



SUBGRADIENTE DIFERENCIÁVEL VIA SUAVIZAÇÃO HIPERBÓLICA

Helder Manoel Venceslau

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Adilson Elias Xavier

Rio de Janeiro
Março de 2013

SUBGRADIENTE DIFERENCIÁVEL VIA SUAVIZAÇÃO HIPERBÓLICA

Helder Manoel Venceslau

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Adilson Elias Xavier, D.Sc.

Prof. Nelson Maculan Filho, D.Habil.

Prof. Abilio Pereira de Lucena Filho, D.Sc.

Prof. Michael Ferreira de Souza, D.Sc.

Prof^a. Fernanda Maria Pereira Raupp, D.Sc.

Prof. Horacio Hideki Yanasse, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2013

Venceslau, Helder Manoel

Subgradiente Diferenciável via suavização hiperbólica/Helder Manoel Venceslau. – Rio de Janeiro: UFRJ/COPPE, 2013.

XV, 94 p.: il.; 29,7cm.

Orientador: Adilson Elias Xavier

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2013.

Referências Bibliográficas: p. 76 – 78.

1. Problema não diferenciável. 2. Subgradiente.
3. Suavização Hiperbólica. I. Xavier, Adilson Elias.
II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

*Aos meus pais, Manoel e
Antonina,
À minha esposa Marilis.*

Agradecimentos

Gostaria de expressar meus agradecimentos a todos os que contribuíram para a realização desta tese.

Primeiramente ao meu orientador, Prof. Adilson Elias Xavier, que literalmente me "adotou" quando eu ainda buscava uma oportunidade de ser aceito na linha de pesquisa de otimização do PESC/COPPE. Tenho para com ele um débito impagável, pois sua fé em mim trouxe-me grande conforto durante a crítica fase de minha vida durante a qual buscava uma reestruturação de minha carreira profissional, após vinte e quatro anos de trabalho na área empresarial.

Aos professores do PESC por sua dedicação e estímulo, dentre os quais faço questão de mencionar:

- Prof. Maculan, que me recebeu sem reservas quando buscava orientações durante a submissão de minha candidatura ao doutorado. Já como doutorando, participar de suas aulas foi sempre uma experiência enriquecedora e seus "desafios intelectuais" muito inspiradores;
- Prof. Paulo Roberto, que me proporcionou um feliz reencontro com a Álgebra Linear e a Análise Numérica, após anos distanciado destas áreas;
- Prof^ª. Suzana Scheimberg, que me apresentou o "mundo convexo", ao qual fiquei permanentemente conectado;
- Prof. Abílio Lucena, que me introduziu no desafiador mundo da Otimização Combinatória. Sua dedicação no preparo do material de aula foi sempre admirável, e suas cobranças mais ainda.

Fica também registrado o reconhecimento pelo apoio prestado pela secretária da linha de pesquisa de otimização, Maria de Fátima, e pelos funcionários da secretaria do PESC, quando o enfrentamento da burocracia se fez necessário. Dentre estes gostaria de destacar em especial a nossa "enfática" Josefina Solange e a Cláudia Prata, que sempre conseguiram soluções para as piores situações.

Aos companheiros do PESC/COPPE, que tornaram mais agradável a participação nas aulas e estudos, ficam meus especiais agradecimentos: Rogério, Vinícius

Xavier, Renan, Jesus, Vinícius Forte, Martagão, Guilherme, Ana Flávia, Virgínia, Aldemir, Max, Arnaldo, Afonso, João Benício, Gardênia e Pedro. Espero ter lembrado de todos.

Agradeço também ao Prof. Michael Souza, que demonstrou grande maestria nas poucas oportunidades em que tive o privilégio de assistir às suas aulas.

No âmbito familiar, que curiosamente possui interseção não vazia com o âmbito acadêmico, agradeço à minha esposa e doutoranda Marilis Venceslau, cujo apoio transcendeu minhas expectativas, principalmente no que concerne à tomada de decisão quanto à minha carreira profissional. Confesso que esta mudança foi muito mais "sugerida" do que simplesmente "apoiada" por ela, o que demonstra seu profundo conhecimento de minhas reais aspirações. Espero ter retribuído em parte ao estimulá-la a entrar também para o doutorado na linha de pesquisa de otimização do PESC/COPPE, o que nos permitiu a satisfação do convívio dentro do ambiente acadêmico durante alguns anos.

Ao meu filho Daniel um agradecimento especial pelos incentivos e providencial ajuda no dia da apresentação de minha tese. Vejo nele a minha herança, que agora também se configura nos mesmos interesses.

Também gostaria de agradecer aos meus pais, que sempre apoiaram meu crescimento profissional, respeitando minhas preferências acadêmicas durante minha juventude, quando para mim teria sido até mais fácil, diante das dúvidas, simplesmente delegar-lhes a responsabilidade por realizar estas escolhas.

À UFRJ pela grande oportunidade de formação acadêmica.

Ao CNPq e à FAPERJ pelo apoio financeiro prestado.

Por fim agradeço a Deus por ter me proporcionado esta preciosa benção de concretizar meus sonhos intelectuais.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

SUBGRADIENTE DIFERENCIÁVEL VIA SUAVIZAÇÃO HIPERBÓLICA

Helder Manoel Venceslau

Março/2013

Orientador: Adilson Elias Xavier

Programa: Engenharia de Sistemas e Computação

Esta tese apresenta o "subgradiente diferenciável", denotado por $\nabla f_s(\tau, x)$, uma alternativa diferenciável para a geração de ε -subgradientes utilizados na resolução de problemas de otimização não diferenciáveis. Na prática, a função objetivo não diferenciável do problema de otimização é suavizada mediante a sua substituição por uma aproximação diferenciável parametrizada por $\tau > 0$, que tende para a função objetivo original quando $\tau \rightarrow 0$.

Quando a função sob análise atender a alguns requisitos, será possível estabelecer uma desigualdade envolvendo ε e τ que garantirá a geração de um ε -subgradiente para qualquer $\varepsilon > 0$, independentemente de um particular $x \in \mathbb{R}^n$. No escopo deste trabalho iremos focar no caso especial de uma função convexa diferenciável com um ponto isolado de não diferenciabilidade, e em sua generalização, as funções ZH convexas.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

DIFFERENTIABLE SUBGRADIENT VIA HIPERBOLIC SMOOTHING

Helder Manoel Venceslau

March/2013

Advisor: Adilson Elias Xavier

Department: Systems Engineering and Computer Science

This thesis presents the "differentiable subgradient", denoted by $\nabla f_s(\tau, x)$, a differentiable alternative for the generation of ε -subgradients used in the resolution of non differentiable optimization problems. Practically speaking, the non differentiable objective function of the optimization problem is smoothed via its replacement by a differentiable approximation parameterized by $\tau > 0$, which tends to the original objective function when $\tau \rightarrow 0$.

Whenever the target function complies with some requirements involving ε and τ , it will be possible to state an inequality that ensures the generation of an ε -subgradient for any $\varepsilon > 0$, independently of the particular $x \in \mathbb{R}^n$. In the context of this work we will focus on the special case of a convex differentiable function possessing an isolated point of non differentiability, and on its generalization, the ZH convex functions.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xiii
1 Introdução	1
2 Revisão Bibliográfica	4
2.1 Discussão sucinta das abordagens	4
2.1.1 Abordagem direta (problema original é mantido)	5
2.1.2 Abordagem suavizada (ou suavizadora)	6
2.2 Definições básicas	8
3 Apresentação do problema e a suavização de funções	11
3.1 Definição da suavização de funções	11
3.2 Regiões de não diferenciabilidade Z_f	13
3.3 Conceitos e nomenclatura	18
3.4 Decomposição DA	19
3.5 Definição da distância à região de não diferenciabilidade Z_f	20
3.6 Suavização da componente f_b da decomposição DA	21
3.7 Região de não diferenciabilidade $Z_f = \{z\}$	24
4 Decomposição DHA da função $f(x)$	25
4.1 Decomposição DHA - primeira etapa	25
4.2 Decomposição DHA - segunda etapa	30
4.2.1 Propriedades da componente homogênea $f_p(h)$	32
4.2.2 Decomposição adicional da componente homogênea $f_p(h)$	34
4.3 Resumo da decomposição DHA	34
4.4 Similaridade com a série de Taylor	34
4.5 Subgradiente médio	35
5 DHA: Suavização da componente homogênea $f_p(h)$	37
5.1 Família de suavizações de $f_p(h)$	37

5.2	Escolha da função ϕ	41
5.3	Alternativas para a função ϕ	43
6	DHA: Condições para geração de ε-subgradientes de $f(x)$	46
6.1	Considerações iniciais	46
6.2	Relação de pertinência em ε -subdiferenciais de $f(x)$	47
6.3	Primeiro exemplo: $f_d(x)$ convexa	48
6.4	Segundo exemplo: $f_d(x)$ não-convexa	50
7	Análise SDHA para $f_d(x)$ convexa	54
7.1	Cálculo de $\partial_\varepsilon f_p(h)$	54
7.2	Cálculo de $\nabla f_q(\tau, h)$	57
7.3	Relação entre ε e τ para que $\nabla f_q(\tau, h)$ seja um ε -subgradiente de $f_p(h)$	57
8	Uso da suavização SDA para gerar ε-subgradientes	61
8.1	SDA na geração de ε -subgradientes para $Z_f = \{z\}$	61
8.2	Comparativo das suavizações SDHA e SDA	65
8.3	SDA na geração de ε -subgradientes para funções ZH convexas	67
9	Resultados e Discussões	69
9.1	A suavização hiperbólica clássica	69
9.2	Suavização de algumas funções	70
9.2.1	Função norma euclidiana	70
9.2.2	Função $f_1(x)$ definida na seção 3.2	70
9.2.3	Função exemplo de Shor	71
9.2.4	Função "unha" definida na seção A.1	71
9.2.5	Função $f_2(x)$ definida na seção 3.2	72
9.3	Transformação de funções convexas gerais em funções ZH convexas	72
10	Conclusões	74
	Referências Bibliográficas	76
A	Alguns resultados básicos importantes	79
A.1	A dimensão do subdiferencial	79
A.2	O subdiferencial $\partial_\infty f$	81
A.3	Aproximação da função máximo	84

Lista de Figuras

2.1	Exemplo no \mathbb{R}^1 onde as inclinações das retas representam um gradiente (no ponto x_a), alguns subgradientes ($f(x)$ não diferenciável na origem) e um ε -subgradiente (no ponto x_b) de uma função convexa $f(x)$	5
3.1	Gráfico de $f_1(x)$ e suas regiões Z_{f_1} e M_{f_1}	15
3.2	Gráfico de $f_2(x)$ e suas regiões Z_{f_2} e M_{f_2}	15
3.3	Gráfico de $f_3(x)$ e suas regiões Z_{f_3} e M_{f_3}	16
3.4	Exemplo de função $f(x)$ no \mathbb{R}^2 , com $G(Z_f) \subset H$, onde H é hiperplano de suporte ao epígrafo de $f(x)$, aqui representado pela função afim $H(x)$	18
4.1	Exemplo de função $f(x)$ no \mathbb{R}^1 e sua derivada direcional $f'(z, x - z)$.	26
4.2	Exemplo de decomposição de $f(x)$ em $f_n(x)$ e $f_d(x)$	26
4.3	$H(x)$ composto por um g arbitrário que pertence a $ri\{\partial f(z)\}$	32
4.4	$f_p(x - z)$ com base na $f(x)$ da figura anterior	32
5.1	Exemplos de funções $\phi(\tau, t)$ admissíveis	39
5.2	Exemplos de funções $\sigma(\tau, t)$ admissíveis	44
6.1	Secção de $f(x) = \ x\ ^2 + 2\ x\ + 2$, com $f_d(x)$ convexa	50
6.2	Secção da função $f_1(x)$ pelo plano $x_2 = 0$	51
6.3	Exemplo de $f(x)$ cuja $f_d(x)$ é não convexa	52
6.4	$f_d(x)$ não convexa da função $f(x)$ apresentada na figura anterior . . .	53
7.1	Gráfico da função $\eta(\tau, s)$	60
8.1	Gráfico que ilustra $\varepsilon_b(x)$ -subgradientes para um dado $x \in \mathbb{R}^n$	64
8.2	Gráfico das versões suavizadas de $f(x) = x^2 + 2 x $	67
9.1	Secção de $f(x)$ e $f(x) - g(x)$ ao longo de x_1	73
A.1	Visões distintas do gráfico da função $f(x_1, x_2) = \sqrt{x_1^2 + \frac{x_2^4}{4}} + \frac{x_2^2}{2}$. . .	80

A.2	Secções do gráfico da função $f(x_1, x_2) = \sqrt{x_1^2 + \frac{x_2^4}{4}} + \frac{x_2^2}{2}$	80
A.3	Exemplo gráfico de aplicação da proposição A.1	82
A.4	Exemplo gráfico de aplicação da proposição A.2	83
A.5	Representação pictórica do contexto da proposição A.3	85
A.6	Gráficos de $r(t)$, $u(t)$ e $\delta(t)$. Todas são nulas para $t < 0$	86
A.7	Gráfico de uma função $u_s(t)$ genérica.	89
A.8	Gráficos de funções $r_s(t)$ e $\delta_s(t)$ genéricas.	89

Lista de Tabelas

5.1	Equivalentes limites das funções $\phi(\tau, t)$	44
8.1	Contexto SDHA (f_d convexa)	66
8.2	Contexto SDA	66

Lista de Símbolos

$aff\{C\}$	envoltória afim do conjunto $C \subset \mathbb{R}^n$, formada por todas as combinações afins de pontos de C
$cl\{C\}$	fecho do conjunto $C \subset \mathbb{R}^n$
$cone\{C\}$	cone gerado pelo conjunto $C \subset \mathbb{R}^n$, definido por $cone\{C\} = \{\alpha x : \alpha \geq 0, x \in C\}$, que é convexo se C for convexo
$conv\{C\}$	fecho da envoltória convexa do conjunto $C \subset \mathbb{R}^n$, formada por todas as combinações convexas de pontos de C
$dim\{C\}$	dimensão do conjunto $C \subset \mathbb{R}^n$, definido como a dimensão do espaço vetorial associado $V_C = aff\{C\} - c$, para um $c \in C$ qualquer
$dom f$	definido como $\{x \in \mathbb{R}^n : f(x) < +\infty\}$
$epi f$	epígrafo da função $f(x)$
$f(x)$	função básica sob análise
$f_a(\tau, x)$	componente de ajuste sobre $f(x)$ a fim de suavizá-la ($f_s(\tau, x) = f(x) + f_a(x)$)
$f_b(h)$	definida como $f(x) - H(x)$, onde $H(x)$ é função afim que representa um hiperplano de suporte ao epígrafo de $f(x)$ e que contém a imagem da região de não diferenciabilidade da função f
$f_c(\tau, h)$	versão suavizada de $f_b(h)$, parametrizada por τ
$f_d(x)$	componente diferenciável de $f(x)$
$f_n(x)$	componente de $f(x)$ não-diferenciável em z
$f_p(h)$	componente de $f(x)$ não-diferenciável em z transladada para a origem ($h = x - z$) e ajustada para contradomínio \mathbb{R}_+
$f_q(\tau, h)$	versão suavizada de $f_p(h)$, parametrizada por τ

$f_r(h)$	definida como $f_d(z+h) - f(z)$
$f_s(\tau, x)$	função $f(x)$ suavizada com parametro τ
$int\{C\}$	interior do conjunto $C \subset \mathbb{R}^n$
$L_c(f)$	conjunto de nível, em situações gerais definido como $L_c(f) = \{x \in dom f : f(x) \leq c\}$
$\bar{\mathbb{R}}$	reais estendidos, definido como a reta real acrescida de $-\infty$ e $+\infty$
$ri\{C\}$	interior relativo do conjunto $C \subset \mathbb{R}^n$, definido como o interior de C relativo a $aff\{C\}$
$V(C, c)$	espaço vetorial $aff\{C\} - c$, onde $c \in C$ qualquer e $C \subset \mathbb{R}^n$
z	ponto isolado onde $f(x)$ não é diferenciável
Z_f	região de não diferenciabilidade da função f
$\Delta f(y, x)$	definida como a diferença $f(y) - f(x)$
$\partial\{C\}$	fronteira do conjunto $C \subset \mathbb{R}^n$, ou seja, $cl\{C\} \setminus int\{C\}$ ($= C \setminus int\{C\}$ se C for fechado)
$\partial f(x)$	subdiferencial de f no ponto x
$\partial_\varepsilon f(x)$	ε -subdiferencial de f no ponto x
$\phi(\tau, t)$	função base para a realização das suavizações
$\kappa(t)$	função moduladora auxiliar na geração de funções $\phi(\tau, t)$ ($\phi(\tau, t) = \kappa(t)\sigma(\tau, t) + t$)
$\sigma(\tau, t)$	função auxiliar na geração de funções $\phi(\tau, t)$ ($\phi(\tau, t) = \sigma(\tau, t) + t$)

Capítulo 1

Introdução

O tema desta tese é a resolução de problemas de otimização não diferenciáveis e irrestritos definidos no \mathbb{R}^n . O foco será a suavização da não diferenciabilidade presente na função objetivo. Em particular, será desenvolvida uma estratégia de suavização das componentes não diferenciáveis presentes em uma formulação. Quando a não diferenciabilidade for constituída por um ponto isolado demonstraremos a possibilidade de construir ε -subgradientes derivados diretamente da função objetivo suavizada. Os gradientes da função suavizada serão então denominados *subgradientes diferenciáveis*.

Os problemas de otimização não diferenciáveis formam uma classe bastante especial dentro do conjunto dos problemas de otimização contínua. Isto porque a grande maioria dos métodos utilizados para a resolução destes problemas foi originalmente projetada para situações nas quais a função objetivo pertence pelo menos à classe C^1 (primeira derivada contínua). Nestes casos o gradiente é uma rica fonte de informações sobre o comportamento da função objetivo nas vizinhanças de um ponto considerado, viabilizando assim a busca pelo ótimo, ainda que restrita a um âmbito local.

Por outro lado, quando a função objetivo for não diferenciável, os métodos de otimização tradicionais não se aplicam e novos métodos devem ser especialmente projetados para que a busca pelo ótimo possa ser realizada de uma forma eficaz¹ e eficiente².

Nesta tese a abordagem utilizada será a suavização da função objetivo utilizando a suavização hiperbólica e será provado que o gradiente da função suavizada parametrizada é um ε -subgradiente da função objetivo original, para qualquer ε dado. Podemos então divisar uma equivalência entre as abordagens diretas (que não alteram o problema original) e as abordagens que utilizam a suavização.

O contexto dos principais desenvolvimentos será o de uma não diferenciabilidade

¹Um ótimo possa ser efetivamente encontrado.

²Em termos de complexidade computacional.

em um ponto isolado, mas estas idéias serão utilizadas para a suavização de funções em um contexto mais amplo.

No capítulo 2 serão apresentadas a revisão bibliográfica e os principais conceitos relacionados com a área de otimização não diferenciável. Este capítulo também inclui uma lista de definições que serão utilizadas como base para os desenvolvimentos posteriores.

No capítulo 3 é apresentada a definição do problema geral não diferenciável ir-restrito e são comentadas algumas de suas características principais. São também apresentados os principais conceitos relacionados com as técnicas de suavização de funções convexas desenvolvidas nesta tese, tais como a região de não diferenciabilidade, as funções ZH convexas, a definição de suavização adotada, a distância à região de não diferenciabilidade, a decomposição DA e a suavização SDA associada.

Nos capítulos 4, 5, 6 e 7 são realizados desenvolvimentos específicos para uma região de não diferenciabilidade constituída por apenas um ponto. Alguns destes desenvolvimentos serão utilizados como base para desenvolvimentos nos capítulos posteriores.

No capítulo 4 são discutidos processos de decomposição da função objetivo com a finalidade de isolar a componente que será o foco do processo de suavização. A decomposição DHA é apresentada, juntamente com suas propriedades.

No capítulo 5 é discutido o problema de suavizar a componente não diferenciável da função objetivo presente na decomposição DHA, culminando com o desenvolvimento teórico da suavização SDHA e da primeira versão do *subgradiente diferenciável*. Esta denominação, talvez considerada "bizarra" para os ouvidos mais acostumados com a nomenclatura tradicional da área, é justificada nos capítulos seguintes. A intenção implícita é justamente provocativa.

No capítulo 6 é apresentado um resultado que demonstra a possibilidade de se produzir ε -subgradientes da função objetivo com base em ε -subgradientes da componente não diferenciável da decomposição DHA. É um passo preparatório para se utilizar subgradientes diferenciáveis na geração de ε -subgradientes da função objetivo.

No capítulo 7 a pertinência de subgradientes diferenciáveis em ε -subdiferenciais é desenvolvida considerando-se a convexidade da componente diferença presente na decomposição DHA da função objetivo. Os resultados demonstram a existência de uma desigualdade entre os parâmetros ε do subgradiente e τ da suavização SDHA (independentemente do ponto $x \in \mathbb{R}^n$ considerado) a qual assegura a existência de um subgradiente diferenciável para qualquer ε dado.

No capítulo 8 não se considera que a componente diferença da decomposição DHA seja convexa e toma-se o caminho alternativo de se tentar desenvolver para a suavização SDA os mesmos resultados auferidos para a suavização SDHA, incluindo

a definição de uma nova versão do subgradiente diferenciável. Curiosamente, o resultado principal desenvolvido no capítulo 7 permanece válido: é possível desenvolver a mesma desigualdade entre os parâmetros ε do subgradiente e τ da suavização SDA (mais uma vez independentemente do ponto $x \in \mathbb{R}^n$ considerado) a qual assegura a existência de um subgradiente diferenciável para qualquer ε dado. Mais ainda, é provado que esta desigualdade é válida para as funções ZH convexas em geral, no sentido de que elas são uma generalização natural das funções convexas com um único ponto de não diferenciabilidade.

No capítulo 9 serão apresentados resultados práticos da aplicação da teoria do subgradiente diferenciável a alguns exemplos discutidos na tese e também a alguns problemas mencionados na literatura da área, os quais exemplificam a diversidade de estruturas não diferenciáveis que podem ser tipicamente encontradas nesta classe de problemas.

O capítulo 10 foi reservado para a apresentação das conclusões deste trabalho, assim como algumas sugestões para a continuação do mesmo.

O apêndice A consiste de alguns resultados importantes que foram utilizados durante os desenvolvimentos desta tese, os quais foram referenciados quando necessário. A razão para a sua alocação no apêndice foi justamente a de não desviar o foco dos desenvolvimentos em curso. Alguns resultados são simplesmente exemplos de casos excepcionais, enquanto outros consolidam propriedades que em conjunto são estrategicamente úteis, facilitando assim a sua utilização no corpo desta tese.

Capítulo 2

Revisão Bibliográfica

A área da otimização não diferenciável é mais geral e complexa, quando comparada com a área da otimização que trata dos problemas diferenciáveis, e demanda a utilização de conceitos especializados.

Uma figura pioneira e que estabeleceu resultados básicos para as áreas da Otimização Não Linear e Análise Convexa foi M. W. Fenchel (FENCHEL, 1953; FENCHEL e BONNESEN, 1934). A definição do subdiferencial para funções convexas não diferenciáveis por R.T. Rockafellar e J.J. Moreau (ROCKAFELLAR, 1970) no começo da década de 1960 foi o ponto de partida para a criação do *Cálculo Subdiferencial*, fruto dos esforços de desenvolvimento da Análise Convexa. Vários conceitos similares foram posteriormente criados (BORWEIN e ZHU, 1999), generalizando a idéia do subdiferencial para funções não convexas, sendo o *gradiente generalizado* de F.H. Clarke (CLARKE, 1983) considerado o mais importante. Daremos continuidade à revisão bibliográfica com base nas abordagens para solucionar problemas não diferenciáveis, apresentadas a seguir.

2.1 Discussão sucinta das abordagens

A área de otimização não diferenciável exige a utilização de métodos especializados. Isto porque, no caso dos problemas diferenciáveis, o gradiente da função objetivo (se disponível) fornece informações tão relevantes para a busca do ótimo que nenhum método de otimização reconhecidamente eficiente neste contexto o despreza¹. É o caso de métodos consagrados tais como os métodos do gradiente, do gradiente-conjugado e quasi-Newton.

A indisponibilidade do gradiente nos problemas não diferenciáveis é portanto uma séria dificuldade e abordagens alternativas foram criadas, não existindo necessariamente uma preferencial, dependendo do problema sob análise. Na Figura 2.1

¹Ainda que seja para a busca de ótimos locais em problemas de otimização global.

apresentamos um exemplo de uma função $f(x)$ definida no \mathbb{R}^1 onde são destacadas as retas que representam as inclinações dadas por gradientes, subgradientes e ε -subgradientes. Os subgradientes e os ε -subgradientes são alternativas ao uso dos gradientes nos casos em que as funções são não diferenciáveis. Esta figura também apresenta uma versão suavizada $f_s(x)$ da função original e um ε -subgradiente de $f(x)$ correspondente ao gradiente de $f_s(x)$ no ponto x_b . É justamente esta equivalência que será desenvolvida nesta tese.

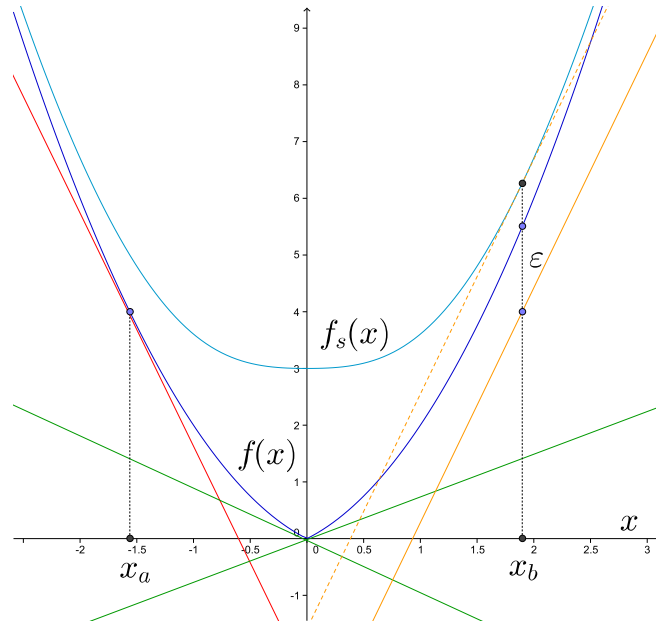


Figura 2.1: Exemplo no \mathbb{R}^1 onde as inclinações das retas representam um gradiente (no ponto x_a), alguns subgradientes ($f(x)$ não diferenciável na origem) e um ε -subgradiente (no ponto x_b) de uma função convexa $f(x)$.

Apresentaremos agora uma breve descrição das abordagens tradicionalmente utilizadas, visando contextualizar os desenvolvimentos posteriores.

2.1.1 Abordagem direta (problema original é mantido)

Nesta abordagem os métodos utilizados nos problemas diferenciáveis são adaptados ou novos métodos são criados visando garantir a convergência para uma solução ótima, apesar da não diferenciabilidade.

As principais opções relacionadas com a adaptação dos métodos tradicionais são o método do subgradiente, o método do ε -subgradiente, o método dos planos cortantes e o método dos feixes (IZMAILOV e SOLODOV, 2007). Quanto aos novos métodos, eles basicamente são da classe "derivative free" (heurísticas, livres de derivadas ou gradientes), tais como os métodos Simplex de Nelder-Mead, "Simulated Annealing", "Genetic Algorithms", "Random Search", etc (CONN *et al.*, 2008).

Com a abordagem direta busca-se resolver o problema em seu formato original, sem tentar simplificá-lo. Manter o problema original implica em fornecer alternativas para uma série de elementos que são utilizados nos algoritmos de otimização irrestritos quando a função objetivo é de classe C^1 . De uma forma geral, os algoritmos para resolução de problemas gerais de programação não-lineares se caracterizam pelos seguintes elementos:

- Escolha de um ponto inicial x_0
- Escolha de uma direção d_k
- Escolha de um passo α_k
- Critério de parada
- Garantias de convergência, pelo menos linear

Normalmente a escolha do ponto inicial x_0 depende do problema em questão, não sendo uma característica intrínseca do algoritmo. Já a escolha da direção d_k , do passo α_k e do critério de parada, assim como as características de convergência, dependem especificamente do algoritmo utilizado. O ideal é que a direção d_k seja uma direção de descida, mas isto nem sempre é possível, como por exemplo no método do subgradiente, onde não existem garantias de que as novas direções geradas pelo algoritmo serão sempre de descida. Por outro lado, o estudo da convergência para estes métodos torna-se mais complexo, como é o caso do método dos feixes, onde são definidos os passos "nulo" e "sério".

Existe uma outra dificuldade nos algoritmos pertencentes a esta abordagem: a possibilidade de convergência para um ponto que não é ótimo (o algoritmo não é fechado). Este tipo de dificuldade ocorre devido a imperfeições na adaptação dos métodos tradicionais que tratam da otimização de funções diferenciáveis. O exemplo mais conhecido é o método do subgradiente operando em uma subregião diferenciável. Como o subdiferencial de qualquer ponto x pertencente a uma destas subregiões constitui-se de um conjunto contendo apenas um subgradiente ($\partial f(x) = \{\nabla f(x)\}$), se os passos forem calculados com base somente nesta informação (método do gradiente), em algumas situações poderá existir uma aparente convergência, mas para um ponto que não necessariamente é ótimo, como apresentado no início do capítulo 2 de SHOR (1985).

2.1.2 Abordagem suavizada (ou suavizadora)

Nesta abordagem a função objetivo original não diferenciável é substituída por uma função diferenciável parametrizada $f_s(x)$ (uma "suavização") que tende para a função objetivo original quando seus parâmetros convergem para valores pré-definidos

(normalmente 0 ou $+\infty$, dependendo da parametrização), e os métodos diferenciáveis tradicionais são então utilizados. A solução final é o limite das soluções encontradas para os problemas modificados quando os valores dos parâmetros convergem para seus valores pré-definidos.

Uma outra forma de se compreender o que ocorre na suavização é considerar que a não-diferenciabilidade presente em um ponto z do domínio da função objetivo é eliminada mediante uma "distribuição" de seus efeitos, concentrados no ponto z , para uma vizinhança do ponto z . Este é um ponto de vista importante e demonstra o real desafio na implementação dos algoritmos para otimização não-diferenciável (parcialmente resolvido em teoria pelo uso dos ε -subgradientes): as informações locais ($f(x)$ e um subgradiente $g \in \partial f(x)$) não incorporam os efeitos "remotos" das não-diferenciabilidades e são portanto fontes limitadas de informação quando utilizadas nos algoritmos tradicionais de otimização.

Nos métodos que envolvem ε -subgradientes, pertencentes às abordagens diretas, busca-se justamente agregar em um ponto os subgradientes de sua vizinhança, o que confere a este método sensibilidade quanto aos efeitos "remotos" das não-diferenciabilidades. No entanto, com a exceção de casos especiais, a geração de ε -subgradientes não é trivial.

O efeito da suavização nos pontos de não diferenciabilidade é normalmente inexpressivo (teoricamente o algoritmo nunca passará por eles, pois estes pontos formam um conjunto com medida de Lebesgue nula). Por outro lado, os efeitos da suavização das não diferenciabilidades dos pontos "remotos" z no ponto x_k da k -ésima iteração do algoritmo de otimização são o que realmente importa, pois a suavização automaticamente incorpora na função suavizada $f_s(x)$ os efeitos "remotos" das não-diferenciabilidades.

Uma das primeiras ferramentas utilizadas nesta abordagem foi o mapeamento proximal, definido por MOREAU (1965), o qual possui propriedades suavizadoras. Já BERTSEKAS (1974) considera problemas nos quais a função objetivo apresenta não diferenciabilidades exclusivamente pela presença de termos da forma $\max\{0, f_i(x)\}$, $i \in I$, chamados de "simple kinks". Posteriormente surgiu a técnica de suavização por funções assintóticas criada por BEN-TAL e TEBoulLE (1988). Em 1982 surgiu a penalização hiperbólica (XAVIER, 1982, 2001), que posteriormente deu origem à técnica da suavização hiperbólica (XAVIER *et al.*, 1989). A suavização hiperbólica foi posteriormente utilizada na resolução de problemas clássicos da literatura (XAVIER, 2009, 2011; XAVIER e DE OLIVEIRA, 2005) e é o objeto de estudo deste trabalho.

Atualmente existem outras alternativas que podem ser encontradas na literatura, mas muitas estão ligadas a problemas específicos, tais como Minimax, etc.

2.2 Definições básicas

Serão agora apresentadas uma série de definições que tem por base as referências bibliográficas, principalmente ROCKAFELLAR (1970) e LIMA (1981).

Iremos utilizar as notações B^n (resp. D^n) para representar a bola aberta (resp. fechada) com centro na origem e raio unitário imersa no espaço \mathbb{R}^n e S^n para representar a esfera com centro na origem e raio unitário imersa no espaço \mathbb{R}^n :

$$B^n = \{x \in \mathbb{R}^n : \|x\| < 1\}, \quad (2.1)$$

$$D^n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}, \quad (2.2)$$

$$S^n = \{x \in \mathbb{R}^n : \|x\| = 1\}. \quad (2.3)$$

É claro que a bola aberta com centro no ponto a e raio r pode então ser representada por $a + rB^n$, onde a soma é no sentido de Minkowski. O mesmo artifício também pode ser utilizado para bolas fechadas e esferas com centros fora da origem e raios quaisquer.

A reta real estendida é a reta real acrescida de $-\infty$ e $+\infty$, e será denotada por $\overline{\mathbb{R}}$.

Um subconjunto $C \subset \mathbb{R}^n$ é denominado convexo se ele contém o segmento ligando dois quaisquer de seus pontos.

Uma função f do \mathbb{R}^n em $\overline{\mathbb{R}}$ é denominada convexa se o seu epígrafo, definido por

$$\text{epi } f = \{(x, \mu) : x \in \mathbb{R}^n, \mu \in \mathbb{R}, f(x) \leq \mu\} \subseteq \mathbb{R}^{n+1} \quad (2.4)$$

for um conjunto convexo.

Um hiperplano H tal que $\text{epi } f$ está totalmente contido em um dos semi-espacos fechados determinados por H e possui pelo menos um ponto em comum com $\text{epi } f$ é chamado de hiperplano de suporte a f .

O domínio efetivo de f é definido como:

$$\text{dom } f = \{x \in \mathbb{R}^n : f(x) < +\infty\}. \quad (2.5)$$

A função convexa f é chamada de *própria* se $\text{dom } f \neq \emptyset$ e $f(x) > -\infty, \forall x \in \text{dom } f$.

Quando uma função f é própria e $\text{dom } f = \mathbb{R}^n$, podemos considerá-la como uma função de \mathbb{R}^n em \mathbb{R} (e não $\overline{\mathbb{R}}$), pois os valores $-\infty$ ou $+\infty$ nunca serão atingidos.

O conjunto de nível da função f é definido como:

$$L_c(f) = \{x \in \mathbb{R}^n : f(x) \leq c\}. \quad (2.6)$$

A função f mencionada nas definições seguintes será sempre uma função convexa própria de \mathbb{R}^n em \mathbb{R} , não necessariamente diferenciável.

Um subgradiente de f no ponto x é definido como qualquer $g \in \mathbb{R}^n$ que satisfaça a seguinte desigualdade:

$$f(y) \geq f(x) + g \cdot (y - x), \forall y \in \mathbb{R}^n. \quad (2.7)$$

O subdiferencial de f no ponto x é definido como o conjunto de todos os subgradientes de f neste ponto:

$$\partial f(x) = \{g \in \mathbb{R}^n : f(y) \geq f(x) + g \cdot (y - x), \forall y \in \mathbb{R}^n\}. \quad (2.8)$$

De forma similar, dado um $\varepsilon > 0$ define-se um ε -subgradiente de uma função f no ponto x como sendo qualquer $g \in \mathbb{R}^n$ que satisfaça a seguinte desigualdade:

$$f(y) \geq f(x) + g \cdot (y - x) - \varepsilon, \forall y \in \mathbb{R}^n. \quad (2.9)$$

E o ε -subdiferencial de f no ponto x é definido como o conjunto de todos os ε -subgradientes deste ponto:

$$\partial_\varepsilon f(x) = \{g \in \mathbb{R}^n : f(y) \geq f(x) + g \cdot (y - x) - \varepsilon, \forall y \in \mathbb{R}^n\}. \quad (2.10)$$

Nas condições em que estamos estudando ($f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexa), podemos afirmar que tanto ∂f quanto $\partial_\varepsilon f$ são conjuntos não-vazios, convexos e compactos. É claro que para todo $x \in \mathbb{R}^n$, $\partial f(x) \subset \partial_\varepsilon f(x)$ e $\partial_0 f(x) = \partial f(x)$.

A derivada direcional (unilateral) sempre existe para funções convexas e é definida para a direção $d \in \mathbb{R}^n$ como:

$$f'(x, d) = \lim_{t \rightarrow 0^+} \frac{f(x + td) - f(x)}{t}. \quad (2.11)$$

Esta definição implica em que a derivada direcional é (positivamente) homogênea de grau um.

A derivada de $f(x)$, que no caso de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é chamada de gradiente e representada por $\nabla f(x)$, existe se o seguinte limite existir:

$$\lim_{h \rightarrow 0} \frac{f(x + h) - f(x) - \nabla f(x) \cdot h}{\|h\|} = 0. \quad (2.12)$$

Como um caso especial, se tivermos

$$\lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{\|h\|} = 0, \quad (2.13)$$

então, pela unicidade do gradiente em x , temos que $\nabla f(x) = 0$.

Os subgradientes estão intimamente relacionados com as derivadas direcionais, como apresentado abaixo:

$$f'(x, d) = \max_{g \in \partial f(x)} g \cdot d, \forall d \in \mathbb{R}^n. \quad (2.14)$$

Em particular, a expressão acima pode ser inserida na desigualdade geral dos subgradientes, resultando em:

$$f(y) \geq f(x) + f'(x, y - x), \forall y \in \mathbb{R}^n. \quad (2.15)$$

Dado que $f'(x, d)$ é homogênea, a igualdade (2.14) pode ser expressa utilizando somente vetores unitários $u \in S^n$:

$$f'(x, u) = \max_{g \in \partial f(x)} g \cdot u. \quad (2.16)$$

Com base nesta observação, facilmente se percebe que:

$$\bigcap_{u \in S^n} \{g \in \mathbb{R}^n : g \cdot u \leq f'(x, u)\} = \partial f(x). \quad (2.17)$$

A função conjugada convexa de $f(x)$ é definida como:

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{x \cdot y - f(x)\}. \quad (2.18)$$

A função conjugada desempenha um papel fundamental na Análise Convexa, similar ao papel da Transformada de Laplace na Análise convencional. Por definição, vale a desigualdade de Fenchel:

$$f(x) + f^*(y) \geq y \cdot x, \quad (2.19)$$

para todo $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^n$.

Capítulo 3

Apresentação do problema e a suavização de funções

Apresentaremos agora o problema básico que será analisado nesta tese:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & x \in \mathbb{R}^n, \end{aligned} \tag{3.1}$$

onde iremos supor que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é própria, convexa (e portanto contínua em seu domínio efetivo, que é o \mathbb{R}^n) e de classe pelo menos C^1 em $\mathbb{R}^n \setminus Z_f$, onde Z_f é um subconjunto fechado de \mathbb{R}^n que representa a região de não diferenciabilidade de f . Desta forma, temos $\nabla f(x)$ bem definido para qualquer ponto pertencente ao conjunto aberto $\mathbb{R}^n \setminus Z_f$. Especificamente nos pontos $z \in Z_f$, no entanto, não temos o gradiente $\nabla f(z)$ definido, mas como f é convexa temos o subdiferencial $\partial f(z)$ sempre bem definido.

A razão para trabalharmos com funções convexas se deve ao fato de que os ε -subdiferenciais somente são definidos para funções desta natureza. Como nosso objetivo é encontrar uma equivalência entre os ε -subgradientes e os gradientes de funções suavizadas (subgradientes diferenciáveis), focaremos em funções com as propriedades gerais apresentadas acima.

3.1 Definição da suavização de funções

A literatura (BEN-TAL e TEBoulLE, 1988; BERTSEKAS, 1974; MOREAU, 1965; XAVIER, 2009, 2011; XAVIER e DE OLIVEIRA, 2005) apresenta vários processos utilizados para realizar a suavização de funções (muitos dos quais são "ad hoc"), ficando muitas vezes implícito o que é a própria suavização de funções. Precisamos então definir formalmente uma suavização paramétrica e assim criar o devido embasamento para os desenvolvimentos posteriores.

Definição 3.1. Seja uma função real $f(x)$ com domínio no \mathbb{R}^n e região de não diferenciabilidade Z_f . Sejam $\tau \in \mathbb{R}^p$ vetor paramétrico, $\tau_0 \in \mathbb{R}^p$ vetor fixo, e $d(x, Z_f)$ uma função distância genérica entre $x \in \mathbb{R}^n$ e Z_f . Uma suavização paramétrica de $f(x)$ é uma função paramétrica real $f_s(\tau, x)$ com o mesmo domínio de $f(x)$ que possui as seguintes propriedades, para todo $x \in \mathbb{R}^n$:

1. $f_s(\tau, x)$ é pelo menos de classe C^1 para $\tau \neq \tau_0$;
2. Para todo $x \in \mathbb{R}^n$, $f_s(\tau, x)$ converge para $f(x)$ quando $\tau \rightarrow \tau_0$ ou seja,

$$\lim_{\tau \rightarrow \tau_0} |f(\tau, x) - f(x)| = 0 \quad \forall x \in \mathbb{R}^n;$$

3. $\nabla f_s(\tau, x) \rightarrow \nabla f(x)$ quando $d(x, Z_f) \rightarrow \infty$, para todo $x \in \mathbb{R}^n \setminus Z_f$.

Em casos mais simples podemos ter uma versão monoparamétrica na qual o vetor τ é constituído por apenas um parâmetro real. Nestes casos o vetor τ_0 é simplesmente um número real.

A segunda condição é natural e indica que podemos fazer $f_s(\tau, x)$ se aproximar de $f(x)$ o quanto quisermos, bastando para isto aproximarmos suficientemente τ de τ_0 .

A terceira condição exprime a noção intuitiva de que a suavização deve ser um processo local à região de não diferenciabilidade Z_f , ou seja, a função suavizada deve ter um comportamento diferencial praticamente idêntico ao da função original em pontos distantes da região de não diferenciabilidade. Esta condição, de forma isolada, permite que o limite da função $f(\tau, x)$ quando $\tau \rightarrow \tau_0$ possua um nível (deslocamento vertical) diferente que o da função original $f(x)$, algo sem consequências em problemas de otimização, sendo esta a razão de estar sob forma diferencial. A motivação para esta condição se origina no fato de que os algoritmos de otimização que trabalham com funções diferenciáveis sempre utilizam o gradiente e ele deve ser praticamente o mesmo que o da função não diferenciável original em pontos distantes da região de não diferenciabilidade. É importante que a função distância $d(x, Z_f)$ esteja bem definida, não sujeita a ambigüidades, e compatível com o processo de suavização almejado.

Esta última condição também nos faz perceber que a suavização deve ser um processo que combine o valor da função $f(x)$, para $x \notin Z_f$ com o valor de $f(z)$ para algum $z \in Z_f$, e os pesos utilizados nesta combinação devem depender da distância $d(x, Z_f)$. Esta será a idéia básica que será desenvolvida nas seções posteriores.

A definição 3.1 é simples e concisa, e será a que iremos utilizar a partir deste ponto.

3.2 Regiões de não diferenciabilidade Z_f

Conforme definido anteriormente, a região de não diferenciabilidade Z_f é constituída pelo conjunto de pontos onde a função convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ não é diferenciável e portanto $\mathbb{R}^n \setminus Z_f$ é o conjunto dos pontos onde f é diferenciável. Por outro lado, a teoria das funções convexas nos afirma que f possui derivadas direcionais bem definidas em cada ponto do domínio efetivo de f (em nosso caso o próprio \mathbb{R}^n) e portanto também temos subdiferenciais bem definidos neste domínio. Podemos então definir explicitamente a região de não diferenciabilidade para f convexa como a união de todos os pontos de seu domínio para os quais o subdiferencial não seja unitário¹:

$$Z_f = \{z \in \mathbb{R}^n \mid \#\partial f(z) > 1\},$$

onde utilizamos " $\#$ " para denotar a cardinalidade de um conjunto.

Iremos agora elaborar um pouco mais o conceito de região de não diferenciabilidade, listando propriedades e impondo algumas restrições simplificadoras para facilitar o desenvolvimento de resultados teóricos.

Propriedade 1: Z_f é um conjunto fechado. Esta propriedade é consequência imediata da definição de derivada 2.12, a qual demanda que para todo ponto de diferenciabilidade x exista um $r > 0$ tal que $x \in (x + rB^n) \subset \mathbb{R}^n \setminus Z_f$. Assim podemos afirmar que $\mathbb{R}^n \setminus Z_f$, visto como união de uma família de bolas abertas, é também um conjunto aberto e conseqüentemente temos que Z_f é um conjunto fechado, algo já mencionado no início deste capítulo.

Hipótese 1: Z_f deve possuir medida de Lebesgue nula. Primeiramente enunciemos o teorema de Rademacher.

Teorema 3.2. (Rademacher, 1919). *Se $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é Lipschitz, então f é diferenciável em quase todo ponto.*

Em virtude deste teorema, se nossa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ for uma função globalmente Lipschitz contínua no \mathbb{R}^n então Z_f é um conjunto de medida de Lebesgue nula.

Sendo f convexa, somente podemos afirmar que ela é *localmente* Lipschitz contínua em seu domínio efetivo \mathbb{R}^n . No entanto, quando $Z_f \subset U \subset \mathbb{R}^n$, sendo U aberto e limitado, temos que f restrita a U é globalmente Lipschitz e podemos então afirmar que Z_f tem medida de Lebesgue nula. Para simplificar o caso geral, no qual Z_f pode não ser compacto (e portanto não estar contido em nenhum aberto U limitado), iremos impor a restrição de que Z_f sempre tenha medida de Lebesgue

¹Para f convexa temos sempre $\partial f(x) \neq \emptyset, \forall x \in \mathbb{R}^n$ e se f é diferenciável em $x \in \mathbb{R}^n$ então $\partial f(x) = \{\nabla f(x)\}$, um conjunto unitário.

nula para as funções que serão tratadas nesta tese. Assim, no \mathbb{R}^2 poderemos ter Z_f constituída por pontos e curvas, mas nenhuma região que possua área positiva. De forma similar, no \mathbb{R}^3 poderemos ter Z_f constituída por pontos, curvas e regiões que possuam área positiva, mas nenhuma região que possua volume positivo.

Mesmo com estas considerações iniciais, uma região de não diferenciabilidade fechada e de medida nula ainda possui um certo grau de liberdade geométrica, isto porque ainda não exploramos totalmente a convexidade de f .

Antes de continuarmos a listar mais características de Z_f , iremos comentar um pouco sobre a função $d(x, Z_f)$, mencionada em nossa definição de suavização paramétrica. Esta distância deverá servir, de alguma forma, para realizar uma combinação suave entre os valores $f(x)$ e os valores $f(z)$, com $z \in Z_f$. A questão é: dado x , qual z escolher? Uma primeira idéia seria trabalharmos com a projeção z de um ponto qualquer $x \in \mathbb{R}^n$ sobre Z_f e definirmos $d(x, Z_f)$ como a distância euclidiana entre x e z .

Seguindo esta linha de raciocínio iremos definir a região M_f , formada pelos pontos de $\mathbb{R}^n \setminus Z_f$ para os quais a distância ao conjunto Z_f é satisfeita por mais de um ponto em Z_f . Consideremos a função conjunto projeção de $x \in \mathbb{R}^n \setminus Z_f$ em Z_f :

$$\Pr(x) = \arg \min_{z \in Z_f} \|z - x\|.$$

Devemos observar que o mínimo sempre existe e é atingido por pelo menos um elemento de Z_f , pois este é um conjunto fechado². Podemos então definir M_f explicitamente como:

$$M_f = \{x \in \mathbb{R}^n \setminus Z_f : \# \Pr(x) > 1\}.$$

Alguns exemplos poderão ilustrar de forma mais efetiva as possíveis configurações geométricas para os conjuntos Z_f e M_f . Sejam as seguintes funções convexas reais com domínio no \mathbb{R}^2 :

$$\begin{aligned} f_1(x) &= \max\{0, \|x\| - 1\}, \\ f_2(x) &= \max\{|x_1|, |x_2|\}, \\ f_3(x) &= \begin{cases} |x_2|, & -1 \leq x_1 \leq 1, \\ \|x - (1, 0)\|, & x_1 > 1, \\ \|x - (-1, 0)\|, & x_1 < -1. \end{cases} \end{aligned}$$

Apresentamos nas Figuras 3.1, 3.2 e 3.3 os gráficos das funções f_1 , f_2 e f_3 , assim como uma representação de suas respectivas regiões (Z_{f_1}, M_{f_1}) , (Z_{f_2}, M_{f_2}) , (Z_{f_3}, M_{f_3}) .

²Conforme apresentado na seção 8.1 de BOYD e VANDENBERGHE (2009).

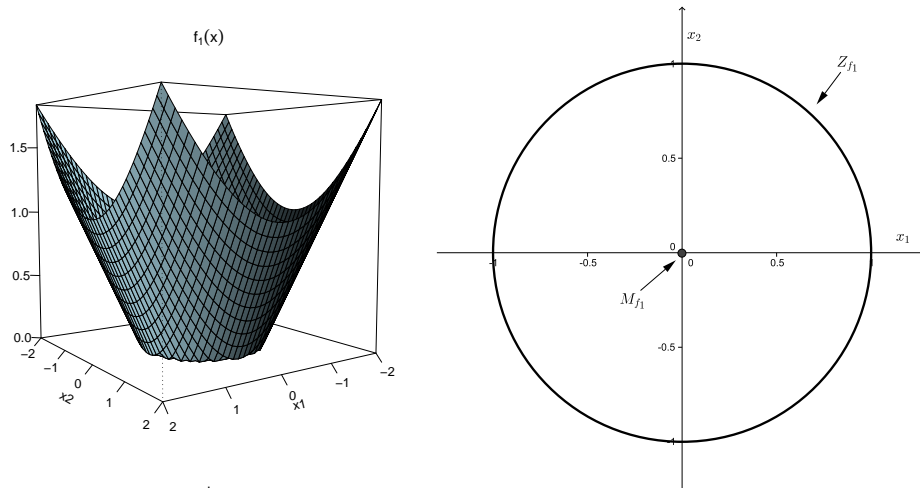


Figura 3.1: Gráfico de $f_1(x)$ e suas regiões Z_{f_1} e M_{f_1}

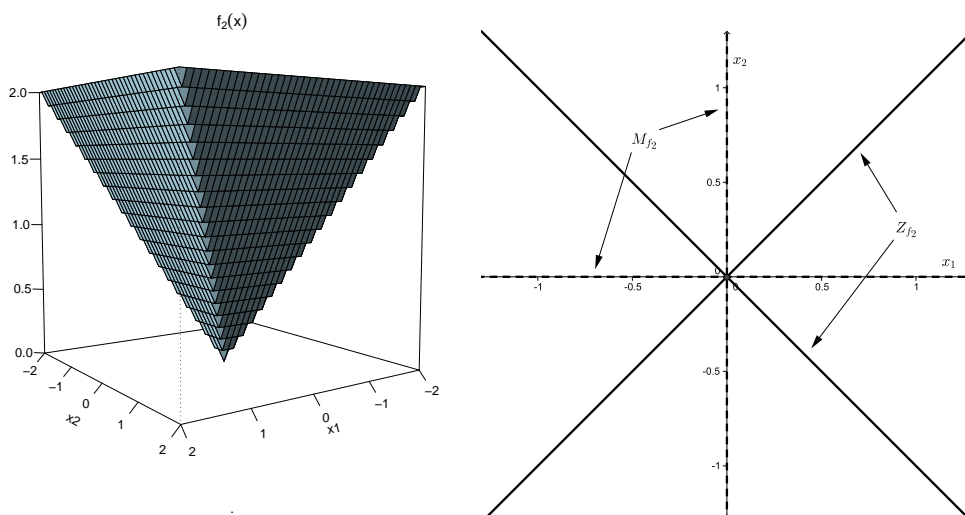


Figura 3.2: Gráfico de $f_2(x)$ e suas regiões Z_{f_2} e M_{f_2}

Como se pode facilmente perceber pela Figura 3.1, a região de não diferenciabilidade Z_{f_1} tem a forma de uma circunferência de raio unitário, sendo portanto conexa, compacta (fechada e limitada), mas não convexa. A região M_{f_1} é constituída apenas pela origem O e para este ponto $\text{Pr}_{Z_f}(O)$ é a própria circunferência, pois todos estes pontos atingem a distância unitária com relação à origem. Assim:

$$\begin{aligned} Z_{f_1} &= \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}, \\ M_{f_1} &= \{(0, 0)\}. \end{aligned}$$

Na Figura 3.2 temos uma região de não diferenciabilidade Z_{f_2} conexa, fechada, ilimitada e não convexa³, formada por duas retas perpendiculares que passam pela origem, ambas com inclinações de 45° em relação aos eixos x_1 e x_2 . A região M_{f_2}

³O serrilhado presente no gráfico de $f_2(x)$ é um "artefato" criado pelo gerador de gráficos.

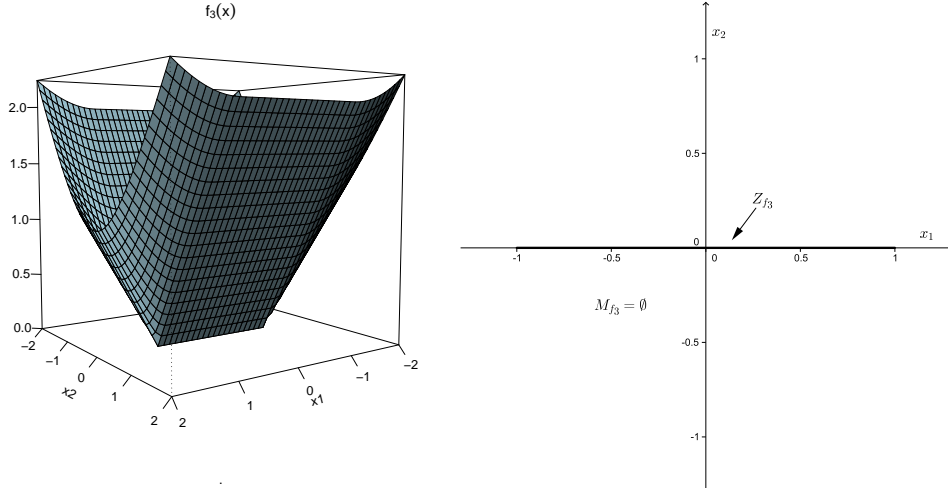


Figura 3.3: Gráfico de $f_3(x)$ e suas regiões Z_{f_3} e M_{f_3}

é constituída por duas retas perpendiculares que coincidem com os eixos x_1 e x_2 e cada ponto de M_{f_2} , exceto a origem, possui duas projeções distintas sobre Z_{f_2} . Assim:

$$\begin{aligned} Z_{f_2} &= \{(x_1, x_2) \in \mathbb{R}^2 : |x_1| = |x_2|\}, \\ M_{f_2} &= \{(x_1, 0) \in \mathbb{R}^2 : |x_1| > 0\} \cup \{(0, x_2) \in \mathbb{R}^2 : |x_2| > 0\}. \end{aligned}$$

Já na Figura 3.3 temos uma região de não diferenciabilidade Z_{f_3} conexa, fechada, limitada e convexa formada por um segmento de reta ligando os pontos $(-1, 0)$ e $(1, 0)$. A região M_{f_3} é vazia pois todo ponto $x \in \mathbb{R}^2 \setminus Z_{f_3}$ possui projeção única. Assim:

$$\begin{aligned} Z_{f_3} &= \{(x_1, 0) \in \mathbb{R}^2 : -1 \leq x_1 \leq 1\}, \\ M_{f_3} &= \emptyset. \end{aligned}$$

Por definição, quando $M_f \neq \emptyset$, não existe unicidade na projeção de um ponto $x \in \mathbb{R}^n$ sobre Z_f . Se o processo de suavização utilizasse o valor de f na projeção z sobre o conjunto Z_f poderíamos ter ambigüidade na escolha de qual $z \in Z_f$ deveria ser utilizado. Como em várias situações poderemos ter $M_f \neq \emptyset$, somos forçados a abandonar a projeção tradicional de x sobre o conjunto Z_f como um componente de nossa metodologia de suavização.

Para tratarmos suavizações de funções convexas mais gerais teremos portanto que definir uma nova função projeção e uma nova função distância de um ponto x ao conjunto Z_f . Poderíamos no entanto prescindir da unicidade de uma função projeção se fosse possível definir uma função distância para a qual pudéssemos ter $d(x, z)$ independente do particular $z \in Z_f$ escolhido. Se de alguma forma tivéssemos

$f(z)$ independente do particular $z \in Z_f$ poderíamos pensar em utilizar a própria função f para definir $d(x, z)$. É com este intuito que enunciaremos mais uma hipótese a ser cumprida por Z_f .

Hipótese 2: Existe hiperplano de suporte ao epígrafo que contém $(z, f(z))$ para todo $z \in Z_f$. Consideremos que o conjunto $G(Z_f) = \{(z, f(z)) : z \in Z_f\}$ esteja contido em um hiperplano de suporte não vertical H do epígrafo de $f(x)$. Seja $H(x)$ a função afim que representa este hiperplano de suporte. Temos que $-H(x)$ é convexa e a soma de funções convexas definida por $f_b(x) = f(x) - H(x)$ também é convexa. Por definição $f_b(x) \geq 0$ e em especial $f_b(z) = 0$ para todo $z \in Z_f$. Como $H(x)$ é infinitamente diferenciável, temos que $f_b(x)$ herda todas as características de diferenciabilidade de $f(x)$, possuindo assim a mesma região de não diferenciabilidade Z_f . Assim, todas as conclusões com relação às propriedades topológicas de Z_f mediante análise de $f_b(x)$ se aplicam diretamente a $f(x)$.

Como $f_b(x)$ é convexa, todos os seus conjuntos de nível são convexos, em particular o conjunto de nível de $f_b(x)$ para o nível zero, $L_0(f_b)$. Se o interior deste conjunto for não vazio, $f_b(x)$ terá que ser diferenciável nestes pontos interiores, pois neste caso $\text{int}\{L_0(f_b)\}$ possui medida plena⁴. Concluimos então que Z_f tem que estar contido na fronteira de $L_0(f_b)$: $Z_f \subset \partial\{L_0(f_b)\}$. Por outro lado, qualquer ponto $x \in \mathbb{R}^n$ pertence a um conjunto de nível $L_c(f_b)$ para algum $c \geq 0$ (basta tomarmos $c \geq f_b(x)$).

Pelo exposto acima, a função $f_b(x)$ parece ser apropriada para utilização na definição de uma distância, pois $f_b(z) = 0$ para todo $z \in Z_f$ (ou seja, independe do z utilizado). A suavização teria então que ser realizada sobre $f_b(x)$ e não diretamente sobre $f(x)$, mas se temos uma suavização de $f_b(x)$ disponível basta somarmos $H(x)$ para obtermos uma suavização de $f(x)$.

Poderíamos nos questionar se a Hipótese 2 não é muito restritiva. Na verdade muitas funções podem ser transformadas de tal forma a se enquadrar nesta hipótese e, após serem suavizadas, passarem pela transformação inversa para obtenção de uma versão suavizada da função original. Esta abordagem será utilizada em exemplos no final desta tese.

Sumarizando, as propriedades e hipóteses sobre o conjunto Z_f são:

- Z_f é um conjunto fechado, não necessariamente compacto;
- Z_f deve ser um conjunto com medida de Lebesgue nula;
- $G(Z_f) = \{(z, f(z)) : z \in Z_f\} \subset H$, onde H é um hiperplano de suporte não vertical H do epígrafo de $f(x)$.

⁴Temos $f_b(x) = 0$ para todo $x \in \text{int}\{L_0(f_b)\}$, e uma função constante definida em um conjunto de medida plena é trivialmente diferenciável.

Uma função convexa não diferenciável cuja região de não diferenciabilidade Z_f atenda às hipóteses acima descritas será doravante denominada uma *função ZH convexa*. Para as funções ZH convexas será possível divisar um procedimento para "redistribuir" a não diferenciabilidade concentrada no conjunto de medida nula Z_f para uma região de medida plena dentro do domínio da função f . A Figura 3.4 ilustra o epígrafo de uma função ZH convexa, destacando sua região de não diferenciabilidade Z_f .

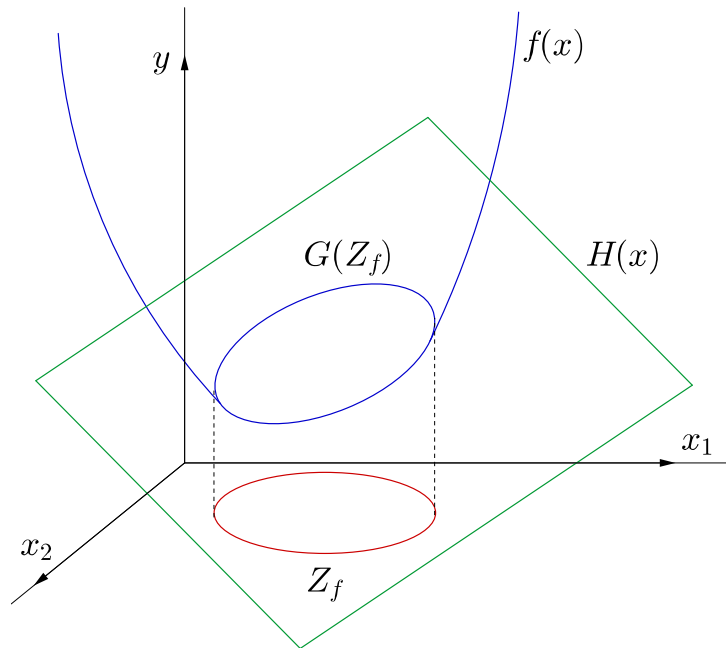


Figura 3.4: Exemplo de função $f(x)$ no \mathbb{R}^2 , com $G(Z_f) \subset H$, onde H é hiperplano de suporte ao epígrafo de $f(x)$, aqui representado pela função afim $H(x)$.

Voltando aos exemplos, devemos notar que a função Z_{f_2} não atende à terceira restrição e a princípio não poderia ser suavizada mediante a aplicação da metodologia desenvolvida. Esta função será discutida no final desta tese.

A partir deste ponto todas as regiões de não diferenciabilidade Z_f consideradas atenderão às propriedades listadas acima. Em outras palavras, iremos sempre trabalhar com funções ZH convexas.

3.3 Conceitos e nomenclatura

Nesta seção firmaremos alguns conceitos e nomenclatura, dentro do escopo de nossas necessidades específicas.

Primeiramente devemos sempre lembrar que estamos trabalhando com f convexa e $\text{dom } f = \mathbb{R}^n$, uma vez que não nos interessa avaliar o comportamento de f nas fronteiras de um domínio restrito. Desta forma não temos que nos preocupar com

singularidades que somente contribuiriam para aumentar a complexidade da análise de Z_f e a consequente suavização de f .

Estamos também considerando que f é própria, ou seja, não atinge o valor $-\infty$. Esta é uma restrição naturalmente imposta na formulação dos problemas de otimização considerados na prática. Em particular, para f própria é sempre possível definir um hiperplano de suporte H não vertical. Por outro lado, como também impomos $\text{dom } f = \mathbb{R}^n$, temos que todos os subdiferenciais são compactos e esta propriedade será fundamental para o desenvolvimento da suavização de f .

Devemos notar que uma função ZH convexa é automaticamente própria, pois um ponto $x \in \mathbb{R}^n$ para o qual $f(x) = -\infty$ implicaria em $f(x) < H(x)$, onde $H(x)$ representa o hiperplano de suporte não vertical H que contém $G(Z_f)$, o que invalidaria o fato de H ser um hiperplano de suporte. Assim, estamos caracterizando completamente nossas funções f sob análise ao exigirmos que sejam ZH convexas e $\text{dom } f = \mathbb{R}^n$.

Adicionalmente, em conformidade com os desenvolvimentos apresentados na seção A.1 do apêndice, iremos utilizar a nomenclatura alternativa $\partial_\infty f$ para a região $\text{dom } f^*$, pois esta é mais adequada para o nosso contexto, onde buscamos equivalências com ε -subdiferenciais. De uma forma geral, cada ponto $x \in \mathbb{R}^n$ do domínio de f é mapeado em um subconjunto $\partial f(x) \subset \partial_\infty f$, assim como dado um $\varepsilon \geq 0$, cada ponto $x \in \mathbb{R}^n$ é mapeado em um subconjunto $\partial_\varepsilon f(x) \subset \partial_\infty f$.

3.4 Decomposição DA

Quando uma função f é a base para a criação de novas funções, faz-se necessária uma profunda análise de seu comportamento. Um exemplo típico é a função distância euclidiana, utilizada em profusão para a criação de funções objetivo encontradas nos problemas de localização. Um esforço na suavização da função base f tem portanto grande repercussão na suavização da formulação original do problema, basicamente substituindo todas as ocorrências da função f por sua versão suavizada.

Conforme já mencionado na seção 3.2, para adequar uma função ZH convexa f ao processo de suavização devemos primeiramente considerar o hiperplano de suporte que contém $G(Z_f) = \{(z, f(z)) : z \in Z_f\}$. Seja $H(x) = \bar{g} \cdot x + b$, com $\bar{g}, b \in \mathbb{R}^n$, a função afim que representa este hiperplano de suporte no \mathbb{R}^{n+1} . Por seu intermédio iremos então definir

$$f_b(x) = f(x) - H(x).$$

Obviamente podemos recompor $f(x)$ como:

$$f(x) = (f(x) - H(x)) + H(x) = f_b(x) + H(x).$$

Este tipo de decomposição será doravante chamada de decomposição DA (Diferença-Afim), e iremos denominar as funções f_b e H como as componentes *diferença* e *afim* da decomposição DA.

Por definição temos:

$$\begin{cases} f_b(x) = 0 & , x \in Z_f, \\ f_b(x) \geq 0 & , x \in \mathbb{R}^n \setminus Z_f. \end{cases}$$

Voltando aos exemplos da seção 3.2, notamos que $H(x) = 0$ para todos os exemplos considerados, os quais podem ser vistos como os resultados deste primeiro passo preparatório para a suavização.

3.5 Definição da distância à região de não diferenciabilidade Z_f

Podemos notar que, de uma forma geral, temos $f_b(x)$ com valores cada vez maiores à medida que x se distancia (no sentido usual) de Z_f , e isto nos sugere adotar o valor de $f_b(x)$ como base para a definição de uma (pseudo) distância. Obviamente temos que verificar se esta idéia faz sentido. Definamos então:

$$d(x, y) = |f_b(x) - f_b(y)|, \quad x, y \in \mathbb{R}^n.$$

Temos portanto, para quaisquer $x, y, w \in \mathbb{R}^n$:

1. $d(x, y) = |f_b(x) - f_b(y)| \geq 0$,
2. $d(x, x) = |f_b(x) - f_b(x)| = 0$,
3. $d(x, y) = |f_b(x) - f_b(y)| = |f_b(y) - f_b(x)| = d(y, x)$,
4. $d(x, y) = |f_b(x) - f_b(y)| = |f_b(x) - f_b(w) + f_b(w) - f_b(y)| \leq |f_b(x) - f_b(w)| + |f_b(w) - f_b(y)| = d(x, w) + d(w, y)$.

E assim provamos que $d(x, y)$ é não negativa, nula para $y = x$, simétrica e atende à desigualdade triangular. Somente não podemos afirmar que $d(x, y) > 0$ para $x \neq y$, pois $d(x, y) = 0$ quando $f(x) = f(y)$. Desta forma $d(x, y)$ atende a todos os axiomas de uma pseudo-distância. Uma distância poderia ser gerada se considerássemos a relação de equivalência: $x \sim y$ se $d(x, y) = 0$. A classe de equivalência de x seria então definida por

$$[x] = \{y \in \mathbb{R}^n : d(y, x) = 0\} = \{y \in \mathbb{R}^n : f(y) = f(x)\},$$

e o conjunto de classes de equivalência seria dado pelo conjunto quociente \mathbb{R}^n / \sim . Como nosso objeto de interesse é o próprio domínio de f , e não o conjunto quociente, nos basta trabalhar com a pseudo-distância, que de forma simplificada chamaremos simplesmente de distância.

Para o desenvolvimento da suavização de uma função f iremos considerar a distância de um ponto $x \in \mathbb{R}^n$ arbitrário à região Z_f . Como $f_b(z) = 0$ para todo $z \in Z_f$, podemos escrever:

$$d(x, z) = |f_b(x) - f_b(z)| = |f_b(x)| = f_b(x), \forall z \in Z_f,$$

onde utilizamos o fato de que $f_b(x) \geq 0, \forall x \in \mathbb{R}^n$. Portanto, como havíamos anteriormente sugerido, o próprio valor da função f_b em um ponto $x \in \mathbb{R}^n$ arbitrário pode ser utilizado como uma distância genérica ao conjunto Z_f , e podemos portanto escrever $d(x, Z_f) = f_b(x), \forall x \in \mathbb{R}^n$. Note que em especial $d(x, Z_f) = 0, \forall x \in L_0(f_b)$.

O fato de f_b ser uma pseudo-distância não é um problema. Pelo contrário, esta definição é perfeitamente compatível com a idéia intuitiva de representar o afastamento do hiperplano que "contém" a parte não diferenciável da função f_b . Esta será portanto a distância utilizada nos desenvolvimentos posteriores.

3.6 Suavização da componente f_b da decomposição DA

Uma vez definida uma distância, resta agora definirmos como aplicá-la ao processo de suavização de f_b . Uma alternativa promissora é a composição de f_b com uma função suavizadora $\phi(t)$ de \mathbb{R}_+ para \mathbb{R}_+ , ou seja, definirmos uma nova função como $f_c(x) = \phi(f_b(x))$. Para que isto funcione, precisamos que:

- $\phi(t)$ se comporte como a identidade para valores altos de t e assim possamos reproduzir $f_b(x)$ quando seus valores forem altos, ou seja, quando estivermos "distantes" de Z_f . Esta restrição também significa que $\phi'(t)$ deverá se aproximar do valor unitário e, como $\nabla f_c(x) = \phi'(f_b(x))\nabla f_b(x)$, teremos também $\nabla f_c(x)$ se aproximando de $\nabla f_b(x)$;
- $\phi'(t)$ deverá se aproximar de zero quando t se aproximar de zero. Isto porque, como $\nabla f_c(x) = \phi'(f_b(x))\nabla f_b(x)$ e $\nabla f_b(z)$ não está definido para $z \in Z_f$, teremos $\nabla f_c(x)$ se aproximando de zero, valor este que pode ser utilizado como um valor que substitui todos os subgradientes pertencentes a Z_f .

Se adicionalmente tivermos $0 \leq \phi'(t) \leq 1$, este conjunto de propriedades pode ser compreendido como imposições sobre $\phi(t)$ para que $1 - \phi'(t)$ e $\phi'(t)$ sejam

utilizados como coeficientes de uma combinação convexa entre o subgradiente nulo, correspondente ao subgradiente que representa Z_f , e o gradiente $\nabla f_b(x)$:

$$\nabla f_c(x) = (1 - \phi'(f_b(x)))0 + \phi'(f_b(x))\nabla f_b(x).$$

O papel de $\phi'(t)$ como coeficiente de uma combinação convexa fica ainda mais claro se considerarmos o gradiente da versão suavizada de $f(x)$, que denotaremos por $f_s(x)$, dada pela substituição de $f_b(x)$ por sua versão suavizada $f_c(x)$:

$$f_s(x) = H(x) + f_c(x) = H(x) + \phi(f(x) - H(x)).$$

A suavização acima será doravante denominada de SDA, ou seja, suavização da decomposição DA. Iremos sempre utilizar a notação geral f_s para a suavização da função f . O tipo de suavização utilizado dependerá do contexto.

Para calcularmos $\nabla f_s(x)$ precisamos antes expressar $\nabla f_c(x)$ em função de $\nabla f(x)$:

$$\nabla f_c(x) = \phi'(f_b(x))\nabla f_b(x) = \phi'(f_b(x))(\nabla f(x) - \bar{g}).$$

Temos portanto:

$$\nabla f_s(x) = \nabla H(x) + \nabla f_c(x) = \bar{g} + \phi'(f_b(x))(\nabla f(x) - \bar{g}),$$

que pode ser reescrito como:

$$\nabla f_s(x) = [1 - \phi'(f_b(x))] \bar{g} + \phi'(f_b(x))\nabla f(x),$$

onde a combinação convexa aparece agora de forma mais explícita. Podemos também expressar a *função de ajuste* que leva $f(x)$ em $f_s(x)$, assim definida:

$$f_a(x) = f_s(x) - f(x) = H(x) + \phi(f(x) - H(x)) - f(x) = \phi(f(x) - H(x)) - (f(x) - H(x)).$$

Se agora definirmos $\sigma(t) = \phi(t) - t$, podemos reescrever a expressão anterior como:

$$f_a(x) = \sigma(f(x) - H(x)),$$

e portanto

$$f_s(x) = f(x) + f_a(x) = f(x) + \sigma(f(x) - H(x)).$$

A função de ajuste $f_a(x)$ é particularmente útil quando tivermos que lidar com uma função $f(x)$ que possa ser expressa sob a forma

$$f(x) = \sum_{i=1}^K f_i(x),$$

onde uma região de não diferenciabilidade distinta suavizável via SDA pode ser alocada a cada função componente $f_i(x)$. Nestas situações a função suavizada pode ser escrita como

$$f_s(x) = \sum_{i=1}^K f_{s,i}(x) = \sum_{i=1}^K (f_i(x) + f_{a,i}(x)) = f(x) + \sum_{i=1}^K f_{a,i}(x),$$

onde $f_{a,i}(x)$ representa a função de ajuste associada a cada função componente $f_i(x)$ e sua respectiva zona de não diferenciabilidade Z_{f_i} . Utilizando a função $\sigma(t)$, podemos reescrever esta expressão como

$$f_s(x) = f(x) + \sum_{i=1}^K \sigma(f_i(x) - H_i(x)).$$

Uma generalização da situação anterior pode ser realizada se considerarmos que $f(x)$ seja dada por

$$f(x) = \Psi(f_1(x), f_2(x), \dots, f_K(x)),$$

onde $\Psi(\xi_1, \xi_2, \dots, \xi_K)$ é uma função de classe C^1 em cada uma de suas variáveis e mais uma vez cada função componente $f_i(x)$ possui sua região de não diferenciabilidade distinta e suavizável via SDA. Notar que na situação analisada acima tínhamos $\Psi(\xi_1, \xi_2, \dots, \xi_K) = \xi_1 + \xi_2 + \dots + \xi_K$. O gradiente de $f(x)$ é dado por

$$\nabla f(x) = \frac{\partial \Psi}{\partial \xi_1} \nabla f_1(x) + \frac{\partial \Psi}{\partial \xi_2} \nabla f_2(x) + \dots + \frac{\partial \Psi}{\partial \xi_K} \nabla f_K(x),$$

e este gradiente não está definido para os pontos pertencentes à $\bigcup Z_{f_i}$. Se definirmos

$$f_s(x) = \Psi(f_{s,1}(x), f_{s,2}(x), \dots, f_{s,K}(x)),$$

temos que o seu gradiente

$$\nabla f_s(x) = \frac{\partial \Psi}{\partial \xi_1} \nabla f_{s,1}(x) + \frac{\partial \Psi}{\partial \xi_2} \nabla f_{s,2}(x) + \dots + \frac{\partial \Psi}{\partial \xi_K} \nabla f_{s,K}(x)$$

está obviamente bem definido para todo $x \in \mathbb{R}^n$. As propriedades de suavização SDA de cada $f_i(x)$ induzem as mesmas propriedades de suavização sobre a função f , dada a diferenciabilidade de Ψ , e assim toda a teoria anteriormente desenvolvida

se aplica naturalmente a esta situação mais geral. Este é um caso em que certamente teríamos o parâmetro τ como um vetor composto pelos τ_i utilizados na suavização de cada f_i , e deveríamos fazer $\tau \rightarrow \tau_0$, onde τ_0 também seria um vetor.

Até o momento não apresentamos nenhuma função $\phi(t)$ que atenda às propriedades listadas anteriormente. Por outro lado, nos problemas práticos torna-se necessário que a função $\phi(t)$ possua pelo menos um parâmetro de controle que permita aproximá-la da função identidade, de tal forma que possamos ter $f_s(x)$ se aproximando de $f(x)$.

3.7 Região de não diferenciabilidade $Z_f = \{z\}$

Visando a derivação de um relacionamento explícito entre o gradiente da função suavizada e os ε -subgradientes da função f em um ponto x , nesta tese trabalharemos de forma especial com uma simplificação da região de não diferenciabilidade, que será definida como $Z_f = \{z\}$, ou seja, constituída por apenas um ponto. Analisaremos este caso formalmente e em profundidade a partir do capítulo seguinte. Veremos mais a frente que esta análise bastante particular já irá implicar no desenvolvimento de uma série de conceitos e propriedades, que por sua vez poderão ser expandidos para as regiões Z_f mais gerais em uma continuação futura do tema desta tese.

Deve ficar claro que as funções convexas f com $\text{dom } f = \mathbb{R}^n$ para as quais $Z_f = \{z\}$ são automaticamente funções ZH convexas. Primeiramente temos que Z_f é fechado e tem medida de Lebesgue nula. Adicionalmente, se tomarmos um $g \in \partial f(z)$ arbitrário, a desigualdade fundamental dos subgradientes nos diz que para todo $x \in \mathbb{R}^n$ vale

$$f(x) \geq f(z) + g \cdot (x - z).$$

Se definirmos $H(x) = f(z) + g \cdot (x - z)$, percebemos que esta função afim representa um hiperplano e a desigualdade acima nos diz que este hiperplano é de suporte ao epígrafo de f . Por outro lado, $G(Z_f) = \{(z, f(z))\} = \{(z, H(z))\}$ e portanto $G(Z_f)$ está contido no hiperplano de suporte representado por $H(x)$.

Temos assim que f é uma função ZH convexa e por simplicidade iremos apenas mencionar que $Z_f = \{z\}$ nos capítulos seguintes, onde iremos desenvolver uma equivalência entre ε -subgradientes e subgradientes diferenciáveis.

Capítulo 4

Decomposição DHA da função $f(x)$

A decomposição DHA será apresentada neste capítulo e utilizada nos capítulos seguintes para o desenvolvimento de um tipo particular de suavização de funções convexas. Aqueles que já trabalharam com a suavização hiperbólica encontrarão na decomposição DHA uma generalização natural para o \mathbb{R}^n da etapa inicial de análise das assíntotas de uma função de domínio real $f(x)$ no ponto isolado de não diferenciabilidade z . A decomposição DHA apresenta, no entanto, uma fragilidade: necessita do atendimento de uma restrição adicional para ser utilizada na derivação do relacionamento explícito entre subgradientes diferenciáveis e ε -subgradientes.

A decomposição DA (Diferença-Afim), descrita na seção 3.4, foi criada posteriormente à decomposição DHA. Além de ser estruturalmente mais simples, a decomposição DA não necessita atender às restrições extras para ser utilizada. No entanto, o estudo da decomposição DHA é bastante enriquecedor, dadas as várias propriedades da componente homogênea desta decomposição, o que permite um aprofundamento na compreensão da suavização de funções. Além disto, uma das virtudes da decomposição DHA é justamente precipitar o desenvolvimento da decomposição DA.

A decomposição DHA (Diferença-Homogênea-Afim) tem por objetivo decompor a função $f(x)$ em funções mais simples, segregando a não diferenciabilidade e alocando-a em sua componente homogênea. Dentre as funções não diferenciáveis em z definidas no \mathbb{R}^n , uma função homogênea pode ser considerada como a mais simples possível. A suavização pode então ser realizada sobre esta componente homogênea.

4.1 Decomposição DHA - primeira etapa

O objetivo da primeira etapa da decomposição DHA (Diferença-Homogênea-Afim) é decompor a função convexa f na soma de duas funções, f_d e f_n , com as seguintes propriedades:

- f_d é diferenciável em \mathbb{R}^n , com $\nabla f_d(z) = 0$, e traz consigo a "complexidade

analítica" de f ;

- f_n é diferenciável em $\mathbb{R}^n \setminus \{z\}$, convexa e extremamente simples de ser analisada. Além disto mostraremos que $\partial f_n(z) = \partial f(z)$, ou seja, f_n traz consigo a não-diferenciabilidade de f .

Como a não-diferenciabilidade de f é transferida para f_n , o estudo da suavização de f será focado primeiramente em f_n (a segunda etapa da decomposição DHA irá alterar um pouco f_n como veremos mais a frente).

A primeira etapa da decomposição DHA é realizada da seguinte forma:

$$f(x) = f(x) - f(z) - f'(z, x - z) + f'(z, x - z) + f(z) = f_d(x) + f_n(x) + f(z), \quad (4.1)$$

onde $f_d(x) = f(x) - f(z) - f'(z, x - z)$ e $f_n(x) = f'(z, x - z)$.

Visando facilitar a compreensão destes conceitos apresentamos nas Figuras 4.1 e 4.2 um exemplo de função $f(x)$ com domínio em \mathbb{R}^1 , e sua decomposição na soma das funções $f_n(x)$ e $f_d(x)$.

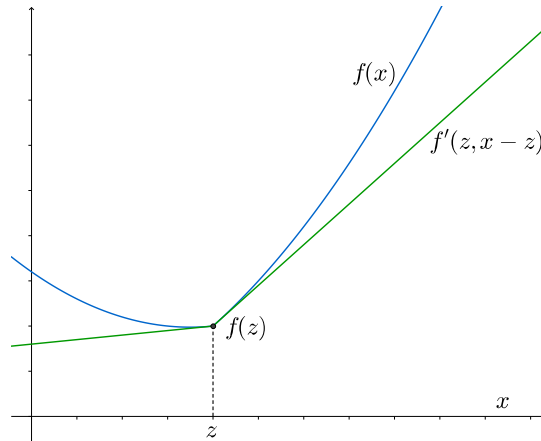


Figura 4.1: Exemplo de função $f(x)$ no \mathbb{R}^1 e sua derivada direcional $f'(z, x - z)$

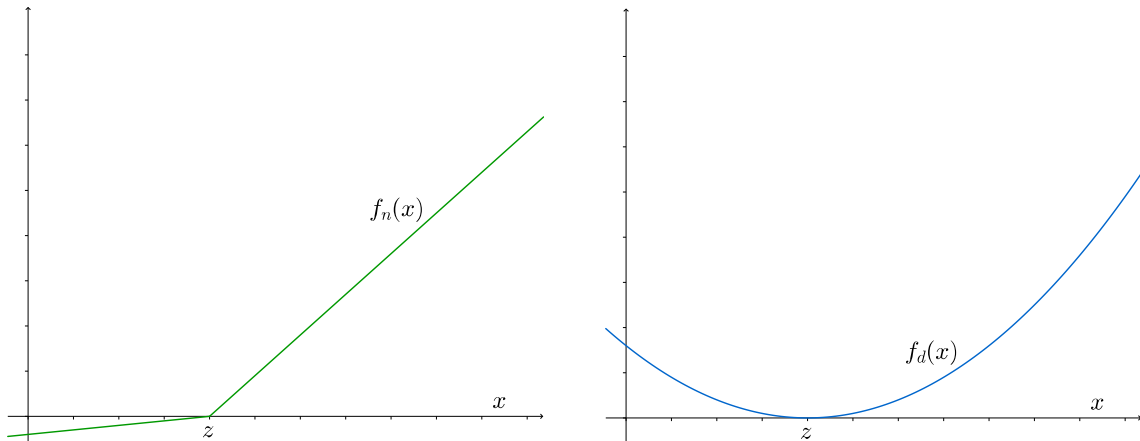


Figura 4.2: Exemplo de decomposição de $f(x)$ em $f_n(x)$ e $f_d(x)$

Iremos agora demonstrar as principais propriedades das funções f_d e f_n por intermédio de um conjunto de proposições, as quais fornecerão os fundamentos que justificam a adequabilidade das definições adotadas.

Proposição 4.1. *A função $f_d(x)$ é diferenciável no ponto $x = z$ e $\nabla f_d(z) = 0$.*

Demonstração. Como $f'(z, h)$ é homogênea em h (e portanto contínua em $h = 0$), temos que:

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{f_d(z+h) - f_d(z)}{\|h\|} = \\ &= \lim_{h \rightarrow 0} \frac{f(z+h) - f(z) - f'(z, z+h-z) - f(z) + f(z) + f'(z, z-z)}{\|h\|} = \\ &= \lim_{h \rightarrow 0} \frac{f(z+h) - f(z) - f'(z, h)}{\|h\|}. \end{aligned}$$

Se agora fizermos $h = tu$, com $u \in S^n$ representando a direção de h e $t = \|h\|$ representando o seu módulo, temos $h \rightarrow 0 \iff t \rightarrow 0+$ (mantendo a liberdade da direção u). Assim:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{f(z+h) - f(z) - f'(z, h)}{\|h\|} &= \lim_{t \rightarrow 0+} \frac{f(z+tu) - f(z) - f'(z, tu)}{t} = \\ &= \left(\lim_{t \rightarrow 0+} \frac{f(z+tu) - f(z)}{t} \right) - f'(z, u) = f'(z, u) - f'(z, u) = 0. \end{aligned}$$

Utilizando (2.13) temos portanto que $f_d(x)$ é diferenciável no ponto $x = z$ e também fica demonstrado que $\nabla f_d(x) = 0$. □

Proposição 4.2. *A função $f_d(x)$ é diferenciável em $x \in \mathbb{R}^n \setminus \{z\}$ e temos que $\nabla f_n(x) = \lim_{t \rightarrow 0} \nabla f(z + t(x-z))$.*

Demonstração. Como $f_d(x) = f(x) - f_n(x) - f(z)$ e $f(x)$ é diferenciável em todo $x \in \mathbb{R}^n \setminus \{z\}$, basta provarmos que $f_n(x) = f'(z, x-z)$ também é diferenciável em todo $x \in \mathbb{R}^n \setminus \{z\}$. Para tal, calcularemos a derivada direcional na direção arbitrária dada pelo vetor unitário $u \in S^n$ e mostraremos que ela é linear em u :

$$\begin{aligned} f'_n(x, u) &= \lim_{s \rightarrow 0+} \frac{f'(z, x+su-z) - f'(z, x-z)}{s} = \\ &= \lim_{s \rightarrow 0+} \frac{\lim_{t \rightarrow 0+} \frac{f(z+tx+tsu-tz) - f(z)}{t} - \lim_{t \rightarrow 0+} \frac{f(z+tx-tz) - f(z)}{t}}{s} = \\ &= \lim_{s \rightarrow 0+} \frac{\lim_{t \rightarrow 0+} \frac{f(z+tx+tsu-tz) - f(z) - f(z+tx-tz) + f(z)}{t}}{s} = \end{aligned}$$

$$\begin{aligned}
&= \lim_{s \rightarrow 0^+} \frac{\lim_{t \rightarrow 0^+} \frac{f(z+tx+tsu-tz) - f(z+tx-tz)}{t}}{s} = \\
&= \lim_{t \rightarrow 0^+} \frac{\lim_{s \rightarrow 0^+} \frac{f(z+tx+stu-tz) - f(z+tx-tz)}{s}}{t} = \lim_{t \rightarrow 0^+} \frac{\nabla f(z + t(x - z)).tu}{t} = \\
&= \lim_{t \rightarrow 0^+} \nabla f(z + t(x - z)).u = \left(\lim_{t \rightarrow 0^+} \nabla f(z + t(x - z)) \right).u = \nabla f_n(x).u,
\end{aligned}$$

onde várias propriedades básicas referentes à existência de limites foram utilizadas. Assim temos que $f'_n(x, u)$ é linear em u , e portanto $f_n(x)$ é diferenciável em todo $x \in \mathbb{R}^n \setminus \{z\}$. Este desenvolvimento adicionalmente demonstra que $\nabla f_n(x) = \lim_{t \rightarrow 0^+} \nabla f(z + t(x - z))$. \square

As Proposições 4.1 e 4.2 conjuntamente nos informam que $f_d(x)$ é diferenciável em todo o domínio \mathbb{R}^n .

Proposição 4.3. $f(x)$ convexa implica $f_n(x) = f'(z, x - z)$ convexa.

Demonstração. Sejam $x_1, x_2 \in \mathbb{R}^n \setminus \{z\}$, com $x_1 \neq x_2$ e $\lambda \in [0, 1]$. Temos então que:

$$\begin{aligned}
&f_n(\lambda x_1 + (1 - \lambda)x_2) = f'(z, \lambda x_1 + (1 - \lambda)x_2 - z) = \\
&= \lim_{t \rightarrow 0^+} \frac{f(z + t\lambda x_1 + t(1 - \lambda)x_2 - tz) - f(z)}{t} = \\
&= \lim_{t \rightarrow 0^+} \frac{f(\lambda z + t\lambda x_1 - t\lambda z + (1 - \lambda)z + t(1 - \lambda)x_2 - t(1 - \lambda)z) - f(z)}{t} = \\
&= \lim_{t \rightarrow 0^+} \frac{f(\lambda(z + tx_1 - tz) + (1 - \lambda)(z + tx_2 - tz)) - f(z)}{t} \leq \\
&\leq \lim_{t \rightarrow 0^+} \frac{\lambda f(z + t(x_1 - z)) + (1 - \lambda)f(z + t(x_2 - z)) - \lambda f(z) - (1 - \lambda)f(z)}{t} = \\
&= \lambda \lim_{t \rightarrow 0^+} \frac{f(z + t(x_1 - z)) - f(z)}{t} + (1 - \lambda) \lim_{t \rightarrow 0^+} \frac{f(z + t(x_2 - z)) - f(z)}{t} = \\
&= \lambda f'(z, x_1 - z) + (1 - \lambda)f'(z, x_2 - z) = \lambda f_n(x_1) + (1 - \lambda)f_n(x_2).
\end{aligned}$$

O que demonstra a convexidade de $f_n(x)$. \square

Proposição 4.4. As funções $f_d(x)$ e $f_n(x)$ assumem o valor zero no ponto z .

Demonstração. Temos que $f_n(z) = f'(z, z - z) = f'(z, 0) = 0$ e daí segue imediatamente que $f_d(z) = f(z) - f(z) + 0 = 0$. \square

Proposição 4.5. Para a função $f_d(x)$ vale que $f_d(x) \geq 0, \forall x \in \mathbb{R}^n$.

Demonstração. Utilizando a desigualdade (2.15) temos

$$f(x) \geq f(z) + f'(z, x - z),$$

ou,

$$f_d(x) = f(x) - f(z) - f'(z, x - z) \geq 0, \forall x \in \mathbb{R}^n.$$

□

Proposição 4.6. *Os subdiferenciais de $f_n(x)$ e $f(x)$ são iguais no ponto z , ou seja, $\partial f_n(z) = \partial f(z)$.*

Demonstração. O subdiferencial $\partial f_n(z)$ existe porque a proposição 4.3 nos afirma que f_n é uma função convexa. Iremos então provar as duas inclusões.

a) $\partial f_n(z) \subset \partial f(z)$: Seja $g \in \partial f_n(z)$. Temos então que para este g vale:

$$f_n(x) \geq f_n(z) + g.(x - z).$$

Somando $f_d(x) + f(z)$ a ambos os lados da inequação obtemos

$$f(x) = f_d(x) + f_n(x) + f(z) \geq f_d(x) + f_n(z) + f(z) + g.(x - z).$$

Utilizando-se as proposições 4.4 e 4.5, temos:

$$f(x) \geq f_d(x) + f(z) + g.(x - z) \geq f(z) + g.(x - z).$$

Ou seja, $g \in \partial f(z)$. Consequentemente $\partial f_n(z) \subset \partial f(z)$.

b) $\partial f_n(z) \supset \partial f(z)$: Como $\partial f(z)$ é um conjunto convexo e compacto, basta provarmos que todos os seus pontos de fronteira pertencem a outro conjunto convexo para provarmos sua inclusão neste outro conjunto. Lembremos que a notação $\partial\{\}$ simboliza a fronteira de um conjunto. Consideremos então $g \in \partial\{\partial f(z)\}$. Iremos supor por absurdo que $g \notin \partial f_n(z)$ e chegaremos a uma contradição. Iremos trabalhar com um $y = z + tg$, para um certo $t > 0$ a ser determinado para atender aos nossos propósitos. Temos então que $f_n(y) = f'(z, y - z) = tf'(z, g)$ e $g.(y - z) = g.tg = t\|g\|^2$. Sabemos que tanto $f'(z, y - z)$ quanto $g.(y - z)$ são homogêneas em $y - z$ (ou lineares em t) e como $g \notin \partial f_n(z)$, é possível negar a desigualdade dos subgradientes de $f_n(x)$, ou seja, para algum $y \neq z$ (ou $t > 0$) temos:

$$f'(z, y - z) = f_n(y) < f_n(z) + g.(y - z) = g.(y - z).$$

Temos então que $g.(y - z) - f'(z, y - z) = t(\|g\|^2 - f'(z, g)) > 0$ e como $t > 0$ temos que $\|g\|^2 - f'(z, g) > 0$. Assim, dado qualquer $\varepsilon > 0$ podemos fazer

$t(\|g\|^2 - f'(z, g)) = \varepsilon$ simplesmente tomando $t = \frac{\varepsilon}{\|g\|^2 - f'(z, g)}$. Como $\nabla f_d(z) = 0$, temos então que $f'_d(z, g) = \nabla f_d(z) \cdot g = 0$. Em outras palavras:

$$\lim_{t \rightarrow 0^+} \frac{f_d(z + tg) - f_d(z)}{t} = \lim_{t \rightarrow 0^+} \frac{f_d(z + tg)}{t} = 0,$$

onde utilizamos $f_d(z) = 0$. Este limite pode ser reescrito, dividindo-se por uma constante:

$$\lim_{t \rightarrow 0^+} \frac{f_d(z + tg)}{t(\|g\|^2 - f'(z, g))} = 0.$$

Como este limite é zero, existe um $\hat{t} > 0$, com um respectivo $\hat{y} = z + \hat{t}g$, tal que:

$$f_d(\hat{y}) = f_d(z + \hat{t}g) < \hat{t}(\|g\|^2 - f'(z, g)) = g \cdot (\hat{y} - z) - f'(z, \hat{y} - z),$$

ou,

$$f(\hat{y}) = f_d(\hat{y}) + f'(z, \hat{y} - z) + f(z) < f(z) + g \cdot (\hat{y} - z).$$

Mas este \hat{y} contradiz a desigualdade dos subgradientes de $\partial f(z)$. Logo concluímos que $g \in \partial\{\partial f(z)\} \Rightarrow g \in \partial f_n(z)$ e portanto temos que $\partial f(z) \subset \partial f_n(z)$.

Agrupando as duas inclusões provadas em (a) e (b) temos então finalmente que $\partial f_n(z) = \partial f(z)$. \square

4.2 Decomposição DHA - segunda etapa

Como acabamos de provar, $\partial f_n(z) = \partial f(z)$. Isto indica que a suavização de $f(x)$ pode ser realizada diretamente em $f_n(x)$, pois toda a informação de não diferenciabilidade do subgradiente de $\partial f(z)$ está presente em $\partial f_n(x)$. No entanto, para que seja possível o pleno desenvolvimento do método de suavização sobre a decomposição DHA que será apresentado mais adiante, duas deficiências básicas de $f_n(x)$ devem ser contornadas:

- $f_n(x)$ não está centrada na origem;
- $f_n(x)$ pode assumir valores negativos.

Ao eliminar estas deficiências passaremos a tratar com uma função positivamente homogênea que não assume valores negativos. Para tal, lembremos que $f'(x, h) =$

$\max_{g \in \partial f(x)} g \cdot h$ e portanto, se tomarmos um $\bar{g} \in ri\{\partial f_n(z)\}$ arbitrário¹ teremos:

$$\begin{aligned} f_n(x) - \bar{g} \cdot (x - z) &= f'(z, x - z) - \bar{g} \cdot (x - z) = \\ &= \max_{g \in \partial f(z)} g \cdot (x - z) - \bar{g} \cdot (x - z) \geq \bar{g} \cdot (x - z) - \bar{g} \cdot (x - z) = 0. \end{aligned} \quad (4.2)$$

Definiremos então $f_p(h) = f'(z, h) - \bar{g} \cdot h$, para $h \in \mathbb{R}^n$, e a inequação anterior demonstra que $f_p(h) \geq 0$. Devemos notar que, pela definição de $f_p(x)$, $f_n(x)$ convexa implica trivialmente $f_p(h)$ também convexa. Temos então a segunda etapa da decomposição DHA, definida por:

$$f_n(x) = (f'(z, x - z) - \bar{g} \cdot (x - z)) + \bar{g} \cdot (x - z) = f_p(x - z) + \bar{g} \cdot (x - z). \quad (4.3)$$

Podemos então descrever a decomposição DHA da função $f(x)$ em sua forma completa como:

$$f(x) = f_d(x) + f_n(x) + f(z) = f_d(x) + f_p(x - z) + f(z) + \bar{g} \cdot (x - z). \quad (4.4)$$

Para simplificar a visualização desta decomposição, iremos definir

$$H(x) = f(z) + \bar{g} \cdot (x - z),$$

que pode ser interpretado como um hiperplano de suporte ao epígrafo de $f(x)$, passando pelo ponto $(z, f(z))$ no espaço \mathbb{R}^{n+1} . Assim, a decomposição DHA pode ser escrita de forma compacta como:

$$f(x) = f_d(x) + f_p(x - z) + H(x).$$

Nesta decomposição iremos denominar as funções f_d como a *componente diferença*, f_p como a *componente homogênea*, e H como a *componente afim*, de uma forma semelhante à denominação utilizada para a decomposição DA. Notar que a componente diferença f_d da decomposição DHA é diferente da componente diferença f_b da decomposição DA. Para um mesmo $H(x)$ aplicado nas duas decomposições temos

$$f_d(x) + f_p(x - z) = f(x) - H(x) = f_b(x),$$

o que demonstra uma relativa proximidade entre as decomposições DHA e DA.

Continuando com o exemplo de uma função $f(x)$ com domínio em \mathbb{R}^1 , a Figura 4.3 apresenta um g arbitrário que, por pertencer claramente a $ri\{\partial f(z)\}$, poderia ser utilizado como o \bar{g} que compõe o $H(x)$ utilizado na decomposição DHA.

¹No Apêndice A.1 é apresentada a necessidade de se tratar com o interior relativo do subdiferencial, $ri\{\partial f(x)\}$, para uma função convexa f arbitrária.

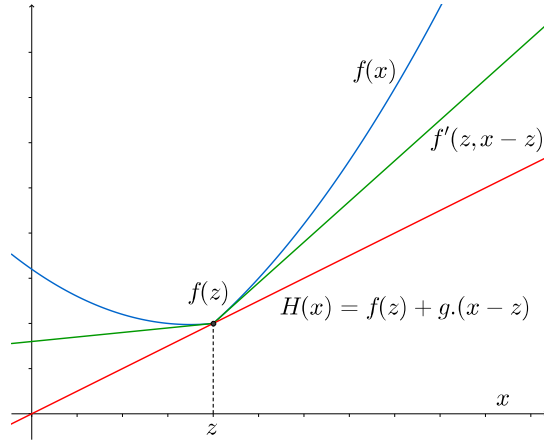


Figura 4.3: $H(x)$ composto por um g arbitrário que pertence a $ri\{\partial f(z)\}$

Com base na figura anterior, a Figura 4.4 apresenta $f_p(x - z) = f'(z, x - z) - \bar{g} \cdot (x - z)$.

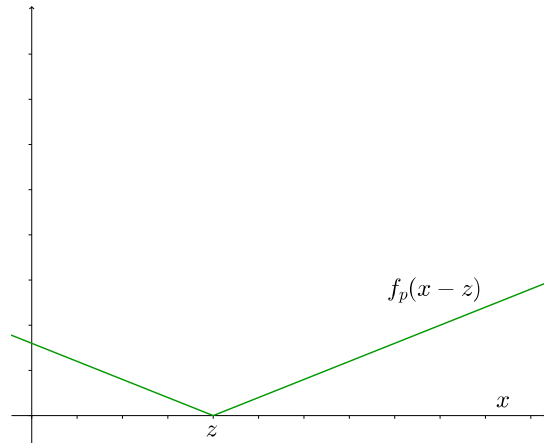


Figura 4.4: $f_p(x - z)$ com base na $f(x)$ da figura anterior

Apesar de arbitrário no âmbito deste desenvolvimento teórico, o valor mais adequado para \bar{g} deve ser escolhido com base nas características específicas de cada problema. Por exemplo, em um problema de minimização de $f(x)$ em que $0 \in \partial f(z)$, o método de escolha deveria apontar para $\bar{g} = 0$. Na seção 4.5 comentaremos sobre a computação de uma escolha adequada para \bar{g} .

4.2.1 Propriedades da componente homogênea $f_p(h)$

A principal propriedade de $f_p(h)$ é justamente o fato de ela ser positivamente homogênea, pois para $\alpha > 0$ e $h \in \mathbb{R}^n$ temos:

$$f_p(\alpha h) = f'(z, \alpha h) - \bar{g} \cdot \alpha h = \alpha(f'(z, h) - \bar{g} \cdot h) = \alpha f_p(h). \quad (4.5)$$

Em especial, temos:

$$f_p(0) = f'(z, 0) - \bar{g}.0 = 0. \quad (4.6)$$

Por outro lado, seja $\dim\{\partial f(z)\} = k \leq n$. Definamos $V(\partial f(z), \bar{g}) = \text{aff}\{\partial f(z)\} - \bar{g}$, ou seja, a envoltória afim de $\partial f(z)$ menos o elemento $\bar{g} \in \partial f(z)$, que por definição resulta em um espaço vetorial, e seja $V^\perp(\partial f(z), \bar{g})$ o seu complementar ortogonal.

Dado $h \in V(\partial f(z), \bar{g})$ arbitrário, podemos escolher $\hat{g} \in \partial f(z)$ tal que $\hat{g} - \bar{g} = \alpha h$, para um determinado $\alpha > 0$, pois $\bar{g} \in \text{ri}\{\partial f_n(z)\}$ e desta forma temos:

$$f_p(h) = f'(z, h) - \bar{g}.h = \max_{g \in \partial f(z)} g.h - \bar{g}.h \geq \hat{g}.h - \bar{g}.h = (\hat{g} - \bar{g}).h = \alpha h.h > 0. \quad (4.7)$$

Consideremos agora $h \in V^\perp(\partial f(z), \bar{g})$. Obviamente $g.h = 0, \forall g \in \partial f(z)$. Temos então:

$$f_p(h) = f'(z, h) - \bar{g}.h = \max_{g \in \partial f(z)} g.h - \bar{g}.h = 0 - 0 = 0. \quad (4.8)$$

Ou de forma resumida, $f_p(h) > 0$ para $h \in V(\partial f(z), \bar{g})$ e $f_p(h) = 0$ para $h \in V^\perp(\partial f(z), \bar{g})$.

Como será visto mais adiante, para a aplicação dos métodos de suavização será necessário calcular $\partial f_p(0)$ e com isso o cálculo de $f'_p(0, h)$ se torna necessário:

$$f'_p(0, h) = \lim_{t \rightarrow 0^+} \frac{f_p(0 + th) - f_p(0)}{t} = \lim_{t \rightarrow 0^+} \frac{t f_p(h)}{t} = f_p(h). \quad (4.9)$$

Para $u \in S^n$ unitário, vale obviamente que $f'_p(0, u) = f_p(u) = f_p(u)||u||$.

Uma identidade que nos será útil relaciona $\nabla f_p(h)$ com $f_p(h)$:

$$f'_p(h, h) = \lim_{t \rightarrow 0^+} \frac{f_p(h + th) - f_p(h)}{t} = \lim_{t \rightarrow 0^+} \frac{t f_p(h)}{t} = f_p(h). \quad (4.10)$$

Mas como $f_p(h) = f'(z, h) - \bar{g}.h$ é diferenciável para $h \neq 0$ (pois a proposição 4.2 nos afirma que $f'(z, h) = f'_n(h + z)$ é diferenciável para $h \neq 0$), temos que $f'_p(h, h) = \nabla f_p(h).h$. Logo:

$$\nabla f_p(h).h = f_p(h). \quad (4.11)$$

As expressões dos dois lados desta igualdade são homogêneas e portanto valem também as seguintes igualdades:

$$\nabla f_p(h).u_h = f_p(u_h). \quad (4.12)$$

$$\nabla f_p(kh) = \nabla f_p(h) = \nabla f_p(u_h), \forall k > 0. \quad (4.13)$$

4.2.2 Decomposição adicional da componente homogênea

$$f_p(h)$$

Por ser homogênea, $f_p(h)$ pode sofrer uma decomposição adicional. Tomando-se $h \neq 0$:

$$f_p(h) = f'(z, h) - \bar{g} \cdot h = \left(f'(z, \frac{h}{\|h\|}) - \bar{g} \cdot \frac{h}{\|h\|} \right) \|h\| = f_p \left(\frac{h}{\|h\|} \right) \|h\| = f_p(u_h) \|h\|, \quad (4.14)$$

onde $u_h = h/\|h\| \in S^n$. Desta forma, $f(x)$ pode ser expressa como:

$$f(x) = f_d(x) + f_p(u_{x-z})\|x - z\| + H(x). \quad (4.15)$$

Esta decomposição é particularmente interessante quando $f_p(u_h)$ é constante. A função não-diferenciável em z dada por $f(x) = k\|x - z\|$, com $k > 0$, é um exemplo trivial, para o qual $f_d(x) \equiv 0$, $\bar{g} \equiv 0$ e $f_p(u_h) \equiv k$. O caso unidimensional $f(x) = |x|$ é um exemplo ainda mais trivial.

De uma forma geral $f_p(h) = f_p(u_h)\|h\|$ nos diz que $\|h\|$ é a componente que expressa a escala radial de $f_p(h)$, enquanto que $f_p(u_h)$ é a componente que expressa a escala global e as mudanças transversais de $f_p(h)$. Desta forma, $f_p(h)$ está completamente definida se definirmos $f_p(u)$ para todo $u \in S^n$.

4.3 Resumo da decomposição DHA

A decomposição DHA nos permite expressar $f(x) = f_d(x) + f_p(x - z) + H(x)$. Já vimos que $f_d(x)$ é diferenciável no espaço \mathbb{R}^n e $H(x)$ também é trivialmente diferenciável no \mathbb{R}^n . Portanto nossa atenção com relação à suavização de $f(x)$ deve se concentrar na suavização da componente homogênea $f_p(h)$, o que dará origem à decomposição SDHA. Este será nosso foco no próximo capítulo.

4.4 Similaridade com a série de Taylor

Um interessante paralelo da decomposição DHA com a série de Taylor pode ser traçado se a descrevermos da seguinte forma:

$$f(x) = f(z) + \bar{g} \cdot (x - z) + f_p(x - z) + f_d(x).$$

Se definirmos $f_r(x - z) = f_d(x)$, esta expressão pode ser reescrita como:

$$f(x) = f(z) + \bar{g} \cdot (x - z) + f_p(x - z) + f_r(x - z). \quad (4.16)$$

Nesta decomposição percebemos que a componente afim² $H(x) = f(z) + \bar{g} \cdot (x - z)$, juntamente com a componente $f_p(x - z)$, respondem pelas variações de primeira ordem de $f(x)$, acompanhadas por um "resto" dado por $f_r(x - z)$. Em uma expansão de uma função analítica não teríamos o termo não-diferenciável $f_p(x - z)$ e \bar{g} seria substituído por $\nabla f(z)$. No entanto, como demonstramos nos desenvolvimentos anteriores, a expansão 4.16 sempre será válida quando $f(x)$ for convexa e de classe C^1 para $x \in \mathbb{R}^n \setminus \{z\}$.

Adicionalmente, se $f_r(x - z)$ puder ser expresso como uma série de Taylor $T_r(x - z)$ dentro de uma certa vizinhança $V(z)$ do ponto z , então $f(x)$ pode sofrer uma decomposição completa em uma série expressa por:

$$f(x) = f(z) + \bar{g} \cdot (x - z) + f_p(x - z) + T_r(x - z),$$

válida para qualquer $x \in V(z)$.

4.5 Subgradiente médio

Um subgradiente adequado para ser utilizado na decomposição DHA é o que chamaremos de "subgradiente médio". O cálculo do subgradiente médio no ponto z , que também denotaremos por \bar{g} , será realizado considerando-se uma distribuição uniforme dos subgradientes dentro do subdiferencial $\partial f(z)$, ou seja, uma ponderação de mesmo peso volumétrico dv para cada subgradiente $g \in \partial f(z)$, dividido pelo volume $Vol(\partial f(z))$ do subdiferencial:

$$\bar{g} = \frac{1}{Vol(\partial f(z))} \oint_{\partial f(z)} g dv. \quad (4.17)$$

Como $\partial f(z)$ é compacto, a integral utilizada pode ser a de Riemann. Notar que esta definição de \bar{g} corresponde à computação do centróide de $\partial f(z)$. A computação efetiva do centróide de um conjunto dentro de um certo grau de aproximação pode ser realizada por vários métodos, dentre os quais os métodos randômico (Monte Carlo) e quasi-randômico (quasi-Monte Carlo) são considerados práticos e eficientes: pontos aleatórios são gerados dentro do conjunto e sua média é calculada. Na prática a configuração geométrica de $\partial f(z)$ irá determinar o nível de dificuldade associado ao cálculo do centróide.

²A componente afim em geral não é única para $Z_f = \{z\}$, pois existe alguma liberdade na escolha de \bar{g} .

Obviamente, se z for um ponto diferenciável de $f(x)$ então definimos $\bar{g} = \nabla f(z)$, para mantermos a consistência.

Capítulo 5

DHA: Suavização da componente homogênea $f_p(h)$

Como já demonstrado no capítulo anterior, a componente homogênea $f_p(h)$ é uma função extremamente simples, apesar de não diferenciável na origem. Iremos portanto nos concentrar na suavização de $f_p(h)$, que é a única componente não diferenciável da decomposição DHA.

5.1 Família de suavizações de $f_p(h)$

Iremos agora definir uma família de suavizações de $f_p(h)$, parametrizada por τ , que representará a noção intuitiva de suavização de $f_p(h)$ e estabeleceremos condições para que uma função $f_q(\tau, h)$ pertença a esta família. Já vimos anteriormente que $f_p(h) = f_p(u_h)||h||$, que $f_p(0) = 0$, que $f_p(h) > 0$ para $h \in V(\partial f(z), \bar{g})$ e $f_p(h) = 0$ para $h \in V^\perp(\partial f(z), \bar{g})$, e ainda que $f_p(h)$ é diferenciável para $h \neq 0$. Ou seja, para $h \in V(\partial f(z), \bar{g}) \setminus \{0\}$, $f_p(h)$ tem valores positivos que aumentam em proporção direta com $||h||$ e esta proporcionalidade é justamente a razão da não-diferenciabilidade de $f_p(h)$ na origem, apesar de não ser um problema para a diferenciabilidade de $f_p(h)$ para $h \neq 0$. Nosso objetivo será portanto suavizar esta proporcionalidade nas vizinhanças da origem, mediante a composição de $f_p(h)$ com uma função suavizadora, ou, em outras palavras, substituir $f_p(h)$ por uma nova função $f_q(\tau, h) = \phi(\tau, f_p(h))$, definida com base em uma função $\phi(\tau, t) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ convenientemente escolhida. Notar que este procedimento de suavização é basicamente o mesmo desenvolvido em um âmbito mais geral na seção 3.6, que tratava da suavização SDA.

Veremos agora as condições que devem ser satisfeitas por $\phi(\tau, t)$. Em primeiro lugar $\phi(\tau, t)$ deve ser obviamente de classe C^1 com relação à variável t , uma vez que desejamos suavizar $f_p(h)$. Por outro lado, devemos encontrar condições para que $f_q(\tau, h)$, composição de uma função $\phi(\tau, t)$ diferenciável com uma função $f_p(h)$ não-

diferenciável na origem, possa ser efetivamente diferenciável na origem. Avaliando seu gradiente para $h \neq 0$:

$$\nabla f_q(\tau, h) = \phi'(\tau, f_p(h)) \nabla f_p(h). \quad (5.1)$$

Em geral, para cada direção definida por h teremos um $\nabla f_p(h)$ diferente e a única forma de *compatibilizá-los* de tal forma que $\nabla f_q(\tau, h)$ possa ser definido na origem é definindo $\phi'(\tau, 0) = 0$. Desta forma, para qualquer $h \neq 0$:

$$\lim_{h \rightarrow 0} \phi'(\tau, f_p(h)) \nabla f_p(h) = \lim_{t \rightarrow 0} \phi'(\tau, t f_p(u_h)) \nabla f_p(u_h) = 0, \quad (5.2)$$

onde utilizamos (4.13) e a homogeneidade de $f_p(h)$. Conseqüentemente, se definirmos $\nabla f_q(\tau, 0) = 0$ (uma vez que $\nabla f_p(h)$ não está definido para $h = 0$), iremos assegurar a continuidade de $\nabla f_q(\tau, 0)$ e assim criando as condições necessárias para que $f_q(\tau, h)$ também pertença à classe C^1 .

Sendo a origem o ponto de convergência de todas as direções de $h \in \mathbb{R}^n$, temos que estabelecer um valor para $\phi(\tau, t)$ na origem e consideraremos que este será o valor do próprio parâmetro τ , ou seja, $\phi(\tau, 0) = \tau$.

Para valores grandes de t devemos fazer com que $\phi(\tau, t)$ seja aproximadamente igual à função identidade, pois $f_p(h)$ já é diferenciável para $h \neq 0$. Em outras palavras, $\phi(\tau, t) - t$ deve tender para zero quando t tende para o infinito. Esta é na verdade uma condição mais forte do que simplesmente impor que o limite de $\phi'(\tau, t)$ quando t tende para o infinito seja igual a 1, o que parece ser o suficiente para engendrar o atendimento ao terceiro requisito da definição 3.1. Veremos porém que este requisito alternativo deriva automaticamente do requisito anterior, o qual irá adicionalmente assegurar a aproximação entre os valores de $f_p(h)$ e sua suavização $\phi(\tau, f_p(h))$ quanto t tende para o infinito, e não somente entre seus respectivos gradientes.

Finalmente, $f_q(\tau, h)$ deve aproximar-se cada vez mais de $f_p(h)$ quando o parâmetro τ tende para zero.

Com base nos requisitos apresentados acima, suponhamos que para um τ arbitrário exista uma função:

$$\begin{aligned} \phi(\tau, \cdot) : \mathbb{R}_+ &\rightarrow \mathbb{R}_+ \\ t &\rightarrow \phi(\tau, t), \end{aligned}$$

de classe C^1 , com as seguintes propriedades:

$$\phi(\tau, 0) = \tau, \quad (5.3)$$

$$\phi'(\tau, 0) = 0, \quad (5.4)$$

$$\lim_{t \rightarrow +\infty} \phi(\tau, t) - t = 0, \quad (5.5)$$

$$\lim_{\tau \rightarrow 0} \phi(\tau, t) = t, \forall t \in \mathbb{R}_+. \quad (5.6)$$

A Figura 5.1 apresenta alguns exemplos de funções $\phi(\tau, t)$ que atendem estas propriedades.

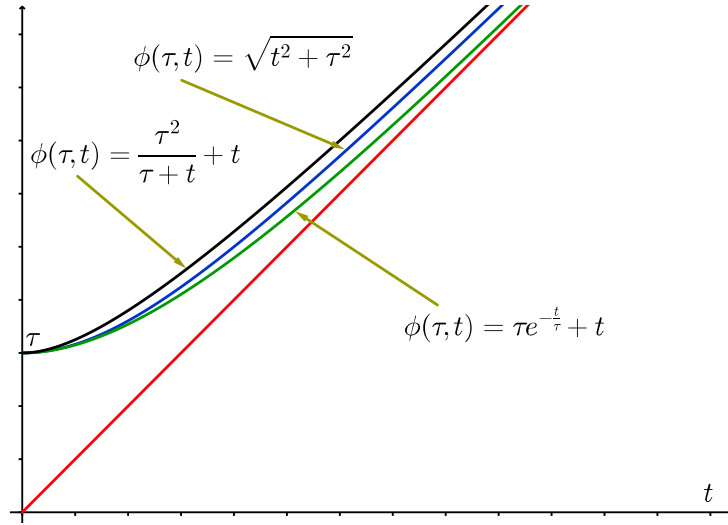


Figura 5.1: Exemplos de funções $\phi(\tau, t)$ admissíveis

Para simplificar, em algumas deduções o parâmetro τ da função $\phi(\tau, t)$ será considerado implícito e escreveremos simplesmente $\phi(t)$.

Da propriedade (5.5) podemos deduzir as seguintes propriedades adicionais, sendo que na segunda utilizamos L'Hôpital:

$$\lim_{t \rightarrow \infty} \frac{\phi(t)}{t} = \lim_{t \rightarrow \infty} \frac{\phi(t) - t + t}{t} = \lim_{t \rightarrow \infty} \frac{\phi(t) - t}{t} + 1 = 1, \quad (5.7)$$

$$\lim_{t \rightarrow \infty} \frac{\phi(t)}{t} = \lim_{t \rightarrow \infty} \phi'(t) = 1. \quad (5.8)$$

Como já mencionado anteriormente, definamos formalmente a seguinte função:

$$f_q(\tau, h) = \phi(\tau, f_p(h)). \quad (5.9)$$

Para demonstrarmos que $f_q(\tau, h)$ é uma suavização de $f_p(h)$, iremos demonstrar o atendimento das condições listadas na definição 3.1. Para nossa definição τ é um parâmetro real e $\tau_0 = 0$.

Proposição 5.1. A função $f_q(\tau, h)$ é de classe C^1 , para todo $\tau \neq 0$.

Demonstração. Seja $\tau \neq 0$. Como $\phi(\tau, t)$ é de classe C^1 , temos para $h \neq 0$,

$$\nabla f_q(\tau, h) = \nabla \phi(\tau, f_p(h)) = \phi'(\tau, f_p(h)) \nabla f_p(h), \quad (5.10)$$

que está bem definido pois $\nabla f_p(h)$ existe para todo $h \neq 0$.

A diferenciabilidade na origem (onde o gradiente deve ser nulo) pode ser provada por:

$$\lim_{h \rightarrow 0} \frac{f_q(\tau, h) - f_q(\tau, 0)}{\|h\|} = \lim_{h \rightarrow 0} \frac{\phi(\tau, f_p(h)) - \phi(\tau, f_p(0))}{\|h\|} = \lim_{h \rightarrow 0} \frac{\phi(\tau, f_p(h)) - \phi(\tau, 0)}{\|h\|}.$$

Fazendo agora $h = tu_h$, com $t = \|h\|$, temos $h \rightarrow 0 \iff t \rightarrow 0$. Assim, utilizando (5.4):

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\phi(\tau, f_p(h)) - \phi(\tau, 0)}{\|h\|} &= \lim_{t \rightarrow 0} \frac{\phi(\tau, f_p(tu_h)) - \phi(\tau, 0)}{t} = \\ &= f_p(u_h) \lim_{t \rightarrow 0} \frac{\phi(\tau, t f_p(u_h)) - \phi(\tau, 0)}{t f_p(u_h)} = f_p(u_h) \phi'(\tau, 0) = 0. \end{aligned}$$

Ou seja, utilizando (2.13) temos que $f_q(\tau, h)$ é diferenciável na origem e:

$$\nabla f_q(\tau, 0) = 0. \quad (5.11)$$

□

Proposição 5.2. A função $f_q(\tau, h)$ converge para $f_p(h)$ quando $\tau \rightarrow 0$.

Demonstração. Seja $h \in \mathbb{R}^n$ arbitrário. Utilizando (5.6) temos:

$$\lim_{\tau \rightarrow 0} f_q(\tau, h) = \lim_{\tau \rightarrow 0} \phi(\tau, f_p(h)) = f_p(h),$$

ou seja,

$$\lim_{\tau \rightarrow 0} f_q(\tau, h) = f_p(h), \quad \forall h \in \mathbb{R}^n, \quad (5.12)$$

e, portanto, a função $f_q(\tau, h)$ é uma aproximação da função $f_p(h)$, que tende para $f_p(h)$ quando $\tau \rightarrow 0$. □

Proposição 5.3. $\nabla f_q(\tau, h) \rightarrow \nabla f_p(h)$ quando $d(h, 0) \rightarrow \infty$.

Demonstração. Com base na definição de distância à região de não diferenciabilidade já apresentada na seção 3.4, temos $d(h, 0) = f_p(h)$. Fazamos $h = tu_h$, sendo $t = \|h\| > 0$ e u_h o vetor unitário na direção de h . Como $f_p(h) = tf_p(u_h)$, fazer $f_p(h) \rightarrow \infty$ equivale a fazer $t \rightarrow \infty$ para quaisquer direções u_h nas quais $f_p(u_h) > 0$, ou seja, para as direções em que nos afastamos da região de não diferenciabilidade. Também sabemos por (4.13) que $\nabla f_p(h) = \nabla f_p(tu_h) = \nabla f_p(u_h)$, e portanto:

$$\lim_{f_p(h) \rightarrow \infty} \nabla f_q(\tau, h) = \lim_{t \rightarrow \infty} \phi'(\tau, f_p(tu_h)) \nabla f_p(tu_h) = \lim_{t \rightarrow \infty} \phi'(\tau, tf_p(u_h)) \nabla f_p(h) = \nabla f_p(h),$$

onde utilizamos (5.8) na última igualdade. \square

As proposições acima nos garantem então que $f_q(\tau, h) = \phi(\tau, f_p(h))$ atende integralmente a definição 3.1 e é portanto uma suavização paramétrica de $f_p(h)$.

Considerando-se disponível a função $\phi(\tau, t)$, podemos definir a função $f_s(\tau, x)$, suavização da decomposição DHA de $f(x)$ e parametrizada por τ , como:

$$f_s(\tau, x) = f_d(x) + f_q(\tau, x - z) + H(x). \quad (5.13)$$

Esta suavização será doravante denominada SDHA. Mais uma vez enfatizamos que f_s é uma denominação geral para a suavização da função f . O tipo de suavização aplicado dependerá do contexto.

Como já provamos que $\lim_{\tau \rightarrow 0} f_q(\tau, h) = f_p(h)$, $\forall h \in \mathbb{R}^n$, torna-se imediato que:

$$\lim_{\tau \rightarrow 0} f_s(\tau, x) = f_d(x) + f_p(x - z) + H(x) = f(x), \forall x \in \mathbb{R}^n. \quad (5.14)$$

Com relação aos gradientes, temos, para $x \neq z$ e utilizando (5.10):

$$\nabla f(x) = \nabla f_d(x) + \nabla f_p(x - z) + \bar{g}, \quad (5.15)$$

$$\nabla f_s(\tau, x) = \nabla f_d(x) + \phi'(\tau, f_p(x - z)) \nabla f_p(x - z) + \bar{g}. \quad (5.16)$$

$\nabla f_s(\tau, x)$ será daqui para a frente chamado de *subgradiente diferenciável* de $f(x)$ no ponto x , com parâmetro τ , neste caso com relação à suavização SDHA.

5.2 Escolha da função ϕ

As propriedades definidoras de $\phi(\tau, t)$ enumeradas na seção anterior não implicam em unicidade, pois uma grande variedade de funções completamente distintas as atendem. Nossa intenção nesta seção é encontrar uma função *suficientemente simples* que possa ser nossa $\phi(\tau, t)$. Para tal, iremos considerar um caso especial, onde acrescentaremos as restrições $\phi(t) \geq t$, $\forall t \in \mathbb{R}_+$, e também que $\phi'(t)$ é monótona não

decrecente, o que por (5.8) implica em $\phi'(t) \leq 1$. Notar que esta condição, juntamente com (5.4), implicam em termos $\phi'(t)$ como o coeficiente de uma combinação convexa, conforme discutido em um contexto mais geral na seção 3.6. Agrupando estas desigualdades temos:

$$\phi'(t) \leq 1 \leq \frac{\phi(t)}{t}.$$

Já sabemos por (5.8) que $\phi'(t)$ e $\frac{\phi(t)}{t}$ tendem para 1 quando $t \rightarrow \infty$. Quando $t \rightarrow 0$ temos pelas propriedades (5.3) e (5.4) que $\phi'(t)$ tende para 0 e $\frac{\phi(t)}{t}$ tende para ∞ . Isto nos sugere adotarmos o número 1 como a média geométrica entre $\phi'(t)$ e $\frac{\phi(t)}{t}$, na esperança de que esta condição adicional nos permitirá definir uma função $\phi(t)$ adequada. Temos portanto:

$$\phi'(t) \frac{\phi(t)}{t} = 1 \implies \phi'(t)\phi(t) - t = 0. \quad (5.17)$$

Resolvendo-se esta equação diferencial:

$$\begin{aligned} \phi'(t)\phi(t) - t &= 0, \\ 2\phi'(t)\phi(t) &= 2t, \\ (\phi(t)^2)' &= (t^2)', \\ \phi(t)^2 &= t^2 + K, \\ \phi(t) &= \sqrt{K + t^2}. \end{aligned}$$

Por (5.3) devemos ter $\phi(0) = \tau$, o que resulta em $K = \tau^2$ e chegamos então à expressão final

$$\phi(\tau, t) = \sqrt{\tau^2 + t^2}, \quad (5.18)$$

definida no domínio \mathbb{R}_+ . Esta função é particularmente simples e conseqüentemente de fácil aplicação. Verifica-se que ela é de classe C^1 (na verdade de classe C^∞), e as condições (5.3), (5.4), (5.5) e (5.6) são trivialmente satisfeitas. Abaixo apresentamos suas derivadas primeira e segunda:

$$\phi'(\tau, t) = \frac{t}{\sqrt{\tau^2 + t^2}}, \quad (5.19)$$

$$\phi''(\tau, t) = \frac{\tau^2}{(\tau^2 + t^2)^{3/2}}. \quad (5.20)$$

É interessante verificarmos como fica então $f_q(\tau, h)$ com base nesta $\phi(t)$. Temos:

$$f_q(\tau, h) = \phi(\tau, f_p(h)) = \sqrt{\tau^2 + f_p(h)^2}, \quad (5.21)$$

que é uma expressão extremamente simples.

A expressão da função suavizada $f_s(\tau, x)$, utilizando (5.13), é então dada por:

$$f_s(\tau, x) = f_d(x) + \sqrt{\tau^2 + f_p(x-z)^2} + H(x), \quad (5.22)$$

que, também pode ser reescrita como:

$$f_s(\tau, x) = f(x) + \left[\sqrt{\tau^2 + f_p(x-z)^2} - f_p(x-z) \right], \quad (5.23)$$

pois $f_d(x) + H(x) = f(x) - f_p(x-z)$. Esta expressão pode ser abreviada como:

$$f_s(\tau, x) = f(x) + f_a(\tau, x). \quad (5.24)$$

O segundo termo da expressão acima pode ser é a função de ajuste de suavização, que depende das derivadas direcionais em z , da distância $x-z$ e do parâmetro τ . Voltando a expressá-lo em função de f_p , temos:

$$f_a(\tau, x) = \sqrt{\tau^2 + f_p(x-z)^2} - f_p(x-z). \quad (5.25)$$

5.3 Alternativas para a função ϕ

A função $\phi(\tau, t)$ desenvolvida na seção anterior é uma entre várias alternativas possíveis. Uma forma geral de gerar estas funções é mediante a escolha de uma função base $\sigma(\tau, t) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ de classe C^1 para a qual sejam válidas as seguintes propriedades:

$$\begin{aligned} \sigma(\tau, 0) &= \tau, \\ \sigma'(\tau, 0) &= -1, \\ \lim_{t \rightarrow +\infty} \sigma(\tau, t) &= 0, \\ \lim_{\tau \rightarrow 0} \sigma(\tau, t) &= 0, \forall t > 0. \end{aligned}$$

A Figura 5.2 apresenta alguns exemplos de funções $\sigma(\tau, t)$ que atendem a estas propriedades, produzidas com base nas funções $\phi(\tau, t)$ exemplificadas na Figura 5.1.

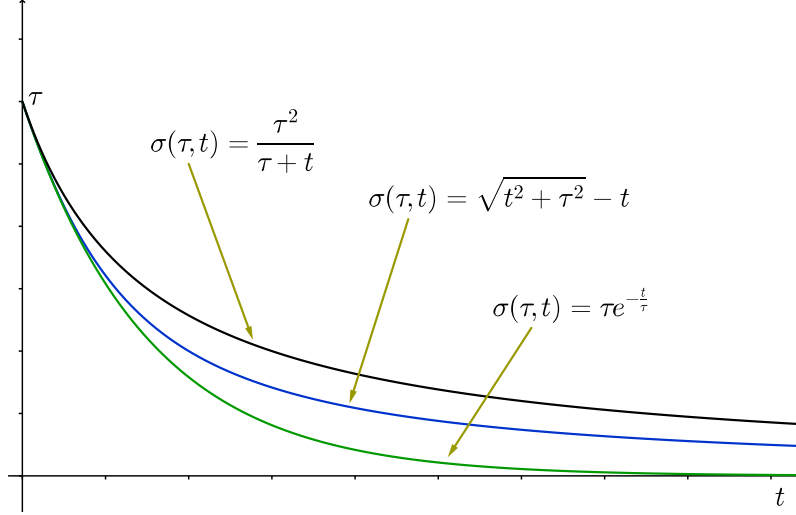


Figura 5.2: Exemplos de funções $\sigma(\tau, t)$ admissíveis

Podemos então definir $\phi(\tau, t) = \sigma(\tau, t) + t$ e é imediato que esta função atende a todas as propriedades requeridas para uma função $\phi(\tau, t)$. A lista abaixo apresenta um conjunto de alternativas baseadas neste método:

- $\phi(\tau, t) = \tau \exp(-\frac{t}{\tau}) + t = \tau \left(\exp(-\frac{t}{\tau}) + \frac{t}{\tau} \right)$;
- $\phi(\tau, t) = \frac{\tau^2}{\tau+t} + t = \tau \left(\left(1 + \frac{t}{\tau}\right)^{-1} + \frac{t}{\tau} \right)$;
- $\phi(\tau, t) = \frac{\tau^{k+1}}{(\tau + \frac{t}{k})^k} + t = \tau \left(\left(1 + \frac{t}{k\tau}\right)^{-k} + \frac{t}{\tau} \right), \forall k \in \mathbb{N} - \{0\}$.

Podemos notar uma certa estrutura nestas funções, pois todas podem ser escritas como $\phi(\tau, t) = \tau \omega(\frac{t}{\tau})$, para uma certa função $\omega : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, incluindo nossa função

$$\phi(\tau, t) = \sqrt{\tau^2 + t^2} = \tau \sqrt{1 + \left(\frac{t}{\tau}\right)^2},$$

conforme apresentado na Tabela 5.1.

Tabela 5.1: Equivalentes limites das funções $\phi(\tau, t)$

$\phi(\tau, t)$	$\omega(t)$	$\omega_\infty(s)$
$\tau \left(\exp(-\frac{t}{\tau}) + \frac{t}{\tau} \right)$	$\exp(-t) + t$	t
$\tau \left(\left(1 + \frac{t}{\tau}\right)^{-1} + \frac{t}{\tau} \right)$	$(1+t)^{-1} + t$	t
$\tau \left(\left(1 + \frac{t}{k\tau}\right)^{-k} + \frac{t}{\tau} \right)$	$\left(1 + \frac{t}{k}\right)^{-k} + t$	t
$\tau \sqrt{1 + \left(\frac{t}{\tau}\right)^2}$	$\sqrt{1 + t^2}$	t

Desta forma, o método de escolha inicial de uma função $\sigma(\tau, t)$ nos leva à cons-

trução de funções $\tau\omega\left(\frac{t}{\tau}\right)$ que tem sempre como limite a função¹ $\omega_\infty(t) = t, t \geq 0$, ou seja:

$$\omega_\infty(t) = \lim_{\tau \rightarrow 0^+} \tau\omega\left(\frac{t}{\tau}\right) = t, \forall t \geq 0.$$

A utilização das chamadas *funções assintóticas* é justamente uma das técnicas de suavização encontradas na literatura, proposta por BEN-TAL e TEBoulLE (1988). Em nosso contexto estamos tratando com a função $\phi(\tau, t)$ com domínio nos reais não negativos, tais que $\phi'(\tau, 0) = 0$, atendendo a restrição (5.4). Mas se estendermos seu domínio para toda a reta real, construindo uma nova função simétrica e diferenciável na origem definida por $\phi_\omega(\tau, t) = \phi(\tau, |t|)$, teremos então sua função assintótica dada por:

$$\lim_{\tau \rightarrow 0^+} \tau\phi_\omega\left(\tau, \frac{t}{\tau}\right) = \lim_{\tau \rightarrow 0^+} \tau\phi\left(\tau, \left|\frac{t}{\tau}\right|\right) = |t| \lim_{\tau \rightarrow 0^+} \frac{\phi\left(\tau, \left|\frac{t}{\tau}\right|\right)}{\left|\frac{t}{\tau}\right|} = |t| \lim_{s \rightarrow +\infty} \frac{\phi(\tau, s)}{s} = |t|,$$

onde utilizamos a propriedade (5.7) no último limite. Percebemos que a função $|t|$ é não diferenciável na origem e temos portanto um conjunto de funções $\phi_\omega(\tau, t)$, diferenciáveis em toda a reta real, e que tendem para a função $|t|$ quando $\tau \rightarrow 0^+$.

Em nossos desenvolvimentos continuaremos a utilizar a função $\phi(\tau, t) = \sqrt{\tau^2 + t^2}$ em virtude de sua extrema simplicidade e estabilidade computacional. Esta é a base da Suavização Hiperbólica clássica, sendo que a função suavizada resultante, conforme já mencionado anteriormente, será denominada de *subgradiente diferenciável*.

¹Utilizamos a nomenclatura de AUSLENDER e TEBoulLE (2003), exceto pelo fato de que estamos trabalhando com os reais não negativos.

Capítulo 6

DHA: Condições para geração de ε -subgradientes de $f(x)$

6.1 Considerações iniciais

Por definição o subdiferencial generaliza para um ponto de não diferenciabilidade z a propriedade da diferencial de uma função convexa diferenciável em um ponto x arbitrário, dada por $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$, $\forall y \in \mathbb{R}^n$, que se transforma na expressão muito similar $f(y) \geq f(x) + g \cdot (y - x)$, $\forall g \in \partial f(z)$, $\forall y \in \mathbb{R}^n$. Permanece uma ligação com o processo de diferenciação, assumindo-se que z seja um ponto isolado de não diferenciabilidade, dada pela expressão $f'(z, d) = \max_{g \in \partial f(z)} g \cdot d$, $\forall d \in \mathbb{R}^n$. Desta forma, $\partial f(z)$ incorpora todos os aspectos do comportamento de primeira ordem de $f(x)$ no ponto z e assim temos que $\partial f(z)$ (ou de forma ainda mais geral, $\partial f(x) = \{\nabla f(x)\}$) contem informações suficientes para a geração de uma aproximação de primeira ordem para $f(x)$ em uma vizinhança diferenciável $x + \varepsilon B^n$ do ponto considerado, pois, para x diferenciável, temos $f(y) \approx f(x) + \nabla f(x) \cdot (y - x)$, $y \in (x + \varepsilon B^n)$, e, para z não diferenciável, temos $f(y) \approx f(z) + f'(z, y - z)$, $y \in (x + \varepsilon B^n)$, onde $f'(z, y - z) = \max_{g \in \partial f(z)} g \cdot (y - z)$.

Por outro lado, tanto $\partial f(z)$ quanto $\partial f(x)$ (∂f , para um ponto arbitrário) ainda estão limitados a um comportamento local, que no caso mais desfavorável será dado por uma derivada direcional¹. Se as informações contidas em ∂f forem utilizadas diretamente em algoritmos de otimização, problemas de convergência podem ocorrer (como a convergência para soluções não ótimas, já comentado anteriormente), pois informações locais (diferenciais) não conseguem expressar comportamentos não diferenciáveis que porventura estejam presentes nas vizinhanças de um determinado ponto. A "solução padrão" para esta situação é substituir os subdiferenciais ∂f pelos ε -subdiferenciais $\partial_\varepsilon f$, constituídos por ε -subgradientes, os quais expressam compor-

¹A derivada direcional sempre estará bem definida em nosso estudo de funções convexas no \mathbb{R}^n

tamentos não somente diferenciais mas também comportamentos não diferenciais presentes em f nas vizinhanças de um ponto. O raciocínio do porque isto acontece é apresentado na proposição A.1 do apêndice.

Não faz parte do escopo deste trabalho discutir a teoria dos ε -subdiferenciais e utilizaremos com liberdade os resultados da análise convexa presentes nas referências.

Nosso principal objetivo é encontrar uma relação entre τ e ε tal que, dado um x qualquer, possamos garantir a pertinência $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$ ou seja, garantir que o subgradiente diferenciável, definido por $\nabla f_s(\tau, x)$ e parametrizado por τ , equivale à escolha de um ε -subgradiente para o ponto x dado. Este é um cenário ideal, pois relacionar diretamente τ e ε de uma forma simples pode não ser possível quando utilizamos a suavização SDHA. No entanto, é possível demonstrar uma relação direta entre τ e ε quando $f_d(x)$ é convexa, o que será desenvolvido a partir deste capítulo.

6.2 Relação de pertinência em ε -subdiferenciais de $f(x)$

Para um $\varepsilon > 0$ arbitrário, mostraremos agora que, se uma determinada restrição for atendida, podemos utilizar um ε -subgradiente de $f_p(x-z)$ para gerar ε -subgradientes de $\partial_\varepsilon f(x)$, lembrando que $f_p(x)$ é convexa.

Seja $g \in \partial_\varepsilon f_p(x-z)$. Então, para todo $y \in \mathbb{R}^n$:

$$\begin{aligned} f_p(y-z) &\geq f_p(x-z) + g \cdot (y-x) - \varepsilon, \\ f_d(y) + f_p(y-z) + H(y) &\geq f_d(y) + f_p(x-z) + f(z) + \bar{g} \cdot (y-z) + g \cdot (y-x) - \varepsilon, \\ f(y) &\geq f_d(y) + f_p(x-z) + f(z) + \bar{g} \cdot (y-z) + g \cdot (y-x) - \varepsilon, \end{aligned}$$

onde utilizamos que $H(y) = f(z) + \bar{g} \cdot (y-z)$. Somando e subtraindo $f_d(x)$ e lembrando que $f(x) = f_d(x) + f_p(x-z) + f(z) + \bar{g} \cdot (x-z)$, temos:

$$\begin{aligned} f(y) &\geq f_d(x) + f_p(x-z) + f(z) + \bar{g} \cdot (y-z) + g \cdot (y-x) - \varepsilon + f_d(y) - f_d(x), \\ f(y) &\geq f(x) - \bar{g} \cdot (x-z) + \bar{g} \cdot (y-z) + g \cdot (y-x) - \varepsilon + f_d(y) - f_d(x), \\ f(y) &\geq f(x) + (g + \bar{g}) \cdot (y-x) - \varepsilon + f_d(y) - f_d(x), \\ f(y) &\geq f(x) + (\nabla f_d(x) + g + \bar{g}) \cdot (y-x) - \varepsilon + f_d(y) - (f_d(x) + \nabla f_d(x)(y-x)). \end{aligned}$$

Suponhamos agora que $f_d(x)$ seja uma função que atenda à seguinte restrição:

$$f_d(y) - (f_d(x) + \nabla f_d(x)(y-x)) \geq -\varepsilon_d(x), \quad (6.1)$$

para todo par $x, y \in \mathbb{R}^n$ e para uma determinada função $\varepsilon_d(x) \geq 0$. Temos então que:

$$f(y) \geq f(x) + (\nabla f_d(x) + g + \bar{g}) \cdot (y - x) - \varepsilon - \varepsilon_d(x), \quad (6.2)$$

ou seja, $\nabla f_d(x) + g + \bar{g} \in \partial_{\varepsilon + \varepsilon_d(x)} f(x)$ e portanto:

$$\nabla f_d(x) + \partial_\varepsilon f_p(x - z) + \bar{g} \subset \partial_{\varepsilon + \varepsilon_d(x)} f(x). \quad (6.3)$$

Em outras palavras, se pudermos gerar subgradientes de $\partial_\varepsilon f_p(x - z)$ iremos automaticamente gerar subgradientes de $\partial_{\varepsilon + \varepsilon_d(x)} f(x)$.

No caso de uma $f_d(x)$ geral não se pode afirmar que exista uma função $\varepsilon_d(x)$ atendendo a condição 6.1. Por outro lado, impor a restrição $\varepsilon_d(x) \equiv 0$ é equivalente a demandar que $f_d(x)$ seja convexa. Para este caso especial, ε -subgradientes de $\partial_\varepsilon f_p(x - z)$ nos permitem gerar automaticamente ε -subgradientes de $\partial_\varepsilon f(x)$ utilizando a expressão simplificada:

$$\nabla f_d(x) + \partial_\varepsilon f_p(x - z) + \bar{g} \subset \partial_\varepsilon f(x). \quad (6.4)$$

Até este momento não havíamos comentado sobre a convexidade de $f_d(x)$. O desenvolvimento apresentado acima nos mostra que quando $f_d(x)$ é convexa existe um relacionamento direto e simples entre $\partial_\varepsilon f_p(x - z)$ e $\partial_\varepsilon f(x)$, e isto implica em uma grande simplificação na derivação do relacionamento entre o subgradiente diferenciável e o ε -subdiferencial de $f(x)$.

Iremos mostrar agora um exemplo simples de $f(x)$, em que podemos facilmente provar que $f_d(x)$ é convexa e depois apresentaremos um exemplo onde $f_d(x)$ não é convexa.

6.3 Primeiro exemplo: $f_d(x)$ convexa

Seja $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ uma função convexa diferenciável não-decrescente e não negativa, tendo por domínio os reais não negativos, tal que $h'(0^+) = k > 0$. Iremos agora provar que $g(t) = h(t) - kt$ tem as mesmas propriedades gerais de $h(t)$. Obviamente $g(t)$ é diferenciável (com $g'(0^+) = 0$) e o domínio continua sendo o dos reais não negativos. Como $h(t)$ é convexa e não negativa temos:

$$h(t) \geq h(0) + kt \Rightarrow g(t) = h(t) - kt \geq h(0) \geq 0.$$

Logo $g(t)$ é não negativa. Por outro lado, a desigualdade básica das funções convexas diferenciáveis nos reais nos diz que $h'(t) \geq h'(0) = k$. Dados $0 < t_1 < t_2$

e utilizando a convexidade de $h(t)$ temos:

$$\begin{aligned} h(t_2) &\geq h(t_1) + h'(t_1)(t_2 - t_1) \geq h(t_1) + k(t_2 - t_1) \Rightarrow \\ &\Rightarrow g(t_2) = h(t_2) - kt_2 \geq h(t_1) - kt_1 = g(t_1), \end{aligned}$$

e portanto $g(t)$ é não-decrescente. Sejam agora $0 \leq \lambda \leq 1$ e $0 < t_1 < t_2$. Mais uma vez utilizando a convexidade de $h(t)$ temos:

$$\begin{aligned} g((1 - \lambda)t_1 + \lambda t_2) &= h((1 - \lambda)t_1 + \lambda t_2) - k((1 - \lambda)t_1 + \lambda t_2) \leq \\ &\leq (1 - \lambda)h(t_1) + \lambda h(t_2) - (1 - \lambda)kt_1 - \lambda kt_2 = (1 - \lambda)g(t_1) + \lambda g(t_2), \end{aligned}$$

e assim $g(t)$ é também convexa.

Resumidamente temos que $g(t)$ é, tal como $h(t)$, uma função convexa diferenciável não-decrescente e não negativa tendo como domínio os reais não negativos.

Definamos agora $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$ como $f(x) = h(\|x\|) = g(\|x\|) + k\|x\|$. Temos claramente que $f_n(x) = k\|x\|$ e portanto $f_d(x) = h(\|x\|) - h(0) - k\|x\| = g(\|x\|) - g(0)$. Sejam x_1 e x_2 dois pontos quaisquer do \mathbb{R}^n e seja $0 \leq \lambda \leq 1$. Temos então:

$$\begin{aligned} f_d((1 - \lambda)x_1 + \lambda x_2) &= g(\|(1 - \lambda)x_1 + \lambda x_2\|) - g(0) \leq \\ &\leq g((1 - \lambda)\|x_1\| + \lambda\|x_2\|) - g(0) \leq \\ &\leq (1 - \lambda)g(\|x_1\|) + \lambda g(\|x_2\|) - (1 - \lambda + \lambda)g(0) = \\ &= (1 - \lambda)f_d(x_1) + \lambda f_d(x_2), \end{aligned}$$

onde na primeira desigualdade utilizamos o fato de que $g(t)$ é não-decrescente e a função norma é convexa, e na segunda desigualdade utilizamos o fato de que $g(t)$ é convexa. A função $f_d(x)$ é portanto convexa.

De forma concreta, podemos tomar como exemplo a função $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ definida por $h(x) = x^2 + 2x + 2$, que pode ser expressa por $h(x) = g(x) + 2x$, com $g(x) = x^2 + 2$ e $k = 2$. Assim temos $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$ definida por $f(x) = \|x\|^2 + 2\|x\| + 2$, com $f_n(x) = 2\|x\|$ e $f_d(x) = g(\|x\|) - g(0) = \|x\|^2$. A Figura 6.1 apresenta uma secção de $f(x)$, $f_d(x)$ e $f_n(x)$ para este caso particular. Notar a simetria gerada por este método.

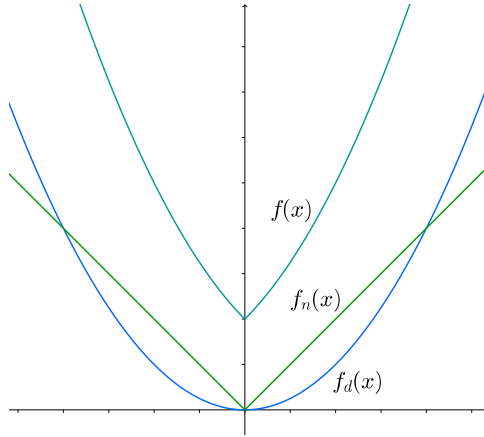


Figura 6.1: Secção de $f(x) = \|x\|^2 + 2\|x\| + 2$, com $f_d(x)$ convexa

Temos portanto um exemplo de $f(x)$ para a qual $f_d(x)$ é convexa.

6.4 Segundo exemplo: $f_d(x)$ não-convexa

Daremos agora um exemplo de uma função convexa $f : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ para a qual $f_d(x)$ não é convexa. Iniciaremos definindo uma primeira função, que chamaremos de $f_1(x)$, como o máximo entre um múltiplo da função norma e uma função afim que não passa pela origem:

$$f_1(x) = \max\{k_1\|x\|, k_2x_1 - a\}, \quad (6.5)$$

onde k_1 , k_2 e a são constantes positivas, com $k_2 > k_1$ para assegurar que haja interseção entre os gráficos das duas funções. A função $f_1(x)$ é convexa por ser o máximo de duas funções convexas. Apresentamos na Figura 6.2 uma secção desta função pelo plano $x_2 = 0$.

Para esta função temos claramente que $z = 0$ é um ponto de não diferenciabilidade. No entanto, a função $f_1(x)$ é também não diferenciável nos pontos do domínio para os quais ocorre a interseção entre os gráficos de suas funções componentes. Para remediar isto, iremos substituir a função máximo por uma suavização exponencial $S(\xi_1, \xi_2) = \ln(e^{\xi_1} + e^{\xi_2})$. No apêndice A.3 é apresentado o desenvolvimento analítico que demonstra a convexidade de $S(\xi_1, \xi_2)$ para $\xi_1(x) = k_1\|x\|$ e $\xi_2(x) = k_2x_1 - a$.

Assim definiremos finalmente nossa função convexa $f(x)$ como:

$$f(x) = \ln(e^{k_1\|x\|} + e^{k_2x_1 - a}). \quad (6.6)$$

A presença da norma $\|x\|$ na composição da função $f(x)$ mantém a não diferenciabilidade em $z = 0$ e como o argumento da função logaritmo é sempre maior que

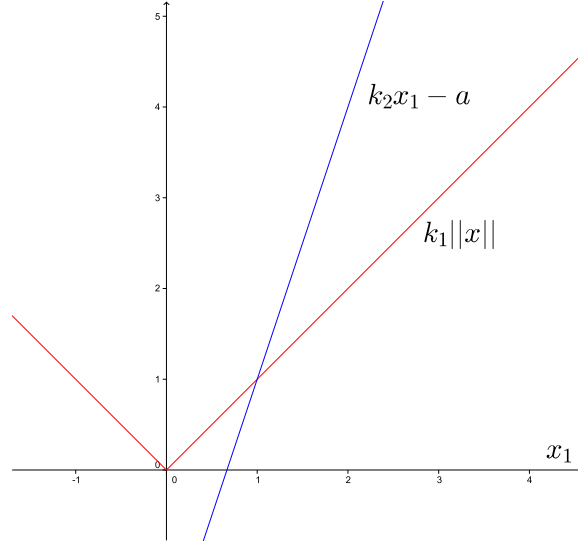


Figura 6.2: Secção da função $f_1(x)$ pelo plano $x_2 = 0$

um, temos que $f(x) > 0, \forall x \in \mathbb{R}^2$, e assim temos a liberdade de escolher $\bar{g} = 0$ para simplificar nossas considerações quanto à decomposição de $f(x)$.

Visando agora completar a decomposição de $f(x)$, calculemos a derivada direcional no ponto $z = 0$:

$$f'(0, d) = \lim_{t \rightarrow 0^+} \frac{f(td) - f(0)}{t} = \lim_{t \rightarrow 0^+} \frac{\ln(e^{k_1 t \|d\|} + e^{k_2 t d_1 - a}) - \ln(1 + e^{-a})}{t}, \quad (6.7)$$

e aplicando-se L'Hôpital,

$$f'(0, d) = \lim_{t \rightarrow 0^+} \frac{k_1 \|d\| e^{k_1 t \|d\|} + k_2 d_1 e^{k_2 t d_1 - a}}{e^{k_1 t \|d\|} + e^{k_2 t d_1 - a}} = \frac{k_1 \|d\| + k_2 d_1 e^{-a}}{1 + e^{-a}}. \quad (6.8)$$

Para nosso exemplo iremos impor que $f'(0, d) > 0$ para toda a direção d . Temos então que assegurar isto pela análise do numerador da expressão de $f'(0, d)$ na pior situação, quando d produz o valor mais negativo possível para d_1 , que trivialmente acontece quando $d = (-\|d\|, 0)$ e então deve valer:

$$k_1 \|d\| - k_2 \|d\| e^{-a} > 0 \Rightarrow a > \ln \left(\frac{k_2}{k_1} \right) \quad (6.9)$$

Se definirmos a constante $\lambda = 1/(1 + e^{-a})$ e lembrarmos que $e^{-a} > 0, \forall a \in \mathbb{R}$, temos que $0 < \lambda < 1$. Podemos então reescrever a expressão de $f'(0, d)$ como:

$$f'(0, d) = \lambda k_1 \|d\| + (1 - \lambda) k_2 d_1. \quad (6.10)$$

Ou seja, a derivada direcional de $f(x)$ no ponto de não diferenciabilidade $z = 0$ é

uma combinação convexa de $k_1||d||$ e de k_2d_1 . Finalmente, temos então:

$$f_d(x) = \ln(e^{k_1||x||} + e^{k_2x_1-a}) - (\lambda k_1||x|| + (1-\lambda)k_2x_1) - \ln(1 + e^{-a}),$$

$$f_p(x-z) = f_p(x) = \lambda k_1||x|| + (1-\lambda)k_2x_1.$$

As Figuras 6.3 e 6.4 representam gráficamente $f(x)$ e $f_d(x)$, para $k_1 = 1$, $k_2 = 3$ e $a = 2$, onde claramente podemos suspeitar da não convexidade de $f_d(x)$.

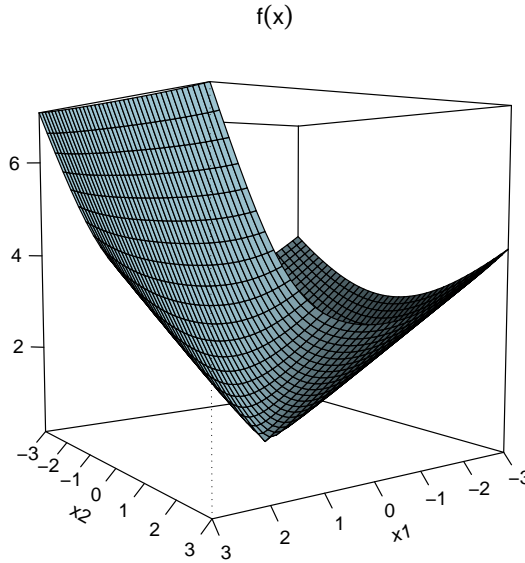


Figura 6.3: Exemplo de $f(x)$ cuja $f_d(x)$ é não convexa

Resta agora provarmos que $f_d(x)$ é não convexa. Consideremos a função $f_d(x)$ ainda com a mesma parametrização utilizada para os gráficos e iremos provar que existem dois pontos distintos x^1 e x^2 pertencentes ao domínio de \mathbb{R}^2 tais que $f_d\left(\frac{x^1}{2} + \frac{x^2}{2}\right) > \frac{f_d(x^1)}{2} + \frac{f_d(x^2)}{2}$. Mas $f_d(x)$ é simétrica com relação à segunda componente do vetor x , pois a expressão é invariante com relação à simetria desta componente, e portanto se tomarmos x^1 e x^2 simétricos com relação ao eixo da primeira componente a desigualdade irá ser simplificada.

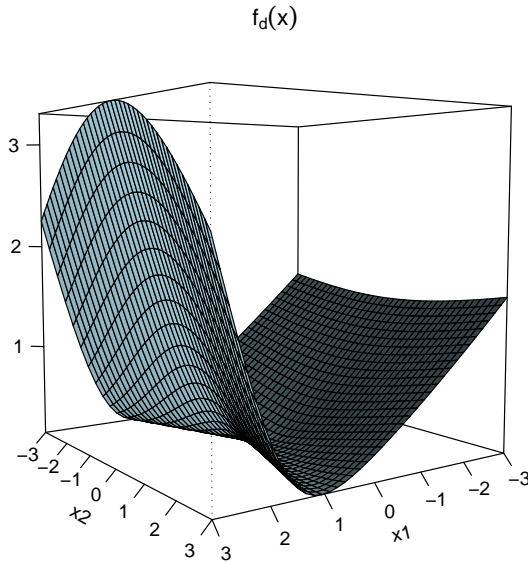


Figura 6.4: $f_d(x)$ não convexa da função $f(x)$ apresentada na figura anterior

Tomemos então $x^1 = (b, b)$ e $x^2 = (b, -b)$, com $b > 0$. Teremos que provar então que para um determinado $b > 0$ vale

$$f_d\left(\frac{x^1}{2} + \frac{x^2}{2}\right) = f_d(b, 0) > \frac{f_d(x^1)}{2} + \frac{f_d(x^2)}{2} = \frac{f_d(b, b)}{2} + \frac{f_d(-b, b)}{2} = f_d(b, b).$$

Como basta um exemplo para provarmos a possibilidade acima mencionada, basta tomarmos $b = 2$ e para este valor teremos $f_d(2, 0) (\approx 1.52) > f_d(2, 2) (\approx 0.94)$, demonstrando que $f_d(x)$ é não convexa.

Capítulo 7

Análise SDHA para $f_d(x)$ convexa

Assumindo-se a convexidade de $f_d(x)$, já demonstramos no capítulo anterior que:

$$\nabla f_d(x) + \partial_\varepsilon f_p(x - z) + \bar{g} \subset \partial_\varepsilon f(x). \quad (7.1)$$

Por outro lado, com base em (5.13), temos:

$$\nabla f_s(\tau, x) = \nabla f_d(x) + \nabla f_q(\tau, x - z) + \bar{g}. \quad (7.2)$$

Comparando-se as duas expressões anteriores, podemos afirmar que:

$$\nabla f_q(\tau, x - z) \in \partial_\varepsilon f_p(x - z) \implies \nabla f_s(\tau, x) \in \partial_\varepsilon f(x). \quad (7.3)$$

Percebemos então que é necessário explicitarmos $\partial_\varepsilon f_p(h)$ para que a implicação acima possa ser explorada.

7.1 Cálculo de $\partial_\varepsilon f_p(h)$

Para calcularmos $\partial_\varepsilon f_p(h)$ comecemos com a equação básica de um ε -subgradiente de $f_p(h)$:

$$f_p(y) \geq f_p(h) + g \cdot (y - h) - \varepsilon, \quad (7.4)$$

ou,

$$g \cdot y - f_p(y) \leq g \cdot h - f_p(h) + \varepsilon. \quad (7.5)$$

Seja agora $y = tu$, com $u \in S^n$ unitário, onde u representa a direção de y e $t \geq 0$ representa o seu módulo ($\|y\| = t$):

$$g \cdot tu - f_p(tu) \leq g \cdot h - f_p(h) + \varepsilon, \quad (7.6)$$

que, por ser f_p homogênea, pode ser reescrito como:

$$t(g.u - f_p(u)) \leq g.h - f_p(h) + \varepsilon. \quad (7.7)$$

Assim, para que g pertença a $\partial_\varepsilon f_p(h)$, a desigualdade acima deve ser válida para todo $u \in S^n$ unitário e t positivo. Como t pode crescer sem limite para qualquer g dado, devemos ter obrigatoriamente $g.u - f_p(u) \leq 0$, ou $g.u \leq f_p(u)$, para todo $u \in S^n$ unitário. Por outro lado, por (4.9) e (2.14) sabemos que:

$$f_p(u) = f'_p(0, u) = \max_{a \in \partial f_p(0)} a.u. \quad (7.8)$$

Assim, para que g pertença a $\partial f_\varepsilon(h)$ deve ser válido, para todo vetor unitário $u \in S^n$, que:

$$g.u \leq f'_p(0, u) = \max_{a \in \partial f_p(0)} a.u. \quad (7.9)$$

A validade desta desigualdade em g para todo vetor unitário $u \in S^n$ pode ser descrita como:

$$g \in \bigcap_{u \in S^n} \{a \in \mathbb{R}^n : a.u \leq f'_p(0, u)\} = \partial f_p(0).$$

Esta última expressão nos diz simplesmente que, para todo vetor unitário $u \in S^n$, $g.u \leq f_p(u)$ se e somente se $g \in \partial f_p(0)$, e esta é justamente a primeira restrição que estávamos procurando. Assim, para que $g \in \partial_\varepsilon f_p(h)$ devemos ter como primeira restrição $g \in \partial f_p(0)$.

Porém, esta restrição sobre g não é a única. Isto porque para $g \in \partial f_p(0)$ a expressão do lado esquerdo, linear em t , é sempre menor ou igual a zero, como acabamos de provar, de forma que sempre atingirá zero como seu valor máximo, para $t = 0$. Desta forma, para que a desigualdade (7.7) seja sempre válida, a expressão do lado direito de (7.7) deve ser sempre maior ou igual a zero, e isto impõe uma restrição adicional sobre g

$$g.h - f_p(h) + \varepsilon \geq 0, \quad (7.10)$$

ou,

$$g.h \geq f_p(h) - \varepsilon. \quad (7.11)$$

Esta é a equação de um semi-espço fechado $H_+ = \{g \in \mathbb{R}^n : g.h \geq f_p(h) - \varepsilon\}$, definido pelo hiperplano $H = \{g \in \mathbb{R}^n : g.h = f_p(h) - \varepsilon\}$ ortogonal à direção dada por h .

Com base nas duas restrições apresentadas acima, temos portanto que $g \in \partial_\varepsilon f_p(h) \Leftrightarrow g \in \partial f_p(0) \cap H_+$.

Pode no entanto ocorrer que, para determinados h e ε , $\partial f_p(0) \subset H_+$ e a restrição

sobre g ficar simplificada para somente $g \in \partial f_p(0)$. Iremos reescrever nossa última desigualdade, buscando agora restrições sobre ε que a tornem automaticamente válida:

$$\varepsilon \geq f_p(h) - g.h. \quad (7.12)$$

Como neste momento estamos considerando h fixo e esta desigualdade deve ser válida para todo g que atende a primeira restrição, ε deve satisfazer:

$$\varepsilon \geq \max_{g \in \partial f_p(0)} \{f_p(h) - g.h\} = f_p(h) + \max_{g \in \partial f_p(0)} g.(-h) = f_p(h) + f'_p(0, -h), \quad (7.13)$$

e com base em (4.9) temos finalmente:

$$\varepsilon \geq f_p(h) + f_p(-h). \quad (7.14)$$

Portanto, se $\varepsilon \geq f_p(h) + f_p(-h)$ a única restrição sobre os ε -subgradientes g de $f_p(h)$ é que $g \in \partial f_p(0)$. Para simplificar expressões posteriores iremos definir $\varepsilon_h = f_p(h) + f_p(-h)$ e a condição para que não exista uma restrição adicional a g para um dado h é que $\varepsilon \geq \varepsilon_h$. Caso $0 \leq \varepsilon < \varepsilon_h$, então as duas restrições devem ser aplicadas.

De forma resumida:

$$g \in \partial_\varepsilon f_p(h) \Leftrightarrow \begin{cases} 0 \leq \varepsilon < f_p(h) + f_p(-h) & : g \in \partial f_p(0) \cap H_+, \\ \varepsilon \geq f_p(h) + f_p(-h) & : g \in \partial f_p(0). \end{cases} \quad (7.15)$$

Antes de continuarmos iremos provar que para qualquer $\varepsilon > 0$ é sempre válido que:

$$\nabla f_p(h) \in \partial_\varepsilon f_p(h). \quad (7.16)$$

Primeiramente temos que $\nabla f_p(h)$ atende a primeira restrição pois, pela equação básica dos subgradientes:

$$\begin{aligned} f_p(y) &\geq f_p(h) + \nabla f_p(h).(y - h) \\ &= \nabla f_p(h).y \\ &= f_p(0) + \nabla f_p(h).(y - 0) \end{aligned} \quad (7.17)$$

onde utilizamos (4.11). Assim concluímos que $\nabla f_p(h) \in \partial f_p(0)$.

Para a segunda restrição temos que para todo $\varepsilon > 0$:

$$\nabla f_p(h).h = f_p(h) \geq f_p(h) - \varepsilon \Rightarrow \nabla f_p(h) \in H_+, \quad (7.18)$$

o que prova a afirmação realizada.

7.2 Cálculo de $\nabla f_q(\tau, h)$

Para o caso particular $h = 0$ temos por (5.11) que $\nabla f_q(\tau, 0) = 0$, independente de qual $\phi(\tau, t)$ for utilizada.

Para $h \neq 0$, lembremos (5.10):

$$\nabla f_q(\tau, h) = \phi'(\tau, f_p(h)) \nabla f_p(h).$$

Como já escolhemos nossa $\phi(\tau, t)$, podemos agora explicitar $\nabla f_q(\tau, h)$ com base na derivada de $\phi(\tau, t)$, conforme apresentada em (5.19):

$$\nabla f_q(\tau, h) = \frac{f_p(h)}{\sqrt{\tau^2 + f_p(h)^2}} \nabla f_p(h). \quad (7.19)$$

De uma forma geral utilizaremos somente a expressão acima, ficando subentendido que para o caso $h = 0$ temos por definição $\nabla f_q(\tau, 0) = 0$, o que assegura a continuidade de $\nabla f_q(\tau, h)$ para todo $h \in \mathbb{R}^n$.

7.3 Relação entre ε e τ para que $\nabla f_q(\tau, h)$ seja um ε -subgradiente de $f_p(h)$

Iremos agora demonstrar que para qualquer $\varepsilon > 0$ existe um $\tau > 0$ tal que $\nabla f_q(\tau, h) \in \partial_\varepsilon f_p(h)$, $\forall h \in \mathbb{R}^n$. Para que esta afirmação seja verdadeira, τ deve ser obtido como uma função de ε e o gradiente $\nabla f_q(\tau, h)$ para este τ deve atender às duas restrições $\nabla f_q(\tau, h) \in \partial f_p(0)$ e $\nabla f_q(\tau, h) \in H_+$, conforme já apresentado anteriormente.

Para que $\nabla f_q(\tau, h) \in \partial f_p(0)$ devemos notar primeiramente que $0 \in \partial f_p(0)$ (pois $f_p(h) \geq 0 \Rightarrow f_p(h) \geq f_p(0) + 0 \cdot (h - 0)$) e, por (7.16), temos também que $\nabla f_p(h) \in \partial f_p(0)$. Como $\partial f_p(0)$ é convexo, para qualquer λ tal que $0 \leq \lambda \leq 1$ vale $(1 - \lambda)0 + \lambda \nabla f_p(h) \in \partial f_p(0)$, ou $\lambda \nabla f_p(h) \in \partial f_p(0)$. Temos também que $f_p(h) \geq 0$ e $0 < \tau \leq \sqrt{\tau^2 + f_p(h)^2}$ e se definirmos

$$\lambda = \frac{f_p(h)}{\sqrt{\tau^2 + f_p(h)^2}},$$

temos imediatamente que $\lambda \geq 0$. Por outro lado, temos:

$$\lambda^2 = \frac{f_p(h)^2}{\tau^2 + f_p(h)^2} < \frac{\tau^2 + f_p(h)^2}{\tau^2 + f_p(h)^2} = 1 \Rightarrow \lambda < 1. \quad (7.20)$$

Concluimos portanto que $\lambda \nabla f_p(h) \in \partial f_p(0)$. Mas por (7.19) temos que $\lambda \nabla f_p(h) = \nabla f_q(\tau, h)$, ou seja, para todo $\tau > 0$ e todo $h \in \mathbb{R}^n$ vale $\nabla f_q(\tau, h) \in$

$\partial f_p(0)$ e a primeira restrição é sempre cumprida.

A segunda restrição, $\nabla f_q(\tau, h) \in H_+$, demanda por (7.11) que tenhamos $\nabla f_q(\tau, h).h \geq f_p(h) - \varepsilon$. Expandindo esta expressão utilizando (7.19) temos:

$$\frac{f_p(h)}{\sqrt{\tau^2 + f_p(h)^2}} \nabla f_p(h).h \geq f_p(h) - \varepsilon. \quad (7.21)$$

Utilizando $\nabla f_p(h).h = f_p(h)$ (4.11) e rearranjando a desigualdade acima:

$$f_p(h) - \frac{f_p(h)^2}{\sqrt{\tau^2 + f_p(h)^2}} \leq \varepsilon. \quad (7.22)$$

Se denominarmos

$$\eta(\tau, s) = s - \frac{s^2}{\sqrt{\tau^2 + s^2}} = s \left(1 - \frac{s}{\sqrt{\tau^2 + s^2}} \right),$$

a restrição anterior pode ser reescrita como $\eta(\tau, f_p(h)) \leq \varepsilon$. Facilmente verificamos que $\eta(\tau, 0) = 0$ e também que para $s > 0$ sempre teremos $s/\sqrt{\tau^2 + s^2} < 1$, como já provado acima, implicando em $\eta(\tau, s) > 0$. Finalmente, temos que $\lim_{s \rightarrow \infty} \eta(\tau, s) = 0$, pois como para $\alpha > 0$ vale que $\sqrt{1 + \alpha} < 1 + \alpha/2$, temos:

$$\begin{aligned} \lim_{s \rightarrow \infty} s - \frac{s^2}{\sqrt{\tau^2 + s^2}} &= \lim_{s \rightarrow \infty} s - \frac{s}{\sqrt{\frac{\tau^2}{s^2} + 1}} = \lim_{s \rightarrow \infty} \frac{s\sqrt{\frac{\tau^2}{s^2} + 1} - s}{\sqrt{\frac{\tau^2}{s^2} + 1}} < \\ &< \lim_{s \rightarrow \infty} \frac{s(\frac{\tau^2}{2s^2} + 1) - s}{\sqrt{\frac{\tau^2}{s^2} + 1}} = \lim_{s \rightarrow \infty} \frac{\frac{\tau^2}{2s}}{\sqrt{\frac{\tau^2}{s^2} + 1}} = 0. \end{aligned}$$

As observações acima, juntamente com o fato de que a função $\eta(\tau, s)$ é claramente de classe C^1 , nos indicam que $\eta(\tau, s)$ possui pelo menos um máximo local em \mathbb{R}_+ . Notemos que se $\eta(\tau, s)$ possuir um valor máximo bem determinado, a segunda restrição poderá ser facilmente atendida simplesmente fazendo $\max_{s \in \mathbb{R}_+} \eta(\tau, s) \leq \varepsilon$, uma relação que independe do valor de s (ou, em nosso caso de interesse, do valor de $f_p(h)$).

Iremos portanto levantar os pontos críticos de $\eta(\tau, s)$, fazendo $\eta'(\tau, s) = 0$:

$$\eta'(\tau, s) = 1 - \frac{2s\tau^2 + s^3}{(\tau^2 + s^2)^{\frac{3}{2}}} = \frac{(\tau^2 + s^2)^{\frac{3}{2}} - 2s\tau^2 - s^3}{(\tau^2 + s^2)^{\frac{3}{2}}} = 0.$$

Trabalhando agora somente com o numerador e desenvolvendo temos:

$$\begin{aligned} s^3 + 2s\tau^2 &= (\tau^2 + s^2)^{\frac{3}{2}} \\ s^6 + 4s^4\tau^2 + 4s^2\tau^4 &= s^6 + 3s^4\tau^2 + 3s^2\tau^4 + \tau^6 \\ s^4 + s^2\tau^2 - \tau^4 &= 0 \end{aligned}$$

Esta última equação é uma biquadrada e pode ser facilmente resolvida com a substituição $y = s^2$:

$$y^2 + y\tau^2 - \tau^4 = 0.$$

Temos portanto:

$$y = s^2 = \frac{-\tau^2 + \sqrt{\tau^4 + 4\tau^4}}{2} = \frac{\sqrt{5} - 1}{2}\tau^2,$$

onde foi utilizada somente a raiz positiva pois $s^2 \geq 0$. Logo:

$$s = \sqrt{\frac{\sqrt{5} - 1}{2}\tau} = \frac{\tau}{\sqrt{\varphi}}, \quad (7.23)$$

onde $\varphi = (\sqrt{5} + 1)/2$ é a razão áurea, solução de $\varphi - 1 = 1/\varphi$, ou equivalentemente, $\varphi^2 = \varphi + 1$, relação esta que será utilizada a seguir.

Como somente temos uma raiz viável, pois também temos que ter $s \geq 0$, o máximo é único. O valor deste máximo como uma função de τ é dado por:

$$\max_{s \in \mathbb{R}_+} \eta(\tau, s) = \eta\left(\tau, \frac{\tau}{\sqrt{\varphi}}\right) = \frac{\tau}{\sqrt{\varphi}} - \frac{\left(\frac{\tau}{\sqrt{\varphi}}\right)^2}{\sqrt{\tau^2 + \left(\frac{\tau}{\sqrt{\varphi}}\right)^2}} = \left(\frac{1}{\sqrt{\varphi}} - \frac{\frac{1}{\varphi}}{\sqrt{1 + \frac{1}{\varphi}}}\right)\tau,$$

e, portanto,

$$\max_{s \in \mathbb{R}_+} \eta(\tau, s) = \left(\frac{1}{\sqrt{\varphi}} - \frac{\frac{1}{\varphi}}{\sqrt{1 + \frac{1}{\varphi}}}\right)\tau = \left(\frac{1}{\sqrt{\varphi}} - \frac{\frac{1}{\varphi}}{\sqrt{\varphi}}\right)\tau = \varphi^{-\frac{5}{2}}\tau.$$

O gráfico da função $\eta(\tau, s)$ é apresentado na Figura 7.1 para uma melhor compreensão de suas principais características.

A segunda restrição será cumprida portanto, se tivermos:

$$\varphi^{-\frac{5}{2}}\tau \leq \varepsilon \quad (7.24)$$

Como já verificamos que a primeira restrição é sempre cumprida, temos que a segunda restrição é a única que deve ser observada. Na prática começamos com um ε dado e desejamos calcular um ε -subgradiente. Devemos então determinar um τ

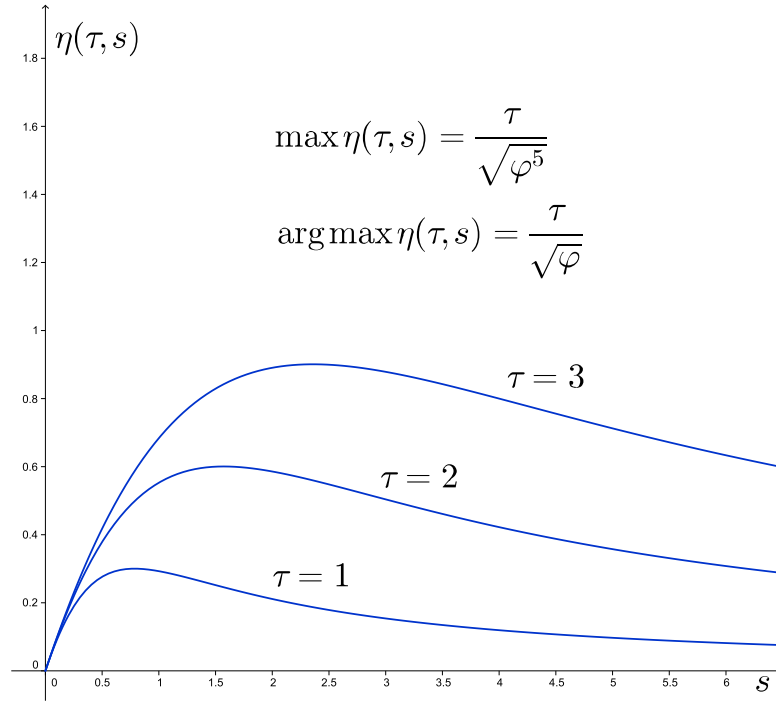


Figura 7.1: Gráfico da função $\eta(\tau, s)$

para um determinado ε dado, e portanto o parâmetro τ deve obedecer à restrição:

$$\tau \leq \varphi^{\frac{5}{2}} \varepsilon = \sqrt{\varphi^5} \varepsilon \quad (7.25)$$

Portanto, dado um $\varepsilon > 0$ escolhamos um $\tau > 0$ atendendo $\tau \leq \sqrt{\varphi^5} \varepsilon$ e então $\nabla f_q(\tau, h)$ será um ε -subgradiente de $f_p(h)$. Concluimos, utilizando (7.3) que, dado $\varepsilon > 0$:

$$0 < \tau \leq \sqrt{\varphi^5} \varepsilon \implies \nabla f_s(\tau, x) = \bar{g} + \nabla f_d(x) + \nabla f_q(\tau, x - z) \in \partial_\varepsilon f(x) \quad (7.26)$$

Ou seja, também é válido que $\nabla f_s(\tau, x)$ é um ε -subgradiente de $f(x)$, como queríamos demonstrar.

Capítulo 8

Uso da suavização SDA para gerar ε -subgradientes

8.1 SDA na geração de ε -subgradientes para $Z_f = \{z\}$

Como vimos no capítulo 7, quando utilizamos a suavização SDHA a convexidade de $f_d(x)$ permite idealizar o relacionamento entre ε e τ , de tal forma a torná-lo independente do particular valor de x considerado. No entanto, muitas vezes pode não ser fácil confirmar a convexidade de f_d e este fato inviabiliza a utilização do subgradiente diferenciável via SDHA.

Para remediar esta situação iremos utilizar a suavização SDA, conforme desenvolvimentos do capítulo 3, para definirmos um novo subgradiente diferenciável $\nabla f_s(\tau, x)$ e encontrar um relacionamento entre τ e ε tal que $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$. O desenvolvimento será realizado sem utilizarmos diretamente os resultados da seção 3.6, visando maior clareza no caso especial $Z_f = \{z\}$. Definamos então a decomposição DA para um ponto isolado z como:

$$f(x) = (f(x) - f(z) - \bar{g} \cdot (x - z)) + f(z) + \bar{g} \cdot (x - z) = (f(x) - H(x)) + H(x),$$

onde $\bar{g} \in \text{ri}\{\partial f(z)\}$, como nos desenvolvimentos anteriores, e denominaremos então

$$f_b(x) = f(x) - H(x) = f_d(x) + f_p(x - z),$$

e assim

$$f(x) = f_b(x) + H(x). \tag{8.1}$$

Por definição temos $f_b(z) = 0$ e como $f(x) \geq f(z) + \bar{g} \cdot (x - z) = H(x)$, temos também $f_b(x) \geq 0, \forall x \in \mathbb{R}^n$. Utilizando as propriedades de f_p e f_d já apresentadas

anteriormente podemos facilmente concluir que f_b é diferenciável em todo ponto, exceto pelo ponto z de não diferenciabilidade¹, e portanto:

$$\nabla f_b(x) = \nabla f(x) - \bar{g}, \forall x \in \mathbb{R}^n \setminus \{z\}.$$

Para suavizar $f_b(x)$ iremos criar a sua composta com a função de suavização $\phi(\tau, t)$, definindo uma nova função:

$$f_c(\tau, x) = \phi(\tau, f_b(x)).$$

Esta função é diferenciável para todo $x \neq z$:

$$\nabla f_c(\tau, x) = \phi'(\tau, f_b(x)) \nabla f_b(x).$$

Para $x = z$ podemos utilizar o limite

$$\lim_{x \rightarrow z} \nabla f_c(\tau, x) = \lim_{x \rightarrow z} \phi'(\tau, f_b(x)) \nabla f_b(x) = \lim_{x \rightarrow z} \phi'(\tau, f_b(x)) (\nabla f(x) - \bar{g}) = 0,$$

pois $\|\nabla f(x) - \bar{g}\| < \infty$, e definir

$$\nabla f_c(\tau, z) = 0,$$

e assim f_c pertence à classe C^1 . Esta função pode portanto ser utilizada para criarmos uma nova versão da suavização de f , substituindo-se f_b em (8.1) por f_c :

$$f_s(\tau, x) = f_c(\tau, x) + H(x),$$

e a (nova) versão do subgradiente diferenciável é:

$$\nabla f_s(\tau, x) = \nabla f_c(\tau, x) + \bar{g} = \phi'(\tau, f_b(x)) (\nabla f(x) - \bar{g}) + \bar{g},$$

ou seja:

$$\nabla f_s(\tau, x) = [1 - \phi'(\tau, f_b(x))] \bar{g} + \phi'(\tau, f_b(x)) \nabla f(x).$$

Já sabemos que $\phi'(\tau, t)$ se comporta como o coeficiente de uma combinação convexa e portanto, para simplificar os desenvolvimentos, iremos denominar $\alpha(x) = \phi'(\tau, f_b(x))$, onde supomos que τ esteja fixo e implícito. Assim podemos escrever de forma mais clara a combinação convexa:

$$\nabla f_s(\tau, x) = (1 - \alpha(x)) \bar{g} + \alpha(x) \nabla f(x).$$

¹Notar que f_b , à semelhança de f_d , tem o próprio ponto z como referência para seu ponto de não diferenciabilidade, diferentemente de f_p , que tem a origem como referência para seu ponto de não diferenciabilidade.

Resta mostrarmos que $\nabla f_s(\tau, x)$ é um ε -subgradiente de $f(x)$ no ponto $x \in \mathbb{R}^n$. Para $x = z$ temos $\nabla f_s(\tau, z) = \bar{g}$ e como $\bar{g} \in \partial f(z)$ temos que $\nabla f_s(\tau, z) \in \partial_\varepsilon f(z)$, para qualquer $\varepsilon \geq 0$.

Para $x \neq z$, lembremos primeiramente que pela convexidade de f :

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x). \quad (8.2)$$

Temos também a seguinte série de desigualdades:

$$\begin{aligned} f(y) &\geq f(z) + \bar{g} \cdot (y - z), \\ f(y) &\geq f(x) + \bar{g} \cdot (y - x + x - z) - f(x) + f(z), \\ f(y) &\geq f(x) + \bar{g} \cdot (y - x) - (f(x) - f(z) - \bar{g} \cdot (x - z)), \\ f(y) &\geq f(x) + \bar{g} \cdot (y - x) - f_b(x). \end{aligned} \quad (8.3)$$

Se agora multiplicarmos a desigualdade (8.2) por $\alpha(x)$ e a desigualdade (8.3) por $(1 - \alpha(x))$ e somarmos teremos:

$$f(y) \geq f(x) + [(1 - \alpha(x))\bar{g} + \alpha(x)\nabla f(x)] \cdot (y - x) - (1 - \alpha(x))f_b(x),$$

que pode ser reescrito como:

$$f(y) \geq f(x) + \nabla f_s(\tau, x) \cdot (y - x) - \varepsilon_b(x), \quad (8.4)$$

onde $\varepsilon_b(x) = (1 - \alpha(x))f_b(x)$. Como $1 - \alpha(x) \geq 0$ e $f_b(x) \geq 0$ sempre teremos $\varepsilon_b(x) \geq 0$. Portanto para $\varepsilon = \varepsilon_b(x)$ é válida a pertinência $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$. Estamos no entanto assumindo que $\varepsilon \geq 0$ é fornecido e por seu intermédio queremos calcular o τ adequado para garantir que $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$.

Já vimos que para $x = z$ sempre teremos $\nabla f_s(\tau, z) = \bar{g} \in \partial_\varepsilon f(z)$, para qualquer $\varepsilon \geq 0$ e portanto neste caso qualquer $\tau \geq 0$ serve².

Suponhamos agora que $x \neq z$. Para $\varepsilon \geq f_b(x) \geq (1 - \alpha(x))f_b(x) = \varepsilon_b(x)$ podemos utilizar (8.4) para escrever

$$f(y) \geq f(x) + \nabla f_s(\tau, x) \cdot (y - x) - \varepsilon_b(x) \geq f(x) + \nabla f_s(\tau, x) \cdot (y - x) - \varepsilon,$$

e segue imediatamente que $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$. Portanto qualquer $\tau \geq 0$ é adequado para esta situação.

Por outro lado, se tivermos $0 \leq \varepsilon < f_b(x)$ basta forçarmos a validade da desi-

²Lembrar que em um problema de minimização a possibilidade de escolher $\bar{g} = 0$ implica em termos z como minimizador global.

gualdade

$$\varepsilon_b(x) = (1 - \alpha(x))f_b(x) = [1 - \phi'(\tau, f_b(x))] f_b(x) \leq \varepsilon$$

e desta forma encontrarmos alguma restrição sobre τ . Utilizando a formulação hiperbólica de $\phi'(\tau, t)$ temos:

$$\left[1 - \frac{f_b(x)}{\sqrt{\tau^2 + f_b(x)^2}} \right] f_b(x) \leq \varepsilon.$$

Mas a expressão à esquerda pode ser simplificada utilizando a função $\eta(\tau, s)$:

$$\eta(\tau, f_b(x)) \leq \varepsilon.$$

Como o máximo da função $\eta(\tau, s)$ é igual a $\varphi^{-\frac{5}{2}}\tau$, podemos estabelecer a seguinte desigualdade, à semelhança do que foi feito na seção 7.3:

$$\tau \leq \varphi^{\frac{5}{2}}\varepsilon = \sqrt{\varphi^5}\varepsilon, \quad (8.5)$$

e com ela garantimos a pertinência $\nabla f_s(\tau, z) \in \partial_\varepsilon f(z)$. Como para a situação $x = z$ e para a situação $x \neq z$ com $\varepsilon \geq f_b(x)$ não obtivemos nenhuma restrição sobre τ , podemos adotar a restrição (8.5) como sendo a *única* restrição a ser aplicada sobre τ , e que permite atender assim a todas as situações analisadas acima.

A Figura 8.1 ilustra como $\varepsilon_b(x)$, para um dado x , permite definir $\varepsilon_b(x)$ -subgradients em uma situação hipotética no \mathbb{R}^1 .

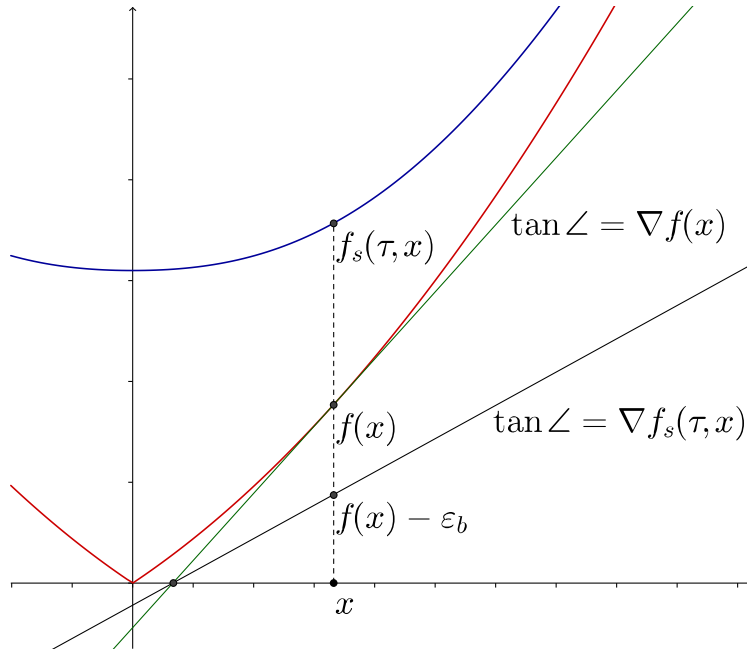


Figura 8.1: Gráfico que ilustra $\varepsilon_b(x)$ -subgradients para um dado $x \in \mathbb{R}^n$

Com o exposto acima provamos que nos casos em que a suavização SDHA não

pode ser utilizada porque $f_d(x)$ não é convexa, é possível utilizarmos a suavização SDA adaptada para $Z_f = \{z\}$ e seu subgradiente diferenciável associado $\nabla f_s(\tau, x)$ para acharmos um τ , atendendo a restrição (8.5), para o qual é válida a pertinência $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$. Este é um resultado bastante notório e interessante.

8.2 Comparativo das suavizações SDHA e SDA

Na seção anterior demonstramos a pertinência $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$ do subgradiente diferenciável gerado pela suavização SDA, sem impor a necessidade de $f_d(x)$ ser convexa. No entanto, para conseguirmos isto tivemos que alterar a definição do subgradiente diferenciável que desenvolvemos para a suavização SDHA. É interessante então compararmos estas duas situações, conforme apresentado nas Tabelas 8.1 e 8.2.

Deve ser observado que no início dos trabalhos relacionados com esta tese a utilização de assíntotas para a análise inicial da suavização hiperbólica de funções não diferenciáveis definidas no domínio real motivou a decomposição DHA aqui apresentada. A simplicidade analítica da função f_p atraiu esforços no sentido de suavizá-la e recompor uma versão suavizada da função original. A convexidade da função f_d proporcionou então uma condição suficiente para que uma desigualdade entre τ e ε pudesse garantir a pertinência $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$.

No entanto, os esforços empenhados no sentido de provar a existência de uma desigualdade similar para o caso de f_d não convexa foram infrutíferos. Buscou-se então verificar se a decomposição DA poderia ser utilizada em uma situação mais geral, para a qual não fosse necessária a imposição de nenhuma restrição técnica especial. Desta forma, quando mencionamos "contexto SDA", estamos na realidade mencionando uma situação mais geral, pois os desenvolvimentos realizados neste contexto são também válidos para quando f_d for convexa.

Recomendamos portanto a utilização da suavização SDA para os casos gerais, por não necessitar de verificações quanto à convexidade de f_d . Para alguns casos particulares em que o subdiferencial $\partial f(z)$ está disponível, e f_d é reconhecidamente convexa, a decomposição do contexto "SDHA (f_d convexa)" pode também ser utilizada.

Tabela 8.1: Contexto SDHA (f_d convexa)

Definições	Formulações
Decomposição	$f(x) = f_d(x) + f_p(x - z) + H(x)$
Suav. $f_p(x)$	$f_q(\tau, h) = \phi(\tau, f_p(h))$
Suav. $f(x)$	$f_s(\tau, x) = f_d(x) + f_q(\tau, x - z) + H(x)$
F. Ajuste	$f_a(\tau, x) = f_s(\tau, x) - f(x) = f_q(\tau, x - z) - f_p(x - z)$
S.G. Dif.	$\nabla f_s(\tau, x) = \nabla f_d(x) + \phi'(\tau, f_p(x - z))\nabla f_p(x - z) + \bar{g}$
Suav. Hip. $f_p(x)$	$f_q(\tau, h) = \sqrt{\tau^2 + f_p(x - z)^2}$
Suav. Hip. $f(x)$	$f_s(\tau, x) = f_d(x) + \sqrt{\tau^2 + f_p(x - z)^2} + H(x)$
F. Ajuste Hip.	$f_a(\tau, x) = f_s(\tau, x) - f(x) = \sqrt{\tau^2 + f_p(x - z)^2} - f_p(x - z)$
S.G. Dif. Hip.	$\nabla f_s(\tau, x) = \nabla f_d(x) + \frac{f_p(x - z)}{\sqrt{\tau^2 + f_p(x - z)^2}}\nabla f_p(x - z) + \bar{g}$
Restr. τ Hip.	$\tau \leq \sqrt{\varphi^5 \varepsilon}$

Tabela 8.2: Contexto SDA

Definições	Formulações
Decomposição	$f(x) = f_b(x) + H(x)$
Suav. $f_b(x)$	$f_c(\tau, x) = \phi(\tau, f_b(x))$
Suav. $f(x)$	$f_s(\tau, x) = f_c(\tau, x) + H(x)$
F. Ajuste	$f_a(\tau, x) = f_s(\tau, x) - f(x) = f_c(\tau, x) - f_b(x)$
S.G. Dif.	$\nabla f_s(\tau, x) = (1 - \alpha(x))\bar{g} + \alpha(x)\nabla f(x), \alpha(x) = \phi'(\tau, f_b(x))$
Suav. Hip. $f_b(x)$	$f_c(\tau, x) = \sqrt{\tau^2 + f_b(x)^2}$
Suav. Hip. $f(x)$	$f_s(\tau, x) = \sqrt{\tau^2 + f_b(x)^2} + H(x)$
F. Ajuste Hip.	$f_a(\tau, x) = f_s(\tau, x) - f(x) = \sqrt{\tau^2 + f_b(x)^2} - f_b(x)$
S.G. Dif. Hip.	$\nabla f_s(\tau, x) = (1 - \alpha(x))\bar{g} + \alpha(x)\nabla f(x), \alpha(x) = \frac{f_b(x)}{\sqrt{\tau^2 + f_b(x)^2}}$
Restr. τ Hip.	$\tau \leq \sqrt{\varphi^5 \varepsilon}$

Para exemplificarmos, tomemos a função $f(x) = x^2 + 2|x|$, com domínio real, baseada na função discutida na seção 6.3, que provamos possuir f_d convexa³. Temos $z = 0$ e tomando-se $\bar{g} = 0$ podemos facilmente identificar as funções:

$$\begin{aligned}
f_d(x) &= x^2, \\
f_p(x - 0) &= 2|x|, \\
f_b(x - 0) &= x^2 + 2|x|, \\
H(x) &= 0.
\end{aligned}$$

³Foi realizado um deslocamento para a origem visando aprimorar a apresentação gráfica.

Com base nestas componentes, temos a versão SDHA (f_d convexa) de $f(x)$ com suavização hiperbólica dada por:

$$f_{s,1}(x) = f_d(x) + f_q(\tau, x - 0) + H(x) = x^2 + \sqrt{\tau^2 + (2|x|)^2},$$

enquanto que a versão SDA de $f(x)$ com suavização hiperbólica é dada por:

$$f_{s,2}(x) = f_c(\tau, x - 0) + H(x) = \sqrt{\tau^2 + (x^2 + 2|x|)^2}.$$

As duas versões suavizadas de $f(x)$ podem ser comparadas na Figura 8.2. Podemos perceber que para este caso elas estão bastante próximas.

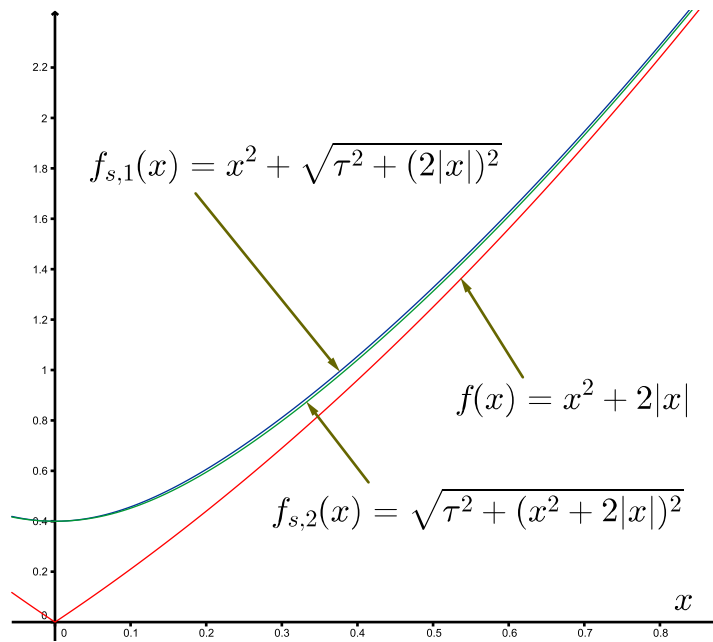


Figura 8.2: Gráfico das versões suavizadas de $f(x) = x^2 + 2|x|$

8.3 SDA na geração de ε -subgradientes para funções ZH convexas

Até este momento estávamos trabalhando com $Z_f = \{z\}$, que é uma função ZH convexa muito especial. No entanto, algumas observações irão demonstrar que é possível ampliarmos a geração de ε -subgradientes para casos mais gerais.

Primeiramente devemos observar que um Hiperplano de suporte ao epígrafo de uma função ZH convexa f pode ser representado pela função afim $H(x) = \bar{g} \cdot x + b$, com $\bar{g}, b \in \mathbb{R}^n$. Como $H(x)$ representa um hiperplano de suporte, temos o seguinte

desenvolvimento de desigualdades:

$$\begin{aligned}
f(y) &\geq H(y), \\
f(y) &\geq \bar{g}.y + b, \\
f(y) &\geq f(x) + \bar{g}.(y - x + x) - f(x) + b, \\
f(y) &\geq f(x) + \bar{g}.(y - x) - (f(x) - \bar{g}.x - b), \\
f(y) &\geq f(x) + \bar{g}.(y - x) - (f(x) - H(x)), \\
f(y) &\geq f(x) + \bar{g}.(y - x) - f_b(x).
\end{aligned}$$

Esta desigualdade final é a mesma apresentada em (8.3), a qual foi desenvolvida para $Z_f = \{z\}$, e que agora provamos ser também válida no contexto de qualquer função ZH convexa. Como a desigualdade (8.2) é sempre válida, podemos realizar o mesmo desenvolvimento apresentado na seção 8.1, incluindo a utilização da formulação hiperbólica de $\phi'(\tau, t)$, para concluir que para qualquer função ZH convexa, dado um $\varepsilon > 0$, podemos gerar um subgradiente diferenciável $\nabla f_s(\tau, x) \in \partial_\varepsilon f(x)$ se for respeitada a restrição

$$\tau \leq \varphi^{\frac{5}{2}}\varepsilon = \sqrt{\varphi^5}\varepsilon.$$

Se considerarmos que as funções ZH convexas são uma generalização natural das funções que possuem não diferenciabilidade localizada em um ponto isolado, o resultado acima parece fazer bastante sentido. Assim, as formulações para a suavização SDA e geração de ε -subgradientes apresentadas na Tabela 8.2 também podem ser utilizadas para o caso de funções ZH convexas.

Capítulo 9

Resultados e Discussões

Os desenvolvimentos apresentados no corpo desta tese são de cunho eminentemente teórico e portanto a realização de testes computacionais não faz parte dos objetivos almejados. Neste capítulo iremos portanto apresentar alguns resultados e discussões referentes ao contexto destes desenvolvimentos em comparação com o que está disponível na literatura.

9.1 A suavização hiperbólica clássica

As expressões desenvolvidas até o momento não parecem a princípio serem idênticas às utilizadas nos trabalhos do Prof. A. E. Xavier, idealizador da Penalização e Suavização Hiperbólicas. Mas como o objetivo desta tese foi justamente estender a aplicação da Suavização Hiperbólica diretamente para dimensões maiores e regiões de não diferenciabilidade mais complexas, é de se esperar que a Suavização Hiperbólica "clássica" seja um caso especial da metodologia aqui desenvolvida.

Desta forma, iremos demonstrar que a função $\varphi(y) = \max\{0, y\}$ utilizada na abordagem clássica possui essencialmente a mesma suavização utilizando qualquer uma das abordagens. Na abordagem clássica a função $\varphi(y)$ dá origem à seguinte função suavizada:

$$\phi_H(y, \tau) = \frac{y + \sqrt{y^2 + \tau^2}}{2}.$$

Notar que, para não causar confusão com nossa definição de função ϕ , estamos utilizando a denominação ϕ_H para representar a função suavizada " ϕ " presente na literatura da Penalização e Suavização Hiperbólicas.

Utilizando o processo de decomposição DHA apresentado nesta tese temos que a expressão suavizada para $\varphi(y)$ pode ser calculada tomando $\bar{g} = \frac{1}{2}$, $f_p(y - 0) =$

$f_p(y) = \frac{|y|}{2}$ e $f_d(y) = 0$. Assim:

$$\phi(y, \tau) = 0 + \sqrt{\left(\frac{|y|}{2}\right)^2 + \tau^2} + \frac{1}{2}y = \frac{y + \sqrt{y^2 + 4\tau^2}}{2}.$$

Esta é uma expressão praticamente idêntica à utilizada na Suavização Hiperbólica clássica, a menos do fator 4, que é imaterial.

9.2 Suavização de algumas funções

Nesta seção apresentaremos algumas funções ZH convexas e suas respectivas suavizações hiperbólicas, com o intuito de demonstrar a abrangência da metodologia aqui apresentada.

9.2.1 Função norma euclidiana

Esta função, já comentada anteriormente, é um dos exemplos mais expressivos, tanto por sua simplicidade como por sua ubiqüidade. Tomando-se $f(x) = \|x\|$ podemos trivialmente fixar $z = 0$, $\bar{g} = 0$, $f_d(x) = 0$ e $f_p(x) = \|x\|$, ou seja, temos somente a componente $f_p(x)$ que é a própria função $\|x\|$. A suavização hiperbólica é dada por:

$$f_s(\tau, x) = f_q(\tau, x) = \sqrt{\|x\|^2 + \tau^2}.$$

Notar que, como $f_d(x) = 0$, temos $f_b(x) = f_p(x)$ e as suavizações SDHA e SDA são idênticas.

9.2.2 Função $f_1(x)$ definida na seção 3.2

A função $f_1(x) = \max\{0, \|x\| - 1\}$ possui a região de não diferenciabilidade dada por $Z_{f_1} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$, constituída por mais de um ponto. Utilizaremos então a suavização SDA. Temos obviamente $H(x) = 0$, ou seja, o hiperplano horizontal que passa pela origem é um hiperplano de suporte para o epígrafo de $f_1(x)$, e portanto $f_b(x) = f_1(x)$. Logo temos simplesmente:

$$f_s(\tau, x) = \sqrt{f_1(x)^2 + \tau^2}.$$

Se compararmos este exemplo com o anterior notamos que os formatos finais das suavizações são:

$$f_s(\tau, x) = \sqrt{f(x)^2 + \tau^2}.$$

Isto sempre acontecerá quando $H(x) \equiv 0$ na SDA (ou quando $f_d(x) \equiv 0$ e $H(x) \equiv 0$ na SDHA).

9.2.3 Função exemplo de Shor

Iremos aqui tratar da suavização de uma função mais complexa. Trata-se da função com domínio no \mathbb{R}^2 apresentada no início do capítulo 2 de (SHOR, 1985), e definida como:

$$f(x) = \begin{cases} 5\sqrt{9x_1^2 + 16x_2^2} & , x_1 > |x_2|, \\ 9x_1 + 16|x_2| & , x_1 \leq |x_2|. \end{cases}$$

Shor demonstra que esta função é diferenciável no semi-plano definido por $x_1 > 0$, ou seja, apesar da aparente distinção da função nas duas regiões dadas por $x_1 > |x_2|$ e $x_1 \leq |x_2|$, a função é contínua e diferenciável nas semi-retas $x_1 = |x_2|$, para $x_1 > 0$. Por outro lado, podemos facilmente perceber que a zona de não diferenciabilidade de f é dada por $Z_f = \{(x_1, 0) : x_1 < 0\}$.

Utilizaremos mais uma vez a suavização SDA, pois a região de não diferenciabilidade é constituída por mais de um ponto¹. Este exemplo é não-trivial pois $H(x)$ não é identicamente nula, e sim dada por $H(x) = 9x_1$. Isto pode ser comprovado facilmente se notarmos que $G(Z_f) = \{(x, f(x)) : x \in Z_f\} = \{(x_1, 0, 9x_1) : x_1 < 0\} = \{(x, H(x)) : x \in Z_f\}$. Assim:

$$f_s(\tau, x) = \sqrt{(f(x) - 9x_1)^2 + \tau^2} + 9x_1.$$

9.2.4 Função "unha" definida na seção A.1

A função "unha" definida por

$$f(x) = f(x_1, x_2) = \sqrt{x_1^2 + \frac{x_2^4}{4}} + \frac{x_2^2}{2}$$

é apresentada na seção A.1. Trata-se de uma função com um único ponto de não diferenciabilidade dado por $z = (0, 0)$. A análise da Hessiana desta função demonstra que ela é convexa e o cálculo do subdiferencial no ponto zero resulta em $\partial f(0) = \{(t, 0) \in \mathbb{R}^2 : -1 \leq t \leq 1\}$. Com base nestas informações podemos então definir:

$$\begin{aligned} H(x) &= 0, \\ f_p(x) &= |x_1|, \\ f_d(x) &= f(x) - |x_1|, \\ f_b(x) &= f(x). \end{aligned}$$

¹Este é um caso de uma região de não diferenciabilidade não compacta.

Podemos facilmente perceber que a desigualdade

$$f_d\left(\frac{(-2, 2)}{2} + \frac{(2, 2)}{2}\right) = f_d(0, 2) = 4 > \frac{f_d(-2, 2)}{2} + \frac{f_d(2, 2)}{2} = 2\sqrt{2} \approx 2.83$$

implica na não convexidade de $f_d(x)$, e portanto não podemos utilizar a suavização SDHA neste caso, se quisermos gerar ε -subgradientes para $f(x)$. Podemos no entanto utilizar livremente a suavização SDA, por intermédio da qual chegamos a

$$f_s(\tau, x) = \sqrt{\left(\sqrt{x_1^2 + \frac{x_2^4}{4} + \frac{x_2^2}{2}}\right)^2 + \tau^2}.$$

9.2.5 Função $f_2(x)$ definida na seção 3.2

Já vimos que a função definida por

$$f_2(x) = \max\{|x_1|, |x_2|\}$$

possui uma região de não diferenciabilidade definida por $Z_{f_2} = \{(x_1, x_2) \in \mathbb{R}^2 : |x_1| = |x_2|\}$, e que $G(Z_{f_2})$ não pode ser contida em um hiperplano de suporte. No entanto, podemos reescrever $f_2(x)$ como

$$f_2(x) = \frac{|x_1 + x_2|}{2} + \frac{|x_1 - x_2|}{2} = f_2^+(x) + f_2^-(x).$$

Temos agora que f_2^+ e f_2^- são funções ZH convexas com regiões de não diferenciabilidade dadas por $Z_{f_2^+} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = -x_2\}$ e $Z_{f_2^-} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = x_2\}$. Portanto podemos formar a suavização SDA de $f_2(x)$ como

$$f_s(\tau, x) = \sqrt{\frac{(x_1 + x_2)^2}{4} + \tau^2} + \sqrt{\frac{(x_1 - x_2)^2}{4} + \tau^2}.$$

9.3 Transformação de funções convexas gerais em funções ZH convexas

A teoria apresentada para a suavização de funções convexas não diferenciáveis se aplica especificamente para as funções ZH convexas. No entanto, pode ser que uma função convexa geral possa ser transformada (de forma diferenciável) em uma função ZH convexa, passando então pelo processo de suavização e, mediante transformada inversa, se converter em uma versão suavizada da função original.

Iremos demonstrar como isto pode ser feito com um exemplo bastante simples.

Seja a função convexa definida por

$$f(x) = \begin{cases} x_1^2 + x_2^2 & , \text{ se } x_1^2 + x_2^2 \leq 1, \\ 4(x_1^2 + x_2^2) - 3 & , \text{ caso contrário.} \end{cases}$$

Se tomarmos esta função e subtrairmos $g(x) = x_1^2 + x_2^2$ passaremos a ter uma função ZH convexa dada por

$$f(x) - g(x) = \begin{cases} 0 & , \text{ se } x_1^2 + x_2^2 \leq 1, \\ 3(x_1^2 + x_2^2) - 3 & , \text{ caso contrário.} \end{cases}$$

A Figura 9.1 apresenta os gráficos destas funções.

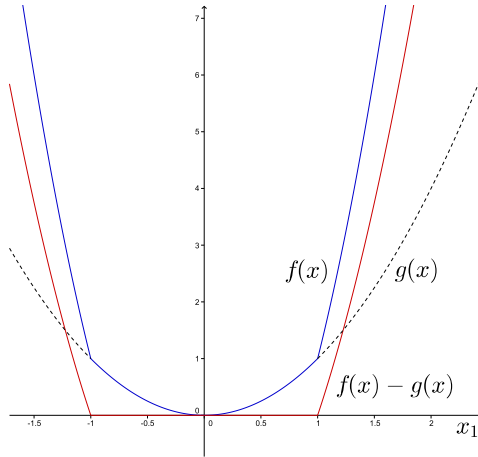


Figura 9.1: Secção de $f(x)$ e $f(x) - g(x)$ ao longo de x_1

Esta nova função pode então passar pela suavização SDA e voltando a somar $g(x)$ obteremos uma versão suavizada da função original. Temos então:

$$f_s(\tau, x) = \sqrt{(f(x) - g(x))^2 + \tau^2} + g(x).$$

Devemos perceber que não podemos afirmar nada com relação à geração de ε -subgradientes quando utilizamos transformações como a apresentada acima. O que estamos fazendo acima é na verdade uma generalização da etapa de subtração de $H(x)$ de $f(x)$, a qual tem por finalidade preparar uma função ZH convexa para a etapa de suavização SDA. Para as funções ZH convexas esta subtração simplifica $f(x)$, gerando $f_b(x)$, e os resultados quanto à geração de ε -subgradientes podem ser transportados linearmente para a função original. No caso de uma transformação diferenciável genérica não podemos afirmar nada quanto aos ε -subgradientes (e de uma forma geral, nem quanto à convexidade), mas sempre teremos como resultado final uma versão suavizada da função original.

Capítulo 10

Conclusões

O principal objetivo desta tese foi demonstrar a possibilidade de geração de ε -subgradientes para uma função convexa não diferenciável $f(x)$ tendo por base a sua suavização hiperbólica¹. Como a suavização hiperbólica clássica é feita para funções que possuem uma não diferenciabilidade isolada em um ponto, e que tem por domínio a reta real, alcançar este objetivo exigiu que algumas contextualizações fossem feitas, as quais resultaram na criação de uma série de conceitos, tais como a região de não diferenciabilidade Z_f , as funções ZH convexas, as decomposições DHA e DA, as suavizações SDHA e SDA, e as funções de suavização $\phi(\tau, t)$, tanto a hiperbólica como as versões alternativas.

Com a ajuda dos conceitos acima mencionados pudemos chegar ao resultado mais forte desta tese: para todo $\varepsilon > 0$ é possível gerar ε -subgradientes para uma função ZH convexa contanto que o parametro τ utilizado na suavização SDA Hiperbólica atenda à restrição $\tau \leq \sqrt{\varphi^5 \varepsilon}$.

Podemos então concluir que as metas inicialmente traçadas foram alcançadas, dentro do particular contexto apresentado nesta tese.

Quanto à continuação deste trabalho podemos divisar alguns objetivos interessantes, de uma forma geral relacionados com funções não convexas, dentre os quais podemos listar:

- Estender a definição de região de não diferenciabilidade para funções não convexas;
- Estender a definição de distância à região de não diferenciabilidade para funções não convexas;
- Desenvolver uma suavização hiperbólica no \mathbb{R}^n para funções não convexas;

¹Deve ficar claro que a suavização hiperbólica, em geral, não tem seu uso restrito às funções convexas.

- Compatibilizar os subgradientes diferenciáveis com o gradiente de Clarke, se isto for possível;
- Explorar as capacidades de convexificação presentes na suavização hiperbólica;
- Estudar as condições para gerar funções ZH convexas a partir de funções convexas gerais.

Espera-se que os desenvolvimentos realizados nesta tese sejam relevantes não somente para aplicações imediatas na suavização de funções convexas não diferenciáveis, como também estimulem novos desenvolvimentos partindo das idéias aqui apresentadas.

Referências Bibliográficas

- AUSLENDER, A., TEBoulLE, M., 2003, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. New York, Springer Verlag.
- BEN-TAL, A., TEBoulLE, M., 1988, “A smoothing technique for nondifferentiable optimization problems”. In: *Optimization - Fifth French-German Conference*, pp. 1–11, Castel-Novel, Out. 1988.
- BERTSEKAS, D., 1974, *Nondifferentiable Optimization Via Approximation*. 1 ed. Urbana, University of Illinois.
- BORWEIN, J. M., ZHU, Q. J., 1999, “A Survey of Subdifferential Calculus with Applications”, *Nonlinear Analysis: Theory, Methods & Applications*, v. 38, pp. 687 – 773.
- BOYD, S., VANDENBERGHE, L., 2009, *Convex Optimization*. 7 ed. Cambridge, Cambridge University Press.
- CLARKE, F., 1983, *Optimization and Nonsmooth Analysis*. 1 ed. New York, Wiley-Interscience.
- CLARKE, F., “Generalized gradients and applications”. In: *Trans. Amer. Math. Soc.*, v. 205, pp. 247–262, 1975.
- CONN, A. R., SCHEINBERG, K., VICENTE, L. N., 2008, *Introduction to Derivative-Free Optimization*. 1 ed. Philadelphia, SIAM.
- FENCHEL, M. W., 1953, *Convex Cones, Sets, and Functions*. Princeton, New Jersey, Princeton University.
- FENCHEL, M. W., BONNESEN, T., 1934, *Theorie der konvexen Körper*. Berlin, Verlag von Julius Springer.
- IZMAILOV, A., SOLODOV, M., 2007, *Otimização Volume 2*. 1 ed. Rio de Janeiro, IMPA.

- LEMARÉCHAL, C., “Nondifferentiable Optimization”. In: *Handbooks in Operations Research and Management Science*, v. 1, *Optimization*, Elsevier Science Publishers B.V., pp. 529–572, 1989.
- LIMA, E. L., 1981, *Curso de Análise Volume 2*. 1 ed. Rio de Janeiro, IMPA.
- LUENBERGER, D. G., YE, Y., 2008, *Linear and Nonlinear Programming*. 3 ed. New York, Springer Science+Business Media.
- MORÉ, J. J., WILD, S. M., 2007, *Benchmarking derivative-free optimization algorithms*. Relatório técnico, Tech. Report ANL/MCS-P1471-1207, Mathematics and Computer Science Division, Argonne National Laboratory, USA.
- MOREAU, J.-J., 1965, “Proximité et dualité dans un espace hilbertien”, *Bulletin de la Société Mathématique de France*, v. 93, pp. 273–299.
- PENOT, J.-P., 1992, “Topologies and convergences on the space of convex functions”, *Nonlinear Anal.*, v. 18, n. 10 (maio), pp. 905–916. ISSN: 0362-546X. doi: 10.1016/0362-546X(92)90128-2. Disponível em: [http://dx.doi.org/10.1016/0362-546X\(92\)90128-2](http://dx.doi.org/10.1016/0362-546X(92)90128-2).
- ROCKAFELLAR, R. T., 1970, *Convex Analysis*. 1 ed. Princeton, Princeton University Press.
- SHOR, N. Z., 1985, *Minimization Methods for Non-Differentiable functions*. 1 ed. New York, Springer-Verlag.
- XAVIER, A. E., 1982, *Penalização Hiperbólica: um novo método para resolução de problemas de otimização*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro.
- XAVIER, A. E., 1992, *Penalização Hiperbólica*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro.
- XAVIER, A. E., MAGALHÃES, P. C., DA SILVA, L. P., 1989, “Uma Técnica de Suavização de Funções Lineares por Parte Aplicada a Calibração de Modelos Chuva-Vazão”. In: *Anais do XXII Simpósio Brasileiro de Pesquisa Operacional*, pp. 166–174, Fortaleza, Out. 1989.
- XAVIER, A. E., 2001, “Hyperbolic penalty: a new method for nonlinear programming with inequalities”, *ITOR*, v. 8, pp. 659–671.
- XAVIER, A. E., 2009, “The hyperbolic smoothing clustering method”, *Pattern Recognition*, v. 43, pp. 731–737.

- XAVIER, A. E., 2011, “Solving the minimum sum-of-squares clustering problem by hyperbolic smoothing and partition into boundary and gravitational regions”, *Pattern Recognition*, v. 44, pp. 70–77.
- XAVIER, A. E., DE OLIVEIRA, A. A. F., 2005, “Optimal Covering of Plane Domains by Circles via Hyperbolic Smoothing”, *Journal of Global Optimization*, v. 31, pp. 493–504.
- XAVIER, A. E., MACULAN, N., 1995, *Hyperbolic Lagrangean: a new method of multipliers*. Relatório técnico, COPPE/UFRJ, Rio de Janeiro.

Apêndice A

Alguns resultados básicos importantes

Neste apêndice serão apresentados alguns resultados mencionados no texto, mas cujas demonstrações poderiam desviar o foco dos desenvolvimentos em curso.

A.1 A dimensão do subdiferencial

Seja uma função real convexa $f(x)$ com domínio igual ao \mathbb{R}^n , com um ponto isolado de não diferenciabilidade z . Uma propriedade do subdiferencial $\partial f(z)$ que pode a princípio parecer não natural, mas que é considerada neste trabalho, é que $\dim\{\partial f(z)\} \leq n$. Como z é um ponto isolado de não diferenciabilidade, poderíamos ser levados a pensar que sempre teríamos $\dim\{\partial f(z)\} = n$. Um exemplo simples nos convence do contrário.

Seja a função convexa com domínio no \mathbb{R}^2 definida por

$$f(x) = f(x_1, x_2) = \sqrt{x_1^2 + \frac{x_2^4}{4}} + \frac{x_2^2}{2},$$

a qual pode ser interpretada como sendo a função módulo $f(x_1, 0) = |x_1|$ na direção do eixo x_1 e como a função parabólica $f(0, x_2) = x_2^2$ na direção do eixo x_2 . A Figura A.1 mostra o gráfico de $f(x)$ visto de duas posições diferentes, o que permite uma melhor visualização do comportamento desta função. A Figura A.2 apresenta as secções do gráfico desta função nos planos x_1z e x_2z , visando deixar bastante clara a não diferenciabilidade na origem. Devido ao seu aspecto geométrico, chamaremos esta função de função "unha".

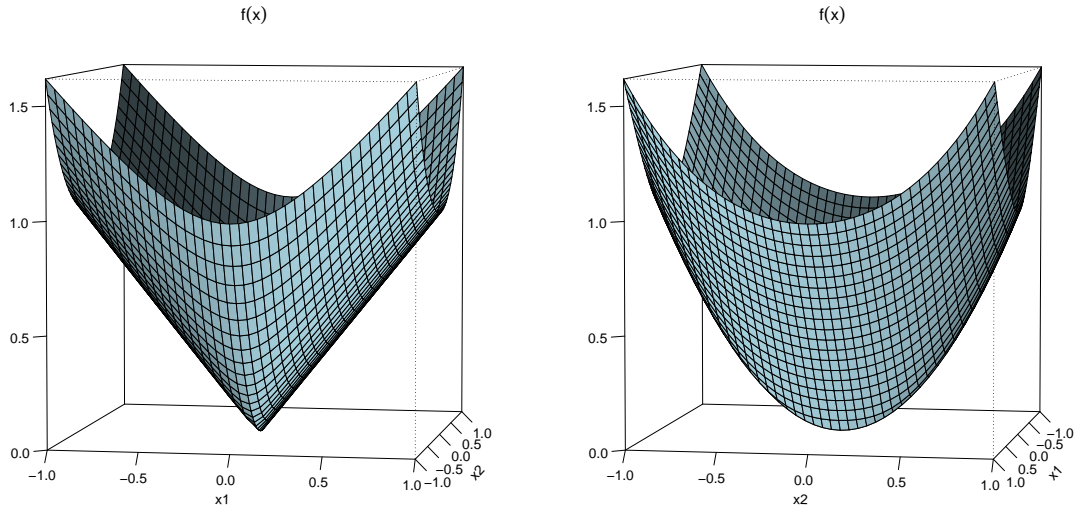


Figura A.1: Visões distintas do gráfico da função $f(x_1, x_2) = \sqrt{x_1^2 + \frac{x_2^4}{4}} + \frac{x_2^2}{2}$

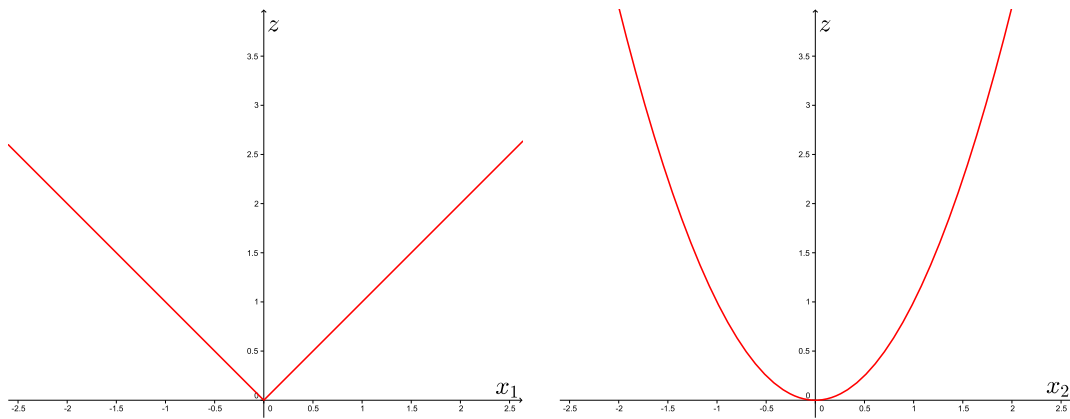


Figura A.2: Secções do gráfico da função $f(x_1, x_2) = \sqrt{x_1^2 + \frac{x_2^4}{4}} + \frac{x_2^2}{2}$

Os componentes do gradiente são:

$$\frac{\partial f}{\partial x_1} = \frac{x_1}{\sqrt{x_1^2 + \frac{x_2^4}{4}}},$$

$$\frac{\partial f}{\partial x_2} = \frac{x_2^3}{2\sqrt{x_1^2 + \frac{x_2^4}{4}}} + x_2,$$

ou seja, $f(x)$ é claramente diferenciável fora da origem. No entanto, apesar de termos $f(x)$ bem definida na origem, assumindo o valor zero, o gradiente não está definido na origem, mas a derivada direcional $f'(0, d)$ pode ser calculada:

$$\lim_{t \rightarrow 0^+} \frac{f(td) - f(0)}{t} = \lim_{t \rightarrow 0^+} \frac{\sqrt{t^2 d_1^2 + \frac{t^4 d_2^4}{4}} + \frac{t^2 d_2^2}{2}}{t} = \lim_{t \rightarrow 0^+} \sqrt{d_1^2 + \frac{t^2 d_2^4}{4}} + \frac{t d_2^2}{2} = |d_1|.$$

Se $d_1 > 0$ então $f'(0, d) = d_1 = d \cdot e_1$, onde e_1 é o vetor unitário correspondente à coordenada x_1 . Se por outro lado $d_1 < 0$ então $f'(0, d) = |d_1| = -d_1 = d \cdot (-e_1)$. Como para f convexa sempre vale que $f'(x, d) = \max_{g \in \partial f(x)} d \cdot g$, para o nosso caso particular vale que $f'(0, d) = |d_1| = \max\{d \cdot e_1, d \cdot (-e_1)\}$, e portanto e_1 e $-e_1$ pertencem a $\partial f(0)$, e bastam para caracterizá-lo. Mas como o subdiferencial é sempre convexo, temos então que $\partial f(0) = \{(t, 0) \in \mathbb{R}^2 : -1 \leq t \leq 1\}$, o qual tem obviamente dimensão um, apesar do domínio de $f(x)$ ser \mathbb{R}^2 e z ser um ponto isolado de não diferenciabilidade.

Este exemplo pode ser facilmente estendido para dimensões superiores a dois, o que nos mostra que de uma forma geral vale que $\dim\{\partial f(z)\} \leq n$. Sendo assim, quando falarmos do "interior" de $\partial f(z)$ o correto será nos referirmos ao "interior relativo" de $\partial f(z)$, $ri\{\partial f(z)\}$, definido como o interior de $\partial f(z)$ relativo à sua envoltória afim.

A.2 O subdiferencial $\partial_\infty f$

Ainda considerando $f : \mathbb{R}^n \rightarrow \mathbb{R}$ como uma função convexa, iremos agora definir o ∞ -subdiferencial de f , representado por $\partial_\infty f$, um conceito importante dentro do contexto da exploração dos ε -subgradientes. Iremos demonstrar então que $\partial_\infty f$ é na verdade uma outra representação para $dom f^*$. Nesta seção trabalharemos com um $dom f$ geral, mas devemos lembrar que nos desenvolvimentos do corpo deste trabalho utilizamos $dom f = \mathbb{R}^n$.

Proposição A.1. *Para qualquer $g \in dom f^*$ e qualquer $x \in dom f$ existe um $\varepsilon \geq 0$ tal que $g \in \partial_\varepsilon f(x)$.*

Demonstração. Sejam $g \in dom f^*$ e $x \in dom f$ arbitrários, o que implica em termos $f(x) < +\infty$ e $f^*(g) < +\infty$. Pela desigualdade de Fenchel (2.19) temos:

$$f^*(g) + f(x) \geq g \cdot x,$$

que pode ser rearranjado como $f^*(g) + f(x) - g \cdot x \geq 0$. Se definirmos $\varepsilon = f^*(g) + f(x) - g \cdot x$, as desigualdades $f(x) < +\infty$ e $f^*(g) < +\infty$ implicam em termos $\varepsilon < +\infty$, e portanto ε é um número real (não estendido). Para este ε podemos escrever $f^*(g) = -f(x) + g \cdot x + \varepsilon$ e temos então, pela definição de f^* :

$$-f(x) + g \cdot x + \varepsilon = f^*(g) = \sup_{y \in \mathbb{R}^n} \{g \cdot y - f(y)\},$$

e portanto $-f(x) + g \cdot x + \varepsilon \geq g \cdot y - f(y), \forall y \in \mathbb{R}^n$. Assim:

$$f(y) \geq f(x) + g \cdot (y - x) - \varepsilon,$$

e isto significa que $g \in \partial_\varepsilon f(x)$. □

A Figura A.3 ilustra no \mathbb{R}^1 um exemplo de situação onde, dados $g \in \text{dom } f^*$ e $x \in \text{dom } f$ arbitrários, se aplica a proposição A.1 para se achar o ε visualizado no gráfico.

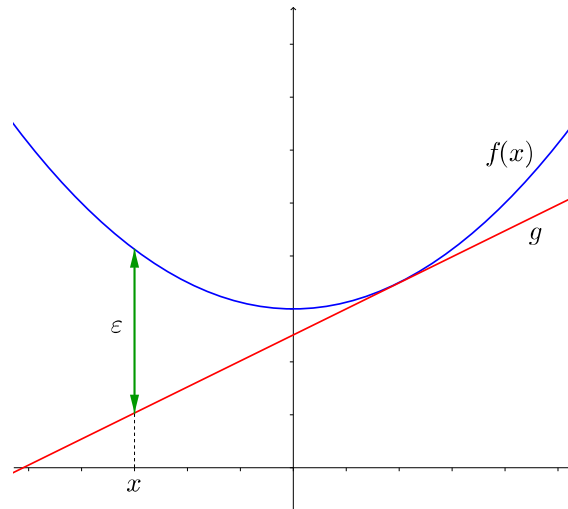


Figura A.3: Exemplo gráfico de aplicação da proposição A.1

Uma proposição similar à esta pode ser enunciada, relacionada com a interseção de dois subdiferenciais.

Proposição A.2. *Dado um $g \in \partial_{\varepsilon_1} f(x_1)$ qualquer, para $x_1 \in \text{dom } f$ e $\varepsilon_1 \geq 0$, e $x_2 \in \text{dom } f$ qualquer, existe $\varepsilon_2 \geq 0$ tal que $g \in \partial_{\varepsilon_2} f(x_2)$, ou seja, $g \in \partial_{\varepsilon_1} f(x_1) \cap \partial_{\varepsilon_2} f(x_2)$.*

Demonstração. Sejam $g \in \partial_{\varepsilon_1} f(x_1)$ e $x_2 \in \text{dom } f$ quaisquer, e por definição temos $f(x_2) < +\infty$. Pela definição de $\partial_{\varepsilon_1} f(x_1)$ temos:

$$f(y) \geq f(x_1) + g \cdot (y - x_1) - \varepsilon_1,$$

para qualquer $y \in \mathbb{R}^n$. Podemos reescrever esta desigualdade como

$$\begin{aligned} f(y) &\geq f(x_2) + g \cdot (y - x_2 + x_2 - x_1) + f(x_1) - f(x_2) - \varepsilon_1, \\ f(y) &\geq f(x_2) + g \cdot (y - x_2) - (f(x_2) - f(x_1) - g \cdot (x_2 - x_1) + \varepsilon_1). \end{aligned}$$

mas como $g \in \partial_{\varepsilon_1} f(x_1)$ temos que $f(x_2) - f(x_1) - g \cdot (x_2 - x_1) + \varepsilon_1 \geq 0$. Denominando $\varepsilon_2 = f(x_2) - f(x_1) - g \cdot (x_2 - x_1) + \varepsilon_1$ temos então

$$f(y) \geq f(x_2) + g \cdot (y - x_2) - \varepsilon_2,$$

e isto significa que $g \in \partial_{\varepsilon_2} f(x_2)$. □

A Figura A.4 ilustra no \mathbb{R}^1 um exemplo de situação onde, dados $g \in \partial_{\varepsilon_1} f(x_1)$ e $x_2 \in \text{dom } f$ arbitrários, se aplica a proposição A.2 para se achar o ε_2 visualizado no gráfico.

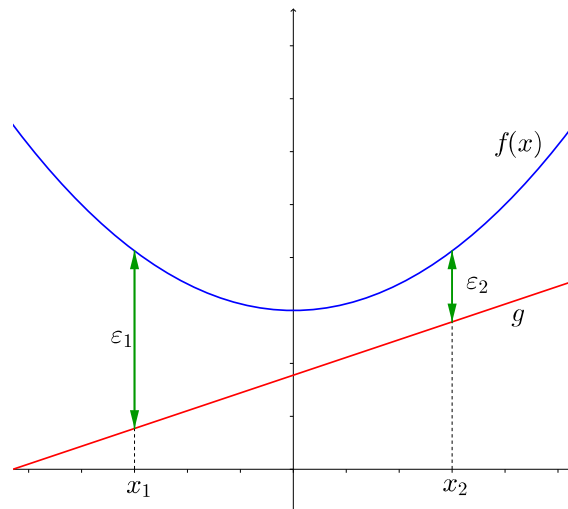


Figura A.4: Exemplo gráfico de aplicação da proposição A.2

A proposição A.1 indica que, dado um $g \in \text{dom } f^*$ qualquer e um $x \in \text{dom } f$ qualquer, é sempre possível conseguirmos um $\varepsilon \geq 0$ tal que possamos assegurar a pertinência $g \in \partial_{\varepsilon} f(x)$, e este fato nos faz perguntar o que acontece com $\partial_{\varepsilon} f(x)$ quando $\varepsilon \rightarrow \infty$. Este questionamento dá origem à nossa proposição seguinte.

Proposição A.3. *O ∞ -subdiferencial de f , definido como o limite quando $\varepsilon \rightarrow \infty$ do conjunto $\partial_{\varepsilon} f(x)$ para um $x \in \text{dom } f$ arbitrário, e representado por $\partial_{\infty} f$, é na verdade uma outra forma de expressar $\text{dom } f^*$.*

Demonstração. Queremos provar duas coisas: que $\partial_{\infty} f = \lim_{\varepsilon \rightarrow \infty} \partial_{\varepsilon} f(x)$ é único, independentemente do $x \in \text{dom } f$ escolhido para o limite, e que $\partial_{\infty} f = \text{dom } f^*$.

Já sabemos que se $0 \leq \varepsilon_{\alpha} < \varepsilon_{\beta}$ então, para um $x \in \text{dom } f$ arbitrário temos que $\partial_{\varepsilon_{\alpha}} f(x) \subseteq \partial_{\varepsilon_{\beta}} f(x) \subseteq \mathbb{R}^n$. Como a sequência de conjuntos $\partial_{\varepsilon} f(x)$ é monótona não decrescente com relação à ordem imposta pela inclusão e a sequência é limitada por \mathbb{R}^n , que é um conjunto fechado, temos que $\partial_{\infty} f = \lim_{\varepsilon \rightarrow \infty} \partial_{\varepsilon} f(x)$ está bem definido.

Seja $g \in \partial_{\infty} f(x_1) = \lim_{\varepsilon \rightarrow \infty} \partial_{\varepsilon} f(x_1)$ para algum $x_1 \in \text{dom } f$. Logo, pela definição de $\partial_{\infty} f(x_1)$ existe $\varepsilon_1 \geq 0$ tal que $g \in \partial_{\varepsilon_1} f(x_1)$. Seja agora um $x_2 \in \text{dom } f$

qualquer. Pela proposição A.2 existe $\varepsilon_2 \geq 0$ tal que $g \in \partial_{\varepsilon_2} f(x_2) \subseteq \partial_{\infty} f(x_2)$. Logo $\partial_{\infty} f(x_1)$ não depende do particular $x_1 \in \text{dom } f$ escolhido e pode ser representado simplesmente por $\partial_{\infty} f$.

Para provarmos que $\partial_{\infty} f = \text{dom } f^*$ consideremos primeiramente $g \in \text{dom } f^*$. Tomemos um $x \in \text{dom } f$ qualquer e utilizando a proposição A.1 temos que existe um $\varepsilon \geq 0$ tal que $g \in \partial_{\varepsilon} f(x) \subseteq \partial_{\infty} f(x) = \partial_{\infty} f$. Logo $g \in \partial_{\infty} f$.

Seja agora $g \in \partial_{\infty} f$. Por definição, dado um $x \in \text{dom } f$ qualquer temos $g \in \partial_{\infty} f(x) = \lim_{\varepsilon \rightarrow \infty} \partial_{\varepsilon} f(x)$ e portanto podemos escolher um $\varepsilon \geq 0$ tal que $g \in \partial_{\varepsilon} f(x)$. Portanto, pela equação básica dos ε -subgradientes:

$$f(y) \geq f(x) + g \cdot (y - x) - \varepsilon,$$

para todo $y \in \mathbb{R}^n$. Rearranjando temos:

$$\varepsilon + g \cdot x - f(x) \geq g \cdot y - f(y),$$

e tomando o supremo com relação a $y \in \mathbb{R}^n$:

$$\infty > \varepsilon + g \cdot x - f(x) \geq \sup_{y \in \mathbb{R}^n} \{g \cdot y - f(y)\} = f^*(g),$$

ou seja, $g \in \text{dom } f^*$.

As duas inclusões provam então que $\partial_{\infty} f = \text{dom } f^*$. □

A Figura A.5 ilustra de forma pictórica o contexto da proposição A.3. Notar que em geral $\text{dom } f^*$ não é o \mathbb{R}^n . É possível portanto que, dado um certo $x \in \text{dom } f$, tenhamos $\text{dom } f^* = \partial_{\varepsilon} f(x)$ para $\varepsilon \geq 0$, sem necessariamente termos que considerar o limite quando $\varepsilon \rightarrow \infty$. Um exemplo trivial é a função norma $f(x) = \|x\|$: dado $x \in \mathbb{R}^n$ basta escolhermos $\varepsilon = 2\|x\|$ e teremos $\text{dom } f^* = \{g \in \mathbb{R}^n : \|g\| \leq 1\} = \partial_{2\|x\|} f(x)$. Em especial, para $x = 0$ temos que $\text{dom } f^* = \partial_0 f(0) = \partial f(0)$, ou seja, $\varepsilon = 0$ basta.

A.3 Aproximação da função máximo

A função máximo entre dois números é frequentemente utilizada na criação de uma função convexa a partir de duas funções convexas iniciais, mas a função resultante pode não ser diferenciável mesmo que as funções componentes o sejam, sendo este um resultado bastante conhecido da Análise Convexa. Na verdade, esta não diferenciabilidade é muitas vezes explorada na modelagem de problemas não diferenciáveis.

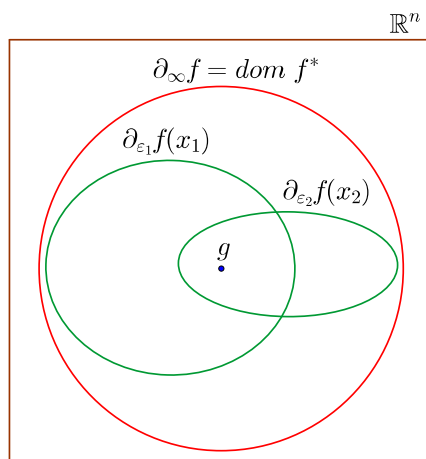


Figura A.5: Representação pictórica do contexto da proposição A.3

Em nosso desenvolvimento começaremos apresentando a seguinte definição clássica da função máximo:

$$\max(\xi_1, \xi_2) = M(\xi_1, \xi_2) = \frac{\xi_1 + \xi_2}{2} + \frac{|\xi_1 - \xi_2|}{2}, \forall \xi_1, \xi_2 \in \mathbb{R},$$

onde utilizaremos $M(\xi_1, \xi_2)$ para simplificar as notações. Esta definição pode ser facilmente verificada, e a não diferenciabilidade da função máximo manifesta-se pela presença da função valor absoluto.

Para nossos propósitos, no entanto, daremos preferência para uma outra versão da função máximo, baseada na função rampa, assim definida:

$$r(t) = \begin{cases} 0, & t < 0, \\ t, & t \geq 0. \end{cases}$$

Esta é uma função contínua e diferenciável em todos os pontos, exceto na origem. Com base na função rampa podemos construir uma versão extremamente compacta da função máximo, dada por:

$$M(\xi_1, \xi_2) = r(\xi_1 - \xi_2) + \xi_2, \forall \xi_1, \xi_2 \in \mathbb{R},$$

que também pode ser facilmente verificada. A derivada da função rampa é a função degrau $u(t)$ e por sua vez a derivada da função degrau é a função (generalizada) delta de Dirac $\delta(t)$. Estas funções são assim definidas, para atender aos nossos

requisitos:

$$u(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{2}, & t = 0, \\ 1, & t > 0. \end{cases}$$

$$\begin{aligned} \delta(t) &= 0, t \neq 0, \\ \int_{-\infty}^{+\infty} \delta(t) dt &= 1. \end{aligned}$$

A razão para definirmos $u(0) = 1/2$ é que isto proporciona a validade da expressão $u(t) + u(-t) = 1$, ou seja, $u(-t) = 1 - u(t)$, $\forall t \in \mathbb{R}$. Além disto é imediato que $0 \leq u(t) \leq 1$, e podemos assim interpretar $u(t)$ como sendo um certo λ de uma combinação convexa, e $u(-t) = 1 - u(t)$ como seu respectivo $1 - \lambda$. A Figura A.6 ilustra os gráficos das funções $r(t)$, $u(t)$ e $\delta(t)$.

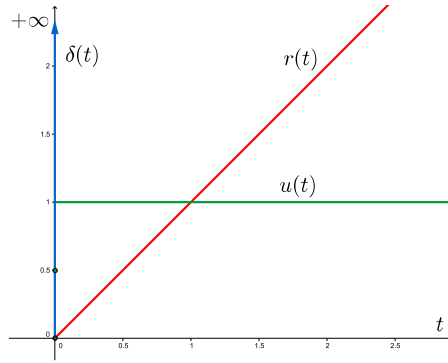


Figura A.6: Gráficos de $r(t)$, $u(t)$ e $\delta(t)$. Todas são nulas para $t < 0$.

Iremos agora considerar que ξ_1 e ξ_2 sejam funções diferenciáveis tendo por domínio o espaço \mathbb{R}^2 , e por simplicidade notacional iremos representá-las por $\xi_1(x, y)$ e $\xi_2(x, y)$ e, com um certo abuso de notação, utilizaremos $M(x, y)$ querendo dizer $M(\xi_1(x, y), \xi_2(x, y))$. As seguintes definições irão facilitar algumas referências posteriores:

$$Z_1 = \{(x, y) : \xi_1(x, y) > \xi_2(x, y)\},$$

$$Z_2 = \{(x, y) : \xi_1(x, y) < \xi_2(x, y)\},$$

$$Z_0 = \{(x, y) : \xi_1(x, y) = \xi_2(x, y)\}.$$

Para os espaços em que estamos trabalhando normalmente teremos Z_1 e Z_2 com dimensão dois e Z_0 com dimensão um. A região Z_0 pode ser imaginada como uma região de transição entre as regiões Z_1 e Z_2 .

Temos então, para $(x, y) \in Z_1 \cup Z_2$, ou seja, para a região onde $\xi_1(x, y) \neq \xi_2(x, y)$:

$$\frac{\partial M}{\partial x} = \frac{\partial M}{\partial \xi_1} \frac{\partial \xi_1}{\partial x} + \frac{\partial M}{\partial \xi_2} \frac{\partial \xi_2}{\partial x} = u(\xi_1 - \xi_2) \frac{\partial \xi_1}{\partial x} + (1 - u(\xi_1 - \xi_2)) \frac{\partial \xi_2}{\partial x},$$

$$\frac{\partial M}{\partial y} = \frac{\partial M}{\partial \xi_1} \frac{\partial \xi_1}{\partial y} + \frac{\partial M}{\partial \xi_2} \frac{\partial \xi_2}{\partial y} = u(\xi_1 - \xi_2) \frac{\partial \xi_1}{\partial y} + (1 - u(\xi_1 - \xi_2)) \frac{\partial \xi_2}{\partial y},$$

e assim podemos escrever o gradiente de M como:

$$\nabla M = u(\xi_1 - \xi_2) \nabla \xi_1 + u(\xi_2 - \xi_1) \nabla \xi_2,$$

onde utilizamos $u(\xi_2 - \xi_1) = 1 - u(\xi_1 - \xi_2)$. Verificamos então que o gradiente de M é uma combinação convexa extrema ($\lambda = 0$ ou $\lambda = 1$) dos gradientes de ξ_1 e ξ_2 , um resultado que é bastante intuitivo. O gradiente no entanto não está definido na região Z_0 , transição entre Z_1 e Z_2 , quando o gradiente de M muda bruscamente entre os gradientes $\nabla \xi_1$ e $\nabla \xi_2$ ao nos deslocarmos da região Z_1 para a região Z_2 (e vice-versa).

O cálculo da Hessiana é facilitado se notarmos que $\delta(t)$ é nula em todo ponto diferente de zero. Nas expressões abaixo utilizaremos $\xi_{ij} = \xi_i - \xi_j$. Assim, para $(x, y) \in Z_1 \cup Z_2$:

$$\frac{\partial^2 M}{\partial x^2} = \frac{\partial}{\partial x} \left(u(\xi_{12}) \frac{\partial \xi_1}{\partial x} + (1 - u(\xi_{12})) \frac{\partial \xi_2}{\partial x} \right) = u(\xi_{12}) \frac{\partial^2 \xi_1}{\partial x^2} + (1 - u(\xi_{12})) \frac{\partial^2 \xi_2}{\partial x^2},$$

$$\frac{\partial^2 M}{\partial y^2} = \frac{\partial}{\partial y} \left(u(\xi_{12}) \frac{\partial \xi_1}{\partial y} + (1 - u(\xi_{12})) \frac{\partial \xi_2}{\partial y} \right) = u(\xi_{12}) \frac{\partial^2 \xi_1}{\partial y^2} + (1 - u(\xi_{12})) \frac{\partial^2 \xi_2}{\partial y^2},$$

$$\frac{\partial^2 M}{\partial x \partial y} = \frac{\partial}{\partial y} \left(u(\xi_{12}) \frac{\partial \xi_1}{\partial x} + (1 - u(\xi_{12})) \frac{\partial \xi_2}{\partial x} \right) = u(\xi_{12}) \frac{\partial^2 \xi_1}{\partial x \partial y} + (1 - u(\xi_{12})) \frac{\partial^2 \xi_2}{\partial x \partial y},$$

e portanto temos que:

$$H_M = u(\xi_1 - \xi_2) H_{\xi_1} + u(\xi_2 - \xi_1) H_{\xi_2}.$$

De forma similar ao gradiente de M , verificamos que a Hessiana de M é uma combinação convexa extrema das Hessianas de ξ_1 e ξ_2 e mais uma vez devemos lembrar que H_M não está definida na região Z_0 , quando a Hessiana de M muda bruscamente entre as Hessianas H_{ξ_1} e H_{ξ_2} ao nos deslocarmos da região Z_1 para a região Z_2 (e vice-versa).

Se tivermos $\xi_1(x, y)$ e $\xi_2(x, y)$ convexas podemos afirmar que $M(x, y)$ também será convexa. Isto porque o epígrafo de $M(x, y)$ é a interseção dos epígrafos (convexos) de $\xi_1(x, y)$ e $\xi_2(x, y)$, e como a interseção de conjuntos convexos também é um conjunto convexo, o epígrafo de $M(x, y)$ também é um conjunto convexo e portanto

$M(x, y)$ é convexa. Tentar mostrar a convexidade de $M(x, y)$ por intermédio das relações analíticas derivadas acima se torna trabalhoso devido à divisão do plano xy nas regiões Z_1 , Z_2 e Z_0 , causada pela não diferenciabilidade de $M(\xi_1, \xi_2)$. No entanto, se substituirmos $M(\xi_1, \xi_2)$ por uma aproximação diferenciável $S(\xi_1, \xi_2)$, o método analítico se torna mais tratável e restrições analíticas podem ser determinadas para garantir a convexidade da composta $S(x, y) = S(\xi_1(x, y), \xi_2(x, y))$ em todo o \mathbb{R}^2 .

Ao intentarmos construir uma aproximação genérica suave (diferenciável) para a função máximo precisamos fornecer uma aproximação suave para a função rampa $r(t)$, base de nossa definição. Primeiramente observemos que $r(t)$ pode ser definida em termos da função $u(t)$ como:

$$r(t) = \int_{-\infty}^t u(\nu) d\nu.$$

Adicionalmente, podemos verificar que a função $u(t)$ possui algumas propriedades que a tornam matematicamente interessante, já comentadas acima, tais como o fato de que $0 \leq u(t) \leq 1$ e que $u(t) + u(-t) = 1$. Iremos portanto partir de uma aproximação suave para a própria função $u(t)$, que chamaremos de $u_s(t)$ e com antiderivada que chamaremos de $r_s(t)$, o que nos permitirá trabalharmos pelo menos com a segunda derivada de $r_s(t)$, que chamaremos de $\delta_s(t)$.

Juntamente com as propriedades herdadas de $u(t)$, iremos impor as seguintes restrições a $u_s(t)$:

- $u_s(t)$ está definida na reta real e $0 \leq u_s(t) \leq 1$, ou seja, $u_s : \mathbb{R} \rightarrow [0, 1]$;
- $u_s(t) \in C^1$, ou seja, $\delta_s(t)$ é contínua;
- $u_s(t) + u_s(-t) = 1$, ou seja, $u_s(t) - 1/2$ é anti-simétrica. Isto implica em termos $u_s(0) = 1/2$;
- $\delta_s(t) > 0$, ou seja, $u_s(t)$ é estritamente crescente;
- $\lim_{t \rightarrow -\infty} u_s(t) = 0$ e $\lim_{t \rightarrow +\infty} u_s(t) = 1$.

A Figura A.7 ilustra o gráfico de uma versão genérica de $u_s(t)$ que atende a todas as restrições listadas acima. Na Figura A.8 são apresentados os gráficos de versões genéricas de $r_s(t)$ e $\delta_s(t)$, construídas de acordo com as restrições impostas à função $u_s(t)$ da qual elas se originam.

Com base nestas novas funções a aproximação suave para a função máximo pode então ser definida como:

$$S(\xi_1, \xi_2) = r_s(\xi_1 - \xi_2) + \xi_2.$$

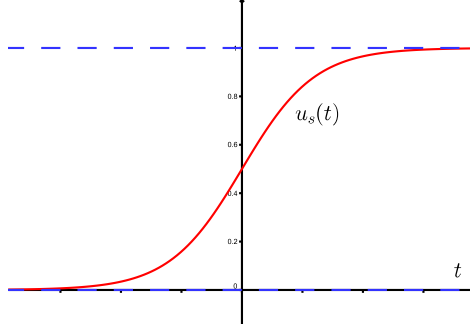


Figura A.7: Gráfico de uma função $u_s(t)$ genérica.

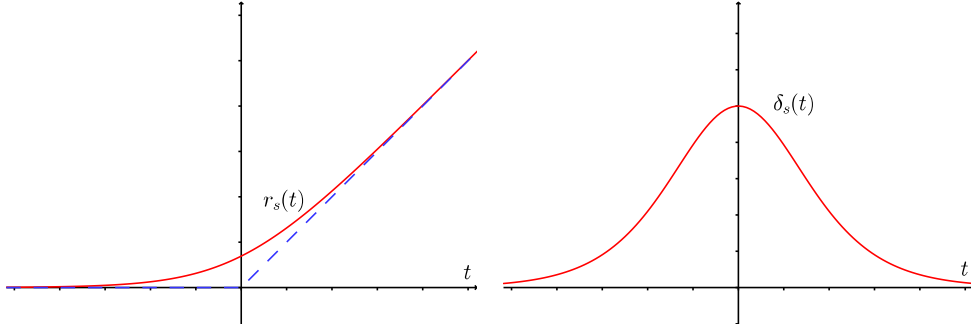


Figura A.8: Gráficos de funções $r_s(t)$ e $\delta_s(t)$ genéricas.

onde mais uma vez estamos considerando que $\xi_1(x, y)$ e $\xi_2(x, y)$ sejam funções diferenciáveis tendo por domínio o espaço \mathbb{R}^2 . Temos então:

$$\frac{\partial S}{\partial x} = \frac{\partial S}{\partial \xi_1} \frac{\partial \xi_1}{\partial x} + \frac{\partial S}{\partial \xi_2} \frac{\partial \xi_2}{\partial x} = u_s(\xi_1 - \xi_2) \frac{\partial \xi_1}{\partial x} + (1 - u_s(\xi_1 - \xi_2)) \frac{\partial \xi_2}{\partial x},$$

$$\frac{\partial S}{\partial y} = \frac{\partial S}{\partial \xi_1} \frac{\partial \xi_1}{\partial y} + \frac{\partial S}{\partial \xi_2} \frac{\partial \xi_2}{\partial y} = u_s(\xi_1 - \xi_2) \frac{\partial \xi_1}{\partial y} + (1 - u_s(\xi_1 - \xi_2)) \frac{\partial \xi_2}{\partial y},$$

e assim podemos escrever o gradiente de S como:

$$\nabla S = u_s(\xi_1 - \xi_2) \nabla \xi_1 + u_s(\xi_2 - \xi_1) \nabla \xi_2,$$

onde utilizamos $u(\xi_2 - \xi_1) = 1 - u(\xi_1 - \xi_2)$. Verificamos então que o gradiente de S é uma combinação convexa dos gradientes de ξ_1 e ξ_2 , como também aconteceu com M , mas no caso de ∇S os coeficientes da combinação convexa são funções suaves, diferentemente dos coeficientes de ∇M , que variam abruptamente entre os únicos valores possíveis, 0 e 1.

Para o cálculo da Hessiana de S , calculemos primeiramente:

$$\frac{\partial^2 S}{\partial x^2} = u_s(\xi_1 - \xi_2) \frac{\partial^2 \xi_1}{\partial x^2} + u_s(\xi_2 - \xi_1) \frac{\partial^2 \xi_2}{\partial x^2} + \delta_s(\xi_1 - \xi_2) \left(\frac{\partial \xi_1}{\partial x} - \frac{\partial \xi_2}{\partial x} \right)^2,$$

$$\frac{\partial^2 S}{\partial y^2} = u_s(\xi_1 - \xi_2) \frac{\partial^2 \xi_1}{\partial y^2} + u(\xi_2 - \xi_1) \frac{\partial^2 \xi_2}{\partial y^2} + \delta_s(\xi_1 - \xi_2) \left(\frac{\partial \xi_1}{\partial y} - \frac{\partial \xi_2}{\partial y} \right)^2,$$

$$\frac{\partial^2 S}{\partial x \partial y} = u_s(\xi_1 - \xi_2) \frac{\partial^2 \xi_1}{\partial x \partial y} + u(\xi_2 - \xi_1) \frac{\partial^2 \xi_2}{\partial x \partial y} + \delta_s(\xi_1 - \xi_2) \left(\frac{\partial \xi_1}{\partial x} - \frac{\partial \xi_2}{\partial x} \right) \left(\frac{\partial \xi_1}{\partial y} - \frac{\partial \xi_2}{\partial y} \right),$$

e portanto temos que:

$$H_S = u_s(\xi_1 - \xi_2)H_{\xi_1} + u_s(\xi_2 - \xi_1)H_{\xi_2} + \delta_s(\xi_1 - \xi_2)(\nabla \xi_1 - \nabla \xi_2)(\nabla \xi_1 - \nabla \xi_2)^T.$$

Mas:

$$(\nabla \xi_1 - \nabla \xi_2)(\nabla \xi_1 - \nabla \xi_2)^T = (\nabla \xi_1 - \nabla \xi_2)^T (\nabla \xi_1 - \nabla \xi_2) \frac{(\nabla \xi_1 - \nabla \xi_2)(\nabla \xi_1 - \nabla \xi_2)^T}{(\nabla \xi_1 - \nabla \xi_2)^T (\nabla \xi_1 - \nabla \xi_2)},$$

e daí resulta:

$$\begin{aligned} (\nabla \xi_1 - \nabla \xi_2)(\nabla \xi_1 - \nabla \xi_2)^T &= \|\nabla \xi_1 - \nabla \xi_2\|^2 \frac{(\nabla \xi_1 - \nabla \xi_2)(\nabla \xi_1 - \nabla \xi_2)^T}{(\nabla \xi_1 - \nabla \xi_2)^T (\nabla \xi_1 - \nabla \xi_2)}, \\ &= \|\nabla \xi_1 - \nabla \xi_2\|^2 \text{Proj}_{\nabla \xi_1 - \nabla \xi_2}, \end{aligned}$$

e finalmente podemos escrever:

$$H_S = u_s(\xi_1 - \xi_2)H_{\xi_1} + u_s(\xi_2 - \xi_1)H_{\xi_2} + \delta_s(\xi_1 - \xi_2)\|\nabla \xi_1 - \nabla \xi_2\|^2 \text{Proj}_{\nabla \xi_1 - \nabla \xi_2},$$

onde $\text{Proj}_{\nabla \xi_1 - \nabla \xi_2}$ representa a matriz de projeção na direção dada pelo vetor $\nabla \xi_1 - \nabla \xi_2$.

Verificamos que a Hessiana de S é uma combinação convexa das Hessianas de ξ_1 e ξ_2 acrescida de um termo de interação $\delta_s(\xi_1 - \xi_2)\|\nabla \xi_1 - \nabla \xi_2\|^2 \text{Proj}_{\nabla \xi_1 - \nabla \xi_2}$. Mas $\text{Proj}_{\nabla \xi_1 - \nabla \xi_2}$ é uma matriz de posto um que possui um único autovalor positivo dado por 1 (como toda matriz de projeção) e qualquer autovetor pode então ser expresso por $v = K(\nabla \xi_1 - \nabla \xi_2)$, para $K \in \mathbb{R} - \{0\}$. Assim, podemos afirmar que $\text{Proj}_{\nabla \xi_1 - \nabla \xi_2}$ é uma matriz positiva semidefinida.

Concluimos então que se H_{ξ_1} ou H_{ξ_2} (ou ambas) for uma matriz definida positiva, então H_S também o será e assim $S(x, y)$ será uma função (estritamente) convexa. Caso H_{ξ_1} e H_{ξ_2} forem matrizes positivas semidefinidas, então a convexidade de $S(x, y)$ deverá ser avaliada nos pontos onde H_S resultar semidefinida positiva.

Iremos agora estudar o caso de nosso interesse imediato, uma suavização exponencial, onde utilizaremos:

$$u_s(t) = \frac{e^t}{1 + e^t},$$

que atende todas as restrições criadas para $u_s(t)$, o que pode ser facilmente verificado.

Assim:

$$r_s(t) = \ln(1 + e^t),$$

$$\delta_s(t) = \frac{e^t}{(1 + e^t)^2},$$

e portanto:

$$\begin{aligned} S(\xi_1, \xi_2) &= \ln(1 + e^{\xi_1 - \xi_2}) + \xi_2 = \ln(e^{\xi_1} + e^{\xi_2}), \\ \nabla S &= \frac{e^{\xi_1}}{e^{\xi_1} + e^{\xi_2}} \nabla \xi_1 + \frac{e^{\xi_2}}{e^{\xi_1} + e^{\xi_2}} \nabla \xi_2, \\ H_S &= \frac{e^{\xi_1}}{e^{\xi_1} + e^{\xi_2}} H_{\xi_1} + \frac{e^{\xi_2}}{e^{\xi_1} + e^{\xi_2}} H_{\xi_2} + \frac{e^{\xi_1 + \xi_2}}{(e^{\xi_1} + e^{\xi_2})^2} \|\nabla \xi_1 - \nabla \xi_2\|^2 \text{Proj}_{\nabla \xi_1 - \nabla \xi_2}. \end{aligned}$$

Na seção 6.4 é apresentado um caso de suavização exponencial utilizando as funções $\xi_1(x, y) = k_1 \sqrt{x^2 + y^2}$ e $\xi_2(x, y) = k_2 x - a$, onde k_1 , k_2 e a são constantes positivas, com $k_2 > k_1$ para assegurar que haja interseção entre os gráficos das duas funções, e $a > \ln\left(\frac{k_2}{k_1}\right)$, para impor que $0 \in \partial S(0, 0)$, ou seja, para que a origem seja o ponto de mínimo de $S(x, y)$. Temos portanto:

$$S(x, y) = \ln\left(e^{k_1 \sqrt{x^2 + y^2}} + e^{k_2 x - a}\right).$$

É notório que ξ_1 é convexa (não estritamente), por ser um múltiplo da função norma, e ξ_2 é uma função afim (e portanto convexa não estritamente). A pergunta é: o que pode ser dito sobre a convexidade de $S(x, y)$? Iremos então analisar a estrutura de H_S e verificar em que condições podemos afirmar que $S(x, y)$ é convexa.

A função ξ_2 é extremamente simples e seu gradiente e Hessiana são dados por:

$$\nabla \xi_2 = \begin{bmatrix} k_2 \\ 0 \end{bmatrix},$$

$$H_{\xi_2} = 0.$$

O gradiente de ξ_1 é trivialmente calculado como:

$$\nabla \xi_1 = \begin{bmatrix} \frac{k_1 x}{\sqrt{x^2 + y^2}} \\ \frac{k_1 y}{\sqrt{x^2 + y^2}} \end{bmatrix},$$

e sua Hessiana:

$$H_{\xi_1} = \begin{bmatrix} \frac{k_1 y^2}{(x^2 + y^2)^{3/2}} & -\frac{k_1 xy}{(x^2 + y^2)^{3/2}} \\ -\frac{k_1 xy}{(x^2 + y^2)^{3/2}} & \frac{k_1 x^2}{(x^2 + y^2)^{3/2}} \end{bmatrix} = \frac{1}{k_1 \sqrt{x^2 + y^2}} \begin{bmatrix} \frac{k_1^2 y^2}{x^2 + y^2} & -\frac{k_1^2 xy}{x^2 + y^2} \\ -\frac{k_1^2 xy}{x^2 + y^2} & \frac{k_1^2 x^2}{x^2 + y^2} \end{bmatrix},$$

que também pode ser escrita como:

$$H_{\xi_1} = \frac{1}{k_1 \sqrt{x^2 + y^2}} \begin{bmatrix} \frac{k_1 y}{\sqrt{x^2 + y^2}} \\ \frac{-k_1 x}{\sqrt{x^2 + y^2}} \end{bmatrix} \begin{bmatrix} \frac{k_1 y}{\sqrt{x^2 + y^2}} & \frac{-k_1 x}{\sqrt{x^2 + y^2}} \end{bmatrix} = \frac{1}{k_1 \sqrt{x^2 + y^2}} (R \nabla \xi_1) (R \nabla \xi_1)^T,$$

onde:

$$R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

representa uma matriz de rotação de 90° na direção horária.

A identidade $(R \nabla \xi_1)^T (R \nabla \xi_1) = (\nabla \xi_1)^T (\nabla \xi_1) = k_1^2$ pode ser facilmente verificada, e disto resulta:

$$H_{\xi_1} = \frac{k_1}{\sqrt{x^2 + y^2}} \frac{(R \nabla \xi_1) (R \nabla \xi_1)^T}{(R \nabla \xi_1)^T (R \nabla \xi_1)} = \frac{k_1}{\sqrt{x^2 + y^2}} \text{Proj}_{R \nabla \xi_1},$$

onde $\text{Proj}_{R \nabla \xi_1}$ representa a matriz de projeção na direção dada pelo vetor $R \nabla \xi_1$.

Por sua vez:

$$\nabla \xi_1 - \nabla \xi_2 = \begin{bmatrix} \frac{k_1 x}{\sqrt{x^2 + y^2}} - k_2 \\ \frac{k_1 y}{\sqrt{x^2 + y^2}} \end{bmatrix},$$

e a matriz do último termo de H_S é a matriz de projeção sobre este vetor. Logo:

$$H_S = \frac{e^{\xi_1}}{e^{\xi_1} + e^{\xi_2}} \frac{k_1}{\sqrt{x^2 + y^2}} \text{Proj}_{R \nabla \xi_1} + \frac{e^{\xi_1 + \xi_2}}{(e^{\xi_1} + e^{\xi_2})^2} \|\nabla \xi_1 - \nabla \xi_2\|^2 \text{Proj}_{\nabla \xi_1 - \nabla \xi_2}.$$

Como H_S é a soma *positivamente* ponderada de duas matrizes de projeção nas direções $R \nabla \xi_1$ e $\nabla \xi_1 - \nabla \xi_2$, e considerando que estamos trabalhando no \mathbb{R}^2 , ela somente será semidefinida positiva se estes vetores apontarem na mesma direção. Portanto a situação que deve ser analisada é quando $R \nabla \xi_1 = K (\nabla \xi_1 - \nabla \xi_2)$, para algum $K \in \mathbb{R} - \{0\}$. Temos portanto o sistema de equações:

$$\begin{aligned} \frac{k_1 y}{\sqrt{x^2 + y^2}} &= K \left(\frac{k_1 x}{\sqrt{x^2 + y^2}} - k_2 \right), \\ \frac{-k_1 x}{\sqrt{x^2 + y^2}} &= K \left(\frac{k_1 y}{\sqrt{x^2 + y^2}} \right), \end{aligned}$$

ou, para $(x, y) \neq (0, 0)$,

$$\begin{aligned} k_1 y &= K \left(k_1 x - k_2 \sqrt{x^2 + y^2} \right), \\ -k_1 x &= K k_1 y, \end{aligned}$$

Notamos imediatamente que se $x < 0$ a primeira equação implica em que y

deve ter sinal contrário ao de K . No entanto, a segunda equação indica que y deve ser diferente de zero e ter o mesmo sinal que K . Assim, é impossível que H_S seja semidefinida para $x < 0$, ou seja, para $x < 0$ sempre teremos H_S definida positiva e portanto estritamente convexa.

Para $x = 0$ a segunda equação implica em termos $y = 0$. Mas a origem está fora do contexto do estudo da Hessiana desta função, pois foi o ponto escolhido para abrigar a única não diferenciabilidade de $S(x, y)$. Portanto H_S é definida positiva e portanto estritamente convexa na reta definida por $\{(0, y) : y \neq 0\}$.

Para $x > 0$, temos pela segunda equação que $-x = Ky$ (logo $y \neq 0$) e portanto $K = -\frac{x}{y}$. Introduzindo na primeira equação:

$$\begin{aligned} k_1 y^2 &= -k_1 x^2 + k_2 x \sqrt{x^2 + y^2}, \\ k_1 (x^2 + y^2) &= k_2 x \sqrt{x^2 + y^2}, \\ k_1 \sqrt{x^2 + y^2} &= k_2 x, \\ k_1^2 x^2 + k_1^2 y^2 &= k_2^2 x^2, \\ y^2 &= \left(\left(\frac{k_2}{k_1} \right)^2 - 1 \right) x^2, \end{aligned}$$

e finalmente temos:

$$y = \pm \left(\sqrt{\left(\frac{k_2}{k_1} \right)^2 - 1} \right) x.$$

Esta equação representa duas semiretas partindo da origem e contidas no semiplano dos $x > 0$, com coeficientes angulares simétricos e sabemos, por este desenvolvimento, que H_S é semidefinida sobre estas semiretas. Como nos pontos onde a Hessiana é semidefinida não podemos fazer afirmações quanto à convexidade, iremos analisar $S(x, y)$ sobre estas semiretas buscando determinar seu comportamento local. A forma mais fácil de fazer isto é substituir y em função de x diretamente na equação de $S(x, y)$:

$$\begin{aligned} S \left(x, \pm \left(\sqrt{\left(\frac{k_2}{k_1} \right)^2 - 1} \right) x \right) &= \ln \left(e^{k_1 \sqrt{x^2 + \left(\left(\frac{k_2}{k_1} \right)^2 - 1 \right) x^2}} + e^{k_2 x - a} \right) = \\ &= \ln \left(e^{k_1 \sqrt{\left(\frac{k_2}{k_1} \right)^2 x^2}} + e^{k_2 x - a} \right) = \ln(e^{k_2 x} + e^{k_2 x - a}) = \ln(1 + e^{-a}) + k_2 x, \end{aligned}$$

que é uma função linear em x , e portanto convexa. Assim $S(x, y)$ é convexa em todos os pontos das duas semiretas, e portanto em todos os pontos do semiplano aberto $\{(x, y) : x > 0\}$.

Como $S(x, y) \geq \ln(1 + e^{-a})$, pois para $a > \ln\left(\frac{k_2}{k_1}\right)$ a origem é o ponto de

mínimo de $S(x, y)$, é possível considerarmos um hiperplano passando pelo ponto do \mathbb{R}^3 dado pelas coordenadas $(0, 0, \ln(1 + e^{-a}))$, paralelo ao plano xy , e deixando todo o epígrafo de $S(x, y)$ em somente um de seus semiespaços, e portanto a origem não viola a convexidade de $S(x, y)$.

A conclusão deste estudo analítico da função $S(x, y)$ é que podemos afirmar que ela é convexa em todo o \mathbb{R}^2 .