On Capacity-Achieving Input Distributions to Additive Vector Gaussian Noise Channels Under Peak and Even Moment Constraints

by

Jonah Sean Eisen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

We investigate the support of the capacity-achieving input distribution to a vector-valued Gaussian noise channel. The input is subject to a radial even-moment constraint and, in some cases, is additionally restricted to a given compact subset of $\mathbb{R}^n$. Unlike much of the prior work in this field, the noise components are permitted to have different variances and the compact input alphabet is not necessarily a ball. Therefore, the problem considered here is not limited to being spherically symmetric, which forces the analysis to be done in n dimensions.

In contrast to a commonly held belief, we demonstrate that the n-dimensional (real-analytic) Identity Theorem can be used to obtain results in a multivariate setting. In particular, it is determined that when the even-moment constraint is greater than n, or when the input alphabet is compact, the capacity-achieving distribution's support has Lebesgue measure 0 and is nowhere dense in $\mathbb{R}^n$. An alternate proof of this result is then given by exploiting the geometry of the zero set of a real-analytic function. Furthermore, this latter approach is used to show that the support is composed of a countable union of submanifolds, each with dimension $n-1$ or less. In the compact case, the support is a finite union of submanifolds.

Acknowledgements

I would like to thank my supervisors, Professors Mazumdar and Mitran, for their constant support throughout this program. I would also like to thank Professors Alajaji and Yüksel from my undergraduate degree at Queen's University as well as my internship supervisor Susan Watson for motivating me to pursue graduate studies in Communications.

Dedication

To my family and friends.

Table of Contents

List of Figures

Chapter 1

Introduction

Noise is known to impede transmissions from sender to receiver, limiting the rate at which information can be sent. Given an additive stochastic model for the noise, the following question arises: how quickly can information be reliably transmitted through the channel? Unsurprisingly, the answer to this question depends also on the limitations of the transmitter. If arbitrary input signals are allowed, then under an additive Gaussian noise model, the rate of reliable information transfer is unlimited [10]. However, practical transmitters are unable to produce arbitrary signals, so the channel model must be adapted to include input constraints.

The scalar Additive White Gaussian Noise (AWGN) channel under an average power constraint is a classic problem. The optimal input distribution is known to be a zero-mean Gaussian distribution with variance corresponding to the maximum allowable average input power [2, 10]. However, since the output of amplifiers used in transmitters is severely distorted when the input is too large, an unbounded input is impractical [3, 8, 22]. For this reason, in addition to an average power constraint, there is interest in considering channels with bounded input amplitudes. Inputs smaller than a certain threshold can also be difficult for transmitters to produce, further motivating restriction of inputs to more general compact sets.

In many situations, it is necessary to consider vector-valued channels. For instance, quadrature channels use complex-valued inputs and Multiple Input Multiple Output (MIMO) inputs can have n complex components (or equivalently, $2n$ real components). Additionally, noise with memory can be represented by correlated noise components in a vector-valued channel.

Under average power constraints, the optimal input to a channel with additive multivariate Gaussian noise is also multivariate Gaussian [10]. The optimal average power allocation amongst the input components is given by so-called “water-filling”. When the allowable average power is small, power is committed to the component with the lowest noise variance. As the average power increases, water-filling seeks to allocate average power such that the output variance is equal across all components. Amplifier saturation and distortion is an issue for the practicality of this input as well, leading to the consideration of inputs restricted to compact subsets of $\mathbb{R}^n$.

This thesis studies vector-valued inputs subject to additive multivariate Gaussian noise with components that are not necessarily identically distributed. The input is subject to an average even moment radial constraint, which is a generalization of the average power constraint. For example, since a second moment constraint limits variance, a 4th moment constraint limits the variance of the variance. The case of the input also being restricted to an arbitrary compact subset of $\mathbb{R}^n$ is considered as well. Note that, due to technical reasons, if the input is allowed to take arbitrary values in $\mathbb{R}^n$, provided it satisfies a $2k$'th moment constraint, then only the case $2k > n$ is considered.

1.1 Achievable Rates and Capacity

Since noise and transmitter limitations are persistent features of a given channel, it was thought prior to the 1940's that a transmitted message could not be received error-free with arbitrarily high probability [10]. Claude Shannon introduced the concept of adding redundancy to a message to make it less susceptible to channel noise. The ratio of information bits to total uses of the channel is called the rate. Suppose a transmitter wishes to send an arbitrary message from a set indexed by $\{1,\ldots,M\}$ at rate R over a channel that accepts inputs from the set $\mathscr{X}$. Since $\log_2 M$ information bits are needed to describe an index from 1 to M, $n = \log_2 M/R$ channel uses are required. Therefore, rather than send a message directly, an associated codeword consisting of n channel symbols from $\mathscr{X}$ is sent.

To utilize codewords, the transmitter and receiver prearrange a deterministic mapping $x^n(\cdot)$ of the M messages to M codewords. Each codeword consists of n channel symbols, so each codeword belongs to $\mathscr{X}^n$. The mapping $x^n(\cdot)$, known as an encoder, is then defined by

$$x^n : \{1,\ldots,M\} \to \mathscr{X}^n \quad (1.1)$$
$$w \mapsto x^n(w). \quad (1.2)$$

Therefore, the transmitter and receiver agree on a one-to-one correspondence from M messages to the index set $\{1,\ldots,M\}$ and from $\{1,\ldots,M\}$ to M codewords in $\mathscr{X}^n$.

To send message w, the transmitter encodes it as $x^n(w)$. The sequence of symbols $Y^n(w)$, a version of $x^n(w)$ possibly corrupted by noise, arrives at the receiver. The receiver uses a deterministic function known as a decoder to make a best guess $\hat{w}$ of which index w was sent. If

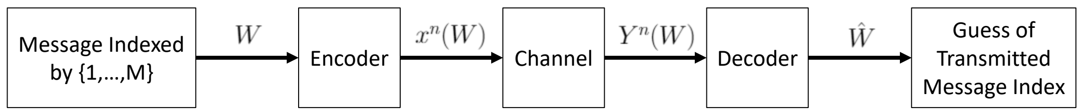

Figure 1.1: Channel Diagram

$\hat{w} \neq w$, then a transmission error is said to have taken place. In general, since the user's behaviour is unknown, message selection and the input codeword are treated as random variables. With the random variable $W \in \{1, \ldots, M\}$ representing the message index, the input is

$$X^n = (X_1, \ldots, X_n) = (x_1^n(W), \ldots, x_n^n(W)). \tag{1.3}$$

The output of the channel is $Y^n = Y^n(W)$ and the message is decoded as $\hat{W} = \hat{W}(Y^n)$. Since the encoding from $\{1, \ldots, M\}$ to $\mathscr{X}^n$ is an injective function, the sequence $W \to X^n \to Y^n$ forms a Markov chain. That is, Y^n is conditionally independent of W given X^n. The channel diagram is illustrated in Figure 1.1.

A rate R is called achievable over a channel if there exists a sequence of codes with increasing block length n and rate R for which the probability of error $P\{W \neq \hat{W}(Y^n)\}$ tends to 0 as n goes to infinity. Shannon showed that for any channel, modeled by the probabilistic relationship between input and output, there exists a fundamental limit to the rate at which reliable communication can be achieved. This limit is called the channel capacity and is denoted by C.

Shannon's Channel Coding Theorem shows that each rate $R < C$ is achievable. It uses the Weak Law of Large Numbers to show that for each $\varepsilon > 0$, provided n is sufficiently large, there exists a coding scheme of rate R for which the probability of error is less than ε. Conversely, any achievable rate must satisfy $R \leq C$. Furthermore, Shannon expressed the capacity as the value obtained by maximizing the mutual information between the channel input X^n and output Y^n [10].

Mutual information is closely related to entropy, which measures the average information one gains by observing the outcome of a discrete random variable. The differential entropy is the analogue of entropy for continuous random variables with densities. The channel model considered in this thesis uses continuous random variables, so our focus is restricted to differential entropy. Section 1.2 introduces mutual information and differential entropy formally.

1.2 Information Measures

The definitions of basic information measures of continuous random variables with densities are briefly recalled here. Let X be a continuous random variable with density p_X on $\mathscr{X} \subseteq \mathbb{R}$. Let Y be a real-valued random variable such that X and Y have joint density $p_{X,Y} = p_X p_{Y|X}$ on $\mathscr{X} \times \mathscr{Y}$. Denote the marginal density of Y by $p(\cdot; p_X)$ to make explicit its dependence on p_X for fixed $p_{Y|X}$. For the purposes of this introduction, it is assumed that each of the integrals in the following discussion are finite.

The differential entropy of X is defined as

$$h(X) \triangleq -\int_{x\in\mathscr{X}} p_X(x) \ln p_X(x)\, dx. \tag{1.4}$$

Differential entropy can be defined with logarithms of any base, resulting in units that differ from each other by constant factors. A base of e is used in this thesis, corresponding to measurement in nats. Since differential entropy is a function only of p_X, it can also be expressed as $h(p_X)$.

The differential entropy of Y conditioned on a particular value $x \in \mathscr{X}$ is given by

$$h(Y \mid X = x) \triangleq -\int_{y\in\mathscr{Y}} p_{Y|X}(y \mid x) \ln p_{Y|X}(y \mid x)\, dy. \tag{1.5}$$

This quantity averaged over $\mathscr{X}$ is called the conditional differential entropy of Y given X:

$$h(Y \mid X) \triangleq -\int_{\mathscr{X}} p_X(x) h(Y \mid X = x)\, dx \tag{1.6}$$

$$= -\int_{(x,y)\in\mathscr{X}\times\mathscr{Y}} p_{X,Y}(x,y) \ln p_{Y|X}(y \mid x)\, dx\, dy. \tag{1.7}$$

The mutual information between X and Y is defined as

$$I(X;Y) = \int_{\mathscr{X}\times\mathscr{Y}} p_{X,Y}(x,y) \ln \frac{p_{X,Y}(x,y)}{p_X(x) p(y; p_X)}\, dx\, dy. \tag{1.8}$$

The assumption of an input density p_X is sometimes too restrictive, as it is in this thesis. Let X have distribution F and suppose that Y has a density induced by F given by

$$p(y;F) = \int_{\mathscr{X}} p_{Y|X}(y \mid x)\, dF(x). \tag{1.9}$$

When $p_{Y|X}$ is sufficiently well-behaved, the mutual information can be expressed as a function of F given by

$$I(F) = I(X;Y) \tag{1.10}$$

$$= \int_{\mathscr{Y}} \int_{\mathscr{X}} p_{Y|X}(y \mid x) \ln \frac{p_{Y|X}(y \mid x)}{p(y;F)} \, dF(x) \, dy \tag{1.11}$$

Mutual information can then be related to differential entropy by

$$I(F) = h(p(\cdot;F)) - h(Y \mid X). \tag{1.12}$$

Consider the AWGN channel given by

$$Y = X + N, \tag{1.13}$$

where $N \sim \mathscr{N}(0, \sigma^2)$ is independent of X. For any $x, y \in \mathbb{R}$, the conditional density of Y given X is $p_{Y|X}(y \mid x) = p_N(y - x)$. Furthermore,

$$h(Y \mid X) = h(X + N \mid X) \tag{1.14}$$

$$= -\int_{\mathscr{X}} \int_{-\infty}^{\infty} p_{N+X|X}(n + x \mid x) \ln p_{N+X|X}(n + x \mid x) \, dn \, dF(x) \tag{1.15}$$

$$= -\int_{\mathscr{X}} \int_{-\infty}^{\infty} p_{N|X}(n \mid x) \ln p_{N|X}(n \mid x) \, dn \, dF(x) \tag{1.16}$$

$$= -\int_{\mathscr{X}} \int_{-\infty}^{\infty} p_N(n) \ln p_N(n) \, dn \, dF(x) \tag{1.17}$$

$$= \int_{\mathscr{X}} h(N) \, dF(x) \tag{1.18}$$

$$= h(N), \tag{1.19}$$

where (1.17) is due to the independence of X and N. The differential entropy of a Gaussian random variable with variance σ^2 is $\frac{1}{2} \ln 2\pi e \sigma^2$ nats and (1.12) becomes

$$I(F) = h(p(\cdot;F)) - \frac{1}{2} \ln 2\pi e \sigma^2. \tag{1.20}$$

1.3 Channel Capacity in Terms of Information Measures

With this understanding of information measures, the discussion regarding the Channel Coding Theorem can be continued. Consider a code of length n transmitted over a memoryless AWGN channel – i.e., for any $x^n, y^n \in \mathbb{R}^n$,

$$p_{Y^n|X^n}(y^n \mid x^n) = \prod_{i=1}^{n} p_{Y|X}(y_i \mid x_i) \tag{1.21}$$

$$= \prod_{i=1}^{n} p_N(y_i - x_i). \tag{1.22}$$

Average and peak power constraints on the input are modeled by restricting input distributions to the set

$$\mathscr{P}(\mathscr{A}, a) = \{F \in \mathscr{F}(\mathbb{R}) \mid F(\mathscr{A}) = 1, \mathbb{E}_{\boldsymbol{X} \sim F}[\|\boldsymbol{X}\|^2] \leq a\}, \tag{1.23}$$

where $\mathscr{F}(\mathbb{R})$ is the set of measures on $\mathbb{R}$, $\mathscr{A} \subseteq \mathbb{R}$ and $a > 0$. To impose a peak power constraint, $\mathscr{A}$ is often chosen to be compact.

The Channel Coding Theorem states that the capacity C of this channel is given by the optimization problem [2, 10]

$$C = \sup_{F \in \mathscr{P}(\mathscr{A}, a)} I(F). \tag{1.24}$$

By (1.20), this can be rewritten as

$$C = \sup_{F \in \mathscr{P}(\mathscr{A}, a)} h(p(\cdot; F)) - \frac{1}{2} \ln 2\pi e \sigma^2 \tag{1.25}$$

It is well known that when $\mathscr{A} = \mathbb{R}$, the capacity-achieving distribution for (1.25) is $X \sim \mathscr{N}(0, a)$ [2, 10]. When $\mathscr{A}$ is compact, this choice of X is inadmissible and the problem becomes far less trivial [28].

A generalization of the scalar AWGN channel is to vector-valued channels. This is motivated by quadrature modulators as well as Multiple Input Multiple Output channels, where multiple antenna are utilized at the transmitter and receiver. The vector AWGN channel takes inputs $\boldsymbol{X} \in \mathbb{R}^n$ and generates outputs according to

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{N}, \tag{1.26}$$

where $\boldsymbol{A}$ is an invertible matrix known to the receiver and $\boldsymbol{N} \sim \mathscr{N}(0, \boldsymbol{\Sigma})$. For some $a > 0$, the input is subject to

$$\mathbb{E}[\|\boldsymbol{X}\|^2] \leq a. \tag{1.27}$$

The extensions of the information measures to vector-valued random variables are as expected and consist of integration over subsets of $\mathbb{R}^n$, rather than $\mathbb{R}$. When $\boldsymbol{\Sigma}$ is non-degenerate,

$$h(\boldsymbol{N}) = \frac{1}{2} \ln |2\pi e \boldsymbol{\Sigma}| \tag{1.28}$$

and the capacity-achieving input is known to be a multivariate Gaussian distribution [10].

This thesis studies the channel given by (1.26), with input possibly restricted to compact sets, and under the constraint

$$\mathbb{E}[\|\boldsymbol{X}\|^{2k}] \leq a, \tag{1.29}$$

where $k \in \mathbb{Z}_{>0}$ and $a > 0$.

Chapter 2 summarizes the related prior work and Chapter 3 provides the main results. Section 3.1 frames the capacity-achieving distribution as the solution to a convex optimization problem. Furthermore, it characterizes the support of the capacity-achieving distribution F^* in terms of the zeros of a particular analytic function, denoted as $s(\cdot; F^*)$. Section 3.2 uses Hilbert space theory and Hermite polynomials to simplify the expression for $s(\cdot; F^*)$. Section 3.3 shows through contradiction that $s(\cdot; F^*)$ is not constant, which implies that the support of the capacity-achieving distribution is not all of $\mathbb{R}^n$. However, the main utility of this observation is as an intermediary result for Sections 3.4 and 3.5. Section 3.4 uses the n-dimensional real-analytic Identity Theorem to show that the support of the capacity-achieving distribution is nowhere dense in $\mathbb{R}^n$ and has Lebesgue measure 0. Section 3.5 shows that the support is contained in a countable union of i-dimensional, $i \in \{0, \ldots, n-1\}$, submanifolds. In the compact case, this union is finite. The geometric results are then used to provide alternate proofs for the results of Section 3.4.

The appendices are used to provide technical support for the discussions of Chapter 3. Appendix A recalls theorems from convex optimization. It also establishes the relationship between the support of the capacity-achieving distribution and the zeros of $s(\cdot; F^*)$. Appendix B shows that the feasible set is convex and compact. Appendix C establishes bounds and integrability results pertaining to the output density induced by the capacity-achieving input. Appendix D uses results from Appendix C to prove the weak continuity, strict concavity and weak differentiability of the optimization problem's objective function. Finally, Appendix E shows that $s(\cdot; F^*)$ is an entire function.

Chapter 2

Prior Work

Dating back to Shannon's work in [25], much of the research on continuous channels has focused on average power (equivalently, second moment) constraints on the input. A transmitter's inability to produce arbitrary peak powers then led to the consideration of additional peak power constraints, modeled by restricting the input almost surely to compact sets.

2.1 Summary of Smith's Contributions

The first major result on amplitude constrained channels was by Smith [28]. The methods used in [28] inspired much of the subsequent literature in the area of channels with amplitude-constrained inputs, including this thesis. A summary of the key ideas is presented here.

In [28], a scalar channel that relates the output Y to input X and Gaussian noise N is given by

$$Y = X + N. \tag{2.1}$$

For some $A > 0$, a peak power constraint is given by $X \in [-A, A]$ almost surely. A second case in the paper considers an additional constraint given by $\mathbb{E}[X^2] \leq \sigma^2$, for some $\sigma^2 > 0$. In each case, Smith determines that the capacity-achieving distribution is discrete with a finite number of mass points. Smith is then able to use this result to numerically search for the capacity under various values of A and σ^2.

Let $\mathscr{P}(A, \sigma^2)$ be the set of distributions that satisfy the amplitude and average cost constraint with parameters A and σ^2, respectively. When the average cost constraint is inactive, the notation $\mathscr{P}(A, \infty)$ will be used.

To avoid assuming the existence of a capacity-achieving density, the mutual information is treated as in (1.10). In retrospect, this more general treatment is evidently necessary since random variables with point masses do not have densities. Substituting the conditional density $p_{Y|X}(y \mid x) = p_N(y-x)$, the mutual information, as a functional on $\mathscr{P}(A,\sigma^2)$, is given by

$$I(F) = \int_{-\infty}^{\infty}\int_{-A}^{A} p_N(y-x)\ln\frac{p_N(y-x)}{p(y;F)}\,dF(x)\,dy, \tag{2.2}$$

where $F \in \mathscr{P}(A,\sigma^2)$. The capacity C can be expressed as

$$C = \sup_{F\in\mathscr{P}(A,\sigma^2)} I(F). \tag{2.3}$$

Consider first the case $\sigma^2 = \infty$. Smith shows that $\mathscr{P}(A,\infty)$ is convex and compact, while $I(\cdot)$ is weak continuous, strictly concave and weak differentiable. By standard convex optimization arguments, there is a unique $F^* \in \mathscr{P}(A,\infty)$ for which $I(F^*) = C$. Moreover, Smith shows that F^* is capacity achieving if and only if for all $F \in \mathscr{P}(A,\infty)$,

$$\int_{-\infty}^{\infty} h(x;F^*)\,dF(x) - C - h(N) \le 0. \tag{2.4}$$

The quantity

$$h(x;F_0) \triangleq -\int_{-\infty}^{\infty} p_N(y-x)\ln p(y;F_0)\,dy. \tag{2.5}$$

is called the marginal entropy density of $F_0 \in \mathscr{P}(A,\infty)$ at $x \in [-A,A]$

From (2.4), Smith proceeds to show that for all $x \in [-A,A]$,

$$h(x;F^*) - C - h(N) \le 0, \tag{2.6}$$

and if x is a point at which F^* increases, then equality holds in (2.6):

$$h(x;F^*) - C - h(N) = 0. \tag{2.7}$$

Therefore, the points at which F^* increases form a subset of the zeros of the function (in x) on the left side of (2.7). This function has an entire extension to the complex plane, so the 1-dimensional Identity Theorem gives conditions on its set of zeros under which it is identically 0 [7].

Theorem 2.1 (1-dimensional Identity Theorem)**.** *Suppose $f(\cdot)$ is analytic on a domain $D \subseteq \mathbb{C}$ and let*

$$Z = \{z \in D \mid f(z) = 0\}. \tag{2.8}$$

If there exists a sequence of distinct points $\{z_i\}_{i=0}^{\infty} \subseteq Z$ *converging to some* $z \in \mathbb{C}$*, then* $Z = D$*.*

In other words, Theorem 2.1 states that if Z has an accumulation point, then $f(\cdot)$ is identically 0 on D.

By way of contradiction, Smith assumes that the zero set Z of the function on the left side of (2.7) has an accumulation point. By the Identity Theorem, this function is identically 0 on $\mathbb{C}$. Using Fourier analysis, Smith shows that this is a contradiction and concludes that Z cannot have an accumulation point. Since the set of points of increase of F^* is a subset of Z, it cannot have an accumulation point either. Furthermore, since all points of increase must be in the compact interval $[-A,A]$, the Bolzano-Weierstrass Theorem implies that F^* must have a finite number of points of increase.

The argument for $\sigma^2 < \infty$ proceeds in a similar fashion to the one above, but requires the use of a Lagrange multiplier. In this case, there exists a constant $\gamma > 0$ such that any $x \in [-A,A]$ must satisfy

$$h(x;F^*) - \gamma(x^2 - \sigma^2) - C - h(N) \leq 0. \tag{2.9}$$

If x is a point of increase of F^*, then

$$h(x;F^*) - \gamma(x^2 - \sigma^2) - C - h(N) = 0. \tag{2.10}$$

Fourier analysis is again used to show that if the extension of the left side of (2.10) to the complex plane is identically 0, then a contradiction arises. It is concluded that F^* has a finite number of points of increase in this case as well.

Determining the capacity-achieving distribution for given A and σ^2 reduces to finding the magnitude and positions of a finite number of point masses. For a given guess of the number of mass points, a finite-dimensional optimization problem is solved. If the resulting solution meets the necessary and sufficient condition of (2.6) and (2.7) (or (2.9) and (2.10) when $\sigma^2 < \infty$), then it is known to be the capacity-achieving distribution. The number of mass points is iterated over until the correct distribution is found.

2.2 Extensions of Smith's Contributions

The results of [28] are extended to channels with 2-dimensional inputs with average and peak radial constraints in [24]. Under the assumption that the noise has independent Gaussian components of equal variance, they note that the phase of the capacity-achieving distribution is uniform over $[0,2\pi)$ and independent of the radius. The authors work in polar coordinates and use the univariate Identity Theorem to find conditions for which the distribution of the input's radius achieves capacity. They conclude that it is optimal to concentrate the input on a finite number of concentric circles.

The case of inputs constrained to arbitrary compact sets and subject to a finite number of quadratic cost constraints as well as non-degenerate multivariate Gaussian noise is considered in [9]. It is concluded that the support of the capacity-achieving distribution must be "sparse". That is, there must exist a non-zero analytic function that is 0 on the support of the capacity-achieving distribution. Assuming otherwise leads to a contradiction by the n-dimensional Identity Theorem and Fourier analysis. These results, while quite general, do not consider either inputs of unbounded support or inputs subject to higher moment constraints. Outside of the special cases of $n = 1$ or spherically symmetric channels, they do not explore a characterization of sparse sets in $\mathbb{R}^n$.

The Identity Theorem for functions of a single complex variable is crucial to Smith's argument and many papers that follow. The theorem can be applied to any univariate entire function that has an accumulation point of zeros. The Identity Theorem in n complex dimensions can only be applied to an analytic function with an open set of zeros in $\mathbb{C}^n$. Therefore, to apply the Identity Theorem for $n > 1$, an analogue of (2.7) would need to hold on an open subset of $\mathbb{C}^n$. It was suspected by some authors that since $\mathbb{R}^n$ is not open in $\mathbb{C}^n$, no topological assumption on the support of the capacity-achieving distribution would be sufficient for this purpose [13, 23, 29]. Therefore, many papers restrict their models to ones that maintain spherical symmetry so that the 1-dimensional results can be exploited by working with the distribution of the input's radius (eg. [12, 14, 23, 24]).

It is shown in [23] that the optimal distribution under peak and quadratic average constraints in $\mathbb{R}^n$ is concentrated on a finite number of concentric shells. This result is obtained by extending the methods of [24]. When the average power constraint is removed, a closed form approximation for capacity is found for n sufficiently large.

In [12, 14], the number and positions of optimal concentric shells under a peak radial constraint are studied. In [12], the properties of subharmonic functions are employed to find the least restrictive amplitude constraint for which the optimal distribution is concentrated on a single sphere. In [14], Karlin's Oscillation Theorem is used in conjunction with conditions like (2.6) and (2.7), to find an upper bound on the number of shells that grows quadratically with the amplitude constraint. A similar result is found for $n = 1$ under an additional average power constraint.

In [15], MIMO channels of the form

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N} \tag{2.11}$$

are considered, where $\boldsymbol{H} \in \mathbb{R}^{m\times n}$ is constant and $\boldsymbol{N} \sim \mathcal{N}(0,\boldsymbol{I}_m)$. The input $\boldsymbol{X}$ is restricted to a compact set, but no average power constraints are considered. Using the Real-Analytic Identity Theorem and steps similar to [28], it is determined that the support of the optimal input distribution is nowhere dense in $\mathbb{R}^n$. The notion of nowhere dense is discussed in detail in relation to the main results of this thesis in Chapter 3. The Real-Analytic Identity Theorem is also employed to show that the support of the optimal input distribution has Lebesgue measure 0. When the marginal entropy density $h(\cdot;F^*)$, defined analogously to that defined in [28], is spherically symmetric, the support is composed of a finite number of concentric shells. With $\boldsymbol{H} = \boldsymbol{I}_n$, this result coincides with [23]. For the case considered in [9] that coincides with this setup, [15] gives an instance of sparsity in terms of subsets of $\mathbb{R}^n$, rather than analytic functions.

Subsequent sections of [15] give upper bounds on the size of the input space and conditions on $\boldsymbol{H}$ for which the optimal distribution is concentrated only on the boundary of the input space. Finally, [15] gives several upper and lower bounds on capacity.

In [17], a scalar channel with input subject to a combination of even moment constraints and restrictions to compact or non-negative subsets of $\mathbb{R}$ is studied. Since the ideas used for the treatment of even moment constraints are expanded to n dimensions in this thesis, a summary is given for that case here. With $f(x) = x^l$, $l \in \mathbb{Z}_{\geq 0}$, and $N \sim \mathcal{N}(0,1)$, the output of the channel is given by

$$Y = f(X) + N. \tag{2.12}$$

When the constraint $\mathbb{E}[X^{2k}] \leq a$ is imposed, optimality conditions give rise to

$$h(f(x);F^*) - C - h(N) - \gamma(x^{2k} - a) \leq 0, \tag{2.13}$$

and, when x is a point of increase of F^*, equality must hold. Now consider the weighted L^2 space

$$L^2_{p_N}(\mathbb{R}) \triangleq \{g : \mathbb{R} \to \mathbb{R} \mid \int_{-\infty}^{\infty} |g(x)|^2 p_N(x)\, dx < \infty\} \tag{2.14}$$

equipped with inner product

$$\langle g, r\rangle = \int_{-\infty}^{\infty} g(x) r(x) p_N(x)\, dx. \tag{2.15}$$

In [28], it is shown that the marginal entropy density of F^* at $f(x)$ can be expressed as

$$h(f(x);F^*) = -\int_{-\infty}^{\infty} p_N(y) e^{-\frac{f(x)^2}{2} + yf(x)} \ln p(y;F^*)\, dy. \tag{2.16}$$

The Hermite polynomials $\{H_m\}_{m=0}^{\infty}$ form an orthogonal basis for $L^2_{p_N}(\mathbb{R})$ and are given by the generating function

$$e^{-\frac{u^2}{2}+yu} = \sum_{m=0}^{\infty} H_m(y)\frac{u^m}{m!}. \tag{2.17}$$

Since, as shown in [28], $\ln p(y;F^*) \in L^2_{p_N}(\mathbb{R})$, $\ln p(y;F^*)$ can be expressed as a linear combination of Hermite polynomials:

$$\ln p(y;F^*) = \sum_{m=0}^{\infty} c_m H_m(y). \tag{2.18}$$

Substituting (2.17) and (2.18) into (2.16) and using the orthogonality of the Hermite polynomials,

$$h(f(x);F^*) = \sum_{m=0}^{\infty} c_m f(x)^m = \sum_{m=0}^{\infty} c_m x^{lm}. \tag{2.19}$$

The existence of an accumulation point of the points of increase of F^* is assumed. Since the left side of (2.13) has an analytic extension to an entire function on $\mathbb{C}$, the Identity Theorem states that this extension is identically 0 on $\mathbb{C}$. Substituting (2.19) into (2.13) allows solving for the values of the constants c_m, $m \in \mathbb{Z}_{\geq 0}$ by coefficient matching.

When $k = l$, the constraint is analogous to a quadratic average cost constraint on X^l. If the input is allowed to take arbitrary values in $\mathbb{R}$, then this channel reflects the classic AWGN channel. In this case, no contradiction arises if X^l can be chosen to have a Gaussian density. This can occur if and only if l is odd, allowing X^l to take both positive and negative values in $\mathbb{R}$.

For other values of the pair (k,l), rather than computing Fourier transforms directly, as is done in [28], a contradiction is derived using Hardy's Theorem. The authors upper bound the Fourier transform of the output density, then use Hardy's Theorem to lower bound the output density. For many pairs (k,l), this lower bound is violated by the values of c_m found by coefficient matching. Therefore, (2.13) cannot hold everywhere and the points of increase of F^* must not have an accumulation point.

In [16], a complex-valued non-dispersive optical channel is considered under average costs that grow super-quadratically in radius, peak constraints, or both. The noise is taken to be circularly symmetric and under these conditions, so is the optimal input. The number of concentric circles composing the support of the distribution is shown to be finite. The results are obtained by looking at the limiting behaviour of the analytically extended optimality condition as the amplitude increases.

In this thesis, we study an n-dimensional channel subject to non-degenerate Gaussian noise. The input is sometimes restricted to a compact subset of $\mathbb{R}^n$ and its norm is subject to even moment constraints. The possible differences in variance between noise components results in a channel that is not spherically symmetric. This problem is addressed through the use of two alternative approaches:

1. application of the real-analytic Identity Theorem in n-dimensions, and
2. utilization of the structure of zero sets of real-analytic functions.

The novel contributions of this thesis are the results concerning the support of capacity-achieving distribution in cases with even moment constraints greater than 2. Similarly to [15], the results are in terms of subsets of $\mathbb{R}^n$, rather than the notion of sparsity used in [9]. It is shown that the support of the capacity-achieving distribution is nowhere dense in $\mathbb{R}^n$ and has Lebesgue measure 0. Furthermore, the geometry of the zero set of a real analytic function is used to prove that the support is contained in a countable union of i-dimensional, $i \in \{0, \ldots, n-1\}$, submanifolds. This union is finite when the input alphabet is compact. Some of the methods employed here have either no or limited exposure in this research community and may be used in future study of spherically asymmetric channels.

Chapter 3

Even-moment Radial Constraints

In this chapter, we consider $\mathbb{R}^n$-valued inputs subject to additive non-degenerate multivariate Gaussian noise. In Section 3.1, the capacity-achieving distribution is established as the objective of an optimization problem. Its support is then framed in terms of the zero set of a certain real-analytic function. Section 3.2 extends to $\mathbb{R}^n$ ideas regarding Hermite bases of Hilbert spaces used in [17] to simplify the real-analytic function obtained in Section 3.1. Section 3.3 shows that, other than in a trivial case, this function is non-constant. It is then shown that the support of the capacity-achieving input has Lebesgue measure 0 and is "nowhere dense". These results are obtained in 2 different ways – Section 3.4 uses the n-dimensional real-analytic Identity Theorem and Section 3.5 makes geometric arguments. Section 3.5 also uses the geometric properties of zero sets of real-analytic functions to show that the support is contained in a countable union of submanifolds of dimensions $0, \ldots, n-1$.

As a first step towards defining the set of feasible input distributions, let $\mathscr{F}(\mathbb{R}^n)$ be the set of finite Borel measures on $\mathbb{R}^n$. Note that $\mathscr{F}(\mathbb{R}^n)$ is contained in the set of finite signed Borel measures on $\mathbb{R}^n$, which has an intrinsic vector space structure and can be equipped with a norm [5]. Since $\mathscr{F}(\mathbb{R}^n)$ lies within a normed vector space, the convexity and compactness of its subsets can be discussed.

The possibility that the transmitter is unable to produce arbitrary signals in $\mathbb{R}^n$ is modeled by restricting the input to an alphabet $\mathscr{A} \subseteq \mathbb{R}^n$. Denote the set of distributions for which the associated random variable is almost surely in $\mathscr{A}$ by

$$\mathscr{F}_n(\mathscr{A}) \triangleq \{F \in \mathscr{F}(\mathbb{R}^n) \mid F(\mathscr{A}) = 1\}. \tag{3.1}$$

Two cases for $\mathscr{A}$ are considered:

1. $\mathscr{A} = \mathbb{R}^n$, and
2. $\mathscr{A} \in \mathscr{C}_n$, where $\mathscr{C}_n$ is the set of compact subsets of $\mathbb{R}^n$.

In addition to the restriction to $\mathscr{A}$, a radial even-moment constraint is associated with the input. For $k \in \mathbb{Z}_{>0}$, the input must belong to the set

$$\mathscr{P}_n(\mathscr{A}, k, a) = \{F \in \mathscr{F}_n(\mathscr{A}) \mid \mathbb{E}_{\boldsymbol{X}\sim F}[\|\boldsymbol{X}\|^{2k}] \leq a\}. \tag{3.2}$$

Due to an integrability condition presented in the appendix (see Lemma C.8), when $\mathscr{A} = \mathbb{R}^n$, we only consider the case that $2k > n$. That is, given $n \geq 1$, we assume throughout this thesis that the pair $(\mathscr{A}, k)$ satisfies one of the following assumptions:

1. $\mathscr{A} \in \mathscr{C}_n$, or
2. $\mathscr{A} = \mathbb{R}^n$ and $2k > n$.

The resulting channel model, with input $\boldsymbol{X} \sim F \in \mathscr{P}_n(\mathscr{A}, k, a)$, is

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{N}, \tag{3.3}$$

where $\boldsymbol{Y}$ and $\boldsymbol{N} \sim \mathscr{N}(0, \boldsymbol{\Sigma})$ are output and noise, respectively and $\boldsymbol{A}$ is an invertible matrix known to the receiver. It is assumed that the noise covariance matrix $\boldsymbol{\Sigma}$ is positive-definite.

Note that by multiplying the output by $\boldsymbol{A}^{-1}$, the receiver obtains

$$\tilde{\boldsymbol{Y}} \triangleq \boldsymbol{A}^{-1}\boldsymbol{Y} \tag{3.4}$$

$$= \boldsymbol{X} + \boldsymbol{A}^{-1}\boldsymbol{N}. \tag{3.5}$$

Since $\boldsymbol{A}^{-1}\boldsymbol{N} \sim \mathscr{N}(0, \boldsymbol{A}^{-1}\boldsymbol{\Sigma}(\boldsymbol{A}^{-1})^T)$ satisfies the assumptions made on the noise, no generality is lost by setting $\boldsymbol{A} = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Substituting this into (3.3) yields the channel model

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}. \tag{3.6}$$

3.1 Optimization Problem

Motivated by the Channel Coding Theorem[10], we would like to solve the optimization problem

$$C = \sup_{\boldsymbol{X}\sim F \in \mathscr{P}_n(\mathscr{A},k,a)} I(\boldsymbol{X};\boldsymbol{Y}) = \sup_{\substack{F \in \mathscr{F}_n(\mathscr{A}) \\ \mathbb{E}_{\boldsymbol{X}\sim F}[\|\boldsymbol{X}\|^{2k}] \leq a}} I(\boldsymbol{X};\boldsymbol{Y}). \tag{3.7}$$

Since $\boldsymbol{N}$ has a fixed distribution and the function relating $\boldsymbol{Y}$ to $\boldsymbol{X}$ and $\boldsymbol{N}$ is deterministic, the mutual information is a function of the distribution of $\boldsymbol{X}$ alone. Thus, the mutual information induced between $\boldsymbol{X} \sim F$ and $\boldsymbol{Y}$ will be denoted by $I(F)$. Similarly, it is useful to think of the even moment constraint in terms of a functional $g : \mathscr{F}_n(\mathscr{A}) \to \mathbb{R} \cup \{\infty\}$ given by

$$g(F) \triangleq \int_{\mathscr{A}} \|\boldsymbol{x}\|^{2k} dF(\boldsymbol{x}) - a. \tag{3.8}$$

Note that $g(F) \leq 0$ is equivalent to $\mathbb{E}_{\boldsymbol{X} \sim F}[\|\boldsymbol{X}\|^{2k}] \leq a$. Rewriting (3.7) in terms of $I(\cdot)$ and $g(\cdot)$ yields

$$C = \sup_{F \in \mathscr{P}_n(\mathscr{A},k,a)} I(F) = \sup_{\substack{F \in \mathscr{F}_n(\mathscr{A}) \\ g(F) \leq 0}} I(F). \tag{3.9}$$

Much of the appendix is dedicated to understanding properties of the problem presented in (3.9). It is shown in Theorem B.1 that $\mathscr{P}_n(\mathscr{A},k,a)$ is convex and compact. Furthermore, by Theorems D.1 and D.2, $I(\cdot)$ is a weak continuous and strictly concave function on $\mathscr{P}_n(\mathscr{A},k,a)$. Therefore, by Theorem A.1, the supremum is achieved by a unique input distribution $F^* \in \mathscr{P}_n(\mathscr{A},k,a)$. That is,

$$C = \max_{F \in \mathscr{P}_n(\mathscr{A},k,a)} I(F) = \max_{\substack{F \in \mathscr{F}_n(\mathscr{A}) \\ g(F) \leq 0}} I(F) = I(F^*). \tag{3.10}$$

We use the notation $\boldsymbol{X}^*$ to describe a capacity-achieving input directly (ie. $\boldsymbol{X}^* \sim F^*$).

In the remainder of this section, we derive, via the optimality condition of Theorem A.3, the condition on the support of F^* given by (**P.2**) of Theorem A.4. However, applying Theorem A.3 to the problem in (3.10) yields an analogue of (**P.1**) which holds only for $F \in \mathscr{P}_n(\mathscr{A},k,a)$ and does not necessarily imply (**P.2**). That is, the constraint $g(F) \leq 0$ is too strict to proceed with (3.10) directly.

The theory of Lagrange multipliers can be utilized to reformulate (3.10) as an unconstrained problem over a larger space. However, the optimality condition of Theorem A.3 requires the weak differentiability of the functionals $I(\cdot)$ and $g(\cdot)$ in this larger space. When $\mathscr{A} = \mathbb{R}^n$, this requirement is not satisfied by $I(\cdot)$ and $g(\cdot)$ on $\mathscr{F}_n(\mathbb{R}^n)$. Instead, we define

$$\mathscr{Q}_n(\mathscr{A},k) \triangleq \bigcup_{b \geq a} \mathscr{P}_n(\mathscr{A},k,b) \tag{3.11}$$

$$= \{F \in \mathscr{F}_n(\mathscr{A}) \mid g(F) < \infty\}. \tag{3.12}$$

and consider the problem given by

$$C = \max_{\substack{F \in \mathscr{Q}_n(\mathscr{A},k) \\ g(F) \leq 0}} I(F), \tag{3.13}$$

where we note that (3.10) and (3.13) have the same objective value and maximizing F^*.

To apply Theorem A.2, note that by Theorem D.3, $g(\cdot)$ is convex. Moreover, letting $F_s \in \mathscr{Q}_n(\mathscr{A},k)$ be a Heaviside step function at $\mathbf{0} \in \mathbb{R}^n$, F_s satisfies the Slater condition since $g(F_s) = -a < 0$. Then, Theorem A.2 implies the existence of $\gamma \geq 0$ such that

$$C = \max_{F \in \mathscr{Q}_n(\mathscr{A},k)} \{J_\gamma(F)\} = J_\gamma(F^*), \tag{3.14}$$

where

$$J_\gamma(F) = I(F) - \gamma g(F). \tag{3.15}$$

Furthermore, for arbitrary $b \geq a$, $F^* \in \mathscr{P}_n(\mathscr{A},k,b)$ and $\mathscr{P}_n(\mathscr{A},k,b) \subseteq \mathscr{Q}_n(\mathscr{A},k)$. Therefore, for this choice of γ, we also have

$$C = \max_{F \in \mathscr{P}_n(\mathscr{A},k,b)} \{J_\gamma(F)\} = J_\gamma(F^*). \tag{3.16}$$

We derive the optimality condition for (3.16) and use the result to apply Theorem A.4, but we must first define some notation: for $F_0 \in \mathscr{F}_n(\mathscr{A})$, the output entropy is given by

$$h_{\boldsymbol{Y}}(F_0) \triangleq -\int_{\mathbb{R}^n} p(\boldsymbol{y};F_0) \ln p(\boldsymbol{y};F_0)\, d\boldsymbol{y} \tag{3.17}$$

and the marginal entropy density at $\boldsymbol{x} \in \mathscr{A}$ is given by

$$h(\boldsymbol{x};F_0) \triangleq -\int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \ln p(\boldsymbol{y};F_0)\, d\boldsymbol{y}, \tag{3.18}$$

whenever the integrals exist. Now, Theorems D.4 and D.5 show that $J_\gamma(\cdot)$ has a weak derivative at F^* in the direction of $F \in \mathscr{P}_n(\mathscr{A},k,b)$ given by

$$J'_\gamma(F^*,F) = I'(F^*,F) - \gamma g'(F^*,F) \tag{3.19}$$

$$= \int_{\mathbb{R}^n} h(\boldsymbol{x};F^*)\, dF(\boldsymbol{x}) - h_{\boldsymbol{Y}}(F^*) - \gamma(g(F) - g(F^*)), \tag{3.20}$$

The expression given in (3.20) can be simplified by noting that the Lagrange multiplier γ

given in Theorem A.2 must satisfy $\gamma g(F^*) = 0$. Observing, in addition, that $C = h_{\boldsymbol{Y}}(F^*) - h(\boldsymbol{N})$ gives

$$J'_\gamma(F^*,F) = \int_{\mathbb{R}^n} h(\boldsymbol{x};F^*)\,dF(\boldsymbol{x}) - C - h(\boldsymbol{N}) - \gamma g(F). \tag{3.21}$$

Note that the differential entropy of the noise, $h(\boldsymbol{N}) = \frac{1}{2}\ln|2\pi e \boldsymbol{\Sigma}|$, is finite since $\boldsymbol{\Sigma}$ is positive definite.

By Theorems D.2 and D.3 and since $\gamma \geq 0$, $J_\gamma(\cdot)$ is the difference between a strictly concave function and a convex function. Therefore, $J_\gamma(\cdot)$ is strictly concave. By the optimality necessary and sufficient condition for optimality presented in Theorem A.3, F^* is optimal if and only if for all $F \in \mathscr{P}_n(\mathscr{A},k,b)$,

$$J'_\gamma(F^*,F) = \int_{\mathbb{R}^n} h(\boldsymbol{x};F^*)\,dF(\boldsymbol{x}) - C - h(\boldsymbol{N}) - \gamma g(F) \leq 0. \tag{3.22}$$

However, $b \geq a$ is arbitrary and each $F \in \mathscr{Q}_n(\mathscr{A},k)$ satisfies $F \in \mathscr{P}_n(\mathscr{A},k,b)$ for some $b \geq a$. Therefore, we observe that F^* is optimal if and only if for all $F \in \mathscr{Q}_n(\mathscr{A},k)$,

$$J'_\gamma(F^*,F) = \int_{\mathbb{R}^n} h(\boldsymbol{x};F^*)\,dF(\boldsymbol{x}) - C - h(\boldsymbol{N}) - \gamma g(F) \leq 0. \tag{3.23}$$

Before proceeding, we must formally introduce the notion of the support of a random variable.

Definition 3.1. Let $\boldsymbol{V}$ be a random variable with alphabet $\mathscr{A} \subseteq \mathbb{R}^n$. Then the support of $\boldsymbol{V}$ is the set given by

$$supp(\boldsymbol{V}) \triangleq \{\boldsymbol{x} \in \mathscr{A} \mid \forall r > 0, P\{\boldsymbol{V} \in B_r(\boldsymbol{x})\} > 0\}, \tag{3.24}$$

where $B_r(\boldsymbol{x})$ is the ball of radius r in $\mathbb{R}^n$ centered at $\boldsymbol{x}$. If $\boldsymbol{V}$ has distribution $F_{\boldsymbol{V}}$, we may alternatively refer to $supp(F_{\boldsymbol{V}}) = supp(\boldsymbol{V})$.

Similarly to the 1-dimensional channels considered in [1, 17, 28], we use the optimality condition to obtain a characterization of $supp(F^*)$. By (3.23), F^* satisfies (**P.1**) of Theorem A.4, so for all $\boldsymbol{x} \in \mathscr{A}$,

$$\gamma(\|\boldsymbol{x}\|^{2k} - a) + C + h(\boldsymbol{N}) + \int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y} \geq 0, \tag{3.25}$$

and if $\boldsymbol{x} \in supp(F^*)$, then

$$\gamma(\|\boldsymbol{x}\|^{2k} - a) + C + h(\boldsymbol{N}) + \int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y} = 0. \tag{3.26}$$

In other words, (3.26) provides a necessary condition for $\boldsymbol{x} \in supp(F^*)$ in terms of the zeros of a function of $\boldsymbol{x}$. Furthermore, this function's extension to $\mathbb{C}^n$ is entire since $\|\cdot\|^{2k}$ is entire and, by Lemma E.1, so is the integral term. Therefore, the function $s(\cdot;F^*) : \mathbb{R}^n \to \mathbb{R}$ given by

$$s(\boldsymbol{x};F^*) \triangleq \gamma(\|\boldsymbol{x}\|^{2k} - a) + C + h(\boldsymbol{N}) + \int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \ln p(\boldsymbol{y};F^*)\, d\boldsymbol{y} \tag{3.27}$$

is real-analytic in $\boldsymbol{x}$ on $\mathbb{R}^n$ and

$$supp(F^*) \subseteq \{\boldsymbol{x} \in \mathbb{R}^n \mid s(\boldsymbol{x};F^*) = 0\} \cap \mathscr{A}. \tag{3.28}$$

The rest of this chapter is dedicated to studying $supp(F^*)$ through the properties of $s(\cdot;F^*)$ and, in particular, its zero set.

3.2 Hilbert Space and Hermite Polynomial Representation

To examine (3.27) further, it is helpful to think of the integral term as an inner product in a Hilbert space. For Borel-measurable weighting function $w : \mathbb{R}^n \to \mathbb{R}$, define

$$L^2_w(\mathbb{R}^n) \triangleq \left\{ \xi : \mathbb{R}^n \to \mathbb{R} \,\middle|\, \int_{\mathbb{R}^n} \xi^2(\boldsymbol{x}) w(\boldsymbol{x})\, d\boldsymbol{x} < \infty \right\}, \tag{3.29}$$

equipped with inner product

$$\langle \xi, \psi \rangle \triangleq \int_{\mathbb{R}^n} \xi(\boldsymbol{x}) \psi(\boldsymbol{x}) w(\boldsymbol{x})\, d\boldsymbol{x}. \tag{3.30}$$

To simplify notation in this section, we define the following operations on vectors: for $\boldsymbol{m} \in \mathbb{Z}^n_{\geq 0}$ and $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{m}! \triangleq m_1! \cdots m_n! \tag{3.31}$$

and

$$\boldsymbol{x}^{\boldsymbol{m}} \triangleq x_1^{m_1} \cdots x_n^{m_n}. \tag{3.32}$$

The notion of orthogonality in a Hilbert space is crucial for representing an unknown function in terms of known ones with nice properties. When $n = 1$, a complete orthogonal family of polynomials can be found by orthogonalizing the monomials $\{x^m\}_{m \in \mathbb{Z}_{\geq 0}}$. This sequence has an essentially unique ordering. In contrast, when $n > 1$, the monomials of $L^2_w(\mathbb{R}^n)$ are $\{\boldsymbol{x}^{\boldsymbol{m}}\}_{\boldsymbol{m} \in \mathbb{Z}^n_{\geq 0}}$

and form a so-called multisequence. Such a multisequence does not have a unique ordering and distinct orderings yield different orthogonal systems that are asymmetric in their arguments [4]. As a result, much of the analysis of Hilbert spaces of functions of several variables is done with biorthogonal systems.

A biorthogonal system comprised of $\{\xi_{\boldsymbol{m}}\}_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}}$ and $\{\psi_{\boldsymbol{m}}\}_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}}$ is one for which there exist some non-zero constants $\{\kappa_{\boldsymbol{m}}\}_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}}$ such that

$$\langle \xi_{\boldsymbol{m}}, \psi_{\boldsymbol{n}} \rangle = \begin{cases} \kappa_{\boldsymbol{m}}, & \text{if } \boldsymbol{m} = \boldsymbol{n} \\ 0, & \text{otherwise.} \end{cases} \tag{3.33}$$

The flexibility of a biorthogonal system allows us to find basis functions that are symmetrical in their arguments.

Consider $L^2_{p_{\boldsymbol{N}}}(\mathbb{R}^n)$, where

$$p_{\boldsymbol{N}}(\boldsymbol{n}) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\boldsymbol{n}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{n}} \tag{3.34}$$

is the density of $\boldsymbol{N}$. For the matrix $\boldsymbol{\Sigma}^{-1}$, define the Hermite polynomials $\{H_{\boldsymbol{m}}\}_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}}$ and $\{G_{\boldsymbol{m}}\}_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}}$ respectively by the generating functions

$$e^{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}} = \sum_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}} \frac{\boldsymbol{x}^{\boldsymbol{m}}}{\boldsymbol{m}!} H_{\boldsymbol{m}}(\boldsymbol{y}) \tag{3.35}$$

and

$$e^{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{y}} = \sum_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}} \frac{\boldsymbol{x}^{\boldsymbol{m}}}{\boldsymbol{m}!} G_{\boldsymbol{m}}(\boldsymbol{y}). \tag{3.36}$$

Then $\{H_{\boldsymbol{m}}\}_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}}$ and $\{G_{\boldsymbol{m}}\}_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}}$ are complete and biorthogonal with [4]

$$\langle G_{\boldsymbol{m}}, H_{\boldsymbol{n}} \rangle = \begin{cases} \boldsymbol{m}!, & \text{if } \boldsymbol{m} = \boldsymbol{n} \\ 0, & \text{otherwise.} \end{cases} \tag{3.37}$$

Furthermore, for each $\boldsymbol{m} \in \mathbb{Z}^n_{\geq 0}$, $H_{\boldsymbol{m}}(\boldsymbol{x})$ and $G_{\boldsymbol{m}}(\boldsymbol{x})$ are of degree m_i in x_i. It should be noted that these Hermite polynomials are dependent on the matrix $\boldsymbol{\Sigma}^{-1}$. However, the discussion here is limited to a fixed system generated by a fixed matrix, so this dependence is omitted from the chosen notation.

Since, by Theorem C.1, $\ln p(\cdot;F^*) \in L^2_{p_{\boldsymbol{N}}}(\mathbb{R}^n)$, there exist constants $\{c_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{Z}^n_{\geq 0}}$ for which

$$\ln p(\boldsymbol{y};F^*) = \sum_{\boldsymbol{k}\in\mathbb{Z}^n_{\geq 0}} c_{\boldsymbol{k}} G_{\boldsymbol{k}}(\boldsymbol{y}). \tag{3.38}$$

Returning to the integral term in (3.27),

$$\int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \ln p(\boldsymbol{y};F^*)\, d\boldsymbol{y} = \int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}) e^{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}} \sum_{\boldsymbol{k}\in\mathbb{Z}^n_{\geq 0}} c_{\boldsymbol{k}} G_{\boldsymbol{k}}(\boldsymbol{y})\, d\boldsymbol{y} \tag{3.39}$$

$$= \int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}) \sum_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}} \frac{\boldsymbol{x}^{\boldsymbol{m}}}{\boldsymbol{m}!} H_{\boldsymbol{m}}(\boldsymbol{y}) \sum_{\boldsymbol{k}\in\mathbb{Z}^n_{\geq 0}} c_{\boldsymbol{k}} G_{\boldsymbol{k}}(\boldsymbol{y})\, d\boldsymbol{y} \tag{3.40}$$

$$= \sum_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}} \frac{\boldsymbol{x}^{\boldsymbol{m}}}{\boldsymbol{m}!} \sum_{\boldsymbol{k}\in\mathbb{Z}^n_{\geq 0}} c_{\boldsymbol{k}} \langle H_{\boldsymbol{m}}(\boldsymbol{y}), G_{\boldsymbol{k}}(\boldsymbol{y})\rangle \tag{3.41}$$

$$= \sum_{\boldsymbol{m}\in\mathbb{Z}^n_{\geq 0}} c_{\boldsymbol{m}} \boldsymbol{x}^{\boldsymbol{m}}. \tag{3.42}$$

This simplification to a polynomial is particularly helpful since the cost function associated with the even moment constraint is also a polynomial. This relationship is exploited in Section 3.3.

3.3 Non-constancy of $s(\cdot;F^*)$

Since any $\boldsymbol{x} \in supp(F^*)$ must satisfy $s(\boldsymbol{x};F^*) = 0$, where $s(\cdot;F^*)$ is defined in (3.27), it is useful to study the zeros of $s(\cdot;F^*)$. Note that since $supp(F^*) \neq \emptyset$,

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid s(\boldsymbol{x};F^*) = 0\} \neq \emptyset. \tag{3.43}$$

Therefore, $s(\cdot;F^*)$ being constant is equivalent to

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid s(\boldsymbol{x};F^*) = 0\} = \mathbb{R}^n. \tag{3.44}$$

In this section, it is shown through contradiction that, in all non-trivial cases, (3.44) does not hold. To that end, suppose that for all $\boldsymbol{x} \in \mathbb{R}^n$, $s(\boldsymbol{x};F^*) = 0$. Substituting (3.42) into (3.27), this is equivalent to

$$[\gamma a - C - h(\boldsymbol{N})] - \gamma\left(\sum_{i=1}^{n} x_i^2\right)^k = \sum_{\boldsymbol{m}\in\mathbb{Z}_{\geq 0}^n} c_{\boldsymbol{m}} x_1^{m_1}\dots x_n^{m_n}, \tag{3.45}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. The discussion proceeds in two cases: $k = 1$ and $k > 1$.

Case: $k = 1$**.** Note that for $n > 1$, since we require $2k > n$ when $\mathscr{A} = \mathbb{R}^n$, we only consider $\mathscr{A} \in \mathscr{C}_n$. However, the case of $\mathscr{P}_n(\mathbb{R}^n, 1, a)$ is a classic problem and $\boldsymbol{X}^*$ is known to be Gaussian [2]. Therefore, this case comprises 2 subcases for which most of the analysis is identical:

S.1 $n \geq 1$ and $\mathscr{A} \in \mathscr{C}_n$, and

S.2 $n = k = 1$ and $\mathscr{A} = \mathbb{R}$

Note that subcase (**S.2**) is the classic AWGN channel and is excluded from our main results, but it is mentioned briefly here for completeness.

Let $\boldsymbol{e}_i$ be the i'th row of the $n \times n$ identity matrix and let $\boldsymbol{0} \in \mathbb{Z}_{\geq 0}^n$ be the all zero vector. Since (3.45) holds for all $\boldsymbol{x} \in \mathbb{R}^n$, matching coefficients gives

$$c_{\boldsymbol{m}} = \begin{cases} \gamma a - C - h(\boldsymbol{N}), & \text{if } \boldsymbol{m} = \boldsymbol{0} \\ -\gamma, & \text{if } \boldsymbol{m} = 2\boldsymbol{e}_i,\ i \in \{1,\dots,n\} \\ 0, & \text{otherwise.} \end{cases} \tag{3.46}$$

Substituting this into (3.38) yields

$$\ln p(\boldsymbol{y}; F^*) = c_{\boldsymbol{0}} G_{\boldsymbol{0}}(\boldsymbol{y}) + \sum_{i=1}^{n} c_{2\boldsymbol{e}_i} G_{2\boldsymbol{e}_i}(\boldsymbol{y}) \tag{3.47}$$

$$= -\gamma \sum_{i=1}^{n} y_i^2 + \ln\kappa \tag{3.48}$$

for some normalizing constant κ. Equivalently,

$$p(\boldsymbol{y}; F^*) = \kappa e^{-\gamma\|\boldsymbol{y}\|^2}. \tag{3.49}$$

By definition, $\gamma \geq 0$, but $\gamma = 0$ results in a constant density on $\mathbb{R}^n$, which is invalid. Thus, the output achieved by $\boldsymbol{X}^*$, $\boldsymbol{Y}^* \triangleq \boldsymbol{X}^* + \boldsymbol{N}$, has independent Gaussian components. That is, $\boldsymbol{Y}^*$ has zero mean and covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \frac{1}{2\gamma}\boldsymbol{I}_n, \tag{3.50}$$

where $\boldsymbol{I}_n$ is the n-dimensional identity matrix. Since $\boldsymbol{X}^*$ and $\boldsymbol{N}$ are independent and $\boldsymbol{N}$ is n-dimensional Gaussian, $\boldsymbol{X}^*$ is also zero-mean n-dimensional Gaussian.

For subcase (S.2), $\boldsymbol{X}^* \sim \mathscr{N}(0, \frac{1}{2\gamma} - \boldsymbol{\Sigma})$ is consistent with $\mathscr{A} = \mathbb{R}^n$. Otherwise, in subcase (S.1), this result is inconsistent with our stipulation that $\mathscr{A} \in \mathscr{C}_n$, so the assumption that $s(\cdot; F^*)$ is identically 0 has led to a contradiction.

Case: $k > 1$**.** Similarly to [17], for this case, we make use of results on the rate of decay of a function compared to that of its Fourier transform.

Lemma 3.1. *Let $\boldsymbol{U} \in \mathbb{R}^n$ have characteristic function satisfying, for all $\boldsymbol{\omega} \in \mathbb{R}^n$,*

$$|\phi_{\boldsymbol{U}}(\boldsymbol{\omega})| \triangleq |\mathbb{E}[e^{i\boldsymbol{\omega}^T\boldsymbol{U}}]| \leq e^{-\frac{\beta\|\boldsymbol{\omega}\|^2}{2}} \tag{3.51}$$

for some $\beta > 0$. Let $\boldsymbol{V}$ be a random variable independent of $\boldsymbol{U}$. Then the characteristic function of $\boldsymbol{W} \triangleq \boldsymbol{U} + \boldsymbol{V}$ satisfies, for all $\boldsymbol{\omega} \in \mathbb{R}^n$,

$$|\phi_{\boldsymbol{W}}(\boldsymbol{\omega})| \leq e^{-\frac{\beta\|\boldsymbol{\omega}\|^2}{2}}.$$

Proof. By the independence of $\boldsymbol{U}$ and $\boldsymbol{V}$ and the fact that characteristic functions have pointwise moduli upper bounded by 1,

$$\begin{aligned}
|\phi_{\boldsymbol{W}}(\omega)| &= |\phi_{\boldsymbol{U}}(\omega)||\phi_{\boldsymbol{V}}(\omega)| \\
&\leq |\phi_{\boldsymbol{V}}(\omega)|e^{-\frac{\beta\|\boldsymbol{\omega}\|^2}{2}} \\
&\leq e^{-\frac{\beta\|\boldsymbol{\omega}\|^2}{2}}.
\end{aligned}$$

□

Lemma 3.2. *Let $\boldsymbol{U} \in \mathbb{R}^n$ have characteristic function satisfying, for all $\boldsymbol{\omega} \in \mathbb{R}^n$,*

$$|\phi_{\boldsymbol{U}}(\boldsymbol{\omega})| \triangleq |\mathbb{E}[e^{i\boldsymbol{\omega}^T\boldsymbol{U}}]| \leq Ke^{-\frac{\beta\|\boldsymbol{\omega}\|^2}{2}} \tag{3.52}$$

for some positive constants β and K. Let $\boldsymbol{V}$ be a random variable independent of $\boldsymbol{U}$ and $\boldsymbol{W} \triangleq$

$\boldsymbol{U}+\boldsymbol{V}$ have density $p_{\boldsymbol{W}}(\cdot)$. If $\alpha > 0$ is such that, for all $\boldsymbol{x} \in \mathbb{R}^n$,

$$p_{\boldsymbol{W}}(\boldsymbol{x}) \leq K e^{-\alpha \|\boldsymbol{x}\|^2}, \tag{3.53}$$

then $\alpha\beta \leq 0.5$.

Proof. Apply Lemma 3.1 and Theorem 4 of [27], noting that an identically 0 function cannot be a density. □

We make use of Lemma 3.2 by setting $\boldsymbol{U} = \boldsymbol{N}$, $\boldsymbol{V} = \boldsymbol{X}$ and $\boldsymbol{W} = \boldsymbol{Y}$ and using (3.45) to obtain a contradiction. Note that, using Rayleigh quotients, the modulus of the characteristic function of $\boldsymbol{N}$ can be upper-bounded for any $\boldsymbol{\omega} \in \mathbb{R}^n$ by

$$|\mathbb{E}[e^{i\boldsymbol{\omega}^T\boldsymbol{N}}]| = e^{-\frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\Sigma}\boldsymbol{\omega}} \leq e^{-\frac{1}{2}\lambda_0^2\|\boldsymbol{\omega}\|^2}, \tag{3.54}$$

where $\lambda_0 > 0$ is the smallest eigenvalue of $\boldsymbol{\Sigma}$. That is, $\boldsymbol{N}$ satisfies (3.51).

To complete the contradiction, we show that for $\alpha > 0$ sufficiently large, $p(\cdot;F^*)$ satisfies the bound in (3.53). Substituting the Multinomial Theorem in (3.45),

$$[\gamma a - C - h(\boldsymbol{N})] - \gamma \sum_{k_1+\ldots+k_n=k} \frac{k!}{k_1!\cdots k_n!} x_1^{2k_1}\cdots x_n^{2k_n} = \sum_{\boldsymbol{m}\in\mathbb{Z}_{\geq 0}^n} c_{\boldsymbol{m}} x_1^{m_1}\cdots x_n^{m_n}. \tag{3.55}$$

Once again, $\gamma = 0$ results in a constant output density over $\mathbb{R}^n$ and can be disregarded as a possibility. By coefficient matching in (3.55), the set of non-zero coefficients, other than $c_{\boldsymbol{0}}$, is indexed by the set

$$B \triangleq \left\{ \boldsymbol{b} \in \mathbb{Z}_{\geq 0}^n \,\middle|\, \sum_{i=1}^n b_i = 2k \text{ and } b_i \text{ is even } \forall i \in \{1,\ldots,n\} \right\}. \tag{3.56}$$

Furthermore, $c_{\boldsymbol{m}} < 0$ for each $\boldsymbol{m} \in B$ and, in particular, $c_{2k\boldsymbol{e}_i} = -\gamma$ for each $i \in \{1,\ldots,n\}$. Therefore, setting $\kappa = \ln(\gamma a - C - h(\boldsymbol{N}))$,

$$p(\boldsymbol{y};F^*) = \kappa e^{\sum_{\boldsymbol{m}\in B} c_{\boldsymbol{m}} G_{\boldsymbol{m}}(\boldsymbol{y})} \tag{3.57}$$

$$= \kappa e^{\sum_{i=1}^n c_{2k\boldsymbol{e}_i} G_{2k\boldsymbol{e}_i}(\boldsymbol{y}) + q(\boldsymbol{y})} \tag{3.58}$$

$$= \kappa e^{-\gamma \sum_{i=1}^n y_i^{2k} + q(\boldsymbol{y})}, \tag{3.59}$$

where $q(\boldsymbol{y})$ has total degree $2k$ and has degree less than $2k$ in each y_i.

Now for $\alpha > 0$, we consider the rate of decay of $p(\boldsymbol{y};F^*)$ relative to that of $e^{-\alpha\|\boldsymbol{y}\|^2}$ through the ratio

$$\frac{p(\boldsymbol{y};F^*)}{e^{-\alpha\|\boldsymbol{y}\|^2}} = \kappa e^{-\gamma\sum_{i=1}^n y_i^{2k}+q(\boldsymbol{y})+\alpha(\sum_{i=1}^n y_i^2)} \tag{3.60}$$

$$= \kappa e^{-\gamma\sum_{i=1}^n y_i^{2k}+q_0(\boldsymbol{y})}, \tag{3.61}$$

where

$$q_0(\boldsymbol{y}) \triangleq q(\boldsymbol{y}) + \alpha(\sum_{i=1}^n y_i^2). \tag{3.62}$$

Since $k > 1$, $q_0(\boldsymbol{y})$ also has total degree $2k$ and degree less than $2k$ in each y_i. Therefore, $D(\alpha) > 0$ can be chosen such that for all $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_{D(\alpha)}}(0)$,

$$p(\boldsymbol{y};F^*) \leq e^{-\alpha\|\boldsymbol{y}\|^2}, \tag{3.63}$$

where the bar over a set denotes closure.

Let $\alpha = 1/(\lambda_0^2)$ and choose $D > 0$ large enough to satisfy (3.63) for all $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$. Note that the continuous function $p(\cdot;F^*)$ attains a maximum M on the compact set $\overline{B_D}(0)$. Let $K = \max\{1, Me^{\alpha D^2/2}\}$. Then we have that for all $\boldsymbol{y} \in \mathbb{R}^n$,

$$p(\boldsymbol{y};F^*) \leq Ke^{-\alpha\|\boldsymbol{y}\|^2}. \tag{3.64}$$

That is, $p(\boldsymbol{y};F^*)$ satisfies (3.53) with this choice of α. However, since $\alpha\beta = 1 > 0.5$, (3.54) and (3.64) contradict Lemma 3.2. Therefore, (3.55) cannot hold for all $\boldsymbol{x} \in \mathbb{R}^n$ and we conclude that for $k > 1$, $s(\cdot;F^*)$ cannot be identically 0 on $\mathbb{R}^n$.

Summary: We summarize the results of the 2 above cases in a theorem.

Theorem 3.1. *Suppose that either*

1. $\mathscr{A} \in \mathscr{C}_n$, *or*
2. $\mathscr{A} = \mathbb{R}^n$, *with* $2k > n$ *and* $(n,k) \neq (1,1)$.

Then

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid s(\boldsymbol{x};F^*) = 0\} \subsetneq \mathbb{R}^n. \tag{3.65}$$

An immediate consequence of Theorem 3.1 is that $supp(F^*)$ is a strict subset of $\mathbb{R}^n$. Theorem 3.1 also shows that $supp(F^*)$ is "sparse" in the sense used by [9]; $s(\cdot;F^*)$ is a non-zero function with an analytic extension to $\mathbb{C}^n$ that is zero on $supp(F^*)$. However, the main importance of Theorem 3.1 is as an intermediary result that is used in Sections 3.4 and 3.5 to obtain a better understanding of the structure of $supp(F^*)$.

3.4 Results Due to the Identity Theorem

In the spirit of prior work (see e.g. [17, 23, 24, 28]), we obtain results in this section through contradiction of the Identity Theorem. However, unlike this prior work, the channel model considered in this thesis uses the multivariate real-analytic Identity Theorem. That is, rather than extending $s(\cdot;F^*)$ to $\mathbb{C}^n$, we work directly in $\mathbb{R}^n$, as is done in [15].

The Identity Theorem for functions of a single complex variable states that an entire function whose zero set has an accumulation point is identically 0 on the complex plane. This allows the authors of [17, 23, 24, 28] to conclude that if the optimal input's support has an accumulation point on the real line, then $s(\cdot;F^*)$ must be identically 0 on the complex plane.

In contrast, the Identity Theorem for functions of $n > 1$ complex variable states that if an entire function is identically 0 on an open subset of $\mathbb{C}^n$, then it is identically 0 on all of $\mathbb{C}^n$. Since no subset of $\mathbb{R}^n$ is an open set in $\mathbb{C}^n$, no assumption on $supp(F^*)$ leads to a direct application of the Identity Theorem for $n > 1$. For this reason, the authors of [13, 23, 29] limit their consideration to spherically symmetric channels, in which the capacity-achieving distribution is only a function of radius. This simplification converts the problem from n dimensions to 1 dimension. However, results can be obtained from the real-analytic Identity Theorem, or even the multivariate complex Identity Theorem after proper consideration. We will only discuss the real-analytic approach here.

We now present the real-analytic Identity Theorem [15].

Theorem 3.2 (Real-analytic Identity Theorem)**.** *Let $U \subseteq \mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be a real-analytic function that is zero on $A \subseteq U$. If A has positive Lebesgue measure then $f(\cdot)$ is identically 0 on U.*

Remark. *Since any open set has positive Lebesgue measure, if $f(\cdot)$ is* 0 *on some open set A, then by Theorem 3.2, it is identically* 0 *on U.*

In addition to showing that $supp(F^*)$ has Lebesgue meausure 0, we show a stronger notion than $supp(F^*)$ containing no open subsets of $\mathbb{R}^n$. We require the following definitions before stating our results.

Definition 3.2. A set $A \subseteq B$ is called dense in B if for every $b \in B$, there exists a sequence $\{a_i\}_{i=0}^{\infty} \subseteq A$ that converges to b.

Definition 3.3. A set $A \subseteq B$ is called nowhere dense in B if for every open set $U \subseteq B$, $A \cap U$ is not dense in U.

Theorems 3.1 and 3.2 are used to obtain the main result of this section.

Theorem 3.3. *Suppose that either*

1. $\mathscr{A} \in \mathscr{C}_n$, *or*
2. $\mathscr{A} = \mathbb{R}^n$, *with* $2k > n$ *and* $(n,k) \neq (1,1)$.

Then $supp(F^*)$ *is nowhere dense in* $\mathbb{R}^n$ *and has Lebesgue measure 0.*

Proof. First note that, by (3.28),

$$supp(F^*) \subseteq \{\boldsymbol{x} \in \mathbb{R}^n \mid s(\boldsymbol{x};F^*) = 0\} \cap \mathscr{A}. \tag{3.66}$$

Therefore, it suffices to show that

$$S \triangleq \{\boldsymbol{x} \in \mathbb{R}^n \mid s(\boldsymbol{x};F^*) = 0\} \tag{3.67}$$

is nowhere dense in $\mathbb{R}^n$ and has Lebesgue measure 0.

Recall that $s(\cdot;F^*)$ is real-analytic and, by Theorem 3.1, $s(\cdot;F^*)$ is not identically 0 on $\mathbb{R}^n$. Then by Theorem 3.2 and the subsequent remark,

P.1 S has Lebesgue measure 0, and

P.2 $s(\cdot;F^*)$ is not identically 0 on any open subset of $\mathbb{R}^n$.

To complete the proof, we will use (**P.2**) to show that S is nowhere dense. Note that a subset of $\mathbb{R}^n$ is nowhere dense if and only if the interior of its closure is the empty set. Furthermore, since $s(\cdot;F^*)$ is continuous (it is real-analytic), S is closed. Therefore, S being nowhere dense is equivalent to its interior being empty, which holds by (**P.2**).

□

3.5 Geometry of a Real-analytic Function's Zero Set

In this section, we use geometry to further investigate the properties of $supp(F^*)$ and an alternate proof for Theorem 3.3, which does not use the Identity Theorem, is given. These discussions consider subsets of a vector's components, so for $\boldsymbol{x} \in \mathbb{R}^n$ and $i \in \{1,\ldots,n\}$, we introduce the notation

$$\boldsymbol{x}^i \triangleq (x_1,\ldots,x_i) \in \mathbb{R}^i. \tag{3.68}$$

Recall that by (3.28),

$$supp(F^*) \subseteq \{\boldsymbol{x} \in \mathbb{R}^n \mid s(\boldsymbol{x};F^*) = 0\} \cap \mathscr{A} \tag{3.69}$$

$$= \{\boldsymbol{x} \in \mathscr{A} \mid s(\boldsymbol{x};F^*) = 0\}. \tag{3.70}$$

Since $s(\cdot;F^*)$ is real-analytic, it is in our interest to study the geometry of zero sets of real-analytic functions. We turn to established results in analysis and state Theorem 6.3.3 of [20] to the level that it is needed in this thesis.

Theorem 3.4 (Structure Theorem)**.** *Let $\psi(\cdot) : \mathbb{R}^n \to \mathbb{R}$ be a real-analytic function, where $\psi(0,\ldots,0,x_n)$ is not identically 0 in x_n. After a rotation of the coordinates $x_1,\ldots,x_{n-1}$, there exist constants δ_m, $m \in \{1,\ldots,n\}$, such that with*

$$Q \triangleq \{\boldsymbol{x} \in \mathbb{R}^n \mid |x_m| < \delta_m \quad \forall m \in \{1,\ldots,n\}\}, \tag{3.71}$$

we have

$$\{\boldsymbol{x} \in Q \mid \psi(\boldsymbol{x}) = 0\} = \bigcup_{i=0}^{n-1} V_i, \tag{3.72}$$

where V_0 is either empty or contains only the origin and V_i, $i \in \{1,\ldots,n-1\}$, is a finite disjoint union of i-dimensional submanifolds. That is, for each $i \in \{1,\ldots,n-1\}$, there exists n_i for which

$$V_i = \bigcup_{j=0}^{n_i} \Gamma_i^j, \tag{3.73}$$

where each Γ_i^j is an i-dimensional submanifold. Furthermore, letting

$$Q_i \triangleq \{\boldsymbol{x}^i \in \mathbb{R}^i \mid |x_m| < \delta_m \quad \forall m \in \{1,\ldots,i\}\}, \tag{3.74}$$

there exist an open set $\Omega_i^j \subseteq Q_i$ and real-analytic functions $\alpha_i^{j,m}(\cdot)$, $m \in \{i+1,\ldots,n\}$, on Ω_i^j for *which*

$$\Gamma_i^j = \{(\boldsymbol{x}^i, \alpha_i^{j,i+1}(\boldsymbol{x}^i), \ldots, \alpha_i^{j,n}(\boldsymbol{x}^i)) \in \mathbb{R}^n \mid \boldsymbol{x}^i \in \Omega_i^j\}. \tag{3.75}$$

We apply Theorem 3.4 to characterize the zero set of $s(\cdot;F^*)$ in the form of (3.72) and obtain the following result.

Theorem 3.5. *Suppose that either*

1. $\mathscr{A} \in \mathscr{C}_n$*, or*
2. $\mathscr{A} = \mathbb{R}^n$*, with* $2k > n$ *and* $(n,k) \neq (1,1)$*.*

Then

$$supp(F^*) \subseteq \{\boldsymbol{x} \in \mathscr{A} \mid s(\boldsymbol{x};F^*) = 0\} = \mathscr{A} \cap \left(\bigcup_{i=0}^{n-1} T_i\right), \tag{3.76}$$

where T_0 *is a countable union of isolated points and* T_i*,* $i \in \{1,\ldots,n-1\}$*, is a countable disjoint union of i-dimensional submanifolds. Furthermore, if* $\mathscr{A} \in \mathscr{C}_n$*, then these unions are finite.*

Proof. First note that, by Theorem 3.1, $s(\cdot;F^*)$ is not identically 0 on $\mathbb{R}^n$. Therefore, for any $\boldsymbol{q} \in \mathbb{Q}^n$ we can translate $s(\cdot;F^*)$ by $\boldsymbol{q}$ and rotate its coordinate system to apply Theorem 3.4. That is, there exists a sufficiently small neighbourhood $Q^{\boldsymbol{q}}$ around $\boldsymbol{q}$ such that

$$\{\boldsymbol{x} \in Q^{\boldsymbol{q}} \mid s(\boldsymbol{x};F^*) = 0\} = \bigcup_{i=0}^{n-1} V_i^{\boldsymbol{q}}, \tag{3.77}$$

where the V_i's are as in Theorem 3.4.

Since $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$,

$$\mathscr{A} \subseteq \bigcup_{\boldsymbol{q} \in \mathbb{Q}^n} Q^{\boldsymbol{q}} \tag{3.78}$$

Furthermore, if $\mathscr{A} = \mathscr{C} \in \mathscr{C}_n$, then this open cover has a finite subcover $\{Q^{\boldsymbol{q}_j}\}_{j=1}^m$. That is,

$$\mathscr{C} \subseteq \bigcup_{j=1}^{m} Q^{\boldsymbol{q}_j}. \tag{3.79}$$

Defining the index set

$$\mathscr{M} \triangleq \begin{cases} \mathbb{Q}^n, & \mathscr{A} = \mathbb{R}^n \\ \{\boldsymbol{q}_j\}_{j=1}^m, & \mathscr{A} \in \mathscr{C}_n, \end{cases} \tag{3.80}$$

we obtain

$$\{\boldsymbol{x} \in \mathscr{A} \mid s(\boldsymbol{x};F^*) = 0\} = \mathscr{A} \cap \left(\bigcup_{\boldsymbol{q} \in \mathscr{M}} \{\boldsymbol{x} \in Q^{\boldsymbol{q}} \mid s(\boldsymbol{x};F^*) = 0\} \right) \tag{3.81}$$

$$= \mathscr{A} \cap \left(\bigcup_{\boldsymbol{q} \in \mathscr{M}} \bigcup_{i=0}^{n-1} V_i^{\boldsymbol{q}} \right) \tag{3.82}$$

$$= \mathscr{A} \cap \left(\bigcup_{i=0}^{n-1} \bigcup_{\boldsymbol{q} \in \mathscr{M}} V_i^{\boldsymbol{q}} \right). \tag{3.83}$$

Since for each $\boldsymbol{q} \in \mathscr{M}$, $V_0^{\boldsymbol{q}}$ is either empty or a single point,

$$T_0 \triangleq \bigcup_{\boldsymbol{q} \in \mathscr{M}} V_0^{\boldsymbol{q}} \tag{3.84}$$

is a countable set of points and is finite when $\mathscr{A} \in \mathscr{C}_n$. Furthermore, each $V_i^{\boldsymbol{q}}$, where $i \in \{1, \ldots, n-1\}$, is itself a finite union of i-dimensional submanifolds. The countable union of a finite union is countable, so

$$T_i \triangleq \bigcup_{\boldsymbol{q} \in \mathscr{M}} V_i^{\boldsymbol{q}} \tag{3.85}$$

is a countable union of i-dimensional submanifolds. When $\mathscr{A} \in \mathscr{C}_n$, this union is also finite. □

Note that Theorem 3.5 agrees with the results of [23] when the cases overlap. That is, when $\mathscr{A} \in \mathscr{C}_n$ and $k = 1$, [23] shows that the capacity-achieving distribution is supported on a finite number of concentric $(n-1)$-spheres. Each $(n-1)$-sphere is an $n-1$ dimensional submanifold.

In the next 2 theorems, we demonstrate that the results of Theorem 3.3 can be recovered through Theorem 3.4 in a manner which does not use the Identity Theorem.

Theorem 3.6. *Suppose that either*

1. $\mathscr{A} \in \mathscr{C}_n$, *or*
2. $\mathscr{A} = \mathbb{R}^n$, *with* $2k > n$ *and* $(n,k) \neq (1,1)$.

Let $\mu(\cdot)$ *denote the n-dimensional Lebesgue measure. Then,*

$$\mu(supp(F^*)) = 0. \tag{3.86}$$

Proof. By Theorem 3.5,

$$supp(F^*) \subseteq \mathscr{A} \cap \left(\bigcup_{i=0}^{n-1} T_i \right) \tag{3.87}$$

$$\subseteq \bigcup_{i=0}^{n-1} \bigcup_{\boldsymbol{q} \in \mathscr{M}} V_i^{\boldsymbol{q}}, \tag{3.88}$$

where $\mathscr{M}$ is countable. Note that for each $\boldsymbol{q} \in \mathscr{M}$, $V_0^{\boldsymbol{q}}$ is either empty or a single point, so $\mu(V_0^{\boldsymbol{q}}) = 0$. Furthermore, for each $\boldsymbol{q} \in \mathscr{M}$ and $i \in \{1, \ldots, n-1\}$, $V_i^{\boldsymbol{q}}$ is a finite disjoint union of $n_i^{\boldsymbol{q}}$ i-dimensional submanifolds. Therefore, it suffices to show that any of these i dimensional submanifolds $\Gamma_i^{\boldsymbol{q},j}$, where $j \in \{1, \ldots, n_i^{\boldsymbol{q}}\}$, satisfies $\mu(\Gamma_i^{\boldsymbol{q},j}) = 0$. By (3.75), for each triple $(i, \boldsymbol{q}, j)$, there exist an open set $\Omega_i^{\boldsymbol{q},j} \subseteq \mathbb{R}^i$ and functions $\alpha_i^{\boldsymbol{q},j,m}(\cdot)$, $m \in \{i+1, \ldots, n\}$, on $\Omega_i^{\boldsymbol{q},j}$ for which

$$\Gamma_i^{\boldsymbol{q},j} = \{(\boldsymbol{x}^i, \alpha_i^{\boldsymbol{q},j,i+1}(\boldsymbol{x}^i), \ldots, \alpha_i^{\boldsymbol{q},j,n}(\boldsymbol{x}^i)) \in \mathbb{R}^n \mid \boldsymbol{x}^i \in \Omega_i^{\boldsymbol{q},j}\}. \tag{3.89}$$

Then with $\mathbf{1}_{\Gamma_i^{\boldsymbol{q},j}}(\cdot)$ denoting the indicator function of $\Gamma_i^{\boldsymbol{q},j}$,

$$\mu(\Gamma_i^{\boldsymbol{q},j}) = \int_{\mathbb{R}^n} \mathbf{1}_{\Gamma_i^{\boldsymbol{q},j}}(\boldsymbol{x})\, d\boldsymbol{x} \tag{3.90}$$

$$= \int_{\Omega_i^{\boldsymbol{q},j}} \int_{\alpha_i^{\boldsymbol{q},j,i+1}(\boldsymbol{x}^i)}^{\alpha_i^{\boldsymbol{q},j,i+1}(\boldsymbol{x}^i)} \cdots \int_{\alpha_i^{\boldsymbol{q},j,n}(\boldsymbol{x}^i)}^{\alpha_i^{\boldsymbol{q},j,n}(\boldsymbol{x}^i)} dx_n \ldots dx_{i+1}\, d\boldsymbol{x}^i \tag{3.91}$$

$$= \int_{\Omega_i^{\boldsymbol{q},j}} 0\, d\boldsymbol{x}^i \tag{3.92}$$

$$= 0. \tag{3.93}$$

□

Theorem 3.7. *Suppose that either*

1. $\mathscr{A} \in \mathscr{C}_n$, *or*
2. $\mathscr{A} = \mathbb{R}^n$, *with* $2k > n$ *and* $(n,k) \neq (1,1)$.

Then $supp(F^*)$ *is nowhere dense in* $\mathbb{R}^n$.

Proof. Note that by Theorem 3.1, the set

$$A \triangleq \{\boldsymbol{x} \in \mathbb{R}^n \mid s(\boldsymbol{x}; F^*) = 0\} \tag{3.94}$$

is a strict subset of $\mathbb{R}^n$. Let $U \subseteq \mathbb{R}^n$ be a non-empty open set – we will show the result by proving that $A \cap U$ is not dense in U.

Fix $\boldsymbol{x} \in U$. Translating $s(\cdot;F^*)$ by $\boldsymbol{x}$, rotating the coordinate system and applying Theorem 3.4 shows that there exists a sufficiently small open set Q containing $\boldsymbol{x}$ on which

$$A \cap Q = \bigcup_{i=0}^{n-1} V_i \tag{3.95}$$

$$= V_0 \cup \left(\bigcup_{i=1}^{n-1} \bigcup_{j=0}^{n_i} \Gamma_i^j \right). \tag{3.96}$$

First note that since U and Q are open, their intersection $U \cap Q$ is open as well. Then for any $\boldsymbol{y} \in U \cap Q$, any sequence converging to $\boldsymbol{y}$ has a subsequence in $U \cap Q$. It suffices to show that there exists a point in $U \cap Q$ that is not the limit of any sequence in $A \cap U \cap Q$.

We proceed by using the parameterization of Γ_i^j given by (3.75). We will show the existence of a point of the form $(\boldsymbol{x}^{n-1}, u_n) \in U \cap Q$ that is not the limit of any sequence whose first $n-1$ components converge to $\boldsymbol{x}^{n-1}$. Noting that any sequence whose first $n-1$ components do not converge to $\boldsymbol{x}^{n-1}$ cannot converge to $(\boldsymbol{x}^{n-1}, u_n)$ will complete the proof.

Let $\{(\boldsymbol{y}^{n-1}, y_n)_m\}_{m=0}^{\infty} \subseteq A \cap U \cap Q$ be a sequence converging to $\boldsymbol{y} \triangleq (\boldsymbol{x}^{n-1}, y_n)$. Using the parameterization from (3.75), the n'th component of sequence index m satisfies one of the following:

1. $(y_n)_m \in \{v_n \mid \boldsymbol{v} \in V_0\}$, or
2. for some $i \in \{1, \ldots, n-1\}$ and $j \in \{1, \ldots, n_i\}$,

$$(y_n)_m = \alpha_i^{j,n}(\boldsymbol{y}_m^i). \tag{3.97}$$

Since $\alpha_i^{j,n}(\cdot)$ is real-analytic, it is continuous. Then for $(y_n)_m$ satisfying (3.97), we have

$$\lim_{m\to\infty} (y_n)_m = \lim_{m\to\infty} \alpha_i^{j,n}(\boldsymbol{y}_m^i) \tag{3.98}$$

$$= \alpha_i^{j,n}(\lim_{m\to\infty} \boldsymbol{y}_m^i) \tag{3.99}$$

$$= \alpha_i^{j,n}(\boldsymbol{x}^i). \tag{3.100}$$

Since V_0 is either empty or a single point, the number of possible values for $\lim_{m\to\infty}(y_n)_m$ is at most

$$|V_0| + \sum_{i=1}^{n-1} n_i < \infty. \tag{3.101}$$

However, since $U \cap Q$ is open and $\boldsymbol{x} \in U \cap Q$, the set $\{t \in \mathbb{R} \mid (\boldsymbol{x}^{n-1}, x_n + t) \in U \cap Q\}$ is uncountable. Thus, there exists t such that $(\boldsymbol{x}^{n-1}, x_n + t) \in U \cap Q$ is not the limit of any sequence in $A \cap U \cap Q$.

□

Chapter 4

Conclusion

4.1 Summary

This thesis has considered vector-valued channels with additive Gaussian noise. The support of the capacity-achieving input distribution was discussed when inputs were subjected to an average even moment radial constraint and, in some cases, an additional restriction to a compact set. When the input alphabet was the entire space $\mathbb{R}^n$, the moment constraint was assumed to satisfy $2k > n$ and the classic case $n = k = 1$ was omitted from most discussion.

Unlike much of the prior work in this area, the noise was not limited to having independent or identically distributed components. Therefore, it could not be concluded that the optimal input distribution was a function of only radius. This precluded the use of the 1-dimensional Identity Theorem. Instead, the n-dimensional real-analytic Identity Theorem was required when following that approach.

The optimal output density was expressed using the generalization of the 1-dimensional approach of [17]. Hermite polynomials in n-dimensions were used to simplify expressions involving the density of a multivariate Gaussian noise density with an arbitrary non-singular covariance matrix. It was determined that the support of the capacity-achieving distribution has Lebesgue measure 0 and is nowhere dense in $\mathbb{R}^n$.

Finally, the geometry of $supp(F^*)$ was studied by using the fact that it is a subset of the zero set of a real analytic function. It was determined that the support is contained in a countable union of single points and submanifolds of dimensions $1, \ldots, n-1$. When the alphabet is compact, this union is finite. Furthermore, geometric arguments were used to provide an alternate proof that the capacity-achieving distribution has Lebesgue measure 0 and is nowhere dense in $\mathbb{R}^n$.

This thesis is an expansion of work concerning even moment input constraints in [17] to vector-valued channels that are not necessarily spherical symmetry. Viewed as a generalization of [9], it considers $2k$'th rather than 2nd moment constraints and discusses an alphabet of $\mathbb{R}^n$. It also provides further detail regarding the supports of capacity-achieving input distributions. Geometric discussions of capacity-achieving input support in prior work are limited to concentric shells in spherically symmetric channels. This thesis uses submanifolds to geometrically describe the supports of capacity-achieving inputs to more general channels.

4.2 Further Research

Given the geometric results of Theorem 3.5, the number of submanifolds forming a superset of $supp(F^*)$ can be explored for compact alphabets. A similar problem for channels under amplitude constraints is studied in [12, 14]. Results of this type could lead to computational methods for finding $supp(F^*)$ and the corresponding capacity for a given choice of $(\mathscr{A}, k, a)$.

The Hermite polynomial approach is well-suited for even moment constraints on inputs to Gaussian noise channels, but it does not appear to generalize easily to other noise models. One potential approach is to define a Hilbert space with weight function given by a different noise density. However, the coefficients in the expansion of $\ln p(\cdot; F^*)$ in (3.38) can only be easily found when this Hilbert space gives rise to a Riesz basis. Furthermore, the noise distribution must decay sufficiently quickly to ensure that the supporting technical results hold.

In [18], scalar channels under a wide variety of input-output functions, cost functions and noise densities are studied. Results are obtained by examining these functions' relative growth rates. The results of this thesis could be generalized by adopting that approach in conjunction with the observations that are made here regarding analytic functions of several variables.

References

[1] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1290–1301, 2001.

[2] F. Alajaji and P. Chen, *An Introduction to Single-User Information Theory*, ser. Springer Undergraduate Texts in Mathematics and Technology. Springer Singapore, 2018. [Online]. Available: https://books.google.ca/books?id=5cRXDwAAQBAJ

[3] S. Alireza Banani and R. G. Vaughan, "Compensating for non-linear amplifiers in MIMO communications systems," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 700–714, Feb 2012.

[4] H. Bateman, *Higher Transcendental Functions [Volumes I-III]. Vol.I-III*. New York: McGraw-Hill Book Company, 1953.

[5] K. Bhaskara Rao and M. Bhaskara Rao, *Theory of Charges*, ser. PURE AND APPLIED MATHEMATICS. New York: Academic Press, 1983.

[6] P. Billingsley, *Convergence of Probability Measures*, ser. Wiley Series in Probability and Statistics. Wiley, 2013. [Online]. Available: https://books.google.ca/books?id=6ItqtwaWZZQC

[7] D. Burdick and F. D. Lesley, "Some uniqueness theorems for analytic functions," *The American Mathematical Monthly*, vol. 82, no. 2, pp. 152–155, 1975. [Online]. Available: http://www.jstor.org/stable/2319660

[8] C.-P. Liang, J.-H. Jong, W. E. Stark, and J. R. East, "Nonlinear amplifier effects in communications systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1461–1466, Aug 1999.

[9] T. H. Chan, S. Hranilovic, and F. R. Kschischang, “Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs,” *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2073–2088, 2005.

[10] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2012. [Online]. Available: https://books.google.ca/books?id=VWq5GG6ycxMC

[11] P. Dupuis and R. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, ser. Wiley Series in Probability and Statistics. Wiley, 1997. [Online]. Available: https://books.google.ca/books?id=e2iJB_O3rS0C

[12] A. Dytso, M. Al, H. V. Poor, and S. Shamai Shitz, “On the capacity of the peak power constrained vector Gaussian channel: An estimation theoretic perspective,” *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3907–3921, 2019.

[13] A. Dytso, M. Goldenbaum, S. Shamai, and H. V. Poor, “Upper and lower bounds on the capacity of amplitude-constrained MIMO channels,” in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[14] A. Dytso, S. Yagli, H. V. Poor, and S. Shamai Shitz, “The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points,” *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2006–2022, 2020.

[15] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), “Amplitude constrained MIMO channels: Properties of optimal input distributions and bounds on the capacity,” *Entropy*, vol. 21, no. 2, 2019. [Online]. Available: https://www.mdpi.com/1099-4300/21/2/200

[16] J. Fahs, A. Tchamkerten, and M. I. Yousefi, “On the optimal input of the nondispersive optical fiber,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 131–135.

[17] J. J. Fahs and I. C. Abou-Faycal, “Using Hermite bases in studying capacity-achieving distributions over AWGN channels,” *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5302–5322, 2012.

[18] J. Fahs and I. Abou-Faycal, “On properties of the support of capacity-achieving distributions for additive noise channel models with input cost constraints,” *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1178–1198, 2018.

[19] R. Gunning and H. Rossi, *Analytic Functions of Several Complex Variables*, ser. Ams Chelsea Publishing. AMS Chelsea Pub., 2009. [Online]. Available: https://books.google.ca/books?id=wsqFAwAAQBAJ

[20] S. Krantz and H. Parks, *A Primer of Real Analytic Functions*, ser. Advanced Texts Series. Birkhäuser Boston, 2002. [Online]. Available: https://books.google.ca/books?id=i4vw2STJl2QC

[21] A. Muleshkov and T. Nguyen, "Easy proof of the Jacobian for the n-dimensional polar coordinates," *Pi Mu Epsilon Journal*, vol. 14, pp. 269–273, 2016.

[22] R. Raich and G. T. Zhou, "On the modeling of memory nonlinear effects of power amplifiers for communication applications," in *Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop.*, Oct 2002, pp. 7–10.

[23] B. Rassouli and B. Clerckx, "On the capacity of vector Gaussian channels with bounded inputs," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 6884–6903, 2016.

[24] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Transactions on Information Theory*, vol. 41, no. 4, pp. 1060–1071, 1995.

[25] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[26] A. Shiryaev and D. Chibisov, *Probability-1*, ser. Graduate Texts in Mathematics. Springer New York, 2016. [Online]. Available: https://books.google.ca/books?id=WTGmDAAAQBAJ

[27] A. Sitaram, M. Sundari, and S. Thangavelu, "Uncertainty principles on certain Lie groups," *Proceedings of the Indian Academy of Sciences - Mathematical Sciences*, vol. 105, no. 2, pp. 135–151, May 1995. [Online]. Available: https://doi.org/10.1007/BF02880360

[28] J. G. Smith, "The information capacity of amplitude- and variance-constrained sclar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019995871903469

[29] J. Sommerfeld, I. Bjelakovic, and H. Boche, "On the boundedness of the support of optimal input measures for Rayleigh fading channels," in *2008 IEEE International Symposium on Information Theory*, 2008, pp. 1208–1212.

APPENDICES

Appendix A

Theorems From Convex Optimization

The following theorems are standard concepts from convex optimization theory and are used in [17]. They are stated here for reference.

Theorem A.1. *If $I_0(\cdot)$ is a weak continuous function on a compact set $\mathscr{P} \subseteq \mathscr{F}$, then $I_0(\cdot)$ achieves its maximum on $\mathscr{P}$. If, in addition, $\mathscr{P}$ is convex and $I_0(\cdot)$ strictly concave, then the maximum is achieved by a unique element $F^* \in \mathscr{P}$.*

Theorem A.2 (Lagrange)**.** *Let $\mathscr{F}$ be a vector space and $\Omega \subseteq \mathscr{F}$ be convex. Let $I_0(\cdot)$ be a concave real functional on Ω and let $g_0(\cdot)$ be a convex real functional on Ω. Suppose a Slater condition is satisfied; that is, suppose there exists $F_0 \in \Omega$ with $g_0(F_0) < 0$. Let*

$$C_0 = \sup_{\substack{F \in \Omega \\ g_0(F) \leq 0}} I_0(F) \tag{A.1}$$

be finite. Then there exists $\lambda \geq 0$ for which

$$C_0 = \sup_{F \in \Omega} \{I_0(F) - \lambda g_0(F)\}. \tag{A.2}$$

Furthermore, if C_0 is achieved by F^* in (A.1), then it is also achieved by F^* in (A.2) and

$$\lambda g_0(F^*) = 0. \tag{A.3}$$

Remark 1. *The Slater condition being satisfied is absolutely essential for Theorem A.2 to hold.*

Remark 2. *In Theorem A.2, λ is known as the Lagrange multiplier and the condition given by (A.3) is known as complementary slackness.*

Theorem A.3 (Optimality Condition)**.** *Let $\mathscr{F}$ be a normed vector space. Suppose that a weakly differentiable concave function $J(\cdot)$ achieves its maximum on a compact and convex set $\Omega \subseteq \mathscr{F}$. Then $J(\cdot)$ achieves its maximum at F^* if and only if $J'(F^*, F) \leq 0$ for all $F \in \Omega$ [28].*

We present here the adaptation to our purposes of a technique that has been in several papers to convert the optimality condition of Theorem A.3 to a condition on individual points (eg. [1, 28]). It is by this theorem that the necessary and sufficient conditions for optimality can be expressed as in (3.25).

Recall the notation

$$h_{\boldsymbol{Y}}(F) \triangleq -\int_{\mathbb{R}^n} p(\boldsymbol{y};F) \ln p(\boldsymbol{y};F)\, d\boldsymbol{y} \tag{A.4}$$

to describe the entropy of the output induced by an input with distribution F. Similarly to [28], denote the marginal entropy density

$$h(\boldsymbol{x};F) \triangleq -\int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \ln p(\boldsymbol{y};F)\, d\boldsymbol{y}. \tag{A.5}$$

Theorem A.4. *Define*

$$\mathscr{Q}_n(\mathscr{A},k) \triangleq \bigcup_{b \geq a} \mathscr{P}_n(\mathscr{A},k,b) \tag{A.6}$$

$$= \{F \in \mathscr{F}_n(\mathscr{A}) \mid g(F) < \infty\}. \tag{A.7}$$

Suppose that F^* satisfies (3.14), where $\gamma \geq 0$ is the Lagrange multiplier corresponding to the problem in (3.10). Then the following are equivalent:

P.1 *For every $F \in \mathscr{Q}_n(\mathscr{A},k)$,*

$$\int_{\mathscr{A}} h(\boldsymbol{x};F^*)\, dF(\boldsymbol{x}) \leq \gamma\Big(\int_{\mathscr{A}} \|\boldsymbol{x}\|^{2k}\, dF(\boldsymbol{x}) - a\Big) + C + h(\boldsymbol{N}). \tag{A.8}$$

P.2 *For all $x \in \mathscr{A}$,*

$$h(\boldsymbol{x};F^*) \leq \gamma(\|\boldsymbol{x}\|^{2k} - a) + C + h(\boldsymbol{N}), \tag{A.9}$$

and if $\boldsymbol{x} \in supp(F^*)$, then

$$h(\boldsymbol{x};F^*) = \gamma(\|\boldsymbol{x}\|^{2k} - a) + C + h(\boldsymbol{N}). \tag{A.10}$$

Proof. This proof proceeds in a similar manner to Theorem 4 of [1]. For any $F \in \mathscr{Q}_n(\mathscr{A},k)$, integrating both sides of (A.9) with respect to $dF(\cdot)$ yields that (**P.2**) implies (**P.1**).

It remains to show that (**P.1**) implies (**P.2**). Suppose this implication is false – that is, (**P.1**) holds but there either exists $\boldsymbol{v} \in \mathscr{A}$ for which

$$h(\boldsymbol{v};F^*) > \gamma(\|\boldsymbol{v}\|^{2k} - a) + C + h(\boldsymbol{N}), \tag{A.11}$$

or there exists $\boldsymbol{w} \in supp(F^*)$ such that

$$h(\boldsymbol{w};F^*) \neq \gamma(\|\boldsymbol{w}\|^{2k} - a) + C + h(\boldsymbol{N}). \tag{A.12}$$

If (A.11) holds, then let $b = \max\{\|\boldsymbol{v}\|^{2k}, a\}$ and let $F(\boldsymbol{x}) = \prod_{i=1}^{n} u_1(x_i - v_i)$, where $u_1(\cdot)$ is the Heaviside step function. Then $F \in \mathscr{P}_n(\mathscr{A},k,b) \subseteq \mathscr{Q}_n(\mathscr{A},k)$ and

$$\int_{\mathscr{A}} h(\boldsymbol{x};F^*)\,dF(\boldsymbol{x}) - \gamma(\int_{\mathscr{A}} \|\boldsymbol{x}\|^{2k}\,dF(\boldsymbol{x}) - a) = h(\boldsymbol{v};F^*) - \gamma(\|\boldsymbol{v}\|^{2k} - a) \tag{A.13}$$

$$> C + h(\boldsymbol{N}), \tag{A.14}$$

contradicting (A.8). Therefore, (A.11) cannot be satisfied for any $\boldsymbol{x} \in \mathscr{A}$ and we are left with the alternative that there exists $\boldsymbol{w} \in supp(F^*) \subseteq \mathscr{A}$ for which (A.12) holds. That is,

$$h(\boldsymbol{w};F^*) < \gamma(\|\boldsymbol{w}\|^{2k} - a) + C + h(\boldsymbol{N}). \tag{A.15}$$

Note that since $\gamma g(F^*) = 0$ by the complementary slackness condition of (A.3),

$$\gamma a = \gamma \int_{\mathscr{A}} \|\boldsymbol{x}\|^{2k}\,dF^*(\boldsymbol{x}). \tag{A.16}$$

Since $h(\cdot;F)$ and $\|\cdot\|^{2k}$ are continuous, (A.15) is satisfied on $B_\delta(\boldsymbol{w})$ for some $\delta > 0$. Observe that $\boldsymbol{w} \in supp(F^*)$, so there exists ε such that $P\{\boldsymbol{X}^* \in B_\delta(\boldsymbol{w})\} = \varepsilon > 0$. Then substituting (A.16),

$$C + h(\boldsymbol{N}) - \gamma a = h_{\boldsymbol{Y}}(F^*) - \gamma a \tag{A.17}$$

$$= \int_{\mathscr{A}} \left[h(\boldsymbol{x};F^*) - \gamma\|\boldsymbol{x}\|^{2k}\right] dF^*(\boldsymbol{x}) \tag{A.18}$$

$$= \int_{B_\delta(\boldsymbol{w})} \left[h(\boldsymbol{x};F^*) - \gamma\|\boldsymbol{x}\|^{2k}\right] dF^*(\boldsymbol{x}) + \int_{\mathscr{A}\setminus B_\delta(\boldsymbol{w})} \left[h(\boldsymbol{x};F^*) - \gamma\|\boldsymbol{x}\|^{2k}\right] dF^*(\boldsymbol{x}) \tag{A.19}$$

$$< \varepsilon\left[C + h(\boldsymbol{N}) - \gamma a\right] + (1-\varepsilon)\left[C + h(\boldsymbol{N}) - \gamma a\right] \tag{A.20}$$

$$= C + h(\boldsymbol{N}) - \gamma a, \tag{A.21}$$

where the first term of (A.20) is due to (A.15) and the second is due to (A.9). The above is a contradiction, which completes the proof. □

Appendix B

Convexity and Compactness of Optimization Space

Theorem B.1. *$\mathscr{F}_n(\mathscr{A})$ is convex and $\mathscr{P}_n(\mathscr{A},k,a)$ is convex and compact.*

Proof. We first show the convexity of $\mathscr{F}_n(\mathscr{A})$. Let $F_1, F_2 \in \mathscr{F}_n(\mathscr{A})$, $\lambda \in [0,1]$ and $F_\lambda = \lambda F_1 + (1-\lambda)F_2$. Then since

$$supp(F_\lambda) \subseteq supp(F_1) \cup supp(F_2) \subseteq \mathscr{A}, \tag{B.1}$$

$\mathscr{F}_n(\mathscr{A})$ is convex.

To show convexity of $\mathscr{P}_n(\mathscr{A},k,a)$, let $F_1, F_2 \in \mathscr{P}_n(\mathscr{A},k,a)$, $\lambda \in [0,1]$ and $F_\lambda = \lambda F_1 + (1-\lambda)F_2$. Then

$$\int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k} \, dF_\lambda(\boldsymbol{x}) = \lambda \int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k} \, dF_1(\boldsymbol{x}) + (1-\lambda) \int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k} \, dF_2(\boldsymbol{x}) \tag{B.2}$$

$$\leq \lambda a + (1-\lambda)a = a. \tag{B.3}$$

Thus, F_λ satisfies both the moment constraint and the relation given in (B.1) and we conclude that $\mathscr{P}_n(\mathscr{A},k,a)$ is convex.

It remains to show the compactness of $\mathscr{P}_n(\mathscr{A},k,a)$. Note that the Lévy-Prokhorov metric metrizes weak convergence in $\mathscr{F}(\mathbb{R}^n)$ [26], so sequential compactness is equivalent to compactness. To prove compactness of $\mathscr{P}_n(\mathscr{A},k,a)$, we first show relative compactness, which allows us to conclude that any sequence in $\mathscr{P}_n(\mathscr{A},k,a)$ has a subsequence that converges to some $F \in \mathscr{F}_n(\mathscr{A})$. Showing further that $F \in \mathscr{P}_n(\mathscr{A},k,a)$ will complete the proof.

Observe that each $F \in \mathscr{F}(\mathbb{R}^n)$ is defined on the complete separable metric space $\mathbb{R}^n$ equipped with Euclidean distance. By Prokhorov's Theorem (Theorem 3.2.1 of [26]), relative compactness of $\mathscr{P}_n(\mathscr{A}, k, a)$ is equivalent to tightness of $\mathscr{P}_n(\mathscr{A}, k, a)$, so we will prove the latter.

To show tightness of $\mathscr{P}_n(\mathscr{A}, k, a)$, let $\boldsymbol{X} \sim F \in \mathscr{P}_n(\mathscr{A}, k, a)$, $\varepsilon > 0$ and $D = (a/\varepsilon)^{1/2k}$. Then applying Markov's inequality,

$$\begin{aligned} P\{\boldsymbol{X} \in \mathbb{R}^n \setminus \overline{B_D}(0)\} &= P\{\|\boldsymbol{X}\| > D\} && \text{(B.4)} \\ &\leq P\{\|\boldsymbol{X}\|^{2k} \geq D^{2k}\} && \text{(B.5)} \\ &\leq \frac{\mathbb{E}[\|\boldsymbol{X}\|^{2k}]}{D^{2k}} && \text{(B.6)} \\ &\leq \frac{a}{D^{2k}} && \text{(B.7)} \\ &= \varepsilon. && \text{(B.8)} \end{aligned}$$

This is a uniform upper-bound for $F \in \mathscr{P}_n(\mathscr{A}, k, a)$, so $\mathscr{P}_n(\mathscr{A}, k, a)$ is tight and as a result, relatively compact.

By the relative compactness of $\mathscr{P}_n(\mathscr{A}, k, a)$, any sequence $\{F_m\}_{m=0}^{\infty} \subseteq \mathscr{P}_n(\mathscr{A}, k, a)$ has a subsequence $\{F_{m_j}\}_{m_j=0}^{\infty}$ that converges weakly to $F \in \mathscr{F}(\mathbb{R}^n)$. To show compactness, we must show that $F \in \mathscr{P}_n(\mathscr{A}, k, a)$.

Since each $F_{m_j} \in \mathscr{P}_n(\mathscr{A}, k, a)$,

$$\int_{\mathscr{A}} \|\boldsymbol{x}\|^{2k} \, dF_{m_j}(\boldsymbol{x}) \leq a. \tag{B.9}$$

By Theorem A.3.12 of [11], since $\|\boldsymbol{x}\|^{2k}$ is non-negative and lower semicontinuous by virtue of being continuous,

$$\int_{\mathscr{A}} \|\boldsymbol{x}\|^{2k} \, dF(\boldsymbol{x}) \leq \liminf_{j \to \infty} \int_{\mathscr{A}} \|\boldsymbol{x}\|^{2k} \, dF_{m_j}(\boldsymbol{x}) \leq a. \tag{B.10}$$

Since $\|\boldsymbol{x}\|^{2k} \geq 0$ for all $\boldsymbol{x} \in \mathscr{A}$,

$$\int_{\mathscr{A}} \|\boldsymbol{x}\|^{2k} \, dF(\boldsymbol{x}) \geq 0. \tag{B.11}$$

By (B.10) and (B.11), the limiting distribution F satisfies the even moment constraint imposed by $\mathscr{P}_n(\mathscr{A}, k, a)$. Therefore, when $\mathscr{A} = \mathbb{R}^n$, we conclude that $F \in \mathscr{P}_n(\mathscr{A}, k, a)$.

For the case $\mathscr{A} = \mathscr{C} \in \mathscr{C}_n$, we must also show that $\boldsymbol{X} \in \mathscr{C}$ almost surely. For any index of the subsequence m_j,

$$\int_{\mathscr{C}} dF_{m_j}(\boldsymbol{x}) = 1. \tag{B.12}$$

By the Portmanteau Theorem [6], since $\mathscr{C}$ is closed,

$$\int_{\mathscr{C}} dF(\boldsymbol{x}) \geq \limsup_{j\to\infty} \int_{\mathscr{C}} dF_{m_j}(\boldsymbol{x}) = 1. \tag{B.13}$$

□

Appendix C

Properties of the Output Density

In Section C.1, the output density is shown to be continuous in $\boldsymbol{y}$ and bounded above by a constant. In Section C.2, we present results related to the rate of decay of $p(\cdot;F)$, which are used to show the integrability results of Section C.3. The findings of Section C.3 are then used in Chapter 3 and Appendix D.

Throughout the remaining appendices, we make use of the eigenvalues of $\mathbf{\Sigma}^{-1}$ in order to establish upper and lower bounds. We let η_0 and η_{n-1} be the respectively minimal and maximal eigenvalues of $\mathbf{\Sigma}^{-1}$. Recall that since $\mathbf{\Sigma}$ is positive-definite, $\mathbf{\Sigma}^{-1}$ is as well and $\eta_{n-1} > \eta_0 > 0$.

C.1 Constant Upper Bound and Continuity in $\mathbb{R}^n$

Lemma C.1. *For any $F \in \mathscr{F}_n(\mathscr{A})$ and $\boldsymbol{y} \in \mathbb{R}^n$,*

$$p(\boldsymbol{y};F) \leq \frac{1}{2\pi\sqrt{|\Sigma|}}. \tag{C.1}$$

Proof.

$$\begin{aligned} p(\boldsymbol{y};F) &= \int_{\mathscr{A}} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x})\,dF(\boldsymbol{x}) && \text{(C.2)}\\ &\leq \frac{1}{2\pi\sqrt{|\Sigma|}}\int_{\mathscr{A}} dF(\boldsymbol{x}) && \text{(C.3)}\\ &= \frac{1}{2\pi\sqrt{|\Sigma|}}. && \text{(C.4)} \end{aligned}$$

□

Lemma C.2. *For any $F \in \mathscr{F}_n(\mathscr{A})$, $p(\cdot;F)$ is continuous in $\boldsymbol{y}$.*

Proof. Let $\{\boldsymbol{y}_i\}_{i=0}^{\infty} \subseteq \mathbb{R}^n$ be a sequence converging to $\boldsymbol{y} \in \mathbb{R}^n$. Then,

$$\lim_{i\to\infty} p(\boldsymbol{y}_i;F) = \lim_{i\to\infty} \int_{\mathscr{A}} p_{\boldsymbol{N}}(\boldsymbol{y}_i - \boldsymbol{x})\, dF(\boldsymbol{x}) \tag{C.5}$$

$$= \int_{\mathscr{A}} \lim_{i\to\infty} p_{\boldsymbol{N}}(\boldsymbol{y}_i - \boldsymbol{x})\, dF(\boldsymbol{x}) \tag{C.6}$$

$$= \int_{\mathscr{A}} p_{\boldsymbol{N}}(\boldsymbol{y} - \boldsymbol{x})\, dF(\boldsymbol{x}) \tag{C.7}$$

$$= p(\boldsymbol{y};F), \tag{C.8}$$

where (C.6) is by the Dominated Convergence Theorem and Lemma C.1, while (C.7) is due to the continuity of $p_{\boldsymbol{N}}(\cdot)$. □

C.2 Upper and Lower Bounds on the Output Density

Lemma C.3. *Let $\eta_0 > 0$ be the minimal eigenvalue of $\boldsymbol{\Sigma}^{-1}$ and let $\mathscr{C} \in \mathscr{C}_n$. Then there exists $D > 0$ such that for any $F \in \mathscr{P}_n(\mathscr{C},k,a)$ and any $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,*

$$p(\boldsymbol{y};F) \leq \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{8}\eta_0 \|\boldsymbol{y}\|^2}. \tag{C.9}$$

Proof. Since $\mathscr{C}$ is compact, there exists $D_0 > 0$ such that $\mathscr{C} \subseteq \overline{B_{D_0}}(0)$. Note that the compactness of $\overline{B_{D_0}}(0)$ and the continuity of $p_{\boldsymbol{N}}(\cdot)$ ensure the existence of

$$\max_{\boldsymbol{x} \in \overline{B_D}(0)} p_{\boldsymbol{N}}(\boldsymbol{y} - \boldsymbol{x}). \tag{C.10}$$

Furthermore, from the theory of Rayleigh quotients, for $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_{D_0}}(0)$,

$$\max_{\boldsymbol{x}\in\overline{B_{D_0}}(0)} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) = \max_{\boldsymbol{x}\in\overline{B_{D_0}}(0)} \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{x})} \tag{C.11}$$

$$= \max_{\boldsymbol{x}\in\overline{B_{D_0}}(0)} \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\frac{(\boldsymbol{y}-\boldsymbol{x})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{x})}{\|\boldsymbol{y}-\boldsymbol{x}\|^2}\|\boldsymbol{y}-\boldsymbol{x}\|^2} \tag{C.12}$$

$$\leq \max_{\boldsymbol{x}\in\overline{B_{D_0}}(0)} \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\eta_0\|\boldsymbol{y}-\boldsymbol{x}\|^2} \tag{C.13}$$

$$= \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\eta_0\|\boldsymbol{y}-\frac{D_0}{\|\boldsymbol{y}\|}\boldsymbol{y}\|^2} \tag{C.14}$$

$$= \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\eta_0(1-\frac{D_0}{\|\boldsymbol{y}\|})^2\|\boldsymbol{y}\|^2}. \tag{C.15}$$

Setting $D = 2D_0$ yields for $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,

$$\max_{\boldsymbol{x}\in\overline{B_{D_0}}(0)} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \leq \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\eta_0(1-\frac{D_0}{2D_0})^2\|\boldsymbol{y}\|^2} \tag{C.16}$$

$$= \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{8}\eta_0\|\boldsymbol{y}\|^2}. \tag{C.17}$$

Then we have for $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,

$$p(\boldsymbol{y};F) = \int_{\mathscr{C}} p_{\boldsymbol{N}}(\boldsymbol{y}-\hat{\boldsymbol{x}})\,dF(\hat{\boldsymbol{x}}) \tag{C.18}$$

$$\leq \int_{\mathscr{C}} \max_{\boldsymbol{x}\in\mathscr{C}} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x})\,dF(\hat{\boldsymbol{x}}) \tag{C.19}$$

$$= \max_{\boldsymbol{x}\in\mathscr{C}} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \tag{C.20}$$

$$\leq \max_{\boldsymbol{x}\in\overline{B_{D_0}}(0)} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \tag{C.21}$$

$$\leq \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{8}\eta_0\|\boldsymbol{y}\|^2}. \tag{C.22}$$

□

Lemma C.4. *Let η_{n-1} be the maximal eigenvalue of $\boldsymbol{\Sigma}^{-1}$ and let $\mathscr{C} \in \mathscr{C}_n$. Then there exists*

$D > 0$ *such that for any* $F \in \mathscr{P}_n(\mathscr{C}, k, a)$ *and any* $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,

$$p(\boldsymbol{y}; F) \geq \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-2\eta_{n-1}\|\boldsymbol{y}\|^2} \tag{C.23}$$

Proof. Since $\mathscr{C}$ is compact, there exists $D > 0$ such that $\mathscr{C} \subseteq \overline{B_D}(0)$. Similarly to the proof of Lemma C.3, Rayleigh quotients give

$$\min_{\boldsymbol{x} \in \overline{B_D}(0)} p_{\boldsymbol{N}}(\boldsymbol{y} - \boldsymbol{x}) \geq \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\eta_{n-1}(1+\frac{D}{\|\boldsymbol{y}\|})^2\|\boldsymbol{y}\|^2} \tag{C.24}$$

Then for any $\boldsymbol{y} \in \mathbb{R}^n$,

$$\begin{aligned}
p(\boldsymbol{y}; F) &= \int_{\mathscr{C}} p_{\boldsymbol{N}}(\boldsymbol{y} - \hat{\boldsymbol{x}})\, dF(\hat{\boldsymbol{x}}) && \text{(C.25)}\\
&\geq \int_{\mathscr{C}} \min_{\boldsymbol{x} \in \mathscr{C}} p_{\boldsymbol{N}}(\boldsymbol{y} - \boldsymbol{x})\, dF(\hat{\boldsymbol{x}}) && \text{(C.26)}\\
&= \min_{\boldsymbol{x} \in \mathscr{C}} p_{\boldsymbol{N}}(\boldsymbol{y} - \boldsymbol{x}) && \text{(C.27)}\\
&\geq \min_{\boldsymbol{x} \in \overline{B_D}(0)} p_{\boldsymbol{N}}(\boldsymbol{y} - \boldsymbol{x}) && \text{(C.28)}\\
&= \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\eta_{n-1}(1+\frac{D}{\|\boldsymbol{y}\|})^2\|\boldsymbol{y}\|^2} && \text{(C.29)}
\end{aligned}$$

Now, for any $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,

$$\left(1 + \frac{D}{\|\boldsymbol{y}\|}\right) \leq 2, \tag{C.30}$$

and we have

$$p(\boldsymbol{y}; F) \geq \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-2\eta_{n-1}\|\boldsymbol{y}\|^2}. \tag{C.31}$$

□

Lemma C.5. *Let* $D = (2a)^{\frac{1}{2k}}$. *Then for any* $F \in \mathscr{P}_n(\mathbb{R}^n, k, a)$ *and* $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,

$$p(\boldsymbol{y}; F) \geq \frac{1}{4\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-2\eta_{n-1}\|\boldsymbol{y}\|^2}. \tag{C.32}$$

Proof. For any $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,

$$p(\boldsymbol{y};F) \geq \int_{\overline{B_D}(0)} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x})\,dF(\boldsymbol{x}) \tag{C.33}$$

$$\geq P\{\|\boldsymbol{X}\| \leq D\} \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\eta_{n-1}\|\boldsymbol{y}+\frac{D}{\|\boldsymbol{y}\|}\boldsymbol{y}\|^2} \tag{C.34}$$

$$\geq (1 - P\{\|\boldsymbol{X}\|^{2k} \geq 2a\}) \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\eta_{n-1}(1+\frac{D}{\|\boldsymbol{y}\|})^2\|\boldsymbol{y}\|^2} \tag{C.35}$$

$$\geq (1 - \frac{\mathbb{E}[\|\boldsymbol{X}\|^{2k}]}{2a}) \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\eta_{n-1}(1+\frac{D}{\|\boldsymbol{y}\|})^2\|\boldsymbol{y}\|^2} \tag{C.36}$$

$$\geq (1 - \frac{a}{2a}) \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-2\eta_{n-1}\|\boldsymbol{y}\|^2} \tag{C.37}$$

$$= \frac{1}{4\pi\sqrt{|\Sigma|}} e^{-2\eta_{n-1}\|\boldsymbol{y}\|^2}, \tag{C.38}$$

where (C.36) is due to Markov's Inequality. □

Lemma C.6. *There exist constants $D > 0$ and κ such that for every $F \in \mathscr{P}_n(\mathscr{A}, k, a)$ and any $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,*

$$|\ln p(\boldsymbol{y};F)| \leq 2\eta_{n-1}\|\boldsymbol{y}\|^2 + \kappa. \tag{C.39}$$

Proof. By either Lemma C.4 or C.5, depending on $\mathscr{A}$, there exist constants $D > 0$ and $M > 0$ such that for any $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,

$$c_{\boldsymbol{y}} \triangleq M e^{-2\eta_{n-1}\|\boldsymbol{y}\|^2} \leq p(\boldsymbol{y};F). \tag{C.40}$$

Furthermore, by Lemma C.1, for any $\boldsymbol{y} \in \mathbb{R}^n$,

$$d \triangleq \frac{1}{2\pi\sqrt{|\Sigma|}} \geq p(\boldsymbol{y};F). \tag{C.41}$$

That is, for any $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$, we have that $p(\boldsymbol{y};F) \in [c_{\boldsymbol{y}}, d]$.

Observe that $f(t) \triangleq |\ln(t)|$ is continuous on $(0,\infty)$ with derivative

$$\frac{df(t)}{dt} = \begin{cases} -\frac{1}{t}, & t \in (0,1) \\ \frac{1}{t}, & t \in (1,\infty) \\ \text{undefinded}, & t = 1. \end{cases} \tag{C.42}$$

The only critical point of $f(t)$ occurs at $t=1$, but $f(1)=0$, which is a global minimum. Therefore, $f(t)$ attains its maximum on $[c_{\boldsymbol{y}},d]$ at one of the endpoints. Then for $t \in [c_{\boldsymbol{y}},d]$,

$$\begin{aligned}
f(t) &\le \max\{f(c_{\boldsymbol{y}}), f(d)\} && \text{(C.43)}\\
&\le f(c_{\boldsymbol{y}}) + f(d) && \text{(C.44)}\\
&= |-2\eta_{n-1}\|\boldsymbol{y}\|^2 + \ln M| + |\ln(2\pi\sqrt{|\Sigma|})| && \text{(C.45)}\\
&\le 2\eta_{n-1}\|\boldsymbol{y}\|^2 + |\ln M| + |\ln(2\pi\sqrt{|\Sigma|})|, && \text{(C.46)}
\end{aligned}$$

where (C.44) is due to $f(t)$ being non-negative on its domain. Setting

$$\kappa \triangleq |\ln M| + |\ln(2\pi\sqrt{|\Sigma|})| \tag{C.47}$$

yields the result. □

C.3 Integrability Results

The results in this section show that the output density is sufficiently well behaved in terms of integrability. Theorem C.1 is used in Chapter 3. The other results are used in Appendix D to prove the weak continuity and weak differentiability of the convex optimization objective function $J_\gamma(\cdot)$.

Due to the spherically symmetric form of the upper bounds in Section C.2, we will make use of the polar coordinates $(\|\boldsymbol{y}\|, \theta_1, \ldots, \theta_{n-1}) \in [0,\infty) \times [0,\pi]^{n-2} \times [0,2\pi]$ when integrating in this section. The absolute value of the Jacobian for n-dimensional polar coordinates satisfies [21]

$$|J_n| \le \|\boldsymbol{y}\|^{n-1}. \tag{C.48}$$

Define

$$\begin{aligned}
\beta_n &\triangleq \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{2\pi} d\theta_{n-1} \ldots d\theta_2\, d\theta_1 && \text{(C.49)}\\
&= 2\pi^{n-1}. && \text{(C.50)}
\end{aligned}$$

Note that for a function $f(\boldsymbol{y}) = f(\|\boldsymbol{y}\|)$ and some $D \geq 0$,

$$\int_{\mathbb{R}^n \setminus \overline{B_D}(0)} f(\|\boldsymbol{y}\|)\, d\boldsymbol{y} \leq \int_D^\infty \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} f(\|\boldsymbol{y}\|)\|\boldsymbol{y}\|^{n-1}\, d\theta_{n-1} \dots d\theta_2\, d\theta_1\, d\|\boldsymbol{y}\| \tag{C.51}$$

$$= \beta_n \int_D^\infty f(\|\boldsymbol{y}\|)\|\boldsymbol{y}\|^{n-1} d\|\boldsymbol{y}\|. \tag{C.52}$$

Theorem C.1. *For any $F \in \mathscr{P}_n(\mathscr{A}, k, a)$, $\ln p(\cdot;F) \in L^2_{p_{\boldsymbol{N}}}(\mathbb{R}^n)$*

Proof. By Lemma C.6, there exist constants $D > 0$ and κ such that

$$\ln^2 p(\boldsymbol{y};F) \leq 4\eta_{n-1}^2 \|\boldsymbol{y}\|^4 + 4\kappa\eta_{n-1}\|\boldsymbol{y}\|^2 + \kappa^2 \tag{C.53}$$

$$\triangleq M(\|\boldsymbol{y}\|). \tag{C.54}$$

Integrating against $p_N(\cdot)$,

$$\int_{\mathbb{R}^n \setminus \overline{B_D}(0)} \ln^2 p(\boldsymbol{y};F) p_{\boldsymbol{N}}(\boldsymbol{y})\, dy \leq \int_{\mathbb{R}^n \setminus \overline{B_D}(0)} M(\|\boldsymbol{y}\|) \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\boldsymbol{y}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{y}}\, dy \tag{C.55}$$

$$\leq \int_{\mathbb{R}^n \setminus \overline{B_D}(0)} M(\|\boldsymbol{y}\|) \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\eta_0\|\boldsymbol{y}\|^2}\, dy \tag{C.56}$$

$$\leq \beta_n \int_D^\infty M(\|\boldsymbol{y}\|)\|\boldsymbol{y}\|^{n-1} \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\eta_0\|\boldsymbol{y}\|^2} d\,\|\boldsymbol{y}\| \tag{C.57}$$

$$< \infty, \tag{C.58}$$

since $\eta_0 > 0$ and $M(\|\boldsymbol{y}\|)\|\boldsymbol{y}\|^{n-1}$ is a polynomial in $\|\boldsymbol{y}\|$. Noting that $\ln^2 p(\cdot;F)$ is continous and, thus integrable on the compact set $\overline{B_D}(0)$, completes the proof. □

Lemma C.7. *Let $\mathscr{C} \in \mathscr{C}_n$ and let $F_0, F_1 \in \mathscr{P}_n(\mathscr{C}, k, a)$. Then*

$$\int_{\mathbb{R}^n} |p(\boldsymbol{y};F_0) \ln p(\boldsymbol{y};F_1)|\, d\boldsymbol{y} < \infty. \tag{C.59}$$

Proof. Since $|p(\boldsymbol{y};F_0)\ln p(\boldsymbol{y};F_1)|$ is continuous, it is integrable on any closed ball $\overline{B_D}(0)$, where $D > 0$. Therefore, we need only show that there exists $D > 0$ for which $|p(\boldsymbol{y};F_0)\ln p(\boldsymbol{y};F_1)|$ is integrable on $\mathbb{R}^n \setminus \overline{B_D}(0)$.

Let $\eta_{n-1} \geq \eta_0 > 0$ be the maximal and minimal eigenvalues of $\boldsymbol{\Sigma}^{-1}$. By Lemma C.3, there exists $D_0 > 0$ such that for all $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_{D_0}}(0)$,

$$p(\boldsymbol{y};F_0) \leq \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{8}\eta_0\|\boldsymbol{y}\|^2}. \tag{C.60}$$

Furthermore, by Lemma C.6, there exist constants $D_1 > 0$ and κ such that for all $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_{D_1}}(0)$,

$$|\ln p(\boldsymbol{y};F_1)| \leq 2\eta_{n-1}\|\boldsymbol{y}\|^2 + \kappa. \tag{C.61}$$

Let $D = \max\{D_0, D_1\}$. Then by (C.61) and (C.60),

$$\int_{\mathbb{R}^n \setminus \overline{B_D}(0)} |p(\boldsymbol{y};F_0) \ln p(\boldsymbol{y};F_1)| \, d\boldsymbol{y} \leq \int_{\mathbb{R}^n \setminus \overline{B_D}(0)} \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{8}\eta_0\|\boldsymbol{y}\|^2} (2\eta_{n-1}\|\boldsymbol{y}\|^2 + \kappa) \, d\boldsymbol{y} \tag{C.62}$$

$$\leq \beta_n \int_D^\infty \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{8}\eta_0\|\boldsymbol{y}\|^2} (2\eta_{n-1}\|\boldsymbol{y}\|^2 + \kappa) \|\boldsymbol{y}\|^{n-1} \, d\|\boldsymbol{y}\| \tag{C.63}$$

$$< \infty. \tag{C.64}$$

Therefore, $|p(\boldsymbol{y};F_0) \ln p(\boldsymbol{y};F_1)|$ is integrable on $\mathbb{R}^n \setminus \overline{B_D}(0)$, which completes the proof. □

Lemma C.8. *For any $F \in \mathscr{P}_n(\mathbb{R}^n, k, a)$,*

$$\int_{\mathbb{R}^n} |p(\boldsymbol{y};F) \ln p(\boldsymbol{y};F)| \, d\boldsymbol{y} < \infty. \tag{C.65}$$

Proof. This proof follows along the lines of the justification for the use of the Dominated Convergence Theorem in (19) of Appendix E in [17]. The most obvious difference is the domain of integration – the volume of a ball of radius r in $\mathbb{R}^n$ grows proportionally to r^n and forces a stricter upper-bound than would be needed for an integral over $\mathbb{R}$. This is the reasoning for the restriction $2k > n$ when $\mathscr{A} = \mathbb{R}^n$.

We proceed by partitioning $\mathbb{R}^n$ into 3 sets and showing that $|p(\boldsymbol{y};F) \ln p(\boldsymbol{y};F)|$ is integrable over each of them. Let $\eta_0 > 0$ be the smallest eigenvalue of $\boldsymbol{\Sigma}^{-1}$ and note that there exists $D > 0$ such that for all $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$,

$$e^{-\eta_0\frac{1}{4}\|\boldsymbol{y}\|^2} \leq \|\boldsymbol{y}\|^{-2k}. \tag{C.66}$$

Let $F \in \mathscr{P}_n(\mathscr{A},k,a)$ and let

$$A_1 \triangleq \{\boldsymbol{y} \in \mathbb{R}^n \mid \ln p(\boldsymbol{y};F) > 0\}, \tag{C.67}$$

$$A_2 \triangleq \mathbb{R}^n \setminus A_1 \setminus \overline{B_D}(0), \tag{C.68}$$

$$A_3 \triangleq \overline{B_D}(0) \setminus A_1. \tag{C.69}$$

The continuity of $\ln p(\boldsymbol{y};F)$ implies that A_1 is Lebesgue-measurable and the integral can be split up as

$$\int_{\mathbb{R}^n} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|\,d\boldsymbol{y} = \sum_{i=1}^{3} \int_{A_i} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|\,d\boldsymbol{y} \tag{C.70}$$

Since by Lemma C.1, $p(\boldsymbol{y};F)$ is bounded above, we have for some $M \geq 1$,

$$\int_{A_1} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|\,d\boldsymbol{y} = \int_{A_1} p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)\,d\boldsymbol{y} \tag{C.71}$$

$$\leq \int_{A_1} p(\boldsymbol{y};F)\ln M\,d\boldsymbol{y} \tag{C.72}$$

$$\leq \ln M. \tag{C.73}$$

To see that the integral over A_3 is finite, observe that $|p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|$ is continuous in $\boldsymbol{y}$ and therefore, integrable over the compact set $\overline{B_D}(0)$. Furthermore, since the integrand is non-negative,

$$\int_{A_3} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|\,d\boldsymbol{y} = \int_{\overline{B_D}(0)\setminus A_1} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|\,d\boldsymbol{y} \tag{C.74}$$

$$\leq \int_{\overline{B_D}(0)} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|\,d\boldsymbol{y} \tag{C.75}$$

$$< \infty. \tag{C.76}$$

It remains to show integrability over A_2. For $\boldsymbol{y} \in A_2$,

$$p(\boldsymbol{y};F) = \int_{\overline{B_{\|\boldsymbol{y}\|/2}}(0)} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x})\,dF(\boldsymbol{x}) + \int_{\mathbb{R}^n \setminus \overline{B_{\|\boldsymbol{y}\|/2}}(0)} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x})\,dF(\boldsymbol{x}) \tag{C.77}$$

$$\leq e^{-\eta_0\|\frac{1}{2}\boldsymbol{y}\|^2} P\left\{\|\boldsymbol{X}\| \leq \frac{\|\boldsymbol{y}\|}{2}\right\} + p_{\boldsymbol{N}}(0) P\left\{\|\boldsymbol{X}\| > \frac{\|\boldsymbol{y}\|}{2}\right\} \tag{C.78}$$

$$\leq e^{-\eta_0\frac{1}{4}\|\boldsymbol{y}\|^2} + p_{\boldsymbol{N}}(0) P\left\{\|\boldsymbol{X}\|^{2k} > \frac{\|\boldsymbol{y}\|^{2k}}{2^{2k}}\right\} \tag{C.79}$$

$$\leq e^{-\eta_0\frac{1}{4}\|\boldsymbol{y}\|^2} + p_{\boldsymbol{N}}(0) 2^{2k} \mathbb{E}[\|\boldsymbol{X}\|^{2k}]\|\boldsymbol{y}\|^{-2k} \tag{C.80}$$

$$\leq e^{-\eta_0\frac{1}{4}\|\boldsymbol{y}\|^2} + p_{\boldsymbol{N}}(0) 2^{2k} a\|\boldsymbol{y}\|^{-2k} \tag{C.81}$$

$$\leq \left(\frac{2^{2k-1}a}{\pi|\boldsymbol{\Sigma}|} + 1\right)\|\boldsymbol{y}\|^{-2k} \tag{C.82}$$

In the above, (C.80) is due to Markov's inequality and (C.82) comes from $\boldsymbol{y} \in \mathbb{R}^n \setminus \overline{B_D}(0)$.

For any $0 < \delta < 1$ and $0 < x \leq 1$, consider the inequality (see [17])

$$|x \ln x| \leq \frac{x^\delta}{1-\delta}. \tag{C.83}$$

Since $\boldsymbol{y} \notin A_1$, (C.83) may be applied to $p(\boldsymbol{y};F)$, yielding

$$|p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)| \leq \frac{p^\delta(\boldsymbol{y};F)}{1-\delta} \tag{C.84}$$

$$\leq \kappa\|\boldsymbol{y}\|^{-2k\delta}, \tag{C.85}$$

where

$$\kappa = \frac{1}{1-\delta}\left(\frac{2^{2k-1}a}{\pi|\boldsymbol{\Sigma}|} + 1\right)^\delta. \tag{C.86}$$

Then, using polar coordinates,

$$\int_{A_2} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|\,d\boldsymbol{y} \le \kappa \int_{A_2} \|\boldsymbol{y}\|^{-2k\delta}\,d\boldsymbol{y} \tag{C.87}$$

$$= \kappa \int_{\mathbb{R}^n \setminus A_1 \setminus \overline{B_D}(0)} \|\boldsymbol{y}\|^{-2k\delta}\,d\boldsymbol{y} \tag{C.88}$$

$$\le \kappa \int_{\mathbb{R}^n \setminus \overline{B_D}(0)} \|\boldsymbol{y}\|^{-2k\delta}\,d\boldsymbol{y} \tag{C.89}$$

$$\le \kappa\beta_n \int_D^\infty \|\boldsymbol{y}\|^{-2k\delta}\|\boldsymbol{y}\|^{n-1} d\,\|\boldsymbol{y}\| \tag{C.90}$$

$$= \kappa\beta_n \int_D^\infty \|\boldsymbol{y}\|^{-2k\delta+n-1} d\,\|\boldsymbol{y}\| \tag{C.91}$$

The integral in (C.91) is finite if and only if $-2k\delta+n-1<-1$, or equivalently when

$$k > \frac{n}{2\delta}. \tag{C.92}$$

Since $\delta < 1$, this condition implies that $k > n/2$. Furthermore, if for some $\varepsilon > 0$, $k = n/2+\varepsilon$, then

$$\delta = \frac{n}{n+\varepsilon}, \tag{C.93}$$

satisfies (C.92). Therefore, the existence of $\delta \in (0,1)$ satisfying (C.92) is equivalent to

$$k > \frac{n}{2}. \tag{C.94}$$

Therefore, when $2k > n$,

$$\int_{A_2} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F)|\,d\boldsymbol{y} < \infty. \tag{C.95}$$

Together with (C.73) and (C.76), this gives the result. □

Appendix D

Properties of the Objective Functional

The aim of this section is to discuss the weak continuity, strict concavity and weak differentiability of the objective function,

$$J_\gamma(F) \triangleq I(F) - \gamma g(F), \tag{D.1}$$

for the optimization problem posed in (3.14). These properties are instrumental in the establishment and subsequent analysis of the convex optimization problem considered in Chapter 3.

To deal with a technical point in the proof of Theorem A.4, we introduce $b \in \mathbb{R}_{>0}$ and prove concavity, convexity and weak differentiability on $\mathscr{P}_n(\mathscr{A}, k, b)$.

Weak Continuity: The first property we examine is the weak continuity of $I(\cdot)$, which is necessary for the application of Theorem A.1.

Theorem D.1. *$I(\cdot)$ is weak continuous on $\mathscr{P}_n(\mathscr{A}, k, a)$.*

Proof. The proof here largely follows the proof in Appendix E of [17], the main difference being the justification of the applicability of the Dominated Convergence Theorem in (19) of [17]. This step requires that, for any $F \in \mathscr{P}_n(\mathscr{A}, k, a)$,

$$\int_{\mathbb{R}^n} |p(\boldsymbol{y}; F) \ln p(\boldsymbol{y}; F)| \, d\boldsymbol{y} < \infty, \tag{D.2}$$

which holds by either Lemma C.7 or Lemma C.8, depending on $\mathscr{A}$. $\square$

Concavity and Convexity: The strict concavity of $I(\cdot)$ and the convexity of $g(\cdot)$ are needed for the theorems in Appendix A. Note that since $\gamma \geq 0$, these results imply that $J_\gamma(\cdot)$ is strictly concave.

Theorem D.2. *For any $b \in \mathbb{R}_{>0}$, $I(\cdot)$ is strictly concave on $\mathscr{P}_n(\mathscr{A}, k, b)$.*

Proof. See Appendix E of [17]. □

Theorem D.3. *For any $b \in \mathbb{R}_{>0}$, $g(\cdot)$ is convex on $\mathscr{P}_n(\mathscr{A}, k, b)$.*

Proof. Let $\lambda \in [0,1]$ and $F_0, F_1 \in \mathscr{P}_n(\mathscr{A}, k, b)$. Then,

$$g(\lambda F_0 + (1-\lambda)F_1) = \int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k} d\lambda F_0(\boldsymbol{x}) + \int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k} d(1-\lambda)F_1(\boldsymbol{x}) - a \tag{D.3}$$

$$= \lambda(\int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k} dF_0(\boldsymbol{x}) - a) + (1-\lambda)(\int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k} dF_1(\boldsymbol{x}) - a) \tag{D.4}$$

$$= \lambda g(F_0) + (1-\lambda) g(F_1). \tag{D.5}$$

□

Weak Differentiability: We make use of the following notion of a derivative of a function defined on a convex set Ω[17].

Definition D.1. Define the weak derivative of $L : \mathscr{F} \to \mathbb{R}$ at F_0 in the direction F by

$$L'(F_0, F) = \lim_{\lambda \downarrow 0} \frac{L((1-\lambda)F_0 + \lambda F) - L(F_0)}{\lambda}, \tag{D.6}$$

whenever it exists.

Weak differentiability of $J_\gamma(\cdot)$ is necessary for the application of the optimality condition in Theorem A.3.

Theorem D.4. *For any $b \in \mathbb{R}_{>0}$, $g(\cdot)$ is weakly differentiable on $\mathscr{P}_n(\mathscr{A}, k, b)$. Furthermore, for any $F_0, F \in \mathscr{P}_n(\mathscr{A}, k, b)$, the weak derivative is finite and given by*

$$g'(F_0, F) = g(F) - g(F_0). \tag{D.7}$$

Proof. Let $\lambda \in [0,1]$ and $F_0, F \in \mathscr{P}_n(\mathscr{A},k,b)$. Then,

$$g(\lambda F+(1-\lambda)F_0)-g(F_0) = g(\lambda(F-F_0)+F_0)-g(F_0) \tag{D.8}$$

$$= \int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k}\, d\Big[\lambda(F(\boldsymbol{x})-F_0(\boldsymbol{x}))+F_0(\boldsymbol{x})\Big]-a - \left(\int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k}\, dF_0(\boldsymbol{x})-a\right) \tag{D.9}$$

$$= \lambda\left(\int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k}\, dF(\boldsymbol{x}) - \int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k}\, dF_0(\boldsymbol{x})\right) \tag{D.10}$$

$$= \lambda\left[\int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k}\, dF(\boldsymbol{x}) - a - \left(\int_{\mathbb{R}^n} \|\boldsymbol{x}\|^{2k}\, dF_0(\boldsymbol{x}) - a\right)\right] \tag{D.11}$$

$$= \lambda(g(F)-g(F_0)). \tag{D.12}$$

Dividing by λ and taking the limit as it goes to 0 gives (D.7). Finally, since $F_0, F \in \mathscr{P}_n(\mathscr{A},k,b)$, we have that $g(F_0), g(F) \in [-b,0]$, so this quantity is finite. □

Theorem D.5. *For any $b \in \mathbb{R}_{>0}$, $I(\cdot)$ is weakly-differentiable on $\mathscr{P}_n(\mathscr{A},k,b)$. Furthermore, for any $F \in \mathscr{P}_n(\mathscr{A},k,b)$, the weak derivative at F^* is given by*

$$I'(F^*,F) = \int_{\mathbb{R}^n} h(\boldsymbol{x};F^*)\, dF(\boldsymbol{x}) - h_{\boldsymbol{Y}}(F^*). \tag{D.13}$$

Proof. This proof largely follows Appendix E from [17]. The step that requires special attention is the application of the Dominated Convergence Theorem in (27) of [17]. That is, we would like to show the integrability of

$$|(p(\boldsymbol{y};F)+p(\boldsymbol{y};F^*))\ln p(\boldsymbol{y};F^*)| \le |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)| + |p(\boldsymbol{y};F^*)\ln p(\boldsymbol{y};F^*)| \tag{D.14}$$

Case 1: $\mathscr{A} \in \mathscr{C}_n$. The result follows by Lemma C.7.

Case 2: $\mathscr{A} = \mathbb{R}^n$. The second term on the right side of (D.14) is integrable by Lemma C.8. Therefore, it remains only to show that $|p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)|$ is integrable. Define

$$S = \{\boldsymbol{y} \in \mathbb{R}^n \mid \ln p(\boldsymbol{y};F^*) \ge 0\}. \tag{D.15}$$

Then

$$\int_{\mathbb{R}^n} |p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)|\,d\boldsymbol{y} = \int_S p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y} - \int_{\mathbb{R}^n\setminus S} p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y} \quad \text{(D.16)}$$

$$= \int_S p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y} - \Bigg(\int_{\mathbb{R}^n} p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y} - \int_S p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y}\Bigg) \quad \text{(D.17)}$$

$$= 2\int_S p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y} - \int_{\mathbb{R}^n} p(\boldsymbol{y};F)\ln p(\boldsymbol{y};F^*)\,d\boldsymbol{y}. \quad \text{(D.18)}$$

The first term is finite by Lemma C.1 applied to $\ln p(\boldsymbol{y};F^*)$, so it remains to show that the second term is finite.

Using a similar method to that of [17], let $u(\cdot)$ be the n-dimensional Heaviside step function given by

$$u(\boldsymbol{x}) = \prod_{i=1}^{n} u_1(x_i), \quad \text{(D.19)}$$

where $u_1(\cdot)$ is the univariate Heaviside step function. For $\boldsymbol{x}_s \in \mathbb{R}^n$ with $\|\boldsymbol{x}_s\| \geq b^{\frac{1}{2k}}$, let

$$F_s(\boldsymbol{x}) = (1 - \frac{b}{\|\boldsymbol{x}_s\|^{2k}})u(\boldsymbol{x}) + \frac{b}{\|\boldsymbol{x}_s\|^{2k}}u(\boldsymbol{x}-\boldsymbol{x}_s). \quad \text{(D.20)}$$

Then since $g(F_s) = 0$, $F_s \in \mathscr{P}_n(\mathscr{A},k,b)$. We also have

$$p(\boldsymbol{y};F_s) = (1 - \frac{b}{\|\boldsymbol{x}_s\|^{2k}})p_{\boldsymbol{N}}(\boldsymbol{y}) + \frac{b}{\|\boldsymbol{x}_s\|^{2k}}p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}_s). \quad \text{(D.21)}$$

Using the upper bound provided by Lemma C.6 on $|\ln p(\cdot;F^*)|$, we have that

$$\int_{\mathbb{R}^n} |p(\boldsymbol{y};F_s)\ln p(\boldsymbol{y};F^*)|\,d\boldsymbol{y} < \infty. \quad \text{(D.22)}$$

Therefore, the Dominated Convergence step in (27) of [17] goes through for $I'(F^*,F_s)$ and hence

$$I'(F^*,F_s) = \int_{\mathbb{R}^n} h(\boldsymbol{x};F^*)\,dF_s(\boldsymbol{x}) - h_{\boldsymbol{Y}}(F^*). \quad \text{(D.23)}$$

By the complementary slackness of Theorem A.2, $\gamma g(F^*) = 0$. It follows from Theorem D.4 that

$$\gamma g'(F_0, F) = \gamma g(F_s) - \gamma g(F^*) = 0 \tag{D.24}$$

and $J'(F^*, F_s) = I'(F^*, F_s)$.

To get an upper bound on $h(\cdot, F^*)$, let

$$\Omega = \{F_\lambda \mid \lambda \in [0,1], F_\lambda = \lambda F^* + (1-\lambda)F_s\}. \tag{D.25}$$

Note that $\Omega \subseteq \mathscr{P}_n(\mathscr{A}, k, b)$ is convex and, for $F_\lambda \in \Omega$,

$$p(\boldsymbol{y}; F_\lambda) = \lambda p(\boldsymbol{y}; F^*) + (1-\lambda) p(\boldsymbol{y}; F_s). \tag{D.26}$$

Since both $|p(\cdot; F_s) \ln p(\cdot; F^*)|$ and $|p(\cdot; F^*) \ln p(\cdot; F^*)|$ are integrable, so is $|p(\cdot; F_\lambda) \ln p(\cdot; F^*)|$. Therefore, $I(\cdot)$ is weak differentiable on Ω. Furthermore, since $\Omega \subseteq \mathscr{P}_n(\mathscr{A}, k, b)$, F^* is optimal on Ω. By Theorem A.3 applied on Ω, $I'(F^*, F_s) = J'(F^*, F_s) \leq 0$, so that

$$\int_{\mathbb{R}^n} h(\boldsymbol{x}; F^*)\, dF_s(\boldsymbol{x}) \leq h_{\boldsymbol{Y}}(F^*). \tag{D.27}$$

Substituting F_s and rearranging, yields

$$h(\boldsymbol{x}_s; F^*) \leq \frac{h_{\boldsymbol{Y}}(F^*) - h(0; F^*)}{b} \|\boldsymbol{x}_s\|^{2k} + h(0; F^*). \tag{D.28}$$

Recalling that the above is true for any $\boldsymbol{x}_s \in \mathbb{R}^n \setminus B_{b^{1/2k}}(0)$, then for any $F \in \mathscr{P}_n(\mathscr{A}, k, b)$,

$$\int_{\mathbb{R}^n} h(\boldsymbol{x}; F^*)\, dF(\boldsymbol{x}) = \int_{\overline{B_{b^{1/2k}}}(0)} h(\boldsymbol{x}; F^*)\, dF(\boldsymbol{x}) + \int_{\mathbb{R}^n \setminus \overline{B_{b^{1/2k}}}(0)} h(\boldsymbol{x}; F^*)\, dF(\boldsymbol{x}) \tag{D.29}$$

$$\leq \int_{\overline{B_{b^{1/2k}}}(0)} h(\boldsymbol{x}; F^*)\, dF(\boldsymbol{x}) + \int_{\mathbb{R}^n \setminus \overline{B_{b^{1/2k}}}(0)} \left[\frac{h_{\boldsymbol{Y}}(F^*) - h(0; F^*)}{b} \|\boldsymbol{x}\|^{2k} + h(0; F^*) \right] dF(\boldsymbol{x}) \tag{D.30}$$

$$\leq \int_{\overline{B_{b^{1/2k}}}(0)} h(\boldsymbol{x}; F^*)\, dF(\boldsymbol{x}) + h_{\boldsymbol{Y}}(F^*). \tag{D.31}$$

By Lemma C.8, $h_{\boldsymbol{Y}}(F^*)$ is finite and the continuity of $h(\boldsymbol{x}; F^*)$ ensures the integrability of the first term on the compact set $\overline{B_{b^{1/2k}}}(0)$. Then we have

$$
\begin{aligned}
\infty &> \int_{\mathbb{R}^n} h(\boldsymbol{x}; F^*)\, dF(\boldsymbol{x}) && \text{(D.32)} \\
&= -\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \ln p(\boldsymbol{y}; F^*)\, d\boldsymbol{y}\, dF(\boldsymbol{x}) && \text{(D.33)} \\
&= -\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \ln p(\boldsymbol{y}; F^*)\, dF(\boldsymbol{x})\, d\boldsymbol{y} && \text{(D.34)} \\
&= -\int_{\mathbb{R}^n} p(\boldsymbol{y}; F) \ln p(\boldsymbol{y}; F^*)\, d\boldsymbol{y}, && \text{(D.35)}
\end{aligned}
$$

where (D.34) is justified by Lemma C.6 and Fubini's Theorem. By (D.35), (D.18) is finite.

□

Appendix E

Analycity of Marginal Entropy Density

Lemma E.1. *For any* $F \in \mathscr{P}_n(\mathscr{A}, k, a)$*,* $h(\boldsymbol{x}; F)$ *has an analytic extension to an entire function on* $\mathbb{C}^n$*.*

Proof. For convenience of notation, we will prove the the case of $n = 2$ here.

We proceed by applying an orthogonal transformation to $\boldsymbol{N}$ to shift the analysis from 2 complex variables to 1. Since $\boldsymbol{\Sigma}$ is a real symmetric matrix, it can be diagonalized by orthogonal matrices $\boldsymbol{Q}$ and $\boldsymbol{Q}^T$. That is, letting $\lambda_1 \geq \lambda_0 > 0$ be the eigenvalues of $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Lambda} \triangleq \begin{bmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{bmatrix} = \boldsymbol{Q}^T \boldsymbol{\Sigma} \boldsymbol{Q}. \tag{E.1}$$

Moreover,

$$\boldsymbol{U} \triangleq \boldsymbol{Q}^T \boldsymbol{N} \sim \mathscr{N}(0, \boldsymbol{\Lambda}). \tag{E.2}$$

Therefore, the density of $\boldsymbol{U}$ is given by

$$p_{\boldsymbol{U}}(\boldsymbol{u}) = p_{U_1}(u_1) p_{U_2}(u_2) = \frac{1}{\sqrt{2\pi\lambda_0}} e^{-\frac{u_1^2}{2\lambda_0}} \frac{1}{\sqrt{2\pi\lambda_1}} e^{-\frac{u_2^2}{2\lambda_1}}. \tag{E.3}$$

Consider the extension of $h(\boldsymbol{x};F)$ to $\mathbb{C}^2$:

$$h(\boldsymbol{z};F) = -\int_{\mathbb{R}^2} p_N(\boldsymbol{y}-\boldsymbol{z}) \ln p(\boldsymbol{y};F)\, d\boldsymbol{y} \tag{E.4}$$

$$= -\int_{\mathbb{R}^2} \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{z})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{z})} \ln p(\boldsymbol{y};F)\, d\boldsymbol{y} \tag{E.5}$$

$$= -\int_{\mathbb{R}^2} \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{z})^T \boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T(\boldsymbol{y}-\boldsymbol{z})} \ln p(\boldsymbol{y};F)\, d\boldsymbol{y}. \tag{E.6}$$

Applying the change of variables $\boldsymbol{v} = \boldsymbol{Q}^T\boldsymbol{y}$ and $\boldsymbol{w} = \boldsymbol{Q}^T\boldsymbol{z}$ gives

$$h(\boldsymbol{z};F) = h(\boldsymbol{Q}\boldsymbol{w};F) \tag{E.7}$$

$$= -\int_{\mathbb{R}^2} \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{v}-\boldsymbol{w})^T \boldsymbol{\Lambda}^{-1}(\boldsymbol{v}-\boldsymbol{w})} \ln p(Q\boldsymbol{v};F)\, d\boldsymbol{v} \tag{E.8}$$

$$= -\int_{\mathbb{R}^2} \frac{1}{\sqrt{2\pi\lambda_0}} e^{-\frac{(v_1-w_1)^2}{2\lambda_0}} \frac{1}{\sqrt{2\pi\lambda_1}} e^{-\frac{(v_2-w_2)^2}{2\lambda_1}} \ln p(Q\boldsymbol{v};F)\, d\boldsymbol{v} \tag{E.9}$$

$$= -\int_{\mathbb{R}^2} p_{U_1}(v_1-w_1)\, p_{U_2}(v_2-w_2) \ln p(Q\boldsymbol{v};F)\, d\boldsymbol{v}. \tag{E.10}$$

Therefore, the analycity of $h(\boldsymbol{z};F)$ on $\mathbb{C}^2$ is equivalent to the analycity of

$$h_{\boldsymbol{U}}(\boldsymbol{w};F) \triangleq -\int_{\mathbb{R}^2} p_{U_1}(v_1-w_1)\, p_{U_2}(v_2-w_2) \ln p(Q\boldsymbol{v};F)\, d\boldsymbol{v} \tag{E.11}$$

on $\mathbb{C}^2$.

Proof that $h_{\boldsymbol{U}}(\cdot;F)$ is an entire function on $\mathbb{C}^2$: We will first exploit the independence of U_1 and U_2 to prove that $h_{\boldsymbol{U}}((\cdot,w_2);F)$ is an entire function on $\mathbb{C}$. The symmetry of the problem will allow us to see that $h_{\boldsymbol{U}}((w_1,\cdot);F)$ is entire as well. Finally, we will use Hartog's Theorem to conclude that $h_{\boldsymbol{U}}(\cdot;F)$ is an entire function on $\mathbb{C}^2$.

We will use Morera's Theorem to show that $h_{\boldsymbol{U}}((\cdot,w_2);F)$ is entire. To that end, we would like to show that for any closed triangle $\Delta \subseteq \mathbb{C}$ with boundary $\partial\Delta$ and any fixed $w_2 \in \mathbb{C}$,

$$\oint_{\partial\Delta} h_{\boldsymbol{U}}(\boldsymbol{w};F)\,dw_1 = -\oint_{\partial\Delta}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\kappa_1 e^{-\frac{(v_1-w_1)^2}{2\sigma_1^2}} p_{U_2}(v_2-w_2)\ln p(\boldsymbol{Q}(v_1,v_2)^T;F)\,dv_1 dv_2 dw_1 \tag{E.12}$$

$$= 0. \tag{E.13}$$

We do this by using Fubini's Theorem to justify swapping the order of integration and using the analycity of $p_{U_1}(z)$ on $\mathbb{C}$.

To upper bound the integrand on the right side of (E.12), let $\alpha_0, \beta_0 > 0$ be sufficiently large such that

$$\partial\Delta \subseteq \{z \in \mathbb{C} \mid Re\{z\} \in [-\alpha_0, \alpha_0], Im\{z\} \in [-\beta_0, \beta_0]\}. \tag{E.14}$$

By Lemma C.6, there exist constants $D > \alpha_0$ and κ such that for all $\boldsymbol{v} \in \mathbb{R}^2 \setminus \overline{B_D}(0)$,

$$|\ln p(\boldsymbol{Q}\boldsymbol{v};F)| \leq 2\eta_{n-1}\|\boldsymbol{Q}\boldsymbol{v}\|^2 + \kappa. \tag{E.15}$$

Since $\boldsymbol{Q}$ is orthogonal, $\|\boldsymbol{Q}\boldsymbol{v}\| = \|\boldsymbol{v}\|$ and (E.15) can be rewritten as

$$|\ln p(\boldsymbol{Q}(v_1,v_2)^T;F)| \leq 2\eta_{n-1}(v_1^2+v_2^2)+\kappa \tag{E.16}$$

We proceed by splitting the integral with respect to v_1 into the intervals

$$A_1 \triangleq (-\infty, D), \tag{E.17}$$

$$A_2 \triangleq [-D, D], \tag{E.18}$$

$$A_3 \triangleq (D, \infty). \tag{E.19}$$

Then for $w_1 = \alpha + i\beta \in \partial\Delta$ and $v_1 \in A_3$,

$$\left|e^{-\frac{(v_1-w_1)^2}{2\sigma_1^2}}\right| = \left|e^{-\frac{1}{2\sigma_1^2}((v_1-\alpha)^2-\beta^2-2i(v_1-\alpha)\beta)}\right| \tag{E.20}$$

$$= e^{\frac{1}{2\sigma_1^2}\beta^2} e^{-\frac{1}{2\sigma_1^2}(v_1-\alpha)^2} \tag{E.21}$$

$$\leq e^{\frac{1}{2\sigma_1^2}\beta_0^2} e^{-\frac{1}{2\sigma_1^2}(v_1-\alpha_0)^2}. \tag{E.22}$$

Therefore, substituting (E.16) and (E.22), for any $w_1 \in \partial\Delta$,

$$\int_D^\infty \left|\kappa_1 e^{-\frac{(v_1-w_1)^2}{2\sigma_1^2}} \ln p(\boldsymbol{Q}(v_1,v_2)^T;F)\right| dv_1 \leq \kappa_1 e^{\frac{1}{2\sigma_1^2}\beta_0^2} \int_D^\infty e^{-\frac{1}{2\sigma_1^2}(v_1-\alpha_0)^2} (2\eta_{n-1}(v_1^2+v_2^2)+\kappa)\, dv_1 \tag{E.23}$$

$$= c_0 v_2^2 + c_1, \tag{E.24}$$

for some constants $c_0, c_1 < \infty$. Applying similar reasoning when $v_1 \in A_1$, shows that there are constants $c_2, c_3 < \infty$ for which

$$\int_{-\infty}^{-D} \left|\kappa_1 e^{-\frac{(v_1-w_1)^2}{2\sigma_1^2}} \ln p(\boldsymbol{Q}(v_1,v_2)^T;F)\right| dv_1 \leq c_2 v_2^2 + c_3. \tag{E.25}$$

Finally, since the integrand is continuous in v_1, it is integrable on the compact set A_2. Since $\ln p(\boldsymbol{Q}(v_1,\cdot)^T;F)$ is continuous in v_2, there exists a continuous real-valued function $\phi(\cdot)$ such that for any $v_2 \in \mathbb{R}$,

$$\phi(v_2) = \int_{-D}^{D} \left| e^{-\frac{(v_1-w_1)^2}{2\sigma_1^2}} \ln p(\boldsymbol{Q}(v_1,v_2)^T;F)\right| dv_1. \tag{E.26}$$

Furthermore, by Lemma C.6 and the orthogonality of $\boldsymbol{Q}$, for any $v_2 \notin [-D,D]$,

$$\phi(v_2) \leq \int_{-D}^{D} \left| e^{-\frac{(v_1-w_1)^2}{2\sigma_1^2}} 2\eta_{n-1}(v_1^2+v_2^2)+\kappa)\right| dv_1 \tag{E.27}$$

$$= c_4 v_2^2 + c_5, \tag{E.28}$$

for some constants $c_4, c_5 < \infty$.

We proceed to integration with respect to v_2 using the same intervals A_i, $i \in \{1,2,3\}$, that were used for v_1. For any $w_2 = t + ir \in \mathbb{C}$ and $w_1 \in \partial\Delta$,

$$\psi(w_1,w_2) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|\kappa_1 e^{-\frac{(v_1-w_1)^2}{2\sigma_1^2}} p_{U_2}(v_2-w_2) \ln p(\boldsymbol{Q}(v_1,v_2)^T;F)\right| dv_1 dv_2 \tag{E.29}$$

$$= \int_{-\infty}^{\infty} \sum_{i-1}^{3} \int_{A_i} \left|\kappa_1 e^{-\frac{(v_1-w_1)^2}{2\sigma_1^2}} \kappa_2 e^{-\frac{(v_2-w_2)^2}{2\sigma_2^2}} \ln p(\boldsymbol{Q}(v_1,v_2)^T;F)\right| dv_1 dv_2 \tag{E.30}$$

$$\leq \kappa_2 e^{\frac{r^2}{2\sigma_2^2}} \int_{-\infty}^{\infty} e^{-\frac{(v_2-t)^2}{2\sigma_2^2}} \left[c_0 v_2^2 + c_1 + c_2 v_2^2 + c_3 + \phi(v_2)\right] dv_2. \tag{E.31}$$

Since $\phi(\cdot)$ is continuous and A_2 compact,

$$\int_{A_2} e^{-\frac{(v_2-t)^2}{2\sigma_2^2}} \left[c_0 v_2^2 + c_1 + c_2 v_2^2 + c_3 + \phi(v_2)\right] d v_2 < \infty. \tag{E.32}$$

Let $c_6 = c_0 + c_2 + c_4$ and $c_7 = c_1 + c_3 + c_5$. Then by (E.28),

$$\int_{\mathbb{R}\setminus A_2} e^{-\frac{(v_2-t)^2}{2\sigma_2^2}} \left[c_0 v_2^2 + c_1 + c_2 v_2^2 + c_3 + \phi(v_2)\right] d v_2 \leq \int_{\mathbb{R}\setminus A_2} e^{-\frac{(v_2-t)^2}{2\sigma_2^2}} \left[c_6 v_2^2 + c_7\right] d v_2 \tag{E.33}$$

$$< \infty. \tag{E.34}$$

Therefore, $\psi(w_1, w_2)$ is finite for every $w_1 \in \partial\Delta$ and

$$\oint_{\partial\Delta} \psi(w_1, w_2) d w_1 < \infty. \tag{E.35}$$

This justifies the use of Fubini's Theorem in (E.12) so that

$$\oint_{\partial\Delta} h_{\boldsymbol{U}}(\boldsymbol{w}; F)\, dw_1 = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\oint_{\partial\Delta} p_{U_1}(v_1 - w_1) p_{U_2}(v_2 - w_2) \ln p(\boldsymbol{Q}(v_1, v_2)^T; F) d w_1 d v_1 d v_2 \tag{E.36}$$

$$= 0, \tag{E.37}$$

where equality to 0 is due to the analycity of $p_{U_1}(v_1 - w_1)$.

By Morera's Theorem, $h_{\boldsymbol{U}}((w_1, w_2); F)$ is an entire function of w_1 for fixed w_2. Applying similar logic to the above, $h_{\boldsymbol{U}}((w_1, w_2); F)$ is also an entire function of w_2 for fixed w_1. Therefore, by Hartog's Theorem, $h_{\boldsymbol{U}}((w_1, w_2); F)$ is entire on $\mathbb{C}^2$ [19].

□