. Although this aliasing reduces the sparsity of the frequency content, recovery methods can perform (hard) denoising using thresholding-based algorithms, e.g. Accelerated Iterative Hard Thresholding (AIHT) [12].

Here, the PPP plays a major role. In addition to providing a model for random deployments, the randomness of the PPP makes the smeared aliasing appear as uncorrelated noise with no clear dominant alias frequencies [120]. Equispaced frequencies produce less aliasing; hence, the NDFT-3 with measurements following a PPP is selected as the sparsifying matrix Ψ. In the following three analyses are performed to study the PPP as a component of the matrix $\Omega = \Phi\Psi$.

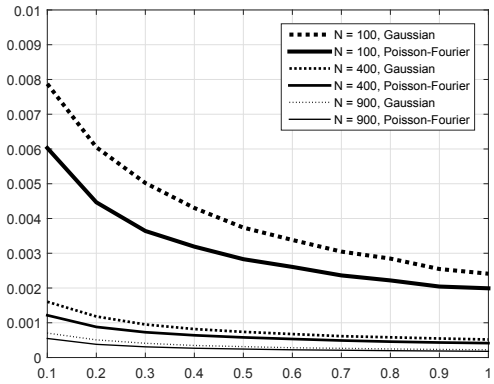
Moderately large experiments (see Figure 5.2(a)) indicate that the RDFT matrix produces a smaller RIC than the Gaussian and Bernoulli ensembles, which are two common compressed sensing matrices. They also indicate (see Figure 5.2(b)) that the RDFT possesses a smaller MIC than the Gaussian and Bernoulli ensembles. Figure 5.2(c) confirms this result by measuring the recovery error of the same matrices.



(a) RIC (*x-axis: realization index: y-axis: RIC+1*)

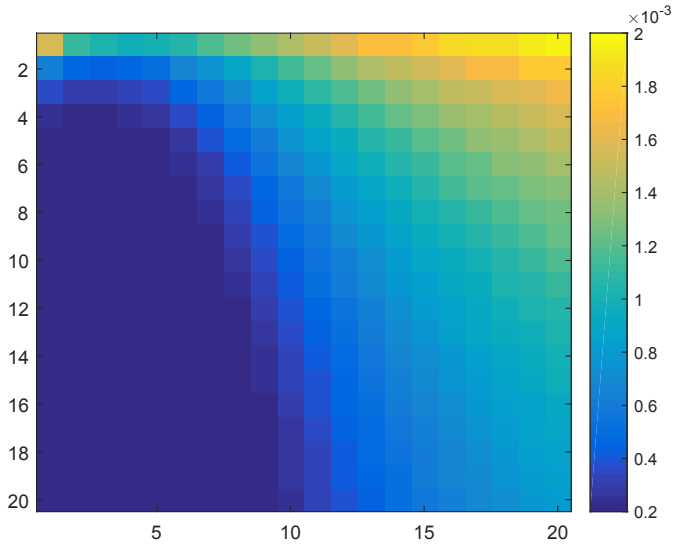


(b) MIC (*x-axis: realization index; y-axis: MIC*)

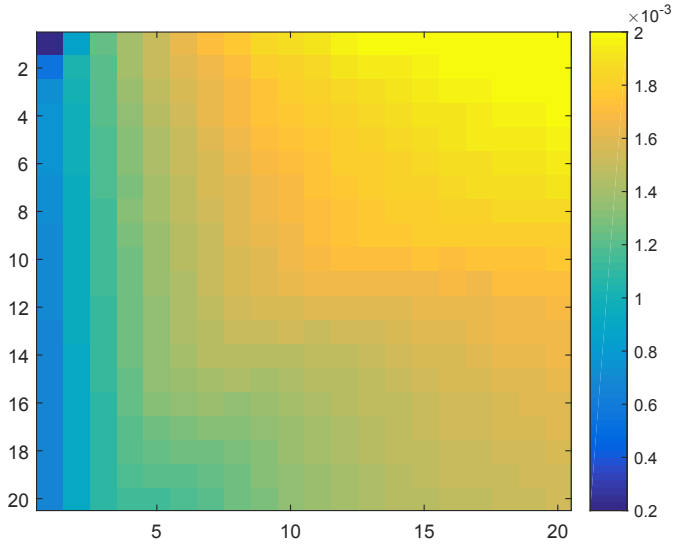


(c) Recovery error (*x-axis:* m/n*; y-axis: recovery error*)

Figure 5.2. Compression analysis of different compressed sensing matrices (over 5000 repetitions): (a) RIC computation; (b) MIC computation; and (c) mean recovery error (infinity-norm sense).



(a) FASTA (*x-axis:* m/n*; y-axis:* k/m)



(b) AIHT (*x-axis:* m/n*; y-axis:* k/m)

Figure 5.3. Phase transition-like analysis of RDFT using algorithm (a) FASTA and (b) AIHT. The $\frac{k}{m}$ is the sparsity fraction, $\frac{m}{n}$ is the undersampling fraction.

The Phase Transition

A phase transition analysis indicates the notorious change in the error recovery of a compressed sensing matrix Ω for all sparsity levels and compression ratios. Commonly, the probability of *perfect recovery* is adopted as the error metric. This probability is given by

$$\mathbb{P}(\textit{perfect recovery}) = \mathbb{P}(\|\mathbf{w} - \mathbf{w}^{\#}\|_{\infty} < 10^{-3}). \tag{5.9}$$

Figure 5.3 is generated by exhaustive experiments for two fast sparse recovery algorithms: Fast Adaptive Shrinkage/Thresholding Algorithm (FASTA) [121] (Figure 5.3(a)) and AIHT [12] (Figure 5.3(b)). Since AIHT recovers only the largest k frequency components, FASTA presents a slightly better performance. However, AIHT is more adequate for S-CDA given its hard denoising property.

5.2.3 Communication Analysis

This section finds the Medium Access Probability (MAP) of independent and uniformly distributed sensors using a CSMA-type protocol. The independence between the positions and time marks of sensors will allow the assumption of independence between the point processes defined by time marks. In addition, the effect of the interference is also computed. It will be shown that the effect of interference is negligible.

Consider the density of successful transmissions, i.e. the mean number of sensors that successfully transmit per square meter,

$$\lambda_{\mathsf{suc}} = \lambda_{\mathsf{on}} p_{\mathsf{map}}(\lambda_{\mathsf{on}}, \nu) p_{\mathsf{suc}}(\lambda_{\mathsf{on}} p_{\mathsf{map}}, \gamma), \tag{5.10}$$

where p_{map} and p_{suc} are the MAP and transmission success probability, respectively, and ν and γ are the carrier sensing and interference threshold parameters, respectively. This density characterizes not only the level of spatial (packing) density $\lambda_{\Omega} = \lambda_{\mathsf{on}} p_{\mathsf{map}}$ or intensity of Θ_{Ω}, but also the quality of transmissions through p_{suc}, which captures the interactions through interference among spatially distributed sensors [33].

The neighborhood or set of contenders of a sensor $\mathbf{x}_i \in \Theta_{\mathsf{on}}$ is defined as

$$\mathcal{N}_i = \{\mathbf{x}_j \in \Theta_{\mathsf{on}} \backslash \{\mathbf{x}_i\} : \xi_i \ell(\|\mathbf{x}_i - \mathbf{x}_j\|) \geq \nu\}. \tag{5.11}$$

This is a very simple contention criterion. It is similar to the so-called protocol model [122], except that fading is considered.

The event "sensor $\mathbf{x}_j$ does not contend with sensor $\mathbf{x}_i$" can be written as

$$u_j^i = \{\mathbf{x}_j \notin \mathcal{N}_i\}, \tag{5.12}$$

while the complementary event is $v_j^i = \{\mathbf{x}_j \in \mathcal{N}_i\}$. Then, the following result characterizes p_{map}.

Theorem 5.2.1 *Let Θ_{on} be a PPP with an intensity λ_{on} modeling a network whose sensors employ a CSMA-type protocol with time marks independent and uniformly distributed over $[0,1]$. The time-frequency with which a sensor is granted permission to transmit by the protocol, denoted $\mathbf{x}_i \in \Theta_\Omega$, given that sensors $\{\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}\}$ are simultaneously transmitting, is*

$$\mathbb{P}(\mathbf{x}_i \in \Theta_\Omega | \mathbf{x}_{[i-1]} \in \Theta_\Omega) \tag{5.13}$$

$$= \frac{1}{\lambda_{\mathsf{on}}} \int_0^{\lambda_{\mathsf{on}}} G_s(1 - \mathbb{P}(u_\cdot^{[i-1]}, v_\cdot^i))ds, \tag{5.14}$$

where the collection $[i-1] = \{1, 2, \ldots, i-1\}$, and $G_s(.)$ is the PGFL (Definition 2.2.1) of a PPP with intensity s.

Theorem 5.2.1 can be written in the more intuitive form.

Theorem 5.2.2 *Under the conditions of Theorem 5.2.1,*

$$\mathbb{P}(\mathbf{x}_i \in \Theta_\Omega | \mathbf{x}_{[i-1]} \in \Theta_\Omega) = \frac{\mathbb{P}(|\mathcal{N}_i \backslash \mathcal{N}_{[i-1]}| > 0)}{\mathbb{E}[|\mathcal{N}_i \backslash \mathcal{N}_{[i-1]}|]}. \tag{5.15}$$

which is the probability that a sensor $\mathbf{x}_i$ contends for the medium, divided by the expected effective number of contenders.

Theorem 5.2.1 provides a highly adaptive p_{map}, as sensors are entering Θ_Ω. However, that level of cognition is not necessary in S-CDA, as it would affect its random sensing. Then, for $i = 1$, the time-frequency for a sensor $\mathbf{x}_i$ to be granted permission to transmit is (cf. [52], eq. (5))

$$p_{\mathsf{map}} = \frac{1 - e^{-\lambda_{\mathsf{on}} c}}{\lambda_{\mathsf{on}} c}, \text{ with } c = \int_{\mathcal{B}} \mathbb{P}(v_z^1) dz, \tag{5.16}$$

where c is the mean neighborhood size normalized to unit density, and $\mathcal{B}$ is the network region, i.e. $\Theta \subset \mathcal{B}$. Then, the point process Θ_Ω can be understood as the thinning of the PPP Θ_{on} with a constant retention function p_{map}, i.e. the point process Θ_Ω is approximated by a PPP with an intensity $\lambda_\Omega = \lambda_{\mathsf{on}} p_{\mathsf{map}}$ (the modified Matérn thinning type II). A similar result holds for the Random Sequential Adsorption (RSA) model or Matérn type III point process with an infinite network area [123].[1]

Since the network geometry, and specifically the set of contenders, defines the integration domain of (5.16), closed-form expressions for the

[1] In [123], event $v_\cdot^1$ should be replaced by $u_\cdot^1$ as in (5.16).

p_{map} can be obtained. This holds especially for regular area networks, integrable path loss functions and fading distributions [41]. For instance, consider a sensor $\mathbf{x}_i$ located in the center of a circular sector neighborhood, with a radius R_i and and opening angle β_i (in radians). Further, assume that the channel of this sensor is characterized by the simplified attenuation path loss

$$\ell(\|\mathbf{x}_i - \mathbf{x}_j\|) = \|\mathbf{x}_i - \mathbf{x}_j\|^{-\alpha}, \tag{5.17}$$

and Rayleigh fading, i.e. ξ_i exponentially distributed with inverse mean parameter η,

$$\xi_i \sim \text{exponential}(\eta). \tag{5.18}$$

Then,

$$c = -\frac{\beta_i R_i^2}{\alpha} E_{\frac{\alpha-2}{\alpha}}(\eta\nu R_i^{\alpha}), \tag{5.19}$$

where $E_c(z)$ is the exponential integral function, $E_c(z) = \int_1^{\infty} \frac{\exp(-zy)}{y^n} dy$. This expression suggests that the p_{map} will decrease as the number of potential contenders increases with R_i, and it will increase as the mean transmitted power decreases, i.e. the effect of the contenders vanishes when η increases [41].

Following the random sensing of S-CDA, take $\beta_i = 2\pi$ and $R_i \to \infty$ (large deployment configuration [33]), $\forall i$,

$$c(\nu) = \frac{2\pi\Gamma(\frac{2}{\alpha})}{\alpha(\eta\nu)^{\frac{2}{\alpha}}}, \tag{5.20}$$

which gives the spatial density of S-CDA,

$$\lambda_{\Omega} = \frac{\alpha(\eta\nu)^{\frac{2}{\alpha}}}{2\pi\Gamma(\frac{2}{\alpha})} \left(1 - e^{-\lambda_{\mathsf{on}} \frac{2\pi\Gamma(\frac{2}{\alpha})}{\alpha(\eta\nu)^{\frac{2}{\alpha}}}}\right). \tag{5.21}$$

Note that this lower bound of the global density of the network allows a uniform sensing of the field.

Now, using the simple relations,

$$\frac{\lambda_{\Omega}}{m} = \frac{\lambda}{n}, \tag{5.22}$$

$$m \propto \delta_{\mathsf{MIC}} k \log n, \tag{5.23}$$

where the $\delta_{\mathsf{MIC}} = 1$ since RDFT is a Fourier transform; and together with (5.21) and (5.20), the parameters, ρ and ν, can be optimized from,

$$\frac{1 - e^{-\rho\lambda c(\nu)}}{c(\nu)} = \frac{\lambda}{n} c_{\Omega} k \log n, \tag{5.24}$$

where c_{Ω} is a constant which depends mainly on the compressed sensing matrix, Ω. This can be found empirically as in [25], then $c_{\Omega} \in [3, 5]$.

Figure 5.4 plots the level sets of both the RHS and LHS of (5.24). For instance, for given density, λ; if the dimension, $n \approx 6000$, and the estimated sparsity, $k \approx 25$, then, the level set in Figure 5.4(a) is approx. 0.2. This means that $\rho \in [0.2, 0.3]$ in Figure 5.4(b). Now, since it is advisable to have less active sensors, $\rho \approx 0.2$. Finally, a carrier sensing threshold, $\nu \approx 60$ will do achieve to transmit $m \approx 900$ samples, to guarantee the recovery of signal $\mathbf{v}$.

Now interference is discussed in the model. Thereupon, the transmission success probability of a receiver associated with a typical active transmitter is computed. This is equivalent to the success probability of the (now) transmitter $\mathbf{x}_0$, given $\mathbf{x}_0 \in \Theta_\Omega$, with the receiver at $(0, r)$, which under independent fading leads to

$$I^{!}_{\Theta_\Omega} = \sum_{m:\mathbf{x}_j \in \Theta_\Omega \setminus \{\mathbf{x}_0\}} \xi_j \|\mathbf{x}_j - (0, r)\|^{-\alpha}. \tag{5.25}$$

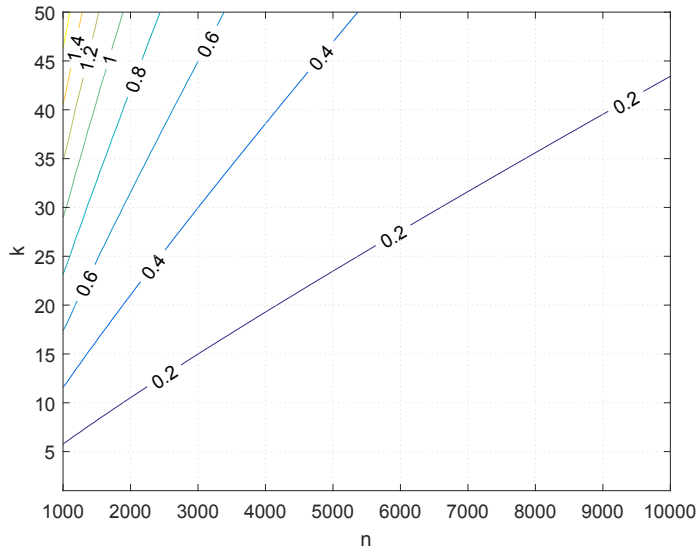
Then,

$$p_{\mathsf{suc}} = \mathbb{E}_{\xi, r}\left[\mathbb{P}^{!}\left(I^{!}_{\Theta_\Omega} < \frac{\xi r^{-\alpha}}{\bar{\gamma}}\right) | \xi, r\right]. \tag{5.26}$$

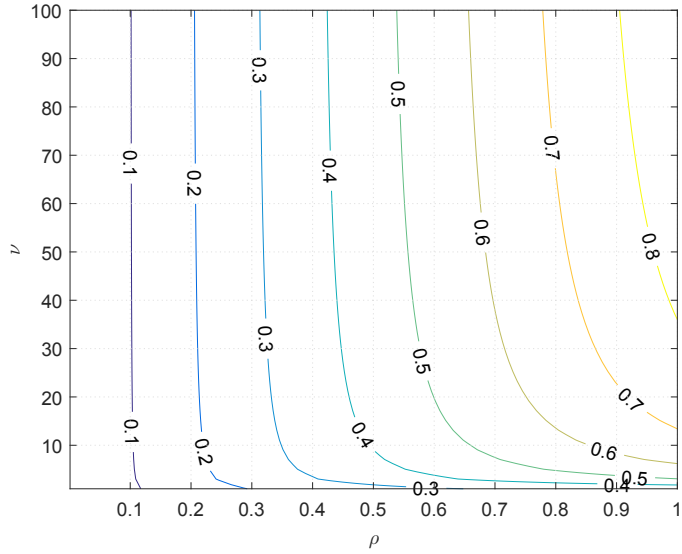
Note that $I^{!}_{\Theta_\Omega}$ is the interference from a point process induced by the CSMA protocol and potentially other communication mechanisms, which has a dependency among sensor locations. It is called the Matérn CSMA process. In [33], this interference is approximated by a shot noise from a (distance-dependent) non-homogeneous PPP.

However, the resulting expression is difficult to compute. Moreover, recent research reveals that under general assumptions, the interference from many kinds of networks will present a heavy tail and be strongly asymmetric and far from normal [38]. Then, methods need to rely on numerical approximations in order to gain some insight [93].

The method based on asymptotic expansions described in Section 4 is adopted to obtain numerically the outage probability. As such, Figure 5.5 shows the results for different interference thresholds. The effect of interference in the number of successfully received packets (in terms of the resulting density after a MAC realization) is low, but this increases with the density λ and carrier sensing ν as the size of neighborhoods decreases and more sensors are retained by the protocol.

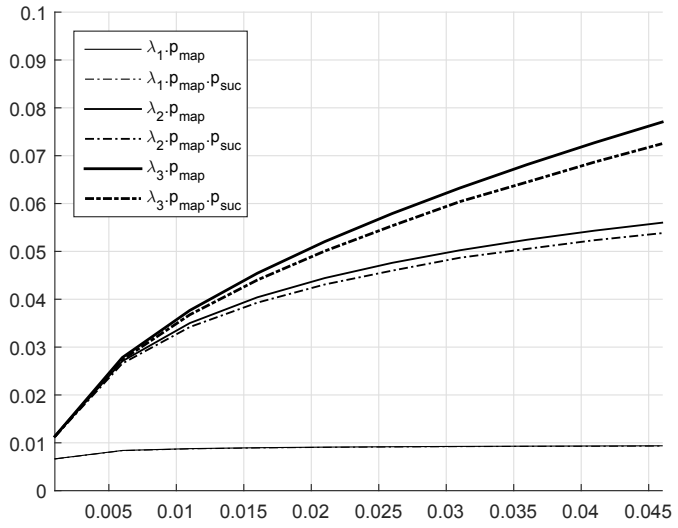


(a) LHS of (5.24) (*x-axis: dimension n; y-axis: sparsity k*)

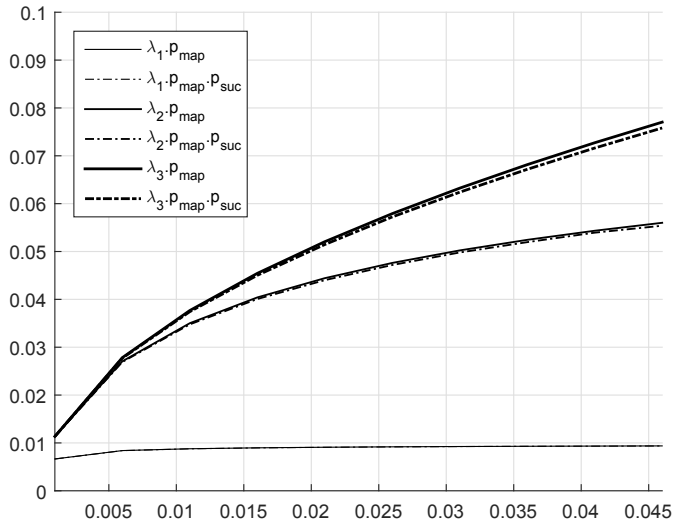


(b) RHS of (5.24) (*x-axis: activation rate ρ; y-axis: carrier threshold ν*)

Figure 5.4. Communication analysis. The level sets of both the RHS and LHS of (5.24). For instance, for given density, λ; if the dimension, $n \approx 6000$, and the estimated sparsity, $k \approx 25$, then, the value of the LHS is approximately equal to 0.2. This means that $\rho \in [0.2, 0.3]$ in the RHS. Now, since it is advisable to have less active sensors, $\rho \approx 0.2$. Finally, a carrier sensing threshold, $\nu \approx 60$ will do achieve to transmit $m \approx 900$ samples, to guarantee the recovery of signal, v.



(a) Density λ_Ω with threshold $\gamma = 1.0$ (*x-axis: carrier sensing* ν)



(b) Density λ_Ω with threshold $\gamma = 2.0$ (*x-axis: carrier sensing* ν)

Figure 5.5. Density of point process Θ_Ω and the effect of interference in terms of the carrier sensing threshold ν, for different densities ($\lambda_1 = 0.01$, $\lambda_2 = 0.1$, and $\lambda_3 = 1$), and interference threshold levels γ.

5.3 Discussion

This section discusses a key technical assumption of S-CDA and the implications of the random sensing of compressed sensing in other advanced communication strategies.

5.3.1 Sparsity Tracking

The sparse recovery mechanism of S-CDA is based on AIHT, a thresholding algorithm, which sets all but the k largest (in magnitude) coefficients to zero. Hence, a key assumption of S-CDA is the knowledge in real-time of the sparsity level of $\mathbf{w}$.

Sparsity tracking is achieved in two steps using core sparsity (with parameters $p = 1$, $q = 2$), the NDFT-3 and results from [89]:

1. First, the initial sparsity level k_0 is computed using packets $\mathbf{v}_0$ from all the sensors or until a desired spatial fidelity is reached. Packets from the whole network build the sensing field with the highest possible resolution and provide major details on the frequency content, $\mathbf{w}_0$,

$$\mathbf{v}_0 \overset{\Omega_0}{\mapsto} \mathbf{w}_0 \mapsto k_0 = -\frac{\|\mathbf{w}_0\|_1}{\|\mathbf{w}_0\|_2}. \tag{5.27}$$

   For less demanding applications, partial measurements can be considered.

2. Second, the sparsity level k_Ω is estimated by computing the sparsity of the coefficients $\mathbf{w}_\Omega$ from the gathered packets $\mathbf{v}_\Omega$ [89],

$$\mathbf{v}_\Omega \overset{\Omega}{\mapsto} \mathbf{w}_\Omega \mapsto k_\Omega = -\frac{\|\mathbf{w}_\Omega\|_1}{\|\mathbf{w}_\Omega\|_2}. \tag{5.28}$$

Figure 5.6 illustrates the results (normalized to [0,1]) of the experiments for $n = 256$, $m = 5k \log n$, and $k = 1, \ldots, 9$. The results indicate that core sparsity slightly overestimates the sparsity approximately by one effective additional component, i.e. $k_\Omega \approx k + 1$. This additional component could be removed, but it will not cause harm to the recovery process since AIHT is robust to a slight overestimation of k.

5.3.2 A Remark on Correlated Signals and Channels

The correlation of signals can be utilized to design suitable sparsifying matrices, e.g. treelets and diffusion wavelets. These matrices could

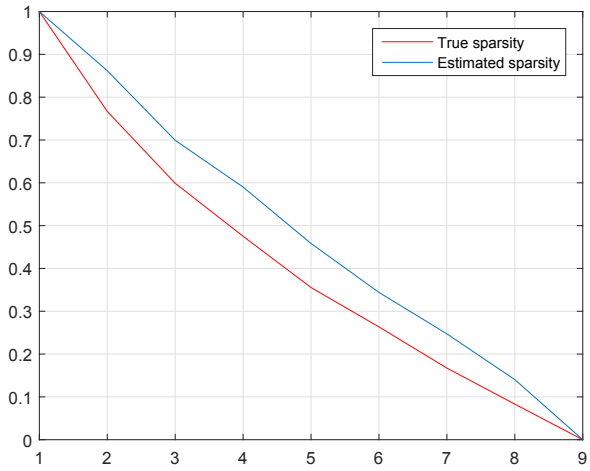


Figure 5.6. Sparsity estimation and tracking using [0,1]-normalized core sparsity k, with parameters $p = 1$ and $q = 2$, in terms of the number of frequency components k (*x-axis: number of components (strict sparsity); y-axis: core sparsity* k).

improve the fidelity of representation and hence recovery of signals. However, they will also require a high level of control to which the associated transmission will provoke major energy consumption.

The correlation of channels can be used to design advanced MAC protocols supported by the cognition capabilities (channel ranking) of the sensors, e.g. opportunistic-ALOHA [124] or opportunistic-CSMA [33]. However, this level of cognition would break the random sensing of compressed sensing. Note that ALOHA also introduces correlation due to channels. For instance, sensors with the best channels may correspond to the nearest sensors to the sink. If transmission is only granted to these sensors, then the information from further locations will be never received. This would prohibit the recovery of the entire sensing field. A similar case which illustrates this scenario occurs in matrix completion [125]. For a 3×3 matrix, if no information is known about one row (or column) of the matrix, then the recovery of that row (or column) is not possible.

5.4 Summary

This chapter introduced S-CDA and developed a framework for its joint compression and communication evaluation. The framework is based on the randomness of the sampling process drawn by S-CDA. S-CDA per-

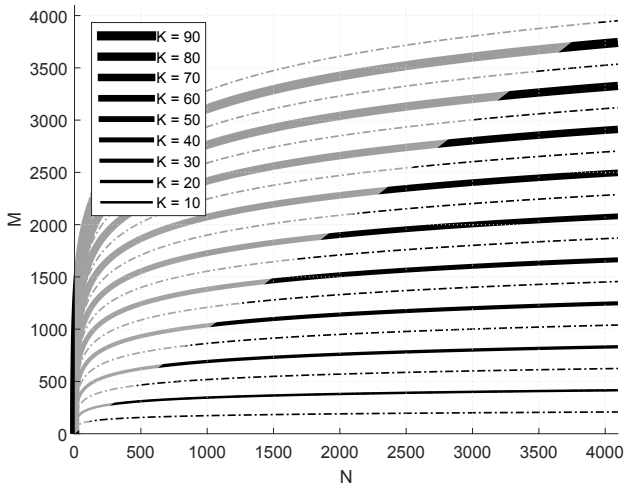
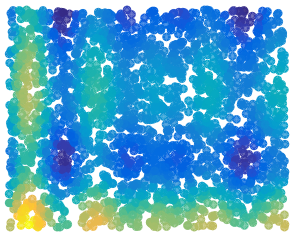
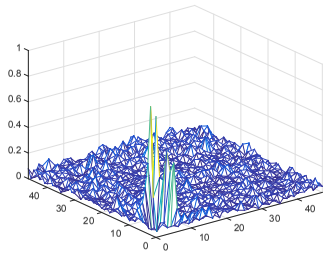


Figure 5.7. The number of samples m in terms of the dimension n, for different frequency components k (*x-axis: n; y-axis: m*).

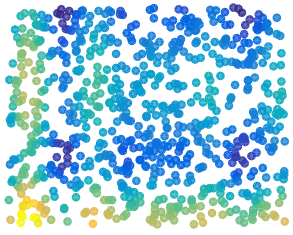
forms a thinning to the PPP which models the WSN. The resulting point process is no longer a PPP; however, it can be approximated by a new PPP with a different intensity measure. This new PPP defines the sparsifying domain from the non-uniform Fourier transform. As a result, the optimal number of samples m is of the form $5k\log(n)$. This relation between the required number of samples m, the sparsity k and the dimension of the sensing field n is found empirically as in [25]. This relationship is described in Figure 5.7. In the case of a network distributed as a regular grid, i.e. where the sparsifying basis is the uniform Fourier transform, the number of required samples m is $4k\log(n)$ [25]. The additional samples in the non-uniform case is the price for including the random network model, which defines a non-invertible basis. Figure 5.8 shows and example of S-CDA. In this experiment, a network of $n = 4000$ sensors is randomly deployed to monitor a sensing field of $k = 20$ frequency components. The required number of samples for recovery is computed and it is approximately $m = 750$ samples, which means that less than 20% of the WSN should be active and contend for the medium.



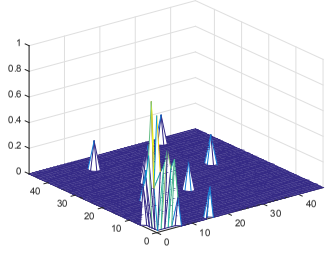
(a) Ground signal $\mathbf{v}$



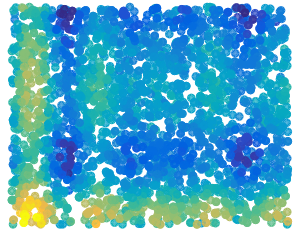
(b) True coefficients $\mathbf{w}$



(c) Sampled signal $\mathbf{y}$



(d) Recovered coefficients $\mathbf{w}^{\#}$



(e) Recovered signal $\mathbf{v}^{\#}$

Figure 5.8. Example of S-CDA in action. A large WSN is randomly deployed to monitor a sensing field (a) of sparse representation in frequency (b). In S-CDA, a small portion of the network is active and have permission to transmit (c). Using the sample signal, S-CDA recovers the frequency content (d) and the sensing field (e).

6. Conclusions and Future Work

6.1 Conclusions

There is considerable recent interest in sparse and heavy-tail distributed signals. In the context of this thesis, sparse signals characterize common real data (in some suitable domain) and heavy-tail distributed signals characterize interference traces in communication networks. Consequently, this thesis studied these two types of signals with the aim of developing a data aggregation protocol for WSN under random deployment.

A complete characterization of sparsity and entropy offers new functions and tools to solve efficiently sparse recovery problems. To clarify this, Chapter 3 proposed a formalism, a joint axiomatic characterization for sparsity and entropy. The proposed set of axioms is constructive and allows for the derivation of *core functions*, core sparsity, and core entropy. Both core functions consist of functional forms of well-known sparsity and entropy functions.

Then, Chapter 4 extended the usability of the Edgeworth and Cornish-Fisher expansions to analogous asymptotic expansions, denoted as EEL and GCFEL, which are based on elements of a formalism called log-statistics. The proposed expansions inherit the capability of log-statistics to characterize heavy-tailed distributions. It is important to note that the relationships of classic statistics remain valid for log-statistics; hence, the extension of other traditional methods should be straightforward. As the main objective of the chapter, the EEL and GCFEL are successfully applied to characterize interference traces drawn using models from stochastic geometry.

Finally, Chapter 5 introduced S-CDA and developed a framework for

its joint compression and communication analysis. This chapter studied the implications of the random sensing of compressed sensing in the data aggregation scenario. Techniques from stochastic geometry, compressed sensing and asymptotic expansions are employed to develop S-CDA. The techniques from stochastic geometry allow for the analysis of the network configuration (location of sensors) and channel models to design and evaluate a MAC protocol. This protocol will ensure the arrival of a given number of packets to reach a desired level of fidelity. This number is described in terms of the density of the resulting network after the application of the protocol. A posterior interference computation shows that the effect of interference is negligible on the effective density. This computation could be utilized to slightly refine the density found by S-CDA and deliver a protocol, that does not only avoid collisions, but that is also robust to interference. Hereafter, the thesis moves on to the compression analysis where, using the location of the network, a compressed sensing matrix is constructed. This matrix satisfies key criteria reported in the compressed sensing literature. With this matrix, which is dependent on the positions of the sensors, the estimated sparsity, and the samples collected, a thresholding algorithm recovers the desired signal. Although the protocol is described in the single-hop scenario, the results directly translate into the multi-hop case. In this case, instead of sending the packages to the sink, the sensors transmit their data to other sensors along the WSN.

6.2 On the Assumptions and Technical Limitations

S-CDA is closely related to WCS [24], RACS [25], ME-CDA [22], and TC-CDA [23]. These protocols are based on compressed sensing, and they all partly perform a compression or communication analysis of CDA. From these, it is possible to recognize a few technical limitations of the models adopted by S-CDA.

On the compression side, the limitation of S-CDA is that its sparsifying domain is given by the non-uniform Fourier transform. This means that the sensing field should be sparse in the frequency domain. Although this holds for a large set of smooth fields, less structured sensing fields or ones varying with time will require a very adaptive sparsifying basis. This may reduce the number of samples by offering suitable transformation domains at the price of additional control packets.

On the communication side, the limitation of S-CDA is the PPP network model. Although this is well justified as sensors can be considered randomly deployed, the nature of the sensing field could demand different deployments. For instance, clustered models would help model monitoring scenarios where highly detailed information is required in specific areas. However, for clustered models, the smeared aliasing can not be considered as uncorrelated noise.

6.3 Future Work and Extensions

This section briefly discusses the directions for future work. Two technical results were developed, sparsity functions and asymptotic expansions for distributions, both of which were adopted to develop a data aggregation protocol for WSNs.

Concerning core sparsity functions, future work will need to study these functions in a general optimization framework, where they can be "tuned" to match the optimal recovery strategy according to the sparsity level. Concerning core entropy, the compatibility of the proposed axioms and specific axioms of entropy, especially those related to relative and conditional entropy, should be studied. Some of these specific axioms were already verified here.

Concerning asymptotic expansions, future work will need to study the use of these expansions to characterize interference in terms of the network parameters. Much of the current results on interference analysis, especially from the stochastic geometry perspective, adopt light-tailed models. Although these models may estimate and adjust the parameters of interference, it does not mean that the estimation of the distribution is correct. This is especially true, since it has been reported that interference in a general setting will be heavy-tailed. Then, the proposed expansion would help to provide insight on network performance using adequate models.

Finally, the proposed S-CDA developed in this thesis is based on the random sensing of compressed sensing. However, recent research indicates that the deterministic and careful design of compressed sensing matrices (and the use of correlated models) would slightly improve the efficiency by reducing the number of samples. These new proposals will add large complexity to all the aspects of data aggregation, such as the design of advanced MAC protocols and of compressed sensing matrices.

In the data aggregation scenario, the deterministic process refers to the selection of specific sensors or rows of the RDFT (or measurements) such that the MIC is minimized. This opens new paths of research that search compressed sensing matrices that derive more advanced MAC protocols not necessarily leading to random sensing. In this case, the level of fairness or non-sparsity of the network with respect to the energy consumption should be studied.

A. Appendix

A.1 Proofs of Chapter 3

In this section Lemmas 3.3.1 to 3.3.8 are proved, which derive attributes 3.3.1 to 3.3.8 from axioms 3.2.1 to 3.2.5. The section also contains the proofs listed in Tables 3.4 and 3.6 where functions, applied to non-zero vectors, are evaluated using axioms and attributes of Section 3.2. These proofs give a refinement of the formalism compared to proofs of [17] of which some proofs are corrected.

Tables 1.1 and 1.2 describe the notation adopted in the proofs related to sparsity and entropy functions, respectively. The proofs are based on the following key results of [70].

Theorem A.1.1 *[70] Let* f, g *be defined on* $\mathcal{W}$ *convex, and*

$$\mathsf{s}(\mathbf{w}) = \frac{\mathsf{f}(\mathbf{w})}{\mathsf{g}(\mathbf{w})}. \tag{1.1}$$

If any of the following properties hold

1. f *non-negative convex,* g *positive concave;*
2. f *non-positive convex,* g *positive convex;*
3. f *convex,* g *positive affine,*

then, s *is semi-strictly quasi-convex on* $\mathcal{W}$.

Theorem A.1.2 *[70] (also [61]) Let* s *be semi-strictly quasi-convex (quasi-concave) defined on* $\mathcal{W} \subseteq \mathbb{R}^n$ *convex. Let* $\mathsf{a} : A \to \mathbb{R}$ *be increasing, with* $\mathsf{s}(\mathcal{W}) \subseteq A$. *Then,* $\mathsf{a} \circ \mathsf{s}$ *is semi-strictly quasi-convex (-concave) on* $\mathcal{W}$, *where* $\circ$ *denotes composition of functions.*

Examples of function $\mathsf{a}(.)$ are the $\log(.)$ and $-\frac{1}{.}$ (which changes the sign and allows to prove Lemma 3.3.8).

A.1.1 Proofs of Lemmas 3.3.1 to 3.4.4

Proof A.1.3 *[Lemma 3.3.1] Let* $\mathbf{w}$ *arbitrary with* $\|\mathbf{w}\|_1 = 1$. *Construct* $\hat{\mathbf{w}} = \mathbf{0}_{n-1}\|1$ *and* $\tilde{\mathbf{w}} = \frac{1}{n}\mathbf{1}_n$ *from* $\mathbf{w}$ *via concentration actions 3.2.3. Then,*

$$\mathsf{s}(\tilde{\mathbf{w}}) \overset{3.2.3}{\leq} \mathsf{s}(\mathbf{w}) \overset{3.2.3}{\leq} \mathsf{s}(\hat{\mathbf{w}}). \tag{1.2}$$

and by scaling 3.2.4,

$$\mathsf{s}(\alpha\mathbf{1}_n) \overset{3.2.4}{=} \mathsf{s}(\tilde{\mathbf{w}}), \tag{1.3}$$

$$\mathsf{s}(\hat{\mathbf{w}}) \overset{3.2.4}{=} \mathsf{s}(\mathbf{0}_{n-1}\|\alpha). \tag{1.4}$$

■

Proof A.1.4 *[Lemma 3.3.2] Let* s *continuous 3.2.1, symmetric 3.2.2, scale-invariant 3.2.4 and non-negative. Then,* s *can be written as* $\frac{\mathsf{f}}{\mathsf{g}}$, *with* f *and* g *continuous, symmetric, homogeneous of degree 1 and non-negative. Construct* $\tilde{\mathbf{w}}$ *from* $\mathbf{w}$ *via a concentration action 3.2.3. Then, it holds that*

$$\mathsf{f}(\tilde{\mathbf{w}}) > \mathsf{f}(\mathbf{w}), \quad \mathsf{g}(\tilde{\mathbf{w}}) < \mathsf{g}(\mathbf{w}). \tag{1.5}$$

For function f *differentiable*

$$\mathsf{f}(\mathbf{w}) - \mathsf{f}(\tilde{\mathbf{w}}) \approx -\alpha\nabla_{\mathbf{e_{i_1}}-\mathbf{e}_{i_2}}\mathsf{f}(\mathbf{w}) \tag{1.6}$$

$$= \alpha\left(\frac{\partial \mathsf{f}}{\partial w_{i_2}}(\mathbf{w}) - \frac{\partial \mathsf{f}}{\partial w_{i_1}}(\mathbf{w})\right). \tag{1.7}$$

Then, by the symmetry of f *and (1.5), each component function* f_k *of* f *satisfies*

$$\lim_{\alpha\to 0}\frac{f_k(w_{i_2}) - f_k(w_{i_2}+\alpha)}{\alpha} < \lim_{\alpha\to 0}\frac{f_k(w_{i_1}) - f_k(w_{i_1}+\alpha)}{\alpha} \tag{1.8}$$

with $w_{i_1} > w_{i_2}$, *which implies that* f *is convex. Similarly,* g *is concave. Now, by Theorem A.1.1,* $\mathsf{s} = \frac{\mathsf{f}}{\mathsf{g}}$ *is semi-strictly quasi-convex. A similar proof holds if* s *non-positive.*

■

Proof A.1.5 *[Lemma 3.3.3] Construct* $\tilde{\mathbf{w}} = (\tilde{w}, \alpha - \tilde{w})$ *from* $\mathbf{w} = (w, \alpha - w)$ *via concentration actions 3.2.3, with* $\frac{\alpha}{2} \leq w \leq \tilde{w} \leq \alpha$.

■

Proof A.1.6 *[Lemma 3.3.4] [17] Let* $\mathbf{w}$ *arbitrary. Then,*

$$\mathsf{s}(\mathbf{w}) \overset{3.2.5}{=} \mathsf{s}(\underbrace{\mathbf{w}\|\ldots\|\mathbf{w}}_{k+1\,\text{times}}) \tag{1.9}$$

$$\overset{3.2.3}{<} s\left(\left(1+\frac{1}{k}\right)\mathbf{w}\|\ldots\|\mathbf{w}\|\mathbf{0}_k\right) \tag{1.10}$$

$$\overset{3.2.4}{=} \mathsf{s}(\mathbf{w}\|\ldots\|\mathbf{w}\|\mathbf{0}_k) \tag{1.11}$$

$$\overset{3.2.5}{=} \mathsf{s}(\mathbf{w}\|0). \tag{1.12}$$

■

Proof A.1.7 *[Lemma 3.3.5] [17] Let* $\tilde{\mathbf{w}} = [w_1,\ldots,w_{i_1}+\beta,\ldots,w_n]$*. Construct* $\hat{\mathbf{w}}$ *from* $\mathbf{w} = \epsilon\tilde{\mathbf{w}}$*,* $0 < \epsilon = 1+\delta$*, via concentration actions 3.2.3,*

$$\hat{w}_{i_2} = \epsilon w_{i_2} - \delta w_{i_2} = w_{i_2},\ i_2 \neq i_1, \tag{1.13}$$

$$\hat{w}_{i_1} = \epsilon(w_{i_1}+\beta) + \sum_{i_2\neq i_1}\delta w_{i_2} \tag{1.14}$$

$$= w_{i_1} + \beta + \alpha, \tag{1.15}$$

where $\alpha = \delta(\beta + \sum_i w_i)$*. Then,*

$$\mathsf{s}(\mathbf{w}) \overset{3.2.4}{=} \mathsf{s}(\tilde{\mathbf{w}}) \tag{1.16}$$

$$\overset{3.2.3}{<} \mathsf{s}(\hat{\mathbf{w}}). \tag{1.17}$$

■

Proof A.1.8 *[Lemma 3.3.6] Let* $\mathbf{w}$ *arbitrary,* $\tilde{\mathbf{w}} = \mathbf{1}_n$ *and* $\alpha < 1$*. Then,*

$$s\left(\mathbf{w} + \frac{1-\alpha}{\alpha}\tilde{\mathbf{w}}\right) \overset{3.2.4}{=} \mathsf{s}(\alpha\mathbf{w} + (1-\alpha)\tilde{\mathbf{w}}) \tag{1.18}$$

$$\overset{3.3.2}{<} \max\{\mathsf{s}(\mathbf{w}), \mathsf{s}(\tilde{\mathbf{w}})\} \tag{1.19}$$

$$\overset{3.3.1}{=} \mathsf{s}(\mathbf{w}). \tag{1.20}$$

■

Proof A.1.9 *[Lemma 3.3.8] Let* $\mathbf{w}$ *and* $\tilde{\mathbf{w}}$ *arbitrary and* $\alpha < 1$*,*

$$\mathsf{s}(\alpha\mathbf{w} + (1-\alpha)\tilde{\mathbf{w}}) \overset{3.3.2}{\leq} \max\{\mathsf{s}(\mathbf{w}), \mathsf{s}(\tilde{\mathbf{w}})\} \tag{1.21}$$

$$\overset{s>0}{\leq} \mathsf{s}(\mathbf{w}) + \mathsf{s}(\tilde{\mathbf{w}}) \tag{1.22}$$

$$\overset{3.2.4}{=} \mathsf{s}(\alpha\mathbf{w}) + \mathsf{s}((1-\alpha)\tilde{\mathbf{w}}). \tag{1.23}$$

■

Proof A.1.10 *[Lemmas 3.4.3 and 3.4.4] Lemmas 3.4.3 is trivial. Then,*

$$\mathsf{h}_{pq}^{b_1}(\mathbf{1}_{b_1}) \overset{3.12}{=} \mathsf{h}_{pq}^{b_2}(\mathbf{1}_{b_1}\|\mathbf{0}_{b_2-b_1}) \tag{1.24}$$

$$\overset{3.3.4}{<} \mathsf{h}_{pq}^{b_2}(\mathbf{1}_{b_2}). \tag{1.25}$$

■

A.1.2 Proofs of Theorems 3.4.1 and 3.4.2

Proof A.1.11 *[Theorems 3.4.1 and 3.4.2] Since concentration 3.2.3 implies quasi-convexity 3.3.2, Theorems A.1.1 and A.1.2, with* $\mathsf{a}(.) = -\frac{1}{.}$ *increasing, state that sparsity* s *can be written as*

$$\mathsf{s}(\mathbf{w}) = -\mathsf{d}(n)\frac{\mathsf{g}(\mathbf{w})}{\mathsf{f}(\mathbf{w})} \tag{1.26}$$

with $\mathsf{d}(n) > 0$, $\forall n$, *responsible of the dimension-invariance of* s *or replication 3.2.5; and* $\mathsf{g} > 0$ *concave and* $\mathsf{f} > 0$ *convex, and both functions continuous 3.2.1, symmetric 3.2.2 and homogeneous of degree 1, such that* s *is scale-invariant 3.2.4. The previous description of* f *and* g *resembles vector norms and leads to the pair of candidate functions*

$$\mathsf{f}(.) = \|.\|_q,\ q \geq 1, \tag{1.27}$$

$$\mathsf{g}(.) = \|.\|_p,\ p \leq 1, \tag{1.28}$$

with $p \neq q$. *Finally, normalization by* $\mathsf{d}(n) = n^{\frac{1}{q}-\frac{1}{p}}$ *enables* s *to satisfy replication 3.2.5. A similar proof holds for core entropy.*

■

A.1.3 Evaluation of Sparsity Functions

Proof A.1.12 *[Continuity, symmetry] By the properties of* ℓ_p*-norms.*

■

Proofs for "*ℓ_0*" (denoted* s_0*)

Proof A.1.13 *[Concentration] (cf. correction of [17].)*

$$\|\tilde{\mathbf{w}}\|_0 = \begin{cases} \|\mathbf{w}\|_0 - 1, & \textit{new null component,} \\ \|\mathbf{w}\|_0, & \textit{otherwise,} \end{cases} \tag{1.29}$$

i.e. $\mathsf{s}_0(\tilde{\mathbf{w}}) \geq \mathsf{s}_0(\mathbf{w})$.

■

Proof A.1.14 *[Scaling] (cf. correction of [17].)*

$$\|\tilde{\mathbf{w}}\|_0 = \|\mathbf{w}\|_0, \tag{1.30}$$

i.e. $\mathsf{s}_0(\tilde{\mathbf{w}}) = \mathsf{s}_0(\mathbf{w})$.

■

Table 1.1. Summary of notation used in proofs related to sparsity functions. Operator π denotes permutation.

	$\tilde{\mathbf{w}}$	$\tilde{n}$	$\hat{\mathbf{w}}$	$\hat{n}$	Assumptions	Desired result
Continuity		n			$\tilde{\mathbf{w}} \to \mathbf{w}$	$\mathsf{s}(\tilde{\mathbf{w}}) \to \mathsf{s}(\mathbf{w})$
Symmetry	$\pi(\mathbf{w})$	n				$\mathsf{s}(\tilde{\mathbf{w}}) = \mathsf{s}(\mathbf{w})$
Concentration	$\tilde{w}_{i_1} = w_{i_1} + \alpha$, $\tilde{w}_{i_2} = w_{i_2} - \alpha$	n			$w_{i_1} \geq w_{i_2}$	$\mathsf{s}(\tilde{\mathbf{w}}) > \mathsf{s}(\mathbf{w})$
Scaling	$\alpha\mathbf{w}$	n				$\mathsf{s}(\tilde{\mathbf{w}}) = \mathsf{s}(\mathbf{w})$
Replication	$\mathbf{w} \Vert \ldots \Vert \mathbf{w}$	kn			$n \to \infty$	$\mathsf{s}(\tilde{\mathbf{w}}) \to \mathsf{s}(\mathbf{w})$
Bounds	$\alpha \mathbf{1}_n$	n	$\mathbf{0}_{n-1} \Vert \alpha$	n		$\mathsf{s}(\tilde{\mathbf{w}}) < \mathsf{s}(\hat{\mathbf{w}})$
Quasi-convex		n	$\alpha\mathbf{w} + (1-\alpha)\tilde{\mathbf{w}}$	n	$\alpha < 1$, $\mathsf{s}(\mathbf{w}) \leq \mathsf{s}(\tilde{\mathbf{w}})$	$\mathsf{s}(\tilde{\mathbf{w}}) > \mathsf{s}(\hat{\mathbf{w}})$
Monotonicity	$[\tilde{w}, \alpha - \tilde{w}]$	2	$[\hat{w}, \alpha - \hat{w}]$	2	$\frac{\alpha}{2} \leq \tilde{w} < \hat{w} \leq \alpha$	$\mathsf{s}(\tilde{\mathbf{w}}) < \mathsf{s}(\hat{\mathbf{w}})$
Completeness	$\mathbf{w} \Vert 0$	$n+1$				$\mathsf{s}(\tilde{\mathbf{w}}) > \mathsf{s}(\mathbf{w})$
Regularity	$\tilde{w}_{i_1} = w_{i_1} + \beta$	n	$\hat{w}_{i_1} = w_{i_1} + \beta + \alpha$	n		$\mathsf{s}(\tilde{\mathbf{w}}) < \mathsf{s}(\hat{\mathbf{w}})$
Hom. growth	$\mathbf{w} + \alpha \mathbf{1}_n$	n				$\mathsf{s}(\tilde{\mathbf{w}}) < \mathsf{s}(\mathbf{w})$

Proof A.1.15 *[Replication] (cf. correction of [17].)*

$$\|\tilde{\mathbf{w}}\|_0 = k\|\mathbf{w}\|_0, \tag{1.31}$$

i.e. $\mathsf{s}_0(\tilde{\mathbf{w}}) = \mathsf{s}_0(\mathbf{w})$.

■

Proof A.1.16 *[Bounds]*

$$\mathsf{s}_0(\tilde{\mathbf{w}}) = -1, \quad \textit{and} \quad \mathsf{s}_0(\hat{\mathbf{w}}) = -\frac{1}{n}, \tag{1.32}$$

i.e. $\mathsf{s}_0(\tilde{\mathbf{w}}) < \mathsf{s}_0(\hat{\mathbf{w}})$.

■

Proof A.1.17 *[Quasi-convexity]*

$$\|\hat{\mathbf{w}}\|_0 \geq \min\{\|\mathbf{w}\|_0, \|\tilde{\mathbf{w}}\|_0\}, \tag{1.33}$$

i.e. $\mathsf{s}_0(\hat{\mathbf{w}}) \leq \max\{\mathsf{s}_0(\mathbf{w}), \mathsf{s}_0(\tilde{\mathbf{w}})\}$.

■

Proof A.1.18 *[Monotonicity]*

$$\|\tilde{\mathbf{w}}\|_0 = 2, \tag{1.34}$$

$$\|\hat{\mathbf{w}}\|_0 = \begin{cases} 2, & \hat{w} < \alpha, \\ 1, & \textit{otherwise}, \end{cases} \tag{1.35}$$

i.e. $\mathsf{s}_0(\tilde{\mathbf{w}}) \leq \mathsf{s}_0(\hat{\mathbf{w}})$.

■

Proof A.1.19 *[Completeness] (cf. correction of [17].)*

$$\|\tilde{\mathbf{w}}\|_0 = \|\mathbf{w}\|_0, \tag{1.36}$$

i.e. $\mathsf{s}_0(\tilde{\mathbf{w}}) > \mathsf{s}_0(\mathbf{w})$.

■

Proof A.1.20 *[Regularity] (cf. correction of [17].)*

$$\|\tilde{\mathbf{w}}\|_0 = \|\hat{\mathbf{w}}\|_0, \tag{1.37}$$

i.e. $\mathsf{s}_0(\tilde{\mathbf{w}}) = \mathsf{s}_0(\hat{\mathbf{w}})$.

■

Proof A.1.21 *[Hom. growth] (cf. correction of [17].)*

$$\|\tilde{\mathbf{w}}\|_0 = \begin{cases} \|\mathbf{w}\|_0 + k, & \textit{(k) new non-null components,} \\ \|\mathbf{w}\|_0, & \textit{otherwise,} \end{cases} \tag{1.38}$$

i.e. $\mathsf{s}_0(\tilde{\mathbf{w}}) \leq \mathsf{s}_0(\mathbf{w})$.

■

Proofs for "ℓ_1" (denoted s_1)

Proof A.1.22 *[Concentration] (cf. counterexample in [17].)*

$$\|\tilde{\mathbf{w}}\|_1 = \|\mathbf{w}\|_1, \tag{1.39}$$

i.e. $\mathsf{s}_1(\tilde{\mathbf{w}}) = \mathsf{s}_1(\mathbf{w})$.

■

Proof A.1.23 *[Scaling] (cf. counterexample in [17].)*

$$\|\tilde{\mathbf{w}}\|_1 = \alpha\|\mathbf{w}\|_1, \tag{1.40}$$

i.e. $\mathsf{s}_1(\tilde{\mathbf{w}}) = \alpha\mathsf{s}_1(\mathbf{w})$.

■

Proof A.1.24 *[Replication] (Correction of [17].)*

$$\|\tilde{\mathbf{w}}\|_1 = k\|\mathbf{w}\|_1, \tag{1.41}$$

i.e. $\mathsf{s}_1(\tilde{\mathbf{w}}) = \mathsf{s}_1(\mathbf{w})$.

■

Proof A.1.25 *[Bounds]*

$$\mathsf{s}_1(\tilde{\mathbf{w}}) = \mathsf{s}_1(\hat{\mathbf{w}}) = -\frac{\alpha}{n}, \tag{1.42}$$

i.e. $\mathsf{s}_1(\tilde{\mathbf{w}}) = \mathsf{s}_1(\hat{\mathbf{w}})$.

■

Proof A.1.26 *[Quasi-convexity] Assume* $\|\tilde{\mathbf{w}}\|_1 \leq \|\mathbf{w}\|_1$*, then, it holds that* $\mathsf{s}_1(\tilde{\mathbf{w}}) \geq \mathsf{s}_1(\mathbf{w})$*. Since* $\mathbf{w}$*,* $\tilde{\mathbf{w}}$ *non-negative,*

$$\|\tilde{\mathbf{w}}\|_1 \leq \alpha\|\tilde{\mathbf{w}}\|_1 + (1-\alpha)\|\mathbf{w}\|_1 = \|\hat{\mathbf{w}}\|_1, \tag{1.43}$$

i.e. $\mathsf{s}_1(\hat{\mathbf{w}}) \leq \mathsf{s}_1(\tilde{\mathbf{w}}) = \max\{\mathsf{s}_1(\mathbf{w}), \mathsf{s}_1(\tilde{\mathbf{w}})\}$.

■

Proof A.1.27 *[Monotonicity]*

$$\|\tilde{\mathbf{w}}\|_1 = \|\hat{\mathbf{w}}\|_1, \tag{1.44}$$

i.e. $\mathsf{s}_1(\tilde{\mathbf{w}}) = \mathsf{s}_1(\hat{\mathbf{w}})$.

■

Proof A.1.28 *[Completeness] (Correction of [17].)*

$$\|\tilde{\mathbf{w}}\|_1 = \|\mathbf{w}\|_1, \tag{1.45}$$

i.e. $\mathsf{s}_1(\tilde{\mathbf{w}}) > \mathsf{s}_1(\mathbf{w})$.

■

Proof A.1.29 *[Regularity] (cf. counterexample in [17].)*

$$\|\hat{\mathbf{w}}\|_1 = \|\tilde{\mathbf{w}}\|_1 + \alpha, \tag{1.46}$$

i.e. $\mathsf{s}_1(\tilde{\mathbf{w}}) > \mathsf{s}_1(\hat{\mathbf{w}})$.

■

Proof A.1.30 *[Hom. growth] (cf. obvious in [17].)*

$$\|\tilde{\mathbf{w}}\|_1 = \|\mathbf{w}\|_1 + n\alpha, \tag{1.47}$$

i.e. $\mathsf{s}_1(\tilde{\mathbf{w}}) < \mathsf{s}_1(\mathbf{w})$.

■

Proof A.1.31 *[Triangle ineq.] Notice that* $\nexists \alpha, \beta$ *such that* $\alpha \mathsf{s}_1(\mathbf{u}) + \beta \geq 0$, $\forall \mathbf{u}$ ***without knowledge of the maximum of*** $\mathbf{w}$ *(*$\sup|\mathbf{w}| = \|\mathbf{w}\|_\infty$*). Then,*

$$\|\hat{\mathbf{w}}\| \leq \|\mathbf{w}\|_1 + \|\tilde{\mathbf{w}}\|_1, \tag{1.48}$$

i.e. $\mathsf{s}_1(\hat{\mathbf{w}}) \geq \mathsf{s}_1(\mathbf{w}) + \mathsf{s}_1(\tilde{\mathbf{w}})$.

■

Proofs for Hoyer

Proof A.1.32 *[Concentration] (Correction of [17].)*

$$\mathrm{g}(\alpha) = \mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}(\alpha)) - \mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w}), \tag{1.49}$$

where $\mathrm{g}(0) = 0$*. Then,*

$$\mathrm{g}'(\alpha) = \frac{\|\mathbf{w}\|_1}{\sqrt{n}-1}\left(\frac{-1}{\|\tilde{\mathbf{w}}(\alpha)\|_2}\right)' \tag{1.50}$$

$$= \frac{1}{\sqrt{n}-1}\frac{\|\mathbf{w}\|_1}{\|\tilde{\mathbf{w}}\|_2^3}(w_{i_1} - w_{i_2} + 2\alpha) > 0, \tag{1.51}$$

since $w_{i_1} + \alpha > w_{i_2} - \alpha$*, i.e.* $\mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}) > \mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w})$.

■

Proof A.1.33 *[Scaling] (cf. obvious in [17].)*

$$\frac{\|\tilde{\mathbf{w}}\|_1}{\|\tilde{\mathbf{w}}\|_2} = \frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2}, \tag{1.52}$$

i.e. $\mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}) = \mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w})$.

■

Proof A.1.34 *[Replication] (cf. counterexample in [17].)*

$$\mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}) = \frac{\sqrt{n}-1}{\sqrt{n}-\frac{1}{\sqrt{m}}}\mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w}), \tag{1.53}$$

where $\frac{\sqrt{n}-1}{\sqrt{n}-\frac{1}{\sqrt{m}}} \to 1$ *as* $n \to \infty$*, i.e.* $\mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}) \to \mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w})$.

■

Proof A.1.35 *[Bounds]*

$$\mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}) = 0, \tag{1.54}$$

$$\mathsf{s}_{\mathsf{Hoyer}}(\hat{\mathbf{w}}) = 1, \tag{1.55}$$

i.e. $\mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}) < \mathsf{s}_{\mathsf{Hoyer}}(\hat{\mathbf{w}})$.

■

Proof A.1.36 *[Quasi-convexity] Let* $p = 1$*,* $q > 1$*, e.g. Hoyer with* $q = 2$*,* $\mathsf{s}_{1\infty}$ *with* $q = \infty$*, then*

- $\mathsf{f}(.) = \|.\|_q$ *convex,*
- $\mathsf{g}(.) = \|.\|_p$ *affine (or concave),*
- *by Theorem A.1.1(3 (or 1)),* $\mathsf{z}(.) = \frac{\mathsf{f}(.)}{\mathsf{g}(.)}$ *is semi-strictly quasi-convex,*
- *and, by Theorem A.1.2 with* $\mathsf{a}(.) = -\frac{1}{.}$*,* $\mathsf{a} \circ \mathsf{z}(.) = -\frac{\mathsf{g}(.)}{\mathsf{f}(.)}$ *is semi-strictly quasi-convex.*

■

Proof A.1.37 *[Monotonicity]*

$$\mathsf{g}(\alpha) = \mathsf{s}_{\mathsf{Hoyer}}(\hat{\mathbf{w}}(\alpha)) - \mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}(\alpha)) \tag{1.56}$$

$$= \frac{\alpha}{\sqrt{n}-1}\left(\frac{1}{\|\tilde{\mathbf{w}}\|_2} - \frac{1}{\|\hat{\mathbf{w}}\|_2}\right) > 0, \tag{1.57}$$

since $\|.\|_2$ *convex and* $\frac{\alpha}{2} \leq \tilde{w} < \hat{w} \leq \alpha$*, i.e.* $\mathsf{s}_{\mathsf{Hoyer}}(\hat{\mathbf{w}}) > \mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}})$.

■

Proof A.1.38 *[Completeness] (cf. obvious in [17].)*

$$\mathsf{g}(\mathbf{w}) = \mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}(\mathbf{w})) - \mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w}) \tag{1.58}$$

$$= \frac{\sqrt{n+1}-\sqrt{n}}{(\sqrt{n+1}-1)(\sqrt{n}-1)} \left(\frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2} - 1 \right) > 0, \tag{1.59}$$

since $\|\mathbf{u}\|_2 < \|\mathbf{u}\|_1$*, i.e.* $\mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}) > \mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w})$.

■

Proof A.1.39 *[Regularity] (Correction of [17].)*

$$\mathsf{g}(\alpha) = \mathsf{s}_{\mathsf{Hoyer}}(\hat{\mathbf{w}}(\alpha)) - \mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}), \tag{1.60}$$

where $\mathsf{g}(0) = 0$*. Then,*

$$\mathsf{g}'(\alpha) = \frac{-1}{\sqrt{n}-1} \left(\frac{\|\tilde{\mathbf{w}}\|_1 + \alpha}{\|\hat{\mathbf{w}}(\alpha)\|_2} \right)' \tag{1.61}$$

$$= \frac{\|\tilde{\mathbf{w}}\|_1(\|\tilde{\mathbf{w}}\|_\infty + \alpha) - \|\tilde{\mathbf{w}}\|_2^2 - \|\tilde{\mathbf{w}}\|_\infty \alpha}{(\sqrt{n}-1)\|\hat{\mathbf{w}}\|_2^3} > 0, \tag{1.62}$$

since $\|\mathbf{u}\|_2^2 < \|\mathbf{u}\|_\infty \|\mathbf{u}\|_1$ *and* $\|\mathbf{u}\|_1 > \|\mathbf{u}\|_\infty$*, i.e.* $\mathsf{s}_{\mathsf{Hoyer}}(\hat{\mathbf{w}}) > \mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}})$.

■

Proof A.1.40 *[Hom. growth] (cf. [17].)*

$$\mathsf{g}(\alpha) = \mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w}) - \mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}}(\alpha)), \tag{1.63}$$

where $\mathsf{g}(0) = 0$*. Then,*

$$\mathsf{g}'(\alpha) = \frac{1}{\sqrt{n}-1} \left(\frac{\|\mathbf{w}\|_1 + n\alpha}{\|\tilde{\mathbf{w}}(\alpha)\|_2} \right)' \tag{1.64}$$

$$= \frac{1}{\sqrt{n}-1} \frac{n\|\mathbf{w}\|_2^2 - \|\mathbf{w}\|_1^2}{\|\tilde{\mathbf{w}}\|_2^3} > 0, \tag{1.65}$$

since $\|\mathbf{u}\|_1 < \sqrt{n}\|\mathbf{u}\|_2$*, i.e.* $\mathsf{s}_{\mathsf{Hoyer}}(\mathbf{w}) > \mathsf{s}_{\mathsf{Hoyer}}(\tilde{\mathbf{w}})$.

■

Proofs for Gini

Recall that the order of indexes changes for this function, i.e. $w_{i_1} > w_{i_2}$ as $i_1 > i_2$.

Proof A.1.41 *[Concentration] (cf. [17].)*

$$\mathsf{g}(\alpha) = \mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}(\alpha)) - \mathsf{s}_{\mathsf{Gini}}(\mathbf{w}), \tag{1.66}$$

where $\mathsf{g}(0) = 0$. *Then,*

$$\mathsf{g}'(\alpha) = \frac{2}{(1+n)\|\mathbf{w}\|_1}(\|\tilde{\mathbf{w}}(\alpha)\|_{1^*})' \tag{1.67}$$

$$= \frac{2}{(1+n)\|\mathbf{w}\|_1}(i_1^*(w_{i_1}+\alpha) + i_2^*(w_{i_2}-\alpha))' \tag{1.68}$$

$$= \frac{2}{(1+n)\|\mathbf{w}\|_1}(i_1^* - i_2^*) > 0, \tag{1.69}$$

where $i_1^* \geq i_1$ *and* $i_2^* \leq i_2$ *are the new indexes (and weights) of components* $w_{i_1} + \alpha$ *and* $w_{i_2} - \alpha$, *respectively; i.e.* $\mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}) > \mathsf{s}_{\mathsf{Gini}}(\mathbf{w})$.

■

Proof A.1.42 *[Scaling] (cf. [17].)*

$$\frac{\|\tilde{\mathbf{w}}\|_{1^*}}{\|\tilde{\mathbf{w}}\|_1} = \frac{\|\mathbf{w}\|_{1^*}}{\|\mathbf{w}\|_1}, \tag{1.70}$$

i.e. $\mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}) = \mathsf{s}_{\mathsf{Gini}}(\mathbf{w})$.

■

Proof A.1.43 *[Replication] (Correction of [17].)*

$$\mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}) = \frac{2}{1+mn}\frac{\sum_{i=1}^{n}\sum_{i'=(i-1)k+1}^{ik} w_i}{k\|\mathbf{w}\|_1} - 1 \tag{1.71}$$

$$= \frac{(1+n)k}{1+kn}\mathsf{s}_{\mathsf{Gini}}(\mathbf{w}), \tag{1.72}$$

where $\frac{(1+n)k}{1+kn} \to 1$ *as* $n \to \infty$, *i.e.* $\mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}) \to \mathsf{s}_{\mathsf{Gini}}(\mathbf{w})$.

■

Proof A.1.44 *[Bounds]*

$$\mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}) = 0, \tag{1.73}$$

$$\mathsf{s}_{\mathsf{Gini}}(\hat{\mathbf{w}}) = \frac{n-1}{n+1}, \tag{1.74}$$

i.e, $\mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}) < \mathsf{s}_{\mathsf{Gini}}(\hat{\mathbf{w}})$.

■

Proof A.1.45 *[Quasi-convexity] Let* $p = 1^*$, $q = 1$, *e.g. Gini, then*

- $\mathsf{f}(.) = \|.\|_p$ *affine (and convex)*
- $\mathsf{g}(.) = \|.\|_q$ *affine*
- *by Theorem A.1.1(3),* $\mathsf{z}(.) = \frac{\mathsf{f}(.)}{\mathsf{g}(.)}$ *is semi-strictly quasi-convex.*

■

Proof A.1.46 *[Monotonicity]*

$$\mathsf{g}(\alpha) = \mathsf{s}_{\mathsf{Gini}}(\hat{\mathbf{w}}(\alpha)) - \mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}(\alpha)) \tag{1.75}$$

$$= \frac{2(\hat{w} - \tilde{w})}{\alpha(1+n)} > 0, \tag{1.76}$$

since $\hat{w} > \tilde{w}$*, i.e.* $\mathsf{s}_{\mathsf{Gini}}(\hat{\mathbf{w}}) > \mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}})$.

■

Proof A.1.47 *[Completeness] (cf. [17].)*

$$\mathsf{g}(\mathbf{w}) = \mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}) - \mathsf{s}_{\mathsf{Gini}}(\mathbf{w}), \tag{1.77}$$

$$= \frac{2}{1+n}\left(\frac{\sum_{i=1}^{n}(i+1)w_i}{\|\mathbf{w}\|_1} - \frac{\sum_{i=1}^{n} i w_i}{\|\mathbf{w}\|_1}\right) \tag{1.78}$$

$$= \frac{2}{1+n} > 0, \tag{1.79}$$

i.e. $\mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}) > \mathsf{s}_{\mathsf{Gini}}(\mathbf{w})$.

■

Proof A.1.48 *[Regularity] (cf. [17].)*

$$\mathsf{g}(\alpha) = \mathsf{s}_{\mathsf{Gini}}(\hat{\mathbf{w}}(\alpha)) - \mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}), \tag{1.80}$$

where $\mathsf{g}(0) = 0$*. Then,*

$$\mathsf{g}'(\alpha) = \frac{2}{1+n}\left(\frac{\|\tilde{\mathbf{w}}\|_{1^*} + n\alpha}{\|\tilde{\mathbf{w}}\|_1 + \alpha}\right)' \tag{1.81}$$

$$\frac{2}{(1+n)\|\hat{\mathbf{w}}\|_1^2}(n\|\tilde{\mathbf{w}}\|_1 - \|\tilde{\mathbf{w}}\|_{1^*}) > 0, \tag{1.82}$$

since $n\|\mathbf{u}\|_1 > \|\mathbf{u}\|_{1^*}$*, i.e.* $\mathsf{s}_{\mathsf{Gini}}(\hat{\mathbf{w}}) > \mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}})$.

■

Proof A.1.49 *[Hom. growth] (cf. [17].)*

$$\mathsf{g}(\alpha) = \mathsf{s}_{\mathsf{Gini}}(\mathbf{w}) - \mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}}(\alpha)), \tag{1.83}$$

where $\mathsf{g}(0) = 0$*. Then,*

$$\mathsf{g}'(\alpha) = \frac{-2}{n+1}\left(\frac{\|\mathbf{w}\|_{1^*} + \sum_{i=1}^{n} i\alpha}{\|\mathbf{w}\|_1 + n\alpha}\right)' \tag{1.84}$$

$$= \frac{2n}{(n+1)\|\tilde{\mathbf{w}}\|_1^2}\sum_{i=1}^{n} w_i(2i - 1 - n) \tag{1.85}$$

$$= \frac{2n\|\tilde{\mathbf{w}}\|_1^{-2}}{(n+1)}\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor}(w_{n-i+1} - w_i)(n+1-2i) > 0, \tag{1.86}$$

since $w_{n-i+1} > w_i$ *and* $n + 1 > 2i$ *with* $i \leq \lfloor \frac{n}{2} \rfloor$*, i.e.* $\mathsf{s}_{\mathsf{Gini}}(\mathbf{w}) > \mathsf{s}_{\mathsf{Gini}}(\tilde{\mathbf{w}})$.

■

Proofs for $\mathsf{s}_{1\infty}$

Proof A.1.50 *[Concentration]*

$$\mathsf{g}(\alpha) = \mathsf{s}_{1\infty}(\tilde{\mathbf{w}}(\alpha)) - \mathsf{s}_{1\infty}(\mathbf{w}) \tag{1.87}$$

$$= \frac{\|\mathbf{w}\|_1}{n}\left(\frac{1}{\|\mathbf{w}\|_\infty} - \frac{1}{\|\tilde{\mathbf{w}}\|_\infty}\right) \geq 0, \tag{1.88}$$

since $\|\tilde{\mathbf{u}}\|_\infty \geq \|\mathbf{u}\|_\infty$*, i.e.* $\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) \geq \mathsf{s}_{1\infty}(\mathbf{w})$.

■

Proof A.1.51 *[Scaling]*

$$\frac{\|\tilde{\mathbf{w}}\|_1}{\|\tilde{\mathbf{w}}\|_\infty} = \frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_\infty}, \tag{1.89}$$

i.e. $\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) = \mathsf{s}_{1\infty}(\mathbf{w})$.

■

Proof A.1.52 *[Replication]*

$$\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) = \mathsf{s}_{1\infty}(\mathbf{w}). \tag{1.90}$$

■

Proof A.1.53 *[Bounds]*

$$\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) = 0 \tag{1.91}$$

$$\mathsf{s}_{1\infty}(\hat{\mathbf{w}}) = 1 - \frac{1}{n}, \tag{1.92}$$

i.e. $\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) < \mathsf{s}_{1\infty}(\hat{\mathbf{w}})$. *Further, let* $\bar{\mathbf{w}}$ *with* $\|\bar{\mathbf{w}}\|_1 = \beta$ *and* $\|\bar{\mathbf{w}}\|_\infty = \lambda\beta$*,* $\frac{1}{n} \leq \lambda \leq 1$*. Then,*

$$\mathsf{s}_{1\infty}(\bar{\mathbf{w}}) = 1 - \frac{1}{\lambda n}, \tag{1.93}$$

i.e. $\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) \leq \mathsf{s}_{1\infty}(\bar{\mathbf{w}}) \leq \mathsf{s}_{1\infty}(\hat{\mathbf{w}})$.

■

Proof A.1.54 *[Monotonicity]*

$$\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) = 1 - \frac{\alpha}{2\tilde{w}}, \tag{1.94}$$

$$\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) = 1 - \frac{\alpha}{2\hat{w}}, \tag{1.95}$$

i.e. $\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) < \mathsf{s}_{1\infty}(\hat{\mathbf{w}})$.

■

Proof A.1.55 *[Completeness]*

$$\mathsf{g}(\mathbf{w}) = \mathsf{s}_{1\infty}(\tilde{\mathbf{w}}(\mathbf{w})) - \mathsf{s}_{1\infty}(\mathbf{w}) \tag{1.96}$$

$$= \frac{1}{n(n+1)} \frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_\infty} > 0, \tag{1.97}$$

i.e. $\mathsf{s}_{1\infty}(\tilde{\mathbf{w}}) > \mathsf{s}_{1\infty}(\mathbf{w})$.

■

Proof A.1.56 *[Regularity]*

$$\mathsf{g}(\alpha) = \mathsf{s}_{1\infty}(\hat{\mathbf{w}}(\alpha)) - \mathsf{s}_{1\infty}(\tilde{\mathbf{w}}), \tag{1.98}$$

$$= \frac{\alpha(\|\tilde{\mathbf{w}}\|_1 - \|\tilde{\mathbf{w}}\|_\infty)}{n\|\tilde{\mathbf{w}}\|_\infty \|\hat{\mathbf{w}}\|_\infty} > 0, \tag{1.99}$$

since $\|\mathbf{u}\|_1 > \|\mathbf{u}\|_\infty$*, i.e.* $\mathsf{s}_{1\infty}(\hat{\mathbf{w}}) > \mathsf{s}_{1\infty}(\tilde{\mathbf{w}})$.

■

Proof A.1.57 *[Hom. growth]*

$$\mathsf{g}(\alpha) = \mathsf{s}_{1\infty}(\mathbf{w}) - \mathsf{s}_{1\infty}(\tilde{\mathbf{w}}(\alpha)) \tag{1.100}$$

$$= \frac{\alpha(n\|\mathbf{w}\|_\infty - \|\mathbf{w}\|_1)}{n\|\mathbf{w}\|_\infty \|\tilde{\mathbf{w}}\|_\infty} > 0, \tag{1.101}$$

since $n\|\mathbf{u}\|_\infty > \|\mathbf{u}\|_1$*, i.e.* $\mathsf{s}_{1\infty}(\mathbf{w}) > \mathsf{s}_{1\infty}(\tilde{\mathbf{w}})$.

■

Proofs for s_{pq}***,*** $p \leq 1 < q$

Proof A.1.58 *[Concentration] (Correction of [17].)*

$$\mathsf{g}(\alpha) = \mathsf{s}_{pq}(\tilde{\mathbf{w}}(\alpha)) - \mathsf{s}_{pq}(\mathbf{w}) \tag{1.102}$$

$$\approx \alpha \nabla_{\mathbf{e}_{i_1} - \mathbf{e}_{i_2}} \mathsf{s}_{pq}(\mathbf{w}), \tag{1.103}$$

where $\mathsf{g}(0) = 0$*,* $\nabla_{\mathbf{e}_{i_1} - \mathbf{e}_{i_2}} \mathsf{f}(\mathbf{u}) = \frac{\partial \mathsf{f}}{\partial u_{i_1}}(\mathbf{u}) - \frac{\partial \mathsf{f}}{\partial u_{i_1}}(\mathbf{u})$*, and*

$$\frac{\partial \mathsf{s}_{pq}}{\partial w_i}(\mathbf{w}) = n^{\frac{1}{q} - \frac{1}{p}} \frac{\|\mathbf{w}\|_p}{\|\mathbf{w}\|_q} \left(\frac{w_i^{q-1}}{\|\mathbf{w}\|_q^q} - \frac{w_i^{p-1}}{\|\mathbf{w}\|_p^p} \right) \tag{1.104}$$

positive, since $u_i^{q-1}\|\mathbf{u}\|_p^p > u_i^{p-1}\|\mathbf{u}\|_q^q$ *(expand to see); i.e.* $\mathsf{s}_{pq}(\tilde{\mathbf{w}}) > \mathsf{s}_{pq}(\mathbf{w})$.

■

Proof A.1.59 *[Scaling] (cf. [17].)*

$$\frac{\|\tilde{\mathbf{w}}\|_p}{\|\tilde{\mathbf{w}}\|_q} = \frac{\|\mathbf{w}\|_p}{\|\mathbf{w}\|_q}, \tag{1.105}$$

i.e. $\mathsf{s}_{pq}(\tilde{\mathbf{w}}) = \mathsf{s}_{pq}(\mathbf{w})$.

■

Proof A.1.60 *[Replication] (cf. [17].)*

$$\mathsf{s}_{pq}(\tilde{\mathbf{w}}) = \mathsf{s}_{pq}(\mathbf{w}). \tag{1.106}$$

■

Proof A.1.61 *[Bounds]*

$$\mathsf{s}_{pq}(\tilde{\mathbf{w}}) = -1, \tag{1.107}$$

$$\mathsf{s}_{pq}(\hat{\mathbf{w}}) = -n^{\frac{1}{q}-\frac{1}{p}}, \tag{1.108}$$

i.e. $\mathsf{s}_{pq}(\tilde{\mathbf{w}}) < \mathsf{s}_{pq}(\hat{\mathbf{w}})$.

■

Proof A.1.62 *[Quasi-convexity] Let* $p \leq 1$, $q > 1$, *e.g. Hoyer with* $q = 2$, $\mathsf{s}_{1\infty}$ *with* $q = \infty$, *then*

- $\mathsf{f}(.) = -\|.\|_p$ *non-positive convex,*
- $\mathsf{g}(.) = \|.\|_q$ *positive convex,*
- *by Theorem A.1.1(2),* $\mathsf{z}(.) = \frac{\mathsf{f}(.)}{\mathsf{g}(.)}$ *is semi-strictly quasi-convex;*

■

Proof A.1.63 *[Monotonicity] Let* $\beta = \hat{w} - \tilde{w}$,

$$\mathsf{g}(\beta) = \mathsf{s}_{pq}(\hat{\mathbf{w}}) - \mathsf{s}_{pq}(\tilde{\mathbf{w}}) \tag{1.109}$$

$$\approx \beta \nabla_{\mathbf{e}_{i_1} - \mathbf{e}_{i_2}} \mathsf{s}_{pq}(\tilde{\mathbf{w}}), \tag{1.110}$$

where $\mathsf{g}(0) = 0$, $\nabla_{\mathbf{e}_{i_1} - \mathbf{e}_{i_2}} \mathsf{f}(\mathbf{u}) = \frac{\partial \mathsf{f}}{\partial u_{i_1}}(\mathbf{u}) - \frac{\partial \mathsf{f}}{\partial u_{i_2}}(\mathbf{u})$, *and*

$$\frac{\partial \mathsf{s}_{pq}}{\partial \tilde{w}_i}(\tilde{\mathbf{w}}) = n^{\frac{1}{q}-\frac{1}{p}} \frac{\|\tilde{\mathbf{w}}\|_p}{\|\tilde{\mathbf{w}}\|_q} \left(\frac{\tilde{w}_i^{q-1}}{\|\tilde{\mathbf{w}}\|_q^q} - \frac{\tilde{w}_i^{p-1}}{\|\tilde{\mathbf{w}}\|_p^p} \right) \tag{1.111}$$

positive, since $u_i^{q-1}\|\mathbf{u}\|_p^p > u_i^{p-1}\|\mathbf{u}\|_q^q$ *(expand to see); i.e.* $\mathsf{s}_{pq}(\hat{\mathbf{w}}) > \mathsf{s}_{pq}(\tilde{\mathbf{w}})$.

■

Proof A.1.64 *[Completeness] (cf. [17].)*

$$\mathsf{g}(\mathbf{w}) = \mathsf{s}_{pq}(\tilde{\mathbf{w}}(\mathbf{w})) - \mathsf{s}_{pq}(\mathbf{w}) \tag{1.112}$$

$$= \frac{\|\mathbf{w}\|_p}{\|\mathbf{w}\|_q}(n^{\frac{1}{q}-\frac{1}{p}} - (n+1)^{\frac{1}{q}-\frac{1}{p}}) > 0, \tag{1.113}$$

i.e. $\mathsf{s}_{pq}(\tilde{\mathbf{w}}) > \mathsf{s}_{pq}(\mathbf{w})$.

■

Proof A.1.65 *[Regularity] (cf. approximation in [17] under minor corrections.)*

$$\mathsf{g}(\alpha) = \mathsf{s}_{pq}(\hat{\mathbf{w}}) - \mathsf{s}_{pq}(\tilde{\mathbf{w}}) \tag{1.114}$$

$$\approx \alpha \nabla_{\mathbf{e}_i} \mathsf{s}_{pq}(\tilde{\mathbf{w}}), \tag{1.115}$$

where $\mathsf{g}(0) = 0$, $\nabla_{\mathbf{e}_i}\mathsf{f}(\mathbf{u}) = \frac{\partial \mathsf{f}}{\partial u_i}(\mathbf{u})$, *and*

$$\frac{\partial \mathsf{s}_{pq}}{\partial \tilde{w}_i}(\tilde{\mathbf{w}}) = n^{\frac{1}{q}-\frac{1}{p}} \frac{\|\tilde{\mathbf{w}}\|_p}{\|\tilde{\mathbf{w}}\|_q} \left(\frac{\tilde{w}_i^{q-1}}{\|\tilde{\mathbf{w}}\|_q^q} - \frac{\tilde{w}_i^{p-1}}{\|\tilde{\mathbf{w}}\|_p^p} \right) \tag{1.116}$$

positive, since $u_i^{q-1}\|\mathbf{u}\|_p^p > u_i^{p-1}\|\mathbf{u}\|_q^q$ *(expand to see); i.e.* $\mathsf{s}_{pq}(\hat{\mathbf{w}}) > \mathsf{s}_{pq}(\tilde{\mathbf{w}})$.

■

Proof A.1.66 *[Hom. growth] (cf. approximation in [17] which divides by non-necessarily non-null components.)*

$$\mathsf{g}(\alpha) = \mathsf{s}_{pq}(\mathbf{w}) - \mathsf{s}_{pq}(\tilde{\mathbf{w}}(\alpha)), \tag{1.117}$$

where $\mathsf{g}(0) = 0$. *Then,*

$$\mathsf{g}'(\alpha) = n^{\frac{1}{q}-\frac{1}{p}} \left(\frac{\|\mathbf{w} + \alpha \mathbf{1}_n\|_p}{\|\mathbf{w} + \alpha \mathbf{1}_n\|_q} \right)' \tag{1.118}$$

$$= n^{\frac{1}{q}-\frac{1}{p}} \frac{\|\tilde{\mathbf{w}}\|_p}{\|\tilde{\mathbf{w}}\|_q} \left(\frac{\|\tilde{\mathbf{w}}\|_{p-1}^{p-1}}{\|\tilde{\mathbf{w}}\|_p^p} - \frac{\|\tilde{\mathbf{w}}\|_{q-1}^{q-1}}{\|\tilde{\mathbf{w}}\|_q^q} \right) > 0, \tag{1.119}$$

since $\|\mathbf{u}\|_{p-1}^{p-1}\|\mathbf{u}\|_q^q > \|\mathbf{u}\|_p^p\|\mathbf{u}\|_{q-1}^{q-1}$ *(expand to see), i.e.* $\mathsf{s}_{pq}(\mathbf{w}) > \mathsf{s}_{pq}(\tilde{\mathbf{w}})$.

■

A.1.4 Evaluation of Entropy Functions

Regularity 3.3.5 and homogeneous growth 3.3.6 use a histogram model, e.g. if mass q_j is added to mass p_j of atom a_j, then PMF $\tilde{\mathbf{p}}$ is obtained after normalization $\|.\|_1 = 1$ (using the new number of realizations $\tilde{n} = n + q_j$). Completeness 3.3.4 requires normalization by the number of atoms. The simple trial $\frac{1}{b}$ (instead of $b^{\frac{1}{q}-\frac{1}{p}}$) will do it.

Table 1.2. Summary of notation used in proofs related to entropy functions. Operator π denotes permutation.

	$\tilde{\mathbf{p}}$	$\tilde{b}, \tilde{n}$	$\hat{\mathbf{p}}$	$\hat{b}, \hat{n}$	Assumptions	Desired result
Continuity		b, n			$\tilde{\mathbf{p}} \to \mathbf{p}$	$\mathsf{h}^b(\tilde{\mathbf{p}}) \to \mathsf{h}^b(\mathbf{p})$
Symmetry	$\pi(\mathbf{p})$	b, n				$\mathsf{h}^b(\tilde{\mathbf{p}}) = \mathsf{h}^b(\mathbf{p})$
Concentration	$\tilde{p}_{j_1} = p_{j_1} + \alpha$, $\tilde{p}_{j_2} = p_{j_2} - \alpha$	b, n			$p_{j_1} \geq p_{j_2}$	$\mathsf{h}^b(\tilde{\mathbf{p}}) < \mathsf{h}^b(\mathbf{p})$
Scaling	$\alpha\mathbf{p}$	$b, \alpha n$				$\mathsf{h}^b(\tilde{\mathbf{p}}) = \mathsf{h}^b(\mathbf{p})$
Replication	$\mathbf{p} \Vert \ldots \Vert \mathbf{p}$	b, kn			$b \to \infty$	$\mathsf{h}^b(\tilde{\mathbf{p}}) \to \mathsf{h}^b(\mathbf{p})$
Bounds	$\frac{1}{b}\mathbf{1}_b$	b, n	$\mathbf{0}_{b-1} \Vert 1$	b, n		$\mathsf{h}^b(\tilde{\mathbf{p}}) > \mathsf{h}^b(\hat{\mathbf{p}})$
Quasi-concave		b, n	$\alpha\mathbf{p} + (1-\alpha)\tilde{\mathbf{p}}$	b, n	$\alpha < 1$, $\mathsf{h}^b(\mathbf{p}) \geq \mathsf{h}^b(\tilde{\mathbf{p}})$	$\mathsf{h}^b(\tilde{\mathbf{p}}) < \mathsf{h}^b(\hat{\mathbf{p}})$
Monotonicity	$[\tilde{p}, 1-\tilde{p}]$	$2, n$	$[\hat{p}, 1-\hat{p}]$	$2, n$	$\frac{1}{2} \leq \tilde{p} < \hat{p} \leq 1$	$\mathsf{h}^b(\tilde{\mathbf{p}}) > \mathsf{h}^b(\hat{\mathbf{p}})$
Completeness	$\mathbf{p} \Vert 0$	$b+1, n$				$\mathsf{h}^b(\tilde{\mathbf{p}}) < \mathsf{h}^b(\mathbf{p})$
Regularity	$\tilde{p}_{j_1} = p_{j_1} + \beta$	$b, n+\beta$	$\hat{p}_{j_1} = p_{j_1} + \beta + \alpha$	$b, n+\beta+\alpha$		$\mathsf{h}^b(\tilde{\mathbf{p}}) > \mathsf{h}^b(\hat{\mathbf{p}})$
Hom. growth	$\mathbf{p} + \alpha\mathbf{1}_n$	$b, n + b\alpha$				$\mathsf{h}^b(\tilde{\mathbf{p}}) > \mathsf{h}^b(\mathbf{p})$

Proof A.1.67 *[Continuity, symmetry] By the continuity and symmetry of ℓ_p-norms, and the continuity of the logarithm and power functions.*

■

Proofs for Shannon entropy

Proof A.1.68 *[Concentration] (cf. [62].)*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{p}) - \mathsf{h}^b_{\mathsf{Shannon}}(\tilde{\mathbf{p}}(\alpha)), \tag{1.120}$$

where $\mathsf{g}(0) = 0$. *Then,*

$$\mathsf{g}'(\alpha) = (\log_2(p_{j_2} - \alpha)^{p_{j_2} - \alpha} + \log_2(p_{j_1} + \alpha)^{p_{j_1} + \alpha})' \tag{1.121}$$

$$= \log_2 \frac{p_{j_1} + \alpha}{p_{j_2} - \alpha} > 0, \tag{1.122}$$

since $p_{j_1} + \alpha > p_{j_2} - \alpha$, *i.e.* $\mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{w}) > \mathsf{h}^b_{\mathsf{Shannon}}(\tilde{\mathbf{w}})$.

■

Proof A.1.69 *[Scaling] By the ℓ_1 normalization of* $\mathbf{p}$, *i.e.* $\|\mathbf{p}\|_1 = 1$.

■

Proof A.1.70 *[Replication] Since each replica of the histogram allocates proportional occurrences to atoms.*

■

Proof A.1.71 *[Bounds] Shannon entropy attains its maximum with the uniform distribution and its minimum with a constant process [85].*

$$\mathsf{h}^b_{\mathsf{Shannon}}(\tilde{\mathbf{p}}) = \log_2 b, \tag{1.123}$$

$$\mathsf{h}^b_{\mathsf{Shannon}}(\hat{\mathbf{p}}) = 0, \tag{1.124}$$

i.e. $\mathsf{h}^b_{\mathsf{Shannon}}(\tilde{\mathbf{p}}) > \mathsf{h}^b_{\mathsf{Shannon}}(\hat{\mathbf{p}})$.

■

Proof A.1.72 *[Quasi-convexity] Shannon entropy is concave [85].*

■

Proof A.1.73 *[Monotonicity] By the Shannon entropy of the Bernoulli distribution [85].*

■

Proof A.1.74 *[Completeness]*

$$\mathsf{g}(\mathbf{w}) = \mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{p}) - \mathsf{h}^b_{\mathsf{Shannon}}(\tilde{\mathbf{p}}(\mathbf{p})) \tag{1.125}$$

$$= \frac{1}{b+1}\mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{p}) \geq 0, \tag{1.126}$$

since $\mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{p}) \geq 0$*, i.e.* $\mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{p}) \geq \mathsf{h}^b_{\mathsf{Shannon}}(\tilde{\mathbf{p}})$.

■

Proof A.1.75 *[Regularity]*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Shannon}}(\hat{\mathbf{p}}(\alpha)) - \mathsf{h}^b_{\mathsf{Shannon}}(\hat{\mathbf{p}}), \tag{1.127}$$

where $\mathsf{g}(0) = 0$*. Then,*

$$\mathsf{g}'(\alpha) = \sum_{\substack{j=1 \\ j\neq j_1}}^{b} \left(\frac{q_j}{\tilde{n}+\alpha}\log_2\frac{q_j}{\tilde{n}+\alpha}\right)' \dots \tag{1.128}$$

$$\dots + \left(\frac{\tilde{q_{j_1}}+\alpha}{\tilde{n}+\alpha}\log_2\frac{\tilde{q_{j_1}}+\alpha}{\tilde{n}+\alpha}\right)' \tag{1.129}$$

$$= \frac{1}{\hat{n}}(\mathsf{h}^b_{\mathsf{Shannon}}(\hat{\mathbf{p}}) - \mathsf{h}^b_{1\infty}(\hat{\mathbf{p}})) > 0, \tag{1.130}$$

since $\mathsf{h}^b_{\mathsf{Shannon}}(\hat{\mathbf{p}}) > \mathsf{h}^b_{1\infty}(\hat{\mathbf{p}})$*, i.e.* $\mathsf{h}^b_{\mathsf{Shannon}}(\hat{\mathbf{p}}) > \mathsf{h}^b_{\mathsf{Shannon}}(\hat{\mathbf{p}})$ *(cf. Figures 3.3 and 3.4).*

■

Proof A.1.76 *[Hom. growth]*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Shannon}}(\tilde{\mathbf{p}}(\alpha)) - \mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{p}), \tag{1.131}$$

where $\mathsf{g}(0) = 0$*. Then,*

$$\mathsf{g}'(\alpha) = -\sum_{j=1}^{b}\left(\frac{q_j+\alpha}{n+b\alpha}\log_2\frac{q_j+\alpha}{n+b\alpha}\right)' \tag{1.132}$$

$$= \sum_{j=1}^{b}\frac{q_j-n}{\tilde{n}}\log_2\frac{\tilde{q_j}}{\tilde{n}} > 0, \tag{1.133}$$

i.e. $\mathsf{h}^b_{\mathsf{Shannon}}(\tilde{\mathbf{w}}) \geq \mathsf{h}^b_{\mathsf{Shannon}}(\mathbf{w})$.

■

Proofs for Rényi entropy

Proof A.1.77 *[Concentration]*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Renyi}}(\mathbf{p}) - \mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}(\alpha)), \tag{1.134}$$

where $\mathsf{g}(0) = 0$. *Let* $\mathsf{d} = \frac{-1}{(1-q)\log 2}$. *Then,*

$$\mathsf{g}'(\alpha) = \mathsf{d}\left(\log\left(\sum_{\substack{j=1\\ j\neq j_1,j_2}}^{b} p_j^q + (p_{j_1}+\alpha)^q + (p_{j_2}-\alpha)^q\right)\right)' \tag{1.135}$$

$$= \frac{\mathsf{d}}{\|\tilde{\mathbf{p}}\|_q^q}((p_{j_1}+\alpha)^{q-1} - (p_{j_2}-\alpha)^{q-1}) > 0, \tag{1.136}$$

since $p_{j_1} + \alpha > p_{j_2} - \alpha$, *i.e.* $\mathsf{h}^b_{\mathsf{Renyi}}(\mathbf{p}) > \mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}})$.

■

Proof A.1.78 *[Bounds]*

$$\mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}) = \log_2 b, \tag{1.137}$$

$$\mathsf{h}^b_{\mathsf{Renyi}}(\hat{\mathbf{p}}) = 0, \tag{1.138}$$

i.e. $\mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}) > \mathsf{h}^b_{\mathsf{Renyi}}(\hat{\mathbf{p}})$.

■

Proof A.1.79 *[Quasi-concavity] By Theorem A.1.2, with* $\log_2$ *increasing,*

$$\mathsf{h}^b_{\mathsf{Renyi}}(\hat{\mathbf{p}}) = \underbrace{\frac{q}{1-q}}_{\substack{>0,\, q<1\\ <0,\, q>1}} (\log_2 \circ \underbrace{\|.\|_q}_{\substack{\text{concave},\, q<1\\ \text{convex},\, q>1}})(\mathbf{p}). \tag{1.139}$$

■

Proof A.1.80 *[Monotonicity]*

$$\mathsf{g}(1) = \mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}) - \mathsf{h}^b_{\mathsf{Renyi}}(\hat{\mathbf{p}}) \tag{1.140}$$

$$= \frac{1}{(1-q)} \log_2 \frac{\|\tilde{\mathbf{p}}\|_q}{\|\hat{\mathbf{p}}\|_q}, \tag{1.141}$$

since $\|\tilde{\mathbf{p}}\|_q > (<)\|\hat{\mathbf{p}}\|_q$ *by the concavity (convexity) when* $q < (>)1$, *i.e.* $\mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}) < \mathsf{h}^b_{\mathsf{Renyi}}(\hat{\mathbf{p}})$.

■

Proof A.1.81 *[Completeness]*

$$\mathsf{g}(\mathbf{p}) = \mathsf{h}^b_{\mathsf{Renyi}}(\mathbf{p}) - \mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}(\mathbf{p})) \tag{1.142}$$

$$= \frac{1}{\bar{b}} \mathsf{h}^b_{\mathsf{Renyi}}(\mathbf{p}) \geq 0, \tag{1.143}$$

since $\mathsf{h}^b_{\mathsf{Renyi}}(\mathbf{p}) \geq 0$, *i.e.* $\mathsf{h}^b_{\mathsf{Renyi}}(\mathbf{p}) > \mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}})$.

■

Proof A.1.82 *[Regularity]*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}) - \mathsf{h}^b_{\mathsf{Renyi}}(\hat{\mathbf{p}}(\alpha)), \tag{1.144}$$

where $\mathsf{g}(0) = 0$. *Then,*

$$\mathsf{g}'(\alpha) = \frac{-1}{(1-q)\log 2} \sum_{\substack{j=1 \\ j\neq j_1}}^{b} \left(\frac{q_j}{\tilde{n}+\alpha}\right)' + \left(\frac{\tilde{q}_j+\alpha}{\tilde{n}+\alpha}\right)' \tag{1.145}$$

$$= \frac{q}{(1-q)\log 2 \|\hat{\mathbf{p}}\|_q^q \hat{n}} (\|\hat{\mathbf{p}}\|_q^q - \|\hat{\mathbf{p}}\|_\infty^{q-1}) > 0, \tag{1.146}$$

since $\|\mathbf{u}\|_q^q > (<)\|\mathbf{u}\|_\infty^{q-1}$ *when* $q < (>)1$, *i.e.* $\mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}) > \mathsf{h}^b_{\mathsf{Renyi}}(\hat{\mathbf{p}})$.

■

Proof A.1.83 *[Hom. growth]*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}) - \mathsf{h}^b_{\mathsf{Renyi}}(\mathbf{p}), \tag{1.147}$$

where $\mathsf{g}(0) = 0$. *Then,*

$$\mathsf{g}'(\alpha) = \frac{1}{(1-q)\log_2} \sum_{j=1}^{b} \left(\left(\frac{q_j+\alpha}{n+b\alpha}\right)^q\right)' \tag{1.148}$$

$$= \frac{q}{(1-q)\log 2 \|\mathbf{p}\|_q^q \tilde{n}} \sum_{j=1}^{b} (\tilde{p}_j^{q-1} - 1) > 0, \tag{1.149}$$

since $\tilde{p}_j^{q-1} < (>)1$ *with* $q > (<)1$, *i.e.* $\mathsf{h}^b_{\mathsf{Renyi}}(\tilde{\mathbf{p}}) > \mathsf{h}^b_{\mathsf{Renyi}}(\hat{\mathbf{p}})$.

■

Proofs for Tsallis entropy

Proof A.1.84 *[Concentration]*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Tsallis}}(\mathbf{p}) - \mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}(\alpha)), \tag{1.150}$$

where $\mathsf{g}(0) = 0$. *Then,*

$$\mathsf{g}'(\alpha) = \frac{1}{(q-1)}((p_{j_1}+\alpha)^q + (p_{j_2}-\alpha)^q)' \tag{1.151}$$

$$= \frac{q}{(q-1)}((p_{j_1}+\alpha)^{q-1} - (p_{j_2}-\alpha)^{q-1}) > 0, \tag{1.152}$$

since $p_{j_1} + \alpha > p_{j_2} - \alpha$, *i.e.* $\mathsf{h}^b_{\mathsf{Tsallis}}(\mathbf{p}) > \mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}})$.

■

Proof A.1.85 *[Bounds]*

$$\mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}) = \frac{1-b^{1-q}}{(q-1)}, \tag{1.153}$$

$$\mathsf{h}^b_{\mathsf{Tsallis}}(\hat{\mathbf{p}}) = 0, \tag{1.154}$$

i.e. $\mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}) > \mathsf{h}^b_{\mathsf{Tsallis}}(\hat{\mathbf{p}})$.

■

Proof A.1.86 *[Quasi-concavity] By Theorem A.1.2, with* $(.)^q$ *increasing,*

$$\mathsf{h}^b_{\mathsf{Tsallis}}(\hat{\mathbf{p}}) = \underbrace{\frac{q}{q-1}}_{\substack{<0, q<1 \\ >0, q>1}} (1-(.)^q \circ \underbrace{\|.\|_q}_{\substack{\text{concave}, q<1 \\ \text{convex}, q>1}})(\mathbf{p}). \tag{1.155}$$

■

Proof A.1.87 *[Monotonicity]*

$$\mathsf{g}(1) = \mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}) - \mathsf{h}^b_{\mathsf{Tsallis}}(\hat{\mathbf{p}}) \tag{1.156}$$

$$= \frac{\|\tilde{\mathbf{p}}\|_q^q}{(q-1)} \left(\frac{\|\hat{\mathbf{p}}\|_q^q}{\|\tilde{\mathbf{p}}\|_q^q} - 1 \right), \tag{1.157}$$

since $\|\tilde{\mathbf{p}}\|_q > (<) \|\hat{\mathbf{p}}\|_q$ *by the concavity (convexity) when* $q < (>) 1$*, i.e.* $\mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}) < \mathsf{h}^b_{\mathsf{Tsallis}}(\hat{\mathbf{p}})$.

■

Proof A.1.88 *[Completeness]*

$$\mathsf{g}(\mathbf{p}) = \mathsf{h}^b_{\mathsf{Tsallis}}(\mathbf{p}) - \mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}(\mathbf{p})) \tag{1.158}$$

$$= \frac{1}{b+1} \mathsf{h}^b_{\mathsf{Tsallis}}(\mathbf{p}) \geq 0, \tag{1.159}$$

since $\mathsf{h}^b_{\mathsf{Tsallis}}(\mathbf{p}) \geq 0$*, i.e.* $\mathsf{h}^b_{\mathsf{Tsallis}}(\mathbf{p}) > \mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}})$.

■

Proof A.1.89 *[Regularity]*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}) - \mathsf{h}^b_{\mathsf{Tsallis}}(\hat{\mathbf{p}}(\alpha)), \tag{1.160}$$

where $\mathsf{g}(0) = 0$*. Then,*

$$\mathsf{g}'(\alpha) = \frac{1}{(1-q)} \sum_{\substack{j=1 \\ j\neq j_1}}^{b} \left(\left(\frac{q_j}{\tilde{n}+\alpha} \right)^q + \left(\frac{\tilde{q}_{j_1}+\alpha}{\tilde{n}+\alpha} \right)^q \right)' \tag{1.161}$$

$$\frac{q}{(1-q)\tilde{n}} (\|\hat{\mathbf{p}}\|_\infty^{q-1} - \|\hat{\mathbf{p}}\|_q^q) > 0, \tag{1.162}$$

since $\|\mathbf{u}\|_\infty^{q-1} > (<) \|\mathbf{u}\|_q^q$ *with* $q < (>) 1$, *i.e.* $\mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}) > \mathsf{h}^b_{\mathsf{Tsallis}}(\hat{\mathbf{p}})$.

■

Proof A.1.90 *[Hom. growth]*

$$\mathsf{g}(\alpha) = \mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}) - \mathsf{h}^b_{\mathsf{Tsallis}}(\mathbf{p}), \tag{1.163}$$

where $\mathsf{g}(0) = 0$. *Then,*

$$\mathsf{g}'(\alpha) = \frac{-1}{(1-q)} \sum_{j=1}^{b} \left(\left(\frac{q_j+\alpha}{n+b\alpha} \right)^q \right)' \tag{1.164}$$

$$= \frac{-q}{(q-1)\tilde{n}} (\|\tilde{\mathbf{p}}\|_{q-1}^{q-1} - b\|\tilde{\mathbf{p}}\|_q^q) > 0, \tag{1.165}$$

since $\|\mathbf{u}\|_{q-1}^{q-1} \gtrsim (\lesssim) \|\mathbf{u}\|_0 \|\mathbf{u}\|_q^q$ *with* $q \lesssim (\gtrsim) 1$, *i.e.* $\mathsf{h}^b_{\mathsf{Tsallis}}(\tilde{\mathbf{p}}) > \mathsf{h}^b_{\mathsf{Tsallis}}(\hat{\mathbf{p}})$.

■

A.2 Proofs of Chapter 4

A.2.1 Proof of Theorem 4.3.1:

Let $\tilde{s}_{Y,k} = \frac{\tilde{k}_{Y,k}}{k!\delta^{k-2}}$ with δ arbitrary. Now, (4.28) can be written as

$$\tilde{\varphi}_Y(v) = e^{\frac{(v-1)^2}{2}} e^{\sum_{k\geq 1} \tilde{s}_{Y,k+2}\delta^k (v-1)^{k+2}}, \tag{1.166}$$

Note that, by the definition of $\tilde{s}_{Y,k}$, δ could be deleted in (1.166). However, this term is introduced to perform an expansion as follows: first, write $f(t) \equiv \exp(t)$ and $g(\delta) \equiv \sum_{k\geq 1} \tilde{s}_{Y,k+2}\delta^k (v-1)^{k+2}$. Then, the Taylor series of $f \circ g(\delta)$ around $\delta = 0$ is $\sum_{s\geq 0} \tilde{\mathcal{P}}_s (v-1)\delta^s$ with $\tilde{\mathcal{P}}_s(v-1) = \frac{1}{s!} D^s_\delta (f \circ g)(0)$,

$$\tilde{\varphi}_Y(v) = e^{\frac{(v-1)^2}{2}} \left(1 + \sum_{s\geq 1} \delta^s \sum_{\mathcal{K}_s} \tilde{\omega}_{Y,s} (v-1)^{s+2r} \right), \tag{1.167}$$

and which inverse Mellin transform recovers the density of Y. The term $v-1$ for the Mellin transform plays a similar role as the term iu for

the Fourier transform. Note that $e^{\frac{(v-1)^2}{2}}$ is the Mellin transform of the standard log-normal distribution. Further, note that the Mellin transform can be written in terms of the Fourier transform as $\acute{f}(v) = \widehat{f \circ e^{\cdot}}(v)$. This leads to long calculations. Instead, by the change of variables $iu = v - 1$ and $Z = \mathcal{L}(Y)$, which lead to

$$\tilde{\varphi}_Y(v) = \mathbb{E}_Y[Y^{v-1}] = \mathbb{E}_Y[Y^{iu}] = \mathbb{E}_Z[e^{iuZ}] = \varphi_Z(u), \tag{1.168}$$

(1.167) reads

$$\varphi_Z(u) = e^{\frac{(iu)^2}{2}} \left(1 + \sum_{s\geq 1} \delta^s \sum_{\mathcal{K}_s} \tilde{\omega}_{Y,s} (iu)^{s+2r}\right). \tag{1.169}$$

The result follows now by the key relation between the standard normal distribution and Hermite polynomials, i.e. $f_{\mathcal{N}}(x)\mathcal{H}_k(x) = (-1)^k D_x^k f_{\mathcal{N}}(x)$, and by the properties of the Fourier transform of the k^{th} order derivatives of a function. Finally, from the density of $Z \in \mathcal{N}(0,1)$ in (4.34) the density of RV $Y \in \mathcal{L}^{-1}\mathcal{N}(0,1)$ is easily recovered by reversing the logarithmic transformation.

■

A.2.2 Proof of Lemma 4.3.4:

The Taylor series around $u = 1$ of the second characteristic log-function is

$$\tilde{\psi}_X(v) = \underbrace{\tilde{\psi}_X(1)}_{=0} + \sum_{k\geq 1} \frac{D_v^k \tilde{\psi}_X(1)}{k!}(v-1)^k \tag{1.170}$$

$$= \sum_{k\geq 1} \frac{\tilde{\kappa}_{X,k}}{k!}(v-1)^k. \tag{1.171}$$

Then,

$$\tilde{\varphi}_X(v) = \exp\left(\sum_{k\geq 1} \frac{\tilde{\lambda}_k}{k!}(v-1)^k\right) \tilde{\varphi}_U(v). \tag{1.172}$$

Recall $\tilde{\varphi}_X(v) = \mathbb{E}_X[X^{v-1}]$ and $\varphi_X(u) = \mathbb{E}_X[e^{iuX}]$. Thus, the change of variables $iu = v - 1$ leads to

$$\mathbb{E}_X[e^{iu\log X}] = \exp\left(\sum_{k\geq 1} \frac{\tilde{\lambda}_k}{k!}(iu)^k\right) \mathbb{E}_U(e^{iu\log U}), \tag{1.173}$$

which in terms of characteristic functions of RVs Y and V reads

$$\varphi_Y(u) = \exp\left(\sum_{k\geq 1} \frac{\tilde{\lambda}_k}{k!}(iu)^k\right) \varphi_V(u). \tag{1.174}$$

Now the result follows by the existing Gram-Charlier series (in terms of cumulants).

■

A.2.3 Proof of Theorem 4.3.5:

The proof can be traced from the combinatorial arguments of [105] (equations (32)-(39)).

■

A.3 Proofs of Chapter 5

A.3.1 Proof of Theorem 5.2.1:

Let $\tau(\mathbf{x}_i) = t_i$ be the time mark of sensor $\mathbf{x}_i$ and assume

$$0 = t_{n+1} \leq t_i \leq \cdots \leq t_1 \leq t_0 = 1. \tag{1.175}$$

For each k, define the sets

$$\Theta_{\lambda_{\mathsf{on}}[t_{k+1},t_k)} = \{\mathbf{x}_j \in \Theta_{\mathsf{on}} : \tau(\mathbf{x}_j) \in [t_{k+1}, t_k)\}. \tag{1.176}$$

Then, $\mathbf{x}_1, \ldots, \mathbf{x}_i \in \Theta_\Omega$ if the following $n-1$ conditions hold

$$C_k = \{\Theta_{\lambda_{\mathsf{on}}[t_{k+1},t_k)}(\cup_{m\leq k}\mathcal{N}_m) = 0\},\ k = 2, \ldots, n \tag{1.177}$$

which, for each k, represents the event that sensors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ do not contend with sensors with smaller (hence better) marks. These conditions are independent since $\Theta_{\lambda_{\mathsf{on}}[t_{k+1},t_k)}$ and $\mathbf{x}_k$ are independent PPPs and RVs, respectively. Then,

$$\begin{aligned}
\mathbb{P}(C_k|t_{k+1}, t_k) &= \mathbb{P}(u_z^{[k]}, \forall z \in \Theta_{\lambda_{\mathsf{on}}[t_{k+1},t_k)}) \\
&= \mathbb{P}(\Theta_{\lambda_{\mathsf{on}}[t_{k+1},t_k)}(\mathcal{N}_{[k]}) = 0) \\
&\overset{(2.8)}{=} \exp\left(-\lambda_{\mathsf{on}}(t_k - t_{k+1}) \int_{\mathbb{R}^2} (1 - \mathbb{P}(u_z^{[k]}))dz\right) \\
&\overset{(2.9)}{=} G_{\lambda_{\mathsf{on}}(t_k - t_{k+1})}(b_k), \quad \text{with } b_k = \mathbb{P}(u_z^{[k]}).
\end{aligned} \tag{1.178}$$

Note that

$$B_k = \int_{\mathbb{R}^2} (1 - \mathbb{P}(u_z^{[k]}))dz \tag{1.179}$$

is the size of neighbors of sensors $\mathbf{x}_{[k]}$ when the density is normalized to 1. Now, let $T = \{t_1, \ldots, t_i\}$. Then,

$$\mathbb{P}(\mathbf{x}_{[n]} \in \Theta_\Omega | T) = \mathbb{P}(\cap_{k<n} C_k | T) \tag{1.180}$$

and we just need deconditioning in the marks as follows.

$$\begin{aligned}
\mathbb{P}(\mathbf{x}_{[n]} \in \Theta_\Omega) &= \int_{[0,1]^n} \mathbb{P}(\cap_{k<n} C_k | T) d(\otimes_{k\leq n} t_k) \\
&\overset{(1.178)}{=} \int_{[0,1]^n} \prod_{k\leq n} G_{\lambda_{\mathsf{on}}(t_k - t_{k+1})}(b_k) dt_k \\
&\overset{(1.179)}{=} \int_{[0,1]^n} \prod_{k\leq n} \exp(-\lambda_{\mathsf{on}}(t_k - t_{k+1}) B_k) dt_k \\
&= \int_{[0,1]^n} \prod_{k\leq n} \exp(-\lambda_{\mathsf{on}} t_k (B_k - B_{k-1})) dt_k, \quad B_0 = 0 \\
&= \int_{[0,\lambda_{\mathsf{on}}]^n} \prod_{k\leq n} \exp(-s_k (B_k - B_{k-1})) \frac{ds_k}{\lambda_{\mathsf{on}}}, \quad \lambda_{\mathsf{on}} t_k = s_k \\
&= \prod_{k\leq n} \frac{1}{\lambda_{\mathsf{on}}} \int_0^{\lambda_{\mathsf{on}}} \exp(-s(B_k - B_{k-1})) ds \\
&= \prod_{k\leq n} \frac{1}{\lambda_{\mathsf{on}}} \int_0^{\lambda_{\mathsf{on}}} \exp(-s \int_{\mathbb{R}^2} \mathbb{P}(u_z^{[k-1]}, v_z^k) dz) ds \\
&\overset{(2.9)}{=} \prod_{k\leq n} \underbrace{\frac{1}{\lambda_{\mathsf{on}}} \int_0^{\lambda_{\mathsf{on}}} G_s(1 - \mathbb{P}(u_\cdot^{[k-1]}, v_\cdot^k)) ds}_{\mathbb{P}(\mathbf{x}_k \in \Theta_\Omega | \mathbf{x}_{[k-1]} \in \Theta_\Omega)}.
\end{aligned}$$

■

A.3.2 Proof of Theorem 5.2.2:

From Theorem 5.2.1, simple calculations show that

$$\mathbb{P}(\mathbf{x}_i \in \Theta_\Omega | \mathbf{x}_{[n-1]} \in \Theta_\Omega) = \frac{1 - e^{-\lambda_{\mathsf{on}} c_i}}{\lambda_{\mathsf{on}} c_i},$$

where

$$\lambda_{\mathsf{on}} c_i = \int_{\mathbb{R}^2} \mathbb{P}(u_j^{[n-1]}, v_j^n) \lambda_{\mathsf{on}} dm = \mathbb{E}[|\mathcal{N}_i \backslash \mathcal{N}_{[n-1]}|]$$

is the expected number of sensors which are in the neighborhood of sensor $\mathbf{x}_i$ and are not in the neighborhoods of sensors $\mathbf{x}_{[n-1]}$. It can also be understood as the expected number of remaining contenders of sensor $\mathbf{x}_i$, i.e. those sensors which have not lost the contention with sensors $\mathbf{x}_{[n-1]}$ in the current MAC frame.

Now, since $|\mathcal{N}_i \backslash \mathcal{N}_{[n-1]}|$ is a Poisson RV [126] we have that

$$\mathbb{P}(|\mathcal{N}_i \backslash \mathcal{N}_{[n-1]}| = 0) \equiv \exp(-\lambda_{\mathsf{on}} c_i),$$

and we have the result.

■

Bibliography

[1] J. P. Conti, "The internet of things," *Communications Engineer*, vol. 4, no. 6, pp. 20–25, Dec 2006. *Cited on page(s): 1*

[2] F. J. Alexander, A. Hoisie, and A. Szalay, "Big data," *Computing in Science & Engineering*, vol. 13, no. 6, pp. 10–13, Nov-Dec 2011. *Cited on page(s): 1*

[3] R. Rajagopalan and P. K. Varshney, "Data-aggregation techniques in sensor networks: A survey," *Communications Surveys and Tutorials*, vol. 8, no. 4, pp. 48–63, Fourth Quarter 2006. *Cited on page(s): 1*

[4] M. Martinelli, L. Ioriatti, F. Viani, M. Benedetti, and A. Massa, "A WSN-based solution for precision farm purposes," in *Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, Jul 2009, pp. V–469–V–472. *Cited on page(s): 1*

[5] P. Corke, T. Wark, R. Jurdak, W. Hu, P. Valencia, and D. Moore, "Environmental wireless sensor networks," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1903–1917, Nov 2010. *Cited on page(s): 1*

[6] A. Mahmood, *Enabling Time-Synchronized and Interference-Aware Initialization of Wireless Sensor Networks*. Aalto University publication series DOCTORAL DISSERTATIONS, 8/2014, 2014, ch. Probability Functions. *Cited on page(s): 2*

[7] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006. *Cited on page(s): 2, 3, 5, 9, 10, 61*

[8] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 797–829, 2004. *Cited on page(s): 2, 3, 10*

[9] Y. Kim, M. S. Nadar, and A. Bilgin, "Dynamic compressive magnetic resonance imaging using a Gaussian scale mixtures model," in *Proceedings of the 2011 IEEE International Conference on Image Processing (ICIP)*, Sept 2011, pp. 2293–2296. *Cited on page(s): 2*

[10] L. Anitori, W. van Rossum, M. Otten, A. Maleki, and R. Baraniuk, "Compressive sensing radar: Simulation and experiments for target detection," in *Proceedings of the 2013 European Signal Processing Conference (EUSIPCO)*, Sept 2013, pp. 1–5. *Cited on page(s): 2*

[11] E. Candes and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007. *Cited on page(s): 2, 3*

[12] T. Blumensath, "Accelerated iterative hard thresholding," *Signal Processing*, vol. 92, no. 3, pp. 752–756, 2012. *Cited on page(s): 2, 3, 63, 64, 67*

[13] T. Goldstein, C. Studer, and R. G. Baraniuk, "FASTA: A generalized implementation of forward-backward splitting," *CoRR*, vol. abs/1501.04979, 2015. *Cited on page(s): 2, 3*

[14] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec 2006. *Cited on page(s): 2, 3*

[15] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, p. 969, 2007. *Cited on page(s): 2, 3, 5, 11*

[16] G. Xu and Z. Xu, "Compressed sensing matrices from Fourier matrices," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 469–478, Jan 2015. *Cited on page(s): 2, 3, 11*

[17] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723–4741, Oct 2009. *Cited on page(s): 2, 3, 10, 17, 18, 23, 25, 26, 81, 83, 84, 86, 87, 88, 89, 90, 91, 92, 94, 95, 96*

[18] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processsing Letters*, vol. 14, no. 10, pp. 707–710, Oct 2007. *Cited on page(s): 2, 3, 29*

[19] G. Zhao, F. Shen, Z. Wang, and G. Shi, "Cauchy diversity measures: A novel methodology for enhancing sparsity in compressed sensing," *IET Signal Processing*, vol. 7, no. 9, pp. 791–799, Dec 2013. *Cited on page(s): 2, 3*

[20] S. J. Baek, G. de Veciana, and X. Su, "Minimizing energy consumption in large-scale sensor networks through distributed data compression and hierarchical aggregation," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1130–1140, Aug 2004. *Cited on page(s): 2, 3*

[21] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2006, pp. 134–142. *Cited on page(s): 2, 3*

[22] L. Xiang, J. Luo, and C. Rosenberg, "Compressed data aggregation: Energy-efficient and high-fidelity data collection," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1722–1735, Dec 2013. *Cited on page(s): 2, 3, 59, 60, 78*

[23] C. Zhao, W. Zhang, Y. Yang, and S. Yao, "Treelet-based clustered compressive data aggregation for wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4257–4267, 2015. *Cited on page(s): 2, 3, 59, 60, 78*

[24] G. Yang, V. Y. F. Tan, C. K. Ho, S. H. Ting, and Y. L. Guan, "Wireless compressive sensing for energy harvesting sensor nodes," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4491–4505, Sept 2013. *Cited on page(s): 2, 3, 59, 78*

[25] F. Fazel, M. Fazel, and M. Stojanovic, "Random access compressed sensing over fading and noisy communication channels," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2114–2125, May 2013. *Cited on page(s): 2, 3, 59, 61, 63, 69, 75, 78*

[26] J. F. C. Kingman, *Poisson Processes*, ser. Oxford Studies in Probability. Clarendon Press, 1992. *Cited on page(s): 3*

[27] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, ser. Wiley Series in Probability and Statistics. Wiley, 2013. *Cited on page(s): 3*

[28] A. Ghasemi and E. S. Sousa, "Interference aggregation in spectrum-sensing cognitive wireless networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 41–56, Feb 2008. *Cited on page(s): 3, 40, 51*

[29] M. Aljuaid and H. Yanikomeroglu, "Investigating the Gaussian convergence of the distribution of the aggregate interference power in large wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4418–4424, Nov 2010. *Cited on page(s): 3, 39*

[30] G. L. Torrisi and E. Leonardi, "Simulating the tail of the interference in a Poisson network model," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1773–1787, Mar 2013. *Cited on page(s): 3, 40*

[31] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler, "Stochastic analysis of spatial and opportunistic ALOHA," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1105–1119, Sept 2009. *Cited on page(s): 3*

[32] M. Kaynia, N. Jindal, and G. E. Oien, "Improving the performance of wireless ad hoc networks through MAC layer design," *IEEE Transactions on Wireless Communications*, vol. 10, no. 1, pp. 240–252, Jan 2011. *Cited on page(s): 3*

[33] Y. Kim, F. Baccelli, and G. de Veciana, "Spatial reuse and fairness of ad hoc networks with channel-aware CSMA protocols," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4139–4157, Jul 2014. *Cited on page(s): 3, 67, 69, 70, 74*

[34] S. Srinivasa and M. Haenggi, "Distance distributions in finite uniformly random networks: Theory and applications," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, pp. 940–949, Feb 2010. *Cited on page(s): 3, 12*

[35] V. Suryaprakash, J. Moller, and G. Fettweis, "On the modeling and analysis of heterogeneous radio access networks using a Poisson cluster process," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1035–1047, Feb 2015. *Cited on page(s): 3*

[36] N. Deng, W. Zhou, and M. Haenggi, "The Ginibre point process as a model for wireless networks with repulsion," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 107–121, Jan 2015. *Cited on page(s): 3*

[37] M. Haenggi, "On distances in uniformly random networks," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3584–3586, Oct 2005. *Cited on page(s): 3, 12*

[38] A. Busson, G. Chelius, and J. M. Gorce, "Interference modeling in CSMA multi-hop wireless networks," INRIA, Research Report 6624, 2009. *Cited on page(s): 3, 14, 39, 57, 70*

[39] J. M. Nicolas, "Introduction aux statistiques de deuxieeme espece: Applications des logs-moments et des logs-cumulants a l'analyse des lois d'images radar," *Traitement du Signal*, vol. 19, no. 3, pp. 139–167, 2002. *Cited on page(s): 5, 40, 41, 46*

[40] G. Pastor, I. Mora-Jimenez, R. Jantti, and A. J. Caamano, "Sparsity-based criteria for entropy measures," in *Proceedings of the 2013 IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2013, pp. 1–5. *Cited on page(s): 6, 7, 25, 37*

[41] G. Pastor, I. Mora-Jimenez, A. J. Caamano, and R. Jantti, "Medium access probability in uniform networks with general propagation models," in *Proceedings of the 2013 IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2013, pp. 1–5. *Cited on page(s): 6, 7, 69*

[42] M. I. Chidean, G. Pastor, E. Morgado, J. Ramiro-Bargueno, and A. J. Caamano, "Wireless sensor network for low-complexity entropy determination of human gait," in *Proceedings of the 2013 IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 2644–2648. *Cited on page(s): 6, 7*

[43] G. Pastor, I. Mora-Jimenez, A. J. Caamano, and R. Jantti, "Log-cumulants-based Edgeworth expansion for skew-distributed aggregate interference," in *Proceedings of the 2014 IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2014, pp. 390–394. *Cited on page(s): 6, 7, 46*

[44] ——, "Log-cumulant matching approximation of heavy-tailed-distributed aggregate interference," in *Proceedings of the 2015 IEEE International Conference on Communications (ICC)*, Jun 2015, pp. 4811 – 4815. *Cited on page(s): 6, 7, 40, 51, 54*

[45] G. Pastor, I. Norros, R. Jantti, and A. J. Caamano, "Compressive data aggregation from Poisson point process observations," in *Proceedings of the 2015 IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2015, p. TBP. *Cited on page(s): 6, 7*

[46] G. Pastor, I. Mora-Jimenez, A. J. Caamano, and R. Jantti, "Asymptotic expansions for heavy-tailed data," *IEEE Signal Processing Letters (accepted)*, 2016. *Cited on page(s): 6*

[47] E. Candes, "Mathematics of sparsity (and a few other things)," in *Proceedings of the 2014 International Congress of Mathematicians*, 2014. *Cited on page(s): 10, 60*

[48] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1248–1259, Feb 2014. *Cited on page(s): 10*

[49] B. Adcock, A. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: Asymptotic incoherence and asymptotic sparsity in compressed sensing," *CoRR*, vol. abs/1302.0561, 2013. *Cited on page(s): 10*

[50] A. Bastounis and A. C. Hansen, "On the absence of the RIP in real-world applications of compressed sensing and the RIP in levels," *CoRR*, vol. abs/1411.4449, 2014. *Cited on page(s): 10*

[51] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks: Volume I Theory*, ser. Foundations and Trends in Networking. Now Publishers Inc., Mar 2009, vol. 3, no. 3-4. *Cited on page(s): 11, 12*

[52] H. Q. Nguyen, F. Baccelli, and D. Kofman, "A stochastic geometry analysis of dense IEEE 802.11 networks," in *Proceedings of the 2007 IEEE International Conference on Computer Communications (INFOCOM)*, May 2007, pp. 1199–1207. *Cited on page(s): 12, 68*

[53] J. Salo, L. Vuokko, and P. Vainikainen, "Why is shadow fading lognormal?" in *Proceedings of the International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Sept 2005, pp. 522–526. *Cited on page(s): 13*

[54] R. K. Ganti and M. Haenggi, "Interference in ad hoc networks with general motion-invariant node distributions," in *Proceedings of the 2008 IEEE International Symposium on Information Theory (ISIT)*, Jul 2008, pp. 1–5. *Cited on page(s): 14, 39*

[55] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," in *Proceedings of the 2010 IEEE International Conference on Computer Communications (INFOCOM)*, Mar 2010, pp. 1–9. *Cited on page(s): 17, 18, 21, 22*

[56] H. Suyari, "Generalization of Shannon-Khinchin axioms to nonextensive systems and the uniqueness theorem for the nonextensive entropy," *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1783–1787, Aug 2004. *Cited on page(s): 17, 18*

[57] I. Csiszar, "Axiomatic characterizations of information measures," *Entropy*, vol. 10, no. 3, pp. 261–273, 2008. *Cited on page(s): 17, 18, 21, 35*

[58] N. Ebrahimi, E. Maasoumi, and E. Soofi, "Ordering univariate distributions by entropy and variance," *Journal of Econometrics*, vol. 90, no. 2, pp. 317–336, Jun 1999. *Cited on page(s): 17, 33*

[59] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul 1948. *Cited on page(s): 18, 21, 22*

[60] N. R. Pal and J. C. Bezdek, "Measuring fuzzy uncertainty," *IEEE Transactions on Fuzzy Systems*, vol. 2, no. 2, pp. 107–118, May 1994. *Cited on page(s): 18, 22*

[61] J. Martin, G. Mayor-Forteza, and J. Suner, "On dispersion measures," *Mathware and Soft Computing*, vol. 8, no. 3, pp. 227–237, 2001. *Cited on page(s): 18, 21, 22, 81*

[62] L. Yuan and H. Kesavan, "Minimum entropy and information measure," *IEEE Transactions on Human-Machine Systems*, vol. 28, no. 3, pp. 488–491, Aug 1998. *Cited on page(s): 18, 21, 98*

[63] H. Dalton, "The measurement of the inequity of incomes," *Economic Journal*, vol. 30, no. 119, pp. 348–361, Sept 1920. *Cited on page(s): 18, 21, 22, 23*

[64] P. Cheridito and E. Kromer, "Reward-risk ratios," *Journal of Investment Strategies*, Dec 2013. *Cited on page(s): 19, 22*

[65] K. Pearson, "Contribution to the mathematical theory of evolution," *Philosophical Transactions of the Royal Society A*, vol. 185, pp. 71–110, 1894. *Cited on page(s): 25, 40*

[66] C. Gini, "On the measure of concentration with special reference to income and statistics," *Colorado College Publication, General Series*, no. 208, pp. 73–79, 1936. *Cited on page(s): 25*

[67] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, Dec 2004. *Cited on page(s): 25*

[68] H. Theil, *Economics and Information Theory*. Amsterdam, North-Holland Publishing Co., 1967. *Cited on page(s): 25*

[69] A. B. Atkinson, "On the measurement of inequality," *Journal of Economic Theory*, vol. 2, no. 3, pp. 244–263, Sept 1970. *Cited on page(s): 25*

[70] A. Cambini and L. Martein, *Generalized Convexity and Optimization*. Springer, 2009, vol. 616. *Cited on page(s): 26, 31, 33, 81*

[71] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Sparse ICA for blind separation of transmitted and reflected images," *International Journal of Imaging Systems and Technology*, vol. 15, pp. 84–91, 2005. *Cited on page(s): 26*

[72] D. Zonoobi, A. A. Kassim, and Y. V. Venkatesh, "Gini index as sparsity measure for signal reconstruction from compressive samples," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 927–932, Sept 2011. *Cited on page(s): 29*

[73] C. M. Akujuobi, O. O. Odejide, A. Annamalai, and G. L. Fudge, "Sparseness measures of signals for compressive sampling," in *Proceedings of the 2007 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Dec 2007, pp. 1042–1047. *Cited on page(s): 29*

[74] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2011, pp. 233–240. *Cited on page(s): 29*

[75] J. Haberl, "On global minima of semistrictly quasiconcave functions," *Optimization Letters*, vol. 3, no. 3, pp. 387–396, 2009. *Cited on page(s): 29*

[76] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, no. 1-2, 1988. *Cited on page(s): 29*

[77] A. Renyi, "On measures of entropy and information," in *1961 Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. UC Press, 1961, pp. 547–561. *Cited on page(s): 29*

[78] R. V. L. Hartley, "Transmission of information," *Bell System Technical Journal*, vol. 7, no. 3, pp. 535–563, 1928. *Cited on page(s): 29*

[79] R. Renner, "Security of quantum key distribution," *International Journal of Quantum Information*, vol. 6, no. 1, 2008. *Cited on page(s): 29*

[80] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 75–81, Jan 1976. *Cited on page(s): 31*

[81] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences (Mathematics)*, vol. 88, no. 6, pp. 2297–2301, Mar 1991. *Cited on page(s): 31*

[82] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology: Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000. *Cited on page(s): 31*

[83] M. D. Esteban and D. Morales, "A summary of entropy statistics," *Kybernetika*, vol. 31, no. 4, pp. 337–346, 1995. *Cited on page(s): 33*

[84] J. Aczel and Z. Daroczy, "Charakterisierung der entropien positiver ordnung und der Shannonschen entropie," *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 14, no. 1, pp. 95–121, Mar 1963. *Cited on page(s): 33*

[85] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006. *Cited on page(s): 33, 98*

[86] A. Cohen and A. Yeredor, "On the use of sparsity for recovering discrete probability distributions from their moments," in *Proceedings of the 2011 IEEE Statistical Signal Processing Workshop (SSP)*, Jun 2011, pp. 753–756. *Cited on page(s): 37*

[87] M. Pilanci, L. El Ghaoui, and V. Chandrasekaran, "Recovery of sparse probability measures via convex programming," in *Neural Information Processing Systems (NIPS), Advances in*, Dec 2012, pp. 2429–2437. *Cited on page(s): 37*

[88] A. Hormati and M. Vetterli, "Distributed compressed sensing: Sparsity models and reconstruction algorithms using annihilating filter," in *Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar 2008, pp. 5141–5144. *Cited on page(s): 37*

[89] M. Lopes, "Estimating unknown sparsity in compressed sensing," in *Proceedings of the International Conference on Machine Learning (ICML)*, vol. 28, no. 3, May 2013, pp. 217–225. *Cited on page(s): 38, 73*

[90] R. Jantti, J. Kerttula, K. Koufos, and K. Ruttik, "Aggregate interference with FCC and ECC white space usage rules: Case study in Finland," in *Proceedings of the 2011 IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, May 2011, pp. 599–602. *Cited on page(s): 39*

[91] M. Matinmikko, H. Okkonen, M. Palola, S. Yrjola, P. Ahokangas, and M. Mustonen, "Spectrum sharing using licensed shared access: The concept and its workflow for LTE-advanced networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 72–79, Apr 2014. *Cited on page(s): 39*

[92] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers," *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1601–1611, Jun 1998. *Cited on page(s): 39*

[93] M. Kountouris and N. Pappas, "Approximating the interference distribution in large wireless networks," in *Proceedings of the 2014 IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2014, pp. 80–84. *Cited on page(s): 40, 70*

[94] A. Rabbachin, T. Q. S. Quek, Hyundong Shin, and M. Z. Win, "Cognitive network interference," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 480–493, Feb 2011. *Cited on page(s): 40*

[95] K. W. Choi, S. Choi, and J. H. Yun, "On the joint distribution of aggregate interference at multiple wireless receivers," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 3, pp. 1355–1362, Mar 2013. *Cited on page(s): 40, 51*

[96] Y. Dominicy and D. Veredas, "The method of simulated quantiles," *Journal of Econometrics*, vol. 172, no. 2, pp. 235–247, Feb 2013. *Cited on page(s): 40*

[97] R. F. Brcich, D. R. Iskander, and A. M. Zoubir, "The stability test for symmetric alpha-stable distributions," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 977–986, Mar 2005. *Cited on page(s): 40*

[98] G. Cottone and M. Di Paola, "On the use of fractional calculus for the probabilistic characterization of random variables," *Probabilistic Engineering Mechanics*, vol. 24, no. 3, pp. 321–330, 2013. *Cited on page(s): 40*

[99] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "On the method of logarithmic cumulants for parametric probability density function estimation," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3791–3806, Oct 2013. *Cited on page(s): 41*

[100] E. E. Kuruoglu, "Density parameter estimation of skewed α-stable distributions," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2192–2201, Oct 2001. *Cited on page(s): 41*

[101] A. Tagliani, "Recovering a probability density function from its Mellin transform," *Applied Mathematics and Computation*, vol. 118, no. 2-3, pp. 151–159, Mar 2001. *Cited on page(s): 41*

[102] D. L. Wallace, "Asymptotic approximations to distributions," *Annals of Mathematical Statistics*, vol. 29, no. 3, pp. 635–654, Sept 1958. *Cited on page(s): 41*

[103] A. Hald, "The early history of the cumulants and the Gram-Charlier series," *International Statistical Review*, vol. 68, no. 2, pp. 137–153, 2000. *Cited on page(s): 41*

[104] S. Blinnikov and R. Moessner, "Expansions for nearly Gaussian distributions," in *Astronomy and Astrophysics Supplement Series*, 1998, pp. 193–205. *Cited on page(s): 41, 43*

[105] G. W. Hill and W. Davis, "Generalized asymptotic expansions of Cornish-Fisher type," *Annals of Mathematical Statistics*, vol. 39, no. 4, pp. 1264–1273, Aug 1968. *Cited on page(s): 41, 44, 105*

[106] R. A. Fisher and E. A. Cornish, "The percentile points of distributions having known cumulants," *Technometrics*, vol. 2, no. 2, pp. 209–225, May 1960. *Cited on page(s): 45*

[107] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1964, ch. Probability Functions. *Cited on page(s): 45*

[108] J. Bertrand, P. Bertrand, and J. Ovarlez, *The Transforms and Applications Handbook*, 2nd ed. CRC Press LLC, 2000, ch. The Mellin Transform. *Cited on page(s): 46*

[109] W. L. Martinez and A. R. Martinez, *Computational Statistics With MATLAB*. CRC Press LLC, 2002. *Cited on page(s): 50*

[110] T. N. Thiele, *Forelaesninger over almindelig Iagttagelseslaere: Sandsynlighedsregning og mindste kvadraters methode*. C.A. Reitzel, 1889. *Cited on page(s): 50*

[111] I. A. Koutrouvelis, "Regression-type estimation of the parameters of stable laws," *Journal of the American Statistical Association*, no. 75, pp. 918 – 928, Dec 1980. *Cited on page(s): 51*

[112] A. Busson, B. Jabbari, A. Babaei, and V. Veque, "Interference and throughput in spectrum sensing cognitive radio networks using point processes," *Journal of Communications and Networks*, vol. 16, no. 1, pp. 67–80, Feb 2014. *Cited on page(s): 51*

[113] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5024–5037, Oct 2012. *Cited on page(s): 54*

[114] O. Cappe, E. Moulines, J. Pesquet, A. Petropulu, and X. Yang, "Long-range dependence and heavy-tail modeling for teletraffic data," *IEEE Signal Processing Magazine*, vol. 19, no. 3, pp. 14–27, May 2002. *Cited on page(s): 55*

[115] C. Reiss, A. Tumanov, G. R. Ganger, R. H. Katz, and M. A. Kozuch, "Heterogeneity and dynamicity of clouds at scale: Google trace analysis," in *Proceedings of the ACM Symposium on Cloud Computing (SoCC)*, no. 7, 2012, pp. 1–13. *Cited on page(s): 55*

[116] P. Wang and I. F. Akyildiz, "On the origins of heavy-tailed delay in dynamic spectrum access networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 2, pp. 204–217, Feb 2012. *Cited on page(s): 55*

[117] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2281–2295, Jun 2011. *Cited on page(s): 55*

[118] B. G. Lindsay and P. Basak, "Moments determine the tail of a distribution (but not much else)," *American Statistician*, vol. 54, no. 4, pp. 248–251, 2000. *Cited on page(s): 56*

[119] K. R. Rao, D. N. Kim, and J. J. Hwang, *Fast Fourier Transform - Algorithms and Applications*, 1st ed. Springer Publishing Company, Incorporated, 2010. *Cited on page(s): 64*

[120] D. M. Bland, T. I. Laakso, and A. Tarczynski, "Analysis of algorithms for nonuniform-time discrete Fourier transform," in *Proceedings of the 1996 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 2, May 1996, pp. 453–456. *Cited on page(s): 64*

[121] T. Goldstein, C. Studer, and R. Baraniuk, "FASTA: A generalized implementation of forward-backward splitting," *CoRR*, vol. abs/1501.04979, 2015. *Cited on page(s): 67*

[122] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, Mar 2000. *Cited on page(s): 67*

[123] T. V. Nguyen and F. Baccelli, "On the spatial modeling of wireless networks by random packing models," in *Proceedings of the 2012 IEEE International Conference on Computer Communications (INFOCOM)*, Mar 2012, pp. 28–36. *Cited on page(s): 68*

[124] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler, "Stochastic analysis of spatial and opportunistic aloha," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1105–1119, September 2009. *Cited on page(s): 74*

[125] E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010. *Cited on page(s): 74*

[126] T. V. Nguyen and F. Baccelli, "A probabilistic model of carrier sensing based cognitive radio," in *Proceedings of the 2010 IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Apr 2010, pp. 1–12. *Cited on page(s): 106*

Aalto-DD 38/2016

This doctoral thesis is conducted under a convention for the joint supervision of thesis at Aalto University, Finland and Rey Juan Carlos University, Spain (Programa de Doctorado Interuniversitario en Multimedia y Comunicaciones, Escuela Internacional de Doctorado).

Universidad Rey Juan Carlos

ISBN 978-952-60-6681-3 (printed)
ISBN 978-952-60-6682-0 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Communications and Networking
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS