Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Departamento de Telemática

Uma Análise da Influência da Curvatura do Espaço em Sistemas de Comunicações

Autor: Rodrigo Gusmão Cavalcante

Orientador: Prof. Dr. Reginaldo Palazzo Jr.

Banca Examinadora:

Prof. Dr. Reginaldo Palazzo Jr.	**DT-FEEC-UNICAMP**
Prof. Dr. Henrique Lazari	**UNESP/RIO CLARO**
Prof. Dr. Renato da Rocha Lopes	**DECOM-FEEC-UNICAMP**
Prof. Dr. Romis Ribeiro de Faissol Attux	**DCA-FEEC-UNICAMP**
Prof. Dr. Weiler Alves Finamore	**CETUC/PUC-RIO**

Tese apresentada na Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica.

Campinas - SP
Maio de 2008

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA
DA ÁREA DE ENGENHARIA E ARQUITETURA- BAE - UNICAMP

C314a

Cavalcante, Rodrigo Gusmão
Uma análise da influência da curvatura do espaço em sistemas de comunicações / Rodrigo Gusmão Cavalcante. – Campinas, SP: [s.n.], 2008.

Orientador: Reginaldo Palazzo Júnior.
Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Curvatura. 2. Materiais óticos. 3. Modulação (Eletrônica). 4. Teoria dos sinais (Telecomunicações). 5. Geometria riemanniana. 6. Superfícies de curvatura constante. I. Palazzo Júnior, Reginaldo. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica de Computação. III. Título

Título em Inglês:	An analisys of the influence of the space curvature in communication systems
Palavras-chave em Inglês:	Communication system, Signal constellation, Twisted modulation, Optical medium, Curvature, Riemannian manifold
Área de concentração:	Telecomunicações e Telemática
Titulação:	Doutor em Engenharia Elétrica
Banca Examinadora:	Henrique Lazari, Renato da Rocha Lopes, Romis Ribeiro de Faissol Attux, Weiler Alves Finamore.
Data da defesa:	09/05/2008
Programa de Pós-Graduação:	Engenharia Elétrica

Uma Análise da Influência da Curvatura do Espaço em Sistemas de Comunicações

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Rodrigo Gusmão Cavalcante e aprovada pela banca examinadora.

Campinas, 9 de maio de 2008.

Prof. Dr. Reginaldo Palazzo Júnior (Presidente): __________

Prof. Dr. Henrique Lazari: __________

Prof. Dr. Weiler Alves Finamore: __________

Prof. Dr. Romis Ribeiro de Faissol Attux: __________

Prof. Dr. Renato da Rocha Lopes: __________

Tese apresentada na Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica.

Campinas - SP
Maio de 2008

Resumo

Em geral, o espaço Euclidiano é utilizado no projeto e na análise de desempenho da maior parte dos sistemas de comunicações atuais. Nesta tese, verificamos que o modelo de um sistema de comunicação não necessariamente está restrito ao espaço Euclidiano, mas sim a uma variedade Riemanniana. Com isso, os sistemas de comunicações podem ser analisados em um contexto mais geral, no qual constatamos que a curvatura do espaço influencia em seus desempenhos. Como exemplo, estudamos a curvatura de meios ópticos e propomos novos perfis de guias de ondas, fibras ópticas e lentes de interesse prático. Além disso, caracterizamos a curvatura de modulações não-lineares (*twisted*) e verificamos que o valor máximo permitido para a energia média do ruído está relacionada ao valor da curvatura da modulação. Neste contexto, as modulações associadas a superfícies mínimas apresentaram bons desempenhos, pois tais modulações são pontos críticos do erro quadrático médio. Mostramos também que o espaço de sinais possui métrica induzida da superfície associada à modulação. Com isso, foi possível demonstrar que os espaços de sinais com curvatura negativa são os que apresentam melhor desempenho segundo a probabilidade média de erro. Dessa forma, alguns exemplos de constelações de sinais geometricamente uniformes foram construídos e analisados em variedades Riemannianas. Finalizamos este trabalho notando que na maioria das vezes que o espaço hiperbólico é utilizado nos blocos de um sistema de comunicações, o desempenho desse sistema tende a se aproximar do ponto ótimo de operação.

Palavras-chave: sistema de comunicações, curvatura, meios ópticos, modulação não-linear, constelação de sinais, variedade Riemanniana.

Abstract

In general, the Euclidian space is used in the design and performance analysis in most of the current communication systems. In this thesis, we note that the model of a communication system is not necessarily restricted to the Euclidian space, more precisely, the model can be linked to Riemannian manifolds. Thus, the communication systems could be analyed in a broader context, in which the curvature of space influence on their performance. As an example, we studied the curvature of optical medium and propose new profiles of waveguides, optical fibers and lenses of practical interest. Moreover, we have characterized the curvature of twisted modulations and found that the maximum value allowed for the average energy of noise is related to the value of the curvature of the modulation. In this context, the modulation associated with minimal surfaces showed good performance, because these modulations are critical points of minimum the mean-square error. We show that the signal space has induced metric associated with surface of the modulation. Thus, we have shown that the signal space with negative curvature is the space where the average error probability decrease a function of the curvature. Thus, some examples of geometrically uniform signal constellations were constructed and analyzed on Riemannian manifolds. Finally we note that most of the time that hyperbolic space is considered in blocks of a communication system, then the performance of this system tends to be closer to the optimum point of operation.

Keywords: communication system, curvature, optical medium, twisted modulation, signal constellation, Riemannian manifold.

À minha família por estar sempre me apoiando e amando.

Dedico

Agradecimentos

Muitas pessoas participaram direta e indiretamente no desenvolvimento deste trabalho. Portanto, peço desculpas àquelas cujos nomes tenham sido momentaneamente esquecidos.

Durante todo o desenvolvimento deste trabalho, pude contar com a atenção, a compreensão, a disponibilidade, a confiança e, principalmente, a paciência do Prof. Dr. Reginaldo Palazzo Jr., que foi muito mais que um orientador, sempre se dedicando e me ajudando a concluir essa importante etapa de minha vida. Meu muito obrigado.

À minha família pelo incentivo e paciência durante o desenvolvimento deste trabalho. Se não fosse pelo apoio de vocês essa jornada em minha vida não seria realidade.

À amiga Wanessa Gazzoni, por ter sido mais que uma colega de laboratório, me ajudando a superar os momentos difíceis que passei durante este período, pela sugestões e revisões realizadas nos textos deste trabalho e, principalmente, por ter me ajudado tanto no doutorado enquanto estive na Bahia. Obrigado por ser a irmã que não tive.

Ao companheirismo dos colegas do laboratório de Telemática (DT) e de Teoria da Informação Aplicada (LTIA) com os quais convivi durante este programa de doutorado na FEEC, fazendo com que este período em Campinas fosse mais agradável.

Aos membros da banca examinadora pela disponibilidade e atenção dispensada ao trabalho, bem como por suas valiosas sugestões.

À comunidade de *Software* Livre, pois este trabalho foi realizado usando programas livres, desde o sistema operacional aos programas de edição e de simulação.

Aos professores e colegas de trabalho do CEFET-BA pelo incentivo, principalmente, à professera Angélica pela força no final deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, pelo apoido financeiro em forma de bolsa de doutorado concedida durante a realização deste trabalho.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Este capítulo apresenta de maneira sucinta uma descrição das principais razões que motivaram a realização deste trabalho. Além disso, para um melhor entendimento deste trabalho apresentamos alguns conceitos básicos sobre geometria diferencial, geometria Riemanniana, óptica geométrica, modulação não-linear (*twisted*) e constelações de sinais. Por fim, realizamos uma descrição geral dos principais resultados obtidos em cada um dos demais capítulos da tese.

1.1 Motivações

Os principais objetivos a serem alcançados na proposta de novos sistemas de comunicações são menor complexidade e melhor desempenho que o apresentado pelos sistemas conhecidos. Nesta direção, iremos considerar cada um dos blocos da Figura 1.1 como sendo constituído basicamente por um conjunto de "pontos" E_i, juntamente com uma métrica, d_i. Tal consideração torna possível a interpretação de cada bloco como um espaço métrico (E_i, d_i).

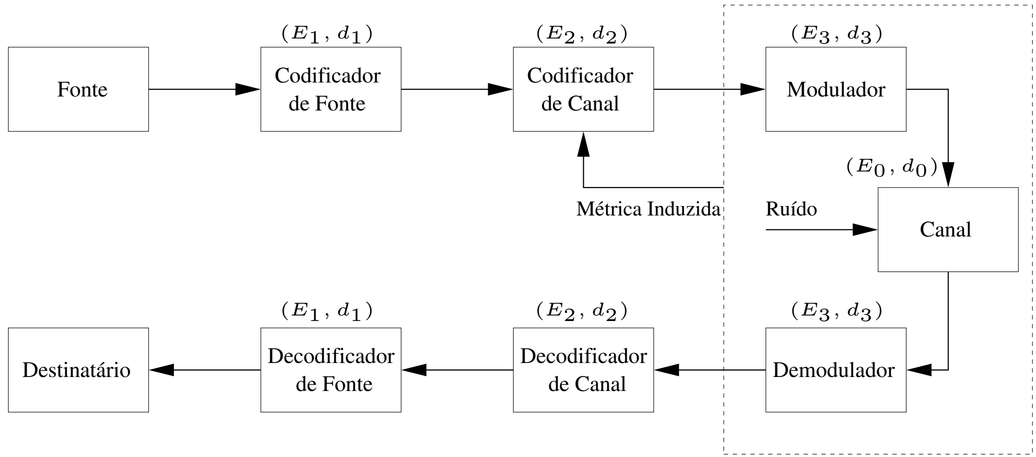


Figura 1.1: Modelo em diagrama de bolcos de um sistema de comunicações digitais.

O que se busca então, é determinar as características geométricas e algébricas associadas aos espaços métricos (E_i, d_i), bem como as propriedades e condições que deverão ser satisfeitas pelas transformações que irão conectá-los, de maneira que se consiga determinar o desempenho do sistema de comunicações sob a menor probabilidade de erro, maior taxa de transmissão, menor potência de transmissão, etc. Naturalmente surge então a pergunta que motivou a realização deste trabalho: *qual o desempenho dos sistemas de comunicações quando a estrutura geométrica associada aos espaços métricos* (E_i, d_i) *não é mais a Euclidiana?*

Neste contexto, as teses [6], [7], [8], [10], [12], [13], [22], [23] e [38] apresentaram esquemas de codificação, decodificação, modulação e demodulação de constelações de sinais em espaços hiperbólicos. Em [29], [30], [31], [32], [33], constelações de sinais foram propostas em variedades Riemannianas, em especial nos espaços de curvatura constante e sobre superfícies mínimas, apresentando bons desempenhos quando o espaço de sinais possui curvatura negativa. Em [34] uma métrica mais geral que as derivadas da métrica Euclidiana foi utilizada para descrever a construção de códigos de bloco lineares em novos espaços métricos. Neste caso, foi verificado que os blocos modulador, canal e demodulador, constituindo um canal discreto sem memória, induz de maneira natural uma métrica no espaço (E_2, d_2) associado ao codificador. Além disso, em [35] canais ópticos de curvatura constante foram propostos para a transmissão de informação, obtendo bons desempenhos quanto ao parâmetro da dispersão modal.

Neste trabalho concentramos nossos esforços, principalmente, na identificação de exemplos de canais de transmissão ou esquemas de modulação com características que pudessem justificar e validar os resultados apresentados no trabalhos citados anteriormente. Além disso, alguns desses exemplos foram utilizados para projetar novos sistemas de comunicações, cujos desempenhos foram medidos em função dos seguintes critérios: menor dispersão da informação transmitida em meios ópticos, menor distorção da informação pelo erro quadrático médio em modulações não-lineares e menor probabilidade média de erro de constelações de sinais projetadas em variedades Riemannianas. Deve-se ressaltar que as modulações não-lineares analisadas neste trabalho foram consideradas, inicialmente, para a transmissão de variáveis aleatórias contínuas, isto é, no contexto de um sistema de comunicações analógico, que pode ser representado pela Figura 1.1 sem os blocos de codificação e decodificação.

No desenvolvimento deste trabalho foi necessário utilizar alguns conceitos matemáticos e físicos não muito utilizados na construção e análise de sistemas de comunicações. Por esse motivo, a seguir vamos apresentar alguns desses conceitos. Contudo, não nos aprofundaremos em nossa descrição mais do que o necessário para o bom entendimento deste trabalho. Por isso, caso necessário, recomendamos a leitura das referências citadas em cada um dos tópicos para um estudo mais avançado. Além disso, alguns conceitos e parâmetros de desempenhos utilizados na construção e na análise dos sistemas de comunicações considerados neste trabalho serão apresentados.

1.2 Alguns Conceitos de Geometria Diferencial

A seguir usaremos [25] para apresentar alguns conceitos introdutórios de geometria diferencial. Neste caso, tem-se que uma curva diferenciável parametrizada $\alpha : (a, b) \in I \to \mathbb{R}^3$ é dita regular se $\alpha'(t) \neq 0$ para todo $t \in I$. O comprimento de arco de uma curva regular parametrizada é definido por

$$L = \int_a^b |\alpha'(t)|\, dt\,,$$

onde $|\cdot|$ é a norma de um vetor e

$$|\alpha'(t)| = \sqrt{(x'(t))^2 + (y'(t))^2 + (z'(t))^2}\,,$$

é o parâmetro de comprimento de arco da curva. Na análise de modulações não-lineares associadas a curvas, um caso muito interessante de parametrização é utilizada, a parametrização pelo comprimento de arco (p.p.c.a), no qual $|\alpha'(t)| = 1$. Para se obter tal parametrização, pode-se reparametrizar a curva $\alpha(t)$ pelo parâmetro s, de maneira que

$$s = \int_0^t |\alpha'(t)|\, dt\,.$$

A curvatura de uma curva plana é definida como

$$k = \frac{x'z'' - x''z'}{[(x')^2 + (z')^2]3/2}\,. \tag{1.1}$$

Uma superfície regular parametrizada, ou simplesmente, uma superfície parametrizada é uma aplicação $X : D \subset \mathbb{R}^2 \to \mathbb{R}^3$, cuja diferencial tem posto dois em todos os pontos do domínio D.

Consequentemente, os vetores $X_u = \frac{\partial X}{\partial u}$ e $X_v = \frac{\partial X}{\partial v}$ são linearmente independentes em todo ponto $p = (u, v)$ do domínio D e o espaço vetorial gerado por esses dois vetores é chamado plano tangente à superfície nesse ponto, denotado por T_pX. Além disso, o vetor normal unitário à superfície em p é dado por

$$N(p) = N(u, v) = \frac{X_u \wedge X_v}{|X_u \wedge X_v|}\,,$$

onde $\wedge$ denota o produto vetorial em $\mathbb{R}^3$. Note que $|X_u \wedge X_v| \neq 0$, pois os vetores X_u e X_v são linearmente independentes.

No contexto deste trabalho, um parâmetro que caracteriza a métrica dos espaços métricos associados aos blocos do sistema de comunicações da Figura 1.1 é a primeira forma fundamental de uma superfície parametrizada $X(u,v)$ dada por

$$E = (X_u \cdot X_u)\,, \quad F = (X_u \cdot X_v) \quad \text{e} \quad G = (X_v \cdot X_v)\;. \tag{1.2}$$

Uma parametrização de interesse neste trabalho ocorre quando uma superfície é parametrizada por parâmetros isotérmicos, isto é, $E = G$ e $F = 0$.

A primeira forma fundamental de uma superfície permite o cálculo do comprimento de um segmento de uma curva regular $\alpha : I \subset \mathbb{R} \to \mathbb{R}^3$, parametrizada por $\alpha(t) = X(u(t), v(t))$, pela seguinte expressão

$$L = \int_a^b \left[(u')^2 E + 2u'v'F + (v')^2 G\right]^{(1/2)} dt\,,$$

onde $u' = \frac{du(t)}{dt}$ e $v' = \frac{dv(t)}{dt}$. Além disso, a área de uma superfície regular pode ser calculada por

$$\mathcal{A} = \iint_D \sqrt{EG - F^2}\, du\, dv\,.$$

Um outro conceito importante é a segunda forma fundamental de uma superfície parametrizada $X(u,v)$, dada por

$$e = \left(X_{uu} \cdot \overrightarrow{N}\right), \quad f = \left(X_{uv} \cdot \overrightarrow{N}\right) \quad \text{e} \quad g = \left(X_{uu} \cdot \overrightarrow{N}\right), \tag{1.3}$$

onde $\overrightarrow{N}$ é o vetor normal à superfície $X(u,v)$.

Na prática existem dois conceitos de curvatura de superfícies, a curvatura média e a curvatura Gaussiana. Neste trabalho, essas curvaturas demonstraram ser de grande interesse para caracterizar o desempenho dos sistemas de comunicações considerados. Neste caso, a curvatura média de uma superfície parametrizada $X(u,v)$ é dada por

$$H = \frac{1}{2}\frac{eG - 2fF + gE}{EG - F^2}\,, \tag{1.4}$$

e a curvatura Gaussiana de uma superfície parametrizada $X(u,v)$ é dada por

$$K = \frac{eg - f^2}{EG - F^2}\,. \tag{1.5}$$

Apesar da expressão da curvatura Gaussiana (1.5) depender da segunda forma fundamental de uma superfície, pode-se mostrar que tal curvatura pode ser calculada apenas em função da primeira forma fundamental da superfície. Tal fato será usado para caracterizarmos alguns sistemas de comunicações apenas por sua métrica e curvatura.

1.2.1 Superfícies mínimas

Exemplos de sistemas de comunicações com bons desempenhos foram associados a superfícies mínimas em todos os capítulos de resultados deste trabalho e, por isso, uma breve descrição de tais superfícies faz-se necessária. Geometricamente, as superfícies mínimas são caracterizadas por minimizarem localmente a área da região determinada por uma curva fechada no espaço. Matematicamente, as superfícies mínimas devem possuir curvatura média (1.4) nula em todos seus pontos. De acordo com [17], exemplos de superfícies mínimas parametrizadas isotermicamente podem ser obtidas com o auxílio da representação de Weierstrass. Nesta representação, uma superfície mínima regular em domínios simplesmente conexos, $X : U \subset \mathbb{R}^2 \to \mathbb{R}^3$, é dada por

$$X(u,v) = Re \int_{z_0}^{z} \frac{1}{2} \left[f(z)(1 - g^2(z)),\ if(z)(1 + g^2(z)),\ f(z)g(z) \right] dz\,, \qquad z_0,\ z \in U\,, \tag{1.6}$$

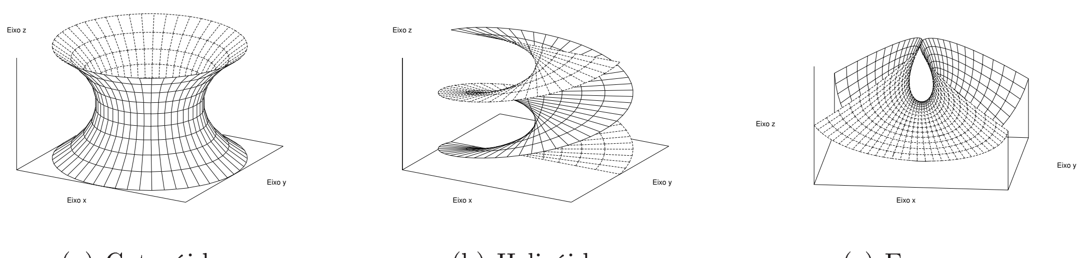
onde $z = u + iv$ e as funções $f(z)$ e $g(z)$ são funções holomorfas que caracterizam a superfície. Além disso, a primeira forma fundamental dessa superfície é dada por

$$E = G = \frac{1}{4}|f(z)|^2(1 + |g(z)|^2)^2 \quad \text{e} \quad F = 0\,. \tag{1.7}$$

Um outro parâmetro importante aos interesses deste trabalho é a curvatura Gaussiana, que de acordo com a representação de Weierstrass, pode ser expressa da seguinte maneira

$$K(u,v) = -\left[\frac{4|g'(z)|}{|f(z)|^2(1 + |g(z)|^2)^2} \right]^2. \tag{1.8}$$

A representação de Weierstrass permite obter uma infinidade de exemplos de superfícies mínimas. Tais superfícies podem ser caracterizadas pelas funções $f(z)$ e $g(z)$, como ilustra a Figura 1.2, contendo as superfícies mínimas catenóide ($f(z) = -e^{-z}$ e $g(z) = -e^{z}$), helicóide ($f(z) = e^{-z}$ e $g(z) = -ie^{z}$) e Enneper ($f(z) = 1$ e $g(z) = z$).



(a) Catenóide. (b) Helicóide. (c) Enneper.

Figura 1.2: Exemplo de superfícies mínimas.

1.3 Alguns Conceitos de Geometria Riemanniana

Nesta seção serão apresentados, de forma sucinta, alguns conceitos de geometria Riemanniana que serão aplicados neste e nos próximos capítulos deste trabalho. Para uma abordagem completa sobre essa importante ferramenta matemática recomendamos a leitura da referência [27].

A geometria Riemanniana surgiu com a necessidade de estender os métodos da geometria diferencial a espaços mais gerais que o $\mathbb{R}^n$. O primeiro exemplo de variedade Riemanniana acessível a nossa experiência é uma superfície regular no $\mathbb{R}^3$, $\mathbf{X} : U \subset \mathbb{R}^2 \to \mathbb{R}^3$. Isto porque, a idéia natural de uma superfície é a de um conjunto bidimensional (U) ao qual se possa aplicar o Cálculo Diferencial do $\mathbb{R}^2$.

Dessa forma, uma variedade diferenciável de dimensão n é um conjunto M^n munido de uma família de aplicações biunívocas, $\mathbf{x}_\alpha : U_\alpha \subset \mathbb{R}^n \to M^n$ de abertos U_α de $\mathbb{R}^n$ em M^n, tais que:

1. $\cup_\alpha \mathbf{x}_\alpha(U_\alpha) = M^n$.

2. Para todo par α e β, com $\mathbf{x}_\alpha(U_\alpha) \cap \mathbf{x}_\beta(U_\beta) = W \neq \phi$, os conjuntos $\mathbf{x}_\alpha^{-1}(W)$ e $\mathbf{x}_\beta^{-1}(W)$ são abertos em $\mathbb{R}^n$ e as aplicações $\mathbf{x}_\beta^{-1} \circ \mathbf{x}_\alpha$ são diferenciáveis.

Um outro conceito igualmente importante aos interesses deste trabalho é o de métrica Riemanniana, que permite medir, por exemplo, comprimentos de curvas, áreas, ângulos, curvaturas, entre outros. A métrica Riemanniana de uma variedade diferenciável é uma forma bilinear simétrica positiva definida como $g_{ij} = \left(\frac{\partial \mathbf{x}}{\partial x_i} \cdot \frac{\partial \mathbf{x}}{\partial x_j}\right)$, no sistema de coordenadas $\mathbf{x} : U \subset \mathbb{R}^n \to M^n$, onde $(\ \cdot\ )$ representa o produto escalar de dois vetores.

Em função da métrica Riemanniana pode-se calcular o comprimento de um segmento da curva $\alpha : I \subset \mathbb{R} \to M^n$, parametrizada por $\alpha(t) = [x_1(t), \ldots, x_n(t)]$, $t \in [a, b]$, da seguinte maneira

$$L = \int_a^b \left(\sum_{i=1}^{n} \sum_{j=1}^{n} x_i' x_j' g_{ij} \right)^{1/2} dt \,, \tag{1.9}$$

onde $x_i' = \frac{\partial x_i}{\partial t}$, para $i = 1, \ldots, n$.

Esta noção de comprimento permite definir em M^n certas curvas especiais, chamadas geodésicas, que possuem a propriedade de minimizar localmente a distância entre quaisquer dois pontos suficientemente próximos. Tais curvas comportam-se em várias situações como se fossem "retas" em M^n e, em [24], estão associadas às trajetórias das ondas eletromagnéticas em meios ópticos. Neste caso, uma curva regular $\gamma : I \subset \mathbb{R} \to M^n$, parametrizada por $\gamma(t) = [x_1(t), \ldots, x_n(t)]$ é uma geodésica em M^n se, e somente se, satisfaz o seguinte sistema de equações diferenciais parciais de segunda ordem

$$x_k'' + \sum_{i,j} \Gamma_{ij}^k x_i' x_j' = 0\,, \quad k = 1, \ldots, n\,, \tag{1.10}$$

onde $x_k'' = \frac{\partial^2 x_k}{\partial t^2}$ e Γ_{ij}^k representa os símbolos de Christoffel de M^n, que dependem apenas da métrica Riemanniana, e podem ser obtidos pela seguinte expressão

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{l=1}^{n} \left[\frac{\partial\, g_{jl}}{\partial x_i} + \frac{\partial\, g_{li}}{\partial x_j} - \frac{\partial\, g_{ij}}{\partial x_k} \right] g^{lk}, \tag{1.11}$$

onde g^{lk} é o elemento da linha l e coluna k da matriz inversa de $G = [g_{ij}]$. Geometricamente, os símbolos de Christoffel estão associados ao conceito de derivada covariante, uma noção da geometria "intrínseca" de superfícies, e, por essa razão, podem ser utilizados para descrever as idéias de geodésicas e de curvatura em variedades Riemannianas, como será realizado a seguir.

A noção de curvatura em uma variedade Riemanniana foi introduzida por Riemann de uma maneira bastante geométrica e de modo que fosse uma generalização natural da curvatura Gaussiana de superfícies no $\mathbb{R}^3$. A curvatura de Riemman pode ser determinada utilizando apenas a métrica de M^n e, intuitivamente, mede o quanto uma variedade Riemanniana deixa de ser Euclidiana. Matematicamente, essa curvatura de Riemann está associada ao tensor curvatura R_{ijks}, que será definido a seguir e utilizado para calcular a curvatura seccional, a curvatura de Ricci e a curvatura escalar de meios ópticos

$$R_{ijkl} = \sum_{m=1}^{n} R_{ijk}^m g_{ml}\,, \tag{1.12}$$

onde R_{ijk}^m é o operador curvatura, que pode ser expresso em termos dos símbolos de Christoffel, equação (1.11), da seguinte forma

$$R_{ijk}^l = \sum_{m=1}^{n} \Gamma_{jk}^m \Gamma_{im}^l - \sum_{m=1}^{n} \Gamma_{ik}^m \Gamma_{jm}^l + \frac{\partial\, \Gamma_{jk}^l}{\partial x_i} - \frac{\partial\, \Gamma_{ik}^l}{\partial x_j}. \tag{1.13}$$

O conceito de curvatura de maior interesse neste trabalho é o de curvatura seccional. Tal curvatura é calculada para um subespaço vetorial bidimensional ν contido em T_pM^n, o espaço tangente de M^n no ponto p. Neste caso, quando o subespaço vetorial ν possuir o conjunto de vetores $\{\frac{\partial \mathbf{x}}{\partial x_i}, \frac{\partial \mathbf{x}}{\partial x_j}\}$ como base, então a curvatura seccional desse subespaço vetorial bidimensional é dada por

$$K_{ij} = -\frac{R_{ijij}}{g_{ii}g_{jj} - g_{ij}^2}. \tag{1.14}$$

Sabemos que, dentre as variedades Riemannianas, aquelas com curvatura seccional constante são as mais simples e possuem um número suficientemente grande de isometrias locais. Tais variedades são conhecidas como espaços homogêneos. No Capítulo

3, foi observado que os espaços ópticos com curvatura seccional constante também demonstraram ser os de maior interesse prático, tanto para o projeto de guias de ondas quanto para o projeto de lentes ópticas.

Além da curvatura seccional, que é uma generalização da curvatura Gaussiana de superfícies, alguns outros conceitos de curvaturas também serão consideradas neste capítulo, como o tensor curvatura de Ricci, dado por

$$R_{ij} = \sum_{k=1}^{n} R^k_{kij}\,, \tag{1.15}$$

e a curvatura escalar,

$$S = \sum_{i,j=1}^{n} g^{ij} R_{ij}\,, \tag{1.16}$$

que também foi considerada anteriormente em [11] e [41] para meios ópticos isotrópicos e anisotrópicos, respectivamente.

Nas próximas duas subseções, as expressões para o cálculo das geodésicas (1.10), do comprimento de um segmento de curva (1.9) e das curvaturas de uma variedade Riemanniana serão novamente obtidas para variedades Riemannianas com parâmetros ortogonais (tensor métrica diagonal) e para variedades Riemannianas com parâmetros isotérmicos (tensor métrica diagonal com elementos iguais). Esses dois casos particulares de variedades Riemannianas serão considerados, pois foram associados a meios ópticos de interesse prático.

1.3.1 Variedades Riemannianas com parâmetros ortogonais

O interesse por variedades Riemannianas com parâmetros ortogonais provém do fato delas terem sido associadas a espaços anisotrópicos em [41]. Tais variedades Riemannianas possuem métrica

$$\begin{cases} g_{ij} = \delta_{ij} g_i \\ g^{ij} = \delta_{ij}/g_i \end{cases}, \quad \text{onde} \quad \delta_{ij} = \begin{cases} 1\,, & \text{se } i = j \\ 0\,, & \text{se } i \neq j \end{cases}. \tag{1.17}$$

Dessa forma, usando a métrica (1.17) em (1.9), tem-se que o comprimento de um segmento de curva em M^n é dado por

$$L = \int_a^b \left(\sum_{i=1}^{n} g_i (x_i')^2 \right)^{1/2} dt\,. \tag{1.18}$$

Além disso, usando a métrica (1.17) em (1.11), pode-se verificar que se os três índices são distintos, então $\Gamma^k_{ij} = 0$, e se pelo menos dois de seus índices são iguais, então

$$\Gamma^i_{ij} = \Gamma^i_{ji} = \frac{g_{i,j}}{2g_i}\,, \quad \Gamma^j_{ii} = -\frac{g_{i,j}}{2g_j}\,, \quad \Gamma^i_{ii} = \frac{g_{i,i}}{2g_i}\,, \tag{1.19}$$

onde $g_{i,j} = \frac{\partial g_i}{\partial x_j}$.

Com isso, pode-se substituir (1.19) em (1.10) para obter que as geodésicas de M^n satisfazem

$$x_k'' + \frac{1}{2g_k}\sum_{i=1}^{n}(2g_{k,i}x_k' - g_{i,k}x_i')\,x_i' = 0\,, \quad k = 1,\ldots,n\,. \tag{1.20}$$

De (1.12) e (1.19) pode-se determinar que o tensor curvatura de uma variedade com métrica (1.17) é diferente de zero apenas para R_{ijij} e R_{ijji}, com $i \neq j$, mais precisamente, $R_{ijij} = -R_{ijji}$ e

$$\begin{aligned} R_{ijij} &= \sum_{l=1}^{n} R_{iji}^{l} g_{lj} = R_{iji}^{j} g_j = \left[\sum_{l=1}^{n}\Gamma_{ji}^{l}\Gamma_{il}^{j} - \sum_{l=1}^{n}\Gamma_{ii}^{l}\Gamma_{jl}^{j} - \frac{\partial}{\partial x_i}\Gamma_{ji}^{j} + \frac{\partial}{\partial x_j}\Gamma_{ii}^{j}\right] g_j \\ &= -\frac{1}{2}\left[\frac{g_{i,i}g_{j,i}}{g_i} + \frac{g_{i,j}g_{j,j}}{g_j} + \frac{(g_{i,j})^2}{2g_i} + \frac{(g_{j,i})^2}{2g_j} - g_{i,jj} - g_{j,ii} - \sum_{l=1}^{n}\frac{g_{i,l}g_{j,l}}{2g_l}\right], \end{aligned} \tag{1.21}$$

onde $g_{j,ii} = \frac{\partial^2 g_j}{\partial x_i^2}$.

Logo, com o auxílio de (1.14) e (1.21), pode-se verificar que a curvatura seccional é igual a

$$K_{ij} = -\frac{R_{ijij}}{g_i g_j}\,, \tag{1.22}$$

e, de (1.15) e (1.19), que a curvatura de Ricci, R_{ij}, é igual a zero se i for diferente de j, e igual a

$$R_{ij} = R_{jj} = -\sum_{k=1}^{n} R_{kjkj} g^k = g_j \sum_{k=1}^{n} K_{kj}\,, \tag{1.23}$$

caso contrário ($i = j$). Com isso, a curvatura escalar (1.16) é dada por

$$S = -\sum_{i,j=1}^{n} R_{ijij} g^i g^j = \sum_{i,j=1}^{n} K_{ij}\,. \tag{1.24}$$

1.3.2 Variedades Riemannianas com parâmetros isotérmicos

As variedades Riemannianas com parâmetros isotérmicos serão consideradas a seguir, pois foram associadas a meios ópticos isotrópicos em [24] e em [11]. Tais variedades Riemannianas possuem métrica com a seguinte característica

$$g_{ij} = \delta_{ij} g\,, \quad g^{ij} = \frac{\delta_{ij}}{g}\,. \tag{1.25}$$

Uma variedade M^n com tal métrica possui curvas cujo comprimento, (1.18), é dado por

$$L = \int_a^b \left(g\sum_{i=1}^{n}(x_i')^2\right)^{1/2} dt\,. \tag{1.26}$$

Além disso, pode-se substituir (1.25) em (1.20) para obter que as geodésicas de M^n, com tal métrica, satisfazem

$$x_k'' + \frac{1}{2g}\sum_{i=1}^{n}\left(2g_{x_i}x_k' - g_{x_k}x_i'\right)x_i' = 0\,, \tag{1.27}$$

onde $k = 1, \ldots, n$ e $g_{x_i} = \frac{\partial g}{\partial x_i}$.

O tensor curvatura de uma variedade Riemanniana com a métrica (1.25) pode ser obtido com o auxílio de (1.21) como sendo igual a

$$R_{ijij} = -\frac{1}{2}\left[\frac{3(g_{x_i})^2}{2g} + \frac{3(g_{x_j})^2}{2g} - g_{x_ix_i} - g_{x_jx_j} - \frac{1}{2g}\sum_{l=1}^{n}(g_{x_l})^2\right], \tag{1.28}$$

onde $g_{x_ix_i} = \frac{\partial^2 g}{\partial x_i^2}$.

Com isso, pode-se obter, com o auxílio de (1.22) e (1.28), que a curvatura seccional seja calculada por

$$K_{ij} = -\frac{R_{ijij}}{g^2}\,. \tag{1.29}$$

A curvatura de Ricci indicada em (1.23) pode ser reescrita nessa métrica da seguinte forma: $R_{ij} = 0$, se $i \neq j$, e caso contrário é dada por

$$R_{jj} = \frac{1}{2g}\left[\frac{4-n}{2g}(\nabla g)^2 - \nabla^2 g\right] + \frac{n-2}{2g}\left[\frac{3}{2g}(g_{x_j})^2 - g_{x_jx_j}\right], \tag{1.30}$$

onde $\nabla = \left[\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right]$ é um operador diferencial conhecido como nabla.

Dessa forma, a expressão da curvatura escalar (1.24) pode ser simplificada para

$$S = \frac{n-1}{g^2}\left[\frac{6-n}{4g}(\nabla g)^2 - \nabla^2 g\right]. \tag{1.31}$$

1.4 Fundamentos de Óptica Geométrica

Alguns conceitos sobre óptica geométrica são necessários para o desenvolvimento deste trabalho. Para tanto, usamos [21] como referência.

Inicialmente, deve-se considerar que as dimensões dos instrumentos ópticos tratados neste trabalho sejam algumas vezes maior do que o comprimento de onda (λ) das ondas eletromagnéticas para os quais esses instrumentos foram projetados. Tal suposição é necessária para que o fenômeno da difração possa ser desprezado e os conceitos da óptica geométrica possam ser aplicados de maneira correta. Além disso, considere que a trajetória de uma onda eletromagnética em um meio óptico isotrópico seja descrita por uma curva regular parametrizada $\mathbf{r}(s) = [x(s), y(s), z(s)]$, onde o parâmetro s representa o comprimento Euclidiano percorrido por sua frente de onda ao longo do meio. Dessa forma, caso o índice de refração desse meio seja variável com a posição, $n = n(x, y, z)$, então a trajetória $\mathbf{r}(s)$ deve satisfazer uma equação diferencial de segunda ordem conhecida como Eikonal, que, no caso particular, é dada por

$$\frac{\partial}{\partial s}\left(n\frac{\partial \mathbf{r}}{\partial s}\right) = \nabla n\,, \quad \text{quando} \quad (x')^2 + (y')^2 + (z')^2 = 1\,, \tag{1.32}$$

onde $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ é o operador diferencial nabla em coordenadas cartesianas. Deve-se ressaltar a necessidade de que o vetor tangente de $\mathbf{r}(s)$ seja realmente unitário, pois, caso contrário a equação Eikonal não descreverá as trajetórias das ondas em meios ópticos isotrópicos. Além disso, a equação Eikonal pode ser derivada do princípio de Fermat, que afirma que o caminho percorrido pela onda no meio é o de menor energia ou de menor tempo de percurso.

Um outro conceito de muita importância para a análise dos instrumentos ópticos é o conceito de comprimento óptico, representado por L_o. Neste caso, o comprimento óptico de uma frente de onda passando por dois pontos quaisquer $P_1 = \mathbf{r}(s_1)$ e $P_2 = \mathbf{r}(s_2)$, é dado por

$$L_o = \int_{s_1}^{s_2} n\, ds\,. \tag{1.33}$$

Com isso, desde que a densidade de energia da frente de onda seja propagada a uma velocidade $v = c/n$ ao longo do meio, temos que

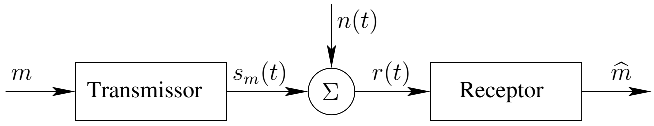
$$n\, ds = \frac{c}{v}\, ds = c\, dt\,,$$

onde dt é o tempo necessário para a energia percorrer uma distância ds ao longo do meio. Em outras palavras, o comprimento ótico L_0 é igual ao produto da velocidade da luz no vácuo pelo tempo necessário para a luz viajar de P_1 a P_2.

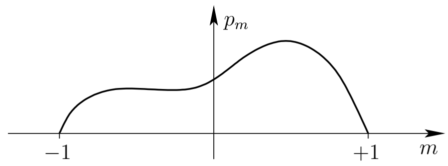
1.5 Introdução à Modulação Não-Linear (*Twisted*)

Em [18], as técnicas de modulação foram associadas a curvas no $\mathbb{R}^N$. Tal interpretação geométrica possibilita o projeto e análise de desempenho das modulações usando algumas ferramentas da geometria diferencial, ou, mais geralmente, conceitos da geometria Riemanniana.

Para entender essa interpretação geométrica, considere um sistema de comunicações cujo objetivo é transmitir uma mensagem representada por uma única variável aleatória contínua, m. Tal sistema é apresentado de maneira simplificada na Figura 1.3(a), onde na entrada do transmissor está presente a variável aleatória m, com função densidade de probabilidade p_m, veja Figura 1.3(b), e em sua saída tem-se o sinal modulado $s_m(t)$ que será transmitido através do canal com ruído aditivo, $n(t)$. Após a observação do sinal recebido, $r(t)$, o receptor estima uma saída $\widehat{m}$, de m.



(a) Diagrama de blocos de um sistema de comunicações.



(b) Exemplo de função densidade de probabilidade *a priori*, p_m.

Figura 1.3: Exemplo de sistema de comunicações para a transmissão de uma variável aleatória contínua.

No sistema de comunicações considerado na Figura 1.3, a técnica de modulação é realizada no bloco transmissor. Nesse bloco, a variável aleatória m é modulada no sinal $s_m(t)$, em geral, para uma nova faixa de freqüência.

A interpretação geométrica para a técnica de modulação proposta em [18] é baseada na possibilidade do sinal modulado $s_m(t)$ ser decomposto em uma base de sinais ortogonais de energia unitária, $\varphi_i(t)$, $i = 1, \ldots, N$, da seguinte forma

$$s_m(t) = s_1(m)\varphi_1(t) + s_2(m)\varphi_2(t) + \cdots + s_N(m)\varphi_N(t), \tag{1.34}$$

onde os sinais $\varphi_i(t)$, no caso particular das modulações, são denominados portadoras e devem satisfazer

$$\int_{-\infty}^{\infty} \varphi_i(t)\,\varphi_j(t)\,dt = \delta_{ij}\,, \quad \forall i, j = \{1, \ldots, N\}, \tag{1.35}$$

para que os sinais sejam ortogonais e de energia unitária, pois $\delta_{ij} = 1$ se $i = j$ e $\delta_{ij} = 0$ se $i \neq j$.

Como as portadoras $\varphi_i(t)$ são sinais ortogonais que possuem energia unitária, então elas podem ser interpretadas como um conjunto de vetores ortonormais $\{\boldsymbol{\varphi}_i\}$, gerando uma base para o $\mathbb{R}^N$. Com isso, o sinal modulado $s_m(t)$ pode ser decomposto nessa base vetorial da seguinte maneira

$$\begin{aligned} \mathbf{s}_m &= s_1(m)\boldsymbol{\varphi}_1 + s_2(m)\boldsymbol{\varphi}_2 + \cdots + s_N(m)\boldsymbol{\varphi}_N \\ &= [s_1(m), s_2(m), \ldots, s_N(m)]. \end{aligned} \tag{1.36}$$

Isto é, o sinal modulado $s_m(t)$, representado vetorialmente por $\mathbf{s}_m$, pode ser interpretado como uma curva parametrizada $\mathbf{s}_m : I \in \mathbb{R} \to \mathbb{R}^N$, $m \in I$, como sugere a interpretação geométrica proposta em [18].

Com o objetivo de analisar e medir o desempenho de sistemas de comunicações como o da Figura 1.3, note que após o processo de recepção, a probabilidade de $\widehat{m}$ ser igual a m é, em geral, zero. Isto ocorre porque m é uma variável aleatória contínua e o ruído por menor que seja, perturba o sistema produzindo pequenas diferenças entre $\widehat{m}$

e m. Por essa razão, o desempenho do sistema de comunicações não pode ser medido pela probabilidade de erro de transmitir m e receber $\hat{m}$, como ocorreria, por exemplo, em sistemas de comunicações que usam constelações de sinais, caso em que a variável aleatória m é discreta. Dessa forma, um bom parâmetro para medir o desempenho do sistema de comunicações considerado é o erro quadrático médio, definido como

$$\overline{\epsilon^2} \triangleq \mathrm{E}\left[(m-\widehat{m})^2\right] = \overline{(m-\widehat{m})^2}\,, \tag{1.37}$$

onde $\mathrm{E}[\,\cdot\,]$ é o operador esperança, medido conjuntamente para todas as entradas permitidas (m) e todos os possíveis valores estimados ($\hat{m}$) pelo receptor após a ação do ruído.

Além de ser útil para medir o desempenho de sistemas de comunicações, pode-se usar o valor do $\overline{\epsilon^2}$ para garantir que o receptor do sistema de comunicações da Figura 1.3 seja projetado de maneira ótima segundo algum critério. Isto é, o valor da estimativa $\hat{m}$ deve ser determinado de maneira que o desempenho do sistema seja o melhor possível, no caso particular, o que minimiza o erro quadrático médio. Neste caso, segundo [18], pode-se mostrar que $\widehat{m}$ deve ser igual ao valor médio de m dado que seja recebido $\mathbf{r}$, ou seja,

$$\hat{m} \triangleq \mathrm{E}\left[m|\mathbf{r}\right] = \overline{m}(\mathbf{r}), \tag{1.38}$$

onde $\mathrm{E}[\,\cdot\,|\,\cdot\,]$ é o operador esperança condicional, calculado em termos da função densidade de probabilidade *a posteriori*, $p_{m|\mathbf{r}}$. Neste momento, deve-se chamar a atenção para o fato de que a estimativa $\widehat{m}$, definida em (1.38), minimiza o $\overline{\epsilon^2}$ independentemente do tipo de modulação (linear ou não-linear) ou do tipo de ruído (Gaussiano ou não-Gaussiano).

Contudo, obter a estimativa $\hat{m}$ através de (1.38), em muitos casos, não é uma tarefa fácil, pois sua determinação depende de uma integração relacionada à função $p_{m|\mathbf{r}}$, que muitas vezes não é conhecida. Portanto, para evitar tais dificuldades, o valor de $\hat{m}$ considerado na análise de desempenho das modulações propostas nesse capítulo é o valor mais provável de m dado que $\mathbf{r}$ seja recebido. Dessa forma, a depender da função densidade de probabilidade *a priori* (p_m) dos sistemas de comunicações considerados, pode-se fazer uso de um dos dois tipos de receptores:

1. **receptor de máxima probabilidade a posteriori (MAP) -** dado que $\mathbf{r}$ seja recebido, então decida pelo valor $\hat{m}$ que satisfaz

$$p_m(\hat{m}|\mathbf{r}) \geq p_m(m|\mathbf{r}) \quad \xrightarrow{\text{Bayes}} \quad p_m(\hat{m})p_{\mathbf{r}}(\mathbf{r}|\hat{m}) \geq p_m(m)p_{\mathbf{r}}(\mathbf{r}|m), \quad \forall m. \tag{1.39}$$

2. **receptor de máxima verossimilhança (ML) -** dado que a função densidade de probabilidade *a priori* seja uniforme e $\mathbf{r}$ seja recebido, então decida pelo valor $\hat{m}$ que satisfaz

$$p_{\mathbf{r}}(\mathbf{r}|\hat{m}) \geq p_{\mathbf{r}}(\mathbf{r}|m), \quad \forall m. \tag{1.40}$$

Além disso, para um melhor entendimento dos resultados descritos neste trabalho, os desempenhos das modulações propostas foram apresentados em função de gráficos da relação sinal ruído do canal de transmissão (CSNR) *versus* a relação sinal ruído da informação transmitida (SNR), definidos como

$$\text{CSNR} \triangleq \frac{\overline{E}_m}{\mathcal{N}} \quad \text{e} \quad \text{SNR} \triangleq \frac{\overline{m^2}}{\overline{\epsilon^2}}, \tag{1.41}$$

onde $\overline{m^2}$ é a energia média (ou variância) da variável aleatória m de média nula, $\mathcal{N}$ é a energia média de ruído presente no canal de transmissão e $\overline{E}_m$ é a energia média de transmissão do sinal $s_m(t)$, definida como

$$\overline{E}_m \triangleq \mathrm{E}\left[|\mathbf{s}_m|^2\right] = \mathrm{E}\left[\int_{-\infty}^{\infty} s_m^2(t)\, dt\right], \tag{1.42}$$

onde $\mathrm{E}[\,\cdot\,]$ é o operador esperança, medido para todos os possíveis valores de m e $|\cdot|$ é a norma de um vetor.

Além disso, as curvas de desempenho (CSNR *versus* SNR) dos sistemas propostos neste capítulo foram comparadas com a curva de um sistema ótimo, denotada pela sigla OPTA e definida em [4] pela seguinte relação

$$\text{SNR} = (1 + \text{CSNR})^{N/M}, \tag{1.43}$$

onde (N/M) é a razão entre as dimensões dos sinais modulado e não modulado, respectivamente. Por exemplo, para o caso particular de modulações associadas a curvas planares essa razão vale 2, pois neste caso $\mathbf{s}_m \subset \mathbb{R}^{N=2}$ e $m \in \mathbb{R}^{M=1}$.

1.6 Visão Geral da Tese

Neste capítulo apresentamos as principais razões que motivaram a realização deste trabalho. Além disso, situamos este trabalho no contexto de uma pesquisa que vem sendo realizada desde 2000 e ressaltamos que os resultados aqui obtidos justificam o desenvolvimento dessa pesquisa. Realizamos também uma breve descrição de alguns tópicos de geometria diferencial, geometria Riemanniana, óptica geométrica e modulação *twisted*.

No Capítulo 2, uma interpretação geométrica que associa a métrica do espaço ao índice de refração de meios ópticos é utilizada para derivar expressões da curvatura de tais meios. Em seguida, um estudo sobre a curvatura de guias de ondas planares é realizado. Neste caso, alguns guias de ondas de interesse prático são associados a espaços de curvatura constante. Com isso, o fenômeno da dispersão modal é relacionado ao valor da curvatura e uma expressão dessa dispersão é derivada em função da curvatura do meio. Além disso, uma associação entre superfícies de rotação e guias de ondas planares é realizada e novos perfis de guias de ondas em meios anisotrópicos com dispersão nula são derivados dessa associação.

Ainda neste capítulo, fibras ópticas foram analisadas segundo sua curvatura e dispersão, sendo que novos perfis foram propostos e comparados com os utilizados comumente para a transmissão de informação. Neste caso, a fibra proposta com perfil toroidal apresentou oito vezes menos dispersão que a fibra com perfil parabólico e os perfis das fibras em meios anisotrópicos associadas a uma hiperesfera apresentaram dispersão nula. Além disso, no contexto de lentes ópticas, um instrumento óptico absoluto conhecido como *"fish-eye"* foi associado a um meio de curvatura positiva. Com isso, novos instrumentos ópticos absolutos foram propostos e associados a espaços hiperbólicos. Neste contexto, a convergência em lentes foi relacionada à curvatura positiva do meio e a divergência à curvatura negativa. Contudo, um novo método de convergência foi proposto para lentes com curvatura negativa, apresentando resultados muito interessantes quanto à forma de convergência, que no caso foi assintótica.

No Capítulo 3, uma associação entre modulação *twisted* e curvas foi generalizada para o contexto das variedades Riemannianas. Neste caso, foi mostrado que além do comprimento de arco da curva ou da métrica da superfície, a curvatura da curva ou da superfície associada à modulação também influencia no desempenho dos sistemas de comunicações. Além disso, a expressão da função densidade de probabilidade *a posteriori* foi obtida em função do comprimento de arco e da curvatura da curva associada à modulação. Verificamos também que é possível realizar um "casamento" entre a função densidade de probabilidade *a priori* e a parametrização da curva, de maneira que o desempenho do sistema de comunicações possa ser melhorado. Ainda no Capítulo 3, um novo esquema de modulação não linear foi proposto, apresentando um desempenho melhor que a modulação PM e a derivada da curva espiral de Arquimedes.

Ainda neste capítulo, com o intuito de diminuir o erro de ponto inicial, foram derivadas relações envolvendo o valor máximo da energia média de ruído e o valor máximo da curvatura da modulação. Além disso, uma expressão do erro quadrático médio foi determinada em função da métrica da superfície, propiciando que técnicas de modulação associadas a superfícies mínimas fossem caracterizadas por serem um ponto crítico de mínimo do erro quadrático médio. Neste contexto, uma nova modulação não-linear associada à superfície mínima helicóide foi proposta.

No Capítulo 4, usamos modulações não-lineares, como as consideradas no capítulo anterior, para realizar a construção e a análise de desempenho de constelações de sinais em espaços mais gerais que o Euclidiano. Neste caso, foi mostrado que, localmente, o espaço de sinais é uma variedade Riemanniana. Mais precisamente, a modulação não-linear induz de maneira natural uma nova métrica no espaço de sinais. Com isso, resultados de trabalhos anteriores puderam ser inseridos neste contexto sem maiores restrições. Além disso, expressamos a probabilidade média de erro em função da curvatura do espaço, e demonstramos que o desempenho de uma sistema de comunicações pode ser melhorado com a diminuição do valor da curvatura do espaço de sinais. Neste contexto, apresentamos a construção e a análise de desempenho de constelações de sinais geometricamente uniformes em espaços de curvatura constante. Neste caso, foram consideradas modulações M-PSK e modulações provenientes de tesselações $\{p, q\}$ dos espaços com curvatura constante. Além disso, constelações de sinais M-PSK em espaços de sinais associados a superfícies mínimas foram construídas e analisadas.

Finalmente, no Capítulo 5 expomos as considerações finais deste trabalho, destacando os resultados mais relevantes de cada capítulo e suas implicações no projeto de novos sistemas de comunicações. Além disso, apresentamos sugestões de trabalhos futuros para complementação e continuidade dos estudos na linha de pesquisa deste trabalho.

Capítulo 2

Um Estudo sobre a Curvatura de Meios Ópticos

Este capítulo apresenta um estudo sobre a curvatura de meios ópticos isotrópicos e anisotrópicos. Para tanto, fizemos uso de uma relação que associa o tensor dielétrico de um meio óptico ao tensor métrica de uma variedade Riemanniana tridimensional. Neste caso, tal associação permitiu que vários conceitos importantes da geometria diferencial e da geometria Riemanniana pudessem ser definidos também no contexto da óptica geométrica. Com isso, foi possível derivar expressões do tensor curvatura, da curvatura seccional e da curvatura de Ricci em meios ópticos e, a partir dessas expressões, caracterizar e interpretar geometricamente tais meios. Como conseqüência dessa interpretação, alguns instrumentos ópticos de interesse prático foram associados a espaços não-Euclidianos, como os espaços de curvatura constante. Além disso, alguns parâmetros de desempenho, como a dispersão, foram calculados em função da curvatura do meio e novos perfis de guias de ondas e lentes ópticas foram propostos de acordo com seus desempenhos e seus valores de curvaturas.

2.1 Introdução

Em [24] e [11], meios ópticos isotrópicos foram associados a espaços não-Euclidianos. Mais precisamente, estes trabalhos verificaram que as trajetórias das ondas eletromagnéticas em meios ópticos estão relacionadas às geodésicas de uma variedade Riemanniana tridimensional. Como conseqüência imediata dessa associação, foi mostrado que o quadrado de índice de refração de um meio isotrópico é igual a métrica de uma variedade Riemanniana.

Em [41], tal resultado foi generalizado para meios ópticos anisotrópicos, de forma que o tensor métrica de uma variedade Riemanniana tridimensional estivesse associado ao tensor dielétrico de um meio óptico. Além disso, tal associação foi utilizada para derivar a expressão da curvatura escalar de meios ópticos. Neste caso, o sinal da curvatura escalar foi relacionada ao fenômeno de convergência e divergência da luz em laser.

Mais recentemente, em [35], novos conceitos de curvatura, como a curvatura seccional e a curvatura de Ricci, foram derivados para meios ópticos anisotrópicos. Com isso, alguns perfis de guias de ondas planares e de fibras ópticas foram propostos e associados a espaços de curvatura seccional constante. Além disso, tais perfis apresentaram um desempenho satisfatório quanto ao fenômeno indesejado da dispersão modal, que ocorre durante a transmissão do pulso de informação, distorcendo sua forma de maneira que sua presença não seja detectada pelo receptor.

Neste capítulo, usamos a associação entre tensor métrica e tensor dielétrico, descrita em [41], para determinar e analisar a curvatura de alguns meios ópticos de interesse prático. O conceito de curvatura de meios ópticos considerado neste trabalho é similar ao conceito de curvatura do espaço-tempo descrito por Einstein, o qual impõe que todos os objetos materiais, desde uma maçã até um planeta, movem-se ao longo de geodésicas no espaço-tempo, a menos que sejam impedidos por uma força exterior. Fenômeno análogo acontece com as trajetórias dos raios de luz em meios ópticos, como demonstrou [24].

Neste caso, para realizarmos um estudo da curvatura de meios ópticos, foi necessário calcular o tensor curvatura, a curvatura seccional, a curvatura de Ricci e a curvatura escalar de tais meios. Além disso, verificamos que os meios ópticos com curvatura seccional constante são os de maior interesse prático, pois possuem propriedades de simetria importantes que podem ser aproveitadas no projeto de instrumentos ópticos.

Apresentamos também uma interpretação geométrica para as lentes e os guias de ondas ópticos, além de determinarmos uma aproximação para a expressão da dispersão modal em função da curvatura seccional do instrumento óptico. Usando esse último resultado, propomos novos guias de ondas em meios isotrópicos com dispersão menor que as de guias conhecidos e, em meios anisotrópicos, guias com dispersão nula.

2.2 A Curvatura de Meios Ópticos

Como mencionado anteriormente, em [41], o tensor métrica $[g_{ij}]$ de uma variedade Riemanniana tridimensional com parâmetros ortogonais, (1.17), foi associado ao tensor dielétrico de um meio anisotrópico não condutor ($\sigma = 0$) e não magnético no sistema de eixos principais do dielétrico.

Neste capítulo propomos uma associação mais geral do que a apresentada em [41] para os tensores métrica e dielétrico. Neste caso, consideramos que o tensor métrica $[g_{ij}]$ de uma variedade Riemanniana tridimensional seja igual ao tensor dielétrico de um meio homogêneo, não condutor e não magnético, no qual o sistema eixos do dielétrico não é necessariamente o principal, como considerado anteriormente, isto é

$$\mathcal{E} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} = G, \tag{2.1}$$

onde $[x, y, z] = [x_1, x_2, x_3]$ e $\mathcal{E}$ é o tensor dielétrico do meio óptico descrito em [21], que, através da relação $\mathbf{D} = \mathcal{E} \cdot \mathbf{E}$, associa a densidade de fluxo elétrico ($\mathbf{D}$) ao vetor campo elétrico ($\mathbf{E}$).

Portanto, podemos calcular o tensor curvatura (1.12), a curvatura seccional (1.14), a curvatura de Ricci (1.15) e a curvatura escalar (1.16) de meios ópticos usando a associação (2.1) e considerando que i, j, k e l variem de 1 até 3.

Como a densidade de energia elétrica, $W_e = (1/8\pi)\mathbf{E}\cdot\mathbf{D}$, é considerada não-negativa nesse meio óptico, então o tensor dielétrico deve ser simétrico, isto é, $\varepsilon_{ij} = \varepsilon_{ji}$. Tal fato é desejável, pois estamos considerando variedades Riemannianas cuja métrica também satisfaz a condição $g_{ji} = g_{ij}$. Com isso, é possível determinar um sistema de eixos principais do dielétrico, de maneira que $\mathcal{E}$ seja reduzido para a forma diagonal. Dessa forma, os cálculos das curvaturas de meios ópticos podem ser simplificados como realizado em [41], pois o tensor dielétrico e o tensor métrica com parâmetros ortogonais podem estar relacionados da seguinte forma

$$\mathcal{E} = \begin{bmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \varepsilon_3 \end{bmatrix} = \begin{bmatrix} n_1^2 & 0 & 0 \\ 0 & n_2^2 & 0 \\ 0 & 0 & n_3^2 \end{bmatrix} = \begin{bmatrix} g_1 & 0 & 0 \\ 0 & g_2 & 0 \\ 0 & 0 & g_3 \end{bmatrix} = G\,, \tag{2.2}$$

onde ε_1, ε_2 e ε_3 são as permissividades principais do dielétrico e n_1, n_2 e n_3 são os índices de refração principais.

Neste caso, considerando a associação (2.2), temos que o tensor curvatura (1.21) pode ser determinado por

$$R_{ijij} = -\frac{1}{2}\left[\frac{g_{i,i}g_{j,i}}{2g_i} + \frac{g_{i,j}g_{j,j}}{2g_j} - \frac{g_{i,k}g_{j,k}}{2g_k} + \frac{(g_{i,j})^2}{2g_i} + \frac{(g_{j,i})^2}{2g_j} - g_{i,jj} - g_{j,ii}\right] \tag{2.3a}$$

$$= \; n_i n_j \left[\left(\frac{n_{i,j}}{n_j}\right)_j + \left(\frac{n_{j,i}}{n_i}\right)_i + \frac{n_{i,k}n_{j,k}}{(n_k)^2}\right], \tag{2.3b}$$

onde $i, j, k = \{1, 2, 3\}$, $i \neq j \neq k \neq i$ e $n_{i,j} = d^2n/dx_i dx_j$.

Além disso, a curvatura seccional pode ser obtida de (2.3a) e (1.22) como sendo

$$K_{ij} = -\frac{1}{n_i n_j}\left[\left(\frac{n_{i,j}}{n_j}\right)_j + \left(\frac{n_{j,i}}{n_i}\right)_i + \frac{n_{i,k}n_{j,k}}{(n_k)^2}\right], \tag{2.4}$$

enquanto as curvaturas de Ricci e escalar são obtidas diretamente da substituição de (2.4) em (1.23) e em (1.24), respectivamente, resultando em

$$R_{jj} = n_j^2 \sum_{k=1}^{3} K_{kj} \quad \text{e} \quad S = 2\left(K_{12} + K_{13} + K_{23}\right).$$

Para o caso isotrópico, considerado em [24] e [11], o tensor métrica (1.25) de uma variedade Riemanniana tridimensional está associado ao índice de refração de um meio isotrópico, isto é, $g = n^2$. Assim, o tensor curvatura (1.28) de um meio isotrópico é dado por

$$\begin{aligned} R_{ijij} &= \frac{g}{2}\nabla^2_{ij}\log(g) + \frac{(g_{x_k})^2}{4g} && (2.5a)\\ &= n^2\nabla^2_{ij}\log(n) - (n_{x_k})^2, && (2.5b)\end{aligned}$$

onde $\nabla^2_{ij} = \left(\frac{\partial^2}{\partial x_j^2} + \frac{\partial^2}{\partial x_i^2}\right)$ e $n_{x_k} = \frac{\partial n}{\partial x_k}$.

Com isso, a curvatura seccional (1.29) é dada por

$$\begin{aligned} K_{ij} &= -\frac{1}{2g}\nabla^2_{ij}\log(g) - \frac{(g_{x_k})^2}{4g^3} && (2.6a)\\ &= -\frac{1}{n^2}\nabla^2_{ij}\log(n) - \frac{(n_{x_k})^2}{n^4}, && (2.6b)\end{aligned}$$

a curvatura de Ricci (1.30) por

$$\begin{aligned} R_{jj} &= \frac{1}{4g^2}\left[(\nabla g)^2 + 3(g_{x_j})^2\right] - \frac{1}{2g}\left[\nabla^2 g + g_{x_jx_j}\right] && (2.7a)\\ &= -\frac{1}{n}\left[\nabla^2 n + n_{x_jx_j}\right] - \frac{(n_{x_j})^2}{2n^2}, && (2.7b)\end{aligned}$$

onde $\nabla = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right]$ e a curvatura escalar (1.31) por

$$\begin{aligned} S &= \frac{2}{g^2}\left[\frac{3}{4g}(\nabla g)^2 - \nabla^2 g\right] && (2.8a)\\ &= \frac{2}{n^4}\left[(\nabla n)^2 - 2n\nabla^2 n\right]. && (2.8b)\end{aligned}$$

Devemos ressaltar que a equação (2.8b), a menos do sinal, havia sido obtida em [11] e associada à análise de foco em laser.

Um outro resultado que pode ser obtido dessa associação entre o tensor dielétrico e o tensor métrica é uma generalização, para meios anisotrópicos, da equação Eikonal (1.32), que caracteriza as trajetórias das ondas em meios isotrópicos. Para obter tal generalização, podemos usar a consideração realizada em [41], de que as trajetórias das ondas em meios anisotrópicos também são geodésicas em meios ópticos. Tal consideração é baseada no princípio de Fermat, que, além de outras implicações, afirma que as ondas percorrem o menor comprimento óptico (ou equivalentemente o menor tempo) em seu percurso no meio óptico. Portanto, uma expressão geral para a equação Eikonal em meios anisotrópicos pode ser obtida reescrevendo a equação das geodésicas (1.20), para o caso em que a métrica da variedade Riemanniana tridimensional seja dada por

(2.2). Com isso, temos que as trajetórias das ondas descritas pela curva parametrizada $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t)]$ devem satisfazer

$$\frac{d}{dt}\left(\frac{d\,\mathbf{x}}{dt} \cdot G\right) = \frac{1}{2}\nabla\left[\frac{d\,\mathbf{x}}{dt} \cdot G \cdot \left(\frac{d\,\mathbf{x}}{dt}\right)^T\right], \tag{2.9}$$

ou, mais precisamente, devem satisfazer o seguinte sistema de equações diferenciais

$$\begin{aligned} n_1^2\, x_1'' + \frac{1}{2}\left[n_{1,1}^2(x_1')^2 - n_{2,1}^2(x_2')^2 - n_{3,1}^2(x_3')^2\right] + n_{1,2}x_1'x_2' + n_{1,3}x_1'x_3' &= 0\,, \\ n_2^2\, x_2'' + \frac{1}{2}\left[n_{2,2}^2(x_2')^2 - n_{1,2}^2(x_1')^2 - n_{3,2}^2(x_3')^2\right] + n_{2,1}x_1'x_2' + n_{2,3}x_2'x_3' &= 0\,, \\ n_3^2\, x_3'' + \frac{1}{2}\left[n_{3,3}^2(x_3')^2 - n_{1,3}^2(x_1')^2 - n_{2,3}^2(x_2')^2\right] + n_{3,1}x_1'x_3' + n_{3,2}x_2'x_3' &= 0\,, \end{aligned} \tag{2.10}$$

onde a variável t pode, por exemplo, representar o tempo.

Portanto, quando o meio é isotrópico ($G = n^2\, I$) podemos usar (2.9) para determinar que as trajetórias das ondas em tal meio devem satisfazer

$$\frac{\partial}{\partial t}\left(n^2\frac{\partial \mathbf{x}}{\partial t}\right) = n\left[(x_1')^2 + (x_2')^2 + (x_3')^2\right]\nabla n\,. \tag{2.11}$$

Para mostrar que a equação Eikonal (1.32) é equivalente à equação (2.11), [24] considerou uma mudança de variáveis dada por $ds = dt/n$, de maneira que (1.32) seja reescrita da seguinte maneira

$$\frac{\partial}{\partial t}\left(n^2\frac{\partial \mathbf{x}}{\partial t}\right) = \frac{\nabla n}{n} \quad \text{com} \quad \left(\frac{\partial \mathbf{x}}{\partial t} \cdot \frac{\partial \mathbf{x}}{\partial t}\right) = \frac{1}{n^2}\,, \tag{2.12}$$

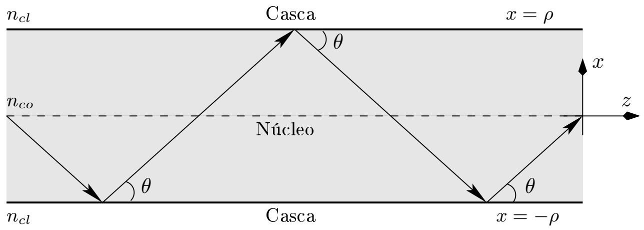
Neste caso, é possível concluir que (2.12) está parametrizada pelo comprimento de arco óptico, enquanto (1.32) está parametrizada pelo comprimento de arco Euclidiano. Tal interpretação é bastante útil quando se realiza uma interpretação ou análise das características geométricas das trajetórias das ondas no meio óptico.

2.3 Análise da Curvatura de Guias de Ondas Planares

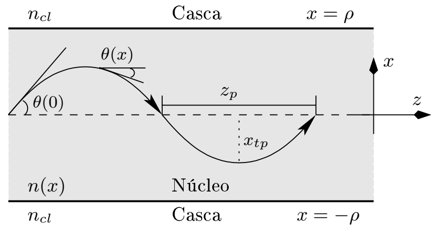
Tipicamente, um guia de ondas consiste de um meio óptico onde a informação em forma de luz é confinada e transmitida, na prática, por um percurso de longa distância. Nesta seção, analisamos a influência da curvatura óptica nessa importante classe de instrumentos ópticos formada pelos guias de ondas planares. Em partitular, serão considerados apenas os guias multimodos, pois supomos que o parâmetro do guia de ondas dado por

$$\mathcal{V} = \frac{2\pi\rho}{\lambda}(n_{co}^2 - n_{cl}^2)^{1/2}\,,$$

seja muito maior que um ($\mathcal{V} \gg 1$), onde λ é a freqüência da onda eletromagnética transmitida nesse guia, ρ é a metade da largura do guia, n_{cl} é o valor do índice de refração da casca que fornece isolação óptica ao guia e n_{co} é o máximo valor do índice de refração do material que compõe o núcleo do guia, geralmente calculado em $x = 0$, como ilustra a Figura 2.1.



(a) Guia com perfil uniforme.



(b) Guia com perfil gradual.

Figura 2.1: Exemplos de modelos de guias de ondas planares.

Em geral, a diferença entre guias de ondas monomodo e multimodos é que o segundo possui núcleo com diâmetro muito maior, permitindo a propagação da luz em vários modos. Além disso, sabemos que os guias multimodos não podem ser utilizados em aplicações de longas distâncias, pois apresentam dispersão modal. Isto é, quando um pulso óptico é transmitido em um guia de ondas multimodos, diversos modos de propagação são excitados, fazendo com que partes do pulso percorram caminhos diferentes no guia. Assim, as componentes do pulso que viajam nos modos de menor distância chegarão ao final da fibra mais rapidamente que as demais, causando um alargamento ou dispersão do pulso transmitido. Portanto, a distância máxima permitida para o uso de um determinado guia de ondas multimodos depende da largura de faixa de freqüência do guia e da taxa de transmissão utilizada.

Nesta seção, o efeito nocivo da dispersão no processo de transmissão será analisado em função da curvatura seccional do meio óptico em questão, que, neste caso, é equivalente à curvatura Gaussiana (K) de uma superfície regular no $\mathbb{R}^3$. Para tanto, suponha que os guias de ondas planares considerados sejam constituídos de uma camada óptica externa (casca), composta por um material de baixo índice de refração (n_{cl}), que envolve uma camada óptica interna (núcleo), fornecendo-lhe isolação óptica. Neste caso, adotamos que o núcleo possua largura 2ρ e que seu índice de refração possa ser uniforme, isto é, $n = n_{co}$, como ilustra a Figura 2.1(a) ou possa variar no eixo $\overrightarrow{ox}$, isto é, $n = n(x)$ como representa a Figura 2.1(b). Este último tipo de guia de ondas é conhecido como gradual e representará o principal objeto de estudo desta seção. Para um estudo das aplicações e dos parâmetros práticos de propagação em guias de ondas recomendamos a referência [2].

Antes de prosseguir com a análise de curvatura dos guias de ondas planares, alguns parâmetros de propagação da onda no guia devem ser apresentados. Tais parâmetros podem ser utilizados para se obter o desempenho dos guias de ondas, principalmente, para se determinar o valor de sua dispersão modal. Esses parâmetros podem ser obtidos com o auxílio da equação Eikonal (1.32), que, no caso particular de $n = n(x)$, pode ser simplificada para o seguinte sistema de equações diferenciais de primeira ordem

$$\begin{cases} x'' = \dfrac{n_x}{n}\left[1-(x')^2\right] \\ z'' = -\dfrac{n_x}{n}x'z' \end{cases} \Rightarrow \begin{cases} (x')^2 = 1-\dfrac{\beta^2}{n^2} \\ z' = \dfrac{\beta}{n} \end{cases}, \tag{2.13}$$

onde o parâmetro $\beta = n(0)z'(0)$ é chamado de invariante de onda e caracteriza a incidência da onda no guia. Com isso, considerando que $x' = \sin(\theta(x))$ e $z' = \cos(\theta(x))$, pois $(x')^2 + (z')^2 = 1$, pode-se obter que o ângulo $\theta(x)$, ilustrado na Figura 2.1(b), seja dado por

$$\theta(x) = \arctan\left(\sqrt{\frac{n^2}{\beta^2}-1}\right).$$

Um outro parâmetro que pode ser encontrado com o auxílio de (2.13) é o ponto de giro (*turning-point*), x_{tp}, que é definido como sendo o máximo valor alcançado pela coordenada $x(s)$ da trajetória da onda no meio. Tal parâmetro pode ser obtido quando $x' = 0$, o que implica em $z' = 1$, isto é, $n(x_{tp}) = \beta$. Além disso, o ponto de giro pode ser utilizado para determinar o valor do maior ângulo de incidência do raio na origem do guia de ondas, de maneira que a luz seja totalmente refletida em seu interior. Esse ângulo é conhecido como ângulo crítico, e neste caso é dado por $\theta_c = \arccos(n(\rho)/n_{co})$, pois foi considerado $x_{tp} = \rho$ e $\beta = n_{co}\cos(\theta_c)$.

Além dos parâmetros do guia de ondas citados anteriormente, neste momento apresentamos mais quatro outros parâmetros de interesse, a saber: o meio período de onda, representado na Figura 2.1(b),

$$z_p = \int_0^s z'(s)\,ds = \int_{-x_{tp}}^{x_{tp}} \frac{\beta}{n(x)}\sqrt{1+\left(\frac{dz}{dx}\right)^2}dx = \int_{-x_{tp}}^{x_{tp}} \frac{\beta}{\sqrt{n^2(x)-\beta^2}}dx, \tag{2.14}$$

o comprimento Euclidiano de meio período de onda

$$L_p = \int_{-x_{tp}}^{x_{tp}} \frac{n(x)}{\sqrt{n^2(x)-\beta^2}}dx, \tag{2.15}$$

o comprimento óptico de meio período de onda (1.33), que é igual a distância geodésica,

$$L_o = \int_{-x_{tp}}^{x_{tp}} \frac{n^2(x)}{\sqrt{n^2(x)-\beta^2}}dx, \tag{2.16}$$

e o tempo transiente de uma onda que percorre um comprimento z no guia de ondas

$$t = \frac{z}{c}\frac{L_o}{z_p}. \tag{2.17}$$

Com isso, a dispersão modal, t_d, que pode ocorrer na transmissão em guias de ondas multimodos, é definida como sendo a máxima variação do tempo transiente (2.17) em função do parâmetro β ou, equivalentemente, em função do ângulo de incidência θ_0. Dessa forma, note que a situação ideal de transmissão acontece quando a expressão (2.17) não depende de β, pois neste caso pode-se verificar que a dispersão modal no guia é nula.

Nesta seção estamos considerando apenas guias de ondas planares ($x_1 = x$ e $x_2 = z$) em meios ópticos isotrópicos, cujos índices de refração dependem apenas de x, então as curvaturas ópticas expressas por (2.6b), (2.7b) e (2.8b) podem ser simplificadas e reescritas da seguinte forma

$$K = K_{12} = K_{21} = \frac{1}{n^4}\left[(n_x)^2 - n_{xx}n\right], \quad \text{(Curvatura Gaussiana)} \tag{2.18a}$$
$$R = R_{11} = R_{22} = n^2 K\,, \qquad \text{(Curvatura de Ricci)} \tag{2.18b}$$
$$S = 2K\,, \qquad \text{(Curvatura escalar)}\,, \tag{2.18c}$$

onde $n_x = \frac{d\,n}{dx}$, $n_{xx} = \frac{d^2\,n}{dx^2}$.

Neste momento, consideramos alguns perfis de guias de ondas planares de interesse prático e calculamos suas curvaturas Gaussianas. Os resultados obtidos são apresentados na Tabela 2.1, sendo que a maioria dos perfis apresentados nesta tabela foram obtidos de [2]. Contudo, devemos chamar a atenção para o fato de que os perfis hiperbólico, Gaussiano e toroidal foram propostos em [35], pois apresentaram valores de dispersão pequenos, como será explicado mais adiante no decorrer desta seção.

Analisando mais detalhadamente as informações apresentadas na Tabela 2.1, podemos verificar que os perfis que possuem curvaturas Gaussiana e escalar constante são os perfis secante hiperbólica, degrau e hiperbólico, com curvaturas positiva, zero e negativa, respectivamente. Tais perfis possuem muitas simetrias, por estarem associados a espaços homogêneos, e como será mostrado mais adiante apresentam valores de dispersão modal nula, positiva e negativa, respectivamente. Além disso, verifique que o guia de ondas com perfil Gaussiano é o único que possui curvatura de Ricci constante, (2.18b).

Como mencionado anteriormente, os guias de ondas planares com curvatura Gaussiana constante estão associados a espaços homogêneos e, por este motivo, serão analisados mais detalhadamente no decorrer desta seção. Como conseqüência dos valores de suas curvaturas, o guia com perfil degrau está associado ao espaço Euclidiano, o guia secante hiperbólica ao espaço elíptico e o guia hiperbólico ao espaço hiperbólico.

Perfil	**Índice de refração ao quadrado (n^2)**	**Curvatura Gaussiana (K)**
Degrau	n_{co}^2	0
Parabólico	$n_{co}^2\left\{1-2\Delta\left(\frac{x}{\rho}\right)^2\right\}$	$\frac{2\Delta\left[1+2\Delta(x/\rho)^2\right]}{\rho^2 n_{co}^2\left[1-2\Delta(x/\rho)^2\right]^3}$
Secante hiperbólica	$n_{co}^2\mathrm{sech}^2\left(\sqrt{2\Delta}\frac{x}{\rho}\right)$	$\frac{2\Delta}{\rho^2 n_{co}^2}$
Clad power law	$n_{co}^2\left\{1-2\Delta\left\lvert\frac{x}{\rho}\right\rvert^q\right\}$	$\frac{q\Delta\lvert x\rvert^q\left\{q-1+2\Delta\lvert x/\rho\rvert^q\right\}}{\rho^q n_{co}^2\lvert x\rvert^2\left\{1-2\Delta\lvert x/\rho\rvert^q\right\}^3}$
Hiperbólico	$n_{co}^2\left(1+\sqrt{2\Delta}\frac{\lvert x\rvert}{\rho}\right)^{-2}$	$-\frac{2\Delta}{\rho^2 n_{co}^2}, \quad x\neq 0$
Gaussiano	$n_{co}^2\exp\left\{-2\Delta\left(\frac{x}{\rho}\right)^2\right\}$	$\frac{2\Delta}{\rho^2 n_{co}^2}\exp\left\{2\Delta\left(\frac{x}{\rho}\right)^2\right\}$
Toroidal	$n_{co}^2\left\{1+\Delta\left(\frac{x}{\rho}\right)^2\right\}^{-2}$	$\frac{2\Delta}{\rho^2 n_{co}^2}\left\{1-\Delta\left(\frac{x}{\rho}\right)^2\right\}$

Tabela 2.1: Curvatura Gaussiana de guias de ondas planares.

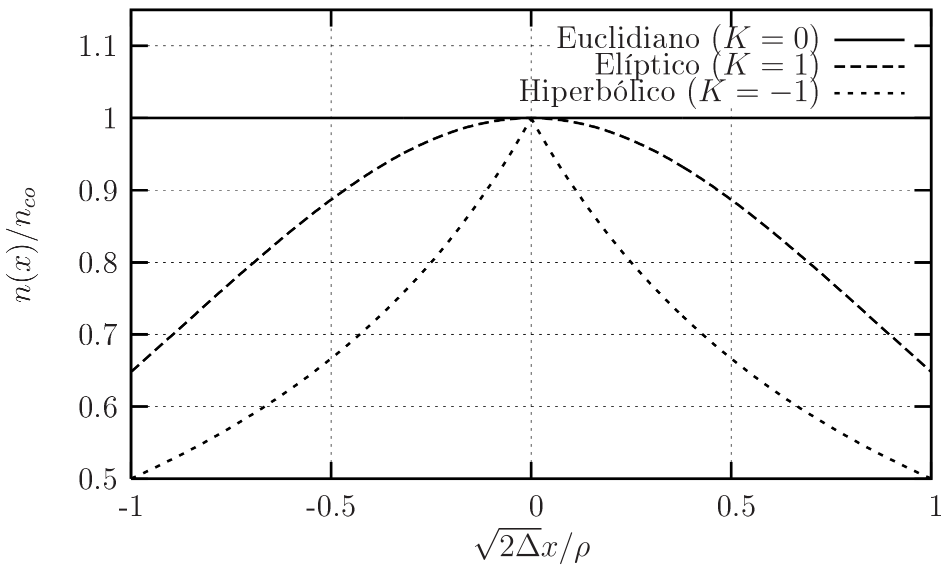


Figura 2.2: Curvas dos índices de refração dos guias de ondas planares Euclidiano, elíptico e hiperbólico.

A Figura 2.2 ilustra cada um desses três perfis, cujos índices de refração são exibidos na Tabela 2.1.

Contudo, existem outros guias de ondas com índices de refração na forma $n(x)$ e com curvatura Gaussiana constante que não foram considerados na Tabela 2.1. Neste caso, podemos usar a expressão (2.18a) para verificar que os índices de refração desses guias de ondas devem satisfazer a seguinte equação diferencial

$$(n_x)^2 - n\, n_{xx} - K\, n^4 = 0\,. \tag{2.19}$$

Com isso, pode-se constatar que, além do perfil secante hiperbólica, não existe uma outra solução para (2.19) quando a curvatura for constante e positiva. Entretanto, quando a curvatura Gaussiana é nula temos apenas mais uma solução, $n(x) = n_{co}\exp(\sqrt{2\Delta}x/\rho)$, e, quando a curvatura Gaussiana for constante e negativa, podemos obter as seguintes soluções para $n(x)$:

$$\frac{n_{co}}{\sqrt{2\Delta}x/\rho}, \quad \frac{n_{co}}{\sin(\sqrt{2\Delta}x/\rho)}, \quad \frac{n_{co}}{\sinh(\sqrt{2\Delta}x/\rho)} \quad \text{e} \quad \frac{n_{co}}{\cos(\sqrt{2\Delta}x/\rho)}\,. \tag{2.20}$$

Porém, observe que apenas o último dos perfis apresentados em (2.20) não possui singularidade em $x = 0$, e talvez possa ser fisicamente realizável. Entretanto, este perfil não confina luz, pois o valor de seu índice de refração aumenta à medida que se afasta da origem ($x = 0$). Por essa razão, esse perfil não tem aplicação como guia de ondas, podendo, contudo, ser usado no projeto de lentes ópticas, como será mostrado na Seção 2.5.

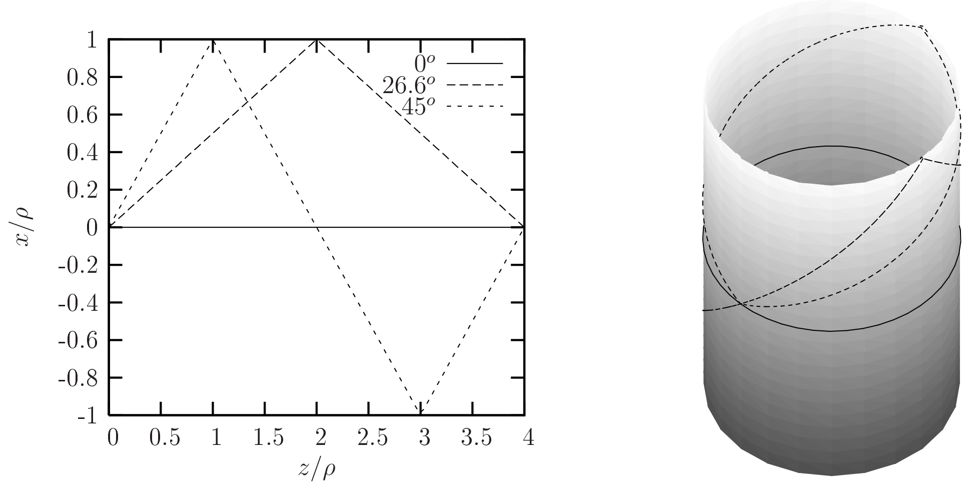
2.3.1 Perfil degrau

Como indicado na Tabela 2.1, o guia de ondas com perfil degrau possui curvatura nula, isto é, o espaço associado a esse guia de ondas é o espaço Euclidiano bidimensional. Neste caso, poderíamos utilizar uma superfície planar para representar esse guia de ondas. Entretanto, vamos representá-lo pela superfície de um cilindro (Figura 2.3), que também possui curvatura Gaussiana nula. Tal decisão foi tomada para manter um padrão na interpretação geométrica dos guias de ondas planares, dado que todos os outros guias de ondas apresentados na Tabela 2.1 puderam ser representados por superfícies de revolução, como será mostrado mais adiante.

Uma possível parametrização regular para a superfície do cilindro associado ao guia de ondas com perfil degrau e métrica igual a n_{co}^2, é dada por

$$X(x,z) = n_{co}\rho\left[\cos(\frac{z}{\rho}), \sin(\frac{z}{\rho}), \frac{x}{\rho}\right].$$

Neste guia de ondas, note que as ondas eletromagnéticas que incidem com ângulos diferentes percorrem caminhos com comprimentos ópticos diferentes. Além disso, verifique que quanto maior for o ângulo de incidência θ_0 maior será a quantidade de reflexões e, portanto, o tempo de percurso, isto é, existe dispersão modal nesse meio. Como o comprimento óptico da trajetória com $\theta_0 = 0$, paralela ao eixo $\overrightarrow{oz}$, é menor do



(a) Guia $K = 0$.

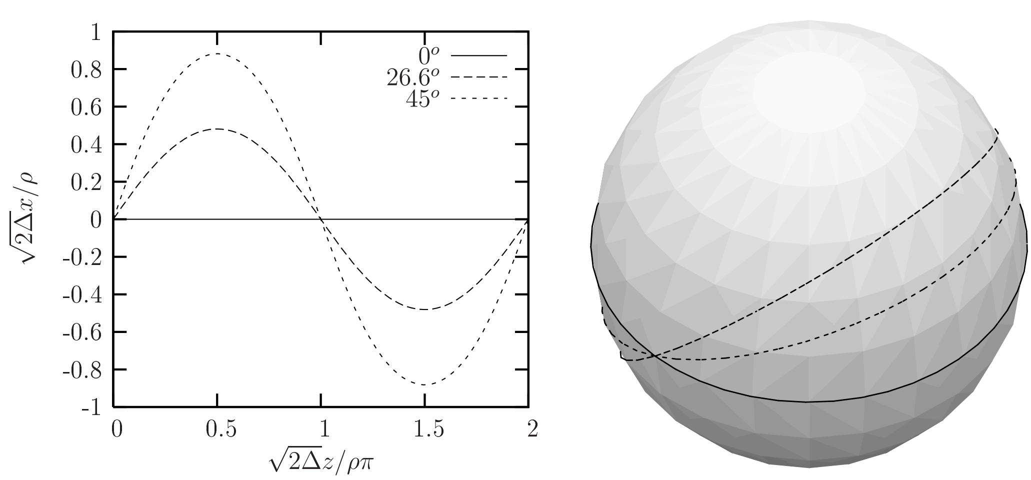
(b) Cilindro.

Figura 2.3: Interpretação geométrica para o guia de ondas com perfil degrau.

que o comprimento óptico das trajetórias com ângulo de incidência diferente de zero, essa dispersão será convencionada como sendo positiva, e caso ocorra o inverso desse fato, a dispersão será dita negativa.

2.3.2 Perfil secante hiperbólica

Da Tabela 2.1, note que o perfil secante hiperbólica possui curvatura constante positiva. Isto significa que podemos associar tal meio óptico à superfície de uma esfera no $\mathbb{R}^3$, como ilustra a Figura 2.4. Neste caso, pode-se afirmar que esse meio óptico é uma variedade Riemanniana bidimensional com métrica induzida de uma superfície esférica.



(a) Guia $K = 1$.

(b) Esfera.

Figura 2.4: Interpretação geométrica para o guia de ondas com perfil secante hiperbólica.

Uma característica de desempenho muito importante desse guia de ondas com perfil secante hiperbólica é que ele possui dispersão nula, como demonstrado em [2]. Este fato pode ser facilmente verificado e visualizado usando a superfície da esfera associada a esse meio óptico. Pois, note que todas as geodésicas que partem de um ponto qualquer da esfera, independentemente da direção de saída, retornam a esse ponto após um mesmo tempo, isto é, os comprimentos geodésicos ou ópticos são todos iguais, não indicando assim a existência de dispersão modal.

Uma possível parametrização regular para a superfície da esfera associada ao guia de ondas com perfil secante hiperbólica e métrica igual a $n_{co}^2 \mathrm{sech}^2(ax)$ é dada por

$$X(x,z) = \frac{n_{co}}{a}\left[\frac{\cos(az)}{\cosh(ax)}, \frac{\sin(az)}{\cosh(ax)}, \tanh(ax)\right],$$

onde $a = \sqrt{2\Delta}/\rho$.

2.3.3 Perfil hiperbólico

Nesta subseção analisaremos o guia de ondas com perfil hiperbólico apresentado na Tabela 2.1. Um fato importante sobre esse guia é que sua curvatura Gaussiana é constante e negativa para $x \neq 0$ e infinita para $x = 0$, pois a derivada de $n(x)$ não é contínua em $x = 0$, devido ao fato de $n(x)$ ser função de $|x|$.

Uma outra característica desse guia é que as trajetórias das ondas eletromagnéticas possuem curvatura constante e positiva, isto é, são semicírculos, como ilustra a Figura 2.5. Este fato é importante para a análise do fenômeno de difração que ocorre em guias de ondas monomodais, porém não será abordado neste trabalho. Para verificar que as trajetórias nesse meio possuem essa propriedade é necessário obter uma expressão para a curvatura das trajetórias em um meio óptico com índice de refração $n(x)$. Para tanto, considere a forma simplificada da equação Eikonal, (2.13), e a equação da curvatura de curvas planas dada em (1.1), para obter que

$$k = -\frac{n_x}{n} z' = -\frac{\beta}{n}\frac{d}{dx}\log(n) = \beta\left(n^{-1}\right)_x . \tag{2.21}$$

Portanto, de (2.21), o valor do módulo da curvatura das trajetórias das ondas no perfil hiperbólico é constante e igual a $\beta\sqrt{2\Delta}/(\rho\, n_{co})$.

Assim, usando a equação (2.13) e o fato das trajetórias nesse guia serem círculos, pode-se obter que tais trajetórias são descritas pela seguinte expressão

$$x = \frac{\rho}{\sqrt{2\Delta}}(-1)^{z_i}\left[\frac{1}{\beta}\sqrt{n_{co}^2 - (n_{co}^2 - \beta^2)(1-2z_f)^2} - 1\right],$$

onde z_i e z_f são as partes inteira e fracionária de z/z_p, respectivamente.

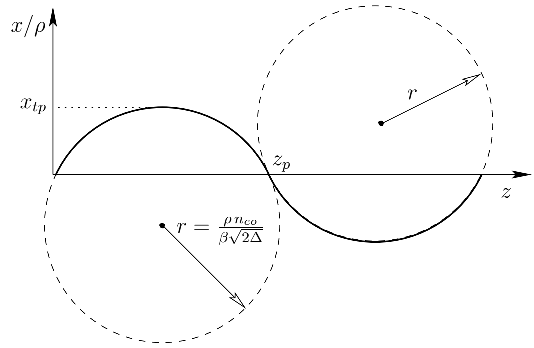


Figura 2.5: Trajetórias das ondas em um guia de ondas planar com perfil hiperbólico.

Além disso, os parâmetros que caracterizam a propagação da onda nesse guia de ondas hiperbólico foram calculados, como o ponto de giro

$$n(x_{tp}) = \beta \quad \Rightarrow \quad x_{tp} = \frac{\rho}{\sqrt{2\Delta}} \left(\frac{n_{co}}{\beta} - 1 \right),$$

o meio período de onda, (2.14), dado por

$$z_p = \frac{2\rho}{\beta\sqrt{2\Delta}} \sqrt{n_{co}^2 - \beta^2},$$

o comprimento Euclidiano de meio período de onda, (2.15), dado por

$$L_p = \frac{2\rho\, n_{co}}{\beta\sqrt{2\Delta}} \left[\frac{\pi}{2} - \sin^{-1}\left(\frac{\beta}{n_{co}} \right) \right],$$

o comprimento óptico de meio período de onda, (2.16), dado por

$$L_o = \frac{2\rho\, n_{co}}{\sqrt{2\Delta}} \operatorname{sech}^{-1}\left(\frac{\beta}{n_{co}} \right),$$

e o tempo transiente, (2.17), dado por

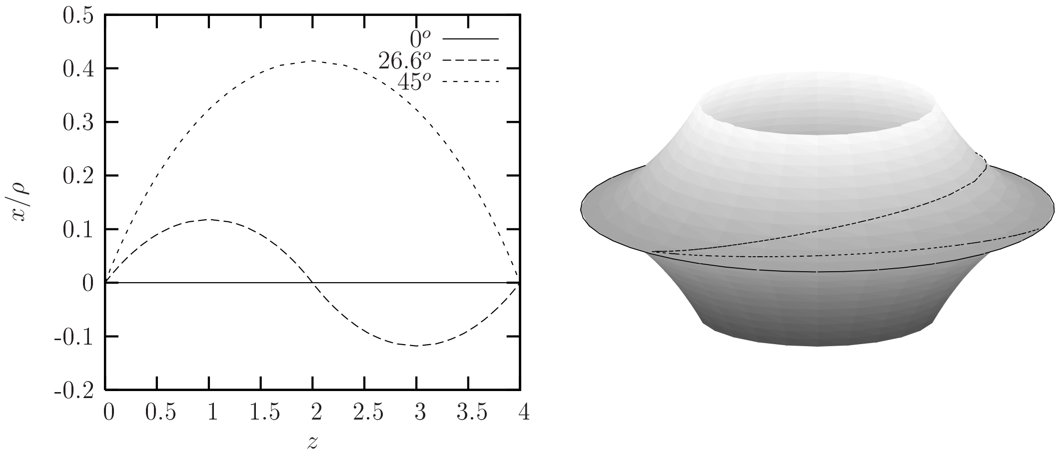
$$t = \frac{z}{c}\frac{L_o}{z_p} = \frac{z\,\beta\, n_{co}}{c\sqrt{n_{co}^2 - \beta^2}} \operatorname{sech}^{-1}\left(\frac{\beta}{n_{co}} \right). \tag{2.22}$$

Um exemplo de superfície do $\mathbb{R}^3$ que pode ser associada ao guia de ondas com perfil hiperbólico é a pseudo-esfera, pois tal superfície possui curvatura Gaussiana constante e negativa. Tal superfície é ilustrada na Figura 2.6, sendo que uma possível parametrização para a pseudo-esfera com métrica igual a $n_{co}^2(1 + \sqrt{2\Delta}|x|/\rho)^{-2}$ é dada por

$$X(x, z) = \frac{n_{co}}{a} \left[\frac{\cos(az)}{\cosh(u)}, \frac{\sin(az)}{\cosh(u)}, u - \tanh(u) \right],$$

onde $a = \sqrt{2\Delta}/\rho$ e $u(x) = \cosh^{-1}(1 + a|x|)$.

Esse guia apresenta uma característica muito especial quanto à dispersão de seus modos, uma dispersão invertida ou negativa, isto é, as ondas que incidem com maior



(a) Guia $K = -1$. (b) Pseudo-esfera.

Figura 2.6: Interpretação geométrica para o guia de ondas com perfil hiperbólico.

inclinação são as que possuem menor tempo de percurso no meio. Este fato pode ser matematicamente comprovado usando (2.22) ou graficamente verificado analisando a Figura 2.6(b).

Com o objetivo de melhor analisar o desempenho desse guia, sua curva de dispersão foi calculada e comparada à curva de dispersão do guia com perfil degrau. A Figura 2.7 apresenta essas curvas de dispersão modal para esses guias em função do ângulo de incidência θ_0 em graus, onde $\beta = n_{co} \cos(\theta_0)$. Observe que a dispersão do guia hiperbólico é menor do que a dispersão do guia Euclidiano.

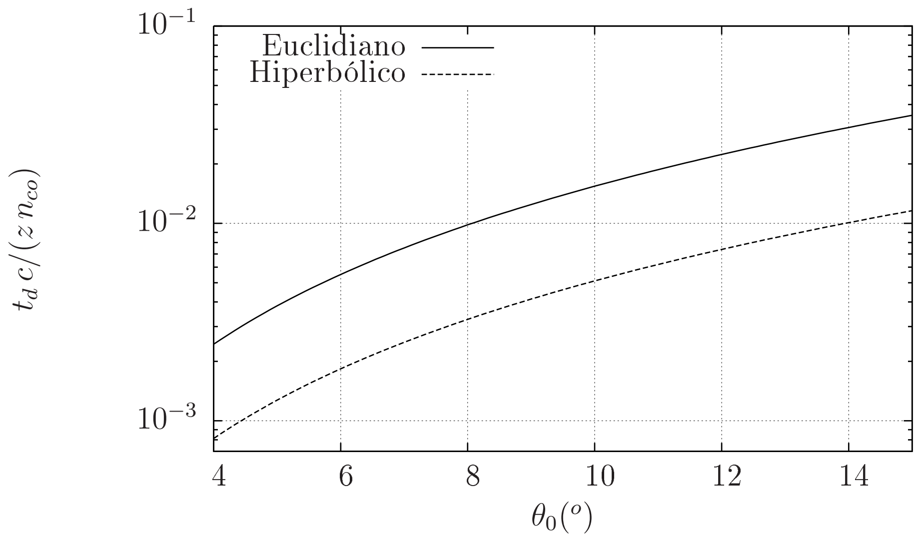


Figura 2.7: Curvas de dispersão para os meios Euclidiano (perfil degrau) e hiperbólico.

O fato do guia de ondas com perfil hiperbólico possuir dispersão negativa pode ser convenientemente aproveitada, pois podemos usar tal guia como um instrumento óptico que minimize os efeitos da dispersão positiva. Por exemplo, podemos usar o guia hiperbólico no final da transmissão realizada em um guia de ondas com dispersão positiva, reduzindo assim a dispersão na transmissão.

2.3.4 Perfis parabólico, Gaussiano, toroidal e *clad power law*

A Tabela 2.1 apresenta os índices de refração e os valores da curvatura Gaussianas dos guias de ondas planares com perfil parabólico, Gaussiano, toroidal e *clad power law*. A Figura 2.8 mostra as curvas dos índices de refração desses guias e do guia com perfil secante hiperbólica, para efeitos de comparação. Neste caso, observe que o guia toroidal é o que mais se aproxima do guia elíptico (perfil secante hiperbólica), que possui dispersão modal nula.

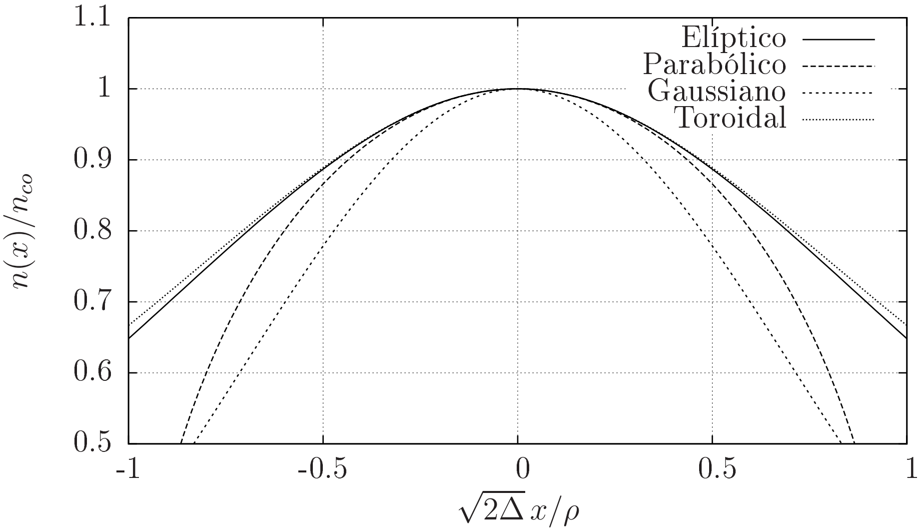


Figura 2.8: Índice de refração dos guias de ondas planares elíptico, parabólico, Gaussiano e toroidal.

Devemos ressaltar que os parâmetros de propagação do guia de ondas com perfil Gaussiano não serão apresentados, pois suas expressões não podem ser obtidas analiticamente, apenas na forma integral, sendo que seus valores são mais facilmente obtidos por integração numérica. Por essa razão, a curva de dispersão desse guia com perfil Gaussiano foi calculada numericamente, mas com uma precisão de aproximadamente 10^{-7}.

A denominação toroidal atribuída a esse guia está relacionada a superfície de revolução toroidal associada a esse guia de ondas (tal fato será melhor explicado na seção 2.3.7). Nosso interesse por tal perfil está associado ao valor de sua dispersão, que, apesar de não ser nula, é menor que a dispersão dos outros guias considerados, principalmente, quando comparada à dispersão do guia parabólico, que é comumente utilizado na prática. As Figuras 2.7 e 2.9 ilustram este fato.

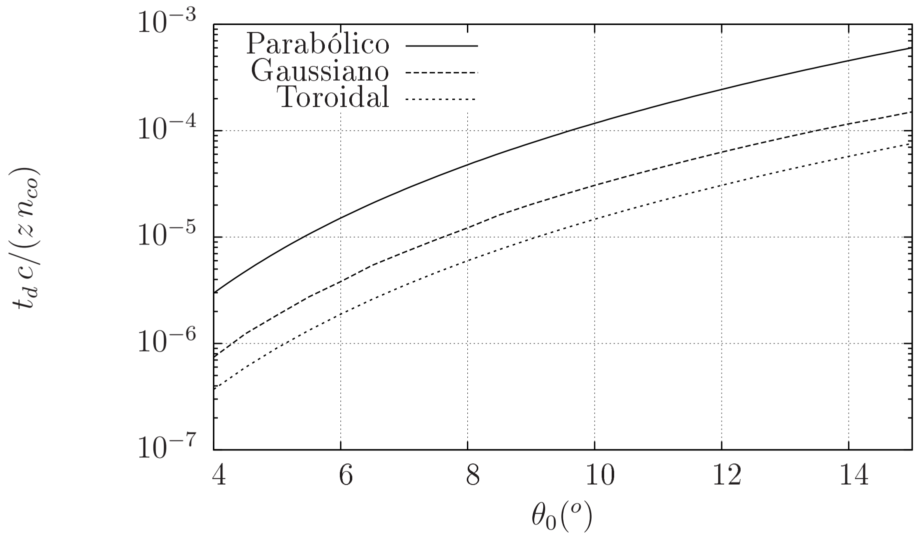


Figura 2.9: Curva de dispersão para os guias de ondas planares com perfis parabólico, Gaussiano e toroidal.

A seguir serão apresentados alguns dos principais parâmetros que caracterizam a propagação da onda em um guia com perfil toroidal. Tais parâmetros podem ser comparados aos dos guias com perfil degrau, parabólico, secante hiperbólica, etc. que foram determinados em [2]. Para o caso do guia toroidal, o ponto de giro é dado por

$$x_{tp} = \frac{\rho}{\sqrt{\Delta}}\sqrt{\frac{n_{co}}{\beta} - 1},$$

o meio período de onda por

$$z_p = \frac{2\rho}{\sqrt{\Delta\beta(\beta + n_{co})}}\left[(\beta + n_{co})\,\mathrm{EllipticE}\left(\sqrt{\frac{\beta - n_{co}}{\beta + n_{co}}}\right) - n_{co}\,\mathrm{EllipticK}\left(\sqrt{\frac{\beta - n_{co}}{\beta + n_{co}}}\right)\right],$$

onde as funções elípticas EllipticE e EllipticK dependem de integrais que só possuem soluções numéricas. Para mais informações sobre essas funções consulte [20].

Além desses parâmetros, para se determinar a dispersão modal desse guia é necessário determinar o comprimento óptico de meio período de onda (2.16), que no caso é dado por

$$L_o = \frac{2\rho n_{co}^2}{\sqrt{\Delta\beta(\beta + n_{co})}}\,\mathrm{EllipticPi}\left(1 + \frac{n_{co}}{\beta}, \sqrt{\frac{\beta - n_{co}}{\beta + n_{co}}}\right),$$

onde EllipticPi é uma outra função elíptica.

Usando estes parâmetros podemos determinar que a dispersão de tal guia é aproximadamente 2 vezes menor do que a dispersão do guia Gaussiano e 8 vezes menor que a dispersão do guia parabólico, como mostra a Figura 2.9. Além disso, a dispersão do guia de ondas toroidal é invertida, assim como ocorreu com a dispersão no guia de ondas hiperbólico.

2.3.5 Influência da curvatura Gaussiana na dispersão modal

A princípio, será determinada uma relação qualitativa entre a dispersão modal e a curvatura Gaussiana dos guias de ondas planares tratados nesta seção. Para tanto, note que o guia de ondas com curvatura constante possui dispersão nula e que os guias de ondas que apresentam alguma dispersão, por sua vez, não têm curvatura constante. Tal fato nos remete a esperar que uma curvatura não constante ou sua variação possa estar relacionada a existência de dispersão.

Para verificar se esta conclusão se aplica aos guias de ondas planares considerados, vamos analisar mais detalhadamente o comportamento das curvaturas desses guias. Para tanto, considere a Figura 2.10, que apresenta os gráficos das curvaturas Gaussianas dos perfis parabólico, Gaussiano e toroidal, cujos valores variam dentro do guia cerca de 8.4%, 2.0% e 1.0%, respectivamente, para $\Delta = 0.01$.

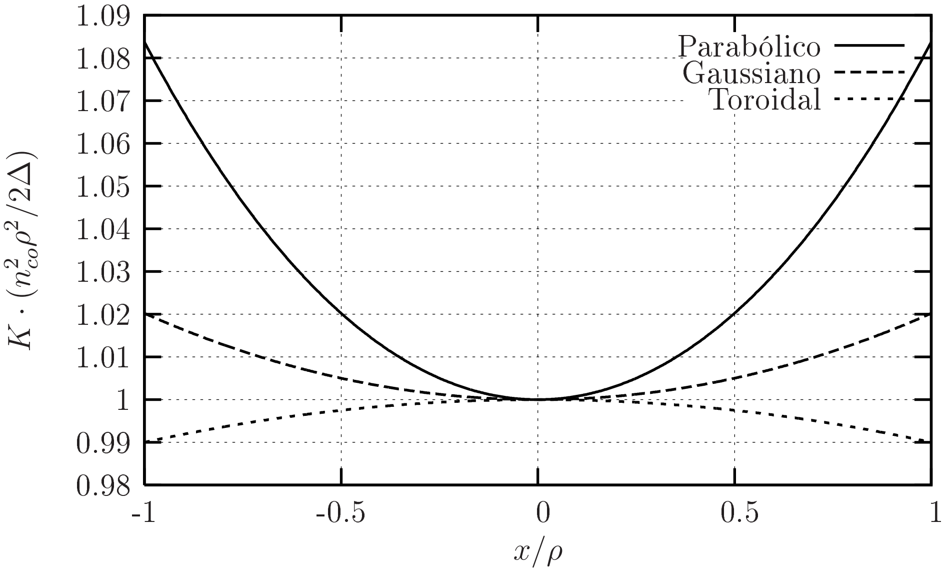


Figura 2.10: Curvatura dos guias de ondas parabólico, Gaussiano e toroidal, com $\Delta = 0.01$.

Neste caso, note que as variações nos valores da curvatura Gaussiana nesses guias de ondas são diretamente proporcionais aos seus valores de dispersão, como ilustra a Figura 2.9. Este fato indica que a curvatura Gaussiana de um guia de ondas certamente influência no valor de sua dispersão modal. Contudo, uma abordagem matemática sobre essa influência da curvatura na dispersão será detalhada apenas mais adiante.

Além disso, a Figura 2.11 apresenta as curvaturas Gaussianas de um meio com perfil *clad power law* para $q = 1.98$, 2 e 2.02, quando $\Delta = 0.01$. Neste caso, note que quando $x \to 0$ a curvatura Gaussiana tende a zero se $q > 2$ e tende a infinito se $q < 2$. Dessa forma, como a variação de curvatura nesse guia de ondas é razoavelmente grande, então é de se esperar que o valor de sua dispersão seja maior do que a dispersão dos guias apresentados.

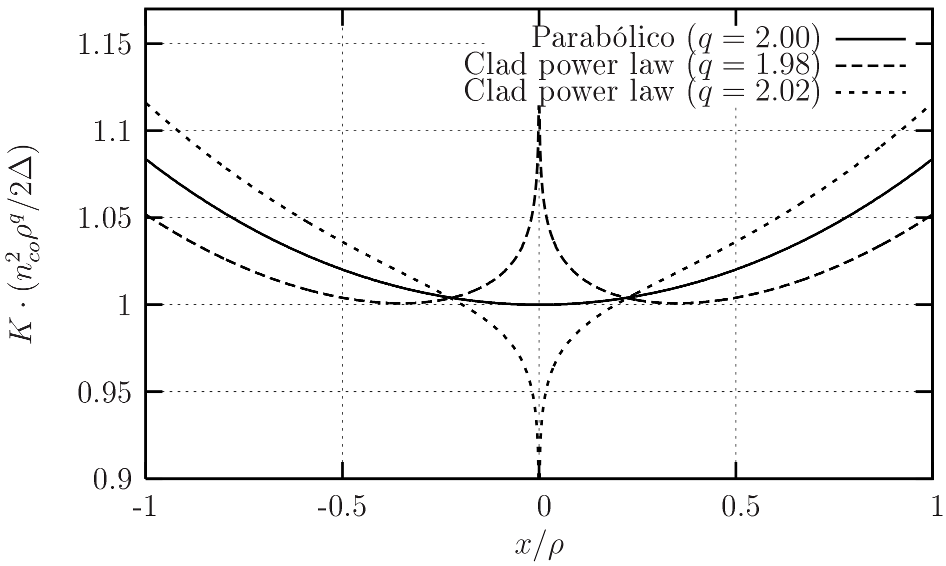


Figura 2.11: Curvatura dos guia de ondas *clad power law* para $q = 1.98$, 2 e 2.02, com $\Delta = 0.01$.

O conjunto de fatos e observações realizadas anteriormente nos remete a pensar que a dispersão está diretamente associada à variação da curvatura no guia. Porém, se isto for verdade, como expressar quantitativamente a dispersão em função da variação da curvatura?

Para responder a essa pergunta, usamos a série de Taylor com aproximação de quinta ordem para derivar uma expressão da dispersão modal em guias de ondas planares, em função de sua curvatura Gaussiana e de suas derivadas em relação a x. A aproximação de quinta ordem foi utilizada para que a dispersão fosse expressa em função da segunda derivada da curvatura. Neste caso, foi mais adequado determinar os parâmetros de propagação da onda, necessários para se obter t_d, em função do comprimento óptico (s), isto é

$$t_d = \frac{z}{c}\left(\frac{s}{z(s) - z(0)} - n_{co}\right) \quad \text{com} \quad z(s) = \sum_{i=0}^{5} \frac{s^i}{i!} z^{(i)}\,, \tag{2.23}$$

onde $z^{(i)}(0)$ é i-ésima derivada de $z(s)$ em relação a s no ponto $s = 0$.

Para tanto, utilizamos o fato de $n_x(0) = 0$ e a expressão (2.12), com $t = s$, para determinar os seguintes valores para as derivadas de $x(s)$ e $z(s)$, em $s = 0$,

$$(x')^2 = \frac{1}{n^2}\left(1 - \frac{\beta^2}{n^2}\right), \quad z' = \frac{\beta}{n^2}, \quad x'' = z'' = 0\,, \quad x^{(3)} = \frac{n_{xx}x'}{n}\left[\frac{1}{n^2} - 2(x')^2\right],$$

$$z^{(3)} = 2n^2(x')^2 z' K\,, \quad z^{(4)} = 0\,, \quad \text{e} \quad z^{(5)} = 2n^2(x')^2 z' \left[\left(K_{xx} + 12n^2K^2\right)(x')^2 - 4K^2\right], \tag{2.24}$$

onde $\beta = n_{co}\cos(\theta_0)$ é o invariante de onda, K é curvatura Gaussiana (2.18a) e K_{xx} é sua segunda derivada em relação a x, dada por

$$K_{xx} = \frac{1}{n^3}\left(-n^{(4)} + 5\frac{(n_{xx})^2}{n}\right),$$

onde $n^{(i)}$ é a i-ésima derivada de $n(x)$ em relação a x.

Portanto, substituindo (2.24) em (2.23) podemos obter a seguinte expressão aproximada para a dispersão modal em guias de ondas planares isotrópicos

$$\begin{aligned} t_d = {} & \frac{z}{c}\frac{n_{co}}{\cos(\theta_0)}\left\{1 + \frac{\cos(\theta_0)\sin^2(\theta_0)}{n_{co}}\frac{s^2}{3}\left[K + \frac{s^2}{20}\left(\frac{\sin^2(\theta_0)}{n_{co}^2}(K_{xx} + 12K^2) - 4K^2\right)\right]\right\}^{-1} \\ - {} & \frac{z}{c}n_{co}\,. \end{aligned} \tag{2.25}$$

Neste caso, note que (2.25) descreve perfeitamente a dispersão de um guia com perfil degrau ($K = 0$) e indica como o valor da curvatura Gaussiana de um guia de ondas influencia em sua dispersão. Além disso, note que quando uma transmissão é realizada em guias de ondas com índices de refração diferentes, porém com os mesmos valores de n_{co}, θ_0 e de K, em $x = 0$, então o termo que determina a diferença de dispersão entre esses meios é K_{xx}.

Para ilustrar tal fato, verifique, com o auxílio das Figuras 2.10 e 2.9, que tanto a curvatura como a dispersão no guia toroidal variam aproximadamente 2 vezes menos do que no guia Gaussiano e 8 vezes menos do que no guia parabólico. Mais precisamente, podemos verificar que a dispersão nesses guias está relacionada à segunda derivada da curvatura Gaussiana no ponto $x = 0$, pois

$$\begin{aligned} K_{xx}(0) &= \frac{32\Delta^2}{\rho^4 n_{co}^2} \quad \text{(parabólico)}, \\ &= \frac{8\Delta^2}{\rho^4 n_{co}^2} \quad \text{(gaussiano)}, \\ &= -\frac{4\Delta^2}{\rho^4 n_{co}^2} \quad \text{(toroidal)}, \end{aligned}$$

que segue a mesma relação encontrada anteriormente entre a curvatura e a dispersão. Isto é, $K_{xx}(0)$ no guia toroidal é 2 vezes menor do que no guia Gaussiano e 8 vezes menor do que no guia parabólico.

Além disso, notamos que a dispersão invertida que ocorre no guia de ondas com perfis toroidal e hiperbólico está relacionada ao sinal de $K_{xx}(0)$, que no caso é negativa, indicando que $K(0)$ é um ponto de máximo global da curvatura, pois $K_x(0) = 0$ devido à simetria do guia.

Neste momento apresentamos uma outra maneira de verificar como a curvatura Gaussiana influencia no desempenho de guias de ondas planares. Para tanto, recorde que, na prática, um guia de ondas com perfil gradual confina luz quando n_{co} é o máximo valor de seu índice de refração, $n(x)$. Este fato implica que as condições $n_x = 0$ e $n_{xx} < 0$ devem ser satisfeitas para $x = 0$. Assim, por substituição direta dessas condições em (2.18a), obtemos que a seguinte condição $K = -n_{xx}/n^4 > 0$, deve ser satisfeita em $x = 0$. Como conseqüência desse fato, podemos concluir que não existe um guia de ondas de interesse prático cuja curvatura seja negativa em todos os seus pontos.

Este fato também pode ser verificado se considerarmos a associação desse meio óptico a uma variedade Riemanniana, pois neste caso podemos aplicar em um meio óptico o conceito de campos de Jacobi e o Teorema de Hadamard. Veja [27] para mais informações sobre estes tópicos da geometria Riemanniana. Dessa forma, como as geodésicas de uma variedade Riemanniana são equivalentes às trajetórias das ondas no meio óptico, então o vetor de Poynting pode ser associado aos campos de Jacobi da correspondente variedade Riemanniana. Com isso, podemos aplicar o Teorema de Hadamard para mostrar que o lugar dos pontos conjugados nessa variedade não é vazio, isto é, existe mais do que uma geodésica que passa por dois pontos quaisquer de M^2. Assim, a aplicação exponencial não é um difeomorfismo local e, conseqüentemente, a curvatura seccional deve ser positiva em algum ponto.

2.3.6 Abertura numérica

Para que a onda eletromagnética seja transmitida ao longo de um guia de ondas, é necessário que exista uma reflexão interna total da onda no guia, consevando assim sua energia durante a transmissão. Neste caso, a luz permanece no núcleo do guia de ondas quando seu ângulo de incidência na fibra seja aceitável. Tal ângulo é chamado de ângulo de aceitação (*acceptance angle*) e caracteriza a abertura numérica (*Numerical Aperture*) que determina a quantidade de luz que entra no guia de ondas. Por esse motivo, esses parâmetros são muito importantes e devem ser analisados em função da curvatura dos guias e considerados no projeto.

Dessa forma, o ângulo máximo de incidência da luz na fibra é dado pelo ângulo crítico (θ_c), que depende dos índices de refração do núcleo (n_{co}) e da casca ($n_{cl} = n(\rho)$), conforme mostra a Figura 2.12, isto é

$$\theta_c = \arccos\left(\frac{n_{cl}}{n_{co}}\right). \tag{2.26}$$

Além disso, em função do ângulo crítico (2.26), pode-se definir a abertura numérica (**NA**) e o ângulo de aceitação ($\widehat{A}$) como

$$\mathbf{NA} = n_{co}\sin(\theta_c) = \sqrt{n_{co}^2 - n_{cl}^2} \quad \text{e} \quad \widehat{A} = 2\theta_c\,.$$

A Tabela 2.2 apresenta os valores da abertura numérica para todos os guias de ondas planares apresentados na Tabela 2.1. Contudo, para uma melhor análise desse

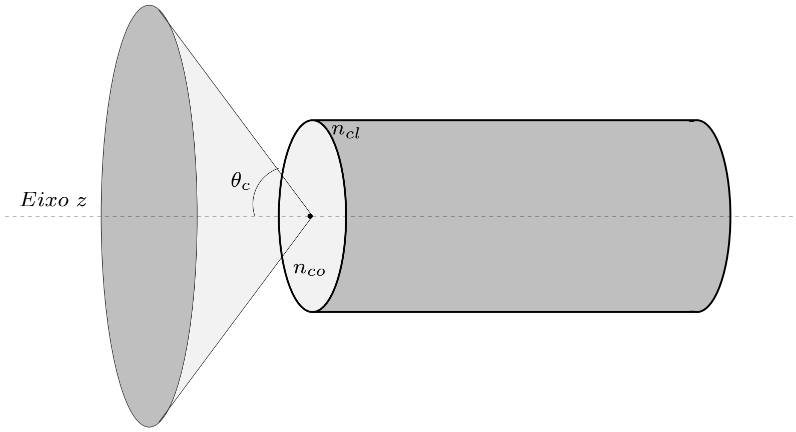


Figura 2.12: Abertura numérica e ângulo de aceitação.

parâmetro em função da curvatura dos guias de ondas planares, as curvas do ângulo crítico versus Δ são apresentadas na Figura 2.13, isto porque tanto K quanto θ_c dependem do parâmetro Δ. Neste caso, podemos perceber que θ_c é praticamente o mesmo para todos os guias exceto para o hiperbólico, que demonstra possuir o maior ângulo de aceitação.

Perfil:	**NA**
Degrau	$\sqrt{n_{co}^2 - n_{cl}^2}$
Parabólico	$n_{co}\sqrt{2\Delta}$
Secante hiperbólica	$n_{co}\tanh(\sqrt{2\Delta})$
Hiperbólico	$n_{co}\sqrt{1-(1+\sqrt{2\Delta})^{-2}}$
Gaussiano	$n_{co}\sqrt{1-\exp(-\sqrt{2\Delta})}$
Toroidal	$n_{co}\sqrt{1-(1+\Delta)^{-2}}$

Tabela 2.2: Abertura numérica de guias de ondas planares.

2.3.7 Guias de ondas planares provenientes de superfícies de revolução

Observe que os guias de ondas com perfis degrau, secante hiperbólica e hiperbólico considerados anteriormente foram interpretados geometricamente e associados a superfícies de revolução, no caso um cilindro, uma esfera e uma pseudoesfera, respectivamente. Apesar de essa associação ter sido justificada pela coincidência de valores das curvaturas Gaussianas dos guias de ondas e das correspondentes superfícies, sabe-se

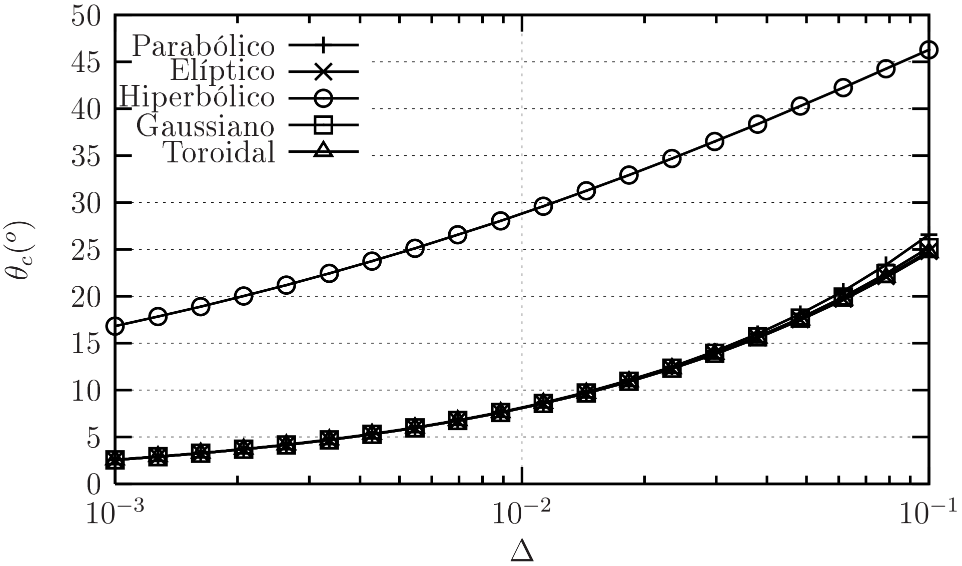


Figura 2.13: Ângulo crítico para os guias de ondas planares.

que existem outras superfícies que também poderiam satisfazer essa condição. Portanto, neste momento, vamos mostrar que todos os guias de ondas planares com índice de refração $n(x)$ podem ser representados por uma única superfície de revolução, cuja determinação pode facilitar a interpretação de alguns parâmetros de propagação da onda no meio.

Para tanto, considere que uma superfície de revolução possa ser parametrizada da seguinte maneira

$$X(x,z) = \frac{1}{a}\left[f(ax)\cos(az), f(ax)\sin(az), g(ax)\right], \tag{2.27}$$

onde $a = \sqrt{2\Delta}/\rho$ e as funções diferenciáveis $f(ax)$ e $g(ax)$ determinam uma parametrização para a geratriz dessa superfície. Com isso, um guia de ondas não necessariamente isotrópico associado à superfície parametrizada (2.27) possui os seguintes índices de refração principais

$$n_1^2 = \left(\frac{\partial X}{\partial x}\cdot\frac{\partial X}{\partial x}\right) = f_x^2 + g_x^2 \quad \text{e} \quad n_2^2 = \left(\frac{\partial X}{\partial z}\cdot\frac{\partial X}{\partial z}\right) = f^2, \tag{2.28}$$

onde $f_x = \frac{1}{a}\frac{\partial f(ax)}{\partial x}$ e $g_x = \frac{1}{a}\frac{\partial g(ax)}{\partial x}$.

Dessa forma, as trajetórias das ondas nesse meio óptico ou, equivalentemente, as geodésicas da superfície de revolução (2.27) devem satisfazer, de (1.20), o seguinte sistema de equações diferenciais parciais de segunda ordem

$$z'' + 2\frac{f_x}{f}z'x' = 0\,, \tag{2.29a}$$

$$x'' - \frac{ff_x}{f_x^2+g_x^2}(z')^2 + \frac{f_xf_{xx}+g_xg_{xx}}{f_x^2+g_x^2}(x')^2 = 0\,. \tag{2.29b}$$

Além disso, quando a geodésica está parametrizada pelo comprimento de arco, $(x')^2(f_x^2+g_x^2)+(z')^2f^2=1$, uma solução para o sistema (2.29) é

$$z'=\frac{\beta}{f^2} \quad \text{e} \quad (x')^2=\left(1-\frac{\beta^2}{f^2}\right)\frac{1}{f_x^2+g_x^2}, \tag{2.30}$$

onde $\beta=z'(0)f^2(x(0))$ será convenientemente denominado invariante de onda, pois podemos assumir $x(0)=0$, $f(0)=n_{co}$ e tomar $x'(0)=\sin(\theta_0)/\sqrt{f_x^2(0)+g_x^2(0)}$ e $z'(0)=\cos(\theta_0)/f(0)$ de maneira a reobter

$$\beta=z'(0)f^2(0)=\frac{\cos(\theta_0)}{f(0)}f^2(0)=n_{co}\cos(\theta_0),$$

onde θ_0 é ângulo de incidência em relação ao eixo x da onda no guia.

Com isso, usando a solução (2.30), podemos determinar os principais parâmetros de propagação da onda no guia anisotrópico, a saber: o ponto de giro, obtido quando $x'=0$,

$$\beta=f(x_{tp}),$$

o meio período de onda

$$z_p=\int_{-x_{tp}}^{x_{tp}}\frac{z'}{x'}\,dx=\beta\int_{-x_{tp}}^{x_{tp}}\frac{\sqrt{f_x^2+g_x^2}}{f\sqrt{f^2-\beta^2}}\,dx,$$

o comprimento Euclidiano de meio período de onda

$$L_p=\int_{-x_{tp}}^{x_{tp}}\sqrt{1-\left(\frac{z'}{x'}\right)^2}\,dx=\int_{-x_{tp}}^{x_{tp}}\sqrt{\frac{f^4+\beta^2(f_x^2+g_x^2-f^2)}{f^2(f^2-\beta^2)}}\,dx,$$

e o comprimento óptico de meio período de onda

$$L_o=\int_{-x_{tp}}^{x_{tp}}\sqrt{f_x^2+g_x^2-f^2\left(\frac{z'}{x'}\right)^2}\,dx=\int_{-x_{tp}}^{x_{tp}}\sqrt{\frac{f^2(f_x^2+g_x^2)}{(f^2-\beta^2)}}\,dx,$$

Além disso, podemos obter, com o auxílio de (1.22), que a curvatura Gaussiana de tal meio óptico é dada por

$$K=\frac{g_x(f_xg_{xx}-g_xf_{xx})}{f(f_x^2+g_x^2)^2}.$$

Observe que, quando o meio óptico é considerado isotrópico, devemos assumir, de (2.28), que

$$n_1=n_2 \quad \Rightarrow \quad n=f=\sqrt{f_x^2+g_x^2},$$

e, neste caso, essas fórmulas apresentadas para o caso anisotrópico coincidem com as expressões dos parâmetros de propagação da onda apresentadas anteriormente para um guia de ondas planar isotrópico.

Neste momento, vamos usar a parametrização (2.27) para obter as superfícies de revolução associadas aos perfis parabólico, Gaussiano e toroidal. Assim, a geratriz do guia de ondas parabólico pode ser calculada por

$$f(ax) = n(x) = n_{co}\sqrt{1-(ax)^2} \quad \text{e}$$

$$g(ax) = a\int_0^{ax} \sqrt{f^2 - f_x^2}\, dx = a\, n_{co}\int_0^{ax} \sqrt{\frac{1-3(ax)^2+(ax)^4}{1-(ax)^2}} dx,$$

onde $a = \sqrt{2\Delta}/\rho$ e $x \in (0, \frac{0.618}{a})$. Como não foi possível obter uma solução analítica para a função $g(ax)$ no guia com perfil parabólico, uma aproximação numérica foi calculada computacionalmente e apresentada na Figura 2.14.

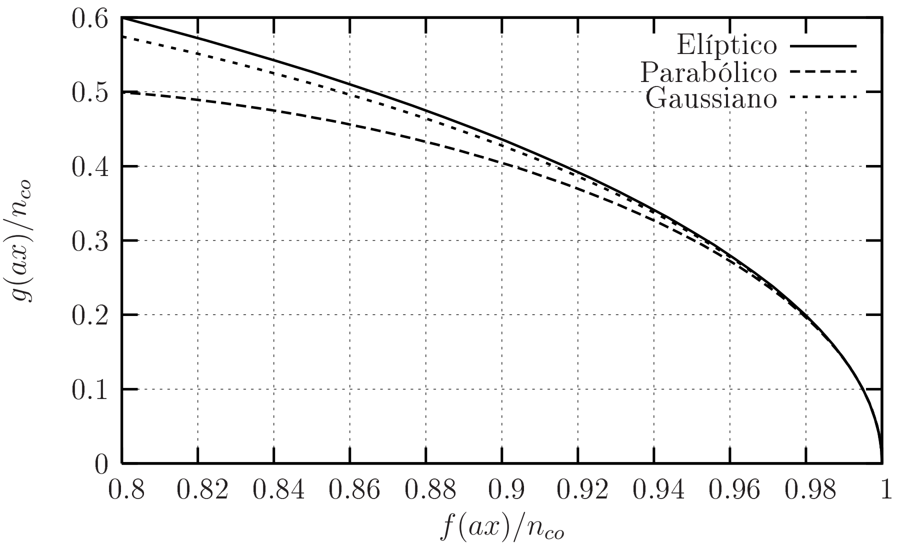


Figura 2.14: Geratrizes para os guias de ondas elíptico, parabólico e Gaussiano.

Além disso, de forma similar, obtemos que o guia de ondas com perfil Gaussiano possui a seguinte geratriz

$$f(ax) = n_{co}e^{-\frac{(ax)^2}{2}} \quad \text{e} \quad g(ax) = a\, n_{co}\int_0^{ax} e^{-\frac{(ax)^2}{2}}\sqrt{1-(ax)^2}dx,$$

onde $a = \sqrt{2\Delta}/\rho$ e $x \in (0, \frac{1}{a})$. Com isso, a geratriz da superfície associada a esse guia de ondas foi determinada numericamente e apresentada na Figura 2.14. Portanto, observe que as geratrizes dos guias parabólico e Gaussiano possuem formas semelhantes à de uma elipse, quando comparadas à geratriz do guia elíptico, que é um círculo.

Para o guia de ondas com perfil toroidal, temos a seguinte geratriz,

$$f(x) = \frac{n_{co}}{1+\Delta(x/\rho)^2} \quad \text{e} \quad g(x) = \frac{n_{co}\sqrt{\Delta}(x/\rho)}{1+\Delta(x/\rho)^2}.$$

Assim, com o auxílio da Figura 2.15 notamos que a geratriz do guia toroidal é um círculo transladado da origem, gerando a superfície de um toro e, dessa forma, explicando a motivação do nome guia de ondas toroidal.

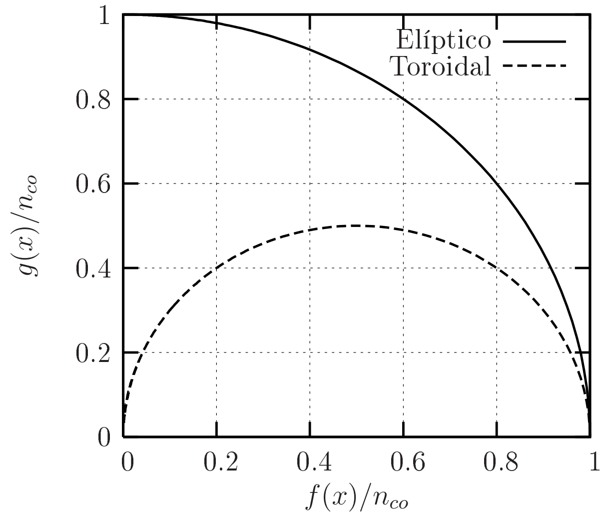


Figura 2.15: Geratrizes para os guias elíptico e toroidal, com $\Delta = 1/2$ e $\rho = 1$.

2.3.8 Guias de ondas planares com dispersão nula em meios anisotrópicos

Como descrito anteriormente, os guias de ondas planares em meios anisotrópicos ou isotrópicos podem ser localmente associados a superfícies de revolução com a parametrização apresentada em (2.27). Certamente, tal interpretação geométrica dos guias de ondas planares facilita sua análise e projeto. Portanto, proporemos novos perfis de guias de ondas em meios anisotrópicos com dispersão modal nula.

Para tanto, a superfície da esfera será novamente associada ao guia de ondas com dispersão nula, pois nesta superfície as geodésicas que partem de um ponto qualquer retornam a este ponto após um mesmo tempo. Neste caso, podemos determinar que uma parametrização adequada ao projeto de guias de ondas para a superfície da esfera na forma (2.27) é dada pelas seguintes funções

$$f(ax) = n_{co}\cos(ax) \quad \text{e} \quad g(ax) = n_{co}\sin(ax)\,,$$

onde $a = \sqrt{2\Delta}/\rho$. Com isso, o meio óptico associado a essa superfície possui, de (2.28), os seguintes índices de refração principais

$$n_1 = n_{co} \quad \text{e} \quad n_2 = n_{co}\cos\left(\sqrt{2\Delta}\frac{x}{\rho}\right). \tag{2.31}$$

Além disso, como as geodésicas de uma esfera são círculos de raio máximo, então a trajetória das ondas nesse meio óptico são dadas por

$$\tan(\sqrt{2\Delta}\frac{x}{\rho}) = \sin(\sqrt{2\Delta}\frac{y}{\rho})\tan(\theta_0)\,,$$

onde θ_0 é o ângulo de incidência do raio em relação ao eixo z. Veja exemplos dessas trajetórias na Figura 2.16. Neste caso, os valores do meio período de onda, do comprimento óptico e do tempo transiente,

$$z_p = \frac{\pi\rho}{\sqrt{2\Delta}}, \quad L_o = n_{co} z_p \quad \text{e} \quad t = \frac{z}{c} n_{co}\,,$$

respectivamente, coincidem com os parâmetros do perfil secante hiperbólica. Contudo, o valor do ponto d giro x_{tp} difere, isto é,

$$x_{tp} = \frac{\theta_0 \rho}{\sqrt{2\Delta}}\,.$$

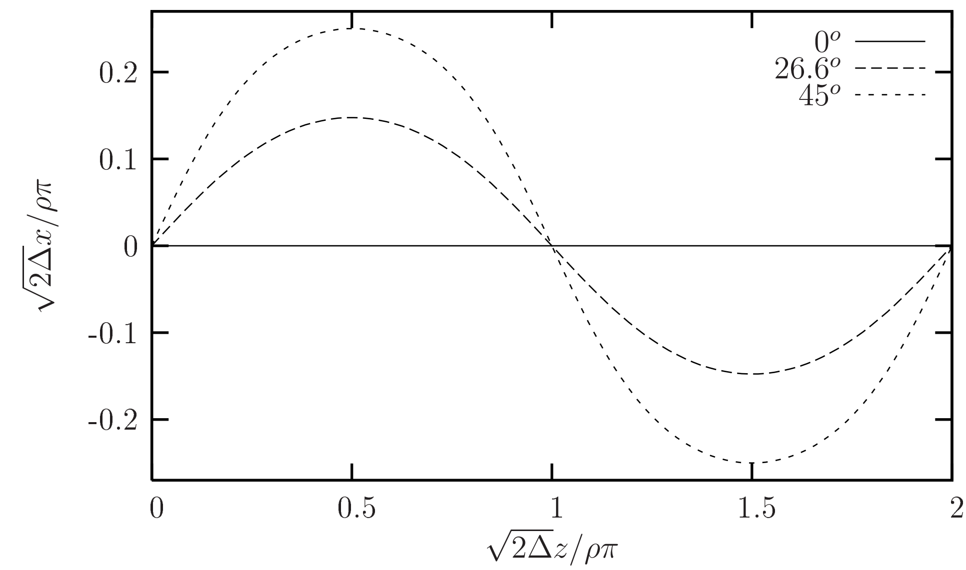


Figura 2.16: Guia de ondas planar anisotrópico com curvatura constante positiva e dispersão nula.

Uma outra parametrização possível para esse guia de ondas planar com curvatura constante é caracterizada pelas seguintes funções

$$f(ax) = n_{co}\sqrt{1-(ax)^2} \quad \text{e} \quad g(ax) = n_{co} ax,$$

cujos índices de refração principais são dados por

$$n_1^2 = \frac{n_{co}^2}{1-(ax)^2} \quad \text{e} \quad n_2^2 = n_{co}^2\left[1-(ax)^2\right]). \tag{2.32}$$

Neste caso, observe a semelhança do índice de refração $n_2(x)$ com o índice de refração do guia de ondas parabólico e que o valor n_{co} não é o máximo valor de $n_1(x)$, como geralmente ocorre para que a luz seja confinada no guia de ondas. Portanto, como ambos os meios anisotrópicos caracterizados por (2.31) e por (2.32) possuem dispersão nula, então a escolha de um deles para fins práticos deve ser determinada pelo nível de complexidade, custo, etc. envolvido em sua fase de construção.

2.4 Análise da Curvatura Seccional de Fibras Ópticas

As fibras ópticas são casos particulares de guias de ondas em três dimensões. Geralmente, o perfil ou o índice de refração das fibras ópticas varia apenas nos eixos $\overrightarrow{ox}$ e $\overrightarrow{oy}$, sendo $\overrightarrow{oz}$ o eixo de propagação. Além disso, os parâmetros de propagação da onda nesses meios ópticos são descritos em coordenas cilíndricas, devido à simetria azimutal que normalmente é assumida.

Nesta fase inicial do estudo da influência da curvatura seccional em fibras ópticas, será considerado que o meio seja isotrópico e que o índice de refração só depende de x e y, isto é, o eixo z é o de propagação. Assim, no intuito de projetar fibras ópticas com dispersão nula, podemos usar o fato de que o guia de ondas planar com dispersão nula foi associado à superfície de uma esfera para afirmar que as fibras ópticas com curvatura constante e associadas a uma hiperesfera no $\mathbb{R}^4$ também possuem dispersão nula. Para comprovar este fato, podemos usar a propriedade de que as geodésicas dessa hipersuperfície serem círculos de raio máximo, e, portando, as geodésicas que partem de um ponto qualquer dessa hipersuperfície retornam a esse ponto após percorrerem a mesma distância, ou seja, após o mesmo tempo.

Entretanto, a pergunta que devemos fazer é: *é possível encontrar uma parametrização para uma hiperesfera cuja métrica ou, equivalentemente, o índice de refração dependa apenas das variáveis x e y?* Para responder a essa pergunta vamos reescrever a expressão (2.6b), para $n = n(x, y)$, da seguinte maneira

$$K_{xy} = \frac{1}{n^4}\left[n_x^2 + n_y^2 - n\left(n_{xx} + n_{yy}\right)\right] , \tag{2.33a}$$

$$K_{xz} = \frac{1}{n^4}\left[n_x^2 - n_y^2 - nn_{xx}\right] , \tag{2.33b}$$

$$K_{yz} = \frac{1}{n^4}\left[n_y^2 - n_x^2 - nn_{yy}\right] . \tag{2.33c}$$

Assim, se fizermos a equação (2.33a) menos as equações (2.33b) e (2.33c) podemos obter a seguinte relação

$$K_{xy} - K_{xz} - K_{yz} = \frac{1}{n^4}\left[n_x^2 + n_y^2\right] . \tag{2.34}$$

Neste caso, observe que o lado direito de (2.34) é sempre não negativo e isso significa que $K_{xy} - K_{xz} - K_{yz} \geq 0$. Por isso, não existe uma fibra óptica com índice de refração

$n(x, y)$ de maneira que suas curvaturas seccionais sejam constantes, positivas e iguais, isto é, associada a uma hiperesfera. Portanto, podemos concluir que não existe uma fibra óptica em um meio óptico isotrópico com índice de refração $n(x, y)$ e dispersão modal nula. Contudo, caso o índice de refração possa variar com x, y e z, então existe meio óptico isotrópico associado à hiperesfera, porém tal meio não possui características de um guia de ondas, mas sim de uma lente óptica, como será mostrado na próxima seção.

Dado que a maioria das fibras ópticas de interesse prático possui simetria azimutal com índice de refração $n(x, y) = n(r)$, onde $r = \sqrt{x^2 + y^2}$, então devemos reescrever a equação da curvatura seccional em coordenadas cilíndricas como

$$K_{xy} = \frac{1}{n^4}\left[n_r^2 - n\left(\frac{n_r}{r} + n_{rr}\right)\right], \tag{2.35a}$$

$$K_{xz} = \frac{1}{r^2 n^4}\left[x^2\left(n_r^2 - n n_{rr}\right) - y^2\left(n_r^2 + \frac{n n_r}{r}\right)\right], \tag{2.35b}$$

$$K_{yz} = \frac{1}{r^2 n^4}\left[y^2\left(n_r^2 - n n_{rr}\right) - x^2\left(n_r^2 + \frac{n n_r}{r}\right)\right]. \tag{2.35c}$$

Dessa maneira, utilizando a expressão da curvatura seccional (2.35), podemos verificar que uma fibra parabólica com índice de refração

$$n(r) = n_{co}\sqrt{1 - 2\Delta\left(\frac{r}{\rho}\right)^2},$$

possui a seguinte curvatura seccional

$$K_{xy} = \frac{4\Delta n_{co}^4}{\rho^2 n^6(r)}, \quad K_{xz} = \frac{K_{xy}}{2}\left[1 - 2\Delta\frac{2y^2 - x^2}{\rho^2}\right] \quad \text{e} \quad K_{yz} = \frac{K_{xy}}{2}\left[1 - 2\Delta\frac{2x^2 - y^2}{\rho^2}\right].$$

Além disso, propomos uma fibra óptica cujo índice de refração foi baseado no perfil de um guia de ondas toroidal. Tal fibra óptica possui o seguinte índice de refração

$$n(r) = \frac{n_{co}}{1 + \Delta\left(\frac{r}{\rho}\right)^2},$$

e a seguinte curvatura seccional

$$K_{xy} = \frac{4\Delta}{\rho^2 n_{co}^2} \quad \text{e} \quad K_{xz} = K_{yz} = \frac{K_{xy}}{2}\left[1 - \frac{\Delta}{2}\left(\frac{r}{\rho}\right)^2\right].$$

Analisando a curvatura dessas duas fibras ópticas apresentadas anteriormente, podemos notar que ambas tendem a zerar o lado esquerdo de (2.34), pois possuem a seguinte propriedade

$$K_{xz} \approx K_{yz} \approx \frac{K_{xy}}{2}.$$

Além disso, note que a curvatura seccional da fibra parabólica não apresenta simetria radial e possui uma variação maior do que a curvatura seccional da fibra toroidal.

Este último fato, como vimos, está associado à dispersão modal. Por esse motivo, a fibra óptica com perfil parabólico possui dispersão maior que a da fibra óptica com perfil toroidal. Contudo, não será necessário determinar as curvas de dispersão dessas fibras ópticas, pois elas coincidem com as curvas de dispersão apresentadas na Figura 2.9, para o caso em que os guias de ondas são planares.

Neste momento, se considerarmos que o meio óptico possa ser anisotrópico, podemos obter algumas fibras com dispersão modal nula. Isto é, a fibra óptica pode ser associada a uma hiperesfera parametrizada por

$$X(x,y,z) = \left[\frac{2(x+y)\cos(z)}{2+x^2+y^2}, \frac{2(x+y)\sin(z)}{2+x^2+y^2}, \frac{2(x-y)}{2+x^2+y^2}, \frac{2(x^2+y^2)}{2+x^2+y^2}\right],$$

cujos índices de refração principais, obtidos da métrica da hipersuperfície, são dados por

$$n_1^2 = n_2^2 = \frac{8}{(2+x^2+y^2)^2} \quad \text{e} \quad n_3^2 = 8\frac{(x+y)^2}{(2+x^2+y^2)^2},$$

ou pela parametrização

$$X(x,y,z) = [\cos(x)\cos(y)\cos(z)\ \cos(x)\cos(y)\sin(z)\,, \cos(x)\sin(y)\,, \sin(x)]\ ,$$

cujos índices de refração principais são dados por

$$n_1^2 = 1, \quad n_2^2 = \cos^2(x) \quad \text{e} \quad n_3^2 = \cos^2(x)\cos^2(y).$$

2.5 Análise da Curvatura de Lentes Ópticas

Uma lente é um dispositivo óptico com simetria axial perfeita ou aproximada que transmite e refrata a luz, podendo fazer com que a luz convirja ou divirja. Normalmente, as lentes são compostas de materiais que refratam ondas eletromagnéticas, como vidro ou plástico transparente. A maioria das lentes de uso prático são ditas esféricas, pois suas superfícies externas são formadas por partes de esferas, de mesmos raios ou não. Além disso, dependendo do sinal da curvatura da superfície externa, essa superfície pode ser chamada de convexa, côncava ou planar, como ilustram os diferentes tipos de lentes apresentados na Figura 2.17. Geralmente, as lentes apresentadas na Figura 2.17 possuem índices de refração constante. Contudo, as lentes consideradas neste trabalho possuem índice de refração gradual e são simples, isto é, são lentes consistindo de um único elemento óptico. Mais informações sobre lentes ópticas podem ser obtidas em [21].

Em geral, as imagens obtidas por lentes não são perfeitas, pois existe sempre um certo grau de distorção ou aberração introduzida pela lente, que faz com que a imagem recebida seja uma réplica imperfeita do objeto real. Por essa razão, o projeto de lentes ópticas, dependendo da aplicação, deve ser realizado de maneira que a distorção

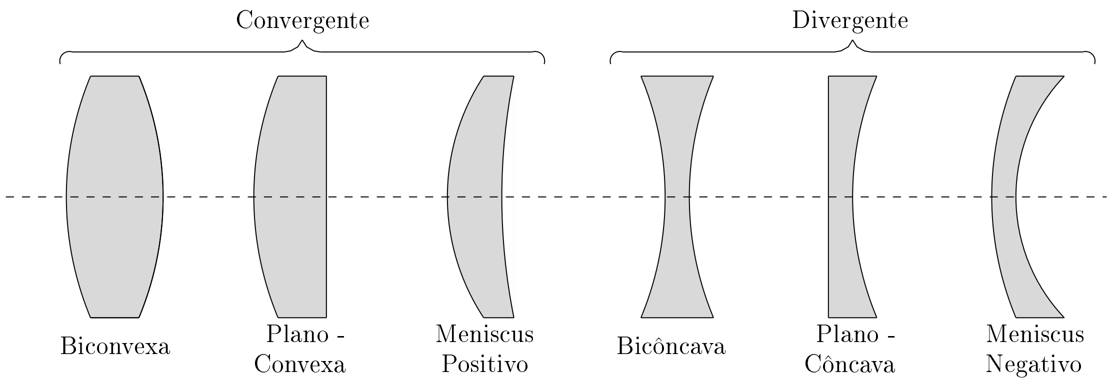
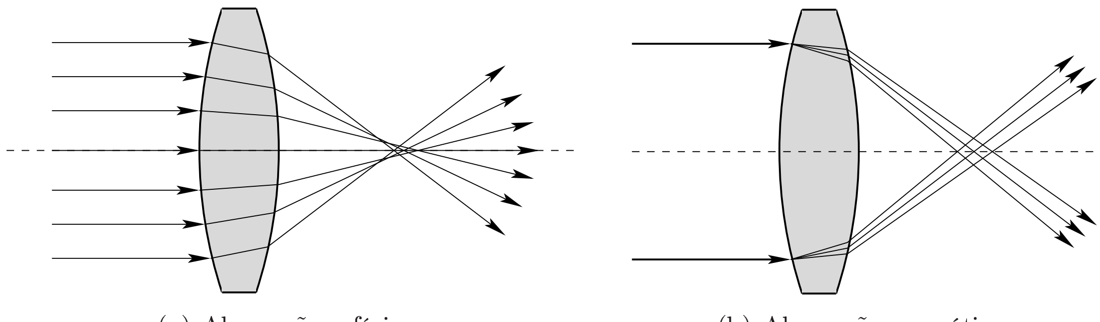


Figura 2.17: Modelos de lentes esféricas em meios ópticos com índice de refração constante.

seja minimizada. Existem vários tipos de distorção que podem afetar a qualidade da imagem, como a aberração esférica e a aberração cromática. A aberração esférica apresentada na Figura 2.18(a) ocorre quando as curvaturas das superfícies esféricas não são ideais, de maneira que a distância do cruzamento dos raios no eixo da lente não são iguais, isto é, não existe um ponto de foco. A aberração cromática apresentada na Figura 2.18(b) ocorre em lentes cujo índice de refração varia para os diferentes valores do comprimento de onda da onda luminosa. Entretanto, neste trabalho analisaremos apenas a influência da curvatura seccional na aberração esférica, desprezando assim a existência de outras distorções.



(a) Aberração esférica.

(b) Aberração cromática.

Figura 2.18: Exemplos de distorções em lentes ópticas.

2.5.1 Instrumentos ópticos absolutos

Nesta subseção consideraremos uma classe especial de lentes, os instrumentos ópticos absolutos descritos em [21]. Nosso interesse por esses instrumentos provém do fato de um instrumento óptico absoluto conhecido como *'fish-eye'*, que foi primeiramente investigado por Maxwell em 1854, possuir curvatura seccional constante positiva.

Considerando um meio óptico como o apresentado na Figura 2.19, note que um número infinito de raios partem de P_0, mas, em geral, somente um número finito deles irá atravessar o instrumento $\mathcal{I}$. Entretanto, em alguns casos especiais, podem passar um número infinito de raios pelo ponto P_1. Neste caso, o ponto P_1 é chamado de imagem de P_0 e ambos são ditos conjugados. Assim, quando P_0 descreve uma curva C_0 no objeto, então P_1 irá descrever uma curva conjugada C_1 na imagem. Geralmente, essas duas curvas não são geometricamente similares. Contudo, caso qualquer curva C_0 no objeto seja geometricamente similar a sua imagem, então $\mathcal{I}$ é um instrumento absoluto.

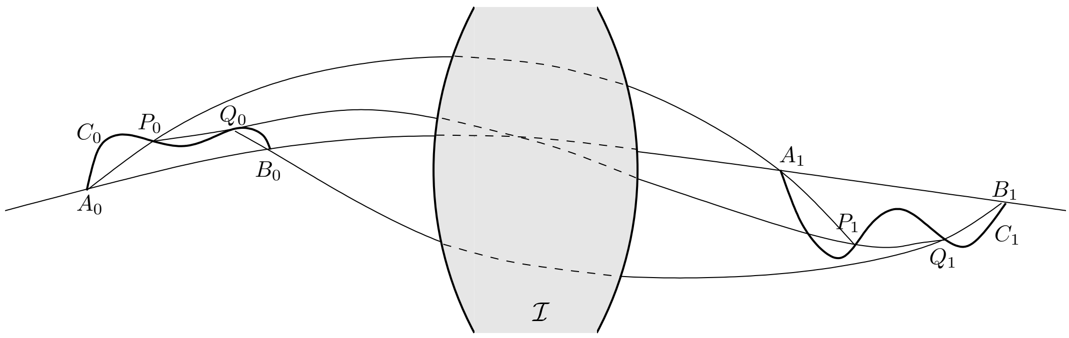


Figura 2.19: Um exemplo de instrumento óptico absoluto.

Neste caso, segundo [21], em um instrumento óptico absoluto o comprimento óptico de qualquer curva no objeto é igual ao comprimento óptico de sua imagem. Esta propriedade foi mostrada no teorema de Maxwell para instrumentos absolutos e, no caso isotrópico, é dada por

$$\int_{C_0} n_0 \, ds_0 = \int_{C_1} n_1 \, ds_1 \, . \tag{2.36}$$

Do teorema de Maxwell a seguinte conclusão pode ser imediatamente obtida. Considere um triângulo pequeno, cujos lados possuem comprimentos $ds_0^{(1)}$, $ds_0^{(2)}$, $ds_0^{(3)}$ e sejam $ds_1^{(1)}$, $ds_1^{(2)}$, $ds_1^{(3)}$ os comprimentos dos lados de sua imagem formada pelo instrumento absoluto. Assim, caso n_0 e n_1 sejam os índices de refração das regiões onde esses triângulos estejam situados, então pelo teorema de Maxwell tem-se que

$$n_0 ds_0^{(1)} = n_1 ds_1^{(1)} \quad n_0 ds_0^{(2)} = n_1 ds_1^{(2)} \quad \text{e} \quad n_0 ds_0^{(3)} = n_1 ds_1^{(3)} \, . \tag{2.37}$$

Com isso, os dois triângulos são similares, pois o comprimento óptico de seus lados e o ângulo formado entre eles no objeto são preservados na imagem. Neste caso, quando a métrica é igual ao índice de refração ao quadrado, o instrumento absoluto realiza uma transformação isométrica, mas, quando o comprimento Euclidiano é considerado, o instrumento absoluto realiza uma transformação conforme.

Além disso, verificamos que as propriedades (2.36) e (2.37) satisfeitas pelos instrumentos ópticos absolutos coincidem com as propriedades de simetria dos espaços de curvatura seccional constante descritas em [27]. Em outras palavras, instrumentos ópticos com curvatura seccional constante são instrumentos absolutos e o uso dessa informação pode facilitar a determinação de novos instrumentos ópticos absolutos.

Neste caso, observe que o índice de refração em coordenadas esféricas que caracteriza o instrumento absoluto *'fish-eye'*

$$n(r) = \frac{n_0}{1+(r/R)^2}\,, \tag{2.38}$$

onde r é a distância à origem, n_0 é o índice de refração em $r = 0$ e R é uma constante, é um caso particular da métrica Riemanniana isotérmica

$$g = n^2 = \frac{1}{\left[1+\frac{K}{4}(x^2+y^2+z^2)\right]^2}\,, \tag{2.39}$$

de variedades Riemannianas com curvatura seccional (K) constante. Com isso, a curvatura seccional de (2.38), obtida de (2.39) ou de (2.6b), é constante e igual a $4/(n_0^2 a^2)$, isto é, tal instrumento é associado ao espaço não-Euclidiano elíptico. Além disso, tal meio óptico está associado a S^3, uma hiperesfera no $\mathbb{R}^4$, parametrizada por

$$\pi^{-1}(x,y,z) = \frac{R^2}{R^2+(x^2+y^2+z^2)}\left[2x, 2y, 2z, R - \frac{x^2+y^2+z^2}{R}\right],$$

onde $\pi(x_1, x_2, x_3, x_4) = (x, y, z)$ é a projeção estereográfica, $\pi : S^3 - \{N\} \to \mathbb{R}^3$, ilustrada na Figura 2.20 para $N = [0, 0, 0, R]$ e R é o raio de S^3.

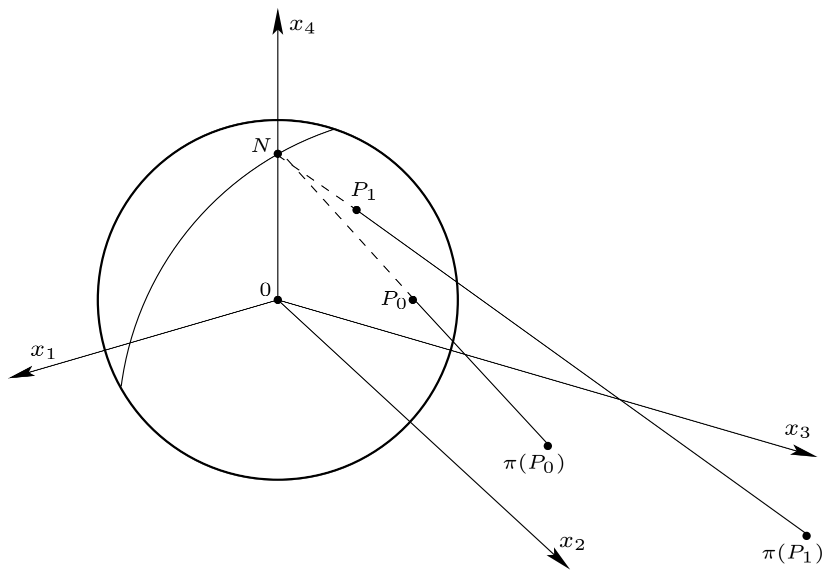


Figura 2.20: Projeção estereográfica de uma hiperesfera no $\mathbb{R}^4$.

O uso dessa informação é de grande relevância para o projeto, a identificação e a interpretação de novas aplicações para esse instrumento óptico, pois podemos usar as simetrias desse espaço ou, mais geralmente, suas transformações isométricas para melhor compreendê-lo. Em outras palavras, o estudo de um instrumento absoluto pode ser auxiliado pela compreensão do seu grupo de automorfismos, caracterizado pelos mapeamentos do espaço em si mesmo, mas preservando a estrutura geométrica essencial do espaço. No caso dos espaços de curvatura constante, esses automorfismos ou mapeamentos são descritos pela transformação de Möbius, dada por

$$f(z) = \frac{az+b}{cz+d}, \tag{2.40}$$

onde a, b, c, d são números complexos satisfazendo $ad - bc \neq 0$.

Exemplos de transformações de Möbius são dilatações, rotações, translações, inversões complexas ou podem ser escritas como composições destas. Contudo, dependendo do espaço (Euclidiano, elíptico ou hiperbólico), nem todas essas transformações são isométricas. Portanto, dependendo do caso, novas restrições devem ser impostas às constantes a, b, c e d para que o mapeamento seja isométrico, como ocorre nos instrumentos ópticos absolutos.

Segundo [21], em um meio com índice de refração com simetria esférica, as trajetórias (geodésicas) das ondas eletromagnéticas descrevem curvas planares pertencentes aos planos que passam pela origem. Tais trajetórias passando pelo ponto $P_0(r_0, \theta_0)$, em coordenadas polares, são dadas pela seguinte equação de um círculo

$$\frac{r^2 - R^2}{r\sin(\theta - \alpha)} = \frac{r_0^2 - R^2}{r\sin(\theta_0 - \alpha)},$$

onde o ângulo α é ilustrado na Figura 2.21.

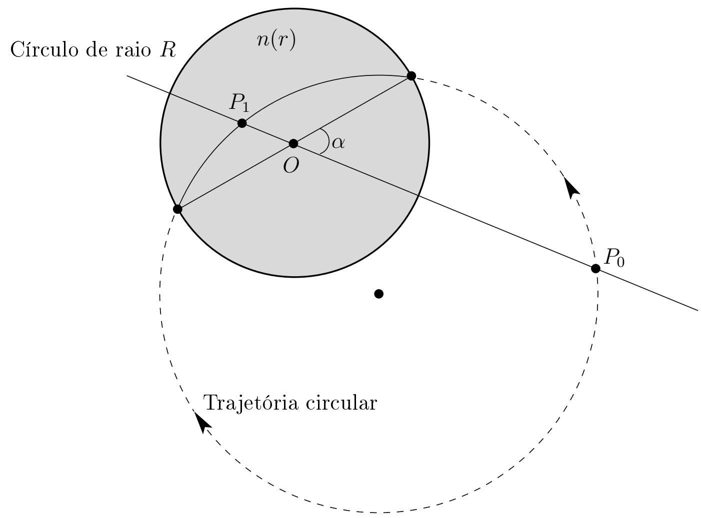
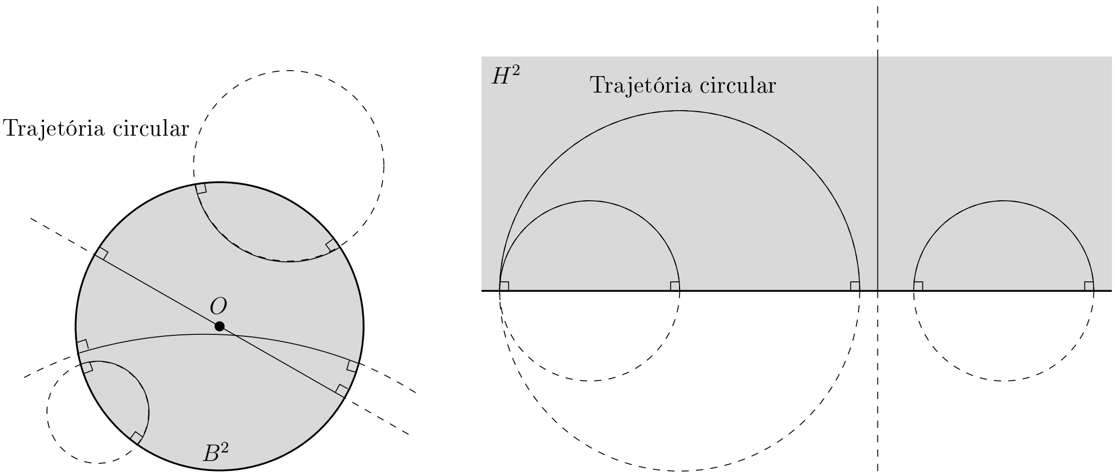


Figura 2.21: Exemplos de trajetórias no instrumento absoluto *'fish-eye'* de Maxwell.

De (2.39) podemos verificar que, além do instrumento absoluto com índice de refração constante ($K = 0$), um novo instrumento absoluto pode ser obtido quando a curvatura seccional K for constante e negativa, a saber, um instrumento absoluto associado ao espaço hiperbólico tridimensional, conhecido como modelo da bola de Poincaré, B^3. Em tal modelo, as trajetórias também são curvas planas e são descritas por retas ou círculos cruzando ortogonalmente o bordo desse espaço, como ilustra a Figura 2.22(a). Este modelo hiperbólico também está associado a uma projeção estereográfica, só que no caso de um hiperbolóide no $\mathbb{R}^4$ projetado no hiperplano $w = 0$ a partir do pólo sul $[0, 0, 0, -1]$.



(a) Modelo do disco de Poincaré. (b) Modelo do semiplano positivo.

Figura 2.22: Exemplos de trajetórias em instrumentos ópticos absolutos hiperbólicos.

Além desse instrumento absoluto hiperbólico, um outro pode ser obtido usando o modelo do semiplano hiperbólico, H^3, cujo índice de refração é dado em coordenadas cartesianas por

$$n^2(x, y, z) = g = \frac{n_0^2}{z^2}\,. \tag{2.41}$$

As geodésicas desse espaço também são curvas planas descritas por retas ou círculos cruzando ortogonalmente o bordo desse espaço, como ilustra a Figura 2.22(b). As transformações isométricas desse espaço são dadas pelas transformações de Möbius (2.40) para a, b, c e d reais satisfazendo $ad - bc > 0$.

As métricas dos espaços de curvatura seccional constante apresentadas em (2.39) e (2.41) são as mais conhecidas. Contudo, existem outras expressões para a métrica de espaços de curvatura constante e, com isso, novos instrumentos ópticos absolutos podem ser determinados por elas. Além disso, existem transformações isométricas entre os espaços com mesma curvatura, possibilitando a análise do instrumento óptico pelo modelo mais conveniente.

Além disso, a aplicação de um instrumento absoluto pode ser determinada pelo sinal de sua curvatura seccional, pois, dependendo do caso, um feixe de raios paralelos pode convergir em lentes com perfil de curvatura positiva e divergir em lentes com perfil de curvatura negativa. Tal fato será analisado nas duas próximas subseções.

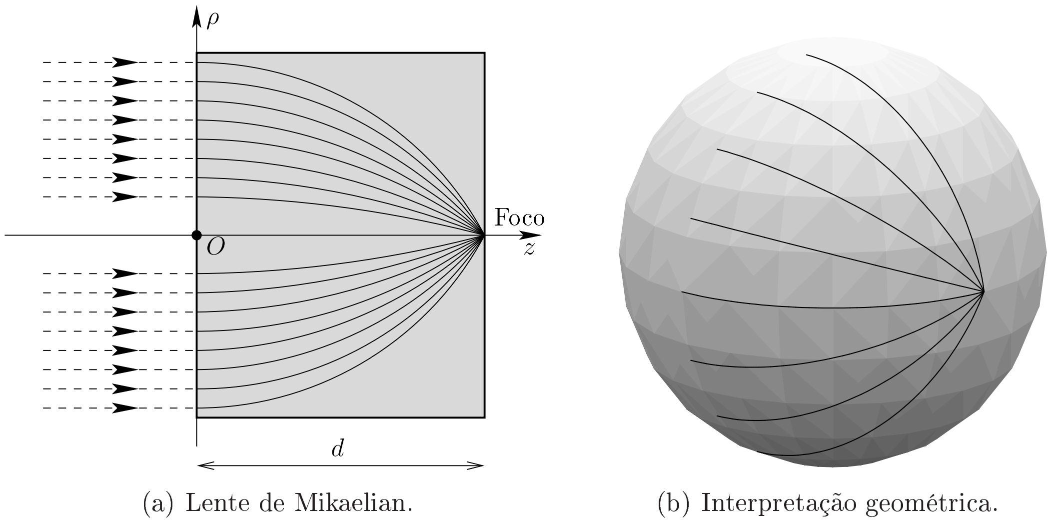
2.5.2 Exemplos de lentes com curvatura positiva

Um estudo sobre alguns dos principais tipos de lentes ópticas de uso prático foi realizado em [36]. Como exemplo, pode-se citar a lente de Mikaelian, cujo índice de refração possui simetria cilíndrica e é dado por

$$n = n_0 \operatorname{sech}\left(\frac{\pi}{2\,d}\sqrt{x^2+y^2}\right) = n_0 \operatorname{sech}\left(\frac{\pi}{2\,d}\rho\right), \tag{2.42}$$

onde n_0 é o índice de refração no eixo óptico e d é a distância da superfície de entrada da lente ao foco, conforme mostra a Figura 2.23(a). Como estamos considerando um feixe de raios paralelos incidindo ortogonalmente na superfície dessa lente, então a sua curvatura pode ser calculada em relação ao plano formado pelos eixos $\overrightarrow{o\rho}$ e $\overrightarrow{oz}$, no caso a curvatura Gaussiana $K(\rho, z)$ ou a curvatura seccional $K_{\rho z}$, que é constante, positiva e igual a

$$K(\rho, z) = \frac{\pi^2}{4n_0^2 d^2}\,.$$



(a) Lente de Mikaelian.

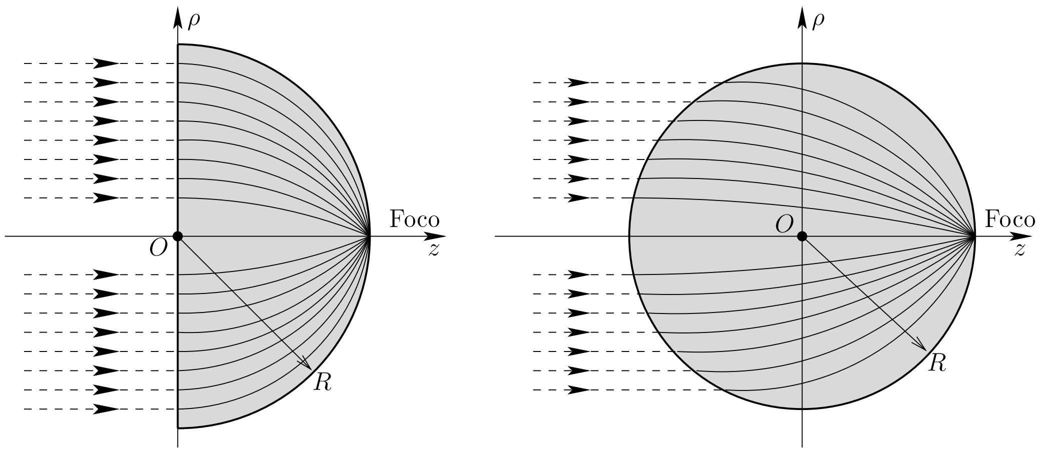
(b) Interpretação geométrica.

Figura 2.23: Exemplo de lente óptica com curvatura constante positiva.

Usando a interpretação geométrica apresentada na subseção 2.3.2, para o guia de ondas com perfil secante hiperbólica, pode-se verificar que a superfície associada ao plano ρz dessa lente é a superfície de uma esfera com raio R é igual a $1/\sqrt{K}$, veja Figura 2.23(b). Com isso, os comprimentos ópticos no interior da lente são iguais e dados por $\pi\,R/2 = d\,n_0$, como esperado, pois os raios são paralelos ao eixo da lente.

Um outro exemplo de lente com curvatura constante positiva é a lente de Maxwell '*fish-eye*'. Nessa lente, para que o foco seja atingido, normalmente, utiliza-se apenas meio hemisfério ($z > 0$), veja ilustração na Figura 2.24(a), caso contrário o foco pode não existir ou situar-se fora da lente. Além disso, devido ao fato dessa lente possuir curvatura seccional constante e positiva, então os comprimentos ópticos dos raio em seu interior serão iguais a $n_0\pi R$, isto é, um quarto do comprimento do círculo máximo de S^3. Contudo, na prática, costuma-se usar também a lente de Lüneburg, Figura 2.24(b), cujo índice de refração também possui simetria esférica e é dado por

$$n = n_0\sqrt{2 - \frac{x^2+y^2+z^2}{R^2}} = n_0\sqrt{2 - \frac{r^2}{R^2}} = n_0\sqrt{2 - \frac{\rho^2+z^2}{R^2}}.$$



(a) Lente de Maxwell '*fish-eye*'. (b) Lente de Lüneburg.

Figura 2.24: Exemplos de lentes ópticas com curvatura positiva e simetria esférica.

Como foi considerado que os raios incidentes na superfície externa da lente de Lüneburg são paralelos ao eixo da lente, $\overrightarrow{oz}$, então a curvatura de interesse prático dessa lente é dada por

$$K_{\rho z} = K(\rho, z) = \frac{4R^4}{n_0^2(2R^2 - \rho^2 - z^2)^3} = \frac{4n_0^4}{n^6 R^2}.$$

Neste caso, note que a curvatura dessa lente não é constante, mas mesmo assim foi possível convergir os raios para um único ponto. Além disso, nessa lente, os raios incidem no ponto focal com uma inclinação em relação ao eixo da lente menor do que a inclinação observada na lente '*fish-eye*'. Entretanto, a lente de Lüneburg não é um instrumento absoluto como a lente de Maxwell, que apresenta uma quantidade grande de isometrias e distorção nula.

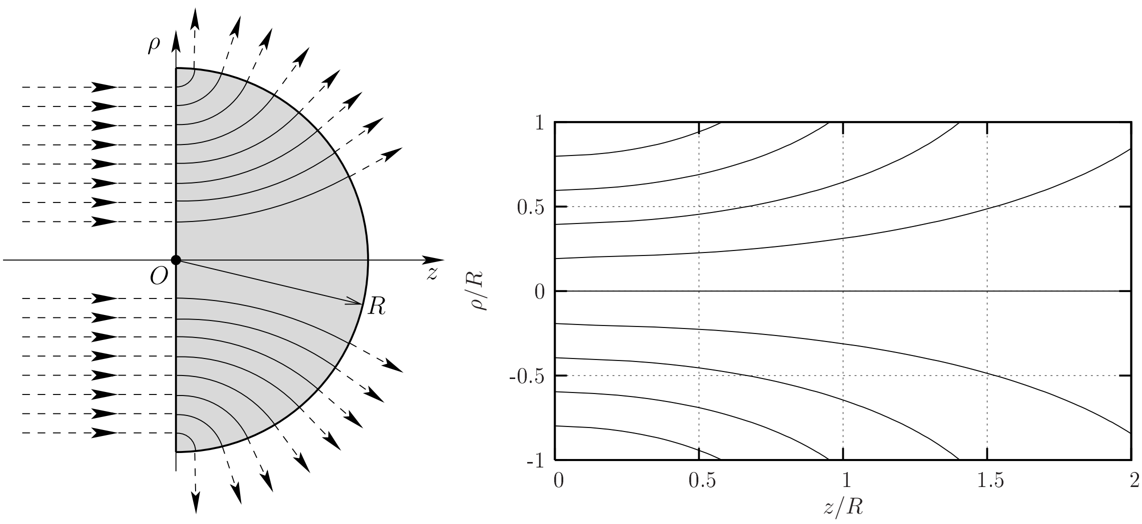
2.5.3 Exemplos de lentes com curvatura negativa

Nesta subseção analisaremos dois exemplos de lentes com curvatura negativa, cujos perfis foram baseados nos índices de refração (2.20) e (2.39). Neste caso, consideramos que o eixo da lente é o $\overrightarrow{oz}$ e que seus índices de refração sejam reescritos em coordenadas cilíndricas da seguinte forma

$$n(\rho, z) = \frac{n_0}{1 - (\rho^2 + z^2)/R^2}, \quad \text{(Lente hiperbólica esférica)}, \tag{2.43a}$$

$$n(\rho, z) = \frac{n_0}{\cos(\rho/R)}, \quad \text{(Lente hiperbólica cilíndrica)}. \tag{2.43b}$$

A lente hiperbólica esférica está associada ao modelo do disco de Poincaré, Figura 2.22(a), e, como mencionado anteriormente, suas trajetórias são retas ou círculos cruzando ortogonalmente o bordo do disco. Assim, considerando que essa lente seja composta por meio hemisfério e como o índice de refração aumenta com ρ, então um feixe de raios paralelos ao eixo da lente diverge ao passar por essa lente, veja ilustração na Figura 2.25(a).



(a) Lente hiperbólica esférica. (b) Lente hiperbólica cilíndrica.

Figura 2.25: Exemplos de lentes ópticas com curvatura negativa constante.

Além disso, verificamos que a lente hiperbólica cilíndrica (2.43b) também diverge um feixe de raios paralelos, como apresentado na Figura 2.25(b). Para comprovar este fato, obtemos que as trajetórias dos raios paralelos no interior dessa lente satisfazendo o sistema (2.13), com as condições iniciais $z(0) = \rho'(0) = 0$, $z'(0) = 1$ e $\rho(0) = \rho_0$, são dadas por

$$e^{z/R} \sin(\rho_0/R) = \sin(\rho/R) + \sqrt{\cos^2(\rho_0/R) - \cos^2(\rho/R)}\,.$$

Portanto, ao analisar as lentes descritas anteriormente podemos verificar que um feixe de raios paralelos ao eixo da lente, em geral, converge ao passar por uma lente com curvatura positiva e diverge ao passar por uma lente com curvatura negativa. Contudo, a caracterização de uma lente quanto à convergência ou divergência está relacionada, principalmente, à variação de $n(\rho, z)$ em relação a ρ, isto é, se $n_\rho < 0$ a lente converge e se $n_\rho > 0$ a lente diverge.

Entretanto, sob determinadas condições, é possível fazer com que exista convergência na lente hiperbólica cilíndrica. Neste caso, consideramos que os raio incidem na superfície lateral da lente a um ângulo θ_0 constante em relação ao eixo da lente. Assim, dependendo do valor desse ângulo de incidência, as trajetórias podem tender assintoticamente para o eixo da lente, como ilustra a Figura 2.26.

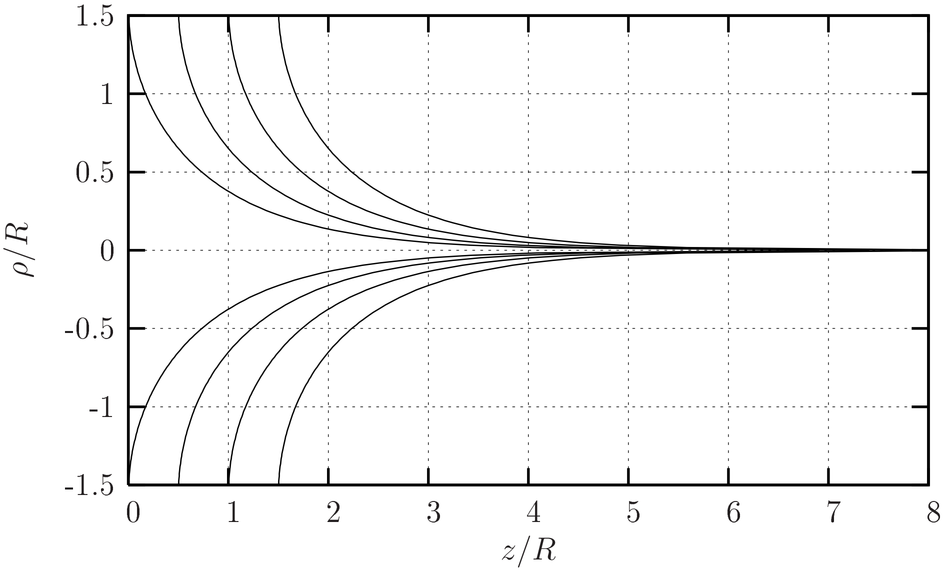


Figura 2.26: Convergência em uma lente hiperbólica cilíndrica.

Para determinarmos o ângulo de incidência correto, de maneira a se ter a convergência, devemos garantir que ρ' tenda a zero quando ρ tender a zero. Neste caso, de (2.13) podemos verificar que se $n^2 = \beta^2$ em $\rho = 0$, essa convergência ocorre. Portanto, como $n(0) = n_0$ e $\beta = n(\rho(0))z'(0) = n(\rho_0)\cos(\theta_0)$, então o ângulo de incidência deve ser igual a

$$\theta_0 = \arccos\left[\frac{n_0}{n(\rho_0)}\right] = \arccos\left[\cos(\rho_0/R)\right] = \frac{\rho_0}{R}. \tag{2.44}$$

Dessa forma, as trajetórias dos raios que incidem na lente com ângulo dado por (2.44), podem ser determinadas por (2.13) e são iguais a

$$z = z_0 + R\log\left[\frac{\sin(\rho_0/R)}{\sin(\rho/R)}\right]. \tag{2.45}$$

A Figura 2.26 apresenta algumas dessas curvas para $\rho_0 = 1.5\,R$ ($\theta_0 \approx 86°$), pois ρ_0 deve ser menor do que $\pi R/2$, senão o raio incidente não converge para o "foco". Além disso, devemos chamar a atenção para o fato de que as trajetórias dessa lente tenderem a um ponto de foco, pois ele não existe, mas sim o eixo $\overrightarrow{oz}$. Este fato é bastante interessante, dado que os raios deixam a lente quase paralelamente e em uma região muito pequena. Esta propriedade não está presente nas lentes de curvatura positiva e, por isso, a lente hiperbólica deve ser considerada em aplicações práticas que requeiram essa importante característica de convergência.

A capacidade da lente hiperbólica de concentrar paralelamente os raios em uma pequena região pode ser medida em função da distância da trajetória ao eixo $\overrightarrow{oz}$. Para tanto, podemos usar a expressão (2.45) para determinar que essa distância diminui exponencialmente com z. Por exemplo, se $z - z_0 = 6R$, então essa distância vale $0.00248R$.

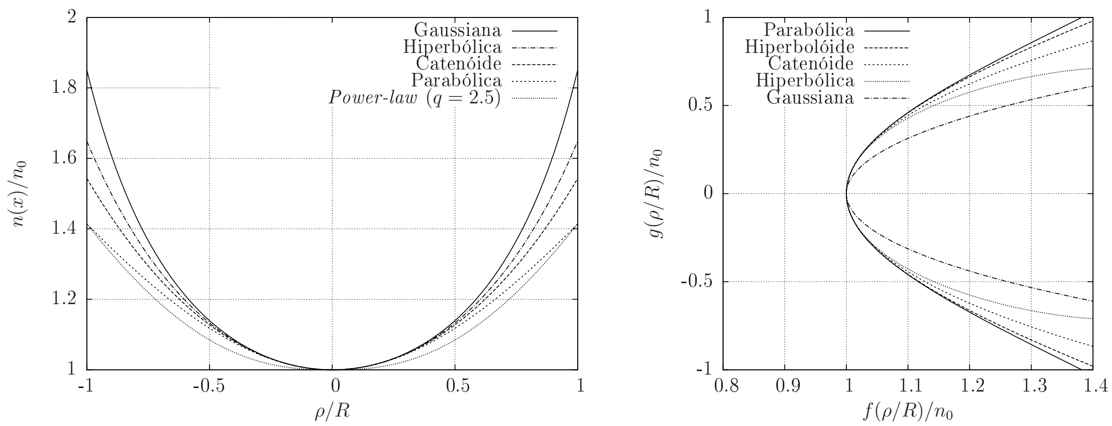
Algumas outros exemplos de perfis de lentes com curvatura negativa que apresentam convergência semelhante ao da lente hiperbólica cilíndrica são apresentados na Tabela 2.3.

Perfil	**Índice de refração ($n(\rho)$)**	**Curvatura (K)**
Hiperbólica	$\dfrac{n_0}{\cos(\rho/R)}$	$-\dfrac{1}{n_0^2R^2}$
Catenóide	$n_0\cosh(\rho/R)$	$-\dfrac{1}{n_0^2R^2\cosh^4(\rho/R)}$
Parabólica	$n_0\left[1+(\rho/R)^2\right]$	$-\dfrac{1-(\rho/R)^2}{n_0^2R^2\left[1+(\rho/R)^2\right]^3}$
Power-law	$n_0\left[1+(\rho/R)^q\right]$	$-\dfrac{q}{2}\left(\dfrac{\rho}{R}\right)^q\dfrac{q-1-(\rho/R)^2}{n_0^2\rho^2\left[1+(\rho/R)^q\right]^3}$
Gaussiana	$n_0\exp\left[\dfrac{(\rho/R)^2}{2}\right]$	$-\dfrac{1}{n_0^2R^2}\exp[-(\rho/R)^2]$

Tabela 2.3: Perfis de lentes com curvatura negativa.

As lentes descritas na Tabela 2.3, cujos índices de refração são mostrados na Figura 2.27(a), são definidas em meios ópticos isotrópicos. Dessa maneira, podemos usar a interpretação geométrica descrita na subseção 2.3.7, a qual associa esses meios ópticos a superfícies de revolução, para determinar as geratrizes associadas a essas lentes, como ilustra a Figura 2.27(b).

Dentre as lentes apresentadas na Tabela 2.3, a com perfil catenóide está associada à classe de superfícies mínimas, subseção 1.2.1, que demonstrou ser de grande interesse neste trabalho, como será mostrado nos próximos capítulos. O comportamento dos raios de luz nessa lente é apresentada na Figura 2.28(a) e sua interpretação geométrica sobre a superfície do catenóide é ilustrada na Figura 2.28(b). Neste caso, observe



(a) Índices de refração. (b) Geratrizes.

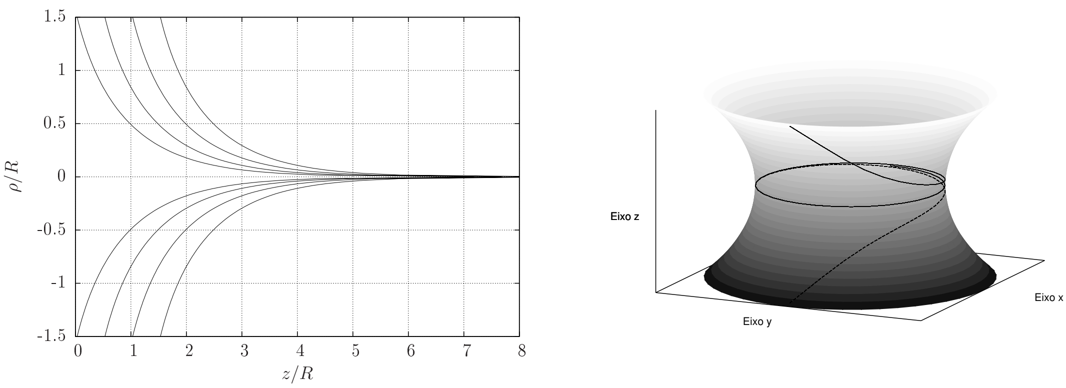
Figura 2.27: Índices de refração e geratrizes das lentes hiperbólica, Gaussiana, parabólica, catenóide e *power-law*.

novamente a convergência assintótica dos raios em torno do eixo de simetria da lente. Da equação (2.11), obtemos que as trajetórias dos raios de luz nesta lente é dada por

$$z = z_0 + 2R\,\mathrm{arctanh}\left[\exp(-|\rho/R|)\right] - 2R\,\mathrm{arctanh}\left[\exp(-|\rho_0/R|)\right] \ ,$$

quando o ângulo de incidência é igual a

$$\theta_0 = \arccos\left[\mathrm{sech}(\rho_0/R)\right] \ .$$



(a) Convergência na lente. (b) Interpretação geométrica.

Figura 2.28: Lente de curvatura negativa associada a superfície mínima catenóide.

De uma maneira semelhante, quando a lente é parabólica, veja Tabela 2.3, podemos obter que se os raios de luz incidem à superfície lateral da lente com um ângulo

$$\theta_0 = \arccos\left\{\left[1 + (\rho_0/R)^2\right]^{-1/2}\right\} ,$$

então suas trajetórias são dadas por

$$z = z_0 + R \log\left(\frac{\rho_0}{\rho}\right).$$

Analisando as lentes propostas na Tabela 2.3, verificamos que praticamente todas elas apresentam a mesma convergência, isto é, após o raio de luz percorrer um mesmo deslocamento no eixo $\overrightarrow{oz}$, a sua distância a esse eixo é aproximadamente a mesma nas lentes consideradas. Portanto, a opção por uma ou outra dessas lentes em uma determinada aplicação vai depender da complexidade de sua realização prática.

Quando o meio óptico é anisotrópico, um exemplo de lente com curvatura negativa cujos raios apresentam convergência semelhante ao da lente hiperbólica cilíndrica é a lente associada à superfície do parabolóide hiperbólico de revolução $x^2 + y^2 - z^2 = 1$, Figura 2.29, parametrizado por

$$X(x,z) = \left[\sqrt{1+x^2}\cos(z), \sqrt{1+x^2}\sin(z), x\right], \tag{2.46}$$

cuja métrica ou índices de refração principais são dados por

$$g_1 = n_1^2 = 1 + x^2 \quad \text{e} \quad g_2 = n_2^2 = \frac{1+2x^2}{1+x^2}.$$

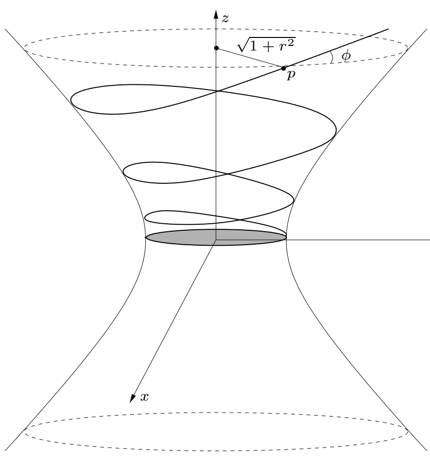


Figura 2.29: Superfície parabolóide hiperbólico de revolução associada a uma lente de curvatura negativa em um meio óptico anisotrópico.

Note que essa lente é constituída de material óptico anisotrópico, cuja curvatura Gaussiana pode ser obtida de (1.22) e vale $K = -(1+2x^2)^{-2}$. Neste caso, observe que sua curvatura é sempre negativa, porém não é constante.

Para analisar as trajetórias dos raios nessa lente, considere uma geodésica que passa pelo ponto p da parte superior de um parabolóide hiperbólico, Figura 2.29, e faz um ângulo ϕ com o paralelo passando por p tal que $\tan(\phi) = x$, onde $\sqrt{1+x^2}$ é a distância do ponto p ao eixo z. De (1.20), podemos verificar que tal geodésica deve satisfazer o seguinte sistema de equações diferenciais

$$\begin{cases} x' = -\sin(\phi)\sqrt{\dfrac{1+x^2}{1+2x^2}} = \dfrac{x}{\sqrt{1+2x^2}}\,, \\ z' = \dfrac{\cos(\phi)}{\sqrt{1+x^2}} = \dfrac{1}{1+x^2}\,, \end{cases} \tag{2.47}$$

cuja solução, quando $x(0) = x_0$ e $z(0) = 0$, Figura 2.30, é dada por

$$z = \left[\tanh^{-1}\left(\frac{1}{\sqrt{1+2x^2}}\right) - \tan^{-1}\left(\sqrt{1+2x^2}\right)\right]\Bigg|_{x_0}^{x}. \tag{2.48}$$

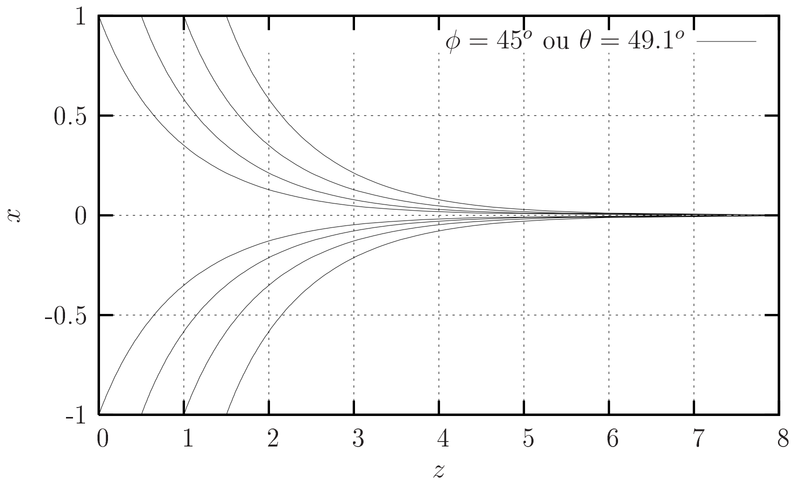


Figura 2.30: Perfil planar de uma lente com curvatura negativa em um meio óptico anisotrópico associado a superfície de um parabolóide hiperbólico.

Portanto, podemos verificar que as lentes ópticas com curvatura seccional negativa possuem propriedades de convergência de interesse prático, principalmente, quando o objetivo é convergir a luz em uma determinada direção por uma longa distância.

Capítulo 3

Análise da Curvatura de Modulações Não-Lineares

Neste capítulo usaremos uma interpretação geométrica que associa esquemas de modulação a curvas e superfícies no espaço Euclidiano para construir e analisar novos esquemas de modulação. Neste caso, as técnicas de modulação foram associadas a imersões em variedades Riemannianas. Com isso, foi possível verificar que, além do comprimento da curva e da métrica Riemanniana, a curvatura é um outro parâmetro de grande relevância no projeto de modulações, pois influencia diretamente em seus desempenhos. Também foi possível determinar algumas relações matemáticas entre parâmetros geométricos e estatísticos da modulação que, quando satisfeitos, melhoram o desempenho do sistema de comunicações. Além disso, verificamos que modulações não-lineares associadas a superfícies mínimas apresentam o melhor desempenho em relação ao erro quadrático médio, pois são pontos críticos de mínimo desse parâmetro.

3.1 Introdução

Inicialmente, as técnicas de modulação presentes na maioria dos sistemas de comunicações foram interpretadas como um método para deslocar e/ou expandir a largura de banda do sinal de informação a ser transmitido. Além disso, em alguns casos práticos é perfeitamente possível projetar modulações para melhorar o desempenho de um sistema de comunicações, seja pelo aumento da energia média de transmissão ou pelo aumento da largura de banda de transmissão. Neste caso, com o objetivo de projetar novos esquemas de modulações, [39] verificou que as técnicas de modulação podem ser interpretadas geometricamente como uma aplicação (linear ou não-linear) que transforma um sinal pertencente ao $\mathbb{R}^M$ em um sinal modulado pertencente ao $\mathbb{R}^N$, com $M < N$. Entretanto, tal interpretação não abordou completamente o problema do projeto "geométrico" de modulações e, por essa razão, tal proposta foi considerada posteriormente em [18] e [5].

Mais precisamente, em [18], as técnicas de modulação foram interpretadas geometricamente como curvas no espaço Euclidiano. Além disso, [18] verificou que o desempenho de um sistema de comunicações melhora quanto maior for o comprimento da curva associada à modulação usada nesse sistema. Tal fato contribuiu para que o projeto de novas técnicas de modulação pudesse ser realizado de maneira geométrica, ou seja, com auxílio de alguns parâmetros geométricos da curva associada à modulação, que influenciam em seu desempenho.

Dessa forma, os trabalhos [19] e [28] usaram tal interpretação geométrica para propor e analisar o desempenho de uma modulação não-linear associada à curva denominada espiral de Arquimedes uniforme. Além disso, [28] generalizou a interpretação geométrica descrita em [18] com intuito de medir o desempenho de modulações que transmitem mais do que uma variável aleatória. Neste caso, foi aplicado um procedimento similar ao usado em [18] para curvas, porém usando superfícies, para desenvolver uma expressão aproximada para o erro quadrático médio em função da métrica de uma variedade Riemanniana associada ao esquema de modulação.

As próximas seções fazem uso das definições e considerações descritas anteriormente, principalmente, da interpretação geométrica proposta em [18], para realizar análises de desempenho de algumas técnicas de modulação conhecidas e outras propostas neste capítulo, bem como determinar os principais parâmetros geométricos de desempenho das técnicas de modulação. Neste caso, consideraremos também que o canal de transmissão seja sempre perturbado por um ruído Gaussiano aditivo branco (AWGN) com densidade espectral de potência $\mathcal{N}_0/2$ por dimensão de transmissão. Além disso, assumiremos que todas as variáveis aleatórias a serem transmitidas nos sistemas de comunicações tratados pertencem ao intervalo $[-1, 1]$.

Por questões didáticas, inicialmente, foram analisadas modulações lineares e não-lineares associadas a curvas, isto é, apenas o caso particular da transmissão de uma única variável aleatória. Posteriormente, foram calculados os desempenhos de modulações não-lineares que transmitem mais de uma variável aleatória, no caso, modulações associadas a superfícies.

3.2 Modulação Linear

A modulação linear foi a primeira técnica de modulação a ser usada na prática e, certamente, é a mais simples de ser entendida e utilizada. Por esse motivo, ela será a primeira técnica de modulação a ser considerada neste capítulo. Nesse esquema de modulação, a informação é transmitida pela variação linear da amplitude da forma de onda $\varphi_1(t)$ em função da entrada m, como ilustra a Figura 3.1. Essa importante classe de modulação inclui os sistemas de modulação em amplitude (AM) e suas derivações com portadora suprimida (SC), com faixa lateral única (SSB) e com faixa lateral dupla e portadora suprimida (DSB-SC). Uma abordagem sobre a modulação linear, mais detalhada do que a apresentada aqui, pode ser obtida consultando [37] e [3].

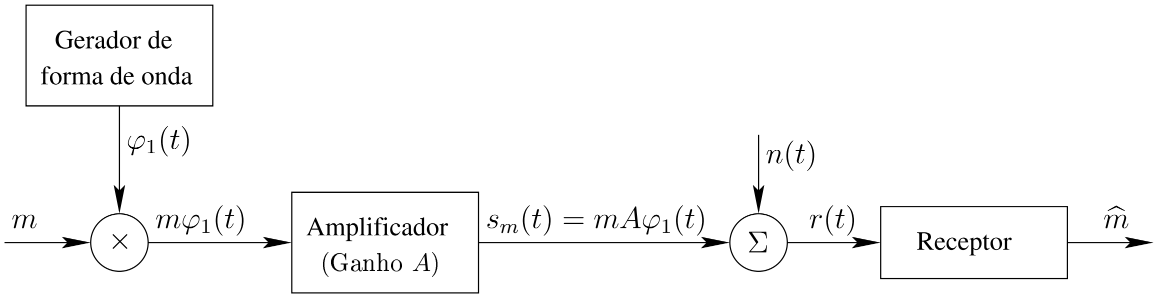


Figura 3.1: Diagrama de bloco de um sistema de comunicações usando modulação linear.

No sistema de comunicações usando modulação linear da Figura 3.1, a forma de onda transmitida (ou o sinal modulado) é dada por

$$s_m(t) = A\,m\varphi_1(t) \quad \text{ou} \quad \mathbf{s}_m = A\,m\boldsymbol{\varphi}_1, \tag{3.1}$$

onde A é o ganho em amplitude do transmissor e $\varphi_1(t)$ é uma forma de onda com energia unitária que caracteriza o sinal da portadora de transmissão.

A Figura 3.2 apresenta uma interpretação geométrica para o processo de recepção de um sistema de comunicações usando uma modulação linear e um receptor de máxima verossimilhança que, no caso particular de um ruído AWGN (representado pelo vetor $\mathbf{n}$), calcula $\widehat{m}$ como sendo o valor que minimiza $|\mathbf{r} - \mathbf{s}_{\hat{m}}|$. Observe que tal modulação pode ser interpretada, geometricamente, como uma curva no espaço Euclidiano, no caso a reta associada ao eixo x variando de $-A$ a A. Além disso, na recepção, a obtenção da estimativa $\hat{m}$ está relacionada a uma projeção do sinal recebido $\mathbf{r}$ sobre a curva associada a modulação, isto porque

$$\min_{\hat{m}} |\mathbf{r} - \mathbf{s}_{\hat{m}}| \quad \Rightarrow \quad (\mathbf{r} - \mathbf{s}_{\hat{m}}) \cdot \mathbf{s}'_{\hat{m}} = 0 \quad \Rightarrow \quad (\mathbf{r} - A\hat{m}\boldsymbol{\varphi}_1) \cdot \boldsymbol{\varphi}_1 = 0 \quad \Rightarrow \quad \hat{m} = \frac{\mathbf{r} \cdot \boldsymbol{\varphi}_1}{A}. \tag{3.2}$$

Com o intuito de comparar o desempenho da modulação linear com o desempenho das modulações não-lineares propostas no presente capítulo, neste momento serão apresentados alguns parâmetros de desempenho da modulação linear que foram obtidos de acordo com [18]. Neste caso, quando a função densidade de probabilidade *a priori* da fonte de informação for uniforme e independente do ruído do canal, pode-se obter, com o auxílio de (1.37), (1.40), (1.41), (3.2) e lembrando que $\mathbf{r} = \mathbf{s}_0 + \mathbf{n}$, que

$$\hat{m} - m_0 = \frac{n_1}{A}, \quad \overline{\epsilon^2} = \frac{\mathcal{N}_0}{2A^2} \quad \text{e} \quad \text{SNR} = \text{CSNR} = \frac{\overline{E}_m}{\mathcal{N}_0/2}, \tag{3.3}$$

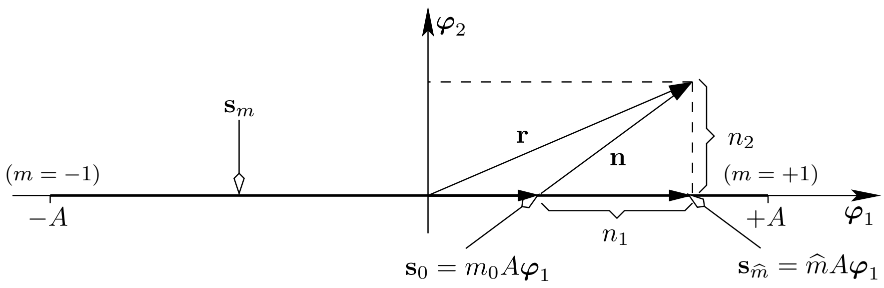


Figura 3.2: Relações geométricas de um sistema de comunicações usando uma modulação linear com uma única variável aleatória transmitida e um receptor de máxima verossimilhança.

onde $\overline{E}_m = A^2\overline{m^2}$ é a energia média do sinal $\mathbf{s}_m$ e $\overline{m^2}$ é a energia média de m.

Com isso, note de (3.3) que o $\overline{\epsilon^2}$ decresce com o aumento de A ou equivalentemente, com o aumento da energia média de transmissão. Portanto, tal parâmetro influencia no desempenho de uma modulação linear e será, convenientemente, chamado de fator de expansão, pois a variável aleatória $m \in [-1, 1]$ é amplificada ou "expandida" para o intervalo $[-A, A]$ na transmissão. Na próxima seção, será mostrado um resultado obtido em [18], no qual o fator de expansão está relacionado ao comprimento da curva associada à modulação e influencia no desempenho de modulações não-lineares, mais precisamente no valor do $\overline{\epsilon^2}$, como acontece na modulação linear.

Ainda segundo [18], quando a transmissão é perturbada pela ação de um ruído AWGN e p_m possui distribuição Gaussiana e independente do ruído do canal, então pode-se obter de (1.37), (1.39) e (1.41) que

$$\hat{m} = \frac{\mathbf{r}\cdot\boldsymbol{\varphi}_1}{A}\frac{\overline{E}_m}{\overline{E}_m + \mathcal{N}_0/2}, \quad \overline{\epsilon^2} = \frac{\overline{m^2}}{1 + 2\overline{E}_m/\mathcal{N}_0} \quad \text{e} \quad \text{SNR} = 1 + \frac{\overline{E}_m}{\mathcal{N}_0/2} = 1 + \text{CSNR}. \tag{3.4}$$

Neste caso, quando os dois sistemas de comunicações descritos em (3.3) e (3.4) são comparados, pode-se verificar que a relação sinal ruído é maior em (3.4) do que em (3.3), principalmente, para valores baixos de CSNR. Contudo, note que tal ganho é conseqüência do uso de um receptor MAP para obter (3.4) e, portanto, de um aumento de complexidade quando comparado com o receptor de máxima verossimilhança usado em (3.3), que possui a vantagem de não usar os valores de $\overline{E}_m$ e de $\mathcal{N}_0$ para calcular a estimativa $\hat{m}$, pois estes geralmente não são determinados com facilidade.

Portanto, note que, quando a função densidade de probabilidade p_m é conhecida, e este fato é usado no processo de recepção (MAP), então o desempenho do sistema de comunicações usando a modulação linear pode ser melhorado, mesmo que às custas de um aumento de complexidade. De forma similar, será mostrado na subseção 3.3.3 que este fato também é válido em sistemas de comunicações usando uma modulação não-linear, e, por essa razão, esse ganho de desempenho deve ser considerado e analisado nas modulações não-lineares, com o intuito de melhorar seus desempenhos.

3.3 Modulação Não-Linear

Na seção anterior, foi mostrado que sistemas de comunicações que usam uma modulação linear em conjunto com um receptor de máxima verossimilhança produzem um erro quadrático médio, (3.3), que pode ser reduzido, em geral, somente pelo incremento da energia média de transmissão ($\overline{E}_m$).

Contudo, sabe-se que, caso a modulação não seja necessariamente linear, então é possível, sob certas condições, decrescer $\overline{\epsilon^2}$ para um dado $\overline{m^2}$ sem incrementar $\overline{E}_m$. Em particular, vários esquemas de modulações não-lineares tais como a modulação por pulso (PPM) e a modulação em freqüência (FM) podem produzir um $\overline{\epsilon^2}$ menor do que o permitido pela modulação linear. Entretanto, tal ganho possui, geralmente, a desvantagem de aumentar a largura de banda de transmissão, quando comparada à largura de banda utilizada pela modulação linear.

Quando a interpretação geométrica (1.36) é utilizada para representar as técnicas de modulação, a largura de banda de transmissão é, primeiramente, estimada pela dimensão do sinal a ser transmitido, que, no caso, pode pertencer ao $\mathbb{R}^N$. Contudo, para que o valor da largura de banda de transmissão seja determinado de uma maneira mais eficiente e precisa, deve-se realizar uma análise espectral do sinal modulado $s(t)$ usando a transformada de Fourier. Entretanto, quando a modulação é não-linear, então o sinal $s(t)$ é uma função não-linear do sinal modulante $m(t)$ e, geralmente, o espectro do sinal modulado não está relacionado de uma maneira simples ao espectro do sinal modulante; pelo contrário, sua determinação é muito mais difícil do que a de um sinal proveniente de uma modulação linear.

Portanto, a obtenção de uma expressão precisa para a largura de banda de transmissão de um sinal proveniente de uma modulação não-linear, cujo sinal modulador $m(t)$ seja qualquer, geralmente, não é possível. Por essa razão, optou-se por usar simulação computacional para determinar a largura de banda de transmissão das modulações não-lineares consideradas neste capítulo. Tal simulação foi realizada com o auxílio da transformada discreta de Fourier (DFT), pois não foi possível usar aproximações do sinal modulado para desenvolver uma expressão analítica que descrevesse a análise espectral do sinal, como usualmente é realizado para algumas modulações não-lineares conhecidas, como por exemplo as modulações em fase (PM) e em freqüência (FM).

De acordo com os propósitos deste capítulo, o primeiro passo para propor novas modulações não-lineares e analisar geometricamente seus desempenhos é identificar quais são os parâmetros geométricos que mais influenciam no desempenho dessas modulações. Para tanto, foi utilizada a interpretação geométrica considerada em [18] e descrita matematicamente por (1.36), pois a mesma possibilita o uso de conceitos da geometria diferencial, tais como métrica e curvatura, como parâmetros geométricos de desempenho das técnicas de modulação.

3.3.1 Considerações geométricas

Como mencionado anteriormente, em [18] foi utilizada a interpretação geométrica que associa modulações a curvas no $\mathbb{R}^N$ para verificar que o elemento de comprimento da curva pode diminuir o erro quadrático médio. Para mostrar tal fato, assuma que a transmissão de $s_m(t)$ esteja sujeita à ação de um ruído suficientemente pequeno, de maneira que $\mathbf{s}_m$ possa ser aproximado, com o auxílio da série de Taylor, da seguinte forma

$$\mathbf{s}_m \approx \mathbf{s}_0 + (m - m_0)\mathbf{s}_0', \quad \text{com} \quad \mathbf{s}_0' = \left.\frac{d\mathbf{s}_m}{dm}\right|_{m=m_0}, \tag{3.5}$$

onde $\mathbf{s}_0'$ é a derivada de $\mathbf{s}_m$ no ponto m_0. Neste caso, note que $\mathbf{s}_m$ pode ser interpretada como a reta tangente à curva $\mathbf{s}_m$ em $\mathbf{s}_0$.

A Figura 3.3 ilustra a recepção por máxima verossimilhança de uma modulação não-linear usando essa aproximação quando $\mathbf{s}_0$ é transmitido. Nesse caso, observe que o receptor obtém a estimativa $\hat{m}$ projetando o sinal recebido $\mathbf{r}$ na direção do vetor tangente à curva, $\mathbf{s}_0'$. Como mencionado anteriormente, tal fato também ocorreu na modulação linear e será matematicamente comprovado a seguir para uma modulação não-linear.

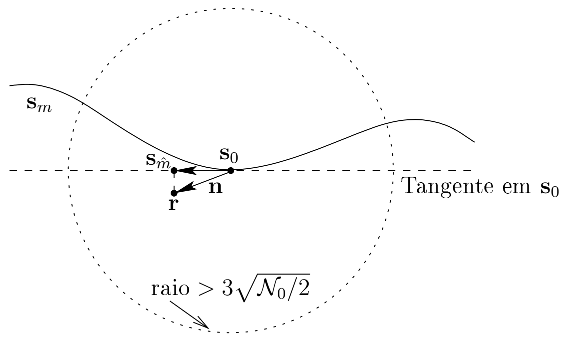


Figura 3.3: Recepção por máxima verossimilhança de um sistema de comunicações usando modulação não-linear quando o ruído de transmissão é AWGN e suficientemente pequeno.

Para tanto, quando $\mathbf{s}_0$ é transmitido pode-se fazer uso da aproximação descrita em (3.5) e do receptor de máxima verossimilhança (1.40) para determinar $\hat{m}$ da seguinte maneira:

$$p_{\mathbf{r}}(\mathbf{r}|\hat{m}) \geq p_{\mathbf{r}}(\mathbf{r}|m) \quad \Rightarrow \quad \frac{1}{\sqrt{\pi\mathcal{N}_0}}\exp\left(-\frac{|\mathbf{r}-\mathbf{s}_{\hat{m}}|^2}{\mathcal{N}_0}\right) \geq \frac{1}{\sqrt{\pi\mathcal{N}_0}}\exp\left(-\frac{|\mathbf{r}-\mathbf{s}_m|^2}{\mathcal{N}_0}\right) \quad \Rightarrow$$

$$|\mathbf{r}-\mathbf{s}_{\hat{m}}| \leq |\mathbf{r}-\mathbf{s}_m| \quad \Rightarrow \quad \hat{m} = \min_m |\mathbf{r}-\mathbf{s}_m| \quad \Rightarrow \quad \frac{d}{dm}|\mathbf{r}-\mathbf{s}_m| = 0 \quad \Rightarrow \quad (\mathbf{r}-\mathbf{s}_m)\cdot\mathbf{s}_0' = 0$$

$$\Rightarrow \quad [\mathbf{r} - \mathbf{s}_0 - (\hat{m} - m_0)\mathbf{s}_0']\cdot\mathbf{s}_0' = 0 \quad \Rightarrow \quad \hat{m} = m_0 + \frac{(\mathbf{r}-\mathbf{s}_0)\cdot\mathbf{s}_0'}{|\mathbf{s}_0'|^2}. \tag{3.6}$$

Para calcular $\overline{\epsilon^2}(m_0)$, o erro quadrático médio, dado que m_0 foi enviado, pode-se usar a estimativa $\hat{m}$ obtida em (3.6) e o fato do ruído ser igual ao sinal recebido menos o sinal transmitido, $\mathbf{n} = \mathbf{r} - \mathbf{s}_0$, como mostra a Figura 3.3, para se obter que

$$\begin{aligned} \overline{\epsilon^2}(m_0) &= \mathrm{E}[(m - \hat{m})^2 | m = m_0] = \frac{1}{|\mathbf{s}_0'|^2} \mathrm{E}\left[\left(\mathbf{n} \cdot \frac{\mathbf{s}_0'}{|\mathbf{s}_0'|}\right)^2\right] = \frac{\mathcal{N}_0/2}{|\mathbf{s}_0'|^2} \\ &= \frac{\mathcal{N}_0/2}{S^2(m_0)}, \end{aligned} \tag{3.7}$$

onde $S(m)$ é o fator de expansão definido como

$$S(m) \triangleq \left|\frac{d\boldsymbol{s}_m}{dm}\right|, \quad \forall\, m \in [-1, 1], \tag{3.8}$$

que generaliza o valor do fator de expansão atribuído anteriormente para a modulação linear, $S(m) = A$, quando $\mathbf{s}_m = mA\boldsymbol{\varphi}_\mathbf{1}$.

Entretanto, observe que o erro quadrático médio obtido em (3.7) depende da variável transmitida m_0. Portanto, de acordo com (1.37), para obter $\overline{\epsilon^2}$, deve-se considerar todos os possíveis valores de entrada, $m_0 \in [-1, 1]$. Neste caso, existem duas situações: uma quando o fator de expansão não é uniforme ($S(m)$ não é constante), o que resulta em

$$\overline{\epsilon^2} = \mathrm{E}\left[\frac{\mathcal{N}_0/2}{S^2(m)}\right] = \int_{-1}^{1} p_m \frac{\mathcal{N}_0/2}{S^2(m)}\, dm\,, \tag{3.9}$$

e outra quando $S(m)$ é uniforme, o que implica que

$$\overline{\epsilon^2} = \frac{\mathcal{N}_0/2}{S^2}. \tag{3.10}$$

Portanto, de (3.10), podemos observar que o fator de expansão influencia no desempenho da modulação, pois o erro quadrático médio é inversamente proporcional ao quadrado de seu valor. Geometricamente, o fator de expansão está associado ao comprimento, L, da curva associada a modulação, isto é,

$$L = \int_{-1}^{1} |\mathbf{s}_m'|\, dm = \begin{cases} 2S\,, & \text{caso } S(m) \text{ seja uniforme,} \\ \int_{-1}^{1} S(m)\, dm, & \text{caso contrário.} \end{cases} \tag{3.11}$$

Por essa razão, quanto maior for o comprimento da curva, menor será o $\overline{\epsilon^2}$ da modulação, como previsto em [18]. Além disso, note que o erro quadrático médio obtido em (3.10) independe da função p_m, fato que motivou [18] a pensar que um fator de expansão uniforme minimizaria o erro quadrático médio produzido por um receptor ML quando p_m fosse escolhido o mais adversamente possível. Contudo, tal fato não garante que não seja possível determinar condições apropriadas para S e p_m, de maneira que o $\overline{\epsilon^2}$ obtido a partir de (3.9) seja menor do que o obtido em (3.10).

Portanto, uma análise mais detalhada de (3.9) deve ser realizada no intuito de responder ao seguinte questionamento: *qualquer que seja p_m, pode existir uma expressão para $S(m)$ de maneira que o valor de $\overline{\epsilon^2}$ obtido por (3.9) seja menor do que o valor apresentado em (3.10)?* Até onde é de nosso conhecimento, uma resposta para tal questionamento é desconhecida. Responder a essa pergunta significa estudar se existe uma relação entre p_m e $S(m)$, de tal forma que o sistema possa ser melhorado. A seguir mostraremos que a resposta para esse questionamente é sim, existe um "casamento" entre p_m e $S(m)$ de maneira que $\overline{\epsilon^2}$ seja menor do que $(\mathcal{N}_0/2)/S^2$.

3.3.2 Casamento entre p_m e $S(m)$

Como mencionado anteriormente, existe uma relação entre a função densidade de probabilidade *a priori* p_m e o fator de expansão $S(m)$ de uma modulação não-linear, cujo objetivo é melhorar o desempenho dos sistemas de comunicações que satisfazem essa relação ou casamento.

Inicialmente, para determinar esse casamento entre p_m e $S(m)$, considere que o erro quadrático médio seja calculado usando (3.9) e que a função densidade de probabilidade *a priori* e o fator de expansão sejam relacionados da seguinte forma

$$p_m = c\, S^2(m), \tag{3.12}$$

onde c é uma constante de normalização determinada por

$$\int_{-1}^{1} c\, S^2(m)\, dm = 1 \quad \Rightarrow \quad c = \frac{1}{\int_{-1}^{1} S^2(m)\, dm}. \tag{3.13}$$

Assim, se (3.12) e (3.13) forem substituídas em (3.9), então o erro quadrático médio proveniente desse sistema de comunicações é dado por

$$\overline{\epsilon^2} = \int_{-1}^{1} p_m \frac{\mathcal{N}_0/2}{S^2(m)}\, dm = \int_{-1}^{1} c\, S^2(m) \frac{\mathcal{N}_0/2}{S^2(m)}\, dm = 2c\mathcal{N}_0/2 = \frac{\mathcal{N}_0}{\int_{-1}^{1} S^2(m)\, dm}. \tag{3.14}$$

Como o objetivo é mostrar que o valor de $\overline{\epsilon^2}$, calculado por (3.14), é menor que o encontrado em (3.10), considere a desigualdade de Schwarz

$$\left(\int_{a}^{b} f g\, dm\right)^2 \leq \int_{a}^{b} f^2\, dm \cdot \int_{a}^{b} g^2\, dm\,, \tag{3.15}$$

com $f = S(m)$ e $g = 1$, para obter com o auxílio de (3.11) que

$$\left(\int_{-1}^{1} S(m)\, dm\right)^2 \leq 2\int_{-1}^{1} S^2(m)\, dm \quad \Rightarrow \quad \int_{-1}^{1} S^2(m)\, dm \geq \frac{L^2}{2} = 2S^2. \tag{3.16}$$

Finalmente, aplicando (3.16) em (3.14), verifica-se que o erro quadrático médio satisfaz a seguinte desigualdade

$$\overline{\epsilon^2} \leq \frac{\mathcal{N}_0/2}{S^2}, \tag{3.17}$$

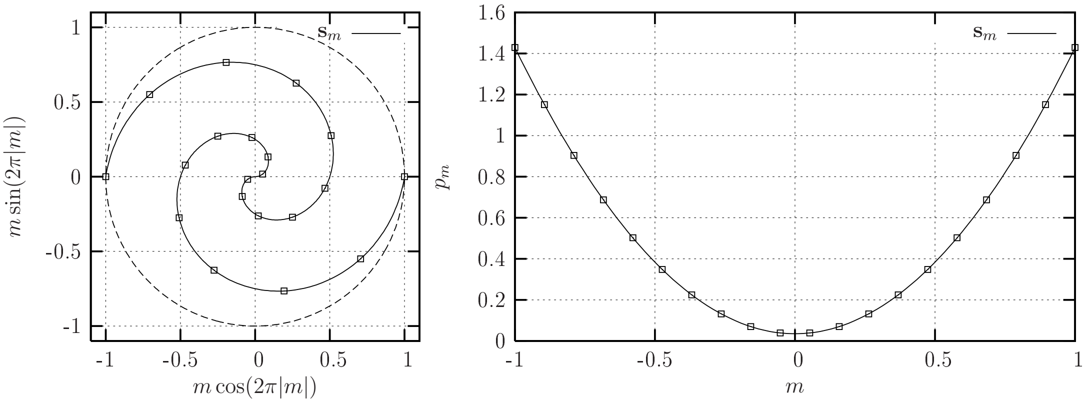
sendo que a igualdade ocorre justamente quando o fator de expansão é uniforme.

Portanto, comparando (3.17) com (3.10) concluímos que o erro quadrático médio de um sistema de comunicações pode ser reduzido usando um receptor ML e energia de ruído suficientemente pequena, desde que p_m satisfaça (3.12). Em outras palavras, um fator de expansão uniforme pode não ser a melhor escolha para minimizar o $\overline{\epsilon^2}$, como sugere [18], pois, dependendo da distribuição de p_m, pode-se determinar um valor para $S(m)$ de maneira a diminuir ainda mais o $\overline{\epsilon^2}$. Além disso, é importante ressaltar que este casamento entre p_m e $S(m)$ pode ser realizado com o auxílio de uma mudança de variável aleatória ou, equivalentemente, por uma reparametrização da curva associada à modulação.

Com o intuito de exemplificar a diminuição do $\overline{\epsilon^2}$ em sistemas de comunicações usando o casamento proposto em (3.12) quando comparados aos sistemas usando $S(m)$ uniforme, considere dois exemplos de sistemas de comunicações usando modulações não-lineares associadas à curva espiral de Arquimedes, cujo sinal modulado pode ser expresso como

$$\mathbf{s}_m = [m\cos(2\pi|m|), m\sin(2\pi|m|)], \quad m \in [-1, 1], \tag{3.18}$$

e representado graficamente na Figura 3.4(a). Devemos chamar a atenção para o fato de que tal esquema de modulação havia sido considerado inicialmente em [19] e posteriormente em [28], sendo que uma análise mais detalhada de seu desempenho é realizada na subseção 3.4.2.



(a) Curva associada a modulação. (b) Função densidade de probabilidade (p_m).

Figura 3.4: Um exemplo de modulação não-linear associada a curva espiral de Arquimedes.

Exemplo 1 Neste primeiro exemplo, será calculado o desempenho de um sistema de comunicações cujo esquema de modulação possui um fator de expansão uniforme.

Neste caso, primeiramente, calculamos o fator de expansão com o auxílio de (3.8) e (3.18), conduzindo a

$$S(m) = \sqrt{1 + 4\pi^2 m^2}, \tag{3.19}$$

que no caso não é uniforme. Em seguida, como o objetivo é obter um fator de expansão uniforme, devemos reparametrizar a curva $\mathbf{s}_m$ dada em (3.18) pelo parâmetro $s \in [-1, 1]$, com $m = m(s)$, de maneira que

$$\frac{L}{2}\, ds = S(m)\, dm \quad \Rightarrow \quad ds = \frac{2}{L}\sqrt{1 + 4\pi^2 m^2}\, dm, \tag{3.20}$$

onde L é o comprimento da curva $\mathbf{s}_m$, podendo ser calculado com o auxílio de (3.11), que no caso particular dessa modulação é igual a 6.766.

Dessa forma, substituindo (3.20) em (3.10) e (3.11) obtemos que

$$S = \frac{L}{2} = 3.383 \quad \text{e} \quad \overline{\epsilon^2} = \frac{\mathcal{N}_0/2}{S^2} = 0.08738\, \mathcal{N}_0/2\,. \tag{3.21}$$

Considerando que a variável aleatória s possua função densidade de probabilidade uniforme, $p_s = 1/2$, podemos obter que $\overline{s^2} = 1/3$. Além disso, substituindo (3.20) em (1.42) podemos obter que a energia média da modulação é dada por

$$\overline{E}_m = \int_{-1}^{1} |\mathbf{s}_m|^2 p_s\, ds = \frac{1}{L}\int_{-1}^{1} m^2\sqrt{1 + 4\pi^2 m^2}\, dm = 0.476\,. \tag{3.22}$$

Os valores de CSNR e SNR podem ser obtidos pela substituição direta de (3.21) e (3.22) em (1.41), como segue

$$\text{CSNR} = \frac{0.476}{\mathcal{N}_0} \quad \text{e} \quad \text{SNR} = \frac{3.815}{\mathcal{N}_0/2} \quad \Rightarrow \quad \text{SNR} = 16.038\, \text{CSNR}. \tag{3.23}$$

Assim, com o auxílio de (3.23), o desempenho desse sistema pode ser medido pela diferença em dB entre SNR e CSNR, que nesse caso é de aproximadamente 12.0 dB.

Exemplo 2 Neste exemplo será considerado um sistema de comunicações cujo esquema de modulação usa o casamento entre p_m e $S(m)$ proposto em (3.12). Inicialmente, foi determinada a função densidade de probabilidade casada ao fator de expansão com o auxílio de (3.19) e (3.12) como segue

$$p_m = c(1 + 4\pi^2 m^2), \quad \text{com } c = \frac{3}{6 + 8\pi^2}. \tag{3.24}$$

Tal função densidade de probabilidade é representada graficamente pela Figura 3.4(b).

Em seguida, (3.19) e (3.24) foram aplicadas em (3.9) e (1.42) para obter que o erro quadrático médio e a energia média desse esquema de modulação são dados por

$$\overline{\epsilon^2} = \frac{3}{3 + 4\pi^2}\mathcal{N}_0/2 \quad \text{e} \quad \overline{E}_m = \overline{m^2} = \frac{5 + 12\pi^2}{15 + 20\pi^2}, \tag{3.25}$$

respectivamente. Finalmente, aplicando (3.25) em (1.41) os seguintes parâmetros de desempenho do sistema são obtidos

$$\text{CSNR} = \frac{5+12\pi^2}{(15+20\pi^2)\mathcal{N}_0} \quad \text{e} \quad \text{SNR} = \frac{5+12\pi^2}{15\mathcal{N}_0/2} \quad \Rightarrow \quad \text{SNR} = 2\left(1+\frac{4}{3}\pi^2\right)\text{CSNR}. \tag{3.26}$$

Assim, com o auxílio de (3.26) é possível verificar que o desempenho desse sistema é de aproximadamente 14.5 dB. Tal desempenho está de acordo com a desigualdade (3.17), pois é 2.5 dB maior do que o ganho do sistema de comunicações do exemplo anterior que usa um fator de expansão uniforme.

Contudo, deve-se ressaltar que tal ganho não está associado apenas ao comprimento da curva, mas também ao casamento que existe entre p_m e $S(m)$, que pode ser interpretado geometricamente com o auxílio da Figura 3.4. Neste caso, observe que o fator de expansão dessa modulação não é uniforme, pois os pontos representados pelos símbolos "quadrados" da Figura 3.4(a) estão associados a valores de m igualmente separados por Δm, porém, a diferença de comprimento entre esses pontos não é constante, sendo que alguns estão mais próximos uns dos outros. Contudo, o casamento proposto faz com que esses pontos mais próximos sejam aqueles menos prováveis de serem transmitidos, como indica p_m na Figura 3.4(b), diminuindo assim o erro quadrático médio do sistema de comunicações do Exemplo 2.

Os desempenhos dos dois sistemas de comunicações dos exemplos anteriores também foram verificados usando simulação computacional para construir as curvas de CSNR *versus* SNR apresentadas na Figura 3.5. Em tal simulação as variáveis aleatórias m e $\mathbf{n} = [n_1, n_2]$ foram obtidas usando um gerador de números aleatórios baseado no método da rejeição. Para obter informação sobre esse método consulte [1]. Além disso, cada um dos pontos das curvas simuladas foi obtido com o envio de aproximadamente 50 mil sinais aleatórios, o que garante uma confiabilidade satisfatória dos resultados encontrados pelo método de Monte Carlo. Com o intuito de comparar o desempenho desses sistemas simulados com o sistema ótimo (1.43), foi traçada na Figura 3.5 a curva OPTA, sendo que, quanto mais próxima estiver as curvas de desempenho simuladas dessa curva OPTA, melhor será o desempenho do sistema.

Além disso, analisando mais detalhadamente a Figura 3.5, é possível constatar que as aproximações (3.21) e (3.25) para o $\overline{\epsilon^2}$ dos sistemas de comunicações considerados podem ser aplicadas para medir seus desempenhos. Contudo, como esperado, tais aproximações são válidas apenas quando a energia do ruído for suficientemente pequena, no caso do exemplo, para valores de CSNR maiores do que 18 dB, pois, nessa situação, as curvas de desempenho estimadas por essas aproximações estão muito próximas das curvas obtidas por simulação.

Neste momento, cabe chamar a atenção para o fato do casamento proposto em (3.12) ter levado em consideração apenas o valor do erro quadrático médio. Contudo,

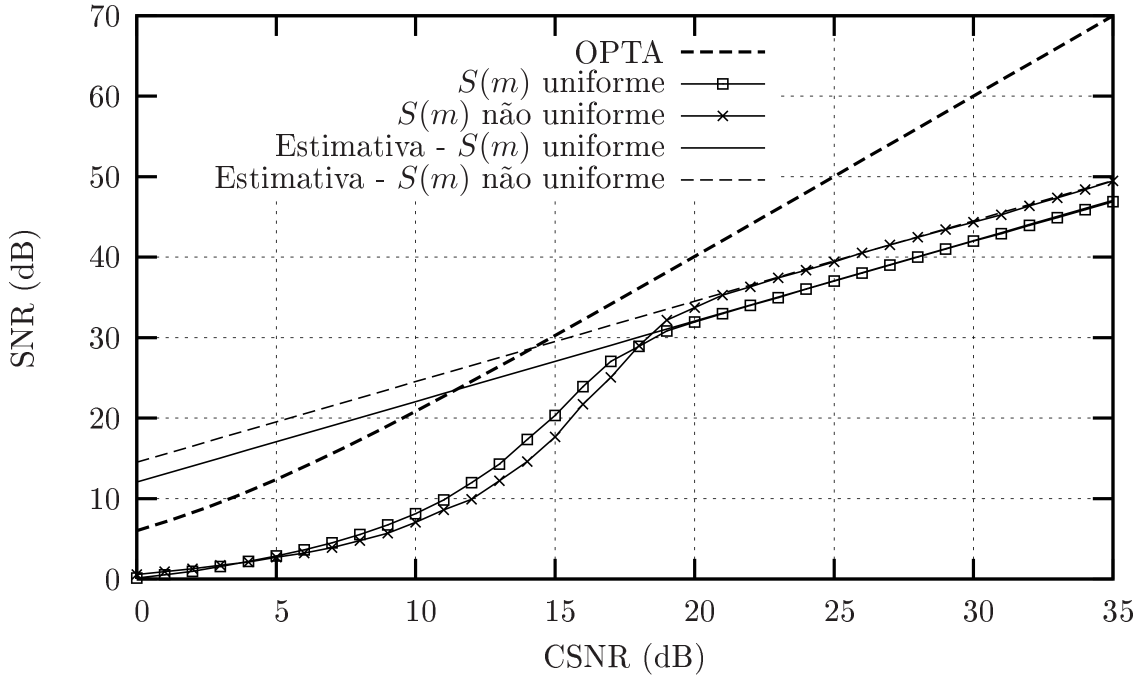


Figura 3.5: Exemplos de curvas de desempenho calculadas usando fatores de expansão uniforme e não uniforme em uma modulação não-linear associada à curva espiral de Arquimedes.

quando se analisa o desempenho de um sistema de comunicações deve-se considerar também os valores da energia média de transmissão e de $\overline{m^2}$, ou mais geralmente o valor da diferença em dB entre CSNR e SNR. Portanto, propomos um novo casamento entre p_m e $S(m)$ cujo objetivo é maximizar tal diferença. Para tanto, as equações (1.41), (1.42) e (3.9) foram usadas para obter que

$$\begin{aligned} SNR - CSNR &= 10\log_{10}\left(\frac{2\,\overline{m^2}}{\overline{E}_m \int_{-1}^{1} 1/S^2(m)p_m\,dm}\right) \\ &= 10\log_{10}\left(\frac{2\int_{-1}^{1} m^2 p_m\,dm}{\int_{-1}^{1} |\mathbf{s}_m|^2 p_m\,dm \cdot \int_{-1}^{1} 1/S^2(m)p_m\,dm}\right), \end{aligned} \qquad (3.27)$$

onde a multiplicação por 2 no numerador decorre da dimensão do espaço no qual a curva associada a modulação está imersa, no caso considerado o $\mathbb{R}^2$. Em outras palavras, essa multiplicação no numerador de (3.27) deve ser por N, caso a curva pertença ao $\mathbb{R}^N$.

Entretanto, determinar uma relação entre p_m, $S(m)$ e m de maneira que o desempenho descrito em (3.27) seja máximo, em geral, não é tão imediado. Porém, para alguns casos particulares, como por exemplo, quando o fator de expansão for uniforme, é possível considerar que $p_m = c/|\mathbf{s}_m|^2$, onde c é uma constante de normalização, para minimizar a energia média de transmissão. Além disso, caso $S(m)$ não seja uniforme, deve-se utilizar p_m como definido em (3.12) para minimizar o erro quadrático médio ou, dependendo do caso, utilizar a desigualdade de Schwarz (3.15) com $f = |\mathbf{s}_m|\sqrt{p_m}$ e $g = \sqrt{p_m}/S(m)$ para obter

$$SNR - CSNR \geq 10\log_{10}\left[\frac{2\int\limits_{-1}^{1} m^2 p_m\, dm}{\left(\int\limits_{-1}^{1} \frac{|\mathbf{s}_m|}{S(m)} p_m\, dm\right)^2}\right], \tag{3.28}$$

em seguida, usar novamente a desigualdade de Schwarz em (3.28), porém com $f = \frac{|\mathbf{s}_m|\sqrt{p_m}}{m\,S(m)}$ e $g = m\sqrt{p_m}$, para determinar que

$$SNR - CSNR \geq 10\log_{10}\left(\frac{2}{\int\limits_{-1}^{1} \frac{|\mathbf{s}_m|^2}{m^2\, S^2(m)} p_m\, dm}\right), \tag{3.29}$$

e, finalmente, para minimizar o denominador de (3.29), a seguinte relação entre a função densidade de probabilidade *a priori* e o fator de expansão deve ser satisfeita

$$p_m = c\frac{m^2\, S^2(m)}{|\mathbf{s}_m|^2}, \tag{3.30}$$

onde c é uma constante de normalização.

Note que, se o casamento proposto em (3.30) for aplicado no sistema de comunicações do Exemplo 2, então o desempenho obtido é o mesmo, pois, nesse caso $|\mathbf{s}_m|^2 = m^2$, o que implica em $p_m = cS^2(m)$, como realizado anteriormente no referido exemplo. Contudo, quando o desempenho do sistema de comunicações do Exemplo 1 é comparado ao desempenho de um sistema de comunicações usando o casamento (3.30) e um fator de expansão uniforme, então o resultado obtido neste último caso é melhor, como mostra o próximo exemplo.

Exemplo 3 Neste exemplo será calculado o desempenho de um sistema de comunicações cujo esquema de modulação está associado à curva espiral de Arquimedes com fator de expansão uniforme e função densidade de probabilidade *a priori* satisfazendo (3.30). Para tanto, considere a mesma reparametrização indicada em (3.20) e, portanto, o mesmo fator de expansão indicado em (3.21). Assim, com o auxílio de (1.41), (1.42) e (3.9), os seguintes parâmetros de desempenho do sistema são obtidos:

$$\overline{\epsilon^2} = 0.0659\mathcal{N}_0/2, \quad \overline{m^2} = 0.4803, \quad \overline{E}_m = 0.6275, \quad \text{e} \quad \text{SNR} = 23.22\,\text{CSNR}\,.$$

Portanto, verifica-se que o desempenho desse sistema de comunicações é de aproximadamente 13.7 dB, que é 1.7 dB melhor que o desempenho do sistema do Exemplo 1, como prevê a desigualdade (3.29).

3.3.3 Uma aproximação para o $\overline{\epsilon^2}$ usando um receptor MAP

Nesta subseção, desenvolveremos uma aproximação para o erro quadrático médio de um sistema de comunicações transmitindo uma variável aleatória Gaussiana m de média nula e variância $\sigma^2 = \overline{m^2}$, sujeito a ação de um ruído AWGN de energia $\mathcal{N}_0/2$ em cada uma das N dimensões, usando um receptor MAP e uma modulação não-linear cujo sinal $\mathbf{s}_m$ será aproximado por (3.5) no processo de recepção.

Para tanto, considere o receptor MAP descrito em (1.39) para determinar a seguinte regra de decisão

$$\begin{gathered}
p_m(\hat{m})p_{\mathbf{r}}(\mathbf{r}|\hat{m}) \geq p_m(m)p_{\mathbf{r}}(\mathbf{r}|m) \quad \Rightarrow \\
\frac{1}{\sqrt{2\pi}\sigma}e^{\left(-\frac{\hat{m}^2}{2\sigma^2}\right)} \cdot \frac{1}{(\pi\mathcal{N}_0)^{N/2}}e^{\left(-\frac{|\mathbf{r}-\mathbf{s}_{\hat{m}}|^2}{\mathcal{N}_0}\right)} \geq \frac{1}{\sqrt{2\pi}\sigma}e^{\left(-\frac{m^2}{2\sigma^2}\right)} \cdot \frac{1}{(\pi\mathcal{N}_0)^{N/2}}e^{\left(-\frac{|\mathbf{r}-\mathbf{s}_m|^2}{\mathcal{N}_0}\right)} \quad \Rightarrow \\
\frac{\hat{m}^2}{2\sigma^2} + \frac{|\mathbf{r}-\mathbf{s}_{\hat{m}}|^2}{\mathcal{N}_0} \leq \frac{m^2}{2\sigma^2} + \frac{|\mathbf{r}-\mathbf{s}_m|^2}{\mathcal{N}_0} \quad \Rightarrow \\
\hat{m} = \min_{m\in(-1,1)}\left\{\frac{m^2}{2\sigma^2} + \frac{|\mathbf{r}-\mathbf{s}_m|^2}{\mathcal{N}_0}\right\}. \qquad (3.31)
\end{gathered}$$

Portanto, a expressão para $\hat{m}$ é obtida a partir da derivada de (3.31) igualada a zero e com o auxílio da aproximação (3.5), o que resulta em

$$m\mathcal{N}_0 - 2\sigma^2(\mathbf{r}-\mathbf{s}_m)\cdot\mathbf{s}'_m = 0 \quad \Rightarrow \quad m\mathcal{N}_0 - 2\sigma^2(\mathbf{r}-\mathbf{s}_0)\cdot\mathbf{s}'_0 + 2\sigma^2(m-m_0)|\mathbf{s}'_0|^2 = 0 \quad \Rightarrow$$

$$\hat{m} = \frac{2\sigma^2\left[(\mathbf{r}-\mathbf{s}_0)\cdot\mathbf{s}'_0 + m_0|\mathbf{s}'_0|^2\right]}{\mathcal{N}_0 + 2\sigma^2|\mathbf{s}'_0|^2} \qquad (3.32)$$

Para calcular $\overline{\epsilon^2}(m_0)$, o erro quadrático médio dado que m_0 foi enviado, pode-se usar a estimativa $\hat{m}$ obtida em (3.32), o fato de o ruído ser igual ao sinal recebido menos o sinal transmitido ($\mathbf{n} = \mathbf{r} - \mathbf{s}_0$) e o fato de que $\mathbf{n}\cdot\mathbf{s}'_0/|\mathbf{s}'_0|$ é uma componente do ruído AWGN em uma única dimensão com média nula e energia $\mathcal{N}_0/2$, para verificar que

$$\begin{aligned}
\overline{\epsilon^2}(m_0) &= \mathrm{E}[(m-\hat{m})^2|m=m_0] = \mathrm{E}\left[\left(\frac{2\sigma^2(\mathbf{r}-\mathbf{s}_0)\cdot\mathbf{s}'_0 - m_0\mathcal{N}_0}{\mathcal{N}_0+2\sigma^2|\mathbf{s}'_0|^2}\right)^2 |m=m_0\right] \\
&= \mathcal{N}_0\frac{2\sigma^4|\mathbf{s}'_0|^2 + m_0^2\mathcal{N}_0}{\left(\mathcal{N}_0+2\sigma^2|\mathbf{s}'_0|^2\right)^2}, \qquad (3.33)
\end{aligned}$$

onde $\mathrm{E}[\,\cdot\,]$ é o operador esperança medido para todos os possíveis valores recebidos $\mathbf{r}$, quando $\mathbf{s}_0$ é transmitido.

Observe que o erro quadrático médio obtido em (3.33) depende da variável m_0. Assim, para encontrar o valor do $\overline{\epsilon^2}$ para o sistema considerado, deve-se calcular a média dos valores de $\overline{\epsilon^2}(m_0)$ para todos os possíveis valores de $m_0 \in [-1, 1]$, isto é,

$$\overline{\epsilon^2} = \mathrm{E}\left[\mathcal{N}_0 \frac{2\sigma^4|\mathbf{s}_0'|^2 + m_0^2\mathcal{N}_0}{(\mathcal{N}_0 + 2\sigma^2|\mathbf{s}_0'|^2)^2}\right] = \mathcal{N}_0 \frac{2\sigma^4 S^2 + \sigma^2\mathcal{N}_0}{(\mathcal{N}_0 + 2\sigma^2 S^2)^2} = \frac{\sigma^2\mathcal{N}_0}{\mathcal{N}_0 + 2\sigma^2 S^2} = \frac{\overline{m^2}\mathcal{N}_0/2}{\overline{m^2}S^2 + \mathcal{N}_0/2}, \tag{3.34}$$

no qual foi considerado um fator de expansão uniforme (independente do valor de m_0). além disso, ressaltamos que a estimativa (3.34), para o $\overline{\epsilon^2}$ usando um receptor MAP, não havia sido obtida anteriormente.

Para uma melhor análise dos desempenhos das modulações não-lineares, são apresentadas a seguir algumas relações entre os parâmetros de desempenho SNR e CSNR de sistemas de comunicações que fazem uso de modulações lineares ou não-lineares e receptores MAP ou ML. Para tanto, foram utilizadas as equações (1.41), (3.3), (3.4), (3.10) e (3.34) com fator de expansão uniforme e $\mathcal{N} = N \cdot \mathcal{N}_0/2$, onde N é a dimensão do espaço que contém a curva associada a modulação, para obter as seguintes relações

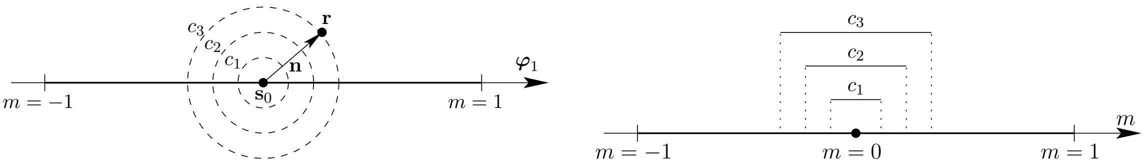
$$\begin{aligned}
\mathrm{SNR} &= N \cdot \mathrm{CSNR}, \quad \text{modulação linear e receptor ML} && (3.35a)\\
\mathrm{SNR} &= \frac{N\,\overline{m^2}S^2}{\overline{E}_m}\mathrm{CSNR}, \quad \text{modulação não-linear e receptor ML} && (3.35b)\\
\mathrm{SNR} &= 1 + N \cdot \mathrm{CSNR}, \quad \text{modulação linear e receptor MAP} && (3.35c)\\
\mathrm{SNR} &= 1 + \frac{N\,\overline{m^2}S^2}{\overline{E}_m}\mathrm{CSNR}, \quad \text{modulação não-linear e receptor MAP} && (3.35d)
\end{aligned}$$

3.3.4 Influência da curvatura no projeto de modulações

Analisando o desempenho dos sistemas de comunicações com o auxílio das relações apresentadas em (3.35), é possível verificar que o uso de uma modulação linear ($\overline{E}_m = S^2\overline{m^2}$) em $\mathbb{R}^N$ produz um ganho, $\mathrm{SNR} = 1 + N\,\mathrm{CSNR}$, diretamente proporcional a N, porém muito abaixo do ganho esperado pelo limitante superior (1.43), que aumenta exponencialmente com a dimensão N. Portanto, o uso de modulações não-lineares demonstra ser uma alternativa para melhorar o desempenho dos sistemas de comunicações, como constatado nos exemplos de modulações não-lineares considerados anteriormente usando a curva espiral de Arquimedes.

Sabendo que a curvatura está associada à variação do vetor tangente de uma curva, então a curvatura de uma modulação linear é nula. Neste caso, é possível afirmar qualitativamente que, se o desempenho de um sistema de comunicações pode ser melhorado usando uma modulação não-linear e, como a curva associada a essa modulação possui curvatura diferente de zero, então a existência de curvatura pode influenciar no desempenho do sistema de comunicação.

Contudo, antes de quantificar a influência da curvatura no projeto de modulações, é necessário introduzir o conceito de erro de ponto inicial (*threshold*) definido em [18]. Para tanto, considere inicialmente a modulação linear apresentada na Figura (3.6(a)), com $\mathbf{s}_0$ transmitido e $\mathbf{r}$ recebido. A Figura 3.6(b) exemplifica o processo de demodulação usando um receptor ML considerando que o sinal $\mathbf{r}$ pertença a um dos círculos c_1, c_2 ou c_3, que caracterizam a região de ação de um ruído AWGN com diferentes valores de energia média. Neste caso, note que o receptor determina três regiões para os possíveis valores de $\hat{m}$, sendo cada uma delas associada a um dos círculos considerados.

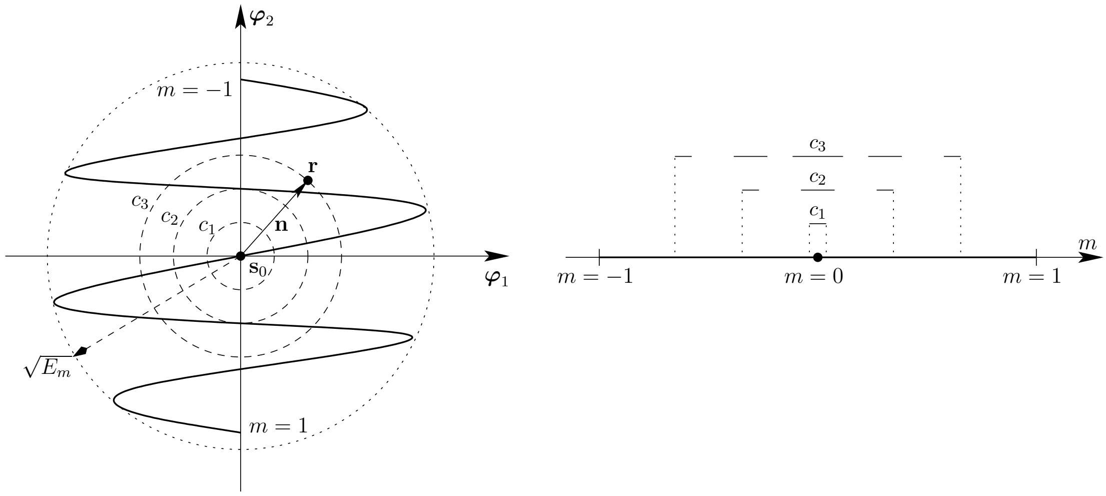


(a) Exemplo na modulação linear. (b) Possíveis valores para $\hat{m}$.

Figura 3.6: Interpretação geométrica do erro de ponto inicial em uma modulação linear sob a ação de um ruído AWGN.

Em seguida, considere a modulação não-linear ilustrada na Figura 3.7(a), também com $\mathbf{s}_0$ transmitido e $\mathbf{r}$ recebido. Nesse caso, observe que o comprimento da curva associada a uma modulação não-linear, em geral, é maior do que o comprimento da reta associada a uma modulação linear de mesma energia média. Contudo, caso a energia de ruído não seja suficientemente pequena, então esse aumento de comprimento das curvas associadas às modulações não-lineares pode causar um novo tipo de erro na recepção, o erro de ponto inicial. Isto porque, com o aumento do comprimento da curva, alguns sinais transmitidos que estão mais afastados em termos do comprimento da curva deveriam estar se distanciando, contudo, tendem a se aproximar cada vez mais, causando uma espécie de "confusão" no receptor. Esse fato ficará mais claro no decorrer desse capítulo, porém, é relevante ressaltar que tal erro é certamente o principal responsável por impedir que o comprimento da curva seja aumentado indefinidamente dentro de uma região limitada com o objetivo de melhorar o desempenho da modulação.

O erro de ponto inicial em modulações não-lineares pode ser interpretado geometricamente com o auxílio da Figura 3.7 da seguinte maneira: caso o sinal recebido $\mathbf{r}$ pertença a região delimitada pelo círculo c_1, então o receptor determinará um único intervalo na vizinhança de $m = 0$ para os possíveis valores de $\hat{m}$, Figura 3.7(b); caso contrário, se $\mathbf{r}$ pertencer às regiões limitadas pelos círculos c_2 ou c_3, então o receptor indicará um conjunto de intervalos para os possíveis valores de $\hat{m}$ que, necessariamente, não estão próximos da vizinhança de $m = 0$. Em outras palavras, o fato do receptor estimar valores de $\hat{m}$ muito diferentes do enviado m_0, quando a energia de ruído é



(a) Exemplo na modulação não-linear. (b) Possíveis valores para $\hat{m}$.

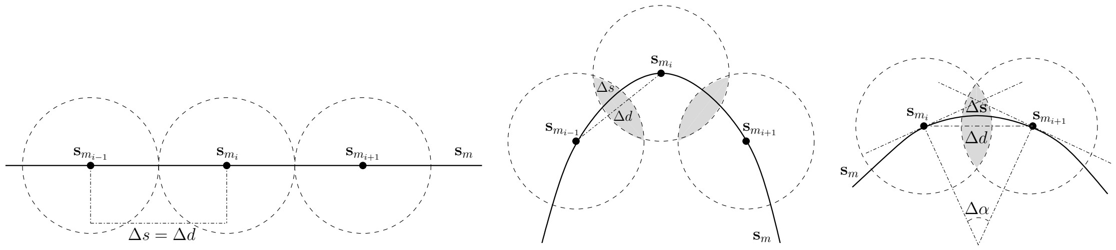
Figura 3.7: Interpretação geométrica do erro de ponto inicial em uma modulação não-linear sob a ação de um ruído AWGN.

"insuficiente" para isso, é que caracteriza a confusão do receptor de não identificar com precisão qual foi o ponto inicial transmitido. Portanto, esse erro pode aumentar significativamente o erro quadrático médio e, conseqüentemente, diminuir o desempenho do sistema, uma vez que, do ponto de vista do sistema de comunicações, é como se a energia do ruído fosse aparentemente maior do que é realmente.

Com isso, pode-se afirmar que, caso a energia de ruído seja suficientemente pequena, então é possível desprezar o erro de ponto inicial e o desempenho da modulação não-linear supera o desempenho da modulação linear, pois o fator de expansão aumenta, o que implica em um $\overline{\epsilon^2}$ menor. Contudo, caso a energia média do ruído seja grande a ponto do erro de ponto inicial ser significativo, então o desempenho da modulação não-linear pode cair rapidamente, atingindo valores inferiores ao da modulação linear, como será mostrado nos exemplos considerados na subseção 3.4.

Além disso, é interessante observar que, para a modulação linear não há erro de ponto inicial. Este fato será associado a seguir com o valor da curvatura da curva associada à modulação, que, no caso da modulação linear, é nula.

Para medir a influência da curvatura no projeto de modulações, considere que a Figura 3.8(a) represente uma modulação linear e a Figura 3.8(b) uma modulação não-linear, ambas sob a ação de um ruído AWGN. Além disso, considere que os círculos apresentados nessas figuras sejam interpretados como a região de ação do ruído dado que os sinais $\mathbf{s}_{m_{i-1}}$, $\mathbf{s}_{m_i}$ e $\mathbf{s}_{m_{i+1}}$, onde $m_{i-1} = m_i - \Delta m$ e $m_{i+1} = m_i + \Delta m$, possam ser transmitidos. Assuma também que o fator de expansão é uniforme em ambas as modulações, o que implica que a diferença de comprimento entre os sinais é igual a $\Delta s = S\Delta m$. Portanto, fazendo uso de tais considerações, é possível realizar uma comparação geométrica entre essas duas modulações e verificar que a distância Euclidiana (Δd) entre os sinais da modulação não-linear é, em geral, menor do que a mesma distância na modulação linear. Tal idéia pode ser melhor compreendida verificando que na modulação não-linear existe uma região de interseção entre os círculos, fato que não ocorre na modulação linear, ou simplesmente notando que $\Delta d < \Delta \mathbf{s}$ na Figura 3.8(b).



(a) Modulação linear ($k = 0$). (b) Modulação não-linear. (c) Δd, $\Delta \mathbf{s}$ e $\Delta \alpha$.

Figura 3.8: Análise da influência da curvatura na modulação s_m.

Em outras palavras, pode-se verificar da Figura 3.8, que quando o ruído for AWGN e o receptor for ML, então o erro de ponto inicial é influenciado pela distância Euclidiana entre os sinais. A seguir, será mostrado que tal distância diminui na modulação não-linear por causa do valor de sua curvatura, sendo portanto essas modulações mais sujeitas à ação do erro de ponto inicial.

Geometricamente, a curvatura está relacionada com a variação do vetor tangente à curva associada a modulação ou, mais precisamente, com a variação do ângulo de $\mathbf{s}'_m$ por comprimento da curva. Neste último caso, a curvatura pode ser definida como

$$k = \lim_{\Delta \mathbf{s} \to 0} \frac{\Delta \alpha}{\Delta s}, \tag{3.36}$$

onde $\Delta \alpha$ é a variação do ângulo de $\mathbf{s}'_m$, como ilustra a Figura 3.8(c). Além disso, dependendo da parametrização da curva associada à modulação, parametrizada pelo comprimento de arco (p.p.c.a, $|\mathbf{s}'_m| = 1$) ou não, a curvatura pode ser calculada por

$$k = |\mathbf{s}''_m|, \qquad \text{(p.p.c.a)} \qquad \text{ou} \tag{3.37a}$$

$$k = \frac{s'_1 s''_2 - s'_2 s''_1}{[(s'_1)^2 + (s'_2)^2]^{(3/2)}}, \quad \text{(caso contrário).} \tag{3.37b}$$

Entretanto, para mostrar a influência da curvatura, não podemos utilizar a aproximação (3.5), pois ela é de primeira ordem e k envolve uma derivada segunda de $\mathbf{s}_m$. Portanto, faz-se necessária a utilização de uma aproximação para $\mathbf{s}_m$ de ordem maior, como a indicada em [25], onde a série de Taylor foi utilizada para realizar uma aproximação de terceira ordem da curva associada a uma modulação com fator de expansão uniforme. Tal aproximação é expressa em função de sua curvatura da seguinte forma

$$\mathbf{s}_m - \mathbf{s}_0 = \left[\Delta s - \frac{k^2}{6}(\Delta s)^3,\ \frac{k}{2}(\Delta s)^2 + \frac{k'}{6}(\Delta s)^3 \right], \tag{3.38}$$

onde $\Delta s = S\,\Delta m$. Como o interesse é medir o quanto a distância entre os sinais diminui de uma modulação não-linear para uma modulação linear, a equação (3.38) foi usada com $\Delta d = |\mathbf{s}_m - \mathbf{s}_0|$ para obter a seguinte relação

$$\frac{\Delta d}{\Delta s} = \sqrt{1 - \frac{k^2}{12}(\Delta s)^2 + \frac{kk'}{6}(\Delta s)^3 + \frac{(k')^2 + k^4}{36}(\Delta s)^4}. \tag{3.39}$$

Portanto, além da verificação geométrica realizada com o auxílio da Figura 3.8, é provado matematicamente a partir de (3.39) que, se a curvatura for zero, então $\Delta d = \Delta s$. Contudo, se a curvatura for diferente de zero, então localmente $\Delta d < \Delta s$, quando $\Delta s \to 0$, condição na qual a aproximação (3.38) é bem definida.

Uma outra expressão matemática que descreve $\Delta d/\Delta s$ em função da curvatura pode ser obtida usando a lei dos senos na Figura 3.8(c), a aproximação $\sin(\theta) \approx \theta - \theta^3/6$ e a definição (3.36) para obter que

$$\frac{\Delta d}{\Delta s} \approx \frac{\sin(\Delta\alpha/2)}{\Delta\alpha/2} \approx 1 - \frac{(\Delta\alpha)^2}{12} = 1 - \frac{k^2}{12}(\Delta s)^2. \tag{3.40}$$

Assim, de (3.39) e (3.40), é possível novamente concluir que a curvatura influencia no desempenho dos sistemas de comunicações, pois se $k \neq 0$, então $\Delta d < \Delta \mathbf{s}$, o que propicia a ocorrência de erro de ponto inicial. Contudo, a perda de desempenho decorrente da existência de erros de ponto inicial, geralmente, é compensada pela diminuição da energia média ou pelo aumento do fator de expansão propiciado pelo uso de uma modulação não-linear. Entretanto, se o valor de k for suficientemente grande, a ponto da distância entre quaisquer dois sinais $\mathbf{s}_0$ e $\mathbf{s}_1$, com $m_0 \gg m_1$, diminuir ou até chegar a zero em um ponto de interseção, então o erro de ponto inicial pode aumentar de maneira significativa, aumentando o erro quadrático médio e prejudicando o desempenho do sistema de comunicações. Por este motivo, a seguir será determinado um limitante superior para o valor da curvatura em função do ruído do canal de transmissão, de maneira que o erro de ponto inicial seja aceitável.

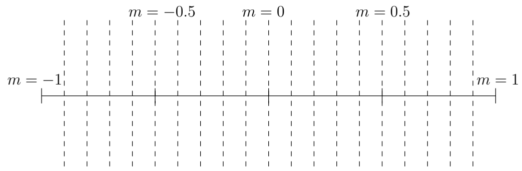
Para obter tal limitante, considere que as estimativas $\hat{m}$ são obtidas usando uma recepção de máxima verossimilhança sob a ação de um ruído AWGN, isto é, $\hat{m}$ são soluções do problema de minimizar $|\mathbf{r} - \mathbf{s}_m|$. Além disso, é interesse desse capítulo analisar a relação entre curvatura e erro de ponto inicial. Para isso, é necessário definir o conceito de "estabilidade" de $\hat{m}$ associada ao seguinte problema: *para quais valores recebidos* $\mathbf{r}$ *a estimativa* $\hat{m}$ *continua sendo a melhor solução para esse problema de otimização?* Em outras palavras, quais são os valores de $\mathbf{r}$ que podem ser estimados por um mesmo valor $\hat{m}$ de maneira que o receptor não cometa erro de ponto inicial? Neste caso, tal estabilidade pode ser calculada pela determinação da condição suficiente para que $\hat{m}$ continue sendo um ponto de mínimo local de $|\mathbf{s}_{\hat{m}} - \mathbf{r}|$, ou, no melhor dos casos, que seja o mais próximo de $\mathbf{r}$.

Geometricamente, o conceito de estabilidade associado à determinação da estimativa $\hat{m}$ pode ser interpretado com o auxílio da Figura 3.9. Neste caso, os possíveis valores de $\mathbf{r}$ que são estimados por um mesmo valor de $\hat{m}$ são representados pelas linhas tracejadas associadas aos valores de $\hat{m}$. Tal fato decorre do problema de otimização considerado, cuja solução deve satisfazer $(\mathbf{s}_{\hat{m}} - \mathbf{r}) \cdot \mathbf{s}'_{\hat{m}} = 0$, implicando que tais retas podem ser parametrizadas por

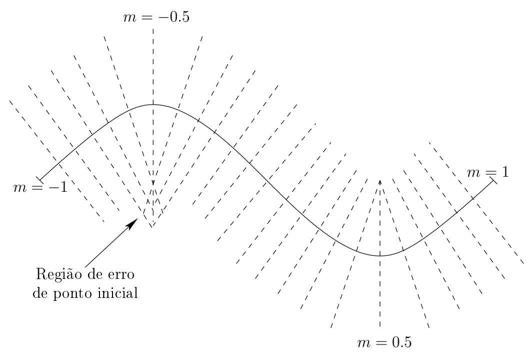
$$\mathbf{r} = \mathbf{s}_{\hat{m}} + h\vec{n}, \tag{3.41}$$

onde $\vec{n}$ é o vetor normal à curva $\mathbf{s}_m$ no ponto $\hat{m}$ e h é a variável de parametrização da reta.

Com isso, note que, quando a modulação é linear, essas retas não se intersectam, vide Figura 3.9(a), implicando que existe apenas um valor de $\hat{m}$ para todo valor $\mathbf{r}$ recebido em sua correspondente linha tracejada. Contudo, caso a modulação seja não-linear, as linhas tracejadas se cruzam e esses pontos de interseção gerados caracterizam mais que um valor para $\hat{m}$ que podem ser demodulados para o mesmo valor de $\mathbf{r}$, fato que é indesejável, veja Figura 3.9(b) na proximidade de $m = -0.5$. Portanto, quando a modulação é não-linear, tal fato pode propiciar erros de ponto inicial dependendo do valor da energia média de ruído.



(a) Em uma modulação linear.



(b) Em uma modulação não-linear.

Figura 3.9: Exemplos de aplicações do conceito de estabilidade em modulações.

Dessa forma, a estabilidade da modulação pode ser medida em função do parâmetro h de cada uma das linhas tracejadas, equação (3.41), associadas a um determinado valor de $\hat{m}$, como ilustra a Figura 3.9. Note que, no caso da modulação ser linear, h varia de $-\infty$ a ∞. Contudo, quando a modulação é não-linear, tal parâmetro pertence a um intervalo mais restrito, que depende da curvatura da modulação, como mostrado a seguir.

Para tanto, considerando um receptor de máxima verossimilhança, pode-se obter que $\hat{m}$ é um ponto de mínimo local de $|\mathbf{r} - \mathbf{s}_m|$ se satisfaz

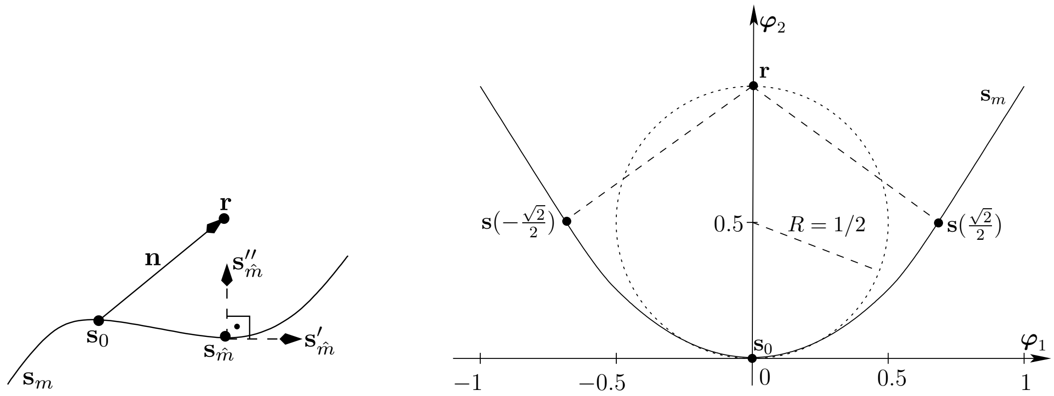
$$|\mathbf{s}'_{\hat{m}}|^2 + (\mathbf{s}_{\hat{m}} - \mathbf{r}) \cdot \mathbf{s}''_{\hat{m}} > 0, \quad \text{quando} \quad (\mathbf{s}_{\hat{m}} - \mathbf{r}) \cdot \mathbf{s}'_{\hat{m}} = 0. \tag{3.42}$$

No caso particular em que a modulação possua fator de expansão uniforme, então $\mathbf{s}'_m \cdot \mathbf{s}''_m = 0$ e a componente $(\mathbf{s}_{\hat{m}} - \mathbf{r})$ é paralela a $\mathbf{s}''_{\hat{m}}$, pois, segundo (3.42), essa componente é ortogonal a $\mathbf{s}'_m$, veja representação na Figura 3.10(a). Assim, com o auxílio de (3.41), a condição descrita em (3.42) pode ser reescrita como

$$|\mathbf{s}'_{\hat{m}}|^2 + h\vec{n} \cdot \mathbf{s}''_{\hat{m}} > 0 \quad \Rightarrow \quad 1 + h\frac{\vec{n} \cdot \mathbf{s}''_{\hat{m}}}{|\mathbf{s}'_{\hat{m}}|^2} > 0 \quad \Rightarrow \quad 1 - h\,k > 0 \quad \Rightarrow \quad h < \frac{1}{k}, \tag{3.43}$$

onde $\vec{n}$ o vetor normal dado por $[s'_2, -s'_1]$, k é a curvatura de $\mathbf{s}_m$ dada por (3.37b) e h pode ser obtida de (3.41) e da aproximação (3.5) como sendo aproximadamente a projeção do ruído $\mathbf{n}$ na direção de $\mathbf{s}''_m$, isto é,

$$h \approx \mathbf{n} \cdot \frac{\mathbf{s}''_m}{|\mathbf{s}''_m|}. \tag{3.44}$$



(a) Exemplo de recepção. (b) Exemplo de estabilidade na parábola.

Figura 3.10: Uma relação geométrica entre estabilidade e curvatura ($k = 1/R$) em uma modulação não-linear, $\mathbf{s}_m$.

Neste caso, note que, se o ruído é AWGN com densidade de potência $\mathcal{N}_0$, então h é aproximadamente uma variável aleatória Gaussiana com média nula e variância $\mathcal{N}_0/2$. Além disso, o valor $1/k$ é conhecido como *raio de curvatura* R, e, portanto, para que $\hat{m}$ seja um mínimo local, h deve ser menor do que o raio de curvatura da curva.

Com isso, observe que a condição apresentada em (3.43) representa um limitante superior para o parâmetro h definido em (3.41). Contudo, tal limitante garante apenas que $\hat{m}$ seja um mínimo local, podendo existir outros mínimos locais que são mais próximos de $\mathbf{r}$ e causar um erro de ponto inicial. Então, para evitar tal erro deve-se restringir o limitante superior para

$$h \leq \min_{m \in (-1,1)} \left\{ \frac{1}{k} \right\}. \tag{3.45}$$

Finalmente, pode-se usar (3.44) e (3.45) para projetar geometricamente uma modulação com tolerância satisfatória ao erro de ponto inicial, caso o raio de curvatura seja sempre menor ou igual a energia média de ruído por dimensão ($\mathcal{N}_0/2$).

Para exemplificar o conceito de estabilidade descrito anteriormente e o uso do limitante (3.45), considere um esquema de modulação cuja curva associada seja uma parábola $\mathbf{s}_m = [m, m^2]$, Figura 3.10(b). Neste exemplo, a curvatura calculada com o auxílio de (3.37b) é dada por

$$k(m) = \frac{2}{(1+4m^2)^{3/2}}. \tag{3.46}$$

Assim, de (3.45) e (3.46), verifica-se que a condição de estabilidade é satisfeita quando $h < 1/2$. Portanto, caso o ponto $s_0 = [0, 0]$ seja enviado e a norma do vetor ruído, $|\mathbf{n}|$, seja menor do que $R = 1/2$, então o receptor identifica apenas um mínimo local de $|\mathbf{r} - \mathbf{s}_m|$ na vizinhança de $\mathbf{s}_0$. Contudo, quando $h = 1 > 1/2 = R$, caso em que $\mathbf{n}$ seja por exemplo $[0, 1]$, então o receptor comete erro de ponto inicial, pois identifica três mínimos locais: $\mathbf{s}(-\frac{\sqrt{2}}{2})$, $\mathbf{s}(0)$ e $\mathbf{s}(\frac{\sqrt{2}}{2})$, com distâncias $\frac{\sqrt{3}}{2}$, 1 e $\frac{\sqrt{3}}{2}$ em relação a $\mathbf{r}$, respectivamente.

Uma análise geométrica do parâmetro h pode ser realizada através da curva $e_h(m) = \mathbf{s}_m + h\vec{n}$, pois dependendo do valor de h, principalmente se o mesmo variar com m, é possível associar tais curvas a um limiar de ocorrência do erro de ponto inicial. A Figura 3.11 apresenta algumas dessas curvas para alguns valores de h quando $\mathbf{s}_m$ é uma parábola. Neste caso, quando $h > 1/2$ a curva possui auto-interseção, mostrando a existência de uma região delimitando pontos de recepção que possivelmente estão sujeitos a erros de ponto inicial.

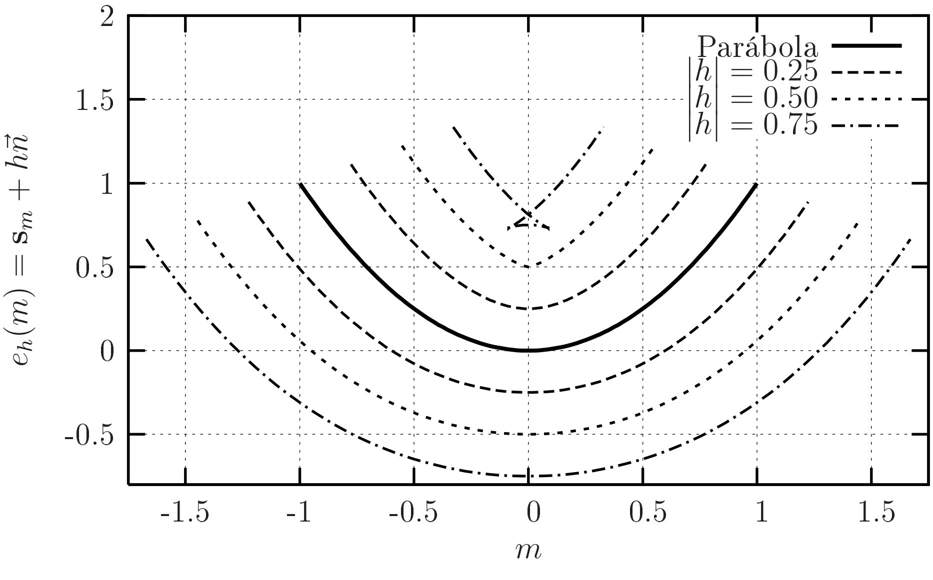


Figura 3.11: Exemplos de curvas $e_h(m) = \mathbf{s}_m + h\vec{n}$ da parábola $\mathbf{s}_m = [m, m^2]$ quando $|h| = 0$, 0.25, 0.5 e 0.75.

Portanto, o valor da curvatura deve ser levado em consideração no projeto das modulações, pois, caso seu valor seja grande, então a energia de ruído deve ser suficientemente pequena para garantir que os erros de ponto inicial não influenciem no desenvolvimento do sistema.

3.3.5 Uma expressão para $p_{\hat{m}|m_0}$ em função de S e de k

Nesta subseção, será obtida uma expressão para a função densidade de probabilidade de $\hat{m}$ dado que m_0 for transmitido, $p_{\hat{m}|m_0}$, em função do fator de expansão (S) e da curvatura (k) da curva associada a uma modulação não-linear. O conhecimento de $p_{\hat{m}|m_0}$ é de grande importância para a análise de desempenho dos sistemas de comunicações, pois tal distribuição pode ser usada para auxiliar no cálculo teórico do erro quadrático médio definido em (1.37). A relevância da obtenção do valor teórico de $\overline{\epsilon^2}$ é verificada quando o mesmo é comparado com o valor aproximado de $\overline{\epsilon^2}$ obtido em (3.7) e que não leva em consideração o erro de ponto inicial (ou a curvatura da modulação), pois, em tal aproximação, foi considerado um ruído suficientemente pequeno em (3.5) e, portanto, na vizinhança do sinal transmitido $\mathbf{s}_0$. Além disso, pode-se usar $p_{\hat{m}|m_0}$ para caracterizar a ação de um ruído aditivo em espaços mais gerais do que o Euclidiano ($k = 0$), mais precisamente o espaço dos sinais cuja métrica é induzida pela curva associada a uma modulação não-linear. Tal fato está associado a dependência de $p_{\hat{m}|m_0}$ às curvaturas (k) e às métricas (S) das curvas associadas às modulações consideradas neste capítulo.

Daremos uma atenção especial à determinação de $p_{\hat{m}|m_0}$ para modulações não-lineares, pois, no caso de uma modulação linear, tal distribuição pode ser obtida facilmente, pois, quando um sistema de comunicações usa uma modulação linear com fator de expansão A e está sujeito a ação de um ruído AWGN com densidade espectral de potência $\mathcal{N}_0/2$, então pode-se verificar de (3.2) que $p_{\hat{m}}$ possui distribuição Gaussiana com média m_0 e variância $\mathcal{N}_0/(2A^2)$. Isto porque, na saída do receptor, $\hat{m}$ assume o valor $m_0 + n_1/A$, onde n_1 é a projeção do ruído na direção de $\boldsymbol{\varphi}_1$.

Contudo, quando se considera uma modulação não-linear, geralmente, a relação entre $\hat{m}$ e m_0 deixa de ser linear e muitas vezes difícil de ser expressa analiticamente, comprometendo assim a obtenção de $p_{\hat{m}|m_0}$. Para amenizar tal problema foi utilizado um artifício matemático (uma mudança de variáveis aleatórias conveniente) que possibilitou expressar $p_{\hat{m}}$ em função do ruído do canal de transmissão, do fator de expansão e da curvatura da curva associada a modulação.

Com o intuito de caracterizar a mudança de variáveis citada anteriormente, considere inicialmente que o ruído seja aditivo, isto é, o sinal recebido é igual ao sinal enviado mais o ruído do canal, $\mathbf{r} = \mathbf{s}_0 + \mathbf{n}$. Além disso, caso o receptor seja de máxima verossimilhança, então a estimativa $\hat{m}$ satisfaz $(\mathbf{r} - \mathbf{s}_{\hat{m}}) \cdot \mathbf{s}'_{\hat{m}} = 0$ e o vetor $(\mathbf{r} - \mathbf{s}_{\hat{m}})$ é normal à curva, podendo ser escrito da forma $\mathbf{r} = \mathbf{s}_{\hat{m}} + h\vec{n}$, como indica (3.41). Portanto, fazendo uso desses resultados, é possível expressar a mudança de variáveis da seguinte maneira,

$$\mathbf{n} = \mathbf{s}_{\hat{m}} - \mathbf{s}_0 + h\vec{n}, \quad \Rightarrow \quad \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} - \begin{bmatrix} s_1(0) \\ s_2(0) \end{bmatrix} + \frac{h}{S}\begin{bmatrix} s_2' \\ -s_1' \end{bmatrix}, \tag{3.47}$$

onde $\vec{n}$ é o vetor normal da curva $\mathbf{s}_m$ em $m = \hat{m}$, h é a variável aleatória contínua citada anteriormente e S é o fator de expansão.

Em outras palavras, a equação (3.47) expressa variáveis aleatórias do ruído (n_1 e n_2) em função das variáveis aleatórias $\hat{m}$ e h. Como a densidade de probabilidade do ruído é geralmente conhecida ($p_{\mathbf{n}}$), então pode-se usá-la para obter a densidade de probabilidade conjunta de $\hat{m}$ e h por $p_{\hat{m},h} = p_{\mathbf{n}}\,|J|$, onde $|J|$ é o Jacobiano dessa mudança de variáreis. Neste caso, podemos obter que o Jacobiano é dado por

$$|J| = \det\begin{bmatrix} \dfrac{d\,n_1}{dh} & \dfrac{d\,n_1}{d\hat{m}} \\ \\ \dfrac{d\,n_2}{dh} & \dfrac{d\,n_2}{d\hat{m}} \end{bmatrix} = \det\begin{bmatrix} \dfrac{s_2'}{S} & s_1' + h\left[\dfrac{s_2''}{S} - \dfrac{s_2'}{S^3}(s_1' + s_2')\right] \\ \\ -\dfrac{s_1'}{S} & s_2' - h\left[\dfrac{s_1''}{S} - \dfrac{s_1'}{S^3}(s_1' + s_2')\right] \end{bmatrix} = S\,(1 + h\,k), \tag{3.48}$$

onde k é a curvatura da curva $\mathbf{s}_m$, definida em (3.37b).

Assim, para o caso particular em que o ruído do canal de transmissão é AWGN, pode-se usar a mudança de variáveis (3.47) e seu correspondente Jacobiano (3.48) para determinar que a função densidade de probabilidade conjunta de $\hat{m}$ e h é dada por

$$p_{\hat{m},h} = \frac{1}{\pi\mathcal{N}_0}\exp\left(-\frac{|\mathbf{n}|^2}{\mathcal{N}_0}\right)|J| = \frac{1}{\pi\mathcal{N}_0}\exp\left(-\frac{|\mathbf{s}_{\hat{m}} - \mathbf{s}_0 + h\vec{n}|^2}{\mathcal{N}_0}\right)S\,(1 + h\,k). \tag{3.49}$$

Para obter a distribuição de $p_{\hat{m}|m_0}$, deve-se integrar $p_{\hat{m},h}$ em função de h, para h variando de h_1 a h_2,

$$p_{\hat{m}|m_0} = \frac{S}{2}\left[\frac{1 - k(\mathbf{s}_{\hat{m}} - \mathbf{s}_0)\cdot\vec{n}}{\sqrt{\pi\mathcal{N}_0}}\exp\left(-\frac{|(\mathbf{s}_{\hat{m}} - \mathbf{s}_0)\cdot\mathbf{s}_{\hat{m}}'|^2}{S^2\mathcal{N}_0}\right)(\mathrm{erf}_2 - \mathrm{erf}_1) + \frac{k}{\pi}\,(\mathrm{e}_1 - \mathrm{e}_2)\right], \tag{3.50}$$

onde

$$\mathrm{e}_i = \exp\left(-\frac{|\mathbf{s}_{\hat{m}} - \mathbf{s}_0 + h_i\vec{n}|^2}{\mathcal{N}_0}\right) \quad \text{e} \quad \mathrm{erf}_i = \mathrm{erf}\left(\frac{h_i + (\mathbf{s}_{\hat{m}} - \mathbf{s}_0)\cdot\vec{n}}{\sqrt{\mathcal{N}_0}}\right),$$

com

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt.$$

Portanto, como proposto, o valor teórico de $p_{\hat{m}|m_0}$ foi expresso em função do fator de expansão (S) e da curvatura (k) da modulação. Além disso, deve-se ressaltar que o resultado obtido em (3.50) é válido mesmo que o fator de expansão S não seja uniforme.

A fim de determinar os limites h_1 e h_2 deste resultado, considere que h delimita o conjunto de valores de $\mathbf{r}$ que são mais próximos de $\mathbf{s}_{\hat{m}}$ do que de qualquer outro valor de $\mathbf{s}_m$. Tal conjunto de pontos será chamado de região de decisão, $\mathcal{R}_m$, para cada valor de $\hat{m}$, e será melhor entendido com o auxílio dos exemplos apresentados a seguir. Contudo, o valor de h, geralmente, não deve variar de $-\infty$ a ∞, senão alguns sinais que não deveriam ser decididos por $\hat{m}$ são computados no cálculo dessa integral, principalmente, aqueles pontos que caracterizam erros de ponto inicial no processo de recepção.

Com o intuito de exemplificar a validade da densidade de probabilidade expressa em (3.50), considere uma modulação linear $\mathbf{s}_m = Am\boldsymbol{\varphi}_1$, na qual os limitantes de h são dados por $h_1 = -\infty$ e $h_2 = \infty$ como ilustra a Figura 3.9(a), para obter

$$p_{\hat{m}|m_0} = \frac{S}{\sqrt{\pi \mathcal{N}_0}} \exp\left(-\frac{|(\mathbf{s}_{\hat{m}} - \mathbf{s}_0) \cdot \mathbf{s}'_{\hat{m}}|^2}{S^2 \mathcal{N}_0}\right) \left(1 - k(\mathbf{s}_{\hat{m}} - \mathbf{s}_0) \cdot \vec{n}\right), \qquad (3.51)$$

mas como $k = 0$ e $S = A$ na modulação linear, então este resultado pode ser reescrito como

$$p_{\hat{m}|m_0} = \frac{A}{\sqrt{\pi \mathcal{N}_0}} \exp\left(-\frac{A^2(\hat{m} - m_0)^2}{\mathcal{N}_0}\right),$$

que como esperado é a densidade de probabilidade Gaussiana com média m_0 e variância $\sigma^2 = \mathcal{N}_0/(2A^2)$, pois $\overline{\epsilon^2} = \overline{(\hat{m} - m_0)^2} = \sigma^2 = \mathcal{N}_0/(2A^2)$. Neste caso, note que esse resultado obtido está de acordo com (3.3).

A seguir, apresentamos um outro exemplo de aplicação da expressão (3.50), só que agora considerando uma modulação não-linear cuja curva associada é a parábola, $\mathbf{s}_m = [m, m^2]$, abordada anteriormente. Neste caso, inicialmente, deve-se determinar os valores de h_1 e h_2 estudando as regiões de decisão representadas pelas retas tracejadas da Figura 3.12 que são associadas aos valores de $\mathbf{r}$ e são estimados pelo mesmo valor de $\hat{m}$.

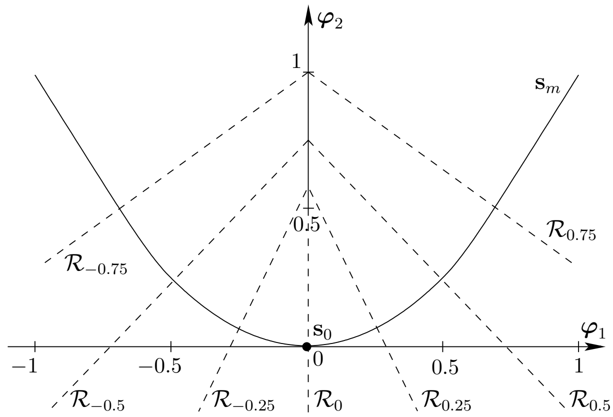
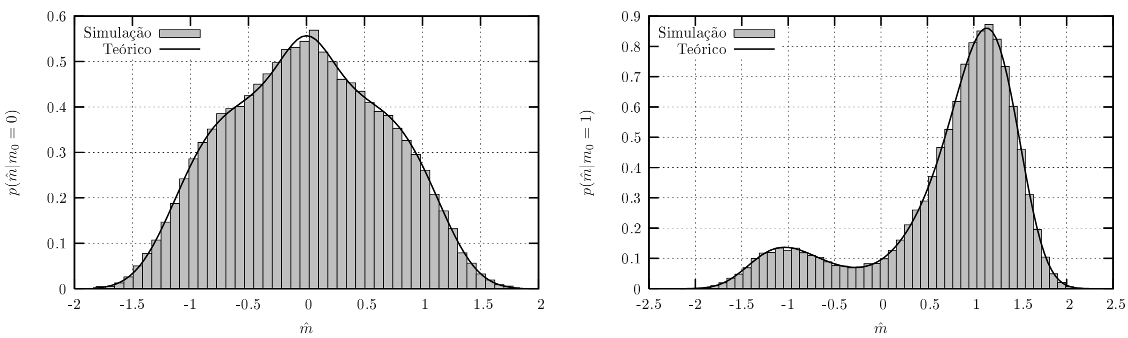


Figura 3.12: Exemplos de regiões de decisão, $\mathcal{R}_m$, para a parábola quando $m_0 = 0$.

Assim, com o auxílio da Figura 3.12, note que, devido à simetria da modulação em relação ao eixo $\overrightarrow{oy}$, pode-se dividir a recepção em duas regiões, isto é, os sinais recebidos com $x > 0$ ($x < 0$) sempre são estimados para valores de $\hat{m} > 0$ ($\hat{m} < 0$). Com base neste fato, é possível usar (3.41) e determinar que

$$h_1 = -\frac{\sqrt{1+4m^2}}{2} \quad \text{e} \quad h_2 = \infty. \tag{3.52}$$

A Figura 3.13 apresenta os gráficos da distribuição de probabilidade $p_{\hat{m}}$ obtidos teoricamente por (3.50) quando (3.52) é satisfeita e via simulação computacional para o exemplo em questão, considerando que foram transmitidos $m_0 = 0$ e $m_0 = 1$ e o sistema está sujeito a ação de um ruído AWGN com $\mathcal{N}_0 = 2$. Observe que os resultados obtidos usando simulação não apresentam grande discrepância quando comparados aos resultados teóricos, indicando assim a aplicabilidade do uso de (3.50) como ferramenta prática no cálculo teórico de $\overline{\epsilon^2}$ e também a validade do uso da simulação para medir esse parâmetro de desempenho.



(a) $m_0 = 0$, $\overline{\hat{m}} = 0$ e $\overline{\epsilon^2}(m_0) = 0.45235$. (b) $m_0 = 1$, $\overline{\hat{m}} = 0.72253$ e $\overline{\epsilon^2}(m_0) = 0.70047$.

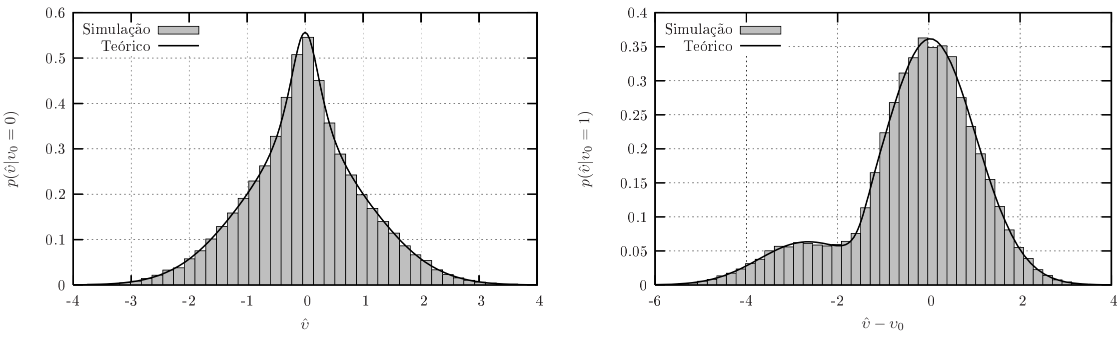
Figura 3.13: Resultados teóricos e de simulação para $p_{\hat{m}|m_0}$ considerando um esquema de modulação associado a parábola, sob a ação de um ruído AWGN com $\mathcal{N}_0 = 2$.

Como neste capítulo também foram realizadas análises de desempenho de sistemas de comunicações usando modulações não-lineares com fator de expansão uniforme, então considere que a parábola tratada anteriormente seja parametrizada pelo comprimento de arco, usando o seguinte parâmetro de reparametrização

$$\upsilon = \frac{m}{2}\sqrt{1+4m^4} + \frac{1}{4}\sinh^{-1}(2m).$$

Nestas condições, as distribuições de $p_{\hat{\upsilon}}$ para os mesmos sinais de transmissão e mesmas condições de ruído considerados no sistema anterior são apresentadas na Figura 3.14.

Observe que, neste caso, também foi verificado a validade do uso de (3.50) para determinar a densidade de probabilidade de $p_{\hat{m}|m_0}$ e conseqüentemente do desempenho do sistema de comunicações, pois foi possível calcular $\overline{\epsilon^2}$ em ambos os casos considerados.



(a) $v_0 = 0$, $\overline{\hat{v}} = 0$ e $\overline{\epsilon^2}(v_0) = 1$. (b) $v_0 = 1.48$, $\overline{\hat{v}} = -0.29193$ e $\overline{\epsilon^2}(v_0) = 2.0393$.

Figura 3.14: Resultados teóricos e de simulação para $p_{\hat{v}|v_0}$ considerando um esquema de modulação associado a parábola p.p.c.a, sob a ação de um ruído AWGN com $\mathcal{N}_0 = 2$.

3.4 Exemplos de Modulações Não-Lineares

Nas seções anteriores, foi realizado um estudo sobre a influência de alguns parâmetros geométricos no desempenho de modulações não-lineares. Nesta ocasião, foi apresentado, assim como em [18], que o fator de expansão sob determinadas condições pode diminuir o erro quadrático médio na transmissão. Além disso, foi mostrado que a curvatura é um outro parâmetro relevante no desempenho das modulações consideradas, principalmente quando o objetivo é diminuir o erro de ponto inicial ou, equivalentemente, o erro quadrático médio.

Nesta seção, consideramos a interpretação geométrica (1.36), que associa modulações a curvas do espaço Euclidiano, para analisar o desempenho de alguns sistemas de comunicações usando modulações não-lineares. Neste caso, será analisada a modulação angular (PM e FM), uma modulação associada à curva espiral de Arquimedes, veja [19] e [28], e uma nova técnica de modulação que foi proposta levando em consideração o comprimento, a curvatura e a minimização da energia média de transmissão, sempre com o intuito de obter um sistema de comunicações melhor do que os citados anteriormente.

3.4.1 Modulação angular

Uma maneira bastante conhecida de modular uma portadora senoidal é usando uma modulação angular. Nesta modulação, a amplitude da portadora é mantida constante e seu ângulo é variado de acordo com o sinal em banda base, $m(t)$. Este método de modulação não-linear possui a importante característica de fornecer uma maior imunidade ao ruído quando comparada com a modulação em amplitude. Entretanto, o ganho de desempenho é obtido às custas do aumento da largura de banda de transmissão.

Neste caso, considere $\theta(t)$ o ângulo de uma portadora senoidal modulada, então o sinal modulado pode ser expresso como

$$s(t) = A\cos[\theta(t)]\,, \tag{3.53}$$

onde A é a amplitude da portadora. Além disso, a freqüência instantânea, $f_i(t)$, desse sinal pode ser definida como sendo $\frac{1}{2\pi}\frac{d\theta(t)}{dt}$, caso $\theta(t)$ varie continuamente com o tempo.

Existe uma infinidade de maneiras distintas pelas quais o ângulo $\theta(t)$ pode ser variado em função do sinal $m(t)$. Entretanto, na prática, são usados apenas dois métodos: modulação em fase (PM) e modulação em freqüência (FM).

Na modulação em fase, o ângulo $\theta(t)$ varia linearmente com o sinal $m(t)$, isto é

$$\theta(t) = 2\pi f_c t + k_p m(t),$$

onde f_c é a freqüência da portadora e k_p é desvio de fase do modulador. Portanto, o sinal modulado em fase $s_m(t)$ é descrito por

$$s_m(t) = A\cos\left[2\pi f_c t + k_p m(t)\right]. \tag{3.54}$$

Na modulação em freqüência, é a freqüência instantânea $f_i(t)$ que varia linearmente com o sinal $m(t)$, isto é

$$f_i(t) = f_c + k_f m(t) \quad \Rightarrow \quad \theta(t) = 2\pi f_c t + 2\pi k_f \int_0^t m(t)\,dt,$$

onde f_c é novamente a freqüência da portadora e k_f é o desvio de freqüência do modulador. Portanto, o sinal modulado em freqüência $s_m(t)$ é descrito por

$$s_m(t) = A\cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(t)\,dt,\right]. \tag{3.55}$$

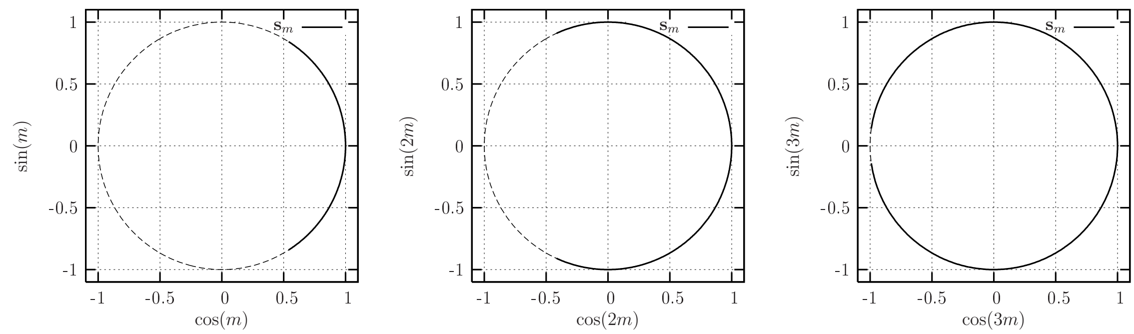
No intuito de analisar o desempenho da modulação angular consideraremos a interpretação geométrica descrita em (1.36) para o caso particular da modulação em fase, pois pode-se generalizar os resultados obtidos para a modulação em freqüência considerando na modulação PM que $m(t)$ seja dado por $c\int g(t)dt$, onde c é uma constante e $g(t)$ é o sinal modulante FM.

Com isso, usando uma expansão trigonométrica simples, o sinal (3.54) pode ser representado na forma descrita por (1.34) com portadoras $\varphi_1(t) = \cos(2\pi f_c t)$ e $\varphi_2(t) = \sin(2\pi f_c t)$ da seguinte maneira

$$s_m(t) = A\cos\left[k_p m(t)\right]\varphi_1(t) - A\sin\left[k_p m(t)\right]\varphi_2(t), \tag{3.56}$$

ou, equivalentemente, na forma vetorial (1.36), como mostrado a seguir

$$\mathbf{s}_m = A\cos(k_p m)\boldsymbol{\varphi}_1 - A\sin(k_p m)\boldsymbol{\varphi}_2 = A\left[\cos(k_p m), -\sin(k_p m)\right]. \tag{3.57}$$

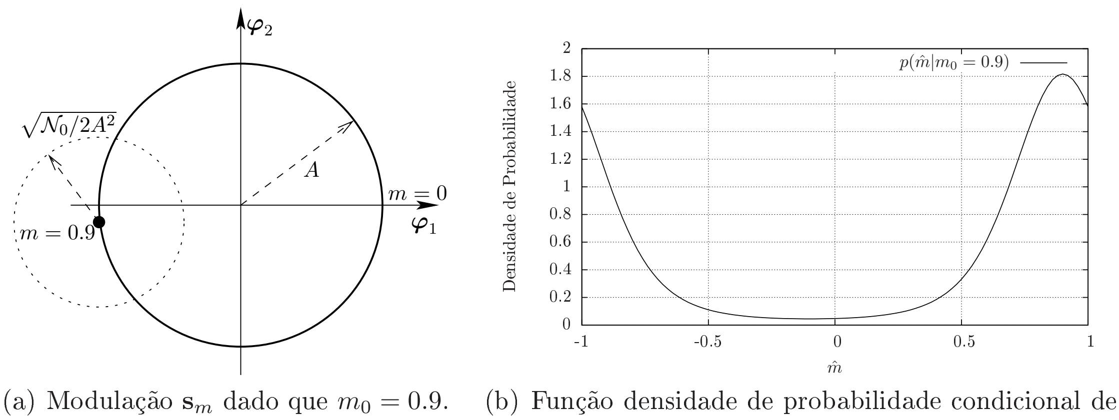


(a) $k_p = 1$. (b) $k_p = 2$. (c) $k_p = 3$.

Figura 3.15: Representação geométrica da modulação em fase (PM), para $k_p = 1, 2, 3$ e $A = 1$.

Assim, de (3.57) verifica-se que tal equação representa a parametrização de um círculo de raio A, veja Figura 3.15. Tal interpretação era esperada, pois, como mencionado anteriormente, a amplitude da onda modulada é constante e igual a A.

Na prática, uma conseqüência não desejável de permitir que o ângulo $\theta(t)$ da portadora se torne dependente do sinal de mensagem $m(t)$ é que os "cruzamentos de zero" de um sinal PM ou FM podem ocasionar uma descontinuidade na fase estimada pelo receptor; cruzamentos de zero referem-se aos instantes de tempo em que a curva associada a modulação cruza o eixo ox mudando de um valor negativo para um positivo ou vice-versa. Tal fato, certamente, está associado às auto-interseções da curva $\mathbf{s}_m$ quando $|k_p| > \pi$ e ao aumento do erro de ponto inicial, veja Figura 3.16. Neste caso, $m_0 = 0.9$ é enviado, Figura 3.16(a), contudo, o receptor possui uma probabilidade significativa de estimar um valor para $\hat{m}$ próximo de -1, veja a função densidade de probabilidade condicional de $\hat{m}$ dado que $m_0 = 0.9$, Figura3.16(b), calculada por (3.50).



(a) Modulação $\mathbf{s}_m$ dado que $m_0 = 0.9$. (b) Função densidade de probabilidade condicional de $\hat{m}$ dado que $m_0 = 0.9$.

Figura 3.16: Análise de erro de ponto inicial para uma modulação PM quando $k_p = \pi$, $\mathcal{N}_0/A^2 = 1$ e $m_0 = 0.9$.

Para minimizar as conseqüências dos cruzamentos de zero no processo de demodulação, geralmente, considera-se que o sinal $\theta(t)$ (ou $m(t)$) seja contínuo no tempo. Assim, pode-se projetar detectores de fase ou freqüência que utilizam essa informação de maneira coerente, isto é, que saibam diferenciar por exemplo $\theta(t) = \pi$ de $\theta(t) = -\pi$, dado que ambos os sinas são representados pelo mesmo ponto sobre a curva associada à modulação, $[-A, 0]$.

Uma análise espectral para a modulação angular pode ser encontrada, por exemplo, em [18] e em [37] para o caso particular de uma modulação de uma única freqüência, cujo sinal modulante senoidal é definido por

$$m(t) = A_m \cos(2\pi f_m t), \tag{3.58}$$

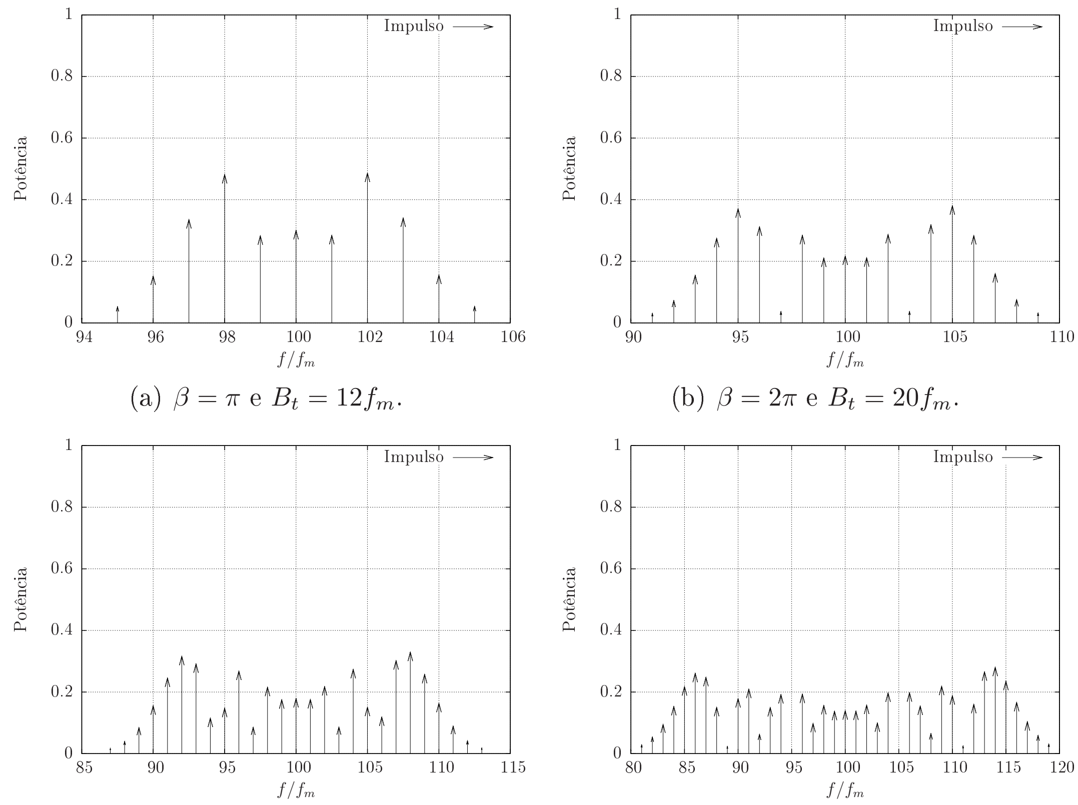
onde A_m é a amplitude do sinal modulante, que no nosso caso vale 1, pois foi considerado que $m \in (-1, 1)$.

Neste caso, a análise espectral da modulação angular foi realizada em função da razão do desvio de freqüência pela freqüência modulante, que é comumente chamada de índice de modulação β e está associada à dimensão efetiva da modulação. Na modulação FM, $\beta = k_f/f_m$ representa o desvio de fase do sinal $s_m(t) = A\cos[2\pi f_c t + \beta \sin(2\pi f_m t)]$, isto é, a variação máxima do ângulo $\theta(t)$ em relação ao ângulo $2\pi f_c t$ da portadora não-modulada; portanto, β é medido em radianos. Na modulação PM, o valor do índice de modulação é igual a k_p.

Dessa forma, a análise espectral descrita em [18] e [37] foi realizada primeiramente para um sinal modulado de banda estreita (β é pequeno em relação a um radiano), o qual se assemelha a um sinal AM ($B_t = 2f_m$), e, posteriormente, para um sinal modulado de banda larga (β é grande em relação a um radiano). Neste último caso, o sinal FM foi aproximado pela série exponencial de Fourier, cujos coeficientes dependem da função de Bessel e do índice de modulação β.

Seria possível prosseguir no estudo e realizar uma análise espectral do sinal FM ou PM para o caso mais geral de um sinal modulado por um sinal de várias freqüências. Entretanto, decidiu-se por não fazê-lo, pois seria um aumento de complexidade desnecessário, dado que a estimativa da largura de banda de transmissão (B_t) de um sinal FM de uma única freqüência é obtida empiricamente por meio da regra de Carson $B_t \approx 2f_m(1+\beta)$, veja [37], e é suficiente para determinar na prática a largura de banda dos sinais.

A Figura 3.17 apresenta alguns gráficos da análise espectral do sinal PM, $\mathbf{s}_m = A\cos[2\pi f_c t + \beta\cos(2\pi f_m t)]$. Essa análise foi obtida com o auxílio da transformada discreta de Fourier (DFT), implementada computacionalmente e simulada considerando $f_c = 100 f_m$, para alguns valores de β. Além disso, deve-se ressaltar que, na análise dos resultados da simulação, foram considerados apenas os impulsos com amplitudes normalizadas maiores ou iguais a 1%. Como era de se esperar, quanto maior for o valor do



(a) $\beta = \pi$ e $B_t = 12f_m$. (b) $\beta = 2\pi$ e $B_t = 20f_m$.

(c) $\beta = 3\pi$ e $B_t = 26f_m$. (d) $\beta = 5\pi$ e $B_t = 40f_m$.

Figura 3.17: Espectros de amplitudes discretas de um sinal PM, normalizados em relação à amplitude da portadora, para o caso do sinal modulante senoidal de freqüência fixa f_m. São mostrados somente os espectros correspondentes às freqüências positivas.

índice de modulação maior é a largura de banda, pois como mencionado anteriormente, β está associada a dimensão efetiva da modulação.

A análise espectral apresentada na Figura 3.17 foi comparada com resultados teóricos descritos em [37] e demonstrou estar bastante próxima do esperado (discrepância menor que 0.1%). Dessa forma, como a análise espectral em modulações não-lineares, geralmente, é um processo trabalhoso e com muitas aproximações, então a DFT foi implementada computacionalmente para analisar os espectros das modulações não-lineares propostas neste capítulo, dado que os resultados obtidos por simulação demonstraram ser satisfatórios aos nossos propósitos de obter a largura de banda de transmissão do sinal modulado.

A seguir, proseguiremos com a análise geométrica do desempenho desse esquema de modulação angular associada à curva parametrizada em (3.57). Para tanto, foi calculado o fator de expansão (3.8) e a curvatura (3.37b) dessa curva que, como visto anteriormente, são parâmetros que influenciam no cálculo do erro quadrático médio

(3.10).

$$S = Ak_p = A\beta\,, \quad \overline{\epsilon^2} = \frac{\mathcal{N}_0}{2A^2k_p^2} = \frac{\mathcal{N}_0}{2A^2\beta^2} \quad \text{e} \quad k = \frac{1}{A}. \tag{3.59}$$

Portanto, a partir de (3.59), podemos verificar que, quanto maior for o valor de $A\beta$, melhor será o desempenho do sistema. Em outras palavras, o sistema pode ser melhorado com o aumento da energia de transmissão, $E_s = A^2$, ou com o aumento da largura de banda do sinal, isto é, com o aumento de β, que é a dimensão efetiva do sinal modulado. Este fato caracteriza bem o artifício de trocar energia média de transmissão por banda de transmissão que é realizado na prática.

O desempenho do sistema de comunicações usando uma modulação em fase (PM) foi calculado com o auxílio de (1.41), (3.57) e (3.59), como segue:

$$\text{CSNR} = \frac{\overline{E}_m}{\mathcal{N}} = \frac{A^2}{\mathcal{N}_0}\,, \quad \text{SNR} = \frac{\overline{m^2}}{\overline{\epsilon^2}} = \frac{2A^2k_p^2\overline{m^2}}{\mathcal{N}_0} \quad \text{e} \quad \text{SNR} = 2k_p^2\overline{m^2}\text{CSNR}.$$

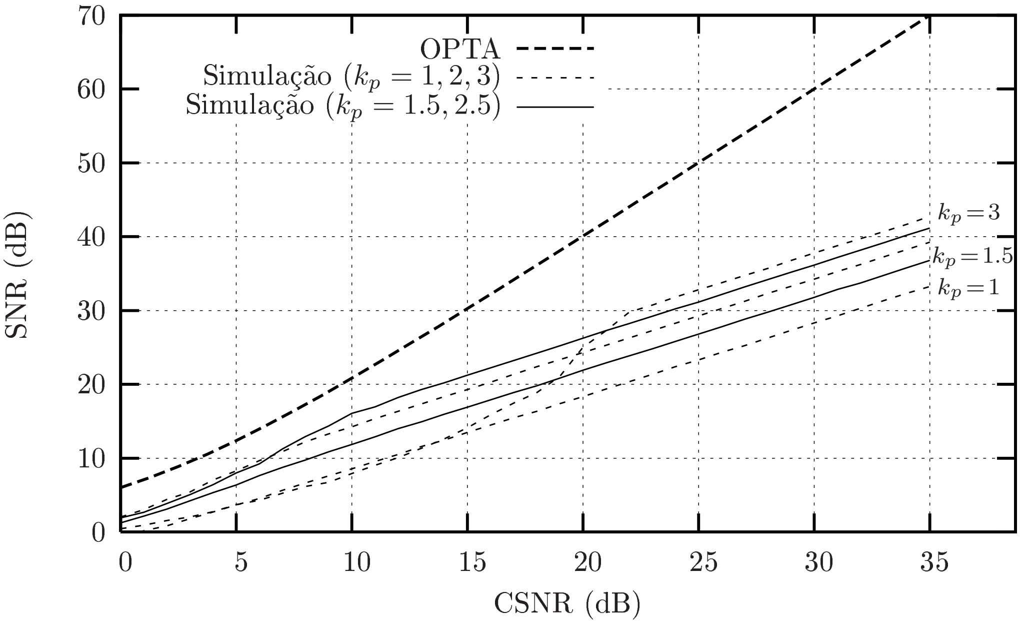


Figura 3.18: Curvas de desempenho da modulação em fase (PM), para $k_p = 1$, 1.5, 2, 2.5 e 3.

As curvas de CSNR *versus* SNR do sistema considerado são apresentadas na Figura 3.18. Tais curvas foram obtidas usando simulação computacional, semelhante a realizada na obtenção da cuva apresentada na Figura 3.5 para $k_p = 1$, 1.5, 2, 2.5 e 3 quando m é uma variável aleatória contínua com densidade de probabilidade uniforme. Como esperado, foi comprovado que, quanto maior o valor de k_p, melhor o desempenho, Porém, quando o sistema possui $k_p = 3$ e CSNR menor do que 20 dB pode-se observar que seu desempenho é inferior ao do sistema com $k_p < 3$. Tal fato está certamente associado ao erro de ponto inicial que ocorre na vizinhança do ponto $[-1, 0]$ da

Figura 3.15(c), na qual valores enviados de m próximos de -1 são demodulados para valores próximos de 1 e vice-versa.

3.4.2 Modulação espiral de Arquimedes

A modulação associada à curva espiral de Arquimedes uniforme foi inicialmente analisada em [19] e posteriormente em [28]. Tal modulação apresentou um desempenho satisfatório e é um dos primeiros exemplos, não típicos, da aplicação da interpretação geométrica (1.36). Por esse motivo, tal modulação será apresentada neste momento. Além disso, uma análise espectral dessa modulação será apresentada, bem como seu desempenho será analisado em função dos parâmetros CSNR e SNR, e serão feitas comparações com os sistemas de comunicações considerados neste capítulo.

O sinal modulado proveniente desse esquema de modulação será definido de duas formas diferentes, a saber

$$\begin{aligned} s_m(t) &= A\, m(t) \cos\left[2\pi f_c t + k_p |m(t)|\right], \qquad \text{(AM-PM)} && \text{(3.60a)} \\ &= A\, m(t) \cos\left[2\pi f_c t + 2\pi k_f \left|\int_0^t m(t)\, dt\right|\right], \qquad \text{(AM-FM)} && \text{(3.60b)} \end{aligned}$$

onde k_p e k_f são constantes que medem o desvio da modulação em relação à fase e à freqüência, respectivamente.

Comparando o sinal modulado (3.60) com os sinais derivados das modulações linear (3.1) e angular (3.53), podemos verificar que a modulação espiral de Arquimedes possui simultaneamente características desses dois esquemas de modulação. Mais precisamente, (3.60a) possui simultaneamente características das modulações AM e PM, enquanto (3.60b) possui simultaneamente características das modulações AM e FM. A Figura 3.19 apresenta um exemplo de geração de um sinal modulado do tipo (3.60a).

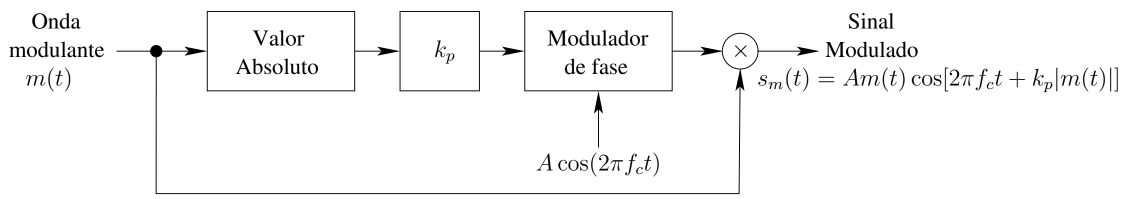
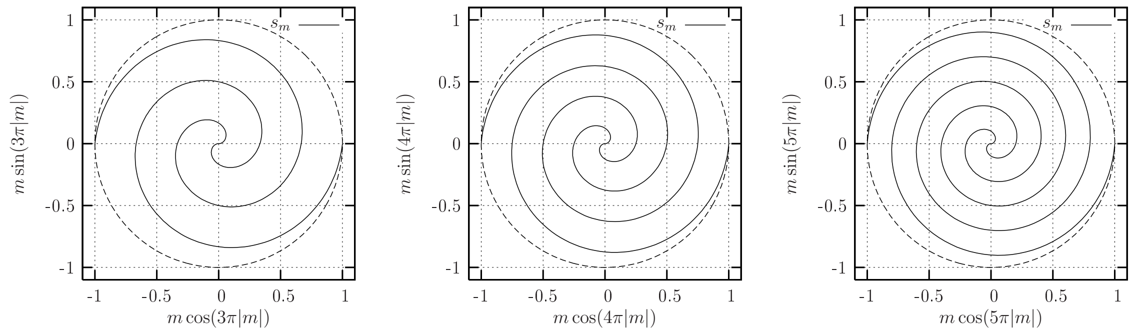


Figura 3.19: Esquema para gerar um sinal modulado associado à curva espiral de Arquimedes utilizando um modulador de fase e um multiplicador de sinais.

Com o intuito de interpretar geometricamente esse esquema de modulação, (3.60a) foi reescrita em sua forma vetorial considerando $\varphi_1(t) = \cos(2\pi f_c t)$ e $\varphi_2(t) = \sin(2\pi f_c t)$ como sua base de sinais ortogonais. Com isso, foi verificado que essa modulação está associada a uma curva do $\mathbb{R}^2$ com a seguinte parametrização

$$\mathbf{s}_m = A\, m\left[\cos(k_p\, |m|), -\sin(k_p |m|)\right]. \qquad (3.61)$$

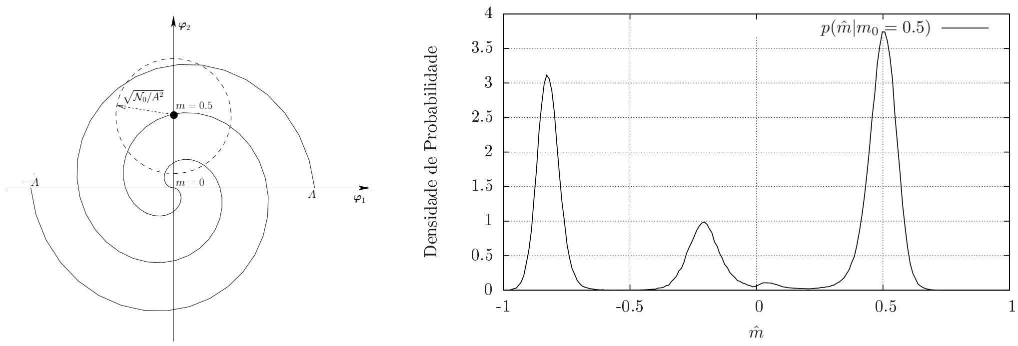
A curva parametrizada por (3.61) é conhecida como espiral de Arquimedes uniforme e sua representação gráfica é mostrada na Figura 3.20 para alguns valores de k_p.



(a) $k_p = 3\pi$.　(b) $k_p = 4\pi$.　(c) $k_p = 5\pi$.

Figura 3.20: Representação geométrica da modulação espiral de Arquimedes (AM-PM), para $k_p = \pi, 2\pi, 3\pi$ e $A = 1$.

Observe que, geometricamente, o valor de k_p indica a quantidade de voltas da espiral e, conseqüentemente, quanto maior for o valor de k_p maior será o comprimento da curva. Portanto, de (3.10), o erro quadrático médio de um sistema de comunicações usando essa modulação pode ser diminuído com o aumento de k_p, pois $\overline{\epsilon^2}$ é inversamente proporcional ao quadrado do fator de expansão da curva. Contudo, caso o valor de A seja mantido constante e o valor de k_p seja aumentado, então a distância entre as espirais certamente irá diminuir ($d = \pi A/k_p$), aumentando assim o erro de ponto inicial pela "aproximação" dos sinais. Tal fato pode ser verificado com o auxílio da Figura 3.21, que representa o envio de $m_0 = 0.5$ em um canal ruidoso, cujo raio de ação do ruído é caracterizado pelo círculo tracejado da Figura 3.21(a). Neste caso, o sinal enviado possui grande probabilidade de ser estimado para $\hat{m} = -0.2$ e $\hat{m} = -0.83$, como sugere $p(\hat{m}|m_0 = 0.5)$ na Figura 3.21(b).



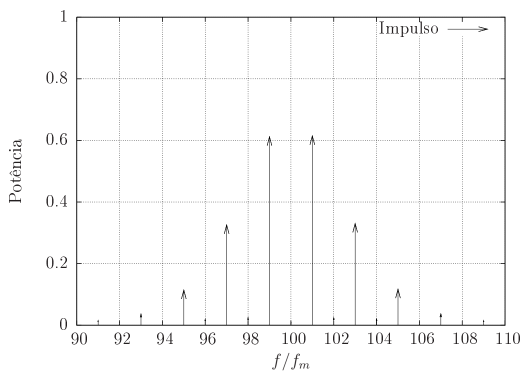
(a) $\mathbf{s}_m$ com $m_0 = 0.5$.　(b) $p(\hat{m}|m_0 = 0.5)$ e $\overline{\epsilon^2} = 0.665$.

Figura 3.21: Análise de erro de ponto inicial para uma modulação PM quando $k_p = 3\pi$, $\mathcal{N}_0/A^2 = 1/8$ e $m_0 = 0.5$.

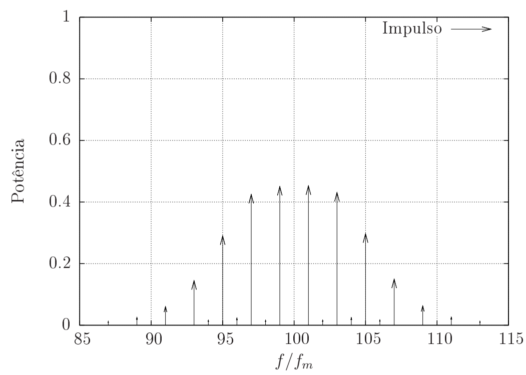
A seguir, a análise espectral do sinal modulado (3.60b) será obtida de uma maneira bastante semelhante à que foi realizada anteriormente para a modulação angular, utilizando a transformada discreta de Fourier. Neste caso, considera-se novamente um sinal modulante de uma única freqüência como o apresentado em (3.58) e os índices de modulação $\beta = k_p$ e $\beta = k_f/f_m$ para as modulações AM-PM e AM-FM, respectivamente. Tais considerações são necessárias para facilitar a comparação dessa modulação com a modulação angular. Nestas condições, o sinal modulado AM-PM é dado por

$$s(t) = A\cos(2\pi f_m t)\cos\left[2\pi f_c t + \beta|\cos(2\pi f_m t)|\right]. \tag{3.62}$$

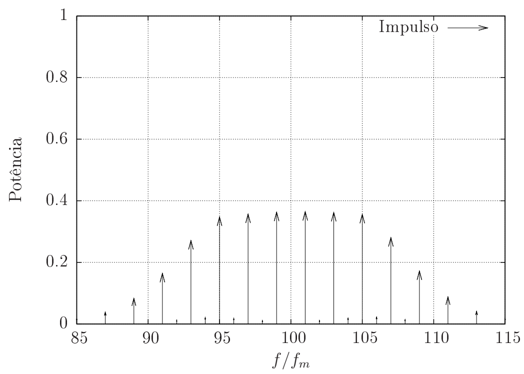
Com o auxílio da transformada discreta de Fourier, foi realizada uma simulação computacional para obter o espectro do sinal (3.62) considerando que $f_c = 100 f_m$, veja Figura 3.22. Assim como ocorreu na modulação angular, o valor de β também aumenta a largura de banda de transmissão do sinal AM-PM. Contudo, podemos observar que, para o mesmo valor de β, a modulação AM-PM necessita de um pouco mais de largura de banda do que o sinal modulado em PM.



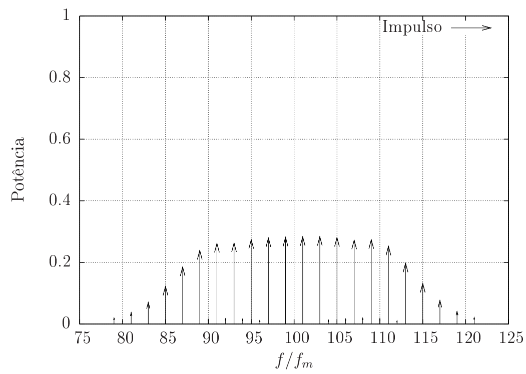
(a) $\beta = \pi$ e $B_t = 18 f_m$.



(b) $\beta = 2\pi$ e $B_t = 26 f_m$.



(c) $\beta = 3\pi$ e $B_t = 30 f_m$.



(d) $\beta = 5\pi$ e $B_t = 42 f_m$.

Figura 3.22: Espectros de amplitudes discretas de um sinal AM-PM associado à curva espiral de Arquimedes, normalizados em relação à amplitude A, para o caso do sinal modulante senoidal de freqüência fixa f_m. São mostrados somente os espectros correspondentes às freqüências positivas.

Pode-se verificar que a análise espectral apresentada na Figura 3.22 permite estimar a largura de banda de transmissão da modulação AM-FM, que é um parâmetro prático relevante na implementação dessa técnica de modulação.

Para a modulação AM-PM, o fator de expansão, o erro quadrático médio e a curvatura são dados respectivamente por

$$S = A\sqrt{1 + k_p^2 m^2}, \quad \overline{\epsilon^2} = \frac{\mathcal{N}_0}{2A^2(1 + k_p^2 m^2)} \quad \text{e} \quad k = -\frac{k_p(2 + k_p^2 m^2)}{A(1 + k_p^2 m^2)^{3/2}}. \tag{3.63}$$

Analisando a curvatura dessa modulação pode-se verificar que, próximo à origem do sistema de coordenadas (quando $m = 0$) o sistema está mais sujeito à ação danosa do ruído, mais precisamente à ocorrência de erro de ponto inicial. Este fato, como explicado anteriormente, está associado ao valor do módulo da curvatura neste ponto, que, no caso vale, $2k_p/A$ e coincide com seu valor máximo. Além disso, quando $k_p \to \infty$ o módulo da curvatura dessa modulação tende a $1/A$, se aproximando do valor da curvatura da modulação PM, e quando $k_p \to 0$, então $k \to 0$, se aproximando do valor da curvatura da modulação AM.

Substituindo os parâmetros S e $\overline{\epsilon^2}$ de (3.63) em (1.41), obtemos que o desempenho de um sistema de comunicações usando a modulação AM-PM é dado por

$$\text{CSNR} = \frac{A^2\overline{m^2}}{\mathcal{N}_0}, \quad \text{SNR} = \frac{2A^2\overline{m^2}(1 + k_p^2 m^2)}{\mathcal{N}_0} \quad \text{e} \quad \text{SNR} = \text{CSNR}(1 + k_p^2 m^2) \tag{3.64}$$

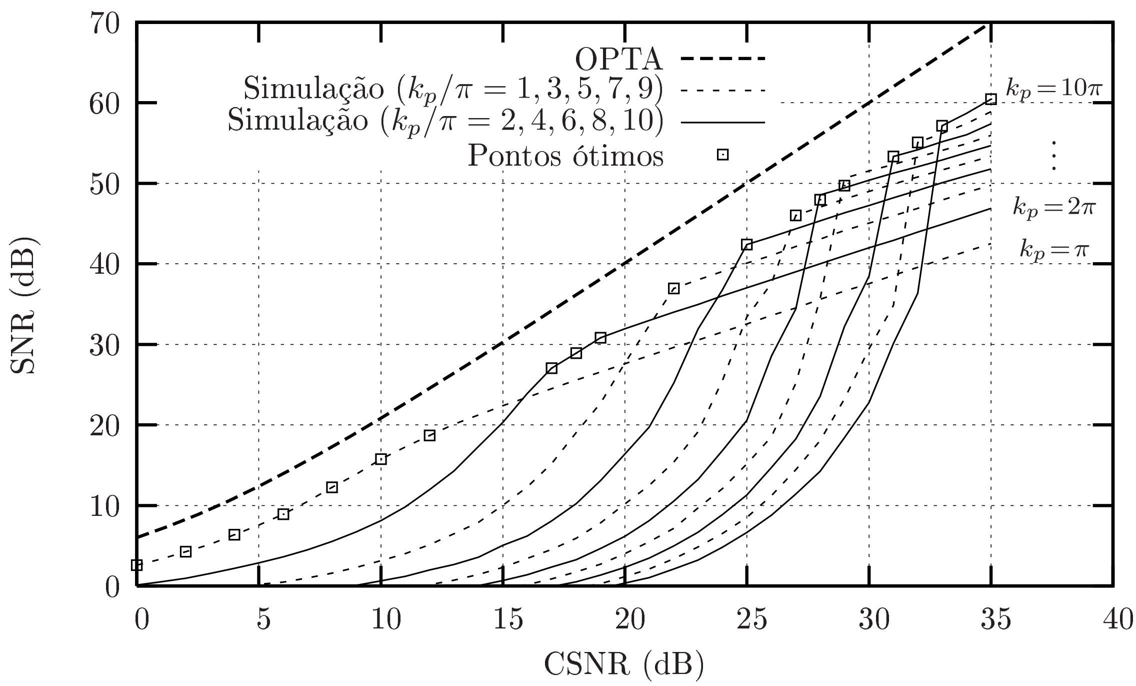


Figura 3.23: Curvas de desempenho da modulação espiral de Arquimedes (AM-PM) para $k_p = \pi, 2\pi, \ldots, 10\pi$.

O desempenho desse esquema de modulação foi calculado também por simulação computacional. Neste caso, novamente foi considerado que m possui densidade de probabilidade uniforme e que o fator expansão seja uniforme. As curvas simuladas de CSNR *versus* SNR são mostradas na Figura 3.23 e essas curvas obtidas se assemelham aos resultados apresentados em [28].

Observe que para todas as curvas apresentadas na Figura 3.23 existe um ponto ótimo de desempenho, no qual o aumento do valor de CSNR não aumenta o ganho em dB do sistema, pois, acima desse ponto, a curva se aproxima da reta descrita em (3.64). Além disso, para valores de CSNR abaixo desse ponto ótimo, o desempenho cai rapidamente, pois, neste caso, o erro de ponto inicial se torna cada vez mais significativo, prejudicando o sistema. Tal fato indica que, dependendo do valor da energia média de ruído, existe um valor ideal para k_p que pode maximizar o desempenho do sistema com a diminuição dos efeitos prejudiciais do erro de ponto inicial, mas aproveitando ao máximo o aumento do comprimento da curva para diminuir o $\overline{\epsilon^2}$.

Tal fato pode ser melhor entendido com a análise da Tabela 3.1. Neste caso, note que o ponto ótimo de desempenho do sistema AM-PM pode ser obtido quando o valor de $\sqrt{\mathcal{N}_0/2A^2}$ for aproximadamente duas vezes menor do que A/S e d/A, que são parâmetros que influenciam no $\overline{\epsilon^2} = \mathcal{N}_0/2S^2$ e no erro de ponto inicial, respectivamente.

k_p (rad)	CSNR (dB)	$\sqrt{\mathcal{N}_0/2A^2}$	A/S	$d/A = \pi/2k_p$
π	12	0.3076	0.5142	0.5000
2π	19	0.1374	0.2956	0.2500
3π	22	0.0973	0.2043	0.1667
4π	25	0.0689	0.1555	0.1250
5π	27	0.0547	0.1253	0.1000
6π	28	0.0488	0.1049	0.0833
7π	29	0.0435	0.0901	0.0714
8π	31	0.0345	0.0790	0.0625
9π	32	0.0308	0.0703	0.0556
10π	33	0.0274	0.0634	0.0500

Tabela 3.1: Análise dos pontos ótimos de desempenho da modulação AM-PM em relação a k_p.

3.4.3 Modulação senoidal

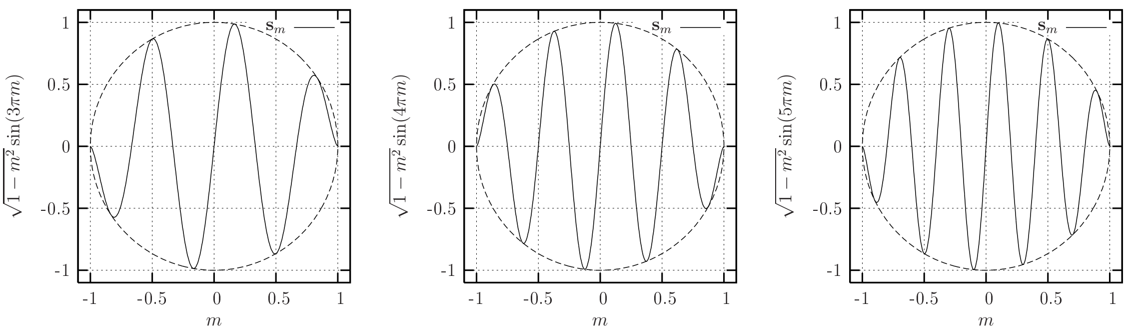
A modulação apresentada a seguir foi projetada com o auxílio da interpretação geométrica (1.36) com o intuito de obter um novo esquema de modulação, cujo desempenho seja melhor do que os apresentados nas modulações consideradas anteriormente. Para obter uma modulação com esse desempenho, foi necessário exigir que a curva associada a essa modulação apresentasse algumas características geométricas importantes, tais como: um comprimento elevado para diminuir o erro quadrático médio, não possuir auto-interseção e ter uma curvatura pequena para minimizar as ocorrências de erros de ponto inicial e também uma energia média de transmissão reduzida para que o aumento no desempenho do sistema possa ser atingido.

Contudo, determinar uma curva com tais propriedades a partir da formulação e resolução de um problema de otimização satisfazendo essas condições, certamente, não é um trabalho fácil. Para facilitar a obtenção dessa curva, foi considerado que geometricamente ela seja similar à curva apresentada na Figura 3.7(a), que, graficamente, aparenta satisfazer as condições impostas à curva desejada.

Com isso, podemos determinar empiricamente que uma possível parametrização para essa curva seja dada por

$$\mathbf{s}_m = A[m, \sqrt{1-m^2}\sin^{\gamma}(\beta m)], \quad m \in (-1,1), \tag{3.65}$$

onde $\beta \in \mathbb{R}$ e $\gamma \in [0,1]$ são constantes desse esquema de modulação que influenciam em seu desempenho. A Figura 3.24 apresenta algumas curvas da parametrização (3.65) para alguns valores de β e $\gamma = 1$. Neste caso, observe que essas curvas são senóides de diferentes freqüências e amplitudes limitada por um círculo de raio A. Por esse motivo, a modulação em questão será denominada modulação senoidal.

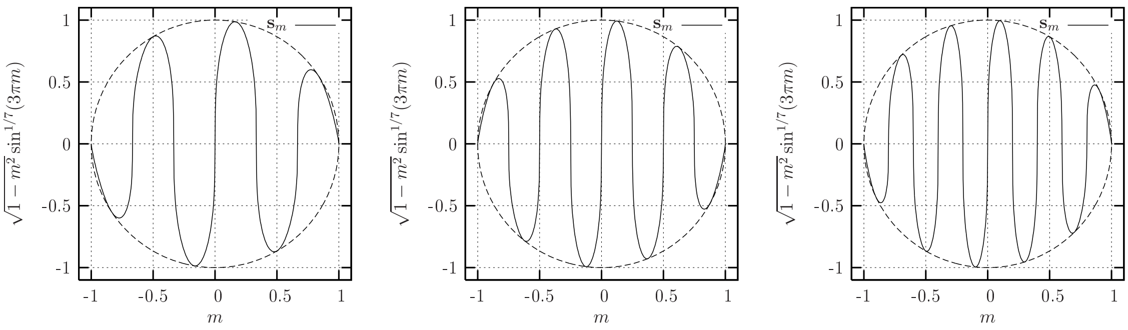


(a) $\beta = 3\pi$. (b) $\beta = 4\pi$. (c) $\beta = 5\pi$.

Figura 3.24: Representação geométrica para a modulação senoidal, quando $\beta = 3\pi$, 4π e 5π, $\gamma = 1$ e $A = 1$.

A Figura 3.25 apresenta algumas outras curvas calculadas usando a parametrização (3.65) para $\gamma = 1/7$. Comparando a Figura 3.25 com a Figura 3.24, pode-se observar que as curvas com $\gamma = 1/7$ apresentam menor curvatura e maior comprimento do que as curvas com $\gamma = 1$. Este fato será provado matematicamente mais adiante e relacionado ao desempenho dessa modulação.

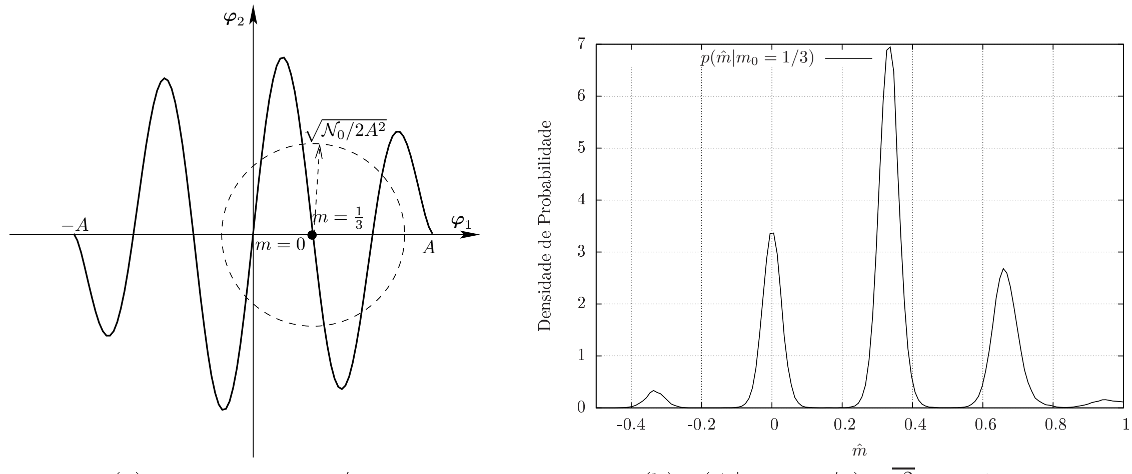
Além disso, no decorrer da análise de desempenho da modulação senoidal, as constantes β e γ serão associadas aos seguintes parâmetros de um sistema de comunicações: largura de banda de transmissão, erro quadrático médio e erro de ponto inicial. Contudo, primeiramente, é interessante notar de (3.65) que, quando $\beta = 0$, essa modulação é semelhante à modulação linear (a curva é uma reta) e, quando $\gamma = 0$, essa modulação é semelhante à modulação PM com $k_p = \pi/2$ (a curva é um círculo).



(a) $\beta = 3\pi$. (b) $\beta = 4\pi$. (c) $\beta = 5\pi$.

Figura 3.25: Representação geométrica para a modulação senoidal, quando $\beta = 3\pi$, 4π e 5π, $\gamma = 1/7$ e $A = 1$.

Observe que, geometricamente, o valor de β indica a quantidade de "picos" da senóide. Conseqüentemente, quanto maior for o valor de β, maior será o comprimento da respectiva curva. Com isso, o erro quadrático médio de um sistema de comunicações usando essa modulação pode ser diminuído com o aumento de β, pois $\overline{\epsilon^2}$ é inversamente proporcional ao quadrado do fator de expansão da curva, como indica a equação (3.10). Contudo, caso o valor de A seja mantido constante e o valor de β seja aumentado, então a distância entre dois ramos da senóide certamente irá diminuir ($d = \pi A/\beta$), aumentando assim o erro de ponto inicial pela "aproximação" dos sinais. Tal fato pode ser melhor compreendido com o auxílio da Figura 3.26, que representa o envio de $m_0 = 1/3$ em um canal ruidoso, cujo raio de ação do ruído é caracterizado pelo círculo tracejado da Figura 3.26(a). Neste caso, o sinal enviado possui grande probabilidade de ser estimado para $\hat{m} = 0$, $\hat{m} = 1/3$ e $\hat{m} = 2/3$, como sugere $p(\hat{m}|m_0 = 1/3)$ na Figura 3.26(b).



(a) $\mathbf{s}_m$ com $m_0 = 1/3$. (b) $p(\hat{m}|m_0 = 1/3)$ e $\overline{\epsilon^2} = 0.0703$.

Figura 3.26: Análise de erro de ponto inicial para uma modulação senoidal quando $\beta = 3\pi$, $\gamma = 1$, $\mathcal{N}_0/A^2 = 1/2$ e $m_0 = 1/3$.

Neste momento, um fato que deve ser ressaltado é que o erro de ponto inicial na modulação senoidal não aumenta o valor do erro quadrático médio tanto quanto o mesmo erro de ponto inicial na modulação AM-PM. Para verificar tal fato, compare as Figuras 3.21 e 3.26 e seus respectivos valores de $\overline{\epsilon^2}$ para os pontos transmitidos. Neste caso, note que um erro de ponto inicial na modulação AM-PM, geralmente, implica em uma estimativa $\hat{m}$ de sinal contrário ao do valor enviado m_0, fato que normalmente não ocorre na modulação senoidal, pois o segmento da curva referente a $m > 0$ está totalmente contida no semiplano $x > 0$. Portanto, quando ocorre um erro de ponto inicial na modulação senoidal, o valor de $(\hat{m} - m_0)^2$ é menor do que o mesmo valor na modulação AM-PM, fato que certamente indica um melhor desempenho da modulação senoidal em relação a modulação AM-PM.

Além disso, para efeito de comparação da largura de banda de transmissão da modulação senoidal com a largura de banda das outras modulações apresentadas, considere novamente o sinal modulante de uma única freqüência, (3.58). Neste caso, o sinal modulado usando a modulação senoidal é dado por

$$s_m(t) = A\left[\cos(2\pi f_m t)\cos(2\pi f_c t) + |\sin(2\pi f_m t)|\sin(\beta\cos(2\pi f_m t))\sin(2\pi f_c t)\right]. \quad (3.66)$$

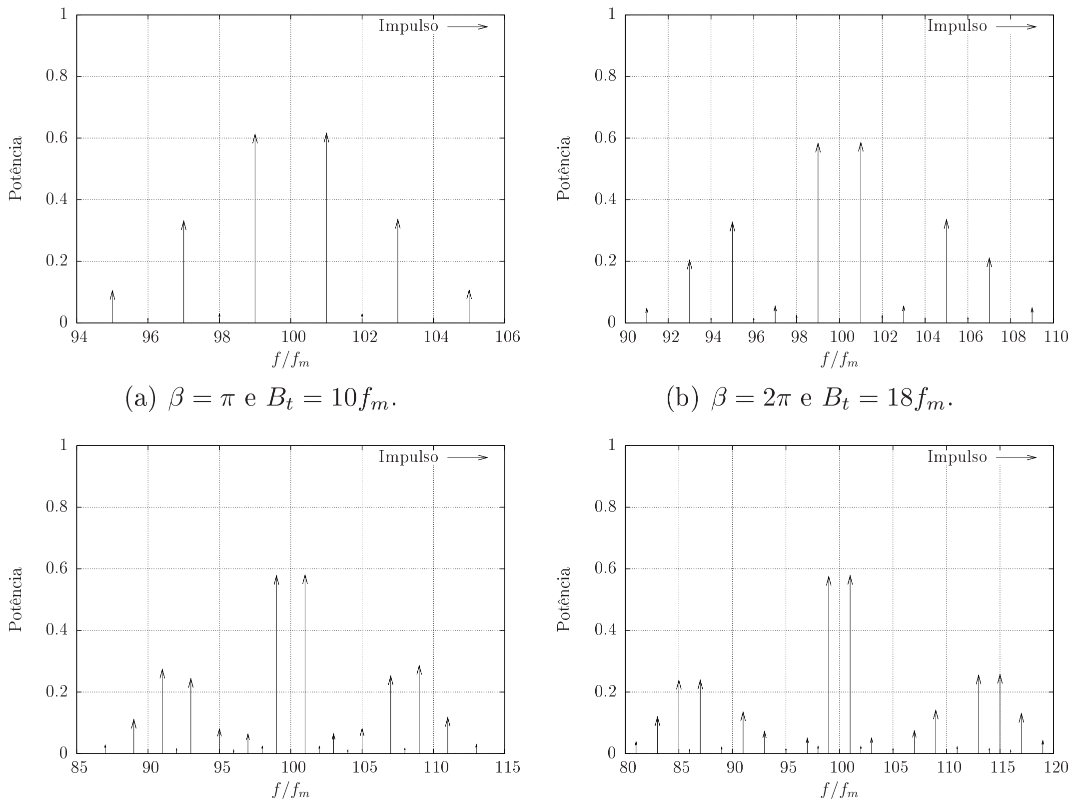
Com o auxílio da transformada discreta de Fourier foi realizada uma simulação computacional para obter a análise espectral do sinal (3.66) considerando que $f_c = 100 f_m$, veja Figura 3.27. Note que, assim como ocorreu anteriormente, o valor de β também aumenta a largura de banda de transmissão da modulação senoidal. Contudo, podemos observar que para o mesmo valor de β a modulação senoidal necessita de um pouco menos de largura de banda que o sinal modulado em PM.

A análise espectral apresentada na Figura 3.27 para a modulação senoidal possibilita determinar uma estimativa da largura de banda de transmissão dessa modulação. Entretanto, deve-se chamar a atenção para o fato dessa análise ter sido realizada apenas para $\gamma = 1$, mas foi verificado que quando este valor diminui, então o valor da largura de banda de transmissão aumenta. Contudo, quando a análise espectral é realizada considerando um fator de expansão uniforme, então o parâmetro γ não influencia tanto no cálculo de B_t.

Para realizar uma análise de desempenho da modulação senoidal, é necessário determinar alguns parâmetros geométricos da curva associada a essa modulação. O primeiro desses parâmetros é o fator de expansão, pois o erro quadrático médio é inversamente proporcional ao quadrado de seu valor.

$$S^2 = 1 + \sin(\beta m)^{2\gamma}\left(-\frac{m}{\sqrt{1-m^2}} + \frac{\gamma\beta\sqrt{1-m^2}}{\tan(\beta m)}\right)^2. \quad (3.67)$$

Como a análise de desempenho dessa modulação foi realizada considerando um fator de expansão uniforme, e apresentar analiticamente uma reparametrização para (3.65)



(a) $\beta = \pi$ e $B_t = 10f_m$.

(b) $\beta = 2\pi$ e $B_t = 18f_m$.

(c) $\beta = 3\pi$ e $B_t = 26f_m$.

(d) $\beta = 5\pi$ e $B_t = 38f_m$.

Figura 3.27: Espectros de amplitudes discretas de um sinal associado à modulação senoidal, normalizados em relação à amplitude A, para o caso do sinal modulante senoidal de freqüência fixa f_m. São mostrados somente os espectros correspondentes às freqüências positivas.

não é viável, então alguns valores de S uniformes foram calculados e apresentados na Tabela 3.2 com o auxílio de (3.11), para $\beta = \pi,\ 2\pi, \ldots, 9\pi$ e $\gamma = 1$ e $1/7$.

S	$\beta = \pi$	$\beta = 2\pi$	$\beta = 3\pi$	$\beta = 4\pi$	$\beta = 5\pi$	$\beta = 6\pi$	$\beta = 7\pi$	$\beta = 8\pi$	$\beta = 9\pi$
$\gamma = 1$	2.0894	3.5195	5.0200	6.5472	8.0878	9.6361	11.189	12.746	14.305
$\gamma = 1/7$	2.3293	3.8426	5.3544	6.8762	8.4066	9.9436	11.486	13.032	14.581

Tabela 3.2: Alguns valores para o fator de expansão uniforme da modulação senoidal para $\beta = \pi, 2\pi, \ldots, 9\pi$ e $\gamma = 1$ e $1/7$.

Como indicado anteriormente, um outro parâmetro geométrico que influencia no desempenho de modulações é a curvatura. Contudo, devido à complexidade de sua expressão analítica para a modulação senoidal, é mais interessante comparar algumas de suas curvas para alguns valores de γ e verificar sua influência no desempenho da modulação. A Figura 3.28 mostra dois gráficos da curvatura dessa modulação para $\beta = 3\pi$ e $\gamma = 1$ e $1/7$. Neste caso, observe que um valor de γ menor do que 1 diminui a curvatura da modulação, fato extremamente desejável para diminuir a perda de desempenho causada pelo erro de ponto inicial.

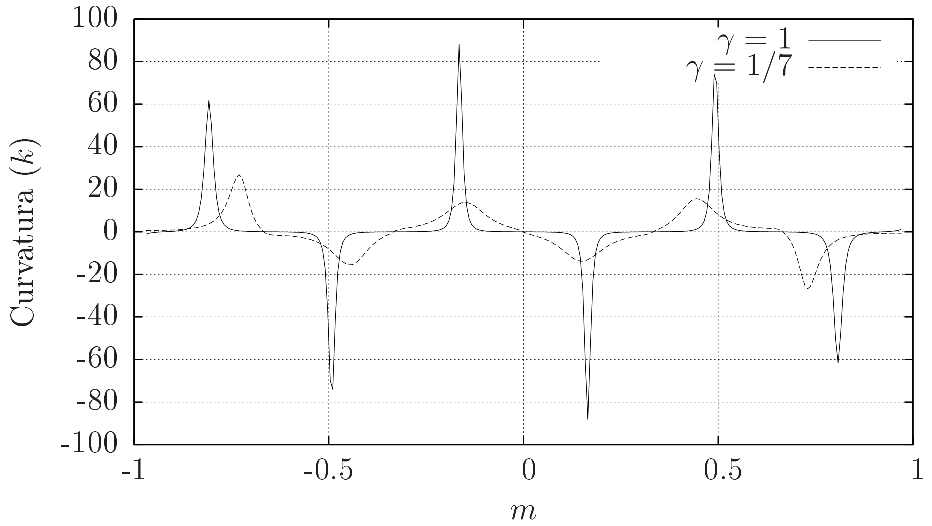


Figura 3.28: Curvatura da modulação senoidal para $\beta = 3\pi$ e $\gamma = 1$ e $1/7$.

De posse dessas informações, o desempenho da modulação senoidal foi calculado também por simulação computacional. Neste caso, novamente foi considerado que m possui densidade de probabilidade uniforme e que o fator de expansão é uniforme. As curvas simuladas de CSNR *versus* SNR são mostradas na Figura 3.29.

Dessa forma, comparando as curvas de desempenho das Figuras 3.23 e 3.29, podemos verificar que os pontos ótimos de desempenho da modulação senoidal possuem um valor de SNR menor do que o apresentado na modulação AM-PM. Contudo, analisando as curvas de desempenho da modulação senoidal, note que seus desempenhos abaixo dos pontos ótimos não caem tão rapidamente como na Figura 3.23. Este fato certamente está associado à diminuição da influência do erro de ponto inicial no valor de $\overline{\epsilon^2}$, devido à geometria da curva associada à modulação, como descrito na explicação da Figura 3.26.

Além do desempenho apresentado na Figura 3.29 para a modulação senoidal, outras curvas de CSNR *versus* SNR considerando $\gamma = 1/7$ foram determinadas para essa modulação. Tais curvas são apresentadas na Figura 3.30 para $\beta = \pi, 2\pi, \ldots, 10\pi$. Neste caso, note que o desempenho do sistema de comunicações usando essa modulação com tais parâmetros aumentou significativamente, a ponto de alcançar o desempenho da modulação AM-PM. Sobretudo, deve-se ressaltar que seu desempenho para valores abaixo dos pontos ótimos não decai tão rapidamente como na modulação AM-PM, demonstrando assim a superioridade da senoidal.

Neste momento, é importante chamar a atenção para o fato do aumento de desempenho apresentado na Figura 3.29 não ser decorrente do aumento do fator de expansão, que, como mostra a Tabela 3.2, não é significativo, mas sim a diminuição do valor de sua curvatura, como indica a Figura 3.28. Este fato novamente comprova a importância de se considerar a curvatura como um parâmetro relevante no projeto de modulações.

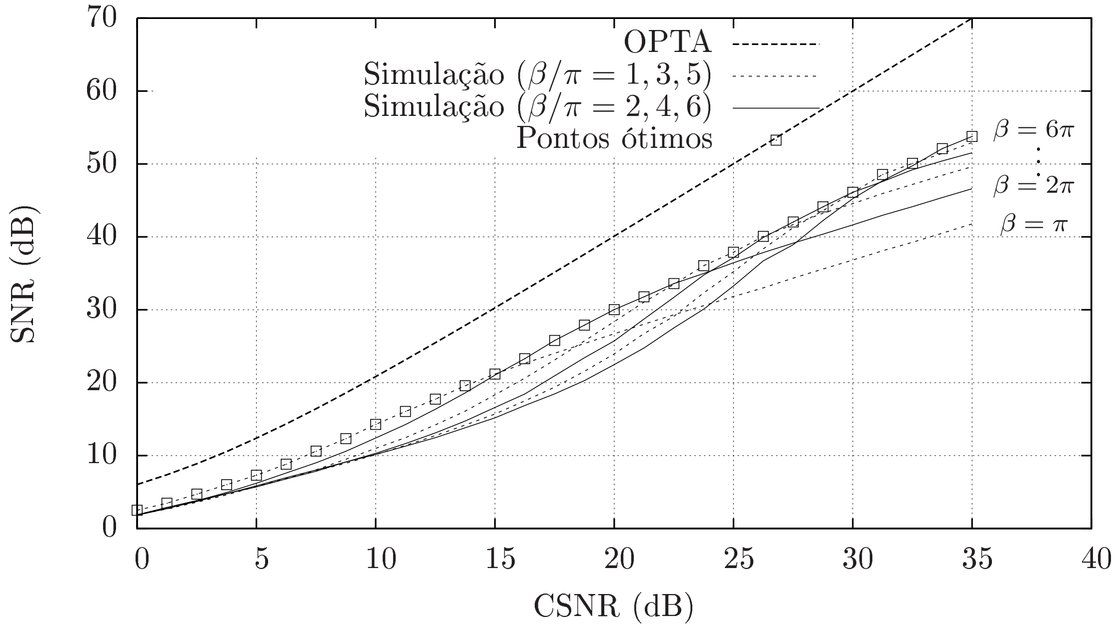


Figura 3.29: Curvas de desempenho da modulação senoidal para $\beta = \pi, 2\pi, \ldots, 6\pi$ e $\gamma = 1$.

A Figura 3.31 mostra as curvas ótimas de desempenho das modulações AM-PM e senoidal com o objetivo de comparar seus desempenhos. Observe que os desempenhos dessas modulações estão muito próximos, desde que $\gamma = 1/7$ na modulação senoidal.

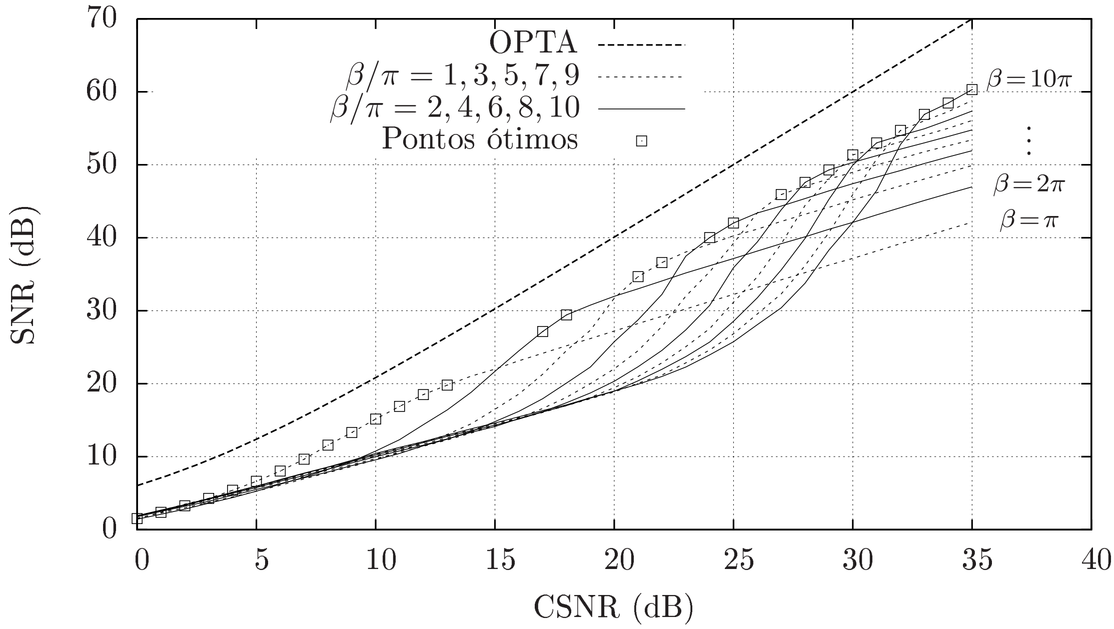


Figura 3.30: Curvas de desempenho da modulação senoidal para $\beta = \pi, 2\pi, \ldots, 10\pi$ e $\gamma = 1/7$.

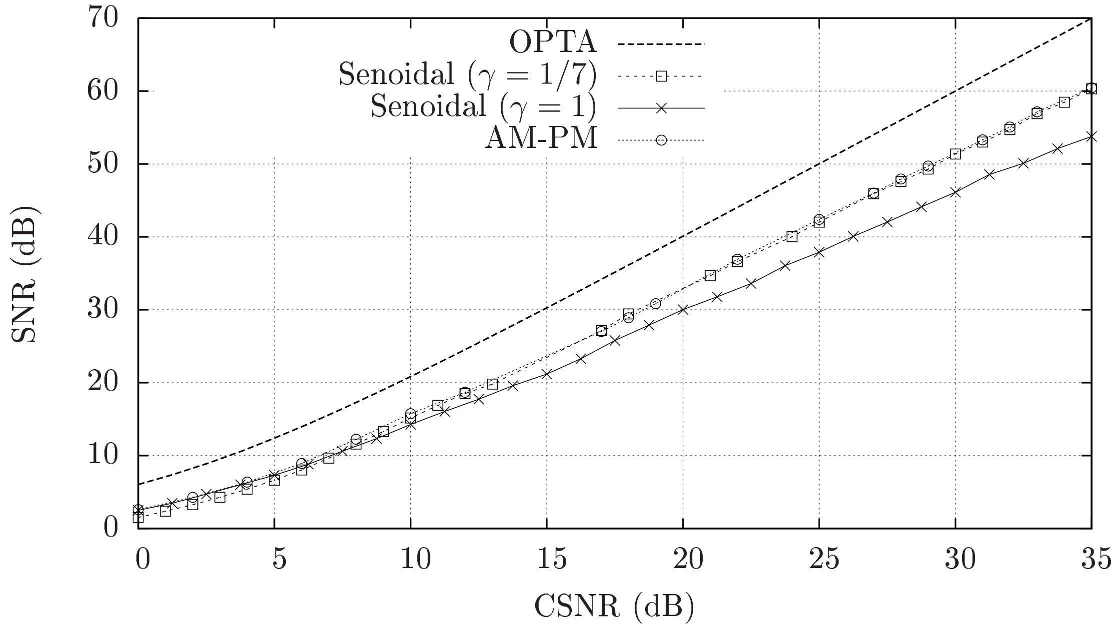


Figura 3.31: Comparação entre os desempenhos das modulações AM-PM e senoidal.

3.5 Modulação Não-Linear $(M : N)$

Apresentaremos nesta seção um esquema de modulação não-linear que transmite um vetor de variáveis aleatórias $\mathbf{m} = [m_1, m_2, \ldots, m_M]$, $m_i \in (-1, 1)$, através do seguinte sinal modulado

$$\begin{aligned} s_{\mathbf{m}}(t) &= s_1(\mathbf{m})\varphi_1(t) + s_2(\mathbf{m})\varphi_2(t) + \cdots + s_N(\mathbf{m})\varphi_N(t) && (3.68a) \\ &= s_1(\mathbf{m})\boldsymbol{\varphi}_1 + s_2(\mathbf{m})\boldsymbol{\varphi}_2 + \cdots + s_N(\mathbf{m})\boldsymbol{\varphi}_N \\ &= [s_1(\mathbf{m}), s_2(\mathbf{m}), \ldots, s_N(\mathbf{m})]\,, && (3.68b) \end{aligned}$$

onde $\varphi_i(t)$, $i = 1, \ldots, N$, é um conjunto de sinais ortogonais de energia unitária e $\boldsymbol{\varphi}_i$, $i = 1, \ldots, N$, é uma base de vetores ortonormais do espaço Euclidiano N-dimensional. Por este motivo, a modulação não-linear tratada nesta seção será denotada por $(M : N)$, enquanto os esquemas de modulação determinados por (1.36) poderiam ter sido denotados por $(1 : N)$.

Observe que, geometricamente, a parametrização (3.68b) representa uma imersão do $\mathbb{R}^M$ no $\mathbb{R}^N$, veja [27]. Exemplos comuns de imersões são as superfícies (ou hipersuperfícies) e as curvas (ou hipercurvas) no $\mathbb{R}^N$. Como as modulações consideradas nessa seção estão associadas a espaços mais gerais do que o Euclidiano, faz-se necessário o uso de alguns conceitos da geometria Riemanniana. Tais conceitos foram descritos no capítulo de introdução e podem ser obtidos também em [27].

Por ser mais adequado, neste momento, serão analisados apenas os desempenhos de modulações associadas a superfícies regulares no $\mathbb{R}^3$, sendo que o principal objetivo é determinar quais são os parâmetros geométricos que mais influenciam em seus desempenhos. Exemplos de desempenhos de modulações associadas a hipersuperfícies ou hipercurvas serão consideradas nas próximas subseções, momento este em que o leitor esteja mais familiarizado com os novos conceitos geométricos expostos nesta primeira análise.

Para tanto, considere que o sistema de comunicações esteja sujeito a ação de um ruído AWGN de energia média $\mathcal{N} = N\mathcal{N}_0/2$, sendo $\sigma^2 = \mathcal{N}_0/2$ a energia média de ruído por dimensão. Além disso, considere que, quando um receptor de máxima verossimilhança é usado, então a estimativa $\hat{\mathbf{m}}$ é obtida como sendo o valor de $\mathbf{m}$ que minimiza $|\mathbf{r} - \mathbf{s_m}|$, isto é, o receptor estima pelo sinal $\mathbf{s}_{\hat{\mathbf{m}}}$ mais próximo de $\mathbf{r}$. Matematicamente, a estimativa $\hat{\mathbf{m}}$ deve satisfazer o seguinte sistema de equações

$$\hat{\mathbf{m}} = \min_{\mathbf{m}} |\mathbf{r} - \mathbf{s_m}| \quad \Rightarrow \quad \begin{cases} (\mathbf{r} - \mathbf{s}_{\hat{\mathbf{m}}}) \cdot \left.\dfrac{\partial\, \mathbf{s_m}}{\partial m_1}\right|_{\mathbf{m}=\hat{\mathbf{m}}} = 0 \\ (\mathbf{r} - \mathbf{s}_{\hat{\mathbf{m}}}) \cdot \left.\dfrac{\partial\, \mathbf{s_m}}{\partial m_2}\right|_{\mathbf{m}=\hat{\mathbf{m}}} = 0 \\ \quad\vdots \\ (\mathbf{r} - \mathbf{s}_{\hat{\mathbf{m}}}) \cdot \left.\dfrac{\partial\, \mathbf{s_m}}{\partial m_M}\right|_{\mathbf{m}=\hat{\mathbf{m}}} = 0 \end{cases} \quad \Rightarrow \quad (\mathbf{r} - \mathbf{s}_{\hat{\mathbf{m}}}) \cdot J_{\hat{\mathbf{m}}} = 0, \tag{3.69}$$

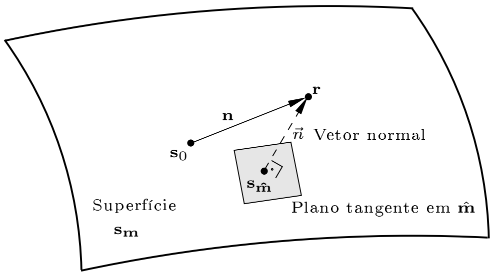
onde $J_{\hat{\mathbf{m}}}$ é a matriz Jacobiana de $\mathbf{s_m}$ no ponto $\mathbf{m} = \hat{\mathbf{m}}$, dada por

$$J = \begin{bmatrix} \dfrac{\partial\, \mathbf{s_m}}{\partial m_1} \\ \dfrac{\partial\, \mathbf{s_m}}{\partial m_2} \\ \vdots \\ \dfrac{\partial\, \mathbf{s_m}}{\partial m_M} \end{bmatrix}^T = \begin{bmatrix} \dfrac{\partial\, s_1}{\partial m_1} & \dfrac{\partial\, s_1}{\partial m_2} & \cdots & \dfrac{\partial\, s_1}{\partial m_M} \\ \dfrac{\partial\, s_2}{\partial m_1} & \dfrac{\partial\, s_2}{\partial m_2} & \cdots & \dfrac{\partial\, s_2}{\partial m_M} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial\, s_N}{\partial m_1} & \dfrac{\partial\, s_N}{\partial m_2} & \cdots & \dfrac{\partial\, s_N}{\partial m_M} \end{bmatrix}.$$

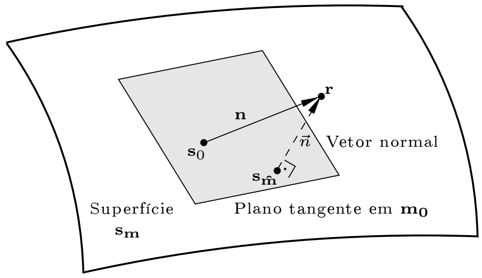
Neste caso, observe que o vetor $\mathbf{r} - \mathbf{s}_{\hat{\mathbf{m}}}$ é simultaneamente ortogonal aos vetores tangentes $\frac{\partial \mathbf{s_m}}{\partial m_i}$, $i = 1, \ldots, M$ em $\mathbf{m} = \hat{\mathbf{m}}$, o que implica que $\hat{\mathbf{m}}$ é obtido quando $\mathbf{r} - \mathbf{s}_{\hat{\mathbf{m}}}$ é normal a $\mathbf{s_m}$. Para uma melhor interpretação desse resultado veja a Figura 3.32(a), no caso particular em que $\mathbf{s_m}$ é uma superfície do $\mathbb{R}^3$. Dessa forma, o sinal recebido pode ser escrito da seguinte forma

$$\mathbf{r} = \mathbf{s}_{\hat{\mathbf{m}}} + h_1\vec{n}_1 + h_2\vec{n}_2 + \cdots + \vec{n}_{N-M}, \tag{3.70}$$

onde $\vec{n}_i$, $i = 1, \ldots, N - M$ são vetores ortonormais que geram uma base do subespaço normal de $\mathbf{s_m}$ e h_i, $i = 1, \ldots, N - M$ são constantes reais que posteriormente serão associadas a variáveis aleatórias, como realizado anteriormente na obtenção de $p_{\hat{m}|m_0}$ para modulações associadas a curvas.



(a) Projeção na superfície.



(b) Projeção no plano tangente aproximado.

Figura 3.32: Exemplo de recepção por máxima verossimilhança de um sistema de comunicações usando modulação não-linear associada a uma superfície regular.

Quando a energia média de ruído for suficientemente pequena, de maneira que o sinal recebido $\mathbf{r}$ esteja na vizinhança do sinal enviado $\mathbf{s_0}$, então o sinal $\mathbf{s_{\hat{m}}}$ pode ser localmente aproximado por

$$\mathbf{s_m} = \mathbf{s_0} + (\mathbf{m} - \mathbf{m_0}) \cdot J_{\mathbf{m_0}}^T, \tag{3.71}$$

veja Figura 3.32(b) para melhor entendimento. Neste caso, observe que $\mathbf{s_{\hat{m}}}$ é obtido por uma projeção de $\mathbf{r}$ no espaço tangente de $\mathbf{s_m}$ em $\mathbf{m_0}$. Tal fato implica que quando a energia média de ruído for suficientemente pequena, então o receptor pode usar a geometria intrínseca da superfície para determinar $\hat{\mathbf{m}}$, mais precisamente, pode-se considerar que o conjunto de sinais $\mathbf{m}$ possui métrica induzida de $\mathbf{s_m}$. Essa consideração pode ser melhor entendida com a análise de desempenho descrita a seguir.

Usando a aproximação (3.71) no sistema de equações (3.69), é possível determinar uma aproximação para o erro quadrático médio dos sistemas de comunicações que usam uma modulação não-linear $(M : N)$. Neste caso, como $J_{\hat{\mathbf{m}}} = J_{\mathbf{m_0}}$ devido à aproximação realizada, então a estimativa $\hat{\mathbf{m}}$ pode ser obtida da seguinte maneira

$$\begin{aligned} & \left[\mathbf{r} - \mathbf{s_0} - (\hat{\mathbf{m}} - \mathbf{m_0}) \cdot J_{\mathbf{m_0}}^T\right] \cdot J_{\mathbf{m_0}} = 0 \\ \Rightarrow \quad & (\mathbf{r} - \mathbf{s_0}) \cdot J_{\mathbf{m_0}} = (\hat{\mathbf{m}} - \mathbf{m_0}) \cdot G_{\mathbf{m_0}} && (3.72a) \\ \Rightarrow \quad & \hat{\mathbf{m}} = \mathbf{m_0} + (\mathbf{r} - \mathbf{s_0}) \cdot J_{\mathbf{m_0}} \cdot G_{\mathbf{m_0}}^{-1}, && (3.72b) \end{aligned}$$

onde $G_{\mathbf{m_0}} = J_{\mathbf{m_0}}^T \cdot J_{\mathbf{m_0}}$ é a matriz métrica ou o tensor métrica Riemanniana de $\mathbf{s_m}$ em $\mathbf{m_0}$, dado por

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1M} \\ g_{21} & g_{22} & \cdots & g_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ g_{M1} & g_{M2} & \cdots & g_{MM} \end{bmatrix},$$

onde $g_{ij} = g_{ji} = \frac{\partial \mathbf{s_m}}{\partial m_i} \cdot \frac{\partial \mathbf{s_m}}{\partial m_j}$, para $i, j = 1, \ldots, M$.

Com o auxílio da Figura 3.32, note que o ruído do canal de transmissão é dado por $\mathbf{n} = \mathbf{r} - \mathbf{s_0}$. Então, da equação (3.72b), podemos verificar que, para o receptor, o erro cometido em $\hat{\mathbf{m}} - \mathbf{m_0}$ é decorrente da projeção do ruído, $\mathbf{n} \cdot J_{\mathbf{m_0}}$, no espaço tangente de $\mathbf{s_m}$ em $\mathbf{m_0}$. Dessa forma, podemos assumir que, para o receptor, é como se o sinal recebido $\mathbf{r}$ estivesse sempre sobre a superfície $\mathbf{s_m}$. Pois, neste caso, a parcela da energia do ruído que age em $\mathbf{s_0}$ de maneira que $\mathbf{r}$ não pertença a $\mathbf{s_m}$ foi desprezada na projeção descrita anteriormente.

Como foi considerado que o ruído no canal de transmissão é AWGN, então $\mathbf{n}$ é um vetor de variáveis aleatórias Gaussianas independentes, de média nula e variância $\sigma^2 = \mathcal{N}_0/2$. Neste caso, com o auxílio de (3.72b), podemos obter uma expressão aproximada para o erro quadrático médio, dado que $\mathbf{m_0}$ tenha sido transmitido, como sendo

$$\begin{aligned}
\overline{\epsilon^2}(\mathbf{m_0}) &= \frac{1}{M} \left|\hat{\mathbf{m}} - \mathbf{m_0}\right|^2 = \frac{1}{M}(\mathbf{r} - \mathbf{s_0}) \cdot J_{\mathbf{m_0}} \cdot (G_{\mathbf{m_0}}^{-1})^2 \cdot J_{\mathbf{m_0}}^T \cdot (\mathbf{r} - \mathbf{s_0})^T && \text{(3.73a)} \\
&= \frac{\sigma^2}{M} \sum_{i=1}^{M} \sum_{j=1}^{M} g_{ij} \sum_{k=1}^{M} g_{ik}^{-1} g_{kj}^{-1} = \frac{\sigma^2}{M \det(G)} \sum_{i=1}^{M} G_{ii} && \text{(3.73b)} \\
&= \frac{\sigma^2}{2} \frac{g_{11} + g_{22}}{g_{11}g_{22} - (g_{12})^2}, \quad \text{para } M = 2 && \text{(3.73c)} \\
&= \frac{\sigma^2}{3} \frac{g_{11}g_{22} + g_{11}g_{33} + g_{22}g_{33} - (g_{12})^2 - (g_{13})^2 - (g_{23})^2}{g_{11}g_{22}g_{33} - g_{11}(g_{23})^2 - g_{22}(g_{13})^2 - g_{33}(g_{12})^2 + 2g_{12}g_{13}g_{23}}, \quad \text{para } M = 3
\end{aligned}$$

onde G_{ii}, $i = 1, \ldots, M$ são os cofatores dos elementos da diagonal de G e os elementos g_{ij} são calculados para $\mathbf{m} = \mathbf{m_0}$.

Neste caso, é importante informar que a expressão (3.73a) havia sido obtida em [28]. Contudo, tal expressão não informa claramente alguns pontos relevantes que são facilmente percebidos na expressão (3.73b), aqui desenvolvida. Por exemplo, o $\overline{\epsilon^2}$ é inversamente proporcional ao quadrado do elemento de volume da variedade Riemanniana associada a $\mathbf{s_m}$ e diretamente proporcional à soma dos quadrados dos elementos de volume das subvariedades Riemannianas de dimensão $M-1$. Além disso, tal expressão é mais adequada para a realização de técnicas de otimização visando a identificação de uma classe de hipersuperfícies que melhor se adequa ao processo de modulação.

Como mostrado anteriormente, o $\overline{\epsilon^2}(m_0)$ de modulações associadas a curvas é inversamente proporcional ao quadrado de seu fator de expansão, (3.7), que, dependendo da parametrização da curva, pode ser uniforme (não depende de m_0) para melhorar o desempenho da modulação. Contudo, quando consideram-se modulações associadas a superfícies, o $\overline{\epsilon^2}(\mathbf{m_0})$ calculado por (3.73b) é função da métrica g_{ij}, que, geralmente, depende do valor de $\mathbf{m_0}$. Isto implica que, na maioria dos casos, o erro quadrático médio seja calculado por

$$\overline{\epsilon^2} = \mathrm{E}\left[\overline{\epsilon^2}(\mathbf{m_0})\right] = \int_{-1}^{1} \cdots \int_{-1}^{1} \frac{\sigma^2}{M\ \det(G)} \sum_{i=1}^{M} G_{ii}\, p_{\mathbf{m}}\, dm_1 \cdots dm_M, \tag{3.74}$$

onde $p_{\mathbf{m}}$ é a função densidade de probabilidade conjunta *a priori* do vetor $\mathbf{m}$.

Dessa forma, assim como ocorreu no caso de modulações associadas a curvas, note que a parametrização da superfície associada à modulação também pode influenciar no desempenho dos sistemas de comunicações. Por isso, esperamos que, quanto menos complexa seja a métrica, mais fácil também seja a análise de desempenho do sistema. Dessa maneira, a parametrização isotérmica para superfícies aparenta ser muito interessante aos propósitos deste capítulo. Nesta parametrização, a métrica é caracterizada por $g_{ii} = g$ e $g_{ij} = 0$, se $i \neq j$. Com isso, o erro quadrático médio (3.73b) pode ser simplificado para

$$\overline{\epsilon^2}(\mathbf{m_0}) = \frac{\sigma^2}{g} = \frac{\mathcal{N}_0/2}{g}, \tag{3.75}$$

que é muito semelhante à expressão (3.7) do $\overline{\epsilon^2}$ para o caso de modulações associadas a curvas. Tal resultado, de certa forma, é bem conveniente, pois tanto o fator de expansão (S) quanto a raiz quadrada da métrica ($\sqrt{g}$) estão relacionados ao elemento de comprimento de $\mathbf{s_m}$, fato que, posteriormente, será associado ao comprimento geodésico e à curvatura seccional de variedades Riemannianas associadas a esquemas de modulação.

Como exemplo de aplicação da expressão (3.75), considere um esquema de modulação baseada em uma variedade Riemanniana com curvatura seccional constante, cuja métrica seja definida como

$$g = \frac{1}{\left(1 + \frac{K}{4}\sum_{i=1}^{M} m_i^2\right)^2},$$

onde K é o valor da curvatura seccional da variedade Riemanniana. Se $K < 0$, a métrica g_{ij} é definida em uma bola de raio $\sqrt{-4/K}$. Portanto, de (3.73b), obtém-se que o erro quadrático médio é dado por

$$\overline{\epsilon^2}(\mathbf{m_0}) = \mathcal{N}_0/2\left(1 + \frac{K}{4}\sum_{i=1}^{M} m_i^2\right)^2, \tag{3.76}$$

o qual decresce com a diminuição do valor de K, veja Figura 3.33. Tal observação é muito importante, pois afirma que, dentre as modulações associadas a espaços de curvatura constantes, aquelas associadas a espaços de curvatura negativa produzem um erro quadrático médio menor. Um resultado semelhante a esse também foi obtido em [29], quando foi analisado o desempenhos de constelações de sinais em variedades Riemannianas.

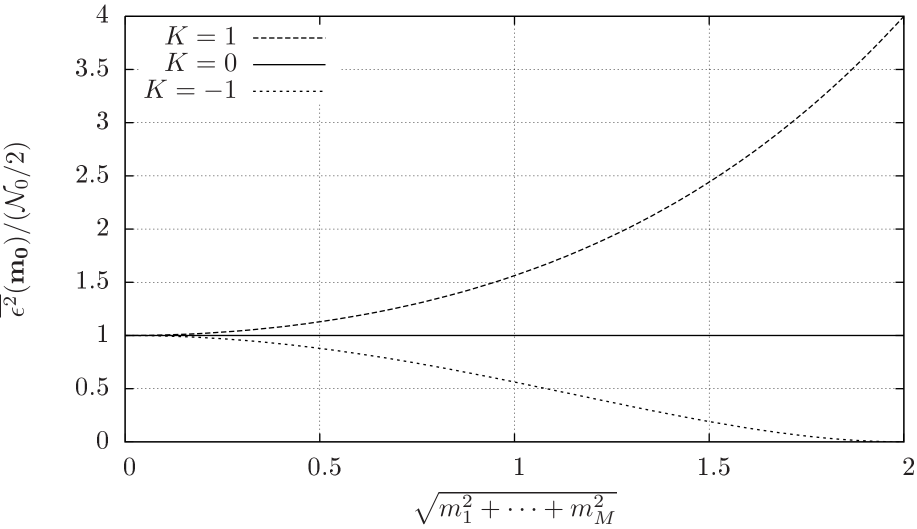


Figura 3.33: Erro quadrático médio de modulações associadas a espaços de curvatura seccional constante, $K = -1, 0$ e 1.

3.5.1 Uma expressão para $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$

Neste momento, será apresentada uma expressão para a probabilidade condicional de $\hat{\mathbf{m}}$ dado que $\mathbf{m_0}$ foi enviado. Para tanto, como considerado anteriormente, a energia média de ruído do canal de transmissão deve ser suficientemente pequena de maneira que a aproximação (3.71) seja válida. Neste caso, quando um receptor de máxima verossimilhança é usado, então a expressão (3.72a) é utilizada para a determinação da expressão de $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ na vizinhança de $\mathbf{s_0}$. Além disso, um fato muito importante que deve ser esclarecido é que a expressão obtida de $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ é aproximada, sendo válida apenas para valores de $\hat{\mathbf{m}}$ próximos de $\mathbf{m_0}$, e mais precisa quanto menor for a energia média de ruído.

Dessa forma, como ilustra a Figura 3.32(b), o processo de recepção é realizado no plano tangente de $\mathbf{s_m}$ em $\mathbf{s_0}$. Por essa razão, para obter $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$, foi necessário realizar uma transformação de variáveis aleatórias com o objetivo de determinar a função densidade de probabilidade do ruído $\mathbf{n}$ projetado no plano tangente de $\mathbf{s_m}$ em $\mathbf{s_0}$. Para tanto, foi considerado que, quando o ruído do canal de transmissão é AWGN, então o mesmo pode ser decomposto como $\mathbf{n} = \mathbf{n}^\top + \mathbf{n}^\perp$, onde $\mathbf{n}^\top$ e $\mathbf{n}^\perp$ são projeções do ruído em duas bases de vetores ortonormais que geram o espaço tangente e o espaço normal de $\mathbf{s_m}$ em $\mathbf{s_0}$, respectivamente.

Neste caso, note que tal decomposição do ruído foi escolhida por facilitar a caracterização da função densidade de probabilidade do vetor $\mathbf{n}^\top$, que no caso é Gaussiana circular de média nula. Com isso, a transformação de variáveis aleatórias descrita anteriormente pode ser obtida considerando o lado esquerdo da igualdade (3.72a)

$$(\mathbf{r} - \mathbf{s_0}) \cdot J_{\mathbf{m_0}} = \mathbf{n} \cdot J_{\mathbf{m_0}} = (\mathbf{n}^\top + \mathbf{n}^\perp) \cdot J_{\mathbf{m_0}} =$$

$$\mathbf{n}^\top \cdot J_{\mathbf{m_0}} = \overline{\mathbf{n}}\,, \tag{3.77}$$

onde $\overline{\mathbf{n}}$ é a projeção de $\mathbf{n}^\top$ na base de vetores tangentes de $\mathbf{s_m}$ em $\mathbf{m} = \mathbf{m_0}$ e considerando o processo de ortogonalização de Gram-Schmidt que usa a base de vetores linearmente independente $\{\mathbf{s}_\mathbf{m}^{(1)} = \frac{\partial \mathbf{s_m}}{\partial m_1}, \ldots, \mathbf{s}_\mathbf{m}^{(M)} = \frac{\partial \mathbf{s_m}}{\partial m_M}\}$ para obter uma base de vetores ortonormais $\{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_M\} \subset \mathbb{R}^N$, na qual $\mathbf{n}^\top = n_1^\top \vec{b}_1 + \cdots + n_M^\top \vec{b}_M$. No caso, o processo de ortogonalização de Gram-Schmidt é dado por

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{s}_\mathbf{m}^{(1)}\,, & \vec{b}_1 &= \frac{\mathbf{u}_1}{|\mathbf{u}_1|} \\
\mathbf{u}_2 &= \mathbf{s}_\mathbf{m}^{(2)} - \frac{\mathbf{s}_\mathbf{m}^{(2)} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1\,, & \vec{b}_2 &= \frac{\mathbf{u}_2}{|\mathbf{u}_2|} \\
&\vdots & &\vdots \\
\mathbf{u}_M &= \mathbf{s}_\mathbf{m}^{(M)} - \sum_{j=1}^{M-1} \frac{\mathbf{s}_\mathbf{m}^{(M)} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j}\mathbf{u}_j\,, & \vec{b}_M &= \frac{\mathbf{u}_M}{|\mathbf{u}_M|}\,.
\end{aligned}$$

Portanto, para o caso particular em que $M = 2$, a transformação de variáveis aleatórias, $\mathbf{n}^\top$ em função de $\overline{\mathbf{n}}$, é calculada pela seguinte mudança de base, obtida do processo de ortogonalização de Gram-Schmidt,

$$\vec{b}_1 = \frac{\mathbf{s}_\mathbf{m}^{(1)}}{\sqrt{g_{11}}} \quad \text{e} \quad \vec{b}_2 = \frac{\sqrt{g_{11}}}{\sqrt{g_{11}g_{22} - (g_{12}^2)}}\left(\mathbf{s}_\mathbf{m}^{(2)} - \frac{g_{12}}{g_{11}}\mathbf{s}_\mathbf{m}^{(1)}\right),$$

na qual o vetor $\overline{\mathbf{n}}$ é determinado com o auxílio de (3.77) e $\mathbf{n}^\top = n_1^\top \vec{b}_1 + n_2^\top \vec{b}_2$ da seguinte forma

$$\overline{n}_1 = \mathbf{n}^\top \cdot \mathbf{s}_\mathbf{m}^{(1)} = n_1^\top \sqrt{g_{11}} \quad \text{e} \quad \overline{n}_2 = \mathbf{n}^\top \cdot \mathbf{s}_\mathbf{m}^{(2)} = \frac{n_1^\top g_{12}}{\sqrt{g_{11}}} + \frac{n_2^\top}{\sqrt{g_{11}}}\sqrt{g_{11}g_{22} - (g_{12})^2}\,,$$

e, finalmente, a transformação de variáveis é dada por

$$n_1^\top = \frac{\overline{n}_1}{\sqrt{g_{11}}}\,, \quad n_2^\top = \frac{\sqrt{g_{11}}}{\sqrt{g_{11}g_{22} - (g_{12})^2}}\left(\overline{n}_2 - \frac{g_{12}}{g_{11}}\overline{n}_1\right) \quad \text{e}$$

$$\left|\frac{\partial(n_1^\top, n_2^\top)}{\partial(\overline{n}_1, \overline{n}_2)}\right| = \frac{1}{\sqrt{g_{11}g_{22} - (g_{12})^2}}. \tag{3.78}$$

Dessa forma, a função densidade de probabilidade de $\overline{\mathbf{n}}$ é obtida com o auxílio de (3.5.1) da seguinte forma

$$\begin{aligned}
p_{\overline{\mathbf{n}}} &= p_{\mathbf{n}^\top}(\overline{\mathbf{n}})\left|\frac{\partial(n_1^\top, n_2^\top)}{\partial(\overline{n}_1, \overline{n}_2)}\right| = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(n_1^\top)^2 + (n_2^\top)^2}{2\sigma^2}\right)\left|\frac{\partial(n_1^\top, n_2^\top)}{\partial(\overline{n}_1, \overline{n}_2)}\right| \\
&= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\frac{(\overline{n}_1)^2 g_{22} - 2\overline{n}_1\overline{n}_2 g_{12} + (\overline{n}_2)^2 g_{11}}{g_{11}g_{22} - (g_{12})^2}\right)\frac{1}{\sqrt{g_{11}g_{22} - (g_{12})^2}} \\
&= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\overline{\mathbf{n}} \cdot G^{-1} \cdot \overline{\mathbf{n}}^T}{2\sigma^2}\right)\frac{1}{\sqrt{\det(G)}}
\end{aligned} \tag{3.79}$$

Com isso, a função densidade de probabilidade de $\hat{\mathbf{m}}$ dado que $\mathbf{m_0}$ tenha sido enviada pode ser obtida através de uma nova transformação de variáveis aleatórias, determinada com o auxílio das equações (3.72a) e (3.77), da seguinte forma

$$\overline{\mathbf{n}} = (\hat{\mathbf{m}} - \mathbf{m_0}) \cdot G_{\mathbf{m_0}} \quad \text{e} \quad \left| \frac{\partial(\overline{\mathbf{n}})}{\partial(\hat{\mathbf{m}})} \right| = \det(G_{\mathbf{m_0}}). \tag{3.80}$$

Assim, usando essa última transformação de variáveis em (3.79) a função $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ é dada por

$$p_{\hat{\mathbf{m}}|\mathbf{m_0}} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(\hat{\mathbf{m}} - \mathbf{m_0}) \cdot G \cdot (\hat{\mathbf{m}} - \mathbf{m_0})^T}{2\sigma^2} \right) \sqrt{\det(G)}\,, \tag{3.81}$$

onde a matriz métrica G é calculada em $\mathbf{m_0}$. Por essa razão, a expressão de $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ obtida em (3.81) é válida na vizinhança de $\mathbf{m_0}$.

Apesar da expressão (3.81) ter sido obtida para o caso particular de $M = 2$, pode-se verificar de maneira similar que tal função densidade de probabilidade continua valendo para outros valores de M, apenas pela substituição da constante $2\pi\sigma^2$ por $(2\pi\sigma^2)^{(M/2)}$. Com isso, para o caso particular em que a métrica é isotérmica ($g_{ij} = g\,\delta_{ij}$), a expressão do $\overline{\epsilon^2}$ (3.75) pode facilmente ser reobtida com o auxílio de (3.81), pois neste caso $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ é Gaussiana de média $\mathbf{m_0}$ e variância $M\,\sigma^2/g$, logo $\overline{\epsilon^2}(\mathbf{m_0}) = \sigma^2/g$ como esperado. Além disso, a expressão (3.73b) também pode ser obtida de (3.81), desde que se use uma mudança de coordenadas de maneira que $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ seja interpretada como uma Gaussiana, facilitando, assim, a obtenção de sua variância.

Em [29], uma expressão de $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$, similar à obtida em (3.81), foi utilizada para medir o desempenho de constelações de sinais em variedades Riemannianas. Sendo que o termo $(\hat{\mathbf{m}} - \mathbf{m_0}) \cdot G \cdot (\hat{\mathbf{m}} - \mathbf{m_0})^T$ foi substituído pela distância geodésica ao quadrado entre os pontos $\hat{\mathbf{m}}$ e $\mathbf{m_0}$ e o valor de $\det(G)$ era função de $\hat{\mathbf{m}}$. Por essa razão, duas novas constantes foram inseridas na densidade $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ de [29] para que a sua integral fosse igual a 1.

Neste caso, quando a energia média de ruído for suficientemente pequena, pode-se verificar que as substituições realizadas por [29] em (3.81) são possíveis e adequadas, justificando assim a sua utilização nas análises de desempenhos de constelações de sinais em variedades Riemannianas, como realizado nos trabalhos [29], [30], [31], [32] e [33].

Além da expressão (3.81), obtida em função da métrica g_{ij} e de $(\hat{\mathbf{m}} - \mathbf{m_0})$, uma outra fórmula mais precisa para $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ pode ser obtida em função da métrica g_{ij}, de $(\mathbf{s}_{\hat{\mathbf{m}}} - \mathbf{s_0})$ e dos vetores $\mathbf{s}_{\mathbf{m}}^{(ij)} = \frac{\partial^2 \mathbf{s_m}}{\partial m_i \partial m_j}$ em $\mathbf{m} = \hat{\mathbf{m}}$. Para determinar tal expressão, considere $\mathbf{r} = \mathbf{s_0} + \mathbf{n}$ em (3.69) para obter a seguinte relação

$$(\mathbf{s_0} + \mathbf{n} - \mathbf{s}_{\hat{\mathbf{m}}}) \cdot J_{\hat{\mathbf{m}}} = 0 \quad \Rightarrow \quad \overline{\mathbf{n}} = \mathbf{n} \cdot J_{\hat{\mathbf{m}}} = (\mathbf{s}_{\hat{\mathbf{m}}} - \mathbf{s_0}) \cdot J_{\hat{\mathbf{m}}}. \tag{3.82}$$

Neste caso, a função densidade de probabilidade para o ruído $\overline{\mathbf{n}}$, representado pelo lado esquerdo da equação (3.82), foi desenvolvida em (3.79), com a diferença que a matriz G agora é calculada em $\mathbf{m} = \hat{\mathbf{m}}$ em vez de $\mathbf{m_0}$, como realizado anteriormente. Portanto, $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ pode ser determinada usando da transformação de variáveis aleatórias de $\overline{\mathbf{n}}$ em função de $\mathbf{s}_{\hat{\mathbf{m}}}$ e $\mathbf{s_0}$, obtida com o auxílio de (3.82) e caracterizada pelo seguinte Jacobiano

$$|J^+| = \det \begin{bmatrix} g_{11} + (\mathbf{s}_{\hat{\mathbf{m}}} - \mathbf{s_0}) \cdot \mathbf{s}_{\mathbf{m}}^{(11)} & \cdots & g_{1M} + (\mathbf{s}_{\hat{\mathbf{m}}} - \mathbf{s_0}) \cdot \mathbf{s}_{\mathbf{m}}^{(1M)} \\ \vdots & \ddots & \vdots \\ g_{M1} + (\mathbf{s}_{\hat{\mathbf{m}}} - \mathbf{s_0}) \cdot \mathbf{s}_{\mathbf{m}}^{(M1)} & \cdots & g_{MM} + (\mathbf{s}_{\hat{\mathbf{m}}} - \mathbf{s_0}) \cdot \mathbf{s}_{\mathbf{m}}^{(MM)} \end{bmatrix}. \tag{3.83}$$

Com isso, de (3.79) e (3.83), $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ é dada por

$$p_{\hat{\mathbf{m}}|\mathbf{m_0}} = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(- \frac{(\mathbf{s}_{\hat{\mathbf{m}}} - \mathbf{s_0}) \cdot J_{\hat{\mathbf{m}}} \cdot G^{-1} \cdot J_{\hat{\mathbf{m}}}^T \cdot (\mathbf{s}_{\hat{\mathbf{m}}} - \mathbf{s_0})^T}{2\sigma^2} \right) \frac{|J^+|}{\sqrt{\det(G)}}, \tag{3.84}$$

onde a matriz métrica G é calculada em $\hat{\mathbf{m}}$.

Com o objetivo de ilustrar uma aplicação para a expressão (3.84), considere o caso em que $M = 1$, mais precisamente, a modulação é associada a uma curva do espaço Euclidiano n-dimensional com parametrização $\mathbf{s}_m$. Neste caso particular, $J = s'_m$ e $g_{11} = |s'_m|^2$, então $p_{\hat{m}|m_0}$ obtida de (3.84) é igual a

$$p_{\hat{m}|m_0} = \frac{1}{\sqrt{2\pi\sigma^2 |\mathbf{s}'_m|^2}} \exp\left(- \frac{|(\mathbf{s}_{\hat{m}} - \mathbf{s}_0) \cdot \mathbf{s}'_m|^2}{2\sigma^2 \, |\mathbf{s}'_m|^2} \right) \left(|\mathbf{s}'_m|^2 + (\mathbf{s}_{\hat{m}} - \mathbf{s}_0) \cdot \mathbf{s}''_m \right),$$

que é similar à expressão (3.51), obtida anteriormente usando um procedimento diferente do aplicado aqui. Pois, em (3.51), tem-se que $S = |\mathbf{s}'_m|$, $\vec{n} = -\mathbf{s}''_m / |\mathbf{s}''_m|$ e $k = |\mathbf{s}''_m|$, o que comprova a afirmação anterior.

Dessa forma, a expressão (3.84) certamente caracteriza bem a função $p_{\hat{\mathbf{m}}|\mathbf{m_0}}$ de modulações não-lineares associadas a curvas ou a superfícies. A única desvantagem da aplicação de (3.84) é que em sua obtenção, a partir de (3.82), foram considerados todos os pontos críticos da função $|\mathbf{r} - \mathbf{s_m}|$, isto é, como se o receptor estimasse $\hat{m}$ como sendo todos os pontos críticos de mínimo ou de máximo local do problema de otimização em questão, e não apenas o valor $\mathbf{m} = \hat{\mathbf{m}}$ que caracteriza a menor distância entre $\mathbf{r}$ e $\mathbf{s_m}$, como normalmente é realizado no processo de demodulação usando um receptor ML.

3.5.2 Primeira variação do $\overline{\epsilon^2}$

Nesta subseção, utilizaremos um método de otimização no qual foi possível identificar uma classe de superfícies que, quando associadas a esquemas de modulação, podem minimizar o erro quadrático médio. Contudo, nesta análise, foram consideradas apenas modulações associadas a superfícies ($M = 2$), pois o tratamento de um caso mais geral é muito mais complexo. Porém, espera-se que um resultado semelhante ao encontrado aqui possa ser generalizado para o caso de modulações associadas a hipersuperfícies.

Dessa forma, considere que o esquema de modulação em questão esteja associado a uma superfície regular parametrizada, $X : D \subset \mathbb{R}^2 \to \mathbb{R}^3$, na qual os parâmetros a serem transmitidos são (u, v) e pertencem ao domínio D, em vez de $(m_1, m_2) \in (-1, 1) \times (-1, 1)$ como realizado anteriormente. Essa mudança de notação foi necessária para que alguns resultados da geometria diferencial apresentados em [25] pudessem ser aqui aplicados com maior familiaridade.

Com o objetivo de especificar a técnica de otimização utilizada, considere uma função diferenciável $\eta : D \subset \mathbb{R}^2 \to \mathbb{R}$ e uma variação normal da superfície $X(u, v)$ determinada por $\eta(u, v)$, caracterizada pela aplicação $X^t : D \times (-\epsilon, \epsilon) \to \mathbb{R}^3$, definida como

$$X^t(u, v) = X(u, v) + t\eta(u, v)N(u, v), \quad (u, v) \in D,\ t \in (-\epsilon, \epsilon), \tag{3.85}$$

onde $N(u, v)$ é o vetor normal unitário da superfície $X(u, v)$. Observe que tal variação de $X(u, v)$ representa, para cada valor de t, uma nova superfície gerada a partir de uma "variação" contínua e normal da superfície original. Portanto, como $X^t(u, v)$ descreve uma infinidade de possíveis superfícies parametrizadas, então essa variação pode ser considerada para se determinar quais características $X^0(u, v) = X(u, v)$ deve possuir para que o erro quadrático médio de uma modulação associada a essa parametrização possa ser minimizado.

Dessa forma, é necessário determinar a primeira forma fundamental de $X^t(u, v)$ para calcular o $\overline{\epsilon^2}$ usando (3.73c), isto é,

$$\begin{aligned}
E^t &= \left(X_u^t \cdot X_u^t\right) = E + 2t\eta\left(X_u \cdot N_u\right) + t^2\eta^2\left(N_u \cdot N_u\right) + t^2\eta_u^2\,, \\
F^t &= \left(X_u^t \cdot X_v^t\right) = F + t\eta\left[\left(X_u \cdot N_v\right) + \left(X_v \cdot N_u\right)\right] + t^2\eta^2\left(N_u \cdot N_v\right) + t^2\eta_u\eta_v\,, \\
G^t &= \left(X_v^t \cdot X_v^t\right) = G + 2t\eta\left(X_v \cdot N_v\right) + t^2\eta^2\left(N_v \cdot N_v\right) + t^2\eta_v^2\,.
\end{aligned} \tag{3.86}$$

Com isso, o erro quadrático médio de uma modulação associada a $X^t(u, v)$ pode ser obtido com o auxílio de (3.74) e (3.86), como

$$\overline{\epsilon^2} = \frac{\mathcal{N}_0}{4}\frac{E^t + G^t}{E^tG^t - (F^t)^2}. \tag{3.87}$$

Assim, para caracterizar o ponto crítico de (3.87), deve-se derivar $\overline{\epsilon^2}$ em relação a t e igualá-la a zero em $t = 0$. Contudo, como essa integral é realizada em função das variáveis u e v e a derivada em questão é em relação a t, então é possível e conveniente considerar a derivada em relação a t apenas do integrando de (3.87), isto é,

$$\frac{1}{E^tG^t - (F^t)^2}\left(\frac{\partial E^t}{\partial t} + \frac{\partial G^t}{\partial t}\right) - \frac{E^t + G^t}{[E^tG^t - (F^t)^2]^2}\left(G^t\frac{\partial E^t}{\partial t} + E^t\frac{\partial G^t}{\partial t} - 2F^t\frac{\partial F^t}{\partial t}\right)\Bigg|_{t=0} = 0\,. \tag{3.88}$$

Com o auxílio de (3.86), pode-se verificar que, em $t = 0$, a derivada em relação a t dos coeficientes da primeira forma fundamental de $X^t(u, v)$, são iguais a

$$\begin{aligned} \left.\frac{\partial E^t}{\partial t}\right|_{t=0} &= 2\eta\,(X_u \cdot N_u) = -2\eta e\,, \\ \left.\frac{\partial F^t}{\partial t}\right|_{t=0} &= \eta\,[(X_u \cdot N_v) + (X_v \cdot N_u)] = -2\eta f\,, \\ \left.\frac{\partial G^t}{\partial t}\right|_{t=0} &= 2\eta\,(X_v \cdot N_v) = -2\eta g\,, \end{aligned} \tag{3.89}$$

onde e, f e g são os coeficientes da segunda forma fundamental de $X(u, v)$, como descrito anteriormente. Então, usando (3.89) a equação (3.88) pode ser reescrita como

$$-2\eta\frac{e+g}{EG-F^2} + 2\eta\frac{(E+G)(eG-2fF+gE)}{(EG-F^2)^2} = 0,$$

que é satisfeita quando $X(u, v)$ for uma parametrização isotérmica ($E = G$ e $F = 0$) e sua curvatura média H, (1.4), for igual a zero, isto é,

$$H = \frac{1}{2}\frac{eG-2fF+gE}{EG-F^2} = \frac{e+g}{2E} = 0 \quad \text{e} \quad e = -g\,. \tag{3.90}$$

Portanto, um esquema de modulação associado a uma superfície mínima ($H = 0$) parametrizada isotermicamente é um ponto crítico para o erro quadrático médio. As superfícies mínimas formam uma classe de superfícies muito importante e estudada na matemática. Tais superfícies são soluções do problema de Plateau, isto é, dada uma curva fechada simples (sem auto-interseção), a superfície de área mínima que tem essa curva como bordo é uma superfície mínima. Um breve estudo sobre as superfícies mínimas pode ser encontrado em [29], enquanto que uma leitura mais avançada pode ser obtida em [17].

Neste caso, uma técnica que pode ser utilizada para se obter superfícies mínimas parametrizadas isotermicamente é a representação de Weierstrass. Para obter mais informações sobre essa representação veja a subseção 1.2.1 e as referências [17] e [29].

Além disso, as superfícies mínimas haviam sido consideradas anteriormente em [29], [31] e [14] quando constelações de sinais foram projetadas e analisadas sobre tais superfícies. Além disso, em [29], foi mostrado que, assim como ocorreu aqui para o $\overline{\epsilon^2}$, as superfícies mínimas também são pontos críticos da probabilidade média de erro em sistemas de comunicações digitais.

Contudo, um fato muito importante deve ser esclarecido neste momento. Para tanto, é necessário recordar que o elemento de área de uma superfície é dado por $\sqrt{EG - F^2}$ e que o erro quadrático médio dado em (3.73c) é inversamente proporcional ao quadrado desse elemento de área. Assim, se as superfícies mínimas são pontos críticos de mínimo desse elemento de área, então as modulações associadas a superfícies mínimas deveriam maximizar o $\overline{\epsilon^2}$. Porém, na próxima seção, será mostrado que tal fato na maioria das vezes não é verdade, pois, quando a variação $X^t(u,v)$ está sujeita a determinadas restrições como tipo de variação, energia média de transmissão mínima ou constante, então a superfície mínima é um ponto crítico de mínimo para o erro quadrático médio, ao contrário da conclusão anteriormente sugerida.

3.5.3 Segunda variação do $\overline{\epsilon^2}$

Analisaremos, em $t = 0$, a segunda derivada em relação a t do erro quadrático médio de uma modulação associada a variação $X^t(u,v)$ definida em (3.85), isto é,

$$\begin{aligned}\frac{\partial^2\overline{\epsilon^2}(\mathbf{m_0})}{\partial t^2} &= \frac{\mathcal{N}_0}{4}\left\{\frac{\frac{\partial^2 E^t}{\partial t^2}+\frac{\partial^2 G^t}{\partial t^2}}{E^tG^t-(F^t)^2} - 2\frac{\left(\frac{\partial E^t}{\partial t}+\frac{\partial G^t}{\partial t}\right)\left(G^t\frac{\partial E^t}{\partial t}+E^t\frac{\partial G^t}{\partial t}-2F^t\frac{\partial F^t}{\partial t}\right)}{\left[E^tG^t-(F^t)^2\right]^2}\right.\\ &\quad -\frac{(E^t+G^t)\left[G^t\frac{\partial^2 E^t}{\partial t^2}+2\frac{\partial E^t}{\partial t}\frac{\partial G^t}{\partial t}+E^t\frac{\partial^2 G^t}{\partial t^2}-2\left(\frac{\partial^2 F^t}{\partial t^2}\right)^2-2F^t\frac{\partial^2 F^t}{\partial t^2}\right]}{\left[E^tG^t-(F^t)^2\right]^2}\\ &\quad \left.+\frac{(E^t+G^t)\left(G^t\frac{\partial E^t}{\partial t}+E^t\frac{\partial G^t}{\partial t}-2F^t\frac{\partial F^t}{\partial t}\right)^2}{\left[E^tG^t-(F^t)^2\right]^3}\right\}. \end{aligned} \tag{3.91}$$

Com o auxílio de (3.86), pode-se verificar que, em $t = 0$, a segunda derivada em relação a t dos coeficientes da primeira forma fundamental de $X^t(u,v)$ são iguais a

$$\begin{aligned}\left.\frac{\partial^2 E^t}{\partial t^2}\right|_{t=0} &= 2\eta_u^2,\\ \left.\frac{\partial^2 F^t}{\partial t^2}\right|_{t=0} &= 2\eta_u\eta_v + 2\eta^2\left(N_u\cdot N_v\right),\\ \left.\frac{\partial^2 G^t}{\partial t^2}\right|_{t=0} &= 2\eta_v^2. \end{aligned} \tag{3.92}$$

Dessa forma, usando as derivadas da primeira forma fundamental descritas em (3.89) e em (3.92) e utilizando as condições (3.90) necessárias para $X(u,v)$ ser um ponto crítico de $\overline{\epsilon^2}$, então a equação (3.91) pode ser reescrita, em $t = 0$, como

$$\left.\frac{\partial^2\overline{\epsilon^2}(\mathbf{m_0})}{\partial t^2}\right|_{t=0} = -2\frac{\eta_u^2+\eta_v^2}{E^2} - 16\eta^2\frac{eg-f^2}{E^3} = -\frac{2}{E}\left(\frac{\eta_u^2+\eta_v^2}{E}+8\eta^2 K\right), \tag{3.93}$$

onde K é a curvatura Gaussiana da superfície $X(u,v)$, dada em (1.5).

Com isso, note que, se a superfície $X(u,v)$ possuir curvatura Gaussiana igual a zero, caso da modulação linear associada a uma superfície planar, então pode-se verificar de (3.93) que a segunda derivada em relação a t de $\overline{\epsilon^2}$ é negativa. Isto é, a modulação linear é um ponto crítico de máximo do erro quadrático médio. Além disso, uma vez que a curvatura Gaussiana seja negativa, então é possível que essa derivada seja positiva, caracterizando assim um ponto crítico de mínimo para $\overline{\epsilon^2}$. Neste caso, as superfícies mínimas se adequam perfeitamente, pois suas curvaturas Gaussianas são sempre negativas, exceto quando a superfície é o plano.

Portanto, a condição para que (3.93) seja positiva, além de depender de K, depende também da função $\eta(u,v)$. Neste caso, é necessário então caracterizar ou restringir as variações normais $X^t(u,v)$, de maneira que uma superfície mínima com parâmetros isotérmicos seja um ponto crítico de mínimo do erro quadrático médio. Um caso interessante para a variação $X^t(u,v)$ ocorre quando $\eta(u,v)$ é constante, pois $\eta_u = \eta_v = 0$ e como $K < 0$, então a segunda derivada em relação a t de $\overline{\epsilon^2}$ é positiva. Além disso, quando $\eta(u,v)$ for suave (η_u/η e $\eta_v/\eta \ll 1$), propiciando assim pequenas variações em $X^t(u,v)$, então a superfície mínima $X(u,v)$ parametrizada isotermicamente minimiza o erro quadrático médio. Em outras palavras, sob pequenas perturbações, $X(u,v)$ é estável quanto a sua minimização do $\overline{\epsilon^2}$.

Uma outra maneira adequada de caracterizar ou restringir a função $\eta(u,v)$, é supor que a energia média de transmissão permaneça praticamente constante durante a variação $X^t(u,v)$, ou que a mesma seja minimizada para algum valor de t. Para tanto, quando o valor médio de $X(u,v)$ for igual a origem do sistema de coordenadas, considere a seguinte fórmula para a energia média de transmissão

$$\overline{E}_m = \iint_D \left(X^t \cdot X^t\right) p_{uv}\, du\, dv = \iint_D \left[(X \cdot X) + 2t\,\eta\,(X \cdot N) + t^2\eta^2\right] p_{uv}\, du\, dv\,. \quad (3.94)$$

Com isso, o ponto crítico da energia média de transmissão (3.94) pode ser obtido por

$$\frac{\partial \overline{E}_m}{\partial t} = 0 \quad \Rightarrow \quad \iint_D 2\eta\,[(X \cdot N) + t\,\eta]\, p_{uv}\, du\, dv = 0 \quad \Rightarrow \quad t\,\eta = -\,(X \cdot N)\;. \quad (3.95)$$

Pode-se verificar facilmente que tal ponto crítico é de mínimo, pois

$$\frac{\partial^2 \overline{E}_m}{\partial t^2} = \iint_D 2\eta^2 p_{uv}\, du\, dv > 0\,.$$

Além disso, pode-se verificar que quando $t\,\eta = -2\,(X,N)$, a energia média de transmissão (3.94) é igual à inicial. Dessa forma, é adequado escolher uma variação $X^t(u,v)$, $t = (-\epsilon, \epsilon)$, com função $\eta(u,v) = c\,(X,N)$, onde c pode ser, por exemplo, igual a $-2/\epsilon$, para que essa variação possa contemplar superfícies com energia média de transmissão maiores ($-\epsilon < t < 0$), igual ($t = \epsilon$) ou menores ($0 < t < \epsilon$) do que a inicial ($t = 0$).

Portanto, os valores de η_u e η_v, necessários para analisar a equação (3.93), podem ser obtidos da consideração anterior $\eta(u,v) = c\,(X,N)$ e de algumas relações descritas em [25] (pág. 154), da seguinte forma

$$
\begin{aligned}
\eta_u &= c\,[(X_u \cdot N) + (X \cdot N_u)] = c\,(X \cdot N_u) = -\frac{c}{E}\,[e\,(X \cdot X_u) + f\,(X \cdot X_v)]\ , \\
\eta_v &= c\,[(X_v \cdot N) + (X \cdot N_v)] = c\,(X \cdot N_v) = -\frac{c}{E}\,[f\,(X \cdot X_u) + g\,(X \cdot X_v)]\ , \\
\eta_u^2 + \eta_v^2 &= -c^2 K \left[(X \cdot X_u)^2 + (X \cdot X_v)^2\right]\ ,
\end{aligned}
\tag{3.96}
$$

onde os coeficientes da segunda forma fundamental de $X(u,v)$ são relacionados por $e = -g$, como indica (3.90), e K é a curvatura Gaussiana dada em (1.5).

Assim, aplicando (3.96) em (3.93) pode-se determinar a seguinte condição para que uma superfície mínima parametrizada isotermicamente seja um ponto crítico de mínimo do erro quadrático médio,

$$
\left.\frac{\partial^2 \overline{\epsilon^2}(\mathbf{m_0})}{\partial t^2}\right|_{t=0} > 0 \quad \Rightarrow \quad -\frac{2}{E}\left\{-c^2 \frac{K}{E}\left[(X \cdot X_u)^2 + (X \cdot X_v)^2\right] + 8\eta^2 K\right\} > 0 \quad \Rightarrow
$$

$$
\frac{2c^2 K}{E}\left[\left(X \cdot \frac{X_u}{|X_u|}\right)^2 + \left(X \cdot \frac{X_v}{|X_v|}\right)^2 - 8\,(X \cdot N)^2\right] > 0\,,
$$

mas como $K < 0$ e os vetores $\frac{X_u}{|X_u|}$, $\frac{X_v}{|X_v|}$ e N são ortonormais, então essa condição pode ser reescrita como

$$
\frac{2c^2 K}{E}\left[(X \cdot X) - 9\,(X \cdot N)^2\right] > 0 \quad \Rightarrow \quad |X|^2 - 9|X|^2 \cos^2(\theta) < 0 \quad \Rightarrow \quad |\cos(\theta)| > \frac{1}{3}\,,
\tag{3.97}
$$

onde θ é o menor ângulo formado entre os vetores X e N.

Portanto, como $\arccos(1/3) \approx 70.5°$ então pode-se afirmar que mesmo sujeita a variações significativas, desde que $\eta(u,v) = c\,(X, N)$, a superfície mínima $X(u,v)$ ainda é estável quanto a minimização do $\overline{\epsilon^2}$. Por esse motivo, nas próximas seções novos esquemas de modulações sobre superfícies mínimas serão propostos e analisados. Além disso, deve-se ressaltar que, apesar de projetarmos modulações baseadas em superfícies mínimas, caracterizadas por sua curvatura média igual a zero, a sua curvatura Gaussiana também será considerada. Tal fato é necessário por que o valor da curvatura Gaussiana, assim como a curvatura de uma curva, também influencia na probabilidade de ocorrência de erro de ponto inicial, como mostrado na próxima seção.

3.5.4 Análise da curvatura Gaussiana em esquemas de modulação

Na subseção 3.3.4 obtivemos a expressão (3.43), que relaciona o valor da energia média de ruído ao valor da curvatura da curva $\mathbf{s}_m$. Naquela ocasião, foi possível determinar que a ocorrência de erro de ponto inicial pode ser diminuída caso a energia média de ruído seja menor do que o menor raio de curvatura da curva associada à modulação.

Nesta subseção, um resultado similar a esse será obtido. Mais precisamente, mostraremos que tanto a curvatura média quanto a curvatura Gaussiana podem influenciar na ocorrência de erros de ponto inicial. Para tanto, considere uma superfície regular parametrizada $X(u,v)$, onde u e v são as variáveis a serem transmitidas. Dessa forma, quando o ruído é AWGN e o receptor é de máxima verossimilhança, então uma solução para o problema $\min|\mathbf{r} - X(u,v)|$, como descrito em (3.69), deve satisfazer o seguinte sistema de equações

$$\begin{cases} (\mathbf{r} - X) \cdot X_u = 0\,, \\ (\mathbf{r} - X) \cdot X_v = 0\,. \end{cases} \tag{3.98}$$

Contudo, para que as soluções de (3.98) sejam pontos críticos de mínimo, então os autovalores da seguinte matriz devem ser positivos

$$\mathbf{H} = \begin{bmatrix} -(\mathbf{r} - X) \cdot X_{uu} + (X_u \cdot X_u) & -(\mathbf{r} - X) \cdot X_{uv} + (X_u \cdot X_v) \\ -(\mathbf{r} - X) \cdot X_{uv} + (X_u \cdot X_v) & -(\mathbf{r} - X) \cdot X_{vv} + (X_v \cdot X_v) \end{bmatrix}. \tag{3.99a}$$

Entretanto, antes de obtermos os autovalores de (3.99a), é necessário verificar de (3.98) que o vetor $(\mathbf{r} - X)$ é ortogonal a X_u e a X_v. Isto é, $(\mathbf{r} - X)$ possui a direção do vetor normal à superfície $X(u,v)$ e, portanto, pode ser escrito da seguinte maneira $(\mathbf{r} - X) = h\,N$, como sugerido na expressão (3.70). Dessa forma, a matriz Hessiana (3.99a) pode ser escrita como

$$\mathbf{H} = \begin{bmatrix} -h\,e + E & -h\,f + F \\ -h\,f + F & -h\,g + G \end{bmatrix}, \tag{3.99b}$$

onde os coeficientes E, F e G são dados por (1.2) e os coeficientes e, f e g são dados por (1.3).

Com isso, podemos obter os autovalores de (3.99b) como sendo as raízes do seguinte polinômio característico

$$\begin{aligned} \det(\mathbf{H} - \lambda I) &= (-h\,e + E - \lambda)\,(-h\,g + G - \lambda) - (-h\,f + F)^2 \\ &= \lambda^2 - \lambda\,[E + G - h(e+g)] + (EG - F^2)(h^2K - 2h\,H + 1)\,, \end{aligned} \tag{3.100}$$

onde H e K são as curvaturas média e Gaussiana dadas por (3.90) e (1.5), respectivamente. Além disso, quando os parâmetros são isotérmicos $E = G$ e $F = 0$, esse polinômio pode ser escrito da seguinte forma

$$\det(\mathbf{H} - \lambda\mathbf{I}) = \lambda^2 - 2\lambda E\,(1 - h\,H) + E^2(h^2K - 2h\,H + 1)\,,$$

e suas raízes são dadas por

$$\lambda = E\left(1 - h\,H \pm |h|\sqrt{H^2 - K}\right). \tag{3.101}$$

Neste caso, como $\sqrt{H^2 - K}$ é não negativo e E é sempre positivo, então a condição necessária para que os autovalores (3.101) sejam positivos pode ser expressa pela seguinte desigualdade

$$h\,H + |h|\sqrt{H^2 - K} < 1\,, \tag{3.102}$$

que, no caso particular de $X(u, v)$ ser uma superfície mínima ($H = 0$), pode ser convenientemente reescrita como

$$|h| < \frac{1}{\sqrt{-K}}\,. \tag{3.103}$$

Portanto, note que a condição (3.103) é mais geral que a condição (3.43) obtida anteriormente. Além disso, o parâmetro h pode ser associado ao ruído de transmissão com o auxílio da Figura 3.32, da aproximação (3.71) e da equação (3.70) da seguinte forma

$$\begin{cases} \mathbf{r} = X_0 + \mathbf{n} \\ \hat{X} = X_0 + (\hat{u} - u_0)X_u + (\hat{v} - v_0)X_v \\ \mathbf{r} = \hat{X} + h\,N \end{cases} \quad \Rightarrow \quad X_0 + \mathbf{n} = X_0 + (\hat{u} - u_0)X_u + (\hat{v} - v_0)X_v$$

$$(\mathbf{n} \cdot N) = (\hat{u} - u_0)\,(X_u \cdot N) + (\hat{v} - v_0)\,(X_v \cdot N) + h\,(N \cdot N) \quad \Rightarrow \quad h = (\mathbf{n} \cdot N)\,. \tag{3.104}$$

Dessa forma, quando o ruído do canal de transmissão é AWGN, então o parâmetro h pode ser considerado como uma variável aleatória Gaussiana de média nula e variância $\mathcal{N}_0/2$. Assim, com o auxílio de (3.103) e de (3.104), a ocorrência de erros de ponto inicial pode ser minimizada quando a seguinte condição for satisfeita

$$\frac{\mathcal{N}_0}{2} \leq \min\left\{\frac{1}{-K}\right\}. \tag{3.105}$$

Com isso, o valor da curvatura Gaussiana deve ser considerado no projeto de esquemas de modulações, principalmente, quando essas modulações são associadas a superfícies mínimas com parametrização isotérmica, que no caso também minimizam o erro quadrático médio, como demonstrado nas seções anteriores.

3.5.5 Exemplos de modulações associadas a superfícies mínimas

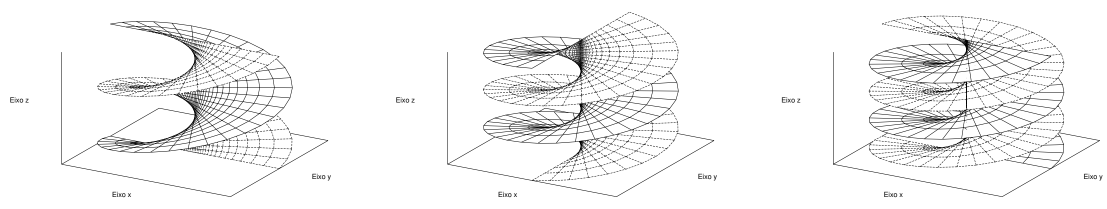
Nesta subseção, serão apresentados alguns exemplos de esquemas de modulação não-lineares associadas a superfícies mínimas no $\mathbb{R}^3$. A decisão por usar essa classe de superfícies está relacionada a suas propriedades de minimização do erro quadrático médio e de estabilidade da estimativa $\hat{\mathbf{m}}$, determinadas anteriormente. Exemplos de superfícies mínimas parametrizadas isotermicamente podem ser obtidas pela representação Weierstrass (1.6).

Da mesma maneira que as características geométricas de uma curva associada a um esquema de modulação influenciam em seu desempenho, a determinação de uma superfície entre as infinitas possibilidades de superfícies mínimas deve ser realizada levando-se em consideração alguns critérios geométricos relevantes ao processo de transmissão. Tais critérios, certamente, nos levam a escolher uma superfície sem auto-interseções, para diminuir a ocorrência de erros de ponto inicial, e que possua uma área relativamente grande, para uma dada energia média de transmissão, diminuindo assim o erro quadrático médio.

Segundo estes critérios, dentre as superfícies mínimas consideradas neste capítulo, a helicóide foi a que apresentou melhor desempenho. Para uma melhor análise da influência da parametrização da superfície $X(u,v)$ no desempenho dos sistemas de comunicações, serão consideradas três diferentes parametrizações para a superfície mínima helicóide.

$$\begin{aligned}
X(u,v) &= [u\cos(\pi\beta\, v), u\sin(\pi\beta\, v), v]\ , && (3.106a)\\
&= [\sinh(u)\cos(\pi\beta\, v), \sinh(u)\sin(\pi\beta\, v), v]\ , && (3.106b)\\
&= [\sinh(\pi\beta\, u)\cos(\pi\beta\, v), \sinh(\pi\beta\, u)\sin(\pi\beta\, v), \pi\beta\, v]\ , && (3.106c)
\end{aligned}$$

onde β é um parâmetro da modulação que influencia em seu desempenho. Geometricamente o parâmetro β está relacionado à quantidade de voltas da helicóide, como indica os gráficos da superfície helicóide para $\beta = 1, 2$ e 3 na Figura 3.34. Além disso, quando $\beta = 0$ na parametrização (3.106a), a superfície em questão é o plano e conseqüentemente a modulação associada é a linear. Em outras palavras, quando $\beta > 0$, a modulação deixa de ser linear.



(a) $\beta = 1.0$. (b) $\beta = 1.5$. (c) $\beta = 2.0$.

Figura 3.34: Superfície mínima helicóide para, para $\beta = 1, 1.5$ e 2.

Neste caso, observamos que, apesar das três parametrizações representarem a mesma superfície, os desempenhos obtidos para cada uma dessas modulações foram diferentes. Tal fato está relacionado aos seguintes valores da primeira forma fundamental de cada uma dessas parametrizações,

$$E = 1\,, \quad G = 1 + (\pi\beta\, u)^2 \quad \text{e} \quad F = 0\,, \tag{3.107a}$$

$$E = \cosh^2(u)\,, \quad G = 1 + \sinh^2(u)(\pi\beta)^2 \quad \text{e} \quad F = 0\,, \tag{3.107b}$$

$$E = G = (\pi\beta)^2 \cosh^2(\pi\beta\, u) \quad \text{e} \quad F = 0\,, \tag{3.107c}$$

respectivamente, pois, segundo (3.73c), o erro quadrático médio dos sistemas de comunicações cujas modulações possuem as métricas (3.107), pode ser expresso por

$$\overline{\epsilon^2} = \frac{\mathcal{N}_0}{4}\frac{E+G}{EG} = \frac{\mathcal{N}_0}{4}\left(\frac{1}{E} + \frac{1}{G}\right)\,,$$

que, no caso, decresce da modulação (3.107a) para a modulação (3.107b) e dessa para a modulação (3.107c), como ilustra a Figura 3.35.

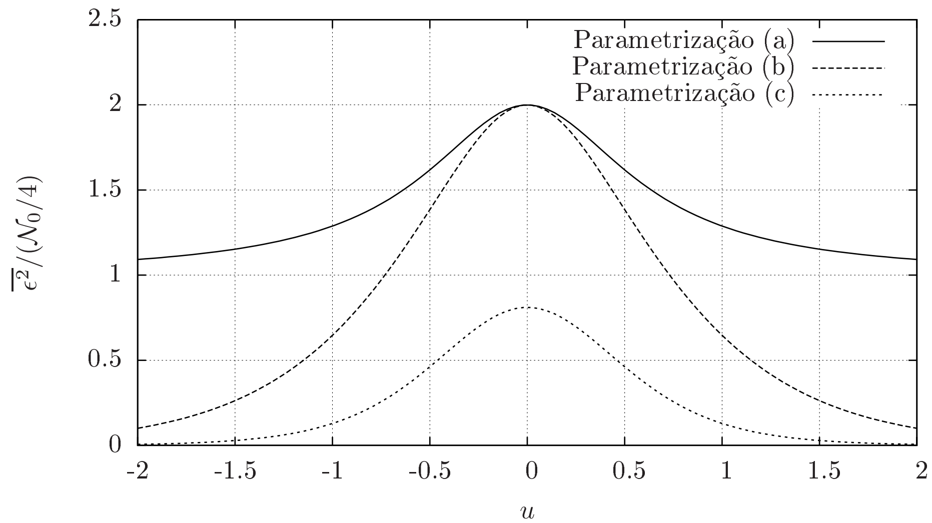


Figura 3.35: Curvas do erro quadrático médio da modulação associada a superfície mínima helicóide, para cada uma das três parametrizações consideradas quando $\beta = 1$.

Tal fato pode ser explicado com o auxílio da subseção 3.5.2, cujo resultado afirma que uma superfície mínima é um ponto crítico de mínimo do erro quadrático médio se estiver parametrizada isotermicamente, veja condição (3.90). Neste caso, segundo (3.107), a única parametrização com $E = G$ e $F = 0$ é a (3.106c), e por este motivo foi a que apresentou menor $\overline{\epsilon^2}$.

Além disso, foi realizada uma simulação computacional para determinar as curvas de desempenho (CSNR *versus* SNR) do sistema de comunicações associado à superfície mínima helicóide, parametrizadas por (3.106). Tais curvas de desempenho são apresentadas nas Figuras 3.36, 3.37 e 3.38, sendo que, como esperado, o sistema associado à parametrização isotérmica (3.106c) foi o que apresentou melhor desempenho, veja Figura 3.39 para melhor comparação. Além disso, observe de Figura 3.39 que o desempenho dessa modulação não-linear proposta está bastante próximo do desempenho ótimo, previsto pela curva OPTA.

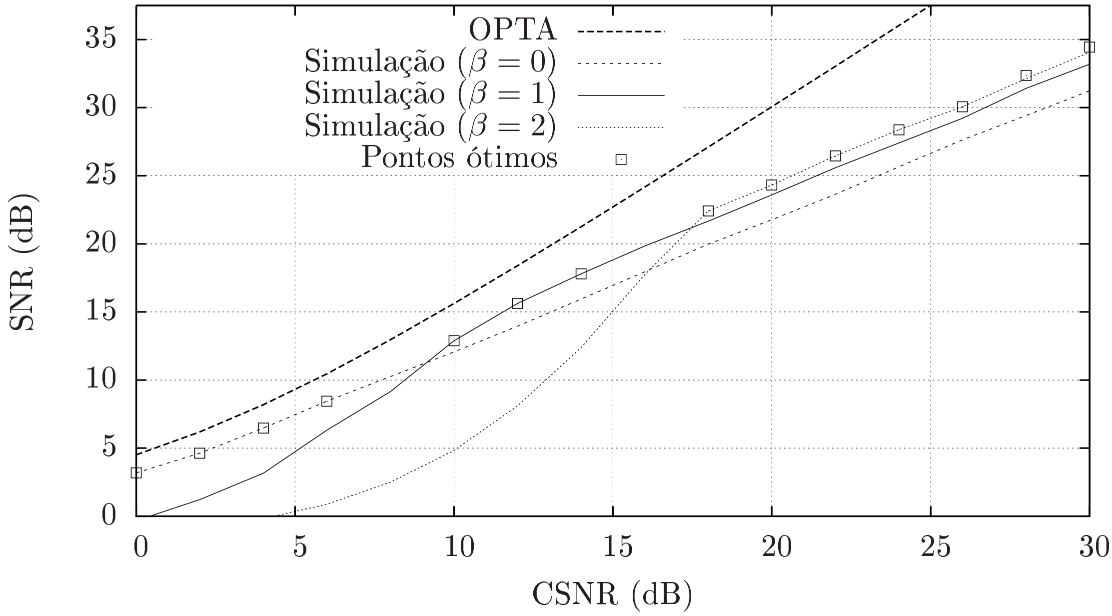


Figura 3.36: Curvas de desempenho de uma modulação não-linear associada à superfície mínima helicóide parametrizada por (3.106a), para $\beta = 1, 2$ e 3.

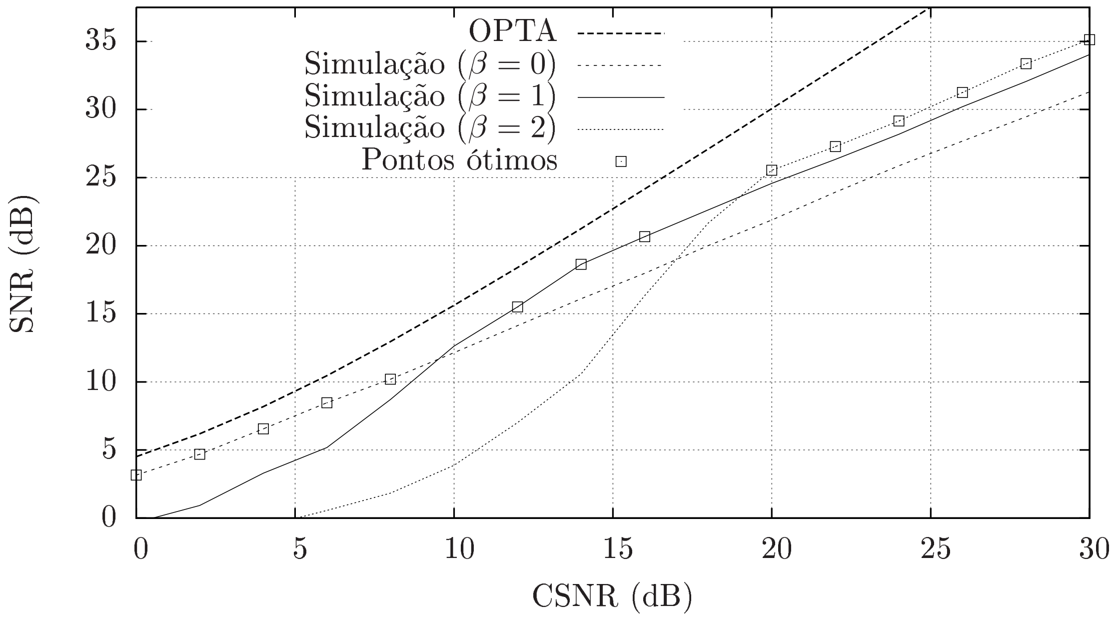


Figura 3.37: Curvas de desempenho de uma modulação não-linear associada à superfície mínima helicóide parametrizada por (3.106b), para $\beta = 1, 2$ e 3.

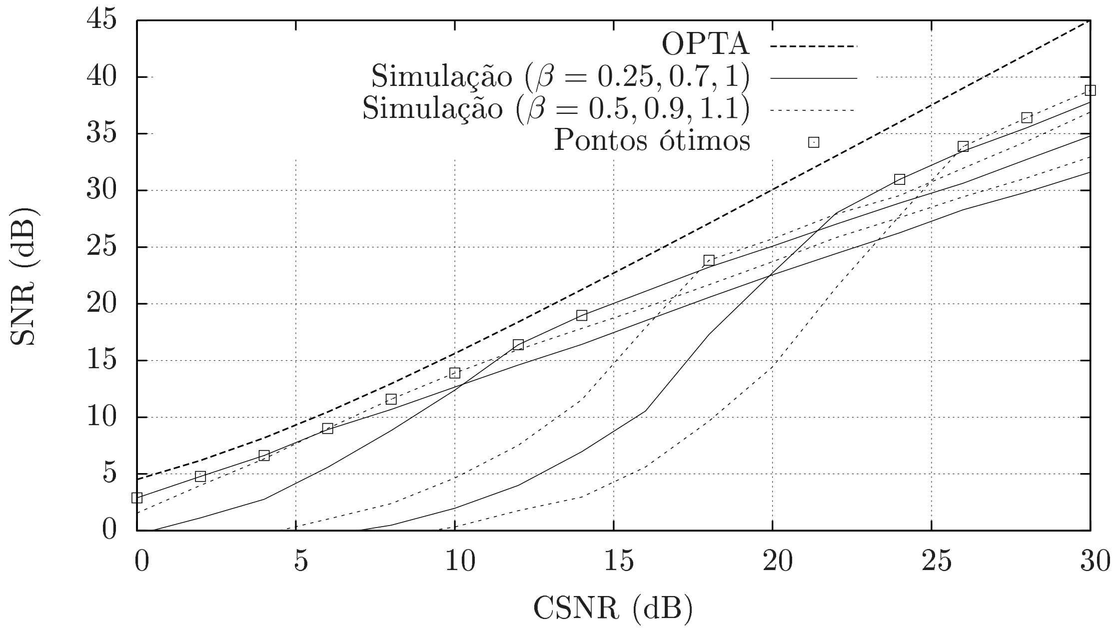


Figura 3.38: Curvas de desempenho de uma modulação não-linear associada à superfície mínima helicóide parametrizada por (3.106c), para $\beta = 0.25$, 0.5, 0.7, 0.9, 1 e 1.1.

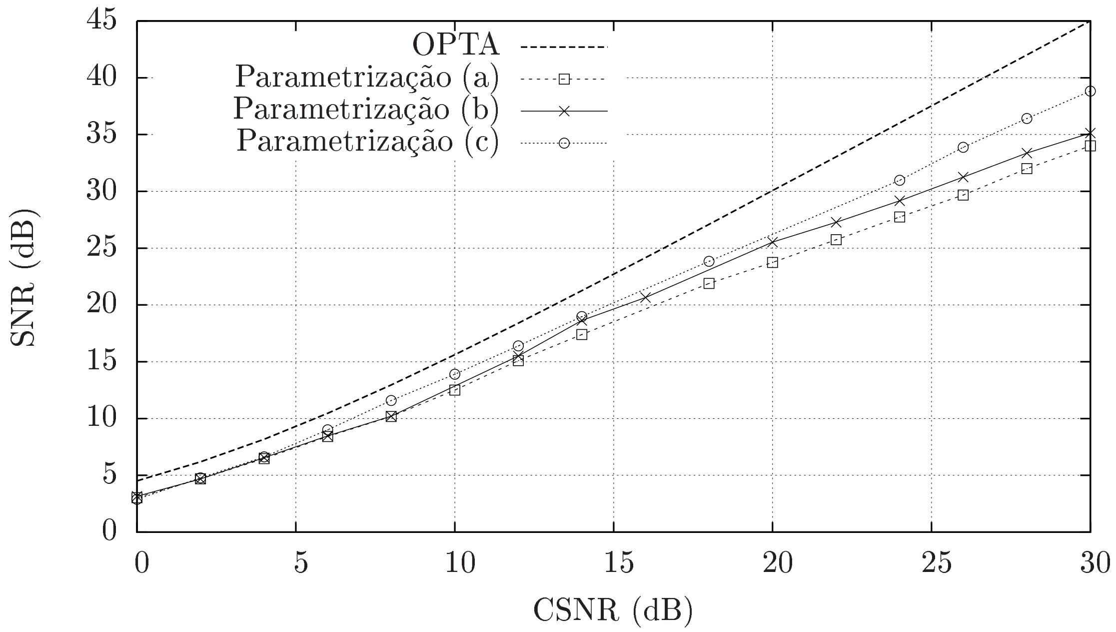


Figura 3.39: Comparação entre as curvas de desempenho das três modulações associadas à superfície mínima helicóide.

Dessa forma, é possível verificar que os resultados obtidos nesta subseção via simulação computacional estão de acordo com os resultados teóricos obtidos anteriormente, com o auxílio da aproximação (3.71), para modulações não-lineares associadas a superfícies. Tal fato confirma que as características geométricas das modulações não-lineares que foram destacadas nesta seção são suficientes para o projeto geométrico eficiente de modulações não-lineares, desde que a superfície associada a modulação satisfaça as características geométricas aqui apresentadas.

Com isso, nesta seção, podemos verificar que a interpretação geométrica que associa modulações a superfícies no espaço Euclidiano permitiu o uso de conceitos da geometria diferencial e Riemanniana no contexto de projetos de modulações. Além disso, a expressão aproximada para o erro quadrático médio que foi derivada e aplicada a espaços de curvatura seccional constante, indica que o uso de espaços hiperbólicos no projeto de modulações pode melhorar o desempenho dos sistemas de comunicações. Não obstante, verificamos que as modulações não-lineares associadas a superfícies mínimas minimizam o erro quadrático médio, mostramos também que as curvaturas média e Gaussiana influenciam no erro de ponto inicial e relacionamos essas curvaturas a energia média do ruído. Finalmente, constatamos que modulações não-lineares com bons desempenhos podem ser projetadas geometricamente.

Capítulo 4

Influência da Curvatura no Desempenho de Constelações de Sinais

Neste capítulo faremos uso de modulações não-lineares, como as consideradas no capítulo anterior, para realizar a construção e a análise de desempenho de constelações de sinais em espaços mais gerais que o Euclidiano. Neste caso, verificamos que a superfície associada a uma técnica de modulação não-linear induz de maneia natural uma nova métrica no espaço de sinais. Tal métrica, geralmente, não é Euclidiana, e a sua utilização no espaço de sinais propicia a construção de novas constelações de sinais com desempenhos, em termos da probabilidade média de erro, melhores que os desempenhos dos sistemas de comunicações tradicionais usando uma modulação linear, isto é, a curvatura é nula e o espaço de sinais é o Euclidiano. Para tanto, será necessário o uso da geometria Riemanniana para estender os principais parâmetros de desempenho das constelações de sinais para esse contexto mais geral. Dessa forma, como os espaços de sinais serão interpretados como variedades Riemannianas, então alguns invariantes geométricos dessas variedades, tal como a curvatura seccional, serão associados ao desempenho das constelações de sinais. Mais precisamente, as constelações de sinais projetadas em espaços de sinais com curvatura seccional constante e negativa, como o espaço hiperbólico, foram as que apresentaram as melhores curvas de desempenho.

4.1 Introdução

Os trabalhos [7] e [6] tratam da construção e da análise de desempenho de constelações de sinais do tipo M-PSK definidas em espaços hiperbólicos. Na análise de desempenho das constelações de sinais destes trabalhos, a métrica do espaço foi utilizada para definir uma função densidade de probabilidade do tipo Gaussiana para representar o ruído do canal de transmissão. Em [29], [30], [31] e [32], uma abordagem semelhante a essa foi realizada, mas considerando uma diversidade maior de tipos de constelações de sinais definidas em variedades Riemannianas, como, por exemplo, constelações de sinais geometricamente uniformes em espaços de curvatura seccional constante e em espaços com métrica induzida de superfícies mínimas.

Entretanto, nenhum destes trabalhos demonstrou precisamente a procedência da expressão da função densidade de probabilidade Gaussiana usada na caracterização do ruído do canal de transmissão, ou descreveu o sistema de comunicações cujo ruído tivesse essa função densidade de probabilidade do ruído. Contudo, devemos ressaltar que este fato não compromete os resultados encontrados nestes trabalhos, pois mostraremos na próxima seção que essa expressão da função densidade de probabilidade está correta, desde que seja considerada uma relação sinal ruído suficientemente grande. Este fato era esperado, visto que a função densidade de probabilidade do ruído usada nestes trabalhos foi rigorosamente definida, de modo que seu comportamento descrevesse um ruído Gaussiano do ponto de vista do espaço não Euclidiano considerado. Neste caso, alguns parâmetros geométricos foram usados, como a métrica e a curvatura do espaço, as áreas das regiões de decisão e as distâncias entre os sinais, que no caso foram associadas ao comprimento geodésico.

Dessa forma, um dos primeiros resultados deste capítulo é mostrar que a função densidade de probabilidade do ruído considerada nos trabalhos anteriores está associada à probabilidade de erro no espaço de sinais de um sistema de comunicações sujeito a ação de um ruído aditivo Gaussiano branco e usando uma modulação não-linear. Mais geralmente, iremos mostrar que o espaço de sinais está associado a uma variedade Riemanniana quando a energia média do ruído do canal de transmissão é suficientemente pequena. Com isso, usando essa associação iremos propor a construção de constelações de sinais em variedades Riemannianas, bem como realizar as correspondentes análises de desempenhos.

4.2 Função Densidade de Probabilidade Condicional, $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0)$

No Capítulo 3, as técnicas de modulação não-lineares foram associadas a curvas ou a superfícies no $\mathbb{R}^n$. Com isso, considere que as variáveis $(u, v) \in D$ possam ser transmitidas usando uma modulação associada a uma superfície regular $X : D \subset \mathbb{R}^2 \to \mathbb{R}^3$. Neste caso, uma constelação de sinais usando essa técnica de modulação pode ser representa pelo seguinte conjunto de pontos

$$\mathcal{X} = \{\mathbf{x}_1 = [u_1, v_1], \ldots, \mathbf{x}_m = [u_m, v_m]\},$$

no espaço bidimensional D. Além disso, esses sinais também pertencem à superfície $X(u, v)$ e, por esse motivo, também podem ser representados por pontos no $\mathbb{R}^3$.

Ainda no Capítulo 3 foi derivada uma expressão de $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0)$, a função densidade de probabilidade condicional de que o sinal $\mathbf{y} = [u, v]$ seja recebido dado que $\mathbf{x}_0 = [u_0, v_0]$ tenha sido enviado. Tal expressão pode ser obtida de (3.81) para o caso em que a energia média do ruído seja suficientemente pequena ou por (3.84) para um caso mais geral, onde não haja restrição quanto a energia média de ruído. Contudo, neste último caso, alguns valores de $\mathbf{y}$ são computados mais de uma vez devido à solução do problema de otimização considerado em sua obtenção.

A expressão (3.80) apresenta uma mudança de variáveis aleatórias que, nesse caso particular, relaciona o vetor ruído do canal, $\mathbf{n}$, ao vetor erro $\mathbf{y} - \mathbf{x}_0$, isto é,

$$\mathbf{n} = (\mathbf{y} - \mathbf{x}_0) \cdot G\,, \tag{4.1}$$

onde G é a matriz métrica calculada no ponto $\mathbf{x}_0$. Neste caso, note que essa expressão depende da métrica da superfície e este fato demonstra como a métrica no espaço de sinais é induzida pela métrica da superfície associada à modulação. Entretanto, devemos chamar a atenção para o fato de que a expressão (4.1) é válida para uma vizinhança de $\mathbf{x}_0$ ou, equivalentemente, quando a energia média do ruído for suficientemente pequena.

Neste caso, a função densidade de probabilidade condicional de $\mathbf{y}$ dado $\mathbf{x}_0$ pode ser determinada com o auxílio de (3.81) da seguinte forma

$$p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(\mathbf{y} - \mathbf{x}_0) \cdot G \cdot (\mathbf{y} - \mathbf{x}_0)^T}{2\sigma^2}\right] \sqrt{\det(G)}\,. \tag{4.2}$$

De acordo com (3.71), quando a energia média do ruído for suficientemente pequena, a seguinte aproximação pode ser utilizada

$$X(u, v) = X(u_0, v_0) + (\mathbf{y} - \mathbf{x}_0) \cdot J^T\,,$$

onde J é a matriz Jacobiana de $X(u, v)$. Neste caso, como $G = J^T \cdot J$, então o termo da exponencial de (4.2) é igual à distância Euclidiana ao quadrado entre os pontos $X(u, v)$ e $X(u_0, v_0)$, que, localmente, é uma boa aproximação para a distância geodésica ao quadrado entre esses pontos, $d^2(\mathbf{y}, \mathbf{x}_0)$. Por essa razão, a expressão da função densidade de probabilidade

$$p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0) = k_1 e^{-k_2\, d^2(\mathbf{y}, \mathbf{x}_0)} \sqrt{\det(G)}\,, \tag{4.3}$$

utilizada nos trabalhos [7], [6], [29], [30], [31] e [32] está bem definida. As constantes k_1 e k_2 de (4.3) são positivas e são necessárias para que a integral dessa expressão em todo o espaço de sinais seja igual a 1 e para que a variância de $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0)$ possa ser definida em função dessas constantes.

Neste contexto, pode-se considerar que a superfície regular parametrizada $X : D \subset \mathbb{R}^2 \to \mathbb{R}^3$ induz de maneira natural uma métrica em D, isto é, dados $p \in$

D e $\mathbf{v}_1, \mathbf{v}_2 \in T_pX$, o plano tangente à superfície no ponto p, então $(\mathbf{v}_1 \cdot \mathbf{v}_2)_p = (dX_p(\mathbf{v}_1) \cdot dX_p(\mathbf{v}_2))_{X(p)}$, onde $dX_p(\mathbf{v}_1)$ é a diferencial da aplicação X no ponto p segundo o vetor $\mathbf{v}_1$. Observe que, nessa situação, X passa a ser uma aplicação isométrica de D em $X(D)$ e que esse espaço bidimensional com métrica induzida da superfície $X(u, v)$ é uma variedade riemanniana bidimensional, M^2, com sistema de coordenadas (D, X).

Portanto, como o espaço de sinais pode ser interpretado como uma variedade Riemanniana, então a construção e a análise de desempenho de constelações de sinais deve ser realizada nesse novo contexto. Para tanto, alguns parâmetros de desempenho das constelações de sinais devem ser redefinidos, como, por exemplo, a probabilidade de o demodulador decidir erroneamente dado que o sinal $\mathbf{x}_i$ seja transmitido,

$$P_{e,i} = 1 - \int_{R_i} p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i)\, du\, dv\,, \tag{4.4}$$

onde R_i é a região de decisão do sinal $\mathbf{x}_i$, ou seja, o conjunto representando todos os sinais recebidos que são decididos como $\mathbf{x}_i$. Devemos ressaltar que a probabilidade de erro (4.4) é invariante por isometrias e que a probabilidade média de erro, P_e, de uma constelação de sinais é dada por

$$P_e = \sum_{i=1}^{m} p_i P_{e,i}\,, \tag{4.5}$$

onde p_i é a probabilidade de ocorrência do sinal x_i.

Além disso, a energia média de transmissão de uma constelação de sinais $\mathcal{X}$ é definida como

$$\overline{E}_t = \sum_{i=1}^{m} p_i d^2(\mathbf{x}_i, \bar{\mathbf{x}})\,, \tag{4.6}$$

onde $d^2(\mathbf{x}_i, \bar{\mathbf{x}})$ é a distância geodésica ao quadrado entre o sinal $\mathbf{x}_i$ e o baricentro da constelação $\bar{\mathbf{x}}$. Sabemos que o baricentro de uma constelação de sinais é o ponto que minimiza a energia média da constelação de sinais. Portanto, para determinar $\bar{\mathbf{x}}$ devemos calcular a primeira derivada da energia média em relação a u e v, e supor que a mesma seja zero em $\bar{\mathbf{x}}$. Dessa maneira, temos que $\bar{\mathbf{x}}$ é um ponto que satisfaz as seguintes equações diferenciais parciais

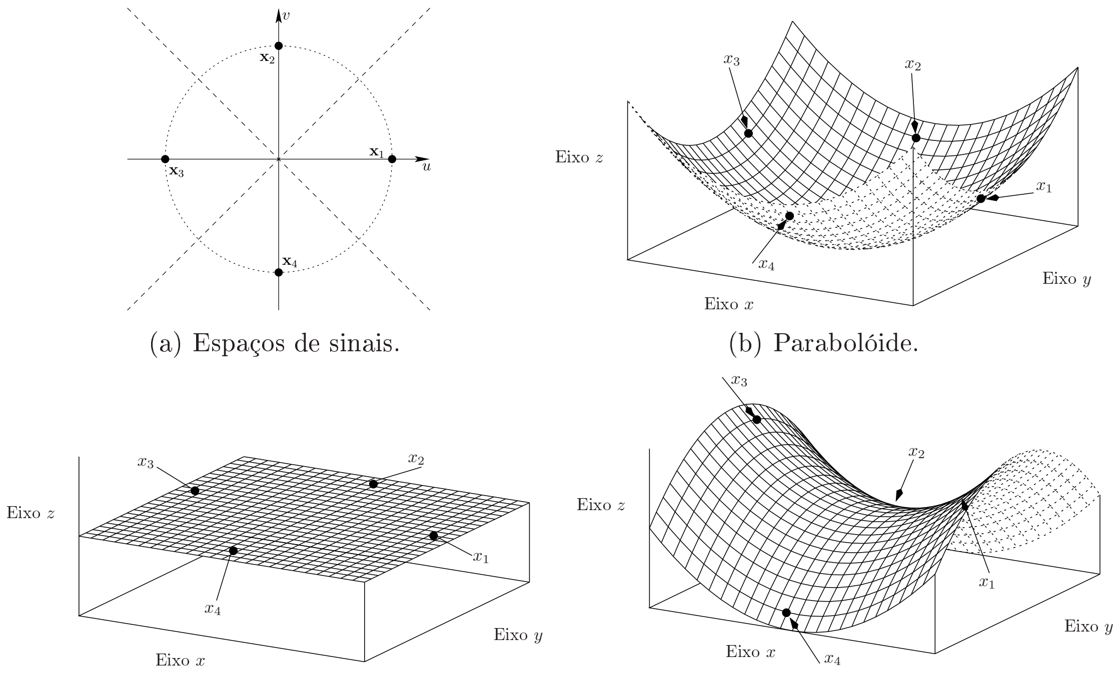
$$\begin{aligned} \left.\frac{\partial \overline{E}_t}{\partial u}\right|_{\mathbf{x}=\bar{\mathbf{x}}} &= \sum_{i=1}^{m} p_i \left.\frac{\partial\, d^2(\mathbf{x}_i, \mathbf{x})}{\partial u}\right|_{\mathbf{x}=\bar{\mathbf{x}}} = 0, \\ \left.\frac{\partial \overline{E}_t}{\partial v}\right|_{\mathbf{x}=\bar{\mathbf{x}}} &= \sum_{i=1}^{m} p_i \left.\frac{\partial\, d^2(\mathbf{x}_i, \mathbf{x})}{\partial v}\right|_{\mathbf{x}=\bar{\mathbf{x}}} = 0. \end{aligned} \tag{4.7}$$

No sentido de explicitar analiticamente a relação sinal-ruído no sistema de comunicações, a energia média do ruído no espaço de sinais, dado que $\mathbf{x}_i$ foi transmitido, é definida como

$$\sigma^2 = \int_D d^2(\mathbf{y}, \mathbf{x}_i) p(\mathbf{y}|\mathbf{x} = \mathbf{x}_i)\, du\, dv\,. \tag{4.8}$$

Com o objetivo de analisar o desempenho de sistemas de comunicações usando constelações de sinais e modulações não-lineares, vamos considerar uma modulação do tipo 4-PSK sobre as superfícies de um parabolóide, de um plano e de uma sela, Figura 4.1, cujas parametrizações são dadas por

$$\begin{aligned} X(u,v) &= [u, v, u^2 + v^2] \quad \text{(Parabolóide)}\,, & (4.9a)\\ X(u,v) &= [u, v, 0] \quad \text{(Plano)}\,, & (4.9b)\\ X(u,v) &= [u, v, u^2 - v^2] \quad \text{(Sela)}\,. & (4.9c) \end{aligned}$$



(a) Espaços de sinais. (b) Parabolóide. (c) Plano. (d) Sela.

Figura 4.1: Exemplos de modulações 4-PSK em superfícies.

Neste exemplo, consideramos que as modulações não-lineares associadas a essas superfícies possuem a mesma energia média de transmissão, (1.42). Dessa maneira, caso as variáveis u e v pertençam ao intervalo $[-a, a]$, então os valores das energias médias de transmissão dessas modulações são dadas por

$$\overline{E}_m = \frac{2}{45} a^2 (15 + 4a^2) \quad \text{(Parabolóide e Sela)} \quad \text{e} \quad \overline{E}_m = \frac{2}{3} a^2 \quad \text{(Plano)},$$

mas como a energia média da modulação linear (associada ao plano) é menor que as demais, então um fator de correção deve ser realizado no momento da análise de desempenho da constelação 4-PSK associada a essa modulação.

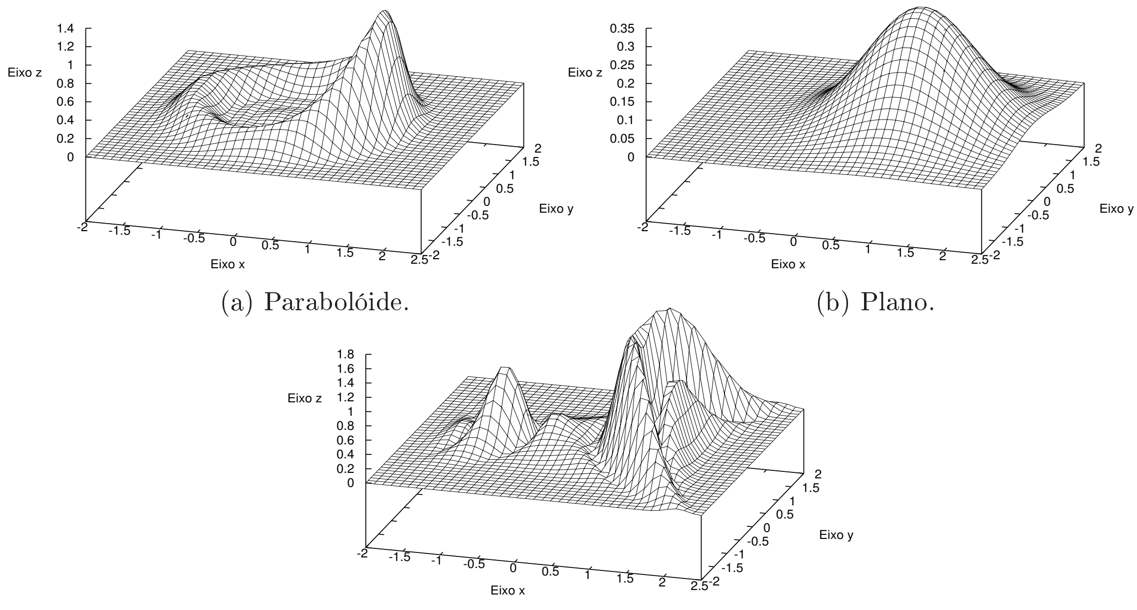
As métricas ou as primeiras formas fundamentais dessas superfícies podem ser obtidas de (1.2) como segue

$$E = 1 + 4u^2\,, \quad F = 4uv \quad \text{e} \quad G = 1 + 4v^2 \quad \text{(Parabolóide)} \tag{4.10a}$$

$$E = 1\,, \quad F = 0 \quad \text{e} \quad G = 1 \quad \text{(Plano)} \tag{4.10b}$$

$$E = 1 + 4u^2\,, \quad F = -4uv \quad \text{e} \quad G = 1 + 4v^2 \quad \text{(Sela)} \tag{4.10c}$$

Com isso, podemos usar as parametrizações (4.9) e as métricas (4.10) para determinar $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0)$ usando (3.84), veja a Figura 4.2. Optamos pelo uso de (3.84) devido ao fato dessa expressão apresentar resultados mais precisos quando a energia média do ruído não é suficientemente pequena, no caso foi utilizado $\sigma^2 = 1/2$ e $\mathbf{x}_0 = [1, 0]$ em todas as modulações.



(a) Parabolóide.

(b) Plano.

(c) Sela.

Figura 4.2: Funções densidades de probabilidades, $p(\mathbf{y}|\mathbf{x}_0)$, de exemplos de modulações 4-PSK em superfícies, calculadas para $\sigma^2 = 1/2$ e $\mathbf{x}_0 = [1, 0]$.

Analisando a Figura 4.2, verificamos que mesmo com um ruído AWGN no canal de transmissão, apenas a modulação associada à superfície planar (modulação linear) apresenta um ruído Gaussiano circular no espaço de sinais. Tal resultado já era esperado, pois a métrica induzida por essa superfície é a métrica Euclidiana, a mesma métrica em que o ruído foi definido para ser Gaussiano. Entretanto, as densidades de probabilidades apresentadas nas Figuras 4.2(a) e 4.2(c) não se assemelham a distribuições Gaussianas, mas para o espaço de sinais das modulações não-lineares associadas às superfícies parabolóide e sela esse ruído possui um comportamento Gaussiano.

As curvas de desempenho (P_e *versus* SNR) das modulações 4-PSK associadas às superfícies parabolóide, plano e sela são apresentadas na Figura 4.3. Devemos ressaltar que, na determinação dessas curvas de desempenho, foi considerado que as modulações não-lineares possuem a mesma energia média de transmissão. Por essa razão, a curva de desempenho associada ao plano possui um desempenho melhor do que a curva de desempenho associada ao parabolóide, caso contrário seu desempenho seria igual ao do parabolóide. Além disso, note que a modulação 4-PSK associada à superfície da sela foi a que apresentou melhor curva de desempenho. Este fato pode ser melhor compreendido considerando o volume de sua função densidade de probabilidade, Figura 4.2(c), na vizinhança do ponto $\mathbf{x}_0$, que no caso é maior que o volume das outras funções densidades de probabilidades. Além disso, este ganho de desempenho está associado ao valor de sua curvatura Gaussiana, que é negativa, contudo tal fato só será esclarecido na próxima seção.

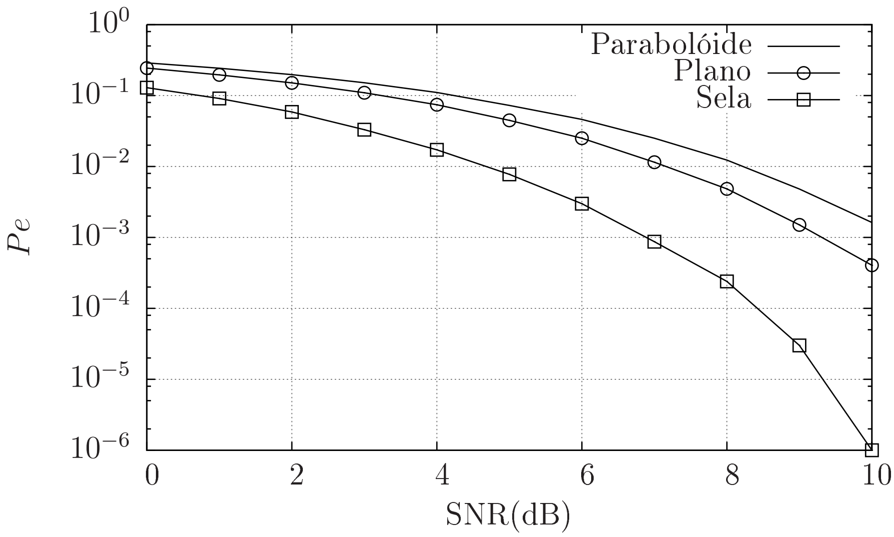


Figura 4.3: Curvas de desempenho de modulações 4-PSK associadas às superfícies parabolóide, plano e sela.

As curvaturas Gaussianas dos espaços de sinais associados às superfícies do parabolóide, do plano e da sela são dadas por

$$K = \frac{4}{(1 + 4x^2 + y^2)^{3/2}}\,, \quad K = 0 \quad \text{e} \quad K = -\frac{4}{(1 + 4x^2 + y^2)^{3/2}}\,,$$

respectivamente.

4.3 Redução da Probabilidade Média de Erro pela Consideração da Curvatura

No exemplo anterior e em [7], [6], [29], [30], [31] e [32], foi observado que o desempenho de constelações de sinais definidas em espaços com curvatura constante e negativa, quanto a probabilidade média de erro, é melhor que o desempenho das constelações de sinais em espaços com curvatura nula ou constante positiva. Nesta seção mostraremos que a probabilidade média de erro diminui com a redução da curvatura. Para tanto, vamos usar o conceito de sistema de coordenadas polar geodésico descrito em [25] para estudar o comportamento da função densidade de probabilidade $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0)$, dada em (4.2), em função da curvatura Gaussiana (K) do espaço de sinais.

Um sistema de coordenadas polar geodésico determina um sistema de coordenadas polar no plano tangente à superfície $X(\rho, \theta)$ no ponto p, T_pX, como ilustra a Figura 4.4. Dessa forma, note que as geodésicas partindo de p são radiais ($\theta = \text{cte}$) e um círculo geodésico ($\rho = \text{cte}$), que geralmente não é uma geodésica, pode ser obtido. Com isso, assumindo que o sinal transmitido $\mathbf{x}_0$ esteja relacionado ao ponto p de um sistema de coordenadas polar geodésico, então podemos associar o comprimento das geodésicas radiais (ρ) à distância desse sinal aos demais sinais da constelação, enquanto que os círculos geodésicos podem ser relacionados à região de decisão do sinal $\mathbf{x}_0$. Neste caso, quanto maior for a área desse círculo, em geral, menor será a probabilidade de erro de transmissão desse sinal.

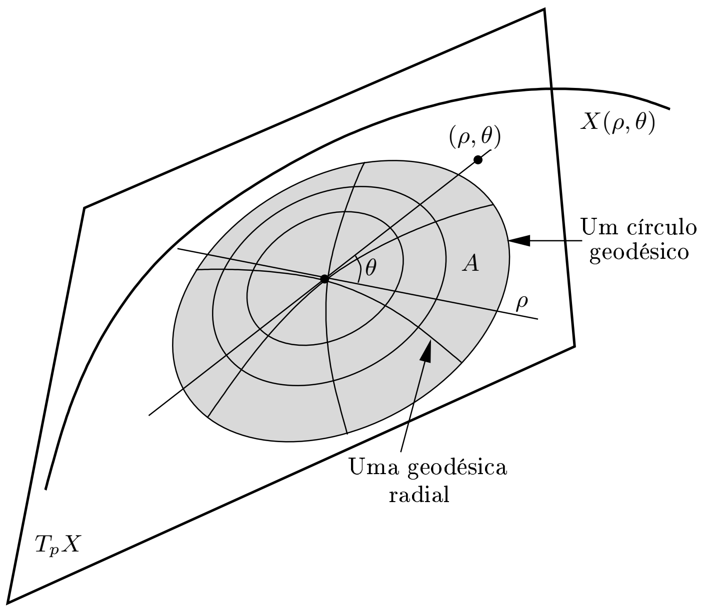


Figura 4.4: Exemplo de um sistema de coordenadas polar geodésico.

De acordo com [25], a primeira forma fundamental nesse sistema de coordenadas de uma superfície $X(\rho,\theta)$ satisfaz as condições

$$E = 1\,, \quad F = 0 \quad \lim_{\rho\to 0} G = 0 \quad \lim_{\rho\to 0}(\sqrt{G})_\rho = 1 \quad \text{e} \quad (\sqrt{G})_{\rho\rho} + K\sqrt{G} = 0\,, \tag{4.11}$$

onde K é a curvatura Gaussiana da superfície associada ao esquema de modulação. Com isso, usando uma expansão em série de Taylor, pode-se determinar uma expressão aproximada para $\sqrt{G}$ em função de ρ e K, como segue

$$\sqrt{G(\rho,\theta)} \approx \rho - \frac{\rho^3}{12}K(\rho,\theta)\,. \tag{4.12}$$

Portanto, podemos usar esse sistema de coordenadas polar geodésico para reescrever a expressão (4.2) da seguinte forma

$$p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0) = \frac{1}{2\pi\sigma^2}e^{-\frac{\rho^2}{2\sigma^2}}\sqrt{EG - F^2} \approx \frac{1}{2\pi\sigma^2}e^{-\frac{\rho^2}{2\sigma^2}}\sqrt{\rho - \frac{\rho^3}{12}K}\,. \tag{4.13}$$

Note que a expressão (4.13) explicita a dependência da função densidade de probabilidade condicional, $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0)$, relativamente à curvatura do espaço. Deve-se ressaltar que, até onde conhecemos, uma expressão similar a essa não havia sido obtida anteriormente. Em [29], as funções densidades de probabilidades foram determinadas para uma métrica específica e a influência da curvatura não era percebida tão claramente como em (4.13).

Como a integral de $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_0)$ determina a probabilidade de acerto, então podemos concluir que, caso a curvatura do espaço de sinais seja negativa, então a probabilidade de erro do sistema de comunicações é menor que em seu sistema de comunicações equivalente em espaços com curvatura nula ou positiva. Além disso, note que, quanto menor for a curvatura do espaço, menor será a probabilidade de erro.

O ganho de desempenho indicado pela expressão (4.13), quando a curvatura é negativa, pode ser interpretado geometricamente pelo o aumento da área da região de decisão do sinal, pois, neste caso, o elemento de área dessa região é igual a $\sqrt{G}$, que aumenta quando $K < 0$.

Uma outra maneira de verificar que o sinal da curvatura influencia no desempenho dos sistemas de comunicações é usar o teorema de Rauch, [26]. Contudo, tal influência é descrita de maneira qualitativa, pois esse teorema, intuitivamente, exprime o fato de que, se a curvatura diminui, então os comprimentos geodésicos entre dois pontos fixos quaisquer no espaço aumentam. Este fato implica que as distâncias entre os sinais devem aumentar com a diminuição da curvatura, sendo que este último fato faz com que a probabilidade de acerto aumente, observe a expressão (4.3).

4.4 Constelações de Sinais em Espaços com Curvatura Constante

Nesta seção apresentaremos a construção e a análise de desempenho de constelações de sinais geometricamente uniformes em espaços de curvatura constante. Inicialmente, vamos considerar constelações de sinais M-PSK e, posteriormente, constelações de sinais provenientes de tesselações $\{p, q\}$ em espaços bidimensionais com curvatura constante. Tais constelações de sinais foram analisadas nos trabalhos [30] e [32]. Por essa razão, a descrição dos resultados desses trabalhos será realizada de maneira sucinta, dando maior enfoque apenas aos resultados de interesse deste capítulo.

Segundo [27], dentre as variedades Riemannianas, os espaços de curvatura constante são os mais simples de serem estudados e possuem um número suficientemente grande de isometrias locais. Isto significa que é sempre possível "deslocar" isometricamente dois triângulos pequenos colocados em posições diferentes e verificar que podem ser sobrepostos. No projeto de constelações de sinais nesses espaços devemos fazer uso dessas isometrias, caso contrário, o ruído do canal de transmissão afetaria cada sinal a ser transmitido de uma maneira diferente. Além disso, essas isometrias garantem a existência de constelações de sinais geometricamente uniformes nos espaços de curvatura constante, pois um grupo de simetrias pode ser definido nesses espaços.

Inicialmente, considere a mesma métrica que foi utilizada nos capítulos anteriores, mais precisamente nas equações (2.39) e (3.76), para caracterizar os espaços de curvatura constante K, isto é

$$g = g_{11} = g_{22} = \frac{1}{\left[1 + \frac{K}{4}(u^2 + v^2)\right]^2} \quad \text{e} \quad g_{12} = g_{21} = 0\,. \tag{4.14}$$

Portanto, se a métrica (4.14) for usada na análise de $p(\mathbf{y}|\mathbf{x}_0)$ dada por (4.2), então podemos obter que

$$p(\mathbf{y}|\mathbf{x}_0) = \frac{g}{2\pi\sigma^2} \exp\left(-g\frac{|\mathbf{y} - \mathbf{x}_0|^2}{2\sigma^2}\right) = \frac{1}{2\pi\sigma_g^2} \exp\left(-\frac{|\mathbf{y} - \mathbf{x}_0|^2}{2\sigma_g^2}\right), \tag{4.15}$$

onde $\sigma_g^2 = \sigma^2/g$. Isto é, $p(\mathbf{y}|\mathbf{x}_0)$ é uma função densidade de probabilidade Gaussiana de média $\mathbf{x}_0$ e variância σ_g^2. Portanto, a energia média do ruído no espaço de sinais é inversamente proporcional a métrica g, ou seja,

$$\mathcal{N} = 2\sigma_g^2 = 2\sigma^2 \left[1 + \frac{K}{4}(u^2 + v^2)\right]^2. \tag{4.16}$$

Dessa forma, como esperado, caso a curvatura do espaço seja negativa, então a influência do ruído do canal de transmissão no espaço de sinais será menor, bem como a probabilidade média de erro. Além disso, a expressão (4.16) indica uma relação entre energia média do ruído e curvatura, que pode ser utilizada no cálculo da relação sinal-ruído para determinar o ganho de energia média de transmissão em função da curvatura, para um dado valor de P_e.

Como este último resultado foi obtido recentemente, então os trabalhos [30] e [32] não utilizaram a métrica (4.14) e a relação (4.16) na construção e na análise de desempenho de constelações de sinais. Contudo, esses trabalhos obtiveram resultados semelhantes, pois utilizaram a expressão (4.3) e a métrica

$$G(\rho,\theta) = \begin{cases} \frac{1}{K}\sin^2(\rho\sqrt{K})\,, & \text{se } K > 0\,, \\ \rho^2\,, & \text{se } K = 0\,, \\ -\frac{1}{K}\sinh^2(\rho\sqrt{-K})\,, & \text{se } K < 0\,, \end{cases} \tag{4.17}$$

definida em um sistema de coordenadas polar geodésico. O fato das métricas (4.14) e (4.17) serem diferentes não altera o desempenho das constelações de sinais definidas nesses espaços, dado que existe uma transformação isométrica entre esses espaços e entre essas constelações de sinais. Tal conclusão poder ser melhor compreendida verificando que as expressões de $p(\mathbf{y}|\mathbf{x}_0)$, dadas por (4.2) e (4.3), dependem exclusivamente da métrica do espaço.

Como nesta seção iremos considerar a construção de constelações de sinais em espaços com curvatura seccional constante, K, então a seguinte denominação será utilizada: a superfície de uma esfera $\mathbb{S}^2$ representará o espaço com $K > 0$; o plano Euclidiano $\mathbb{R}^2$ representará o espaço com $K = 0$; e o plano hiperbólico $\mathbb{H}^2$ representará o espaço com $K < 0$.

É importante ressaltar que os espaços $\mathbb{S}^2$ e $\mathbb{H}^2$ possuem uma estrutura geométrica diferente da Euclidiana. Isto significa, por exemplo, que a distância entre quaisquer dois pontos será obtida pelo menor comprimento geodésico entre eles. Esse e outros conceitos sobre os espaços com curvatura constante e variedades Riemannianas foram apresentados na Seção 1.3 e também podem ser obtidos em [25], [26] e [40].

Neste caso, em $\mathbb{R}^2$ a distância entre quaisquer dois pontos $z_1, z_2 \in \mathbb{R}^2$ é dada por

$$d_{\mathbb{R}} = |z_1 - z_2|\,, \tag{4.18}$$

onde $|\cdot|$ denota o módulo de um número complexo e $z_i = r_i e^{j\theta_i}$, para $i = 1, 2$.

Em $\mathbb{S}^2$, a menor distância geodésica entre quaisquer dois pontos $z_1, z_2 \in \mathbb{S}^2$ é dada por

$$d_{\mathbb{S}} = \frac{2\pi l}{\sqrt{K}} \pm \frac{1}{j\sqrt{K}} \ln \frac{|1 + z_1\bar{z}_2| + j|z_1 - z_2|}{|1 + z_1\bar{z}_2| - j|z_1 - z_2|}\,, \tag{4.19}$$

onde l é o número de vezes que a geodésica passa pelo ponto z_1 ou pelo seu antípoda, até chegar ao ponto z_2, $z_i = r_i e^{j\theta_i}$ e $r_i = -j(e^{j\rho_i\sqrt{K}} - 1)/(e^{j\rho_i\sqrt{K}} + 1)$, para $i = 1, 2$.

E em $\mathbb{H}^2$ a distância geodésica entre quaisquer dois pontos $z_1, z_2 \in \mathbb{H}^2$ é dada por

$$d_{\mathbb{H}} = \frac{1}{\sqrt{-K}} \ln \frac{|1 - z_1\bar{z}_2| + |z_1 - z_2|}{|1 - z_1\bar{z}_2| - |z_1 - z_2|}\,, \tag{4.20}$$

onde $z_i = r_i e^{j\theta_i}$ e $r_i = (e^{\rho_i\sqrt{-K}} - 1)/(e^{\rho_i\sqrt{-K}} + 1)$, para $i = 1, 2$.

Para analisar o desempenho das constelações de sinais nos espaços de curvatura constante com métrica dada por (4.17), vamos utilizar (4.3) juntamente com as distâncias (4.18), (4.19) e (4.20), associadas às respectivas curvaturas do espaço. Neste caso, assim como foi realizado em [29], podemos determinar as seguintes funções densidades de probabilidade no espaço de sinais associado ao $\mathbb{R}^2$, $\mathbb{H}^2$ e $\mathbb{S}^2$, em um sistema de coordenadas polar geodésico.

1. Para $K = 0$, temos

$$p_{\mathbb{R}}(\rho, \theta) = k_1 e^{-k_2\rho^2}\rho\,, \tag{4.21}$$

onde $k_1 = k_2/\pi$ e a energia média do ruído é dada por $\mathcal{N} = 1/k_2$, como esperado.

2. Para $K < 0$, temos

$$p_{\mathbb{H}}(\rho, \theta) = \frac{k_1}{\sqrt{-K}} e^{-k_2\rho^2} \sinh(\sqrt{-K}\rho)\,, \tag{4.22}$$

com

$$k_1 = \frac{\pi^{-3/2} e^{K/4k_2}\sqrt{-Kk_2}}{\operatorname{erf}(\sqrt{-K}/2\sqrt{k_2})}\,,$$

onde $\operatorname{erf}(z)$ denota a função erro, definida para todo $z \in \mathbb{C}$ como

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-w^2} dw\,.$$

Neste caso, a energia média do ruído é dada por

$$\mathcal{N} = \frac{2\sqrt{-Kk_2}e^{K/4k_2} + \sqrt{\pi}\operatorname{erf}(\sqrt{-K/4k_2})(2k_2 - K)}{4k_2^2\sqrt{\pi}\operatorname{erf}(\sqrt{-K/4k_2})}\,.$$

3. Para $K > 0$, temos

$$p_{\mathbb{S}}(\rho, \theta) = \frac{k_1}{\sqrt{K}} e^{-k_2\rho^2} |\sin(\sqrt{K}\rho)|\,, \tag{4.23}$$

com

$$k_1 = \frac{\sqrt{K}}{2\pi} \left(\sum_{i=0}^{\infty} \int_{i\pi}^{(i+1)\pi} (-1)^i e^{-k_2\rho^2} \sin(\sqrt{K}\rho)\, d\rho \right)^{-1}.$$

Assim, quando $k_2 \gg K$, temos que a constante k_1 pode ser aproximada por

$$k_1 \approx \frac{i\pi^{-3/2} e^{K/4k_2}\sqrt{Kk_2}}{\operatorname{erf}(i\sqrt{K}/2\sqrt{k_2})}\,.$$

Conseqüentemente, a energia média do ruído é dada por

$$\mathcal{N} \approx \frac{2i\sqrt{Kk_2}e^{K/4k_2} + \sqrt{\pi}\mathrm{erf}(i\sqrt{K/4k_2})(2k_2 - K)}{4k_2^2\sqrt{\pi}\mathrm{erf}(i\sqrt{K/4k_2})} .$$

Em [29] e [30], as funções densidades de probabilidades (4.21), (4.22) e (4.23) foram analisadas para alguns valores de energia média de ruído e, como esperado de (4.15), representaram funções densidades de probabilidades Gaussianas nos espaços de sinais. Este fato é muito importante para verificarmos que o modelo matemático utilizado está descrevendo corretamente os parâmetros de desempenho do sistema de comunicação em questão. Por este motivo, a seguir, iremos apresentar os principais resultados dos trabalhos [30] e [32].

4.4.1 Constelações de sinais M-PSK em espaços de curvatura constante

Uma constelação de sinais M-PSK no espaço de sinais com métrica (4.17) é dada por

$$\mathbf{x}_i = \rho\, e^{j\theta_i}, \quad i = 1, 2, \ldots, M ,$$

onde $\theta_i = (2i-1)\pi/M$ e ρ é a distância geodésica entre o sinal $\mathbf{x}_i$ e a origem $(0,0)$ do espaço de sinais.

Usando as funções densidades de probabilidades (4.21), (4.22) e (4.23) e as distâncias (4.18), (4.19) e (4.20) podemos determinar as curvas de desempenho das constelações de sinais 4-PSK nos espaços de curvatura constante. Tais desempenhos são apresentados na Figura 4.5 para $K = 1, 0, -1$ e -2. Note que a constelação de sinais definida no espaço hiperbólico é a que apresenta melhor desempenho. Além disso, observe que a probabilidade de erro do sistema diminui quando a curvatura do espaço de sinais passa de $K = -1$ para $K = -2$. Tal resultado pode ser explicado matematicamente pela análise da expressão (4.15).

O desempenho de constelações de sinais 8-PSK e 16-PSK também foram calculados para os espaços de curvatura constante, como ilustra a Figura 4.6. Contudo, a curva de desempenho para o espaço de curvatura positiva não será apresentado, pois seu desempenho foi muito inferior aos dos demais sistemas considerados.

Além dessas curvas de desempenho, uma outra curva determinando a variação da probabilidade média de erro em função da curvatura do espaço de sinais foi determinada. A Figura 4.7 mostra essa curva para as constelações de sinais 4-PSK e 8-PSK, quando a relação sinal-ruído permanece constante em torno de 6 dB. Neste caso, note que a probabilidade de erro diminui exponencialmente com a redução da curvatura do espaço. Esse fato indica que, dada uma energia de transmissão fixa, o desempenho do sistema pode ser melhorando com a redução da curvatura. Esse resultado é semelhante

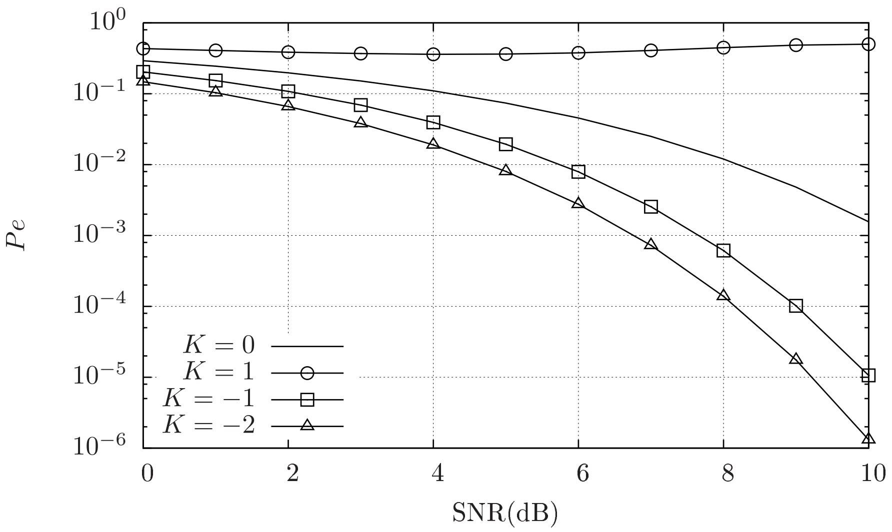
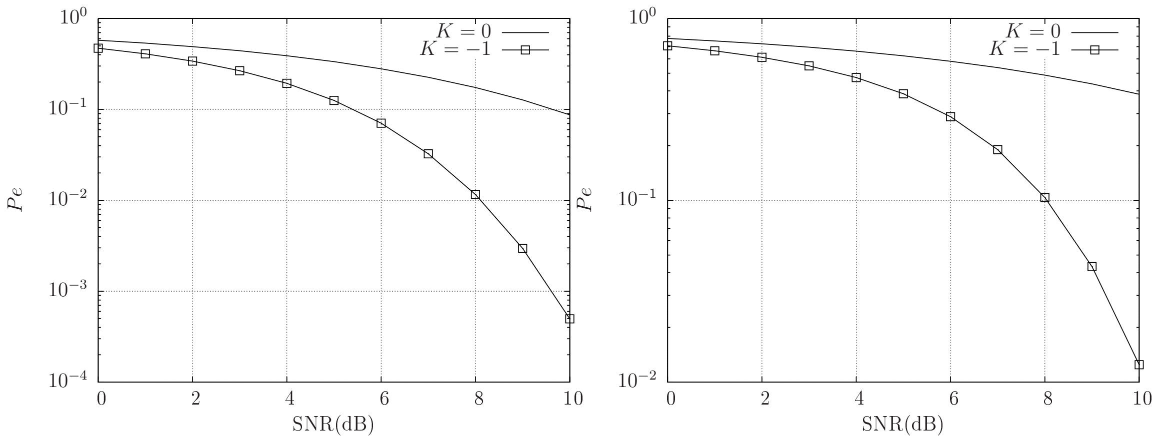


Figura 4.5: Curvas de desempenho de constelações de sinais 4-PSK em espaços de curvatura constante, $K = 1, 0, -1$ e -2.



(a) Constelação 8-PSK.

(b) Constelação 16-PSK.

Figura 4.6: Curvas de desempenho de constelações de sinais 8-PSK e 16-PSK em espaços de curvatura constante, $K = 0$ e $K = -1$.

à troca de energia média de transmissão pela largura de banda, que é freqüentemente utilizadas em sistemas de comunicações.

4.4.2 Constelações de sinais provenientes de tesselações $\{p, q\}$ em espaços de curvatura constante

Nesta subseção, construiremos e analisaremos o desempenho de constelações de sinais geometricamente uniformes provenientes de tesselações $\{p, q\}$ em espaços bidimensionais com curvatura seccional constante. Neste caso, verificamos novamente que as constelações de sinais em espaços com curvatura negativa apresentam os melhores

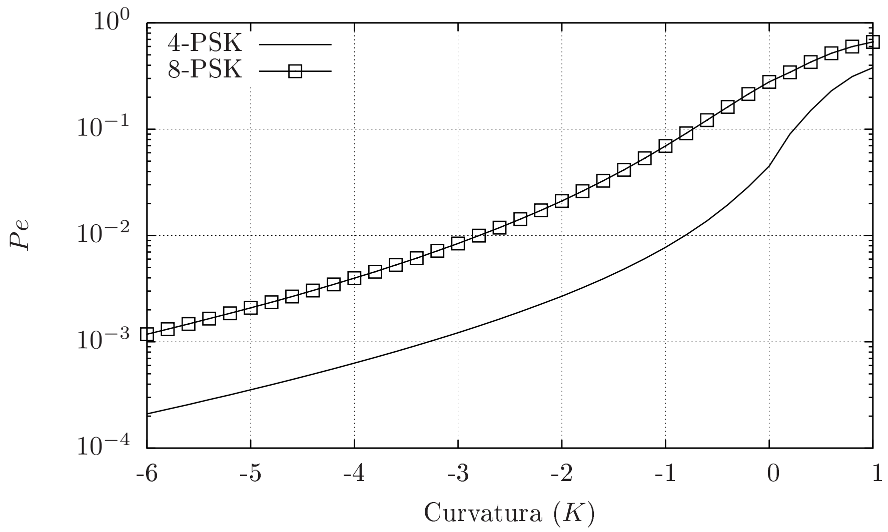


Figura 4.7: Variação da probabilidade média de erro em função da curvatura do espaço de sinais para constelações de sinais 4-PSK e 8-PSK e SNR fixo e igual a 6 dB.

desempenhos em termos da probabilidade de erro quando comparadas com as constelações de sinais em espaços com curvatura maior ou igual a zero. Para tanto, foi utilizada uma medida de energia normalizada de uma tesselação, definida em [32], que auxiliou na análise de desempenho das constelações de sinais nos diferentes espaços com curvatura constante.

O interesse pela teoria dos reticulados, no contexto de sistemas de comunicações digitais, foi estimulado pela sua conexão com a teoria dos números e de códigos corretores de erros. Neste caso, devemos ressaltar que a teoria dos reticulados demonstrou ser uma ferramenta de grande importância para o problema de empacotamento de esferas, auxiliando na construção de boas constelações de sinais ou códigos ótimos dentro do contexto dos trabalhos de Nyquist e Shannon. Para os leitores que desejam um maior entendimento sobre a teoria dos reticulados e do problema de empacotamento de esferas no espaço Euclidiano n-dimensional, recomendamos a referência [16].

Devido ao grande número de simetrias existente nos espaços de curvatura constante, é sempre possível construir um recobrimento (ladrilhamento ou tesselação) dos mesmos. Tal recobrimento determina, de maneira natural, constelações de sinais geometricamente uniformes, [9], formadas pelos baricentros das regiões fundamentais (região de Voronoi) da tesselação. Além disso, à região fundamental está associado o correspondente grupo de simetrias. Portanto, nesses espaços, é sempre possível construir constelações de sinais com espectro de distância independente do sinal considerado e regiões de decisão congruentes, isto é, constelações geometricamente uniformes.

As tesselações $\{p, q\}$ são caracterizadas por polígonos regulares de p lados, sendo que cada vértice é recoberto por q desses polígonos. Como exemplo, considere o reticulado $\{6, 3\}$ que recobre o plano Euclidiano por hexágonos cujos vértices são recobertos por três hexágonos, veja Figura 4.8. Para uma abordagem complementar sobre as tesselações $\{p,q\}$ em espaços com curvatura constante, referimos o leitor para [22] e [29].

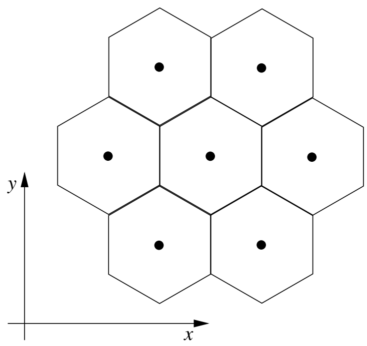
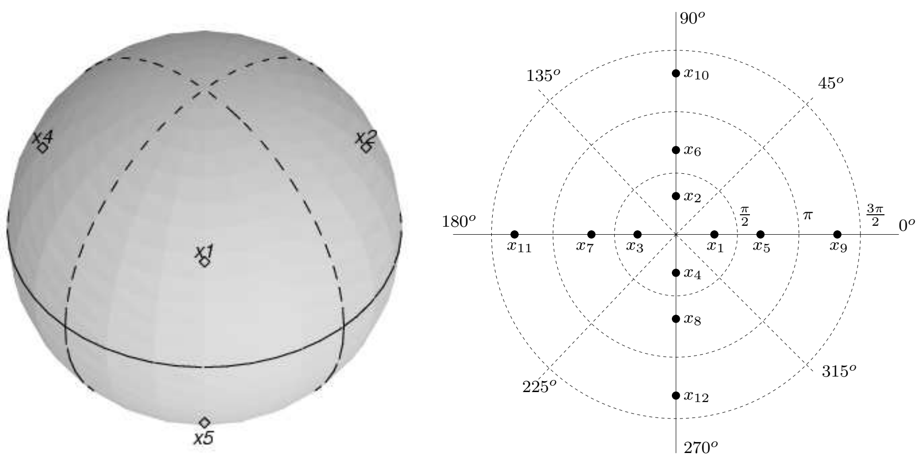


Figura 4.8: Recobrimento de $\mathbb{R}^2$ pela tesselação $\{6,3\}$.

Através do Teorema de Gauss-Bonnet, [25], e de [29], pode-se verificar que as tesselações $\{p,q\}$ em $\mathbb{S}^2$, $\mathbb{R}^2$ e $\mathbb{H}^2$ satisfazem as seguintes condições

$$\begin{cases} (p-2)(q-2) < 4\,, & \text{se } K > 0\,, \\ (p-2)(q-2) = 4\,, & \text{se } K = 0\,, \\ (p-2)(q-2) > 4\,, & \text{se } K < 0\,. \end{cases} \tag{4.24}$$

Como exemplo, considere a tesselação $\{3,4\}$ na superfície de uma esfera como ilustra a Figura 4.9(a). Neste caso, note que os sinais são os vértices de um cubo e as regiões fundamentais são as faces de um octaedro. Este fato pode ser generalizado da seguinte forma: toda tesselação $\{p,q\}$ tem uma tesselação dual $\{q,p\}$ e os vértices da tesselação $\{p,q\}$ constituem os baricentros da tesselação dual $\{q,p\}$ e vice-versa.



(a) Visualização na esfera. (b) No espaço de sinais.

Figura 4.9: Tesselação $\{3,4\}$ em $\mathbb{S}^2$ em um sistema de coordenadas polar geodésico.

A correspondente tesselação $\{3,4\}$ em coordenadas polares geodésicas é como mostrada na Figura 4.9(b). Note que devemos tesselar a superfície esférica por triângulos regulares (regiões de Voronoi) cujos vértices possuem quatro vizinhos. Note também que os lados dos triângulos são formados pelas geodésicas nesta superfície, e por esse motivo a região de Voronoi do sinal é também conhecida como triângulo geodésico.

Dessa forma, através da equação (4.24), podemos encontrar todas as tesselações em $\mathbb{S}^2$ e verificar que as mesmas são formadas pelos sólidos Platônicos em $\mathbb{R}^3$, veja Tabela 4.1. Além disso, essas tesselações geram os códigos de Slepian.

$\{p,q\}$	**Número de Regiões**	**Poliedro**
$\{3,3\}$	4	tetraedro
$\{3,4\}$	8	octaedro
$\{3,5\}$	20	icosaedro
$\{4,3\}$	6	cubo
$\{5,3\}$	12	dodecaedro

Tabela 4.1: Exemplos de tesselações em $\mathbb{S}^2$ recobrindo a superfície de uma esfera por polígonos regulares.

Neste ponto convém introduzir alguns parâmetros relevantes na construção e análise das tesselações $\{p,q\}$ em espaços com curvatura constante. Para tanto, considere uma região fundamental representada por um polígono regular de p lados como mostrado na Figura 4.10. Nesta situação, os ângulos α e β devem valer $2\pi/p$ e π/q, respectivamente, enquanto que os raios externo (ρ_l) e interno (ρ_p) deverão ser obtidos através das equações (4.18), (4.19) e (4.20) e dos correspondentes valores de α e β. A determinação dos valores de ρ_l e ρ_p implica na utilização dos mesmos no cálculo da densidade de empacotamento, da energia média, da probabilidade de erro, dentre outros.

Para o caso particular do plano Euclidiano, temos que ρ_l pode assumir qualquer valor positivo. Em outras palavras, podemos construir as tesselações Euclidianas $\{4,4\}$, $\{3,6\}$ e $\{6,3\}$ para qualquer valor positivo de ρ_l, uma vez que a propriedade de similaridade (homotetia) é válida.

Contudo, quando da construção de tesselações em $\mathbb{S}^2$ e $\mathbb{H}^2$, verificamos que, para valores fixados de p, q e K, existe apenas um único valor de ρ_l, isto porque em $\mathbb{S}^2$ e $\mathbb{H}^2$ a medida de distância está associada a ângulos (transformação conforme). Portanto, em $\mathbb{S}^2$, ρ_l é dado por

$$\rho_l = \frac{1}{j\sqrt{K}} \ln \frac{1+jr_l}{1-jr_l}, \tag{4.25}$$

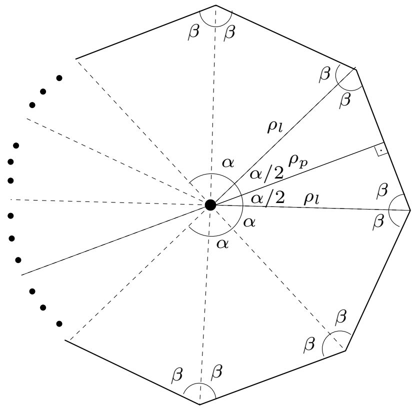


Figura 4.10: Região fundamental de uma tesselação $\{p,q\}$.

onde

$$r_l = \sqrt{\frac{\cos^2(\alpha/2) - \sin(\beta)\sin(\alpha+\beta)}{\sin^2(\beta) - \cos^2(\alpha/2)}}\,.$$

No modelo do disco de Poincaré para o espaço hiperbólico, o valor de ρ_l é dado por

$$\rho_l = \frac{1}{\sqrt{-K}} \ln \frac{1+r_l}{1-r_l}\,, \tag{4.26}$$

onde

$$r_l = \sqrt{\frac{\sin^2(\alpha/2) + \cos(\beta)\cos(\alpha+\beta)}{\cos^2(\beta) - \sin^2(\alpha/2)}}\,.$$

O valor do raio interno (ou de empacotamento), ρ_p, pode ser facilmente obtido utilizando a regra do seno em $\mathbb{S}^2$, $\mathbb{R}^2$ e $\mathbb{H}^2$, conduzindo a

$$\rho_p = \begin{cases} \frac{1}{\sqrt{K}} \arcsin\left(\sin\left(\rho_l\sqrt{K}\right)\sin(\beta)\right)\,, & \text{se } K > 0\,, \\ \rho_l \sin(\beta)\,, & \text{se } K = 0\,, \\ \frac{1}{\sqrt{-K}} \operatorname{arcsinh}\left(\sinh\left(\rho_l\sqrt{-K}\right)\sin(\beta)\right), & \text{se } K < 0\,. \end{cases} \tag{4.27}$$

Uma outra medida importante, para o desenvolvimento deste trabalho, é a área de um disco de raio ρ, denotada por $A_d(\rho)$, em $\mathbb{S}^2$, $\mathbb{R}^2$ e $\mathbb{H}^2$. Através do Teorema de Gauss-Bonnet concluímos que

$$A_d(\rho) = \begin{cases} \frac{4\pi}{K}\sin^2\left(\frac{\rho\sqrt{K}}{2}\right), & \text{se } K > 0\,, \\ \pi\rho^2\,, & \text{se } K = 0\,, \\ -\frac{4\pi}{K}\sinh^2\left(\frac{\rho\sqrt{-K}}{2}\right), & \text{se } K < 0\,. \end{cases} \tag{4.28}$$

Como mostrado em [29], a área, A_l, da região fundamental de uma tesselação $\{p,q\}$ em um espaço com K constante, é dada por

$$A_l = \begin{cases} \pi\frac{4-(p-2)(q-2)}{qK}\,, & \text{se } K \neq 0\,, \\ p\rho_p^2\tan\left(\frac{\alpha}{2}\right), & \text{se } K = 0\,. \end{cases} \tag{4.29}$$

Conhecendo os valores dos raios ρ_l e ρ_p e das áreas A_d e A_l podemos definir Δ, a densidade de empacotamento da tesselação, e Θ, a densidade de empacotamento da tesselação dual (com superposição de esferas), veja [16], como sendo

$$\Delta = \frac{A_d(\rho_p)}{A_l} \qquad \text{e} \qquad \Theta = \frac{A_d(\rho_l)}{A_l}. \tag{4.30}$$

Os valores das densidades Δ e Θ das tesselações $\{p,q\}$ são mostradas nas Tabelas 4.2 e 4.3. Note que o valor da densidade de uma tesselação no plano hiperbólico, por exemplo, não depende da curvatura valer -1 ou -10 e que a tesselação mais densa é a $\{\infty, 3\}$.

$\boldsymbol{\Delta}$	$\boldsymbol{q=3}$	$\boldsymbol{q=4}$	$\boldsymbol{q=5}$	$\boldsymbol{q=6}$	$\boldsymbol{q=7}$	$\boldsymbol{q=8}$	$\cdots$	$\boldsymbol{q=\infty}$
$\boldsymbol{p=3}$	0.8453	0.7340	0.6583	0.6046	0.5649	0.5344	$\cdots$	0.3094
$\boldsymbol{p=4}$	0.8787	0.7854	0.7206	0.6742	0.6397	0.6131	$\cdots$	0.4142
$\boldsymbol{p=5}$	0.8961	0.8120	0.7528	0.7101	0.6781	0.6535	$\cdots$	0.4675
$\boldsymbol{p=6}$	0.9069	0.8284	0.7725	0.7321	0.7017	0.6782	$\cdots$	0.5000
$\boldsymbol{p=7}$	0.9143	0.8396	0.7860	0.7470	0.7177	0.6950	$\cdots$	0.5219
$\boldsymbol{p=8}$	0.9197	0.8478	0.7958	0.7578	0.7293	0.7071	$\cdots$	0.5377
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\boldsymbol{p=\infty}$	0.9549	0.9003	0.8584	0.8270	0.8030	0.7842	$\cdots$	0.6366

Tabela 4.2: Valores da densidade Δ para uma tesselação $\{p,q\}$.

$\boldsymbol{\Theta}$	$\boldsymbol{q=3}$	$\boldsymbol{q=4}$	$\boldsymbol{q=5}$	$\boldsymbol{q=6}$	$\boldsymbol{q=7}$	$\boldsymbol{q=8}$	$\cdots$	$\boldsymbol{q=\infty}$
$\boldsymbol{p=3}$	1.3333	1.6906	2.0535	2.4184	2.7843	3.1508	$\cdots$	∞
$\boldsymbol{p=4}$	1.2679	1.5708	1.8819	2.1962	2.5119	2.8284	$\cdots$	∞
$\boldsymbol{p=5}$	1.2321	1.5055	1.7889	2.0759	2.3648	2.6547	$\cdots$	∞
$\boldsymbol{p=6}$	1.2092	1.4641	1.7300	2.0000	2.2721	2.5452	$\cdots$	∞
$\boldsymbol{p=7}$	1.1933	1.4354	1.6892	1.9475	2.2080	2.4696	$\cdots$	∞
$\boldsymbol{p=8}$	1.1815	1.4142	1.6592	1.9089	2.1609	2.4142	$\cdots$	∞
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\boldsymbol{p=\infty}$	1.1027	1.2733	1.4604	1.6540	1.8508	2.0493	$\cdots$	∞

Tabela 4.3: Valores da densidade Θ para uma tesselação $\{p,q\}$.

Para realizar a análise de desempenho das constelações de sinais é necessário conhecer a relação sinal-ruído do sistema. Contudo, para determinar esse parâmetro, devemos conhecer a energia média da constelação. Porém uma tesselação vista como uma constelação sinais possui infinitos sinais e, conseqüentemente, possui energia média infinita.

Neste caso, poderíamos tentar analisar o desempenho das constelações de sinais através da determinação das probabilidades de erros condicionais associadas às correspondentes regiões fundamentais de cada tesselação, considerando um mesmo valor de ρ_l (metade da distância mínima da constelação de sinais) e uma mesma energia média do ruído $\mathcal{N}$.

Assim, a Figura 4.11 exibe a probabilidade de erro média das regiões fundamentais das tesselações Euclidianas $\{3,6\}$, $\{4,4\}$ e $\{6,3\}$ em função da distância mínima ($d_{min} = 2\rho_l$) das constelações, para $\mathcal{N} = 1$.

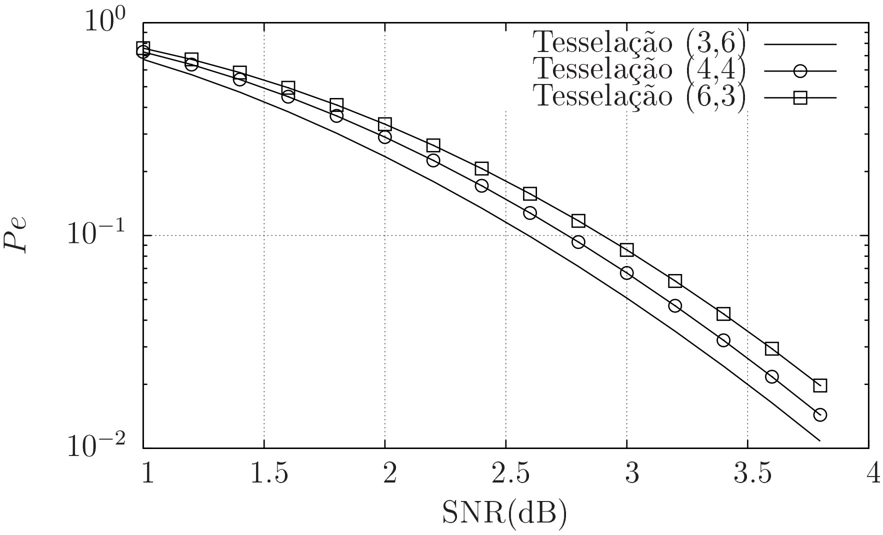


Figura 4.11: Probabilidade de erro *versus* distância mínima das tesselações $\{3,6\}$, $\{4,4\}$ e $\{6,3\}$, para $\mathcal{N} = 1$.

Note na Figura 4.11 que a região fundamental da tesselação $\{3,6\}$ é a que apresenta o menor valor da probabilidade de erro. Este fato pode ser melhor entendido ao analisar a Figura 4.12, pois verifica-se que a região fundamental da tesselação $\{3,6\}$ é a de maior área e, por esse motivo, a probabilidade de acerto é maior (maior região de integração), uma vez que cada uma dessas regiões está sob a ação de um mesmo tipo de ruído, o AWGN.

Entretanto, o procedimento de análise de desempenho apresentado nas Figuras 4.11 e 4.12 não está completo. Isto é, duas constelações de sinais, usando aproximadamente a mesma quantidade de sinais, provenientes das tesselações $\{3,6\}$ e $\{6,3\}$ não têm a mesma energia média, pois a tesselação $\{3,6\}$ possui a menor densidade de empacotamento dentre as três tesselações. Em outras palavras, deve estar claro que a constelação proveniente da tesselação $\{6,3\}$ tem menor energia, pois apresenta a maior

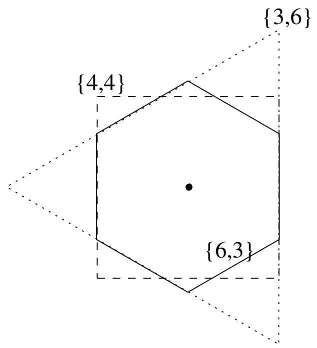


Figura 4.12: Regiões fundamentais das tesselações $\{3,6\}$, $\{4,4\}$ e $\{6,3\}$, para um mesmo valor de ρ_l.

densidade de empacotamento Δ. Portanto, para poder comparar os desempenhos das constelações de sinais de maneira correta, iremos calcular as probabilidades de erro das regiões fundamentais das tesselações $\{p,q\}$ considerando que as constelações de sinais tenham aproximadamente a mesma quantidade de sinais e estão sujetas a ação de um ruído com energia média $\mathcal{N}$.

Para isso, se faz necessário definir a energia média normalizada de uma constelação de sinais proveniente de uma tesselação $\{p,q\}$. Para tanto, inicialmente, considere que a densidade de empacotamento Δ de um reticulado no $\mathbb{R}^n$ seja dada por

$$\Delta = \lim_{r\to\infty} \left(\frac{\rho}{r}\right)^n \sum_{m\leq r^2} M_m \,, \tag{4.31}$$

onde $\rho = \rho_l$ é o raio de empacotamento e M_m é o número de sinais com energia igual a m, veja [16].

Além disso, se considerarmos constelações de sinais com cardinalidades altas, então podemos aproximar (4.31) por

$$\Delta \simeq \left(\frac{\rho}{r}\right)^2 \sum_{m\leq r^2} M_m \quad \Rightarrow \quad r^2 = \frac{\rho^2}{\Delta} \sum_{m\leq r^2} M_m \,. \tag{4.32}$$

Neste caso, note que, para uma esfera de raio r, o valor de $\sum_{m\leq r^2} M_m$ corresponde ao número total de sinais no disco de raio r. Conseqüentemente, a energia total da constelação de sinais é dada por $\sum_{m\leq r^2} mM_m$. Portanto, sua energia média é dada por

$$E = \frac{\sum_{m\leq r^2} mM_m}{\sum_{m\leq r^2} M_m} \leq r^2 \frac{\sum_{m\leq r^2} M_m}{\sum_{m\leq r^2} M_m} = r^2 \,, \tag{4.33}$$

onde utilizamos r^2 como o limitante máximo para cada m.

Portanto, se considerarmos duas constelações de sinais provenientes de tesselações com densidades Δ_1 e Δ_2, raios de empacotamentos ρ_1 e ρ_2 e números de sinais aproximadamente iguais,

$$\sum_{m\leq r_1^2} M_m^1 \simeq \sum_{m\leq r_2^2} M_m^2 \,, \tag{4.34}$$

podemos usar (4.32), (4.33) e (4.34) para determinar a seguinte relação

$$\frac{E_1}{E_2} \leq \frac{\Delta_2}{\Delta_1} \frac{\rho_1^2}{\rho_2^2}. \tag{4.35}$$

A inequação (4.35) indica o máximo ganho de energia que uma constelação de sinais pode alcançar usando uma tesselação com densidade Δ_1 ao invés da tesselação com densidade Δ_2. Esta inequação também é importante no sentido de que podemos definir uma energia média normalizada de uma tesselação em relação a uma outra, um conceito que se mostrou muito útil na análise de desempenho das constelações de sinais consideradas nesta subseção.

A Figura 4.13 apresenta um exemplo da aplicação da inequação (4.35) para as tesselações Euclidianas com energia normalizada em relação à tesselação $\{4,4\}$ com $\rho_1 = \rho_2 = d_{min}/2$, $\Delta_1 = 0.7854$ e $\mathcal{N}_1 = 1$. Neste caso, consideramos $\mathcal{N}_2 = \Delta_1 \mathcal{N}_1 / \Delta_2$.

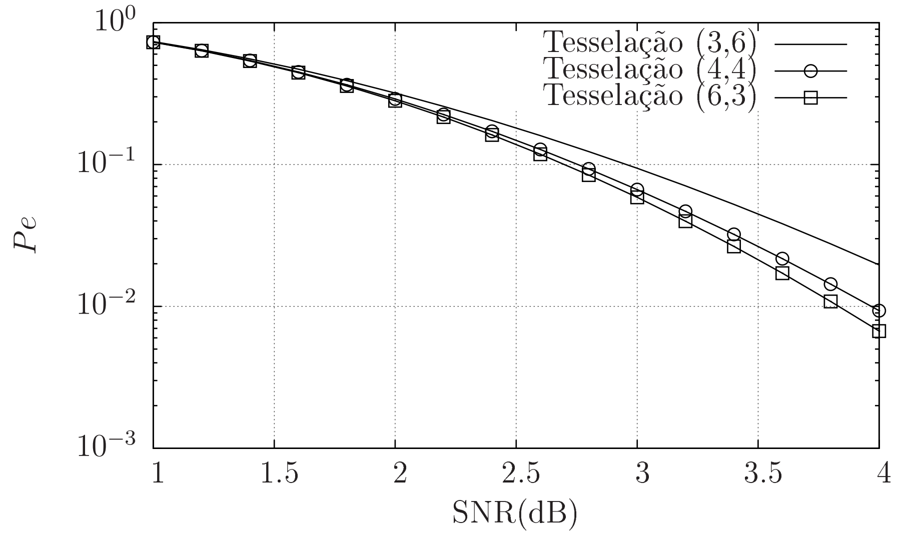


Figura 4.13: Probabilidade de erro *versus* distância mínima das tesselações $\{3,6\}$, $\{4,4\}$ e $\{6,3\}$, para uma energia normalizada em relação à tesselação $\{4,4\}$.

Analisando a Figura 4.13 podemos constatar que a tesselação mais densa atingiu o melhor desempenho em termos da probabilidade de erro. Este resultado é consistente e motiva a aplicação da inequação (4.35) na análise de desempenho das demais constelações de sinais provenientes das tesselações $\{p,q\}$.

Na Tabela 4.4 constata-se novamente que as tesselações mais densas e pertencentes aos espaços de curvatura negativa apresentam o melhor desempenho em termos da probabilidade de erro. Este resultado reproduz os casos conhecidos em $\mathbb{R}^2$ e vai além deles no sentido de confirmar que a curvatura do espaço desempenha um papel importante no projeto e na análise de desempenho das constelações de sinais e, conseqüentemente, dos sistemas de comunicações digitais.

$10^2 \times Pe$	$q = 3$	$q = 4$	$q = 5$	$q = 6$	$q = 7$	$q = 8$	$q = 9$
$p = 3$	0.7318	1.0872	1.5170	1.9618	2.3947	2.8036	3.1841
$p = 4$	0.6909	0.9334	1.2193	1.5066	1.7791	2.0312	2.2617
$p = 5$	0.6764	0.8738	1.1047	1.3342	1.5497	1.7474	1.9269
$p = 6$	0.6702	0.8442	1.0468	1.2469	1.4340	1.6048	1.7594
$p = 7$	0.6673	0.8275	1.0131	1.1958	1.3660	1.5210	1.6610
$p = 8$	0.6661	0.8174	0.9918	1.1630	1.3222	1.4670	1.5974
$p = 9$	0.6657	0.8109	0.9776	1.1408	1.2923	1.4299	1.5538

Tabela 4.4: Valores da Probabilidade de erro média, Pe, para uma tesselação $\{p, q\}$ com energia normalizada em relação à tesselação $\{4, 4\}$, $\rho_1 = 2$ e $\mathcal{N}_1 = 1$.

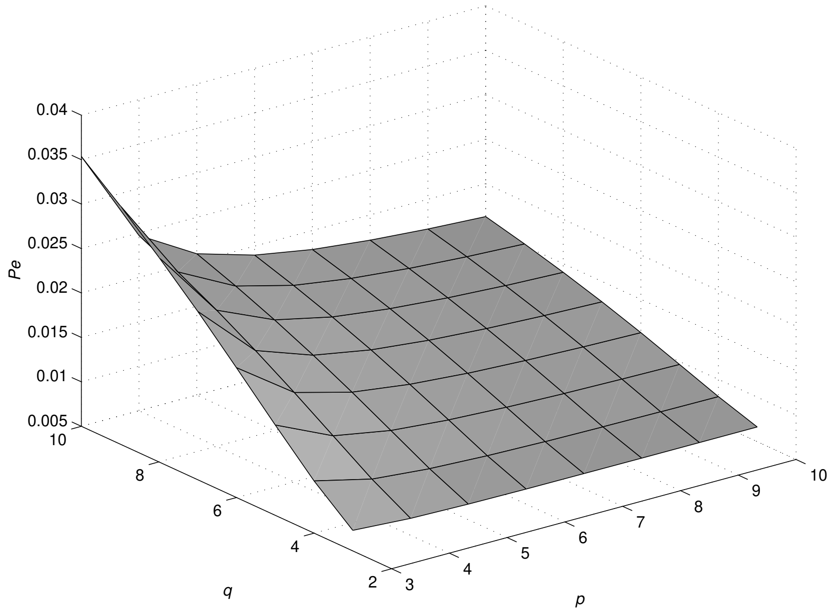


Figura 4.14: Probabilidade de erro média, Pe, para uma tesselação $\{p, q\}$, veja Tabela 4.4.

Na Figura 4.15, calculamos a variação da probabilidade de erro média, em função de d_{min}, das constelações de sinais provenientes das tesselações $\{5, 3\}$, $\{4, 4\}$ e $\{9, 3\}$ com energia média normalizada pela tesselação $\{4, 4\}$, $\mathcal{N}_1 = 1$ e $\rho_1 = 1$. Como era de se esperar, a tesselação $\{9, 3\}$ apresenta a menor probabilidade de erro entre as três tesselações consideradas, pois a mesma ladrilha o espaço hiperbólico.

Nesta seção, mostramos que, quanto menor a curvatura seccional do espaço, menor também é a probabilidade de erro de uma constelação M-PSK. Contudo, nas constelações de sinais provenientes de tesselações $\{p, q\}$, pode-se verificar que a probabilidade de erro não depende, por exemplo, da curvatura valer $K = -1$ ou $K = -10$, mas sim do fato da curvatura ser negativa. Isto se deve ao fato de que, se aumentarmos a

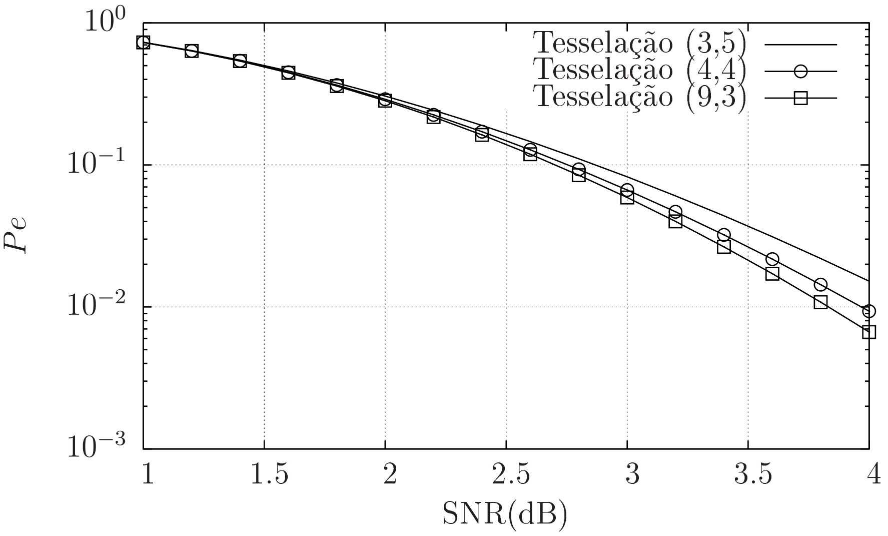


Figura 4.15: Probabilidade de erro *versus* distância mínima das tesselações $\{5, 3\}$, $\{4, 4\}$ e $\{9, 3\}$, para uma energia normalizada em relação à tesselação $\{4, 4\}$ e $\rho_1 = d_{min}/2$.

distância por uma constante c, a curvatura será multiplicada por $1/c$.

4.5 Constelações de Sinais M-PSK em Superfícies Mínimas

Neste seção, apresentaremos uma proposta de construção e de análise de desempenho de constelações de sinais M-PSK usando modulações não-lineares associadas a superfícies mínimas. Em [30], essas constelações de sinais foram estudas para a família de superfícies mínimas Enneper. Nesta seção, consideraremos constelações de sinais M-PSK também sobre a superfície mínima catenóide. Além disso, verificamos que o desempenho dos sistemas usando constelações de sinais sobre superfícies mínimas é melhor do que o desempenho dos sistemas de comunicações tradicionais utilizando a modulação M-PSK. Tal fato está novamente associado ao valor da curvatura do espaço de sinais desse sistema, que, no caso, é negativa.

Uma breve descrição sobre as superfícies mínimas foi realizada no capítulo de introdução deste trabalho e foram associadas a técnicas de modulação não-linear com bons desempenhos. Neste caso, um método de obter exemplos de superfícies mínimas, denominado representação de Weierstrass, foi introduzido pela expressão (1.6). A representação de Weierstrass desempenha um papel essencial na investigação teórica das superfícies mínimas. No caso particular do projeto de constelações de sinais, tal representação permite o uso do princípio de Reflexão de Schwarz, que, aplicado à teoria das superfícies mínimas, afirma que, se uma superfície mínima contém uma reta (respectivamente, uma geodésica plana), então a rotação de π radianos em torno dessa reta (respectivamente, a reflexão em relação ao plano da geodésica) é uma simetria da superfície.

Neste estudo, verificamos que as simetrias de uma superfície mínima são de grande importância na construção e na análise de desempenho das constelações de sinais, pois podemos utilizar tais simetrias para garantir a existência de estruturas algébricas associadas a certas classes de constelações de sinais e também facilitar os cálculos da probabilidade média de erro e da energia média dessas constelações.

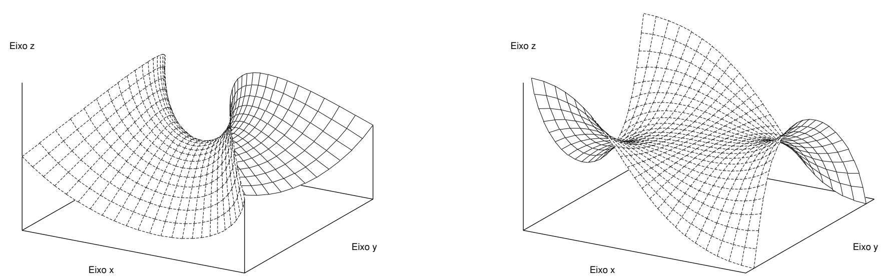
4.5.1 Análise de desempenho na família de superfícies mínimas de Enneper

A família de superfícies mínimas Enneper é obtida da representação de Weierstrass, (1.6), com $f(z) = 1$ e $g(z) = z^n = (u + iv)^n$, onde n é um número real não negativo. Chamamos a atenção ao fato de que, quando $n = 0$, a superfície mínima gerada é o plano Euclidiano e, conseqüentemente, recaímos no modelo tradicional de análise de um sistema de comunicações.

A Figura 4.16 ilustra dois exemplos da família de superfícies mínimas associada à superfície de Enneper. Para $n = 1$, a superfície mínima gerada é a de Enneper (Figura 4.16(a)), cuja parametrização é dada por

$$X(u, v) = \frac{1}{2}\left[u - \frac{u^3}{3} + uv^2, -v + \frac{v^3}{3} - u^2v, u^2 - v^2\right],$$

para $(u, v) \in \mathbb{R}^2$.



(a) Superfície de Enneper, $n = 1$ (b) Superfície de Enneper, $n = 3$

Figura 4.16: Família de superfícies mínimas Enneper.

A escolha desta família de superfícies para o projeto de constelações M-PSK é devido, principalmente, ao fato de que todas as geodésicas que partem do ponto $p = (0, 0)$ são geodésicas planas e, mais do que isso, usando o princípio de Reflexão de Schwarz, temos que as rotações em torno de p constituem o subgrupo das simetrias das geodésicas, conseqüentemente formando o correspondente grupo de isometrias. Como os elementos desse grupo podem ser associados a um conjunto de regiões, temos que, se a cada uma dessas regiões escolhermos um ponto como o representante da “classe de equivalência”, e mais, que esses pontos tenham o mesmo valor de comprimento de arco do ponto $p = (0, 0)$, então o conjunto desses pontos congruentes forma uma constelação de sinais geometricamente uniforme do tipo M-PSK. Chamamos a atenção ao fato de que a propriedade de congruência de regiões não é típica em superfícies mínimas, visto que as mesmas não possuem curvatura constante.

De (1.2), pode-se obter que os coeficientes da primeira forma fundamental dessa família de superfícies mínimas são dados por

$$E = G = \lambda^2 = \frac{1}{4}\left[1 + (u^2 + v^2)^n\right]^2 \quad \text{e} \quad F = 0\,. \tag{4.36}$$

A curvatura Gaussiana da família de superfícies mínimas Enneper, decorrente da substituição de (4.36) em (1.8), é dada por

$$K(u, v) = -16\frac{n^2(u^2 + v^2)^{n-1}}{[1 + (u^2 + v^2)^n]^4}\,. \tag{4.37}$$

onde explicitamos a dependência de $K(u, v)$ com n.

Note que, substituindo $n = 0$ em (4.37), a curvatura Gaussiana é igual a zero para todos os pontos $(u, v) \in \mathbb{R}^2$: isto é consequência do fato já mencionado de que a superfície mínima gerada é o plano Euclidiano. Observe que, para cada valor de n em $K(u, v)$, obtemos uma superfície de revolução, como ilustrado pelas geratrizes para $n = 0, 1$ e 3 na Figura 4.17.

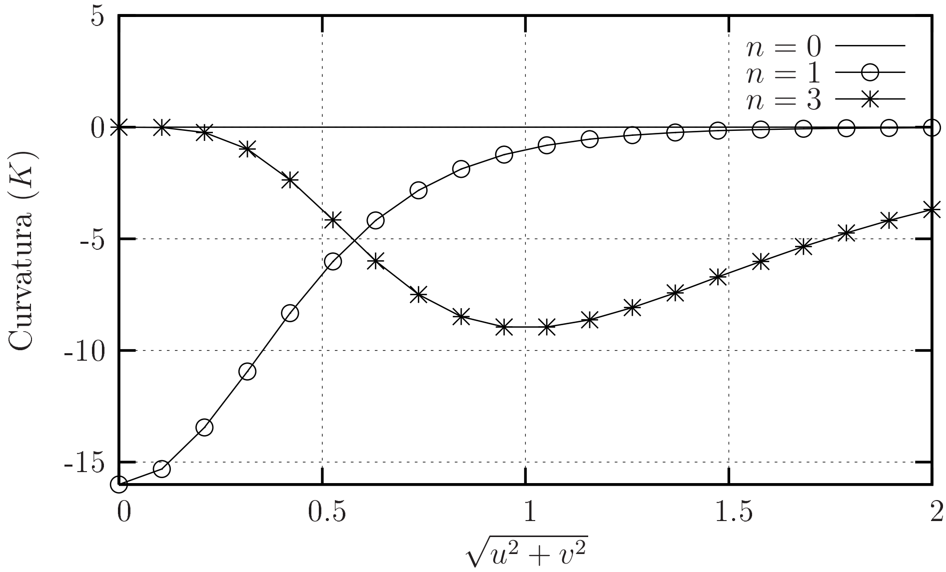


Figura 4.17: Geratrizes da curvatura Gaussiana da família Enneper para $n = 0, 1, 3$.

Da equação (1.27), as geodésicas da família de superfícies mínimas Enneper são as soluções do seguinte sistema de equações diferenciais parciais

$$\begin{aligned} u'' + \tfrac{2n(u^2+v^2)^{n-1}}{1+(u^2+v^2)^n}\left\{u\left[(v')^2-(u')^2\right]+2u'v'v\right\} = 0,\\ v'' + \tfrac{2n(u^2+v^2)^{n-1}}{1+(u^2+v^2)^n}\left\{v\left[(u')^2-(v')^2\right]+2u'v'u\right\} = 0. \end{aligned} \tag{4.38}$$

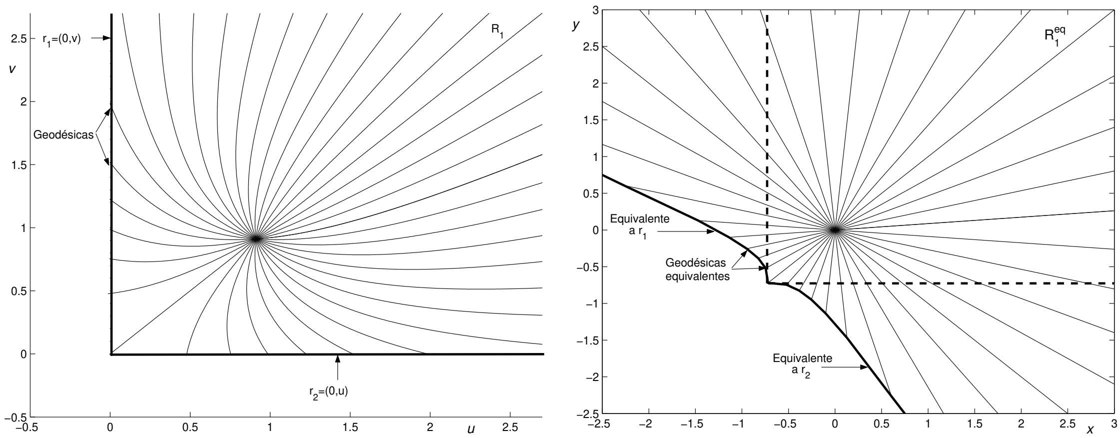
Uma constelação de sinais M-PSK nas superfícies da família de Enneper consiste do conjunto de pontos dados por

$$\mathbf{x}_i = \rho e^{j\theta_i}\,, \quad i = 1, 2, \ldots, M\,,$$

onde $\theta_i = (2i-1)\pi/M$ e ρ é a menor distância geodésica entre um ponto $(u, v) \in \mathbb{R}^2$ e a origem (0,0), que pode ser obtida de (1.26) e dada por

$$d(u,v) = \frac{\sqrt{u^2+v^2}}{2}\left[1 + \frac{(u^2+v^2)^n}{2n+1}\right]. \tag{4.39}$$

Considere, por exemplo, uma constelação 4-PSK com energia média unitária no espaço de sinais, M^2, com métrica induzida da superfície mínima de Enneper. Neste caso, de (4.39), temos que $|u| = |v| = 0.91069$ e a Figura 4.18(a) ilustra a região de decisão do sinal $\mathbf{x}_1 \in \mathcal{X}$, em M^2, como sendo a área limitada pelas retas $r_1 = (u, 0)$ e $r_2 = (0, v)$, para $u > 0$ e $v > 0$.



(a) No espaço de sinais. (b) No plano tangente $T_{\mathbf{x}_1}M$.

Figura 4.18: Região de decisão do sinal $\mathbf{x}_1$ da constelação 4-PSK para $n = 1$.

Devido à possibilidade de interpretar o esquema de modulação como pontos sobre uma superfície, de modo que localmente o ruído possa ser projetado em seu plano tangente, então podemos encontrar uma região de decisão equivalente ao sinal $\mathbf{x}_1$, R_1^{eq}, no plano tangente $T_{\mathbf{x}_1}X$, Figura 4.18(b), como sendo a área limitada pelas curvas equivalentes em $T_{\mathbf{x}_1}X$ às retas r_1 e r_2 em M^2.

Esse conceito de região de decisão equivalente é um artifício muito importante, principalmente quando a variedade M^2 não for de curvatura Gaussiana constante, pois o cálculo da probabilidade de erro de símbolo, $P_{e,m}$, em M^2 pode se tornar um tanto quanto difícil. Sabemos que a única superfície mínima com curvatura Gaussiana constante é o plano, $K_n(u,v) = 0$, e, portanto, a análise de desempenho das constelações de sinais em superfícies mínimas, exceto o plano, apresenta grande complexidade.

A Figura 4.19 ilustra as curvas da probabilidade de erro *versus* a relação sinal ruído de uma constelação 4-PSK nos espaços de sinais com métrica induzida da família de Enneper, para $n = 0, 1$ e 3. Observe que as constelações de sinais definidas para $n = 1$ e 3 possuem desempenhos superiores aos da constelação 4-PSK no plano Euclidiano ($n = 0$). Esse fato pode ser melhor entendido quando observamos a Figura 4.18(b), pois a região de decisão equivalente, R_1^{eq}, contém a região de um mesmo sinal 4-PSK quando a superfície é o plano. Além disso, as curvaturas Gaussianas das superfícies mínimas, exceto o plano, são sempre negativas, e este fato também justifica o ganho de desempenho dessas constelações.

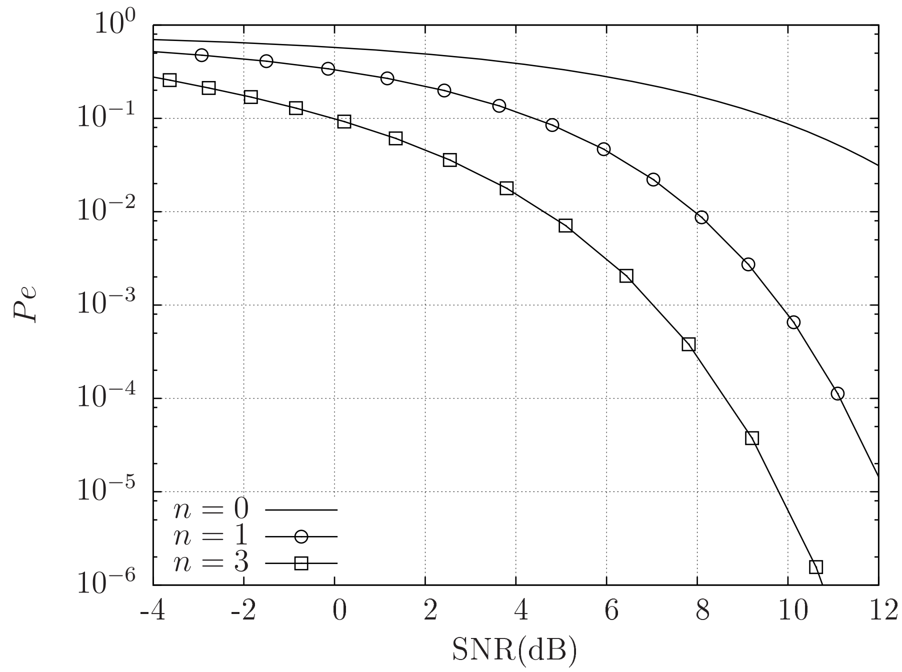
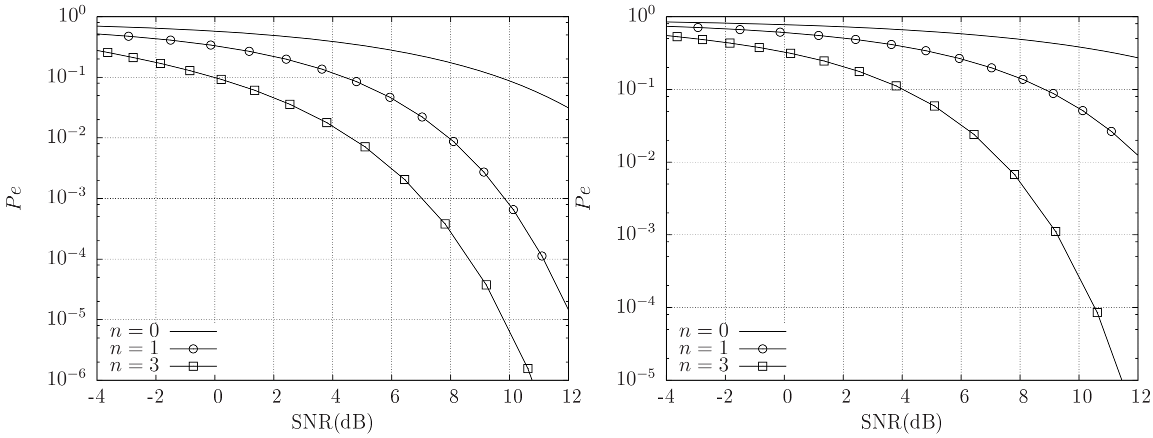


Figura 4.19: $P_e \times SNR$ de uma constelação de sinais 4-PSK na família de superfícies mínimas Enneper, para $n = 0, 1$ e 3.

Da mesma forma que a constelação 4-PSK, as constelações 8-PSK e 16-PSK apresentam desempenhos superiores nos espaços bidimensionais para $n = 1, 3$, em comparação com as constelações bidimensionais 8-PSK e 16-PSK para $n = 0$. As Figuras 4.20(a) e 4.20(b) ilustram tal afirmação.



(a) 8-PSK. (b) 16-PSK.

Figura 4.20: $P_e \times SNR$ de uma constelação de sinais 8-PSK e 16-PSK na família de superfícies mínimas Enneper, para $n = 0, 1$ e 3.

4.5.2 Análise de desempenho na superfície mínima catenóide

Nesta subseção, apresentaremos um exemplo de construção e análise de desempenho da constelação de sinais 4-PSK sobre a superfície mínima catenóide, Figura 1.2(a). A catenóide pode ser obtida da representação de Weierstrass, (1.6), com $f(z) = -e^{-z}$ e $g(z) = -e^{z}$. Neste caso, sua parametrização é dada por

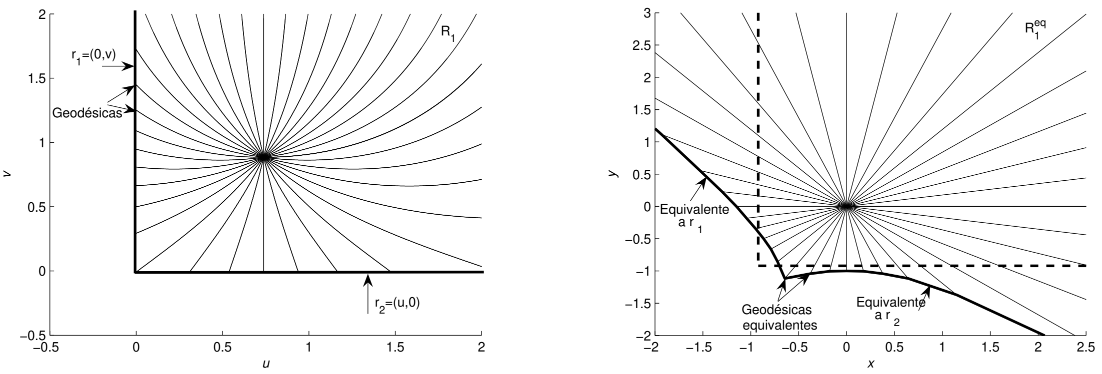
$$X(u, v) = [\cosh(u)\cos(v), \cosh(u)\sin(v), u]\ .$$

Da equação (1.2), seus coeficientes da primeira forma fundamental são dados por $E = G = \lambda^2 = \cosh^2(v)$ e F=0, e, da equação (1.8), sua curvatura Gaussiana vale $K = -\cosh(v)^{-4}$. Além disso, as geodésicas da catenóide devem satisfazer, de (1.27), o seguinte sistema de equações diferenciais

$$\begin{aligned} u'' + 2u'v'\tanh(v) &= 0 \\ v'' + \left[(v')^2 - (u')^2\right] &= 0 \end{aligned} \qquad (4.40)$$

Neste exemplo, a constelação 4-PSK possui sinais dados por $|u| = 0.7383$ e $|v| = 0.8843$, com distância mínima entre os sinais igual a 2 e energia média de transmissão igual a 1.7 no espaço de sinais M^2 com métrica induzida da superfície mínima catenóide. Dessa forma, a Figura 4.21(a) ilustra a região de decisão do sinal $\mathbf{x}_1 \in \mathcal{X}$, em M^2, como sendo a área limitada pelas retas $r_1 = (u, 0)$ e $r_2 = (0, v)$, para $u > 0$ e $v > 0$.

Usando o conceito de região de decisão equivalente ao sinal x_1, R_1^{eq}, no plano tangente $T_{x_1}X$, Figura 4.21(b), como sendo a área limitada pelas curvas equivalentes em $T_{\mathbf{x}_1}M$ às retas r_1 e r_2 em M^2, então podemos calcular sua curva de desempenho, Figura 4.22. Neste caso, verifica-se que o desempenho desse sistema é melhor do que o tradicional (no espaço Euclidiano), porém é inferior ao desempenho da mesma constelação sobre a superfície de Enneper.



(a) No espaço de sinais. (b) No plano tangente $T_{\mathbf{x}_1}X$.

Figura 4.21: Região de decisão do sinal $\mathbf{x}_1$ da constelação 4-PSK para a catenóide.

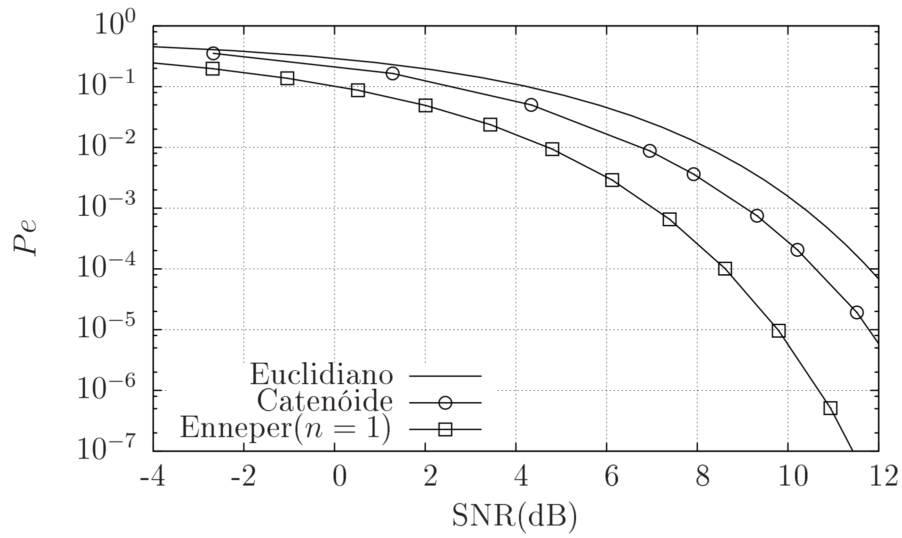


Figura 4.22: $P_e \times SNR$ de uma constelação 4-PSK no espaço Euclidiano, na catenóide e na superfície de Enneper.

Capítulo 5

Conclusões

O estudo de um sistema de comunicações como o apresentado na Figura 1.1, no qual cada um de seus blocos pode ser representado por um espaço métrico, propiciou a análise desse sistema em um contexto mais geral do que é habitualmente tratado. Neste caso, foi possível fazer uso da geometria Riemanniana para caracterizar cada um desses espaços métricos do sistema de comunicações como uma variedade Riemanniana. Dessa forma, alguns parâmetros geométricos como a métrica e a curvatura da variedade puderam ser considerados na análise de desempenho desses sistemas. Com isso, verificamos que a curvatura do espaço influencia no desempenho dos sistemas de comunicações considerados neste trabalho. Por essa razão, quando o objetivo a ser alcançado na proposta de novos sistemas de comunicações é o melhor desempenho, então o estudo da influência da curvatura do espaço em sistemas de comunicações deve ser considerado. Entretanto, deve-se chamar a atenção para o fato de que este trabalho apresenta apenas uma introdução a esse estudo e que foram considerados somente alguns casos particulares de sistemas de comunicações.

Além disso, verificamos que a curvatura do espaço não influencia apenas no desempenho dos sistemas de comunicações, mas também em sua complexidade. Pois, quando o espaço possui curvatura constante, é possível utilizar as várias isometrias e simetrias desses espaços para poder simplificar a construção e análise dos sistemas de comunicações, como ocorreu, por exemplo, no projeto dos guias de ondas, das lentes ópticas e das constelações de sinais geometricamente uniformes. Tal fato é extremamente desejável e implica que os espaços de curvatura constante devam ser, inicialmente, privilegiados na proposta de novos sistemas de comunicações.

Um outro ponto que deve ser ressaltado, é que quando a curvatura do espaço for negativa, geralmente, o desempenho dos sistemas de comunicações considerados neste trabalho foram melhores. Portanto, podemos concluir que na maioria das vezes em que os espaços métricos associados aos blocos do sistema de comunicações da Figura 1.1 utilizam métrica derivadas de espaços hiperbólicos, então o desempenho dos sistemas de comunicações tendem a se aproximar do ponto ótimo de operação.

Neste trabalho identificamos que a curvatura não é um parâmetro como a dispersão, ou o erro quadrático médio ou a probabilidade média de erro, que são utilizados diretamente no cálculo de desempenho dos sistemas comunicações. Por essa razão é que a influência da curvatura no desempenho de sistemas de comunicações, geralmente, não é percebida de maneira imediata e sua desconsideração na análise desempenho é muitas vezes realizada. Contudo, como demonstram, por exemplo, as expressões (2.25), (3.76), (3.105) e (4.13) a curvatura do espaço influencia diretamente nos valores dos parâmetros de desempenho, portanto, no desempenho final do sistema. Este fato sugere algo muito interessante, que a curvatura do espaço ainda é muito pouco explorada nos sistemas de comunicações e seu estudo, neste contexto, ainda pode ser considerado um problema em aberto.

No Capítulo 2, a influência da curvatura foi, inicialmente, percebida no guia de ondas planar com perfil secante hiperbólica (dispersão nula) e no instrumento óptico absoluto "*fish-eye*" (imagem perfeita), ambos com curvatura constante positiva. A partir dessa observação, propusemos nosvos instrumentos ópticos com base na análise de suas curvaturas, demonstrando assim a importância do estudo da curvatura em meios ópticos.

No Capítulo 3, os valores da curvatura da curva, da curvatura Gaussiana e da curvatura média foram utilizados convenientemente para determinar o valor máximo da energia média de ruído de maneira que o erro quadrático médio fosse pouco influenciado pelo erro de ponto inicial e o ganho de desempenho da modulação não-linear fosse preservado. Além disso, a curvatura média de uma superfície associada a uma modulação foi usada para caracterizar as classes de modulações com menor erro quadrático médio, no caso as superfícies mínimas.

No Capítulo 4, verificamos que o espaço de sinais utilizados no projeto de constelações de sinais possui métrica induzida da superfície associada a modulação não-linear. Com isso, constatamos que a curvatura do espaço de sinais influencia no desempenho dos sistemas de comunicações, mais precisamente na probabilidade média de erro. Além disso, concluímos que os espaços de sinais com curvatura negativa são os que apresentam melhor desempenho. Por isso, propusemos e analisamos algumas constelações de sinais M-PSK e geometricamente uniformes em tais espaços, como os espaços de curvatura constante e outros derivados das superfícies mínimas.

5.1 Sugestões de Trabalhos Futuros

Nesta seção apresentaremos alguns caminhos para uma possível continuidade e complementação deste trabalho.

- Busca por novos exemplos de sistemas de comunicações, cujos desempenhos possam ser relacionados aos valores da curvatura dos espaços associados aos sistemas.

- Caracterização das transformações que interligam os blocos de um sistema de comunicações, de maneira que estes blocos sejam "casados" e o máximo desempenho possa ser alcançado.

- Determinar a influência da curvatura no desempenho de fibras ópticas monomodo.

- Expressar a taxa de variação da energia da frente de onda em meios ópticos em função da curvatura do meio. Tal taxa pode propiciar a construção de novos instrumentos ópticos.

- Definir a modulação *twisted* não apenas associada a imersões no $\mathbb{R}^n$, mas também em variedades Riemannianas. Tais modulações seriam equivalentes a uma modulação de uma modulação.

- Realizar o cálculo de variações n-dimensional para verificar se uma imersão mínima é uma ponto crítico de mínimo do erro quadrático médio.

- A construção de novos esquemas de modulações não-lineares associadas às superfícies mínimas, com o auxílio do princípio da reflexão de Schwarz.

- A construção de uma modulação associada à hiperesfera, seja com $M = 1$ ou $M > 1$.

- Construção e análise de desempenho de constelações de sinais em variedades Riemannianas de dimensão maior que dois.

Bibliografia

[1] A.L. Garcia, *Probability and Random Processes for Electrical Enginnering*, Addison Wesley Longman, 2nd ed., 1994.

[2] A.W. Snyder and J. Love, *Optical Waveguide Theory*, Chapman & Hall, London, 1983.

[3] B.P. Lathi, *Modern Digital and Analog Communication Systems*, New York: Oxford University Press, 3rd Edition, 1998.

[4] C.M. Thomas, C. L. May and G. R. Welti, "Hybride amplitude-and-phase modulation for analog data transmission," *IEEE Transactions on Communications*, vol. COM-23, no. 6, pp. 634-645. June 1975.

[5] D.J. Sakrison, *Communication Theory: Transmission of Waveforms and Digital Information.* New York: John Wiley & Sons, Inc, 1968.

[6] E. Agustini, *Constelações de Sinais em Espaços Hiperbólicos*, Tese de Doutorado, IMECC-UNICAMP, Campinas, Brasil, 2002.

[7] E.B. da Silva, *Constelação de Sinais e Análise de Desempenho no Plano Hiperbólico*, Tese de Doutorado, FEEC-UNICAMP, Campinas, Brasil, 2000.

[8] E.D. de Carvalho, *Construção e Rotulamento de Constelações de Sinais Geometricamente Uniformes em Espaços Euclidianos e Hiberbolicos*, Tese de Doutorado, FEEC-UNICAMP, Campinas, Brasil, 2002.

[9] G.D. Forney, Jr., "Geometrically Uniform Codes," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1241-1260, Sept. 1991.

[10] G.O. Santos, *Caracterização Geométrica do Processo de Decodificação da Classe dos Códigos Alternantes Cíclicos através de Polinômios Absolutamente Irredutíveis*, Tese de Doutorado, FEEC-UNICAMP, Campinas, Brasil, 2003.

[11] H. Guo and X. Deng, "Differential geometrical methods in the study of optical transmission (scalar theory). I. Static transmission case," *J. Opt. Soc. Am.*, A **12**, 600-606 (1995).

[12] H. Lazari, *Uma Contribuição a Teoria dos Códigos Geometricamente Uniformes Hiperbólicos*, Tese de Doutorado, FEEC-UNICAMP, Campinas, Brasil, 2000.

[13] J.D. Lima, *Identificação e Estrutura Algébrica das Superfícies Compactas Com e Sem Bordos, Provenientes de Mergulhos de Canais Discretos Sem Memorias*, Tese de Doutorado, FEEC-UNICAMP, Campinas, Brasil, 2002.

[14] J.D. Lima, e R. Palazzo Jr., "Embedding discrete memoryless channels on compact and minimal surfaces," *IEEE Information Theory Workshop*, Bangalore, India, Oct. 20-25, 2002.

[15] J.G. Proakis, *Digital Communications*, 2nd edition, McGraw-Hill, 1989.

[16] J.H. Conway, N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, 2nd edition, Springer-Verlag, 1998.

[17] J. Lucas M. Barbosa, A. Gervásio Colares, *Minimal Surfaces in* R^3, Lecture Notes in Mathematics, No. 1195, Springer-Verlag, Berlin, 1984.

[18] J.M. Wozencraft and I. M. Jacobs, *Principles of Comunication Engineering*, New York: John Wiley & Sons, Inc, 1965.

[19] K.H. Lee and D. P. Petersen, "Optimal linear coding for vector channels,"*IEEE Transactions on Communications*, vol. COM-24, n0. 12, pp. 1283-1290, December 1976.

[20] M. Abramowitz and I. Stegun, *Elliptic Integrals*, Chapter 17, Dover Publications Inc., New York, 1046 p., (1965).

[21] M. Born and E. Wolf, *Principles of Optics*, 7th expaned ed., Cambridge, 2003.

[22] M.B. Faria, *Empacotamento de Esferas em Espaços Hiperbólicos*, Dissertação de Mestrado, IMECC-UNICAMP, Campinas, Brasil, 2001.

[23] M.J. Souza, *Realizações de Constelações de Sinais Hiperbólicas Densas Associadas a Sistemas Lineares através das Funções Automorfas*, Tese de Doutorado, FEEC-UNICAMP, Campinas, Brasil, 2005.

[24] M. Kline and I. W. Kay, *Eletromagnetic Theory and Geometrical Optics*, New York, Interscience Publishers, 1965.

[25] M.P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall Inc., New Jersey, 1976.

[26] M.P. do Carmo, *Geometria Riemanniana*, Impa-Projeto Euclides, 1979.

[27] M.P. do Carmo, *Riemannian Geometry* 1st ed. (Birkhäuser, Boston 1992).

[28] P.A. Floor, and T. A. Ramstad, "Noise analysis for dimension expanding mappings in souce-channel conding," *IEEE - Seventh Workshop on Signal Processing Advances in Wireless Communications*, Cannes, France, July, 2-5, 2006.

[29] R.G. Cavalcante, *Análise de Desempenho de Constelações de Sinais em Variedades Riemannianas*, Dissertação de Mestrado, FEEC-UNICAMP, Campinas, Brasil, 2002.

[30] R.G. Cavalcante, e R. Palazzo Jr., "Performance analysis of M-PSK signal constellations in Riemannian varieties," *Lecture Notes in Computer Science*, Springer-Verlag, 2003.

[31] R.G. Cavalcante, e R. Palazzo Jr. "Construção e análise de desempenho de constelações de sinais M-PSK em superfícies mínimas," *XX Simpósio Brasileiro de Telecomunicações-SBT'03*, Rio de Janeiro, 2003.

[32] R.G. Cavalcante, e R. Palazzo Jr., "Análise de desempenho de constelações de sinais geometricamente uniformes provenientes de reticulados $\{p, q\}$ em espaços bidimensionais com curvatura constante," *XXI Simpósio Brasileiro de Telecomunicações-SBrT'04*, Belém, 2004.

[33] R.G. Cavalcante, H. Lazari, J.D. Lima, and R. Palazzo Jr., "A new approach to the design of digital communication systems," *AMS-DIMACS Series*, pp.1-32, August 2005.

[34] R.G. Cavalcante, e R. Palazzo Jr., "Construção de códigos de bloco lineares sobre $\mathbb{F}_q$ com d_{min} máxima usando programação linear inteira," *XXII Simpósio Brasileiro de Telecomunicações-SBrT'05*, Campinas, 2005.

[35] R.G. Cavalcante, e R. Palazzo Jr., "Um estudo sobre a curvatura de guias de ondas ópticos," *5th International Information and Telecommunication Technologies Synposium-I2TS'2006*, Cuiabá, Brazil, 2006.

[36] R. Ilinsky, "Gradient-index meniscus lens free of spherical aberration", *J. Opt. Soc. Am.*, A **2**, 449-451 (2000).

[37] S. Haykin, *Sistemas de Comunicações: Analógicos e Digitais*, Porto Alegre: Bookman, 4.ed., 2004.

[38] V.L. Vieira, *Grupos Fuchsianos Aritméticos Identificados em Ordens dos Quatérnios para Construção de Constelações de Sinais*, Tese de Doutorado, FEEC-UNICAMP, Campinas, Brasil, 2007.

[39] V.A. Kotel'nikov, *The Theory of Optimum Noise Immunity.* New York: McGraw-Hill Book Company, Inc, 1959.

[40] W. Klingenberg, *A Course in Differential Geometry*, Springer, New York, 1978.

[41] W. Shen, Z. Jufang, S. Zhu, and X. Deng "Fermat's principle, the general Eikonal equation, and space geometry in a static anisotropic medium," *J. Opt. Soc. Am.*, A **14**, 2850-2854 (1997).