UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA

LUCAS DA SILVA

Modelagem geométrica do par roda-trilho com descrição por meio de NURBS

São Paulo

2021

LUCAS DA SILVA

Modelagem geométrica do par roda-trilho com descrição por meio de NURBS

Versão Corrigida

Dissertação apresentada à Escola Politécnica
da Universidade de São Paulo para obtenção
do título de Mestre em Ciências

Área de concentração: Engenharia de
Estruturas

Orientador: Prof. Dr. Alfredo Gay Neto

São Paulo
2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 23 de agosto de 2021

Assinatura do autor

Assinatura do orientador

Catálogo-na-publicação

da Silva, Lucas

Modelagem geométrica do par roda-trilho com descrição por meio de NURBS / L. da Silva -- São Paulo, 2021.
167 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Estruturas e Geotécnica.

Versão corrigida

1.Geometria e modelagem computacional 2.Reconstrução (geometria e modelagem computacional) 3.Curvas e superfícies (geometria e modelagem computacional) 4.Algoritmos geométricos 5.Otimização não linear
I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Estruturas e Geotécnica II.t.

Nome: DA SILVA, Lucas

Título: Modelagem geométrica do par roda-trilho com descrição por meio de NURBS

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Aprovado em:

Banca Examinadora

Prof. Dr.	_____
Instituição:	_____
Julgamento:	_____
Prof. Dr.	_____
Instituição:	_____
Julgamento:	_____
Prof. Dr.	_____
Instituição:	_____
Julgamento:	_____

Dedicatória

À minha querida mãe, Elma Bispo da Silva, a quem devo imensa gratidão por me guiar para e por esta vida

Agradecimentos

Ao Prof. Dr. Alfredo Gay Neto, meu orientador, que apresenta dedicação ímpar para com todos seus orientandos, inclusive para comigo. Foi um grande parceiro nesta jornada. Ajudou-me a manter a disciplina e o interesse no assunto através de questionamentos e conselhos. O melhor orientador que eu poderia querer.

Aos amigos do Laboratório de Mecânica Computacional (LMC) da Escola Politécnica da USP, por manter um clima agradável no espaço de trabalho com os quais tive inúmeras discussões estimulantes às mais variadas formas do pensar. Um agradecimento especial a Guiomar Leite por sua natureza atenciosa e carinho maternal, que sem dúvida foram um suporte emocional inestimável.

Um agradecimento ao meu pai por ter sido uma inspiração. E um agradecimento simbólico a todos os professores e tutores que tive, não tenho dúvida de que há um pouquinho da influência de cada um neste resultado.

À Escola Politécnica da USP pela oportunidade de participar do Programa de Pós-Graduação em Engenharia Civil.

Agradeço à Vale S.A. pelo apoio financeiro através do projeto Cátedra Roda-Trilho.

“A matemática, vista corretamente, possui não apenas verdade, mas também suprema beleza”

Bertrand Russell [1]

Resumo

Parte fundamental da modelagem computacional do contato é a descrição matemática das superfícies dos corpos envolvidos. A proposta deste trabalho é sistematizar a reconstrução de superfícies de rodas e trilhos em ambiente computacional a partir de amostras discretas medidas em elementos ferroviários reais. Estas representações podem, então, ser utilizadas em simulações de elementos finitos de diversas situações de interesse ao contexto ferroviário. Foi desenvolvido um procedimento de ajuste por aproximação de curvas chamadas de splines de arcos, que são compostas por arcos de circunferência conectados de forma suave (continuidade do vetor tangente através dos segmentos). Dada uma sequência de amostras de uma curva plana, propriedades geométricas como comprimento de arco e curvatura são estimadas. Uma primeira curva aproximada é obtida a partir da curvatura estimada. Em seguida, é aplicado um método de otimização por mínimos quadrados que deforma a spline de arcos aprimorando a aproximação. Splines de arcos são um caso particular de curvas racionais NURBS (*Non-Uniform Rational B-splines*). Na segunda parte do trabalho, investiga-se como parametrizar as curvas planas reconstruídas usando funções racionais. Além disso, esquematiza-se como transformar as curvas planas reconstruídas em superfícies NURBS de extrusão e revolução a fim de modelar superfícies de contato de trilhos e rodas respectivamente.

Palavras-chave: Aproximação de curvas. Reconstrução geométrica. Splines de arcos. NURBS. Contato roda-trilho.

Abstract

The mathematical description of the touching bodies surfaces is a fundamental part of computational contact modelling. This work proposal is to systematize the reconstruction of wheel and rail surfaces in computational environment, given a set of discrete samples measured from existing railway elements. The aim is to employ these representations in finite elements simulations of various interesting situations to the railway context. The fitting procedure developed is based on arc-splines, a family of curves composed of circular arcs smoothly joined (the tangent vector is continuous across segments). Given a sequence of points sampled from a planar curve, we estimate geometric properties such as arc-length and curvature. An initial approximation is obtained from the estimated curvature. Then, a better fitting is found deforming the arc-spline by a least-squares method. Arc-splines are rational curves and can be parametrized as NURBS (*Non-Uniform Rational B-splines*). The second half of this work concerns the parametrization of arc-splines using rational functions. As well as the parametrization of NURBS surfaces generated by extrusion and revolution of the approximate curves for modelling the surfaces of rails and wheels.

Keywords: Curve approximation. Geometric reconstruction. Arc-splines. NURBS. Wheel-rail contact.

Lista de siglas

NURBS	<i>Non-Uniform Rational B-Splines</i>
CAD	<i>Computer-Aided Design</i>
CNC	<i>Computer Numeric Control</i>
CAM	<i>Computer-Aided Manufacture</i>
CAGD	<i>Computer-Aided Geometric Design</i>
MEF	Método dos Elementos Finitos

Sumário

Introdução	17
Objetivos.....	24
1 Teoria das curvas planas	25
1.1 Definição de curva plana.....	25
1.2 Vetor tangente	26
1.3 Comprimento de uma curva	28
1.3.1 Exemplo. Parábola	30
1.3.2 Exemplo. Espiral logarítmica	31
1.3.3 Exemplo. Arco de circunferência	32
1.4 Reparametrização por comprimento de arco.....	33
1.5 Curvatura.....	35
1.6 Curvatura orientada	37
1.6.1 Exemplo. Curvatura de uma circunferência	39
1.7 Teorema fundamental das curvas planas.....	40
2 Otimização	42
2.1 Caracterização de um mínimo.....	43
2.2 Considerações sobre algumas funções especiais.....	44
2.2.1 Funções lineares	45
2.2.2 Funções quadráticas	46
2.3 O método de Newton-Raphson	47
2.4 Aplicação ao problema de otimização	49
2.5 O método de Gauss-Newton	51
2.6 Métodos <i>trust-region</i> para otimização	53
2.6.1 O problema local de otimização	54
2.6.2 O passo de Cauchy	55
3 Aproximação de uma curva amostrada por uma spline de arcos	57
3.1 Aproximação de propriedades geométricas a partir das amostras	57

3.1.1	Comprimento de arco	58
3.1.2	Ângulo tangente de partida	58
3.1.3	Curvatura orientada	59
3.2	Arcos de circunferência, biarcos e splines de arcos	63
3.2.1	O conceito de continuidade geométrica.....	66
3.2.2	Conexão geometricamente contínua de arcos.....	67
3.2.3	Determinação do gráfico de curvatura constante por trechos.....	69
3.2.4	Parâmetros da spline de arcos.....	77
3.3	Aprimoramento do ajuste por deformação da curva.....	78
3.3.1	Resíduos e suas derivadas.....	78
3.4	Segmentação	84
4	Resultados.....	87
4.1	Biarco em formato de S	87
4.2	Espiral áurea	90
4.3	Curvas quadráticas.....	91
4.4	Curvas de ordem superior.....	95
4.5	Perfis ferroviários	98
4.5.1	Sobre os parâmetros de entrada do método de ajuste	99
4.5.2	Fluxo de trabalho	101
4.5.3	Trilho usinado	102
4.5.4	Roda ferroviária	106
4.5.5	Perfis de roda incluindo irregularidades	109
5	De curvas polinomiais a NURBS	113
5.1	Curvas polinomiais	113
5.2	Representação de Bézier para curvas polinomiais.....	114
5.2.1	Curvas quadráticas.....	115
5.2.2	Exemplo. Gráfico de uma parábola	116
5.2.3	Curvas de grau n	117
5.3	Curvas racionais.....	119
5.3.1	Curva racional como imagem projetada de uma curva polinomial	121
5.4	Parametrização racional de um arco de circunferência	123
5.4.1	Relação invariante entre os pesos	124

5.5	B-Splines	127
5.5.1	Exemplo. B-Spline quadrática com um único vão	129
5.5.2	Exemplo. B-Spline cúbica.....	131
5.6	NURBS.....	134
5.6.1	Propriedades	134
5.7	Representação NURBS de um biarco	135
5.7.1	Representação com nó interior não duplicado	136
5.7.2	Exemplo. Dois quartos de circunferência com razão 1:2 entre os raios....	141
5.7.3	Exemplo. Semicircunferência de raio unitário	142
5.8	Spline de três ou mais arcos	142
5.9	Representação NURBS de uma circunferência completa.....	144
6	Superfícies para a modelagem de contato	146
6.1	Superfície NURBS	146
6.2	Superfície de extrusão	147
6.2.1	Curvas coordenadas de fronteira	148
6.2.2	Curvas coordenadas intermediárias.....	148
6.2.3	Exemplo. Superfície tipo trilho	149
6.2.4	Efeito da rotação de uma fronteira	150
6.3	Superfície de revolução	151
6.3.1	Exemplo. Esfera	152
6.4	Parametrização NURBS de uma superfície de revolução.....	153
6.4.1	Paralelos, as curvas coordenadas circulares	154
6.4.2	Exemplo. Curva geratriz cúbica	157
6.4.3	Exemplo. Superfície tipo roda ferroviária.....	158
	Conclusão	160
	Trabalhos futuros.....	161
	Referências	162

Introdução

Uma atividade essencial associada à engenharia na criação e produção das soluções do cotidiano é o *design*, que se trata da tarefa de definir a geometria dos objetos a fabricar. Geralmente, a geometria projetada é descrita de forma exata como uma combinação de figuras simples ou formas analíticas conhecidas. Peças automotivas e aeronáuticas, por exemplo, são modeladas atualmente com o auxílio de computador através de softwares de CAD (do inglês, *Computer-Aided Design*), que trabalham em uma base de curvas e superfícies paramétricas racionais. O traçado de uma ferrovia é composto por trechos tangentes retos, curvas circulares e trechos de transição com geometria bem determinada.

Perfis geométricos padrão de roda e trilho também são compostos de segmentos de reta e arcos de circunferência [2, 3]. Por perfil de roda entende-se a curva obtida da intersecção da superfície externa da roda com um plano radial conforme indicado na Figura 1. Por perfil de trilho entende-se a parte superior da seção transversal do trilho conforme indicado na Figura 2. São as regiões onde pode ocorrer contato.

Figura 1 – Roda AAR 1-B de friso largo: à esquerda, a roda seccionada por um plano radial, à direita, o perfil da região de rolamento

Fonte: adaptado de [2]

Figura 2 – Trilho 115 RE: à esquerda, o trilho seccionado por um plano transversal, à direita, o perfil do boleto

Fonte: adaptado de [3]

Os produtos da natureza, no entanto, não têm qualquer compromisso com a “perfeição” idealizada pela mente humana. Mesmo geometrias projetadas são maculadas com imperfeições pelo processo de manufatura. Além disso, a ação do tempo tende a degradar as formas ainda mais, como é o caso de rodas e trilhos, que se desgastam com o uso perdendo o formato original. Em geral, não é possível associar uma figura geométrica simples ou uma combinação de figuras a um objeto real exatamente. Para se trabalhar com uma geometria irregular do mundo real de forma analítica, surge a necessidade de se criar um modelo aproximado que a represente. A modelagem é feita com base em informações obtidas por medição do objeto. Esta questão específica de recriar a geometria de um objeto é referida como problema de reconstrução geométrica.

Essencialmente, a geometria de um corpo é a relação de distâncias e direções entre os infinitos pontos que o compõe, o fim da reconstrução geométrica é a criação de uma forma analítica que tenha aproximadamente as mesmas relações geométricas entre pontos correspondentes. No século XVII, o filósofo e matemático René Descartes propôs um modelo da geometria euclidiana em que as posições dos pontos são dadas por coordenadas medidas usando um sistema de referência formado por eixos ortogonais, a geometria pode então ser traduzida para a linguagem da álgebra, assim como a álgebra, para a linguagem da geometria [4]. A geometria analítica originada pelas ideias de Descartes possibilita a descrição da informação geométrica necessária para a reconstrução do nosso objeto. Como é impossível identificar a posição de infinitos pontos, é feita uma amostragem do objeto, isto é, um conjunto finito de amostras é selecionado e suas posições são mapeadas em um sistema de coordenadas cartesiano. A reconstrução geométrica é aplicada ao conjunto de amostras na esperança de que o todo seja aproximado.

As técnicas de reconstrução podem ser classificadas em dois grupos: interpolação e aproximação. Chamamos de técnicas de interpolação os métodos que criam reconstruções que se encaixam exatamente à informação disponível. As técnicas de aproximação, por outro lado, permitem que a reconstrução tenha um certo desvio do encaixe exato. A aproximação é mais adequada para ajuste de amostras medidas, pois há sempre algum erro introduzido no valor da posição pelo processo de medição. O resultado da interpolação a amostras imprecisas pode conter grandes distorções.

Este trabalho visa o desenvolvimento de um método de ajuste por aproximação de perfis planos de roda e trilho por curvas NURBS (do inglês, *Non-Uniform Rational B-Splines*). Os perfis medidos são dados como listas de pontos. As curvas de ajuste são compostas de arcos de circunferência e segmentos de reta (“arcos de raio infinito”) continuamente conectados, um tipo especial de curva com representação NURBS denominada spline de arcos.

Linhas retas e circulares têm tido extrema importância para a geometria desde suas origens na antiguidade clássica, isso porque são ideias relativamente simples de se conceituar e facilmente realizáveis no mundo material com os simples instrumentos régua e compasso. Até a época atual, da manufatura assistida por computador, os segmentos de reta e os arcos de circunferência se mantêm relevantes por servirem como intermediários entre as formas livres de CAD e o processo de manufatura com controle numérico computadorizado (CNC, do inglês *Computer Numeric Control*) [5, 6, 7, 8], a linguagem G (*G-code*) usada para descrever o trajeto da ferramenta de usinagem em CNC contém instruções para a produção de segmentos retos e circulares.

Além da facilidade em as produzir, splines de arcos fazem a ponte entre a manufatura assistida por computador (CAM, do inglês *Computer-Aided Manufacture*) e o desenho (geométrico) assistido por computador (CA(G)D, do inglês *Computer-Aided (Geometric) Design*) por sua capacidade de aproximar formas diversas como demonstrado por Marciniak e Putz [9]. Em particular, splines de arcos são adequadas para aproximar diversos outros tipos de curvas NURBS [6, 10, 11, 7, 8].

Os perfis de rodas e trilhos reconstruídos como splines de arcos podem ser usados como representações contínuas e suaves das fronteiras desses corpos em simulações computacionais de contato. A computação destes problemas é feita através de métodos da mecânica computacional, área que consiste na aplicação de métodos numéricos para a resolução de problemas de mecânica. Inclusa na mecânica computacional está a mecânica do contato computacional, que lida com problemas em que há interação superficial entre corpos.

A resolução de um problema com contato envolve uma sequência de etapas que se repetem a cada instante de tempo com uma nova configuração do sistema de corpos. Primeiramente é feita uma busca por pontos de contato, a geometria dos corpos tem papel fundamental nesta fase. Conhecidos os pontos de contato, pode-se determinar os esforços resultantes da interação. O contato impõe certas restrições ao movimento (a não penetração entre corpos por exemplo), que são asseguradas através do método dos multiplicadores de Lagrange, ou do método da penalidade, ou de leis constitutivas de interface, entre outras técnicas, o que dá origem a forças reativas, em um contexto de solução de um modelo matemático da mecânica computacional. As forças de contato afetam a evolução do sistema movendo e deformando os corpos que o compõe.

O Método dos Elementos Finitos (MEF) é particularmente popular para resolver problemas com corpos deformáveis, sendo aplicado na mecânica do contato [12] e dinâmica de sistemas multicorpos [13]. Corpos deformáveis são descritos por uma quantidade infinita de graus de liberdade pois cada ponto do corpo pode deslocar-se independentemente dos demais a princípio, e o movimento é governado por equações diferenciais. Através do MEF, o corpo é discretizado e seu movimento passa a ser descrito por um conjunto finito de deslocamentos, e as equações diferenciais são transformadas em equações algébricas com as quais computadores podem lidar facilmente.

Um dos motivos de usar splines de arcos como curva de aproximação para os perfis de rodas e trilhos é sua compatibilidade com o programa de elementos finitos com o qual trabalhamos. O GIRAFFE [14] (do inglês: *Generic Interface Readily Accessible For Finite Elements*, ou Interface Geral de Pronto Acesso Para Elementos Finitos em tradução livre) é uma plataforma de elementos finitos desenvolvida em C++ com propósitos acadêmicos. O código foi escrito tendo em mente a fácil incorporação de novos elementos finitos aos já disponíveis. O programa conta com elementos habituais como barra [15, 16], casca [17] e corpo rígido. A estes elementos podem ser aplicadas várias formas de carregamento como esforços nodais e cargas do ambiente (campo gravitacional e arrasto de fluido). Além disso, o programa permite a interação entre corpos através de uma coleção de métodos de contato (detalhadamente explicados em [18]). A detecção e computação de forças de contato é feita a partir de modelos das fronteiras dos corpos por certas classes de superfícies, dentre as quais se encontram as superfícies de extrusão e revolução baseadas em arcos de circunferência.

Figura 3 – Superfícies obtidas a partir de arcos de circunferência (a) por extrusão (b) por revolução

Fonte: [19]

A superfície de extrusão é função de um parâmetro angular θ , que localiza pontos do arco via funções trigonométricas, e um parâmetro longitudinal ζ . A superfície de revolução é função de dois ângulos: θ , referente ao arco, e ϕ , o ângulo de revolução. Detalhes das parametrizações e do sistema de coordenadas considerado encontram-se em [19].

Qualquer corpo cuja superfície possa ser descrita como de extrusão ou de revolução pode ser modelado desde que o método de ajuste consiga aproximar a curva geratriz. Em particular, o método é usado para modelagem de rodas e trilhos, com os quais se coleta resultados usados na metodologia de análise proposta por Higa [20]. Podemos determinar os pontos de maior propensão ao contato (*loci* de contato) e intensidade das forças que uma população de rodas causa sobre um trecho de via.

A abordagem trabalha com uma simplificação bidimensional do problema que usa apenas os perfis planos e não as superfícies. O Giraffe foi adaptado para tratar o problema de contato entre splines de arcos como um caso especial, pois há solução analítica para o problema de mínima distância entre dois arcos de circunferência.

Um exemplo interessante de aplicação do Giraffe, ilustrado na Figura 4, pode ser encontrado em [19]. Neste exemplo, um transportador aéreo modelado por três corpos rígidos (a parte principal e duas rodas) é sustentado por um trilho-guia flexível.

Figura 4 – Exemplo: transportador aéreo

Fonte: [19]

Há dois tipos de superfícies de contato: as superfícies de contato do trilho-guia são obtidas por extrusão de um arco de raio muito grande (praticamente um segmento de reta) e as das rodas, por revolução do perfil circular das rodas. Os arcos são mostrados na Figura 5.

Com esta ferramenta também é possível tratar o caso especial plano em que as splines de arcos servem de fronteira para as seções planas das rodas e trilhos. No problema de contato entre curvas planas, arcos de circunferência têm a vantagem de que os pares de contato podem ser encontrados analiticamente [20].

Figura 5 – Arcos de contato (a) para o trilho (b) para a roda

Fonte: [19]

Outro programa de uso corrente para a análise de contato é o CONTACT [21]. A resolução de um problema de contato no CONTACT exige que sejam dadas as curvaturas principais das superfícies dos corpos. Estes dados só podem ser obtidos de uma superfície suave, portanto, quando a informação conhecida é um conjunto de pontos discretos, uma reconstrução suave é imprescindível.

Outros programas de análise ferroviária como o GENSYS [22] e o módulo RAIL do SIMPACK [23] (que inclusive podem usar o CONTACT como recurso adicional) têm como entrada listas de pontos. Internamente, estes programas reconstróem um perfil suave a partir da lista de pontos, o SIMPACK, por exemplo, cria uma aproximação NURBS de grau três. No entanto, como o objetivo destes programas é a análise da dinâmica ferroviária, eles não fornecem as parametrizações dos perfis reconstruídos.

A modelagem dos perfis como uma curva composta por um pequeno número de arcos de circunferência facilita também a análise de características geométricas, permitindo fazer inferências que seriam impossíveis a partir do conjunto de pontos discretos diretamente. Além disso, o perfil reconstruído dado por um pequeno conjunto de parâmetros manipuláveis pode ser deformado facilmente em um novo perfil, o que abre espaço para estudos de otimização de perfis.

Neste trabalho, faremos a aproximação por splines de arcos de uma sequência de pontos discretos medidos de um perfil de roda ou trilho. Será estimado o gráfico de curvatura do perfil

a partir dos pontos. Será determinada uma aproximação constante por trechos do gráfico de curvatura e a spline de arcos correspondente será tomada como aproximação do perfil. A qualidade da aproximação será melhorada através de um método de otimização. Por fim, parametrizaremos a solução como curva NURBS e superfícies NURBS serão geradas para modelar o elemento ferroviário original.

O trabalho é dividido em duas partes. À primeira parte concerne o método de ajuste e a segunda parte trata da parametrização NURBS das curvas obtidas.

Os dois primeiros capítulos trazem a fundamentação teórica do método de ajuste. O Capítulo 1 trata dos conceitos de geometria diferencial de curvas planas, sendo os principais a parametrização natural de uma curva por comprimento de arco e a curvatura orientada. A aproximação baseada na curvatura é então aprimorada por uma otimização de mínimos quadrados. O Capítulo 2 dedica-se ao estudo dos conceitos de otimização e aos métodos numéricos ligados a essa área. Em especial, ao método de Newton aplicado à otimização e ao método de Gauss-Newton, uma aproximação do método anterior adequada a problemas de mínimos quadrados.

O Capítulo 3 contém a descrição da metodologia. Inicia com a estimativa das medidas geométricas de uma curva a partir de uma sequência de amostras, que são usadas na produção da curva inicial de ajuste. Em seguida, uma descrição da spline de arcos é dada e relações desta com as amostras são obtidas.

No Capítulo 4 são apresentados os resultados da aplicação da metodologia a uma coleção de exemplos abrangendo curvas paramétricas e perfis de roda e trilho ferroviários.

A meta final é usar as curvas ajustadas em modelagem de contato, o que demanda superfícies parametrizadas. O Capítulo 5 lida com a teoria das curvas NURBS e como splines de arcos podem ser parametrizadas neste esquema. O Capítulo 6 traz a parametrização de superfícies NURBS, em particular, superfícies de extrusão e revolução.

Objetivos

O primeiro objetivo desse trabalho é propor uma metodologia de ajuste de curvas compostas por arcos de circunferência a sequências de pontos medidos em perfis ferroviários reais. O segundo objetivo é parametrizar a spline de arcos obtida pelo procedimento de ajuste como NURBS e usar essa parametrização na produção de superfícies de contato compatíveis com a plataforma de elementos finitos Giraffe.

1 Teoria das curvas planas

A geometria diferencial é o campo da matemática que estuda a geometria de curvas e superfícies, avaliando-se as relações entre pontos de uma vizinhança local pertencente à entidade geométrica. Os conceitos básicos da geometria diferencial são locais e determinados com o emprego do cálculo diferencial em parametrizações que descrevem a forma geométrica. Em curvas, podemos calcular em cada ponto uma reta tangente, que informa a direção de alinhamento entre pontos vizinhos. A variação da reta tangente ao longo da curva é uma quantidade que chamamos de curvatura. Se o alinhamento entre pontos vizinhos se mantém constante, a curva trata-se de uma linha reta, que tem curvatura nula. Quanto maior a curvatura em um ponto, mais a curva se desvia de uma linha reta. A geometria diferencial também estabelece relações de distância entre pontos vizinhos com o uso de derivadas, e com a integração destas distâncias locais é possível determinar o comprimento de uma curva. Podemos dizer que a geometria diferencial é a disciplina que combina a geometria analítica com o cálculo diferencial e integral.

Conceitos similares são definidos em superfícies, a diferença sendo que enquanto, em uma curva, só existe uma direção de alinhamento de vizinhos em cada ponto, em uma superfície, há infinitas direções com vizinhos. A tangente é generalizada em um plano que contém todas as possíveis retas tangentes. A curvatura está relacionada a como o plano tangente varia em cada direção a partir de um ponto. É definida uma métrica local para cálculo de comprimentos, ângulos e áreas. Em tudo é aplicado o aparato do cálculo diferencial e integral.

No que segue, formalizaremos os conceitos da geometria diferencial para curvas planas. Algumas referências com conteúdo sobre o assunto são [24, 25, 26, 27].

1.1 Definição de curva plana

Uma curva plana é uma função contínua $\gamma: I \subset \mathbb{R} \rightarrow \mathbb{R}^2$. A curva é um mapeamento de cada ponto do intervalo I identificado por um parâmetro t a um ponto $\gamma(t)$ no plano. A imagem da curva é chamada de trajetória ou traço. Há duas interpretações interessantes para a curva. Em uma delas, a curva é vista como a deformação do intervalo I em sua imagem $\gamma(I)$. A outra interpretação é física, identificando o parâmetro com tempo, podemos interpretar a função como a posição $\gamma(t)$ no instante t de uma partícula se deslocando sobre a trajetória.

Figura 6 – Representação de uma curva plana genérica

Fonte: autoria própria

Temos associado ao plano um sistema de coordenadas cartesianas, podemos escrever a curva em termos de funções coordenadas $x, y: I \subset \mathbb{R} \rightarrow \mathbb{R}$,

$$\gamma(t) = (x(t), y(t)) = x(t)\mathbf{i} + y(t)\mathbf{j}, \quad (1.1)$$

onde $\mathbf{i}$ e $\mathbf{j}$ são os vetores base de $\mathbb{R}^2$.

Um exemplo típico de curva é o que tem como imagem o gráfico de uma função de x em y . De forma genérica, consideramos uma função $f: I \subset \mathbb{R} \rightarrow \mathbb{R}$ tal que $y = f(x)$. Definimos a curva parametrizada em x

$$\gamma(x) = x\mathbf{i} + f(x)\mathbf{j}, \quad (1.2)$$

A trajetória desta curva passa pelos pontos $(x, f(x))$, que formam o gráfico da função f .

1.2 Vetor tangente

Diferenciando a curva, obtemos um vetor $\dot{\gamma}(t)$ que relaciona a posição de cada ponto $\gamma(t)$ com a do vizinho imediato $\gamma(t + dt)$, tal vetor determina o alinhamento local da curva e é definido como o vetor tangente no ponto. A primeira derivada da Equação (1.1) resulta em

$$\dot{\gamma}(t) = \dot{x}(t)\mathbf{i} + \dot{y}(t)\mathbf{j}. \quad (1.3)$$

Na interpretação física, $\dot{\gamma}(t)$ é o vetor velocidade da partícula no instante t .

Figura 7 – Representação do vetor tangente à curva

Fonte: autoria própria

Para uma curva na forma da Equação (1.2), a primeira derivada com relação a x resulta no vetor tangente

$$\gamma'(x) = 1\mathbf{i} + f'(x)\mathbf{j}. \quad (1.4)$$

A componente na direção $\mathbf{i}$ do vetor tangente sempre é 1 neste caso, do que se conclui que a projeção do ponto $\gamma(x)$ sobre o eixo x tem velocidade uniforme e unitária.

Podemos escrever o vetor tangente como $\dot{\gamma} = v\mathbf{T}$, onde $v = \|\dot{\gamma}\|$ é sua magnitude, que chamaremos de velocidade, e $\mathbf{T}$ é o vetor unitário que indica sua direção. A magnitude é medida com a norma euclidiana associada ao produto escalar, isto é, para um vetor qualquer $\mathbf{u}$,

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}. \quad (1.5)$$

Aplicando o produto escalar à Equação (1.3), chegamos a uma expressão para a velocidade de uma curva com parâmetro genérico t

$$v(t) = \|\dot{\gamma}(t)\| = \sqrt{(\dot{x}(t))^2 + (\dot{y}(t))^2}. \quad (1.6)$$

No caso particular em que y é função de x , com o vetor tangente representado na Equação (1.4), a velocidade em função de x fica

$$v(x) = \|\gamma'(x)\| = \sqrt{1 + f'(x)^2}. \quad (1.7)$$

1.3 Comprimento de uma curva

O módulo do vetor tangente é usado no cálculo do comprimento da curva. Considerando a curva entre dois pontos de parâmetros a e b , podemos aproximar seu comprimento L aproximando a curva por uma sucessão de segmentos de reta conforme a Figura 8.

Figura 8 – Aproximação de uma curva por segmentos de reta

Fonte: autoria própria

Partimos o intervalo $[a, b]$ em um número N de trechos escolhendo valores $t_i, i = 0, 1, 2, \dots, N$, dentro do intervalo, $a = t_0 < t_1 < t_2 < \dots < t_N = b$. No exemplo da Figura 8, temos $N = 7$ trechos. Com a partição do intervalo, construímos uma curva aproximada formada por segmentos de reta ligando os pontos consecutivos da curva, indicados na Figura 8 pelo valor correspondente do parâmetro. A distância entre dois pontos consecutivos é dada pela norma da diferença entre as posições

$$\Delta s_i = \|\gamma(t_i) - \gamma(t_{i-1})\|, i = 1, 2, \dots, N. \quad (1.8)$$

Somando o comprimento de cada segmento, temos o comprimento total da curva aproximada

$$L_N = \sum_{i=1}^N \|\gamma(t_i) - \gamma(t_{i-1})\|. \quad (1.9)$$

Para uma curva γ continuamente diferenciável, definimos o comprimento L entre os valores a e b como o limite da soma na Equação (1.9) quando o número N de trechos tende ao infinito

$$L = \lim_{N \rightarrow \infty} \sum_{i=1}^N \|\gamma(t_i) - \gamma(t_{i-1})\| = \lim_{N \rightarrow \infty} \sum_{i=1}^N \left\| \frac{\gamma(t_i) - \gamma(t_{i-1})}{t_i - t_{i-1}} \right\| (t_i - t_{i-1}). \quad (1.10)$$

Pelo conceito da soma de Riemann, este limite equivale à seguinte integral definida

$$L = \int_a^b \|\dot{\gamma}(t)\| dt . \quad (1.11)$$

A magnitude da primeira derivada da curva $v(t) = \|\dot{\gamma}(t)\|$ está relacionada com a distância entre pontos imediatamente vizinhos

$$ds = \|\gamma(t + dt) - \gamma(t)\| = v(t)dt . \quad (1.12)$$

Assim, $v(t)$ pode ser entendido como um fator métrico da curva.

Fisicamente, ds seria o comprimento infinitesimal percorrido sobre a trajetória pela partícula com velocidade v no instante t durante o intervalo infinitesimal dt . O comprimento total da trajetória percorrida entre os instantes a e b é dado pela integral

$$L = \int_a^b v(t) dt , \quad (1.13)$$

que é idêntica à obtida anteriormente.

Em especial, para curvas em que a coordenada y é escrita em função de x , substituímos a expressão para a velocidade presente na Equação (1.7) para obter a seguinte fórmula para o comprimento da curva

$$L = \int_a^b \sqrt{1 + (f'(x))^2} dx , \quad (1.14)$$

note que a variável de integração é x , o parâmetro deste tipo particular de curva.

A integral da Equação (1.11) é empregada na definição da função comprimento de arco

$$s(t) = \int_a^t \|\dot{\gamma}(\tau)\| d\tau , \quad (1.15)$$

que é o comprimento da curva medido a partir do ponto $\gamma(a)$ até o ponto $\gamma(t)$.

Figura 9 – Comprimento de arco de uma curva entre os pontos $\gamma(a)$ e $\gamma(t)$

Fonte: autoria própria

A cada ponto da curva está associado um, e somente um, valor de s , e este valor é crescente com o parâmetro da curva, de forma que se temos

$$s = f(t), \quad (1.16)$$

ou seja, s como uma função de t , então é possível escrever a função inversa

$$t = f^{-1}(s). \quad (1.17)$$

1.3.1 Exemplo. Parábola

Uma parábola pode ser parametrizada através da função $y(x) = x^2$ da seguinte forma

$$\gamma(x) = x\mathbf{i} + x^2\mathbf{j}. \quad (1.18)$$

Figura 10 – Parábola $y(x) = x^2$ e comprimento de arco calculado a partir do vértice até um ponto $x = b$ genérico

Fonte: autoria própria

Como a imagem da parábola é simétrica por reflexão em torno do eixo y , basta considerar o comprimento de arco a partir do vértice. Aplicando a Equação (1.14) com o limite inferior $a = 0$, correspondente ao vértice, ponto de simetria da parábola, temos

$$s(b) = \int_0^b \sqrt{1 + (2x)^2} dx = \int_0^b \sqrt{1 + 4x^2} dx = \frac{\ln(\sqrt{1 + 4b^2} + 2b)}{4} + \frac{b\sqrt{1 + 4b^2}}{2}, \quad (1.19)$$

válida para valores positivos de b . Para b negativo, usa-se o valor absoluto, pois, devido à simetria, o comprimento calculado à esquerda do vértice é igual ao calculado à direita do vértice. O comprimento entre pontos quaisquer pode ser composto por somas e diferenças usando a fórmula obtida.

1.3.2 Exemplo. Espiral logarítmica

Uma espiral logarítmica é uma trajetória na qual a distância entre a origem do sistema de coordenadas e o ponto sobre a trajetória aumenta, ou diminui, exponencialmente com o ângulo formado entre o raio vetor e o eixo x . Em coordenadas polares, escrevemos $r(\varphi) = e^{k\varphi}$. A Figura 11 traz uma espiral deste tipo com $k = 0,31$ e φ completando uma volta de 0 a 2π .

Figura 11 – Espiral logarítmica $r(\varphi) = e^{0,31\varphi}$, $0 \leq \varphi \leq 2\pi$

Fonte: autoria própria

Adotando φ como parâmetro, podemos escrever a curva em coordenadas cartesianas como

$$\gamma(\varphi) = r(\varphi) \cos \varphi \mathbf{i} + r(\varphi) \sin \varphi \mathbf{j} = e^{k\varphi} \cos \varphi \mathbf{i} + e^{k\varphi} \sin \varphi \mathbf{j}. \quad (1.20)$$

Curvas descritas por coordenadas polares têm uma forma de derivada muito simples em coordenadas cartesianas

$$\gamma'(\varphi) = r'(\varphi)(\cos \varphi \mathbf{i} + \sin \varphi \mathbf{j}) + r(\varphi)(-\sin \varphi \mathbf{i} + \cos \varphi \mathbf{j}) = r'(\varphi)\hat{\mathbf{r}} + r(\varphi)\hat{\boldsymbol{\theta}}, \quad (1.21)$$

em que aparecem dois vetores unitários e ortogonais $\hat{\mathbf{r}} = \cos \varphi \mathbf{i} + \sin \varphi \mathbf{j}$ e $\hat{\boldsymbol{\theta}} = -\sin \varphi \mathbf{i} + \cos \varphi \mathbf{j}$, usualmente chamados de versor radial e versor circunferencial, ou tangencial, respectivamente. As duas parcelas do vetor tangente ficam em quadratura, portanto a norma segue a forma pitagórica

$$\|\gamma'(\varphi)\|^2 = r'(\varphi)^2 + r(\varphi)^2. \quad (1.22)$$

Em particular, para a equação polar da espiral logarítmica $r(\varphi) = e^{k\varphi}$, temos a norma ao quadrado do vetor tangente

$$\|\gamma'(\varphi)\|^2 = (ke^{k\varphi})^2 + (e^{k\varphi})^2 = (1 + k^2)e^{2k\varphi} \quad (1.23)$$

com a qual podemos calcular o comprimento da curva entre dois valores de φ

$$L = \int_{\varphi_1}^{\varphi_2} \|\gamma'(\varphi)\| d\varphi = \sqrt{1+k^2} \int_{\varphi_1}^{\varphi_2} e^{k\varphi} d\varphi = \frac{\sqrt{1+k^2}}{k} (e^{k\varphi_2} - e^{k\varphi_1}). \quad (1.24)$$

Uma propriedade interessante da curva logarítmica é a autosimilaridade. O comprimento da curva cresce com um fator multiplicativo constante a cada segmento com um intervalo angular fixo. Por exemplo, o comprimento do trecho no primeiro quadrante da curva é

$$L_1 = \frac{\sqrt{1+k^2}}{k} (e^{k\frac{\pi}{2}} - 1). \quad (1.25)$$

O trecho no segundo quadrante tem comprimento

$$L_2 = \frac{\sqrt{1+k^2}}{k} (e^{k\pi} - e^{k\frac{\pi}{2}}) = e^{k\frac{\pi}{2}} \frac{\sqrt{1+k^2}}{k} (e^{k\frac{\pi}{2}} - 1) = e^{k\frac{\pi}{2}} L_1. \quad (1.26)$$

Para o trecho do terceiro quadrante

$$L_3 = \frac{\sqrt{1+k^2}}{k} (e^{\frac{3k\pi}{2}} - e^{k\pi}) = e^{k\pi} \frac{\sqrt{1+k^2}}{k} (e^{k\frac{\pi}{2}} - 1) = e^{k\pi} L_1. \quad (1.27)$$

Finalmente, o trecho no quarto quadrante tem comprimento

$$L_4 = \frac{\sqrt{1+k^2}}{k} (e^{2k\pi} - e^{\frac{3k\pi}{2}}) = e^{\frac{3k\pi}{2}} \frac{\sqrt{1+k^2}}{k} (e^{k\frac{\pi}{2}} - 1) = e^{\frac{3k\pi}{2}} L_1. \quad (1.28)$$

De um quadrante para o outro, o intervalo angular é $\pi/2$ e o comprimento é multiplicado pelo fator constante $e^{k\pi/2}$. Isto significa que se a curva for multiplicada por $e^{k\pi/2}$ e sofrer uma rotação de $\pi/2$, a imagem da curva transformada ficará sobreposta à imagem da curva original. A equivalência entre rotação e escala se traduz na autosimilaridade observada, é como se trechos consecutivos fossem cópias escaladas uns dos outros.

Quando o fator de escala para rotações de ângulo reto é a razão áurea φ , a espiral é chamada de áurea, com k obtido pela relação

$$e^{k\frac{\pi}{2}} = \varphi = \frac{1+\sqrt{5}}{2} \Rightarrow k = \frac{2}{\pi} \ln \varphi \approx 0,31. \quad (1.29)$$

A curva representada na Figura 11 é uma aproximação da espiral áurea.

1.3.3 Exemplo. Arco de circunferência

Uma circunferência de raio R centrada na origem é definida pela equação em coordenadas polares $r(\varphi) = R$ (coordenada radial r constante igual ao raio R), a partir da qual se escreve a tradicional parametrização em coordenadas cartesianas

$$\gamma(\varphi) = R \cos \varphi \mathbf{i} + R \sin \varphi \mathbf{j} . \quad (1.30)$$

A “velocidade” desta curva é facilmente obtida como

$$\|\gamma'(\varphi)\| = \|-R \sin \varphi \mathbf{i} + R \cos \varphi \mathbf{j}\| = R , \quad (1.31)$$

uma constante.

O comprimento da circunferência entre dois valores angulares φ_1 e φ_2 é dado pela Equação (1.11):

$$L = \int_{\varphi_1}^{\varphi_2} \|\gamma'(\varphi)\| d\varphi = \int_{\varphi_1}^{\varphi_2} R d\varphi = R(\varphi_2 - \varphi_1) = R\Delta\varphi . \quad (1.32)$$

A fórmula encontrada está de acordo com o conceito de ângulo em radianos que relaciona uma abertura angular $\Delta\varphi$ com o comprimento de arco L subentendido pelo ângulo sendo o raio R o fator de conversão.

1.4 Reparametrização por comprimento de arco

Uma reparametrização da curva é uma composição de funções. Suponha que o parâmetro t possa ser escrito como uma função inversível h de um parâmetro alternativo u

$$t = h(u) . \quad (1.33)$$

A curva reparametrizada é a composição da curva parametrizada em t com a função de reparametrização h :

$$\alpha(u) = \gamma(t) \circ h(u) . \quad (1.34)$$

Trata-se de uma nova função, portanto uma outra curva. A relação biunívoca entre t e u , dada pela inversibilidade de h , implica que ambas as curvas têm o mesmo traço. Também se demonstra que o comprimento calculado entre pontos correspondentes é igual para qualquer parametrização, sendo, portanto, uma propriedade da imagem da curva e não da função em si.

Uma reparametrização de grande importância em curvas é a chamada reparametrização por comprimento de arco, em que usamos como função de reparametrização a inversa expressa na Equação (1.17):

$$\alpha(s) = \gamma(t) \circ f^{-1}(s) . \quad (1.35)$$

A propriedade mais importante da parametrização por comprimento de arco está relacionada ao vetor tangente. Obtemos o vetor tangente derivando a Equação (1.35), o que é feito pela regra da cadeia

$$\alpha'(s) = \left(\frac{d\gamma(t)}{dt} \Big|_{t=f^{-1}(s)} \right) \left(\frac{df^{-1}(s)}{ds} \right). \quad (1.36)$$

O primeiro fator é o vetor tangente à curva $\gamma(t)$ calculado em $t(s)$. O segundo fator, de acordo com o teorema da função inversa, é

$$\frac{df^{-1}(s)}{ds} = \left(\frac{df(t)}{dt} \right)^{-1} \Big|_{t=f^{-1}(s)}, \quad (1.37)$$

contanto que $df(t)/dt \neq 0$. Como a função $f(t)$ é a integral expressa na Equação (1.15), sua derivada se iguala ao integrando

$$\frac{df(t)}{dt} = \|\dot{\gamma}(t)\|. \quad (1.38)$$

Voltando à Equação (1.36), temos

$$\alpha'(s) = \left(\frac{d\gamma(t)}{dt} \Big|_{t=f^{-1}(s)} \right) \frac{df^{-1}(s)}{ds} = \frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|} \Big|_{t=f^{-1}(s)}, \quad (1.39)$$

onde aparece o vetor tangente $\dot{\gamma}(t)$ dividido por sua magnitude $\|\dot{\gamma}(t)\|$, o vetor unitário da direção tangente que denotamos por

$$\mathbf{T}(t) = \frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|}. \quad (1.40)$$

Portanto,

$$\alpha'(s) = \mathbf{T}(s). \quad (1.41)$$

A parametrização por comprimento de arco é tal que, para todo s ,

$$\|\alpha'(s)\| = 1. \quad (1.42)$$

Uma parametrização com a propriedade de velocidade unitária é dita natural, é o caso da parametrização por comprimento de arco.

1.5 Curvatura

Intuitivamente, a curvatura de um traço é a medida do quanto ela se afasta de uma linha reta. Em cada ponto de uma curva suave, a reta tangente é a melhor aproximação retilínea possível para a curva, a mudança de direção do vetor tangente indica o quanto a curva se desvia da reta tangente em cada ponto e, portanto, está relacionada com o que chamamos de curvatura. Quanto maior a curvatura em um ponto, mais intensa é a mudança da direção tangente em relação aos pontos vizinhos. A Figura 12 traz uma ilustração do fenômeno.

Figura 12 – Mudança da direção tangente em uma curva

Fonte: autoria própria

A derivada do vetor unitário $\mathbf{T}$ com relação ao comprimento de arco remete à medida da mudança de direção tangente entre pontos vizinhos. Tendo a parametrização por comprimento de arco, a partir da Equação (1.41), podemos calcular a derivada

$$\frac{d\mathbf{T}(s)}{ds} = \alpha''(s). \quad (1.43)$$

A curvatura κ é definida como a magnitude desta função vetorial

$$\kappa(s) = \left\| \frac{d\mathbf{T}(s)}{ds} \right\| = \|\alpha''(s)\|. \quad (1.44)$$

A expressão da curvatura em coordenadas cartesianas parametrizadas pelo comprimento de arco é escrita na forma pitagórica simples

$$\kappa(s) = \sqrt{x''(s)^2 + y''(s)^2}. \quad (1.45)$$

É importante entender que a curvatura é definida como a derivada do vetor tangente com relação a um parâmetro específico, o comprimento de arco. Se o parâmetro da curva for um t arbitrário, então a regra da cadeia deve ser aplicada à seguinte composição, similar à Equação (1.35),

$$\mathbf{T}(s) = \mathbf{T}(t) \circ f^{-1}(s). \quad (1.46)$$

Derivando a Equação (1.46), temos

$$\frac{d\mathbf{T}(s)}{ds} = \left(\frac{1}{\|\dot{\gamma}(t)\|} \frac{d\mathbf{T}(t)}{dt} \right) \Big|_{t=f^{-1}(s)}. \quad (1.47)$$

A derivada do vetor tangente com relação a t pode ser calculada com base na Equação (1.40):

$$\frac{d\mathbf{T}(t)}{dt} = \frac{d}{dt} \left(\frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|} \right) = \frac{\ddot{\gamma}(t) - (\ddot{\gamma}(t) \cdot \mathbf{T}(t))\mathbf{T}(t)}{\|\dot{\gamma}(t)\|}. \quad (1.48)$$

Interpretando $\ddot{\gamma}(t)$ como o vetor aceleração no instante t da partícula que traça a curva, o numerador do membro direito da Equação (1.48) é a parcela de aceleração na direção ortogonal à trajetória, a aceleração centrípeta. Como $\mathbf{T}(t)$ é a direção da velocidade da partícula, a Equação (1.48) reflete o fato que a aceleração centrípeta é responsável pela mudança de direção. De forma puramente geométrica, a curvatura está relacionada apenas com a mudança de direção do vetor tangente, assim, a variação de magnitude, representada pela parcela $(\ddot{\gamma}(t) \cdot \mathbf{T}(t))\mathbf{T}(t)$, deve ser descontada. No caso da parametrização por comprimento de arco, a magnitude do vetor tangente é constante, por isso a simples derivada indicada na Equação (1.43) é suficiente. Substituindo na Equação (1.47), temos

$$\frac{d\mathbf{T}(s)}{ds} = \left(\frac{\ddot{\gamma}(t) - (\ddot{\gamma}(t) \cdot \mathbf{T}(t))\mathbf{T}(t)}{\|\dot{\gamma}(t)\|^2} \right) \Big|_{t=f^{-1}(s)}, \quad (1.49)$$

equação da qual basta calcular a magnitude de ambos os membros e compor com $s = f(t)$ para obter a curvatura em função do parâmetro t

$$\kappa(t) = \frac{\sqrt{\|\dot{\gamma}(t)\|^2 - (\ddot{\gamma}(t) \cdot \mathbf{T}(t))^2}}{\|\dot{\gamma}(t)\|^2}. \quad (1.50)$$

As Equações (1.43) e (1.44) são versões particulares simplificadas das Equações (1.48) e (1.50) respectivamente. As simplificações obtidas com a parametrização por comprimento de arco mostram quão vantajosa esta é. A generalização da Equação (1.45) é obtida ao se considerar as funções coordenadas da curva em relação a t na Equação (1.50):

$$\kappa(t) = \frac{|\dot{x}(t)\ddot{y}(t) - \dot{y}(t)\ddot{x}(t)|}{(\dot{x}(t)^2 + \dot{y}(t)^2)^{3/2}}. \quad (1.51)$$

Sendo a norma de $\mathbf{T}$ uma constante, vem do cálculo vetorial que sua derivada é ortogonal ao próprio vetor. Portanto, podemos definir um vetor normal à curva como

$$\mathbf{N}(t) = \frac{d\mathbf{T}/dt}{\|d\mathbf{T}/dt\|}. \quad (1.52)$$

É possível demonstrar que a parametrização não tem influência sobre esta direção, sendo, então, uma propriedade da imagem da curva. Escolhendo, em particular, a parametrização por comprimento de arco, chegamos à primeira fórmula de Frenet-Serret

$$\frac{d\mathbf{T}(s)}{ds} = \kappa(s)\mathbf{N}(s). \quad (1.53)$$

A Figura 12 permite a intuição do sentido do vetor $\mathbf{N}$, ele aponta para o “interior” da curva (para o centro de curvatura). Mais precisamente, a reta tangente separa o plano em duas regiões, se a curvatura é não nula, existe uma vizinhança do ponto de tangência contida em somente uma das regiões, o vetor $\mathbf{N}$ aponta para este lado. A orientação do vetor depende da geometria local da curva, podendo inclusive ser indefinida se o ponto tiver curvatura nula.

1.6 Curvatura orientada

No plano, é possível definir um vetor normal $\mathbf{n}$ com orientação fixa em relação ao vetor tangente, permitindo a definição de uma curvatura orientada, ou curvatura com sinal. Define-se $\mathbf{n}$ como o vetor unitário ortogonal a $\mathbf{T}$ tal que o par $(\mathbf{T}, \mathbf{n})$ tenha a mesma orientação do par $(\mathbf{i}, \mathbf{j})$, isto é, existe uma rotação que alinhe $\mathbf{i}$ com $\mathbf{T}$ e $\mathbf{j}$ com $\mathbf{n}$.

Figura 13 – Orientação dos vetores normais em uma curva plana

Fonte: autoria própria

A Figura 13 exemplifica a orientação de $\mathbf{n}$ comparada à de $\mathbf{N}$ em três pontos diferentes. Como foi definido, a orientação de $\mathbf{n}$ é a mesma em todos os pontos, uma rotação de um ângulo

reto no sentido anti-horário em relação a $\mathbf{T}$. O sentido de $\mathbf{N}$ indica o sentido de rotação do vetor tangente, enquanto $\mathbf{n}$ indica a rotação no sentido anti-horário, definido como sentido positivo. No ponto A , $\mathbf{N}$ tem o sentido oposto de $\mathbf{n}$, o que significa que $\mathbf{T}$ tende a girar no sentido negativo, ou sentido horário, neste ponto. No ponto B , não existe o vetor $\mathbf{N}$, pois a curvatura é nula, a tendência local de $\mathbf{T}$ é manter o alinhamento. No ponto C , $\mathbf{n}$ e $\mathbf{N}$ coincidem, pois, a tendência de giro de $\mathbf{T}$ é no sentido positivo.

Definimos a curvatura com sinal como

$$k_s = \frac{d\mathbf{T}}{ds} \cdot \mathbf{n} = (\mathbf{N} \cdot \mathbf{n})\kappa, \quad (1.54)$$

com o produto escalar de ambos os membros da Equação (1.53) por $\mathbf{n}$. De acordo com esta definição, a curvatura é positiva quando $\mathbf{T}$ tende a girar no sentido positivo de rotações, negativa quando o sentido da rotação é negativo e nula quando a tendência é manter o alinhamento (sem curvatura). Na Figura 13, temos os três casos possíveis, no ponto A a curvatura é negativa, em B é nula, e em C , positiva.

Uma expressão em coordenadas cartesianas pode ser obtida para a curvatura orientada. Para um parâmetro arbitrário t escrevemos a Equação (1.54) na forma

$$k_s(t) = \left(\frac{\ddot{\gamma}(t) - (\ddot{\gamma}(t) \cdot \mathbf{T}(t))\mathbf{T}(t)}{\|\dot{\gamma}(t)\|^2} \right) \cdot \mathbf{n}(t), \quad (1.55)$$

tendo em vista que

$$\frac{\ddot{\gamma}(t) - (\ddot{\gamma}(t) \cdot \mathbf{T}(t))\mathbf{T}(t)}{\|\dot{\gamma}(t)\|^2}$$

é a equivalente no parâmetro t da expressão

$$\frac{d\mathbf{T}(s)}{ds}$$

segundo a Equação (1.49).

Pela ortogonalidade dos vetores $\mathbf{T}$ e $\mathbf{n}$, a Equação (1.55) se reduz a

$$k_s(t) = \frac{\ddot{\gamma}(t) \cdot \mathbf{n}(t)}{\|\dot{\gamma}(t)\|^2}. \quad (1.56)$$

Em coordenadas cartesianas,

$$\mathbf{T}(t) = \frac{\dot{x}(t)\mathbf{i} + \dot{y}(t)\mathbf{j}}{\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}}. \quad (1.57)$$

Aplicando uma rotação de $\pi/2$ radianos no sentido anti-horário, obtemos o vetor

$$\mathbf{n}(t) = \frac{-\dot{y}(t)\mathbf{i} + \dot{x}(t)\mathbf{j}}{\sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}}. \quad (1.58)$$

Com isso, a Equação (1.56) fica expressa em coordenadas na forma

$$k_s(t) = \frac{\dot{x}(t)\ddot{y}(t) - \dot{y}(t)\ddot{x}(t)}{(\dot{x}(t)^2 + \dot{y}(t)^2)^{3/2}}. \quad (1.59)$$

No caso da parametrização por comprimento de arco, o denominador é 1 e a fórmula se simplifica para

$$k_s(s) = x'(s)y''(s) - y'(s)x''(s). \quad (1.60)$$

1.6.1 Exemplo. Curvatura de uma circunferência

Sabemos que uma circunferência de raio R centrada na origem pode ser parametrizada como

$$\gamma(\varphi) = R \cos \varphi \mathbf{i} + R \sin \varphi \mathbf{j}. \quad (1.61)$$

Desta forma, o comprimento de arco a partir do ponto $\gamma(0)$, conforme a Equação (1.32), é

$$s = R\varphi, \quad (1.62)$$

o que resulta, respectivamente, na reparametrização por comprimento de arco, sua primeira e segunda derivadas

$$\gamma(s) = R \cos\left(\frac{s}{R}\right) \mathbf{i} + R \sin\left(\frac{s}{R}\right) \mathbf{j}, \quad (1.63)$$

$$\gamma'(s) = -\sin\left(\frac{s}{R}\right) \mathbf{i} + \cos\left(\frac{s}{R}\right) \mathbf{j}, \quad (1.64)$$

e

$$\gamma''(s) = -\frac{1}{R} \cos\left(\frac{s}{R}\right) \mathbf{i} - \frac{1}{R} \sin\left(\frac{s}{R}\right) \mathbf{j}. \quad (1.65)$$

A curvatura é obtida pela Equação (1.44):

$$\kappa = \frac{1}{R}, \quad (1.66)$$

e, aplicando a Equação (1.60), descobrimos que a orientação é positiva

$$k_s = +\frac{1}{R}. \quad (1.67)$$

Com a parametrização dada a circunferência é percorrida no sentido anti-horário, a orientação positiva por convenção. Uma parametrização alternativa da mesma trajetória é obtida trocando as posições de seno e cosseno

$$\gamma(s) = R \sin\left(\frac{s}{R}\right) \mathbf{i} + R \cos\left(\frac{s}{R}\right) \mathbf{j}. \quad (1.68)$$

Verifica-se que a curvatura é negativa neste caso

$$k_s(s) = -\frac{1}{R}, \quad (1.69)$$

a circunferência é percorrida no sentido horário.

1.7 Teorema fundamental das curvas planas

Considere um vetor unitário fixo $\mathbf{u}$ no plano. Seja $\mathbf{v}$ o vetor unitário obtido pela rotação de $\mathbf{u}$ de um ângulo reto no sentido positivo, $\mathbf{u}$ e $\mathbf{v}$ formam uma base ortogonal normalizada do plano.

Seja $\theta(s)$ o ângulo de rotação no sentido positivo necessário para transformar $\mathbf{u}$ em $\mathbf{T}(s)$, então podemos escrever o vetor tangente na base $\{\mathbf{u}, \mathbf{v}\}$ como

$$\mathbf{T}(s) = \cos \theta(s) \mathbf{u} + \sin \theta(s) \mathbf{v}, \quad (1.70)$$

e sua derivada como

$$\mathbf{T}'(s) = -\theta'(s) \sin \theta(s) \mathbf{u} + \theta'(s) \cos \theta(s) \mathbf{v}. \quad (1.71)$$

O vetor $\mathbf{n}$ tem ângulo $\theta(s) + \pi/2$ com relação ao vetor $\mathbf{u}$, portanto

$$\mathbf{n}(s) = \cos\left(\theta(s) + \frac{\pi}{2}\right) \mathbf{u} + \sin\left(\theta(s) + \frac{\pi}{2}\right) \mathbf{v} = -\sin \theta(s) \mathbf{u} + \cos \theta(s) \mathbf{v}. \quad (1.72)$$

Pela definição da curvatura com sinal dada pela Equação (1.54), temos

$$k_s(s) = \mathbf{T}'(s) \cdot \mathbf{n}(s) = \theta'(s), \quad (1.73)$$

o que evidencia a curvatura como velocidade de rotação da direção tangente, ou tendência de girar.

Dada uma função suave $f: I \rightarrow \mathbb{R}$, definindo

$$\theta(s) = \int_{s_0}^s f(t) dt \quad (1.74)$$

e

$$\gamma(s) = \left(\int_{s_0}^s \cos \theta(t) dt, \int_{s_0}^s \sin \theta(t) dt \right), \quad (1.75)$$

para algum $s_0 \in I$, resulta que o vetor tangente à curva γ é

$$\mathbf{T}(s) = \gamma'(s) = \cos \theta(s) \mathbf{i} + \sin \theta(s) \mathbf{j}, \quad (1.76)$$

e $\theta(s)$ é o ângulo que o vetor fixo $\mathbf{i}$ gira para se transformar em $\mathbf{T}(s)$. De acordo com a Equação (1.73), a curvatura orientada de γ é

$$k_s(s) = \theta'(s) = f(s), \quad (1.77)$$

Em conclusão, existe uma curva parametrizada por comprimento de arco cuja curvatura com sinal coincide com uma função suave dada. Este resultado é parte do chamado Teorema Fundamental das Curvas Planas [28]. A segunda parte do teorema atesta que uma segunda curva $\bar{\gamma}(s)$ com a mesma função de curvatura $f: I \rightarrow \mathbb{R}$ pode ser transformada em $\gamma(s)$ por um movimento rígido, a composição de uma rotação com uma translação, isto é, o formato da curva com uma dada curvatura orientada é único.

2 Otimização

Na passagem do século XVIII para o século XIX, havia grande interesse da comunidade científica por fenômenos astronômicos, em particular, a determinação de órbitas de corpos celestes a partir de um conjunto de medições astronômicas pontuais. Teoricamente, desprezando a influência de outros corpos frente ao poder gravitacional do Sol, os corpos do sistema solar seguem órbitas elípticas com o centro de gravidade do Sol em um foco como proposto por Kepler e demonstrado pela Lei da Gravitação Universal de Newton. As órbitas ficam, portanto, determinadas por um conjunto finito de parâmetros. Em 1805, Legendre publica sua obra *Nouvelles Methodes pour la Determination des Orbites des Comètes* contendo um método para ajuste de órbitas em que a soma dos quadrados das diferenças entre os pontos observados e os valores estimados pelo modelo deve ser minimizada, o chamado método dos mínimos quadrados. Gauss é reconhecido como o primeiro a desenvolver o método dos mínimos quadrados por volta de 1795, embora sua publicação tenha sido feita apenas em 1809 [29, 30], uma conquista de Gauss é a demonstração de que o ajuste por mínimos quadrados resulta na melhor estimativa possível partindo do pressuposto que os erros sigam a distribuição normal de probabilidade.

O método dos mínimos quadrados é um exemplo de técnica de otimização. A otimização é a área da matemática que estuda formas de se encontrar os parâmetros corretos para minimizar ou maximizar um valor dependente, no método dos mínimos quadrados este valor é a soma dos quadrados dos erros e os parâmetros determinam a curva de ajuste. O problema de otimização clássico trata da modelagem do custo de um sistema com uma função a valores reais de um conjunto de parâmetros que definam o sistema, e então encontrar os valores dos parâmetros que minimizem o custo. Tal função é denominada objetivo e os parâmetros são chamados de variáveis de otimização. O problema de otimização é enunciado como

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}), \quad (2.1)$$

em que $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo, $\mathbf{x} \in \mathbb{R}^n$ é o vetor de variáveis de otimização contendo n componentes.

Quando as variáveis de otimização podem assumir qualquer valor real, estamos lidando com otimização contínua e irrestrita. Em oposição à otimização contínua está a otimização discreta, caso em que os valores possíveis para as variáveis pertencem a um conjunto contável e, geralmente, finito, por exemplo, um subconjunto dos números inteiros. Também pode ocorrer

a imposição de restrições sobre as variáveis, que podem ser de igualdade ou desigualdade. As restrições têm o poder de limitar o domínio a um subconjunto Ω do $\mathbb{R}^n$.

A seguir, é feita uma breve caracterização da solução do problema de otimização, o ponto minimizador. Veremos que existe uma relação entre a minimização de uma função e os pontos em que seu gradiente é zero, e por isso, os métodos de busca por um minimizador estão relacionados aos de busca por zeros de funções como o método de Newton-Raphson. São descritos os métodos de Newton-Raphson, o método de Gauss-Newton e a aplicação em algoritmos de otimização. Algumas referências para otimização e métodos numéricos são [31, 32, 33].

2.1 Caracterização de um mínimo

Um método de otimização tradicional busca determinar um ponto minimizador no espaço das variáveis $\mathbf{x}^*$. Por definição, um ponto minimizador é aquele que atende à condição

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \Omega, \quad (2.2)$$

isto é, em qualquer outro ponto $\mathbf{x}$ no domínio Ω , a função objetivo não pode ter um valor menor que o avaliado em $\mathbf{x}^*$.

Para variáveis contínuas, a verificação direta de um minimizador pela definição é impossível, mesmo para variáveis discretas ela é ou impossível ou inviável em muitos problemas práticos. A condição expressa na Equação (2.2), é imprescindível, todavia, na demonstração das propriedades básicas dos métodos de otimização.

Com conhecimento suficiente da função objetivo é possível verificar a relação de um dado ponto com outros pertencentes a uma vizinhança sua. Se o ponto é um minimizador, então a função objetivo tem valor igual ou superior ao avaliado no minimizador em todos os outros pontos da vizinhança. Assim, define-se um minimizador local como o $\bar{\mathbf{x}}$ ponto do qual existe uma vizinhança $\mathcal{N} \subset \Omega$ tal que

$$f(\bar{\mathbf{x}}) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathcal{N}. \quad (2.3)$$

A propriedade expressa na Equação (2.2) é uma comparação com todo o domínio caracterizando o minimizador global, onde a função objetivo assume o “mínimo dos mínimos”. Por sua vez, a Equação (2.3) traz uma comparação local, as técnicas do cálculo diferencial são postas a uso na análise do comportamento local da função produzindo características verificáveis de um minimizador local.

Se a função objetivo é continuamente derivável e o domínio é aberto, necessariamente a sua primeira derivada, o vetor gradiente ∇f , deve se anular no ponto $\bar{\mathbf{x}}$:

$$\nabla f(\bar{\mathbf{x}}) = \mathbf{0}. \quad (2.4)$$

Dizemos que $\bar{\mathbf{x}}$ é um ponto crítico da função objetivo, ou que a função objetivo é estacionária em $\bar{\mathbf{x}}$. Caso contrário, existiria uma direção segundo a qual a função poderia ser reduzida, por exemplo, a direção oposta ao gradiente $\mathbf{p} = -\nabla f(\bar{\mathbf{x}})$.

Se a função objetivo tem a segunda derivada contínua, então também é uma condição necessária de minimização que $\nabla^2 f(\bar{\mathbf{x}})$ seja positiva semidefinida, $\nabla^2 f$ denota a segunda derivada. Dizemos que a função objetivo é localmente convexa em $\bar{\mathbf{x}}$. Para uma função escalar, a segunda pode ser representada por uma matriz $\mathbf{H}$ chamada de hessiana. A função é localmente convexa em um ponto se, e somente se,

$$\mathbf{p}^T \mathbf{H} \mathbf{p} \geq 0 \quad (2.5)$$

neste ponto qualquer que seja a direção $\mathbf{p}$. Em particular, tomando $\mathbf{p}$ como um autovetor, a Equação (2.5) mostra que todos os autovalores de $\mathbf{H}$ são não negativos.

Caso todos os autovalores de $\mathbf{H}$ sejam estritamente positivos, a função f é estritamente convexa. Se em um ponto $\bar{\mathbf{x}}$ a função objetivo é estacionária, sua segunda derivada é contínua e a função é estritamente convexa, então o ponto é um minimizador local.

2.2 Considerações sobre algumas funções especiais

Certas funções têm propriedades que permitem caracterizar o mínimo apenas pela Equação (2.4). É o caso, por exemplo, de funções convexas. Uma função convexa com a segunda derivada contínua tem hessiana necessariamente positiva definida em todo ponto. A partir disso demonstra-se que se existe um ponto crítico, então ele é único, e é o minimizador global. O problema de otimização neste caso se torna um sistema de equações, resumido em forma vetorial na Equação (2.4). Outras funções simples com propriedades interessantes são as polinomiais, que são descritas a seguir.

2.2.1 Funções lineares

Uma função linear genérica pode ser escrita como

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n (a_i x_i) + b, \quad (2.6)$$

uma combinação linear das variáveis x_i com coeficientes constantes a_i com adição de uma constante b .

As componentes da primeira derivada da forma linear são

$$\frac{\partial f}{\partial x_j} = \sum_{i=1}^n \left(a_i \frac{\partial x_i}{\partial x_j} \right). \quad (2.7)$$

Inserindo a notação do delta de Kronecker

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases} \quad (2.8)$$

na Equação (2.7), temos

$$\frac{\partial f}{\partial x_j} = \sum_{i=1}^n (a_i \delta_{ij}). \quad (2.9)$$

O delta de Kronecker funciona como um seletor quando aparece em uma soma como a anterior, pois a única parcela não nula ocorre quando $i = j$. Assim, a soma é

$$\frac{\partial f}{\partial x_j} = a_j, \quad (2.10)$$

e o gradiente é escrito na forma vetorial

$$\nabla f = \mathbf{a}, \quad (2.11)$$

onde o vetor $\mathbf{a}$ tem as constantes a_i como componentes.

A menos que todos os a_i sejam nulos, caso em que a função seria constante, não há ponto crítico, portanto, não há mínimos locais em qualquer subconjunto aberto do domínio de otimização. Qualquer mínimo de uma função objetivo linear deve pertencer à fronteira do domínio.

A otimização de funções lineares só faz sentido sob restrições.

2.2.2 Funções quadráticas

Se f é quadrática nas variáveis de otimização, então a Equação (2.4) é um sistema linear de equações com igual número de equações e variáveis escrito em forma vetorial. A este sistema existe associada uma matriz quadrada $\mathbf{A}$ de valores constantes de ordem igual ao número de variáveis.

De fato, podemos escrever uma função quadrática qualquer de n variáveis como

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{2} A_{ij} x_i x_j \right) + \sum_{i=1}^n (-b_i x_i) + c, \quad (2.12)$$

onde A_{ij} , b_i e c são números reais, e vale a simetria $A_{ij} = A_{ji}$.

A derivada parcial de f com relação a uma variável x_k genérica nos dá as componentes do vetor gradiente

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{2} A_{ij} \left(\frac{\partial x_i}{\partial x_k} x_j + x_i \frac{\partial x_j}{\partial x_k} \right) \right) + \sum_{i=1}^n \left(-b_i \frac{\partial x_i}{\partial x_k} \right). \quad (2.13)$$

Considerando o delta de Kronecker e separando a primeira soma em duas, a Equação (2.13) é reescrita como

$$\begin{aligned} \frac{\partial f}{\partial x_k} &= \sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{2} A_{ij} \delta_{ik} x_j \right) + \sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{2} A_{ij} x_i \delta_{jk} \right) + \sum_{i=1}^n (-b_i \delta_{ik}) = \\ &= \frac{1}{2} \sum_{j=1}^n (A_{kj} x_j) + \frac{1}{2} \sum_{i=1}^n (A_{ik} x_i) - b_k. \end{aligned} \quad (2.14)$$

Como os índices i e j das somas são mudos, podemos combiná-las em uma única soma sobre um índice comum. Escolhendo o índice i , a combinação resulta em

$$\frac{\partial f}{\partial x_k} = \frac{1}{2} \sum_{i=1}^n ((A_{ki} + A_{ik}) x_i) - b_k. \quad (2.15)$$

Devido à simetria, $A_{ki} + A_{ik} = 2A_{ki}$, portanto

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^n (A_{ki} x_i) - b_k, \quad (2.16)$$

os termos quadráticos de f quando derivados dão origem a uma soma que é linear nas variáveis de otimização, e os termos lineares resultam em valores constantes que formam um vetor denotado por $\mathbf{b}$.

Em notação matricial, organizamos as variáveis em um vetor coluna

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad (2.17)$$

as componentes A_{ij} formam a matriz quadrada $\mathbf{A}$, e com isso escrevemos a função objetivo como

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{x}^T \mathbf{b} + c, \quad (2.18)$$

onde $(\cdot)^T$ é a operação de transposição.

O vetor gradiente expresso na Equação (2.16) tem a forma matricial

$$\nabla f(\mathbf{x}) = \mathbf{A} \mathbf{x} - \mathbf{b}, \quad (2.19)$$

então o ponto crítico é solução do sistema linear representado por

$$\mathbf{A} \mathbf{x} = \mathbf{b}, \quad (2.20)$$

obtido por anulação do gradiente.

Se os autovalores de $\mathbf{A}$ são todos positivos, então a matriz é positiva definida e a função objetivo é convexa. O ponto crítico é único e o minimizador global.

Outras formas de função f , exceto pelos casos trivial $f = \text{constante}$ ou f linear, resultam em um sistema de equações não lineares.

2.3 O método de Newton-Raphson

É atribuído a Isaac Newton e Joseph Raphson o desenvolvimento de um método baseado em linearização para estimar zeros de uma função. Considerando uma função de uma variável a valores reais $F: \mathbb{R} \rightarrow \mathbb{R}$, um zero de F é um número $\bar{x}$ raiz da equação

$$F(\bar{x}) = 0. \quad (2.21)$$

Se F for uma função linear $F(x) = ax + b$, com $a \neq 0$ e b constantes reais, então a raiz é diretamente calculada pela expressão

$$\bar{x} = -\frac{b}{a}. \quad (2.22)$$

Para resolver uma função genérica, o método faz uma linearização da função. A partir de uma estimativa inicial x_0 da raiz, a função é aproximada por sua reta tangente

$$\tilde{F}(x) = F'(x_0)(x - x_0) + F(x_0). \quad (2.23)$$

O zero da função linearizada é diretamente obtido pela relação

$$x_1 - x_0 = -\frac{F(x_0)}{F'(x_0)}, \quad (2.24)$$

desde que $F'(x_0) \neq 0$, o ponto x_1 é tomado como estimativa seguinte e o processo se repete.

A Figura 14 ilustra duas iterações do método. A reta tangente no ponto $(x_0, F(x_0))$, referente à primeira estimativa, intercepta o eixo x na estimativa seguinte x_1 . A repetição do procedimento leva à estimativa x_2 da raiz.

Figura 14 – Método de Newton-Raphson ilustrado com duas iterações partindo da estimativa inicial x_0

Fonte: autoria própria

O algoritmo produz uma sequência de estimativas $\{x_0, x_1, x_2, \dots, x_n\}$ da qual se espera a convergência para a raiz $\bar{x}$. A convergência do método de Newton-Raphson é quadrática e garantida quando certas condições são atendidas pela função.

Um critério de parada deve ser adotado para encerrar o algoritmo, uma possibilidade é a avaliação de $|F(x_n)|$, se for menor que uma dada tolerância, ou seja, o valor da função é próximo o suficiente de zero, o processo para.

O método pode ser generalizado para funções vetoriais de múltiplas variáveis. Considere uma função vetorial $\mathbf{F}: \mathbb{R}^n \rightarrow \mathbb{R}^n$, a função pode ser separada em n funções componentes de n variáveis $F_i: \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, 2, \dots, n$, um zero de $\mathbf{F}$ é um vetor $\bar{x}$ tal que

$$\begin{cases} F_1(\bar{\mathbf{x}}) = 0 \\ F_2(\bar{\mathbf{x}}) = 0 \\ \vdots \\ F_n(\bar{\mathbf{x}}) = 0 \end{cases} . \quad (2.25)$$

O vetor $\bar{\mathbf{x}}$ é a solução de um sistema possivelmente não linear de n equações, escrito na forma vetorial como

$$\mathbf{F}(\bar{\mathbf{x}}) = \mathbf{0} , \quad (2.26)$$

análoga à vista na Equação (2.21) para o caso de uma função de uma variável.

A linearização da Equação (2.26) em torno de uma estimativa $\mathbf{x}_k$ resulta em

$$\tilde{\mathbf{F}}(\mathbf{x}) = \nabla \mathbf{F}_k(\mathbf{x} - \mathbf{x}_k) + \mathbf{F}_k , \quad (2.27)$$

onde $\nabla \mathbf{F}_k$ denota a derivada de $\mathbf{F}$ com relação ao vetor de variáveis $\mathbf{x}$ calculada no ponto $\mathbf{x}_k$, que pode ser escrita como uma matriz chamada de jacobiana, que será denotada por $\mathbf{H}_k$, e $\mathbf{F}_k$ denota a função $\mathbf{F}$ avaliada no ponto $\mathbf{x}_k$.

A estimativa seguinte é o zero da função linearizada, ou seja,

$$\tilde{\mathbf{F}}(\mathbf{x}_{k+1}) = \mathbf{0} . \quad (2.28)$$

Substituindo $\mathbf{x}_{k+1}$ na Equação (2.27) e escrevendo $\mathbf{x}_{k+1} - \mathbf{x}_k = \Delta \mathbf{x}_k$, o passo da iteração k , temos que este é a solução do sistema linear de equações

$$\mathbf{H}_k \Delta \mathbf{x}_k = -\mathbf{F}_k , \quad (2.29)$$

representado em forma matricial. Portanto, em cada passo do método de Newton-Raphson, um sistema linear de equações deve ser resolvido, o que significa que a matriz $\mathbf{H}_k$ deve ser inversível.

As propriedades de convergência no caso de uma função de uma variável são generalizáveis para n funções de n variáveis quando feitas as devidas analogias.

2.4 Aplicação ao problema de otimização

Um ponto crítico da função f é definido pela Equação (2.4), que a depender da forma da função objetivo representa um sistema não linear de equações. O método de Newton-Raphson pode ser empregado na busca do ponto crítico fazendo a associação $\mathbf{F} = \nabla f$. Neste caso, as funções componentes F_i são as derivadas parciais de f

$$F_i = \frac{\partial f}{\partial x_i}. \quad (2.30)$$

Em certa iteração k , calculamos a linearização da função vetor gradiente em torno da estimativa atual $\mathbf{x}_k$ do ponto crítico

$$\widetilde{\nabla} f(\mathbf{x}) = \nabla^2 f_k(\mathbf{x} - \mathbf{x}_k) + \nabla f_k \quad (2.31)$$

onde surgem (calculadas no ponto $\mathbf{x}_k$) o gradiente ∇f e a segunda derivada da função objetivo $\nabla^2 f$, a matriz hessiana $\mathbf{H}$, que é a matriz jacobiana do vetor gradiente, cujas componentes são as derivadas parciais de segunda ordem de f

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}. \quad (2.32)$$

Se as derivadas parciais são contínuas, então a matriz hessiana é simétrica, segundo o teorema de Clairaut-Schwarz.

O passo de iteração é a solução do sistema linear

$$\mathbf{H}_k \Delta \mathbf{x}_k = -\nabla f_k. \quad (2.33)$$

Assim, a estimativa seguinte é dada pela expressão

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_k^{-1} \nabla f_k. \quad (2.34)$$

Vimos que se f é uma função quadrática, então o ponto crítico é a solução de um sistema linear de equações. Para funções objetivo mais gerais, o sistema de equações pode ser não linear, todavia, o método de Newton-Raphson transforma o problema em uma sequência de sistemas lineares de equações. Concluimos que em todo caso a resolução de sistemas lineares, que é equivalente ao problema de inversão matricial, é um ponto crucial. A literatura sobre esta questão é vasta, e embora seja de grande importância, não será explorada neste trabalho.

A comparação das Equações (2.19) e (2.31) nos leva a uma noção que será essencial posteriormente. A linearização do vetor gradiente em torno da estimativa $\mathbf{x}_k$ é equivalente a aproximar a função objetivo pela forma quadrática presente na Equação (2.18), desde que se façam as substituições $\mathbf{A} \leftarrow \mathbf{H}_k$, $\mathbf{b} \leftarrow -\nabla f_k$, $\mathbf{x} \leftarrow \mathbf{x} - \mathbf{x}_k$, e $c \leftarrow f_k$. Assim, temos o modelo quadrático para a função objetivo

$$\tilde{f}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_k)^T \mathbf{H}_k (\mathbf{x} - \mathbf{x}_k) + (\mathbf{x} - \mathbf{x}_k)^T \nabla f_k + f_k, \quad (2.35)$$

que é na verdade a aproximação obtida truncando a expansão por polinômio de Taylor no termo de segunda ordem.

2.5 O método de Gauss-Newton

O método de Gauss-Newton é uma modificação do método de Newton-Raphson. Na essência, este método trata de uma aproximação da hessiana que é adequada a problemas de mínimos quadrados. A função objetivo deve ser escrita como uma soma de quadrados

$$f(x_1, x_2, \dots, x_n) = \frac{1}{2} \sum_{i=1}^m [r_i(x_1, x_2, \dots, x_n)]^2, \quad (2.36)$$

em que $r_i: \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, 2, \dots, m$, são m funções de n variáveis $x_j, j = 1, 2, \dots, n$, denominadas resíduos.

A primeira derivada de f com relação ao vetor de variáveis resulta no vetor gradiente, que tem componentes

$$\frac{\partial f}{\partial x_j} = \sum_{i=1}^m r_i \frac{\partial r_i}{\partial x_j}. \quad (2.37)$$

A segunda derivada de f com relação ao vetor de variáveis resulta na matriz hessiana, com componentes

$$\frac{\partial^2 f}{\partial x_j \partial x_k} = \sum_{i=1}^m \left(\frac{\partial r_i}{\partial x_j} \frac{\partial r_i}{\partial x_k} + r_i \frac{\partial^2 r_i}{\partial x_j \partial x_k} \right). \quad (2.38)$$

A proposta do método é simplificar as componentes da hessiana eliminando a necessidade das derivadas de segunda ordem dos resíduos. O segundo termo das parcelas da soma é considerado desprezível frente ao primeiro

$$\left| r_i \frac{\partial^2 r_i}{\partial x_j \partial x_k} \right| \ll \left| \frac{\partial r_i}{\partial x_j} \frac{\partial r_i}{\partial x_k} \right|, \quad (2.39)$$

com isso as componentes da hessiana aproximada são

$$B_{jk} = \sum_{i=1}^m \frac{\partial r_i}{\partial x_j} \frac{\partial r_i}{\partial x_k}. \quad (2.40)$$

Em problemas de mínimos quadrados ligados a ciências e engenharia, é típico que as funções de resíduo sejam associadas a erros de medição, que estão supostamente distribuídos

em torno de zero. Assim, nas proximidades da solução, é esperado que as funções de resíduo tenham valor quase nulo. Tal condição é particularmente bem atendida em problemas de reconstrução geométrica, onde as distâncias das amostras até a curva de aproximação são limitadas por um pequeno valor de tolerância. A aproximação é reforçada se as funções de resíduo forem quase lineares, isto é, se as derivadas de segunda ordem forem de pequena magnitude, na vizinhança da solução.

Em notação matricial, os resíduos são organizados em um vetor coluna

$$\mathbf{r} = \begin{bmatrix} r_1(\mathbf{x}) \\ r_2(\mathbf{x}) \\ \vdots \\ r_m(\mathbf{x}) \end{bmatrix}. \quad (2.41)$$

A função objetivo e o vetor gradiente são respectivamente escritos como

$$f(\mathbf{x}) = \frac{1}{2}(\mathbf{r}^T \mathbf{r}) \quad (2.42)$$

e

$$\nabla f = \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_2}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_1} \\ \frac{\partial r_1}{\partial x_2} & \frac{\partial r_2}{\partial x_2} & \cdots & \frac{\partial r_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_1}{\partial x_n} & \frac{\partial r_2}{\partial x_n} & \cdots & \frac{\partial r_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} = \mathbf{J}^T \mathbf{r}, \quad (2.43)$$

onde

$$\mathbf{J} = \nabla \mathbf{r} = \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} & \cdots & \frac{\partial r_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial x_1} & \frac{\partial r_m}{\partial x_2} & \cdots & \frac{\partial r_m}{\partial x_n} \end{bmatrix} \quad (2.44)$$

é o gradiente de $\mathbf{r}(\mathbf{x})$ ou matriz jacobiana da transformação $\mathbf{r}(\mathbf{x})$.

A matriz hessiana aproximada $\mathbf{B}$, de acordo com a Equação (2.40), é obtida pelo produto matricial

$$\mathbf{B} = \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_2}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_1} \\ \frac{\partial r_1}{\partial x_2} & \frac{\partial r_2}{\partial x_2} & \cdots & \frac{\partial r_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_1}{\partial x_n} & \frac{\partial r_2}{\partial x_n} & \cdots & \frac{\partial r_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} & \cdots & \frac{\partial r_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial x_1} & \frac{\partial r_m}{\partial x_2} & \cdots & \frac{\partial r_m}{\partial x_n} \end{bmatrix} = \mathbf{J}^T \mathbf{J}. \quad (2.45)$$

Dada uma aproximação do ponto crítico $\mathbf{x}_k$, a Equação (2.34) é então empregada para atualizar a aproximação trocando a matriz hessiana exata $\mathbf{H}$ pela matriz aproximada $\mathbf{B}$:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{B}^{-1} \nabla f, \quad (2.46)$$

onde $\mathbf{x}_{k+1}$ é a aproximação atualizada a partir do ponto $\mathbf{x}_k$, e $\mathbf{B}$ e ∇f são calculados em $\mathbf{x}_k$.

Substituindo as expressões nas Equações (2.43) e (2.45), escrevemos uma fórmula para a atualização da estimativa em termos apenas das funções resíduo e suas derivadas:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \mathbf{r}. \quad (2.47)$$

2.6 Métodos *trust-region* para otimização

Os métodos vistos até então não garantem a convergência para um minimizador. O método de Newton-Raphson garante convergência quadrática para um ponto crítico desde que a estimativa inicial esteja suficientemente próxima da solução. Já o método de Gauss-Newton tem no máximo convergência quadrática por ser uma aproximação do método de Newton-Raphson, e esta não é garantida dependendo da qualidade da aproximação da hessiana. Em todo caso, ambos os métodos se baseiam em aproximações da função objetivo, o primeiro trata de uma aproximação quadrática da função objetivo, o segundo dá um passo além e troca a hessiana exata $\mathbf{H}$ do modelo quadrático por uma aproximação $\mathbf{B}$. Veremos agora a abordagem da região de confiança (*trust-region*), que tem como motivação reduzir os eventuais impactos negativos causados pelas aproximações e garantir convergência dos algoritmos.

Na aplicação do método de Newton-Raphson ao problema de otimização, identificamos a equivalência entre a linearização do vetor gradiente e a aproximação da função objetivo por uma função modelo quadrática. Em cada iteração, uma aproximação da função objetivo é feita em torno da estimativa atual do ponto minimizador pela forma quadrática

$$\tilde{f}(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{H}_0 \mathbf{x} + \mathbf{x}^T \nabla f_0 + f_0 . \quad (2.48)$$

Tal aproximação polinomial equivale aos três primeiros termos da expansão polinomial de f em torno da estimativa, que é tomada como origem de um sistema de coordenadas local.

Podemos escrever por expansão em polinômio a função f como

$$f(\mathbf{x}) = \tilde{f}(\mathbf{x}) + o(\mathbf{x}^3) , \quad (2.49)$$

a aproximação de f por $\tilde{f}$ é local, pois existem termos de ordem superior, representados pela parcela $o(\mathbf{x}^3)$, que podem se tornar importantes à medida que $\mathbf{x}$ se afasta da estimativa atual, além disso, a expansão polinomial tem um limite de validade para a igualdade, chamado de raio de convergência. Portanto, existe uma região de confiança em torno da estimativa em que a função modelo quadrática é aceitável. Baseados nisso, surgem os métodos *trust-region*, que limitam a atualização do estimador a permanecer dentro da região de confiança.

2.6.1 O problema local de otimização

A abordagem *trust-region* transforma o problema de otimização em uma sequência de problemas locais. Em cada problema local de otimização, é minimizada uma aproximação da função objetivo dentro de uma região restrita.

Considere a função modelo

$$m(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{B} \mathbf{x} + \mathbf{x}^T \nabla f , \quad (2.50)$$

onde o estimador atual é tomado como origem de um sistema de coordenadas local, $\mathbf{B}$ é uma matriz simétrica para aproximar a hessiana da função objetivo e o gradiente ∇f é calculado na origem. O termo constante f_0 presente na Equação (2.48) foi desconsiderado, pois não influencia a busca do minimizador e medidas de redução.

O domínio do modelo é limitado a uma bola fechada de raio Δ , a região de confiança. Consequentemente, existe a minimização da função modelo está sujeita à restrição $\|\mathbf{x}\| \leq \Delta$. Assim, dados uma matriz simétrica $\mathbf{B}$, um vetor ∇f e um número Δ , temos o problema local de otimização, definido com uma restrição:

$$\min \left(\frac{1}{2} \mathbf{x}^T \mathbf{B} \mathbf{x} + \mathbf{x}^T \nabla f \right) \text{ t. q. } \|\mathbf{x}\| \leq \Delta . \quad (2.51)$$

Sendo encontrada uma solução, é necessário verificar a aceitabilidade da função modelo dentro da região de confiança. Isto é feito comparando a redução do modelo com a redução efetiva da função objetivo. Seja $\bar{\mathbf{x}}$ a solução do problema local, definimos como redução relativa à razão

$$\rho = \frac{f_0 - f(\bar{\mathbf{x}})}{m_0 - m(\bar{\mathbf{x}})} = \frac{f(\bar{\mathbf{x}}) - f_0}{m(\bar{\mathbf{x}})}. \quad (2.52)$$

A solução é aceitável quando ρ é próximo da unidade ou maior, o que indica que a redução efetiva é comparável à do modelo ou melhor. Por outro lado, se ρ for próximo de zero ou negativo, a redução efetiva é desprezível comparada à do modelo ou pior, não há redução. Neste último caso, a bola considerada como região de confiança é muito abrangente, a função modelo deixa de representar bem a função objetivo antes de atingir o ponto $\bar{\mathbf{x}}$, portanto o raio Δ deve ser reduzido a um valor menor que $\|\bar{\mathbf{x}}\|$.

Com uma nova região de confiança, uma nova solução deve ser encontrada e a verificação feita. O problema local é finalizado quando uma solução aceitável é determinada, esta então assume a posição de estimador atual.

2.6.2 O passo de Cauchy

A Equação (2.51) traz um problema de otimização com uma restrição de desigualdade. Se a matriz hessiana $\mathbf{B}$ é positiva definida, então o ponto crítico da função modelo é o seu minimizador, caso o minimizador esteja dentro da região de confiança, este é a solução exata do problema, chamada de passo completo. Em outras condições, a solução exata não é tão óbvia, porém, ela não é necessária.

É denominado passo de Cauchy o ponto minimizador da função modelo restrito à região de confiança e ao raio na direção oposta do gradiente. Pode-se escrever

$$\mathbf{p}^c = -\tau \Delta \frac{\nabla f}{\|\nabla f\|}, \quad (2.53)$$

o ponto permanece dentro da região de confiança se, e somente se, $0 < \tau \leq 1$.

Substituindo na Equação (2.50), obtemos m como uma função quadrática τ

$$m(\tau) = \frac{1}{2} \frac{\Delta^2}{\|\nabla f\|^2} (\nabla f^T \mathbf{B} \nabla f) \tau^2 - \Delta \|\nabla f\| \tau. \quad (2.54)$$

Se $\nabla f^T \mathbf{B} \nabla f \leq 0$, então o gráfico de $m(\tau)$ é uma parábola com a concavidade voltada para baixo ou uma reta, em todo caso, o polinômio é monotônico decrescente para $\tau > 0$,

portanto, seu mínimo ocorre no valor limite $\tau = 1$. Caso contrário, o gráfico é uma parábola com a concavidade para cima, o mínimo ocorre no vértice

$$\tau^* = \frac{\|\nabla f\|^3}{\Delta(\nabla f^T \mathbf{B} \nabla f)}, \quad (2.55)$$

além disso, o polinômio é monotônico decrescente para $0 < \tau < \tau^*$ e o mínimo restrito é o menor número entre este valor e 1.

Em conclusão,

$$\tau = \begin{cases} 1, & \text{se } \nabla f^T \mathbf{B} \nabla f \leq 0 \\ \min\{\tau^*, 1\}, & \text{caso contrário} \end{cases} \quad (2.56)$$

substituído em (2.53) resulta no ponto de Cauchy.

A principal propriedade do ponto de Cauchy é a garantia de convergência global dos algoritmos *trust-region* baseados neste passo, por isso a solução exata do problema local de otimização não é necessária para que o método funcione.

Um passo que produza maior redução da função modelo do que a proporcionada pelo passo de Cauchy, também garante o funcionamento. O passo completo é uma alternativa quando a hessiana é positiva definida, além disso, existem métodos que combinam o passo de Cauchy com o passo completo, e métodos que alteram a hessiana de forma a deixá-la positiva definida. [34, 35]

3 Aproximação de uma curva amostrada por uma spline de arcos

As splines de arcos são capazes tanto de aproximar curvas paramétricas quanto dados discretos. Meek e Walton [36] descreveram um método de aproximação de dados discretos com arcos circulares. Pei e Horng [37] e Yang e Du [38] aplicaram arcos à reconstrução de curvas digitais, as quais têm uma natureza ruidosa e discreta. Maier [39] apresentou um método para aproximar uma sequência de pontos usando a quantidade ótima de arcos. Os métodos citados têm um forte caráter geométrico.

Há a possibilidade de se seguir uma via mais algébrica associando métodos de otimização ao ajuste de curvas com splines de arcos. Ong *et al.* [40], por exemplo, empregaram um método de minimização da área calculada entre uma curva original e a spline de aproximação. Parkinson e Moreton [5] utilizaram minimização da “energia” (integral da curvatura ao quadrado) da curva em um método de interpolação para obter o resultado menos “sinuoso” possível. Neste capítulo, é proposto um método de aproximação de dados discretos por splines de arcos através de mínimos quadrados.

Dada uma sequência de pares de coordenadas (x, y) de um conjunto de pontos coplanares amostrados de uma curva suave, queremos construir uma spline de arcos que suficientemente aproxime os dados. A aproximação é aceitável se a maior distância da curva às amostras for menor que um valor de tolerância.

3.1 Aproximação de propriedades geométricas a partir das amostras

A partir da lista de pontos é possível estimar uma parametrização por comprimento de arco, o que resulta em um conjunto de amostras aproximadas $(s, \gamma(s))$. A parametrização aproximada por comprimento de arco será utilizada na estimativa de outras propriedades da curva original. Seja por métodos de diferença ou por aproximação de funções, as derivadas $\gamma'(s)$ e $\gamma''(s)$ podem ser estimadas, o que nos dá o vetor tangente e a curvatura em cada amostra. Estas estimativas serão utilizadas na determinação de uma primeira aproximação às amostras baseada na curvatura. As posições das amostras são posteriormente utilizadas em um método de otimização para ajuste da curva.

3.1.1 Comprimento de arco

A primeira medida que precisamos calcular a partir das amostras é o comprimento de arco, o que resulta em uma aproximação para a parametrização natural da curva original. Dados um ponto P_1 e seu subsequente P_2 , o comprimento do arco entre os dois pontos é estimado como

$$\Delta s = \|P_2 - P_1\|, \quad (3.1)$$

isto é, toma-se o comprimento do segmento reto entre os pontos P_1 e P_2 , a qualidade desta estimativa depende da proximidade dos pontos.

Em termos de coordenadas cartesianas, temos a forma pitagórica

$$\Delta s = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \quad (3.2)$$

Supondo que s_1 seja o parâmetro comprimento de arco associado ao ponto P_1 , o parâmetro associado ao ponto P_2 é

$$s_2 = s_1 + \Delta s. \quad (3.3)$$

Assim, partindo de $s_0 = 0$ para o primeiro ponto do conjunto, podemos determinar o parâmetro s de cada ponto recursivamente conforme

$$\begin{cases} s_i = s_{i-1} + \|P_i - P_{i-1}\|, & \text{if } i > 0 \\ s_0 = 0 \end{cases}. \quad (3.4)$$

A Equação (3.4) é de natureza interpolatória e superestima o comprimento de arco devido ao ruído presente nas amostras. Por outro lado, a aproximação do trecho de curva entre pontos vizinhos por um segmento de reta subestima o comprimento de arco. Nos casos testados, o ruído é pequeno suficiente para não prejudicar a estimativa. Em casos de ruído mais severo ou amostras muito afastadas, recomenda-se a utilização de outro estimador. Alguns estimadores tanto de comprimento de arco quanto tangente e curvatura podem ser encontrados em [36, 41, 42, 43].

3.1.2 Ângulo tangente de partida

A mais importante quantidade empregada no método é a curvatura, ainda assim, há a necessidade de uma orientação para posicionar a curva de aproximação. O primeiro ponto da aproximação será coincidente com a primeira amostra e, a princípio, o vetor tangente inicial será estimado a partir das amostras.

O vetor tangente em um dado ponto P_1 pode ser estimado pela diferença entre este e os vizinhos P_0 e P_2 , respectivamente, o anterior e o posterior. A estimativa pode ser por diferença progressiva

$$\mathbf{T}_1^+ = \frac{P_2 - P_1}{s_2 - s_1}, \quad (3.5)$$

relacionando o ponto de interesse com o posterior, ou por diferença regressiva

$$\mathbf{T}_1^- = \frac{P_1 - P_0}{s_1 - s_0}, \quad (3.6)$$

relacionando com o anterior. Como o ponto de interesse P_1 será a primeira amostra, não existe ponto anterior e a estimativa por avanço $\mathbf{T}_1^+$ é considerada.

O ângulo de orientação do vetor tangente estimado é o medido a partir do eixo x no sentido anti-horário até o vetor, que pode ser calculado como

$$\theta_1 = \text{atan}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) + \phi = \text{atan}\left(\frac{\Delta y}{\Delta x}\right) + \phi, \quad (3.7)$$

se $\Delta x \neq 0$, onde

$$\phi = \begin{cases} \pi, & \text{se } \Delta x < 0 \\ 0, & \text{se } \Delta x > 0 \end{cases} \quad (3.8)$$

corrige θ_1 para o segundo ou terceiro quadrante caso Δx seja negativo, já que os valores da função $\text{atan}(x)$ estão restritos ao conjunto imagem $(-\pi/2, \pi/2)$.

Se $\Delta x = 0$, então

$$\theta_1 = \pm \frac{\pi}{2}, \quad (3.9)$$

sendo o sinal idêntico ao de Δy .

3.1.3 Curvatura orientada

A curvatura como taxa de variação da direção tangente demanda pelo menos três pontos para ser estimada.

Método de interpolação quadrática

O método estima a curvatura em um ponto relacionando sua posição com as posições dos vizinhos, por esta razão, não é possível aplicá-lo às extremidades.

Dado o ponto intermediário P_1 e seus vizinhos P_0 e P_2 , são conhecidas as coordenadas x e y , e uma estimativa para o comprimento de arco s . A coordenada x das amostras é uma função do comprimento de arco $x = x(s)$. Podemos relacionar as coordenadas dos três pontos com as derivadas da função $x(s)$ por uma expansão polinomial

$$x(s_2) \approx x(s_1) + (s_2 - s_1)x'(s_1) + \frac{(s_2 - s_1)^2}{2}x''(s_1) \quad (3.10)$$

e

$$x(s_0) \approx x(s_1) + (s_0 - s_1)x'(s_1) + \frac{(s_0 - s_1)^2}{2}x''(s_1). \quad (3.11)$$

Multiplicando a Equação (3.10) por $l_1 = s_1 - s_0$, o comprimento entre P_0 e P_1 , e a Equação (3.11) por $l_2 = s_2 - s_1$, o comprimento entre P_1 e P_2 , e somando ambas as expressões, pode-se obter

$$l_1x_2 + l_1x_0 = (l_1 + l_2)x_1 + \frac{l_1l_2(l_1 + l_2)}{2}x_1'', \quad (3.12)$$

que resulta na estimativa

$$x_1'' = 2 \frac{l_1x_2 + l_2x_0 - (l_1 + l_2)x_1}{l_1l_2(l_1 + l_2)} \quad (3.13)$$

para a segunda derivada da função coordenada x .

Uma estimativa para a primeira derivada também pode ser obtida do par de Equações (3.10) e (3.11) ao se multiplicar a primeira por l_1^2 e a última por $-l_2^2$ e somando, o resultado é

$$x_1' = \frac{l_1^2(x_2 - x_1) - l_2^2(x_0 - x_1)}{l_1l_2(l_1 + l_2)}. \quad (3.14)$$

As expressões em (3.13) e (3.14) são conhecidas como aproximações por diferenças finitas das derivadas. Quando $l_1 = l_2 = l$ são obtidas as expressões tradicionais

$$x_1' = \frac{x_2 - x_0}{2l}, \quad (3.15)$$

chamada de diferença central, compare com as Equações (3.5) e (3.6), e

$$x_1'' = \frac{x_2 - 2x_1 + x_0}{l^2}. \quad (3.16)$$

Como são decorrentes da aproximação local da função $x(s)$ por truncamento do polinômio de Taylor, é possível avaliar a ordem do erro cometido. As expressões são análogas para a coordenada y .

Interessante de se ressaltar é o fato que as estimativas presentes nas Equações (3.12) e (3.13) produzem uma parábola que interpola os três pontos (s_0, x_0) , (s_1, x_1) e (s_2, x_2) . O mesmo ocorre para a função coordenada y .

A curvatura é calculada pela Equação (1.59) repetida abaixo

$$k_1 = \frac{x'_1 y''_1 - y'_1 x''_1}{((x'_1)^2 + (y'_1)^2)^{3/2}}. \quad (3.17)$$

Método de regressão quadrática

O método anterior, que usa interpolação de três pontos por uma parábola, é suscetível a imperfeições causadas por ruído nas amostras. Os efeitos de ruído tendem a ser reduzidos ao se considerar um conjunto maior de amostras e o ajuste de alguma função no lugar da interpolação.

Novamente, é suposto conhecido o valor da função coordenada $x(s)$ nas amostras. Dado um conjunto de m amostras, buscamos uma função quadrática que se aproxime dos pontos medidos, modelando seu comportamento. A função modelo é da forma

$$\tilde{x}(s) = as^2 + bs + c, \quad (3.18)$$

com a , b e c constantes reais.

O problema é conhecido como regressão quadrática e a solução é tradicionalmente obtida pelo método dos mínimos quadrados, no qual se minimiza a função objetivo

$$f(a, b, c) = \frac{1}{2} \sum_{i=1}^m (x(s_i) - \tilde{x}(s_i))^2 = \frac{1}{2} \sum_{i=1}^m (x(s_i) - (as_i^2 + bs_i + c))^2, \quad (3.19)$$

que é uma função dos parâmetros do modelo $\tilde{x}$. A estimativa de quantidades diferenciais baseada em mínimos quadrados tem um efeito filtrante inerente, que lida bem com o ruído que se espera que amostras medidas tenham [44]. A regressão é tão mais filtrante quanto maior o número de pontos m escolhido.

Neste problema, os resíduos $r_i = x(s_i) - \tilde{x}(s_i)$ são lineares nos parâmetros e a função objetivo é quadrática, portanto, os parâmetros ótimos verificam a anulação do gradiente

$$\frac{\partial f}{\partial a} = \frac{\partial f}{\partial b} = \frac{\partial f}{\partial c} = 0. \quad (3.20)$$

A i -ésima linha da matriz jacobiana expressa na Equação (2.44) tem a forma

$$\mathbf{J}_i = \begin{bmatrix} \frac{\partial r_i}{\partial a} & \frac{\partial r_i}{\partial b} & \frac{\partial r_i}{\partial c} \end{bmatrix} = -[s_i^2 \quad s_i \quad 1], \quad (3.21)$$

o que resulta na matriz hessiana constante

$$\mathbf{B} = \mathbf{J}^T \mathbf{J} = \begin{bmatrix} s_1^2 & s_2^2 & \dots & s_m^2 \\ s_1 & s_2 & \dots & s_m \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} s_1^2 & s_1 & 1 \\ s_2^2 & s_2 & 1 \\ \vdots & \vdots & \vdots \\ s_m^2 & s_m & 1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m s_i^4 & \sum_{i=1}^m s_i^3 & \sum_{i=1}^m s_i^2 \\ \sum_{i=1}^m s_i^3 & \sum_{i=1}^m s_i^2 & \sum_{i=1}^m s_i \\ \sum_{i=1}^m s_i^2 & \sum_{i=1}^m s_i & m \end{bmatrix}, \quad (3.22)$$

conforme a Equação (2.45), e no vetor gradiente

$$\nabla f = \mathbf{J}^T \mathbf{r} = - \begin{bmatrix} s_1^2 & s_2^2 & \dots & s_m^2 \\ s_1 & s_2 & \dots & s_m \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} = - \begin{bmatrix} \sum_{i=1}^m s_i^2 r_i \\ \sum_{i=1}^m s_i r_i \\ \sum_{i=1}^m r_i \end{bmatrix}, \quad (3.23)$$

conforme a Equação (2.43).

Os parâmetros ótimos são solução do sistema linear

$$\mathbf{B}(\mathbf{x} - \mathbf{x}_0) = -\nabla f_0, \quad (3.24)$$

em que $\mathbf{x}$ é o vetor de parâmetros e $\mathbf{x}_0$ um vetor de referência arbitrário, onde o gradiente da função objetivo é calculado. A particular escolha $\mathbf{x}_0 = \mathbf{0}$ implica em $r_i = x(s_i) = x_i$ na expressão do gradiente. O sistema linear final obtido é, em forma matricial,

$$\begin{bmatrix} \sum_{i=1}^m s_i^4 & \sum_{i=1}^m s_i^3 & \sum_{i=1}^m s_i^2 \\ \sum_{i=1}^m s_i^3 & \sum_{i=1}^m s_i^2 & \sum_{i=1}^m s_i \\ \sum_{i=1}^m s_i^2 & \sum_{i=1}^m s_i & m \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m s_i^2 x_i \\ \sum_{i=1}^m s_i x_i \\ \sum_{i=1}^m x_i \end{bmatrix}. \quad (3.25)$$

O mesmo procedimento de regressão é aplicado às coordenadas y do mesmo conjunto de amostras, produzindo duas funções quadráticas de aproximação $\tilde{x}(s)$ e $\tilde{y}(s)$, nas quais é baseado o cálculo da curvatura

$$k_s = \frac{\tilde{x}'(s)\tilde{y}''(s) - \tilde{y}'(s)\tilde{x}''(s)}{\left((\tilde{x}'(s))^2 + (\tilde{y}'(s))^2\right)^{3/2}}. \quad (3.26)$$

A curvatura em um dado ponto P_i é estimada considerando a regressão quadrática a uma vizinhança de amostras do ponto e calculando a curvatura expressa em (3.26) com o valor do comprimento de arco s_i correspondente. Se a vizinhança for formada apenas pelos três pontos P_{i-1} , P_i e P_{i+1} , o resultado será idêntico ao do método anterior, pois a parábola de mínimos quadrados por três pontos é a interpoladora.

A regressão tem capacidade de filtrar ruídos deixando o gráfico de curvatura mais suave. Quanto maior a quantidade de amostras, maior a atenuação do ruído. No entanto, cuidado deve ser tomado para que o número de amostras não seja exageradamente grande, tendo em vista que a regressão quadrática pode perder a representatividade e informação pode ser perdida.

3.2 Arcos de circunferência, biarcos e splines de arcos

Um arco de circunferência é uma curva cujo traço é um segmento extraído de uma circunferência. Dada uma circunferência de raio R e dois pontos não coincidentes, existem dois arcos distintos que têm como extremidades os pontos dados, diremos que estes arcos são complementares.

Figura 15 – Uma circunferência segmentada em dois arcos complementares com extremidades A e B

Fonte: autoria própria

A menos de movimentos de corpo rígido (translação e rotação), um arco fica totalmente determinado por dois valores: o raio da circunferência da qual o arco é extraído, o qual definimos como o raio do arco; e a abertura angular do arco $\Delta\varphi$, o ângulo $\angle ACB$, sendo A e B as extremidades do arco, e C o centro da circunferência, também definido como o centro do arco. O comprimento do arco pode ser dado como alternativa para o ângulo, pois, dado o raio, ambos são dependentes pela relação

$$L = R\Delta\varphi. \quad (3.27)$$

Figura 16 – Arco de circunferência com raio R e comprimento L

Fonte: autoria própria

Como o traço do arco se sobrepõe a uma circunferência, a curvatura em cada ponto é constante e dada por

$$\kappa = \frac{1}{R}, \quad (3.28)$$

a curvatura da circunferência originária, que pode ser dada em lugar do raio como um parâmetro determinante do arco. Um par de parâmetros determina a forma do arco, uma medida de comprimento e uma medida de curvatura.

O posicionamento deste traço em um dado plano depende de mais três parâmetros, dois de posição e um angular. A posição pode ser dada pelas coordenadas de uma extremidade, digamos que A , do arco e a referência angular pode ser dada pela direção da tangente ao arco na mesma extremidade. No entanto, como mostra a Figura 17, existem quatro arcos com

comprimento e curvatura iguais que têm A como extremidade e são tangentes à reta dada em A , a diferença entre eles se encontra na orientação.

Figura 17 – Quatro arcos de mesmo formato com uma extremidade em comum e mesma reta tangente nesta extremidade

Fonte: autoria própria

Devemos informar se a extremidade dada é o ponto de partida ou o ponto de chegada. Parametrizando o arco pelo comprimento, definimos como ponto de partida, ou ponto inicial, o que possui parâmetro $s = 0$, consequentemente, o ponto de chegada, ou ponto final, possui parâmetro $s = L$. Digamos que A seja o ponto de partida, então dos quatro arcos da Figura 17, apenas AB e AC têm o ponto A como ponto inicial.

Figura 18 – Dois arcos espelhados com mesmo ponto de partida e vetor tangente inicial

Fonte: autoria própria

O que diferencia os dois arcos da Figura 18 é a orientação. O arco AB tem curvatura positiva, pois à medida que um ponto progride sobre o arco indo de A para B , o vetor tangente rotaciona no sentido anti-horário. O arco AC , por outro lado, tem curvatura negativa. Assim, não basta conhecer a curvatura, mas também a orientação, ou seja, a curvatura com sinal.

Em conclusão, um arco de circunferência é totalmente determinado por seu ponto de partida, a direção tangente inicial, seu comprimento e sua curvatura com sinal. A partir destas informações é possível calcular o raio do arco, as posições do centro, do ponto final, e pontos intermediários, o vetor tangente final, e o ângulo subentendido.

3.2.1 O conceito de continuidade geométrica

Uma curva é uma função, e como tal é possível caracterizar sua continuidade. Essencialmente, todas as curvas são funções contínuas, isto é, pertencem à classe C^0 . Os traços das curvas não apresentam saltos nem furos.

Uma curva com primeira derivada contínua, ou continuamente derivável, pertence à classe C^1 . A função vetor tangente de curvas C^1 é contínua, o que significa que tanto a direção tangente quanto a magnitude variam suavemente. Uma curva é dita suave quando pertence à classe C^∞ , significando que qualquer derivada é contínua.

No campo do desenho geométrico, mais importante do que a continuidade de uma parametrização é a suavidade da própria imagem. Embora seja frequente a correspondência entre suavidade da parametrização e da imagem traçada, há casos contrários.

Uma curva de classe C^1 não necessariamente tem um traço suave, como por exemplo a curva cicloide. Traçada por um ponto marcado em uma circunferência quando esta rola sem escorregar sobre uma reta, a curva cicloide tem cúspides nas posições em que o ponto marcado toca a reta. O vetor tangente varia continuamente (continuidade C^1), porém, nas cúspides, a direção tangente não pode ser definida, já que o vetor é nulo, e acontece uma inversão do sentido quando considerados os limites laterais (sem continuidade da direção tangente).

Figura 19 – Cicloide $x(t) = t - \sin t, y(t) = 1 - \cos t, 0 \leq t \leq 4\pi$, exemplo de curva suave com uma cúspide na imagem

Fonte: autoria própria

Por outro lado, há casos em que se pode facilmente caracterizar a suavidade da imagem, ainda que não se conheça uma parametrização contínua. É o que acontece com curvas compostas por união de segmentos diversos. Surge o conceito de continuidade geométrica para diferenciar a suavidade da imagem da continuidade paramétrica.

Todas as curvas pertencem à classe G^0 , a primeira ordem de continuidade geométrica, a continuidade de posição. Não é permitido que os traços tenham saltos ou furos, e por isso, os segmentos de uma curva composta devem ser ligados de forma que a extremidade final de um segmento se conecte à extremidade inicial do seguinte. A classe G^1 , primeiro grau ou segunda ordem de continuidade geométrica, engloba as curvas sem cúspides, ou seja, curvas com a

direção tangente contínua. Curvas com vários segmentos são de classe G^1 quando ocorre coincidência das direções tangentes dos pontos conectados. À classe G^2 pertencem as curvas com continuidade de curvatura.

3.2.2 Conexão geometricamente contínua de arcos

É dado um par ordenado de arcos sem posição definida a princípio. Sabe-se qual dos arcos é anterior e são dados apenas comprimento, curvatura e orientação de cada um. Conectando o ponto de chegada do primeiro arco ao ponto de partida do segundo de forma que as direções tangentes na conexão coincidam, uma curva de classe G^1 chamada de biarco é obtida.

A Figura 20 traz dois biarcos como ilustração. O biarco tem formato de C quando as curvaturas têm o mesmo sinal, e formato de S quando os sinais são opostos. Há necessariamente uma descontinuidade de curvatura na conexão, por isso o biarco pertence à classe G^1 .

Figura 20 – Biarco (a) em formato de C e (b) em formato de S

Fonte: autoria própria

No ponto de conexão P_2 , o vetor tangente T_2 coincide com os vetores tangentes dos arcos individuais. Dado o ponto de partida P_1 e a tangente inicial T_1 , o ponto de conexão e sua tangente ficam determinados pelo formato do primeiro arco, estes por sua vez servem de partida para o segundo arco. Assim, a posição do biarco é totalmente determinada por um ponto de partida e uma direção inicial. Uma propriedade que será usada adiante é a colinearidade dos pontos C_1 , C_2 e P_2 .

Biarcos podem ser usados em interpolação, pois sempre existe pelo menos um que tenha como extremidades dois pontos dados. Se além dos pontos, forem dadas as direções tangentes nas extremidades, ainda existe pelo menos um biarco que faça a interpolação respeitando a tangência.

Podemos conectar um terceiro arco a partir do término do segundo, um quarto a partir do terceiro, e assim sucessivamente, tal generalização leva à curva que chamamos de spline de arcos. Uma spline de arcos é uma curva de continuidade G^1 composta por arcos de circunferência. Um exemplo ilustre de spline de arcos é a espiral formada por arcos de abertura $\pi/2$ com raios seguindo a sequência de Fibonacci. Conhecida como espiral de Fibonacci, a curva visualizada na Figura 21 é uma aproximação para a espiral áurea descrita nos exemplos do Capítulo 1.

Figura 21 – Espiral de Fibonacci, spline de arcos cujos raios seguem a sequência de Fibonacci

Fonte: autoria própria

A construção da curva respeita uma ordem. A posição de um dado arco depende do seu antecessor, que, por sua vez, depende do próprio antecessor, portanto, a posição de um arco depende de todos os anteriores ao mesmo, pelo menos indiretamente. Arcos posteriores não influenciam na posição de um dado arco.

Cada arco da spline é descrito por dois números: a curvatura com sinal e o comprimento. Além disso, o formato da curva é rígido, consequência das restrições de continuidade. Portanto, são necessários somente mais três números para determinar a posição da curva no plano. Uma spline de n arcos é determinada por $2n + 3$ parâmetros no total.

O último ponto a ser ressaltado é o formato do gráfico de curvatura de uma spline de arcos. É possível construir uma parametrização por comprimento de arco da curva acumulando o comprimento desde o ponto de partida até o ponto final. No interior de cada segmento a curvatura é constante, e surgem descontinuidades nas posições correspondentes às transições entre arcos.

A Figura 22 mostra um exemplo de spline de arcos formada por quatro arcos de circunferência, o ponto de partida é P_0 , os pontos P_1 , P_2 e P_3 são as transições, e o ponto de chegada é P_4 . O gráfico de curvatura correspondente acompanha a spline, o comprimento de arco no eixo horizontal e a curvatura com sinal no eixo vertical. Para cada arco existe no gráfico

um retângulo, a altura do retângulo equivale à curvatura do arco e a largura, ao comprimento do arco.

Figura 22 – Uma spline de arcos formada por quatro arcos de circunferência e seu gráfico de curvatura correspondente

Fonte: autoria própria

Segundo o teorema fundamental das curvas planas, dada uma função curvatura orientada suave com o comprimento de arco como argumento, existe uma única curva com o perfil de curvatura dado, a menos de movimentos rígidos (translações e rotações). Isso sugere que se encontre uma aproximação por trechos constantes do gráfico de curvatura e que seja tomada a spline de arcos correspondente como primeira curva de ajuste.

3.2.3 Determinação do gráfico de curvatura constante por trechos

Supondo que esteja disponível a função curvatura orientada $k_s: [0, L] \rightarrow \mathbb{R}$, é feita uma partição do domínio $s_0 = 0 < s_1 < \dots < s_i < \dots < s_{n-1} < s_n = L$ em um número n dado de subintervalos $[s_i, s_{i+1}]$, $i = 0, 1, \dots, n-1$.

Dentro de cada subintervalo, um valor representativo $\bar{k}$ para a curvatura é escolhido. A curvatura $\bar{k}$ é tomada como o valor médio da curvatura orientada calculado no intervalo

$$\bar{k} = \frac{1}{b-a} \int_a^b k_s ds. \quad (3.29)$$

O valor médio tem a propriedade de minimizar a quantidade

$$\int_a^b (\bar{k} - k_s)^2 ds, \quad (3.30)$$

a integral do quadrado das diferenças entre a constante $\bar{k}$ e o valor da função k_s . Portanto, dentro de cada subintervalo, a curvatura média é a solução de um problema localizado de mínimos quadrados.

A spline de arcos correspondente deve se ajustar à curva original com erro tão pequeno quanto se queira desde que a partição seja suficientemente refinada. Sabitov e Slovesnov [45] provaram que para uma curva arbitrária de classe C^3 , o erro na posição e na direção tangente da spline tende a zero junto com o comprimento da partição.

Partição ótima

Pei e Horng [37] também constataam que o ajuste com arcos de circunferência é equivalente a produzir uma aproximação por trechos constantes da função de curvatura da curva original. Eles escolhem como pontos de quebra da partição os extremantes da derivada da função de curvatura. A seguir é proposta uma abordagem diferente.

Somando a contribuição de cada parte, temos uma medida total da similaridade entre a aproximação constante por partes e a função de curvatura

$$V(s_1, \dots, s_{N-1}) = \sum_{i=0}^{N-1} \left(\int_{s_i}^{s_{i+1}} (\bar{k}_i - k_s)^2 ds \right). \quad (3.31)$$

Se a partição é fixa, a minimização de V implica na escolha de $\bar{k}_i$ como valor médio no intervalo $[s_i, s_{i+1}]$. Uma redução adicional de V pode ser obtida variando os pontos de quebra da partição $s_i, i = 1, \dots, N - 1$. A partição é dita “ótima” quando minimiza V . A condição necessária de minimização é

$$\frac{\partial V}{\partial s_i} = 0. \quad (3.32)$$

Uma mudança em s_i afeta apenas os dois subintervalos vizinhos $[s_{i-1}, s_i]$ e $[s_i, s_{i+1}]$, sendo que s_i é o ponto intermediário e s_{i-1} e s_{i+1} são fixos para efeitos da derivada parcial.

Considerando uma partição em dois intervalos com extremidades $a < b < c$, temos uma função de uma única variável

$$V(b) = \int_a^b (\bar{k}_1 - k_s)^2 ds + \int_b^c (\bar{k}_2 - k_s)^2 ds, \quad (3.33)$$

onde $\bar{k}_1$ é o valor médio no intervalo $[a, b]$ e $\bar{k}_2$, o valor médio no intervalo $[b, c]$.

O primeiro termo pode ser escrito na forma

$$\int_a^b (\bar{k}_1 - k_s)^2 ds = \int_a^b (k_s^2 - \bar{k}_1^2) ds, \quad (3.34)$$

obtida pelo desenvolvimento do quadrado no integrando explicitado a seguir:

$$\begin{aligned} \int_a^b (\bar{k}_1 - k_s)^2 ds &= \int_a^b (k_s^2 - 2\bar{k}_1 k_s + \bar{k}_1^2) ds = \\ &= \int_a^b (k_s^2) ds - 2\bar{k}_1 \int_a^b (k_s) ds + \int_a^b (\bar{k}_1^2) ds. \end{aligned} \quad (3.35)$$

Substituindo a forma alternativa para a Equação (3.29)

$$\int_a^b k_s ds = \int_a^b \bar{k} ds \quad (3.36)$$

em (3.35) chega-se à Equação (3.34).

Da mesma forma, o segundo termo é escrito como

$$\int_b^c (\bar{k}_2 - k_s)^2 ds = \int_b^c (k_s^2 - \bar{k}_2^2) ds, \quad (3.37)$$

e então

$$V(b) = \int_a^b (k_s^2 - \bar{k}_1^2) ds + \int_b^c (k_s^2 - \bar{k}_2^2) ds. \quad (3.38)$$

Uma mudança Δb no ponto intermediário altera tanto os limites de integração quanto os valores médios calculados. Assim,

$$V(b + \Delta b) = \int_a^{b+\Delta b} (k_s^2 - \bar{\bar{k}}_1^2) ds + \int_{b+\Delta b}^c (k_s^2 - \bar{\bar{k}}_2^2) ds, \quad (3.39)$$

onde $\bar{\bar{k}}_1$ é o valor médio no intervalo $[a, b + \Delta b]$ e $\bar{\bar{k}}_2$, o valor médio no intervalo $[b + \Delta b, c]$.

Subtraindo a Equação (3.38) da Equação (3.39), temos

$$V(b + \Delta b) - V(b) = (\bar{\bar{k}}_1^2 - \bar{k}_1^2)(b - a) + (\bar{\bar{k}}_2^2 - \bar{k}_2^2)\Delta b + (\bar{k}_2^2 - \bar{\bar{k}}_2^2)(c - b). \quad (3.40)$$

A partir da Equação (3.40), a derivada pode ser calculada pela definição

$$\begin{aligned}
 V'(b) &= \lim_{\Delta b \rightarrow 0} \frac{V(b + \Delta b) - V(b)}{\Delta b} = \\
 &\lim_{\Delta b \rightarrow 0} \frac{(\bar{k}_1^2 - \bar{\bar{k}}_1^2)(b - a)}{\Delta b} + \lim_{\Delta b \rightarrow 0} \frac{(\bar{\bar{k}}_2^2 - \bar{k}_2^2)\Delta b}{\Delta b} + \\
 &\lim_{\Delta b \rightarrow 0} \frac{(\bar{k}_2^2 - \bar{\bar{k}}_2^2)(c - b)}{\Delta b}.
 \end{aligned} \tag{3.41}$$

Tendo em vista que $\bar{\bar{k}}_1 = \bar{k}_1$ e $\bar{\bar{k}}_2 = \bar{k}_2$ no limite, a segunda parcela do último membro é avaliada como

$$\lim_{\Delta b \rightarrow 0} \frac{(\bar{\bar{k}}_2^2 - \bar{k}_1^2)\Delta b}{\Delta b} = (\bar{k}_2 - \bar{k}_1)(\bar{k}_2 + \bar{k}_1). \tag{3.42}$$

A primeira parcela pode ser escrita como

$$\begin{aligned}
 \lim_{\Delta b \rightarrow 0} \frac{(\bar{k}_1^2 - \bar{\bar{k}}_1^2)(b - a)}{\Delta b} &= \lim_{\Delta b \rightarrow 0} \frac{(\bar{k}_1 - \bar{\bar{k}}_1)(\bar{k}_1 + \bar{\bar{k}}_1)(b - a)}{\Delta b} = \\
 &-2\bar{k}_1(b - a) \lim_{\Delta b \rightarrow 0} \frac{\bar{\bar{k}}_1 - \bar{k}_1}{\Delta b},
 \end{aligned} \tag{3.43}$$

desde que exista o último limite, que é a derivada do valor médio em relação à extremidade b do intervalo. A diferença entre os valores médios é, em termos de integrais,

$$\begin{aligned}
 \bar{\bar{k}}_1 - \bar{k}_1 &= \frac{1}{b + \Delta b - a} \int_a^{b+\Delta b} k_s ds - \frac{1}{b - a} \int_a^b k_s ds = \\
 &\frac{-\Delta b}{(b + \Delta b - a)(b - a)} \int_a^b k_s ds + \frac{1}{b + \Delta b - a} \int_b^{b+\Delta b} k_s ds.
 \end{aligned} \tag{3.44}$$

De acordo com o teorema do valor médio para integrais, existe β no intervalo $[b, b + \Delta b]$ tal que

$$\int_b^{b+\Delta b} k_s ds = k_s(\beta)\Delta b, \tag{3.45}$$

o que resulta em

$$\bar{\bar{k}}_1 - \bar{k}_1 = \frac{(k_s(\beta) - \bar{k}_1)\Delta b}{b + \Delta b - a}. \tag{3.46}$$

Com isso,

$$\lim_{\Delta b \rightarrow 0} \frac{\bar{\bar{k}}_1 - \bar{k}_1}{\Delta b} = \lim_{\Delta b \rightarrow 0} \frac{k_s(\beta) - \bar{k}_1}{b + \Delta b - a} = \frac{k_s(b) - \bar{k}_1}{b - a}, \quad (3.47)$$

onde é considerado que β tende a b no limite. Analogamente,

$$\lim_{\Delta b \rightarrow 0} \frac{\bar{\bar{k}}_2 - \bar{k}_2}{\Delta b} = \frac{k_s(b) - \bar{k}_2}{b - c}. \quad (3.48)$$

Substituindo a Equação (3.47) em (3.43), temos

$$\lim_{\Delta b \rightarrow 0} \frac{(\bar{k}_1^2 - \bar{\bar{k}}_1^2)(b - a)}{\Delta b} = -2\bar{k}_1(k_s(b) - \bar{k}_1). \quad (3.49)$$

Da mesma forma, o limite da terceira parcela da Equação (3.41) é expresso como

$$\lim_{\Delta b \rightarrow 0} \frac{(\bar{k}_2^2 - \bar{\bar{k}}_2^2)(c - b)}{\Delta b} = 2\bar{k}_2(k_s(b) - \bar{k}_2). \quad (3.50)$$

Finalmente, substituindo as três parcelas calculadas na Equação (3.41) e rearranjando os termos, é obtida uma fórmula para o gradiente de V

$$V'(b) = (\bar{k}_2 - \bar{k}_1) \left(2k_s(b) - (\bar{k}_2 + \bar{k}_1) \right). \quad (3.51)$$

O ponto intermediário ótimo $\bar{b}$ satisfaz a condição

$$V'(\bar{b}) = 0, \quad (3.52)$$

o que, de acordo com Equação (3.51), é equivalente a

$$\bar{k}_1 = \bar{k}_2 \quad (3.53)$$

ou

$$k_s(\bar{b}) = \frac{\bar{k}_1 + \bar{k}_2}{2}. \quad (3.54)$$

Uma partição que satisfaça as Equações (3.53) e (3.54) é dita “crítica” por se tratar de um ponto crítico da função $V(s_1, \dots, s_{N-1})$. Esta partição crítica não parece ter relação com a derivada da função de curvatura, o que a difere da partição de Pei e Horng. A Figura 23 mostra uma partição praticamente crítica de uma função de curvatura aproximada por dois trechos constantes, note como o ponto de quebra localiza-se aproximadamente na posição média entre os patamares constantes.

Figura 23 – Aproximação constante por trechos de uma função com dois subintervalos

Fonte: autoria própria

Para determinar se este ponto é um minimizador, é necessária a segunda derivada da função V . Derivando a Equação (3.51), lembrando que as derivadas dos valores médios são dadas pelos limites nas Equações (3.47) e (3.48), escreve-se

$$V''(b) = \left(\frac{k_s(b) - \bar{k}_2}{b - c} - \frac{k_s(b) - \bar{k}_1}{b - a} \right) (2k_s(b) - (\bar{k}_2 + \bar{k}_1)) + (\bar{k}_2 - \bar{k}_1) \left(2k'_s(b) - \left(\frac{k_s(b) - \bar{k}_2}{b - c} + \frac{k_s(b) - \bar{k}_1}{b - a} \right) \right). \quad (3.55)$$

No caso em que $\bar{k}_1 = \bar{k}_2 = \bar{k}$,

$$V''(\bar{b}) = \left(\frac{k_s(\bar{b}) - \bar{k}}{\bar{b} - c} - \frac{k_s(\bar{b}) - \bar{k}}{\bar{b} - a} \right) (2k_s(\bar{b}) - 2\bar{k}) = \frac{2(k_s(\bar{b}) - \bar{k})^2 (c - a)}{(\bar{b} - c)(\bar{b} - a)}, \quad (3.56)$$

um número negativo devido ao termo $(\bar{b} - c)$ no denominador, mostrando que $V(\bar{b})$ se trata de um máximo. De fato, é como se não houvesse efetivamente uma divisão em dois intervalos, mas um único intervalo $[a, c]$ onde o valor médio é $\bar{k}$, o que é um prejuízo do ponto de vista de aproximar o gráfico de uma função por trechos de funções constantes.

O caso de interesse é então o no qual vale a Equação (3.54), que simplifica (3.55) para a forma

$$V''(\bar{b}) = (\bar{k}_2 - \bar{k}_1) \left(2k'_s(\bar{b}) - \frac{(\bar{k}_2 - \bar{k}_1)(c - a)}{2(c - \bar{b})(\bar{b} - a)} \right). \quad (3.57)$$

Supondo $\bar{k}_2 > \bar{k}_1$,

$$V''(\bar{b}) > 0 \Rightarrow k'_s(\bar{b}) > \frac{(\bar{k}_2 - \bar{k}_1)(c - a)}{4(c - \bar{b})(\bar{b} - a)} > 0. \quad (3.58)$$

Por outro lado, se $\bar{k}_2 < \bar{k}_1$, então

$$V''(\bar{b}) > 0 \Rightarrow k'_s(\bar{b}) < \frac{(\bar{k}_2 - \bar{k}_1)(c - a)}{4(c - \bar{b})(\bar{b} - a)} < 0. \quad (3.59)$$

As duas equações anteriores se resumem na condição

$$\frac{k'_s(\bar{b})}{\bar{k}_2 - \bar{k}_1} > \frac{(c - a)}{4(c - \bar{b})(\bar{b} - a)} > 0 \quad (3.60)$$

Em conclusão, o ponto intermediário ótimo satisfaz a Equação (3.54) e, para que seja um minimizador, a inclinação da reta tangente neste ponto deve atender à Equação (3.60), que impõe um limite inferior para sua magnitude.

Retomando agora o problema com múltiplas partes representado pela soma em (3.31), uma componente genérica do gradiente é dada por

$$\frac{\partial V}{\partial s_i} = (\bar{k}_i - \bar{k}_{i-1}) \left(2k_s(s_i) - (\bar{k}_i + \bar{k}_{i-1}) \right), \quad (3.61)$$

onde $\bar{k}_{i-1}$ e $\bar{k}_i$ são os valores médios nos intervalos respectivos $[s_{i-1}, s_i]$ e $[s_i, s_{i+1}]$ afetados pela mudança em s_i .

A derivação da Equação (3.61) resulta nas componentes da matriz hessiana de V . Como s_{i-1} , s_i e s_{i+1} são os únicos pontos de quebra com influência sobre a Equação (3.61), a matriz hessiana é tridiagonal. As componentes não necessariamente nulas são

$$\frac{\partial^2 V}{\partial s_{i-1} \partial s_i} = -2 \frac{(k_s(s_{i-1}) - \bar{k}_{i-1})(k_s(s_i) - \bar{k}_{i-1})}{s_{i-1} - s_i}, \quad (3.62)$$

$$\begin{aligned} \frac{\partial^2 V}{\partial s_i^2} = & \left(\frac{k_s(s_i) - \bar{k}_i}{s_i - s_{i+1}} - \frac{k_s(s_i) - \bar{k}_{i-1}}{s_i - s_{i-1}} \right) (2k_s(s_i) - (\bar{k}_i + \bar{k}_{i-1})) + \\ & (\bar{k}_i - \bar{k}_{i-1}) \left(2k'_s(s_i) - \left(\frac{k_s(s_i) - \bar{k}_i}{s_i - s_{i+1}} + \frac{k_s(s_i) - \bar{k}_{i-1}}{s_i - s_{i-1}} \right) \right) \end{aligned} \quad (3.63)$$

e

$$\frac{\partial^2 V}{\partial s_{i+1} \partial s_i} = 2 \frac{(k_s(s_{i+1}) - \bar{k}_i)(k_s(s_i) - \bar{k}_i)}{s_{i+1} - s_i}. \quad (3.64)$$

Embora as Equações de (3.61) a (3.64) possibilitem a utilização de um algoritmo *trust-region* com modelo quadrático de hessiana exata a fim de determinar os pontos de quebra ótimos, um esquema bem menos sofisticado foi empregado, tendo em vista que há somente a disponibilidade de valores discretos para s . Dada uma partição inicial, a cada par de subintervalos adjacentes, é calculada a componente do gradiente referente ao ponto intermediário através da Equação (3.61). Se o resultado for positivo, então existe um deslocamento do ponto de quebra para a esquerda que reduz V , o parâmetro da amostra anterior é tomado como novo ponto de quebra. Se, ao contrário, o gradiente for negativo, então o ponto de quebra é atualizado com o valor do parâmetro da amostra seguinte. Este passo é feito até que o gradiente mude de sinal ou se anule ou o ponto de quebra variável alcance um ponto de quebra vizinho. Determinado o ponto de quebra intermediário, passa-se ao próximo par de subintervalos e assim sucessivamente.

Ajustados todos os pontos intermediários, é obtida uma nova partição que não é necessariamente ótima, pois o processo descrito no parágrafo anterior desconsidera efeitos de dependência entre pontos de quebra vizinhos. Espera-se que aplicando as mesmas etapas à partição atual, uma melhor seja encontrada. Esta repetição é feita um número determinado de vezes.

Esse procedimento gerou tanto o resultado da Figura 23 quanto o, mais refinado, da Figura 24. Podemos perceber que a região em que a curvatura varia mais rapidamente concentra mais arcos, enquanto regiões em que a curvatura varia mais gradualmente são aproximadas por menos arcos. Uma característica esperada de uma boa partição. Outra característica do algoritmo é a capacidade de localizar descontinuidades como ilustrado na Figura 25.

Figura 24 – Aproximação constante por trechos de uma função com dez subintervalos

Fonte: autoria própria

Figura 25 – Comportamento do algoritmo de partição na presença de uma descontinuidade (a) partição escolhida arbitrariamente (b) partição obtida

Fonte: autoria própria

3.2.4 Parâmetros da spline de arcos

Do gráfico de curvatura constante por trechos que aproxima o gráfico original deriva-se uma única spline de arcos. A cada subintervalo da partição está associado um arco de circunferência com curvatura orientada igual à curvatura média no intervalo e comprimento de arco igual à largura do intervalo. Desta forma, são garantidas as igualdades de comprimento total e de curvatura total entre a curva de aproximação e a curva original.

A curva de ajuste é posicionada de forma a coincidir seu início com a primeira amostra e as direções tangentes no ponto comum. Assim ficam determinados os $2n + 3$ parâmetros necessários.

3.3 Aprimoramento do ajuste por deformação da curva

A forma da curva de ajuste é determinada pela curvatura e pelo comprimento de cada segmento que a compõe. A variação destes parâmetros gera uma deformação da curva que pode aproximá-la das amostras, minimizando as distâncias das amostras para a curva.

O nível de aproximação é dado pela soma

$$f = \frac{1}{2} \sum_{i=1}^m r_i^2, \quad (3.65)$$

em que $|r_i|$ é a distância de uma amostra até a curva. A aproximação é ótima quando f tem seu valor mínimo, o clássico problema de mínimos quadrados. A função objetivo depende dos parâmetros de forma e posicionamento da spline de arcos.

Serão considerados fixos os comprimentos de cada arco, mantendo o comprimento total da curva de ajuste, suposto igual ao da curva original. O ponto inicial também será mantido como coincidente com a primeira amostra. São variáveis as curvaturas e a direção tangente inicial. Portanto, cada arco pode ser mais ou menos encurvado mantendo seu comprimento e a curva como um todo pode girar em torno do ponto inicial.

A seguir, estudamos como a deformação e rotação da curva afetam os resíduos, e como usar esta informação para minimizar as distâncias.

3.3.1 Resíduos e suas derivadas

Existe uma relação entre os arcos e os subintervalos da partição da curva original. Para cada arco há um correspondente subconjunto fixo de amostras, as que possuem parâmetro pertencente ao intervalo associado ao arco. A distância entre uma amostra e o arco associado a ela é uma medida a ser reduzida no ajuste da curva.

Figura 26 – Distância de um ponto P a um arco com extremidades P_i e P_f , e centro C ; (bloco superior) a reta PC intercepta o arco; (bloco inferior) a reta PC não intercepta o arco

Fonte: autoria própria

A Figura 26 indica como é medida a distância da amostra P até o arco. Quando a reta que liga a amostra ao centro intercepta o arco, a distância do ponto ao arco é igual à distância do ponto à sua projeção ortogonal sobre o arco. No caso (superior-esquerdo), a distância é

$$d = \|C - P\| - R, \quad (3.66)$$

descontando o raio R da distância de C até P . No caso (superior-direito), em que o ponto P está no interior do arco,

$$d = R - \|C - P\|. \quad (3.67)$$

Em ambos, escrevendo o resíduo como

$$r = \|C - P\| - R \quad (3.68)$$

fica satisfeita a condição $|r| = d$.

Nos outros casos, não existe projeção ortogonal de P sobre o arco. No entanto, como se espera que os segmentos da curva solução fiquem muito próximos dos pontos que devem aproximar, o resíduo (3.68) pode ser tomado como um indicador da distância. Neste caso, o ponto P está sendo projetado na extensão do arco.

Pela forma de construção da curva, o ponto inicial de um arco P_i e sua tangente são determinados pelos arcos anteriores, como se a extremidade do arco estivesse engastada. Uma mudança de curvatura deve manter a forma circular, o ponto e a tangente iniciais, enquanto são

alterados o raio, o centro, o ponto e a tangente finais como mostra a Figura 27. A mudança de curvatura consequentemente modifica o resíduo.

Figura 27 – Deformação de um arco mantendo o ponto inicial e sua tangente fixos

Fonte: autoria própria

A razão de mudança de uma quantidade com relação a outra é dada pela derivada parcial da primeira com relação à segunda. Avaliamos a mudança do raio escrevendo

$$R = \left| \frac{1}{k} \right| = \begin{cases} \frac{1}{k}, & k > 0 \\ -\frac{1}{k}, & k < 0 \end{cases}, \quad (3.69)$$

e derivando com relação a k para obter

$$\frac{\partial R}{\partial k} = \begin{cases} -\frac{1}{k^2}, & k > 0 \\ \frac{1}{k^2}, & k < 0 \end{cases} = -\frac{R}{k}. \quad (3.70)$$

Uma pequena mudança de curvatura implica na mudança de raio

$$\Delta R = -\frac{R}{k} \Delta k, \quad (3.71)$$

que pode fazer com que o arco se aproxime da amostra.

A posição do centro também é modificada. A direção normal inicial é constante, assim como a direção tangente inicial, devido a isso, o movimento do centro é restrito à reta normal $P_i C$, o centro modificado C' está condicionado a incidir sobre esta reta. Como indica a Figura 27,

$$\Delta C = C' - C = -N_i(R - R') = N_i \Delta R, \quad (3.72)$$

sendo

$$\mathbf{N}_i = \frac{\mathbf{C} - \mathbf{P}_i}{\|\mathbf{C} - \mathbf{P}_i\|} \quad (3.73)$$

o vetor normal no ponto inicial. Substituindo a Equação (3.71) na Equação (3.72), temos

$$\Delta C = -\frac{R}{k} \mathbf{N}_i \Delta k, \quad (3.74)$$

o que indica que a derivada do centro com relação à curvatura é o vetor

$$\frac{\partial \mathbf{C}}{\partial k} = -\frac{R}{k} \mathbf{N}_i. \quad (3.75)$$

De fato, escrevendo

$$\mathbf{C} - \mathbf{P}_i = R \mathbf{N}_i, \quad (3.76)$$

como $\mathbf{P}_i$ e $\mathbf{N}_i$ são constantes,

$$\frac{\partial \mathbf{C}}{\partial k} = \frac{\partial R}{\partial k} \mathbf{N}_i = -\frac{R}{k} \mathbf{N}_i. \quad (3.77)$$

A deformação do arco também altera a posição do ponto final. A posição do arco seguinte é também modificada pela deformação como mostrado na Figura 28.

Figura 28 – Movimento de um arco devido à deformação do arco anterior sobre o qual ele se apoia

Fonte: autoria própria

O centro do primeiro arco C_1 , o centro do segundo arco C_2 e o ponto intermediário $P_{i,2}$, ponto inicial do segundo arco ou final do primeiro arco, são colineares. Se o par de arcos tem formato de S, como é o caso na Figura 28, então

$$\frac{C_1 - P_{i,2}}{R_1} = -\frac{C_2 - P_{i,2}}{R_2}. \quad (3.78)$$

Como os sinais das curvaturas são opostos,

$$k_1(C_1 - P_{i,2}) = k_2(C_2 - P_{i,2}). \quad (3.79)$$

A última expressão também é válida quando o par tem formato de C.

Os três pontos se mantêm alinhados durante a deformação, portanto a Equação (3.79) pode ser diferenciada com relação à curvatura k_1 , o que resulta em

$$(C_1 - P_{i,2}) + k_1 \left(\frac{\partial C_1}{\partial k_1} - \frac{\partial P_{i,2}}{\partial k_1} \right) = k_2 \left(\frac{\partial C_2}{\partial k_1} - \frac{\partial P_{i,2}}{\partial k_1} \right). \quad (3.80)$$

Pela condição de continuidade, a substituição

$$P_{i,2} = P_{f,1} \quad (3.81)$$

pode ser feita. O ponto final do primeiro arco em termos do centro e da normal é escrito como

$$P_{f,1} = C_1 - R_1 \mathbf{N}_{f,1}. \quad (3.82)$$

Com isso,

$$\frac{\partial C_2}{\partial k_1} = \frac{\partial C_1}{\partial k_1} + \frac{R_1}{k_1} \mathbf{N}_{f,1} + R_1 \left(\frac{k_1}{k_2} - 1 \right) \frac{\partial \mathbf{N}_{f,1}}{\partial k_1}. \quad (3.83)$$

Arcos ainda anteriores indiretamente influenciam a posição de um certo arco. A partir da Equação (3.79) é possível desenvolver um procedimento recursivo para obter a derivada do centro de um dado arco de índice α com relação à curvatura de um arco anterior qualquer de índice $\beta < \alpha$. Primeiramente, reescrevemos a Equação (3.79) como

$$C_\alpha = C_{\alpha-1} - R_{\alpha-1} \mathbf{N}_{f,\alpha-1} + \frac{k_{\alpha-1}}{k_\alpha} R_{\alpha-1} \mathbf{N}_{f,\alpha-1}. \quad (3.84)$$

Derivando com relação à curvatura do arco β , temos

$$\begin{aligned} \frac{\partial C_\alpha}{\partial k_\beta} = & \frac{\partial C_{\alpha-1}}{\partial k_\beta} + \frac{\partial R_{\alpha-1}}{\partial k_\beta} \left(\frac{k_{\alpha-1}}{k_\alpha} - 1 \right) \mathbf{N}_{f,\alpha-1} + \left(\frac{1}{k_\alpha} \frac{\partial k_{\alpha-1}}{\partial k_\beta} \right) R_{\alpha-1} \mathbf{N}_{f,\alpha-1} + \\ & R_{\alpha-1} \left(\frac{k_{\alpha-1}}{k_\alpha} - 1 \right) \frac{\partial \mathbf{N}_{f,\alpha-1}}{\partial k_\beta}. \end{aligned} \quad (3.85)$$

Se $\beta = \alpha - 1$, então a Equação (3.83) é retomada. Se $\beta < \alpha - 1$, então o segundo e o terceiro termos do membro direito são anulados e resta

$$\frac{\partial C_\alpha}{\partial k_\beta} = \frac{\partial C_{\alpha-1}}{\partial k_\beta} + R_{\alpha-1} \left(\frac{k_{\alpha-1}}{k_\alpha} - 1 \right) \frac{\partial \mathbf{N}_{f,\alpha-1}}{\partial k_\beta}. \quad (3.86)$$

A recursão acontece no primeiro termo do membro direito da Equação (3.86), reduzindo uma unidade em cada índice presente a cada repetição

$$\frac{\partial C_{\alpha-1}}{\partial k_\beta} = \frac{\partial C_{\alpha-2}}{\partial k_\beta} + R_{\alpha-2} \left(\frac{k_{\alpha-2}}{k_{\alpha-1}} - 1 \right) \frac{\partial \mathbf{N}_{f,\alpha-2}}{\partial k_\beta}. \quad (3.87)$$

A base da recursão é atingida quando o centro do próprio arco β aparece, esta derivada é conhecida, dada pela Equação (3.75) como

$$\frac{\partial C_\beta}{\partial k_\beta} = -\frac{R_\beta}{k_\beta} \mathbf{N}_{i,\beta}. \quad (3.88)$$

Já o segundo termo do membro direito da Equação (3.86), é obtido ao se considerar o vetor $\mathbf{N}_{f,\alpha-1}$ escrito na base ortogonal composta por $\mathbf{T}_{i,\beta}$ e $\mathbf{N}_{i,\beta}$, um par de vetores fixos com a deformação do arco β

$$\mathbf{N}_{f,\alpha-1} = \sin \Delta\Phi \mathbf{T}_{i,\beta} + \cos \Delta\Phi \mathbf{N}_{i,\beta}. \quad (3.89)$$

Nesta forma, o ângulo $\Delta\Phi$ é obtido pelo acúmulo dos ângulos desde o arco β até o arco $\alpha - 1$ acrescido de $M\pi$, $M \in \mathbb{Z}$, devido a inversões da normal nas transições entre pares de arcos de formato S

$$\begin{aligned} \Delta\Phi &= \Delta\varphi_\beta + \Delta\varphi_{\beta+1} + \dots + \Delta\varphi_{\alpha-1} + M\pi = \\ &= k_\beta L_\beta + k_{\beta+1} L_{\beta+1} + \dots + k_{\alpha-1} L_{\alpha-1} + M\pi. \end{aligned} \quad (3.90)$$

A derivada em relação à curvatura

$$\frac{\partial \mathbf{N}_{f,\alpha-1}}{\partial k_\beta} = L_\beta \cos \Delta\Phi \mathbf{T}_{i,\beta} - L_\beta \sin \Delta\Phi \mathbf{N}_{i,\beta} \quad (3.91)$$

é um vetor ortogonal a $\mathbf{N}_{f,\alpha-1}$ com magnitude igual ao comprimento do arco β .

Com isso, conhecemos as derivadas de todas as quantidades que aparecem no resíduo relativas a todas as curvaturas da spline de arcos. O último parâmetro da curva é o ângulo associado ao vetor tangente inicial. Seja P_0 o ponto inicial e Q qualquer ponto ligado rigidamente à curva, como um dos centros dos arcos por exemplo, denotando por θ o parâmetro angular, vale

$$\frac{\partial Q}{\partial \theta} = \|Q - P_0\| \mathbf{v}_\perp, \quad (3.92)$$

onde $\mathbf{v}_\perp$ é um vetor unitário ortogonal a $Q - P_0$.

Uma ilustração deste fato é apresentada na Figura 29 para uma figura rígida genérica. No limite em que a rotação é infinitesimal, o vetor ΔQ é ortogonal ao vetor $Q - P_0$ e sua magnitude é dada por

$$\|\Delta Q\| = \|Q - P_0\| \Delta\theta . \quad (3.93)$$

Escrevendo a direção como $\mathbf{v}_\perp$, vetor obtido pela rotação do vetor unitário paralelo a $Q - P_0$ de um ângulo reto no sentido positivo, temos

$$\Delta Q = \|Q - P_0\| \mathbf{v}_\perp \Delta\theta , \quad (3.94)$$

de onde se obtém a Equação (3.92).

Figura 29 – Rotação plana de uma figura rígida genérica em torno do polo P_0

Fonte: autoria própria

Voltando aos resíduos expressos na forma da Equação (3.68), se a amostra pertence ao arco de índice α , a derivada do resíduo é dada por

$$\frac{\partial r}{\partial x_j} = \frac{C_\alpha - P}{\|C_\alpha - P\|} \cdot \frac{\partial C_\alpha}{\partial x_j} - \frac{\partial R_\alpha}{\partial x_j} \quad (3.95)$$

com relação a um parâmetro genérico x_j , que pode ser a curvatura de algum arco ou o parâmetro angular.

3.4 Segmentação

Segmentação refere-se a dois pontos do processo de aproximação: determinar o número de segmentos da spline de arcos e qual parte da curva original cada segmento fica encarregado de aproximar. Dado o número de arcos, a segmentação é ditada pela partição do comprimento da curva (conforme Seção 3.2.3). A aplicação do método de aproximação resulta em uma curva

com certo erro, o qual queremos limitar por uma dada tolerância, isso requer um mecanismo de acrescentar arcos.

As estratégias de segmentação podem ser classificadas em duas categorias [36]. O primeiro tipo pode ser dito “ótimo” porque busca uma solução com o menor número de segmentos possível. Também pode ser chamada de “estratégia dos arcos mais longos” por tentar ajustar os arcos de maior comprimento que aproveitem ao máximo a região permitida pela tolerância dada. Os métodos de Marciniak e Putz [9] e de Maier [39] seguem este paradigma. A segunda categoria é a de “bissecção” pois parte de uma solução com certa quantidade de arcos e busca reduzir o erro subdividindo os segmentos até que o critério de tolerância seja atendido. Os métodos de Meek e Walton [11], de Ahn *et al.* [7] e de Yong *et al.* [8], por exemplo, encaixam-se nesta.

O algoritmo de segmentação aqui proposto é baseado em bissecção. É dado um número inicial de arcos. O comprimento estimado da curva é particionado no mesmo número de subintervalos. A partição é manipulada de forma a tornar a aproximação por curvaturas constantes mais similar possível à função de curvatura original – “partição ótima”.

O método de ajuste baseado na função de curvatura é aplicado. Caso a solução satisfaça a tolerância, o algoritmo é encerrado. Caso contrário, a solução baseada em curvatura é utilizada como solução inicial no método de otimização por mínimos quadrados.

Se o método de otimização converge para uma nova solução, então verifica-se se esta atende ao critério de tolerância. No caso afirmativo, a solução foi encontrada. No caso negativo, um número maior de arcos é necessário, então é feita a bissecção do arco com maior erro – “arco mais afastado” – preparando uma nova iteração do método de ajuste. O novo ponto de quebra inserido na partição pela bissecção é ajustado e as curvaturas dos arcos não são alteradas, assim a forma da spline de arcos encontrada é mantida como ponto inicial para a nova iteração do processo de otimização, agora com um arco adicional.

A não convergência do método de otimização indica que a solução inicial não era satisfatória. Neste caso, após a bissecção, o ajuste é completamente reiniciado, isto é, todos os pontos de quebra da partição são reajustados e uma nova aproximação das curvaturas é feita para ser usada como ponto inicial da otimização.

O algoritmo esquematizado desta forma garante que a função objetivo sempre decresça enquanto a otimização convergir e possibilita a continuidade do método caso não ocorra convergência em alguma iteração. Uma visão geral do algoritmo está esquematizada na Figura 30.

Figura 30 – Esquema geral do algoritmo de ajuste

Fonte: autoria própria

4 Resultados

Os métodos descritos no Capítulo 3 foram testados fazendo aproximações de diferentes tipos de curvas. Neste capítulo, resultados de variados testes são apresentados na tentativa de ilustrar diferentes características do procedimento completo.

O primeiro conjunto de exemplos contém curvas paramétricas exatas. Começamos por um exemplo muito simples em que a curva original é um biarco. O segundo exemplo é uma curva espiral (curvatura monótona decrescente ou crescente). A complexidade aumenta nos exemplos com curvas polinomiais.

O segundo conjunto de exemplos é formado por aproximações de perfis ferroviários medidos. Esta seção de exemplos é aproveitada para explorar características práticas do procedimento de ajuste pertinentes à aplicação prática pretendida dos métodos.

4.1 Biarco em formato de S

O exemplo mais simples possível para verificar o funcionamento dos métodos é realizar o ajuste de uma spline de arcos a um biarco amostrado, que deve ser perfeito. A curva amostrada é composta de duas semicircunferências de raio unitário ligadas em forma de S parametrizada como

$$\gamma(t) = \begin{cases} \cos t \mathbf{i} + \sin t \mathbf{j}, & 0 \leq t < \pi \\ (-2 - \cos t) \mathbf{i} + \sin t \mathbf{j}, & \pi \leq t \leq 2\pi \end{cases} \quad (4.1)$$

A curvatura do primeiro segmento é $+1$ e a do segundo segmento, -1 , constantes em cada trecho, o que resulta em um gráfico retangular. A aplicação do algoritmo de aproximação de um perfil de curvatura por retângulos deve colocar o ponto de quebra exatamente no ponto de descontinuidade $s = \pi$.

Figura 31 – Gráfico de curvatura de um biarco e aproximação por trechos constantes desajustada

Fonte: autoria própria

Para testar o funcionamento do algoritmo, foi escolhido um ponto de quebra fora da descontinuidade como indica a Figura 31. A Equação (3.51) aplicada a esta configuração resulta em $V'(b) < 0$, sendo b o ponto de quebra, isto significa que um aumento no valor do ponto de quebra reduz a quantidade V , portanto, o algoritmo desloca o ponto de quebra para a direita. A configuração final é a esperada como mostra a Figura 32.

Figura 32 – Gráfico de curvatura de um biarco e aproximação por trechos constantes ajustada

Fonte: autoria própria

O ajuste da curva é praticamente perfeito com curvaturas $+1,00$ e $-1,00$, veja a Figura

Figura 33 – Biarco em formato de S ajustado

Fonte: autoria própria

Aproveitamos este exemplo para verificar o efeito de uma janela grande de pontos na estimativa do gráfico de curvatura usando a regressão quadrática. A saber, a curva tem mil amostras uniformemente distribuídas e a janela considerada é de cento e uma amostras, a amostra onde se calcula a curvatura mais cinquenta anteriores e cinquenta posteriores. Pode-se observar na Figura 34 a suavização da descontinuidade, o que resulta em curvaturas médias um pouco menores em magnitude.

Figura 34 – Gráfico de curvatura estimado por regressão quadrática

Fonte: autoria própria

O impacto da estimativa da curvatura sobre a curva de ajuste é indicado na Figura 35. Há um desalinhamento devido à curvatura um pouco menor dos arcos de ajuste. A otimização corrige as curvaturas e um ajuste idêntico ao da Figura 33 é obtido.

Figura 35 – Curvas de ajuste a um biarco (a) antes da otimização e (b) depois da otimização

Fonte: autoria própria

4.2 Espiral áurea

Considere a espiral parametrizada por

$$\gamma(t) = e^{kt} \cos t \, \mathbf{i} + e^{kt} \sin t \, \mathbf{j}, k = \frac{2}{\pi} \ln \left(\frac{1 + \sqrt{5}}{2} \right), \quad (4.2)$$

chamada de espiral áurea. É dito que a espiral de Fibonacci é uma boa aproximação para a espiral áurea. O ajuste de uma spline de arcos de quatro arcos a uma volta da espiral pode ajudar na verificação da afirmação anterior.

A curvatura da espiral áurea é dada pela expressão

$$k_s(t) = \frac{1}{\sqrt{e^{2kt}(1 + k^2)}}, \quad (4.3)$$

com gráfico apresentado na Figura 36 junto da aproximação retangular por partes. A partição é tal que cada arco aproxima a imagem de um intervalo de tamanho $\pi/2$ no parâmetro t .

Figura 36 – Gráfico de curvatura da espiral áurea e aproximação retangular com quatro partes

Fonte: autoria própria

A primeira spline de arcos de ajuste, que está associada à aproximação retangular na Figura 36, está sobreposta à espiral logarítmica na Figura 37 (a). O resultado após a otimização é mostrado na Figura 37 (b).

Figura 37 – Curvas de ajuste (a) antes da otimização e (b) depois da otimização

Fonte: autoria própria

Os raios dos quatro arcos antes e após a otimização estão indicados na Tabela 1. Percebe-se que a sequência de raios não difere tanto da sequência de Fibonacci, e o mais interessante é que a razão entre os raios consecutivos é próxima da razão áurea $\varphi \approx 1,6180$ na primeira aproximação. A espiral de Fibonacci é uma boa aproximação para a espiral áurea.

Tabela 1 – Raios dos arcos de ajuste a uma espiral logarítmica

Arco	Número de Fibonacci	Raio antes da otimização	Razão com o raio anterior	Raio depois da otimização	Razão com o raio anterior
1	1	1,3436	-	1,3502	-
2	2	2,1750	1,6188	2,1684	1,6060
3	3	3,5210	1,6189	3,5423	1,6336
4	5	5,6939	1,6171	5,5925	1,5788

4.3 Curvas quadráticas

Walton e Meek [10] e Ahn, Kim e Lee [7] desenvolveram métodos de ajuste com arcos de circunferência válidos para curvas de Bézier quadráticas. Em ambos os trabalhos, as mesmas curvas, ilustradas na Figura 38, são usadas para testar os algoritmos. Os exemplos são replicados aqui com o propósito de comparar a performance do novo método com resultados anteriores.

Figura 38 – Curvas de Bézier quadráticas para exemplos

Fonte: autoria própria

Como o método apresentado lida com conjuntos discretos de pontos, as curvas foram amostradas tomando-se 500 pontos igualmente espaçados no intervalo paramétrico. O erro de ajuste (a máxima distância entre a curva a ser aproximada e a de ajuste) é estimado pelo máximo valor absoluto dentre os resíduos.

Para cada exemplo, são estabelecidos níveis de tolerância a se atender. O número inicial de arcos é 3 em todos os exemplos. O número de arcos mínimo necessário para que o erro do ajuste seja inferior à tolerância é tomado como indicador da performance do algoritmo. A Tabela 2 traz os resultados referentes aos três exemplos. Os resultados indicam que o método proposto não apenas é funcional, pode ser mais eficiente que outras soluções quanto ao número de arcos.

Tabela 2 – Comparação da quantidade de arcos necessária para ajuste de cada exemplo com diferentes tolerâncias usando três métodos

	Tolerância	Número de arcos		
		Método proposto	Walton e Meek [10]	Ahn, Kim e Lee [7]
Exemplo 1	0,0005	8	14	10
Exemplo 2	0,01	3	4	-
	0,001	6	8	6
	0,0001	12	16	-
Exemplo 3	0,01	6	12	-
	0,001	12	18	15
	0,0001	27	46	-

As seguintes figuras ilustram as curvas de aproximação e os perfis de curvatura dos três exemplos. A Figura 39 mostra o Exemplo 1 e sua aproximação com 8 arcos. A tolerância de 0,0005 é pequena o suficiente para que a curva de Bézier e a spline de arco sejam visualmente indistinguíveis. O Exemplo 2 é um segmento espiral (curvatura monótona decrescente) de baixa curvatura. A Figura 40 mostra sua aproximação obtida pelo método proposto com 6 arcos, que atende à tolerância de 0,001.

Figura 39 – Solução da curva quadrática Exemplo 1 com tolerância de 0,0005 e gráfico de curvatura

Fonte: autoria própria

Figura 40 – Solução da curva quadrática Exemplo 2 com tolerância de 0,001 e gráfico de curvatura

Fonte: autoria própria

A Figura 41 mostra o Exemplo 3 e sua aproximação com 12 arcos, que atende à tolerância de 0,001. Este exemplo destaca-se pelo elevado pico de curvatura. Foi particularmente importante neste exemplo que o número inicial de arcos fosse 3, o método apresentou dificuldades de convergência lidando com número inicial de arcos inferior, o que indica a sua sensibilidade à solução inicial. A aproximação com 3 arcos é um bom ponto de partida, o erro desta aproximação é menor que 0,05.

Figura 41 – Solução da curva quadrática Exemplo 3 com tolerância de 0,001 e gráfico de curvatura

Fonte: autoria própria

Uma característica comum às curvas quadráticas é a existência de um único ponto de curvatura extrema (máxima ou mínima). Isto pode ser visto nos gráficos de curvatura dos três exemplos. Ainda sobre os perfis de curvatura, podemos ver como arcos de maior comprimento são, por efeito do algoritmo de partição, usados para aproximar trechos em que a curvatura varia pouco.

4.4 Curvas de ordem superior

Curvas quadráticas têm um número bem limitado de características. Elas têm apenas um ponto de curvatura extrema, o que implica em no máximo dois segmentos espirais. Além disso, a curvatura nunca muda de sinal. Formas mais complexas podem ser desenhadas usando curvas de maior ordem (como cúbicas, quárticas e outras).

Para testar a validade do método para formas mais gerais, replicamos dois exemplos presentes em Meek e Walton [11], uma curva cúbica e uma curva quártica, Figura 42(a) e Figura 42(b) respectivamente.

Figura 42 – Curvas de Bézier de grau maior que 2 para exemplos (a) cúbica (b) quíntica

Fonte: autoria própria

Como nos exemplos anteriores, 500 amostras igualmente espaçadas no intervalo paramétrico das curvas de Bézier foram tomadas. Nestes exemplos, controlamos o número de arcos e avaliamos o erro para visualizar como este diminui conforme arcos são acrescentados. Os resultados obtidos são mostrados na Tabela 3.

Tabela 3 – Comparação do erro de aproximação entre dois métodos aumentando o número de arcos

	Erro de aproximação		Número de arcos
	estimado		
	Método proposto	Meek e Walton [11]	
Cúbica	0,23111	0,32954	4
	0,077348	0,035121	8
	0,0094466	0,0040455	16
	0,0010793	0,00048843	32
Quíntica	2,7389	4,1602	4
	0,74423	0,63756	8
	0,13075	0,053943	16
	0,014687	0,0061082	32

Apesar de o método continuar resultando em aproximações cada vez melhores com o aumento do número de arcos, os resultados não são tão promissores quanto nos exemplos com

curvas quadráticas quando comparados com os da outra solução. Conforme a Tabela 3, o método proposto resulta em uma aproximação com erro menor quando poucos arcos são usados, mas o método de Meek e Walton rapidamente (a partir de 8 arcos) mostra-se mais eficiente.

Os seguintes gráficos (Figura 43 e Figura 44) permitem comparar visualmente o comportamento do erro de aproximação nos dois métodos. Ainda que o método de Meek e Walton tenha um erro menor para o mesmo número de arcos na maior parte dos pontos, as diferenças, em escala logarítmica, não são muito marcantes. Além disso, a última porção dos gráficos é praticamente paralela comparando os métodos em cada exemplo, o que indica uma convergência similar.

Figura 43 – Gráfico do erro de aproximação da curva cúbica pelo número de arcos comparando dois métodos

Fonte: autoria própria

Figura 44 – Gráfico do erro de aproximação da curva quártica pelo número de arcos comparando dois métodos

Fonte: autoria própria

Aproveitamos estes dois exemplos para conferir também como comporta-se a função objetivo do problema de minimização (Equação (3.65)) com o acréscimo de arcos. No gráfico da Figura 45 estão registrados em escala logarítmica todos os valores obtidos para a função

objetivo ao aproximar a curva cúbica com 1 arco a 32 arcos e para aproximar a curva quártica a partir de 4 arcos (a curva quártica exigiu mais arcos para uma boa solução inicial). Embora tenha ocorrido um aumento do erro de aproximação em alguns pontos do processo, o método garante que a função objetivo seja decrescente a cada novo arco enquanto a otimização convergir.

Figura 45 – Gráfico da função objetivo do processo de minimização pelo número de arcos nos dois exemplos

Fonte: autoria própria

O gráfico também apresenta linhas de tendência exponenciais para cada exemplo. A convergência parece similar nos dois exemplos. Em escala logarítmica, as linhas de tendência são retas com as equações $y = 1,54 - 0,202x$ para a cúbica e $y = 2,85 - 0,162x$ para a quártica. As inclinações das linhas de tendência, que são próximas nos dois exemplos, indicam a taxa de convergência. Excluindo os três primeiros pontos do gráfico referente à curva cúbica, a linha de tendência torna-se $y = 0,870 - 0,171x$, as inclinações ficam ainda mais próximas comparando dados no mesmo intervalo de número de arcos.

4.5 Perfis ferroviários

A finalidade prática deste desenvolvimento é o ajuste de perfis de rodas e trilhos em uso em ferrovias possibilitando a modelagem computacional dos elementos ferroviários. A descrição do perfil segmentada em arcos de circunferência é compatível com o programa de elementos finitos Giraffe, que vem sendo usado em investigações numéricas de fenômenos de contato roda-trilho, como impacto por exemplo.

4.5.1 Sobre os parâmetros de entrada do método de ajuste

Para transformar o método de ajuste em um aplicativo de fácil utilização, o número de valores de entrada que devem ser fornecidos pelo usuário foi reduzido tanto quanto possível. Na versão atual, o programa requer quatro parâmetros de entrada: quantidade de pontos usados na estimativa de curvatura, número mínimo e número máximo de arcos, e tolerância. Com a prática, valores padrão tais que o ajuste seja bem-sucedido na maioria dos casos foram determinados.

Quantidade de pontos usados na estimativa de curvatura

O gráfico de curvatura é importante em duas tarefas: escolher os intervalos de cada arco (partição) e fornecer uma aproximação inicial para o método de otimização. A primeira tarefa é prejudicada quando o gráfico é ruidoso. Portanto, a quantidade de pontos deve ser grande o suficiente para que o gráfico seja suave. O número, entretanto, não pode ser excessivamente grande senão o gráfico de curvatura fica descaracterizado e uma aproximação nele baseada pode não ser boa o bastante para fazer convergir a otimização. A quantidade de pontos 41 (o ponto em que se estima mais 20 anteriores e 20 posteriores) levou a bons resultados para a maioria dos casos. Como regra prática, a quantidade escolhida deve ser da ordem de 1/10 do total de amostras (que fica entre 500 e 600).

Figura 46 – Exemplo 1: perfil de trilho com amostragem irregular

Fonte: autoria própria

Um exemplo em que o número 41 é excessivo é representado na Figura 46. Neste caso a amostragem é irregular, pouco densa nas laterais. Não será boa qualquer janela em torno dos

pontos isolados que vá muito além dos seus vizinhos mais próximos. Uma janela de 5 pontos resulta em um gráfico de curvatura mais representativo (veja Figura 47).

Figura 47 – Gráfico de curvatura estimado do exemplo 1 (a) janela de 41 pontos (b) janela de 5 pontos

Fonte: autoria própria

Número mínimo e número máximo de arcos

Desde que a curvatura estimada seja representativa, o número de arcos da primeira solução pode ser tão grande quanto se queira a princípio. Para perfis de roda e trilho, dez arcos costumam ser suficientes e a solução é encontrada em alguns segundos.

O número máximo de arcos é principalmente um limitador para que o problema de otimização não tenha tantas variáveis. Quanto mais arcos, mais variáveis e mais demorada é a resolução do problema de otimização. O mesmo vale para o problema de contato, que mais tempo tende a levar para ser resolvido quanto maior o número de arcos. O número máximo de arcos também serve como um critério de parada caso o problema seja de difícil convergência.

O método de otimização começa a ficar perceptivelmente lento a partir de 30 arcos. Portanto, este foi o número escolhido como padrão de máximo de arcos.

Tolerância

O principal critério usado para escolher a tolerância foi o visual. Quando a razão entre o erro de aproximação e o comprimento da curva fica abaixo de $1/1000$, a curva aproximada fica quase indistinguível visualmente dos pontos na escala real. Tendo em vista que as curvas amostradas têm em torno de 160 mm de comprimento, o erro deve estar abaixo de 0,16 mm. Em alguns casos, porém, o método falhou antes de produzir uma aproximação atendendo a esta tolerância. A tolerância padrão escolhida foi 0,5 mm, que foi atingida na grande maioria dos casos testados. Quando possível, a tolerância é reduzida ao valor “ideal” em torno de 0,16 mm.

Evidentemente, a avaliação visual é subjetiva, não sendo o melhor critério para se determinar se a aproximação é suficientemente boa ou não. Um critério mais objetivo pode ser

relacionado à aplicação em simulações de contato. No futuro, pretendemos investigar a relação entre o erro de aproximação e os resultados das simulações para saber quão sensível ela é.

4.5.2 Fluxo de trabalho

A padronização dos parâmetros de entrada facilita muito o uso da ferramenta. Casos que convergem facilmente são resolvidos com 10 a 15 arcos. Acompanhando a convergência a cada arco adicional, é possível perceber se o método está lidando bem com o conjunto de pontos. Se estiver, há boas chances de o acréscimo de arcos reduzir ainda mais o erro.

Caso não ocorra convergência, verificamos a aproximação baseada na função de curvatura. Se estiver muito distante do conjunto de pontos, a curvatura estimada provavelmente não representa a curva original. A experiência tem mostrado que a redução da quantidade de pontos usados na estimativa facilita a convergência. Em outros casos, simplesmente mudar a quantidade inicial de arcos, não necessariamente aumentando, coloca a curva em uma região do “espaço de soluções possíveis” na qual o procedimento de otimização trabalhe melhor.

O exemplo da Figura 46 converge iniciando com 10 arcos quando a quantidade de pontos usados para estimar curvatura é 3, ainda um pouco menor que o 5, que parecia suficiente pela Figura 47. O resultado é apresentado na Figura 48, o erro de aproximação é de 0,4958 mm.

Figura 48 – Aproximação do exemplo 1 com 15 arcos e gráfico de curvatura

Fonte: autoria própria

Trabalhando com estes dois parâmetros, espera-se conseguir uma aproximação satisfatória. Pode ocorrer do método falhar, caso no qual é necessário relaxar o critério de tolerância e aceitar uma curva com erro superior à tolerância padrão. Nos testes com perfil de roda e trilho, as piores aproximações tiveram erro no máximo de 1% do comprimento da curva (isto é, em torno de 1,6 mm), o que consideramos aceitável pela raridade deste tipo de ocorrência.

É digna de nota a possível presença de *outliers* entre as amostras. Veja a Figura 49, os pontos destacados são claramente anômalos. Como não há mecanismo implementado para lidar com erros de fontes diversas e considera-se que todos os pontos têm comportamento semelhante, o programa não é capaz de providenciar uma aproximação satisfatória automaticamente. Cabe ao usuário a tarefa de identificar e excluir prováveis outliers.

Figura 49 – Exemplo 2: perfil de roda com pontos anômalos

Fonte: autoria própria

Apagando as anomalias, o perfil é facilmente aproximado com os parâmetros padrão. O resultado é mostrado na Figura 50, o erro é 0,44 mm. Há certamente mais arcos do que o necessário no trecho mais retilíneo do perfil (pista da roda). Começando o processo com 5 arcos, uma solução final com 7 e erro 0,46 mm é obtida.

Figura 50 – Aproximação do exemplo 2 com 10 arcos e gráfico de curvatura

Fonte: autoria própria

4.5.3 Trilho usinado

A usinagem de boletos de trilhos é feita com perfis formados por arcos circulares, logo, espera-se que o gráfico de curvatura do contorno da seção de um trilho recém-usinado tenha patamares constantes e descontinuidades. O perfil de trilho exemplo ilustrado na Figura 46 tem

uma sequência de 390 amostras. Como visto na Figura 47 (b), é possível identificar os trechos de curvatura constante do perfil. A partição da curva será feita “manualmente” neste exemplo usando os trechos identificados.

A Figura 51 mostra o gráfico de curvatura do perfil estimado por interpolação quadrática. Existe uma grande perturbação no gráfico correspondente a pontos próximos das extremidades da curva.

Figura 51 – Gráfico de curvatura estimado da amostragem de um trilho recém-usinado por interpolação quadrática

Fonte: autoria própria

A Figura 52 traz uma versão com o intervalo de curvatura limitado a valores razoáveis focando na informação relevante. Duas propriedades interessantes deste gráfico são sua simetria e a formação de patamares constantes característicos de segmentos circulares, desde que sejam ignoradas as pequenas perturbações cuja causa atribuímos ao ruído (erro de medição).

Figura 52 – Gráfico de curvatura estimado da amostragem de um trilho usinado por interpolação quadrática com conjunto imagem limitado a um intervalo razoável

Fonte: autoria própria

O gráfico permite identificar onze trechos no perfil, dos quais cinco são praticamente retos: o início e o fim da curva, as duas faces laterais e o topo do boleto, veja a Figura 53. Tais trechos são condizentes com o que se considera um perfil de trilho típico como o ilustrado na Figura 54.

Figura 53 – Gráfico de curvatura separado nos onze trechos de curvatura quase constante identificados

Fonte: autoria própria

Figura 54 – Perfil de trilho típico

Fonte: [20] (adaptado de [46])

Usando a partição indicada na Figura 53, obtemos a aproximação constante por partes exposta na Figura 55 pelo valor médio de curvatura em cada subintervalo. Esta aproximação constante por trechos resulta na curva de aproximação da Figura 56 (a) e, após a otimização, na curva da Figura 56 (b).

Figura 55 – Gráfico de curvatura com aproximação constante por partes

Fonte: autoria própria

Este exemplo mostra como o gráfico de curvatura pode orientar um usuário a definir manualmente o domínio de cada arco. A definição automática da partição é capaz de identificar os pontos de descontinuidade, porém, não com o número “correto” de onze arcos. Uma das dificuldades que o método enfrenta está na mudança de densidade da amostragem, os pontos das laterais estão isolados das demais amostras. A solução final apresenta a máxima distância de 0,68 mm entre uma amostra e a curva de aproximação.

Figura 56 – Curvas de ajuste a um perfil de trilho recém-usinado (a) antes da otimização e (b) depois da otimização

Fonte: autoria própria

O processo de otimização modifica principalmente os dois últimos arcos. O penúltimo tem o valor absoluto da curvatura elevado e, em compensação, o último arco muda de orientação. As curvaturas obtidas antes e depois da otimização podem ser conferidas na Tabela 4.

Tabela 4 – Curvaturas dos arcos de ajuste a um perfil de trilho recém-usinado

Arco	Curvatura antes da otimização	Curvatura depois da otimização
1	0,007245	0,002304
2	-0,117554	-0,115981
3	-0,018179	-0,014519
4	-0,069479	-0,075909
5	-0,031023	-0,030400
6	-0,004158	-0,003938
7	-0,032255	-0,032191
8	-0,066990	-0,066600
9	-0,023157	-0,015843
10	-0,101381	-0,146587
11	-0,015554	0,067273

4.5.4 Roda ferroviária

Neste exemplo, o ajuste de um perfil de roda ferroviária é explorado detalhadamente. O perfil é dado como uma lista de 502 pontos, cujas posições são mostradas na Figura 57.

Figura 57 – Amostragem de um perfil de roda ferroviária (fora de escala)

Fonte: autoria própria

A curvatura em cada ponto foi obtida por regressão quadrática a subconjuntos de 21 pontos (o ponto central mais dez pontos à esquerda e à direita). A Figura 58 traz o gráfico de curvatura estimada em função do comprimento de arco.

Figura 58 – Gráfico de curvatura estimado da amostragem de um perfil de roda ferroviária por regressão quadrática

Fonte: autoria própria

A geometria típica de uma roda ferroviária está ilustrada na Figura 59. A parte interna da roda tem uma protuberância chamada de friso. A pista de rolamento é o segmento retilíneo central do perfil, sua inclinação define conicidade da roda. Há um segmento de transição entre o friso e a pista, a raiz do friso. Entre o friso e a raiz há um ponto de inflexão. A parte externa tem um chanfro que faz a concordância entre a pista e a face de campo. O perfil tem pelo menos quatro segmentos bem distintos.

Figura 59 – Perfil de roda típica

Fonte: [20] (adaptado de [46])

Iniciamos o procedimento de ajuste com quatro arcos e escolhemos sete arcos como ponto de parada. Na Figura 60, podemos observar a sequência de quatro aproximações obtida e como a spline de arcos aproxima-se do perfil a cada arco criado. A Tabela 5 contém os erros de aproximação em cada caso garantindo que há melhoria. O erro final de 0,24866 mm é igual a aproximadamente 0,16% do comprimento da curva, que é 153 mm.

Figura 60 – Sequência de aproximações ao perfil de roda incrementando o número de arcos

Fonte: autoria própria

Tabela 5 – Erro de aproximação de ajuste de um perfil de roda

Número de arcos	Erro de aproximação (mm)
4	1,6694
5	1,2718
6	0,5488
7	0,2487

Esta sequência de soluções convergiu em todas as etapas. Isso significa que a cada arco acrescentado, uma nova variável de otimização é considerada, entretanto, a forma geométrica do ponto de partida é idêntica à solução anterior.

Uma forma de visualizar a convergência do método de otimização é através do gráfico da norma do gradiente da função objetivo. A Figura 61 apresenta os gráficos nos quatro casos. O método para quando a redução da função modelo (Equação (2.50)) fica abaixo de $1\text{E}-16$ o que resulta em gradientes com magnitude em torno de $1\text{E}-6$ ao fim da otimização.

Figura 61 – Gráfico da norma do gradiente da função objetivo do processo de minimização

Fonte: autoria própria

4.5.5 Perfis de roda incluindo irregularidades

Nesta seção são ilustrados mais exemplos de ajuste de perfis de roda. São oito perfis divididos em três grupos: quatro rodas usinadas, duas rodas de friso fino e duas rodas cavas. Os dois últimos são grupos de rodas desgastadas. Em todos os exemplos, foi usada uma janela de 51 pontos na estimativa do gráfico de curvatura.

Rodas usinadas

As quatro rodas a seguir foram usinadas com base em um único perfil de referência. Portanto, espera-se que as aproximações sejam similares. As quatro aproximações atendem à tolerância padrão de 0,5 mm. O máximo erro observado de cada aproximação é, na ordem em que aparecem na Figura 62, 0,49 mm, 0,37 mm, 0,40 mm e 0,43 mm.

Os raios dos arcos são mostrados na Tabela 6. Raios de arcos identificados como correspondentes são exibidos em uma mesma linha para fácil comparação. As correspondências, principalmente dos arcos 1 a 7, referentes ao friso, indicam a similaridade entre as curvas de ajuste. Os arcos da pista da roda têm raios suficientemente grandes para serem confundidos com segmentos de reta.

Tabela 6 – Raios dos arcos que compõe as aproximações de perfis usinados

Arco	Raio (mm)			
	USI1	USI2	USI3	USI4
1	42,89	30,78	23,38	37,73
2	9,457	8,237	10,62	10,98
3	13,61	14,41		12,67
4	839,4			
5	12,74	17,02	16,11	18,02
6	36,16	36,66	38,15	35,40
7	77,75	128,7	51,72	86,12
8	10800	1578	2052	36390
9	1930	1612	863,9	2090
10	2153	5603	582,7	3439
11	2731	1578	5362	4376
12	545,5	1153	90,16	1,378

Figura 62 – Aproximações de quatro rodas usinadas com tolerância de 0,5 mm

Fonte: autoria própria

Rodas de friso fino

As duas próximas rodas foram classificadas como de friso fino. O desgaste da roda tende a reduzir a largura do friso e aumentar sua altura (por perda de material da pista e da raiz do friso). A tolerância usada é novamente 0,5 mm. O máximo erro observado na aproximação do lado esquerdo da Figura 63 é 0,32 mm e na aproximação do lado direito é 0,43 mm.

Figura 63 – Aproximações de duas rodas de friso fino com tolerância de 0,5 mm

Fonte: autoria própria

O desgaste afeta o gráfico de curvatura dos perfis de forma interessante. A curvatura estimada do perfil mais desgastado (FINO2) é mostrada na Figura 64. Comparando com o gráfico exposto na Figura 50, referente a uma roda usinada, podemos perceber que a curvatura do trecho correspondente ao friso fino é mais acentuada. A saber, o raio do arco do topo do friso (o terceiro arco) é 6,919 mm, sensivelmente menor que o raio do arco do topo do friso do perfil USI3, que tem raio 10,62 mm.

O segundo efeito é o desgaste da pista da roda. O oitavo arco do perfil FINO2 faz parte da pista, porém, é perceptivelmente curvo. O raio deste arco é 168,1 mm. Isto cria uma concavidade na pista da roda.

Figura 64 – Gráfico de curvatura do perfil de roda de friso fino

Fonte: autoria própria

Rodas cavas

As duas últimas rodas foram classificadas como cavas. Rodas cavas são caracterizadas por uma concavidade na pista da roda criada pelo agravamento do desgaste nesta região. A tolerância usada é 0,5 mm. O método de otimização convergiu com o número mínimo de arcos

dado, 10. O máximo erro observado na aproximação do lado esquerdo da Figura 65 é 0,15 mm e na aproximação do lado direito é 0,33 mm.

Figura 65 – Aproximações de duas rodas cavas com tolerância de 0,5 mm acompanhadas dos gráficos de curvatura

Fonte: autoria própria

O perfil CAVA1 tem concavidade na pista semelhante à observada no perfil FINO2 sem apresentar afinamento do friso. O raio do sétimo arco do perfil CAVA1 (o segmento “cavo”) é 152,0 mm. Já o perfil CAVA2 tem sua concavidade mais afastada do friso, vê-se no gráfico de curvatura que há um segmento quase retilíneo na pista antes da concavidade. O raio do sexto arco (segmento quase retilíneo) tem raio 694,1 mm, enquanto os raios dos dois arcos seguintes, onde a concavidade é manifestada, são respectivamente 175,6 mm e 234,8 mm. A concavidade é menos acentuada já que os raios de curvatura são maiores.

5 De curvas polinomiais a NURBS

Neste capítulo, discutiremos o tipo de curva racional chamado de NURBS (*Non-Uniform Rational B-Spline*), definido por um conjunto de pontos de controle com pesos e funções base racionais por partes chamadas de B-Splines racionais. Estas curvas são o padrão usado em CAD/CAGD [47] por sua capacidade tanto de representar formas analíticas (linhas retas e seções cônicas) quanto de desenhar formas livres. Nosso objetivo é dar uma representação NURBS para as splines de arcos.

Partimos de curvas polinomiais. Introduzimos a representação de Bézier para curvas polinomiais. Generalizamos as curvas polinomiais de Bézier em curvas racionais. Por fim, tratamos das curvas polinomiais por partes (B-Splines) e sua generalização racional (NURBS).

Uma apresentação detalhada destes assuntos pode ser encontrada no trabalho de Piegl e Tiller [48]. Cottrell *et al.* [49] dão um bom resumo sobre NURBS por sua importância para a análise isogeométrica como uma base para unificar CAGD e MEF. Rocha [50] também discorre sobre NURBS e sua aplicação em análise isogeométrica. O texto de Araújo [51] aborda curvas utilizadas em CAGD baseado em princípios matemáticos mais gerais focando em curvas de Bézier e B-Splines.

5.1 Curvas polinomiais

Curvas polinomiais formam uma classe particular de curvas de grande interesse. Isso porque são relativamente simples de analisar, já que, como denuncia o nome, suas funções coordenadas são polinômios. Um exemplo particular foi visto na Seção 1, a curva parametrizada (1.18) tem funções coordenadas polinomiais

$$\begin{aligned}x(t) &= t \\ y(t) &= t^2,\end{aligned}\tag{5.1}$$

sendo a coordenada x do primeiro grau em t e a y , do segundo. O grau da curva é igual ao maior grau de polinômio nas funções coordenadas, portanto, dizemos que a parábola é uma curva de segundo grau ou quadrática.

Uma curva quadrática genérica pode ser escrita como

$$\begin{aligned}x(t) &= x_0 + x_1 t + x_2 t^2 \\ y(t) &= y_0 + y_1 t + y_2 t^2,\end{aligned}\tag{5.2}$$

ou ainda

$$\gamma(t) = (x(t), y(t)) = (x_0, y_0) + (x_1, y_1)t + (x_2, y_2)t^2 = P_0 + P_1t + P_2t^2. \quad (5.3)$$

Note que a curva se expressa como um polinômio de segundo grau com coeficientes P_0 , P_1 e P_2 , que não são simples números, mas, em duas dimensões, pares ordenados denominados pontos de controle. Dados os três pontos, temos um elemento particular do conjunto de curvas quadráticas, cada função coordenada sendo elemento do conjunto de polinômios de grau menor ou igual a dois. Este conjunto é um espaço vetorial de dimensão finita caracterizado pela base de três elementos $\{1, t, t^2\}$, chamada de base de potências. A curva é uma combinação linear dos vetores base, expressa pela soma

$$\gamma(t) = \sum_{i=0}^2 P_i t^i. \quad (5.4)$$

A generalização da Equação (5.4) leva às curvas polinomiais de grau até n . Multiplicando cada elemento do conjunto $\{1, t, \dots, t^n\}$ por um ponto de controle P_i e fazendo a soma

$$\gamma(t) = \sum_{i=0}^n P_i t^i = P_0 + P_1t + \dots + P_n t^n. \quad (5.5)$$

O conjunto de $n + 1$ elementos $\{1, t, \dots, t^n\}$ é a base de potências do espaço de polinômios de grau menor ou igual a n , com $n + 1$ dimensões.

5.2 Representação de Bézier para curvas polinomiais

A forma para as curvas polinomiais na base de potências $\{1, t, \dots, t^n\}$ é uma entre infinitas possibilidades. Do ponto de vista geométrico, esta forma não é a mais conveniente, pois em geral não há uma interpretação geométrica clara para os pontos que multiplicam os elementos da base. Existe uma base especial na qual os pontos de controle obtêm propriedades geométricas que justificam chamá-los assim, já que a imagem da curva é controlada pelo posicionamento dos pontos.

Bézier, enquanto trabalhava para a Renault, e de Casteljau, enquanto trabalhava para a Citroën, descobriram independentemente formas equivalentes de representar e calcular curvas polinomiais a partir de “polígonos de controle” [51]. As descobertas de de Casteljau são anteriores, porém, foram mantidas confidenciais. Por isso, essas curvas ficaram conhecidas pelo

nome de Bézier, que as publicou primeiro. O primeiro algoritmo descoberto para computá-las, no entanto, leva o nome de de Casteljau [52].

5.2.1 Curvas quadráticas

A representação de Bézier para curvas quadráticas utiliza a base quadrática de Bernstein $\{B_{0,2}(t) = (1 - t)^2, B_{1,2}(t) = 2t(1 - t), B_{2,2}(t) = t^2\}$. Os gráficos dos polinômios de Bernstein no intervalo $[0, 1]$ são mostrados na Figura 66. Estes polinômios são não nulos em todo o interior do intervalo, entretanto, cada polinômio tem uma região de maior “influência”. O polinômio $B_{0,2}$, por exemplo, tem maior “influência” comparado aos demais na porção inicial da curva. É esta característica da base que nos permite intuir que cada ponto de controle dominará a posição de uma determinada porção da curva. Diz-se que a curva de Bézier tem um controle pseudo-local [51].

Figura 66 – Gráficos dos polinômios quadráticos de Bernstein

Fonte: autoria própria

A curva de Bézier quadrática geral é escrita como

$$\gamma(t) = \sum_{i=0}^2 P_i B_{i,2}(t) = P_0(1 - t)^2 + P_1[2t(1 - t)] + P_2 t^2, t \in [0, 1], \quad (5.6)$$

note que o parâmetro é limitado ao intervalo $[0, 1]$. Um exemplo pode ser visto na Figura 67. É interessante observar que a imagem da curva é uma aproximação suavizada da linha composta pelos segmentos retos que ligam os pontos de controle, chamada de polígono de controle. A base utilizada tem propriedades tais que a forma do polígono de controle determina aspectos geométricos da curva de Bézier, daí os nomes pontos de controle e polígono de controle.

Figura 67 – Curva de Bézier quadrática com pontos de controle $\{(0, 1), (1, 0), (2, 1)\}$

Fonte: autoria própria

Podemos destacar duas propriedades observáveis na Figura 67 e facilmente verificáveis:

- os pontos de controle P_0 e P_2 são as extremidades da curva;
- os segmentos $\overline{P_0P_1}$ e $\overline{P_1P_2}$ tangenciam a curva nas extremidades P_0 e P_2 respectivamente.

A primeira propriedade é verificada por $\gamma(0) = P_0$ e $\gamma(1) = P_2$. A segunda propriedade vem da expressão para o vetor tangente à curva

$$\frac{\gamma'(t)}{2} = (P_1 - P_0)(1 - t) + (P_2 - P_1)t, t \in [0, 1]. \quad (5.7)$$

Na extremidade $\gamma(0) = P_0$, o vetor tangente é paralelo a $P_1 - P_0$, consequentemente, a curva tangencia a reta através dos pontos P_0 e P_1 em P_0 . Da mesma forma, o vetor tangente para $t = 1$ é paralelo a $P_2 - P_1$ e a reta através de P_2 e P_1 é a tangente à curva pela extremidade P_2 . Os pontos de controle P_0 e P_2 determinam onde a curva começa e termina respectivamente, e o ponto P_1 , por ser a intersecção das tangentes, as controla.

5.2.2 Exemplo. Gráfico de uma parábola

Estes dois fatos nos permitem determinar, por exemplo, a representação de Bézier para o gráfico da função $f(x) = x^2$, $0 \leq x \leq 1$. As duas extremidades do gráfico são $P_0 = (0, 0)$ e $P_2 = (1, 1)$. A reta tangente através de P_0 é horizontal (reta $y = 0$), portanto, $P_1 = (x_1, 0)$. A abscissa x_1 é dada pelo encontro da reta tangente através de P_2 com o eixo x , esta reta é dada parametricamente pelas equações

$$\begin{aligned} x(\lambda) &= 1 + \lambda \\ y(\lambda) &= 1 + 2\lambda \end{aligned} \quad (5.8)$$

que para $y = 0$ resulta em $x_1 = 1/2$. Assim, na representação de Bézier, o gráfico de $f(x) = x^2$ é dado por

$$G_f(t) = \left(\frac{1}{2}, 0\right) [2t(1-t)] + (1, 1)t^2 = (t, t^2). \quad (5.9)$$

A parametrização habitual é retomada. A vantagem da representação de Bézier está em ter acesso aos pontos de controle, que possibilitam modificar a imagem intuitivamente os manipulando. O segmento parabólico com os pontos de controle pode ser visto na Figura 68.

Figura 68 – Gráfico da função $f(x) = x^2, 0 \leq x \leq 1$, na representação de Bézier

Fonte: autoria própria

A base de Bernstein tem duas características importantes:

- (não-negatividade) $B_{i,2}(t) \geq 0$, para $t \in [0,1]$;
- (partição da unidade) $B_{0,2}(t) + B_{1,2}(t) + B_{2,2}(t) = 1, \forall t$.

Consequentemente, cada ponto da curva é uma média ponderada pelo valor dos polinômios de Bernstein. Disso decorre a seguinte propriedade das curvas de Bézier:

- (propriedade do fecho convexo) a imagem da curva fica sempre contida no fecho convexo dos pontos de controle, que, para curvas quadráticas, é o triângulo cujos vértices são os pontos de controle.

O fecho convexo de um conjunto de pontos no plano é o menor polígono convexo que contém todos os pontos. Como curvas quadráticas têm apenas três pontos de controle e todo triângulo é convexo, o triângulo cujos vértices são os pontos de controle é o fecho convexo.

5.2.3 Curvas de grau n

A base de Bernstein de grau n tem como elementos os polinômios

$$B_{i,n}(t) = \binom{n}{i} t^i (1-t)^{n-i}, \quad (5.10)$$

que geram a curva

$$\gamma_n(t) = \sum_{i=0}^n P_i B_{i,n}(t), t \in [0, 1]. \quad (5.11)$$

A base tem $n + 1$ elementos e combina-se com $n + 1$ pontos de controle para formar uma curva.

Uma curva de Bézier linear ($n = 1$) demanda dois pontos de controle e é da forma

$$\gamma_1(t) = P_0(1 - t) + P_1 t. \quad (5.12)$$

A imagem é o segmento que liga os pontos P_0 e P_1 .

Curvas de grau 3 são mais interessantes. Em curvas de grau 3 há mais características geométricas realizáveis que em curvas de grau 2 como a presença de um ponto de inflexão Figura 69a ou um laço (autointersecção) Figura 69b. Além disso, com quatro ou mais pontos de controle, é possível produzir uma curva não plana.

Figura 69 – Curvas de Bézier cúbicas (a) com um ponto de inflexão, pontos de controle $\{(0, 0), (1, 1), (1, -1), (2, 0)\}$ (b) com um laço, pontos de controle $\{(1, 0), (3, 1), (0, 1), (2, 0)\}$

Fonte: autoria própria

Os gráficos dos polinômios de Bernstein de grau 3 são apresentados na Figura 70.

Figura 70 – Gráficos dos polinômios cúbicos de Bernstein

Fonte: autoria própria

As propriedades apontadas para curvas quadráticas são válidas para curvas de qualquer grau. Em particular, podemos observar nos dois exemplos da Figura 69 que o primeiro e o último pontos de controle são as extremidades da curva e o polígono de controle tangencia a curva nestes pontos. Além disso, as imagens estão contidas no fecho convexo dos pontos de controle.

5.3 Curvas racionais

Nesta seção, discutiremos uma classe de curvas mais abrangente do que as polinomiais, as curvas racionais. O nome curva racional vem do fato de que suas funções coordenadas são razões entre polinômios. O grau da curva é igual ao maior grau de polinômio presente nas funções coordenadas. Por exemplo, a curva com funções coordenadas

$$\begin{aligned} x(t) &= \frac{x_0 + x_1 t + x_2 t^2}{w_0 + w_1 t + w_2 t^2} \\ y(t) &= \frac{y_0 + y_1 t + y_2 t^2}{w_0 + w_1 t + w_2 t^2} \end{aligned} \quad (5.13)$$

é uma curva racional de segundo grau. Por ser mais abrangente, a classe de curvas racionais é capaz de modelar uma maior variedade de formas.

Seria interessante ter uma representação como a de Bézier para curvas racionais. A forma apresentada na Equação (5.13) sugere que se escreva para uma curva de grau n algo como

$$x(t) = \frac{\sum_{i=0}^n x_i B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)} \quad (5.14)$$

$$\gamma(t) = \frac{\sum_{i=0}^n y_i B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)}$$

ou, com pontos de controle $P_i = (x_i, y_i)$,

$$\gamma(t) = \frac{\sum_{i=0}^n P_i B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)}, \quad (5.15)$$

onde usamos a base de Bernstein para os polinômios dos numeradores e denominadores.

As funções racionais

$$\bar{R}_{i,n}(t) = \frac{B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)} \quad (5.16)$$

assumiriam o papel de base neste caso. O problema desta base é que ela não satisfaz a desejável propriedade de partição da unidade. No entanto, multiplicando por w_i , temos as funções racionais

$$R_{i,n}(t) = \frac{w_i B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)}, \quad (5.17)$$

que claramente formam uma partição da unidade. Estas funções formam a base racional com a qual se escreve

$$\gamma(t) = \sum_{i=0}^n P_i R_{i,n}(t) = \frac{\sum_{i=0}^n P_i w_i B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)}, t \in [0, 1]. \quad (5.18)$$

Os números w_i são chamadas de pesos. Estes devem ser positivos para que a base mantenha a propriedade de não-negatividade dos polinômios de Bernstein. A cada ponto de controle está associado um peso, que indica o quanto o ponto “atrai” a curva para si, quanto maior o peso mais a curva se aproxima do ponto de controle correspondente. Na verdade, o valor individual dos pesos não diz nada, as razões entre pesos é que desempenham um papel no formato da curva. Note que se todos os pesos forem multiplicados pelo mesmo valor, as funções $R_{i,n}$ não se alteram, portanto, é sempre possível arbitrar o valor de um dos pesos, porém, as razões são determinantes.

Figura 71 – Efeito do peso no formato de uma curva cúbica com pontos de controle $\{(0, 0), (1, 0), (1, 2), (2, 0)\}$, $w_0 = w_1 = w_3 = 1$ e w_2 varia dentro do conjunto $\{1/4, 1/2, 1, 2, 4\}$

Fonte: autoria própria

A Figura 71 mostra cinco curvas racionais cúbicas. Em todas, os pesos dos pontos P_0 , P_1 e P_3 são unitários ($w_0 = w_1 = w_3 = 1$) e o peso do ponto P_2 varia dentro do conjunto $\{1/4, 1/2, 1, 2, 4\}$. No caso em que todos os pesos são iguais, $R_{i,n}(t) = B_{i,n}(t)$ e a curva é polinomial. Quando w_2 é maior que os demais pesos ($w_2 = 2$ ou $w_2 = 4$), a curva é atraída para P_2 . Por outro lado, quando w_2 é menor que os demais pesos ($w_2 = 1/4$ ou $w_2 = 1/2$), a curva é repelida de P_2 .

As propriedades de curvas polinomiais destacadas na seção anterior continuam válidas para curvas racionais. Observe na Figura 71 que o primeiro e o último pontos de controle são as extremidades da curva, o polígono de controle é tangente nestes pontos e seu fecho convexo (o triângulo de vértices P_0, P_2, P_3) contém todas as curvas.

5.3.1 Curva racional como imagem projetada de uma curva polinomial

As curvas polinomiais têm como “graus de liberdade” as posições dos pontos de controle. As curvas racionais recebem “graus de liberdade” adicionais (os pesos), o que se reflete em uma capacidade de modelar uma maior variedade de formas. Os pesos w_i cumprem um papel semelhante ao das coordenadas (x_i, y_i) dos pontos de controle. Em certo sentido, cada peso é uma coordenada adicional do ponto de controle. A uma curva racional podemos associar uma curva polinomial com imagem em um espaço com uma dimensão extra.

Considere os pontos de controle localizados no plano $z = 1$, chamado de plano afim, as coordenadas dos pontos de controle são $(x_i, y_i, 1)$ neste caso. Considere os pontos em coordenadas homogêneas $\tilde{P}_i = (x_i w_i, y_i w_i, w_i)$, estes pontos são equivalentes aos pontos de

controle por projeção perspectiva sobre o plano afim através da origem. Nesta projeção, cada coordenada de um ponto é dividida pela coordenada z , assim, a projeção de $\tilde{P}_i$ é o ponto de controle P_i

$$\text{Proj}(\tilde{P}_i) = \text{Proj}(x_i w_i, y_i w_i, w_i) = (x_i, y_i, 1) = P_i. \quad (5.19)$$

A curva de Bézier com pontos de controle $\tilde{P}_i$

$$\tilde{\gamma}(t) = \sum_{i=0}^n \tilde{P}_i B_{i,n}(t) = \left(\sum_{i=0}^n x_i w_i B_{i,n}(t), \sum_{i=0}^n y_i w_i B_{i,n}(t), \sum_{i=0}^n w_i B_{i,n}(t) \right) \quad (5.20)$$

tem imagem no espaço tridimensional das coordenadas homogêneas. A projeção de $\tilde{\gamma}$ no plano afim resulta em

$$\gamma(t) = \left(\frac{\sum_{i=0}^n x_i w_i B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)}, \frac{\sum_{i=0}^n y_i w_i B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)}, 1 \right) = \frac{\sum_{i=0}^n P_i w_i B_{i,n}(t)}{\sum_{j=0}^n w_j B_{j,n}(t)}, \quad (5.21)$$

que é a curva racional com pontos de controle P_i e pesos w_i localizada no plano afim. Portanto, toda curva racional plana tem uma curva polinomial equivalente no espaço tridimensional de coordenadas homogêneas. A Figura 72 ilustra a projeção de uma curva quadrática.

Figura 72 – Projeção perspectiva sobre o plano afim de uma parábola no espaço tridimensional

Fonte: autoria própria

5.4 Parametrização racional de um arco de circunferência

As curvas polinomiais quadráticas são segmentos de parábola, um tipo de seção cônica. As curvas racionais quadráticas permitem representar exatamente segmentos das demais seções cônicas (hipérboles, elipses e circunferências), para isso basta definir corretamente os pesos [53, 47]. Nesta seção, veremos uma forma de representar arcos de circunferência (o nosso segmento de interesse) como curvas racionais quadráticas. Por simplicidade, a circunferência terá raio unitário, o raio pode ser modificado posteriormente como um fator de escala. Com uma transformação de escala não uniforme, arcos de circunferência podem também ser transformados em arcos de elipse.

Considerando a circunferência centrada na origem, as funções coordenadas devem atender à condição

$$[x(t)]^2 + [y(t)]^2 = 1, \forall t. \quad (5.22)$$

Um par de funções racionais adequado é

$$\begin{aligned} x(t) &= \frac{1 - t^2}{1 + t^2} \\ y(t) &= \frac{2t}{1 + t^2} \end{aligned} \quad (5.23)$$

O gráfico destas funções no intervalo $[0, 1]$ é mostrado na Figura 73. A coordenada x decresce de 1 a 0 e a coordenada y cresce de 0 a 1 no intervalo, logo, a imagem traçada é um arco circular indo do ponto $(1, 0)$ ao ponto $(0, 1)$, que é o quarto de circunferência no primeiro quadrante. A imagem da curva é mostrada na Figura 74 acompanhada dos pontos de controle (as extremidades $P_0 = (1, 0)$ e $P_2 = (0, 1)$ e o encontro das tangentes $P_1 = (1, 1)$).

Figura 73 – Gráficos das funções coordenadas de uma parametrização racional de um quarto de circunferência

Fonte: autoria própria

Figura 74 – Quarto de circunferência e pontos de controle

Fonte: autoria própria

Conhecidos os pontos de controle e a parametrização (5.23), podemos determinar os pesos comparando com a Equação (5.18) no caso quadrático. Os pesos são $w_0 = w_1 = 1$ e $w_2 = 2$.

5.4.1 Relação invariante entre os pesos

Para um arco de um dado ângulo θ existe uma condição necessária que os pesos devem atender. Para encontrar tal condição, consideramos o arco no sistema de coordenadas ilustrado na Figura 75, onde estão indicados os pontos de controle $P_0 = (1, 0)$ e $P_2 = (\cos \theta, \sin \theta)$. O ponto P_1 é dado pelo encontro das tangentes. Em P_0 , a reta tangente é a reta vertical $x = 1$, portanto a abscissa de P_1 é $x_1 = 1$. A bissetriz do ângulo θ passa por P_1 . O triângulo OP_0P_1 (sendo O o centro do arco) é retângulo e, portanto, $y_1 = \tan(\theta/2)$.

Figura 75 – Arco de circunferência com ângulo θ e pontos de controle

Fonte: autoria própria

Substituindo os pontos de controle na Equação (5.18), temos as funções coordenadas

$$\begin{aligned} x(t) &= \frac{w_0(1-t)^2 + w_1[2t(1-t)] + \cos \theta w_2 t^2}{w_0(1-t)^2 + w_1[2t(1-t)] + w_2 t^2} \\ y(t) &= \frac{\tan(\theta/2) w_1[2t(1-t)] + \sin \theta w_2 t^2}{w_0(1-t)^2 + w_1[2t(1-t)] + w_2 t^2} \end{aligned} \quad (5.24)$$

que devem obedecer à restrição (5.22) para todo t . Em particular, para $t = 1/2$

$$\begin{aligned} x\left(\frac{1}{2}\right) &= \frac{\frac{w_0}{4} + \frac{w_1}{2} + \frac{\cos \theta w_2}{4}}{\frac{w_0}{4} + \frac{w_1}{2} + \frac{w_2}{4}} \\ y\left(\frac{1}{2}\right) &= \frac{\frac{\tan(\theta/2) w_1}{2} + \frac{\sin \theta w_2}{4}}{\frac{w_0}{4} + \frac{w_1}{2} + \frac{w_2}{4}} \end{aligned} \quad (5.25)$$

e a restrição à circunferência resulta em

$$\begin{aligned} &\left(\frac{\frac{w_0}{4} + \frac{w_1}{2} + \frac{\cos \theta w_2}{4}}{\frac{w_0}{4} + \frac{w_1}{2} + \frac{w_2}{4}} \right)^2 + \left(\frac{\frac{\tan(\theta/2) w_1}{2} + \frac{\sin \theta w_2}{4}}{\frac{w_0}{4} + \frac{w_1}{2} + \frac{w_2}{4}} \right)^2 = 1 \\ \Rightarrow &\left(\frac{w_0}{4} + \frac{w_1}{2} + \frac{\cos \theta w_2}{4} \right)^2 + \left(\frac{\tan(\theta/2) w_1}{2} + \frac{\sin \theta w_2}{4} \right)^2 = \left(\frac{w_0}{4} + \frac{w_1}{2} + \frac{w_2}{4} \right)^2. \end{aligned} \quad (5.26)$$

Após o desenvolvimento dos quadrados e manipulação algébrica, chegamos à seguinte relação entre os pesos

$$\frac{w_1^2}{w_0 w_2} = \cos^2 \left(\frac{\theta}{2} \right). \quad (5.27)$$

Para $\theta = \pi/2$, os pesos devem obedecer

$$\frac{w_1^2}{w_0 w_2} = \frac{1}{2}. \quad (5.28)$$

Podemos arbitrariamente escolher $w_0 = 1$. Além disso, suponhamos que seja feita a reparametrização $u = 1 - t$, que corresponde à inversão do sentido em que a curva é percorrida, o peso w_2 assumiria o papel do peso w_0 , então, por simetria, podemos escolher $w_2 = w_0 = 1$. Daí,

$$w_1 = \frac{\sqrt{2}}{2}. \quad (5.29)$$

Com os três pesos, temos as funções coordenadas

$$x(t) = 1 - \frac{t^2}{1 - (2 - \sqrt{2})t + (2 - \sqrt{2})t^2} \quad (5.30)$$

e

$$y(t) = 1 - \frac{(1-t)^2}{1 - (2 - \sqrt{2})t + (2 - \sqrt{2})t^2}, \quad (5.31)$$

e é possível verificar que elas satisfazem a Equação (5.22) para qualquer $t \in [0, 1]$. A escolha $w_0 = w_1 = 1$ e $w_2 = 2$ também é adequada e resulta na parametrização (5.23), que se diferencia da anterior por não ser simétrica.

Para um θ arbitrário, a escolha simétrica $w_0 = w_2 = 1$ implica

$$w_1 = \cos\left(\frac{\theta}{2}\right). \quad (5.32)$$

Para que se tenha $w_1 > 0$, é necessário que seja $\theta < \pi$. De fato, um arco com $\theta = \pi$ (peso w_1 nulo) tem as retas tangentes nas extremidades paralelas, conseqüentemente não se pode determinar o ponto P_1 de intersecção (Figura 76 (a)). No limite $\theta \rightarrow \pi^-$ (isto é, $\theta < \pi$), o ponto P_1 tende ao infinito de tal forma que o produto $P_1 w_1$ tende ao vetor unitário vertical. Pode-se definir que, para $\theta = \pi$, $w_1 = 0$ e o “ponto de controle no infinito” P_1 é representado pelo “vetor de controle” correspondente [54]. Para um arco com $\theta > \pi$ (peso w_1 negativo), a intersecção das retas tangentes ocorre do “lado incorreto”, a curva violaria a propriedade do fecho convexo (Figura 76 (b)). O uso de pontos de controle no infinito e pesos negativos permite construir arcos quadráticos com $\theta \geq \pi$ [48].

Figura 76 – Arcos não parametrizáveis como curvas racionais quadráticas na representação de Bézier convencional (a) $\theta = \pi$ implica tangentes paralelas e um ponto de controle no infinito (b) $\theta > \pi$ implica o arco complementar como imagem e violação da propriedade de fecho convexo

Fonte: autoria própria

Curvas racionais de ordem mais elevada permitem representar arcos de ângulo maior sem a necessidade de pontos de controle no infinito ou pesos negativos. Curvas racionais cúbicas permitem arcos de ângulo até, não incluso, $4\pi/3$ [55, 56]. Curvas racionais de quarto grau permitem arcos de qualquer ângulo menor que 2π [57].

5.5 B-Splines

As curvas de Bézier são muito convenientes por serem moldáveis intuitivamente movimentando um pequeno conjunto de pontos de controle. Para conferir mais liberdade à forma da curva precisamos acrescentar pontos de controle, o que implica em um aumento de grau. O processamento de curvas de alto grau é menos eficiente e pode apresentar instabilidades numéricas [48]. Uma solução é usar curvas polinomiais por partes.

B-Splines são uma generalização da representação de Bézier. Uma B-Spline é uma curva formada por múltiplos segmentos polinomiais conectados. Este tipo de curva usa o mesmo sistema de pontos de controle, tendo uma expressão análoga à Equação (5.11)

$$\gamma(t) = \sum_{i=0}^n P_i N_{i,p}(t). \quad (5.33)$$

A diferença entre B-Splines e as curvas de Bézier está nas funções base utilizadas. B-Splines têm como base funções polinomiais por partes denotadas por $N_{i,p}$, onde p é o grau dos polinômios. O intervalo paramétrico é particionado em subintervalos dentro dos quais cada

função base tem uma expressão polinomial que se conecta com certo grau de continuidade com o polinômio no subintervalo vizinho. Os pontos entre subintervalos são chamados de nós. A continuidade das funções base é limitada nos nós a no máximo $p - 1$, um grau a menos que o da base, assim, uma função quadrática tem no máximo a primeira derivada contínua através dos subintervalos. A imagem de cada subintervalo é um segmento polinomial. Na Figura 77, temos uma B-Spline cúbica imagem do intervalo paramétrico $[0, 1]$, que foi dividido uniformemente em quatro partes, as cores indicam o segmento polinomial imagem de cada subintervalo.

Figura 77 – B-Spline cúbica com vetor de nós $\{0, 0, 0, 0, 1/4, 1/2, 1/2, 3/4, 3/4, 3/4, 1, 1, 1, 1\}$ e pontos de controle $\{(0, 0), (1, 1), (2, 1), (3, 0), (3, -3), (2, -3), (2, -1), (4, -1), (5, 0), (6, 2)\}$

Fonte: autoria própria

As funções polinomiais por partes que são base de B-Splines são produzidas através da fórmula de recursão de Cox – de Boor. A i -ésima função de grau p é dada por

$$N_{i,p}(t) = \frac{t - u_i}{u_{i+p} - u_i} N_{i,p-1}(t) + \frac{u_{i+1+p} - t}{u_{i+1+p} - u_{i+1}} N_{i+1,p-1}(t), \quad (5.34)$$

nota-se que uma função de grau p depende das funções de grau $p - 1$. A base da recursão são as funções de grau nulo

$$N_{i,0}(t) = \begin{cases} 1, & \text{se } u_i \leq t < u_{i+1} \\ 0, & \text{caso contrário} \end{cases}, \quad (5.35)$$

que são patamares constantes descontínuos em u_i e u_{i+1} .

As funções base dependem da escolha dos números u_i , que são os nós no intervalo paramétrico, ordenados em uma sequência não decrescente $U = \{u_0, u_1, \dots, u_i, \dots, u_{n+p}, u_{n+p+1}\}$, com $u_0 \leq u_1 \leq \dots \leq u_i \leq \dots \leq u_{n+p} \leq u_{n+p+1}$,

denominada vetor de nós (*knot vector*), sendo p o grau e $n + 1$ o número de pontos de controle da curva, o total de nós é $(n + 1) + p + 1$, a soma do número de pontos de controle e do grau mais 1, $p + 1$ é chamado de ordem. Desta forma, o número de pontos de controle para modelar a curva não fica limitado pelo seu grau, pode-se ter tantos pontos de controle quanto se queira desde que o vetor de nós seja de tamanho apropriado, o preço que se paga é a derivabilidade limitada nos nós.

A sequência de nós não é estritamente crescente, a repetição é permitida. Devido a isto, é possível que algumas das diferenças que aparecem nos denominadores da Equação (5.34) sejam nulas, neste caso, a parcela toda é considerada nula. A quantidade de vezes que um nó se repete é chamada de multiplicidade e tem um papel importante na continuidade das funções base e da curva. As funções base não são infinitamente contínuas nos nós, o grau de continuidade em determinado nó é dado por $p - s$, sendo s a multiplicidade do nó, o que limita a continuidade nos nós a no máximo $p - 1$. Consequentemente, a multiplicidade de um nó deve ser menor ou igual ao grau da curva para que ela seja no mínimo necessariamente contínua, com exceção dos nós de extremidade, em que não há preocupação com continuidade, estes podem ter multiplicidade até $p + 1$, a ordem da curva. No caso limite em que ambos os nós de extremidade têm multiplicidade igual à ordem, o vetor de nós é dito aberto, e as curvas resultantes interpolam os pontos de controle extremos. Para nós interiores, se a multiplicidade for p , então as funções base são contínuas, porém, não deriváveis no nó, e a imagem do nó coincide com um ponto de controle.

Chamamos de vão (*knot span*) o intervalo entre nós distintos consecutivos. Cada vão tem como imagem um segmento polinomial. O comprimento de um vão, a diferença entre os nós, não tem grande relevância em si, é na razão entre os comprimentos de vão que há informação sobre o aspecto das funções base, já que são razões deste tipo que estão presentes na fórmula de Cox – de Boor. Consideraremos apenas vetores de nós ditos uniformes, em que todos os vãos têm a mesma dimensão.

5.5.1 Exemplo. B-Spline quadrática com um único vão

Vamos considerar uma curva quadrática com apenas um segmento. Há apenas um vão no vetor de nós e a multiplicidade dos nós deve ser 3 ($= 2 + 1$). Podemos arbitrar que o intervalo paramétrico seja $[0, 1]$, logo, o vetor de nós é $U = \{u_0 = 0, u_1 = 0, u_2 = 0, u_3 = 1, u_4 = 1, u_5 = 1\}$. A quantidade de pontos de controle e de funções base necessárias é 3.

A fórmula de Cox – de Boor é aplicada a este vetor de nós resultando nas funções base. A primeira função de primeiro grau é dada por

$$N_{0,1}(t) = \frac{t - u_0}{u_1 - u_0} N_{0,0}(t) + \frac{u_2 - t}{u_2 - u_1} N_{1,0}(t) , \quad (5.36)$$

como $u_0 = u_1 = u_2 = 0$, ambas as parcelas são anuladas, portanto,

$$N_{0,1}(t) = 0 . \quad (5.37)$$

A segunda função de primeiro grau é dada por

$$N_{1,1}(t) = \frac{t - u_1}{u_2 - u_1} N_{1,0}(t) + \frac{u_3 - t}{u_3 - u_2} N_{2,0}(t) , \quad (5.38)$$

a primeira parcela é anulada, restando

$$N_{1,1}(t) = \frac{u_3 - t}{u_3 - u_2} N_{2,0}(t) = 1 - t . \quad (5.39)$$

A última função linear é

$$N_{2,1}(t) = \frac{t - u_2}{u_3 - u_2} N_{2,0}(t) + \frac{u_4 - t}{u_4 - u_3} N_{3,0}(t) , \quad (5.40)$$

a segunda parcela é anulada pois $u_3 = u_4 = 1$

$$N_{2,1}(t) = \frac{t - u_2}{u_3 - u_2} N_{2,0}(t) = t . \quad (5.41)$$

As funções de segundo grau são

$$N_{0,2}(t) = \frac{t - u_0}{u_2 - u_0} N_{0,1}(t) + \frac{u_3 - t}{u_3 - u_1} N_{1,1}(t) = (1 - t)^2 , \quad (5.42)$$

$$N_{1,2}(t) = \frac{t - u_1}{u_3 - u_1} N_{1,1}(t) + \frac{u_4 - t}{u_4 - u_2} N_{2,1}(t) = 2(1 - t)t \quad (5.43)$$

e

$$N_{2,2}(t) = \frac{t - u_2}{u_4 - u_2} N_{2,1}(t) = t^2 , \quad (5.44)$$

que nada mais são que os polinômios de Bernstein de segundo grau. Vale em geral que um vetor de nós aberto com um único vão produz através da fórmula de Cox – de Boor a base da representação de Bézier. As curvas de Bézier são, portanto, um caso particular de B-Splines.

5.5.2 Exemplo. B-Spline cúbica

As ferramentas de desenho geralmente usam B-Splines cúbicas para construção de formas livres. Isso porque grau 3 é o mínimo necessário para que a curva seja contínua até a segunda derivada em todos os pontos. A continuidade C^2 é suficiente para que a curva seja “visualmente suave”, daí a sua importância.

Neste exemplo, exploraremos mais detalhes da curva da Figura 77. A base usada para desenhar esta curva é calculada a partir do vetor de nós $\{0, 0, 0, 0, 1/4, 1/2, 1/2, 3/4, 3/4, 3/4, 1, 1, 1, 1\}$, que contém 14 nós. A curva é cúbica (grau 3, ou ordem 4). O número de pontos de controle é 10 ($= 14 - 4$).

Figura 78 – Gráficos das funções base cúbicas obtidas a partir do vetor de nós $\{0, 0, 0, 0, 1/4, 1/2, 1/2, 3/4, 3/4, 3/4, 1, 1, 1, 1\}$ (a multiplicidade dos nós é indicada pelo número de marcas verticais que os acompanham)

Fonte: autoria própria

Aplicando a fórmula de Cox – de Boor, obtemos as funções polinomiais cúbicas cujos gráficos são mostrados na Figura 78. As expressões para estas funções, nas partes em que não se anulam, são

$$\begin{aligned}
N_{0,3}(t) &= (1 - 4t)^3, t \in [0, 1/4) \\
N_{1,3}(t) &= \begin{cases} 4t[(1 - 4t)^2 + (1 - 4t)(1 - 2t) + (1 - 2t)^2], t \in [0, 1/4) \\ 2(1 - 2t)^3, t \in [1/4, 1/2) \end{cases} \\
N_{2,3}(t) &= \begin{cases} 8t^2(3 - 8t), t \in [0, 1/4) \\ 2(1 - 2t)^2(8t - 1), t \in [1/4, 1/2) \end{cases} \\
N_{3,3}(t) &= \begin{cases} 16t^3, t \in [0, 1/4) \\ 4t(1 - 2t)(6t - 1) + \frac{1}{2}(3 - 4t)(4t - 1)^2, t \in [1/4, 1/2) \\ \frac{1}{2}(3 - 4t)^3, t \in [1/2, 3/4) \end{cases} \\
N_{4,3}(t) &= \begin{cases} \frac{1}{2}(4t - 1)^3, t \in [1/4, 1/2) \\ \frac{1}{2}(20t - 9)(3 - 4t)^2, t \in [1/2, 3/4) \end{cases} \\
N_{5,3}(t) &= 12(3 - 4t)(2t - 1)^2, t \in [1/2, 3/4) \\
N_{6,3}(t) &= \begin{cases} 8(2t - 1)^3, t \in [1/2, 3/4) \\ 64(1 - t)^3, t \in [3/4, 1] \end{cases} \\
N_{7,3}(t) &= 48(1 - t)^2(4t - 3), t \in [3/4, 1] \\
N_{8,3}(t) &= 12(1 - t)(4t - 3)^2, t \in [3/4, 1] \\
N_{9,3}(t) &= (4t - 3)^3, t \in [3/4, 1]
\end{aligned} \tag{5.45}$$

A partir das expressões das funções polinomiais e seus gráficos, podemos verificar para este caso particular algumas propriedades que são válidas em geral (veja [51, 50] para demonstrações):

- (não-negatividade) $N_{i,p}(t) \geq 0, \forall t$;
- (partição da unidade) $\sum_{i=0}^n N_{i,p}(t) = 1, \forall t$;
- Em cada subintervalo, no máximo $p + 1$ (a ordem) funções base são não nulas, com isso, $p + 1$ pontos de controle influenciam o formato do segmento correspondente; para cúbicas, 4 funções base são não nulas em cada subintervalo;
- (propriedade forte de fecho convexo) cada segmento está contido no fecho convexo dos $p + 1$ pontos de controle que o influenciam (confira a Figura 79 para uma referência visual);
- (suporte local) $N_{i,p}(t)$ é não nula em no máximo p (o grau) subintervalos contíguos; para cúbicas, $N_{i,3}(t)$ é não nula em no máximo 3 subintervalos, isso significa que o movimento de um ponto de controle deforma um trecho limitado da curva, deixando o restante inalterado; pode-se dizer que o controle (local) de

B-Splines é mais eficiente do que o controle (pseudo-local) de curvas de Bézier [48].

Figura 79 – Propriedade forte de fecho convexo ilustrada: cada segmento está contido no polígono convexo formado pelos pontos de controle relevantes

Fonte: autoria própria

O nó $t = 1/2$, que tem multiplicidade 2, é influenciado por apenas dois pontos de controle, P_3 e P_4 , além disso, $N_{3,3}(1/2) = N_{4,3}(1/2) = 1/2$, consequentemente, a imagem do nó é $\gamma(1/2) = (P_3 + P_4)/2$, o ponto médio do segmento $\overline{P_3P_4}$. Mais do que isso, o segmento $\overline{P_3P_4}$ tangencia a curva neste ponto (veja a Figura 77).

Compare os gráficos das funções base (Figura 78) no quarto vão com os gráficos dos polinômios cúbicos de Bernstein (Figura 70). A menos de um fator de escala, eles são idênticos. No intervalo $[3/4, 1]$, as funções $N_{6,3}(t)$, $N_{7,3}(t)$, $N_{8,3}(t)$ e $N_{9,3}(t)$ são os polinômios de Bernstein de grau 3, embora estejam “comprimidos” e “deslocados”, a mudança de variável $s = 4(t - 3/4)$ coloca as expressões vistas na Equação (5.45) na forma (5.10). Consequentemente, o quarto segmento é exatamente uma curva cúbica de Bézier conectada ao restante da B-Spline. Isso acontece em todo vão cujos nós têm a máxima multiplicidade possível (o grau para nós interiores, a ordem para nós de extremidade).

Um último fato interessante sobre esta curva. Os segmentos $\overline{P_6P_7}$ e $\overline{P_5P_6}$ tangenciam a curva em P_6 . Sendo assim, mesmo que as funções base não sejam deriváveis no nó $t = 3/4$, o que gera um bico na imagem em geral, se P_6 for movido de modo a incidir sobre o segmento

$\overline{P_5P_7}$, a imagem terá tangente contínua (continuidade G^1). A Figura 80 mostra como a curva se transforma após alterarmos as posições dos pontos P_6 e P_7 deixando P_5 , P_6 e P_7 colineares. Note que estes pontos modificam somente o terceiro e o quarto segmentos (controle local).

Figura 80 – B-Spline deformada pelo deslocamento dos pontos de controle P_6 e P_7 , o alinhamento dos pontos P_5 , P_6 e P_7 confere suavidade à imagem

Fonte: autoria própria

5.6 NURBS

Da mesma forma que transformamos curvas polinomiais em racionais introduzindo pesos, transformamos B-Splines polinomiais em B-Splines racionais, conhecidas como NURBS (do inglês, *Non-Uniform Rational B-Splines*). Define-se uma curva NURBS como

$$\gamma(t) = \frac{\sum_{i=0}^n P_i w_i N_{i,p}(t)}{\sum_{j=0}^n w_j N_{j,p}(t)}, \quad (5.46)$$

onde $N_{i,p}(t)$ são funções base B-Spline de grau p em um dado vetor de nós.

5.6.1 Propriedades

Podemos escrever a curva na forma

$$\gamma(t) = \sum_{i=0}^n P_i R_{i,p}(t), \quad (5.47)$$

onde

$$R_{i,p}(t) = \frac{w_i N_{i,p}(t)}{\sum_{j=0}^n w_j N_{j,p}(t)} \quad (5.48)$$

são funções base B-Spline racionais.

A base racional conserva todas as propriedades da base polinomial listadas na Seção 5.5.2, resumidas em:

- não-negatividade;
- partição da unidade;
- suporte local.

Consequentemente, a curva NURBS tem as propriedades:

- controle local: cada ponto de controle afeta uma região limitada da curva
- propriedade forte de fecho convexo.

Uma lista mais detalhada de propriedades pode ser encontrada em [47].

Outra propriedade interessante: tanto B-Splines quanto curvas racionais de Bézier são casos particulares de NURBS. B-Splines são NURBS cujos pesos são todos iguais. Curvas racionais de Bézier são NURBS com um único segmento. Então, pode-se pensar que as propriedades de NURBS são herdadas pelas B-Splines e curvas de Bézier. O diagrama a seguir (Figura 81) relaciona as classes de curvas.

Figura 81 – Relações entre as classes de curvas

Fonte: autoria própria

5.7 Representação NURBS de um biarco

Um biarco é a curva composta por dois arcos de circunferência conectados de forma suave. Uma forma de construir um biarco NURBS é simplesmente conectar dois arcos [58]. Neste caso, teremos como possível vetor de nós $U = \{0, 0, 0, 1, 1, 2, 2, 2\}$, com dois vãos: $(0, 1)$ e $(1, 2)$. São 5 pontos de controle no total, sendo que cada arco contribui com 3, porém, dois pontos de controle coincidem. Aplicando a fórmula de Cox – de Boor obtemos as funções base

$$\begin{aligned}
N_{0,2}(t) &= (1-t)^2, t \in [0, 1) \\
N_{1,2}(t) &= 2t(1-t), t \in [0, 1) \\
N_{2,2}(t) &= \begin{cases} t^2, t \in [0, 1) \\ (2-t)^2, t \in [1, 2] \end{cases} \\
N_{3,2}(t) &= 2(t-1)(2-t), t \in [1, 2] \\
N_{4,2}(t) &= (t-1)^2, t \in [1, 2]
\end{aligned} \quad (5.49)$$

No intervalo $[0, 1)$, correspondente ao primeiro arco, $N_{0,2}$, $N_{1,2}$ e $N_{2,2}$ são os polinômios quadráticos de Bernstein na forma (5.10). No intervalo $[1, 2]$, correspondente ao segundo arco, $N_{2,2}$, $N_{3,2}$ e $N_{4,2}$ também são os polinômios quadráticos de Bernstein (deslocados). Desta forma, os dois segmentos seguem a representação de Bézier quase totalmente independentes um do outro, a única restrição é que o ponto de controle P_2 , que é a imagem de $t = 1$, é a extremidade comum dos segmentos. Para garantir a forma circular, os pesos podem ser escolhidos como se os segmentos fossem independentes, por exemplo, $w_0 = w_2 = w_4 = 1$, $w_1 = \cos(\theta_1/2)$ e $w_3 = \cos(\theta_2/2)$, onde θ_1 e θ_2 são os ângulos do primeiro e do segundo arco respectivamente.

Figura 82 – Biarco com 5 pontos de controle, $\theta_1 = \theta_2 = \pi/2$ e pesos $\{1, 1/\sqrt{2}, 1, 1/\sqrt{2}, 1\}$

Fonte: autoria própria

5.7.1 Representação com nó interior não duplicado

A continuidade geométrica do biarco demanda que P_2 esteja alinhado com os pontos P_1 e P_3 . Além disso, os raios dos arcos determinam a posição exata de P_2 no segmento $\overline{P_1P_3}$ (veja a Figura 82). Portanto, o ponto de controle P_2 não acrescenta informação.

A curva seria computacionalmente mais eficiente se o ponto P_2 pudesse ser excluído. Em primeiro lugar porque uma cópia do nó duplicado $t = 1$ também seria eliminada para

manter as quantidades de nós e pontos de controle compatíveis, e tal exclusão tornaria a derivada das funções base contínua, o que automaticamente garante que o biarco seja geometricamente contínuo. Em segundo lugar porque haveria um ponto de controle a menos para ser armazenado. Veremos que é possível remover o ponto P_2 e um nó.

Piegl e Tiller [48] descrevem um algoritmo de remoção de nó, porém a remoção de um nó de multiplicidade dois de uma curva quadrática é aplicável se, e somente se, a curva em coordenadas homogêneas seja contínua na primeira derivada. Como este não é o caso, devemos modificar os pesos a fim de tornar a curva em coordenadas homogêneas adequada para a remoção do nó (como feito em [59]).

Alternativamente, assumimos que um biarco possa ser modelado com quatro pontos de controle (identificados como na Figura 83) e encontramos as razões necessárias para os pesos por considerações geométricas.

Figura 83 – Biarco com 4 pontos de controle, $\theta_1 = \theta_2 = \pi/2$, $R_1 = 1$, $R_2 = 2$ e pesos $\{8/3, 1, 1/2, 2/3\}$

Fonte: autoria própria

No vetor de nós $\{0, 0, 0, 1, 2, 2, 2\}$, as funções base são

$$\begin{aligned} N_{0,2}(t) &= (1-t)^2, t \in [0, 1) \\ N_{1,2}(t) &= \begin{cases} \frac{1}{2}t(4-3t), t \in [0, 1) \\ \frac{1}{2}(2-t)^2, t \in [1, 2] \end{cases} \\ N_{2,2}(t) &= \begin{cases} \frac{1}{2}t^2, t \in [0, 1) \\ \frac{1}{2}(2-t)(3t-2), t \in [1, 2] \end{cases} \\ N_{3,2}(t) &= (t-1)^2, t \in [1, 2] \end{aligned} \quad (5.50)$$

ilustradas na Figura 84.

Figura 84 – Funções base quadráticas no vetor de nós $\{0, 0, 0, 1, 2, 2, 2\}$

Fonte: autoria própria

No nó $t = 1$, apenas as funções base $N_{1,2}$ e $N_{2,2}$ são não nulas, então a imagem do nó, o ponto de conexão N entre os arcos, necessariamente pertence ao segmento $\overline{P_1P_2}$. Além disso, pelas derivadas verifica-se que a curva tangencia o segmento $\overline{P_1P_2}$. A posição de N no segmento é determinada pela razão entre os pesos w_1 e w_2 . Como as únicas funções base não nulas são iguais em $t = 1$,

$$N = \frac{P_1 w_1 + P_2 w_2}{w_1 + w_2} = P_1 + \frac{w_2}{w_1 + w_2} (P_2 - P_1) \quad (5.51)$$

ou

$$(N - P_1) = \frac{w_2}{w_1 + w_2} (P_2 - P_1) \quad (5.52)$$

Conforme a Figura 83, a distância entre P_1 e N é L_1 e a distância entre P_1 e P_2 é $L_1 + L_2$. Logo, calculando a norma de ambos os membros da Equação (5.52), temos

$$L_1 = \frac{w_2}{w_1 + w_2} (L_1 + L_2), \quad (5.53)$$

que equivale à relação inversamente proporcional entre pesos e comprimentos

$$\frac{w_2}{w_1} = \frac{L_1}{L_2}. \quad (5.54)$$

A expressão anterior intuitivamente faz todo sentido, já que o ponto de controle com maior peso “atrai” a curva para si, assim a distância até o nó tem que ser menor.

O comprimento L_1 depende apenas do raio e ângulo do primeiro arco

$$L_1 = R_1 \tan\left(\frac{\theta_1}{2}\right). \quad (5.55)$$

Uma expressão análoga vale para L_2 . Portanto, dados os arcos, a razão entre os pesos w_1 e w_2 é determinada.

A razão entre pesos presente na Equação (5.54) garante que a intersecção da curva com o segmento $\overline{P_1P_2}$ esteja na posição correta. Entretanto, esta razão não diz respeito ao formato dos segmentos, que deve ser circular, isto fica a cargo dos dois pesos remanescentes. O peso w_0 controla o formato do primeiro segmento. Diminuindo-se w_0 , a curva tende a se aproximar do ponto P_1 , pois o peso w_1 “ganha importância”. Por outro lado, aumentando-se w_0 , a curva tende a se afastar de P_1 . Existe um valor de w_0 para o qual o segmento tem o formato circular desejado.

Assumindo que o primeiro segmento seja um arco de circunferência, considere o ponto M de intersecção entre o arco e sua bissetriz (Figura 83). Como M é um ponto do primeiro segmento, existe $u \in [0, 1)$ tal que

$$M = \frac{P_0 w_0 N_{0,2}(u) + P_1 w_1 N_{1,2}(u) + P_2 w_2 N_{2,2}(u)}{w_0 N_{0,2}(u) + w_1 N_{1,2}(u) + w_2 N_{2,2}(u)}. \quad (5.56)$$

Para simplificar a equação acima, vamos considerar um sistema de coordenadas cuja origem coincide com P_1 e no lugar de cada ponto considerar o vetor (equivalente) que liga a origem ao ponto. Assim,

$$\mathbf{M} = \frac{\mathbf{P}_0 \omega_0 N_{0,2}(u) + \mathbf{P}_2 w_2 N_{2,2}(u)}{w_0 N_{0,2}(u) + w_1 N_{1,2}(u) + w_2 N_{2,2}(u)}. \quad (5.57)$$

Reescrevendo os pesos com relação a w_1 , temos

$$\mathbf{M} = \frac{\mathbf{P}_0 \omega_0 N_{0,2}(u) + \mathbf{P}_2 \lambda N_{2,2}(u)}{\omega_0 N_{0,2}(u) + N_{1,2}(u) + \lambda N_{2,2}(u)}, \quad (5.58)$$

onde $\omega_0 = w_0/w_1$ é a razão que buscamos determinar e $\lambda = w_2/w_1 = L_1/L_2$ é uma razão conhecida. Logo, há duas incógnitas: a razão ω_0 e o parâmetro u , o que solicita duas equações.

Calculando o produto escalar de ambos os membros da Equação (5.58) com $\mathbf{P}_0$, temos

$$\mathbf{M} \cdot \mathbf{P}_0 = \frac{\|\mathbf{P}_0\| \omega_0 N_{0,2}(u) + (\mathbf{P}_2 \cdot \mathbf{P}_0) \lambda N_{2,2}(u)}{\omega_0 N_{0,2}(u) + N_{1,2}(u) + \lambda N_{2,2}(u)}. \quad (5.59)$$

De acordo com a Figura 83 (lembrando que P_1 deve ser considerado como origem),

$$\begin{aligned} \mathbf{M} \cdot \mathbf{P}_0 &= \left(\frac{L_1}{\sin\left(\frac{\theta_1}{2}\right)} - R_1 \right) L_1 \cos\left(\frac{\pi}{2} - \frac{\theta_1}{2}\right) = L_1^2 - R_1 L_1 \sin\left(\frac{\theta_1}{2}\right) = L_1^2 \left(1 - \cos\left(\frac{\theta_1}{2}\right)\right) \\ \|\mathbf{P}_0\| &= L_1^2 \\ (\mathbf{P}_2 \cdot \mathbf{P}_0) &= (L_1 + L_2) L_1 \cos(\pi - \theta) = -L_1^2 \left(1 + \frac{1}{\lambda}\right) \cos \theta_1 \end{aligned} \quad (5.60)$$

Substituindo em (5.59) e dividindo ambos os membros pelo fator comum L_1^2 , temos

$$1 - \cos\left(\frac{\theta_1}{2}\right) = \frac{\omega_0 N_{0,2}(u) - (1 + \lambda) \cos \theta_1 N_{2,2}(u)}{\omega_0 N_{0,2}(u) + N_{1,2}(u) + \lambda N_{2,2}(u)}. \quad (5.61)$$

Analogamente, o produto escalar com $\mathbf{P}_2$ resulta na equação

$$1 - \cos\left(\frac{\theta_1}{2}\right) = \frac{(1 + \lambda) N_{2,2}(u) - \cos \theta_1 \omega_0 N_{0,2}(u)}{\omega_0 N_{0,2}(u) + N_{1,2}(u) + \lambda N_{2,2}(u)}. \quad (5.62)$$

Comparando as Equações (5.61) e (5.62), concluímos que

$$\begin{aligned} (1 + \lambda) N_{2,2}(u) - \cos \theta_1 \omega_0 N_{0,2}(u) &= \omega_0 N_{0,2}(u) - (1 + \lambda) \cos \theta_1 N_{2,2}(u) \\ \Rightarrow \omega_0 &= (1 + \lambda) \frac{N_{2,2}(u)}{N_{0,2}(u)} \end{aligned} \quad (5.63)$$

Substituindo ω_0 e as expressões das funções base no intervalo $[0, 1]$ na Equação (5.61) ou (5.62), chegamos a uma equação quadrática em u cujas soluções são $u = 0$ (solução trivial indesejada) e

$$u = \frac{2}{2 + (1 + \lambda) \cos\left(\frac{\theta_1}{2}\right)}. \quad (5.64)$$

Com u dado por (5.64), finalmente, temos

$$\omega_0 = \frac{w_0}{w_1} = \frac{2}{(1 + \lambda) \cos^2\left(\frac{\theta_1}{2}\right)} \quad (5.65)$$

Seguindo um raciocínio análogo para o segundo arco, w_3 , w_2 e θ_2 assumem as posições de w_0 , w_1 e θ_1 respectivamente, e a razão inversa $1/\lambda$ toma o lugar de λ . Chegamos à conclusão que

$$\frac{w_3}{w_2} = \frac{2}{\left(1 + \frac{1}{\lambda}\right) \cos^2\left(\frac{\theta_2}{2}\right)}. \quad (5.66)$$

5.7.2 Exemplo. Dois quartos de circunferência com razão 1:2 entre os raios

Por quarto de circunferência, entende-se um arco com ângulo de 90° . Como os ângulos dos arcos são retos, vale $L_1 = R_1$ e $L_2 = R_2$. Logo,

$$\lambda = \frac{L_1}{L_2} = \frac{R_1}{R_2} = \frac{1}{2}, \quad (5.67)$$

onde foi considerada a razão entre os raios dada.

Podemos arbitrariamente escolher $w_1 = 1$. Das Equações (5.54), (5.65) e (5.66), que relacionam os pesos, temos

$$\frac{w_2}{w_1} = \lambda \Rightarrow w_2 = \frac{1}{2} \quad (5.68)$$

$$\frac{w_0}{w_1} = \frac{2}{(1 + \lambda) \cos^2\left(\frac{\theta_1}{2}\right)} \Rightarrow w_0 = \frac{8}{3} \quad (5.69)$$

$$\frac{w_3}{w_2} = \frac{2}{\left(1 + \frac{1}{\lambda}\right) \cos^2\left(\frac{\theta_2}{2}\right)} \Rightarrow w_3 = \frac{2}{3} \quad (5.70)$$

O parâmetro do ponto M , dado pela Equação (5.64), é

$$u = \frac{8}{8 + 3\sqrt{2}} \cong 0,6534 \quad (5.71)$$

Embora M seja o ponto médio do primeiro arco, não é imagem do ponto médio do intervalo $[0, 1]$, isso evidencia a não uniformidade da parametrização.

5.7.3 Exemplo. Semicircunferência de raio unitário

Podemos construir um arco de 180° quadrático com 4 pontos de controle. Considerando a composição de dois arcos de mesmo ângulo e raio ($\theta_1 = \theta_2 = \pi/2$ e $R_1 = R_2$), encontramos as seguintes condições para os pesos: $w_1 = w_2$ e $w_0 = w_3 = 2w_1$. Uma solução possível é $w_0 = w_3 = 1$ e $w_1 = w_2 = 1/2$. Em um sistema de coordenadas no qual a semicircunferência está centrada na origem e ocupa o 1º e o 2º quadrantes, os pontos de controle devem ser $P_0 = (1, 0)$, $P_1 = (1, 1)$, $P_2 = (-1, 1)$ e $P_3 = (-1, 0)$ (veja a Figura 85).

Figura 85 – Semicircunferência NURBS construída com dois arcos de 90° e 4 pontos de controle

Fonte: autoria própria

5.8 Spline de três ou mais arcos

O raciocínio anterior, em geral, falha quando aplicado a uma spline com três arcos. Um possível vetor de nós uniforme para três segmentos é $U = \{0, 0, 0, 1, 2, 3, 3, 3\}$. As funções de base quadráticas para este vetor de nós são ilustradas na Figura 86.

Figura 86 – Funções base quadráticas no vetor de nós $\{0, 0, 0, 1, 2, 3, 3, 3\}$

Fonte: autoria própria

Comparando com as funções para dois segmentos (Figura 84), observamos o mesmo comportamento nos nós interiores ($t = 1$ e $t = 2$), apenas duas funções não nulas e iguais. Consequentemente, a mesma relação inversamente proporcional de nós com distâncias entre os pontos de conexão e os pontos de controle surge.

$$\begin{aligned} \frac{w_2}{w_1} &= \frac{L_1}{L_2} \\ \frac{w_3}{w_2} &= \frac{L_2}{L_3} \end{aligned} \quad (5.72)$$

Com isto, os três pesos que governam o segmento intermediário ficam comprometidos e sua forma não necessariamente fica circular, como mostra o simples exemplo a seguir (Figura 87). Pretende-se que a NURBS seja uma spline de três arcos de 90° com raios 1, 2 e 1 respectivamente. Foram considerados os pesos $w_0 = w_4 = 8/3$, $w_1 = w_3 = 1$ e $w_2 = 1/2$. Tal escolha de pesos molda o primeiro e o terceiro segmentos como arcos de circunferência. Entretanto, calculando a distância do segundo segmento ao centro do segundo arco, percebemos que, embora o segmento esteja perto de ser um arco circular de raio 2, a distância não é constante como deveria ser (Figura 88).

Não há nada que possa ser feito em questão de pesos para corrigir o formato. Concluimos que não é possível em geral parametrizar uma spline de arcos com mais de dois arcos como NURBS em um vetor uniforme sem nós interiores duplicados. Uma possível solução seria trabalhar com um vetor de nós não uniforme, que modificaria as relações (5.72). Porém, não a exploramos neste trabalho.

Figura 87 – NURBS quadrática com 3 segmentos

Fonte: autoria própria

Figura 88 – Distância do segundo segmento até o centro esperado do segundo arco

Fonte: autoria própria

5.9 Representação NURBS de uma circunferência completa

Uma circunferência pode ser construída de diversas formas conectando-se arcos de circunferência. Cada arco sendo um segmento racional, a circunferência é representada como uma curva NURBS. Ainda que seja possível concatenar arcos de comprimentos diferentes, escolhe-se usar arcos congruentes por gerar uma parametrização melhor [59, 60]. Em grau 2, como é impossível representar um arco com ângulo 180° ou maior, são necessários três arcos de 120° no mínimo.

A Figura 89 (a) mostra a NURBS de três segmentos no vetor de nós $\{0, 0, 0, 1, 1, 2, 2, 3, 3, 3\}$. São 7 pontos de controle que formam um triângulo equilátero circunscrito à circunferência, o último ponto coincide com o primeiro. O conjunto de pesos usado é $\{1, 1/2, 1, 1/2, 1, 1/2, 1\}$.

Uma forma preferível, que é a mais comum, usa quatro arcos de 90° . A Figura 89 (b) mostra a NURBS de quatro segmentos no vetor de nós $\{0, 0, 0, 1, 1, 2, 2, 3, 3, 4, 4, 4\}$. São 9 pontos de controle que formam um quadrado circunscrito à circunferência. O conjunto de pesos

usado é $\{1, 1/\sqrt{2}, 1, 1/\sqrt{2}, 1, 1/\sqrt{2}, 1, 1/\sqrt{2}, 1\}$. Note que o polígono de controle aproxima melhor a circunferência no segundo caso. Uma aproximação ainda melhor acontece com 6 arcos de 60° (polígono hexagonal) [60]. O acréscimo de arcos torna a parametrização e a aproximação pelo polígono de controle melhor ao custo de mais segmentos e descontinuidades [53].

Figura 89 – Circunferência NURBS (a) com 3 arcos e polígono de controle triangular (b) com 4 arcos e polígono de controle quadrado (c) polígono de controle quadrado com dois pontos removidos

Fonte: autoria própria

Considere a circunferência com 9 pontos de controle. Ajustando os pesos de certa forma, é possível remover uma ocorrência do nó 1 e uma ocorrência do nó 3 construindo a circunferência com 7 pontos de controle no vetor de nós $\{0, 0, 0, 1, 2, 2, 3, 4, 4, 4\}$. O resultado são duas semicircunferências de 4 pontos de controle conectadas Figura 89c. O conjunto de pesos obtido é $\{1, 1/2, 1/2, 1, 1/2, 1/2, 1\}$, cada semicircunferência segue o mesmo padrão de pesos do exemplo na Seção 5.7.3. Demonstra-se que é impossível remover o último nó duplicado [59].

6 Superfícies para a modelagem de contato

A modelagem de contato que pretendemos fazer segue o paradigma “*master-master*”, em que não há discretização de nenhuma curva ou superfície de cada par. Contato entre curvas é discutido por Wriggers e Zavarise [61, 62]. Uma generalização usando superfícies é encontrada nos trabalhos de Gay Neto *et al.* [63, 64].

Para análise ferroviária, Higa [20] propôs um modelo plano simplificado da dinâmica de um rodeiro ferroviário em contato com a ferrovia em trechos de tangente e de curva. Higa descreve um contato curva-curva entre splines de arcos, que se reduz a uma análise ponto-ponto entre os centros dos arcos.

Em casos mais gerais, os contatos deixam de ser planos e superfícies são necessárias. Refachinho de Campos [65] transforma os perfis de trilho e roda em superfícies através de extrusão e revolução respectivamente. As superfícies são usadas em uma formulação de corpo rígido em elementos finitos [19]. Gay Neto e Wriggers [66] incluem superfícies NURBS como possibilidade dentro da formulação *master-master*.

Nosso objetivo é parametrizar superfícies de extrusão e revolução como NURBS. Assim, todas as formas de contato para análise ferroviária mencionadas ficam unificadas em uma mesma base. Além disso, existe a oportunidade de empregar a análise isogeométrica [49] na solução do modelo.

6.1 Superfície NURBS

Uma superfície NURBS é dada pela função de dois parâmetros

$$S(u, v) = \frac{\sum_{i=0}^m \sum_{j=0}^n P_{ij} w_{ij} N_{i,p}(u) N_{j,q}(v)}{\sum_{i=0}^m \sum_{j=0}^n w_{ij} N_{i,p}(u) N_{j,q}(v)}, \quad (6.1)$$

uma generalização da Equação (5.46) para curvas. Os pontos de controle P_{ij} formam uma rede com $m + 1$ linhas e $n + 1$ colunas totalizando $(m + 1) \times (n + 1)$ pontos. A cada ponto está associado um produto de duas funções base de B-Splines, diz-se que a base da superfície é um produto tensorial de duas bases de um parâmetro [47, 48, 53]. Cada parâmetro tem o próprio vetor de nós, o vetor de nós do parâmetro u é usado para produzir as funções base $N_{i,p}$ e o vetor de nós do parâmetro v , para produzir as funções base $N_{j,q}$. Os graus das funções (p e q) podem ser diferentes. O número de nós para u é $(m + 1) + (p + 1)$ e o número de nós para v é $(n + 1) + (q + 1)$.

Figura 90 – Transformação de um quadrado no plano paramétrico em uma superfície NURBS cúbica no espaço tridimensional

Fonte: autoria própria

A Figura 90 mostra um exemplo de superfície. Os vetores de nós são $U = V = \{0, 0, 0, 0, 1/2, 1, 1, 1, 1\}$, tendo 9 nós cada. A superfície é cúbica (ordem 4) nos dois parâmetros. A rede de controle, portanto, tem dimensão 5×5 , totalizando 25 pontos de controle. A rede de controle é formada pelos pontos

$$\begin{bmatrix} (1, 5, 1) & (2, 5, -1/2) & (3, 5, 0) & (4, 5, 4/5) & (5, 5, 1) \\ (1, 4, -1) & (2, 4, 0) & (3, 4, 0) & (4, 4, 1) & (5, 4, 4/5) \\ (1, 3, 0) & (2, 3, 0) & (3, 3, 0) & (4, 3, 0) & (5, 3, 0) \\ (1, 2, 1) & (2, 2, 2) & (3, 2, 0) & (4, 2, -1/2) & (5, 2, -4/5) \\ (1, 1, 0) & (2, 1, 1) & (3, 1, 0) & (4, 1, -1) & (5, 1, -1) \end{bmatrix}$$

com todos os pesos iguais (a superfície é polinomial, consequentemente).

6.2 Superfície de extrusão

No caso de superfícies de extrusão, uma das coordenadas é linear e as bordas têm um formato determinado, então não há muita liberdade de escolha quanto a pesos e pontos de controle. Os pesos podem ser iguais, fazendo com que a coordenada linear seja uniforme, ou diferentes. A coordenada linear requer duas colunas de pontos de controle, a primeira coluna de pontos representando o perfil base e a segunda coluna sendo a primeira transformada por uma translação.

Vamos supor que a direção linear só tenha um vão. O índice j só assume os valores $\{0, 1\}$ e as funções base são as funções lineares $N_{j=0}(v) = 1 - v$ e $N_{j=1}(v) = v$. Para todo v , as curvas coordenadas devem ter o mesmo formato.

6.2.1 Curvas coordenadas de fronteira

Com $v = 0$, $N_{j=0}(v = 0) = 1$ e $N_{j=1}(v = 0) = 0$, temos

$$S(u, 0) = C_0(u) = \frac{\sum_{i=0}^m P_{i0} w_{i0} N_{i,p}(u)}{\sum_{i=0}^m w_{i0} N_{i,p}(u)}, \quad (6.2)$$

esta curva é uma NURBS com pontos de controle e pesos (P_{i0}, w_{i0}) .

Com $v = 1$, $N_{j=0}(v = 1) = 0$ e $N_{j=1}(v = 1) = 1$, temos

$$S(u, 1) = C_1(u) = \frac{\sum_{i=0}^m P_{i1} w_{i1} N_{i,p}(u)}{\sum_{i=0}^m w_{i1} N_{i,p}(u)}, \quad (6.3)$$

esta curva é uma NURBS com as mesmas funções base, mas com pontos de controle e pesos (P_{i1}, w_{i1}) .

Ambas as curvas têm que ter o mesmo formato. Logo, C_1 é obtida de C_0 através de um movimento rígido, que pode ser aplicado ao polígono de controle (propriedade de invariância afim [47, 48]). Assim,

$$P_{i1} = \mathbf{R}[P_{i0}] + \mathbf{d}, \quad (6.4)$$

onde $\mathbf{R}$ representa uma rotação e $\mathbf{d}$ é um vetor de deslocamento. Os pesos da segunda curva podem ser iguais aos da primeira ou escalados por um fator constante, o que não altera o formato. Por maior generalidade, escrevemos

$$w_{i1} = \lambda w_{i0}, \quad (6.5)$$

onde λ é um número real positivo.

Para uma superfície de extrusão, não há rotação ($\mathbf{R} = \mathbf{I}$). Vejamos o que acontece com as curvas coordenadas intermediárias quando o polígono de controle é somente deslocado.

6.2.2 Curvas coordenadas intermediárias

Da Equação (6.1), considerando $P_{i1} = P_{i0} + \mathbf{d}$ e $w_{i1} = \lambda w_{i0}$, dado $v = \bar{v}$, temos

$$S(u, \bar{v}) = \frac{\sum_{i=0}^m P_{i0} w_{i0} N_{i,p}(u)(1 - \bar{v}) + (P_{i0} + \mathbf{d}) w_{i1} N_{i,p}(u) \bar{v}}{\sum_{i=0}^m w_{i0} N_{i,p}(u)(1 - \bar{v}) + \lambda w_{i0} N_{i,p}(u) \bar{v}}. \quad (6.6)$$

Distribuindo a soma do numerador e executando algumas manipulações, vem

$$S(u, \bar{v}) = \frac{(1 - \bar{v} + \lambda \bar{v}) \sum_{i=0}^m P_{i0} w_{i0} N_{i,p}(u) + \sum_{i=0}^m \mathbf{d} \lambda w_{i0} N_{i,p}(u) \bar{v}}{(1 - \bar{v} + \lambda \bar{v}) \sum_{i=0}^m w_{i0} N_{i,p}(u)}. \quad (6.7)$$

Como $\mathbf{d}$ e λ são constantes, obtemos a forma

$$S(u, \bar{v}) = \frac{\sum_{i=0}^m P_{i0} w_{i0} N_{i,p}(u)}{\sum_{i=0}^m w_{i0} N_{i,p}(u)} + \frac{\lambda \bar{v}}{(1 - \bar{v} + \lambda \bar{v})} \mathbf{d} \quad (6.8)$$

ou

$$S(u, \bar{v}) = C_0(u) + \frac{\lambda \bar{v}}{(1 - \bar{v} + \lambda \bar{v})} \mathbf{d}. \quad (6.9)$$

Qualquer que seja λ , para $0 \leq \bar{v} \leq 1$, vale

$$0 \leq \frac{\lambda \bar{v}}{(1 - \bar{v} + \lambda \bar{v})} \leq 1. \quad (6.10)$$

Portanto, todas as curvas coordenadas são cópias do perfil base C_0 posicionadas ao longo da linha de deslocamento dada pelo vetor $\mathbf{d}$. O fator λ controla a densidade de curvas coordenadas ao longo da linha de deslocamento:

- $\lambda = 1$: distribuição uniforme (Figura 92 (a));
- $\lambda < 1$: maior concentração próximo da fronteira C_0 (Figura 92 (b));
- $\lambda > 1$: maior concentração próximo da fronteira C_1 (Figura 92 (c)).

6.2.3 Exemplo. Superfície tipo trilho

Consideramos como perfil para extrusão a curva NURBS quadrática de três arcos mostrada na Figura 91, que representa o topo de um trilho ferroviário. O arco central tem raio 10 e comprimento 2,5, os dois outros têm raio 1 e comprimento 0,4354. A curva está contida no plano xy e a linha de deslocamento é ortogonal ao plano. A superfície gerada pode ser vista na Figura 92 em três parametrizações diferentes, cada uma com um certo λ .

Figura 91 – Perfil NURBS quadrático tipo trilho com três arcos

Fonte: autoria própria

Figura 92 – Superfície de extrusão com a curva coordenada $v = 1/2$ destacada (a) $\lambda = 1$: curvas coordenadas uniformemente distribuídas (b) $\lambda = 1/4$: curvas coordenadas mais concentradas em torno de C_0 (c) $\lambda = 4$: curvas coordenadas mais concentradas em torno de C_1

Fonte: autoria própria

6.2.4 Efeito da rotação de uma fronteira

Superfície regrada é toda superfície que pode ser varrida por uma reta (“régua”) movendo-se ao longo de uma curva [67]. Superfície de extrusão é um tipo particular de superfície regrada. Toda superfície NURBS com uma direção linear é uma superfície regrada. Permitindo a rotação do perfil base, obtemos um tipo de superfície regrada que é mais geral do que uma superfície de extrusão.

A Figura 93 mostra duas superfícies cujas curvas coordenadas de fronteira são circunferências. A superfície à esquerda não tem rotação, é um cilindro formado pela extrusão de uma circunferência. A superfície à direita tem uma rotação relativa de 120° entre as circunferências em torno do eixo do cilindro, é um hiperboloide de uma folha. Note que a rotação deforma as curvas coordenadas intermediárias.

Figura 93 – Duas superfícies regradas com bordas circulares (a) cilindro formado por extrusão (b) hiperboloide formado por “torção” do cilindro

Fonte: autoria própria

Um modelo físico deste fenômeno é mostrado na Figura 94. Dois discos são ligados por hastes retas. Torcendo os discos, as hastes permanecem retas e um hiperboloide é formado.

Figura 94 – Modelo físico de um hiperboloide como superfície regrada

Fonte: [68]

6.3 Superfície de revolução

Uma superfície de revolução é a região varrida por uma curva base $G(u)$ chamada de geratriz quando esta sofre uma rotação em torno de um eixo. As curvas coordenadas associadas à direção de revolução são circunferências chamadas de paralelos, cujos centros incidem sobre o eixo de revolução. As curvas associadas à outra coordenada têm o mesmo formato da curva base e são chamadas de meridianos. Seja $S(u, v)$ a superfície de revolução, podemos descrever matematicamente qualquer ponto da superfície pela equação

$$S(u, v) = \mathbf{R}_v[G(u)], \quad (6.11)$$

sendo $\mathbf{R}_v$ um operador de rotação dependente da coordenada v , que é aplicado a cada ponto $G(u)$ da geratriz. Considerando um sistema de coordenadas cartesianas no qual a curva base

pertença ao plano xy e a revolução aconteça em torno do eixo x , obtemos as seguintes representações (em forma matricial)

$$G(u) = \begin{bmatrix} x(u) \\ y(u) \\ 0 \end{bmatrix}, \quad (6.12)$$

$$\mathbf{R}_v = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos v & -\sin v \\ 0 & \sin v & \cos v \end{bmatrix}. \quad (6.13)$$

Nesta representação, v aparece como ângulo de rotação. A forma paramétrica da superfície é dada pelo produto matricial que representa a aplicação de $\mathbf{R}_v$ a $G(u)$.

$$S(u, v) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos v & -\sin v \\ 0 & \sin v & \cos v \end{bmatrix} \begin{bmatrix} x(u) \\ y(u) \\ 0 \end{bmatrix} = \begin{bmatrix} x(u) \\ y(u) \cos v \\ y(u) \sin v \end{bmatrix} \quad (6.14)$$

Podemos notar que, neste sistema de coordenadas, quando u é mantido constante e a coordenada v varia de 0 a 2π , o ponto $S(u, v)$ descreve uma circunferência no plano $x = x(u)$, ortogonal ao eixo x , com raio $\rho = y(u)$. Portanto, confirmamos que os paralelos são circunferências.

6.3.1 Exemplo. Esfera

Se considerarmos como geratriz uma semicircunferência sem as extremidades, a superfície de revolução sobre um eixo através das extremidades obtida é uma esfera sem os polos. Os meridianos são semicircunferências de polo a polo e os paralelos são circunferências ortogonais ao eixo de revolução (veja a Figura 95).

Na parametrização de uma esfera, geralmente se associa os símbolos θ e φ aos parâmetros para denotar ângulos. A parametrização da geratriz é

$$G(\theta) = \begin{bmatrix} R \cos \theta \\ R \sin \theta \\ 0 \end{bmatrix}, 0 < \theta < \pi, \quad (6.15)$$

onde R é o raio da semicircunferência. Consequentemente, a superfície tem componentes

$$S(\theta, \varphi) = \begin{bmatrix} R \cos \theta \\ R \sin \theta \cos \varphi \\ R \sin \theta \sin \varphi \end{bmatrix}. \quad (6.16)$$

Note que todos os pontos da superfície são equidistantes da origem, sendo a distância $\|S(\theta, \varphi)\| = R$ também o raio da esfera.

Figura 95 – Esfera parametrizada

Fonte: autoria própria

Esta parametrização de esfera é usada para mapear a superfície do globo terrestre (Figura 96). Neste caso, os parâmetros recebem os nomes latitude e longitude e a parametrização é chamada de sistema de coordenadas geográficas. Nesta parametrização da superfície praticamente esférica do planeta, os paralelos são as linhas de latitude constante e os meridianos, as linhas de longitude constante.

Figura 96 – Superfície do planeta Terra parametrizada por latitude e longitude

Fonte: autoria própria

6.4 Parametrização NURBS de uma superfície de revolução

A representação exata de circunferências é possível em NURBS, logo, é de se esperar que exista uma forma NURBS para superfícies de revolução, que são, de certo modo, conjuntos infinitos de circunferências justapostas.

Piegl e Tiller [48] sugerem um formato de parametrização de superfície NURBS o qual se verifica corresponder a uma superfície de revolução. A seguir demonstramos de forma mais direta quais devem ser os pontos de controle e pesos de uma superfície de revolução como consequência do fato de que as curvas coordenadas são circunferências.

6.4.1 Paralelos, as curvas coordenadas circulares

Dada a equação geral para uma superfície NURBS (6.1), fixamos $u = \bar{u}$ para estudar um paralelo genérico $C(v)$

$$S(\bar{u}, v) = C(v) = \frac{\sum_{i=0}^m \sum_{j=0}^n P_{ij} w_{ij} N_{i,p}(\bar{u}) N_{j,q}(v)}{\sum_{i=0}^m \sum_{j=0}^n w_{ij} N_{i,p}(\bar{u}) N_{j,q}(v)}. \quad (6.17)$$

Sabemos que a curva $S(\bar{u}, v)$ é uma circunferência, que é representável como NURBS de segundo grau. Portanto, deve ser possível reduzir a Equação (6.17) à forma

$$C(v) = \frac{\sum_{j=0}^n Q_j w_j N_{j,q=2}(v)}{\sum_{j=0}^n w_j N_{j,q=2}(v)}, \quad (6.18)$$

onde evidenciamos o grau $q = 2$ das funções base referentes ao índice j .

Assumindo que

$$w_{ij} = w_i^U w_j^V, \quad (6.19)$$

onde os w_i^U são os pesos da curva geratriz, podemos fatorar a dupla soma no denominador no produto de duas somas

$$S(\bar{u}, v) = \frac{\sum_{j=0}^n \{ [\sum_{i=0}^m P_{ij} w_i^U N_{i,p}(\bar{u})] w_j^V N_{j,q=2}(v) \}}{\sum_{i=0}^m [w_i^U N_{i,p}(\bar{u})] \sum_{j=0}^n [w_j^V N_{j,q=2}(v)]}. \quad (6.20)$$

A fatoração do denominador transforma a superfície em

$$S(u, v) = \sum_{j=0}^n \sum_{i=0}^m P_{ij} R_{i,p}(u) R_{j,q=2}(v), \quad (6.21)$$

onde temos o produto de duas funções base racionais de uma variável,

$$R_{i,p}(u) = \frac{w_i^U N_{i,p}(u)}{\sum_{i=0}^m w_i^U N_{i,p}(u)} \quad (6.22)$$

e

$$R_{j,q=2}(v) = \frac{w_j^V N_{j,q=2}(v)}{\sum_{j=0}^n w_j^V N_{j,q=2}(v)}, \quad (6.23)$$

que é uma função base racional de duas variáveis. Podemos dizer que a base da superfície racional é formada pelo produto tensorial das bases de duas curvas racionais, a geratriz e a trajetória circular.

Definimos

$$Q_j = \frac{\sum_{i=0}^m P_{ij} w_i^U N_{i,p}(\bar{u})}{\sum_{i=0}^m w_i^U N_{i,p}(\bar{u})}, \quad (6.24)$$

isto é, Q_j é o ponto de coordenada $\bar{u}$ de uma curva NURBS de grau p com pontos de controle P_{ij} e pesos w_i^U , veremos adiante que estes pontos de controle estão relacionados com os da curva geratriz. Inspeccionando a Equação (6.20), identificamos a presença do ponto Q_j como definimos e que a expressão da curva coordenada assume o formato desejado

$$S(\bar{u}, v) = C(v) = \frac{\sum_{j=0}^n Q_j w_j^V N_{j,q=2}(v)}{\sum_{j=0}^n w_j^V N_{j,q=2}(v)}, \quad (6.25)$$

A curva geratriz é dada por $S(u, 0)$. Além disso, para $v = 0$, a Equação (6.25) retorna

$$S(\bar{u}, 0) = Q_0. \quad (6.26)$$

Consequentemente, Q_0 é o ponto $G(\bar{u})$ da geratriz. Da definição dos Q_j vem

$$Q_0 = \frac{\sum_{i=0}^m P_{i0} w_i^U N_{i,p}(\bar{u})}{\sum_{i=0}^m w_i^U N_{i,p}(\bar{u})}. \quad (6.27)$$

Na Equação (6.27), identificamos que a primeira coluna da matriz de pontos de controle P_{i0} é o polígono de controle da curva geratriz.

Supondo que o eixo dos x seja o eixo de revolução e a geratriz pertença ao plano xy , a curva C acima (6.25) deve ser uma circunferência cujo raio ρ é a coordenada y do ponto Q_0 . Usando a construção de circunferência com arcos de 90° (Seção 5.9), são necessários 9 pontos de controle Q_j com a sequência de pesos $\{w_0^V = 1, w_1^V = \sqrt{2}/2, w_2^V = 1, w_3^V = \sqrt{2}/2, w_4^V = 1, w_5^V = \sqrt{2}/2, w_6^V = 1, w_7^V = \sqrt{2}/2, w_8^V = 1\}$. Para construir uma circunferência de raio ρ , os pontos são dispostos no quadrado de lado 2ρ que circunscreve a circunferência mostrado na Figura 97.

Figura 97 – Construção NURBS de um paralelo qualquer

Fonte: autoria própria

Da Figura 97, o ponto Q_1 é obtido transladando de ρ o ponto Q_0 na direção z

$$Q_1 = Q_0 + (0, 0, \rho) . \quad (6.28)$$

A questão é quais devem ser os pontos P_{i1} tais que o ponto Q_1 seja o requerido. Uma aparente resposta seria obtida simplesmente por somar ρ à coordenada z dos pontos P_{i0}

$$P_{i1} = P_{i0} + (0, 0, \rho) . \quad (6.29)$$

No entanto, para cada escolha de $\bar{u}$, há um ρ diferente em geral, simplesmente somar um ρ específico não resultaria em uma superfície de revolução. Uma forma de preservar a dependência de $\bar{u}$ é escrevermos ρ como a coordenada y do ponto Q_0 dado pela Equação (6.27), o que resulta na expressão

$$\rho(\bar{u}) = \frac{\sum_{i=0}^m y_{i0} w_i^U N_{i,p}(\bar{u})}{\sum_{i=0}^m w_i^U N_{i,p}(\bar{u})} , \quad (6.30)$$

onde y_{i0} é a ordenada do ponto de controle P_{i0} da curva geratriz.

O ponto Q_1 é, portanto, dado por

$$Q_1 = Q_0 + (0, 0, \rho(\bar{u})) = \frac{\sum_{i=0}^m [P_{i0} + (0, 0, y_{i0})] w_i^U N_{i,p}(\bar{u})}{\sum_{i=0}^m w_i^U N_{i,p}(\bar{u})} \quad (6.31)$$

Na Equação (6.31) identificamos que

$$P_{i1} = P_{i0} + (0, 0, y_{i0}) , \quad (6.32)$$

isto é, cada ponto de controle é deslocado da própria coordenada y na direção z , assim como o ponto Q_0 . Uma outra explicação para este fato é que Q_1 resulta da aplicação sobre Q_0 da transformação linear representada pela matriz

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

e toda transformação linear de uma curva NURBS pode ser transmitida aos pontos de controle diretamente (propriedade de invariância afim [47, 48]).

O ponto Q_2 é dado pela translação de Q_1 na direção negativa do eixo y . Aplicando a transformação correspondente aos pontos de controle, temos

$$P_{i2} = P_{i1} + (0, -y_{i0}, 0). \quad (6.33)$$

Repetindo o raciocínio para os demais pontos Q_j , concluímos que cada linha de pontos de controle na rede P_{ij} deve formar um quadrado de lado $2y_{i0}$ centrado em um ponto do eixo x e paralelo ao plano yz .

6.4.2 Exemplo. Curva geratriz cúbica

Como geratriz escolhemos a curva de Bézier cúbica com os pontos de controle $P_0 = (-1, 2, 0)$, $P_1 = (-1, 1, 0)$, $P_2 = (1, 1, 0)$ e $P_3 = (1, 0, 0)$.

A rede de controle para uma volta completa pode ser organizada em uma matriz de nove colunas. A primeira coluna tem os pontos de controle da geratriz, a segunda coluna é dada por (6.32) e a terceira, por (6.33)

$$\text{Pontos de controle: } \begin{bmatrix} (-1, 2, 0) & (-1, 2, 2) & (-1, 0, 2) \\ (-1, 1, 0) & (-1, 1, 1) & (-1, 0, 1) \\ (1, 1, 0) & (1, 1, 1) & (1, 0, 1) \\ (1, 0, 0) & (1, 0, 0) & (1, 0, 0) \end{bmatrix}.$$

Estas três colunas são suficientes para parametrizar 90° de revolução.

Os pesos são dados por (6.19). Como a geratriz é uma curva de Bézier, $w_i^U = 1$. Portanto, as três primeiras colunas da matriz de pesos são

$$\text{Pesos: } \begin{bmatrix} 1 & 1/\sqrt{2} & 1 \\ 1 & 1/\sqrt{2} & 1 \\ 1 & 1/\sqrt{2} & 1 \\ 1 & 1/\sqrt{2} & 1 \end{bmatrix}.$$

Figura 98 – Parte de uma superfície de revolução NURBS

Fonte: autoria própria

A Figura 98 mostra a imagem da superfície, a rede de controle, as curvas de Bézier produzidas por cada coluna da rede e, em destaque, um certo paralelo com seu polígono de controle. Note que cada ponto de controle Q_j do paralelo incide sobre a curva de Bézier cujos pontos de controle compõem a coluna P_{ij} , estando de acordo com a Equação (6.24).

6.4.3 Exemplo. Superfície tipo roda ferroviária

Vamos montar um perfil usando como modelo o perfil de roda típico ilustrado na Figura 59. O perfil montado é meramente ilustrativo, os raios dos arcos são praticamente arbitrários. O perfil é composto de quatro segmentos:

- um arco para o friso de raio 1 e ângulo 170° (ângulo do friso 10°);
- um arco para a transição do friso para a pista da roda de raio 1,6 e ângulo $77,14^\circ$;
- um segmento reto para a pista de comprimento 7 e inclinação 1:20 ($2,86^\circ$);
- um arco para o chanfro externo de raio 0,8 e ângulo $87,14^\circ$.

A curva geratriz acompanhada de seus pontos de controle é mostrada na Figura 99. O arco de 170° foi dividido em dois para manter os pontos de controle próximos da curva. A linha reta foi desenhada alinhando-se os pontos de controle P_6 , P_7 e P_8 , o trecho do polígono de controle coincide com o segmento.

Figura 99 – Perfil NURBS quadrático tipo roda ferroviária com cinco segmentos

Fonte: autoria própria

A Figura 100 apresenta a quarta parte da superfície de revolução em torno do eixo x . Observando o espaçamento dos meridianos, podemos inferir que a parametrização padrão (pesos $\{1, 1/\sqrt{2}, 1\}$) da direção de revolução não está muito longe da uniformidade, que é uma característica desejável.

Figura 100 – Parte de uma superfície de roda ferroviária

Fonte: autoria própria

Conclusão

Neste trabalho foi desenvolvida uma metodologia de ajuste de curvas a sequências de pontos discretos. O ajuste é feito por aproximação de uma curva composta por arcos de circunferência (spline de arcos) ao conjunto de dados. A aproximação é dividida em duas partes: uma fase preliminar em que uma aproximação baseada em curvaturas é obtida; e uma fase de otimização em que a aproximação é aprimorada através de um método de mínimos quadrados.

Há restrições de continuidade nas conexões entre arcos para que a curva seja tão suave quanto possível. Em primeiro lugar, a extremidade final de um arco deve coincidir com a extremidade inicial do arco seguinte. Em segundo lugar, o vetor tangente deve ser contínuo através dos arcos. Consideramos as restrições de continuidade das splines de arcos como parte da descrição desta família de curvas, assim livramos a otimização de restrições.

Os métodos foram testados por meio de experimentos numéricos com vários tipos de curvas amostradas. Os resultados dão bons indícios de que as splines de arcos são adequadas para aproximação de pontos discretos pelo método de mínimos quadrados.

Em particular, o procedimento mostrou-se eficaz em aproximar listas de pontos medidos de perfis de roda e trilho ferroviários. Foram feitos testes tanto com perfis recém-usinados quanto com perfis desgastados. Há formatos típicos para estes perfis que se repetem frequentemente, o que permitiu que certos valores padrão para parâmetros de controle do procedimento fossem definidos. Mesmo em casos excepcionais, boas aproximações podem ser encontradas sem alteração substancial nos parâmetros.

O segundo objetivo deste trabalho era dar uma representação como NURBS às splines de arcos. Apresentamos diferentes formas de parametrizar splines de arcos como curvas NURBS. Demonstramos como parametrizar um biarco como uma NURBS quadrática de dois segmentos com primeira derivada contínua em toda a curva. Demonstramos através de um exemplo simples como o mesmo tipo de parametrização não é possível com três ou mais arcos.

Foi também apresentado como produzir superfícies de extrusão e revolução NURBS com base em curvas planas. Estas superfícies podem ser usadas para representar trilhos e rodas em simulações computacionais de contato entre estes elementos.

Trabalhos futuros

O critério de tolerância atual do procedimento de ajuste é muito subjetivo. A aproximação é caracterizada como boa a partir do aspecto visual. Critérios mais objetivos podem ser definidos através do comportamento das representações obtidas de rodas e trilhos em simulações computacionais.

Não é necessário que o procedimento de ajuste fique restrito a perfis de roda e trilho. O procedimento pode ser usado também para aproximar com arcos de circunferência o traçado plano de um trecho de via férrea por exemplo, e, com a descrição NURBS, uma representação da superfície inteira do trecho pode ser produzida. O procedimento também pode ser expandido para aproximar curvas tridimensionais.

Uma das deficiências da parametrização NURBS das splines de arcos é a possível descontinuidade da primeira derivada nos pontos de conexão entre os arcos. Usando vetores de nós uniformes, a continuidade da primeira derivada não é garantida, contudo, há a possibilidade de vetores não-uniformes serem capazes de originar parametrizações de splines de arcos com três ou mais segmentos com o máximo de continuidade permitida.

O uso de NURBS no contexto de algoritmos de contato para estudo da interação roda-trilho ainda abre caminho para associação com a análise isogeométrica. Além disso, os resultados podem nortear a otimização dos perfis geométricos dos elementos ferroviários, que é facilitada com NURBS pela maneira intuitiva de deformar superfície através da rede de pontos de controle.

Referências

- [1] B. Russell, *Misticismo e Lógica*, São Paulo: Editora Nacional, 1957.
- [2] “Section G-II: Wheel and Axle Manual,” em *Manual of Standards and Recommended Practices*, Washington, TTCI, 2004.
- [3] *Manual for Railway Engineering*, AREMA, 2014.
- [4] C. B. Boyer e U. C. Merzbach, *História da matemática*, Tradução (de Helena Castro) da 3ª edição americana ed., São Paulo: Blucher, 2012.
- [5] D. Parkinson e D. Moreton, “Optimal biarc-curve fitting,” *Computer-Aided Design*, vol. 419, nº 6, p. 411, 1991.
- [6] D. S. Meek e D. J. Walton, “Approximating quadratic NURBS curves by arc splines,” *Computer-Aided Design*, vol. 25, nº 6, pp. 371 - 376, 1993.
- [7] Y. J. Ahn, H. O. Kim e K.-Y. Lee, “G1 arc spline approximation of quadratic Bézier curves,” *Computer-Aided Design*, vol. 30, nº 8, pp. 615 - 620, 1998.
- [8] J.-H. Yong, S.-M. Hu e J.-G. Sun, “Bisection algorithms for approximating quadratic Bézier curves by G1 arc splines,” *Computer-Aided Design*, vol. 32, nº 4, pp. 253 - 260, 2000.
- [9] K. Marciniak e B. Putz, “Approximation of spirals by piecewise curves of fewest circular arc segments,” *Computer-Aided Design*, vol. 16, nº 2, pp. 87-90, 1984.
- [10] D. J. Walton e D. S. Meek, “Approximation of quadratic Bézier curves by arc splines,” *Journal of Computational and Applied Mathematics*, vol. 54, pp. 107-120, 1994.
- [11] D. S. Meek e D. J. Walton, “Approximating smooth planar curves by arc splines,” *Journal of Computational and Applied Mathematics*, vol. 59, pp. 221-231, 1995.
- [12] P. Wriggers, *Computational Contact Mechanics*, Springer-Verlag Berlin Heidelberg, 2006.

- [13] A. A. Shabana, Dynamics of Multibody Systems, Cambridge University Press, 2005.
- [14] A. Gay Neto, “Giraffe User's Manual - Generic Interface Readily Accessible for Finite Elements,” [Online]. Available: <http://sites.poli.usp.br/p/alfredo.gay/>. [Acesso em Abril 2021].
- [15] A. Gay Neto, C. d. A. Martins e P. M. Pimenta, “Static analysis of offshore risers with a geometrically-exact 3D beam model subjected to unilateral contact,” *Computational Mechanics*, vol. 53, n° 1, pp. 125-145, 2014.
- [16] A. Gay Neto, “Dynamics of offshore risers using a geometrically-exact beam model with hydrodynamic loads and contact with the seabed,” *Engineering Structures*, vol. 125, pp. 438-454, 2016.
- [17] E. M. B. Campelo, P. M. Pimenta e P. Wriggers, “A triangular finite shell element based on a fully nonlinear shell formulation,” *Computational Mechanics*, vol. 31, pp. 505-518, 2003.
- [18] A. Gay Neto, “Modelagem computacional do contato pontual entre corpos: uma visão integrada,” Tese (Livre-Docência) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Estruturas e Geotécnica, São Paulo, 2018.
- [19] P. R. Refachinho de Campos e A. Gay Neto, “Rigid body formulation in a finite element context with contact interaction,” *Computational Mechanics*, vol. 62, n° 6, pp. 1369-1398, 2018.
- [20] D. N. Higa, “Metodologia para quantificação do efeito da pressão de contato em trilhos sujeitos à passagem de uma família de rodas,” Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Estruturas e Geotécnica, São Paulo, 2020.
- [21] E. A. H. Vollebregt, “User guide for CONTACT, Rolling and sliding contact with friction,” VORtech CMCC, Delft, 2019.
- [22] GENSY, “Home Page,” 2021. [Online]. Available: <http://www.gensys.se>.
- [23] Dassault Systemes, “Rail Modules,” 2021. [Online]. Available: <https://www.3ds.com/products-services/simulia/products/simpack/product-modules/rail-modules/>.

- [24] M. P. d. Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, 1976.
- [25] H. Alencar e W. Santos, *Geometria Diferencial das Curvas Planas*.
- [26] M. Itskov, "Curves and Surfaces in Three-Dimensional Euclidean Space," em *Tensor Algebra and Tensor Analysis for Engineers*, 2015, pp. 69-96.
- [27] R. J. Biezuner, *Notas de Aula em Geometria Diferencial*, Belo Horizonte: Instituto de Ciências Exatas da UFMG, 2016.
- [28] T. F. Banchoff e S. T. Lovett, *Differential Geometry of Curves and Surfaces*, 2010.
- [29] M. F. Marinelli, "Métodos de quadrados mínimos," Monografia (Trabalho de Conclusão de Curso para Licenciatura em Matemática) - Universidade Federal de Santa Catarina, UFSC, Florianópolis, 2002.
- [30] R. N. d. Almeida, "O método dos mínimos quadrados: estudo e aplicações para o ensino médio," Dissertação (Mestrado em Matemática) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, UENF, Campos dos Goytacazes, 2015.
- [31] A. F. P. d. C. Humes, I. S. H. d. Melo, L. K. Yoshida e W. T. Martins, "Zeros de Funções," em *Noções de Cálculo Numérico*, São Paulo, McGraw-Hill do Brasil, 1984, pp. 8-35.
- [32] J. Nocedal e S. J. Wright, *Numerical Optimization*, New York: Springer-Verlag, 1999.
- [33] A. R. Conn e N. I. M. G. P. L. Toint, *Trust-Region Methods*, MPS-SIAM, 2000.
- [34] G. A. Shultz, R. B. Schnabel e R. H. Byrd, "A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties," *SIAM Journal on Numerical Analysis*, vol. 22, pp. 47-67, 1985.
- [35] R. H. Byrd, R. B. Schnabel e G. A. Shultz, "Approximate solution of the trust region problem by minimization over two-dimensional subspaces," *Mathematical Programming*, vol. 40, pp. 247-263, 1988.
- [36] D. S. Meek e D. J. Walton, "Approximation of discrete data by G1 arc splines," *Computer-Aided Design*, vol. 24, nº 6, pp. 301-306, 1992.

- [37] S.-C. Pei e J.-H. Horng, "Fittign digital curve using circular arcs," *Pattern Recognition*, vol. 28, nº 1, pp. 107-116, 1995.
- [38] S.-N. Yang e W.-C. Du, "Numerical methods for approximating digitized curves by piecewise circular arcs," *Journal of Computational and Applied Mathematics*, vol. 66, nº 1, pp. 557-569, 1996.
- [39] G. Maier, "Optimal arc spline approximation," *Computer Aided Geometric Design*, vol. 31, nº 5, pp. 211-226, 2014.
- [40] C. J. Ong, Y. S. Wong, H. T. Loh e X. G. Hong, "An optimization approach for biarc curve-fitting of B-spline curves," *Computer-Aided Design*, vol. 28, nº 12, pp. 951-959, 1996.
- [41] J.-O. Lachaud, A. Vialard e F. de Vieilleville, "Analysis and Comparative Evaluation of Discrete Tangent Estimators," em *Discrete Geometry for Computer Imagery, 12th International Conference*, Poitiers, France, 2005.
- [42] J.-P. Salmon, I. Debled-Rennesson e L. Wendling, "A new method to detect arcs and segments from curvature profiles," em *Proceedings - International Conference on Pattern Recognition*, 2006.
- [43] P. Nguyen e I. Debled-Rennesson, "Curvature Estimation in Noisy Curves," em *Proceedings of the 12th international conference on Computer analysis of images and patterns*, 2007.
- [44] N. J. Mitra e A. Nguyen, "Estimating surface normals in noisy point cloud data," em *Proceedings of the nineteenth annual symposium on Computational geometry*, New York, 2003.
- [45] I. K. Sabitov e A. V. Slovesnov, "Approximation of plane curves by circular arcs," *Computational Mathematics and Mathematical Physics*, vol. 50, nº 8, pp. 1279-1288, 2010.
- [46] H. Tournay, "The fundamentals of Vehicle / Track Interaction," em *Guidelines to Best Practices for Heavy Haul Railway Operations: Management of the Wheel and Rail Interface*, Simmons-Boardman Books Inc., 2015.
- [47] L. Piegl, "On NURBS: a survey," *IEEE Computer Graphics and Applications*, vol. 11, nº 1, pp. 55-71, 1991.
- [48] L. Piegl e W. Tiller, *The NURBS Book*, 2 ed., Berlin: Springer, 1997.

- [49] J. A. Cottrell, T. J. Hughes e Y. Bazilevs, *Isogeometric analysis: toward integration of CAD and FEA*, John Wiley & Sons, 2009.
- [50] F. F. Rocha, “NURBS e o Método Isogeométrico,” Dissertação (Mestrado em Matemática) - Universidade Federal do Espírito Santo - UFES, Vitória, 2016.
- [51] D. S. Araújo, “Curvas B-Spline e de Bézier Aplicadas a CAGD,” Dissertação (Mestrado em Matemática) - Instituto Nacional de Matemática Pura e Aplicada, IMPA, Rio de Janeiro, 2016.
- [52] R. Simoni, “Teoria Local das Curvas,” Monografia (Trabalho de Conclusão de Curso para Licenciatura em Matemática) - Universidade Federal de Santa Catarina, UFSC, Florianópolis, 2005.
- [53] L. Piegl e W. Tiller, “Curve and surface constructions using rational B-splines,” *Computer-Aided Design*, vol. 19, nº 9, pp. 485-498, 1987.
- [54] G. Farin, “From conics to NURBS: A tutorial and survey,” *IEEE Computer Graphics and Applications*, vol. 12, nº 5, pp. 78-86, 1992.
- [55] G.-J. Wang, “Rational cubic circular arcs and their application in CAD,” *Computers in Industry*, vol. 16, nº 3, pp. 293-288, 1991.
- [56] J. J. Chou, “Higher order Bézier circles,” *Computer-Aided Design*, vol. 27, nº 4, pp. 303-309, 1995.
- [57] Q.-Q. Hu e G.-J. Wang, “Necessary and sufficient conditions for rational quartic representation of conic sections,” *Journal of Computational and Applied Mathematics*, vol. 203, nº 1, pp. 190-208, 2007.
- [58] L. A. Piegl e W. Tiller, “Data approximation using biarcs,” *Engineering with Computers*, vol. 18, pp. 59-65, 2002.
- [59] L. Piegl e W. Tiller, “A menagerie of rational B-spline circles,” *IEEE Computer Graphics and Applications*, vol. 9, nº 5, pp. 48-56, 1989.
- [60] C. Blanc e C. Schlick, “Accurate parametrization of conics by NURBS,” *IEEE Computer Graphics and Applications*, vol. 16, nº 6, pp. 64-71, 1996.
- [61] P. Wriggers e G. Zavarise, “On contact between three-dimensional beams undergoing large deflections,” *Communications in Numerical Methods in Engineering*, vol. 13, nº 6, pp. 429-438, 1997.

- [62] G. Zavarise e P. Wriggers, "Contact with friction between beams in 3-D space," *International Journal for Numerical Methods in Engineering*, vol. 49, nº 8, pp. 977-1006, 2000.
- [63] A. Gay Neto, P. d. M. Pimenta e P. Wriggers, "A master-surface to master-surface formulation for beam to beam contact. Part I: frictionless interaction," *Computer Methods in Applied Mechanics and Engineering*, vol. 303, pp. 400-429, 2016.
- [64] A. Gay Neto, P. d. M. Pimenta e P. Wriggers, "A master-surface to master-surface formulation for beam to beam contact. Part II: Frictional interaction," *Computer Methods in Applied Mechanics and Engineering*, vol. 319, pp. 146-174, 2017.
- [65] P. R. Refachinho de Campos, "A rigid body and a master-master contact formulation for multibody railway applications," Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Estruturas e Geotécnica, São Paulo, 2018.
- [66] A. Gay Neto e P. Wriggers, "Numerical method for solution of pointwise contact between surfaces," *Computer Methods in Applied Mechanics and Engineering*, vol. 365, 2020.
- [67] F. A. Oliveira e D. M. d. Almeida, "Superfícies regradadas," *C. Q. D. - Revista Eletrônica Paulista de Matemática*, vol. 8, nº 12, pp. 51-75, 2016.
- [68] MATEMATECA IMEUSP, "Superfícies Regradas," [Online]. Available: http://matemateca.ime.usp.br/acervo/superficies_regradadas.html. [Acesso em Novembro 2020].