



PROGRAD
Pró-Reitoria de
Graduação



Notas de estudo

Métodos do Gradiente Espectral Projetado

Samara Viriato Vilar Dias

Vitória da Conquista, 2024

Conteúdo

1	Resultados preliminares	2
1.1	Projeção e Convexidade	2
1.2	Condições de otimalidade	7
2	Buscas lineares	10
2.1	Buscas lineares para o caso com restrições	10
2.2	Busca linear não monótona de Grippo-Lampariello-Lucidi	10
3	Métodos do Gradiente Espectral Projetado	14
3.1	Passo espectral de Barzilai e Borwein	14
3.2	Algoritmos dos Métodos do Gradiente Espectral Projetado	15
3.3	Resultados de convergência	16
3.4	Interpolação quadrática	24
4	Experimentos numéricos	26

1 Resultados preliminares

Teorema 1.1. (Caracterizações de funções convexas diferenciáveis) *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : \Omega \rightarrow \mathbb{R}$ uma função diferenciável em Ω . Então, as propriedades seguintes são equivalentes:*

- (a) *A função f é convexa em Ω ;*
- (b) *para quaisquer $x, y \in \Omega$,*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

Demonstração: Veja a demonstração em [5, Teorema 3.4.30].

Teorema 1.2. (Teorema do Valor Médio (TVM)) *Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é continuamente diferenciável num conjunto convexo e aberto $\Omega \subset \mathbb{R}^n$, então para todo $x, y \in \Omega$ existe $t \in [0, 1]$ tal que*

$$f(y) - f(x) = \langle \nabla f(tx + (1-t)y), y - x \rangle. \quad (1)$$

1.1 Projeção e Convexidade

Teorema 1.3. (Teorema de Weierstrass) *Sejam $D \subset \mathbb{R}^n$ um conjunto compacto não-vazio e $f : D \rightarrow \mathbb{R}$ uma função contínua. Então os problemas de minimizar e maximizar f em D têm soluções globais.*

Demonstração: Sabemos que qualquer problema de maximização

$$\max f(x) \text{ sujeito a } x \in D,$$

pode ser transformado em um problema de minimização equivalente:

$$\min -f(x) \text{ sujeito a } x \in D.$$

Assim, é suficiente provar a existência de um minimizador ou de um maximizador. Mostraremos a existência de um minimizador.

Como a imagem de um conjunto compacto por uma função contínua é compacta, $\{v \in \mathbb{R}; v = f(x) \text{ para algum } x \in D\}$ é compacto. Em particular, este conjunto é limitado inferiormente, ou seja,

$$-\infty < \bar{v} = \inf_{x \in D} f(x).$$

Pela definição de ínfimo, para todo $k \in \mathbb{N}$ existe um $x_k \in D$ tal que

$$\bar{v} \leq f(x_k) \leq \bar{v} + 1/k.$$

Passando ao limite quando $k \rightarrow \infty$, concluímos que

$$\lim_{k \rightarrow \infty} f(x_k) = \bar{v}. \quad (2)$$

Como $\{x_k\} \subset D$ e D é compacto, segue-se que $\{x_k\}$ é uma sequência limitada. Logo, ela possui uma subsequência $\{x_{k_j}\}$ que converge a um ponto de D :

$$\lim_{j \rightarrow \infty} x_{k_j} = \bar{x} \in D.$$

Pela continuidade de f ,

$$\lim_{j \rightarrow \infty} f(x_{k_j}) = f(\bar{x}).$$

Usando (2), temos que $f(\bar{x}) = \bar{v}$, i. e., f assume o valor mínimo em D no ponto $\bar{x} \in D$. Em outras palavras, \bar{x} é um minimizador global do problema de $\min f(x)$ sujeito a $x \in D$.

Definição 1.4. O conjunto de nível da função $f : D \rightarrow \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por $L_{f,D}(c) = \{x \in D; f(x) \leq c\}$.

Corolário 1.5. Sejam $D \subset \mathbb{R}^n$ e $f : D \rightarrow \mathbb{R}$ contínua no conjunto D . Suponhamos que existe $c \in \mathbb{R}$ tal que o conjunto de nível $L_{f,D}(c)$ seja não-vazio e compacto. Então o problema de minimizar f em D possui uma solução global.

Demonstração: Pelo Teorema 1.3, o problema $\min f(x)$ sujeito a $x \in L_{f,D}(c)$ tem uma solução global, digamos \bar{x} . Para todo ponto $x \in D \setminus L_{f,D}(c)$, temos que $f(x) > c \geq f(\bar{x})$, o que mostra que \bar{x} é um minimizador global de f não só em $L_{f,D}(c)$, mas também em D .

Definição 1.6. Dados conjunto $D \subset \mathbb{R}^n$ e ponto $x \in \mathbb{R}^n$, uma projeção (ortogonal) de x sobre D é uma solução global do problema

$$\min \|y - x\| \text{ sujeito a } y \in D. \quad (3)$$

Em outras palavras, a projeção de x sobre D é um dos pontos de D que é mais próximo de x . Dependendo das hipóteses sobre o conjunto D , pode haver mais de um ponto com esta propriedade, um único ponto ou nenhum.

Exemplo 1.7. Casos de existência da projeção:

- Todo ponto da esfera $D = \{y \in \mathbb{R}^n; \|y\| = 1\}$ é uma projeção da origem $x = 0$ sobre D .
- A projeção de $x = 1$ sobre $D = [-1, 0)$ não existe.
- A projeção de $x = 1$ sobre $D = [-1, 0]$ é o único ponto $\bar{x} = 0$.

Teorema 1.8. (Teorema da projeção)

- (a) Seja $D \subset \mathbb{R}^n$ um conjunto fechado. Então para todo ponto $x \in \mathbb{R}^n$, uma projeção de x sobre D existe.
- (b) Se, além de ser fechado, o conjunto D é convexo, então para todo $x \in \mathbb{R}^n$ a projeção de x sobre D , denotada por $P_D(x)$, é única. Além disso, $\bar{x} = P_D(x)$ se, e somente se,

$$\bar{x} \in D, \langle x - \bar{x}, y - \bar{x} \rangle \leq 0, \forall y \in D. \quad (4)$$

Também tem-se que

$$\|P_D(x) - P_D(y)\| \leq \|x - y\|, \forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n. \quad (5)$$

Demonstração:

(a) Fixemos $x \in \mathbb{R}^n$ arbitrário. Seja

$$f : \mathbb{R}^n \rightarrow \mathbb{R}_+, f(x) = \|y - x\|.$$

Evidentemente, f é contínua no \mathbb{R}^n e

$$L_{f,D}(c) = D \cap B(x, c),$$

onde

$$B(x, c) = \{y \in \mathbb{R}^n; \|y - x\| \leq c\} = \{x \in \mathbb{R}^n; f(x) \leq c\},$$

ou seja, o conjunto de nível $L_{f,D}(c)$ é a interseção do conjunto fechado D com o conjunto compacto $B(x, c)$. Portanto, $L_{f,D}(c)$ é compacto. Além disso, para $c > 0$ suficientemente grande, é óbvio que a bola $B(x, c)$ contém pontos de D . Logo, $L_{f,D}(c) \neq \emptyset$. Pelo Corolário 1.5 temos que o problema de minimizar f em D possui uma solução global. Assim, a projeção de x sobre D existe (pela definição de projeção de x sobre D).

(b) Sejam \bar{x} uma solução do problema (3) e $y \in D$ qualquer. Como $\bar{x} \in D$ e D é convexo, temos pela definição de combinação convexa que $(1 - \alpha)\bar{x} + \alpha y = x(\alpha) \in D$ para todo $\alpha \in (0, 1]$. Temos, então, que $\|x - \bar{x}\| \leq \|x - x(\alpha)\| = \|x - ((1 - \alpha)\bar{x} + \alpha y)\| = \|x - \bar{x} + \alpha\bar{x} + \alpha y\|$ e $\|x - x(\alpha)\| \geq \|x - \bar{x}\| \implies \|x - x(\alpha)\|^2 \geq \|x - \bar{x}\|^2$. Assim,

$$\begin{aligned} 0 &\geq \|x - \bar{x}\|^2 - \|x - x(\alpha)\|^2 \\ &= \|x - \bar{x}\|^2 - \|x - \bar{x} + \bar{x} - x(\alpha)\|^2 \\ &= \|x - \bar{x}\|^2 - \|(x - \bar{x}) + (\bar{x} - x(\alpha))\|^2 \\ &= \|x - \bar{x}\|^2 - (\|x - \bar{x}\|^2 + 2\langle(x - \bar{x}), (\bar{x} - x(\alpha))\rangle + \|\bar{x} - x(\alpha)\|^2) \\ &= -2\langle(x - \bar{x}), (\bar{x} - x(\alpha))\rangle - \|\bar{x} - x(\alpha)\|^2 \\ &= 2\langle(x - \bar{x}), (x(\alpha) - \bar{x})\rangle - \|x(\alpha) - \bar{x}\|^2 \\ &= 2\langle(x - \bar{x}), ((1 - \alpha)\bar{x} + \alpha y - \bar{x})\rangle - \|(1 - \alpha)\bar{x} + \alpha y - \bar{x}\|^2 \\ &= 2\langle(x - \bar{x}), (-\alpha\bar{x} + \alpha y)\rangle - \|\alpha y - \alpha\bar{x}\|^2 \\ &= 2\alpha\langle(x - \bar{x}), (y - \bar{x})\rangle - \|\alpha(y - \bar{x})\|^2 \\ &= 2\alpha\langle(x - \bar{x}), (y - \bar{x})\rangle - \alpha^2\|y - \bar{x}\|^2 \end{aligned}$$

Dividindo ambos os lados da desigualdade acima por $2\alpha > 0$, temos

$$0 \geq \langle x - \bar{x}, y - \bar{x} \rangle - \frac{\alpha}{2}\|y - \bar{x}\|^2.$$

Passando ao limite quando $\alpha \rightarrow 0+$, obtemos

$$0 \geq \langle x - \bar{x}, y - \bar{x} \rangle,$$

sendo que $y \in D$ era arbitrário.

Suponhamos agora que um certo \bar{x} satisfaça (4). Então, para todo $y \in D$,

$$\begin{aligned}
0 &\geq \langle x - \bar{x}, y - \bar{x} \rangle \\
&= \langle x, y \rangle - \langle \bar{x}, y \rangle - \langle x, \bar{x} \rangle + \langle \bar{x}, \bar{x} \rangle \\
&= \frac{1}{2}(2\langle x, y \rangle - 2\langle \bar{x}, y \rangle - 2\langle x, \bar{x} \rangle + 2\langle \bar{x}, \bar{x} \rangle) \\
&= \frac{1}{2}(\langle x, x \rangle - \langle x, \bar{x} \rangle - \langle \bar{x}, x \rangle + \langle \bar{x}, \bar{x} \rangle + \langle y, y \rangle - \langle y, \bar{x} \rangle \\
&\quad - \langle \bar{x}, y \rangle + \langle \bar{x}, \bar{x} \rangle - \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle - \langle y, y \rangle) \\
&= \frac{1}{2}(\langle x - \bar{x}, x - \bar{x} \rangle + \langle y, \bar{x}, y - \bar{x} \rangle - \langle x - y, x - y \rangle) \\
&= \frac{1}{2}(\|x - \bar{x}\|^2 + \|y - \bar{x}\|^2 - \|x - y\|^2) \\
&\geq \frac{1}{2}(\|x - \bar{x}\|^2 - \|x - y\|^2).
\end{aligned}$$

Agora, vamos mostrar que a projeção é única. Seja \hat{x} alguma outra solução de (3). Usando (4) para \bar{x} com $y = \hat{x} \in D$ e para \hat{x} com $y = \bar{x} \in D$, temos

$$\langle x - \bar{x}, \hat{x} - \bar{x} \rangle \leq 0 \text{ e } \langle x - \hat{x}, \bar{x} - \hat{x} \rangle \leq 0 \implies -\langle x - \hat{x}, \hat{x} - \bar{x} \rangle \leq 0.$$

Somando, obtemos que

$$\begin{aligned}
0 &\geq \langle x - \bar{x}, \hat{x} - \bar{x} \rangle - \langle x - \hat{x}, \hat{x} - \bar{x} \rangle \\
&= \langle x - \bar{x} - (x - \hat{x}), \hat{x} - \bar{x} \rangle \\
&= \langle \hat{x} - \bar{x}, \hat{x} - \bar{x} \rangle \\
&= \|\hat{x} - \bar{x}\|^2.
\end{aligned}$$

Logo, $\hat{x} = \bar{x}$.

Por fim, vamos provar (5). Como $P_D(x) \in D$ e $P_D(y) \in D$, usando (4) duas vezes (para x e y respectivamente), obtemos

$$\begin{aligned}
\langle x - P_D(x), P_D(y) - P_D(x) \rangle \leq 0 &\implies -\langle x - P_D(y), P_D(x) - P_D(y) \rangle \leq 0, \\
\langle y - P_D(y), P_D(x) - P_D(y) \rangle &\leq 0.
\end{aligned}$$

Somando as duas desigualdades, temos que

$$\begin{aligned}
0 &\geq \langle y - P_D(y), P_D(x) - P_D(y) \rangle - \langle x - P_D(y), P_D(x) - P_D(y) \rangle \\
&= \langle x - P_D(x) - (y - P_D(y)), P_D(y) - P_D(x) \rangle \\
&= \langle (y - x) + (P_D(x) - P_D(y)), P_D(x) - P_D(y) \rangle \\
&= \langle P_D(x) - P_D(y), P_D(x) - P_D(y) \rangle + \langle y - x, P_D(x) - P_D(y) \rangle \\
&= \|P_D(x) - P_D(y)\|^2 + \langle y - x, P_D(x) - P_D(y) \rangle.
\end{aligned}$$

Então temos $0 \geq \|P_D(x) - P_D(y)\|^2 + \langle y - x, P_D(x) - P_D(y) \rangle \implies -\langle y - x, P_D(x) - P_D(y) \rangle \geq \|P_D(x) - P_D(y)\|^2 \implies \langle x - y, P_D(x) - P_D(y) \rangle \geq \|P_D(x) - P_D(y)\|^2 \geq 0$.

Usando agora a desigualdade de Cauchy-Schwarz, obtemos

$$\begin{aligned}\|P_D(x) - P_D(y)\| \|y - x\| &\geq |\langle x - y, P_D(x) - P_D(y) \rangle| \\ &\geq \langle x - y, P_D(x) - P_D(y) \rangle \\ &\geq \|P_D(x) - P_D(y)\|^2.\end{aligned}$$

Se $P_D(x) = P_D(y)$, (5) vale trivialmente. Caso contrário, obtemos (5) dividindo os dois lados da desigualdade acima por $\|P_D(x) - P_D(y)\| > 0$.

Teorema 1.9. *Vale o seguinte*

(a) *Se para $x, y \in \mathbb{R}^n$ uma função $F : \mathbb{R}^n \rightarrow \mathbb{R}^l$ é continuamente diferenciável no intervalo $\{x + ty \mid t \in [0, 1]\}$, então*

$$F(x + y) = F(x) + \int_0^1 F'(x + ty)y dt.$$

(b) *Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é continuamente diferenciável num conjunto convexo e aberto $D \subset \mathbb{R}^n$, então para todo $x, y \in D$ existe $t \in [0, 1]$ tal que*

$$f(y) - f(x) = \langle \nabla f(tx + (1 - t)y), y - x \rangle.$$

Lema 1.10. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no \mathbb{R}^n , com gradiente Lipschitz-contínuo no \mathbb{R}^n com módulo $L > 0$. Então*

$$|f(x + y) - f(x) - \nabla f(x)^T y| \leq \frac{L\|y\|^2}{2}, \quad \forall x, y \in \mathbb{R}^n.$$

Demonstração: Pelo Teorema 1.9 (a), temos que $f(x + y) = f(x) + \int_0^1 \nabla f(x + ty)^T y dt$. Assim,

$$\begin{aligned}|f(x + y) - f(x) - \nabla f(x)^T y| &= \left| f(x) + \int_0^1 \nabla f(x + ty)^T y dt - f(x) - \nabla f(x)^T y \right| \\ &= \left| \int_0^1 \nabla f(x + ty)^T y dt - \nabla f(x)^T y \right| \\ &= \left| \int_0^1 (\nabla f(x + ty) - \nabla f(x))^T y dt \right| \\ &\leq \int_0^1 |(\nabla f(x + ty) - \nabla f(x))^T y| dt.\end{aligned}$$

Pela Desigualdade de Cauchy-Schwarz, temos que

$$\int_0^1 |(\nabla f(x + ty) - \nabla f(x))^T y| dt \leq \int_0^1 \|\nabla f(x + ty) - \nabla f(x)\| \|y\| dt.$$

Além disso, temos (por hipótese) que ∇f é Lipschitz-contínuo no \mathbb{R}^n , ou seja, $\|\nabla f(x + ty) - \nabla f(x)\| \leq L\|(x + ty) - x\| = L\|ty\| = Lt\|y\|$, logo

$$\begin{aligned} \int_0^1 \|\nabla f(x + ty) - \nabla f(x)\| \|y\| dt &\leq \int_0^1 Lt\|y\|^2 dt \\ &= \left[L\frac{t^2}{2}\|y\|^2 \right]_0^1 \\ &= \frac{L\|y\|^2}{2}. \end{aligned}$$

□

1.2 Condições de otimalidade

Teorema 1.11. (*Condição necessária em forma primal*) *Sejam $D \subset \mathbb{R}^n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in D$. Se \bar{x} é uma solução local do problema de minimizar f em D , então*

$$\langle \nabla f(\bar{x}), d \rangle \geq 0, \quad \forall d \in \mathcal{B}_D(\bar{x}). \quad (6)$$

Demonstração: Para $d = 0 \in \mathcal{B}_D(\bar{x})$, teremos $\langle f'(\bar{x}), d \rangle = \langle f'(\bar{x}), 0 \rangle = 0$ e vale a condição (6). Fixemos $d \in \mathcal{B}_D(\bar{x}) - \{0\}$ arbitrário e as sequências associadas $\{t_k\} \subset \mathbb{R}_+ - \{0\}$ e $\{d_k\} \subset \mathbb{R}^n$ tais que $\{t_k\} \rightarrow 0+$, $\{d_k\} \rightarrow d$ ($k \rightarrow \infty$) e $\bar{x} + t_k d_k \in D$ para todo k . Como \bar{x} é um minimizador local, temos que $\bar{x} + t_k d_k \in D$ e $\{\bar{x} + t_k d_k\} \rightarrow \bar{x}$ com ($k \rightarrow \infty$), para todo k suficientemente grande temos que

$$0 \leq f(\bar{x} + t_k d_k) - f(\bar{x}) = t_k \langle \nabla f(\bar{x}), d_k \rangle + o(t_k \|d_k\|)$$

Vamos tomar $t_k \|d_k\| = s_k$ e dividir ambos os lados da desigualdade acima por $t_k = \frac{s_k}{\|d_k\|}$:

$$\begin{aligned} 0 &\leq \frac{s_k}{\|d_k\|} \langle \nabla f(\bar{x}), d_k \rangle + o(s_k) \\ \implies 0 &\leq \langle \nabla f(\bar{x}), d_k \rangle + \|d_k\| \frac{o(s_k)}{s_k} \end{aligned}$$

Passando o limite quando $k \rightarrow \infty$, obtemos

$$0 \leq \langle \nabla f(\bar{x}), d \rangle,$$

que é (6), que queríamos provar.

□

Observamos que se \bar{x} é um minimizador local de f em D , então (6) implica a condição

$$\langle \nabla f(\bar{x}), d \rangle \geq 0, \quad \forall d \in \tau_D(\bar{x}),$$

já que $\tau_D(\bar{x}) \subset \mathcal{B}_D(\bar{x})$. No entanto, esta condição é mais fraca, considerando que nem sempre os dois cones são iguais.

Teorema 1.12. (Condições de otimalidade de primeira ordem no caso de conjunto viável convexo) Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e fechado, e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in D$. Se \bar{x} é um minimizador local de f no conjunto D , então

$$\langle f'(\bar{x}), x - \bar{x} \rangle \geq 0, \quad \forall x \in D, \quad (7)$$

ou, equivalentemente,

$$\bar{x} = P_D(\bar{x} - \alpha f'(\bar{x})), \quad \forall \alpha \in \mathbb{R}_+. \quad (8)$$

Se f é uma função convexa, $\bar{x} \in D$ e vale (7) (ou (8)), então \bar{x} é um minimizador de f em D . Além disso, neste caso (7) e (8) são equivalentes à seguinte condição:

$$\bar{x} = P_D(\bar{x} - \alpha f'(\bar{x})) \text{ para algum } \alpha > 0. \quad (9)$$

Demonstração: Note que $\{d \in \mathbb{R}^n | d = x - \bar{x}, x \in D\}$ é um conjunto de direções viáveis, pois sempre ocorre $x - \bar{x} \in D$, então $\{d \in \mathbb{R}^n | d = x - \bar{x}, x \in D\} \subset \nu_D(\bar{x})$. Por CITAR temos também que $\{d \in \mathbb{R}^n | d = x - \bar{x}, x \in D\} \subset \nu_D(\bar{x}) \subset \mathcal{B}_D(\bar{x})$. Além disso, pelo Teorema 1.11, $\langle f'(\bar{x}), d \rangle \geq 0$ para todo $d \in \mathcal{B}_D(\bar{x})$. Logo, vale

$$\langle f'(\bar{x}), x - \bar{x} \rangle \geq 0, \quad \forall x \in D. \quad (10)$$

Para mostrarmos que (7) é equivalente a (8) utilizaremos também a hipótese de D ser um conjunto fechado. Seja \bar{x} um minimizador local de f no conjunto D . Fixemos $\alpha \in \mathbb{R}_+$ qualquer. Pelo Teorema de Projeção, utilizando a relação (4) com $x = \bar{x} - \alpha \nabla f(\bar{x})$ e $y = \bar{x} \in D$, temos

$$\begin{aligned} & \langle \bar{x} - \alpha \nabla f(\bar{x}) - P_D(\bar{x} - \alpha \nabla f(\bar{x})), \bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x})) \rangle \leq 0 \\ \implies & \langle \bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x})), \bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x})) \rangle \leq -\langle -\alpha \nabla f(\bar{x}), \bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x})) \rangle \end{aligned}$$

De onde obtemos

$$\|\bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x}))\|^2 \leq \alpha \langle \nabla f(\bar{x}), \bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x})) \rangle. \quad (11)$$

Como $P_D(\bar{x} - \alpha \nabla f(\bar{x})) \in D$, pela desigualdade (10) temos que

$$\langle \nabla f(\bar{x}), P_D(\bar{x} - \alpha \nabla f(\bar{x})) - \bar{x} \rangle \geq 0 \implies \langle \nabla f(\bar{x}), \bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x})) \rangle \leq 0.$$

Como $\alpha \geq 0$, de (11) segue que

$$\begin{aligned} \|\bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x}))\|^2 \leq 0 & \implies \|\bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x}))\|^2 = 0 \\ & \implies \|\bar{x} - P_D(\bar{x} - \alpha \nabla f(\bar{x}))\| = 0 \\ & \implies P_D(\bar{x} - \alpha \nabla f(\bar{x})) = \bar{x}. \end{aligned}$$

Se \bar{x} é um minimizador (7) e (9) são satisfeitas por (11) e (8) respectivamente. Suponhamos que valha (7). Usando o item (b) do Teorema 1.1, obtemos que para qualquer $x \in D$,

$$f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq f(\bar{x}),$$

i. e., \bar{x} é um minimizador global.

Suponhamos que (9), i. e., \bar{x} é uma solução do problema

$$\min \psi(x) \text{ sujeito a } x \in D,$$

onde

$$\psi(x) = \frac{1}{2} \|x - (\bar{x} - \alpha f'(\bar{x}))\|^2, \alpha > 0.$$

Por (10), $\forall x \in D$, tem-se que

$$0 \leq \langle \psi'(\bar{x}), x - \bar{x} \rangle = \langle \bar{x} - (\bar{x} - \alpha f'(\bar{x})), x - \bar{x} \rangle = \alpha \langle f'(\bar{x}), x - \bar{x} \rangle.$$

Como $\alpha > 0$, vale (7). Como já mostramos, isto implica que \bar{x} é minimizador global. \square

2 Buscas lineares

2.1 Buscas lineares para o caso com restrições

Busca da minimização unidimensional. No caso da *minimização unidimensional*, α_k é uma solução do problema

$$\min \varphi_k(\alpha) \text{ sujeito a } \alpha \in \mathbb{R}_+, \quad (12)$$

ou do problema

$$\min \varphi_k(\alpha) \text{ sujeito a } \alpha \in [0, \hat{\alpha}],$$

onde $\hat{\alpha} > 0$ é um parâmetro.

Observação 2.1. *No entanto, como no caso irrestrito, as regras de minimização unidimensional não são muito atrativas do ponto de vista prático, pois a resolução de um problema unidimensional em cada iteração de um método, mesmo que seja inexata, é relativamente cara. Mais ainda, cabe frisar que o nosso objetivo eventual tem a ver com minimização de f no \mathbb{R}^n , e não numa semireta (fixa na iteração k). Por isso, não vale “perder tempo” insistindo em resolver o problema unidimensional (12) com precisão. Na prática, outras regras de busca linear (desenvolvidas a seguir) são mais úteis e mais usadas. Essencialmente, elas buscam assegurar decréscimo suficiente da função f em relação ao valor $f(x_k)$, sem se preocupar com a resolução do problema (12) em si.*

Busca de Armijo. Mais interessantes são as extensões da *regra de Armijo*. Isto pode ser feito, por exemplo, pelo mesmo esquema que foi considerado anteriormente, substituindo a condição de Armijo por

$$f(x_k(\alpha)) \leq f(x_k) + \eta \langle \nabla f(x_k), x_k(\alpha) - x_k \rangle \quad (13)$$

$$f(P_D(x_k - \alpha \nabla f(x_k))) \leq f(x_k) + \eta \langle \nabla f(x_k), P_D(x_k - \alpha \nabla f(x_k)) - x_k \rangle.$$

Relembramos que, para começar o processo, devem ser escolhidos um parâmetro $\eta \in (0, 1)$, valor inicial do comprimento do passo $\hat{\alpha} > 0$ e o parâmetro de redução do passo $\theta \in (0, 1)$. Em particular, α_k é calculado como o maior entre todos os números de forma $\alpha = \hat{\alpha} \theta_s$, com $s \in \mathbb{N}$, satisfazendo a desigualdade (13). Como é fácil ver, no caso quando $D = \mathbb{R}^n$, a desigualdade (13) se reduz à condição de Armijo no caso irrestrito.

Outras buscas lineares. Existem também várias modificações da busca de Armijo, por exemplo, similares à *busca de Goldstein*.

Poderíamos considerar também a *busca de comprimento de passo fixo*, onde $\alpha_k = \bar{\alpha}$ para todo k , com $\bar{\alpha} > 0$ fixo.

2.2 Busca linear não monótona de Grippo-Lampariello-Lucidi

O estudo feito nessa seção se baseia em [4]. Neste artigo, é abordado uma generalização da Regra de Armijo ([1]) que é um novo critério para admissão de comprimentos de passo. Em particular, é mostrado que a condição que implica o decréscimo monotônico de $f(x)$ pode ser relaxado e ainda ter convergência global estabelecida.

Essa busca linear depende de um parâmetro inteiro $M \geq 1$ e impõe um decréscimo no valor da função a cada M iterações (se $M = 1$ então esta busca linear reduz-se a uma busca linear monótona). Considere o seguinte resultado:

Teorema 2.2. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função continuamente diferenciável (f é de classe \mathcal{C}^1) e $\{x_k\}$ uma sequência definida por*

$$x_{k+1} = x_k + \alpha_k d_k, \quad d_k \neq 0.$$

Seja $a > 0$, $\sigma, \eta \in (0, 1)$ e M um número inteiro não negativo ($M > 0$). Suponha que:

- (i) *o conjunto de nível $D_0 = \{x \in \mathbb{R}^n; f(x) \leq f(x_0)\}$ é compacto;*
- (ii) *existem números positivos c_1, c_2 tais que:*

$$\nabla f(x_k)^T d_k \leq -c_1 \|\nabla f(x_k)\|^2, \quad (14)$$

$$\|d_k\| \leq c_2 \|\nabla f(x_k)\|; \quad (15)$$

- (iii) *$\alpha_k = \sigma^{h_k} a$, onde h_k é o primeiro número inteiro não negativo h para o qual vale*

$$f(x_k + \sigma^h a d_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \eta \sigma^h a \nabla f(x_k)^T d_k, \quad (16)$$

onde $m(0) = 0$ e $0 \leq m(k) \leq \min\{m(k-1) + 1, M\}$, $k \geq 1$.

Então:

- (a) *a sequência $\{x_k\}$ permanece em D_0 e todo ponto de limite \bar{x} satisfaz $\nabla f(\bar{x}) = 0$;*
- (b) *nenhum ponto limite de $\{x_k\}$ é um máximo local de f ;*
- (c) *se o número de pontos estacionários de f em D_0 for finito, a sequência $\{x_k\}$ converge.*

Demonstração: Dado $k \in \mathbb{N}$, definamos um índice inteiro $l(k) \in [k - m(k), k]$ (ou $k - m(k) \leq l(k) \leq k$) tal que

$$f(x_{l(k)}) = \max_{0 \leq j \leq m(k)} [f(x_{k-j})]. \quad (17)$$

Segue, desta definição e de (16), que

$$f(x_{k+1}) = f(x_k + \sigma^h a d_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \eta \sigma^h a \nabla f(x_k)^T d_k = f(x_{l(k)}) + \eta \sigma^h a \nabla f(x_k)^T d_k$$

Por (14), temos que $f(x_{l(k)}) + \eta \sigma^h a \nabla f(x_k)^T d_k < f(x_{l(k)})$, logo

$$f(x_{k+1}) < f(x_{l(k)}). \quad (18)$$

Como $m(k+1) \leq \min\{m((k+1)-1)+1, M\} = \min\{m(k)+1, M\} \leq m(k)+1$ (se $M > m(k)+1$, $m(k)+1 \geq \min\{m(k)+1, M\}$; se $M < m(k)+1$, $m(k)+1 \geq \min\{m(k)+1, M\}$), podemos escrever que, para todo $k \in \mathbb{N}$,

$$\begin{aligned}
f(x_{l(k+1)}) &= \max_{0 \leq j \leq m(k+1)} [f(x_{(k+1)-j})] \leq \max_{0 \leq j \leq m(k)+1} [f(x_{(k+1)-j})] \\
&= \max_{0 \leq i+1 \leq m(k)+1} [f(x_{(k+1)-(i+1)})] \\
&= \max_{0 \leq i+1 \leq m(k)+1} [f(x_{k-i})] \\
&= \max_{0 \leq i \leq m(k)} [f(x_{k-i})] \\
&= \max_{0 \leq i \leq m(k)} \{f(x_{k-i}), f(x_{k+1})\} \\
&= \max[f(x_{l(k)}), f(x_{k+1})] \\
&= f(x_{l(k)}),
\end{aligned}$$

i. e., temos que

$$f(x_{l(k+1)}) \leq f(x_{l(k)}). \quad (19)$$

Logo, a sequência $\{f(x_{l(k)})\}$ é não crescente. Assim, segue de (17) e de (19) que, para todo $k \in \mathbb{N}$,

$$f(x_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] = f(x_{l(k)}) \leq f(x_{l(0)}) = f(x_0)$$

Portanto, $\{x_k\} \subset D_0$, o que demonstra a primeira afirmação do item (a).

Agora, vamos provar a segunda afirmação do item (a). Pela compacidade de D_0 e pela continuidade e monotonicidade de f , temos que $\{f(x_{l(k)})\}$ admite um limite ℓ quando $k \rightarrow \infty$. Assim, obtemos de (17) e da definição da sequência $\{x_k\}$ que, para $k > M$,

$$\begin{aligned}
f(x_{l(k)}) &= f(x_{l(k)-1} + \alpha_{l(k)-1} d_{l(k)-1}) \\
&\leq \max_{0 \leq j \leq m(l(k)-1)} [f(x_{l(k)-1-j})] + \eta \alpha_{l(k)-1} \nabla f(x_{l(k)-1})^T d_{l(k)-1} \\
&= f(x_{l(l(k)-1)}) + \eta \alpha_{l(k)-1} \nabla f(x_{l(k)-1})^T d_{l(k)-1}.
\end{aligned}$$

Tomando o limite quando $k \rightarrow \infty$ na desigualdade anterior, obtemos

$$\ell \leq \ell + \eta \lim_{k \rightarrow \infty} \alpha_{l(k)-1} \nabla f(x_{l(k)-1})^T d_{l(k)-1} \implies \lim_{k \rightarrow \infty} \alpha_{l(k)-1} \nabla f(x_{l(k)-1})^T d_{l(k)-1} \geq 0.$$

Por outro lado, como $\alpha_k > 0$ e $\nabla f(x_k)^T d_k < 0$, $\forall k \in \mathbb{N}$, segue que

$$\lim_{k \rightarrow \infty} \alpha_{l(k)-1} \nabla f(x_{l(k)-1})^T d_{l(k)-1} \leq 0.$$

Portanto, só pode ocorrer

$$\lim_{k \rightarrow \infty} \alpha_{l(k)-1} \nabla f(x_{l(k)-1})^T d_{l(k)-1} = 0. \quad (20)$$

Pela hipótese (15), temos, para todo $k \in \mathbb{N}$,

$$\begin{aligned} \|d_k\| \leq c_2 \|\nabla f(x_k)\| &\implies \|d_k\|^2 \leq c_2^2 \|\nabla f(x_k)\|^2 \\ &\implies \frac{1}{c_2^2} \|d_k\|^2 \leq \|\nabla f(x_k)\|^2 \\ &\implies \frac{-c_1}{c_2^2} \alpha_k \|d_k\|^2 \geq -c_1 \alpha_k \|\nabla f(x_k)\|^2. \end{aligned}$$

Utilizando a hipótese (14), temos que, para todo k ,

$$\nabla f(x_k)^T d_k \leq -c_1 \|\nabla f(x_k)\|^2 \implies \alpha_k \nabla f(x_k)^T d_k \leq -c_1 \alpha_k \|\nabla f(x_k)\|^2 \leq \frac{-c_1}{c_2^2} \alpha_k \|d_k\|^2.$$

Logo,

$$\alpha_{l(k)-1} \nabla f(x_{l(k)-1})^T d_{l(k)-1} \leq \frac{-c_1}{c_2^2} \alpha_{l(k)-1} \|d_{l(k)-1}\|^2 \leq 0.$$

Tomando o limite quando $k \rightarrow \infty$ e utilizando (20), obtemos

$$\lim_{k \rightarrow \infty} \alpha_{l(k)-1} \|d_{l(k)-1}\|^2 = 0. \quad (21)$$

Suponhamos, por absurdo, que $\lim_{k \rightarrow \infty} \alpha_{l(k)-1} \|d_{l(k)-1}\| \neq 0$. Então, como visto em (21), temos que $\lim_{k \rightarrow \infty} \|d_{l(k)-1}\| = 0$, e logo, como $\alpha_k \leq a$, temos que $\lim_{k \rightarrow \infty} \alpha_{l(k)-1} \|d_{l(k)-1}\| = 0$, o que contradiz a hipótese. Portanto,

$$\lim_{k \rightarrow \infty} \alpha_{l(k)-1} \|d_{l(k)-1}\| = 0. \quad (22)$$

Vamos provar agora que $\lim_{k \rightarrow \infty} \alpha_k \|d_k\| = 0$. Seja

$$\hat{l}(k) := l(k + M + 2), \forall k \in \mathbb{N}.$$

Primeiro vamos provar, por indução em j , que

$$\lim_{k \rightarrow \infty} \alpha_{\hat{l}(k)-j} \|d_{\hat{l}(k)-j}\| = 0 \text{ e} \quad (23)$$

$$\lim_{k \rightarrow \infty} f(x_{\hat{l}(k)-j}) = \lim_{k \rightarrow \infty} f(x_{l(k)}), \quad j = 1, 2, 3, \dots \quad (24)$$

Aqui e na sequência assumimos, sem perda de generalidade, que o índice de iteração k é suficientemente grande para evitar a ocorrência de índices negativos, ou seja, $k \geq j - 1$ (ou $\hat{l}(k) \geq j$).

- Se $j = 1$, como $\{\hat{l}(k)\} \subset \{l(k)\}$, (23) segue de (22). Como D_0 é compacto, f é uniformemente contínua em D_0 , então dado $\varepsilon > 0$, existe $\delta > 0$ tal que $\|x - y\| < \delta$ implica $|f(x) - f(y)| < \varepsilon$, com $x, y \in D_0$. Além disso, por (23), temos que $\|x_{\hat{l}(k)} - x_{\hat{l}(k)-1}\| = \|\alpha_{\hat{l}(k)-1} d_{\hat{l}(k)-1}\| = \alpha_{\hat{l}(k)-1} \|d_{\hat{l}(k)-1}\| \rightarrow 0$. Portanto, $\lim_{k \rightarrow \infty} x_{\hat{l}(k)} - x_{\hat{l}(k)-1} = 0$. Assim, dado $\delta > 0$, existe $k_0 \in \mathbb{N}$ tal que $k \geq k_0$ implica $\|x_{\hat{l}(k)} - x_{\hat{l}(k)-1}\| \leq \delta$. Onde existe $k_0 \in \mathbb{N}$ tal que $|f(x_{\hat{l}(k)}) - f(x_{\hat{l}(k)-1})| < \varepsilon$. Portanto, $\lim_{k \rightarrow \infty} f(x_{\hat{l}(k)}) - f(x_{\hat{l}(k)-1}) = 0$ e (24) também é válida.

3 Métodos do Gradiente Espectral Projetado

Neste seção, nosso objeto de estudo é o Método do Gradiente Espectral Projetado (SPG) para a minimização de funções diferenciáveis em conjuntos não vazios, fechados e convexos (“conjuntos com estrutura simples”). Neste método, estamos interessados na propriedade de manter a viabilidade das iterações projetando frequentemente os pontos de teste no conjunto convexo viável.

Os algoritmos de projeção do gradiente espectral não monótono se aplicam a problemas da forma

$$\text{minimizar } f(x) \text{ sujeito a } x \in \Omega, \quad (25)$$

onde Ω é um conjunto convexo fechado em \mathbb{R}^n e f é uma função definida que possui derivadas parciais contínuas em um conjunto aberto que contém Ω .

O método do Gradiente Espectral Projetado, proposto por [3], representa uma variante do método do Gradiente Projetado. Ele incorpora duas adaptações em relação a este método: a utilização do passo espectral proposto por [2] e uma busca linear não monótona.

3.1 Passo espectral de Barzilai e Borwein

Um artigo pioneiro por Barzilai e Borwein (1988) propõe um método do gradiente para minimização irrestrita de funções diferenciáveis $f : \mathbb{R}^n \rightarrow \mathbb{R}$ que usa uma estratégia nova e fora do padrão para escolha do comprimento de passo.

Dado $x_0 \in \mathbb{R}$, a iteração Barzilai-Borwein (BB) é dada por

$$x_{k+1} = x_k - \lambda_k \nabla f(x_k), \quad (26)$$

onde o comprimento de passo inicial $\lambda_0 > 0$ é arbitrário e, para todo $k = 1, 2, \dots$,

$$\lambda_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} = \frac{(x_k - x_{k-1})^T (x_k - x_{k-1})}{(x_k - x_{k-1})^T (\nabla f(x_k) - \nabla f(x_{k-1}))}, \quad (27)$$

onde $s_{k-1} = x_k - x_{k-1}$ e $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$.

Quando $f(x) = \frac{1}{2}x^T Ax + b^T x + c$ é uma função quadrática e A é uma matriz simétrica definida positiva ($A = A^T$ e $x^T Ax > 0, \forall x \neq 0$), temos que $\nabla f(x) = Ax + b$, $\nabla f(x_k) - \nabla f(x_{k-1}) = Ax_k + b - Ax_{k-1} - b = A(x_k - x_{k-1}) = As_{k-1}$ e $s_{k-1} = x_k - x_{k-1} = (x_{k-1} - \lambda_{k-1} \nabla f(x_{k-1})) - x_{k-1} = -\lambda_{k-1} \nabla f(x_{k-1})$. Assim, o comprimento de passo (27) se torna

$$\lambda_k = \frac{(-\lambda_{k-1} \nabla f(x_{k-1}))^T (-\lambda_{k-1} \nabla f(x_{k-1}))}{(-\lambda_{k-1} \nabla f(x_{k-1}))^T A (-\lambda_{k-1} \nabla f(x_{k-1}))} \quad (28)$$

$$= \frac{\nabla f(x_{k-1})^T \nabla f(x_{k-1})}{\nabla f(x_{k-1})^T A \nabla f(x_{k-1})}. \quad (29)$$

Por isso, o método BB calcula, a cada iteração, o comprimento de passo que minimiza a função quadrática objetivo ao longo da direção negativa do gradiente, mas em vez de usar este comprimento de passo na k -ésima iteração, salva o comprimento de passo para ser usado na próxima iteração. A estratégia proposta por BB para a escolha de um escalar λ

consiste na busca por “corrigir” a escala da direção do gradiente no sentido oposto. Essa estratégia foi originalmente desenvolvida para o caso de uma função quadrática irrestrita, ao observarem a importância que os tamanhos de passo próximos aos autovalores associados a matriz Hessiana exercem sobre a redução das componentes do gradiente.

No Método do Gradiente Espectral, os coeficientes λ_k são devidamente limitados, ou seja, o método usa parâmetros de salvaguarda $0 < \lambda_{\min} < \lambda_{\max} < \infty$ e define, a cada iteração

$$\lambda_k = \max\{\lambda_{\min}, \min\{\frac{s_k^T s_k}{s_k^T y_k}, \lambda_{\max}\}\}. \quad (30)$$

Observação: A ideia principal por trás da escolha espectral de comprimento de passo é que o método de descida mais íngreme é muito lento, mas pode ser acelerado tomando, em vez do tamanho do passo que resulta da minimização da função ao longo do gradiente da iteração atual, aquele que vem da minimização unidimensional no passo anterior.

3.2 Algoritmos dos Métodos do Gradiente Espectral Projetado

No primeiro método do Gradiente Espectral Projetado, que chamaremos de SPG1, utilizaremos a primeira estratégia, proposta por [4], que é realizada ao longo do arco de projeção. Neste caso, dados $\eta \in (0, 1)$ e um inteiro $M > 0$, busca-se determinar um comprimento de passo $\alpha_k > 0$ tal que

$$f(x_k(\alpha_k)) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \eta \nabla f(x_k)^\top d_k, \quad (31)$$

onde $d_k = x_k(\alpha_k) - x_k$.

Nota-se que, em (31), o cálculo do produto $\nabla f(x_k)^\top (x_k(\alpha_k) - x_k)$ é feito para cada ponto teste $x_k(\alpha_k)$. Segundo Birgin et al. [3], este emprego sucessivo de projeções causa um aumento no esforço computacional, o que os levou à formulação de uma nova busca linear não monótona que é empregada em um novo método do Gradiente Espectral Projetado. Essa busca consiste em, dados $\eta \in (0, 1)$ e um inteiro $M > 0$, determinar $\alpha_k > 0$ tal que

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \eta \alpha_k \nabla f(x_k)^\top d_k, \quad (32)$$

com $d_k = x_k(\lambda_k) - x_k$, onde λ_k é o passo espectral. Nessa busca, ao rejeitar o primeiro ponto teste, os próximos são calculados ao longo da mesma direção viável. Isso resulta em um único cálculo da direção d_k e uma única operação de projeção por iteração. Denominaremos essa outra estratégia de SPG2.

O Algoritmo 1 a seguir apresenta formalmente os métodos SPG1 e SPG2.

Algoritmo 1: MÉTODO DO GRADIENTE ESPECTRAL PROJETADO (SPG)

- 1 Considere $x_0 \in \Omega$, $\varepsilon > 0$, $\eta \in (0, 1)$, $1 > \sigma_2 > \sigma_1 > 0$, um inteiro $M > 0$, $\lambda_{\max} > \lambda_{\min} > 0$, $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$ e $k = 0$.
 - 2 Se $\|x_k(1) - x_k\|_\infty < \varepsilon$, então pare e declare que x_k é estacionário.
 - 3 Comece com $\alpha_k = \lambda_k$, calcule α_k que satisfaça (31) e faça $x_{k+1} = x_k(\alpha_k)$ para o SPG1; ou comece com $\alpha_k = 1$, calcule α_k que satisfaça (32) e faça $x_{k+1} = x_k + \alpha_k(x_k(\lambda_k) - x_k)$ para o SPG2.
 - 4 Se (31) não for válida ou (32) não for válida, tome $\alpha_{\text{new}} \in [\sigma_1\alpha_k, \sigma_2\alpha_k]$, faça $\alpha_k = \alpha_{\text{new}}$ e retorne para o passo 3.
 - 5 Calcule $s_k = x_{k+1} - x_k$ e $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.
 - 6 Se $s_k^\top y_k \leq 0$, então faça $\lambda_{k+1} = \lambda_{\max}$. Caso contrário, faça $\lambda_{k+1} = \max\{\lambda_{\min}, \min\{s_k^\top s_k / s_k^\top y_k, \lambda_{\max}\}\}$. Faça $k = k + 1$ e retorne para o passo 2.
-

3.3 Resultados de convergência

Será útil em nossa análise teórica definir o gradiente projetado em escala $\nabla f_t(x)$ como

$$\nabla f_t(x) = [P_\Omega(x - t\nabla f(x)) - x],$$

para todo $x \in \Omega$ e $t > 0$.

- Se x for uma iteração do SPG1 ou do SPG2 e $t = \lambda_k$ (passo espectral BB), então o valor do gradiente projetado em escala é o gradiente espectral projetado (SPG).
- Se $t = 1$, o gradiente projetado em escala é o gradiente projetado contínuo cuja norma infinito $\|\nabla f_1(x)\|_\infty$ é usada para o critério de parada dos algoritmos.

Na verdade, zerar $\nabla f_t(x)$ é equivalente à satisfação de condições estacionárias de primeira ordem. Esta propriedade é enunciada no seguinte lema, cuja prova é consequência da convexidade de Ω .

Lema 3.1. *Para todo $x \in \Omega$ e $t \in (0, \lambda_{\max}]$, vale:*

(i)

$$\nabla f(x)^T \nabla f_t(x) \leq -\frac{1}{t} \|\nabla f_t(x)\|^2 \leq -\frac{1}{\lambda_{\max}} \|\nabla f_t(x)\|^2;$$

(ii) *O vetor $\nabla f_t(\bar{x})$ é nulo se, e somente se, \bar{x} for um ponto estacionário restrito.*

(i) *Demonstração.* Pelo Teorema da Projeção (1.8) e considerando $x = x - t\nabla f(x)$, $y = x$ e $\bar{x} = P_\Omega(x) = P_\Omega(x - t\nabla f(x))$, temos

$$[P_\Omega(x - t\nabla f(x)) - (x - t\nabla f(x))]^T [P_\Omega(x - t\nabla f(x)) - x] \leq 0.$$

Como $\nabla f_t(x) = P_\Omega(x - t\nabla f(x)) - x$, a desigualdade anterior se torna

$$[(P_\Omega(x - t\nabla f(x)) - x) + t\nabla f(x)]^T \nabla f_t(x) \leq 0 \implies [\nabla f_t(x) + t\nabla f(x)]^T \nabla f_t(x) \leq 0.$$

Pela linearidade do produto interno, temos

$$\begin{aligned}\nabla f_t(x)^T \nabla f_t(x) + t \nabla f(x)^T \nabla f_t(x) &\leq 0 \implies (t f(x))^T \nabla f_t(x) \leq -\nabla f_t(x)^T \nabla f_t(x) \\ &\implies f(x)^T \nabla f_t(x) \leq -\frac{1}{t} \nabla f_t(x)^T \nabla f_t(x) \\ &\implies f(x)^T \nabla f_t(x) \leq -\frac{1}{t} \|\nabla f_t(x)\|^2.\end{aligned}$$

Como $t \in (0, \lambda_{\max}]$, $t \leq \lambda_{\max} \implies -\lambda_{\max} \leq -t \implies -\frac{1}{t} \leq -\frac{1}{\lambda_{\max}}$ e temos a seguinte desigualdade:

$$\nabla f(x)^T \nabla f_t(x) \leq -\frac{1}{t} \|\nabla f_t(x)\|^2 \leq -\frac{1}{\lambda_{\max}} \|\nabla f_t(x)\|^2.$$

(ii) *Demonstração.*

(\implies) Se $\nabla f_t(\bar{x})$ é nulo, então $P_\Omega(\bar{x} - t \nabla f(\bar{x})) - \bar{x} = 0 \implies P_\Omega(\bar{x} - t \nabla f(\bar{x})) = \bar{x}$. Daí, vale a condição de ponto estacionário $\nabla f(\bar{x})^T (P_\Omega(\bar{x} - t \nabla f(\bar{x})) - \bar{x}) \leq 0$.

(\impliedby) Se \bar{x} é um estacionário restrito, então pelo Teorema de condição necessária de primeira ordem, $P_\Omega(\bar{x} - t \nabla f(\bar{x})) = \bar{x} \implies P_\Omega(\bar{x} - t \nabla f(\bar{x})) - \bar{x} = 0 \implies \nabla f_t(\bar{x}) = 0$. \square

Lema 3.2. (i) Para todo $x \in \Omega$ e $z \in \mathbb{R}^n$, a função $h : [0, \infty) \rightarrow \mathbb{R}$ dado por

$$h(s) = \frac{\|P_\Omega(x + sz) - x\|}{s}, \text{ para todo } s > 0,$$

é monótona não crescente.

(ii) Para todo $x \in \Omega$ existe $s_x > 0$ tal que para todo $t \in [0, s_x]$ é satisfeito

$$f(P_\Omega(x - t \nabla f(x))) - f(x) \leq \eta \nabla f(x)^T \nabla f_t(x) \quad (33)$$

(i) *Demonstração.* Tomemos dois escalares s_1 e s_2 com $s_1 > 0$ e $s_2 > s_1$, e vamos mostrar que

$$h(s_2) = \frac{\|P_\Omega(x + s_2 z) - x\|}{s_2} \leq \frac{\|P_\Omega(x + s_1 z) - x\|}{s_1} = h(s_1). \quad (34)$$

Definindo

$$y = s_1 z, \gamma = \frac{s_2}{s_1}, a = x + y, b = x + \gamma y,$$

a desigualdade anterior é escrita como

$$\|\bar{b} - x\| \leq \gamma \|\bar{a} - x\|, \quad (35)$$

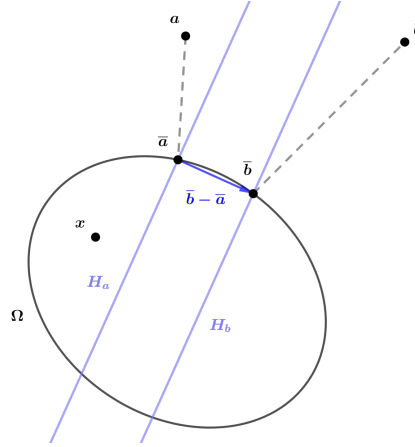
onde \bar{a} e \bar{b} são as projeções de a e b em D , respectivamente. Vamos dividir em casos e provar que a desigualdade anterior vale para qualquer caso.

1. Se $\bar{a} = x$, temos $\|\bar{b} - x\| \leq \gamma \cdot 0 \implies \|\bar{b} - x\| \leq 0 \implies \bar{b} = x$, então a Equação (35) é válida.
2. Se $a \in \Omega$, temos $\bar{a} = a = x + y$, então da Equação (35) se torna $\|\bar{b} - x\| \leq \gamma \|(x + y) - x\| = \gamma \|y\| = \|(x + \gamma y) - x\| = \|b - x\|$, o que é válido porque a projeção é não expansiva.

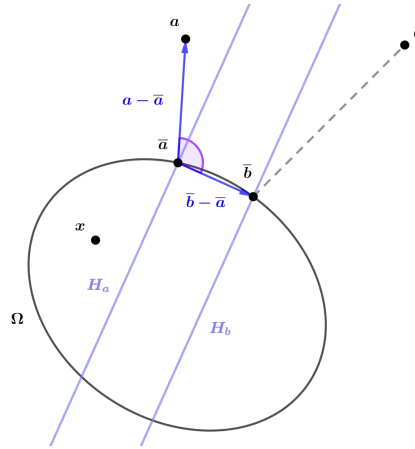
3. Se $\bar{a} = \bar{b}$, então a Equação (35) também vale.

4. Portanto, basta provar para o caso em que $\bar{a} \neq \bar{b}$, $\bar{a} \neq x$, $\bar{b} \neq x$, $a \notin \Omega$.

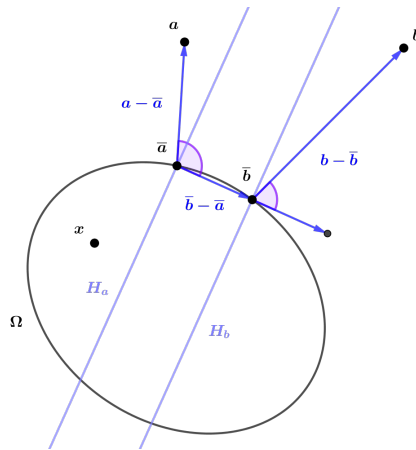
Para provar o item 4, faremos primeiro uma construção geométrica. Sejam H_a e H_b dois hiperplanos ortogonais para $\bar{b} - \bar{a}$ e que passam por \bar{a} e \bar{b} , respectivamente.



Note que, como o ângulo formado entre os vetores $\bar{b} - \bar{a}$ e $a - \bar{a}$ é obtuso, temos que $[\bar{b} - \bar{a}]^T [a - \bar{a}] \leq 0$.

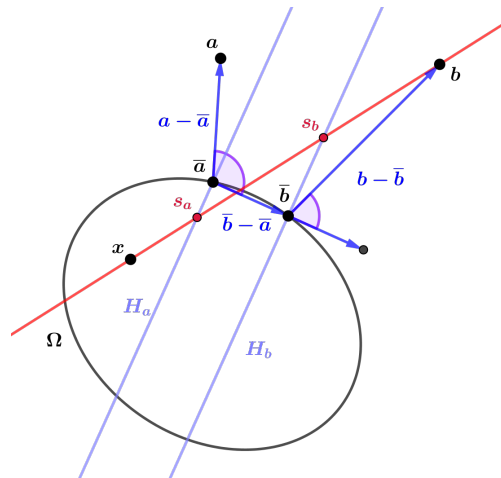


Além disso, tomando um vetor equipolente a $\bar{b} - \bar{a}$ (devido ao sentido do vetor), temos que o ângulo formado entre os vetores $\bar{b} - \bar{a}$ e $b - \bar{b}$ é agudo. Assim, $[\bar{b} - \bar{a}]^T [b - \bar{b}] \geq 0$.

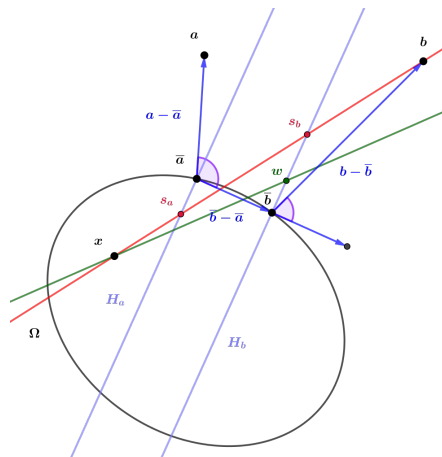


Logo, como $[\bar{b} - \bar{a}]^T [b - \bar{b}] \geq 0$ e $[\bar{b} - \bar{a}]^T [a - \bar{a}] \leq 0$, temos que nem a nem b ficam estritamente entre os dois hiperplanos H_a e H_b . Além disso, x está no mesmo lado de H_a que a , então $x \notin H_a$.

Agora, denote as interseções da reta $\{x + \alpha(b - x); \alpha \in \mathbb{R}\}$ com H_a e H_b por s_a e s_b , respectivamente.



Denote a interseção da reta $\{x + \alpha(\bar{a} - x); \alpha \in \mathbb{R}\}$ com H_b por w .



Temos que

$$\gamma = \frac{\|b - x\|}{\|a - x\|} \geq \frac{\|s_b - x\|}{\|s_a - x\|} = \frac{\|w - x\|}{\|\bar{a} - x\|}$$

Pela semelhança de triângulos

$$\frac{\|w - x\|}{\|\bar{a} - x\|} = \frac{\|w - \bar{a}\| + \|\bar{a} - x\|}{\|\bar{a} - x\|}.$$

Como $[w - \bar{b}]^T[\bar{b} - \bar{a}] = 0$, temos

$$\frac{\|w - \bar{a}\| + \|\bar{a} - x\|}{\|\bar{a} - x\|} \geq \frac{\|\bar{b} - \bar{a}\| + \|\bar{a} - x\|}{\|\bar{a} - x\|}.$$

Por fim, pela desigualdade triangular, temos

$$\gamma \geq \frac{\|\bar{b} - \bar{a}\| + \|\bar{a} - x\|}{\|\bar{a} - x\|} \geq \frac{\|\bar{b} - \bar{a} + \bar{a} - x\|}{\|\bar{a} - x\|} = \frac{\|\bar{b} - x\|}{\|\bar{a} - x\|}.$$

Daí, obtemos

$$\gamma \geq \frac{\|\bar{b} - x\|}{\|\bar{a} - x\|} \implies \|\bar{b} - x\| \leq \gamma \|\bar{a} - x\|.$$

(ii) *Demonstração.* Pelo Teorema da Projeção, para todo $x \in \Omega$ e $s > 0$, temos

$$\begin{aligned} & [x - s\nabla f(x) - P(x - s\nabla f(x))]^\top [x - P(x - s\nabla f(x))] \leq 0 \\ \implies & [-s\nabla f(x) + (x - P(x - s\nabla f(x)))]^\top [x - P(x - s\nabla f(x))] \leq 0 \\ \implies & [-s\nabla f(x)]^\top [x - P(x - s\nabla f(x))] + [x - P(x - s\nabla f(x))]^\top [x - P(x - s\nabla f(x))] \leq 0 \\ \implies & [-s\nabla f(x)]^\top [x - P(x - s\nabla f(x))] + \|x - P(x - s\nabla f(x))\|^2 \leq 0 \\ \implies & -\nabla f(x)^\top [x - P(x - s\nabla f(x))] \leq -\frac{\|x - P(x - s\nabla f(x))\|^2}{s} \\ \implies & \nabla f(x)^\top [x - P(x - s\nabla f(x))] \geq \frac{\|x - P(x - s\nabla f(x))\|^2}{s}. \end{aligned}$$

Se x é estacionário, por (9), temos $\bar{x} = P(\bar{x} - s\nabla f(\bar{x}))$ para algum $s > 0$. Assim, $\|x - P(x - s\nabla f(x))\| = 0$ e a desigualdade é válida com s_x sendo qualquer escalar positivo.

Então, suponha que x não é estacionário e, portanto, $\|x - P(x - s\nabla f(x))\| \neq 0$, para todo $s > 0$. Pelo Teorema do Valor Médio, temos que para todo $x \in \Omega$ e $s \geq 0$,

$$f(x) - f(P(x - s\nabla f(x))) = \nabla f(x)^\top [x - P(x - s\nabla f(x))] + [\nabla f(q_s) - \nabla f(x)]^\top [x - P(x - s\nabla f(x))],$$

onde q_s está no segmento de reta definido por x e $P(x - s\nabla f(x))$.

Portanto, (33) pode ser escrito como

$$\begin{aligned}
& f(P(x - t\nabla f(x))) - f(x) \leq \eta \nabla f(x)^T \nabla f_s(x) \\
\implies & f(x) - f(P(x - t\nabla f(x))) \geq \eta \nabla f(x)^T (-\nabla f_s(x)) \\
\implies & \nabla f(x)^\top [x - P(x - s\nabla f(x))] + [\nabla f(q_s) - \nabla f(x)]^\top [x - P(x - s\nabla f(x))] \\
& \geq \eta \nabla f(x)^\top [x - P(x - s\nabla f(x))] \\
\implies & [\nabla f(q_s) - \nabla f(x)]^\top [x - P(x - s\nabla f(x))] \geq \eta \nabla f(x)^\top [x - P(x - s\nabla f(x))] \\
& \quad - \nabla f(x)^\top [x - P(x - s\nabla f(x))] \\
\implies & [\nabla f(x) - \nabla f(q_s)]^\top [x - P(x - s\nabla f(x))] \leq \nabla f(x)^\top [x - P(x - s\nabla f(x))] \\
& \quad - \eta \nabla f(x)^\top [x - P(x - s\nabla f(x))] \\
\implies & [\nabla f(x) - \nabla f(q_s)]^\top [x - P(x - s\nabla f(x))] \leq -\eta \nabla f(x)^\top [x - P(x - s\nabla f(x))].
\end{aligned}$$

Assim, temos

$$\nabla f(x)^\top [x - P(x - s\nabla f(x))] \geq \frac{\|x - P(x - s\nabla f(x))\|^2}{s}.$$

Para $s \in (0, 1]$ e como, por (i), temos $h(s) = \frac{\|x - P(x - s\nabla f(x))\|}{s}$ não crescente:

$$\begin{aligned}
\nabla f(x)^\top [x - P(x - s\nabla f(x))] & \geq \frac{\|P(x - s\nabla f(x))\|^2}{s} \\
& = h(s) \cdot \|x - P(x - s\nabla f(x))\| \\
& \geq h(1) \cdot \|x - P(x - s\nabla f(x))\| \\
& = \|x - P(x - \nabla f(x))\| \|x - P(x - s\nabla f(x))\|
\end{aligned}$$

Disto, segue que

$$-\nabla f(x)^\top [x - P(x - s\nabla f(x))] \leq -\|x - P(x - \nabla f(x))\| \|x - P(x - s\nabla f(x))\|.$$

Retomando o que tínhamos anteriormente:

$$\begin{aligned}
[\nabla f(x) - \nabla f(q_s)]^\top [x - P(x - s\nabla f(x))] & \leq -\eta \nabla f(x)^\top [x - P(x - s\nabla f(x))] \\
& \leq -\eta \|x - P(x - \nabla f(x))\| \|x - P(x - s\nabla f(x))\| \\
& \leq -\eta \|x - P(x - \nabla f(x))\| \|x - P(x - s\nabla f(x))\| \\
& \quad + \|x - P(x - \nabla f(x))\| \|x - P(x - s\nabla f(x))\| \\
& = (1 - \eta) \|x - P(x - \nabla f(x))\| \|x - P(x - s\nabla f(x))\|.
\end{aligned}$$

Portanto,

$$[\nabla f(x) - \nabla f(q_s)]^\top \frac{[x - P(x - s\nabla f(x))]}{\|x - P(x - s\nabla f(x))\|} \leq (1 - \eta) \|x - P(x - \nabla f(x))\|.$$

Assim, existe $s_x > 0$ tal que a relação acima ocorre e, portanto, as outras equações são satisfeitas para $s \in (0, s_x]$.

□

Teorema 3.3. *O Algoritmo SPG1 está bem definido e qualquer ponto de acumulação da sequência gerada por ele é um ponto estacionário restrito.*

Demonstração. Pelo Lema 3.2 (ii), temos que para todo $\alpha \in [0, \min\{s_{x_k}, \lambda_{\min}\}]$, vale

$$\begin{aligned} f(P(x_k - \alpha \nabla f(x_k))) - \max_{0 \leq j \leq M-1} f(x_{k-j}) &\leq f(P(x_k - \alpha \nabla f(x_k))) - f(x_k) \\ &\leq \gamma \nabla f(x_k)^\top \nabla f_\alpha(x_k) \\ &= \gamma \nabla f(x_k)^\top [P(x_k - \alpha \nabla f(x_k)) - x_k]. \end{aligned}$$

Portanto, um comprimento de passo que satisfaça a busca GLL será encontrado após um número finito de tentativas, e o Algoritmo SPG1 está bem definido.

Seja $\bar{x} \in \Omega$ um ponto de acumulação de $\{x_k\}$ e uma subsequência $\{x_{k_n}\}$ convergente para \bar{x} . Vamos considerar dois casos:

Caso 1. Seja $\{\alpha_k\}$ a sequência dos comprimentos de passo gerada pela Busca GLL. Se $\inf \alpha_k = 0$, então existe uma subsequência $\{x_k\}_K$ tal que $\lim_{k \in K} \alpha_k = 0$ (isto ocorre quando $\{\alpha_k\}$ é limitada inferiormente e decrescente).

Neste caso, como α_k é escolhido pela Busca GLL (31), então existe um índice \bar{k} suficientemente grande tal que para todo $k \geq \bar{k}$, $k \in K$, existe ρ_k de forma que $0 < \sigma_1 \leq \rho_k \leq \sigma_2$ para qual

$$\alpha_{k-1} \sigma_1 \leq \alpha_{k-1} \rho_k \leq \alpha_{k-1} \sigma_2,$$

que implica $\alpha_k = \alpha_{k-1} \rho_k \implies \alpha_{k-1} = \frac{\alpha_k}{\rho_k} = \psi_k > 0$. Temos que ψ_k não satisfaz a Busca GLL, i. e.,

$$\begin{aligned} f(P(x_k - \psi_k \nabla f(x_k))) &> \max_{0 \leq j \leq M-1} f(x_{k-j}) + \gamma \nabla f(x_k)^\top [P(x_k - \psi_k \nabla f(x_k)) - x_k] \\ &\geq f(x_k) + \gamma \nabla f(x_k)^\top [P(x_k - \psi_k \nabla f(x_k)) - x_k]. \end{aligned}$$

Portanto, segue disso que

$$f(P(x_k - \psi_k \nabla f(x_k))) - f(x_k) > \gamma \nabla f(x_k)^\top [P(x_k - \psi_k \nabla f(x_k)) - x_k] = \gamma \nabla f(x_k)^\top \nabla f_{\psi_k}(x_k). \quad (36)$$

Pelo Teorema do Valor Médio (1), obtemos

$$f(P(x_k - \psi_k \nabla f(x_k))) - f(x_k) = \nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) + [\nabla f(\xi_k) - \nabla f(x_k)]^\top \nabla f_{\psi_k}(x_k), \quad (37)$$

onde ξ_k está no segmento de reta definido por x_k e $P(x_k - \psi_k \nabla f(x_k))$. Combinando (36) e (37), obtemos que para todo $k \in K$ suficientemente grande ocorre

$$\begin{aligned} \nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) + [\nabla f(\xi_k) - \nabla f(x_k)]^\top \nabla f_{\psi_k}(x_k) &> \gamma \nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) \\ \nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) - \gamma \nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) &> [\nabla f(x_k) - \nabla f(\xi_k)]^\top \nabla f_{\psi_k}(x_k) \\ (1 - \gamma) \nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) &> [\nabla f(x_k) - \nabla f(\xi_k)]^\top \nabla f_{\psi_k}(x_k). \end{aligned} \quad (38)$$

Pelo Lema 3.1 (i), temos

$$\nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) \leq -\frac{1}{\psi_k} \|\nabla f_{\psi_k}(x_k)\|^2 = -\frac{\|\nabla f_{\psi_k}(x_k)\|}{\psi_k} \|\nabla f_{\psi_k}(x_k)\|. \quad (39)$$

Note que $\lambda_k > \psi_k$, onde λ_k é o comprimento de passo inicial da iterada k . Assim, pelo Lema 2 (i), temos

$$h(\psi_k) = \frac{\|\nabla f_{\psi_k}(x_k)\|}{\psi_k} \geq \frac{\|\nabla f_{\lambda_k}(x_k)\|}{\lambda_k} = h(\lambda_k) \implies -\frac{\|\nabla f_{\psi_k}(x_k)\|}{\psi_k} \leq -\frac{\|\nabla f_{\lambda_k}(x_k)\|}{\lambda_k}.$$

Daí e de (39), temos

$$\nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) \leq -\frac{1}{\psi_k} \|\nabla f_{\psi_k}(x_k)\|^2 \leq -\frac{1}{\lambda_k} \|\nabla f_{\lambda_k}(x_k)\| \|\nabla f_{\psi_k}(x_k)\|. \quad (40)$$

Agora, combinando (38) e (40), temos

$$[\nabla f(x_k) - \nabla f(\xi_k)]^\top \nabla f_{\psi_k}(x_k) < (1 - \gamma) \nabla f(x_k)^\top \nabla f_{\psi_k}(x_k) \leq -(1 - \gamma) \frac{\|\nabla f_{\lambda_k}(x_k)\|}{\lambda_k} \|\nabla f_{\psi_k}(x_k)\|,$$

isto é,

$$\frac{(1 - \gamma)}{\lambda_k} \|\nabla f_{\lambda_k}(x_k)\| \|\nabla f_{\psi_k}(x_k)\| < [\nabla f(\xi_k) - \nabla f(x_k)]^\top \nabla f_{\psi_k}(x_k).$$

Usando a Desigualdade de Schwarz, temos

$$\begin{aligned} \frac{(1 - \gamma)}{\lambda_k} \|\nabla f_{\lambda_k}(x_k)\| \|\nabla f_{\psi_k}(x_k)\| &< [\nabla f(\xi_k) - \nabla f(x_k)]^\top \nabla f_{\psi_k}(x_k) \\ &\leq |[\nabla f(\xi_k) - \nabla f(x_k)]^\top \nabla f_{\psi_k}(x_k)| \\ &\leq \|\nabla f(\xi_k) - \nabla f(x_k)\| \|\nabla f_{\psi_k}(x_k)\|. \end{aligned}$$

Como $\|\nabla f_{\psi_k}(x_k)\| \neq 0$, temos

$$\frac{(1 - \gamma)}{\lambda_k} \|\nabla f_{\lambda_k}(x_k)\| < \|\nabla f(\xi_k) - \nabla f(x_k)\|. \quad (41)$$

Como $\psi_k = \frac{\alpha_k}{\rho_k} \rightarrow 0$ e $x_k \rightarrow \bar{x}$ quando $k \rightarrow \infty$, com $k \in K$, temos $\xi_k \rightarrow \bar{x}$ quando $k \rightarrow \infty$, com $k \in K$.

Tomando uma subsequência $\bar{K} \subseteq K$ tal que $\{\lambda_k\}$ seja convergente para $\bar{\lambda} \in [\lambda_{\min}, \lambda_{\max}]$ e passando o limite em (41) quando $k \rightarrow \infty$, com $k \in \bar{K}$, temos que $\|\nabla f_{\bar{\lambda}}(\bar{x})\| \leq 0$. Portanto, $\nabla f_{\bar{\lambda}}(\bar{x}) = 0$, e \bar{x} é um ponto estacionário restrito.

Caso 2. Assuma que $\inf \alpha_k \geq \rho > 0$. Suponhamos, por absurdo, que \bar{x} não é um ponto estacionário restrito. Portanto, $\|\nabla f_\alpha(\bar{x})\| > 0$ para todo $\alpha \in (0, \lambda_{\max}]$. Pela continuidade e compacidade, temos que existe $\delta > 0$ tal que $\|\nabla f_\alpha(\bar{x})\| \geq \delta > 0$ para todo $\alpha \in [\rho, \lambda_{\max}]$.

Dado $k \in \mathbb{N}$, definamos um índice inteiro $k - m(k) \leq l(k) \leq k$ (ou $l(k) \in [k - m(k), k]$) tal que

$$f(x_{l(k)}) = \max_{0 \leq j \leq m(k)} [f(x_{k-j})]. \quad (42)$$

Pela hipótese (iii) do teorema da Busca GLL, temos $\alpha_k = \theta^{t_k} a$, com $a > 0$ e $\theta \in (0, 1)$, onde t_k é o primeiro inteiro não negativo t tal que

$$f(x_{k+1}) = f(x_k + \theta^t a d_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \gamma \theta^t a \nabla f(x_k)^\top d_k = f(x_{l(k)}) + \gamma \theta^t a \nabla f(x_k)^\top d_k.$$

Além disso, $\nabla f(x_k)^\top d_k < 0$, então $f(x_{l(k)}) + \gamma\theta^t a \nabla f(x_k)^\top d_k < f(x_{l(k)})$. Logo,

$$f(x_{k+1}) < f(x_{l(k)}). \quad (43)$$

Como $m(k+1) \leq \min\{m((k+1)-1)+1, M\} = \min\{m(k)+1, M\} \leq m(k)+1$ (se $M > m(k)+1$, $m(k)+1 \geq \min\{m(k)+1, M\}$; se $M < m(k)+1$, $m(k)+1 \geq \min\{m(k)+1, M\}$), podemos escrever que, para todo $k \in \mathbb{N}$,

$$\begin{aligned} f(x_{l(k+1)}) &= \max_{0 \leq j \leq m(k+1)} [f(x_{(k+1)-j})] \leq \max_{0 \leq j \leq m(k)+1} [f(x_{(k+1)-j})] \\ &= \max_{0 \leq i+1 \leq m(k)+1} [f(x_{(k+1)-(i+1)})] \\ &= \max_{0 \leq i+1 \leq m(k)+1} [f(x_{k-i})] \\ &= \max_{0 \leq i \leq m(k)} [f(x_{k-i})] \\ &= \max_{0 \leq i \leq m(k)} \{f(x_{k-i}), f(x_{k+1})\} \\ &= \max[f(x_{l(k)}), f(x_{k+1})] \\ &= f(x_{l(k)}), \end{aligned}$$

isto é, temos que

$$f(x_{l(k+1)}) \leq f(x_{l(k)}). \quad (44)$$

Logo, a sequência $\{f(x_{l(k)})\}$ é não crescente. De (31), segue que, para $k > M-1$, temos

$$f(x_{l(k)}) \leq f(x_{l(l(k)-1)}) + \gamma \nabla f(x_{l(l(k)-1)})^\top \nabla f_{\alpha_{l(k)-1}}(x_{l(l(k)-1)}).$$

Pela continuidade, para $k \geq \bar{k}$ suficientemente grande, $\|\nabla f_\alpha(\bar{x}_k)\| \geq \delta/2$. Assim, usando o Lema 1, nós obtemos

$$f(x_{l(k)}) \leq f(x_{l(l(k)-1)}) - \frac{\gamma}{\lambda_{\max}} \|\nabla f_{\alpha_{l(k)-1}}(x_{l(l(k)-1)})\|^2 \leq f(x_{l(l(k)-1)}) - \frac{\gamma\delta^2}{4\lambda_{\max}}.$$

Quando $k \rightarrow \infty$, temos $f(x_{l(k)}) \rightarrow -\infty$, o que é uma contradição. De fato, f é uma função contínua e, então, $f(x_k)$ converge para $f(\bar{x})$.

3.4 Interpolação quadrática

Ideia principal: aproximar a função f por um polinômio, utilizando os valores de f (ou de suas derivadas) num certo número de pontos. O minimizador deste polinômio serve como aproximação do minimizador de f .

- **Vantagem:** são algoritmos de determinação do comprimento de passo mais eficientes em termos de números de iterações.
- **Desvantagens:** são algoritmos computacionais mais caros e necessitam de parâmetros de salvaguarda baseados em outras buscas.

Estamos interessados em uma busca linear para otimização irrestrita uni-dimensional, isto é, queremos resolver o problema

$$\min f(x) \text{ sujeito a } x \in [a, b], \quad (45)$$

onde $a, b \in \mathbb{R}$, $a < b$ e $f : [a, b] \rightarrow \mathbb{R}$.

Considerando o problema (45), vamos trabalhar com três pontos: $a_k, b_k, c_k \in [a, b]$, tais que

$$a_k < b_k < c_k, \quad f(a_k) > f(b_k) \text{ e } f(c_k) > f(b_k). \quad (46)$$

Teorema do Valor Extremo. Se f for contínua em um intervalo fechado $[a, b]$, então f assume um valor máximo $f(c)$ e um valor de mínimo $f(d)$ em certos números c e d em $[a, b]$.

Considerando o problema (45), vamos trabalhar com três pontos: $a_k, b_k, c_k \in [a, b]$, tais que

$$a_k < b_k < c_k, \quad f(a_k) > f(b_k) \text{ e } f(c_k) > f(b_k). \quad (47)$$

O minimizador de f pertence ao intervalo $[a_k, c_k]$.

A interpolação quadrática vai aproximar a função f por um polinômio quadrático (fazendo interpolação a partir dos valores de f nos três pontos) e substituir um destes pontos pelo minimizador do polinômio. Apresentaremos a seguir um exemplo de interpolação quadrática abordada em [6].

Algoritmo 2: INTERPOLAÇÃO QUADRÁTICA

- 1 Escolher a_1, b_1 e c_1 satisfazendo às condições (47) escritas com $k = 1$. Tomar $k := 1$.
- 2 Calcular

$$x_k = \frac{1}{2} \cdot \frac{f(a_k)(c_k^2 - b_k^2) + f(b_k)(a_k^2 - c_k^2) + f(c_k)(b_k^2 - a_k^2)}{f(a_k)(c_k - b_k) + f(b_k)(a_k - c_k) + f(c_k)(b_k - a_k)}.$$

- 3 Cond.1: Se $x_k > b_k$ e $f(x_k) < f(b_k)$, então defina $a_{k+1} = b_k$, $b_{k+1} = x_k$ e $c_{k+1} = c_k$.
 Cond.2: Se $x_k > b_k$ e $f(x_k) > f(b_k)$, então defina $a_{k+1} = a_k$, $b_{k+1} = b_k$ e $c_{k+1} = x_k$.
 Cond.3: Se $x_k < b_k$ e $f(x_k) < f(b_k)$, então defina $a_{k+1} = a_k$, $b_{k+1} = x_k$ e $c_{k+1} = b_k$.
 Cond.4: Se $x_k < b_k$ e $f(x_k) > f(b_k)$, então defina $a_{k+1} = x_k$, $b_{k+1} = b_k$ e $c_{k+1} = c_k$.
- 4 Tomar $k := k + 1$ e voltar para o Passo 1.

Apresentaremos a seguir outro exemplo de interpolação quadrática apresentada em [7]. Este é um procedimento de busca linear baseado na interpolação de funções conhecidas e valores da função φ . O objetivo é encontrar um valor de α que satisfaça a condição de diminuição suficiente (armijo ou primeira desigualdade de Wolfe), sem ser “muito pequeno”. Conseqüentemente, será gerada uma sequência decrescente de valores α_i , de modo que cada valor α_i não seja muito menor que seu antecessor α_{i-1} .

Consideramos agora técnicas para encontrar o mínimo da função unidimensional

$$\varphi(\alpha) = f(x_k + \alpha d_k)$$

ou simplesmente para encontrar um comprimento de passo α_k que satisfaça as condições de buscas lineares. Assumimos que d_k é uma direção de descida. Observe que podemos escrever a condição de diminuição suficiente como

$$\varphi(\alpha_k) \leq \varphi(0) + \eta \alpha_k \varphi'(0)^T d_k,$$

onde a constante η é um parâmetro pequeno.

Suponha que seja dado o comprimento de passo inicial α_0 , então

$$\varphi(\alpha_0) \leq \varphi(0) + \eta \alpha_0 \varphi'(0)^T d_k.$$

Se o comprimento do passo α_0 satisfaz a condição, encerramos a busca. Caso contrário, sabemos que o intervalo $[0, \alpha_0]$ contém comprimentos de passo aceitáveis (resultado obtido através da condição de Armijo ou de Wolfe).

Faremos então a aproximação quadrática $\varphi_q(\alpha)$ para φ interpolando as três informações disponíveis — $\varphi(0)$, $\varphi'(0)$ e $\varphi(\alpha_0)$ — para obter

$$\varphi_q(\alpha) = \left(\frac{\varphi(\alpha_0) - \varphi(0) - \alpha_0 \varphi'(0)}{\alpha_0^2} \right) \alpha^2 + \varphi'(0) \alpha + \varphi(0).$$

O novo valor experimental α_1 é definido como o minimizador desta função quadrática, ou seja, obtemos

$$\alpha_1 = - \frac{\alpha_0^2 \varphi'(0)}{2[\varphi(\alpha_0) - \varphi(0) - \alpha_0 \varphi'(0)]}.$$

Se a condição de diminuição suficiente for satisfeita em α_1 , encerramos a busca. Caso contrário, construímos uma função cúbica que interpola as quatro informações $\varphi(0)$, $\varphi'(0)$, $\varphi(\alpha_0)$ e $\varphi(\alpha_1)$.

No método SPG1, vamos considerar a busca GLL (31). Vale ressaltar que, se $\alpha_k = \lambda_k$ não satisfizer essa desigualdade para algum $k \in \mathbb{N}$, será feito um ajuste em α_k por meio de uma interpolação quadrática. Já no método SPG2, vamos considerar a busca BMR (32). Vale ressaltar que, se $\alpha_k = 1$ não satisfizer essa busca para algum $k \in \mathbb{N}$, será feito um ajuste em α_k por meio de uma interpolação quadrática.

Observe o passo 4 do Algoritmo 1: Se (31) não for válida ou (32) não for válida, tome $\alpha_{\text{new}} \in [\sigma_1 \alpha_k, \sigma_2 \alpha_k]$, faça $\alpha_k = \alpha_{\text{new}}$. Como garantir que $\alpha_{\text{new}} \in [\sigma_1 \alpha_k, \sigma_2 \alpha_k]$? Através da seguinte interpolação quadrática: Calcule

$$\alpha_t = - \frac{\alpha^2 \langle \nabla f(x_k), d_k \rangle}{2(f(x_{k+1}) - f(x_k) - \alpha_k (\langle \nabla f(x_k), d_k \rangle))}.$$

Se $\alpha_t \in [\sigma_1, \sigma_2 \alpha_k]$, faça $\alpha_{\text{new}} = \alpha_t$. Se não, ou seja, se $\alpha_t < \sigma_1$ ou $\alpha_t > \sigma_2 \alpha_k$, então faça $\alpha_t = \alpha_t/2$ e $\alpha_{\text{new}} = \alpha_t$.

4 Experimentos numéricos

O critério de diminuição suficiente não monótona, usado no método SPG, depende de uma relação de parâmetro geral $M \geq 1$ e impõe uma diminuição no valor da função a cada M iterações (se $M = 1$ então a busca linear GLL reduz-se a uma busca linear monótona). A busca linear baseia-se numa interpolação quadrática para salvaguarda e visa satisfazer um critério do tipo Armijo com uma diminuição suficiente a partir de um parâmetro $\eta \in (0, 1)$.

Valores sugeridos para os parâmetros: um pequeno parâmetro $\alpha_{\text{min}} > 0$, um parâmetro grande $\alpha_{\text{max}} > \alpha_{\text{min}}$ (ex: $\alpha_{\text{min}} = 1.e - 30$ e $\alpha_{\text{max}} = 1.e + 30$), $\varepsilon = 1.e - 6$, parâmetro de decrescimento suficiente $\eta \in (0, 1)$ (ex: $\eta = 1.e - 4$), parâmetros de “salvaguarda” $0 < \sigma_1 <$

$\sigma_2 < 1$ (ex: $\sigma_1 = 0.1$ e $\sigma_2 = 0.9$). Além disso, um inteiro $M \geq 1$, sendo que se $M = 1$ então a busca linear GLL reduz-se a uma busca linear monótona.

Uma situação importante na qual a projeção é trivial ocorre quando D é uma caixa n -dimensional, possivelmente com alguns limites infinitos. Assim, foram implementados os Métodos do Gradiente Espectral Projetado para este tipo de problema. No total, foram resolvidos 50 problemas. Os problemas CUTE são divididos em 8 classes de acordo com o tipo de função objetivo (quadrática, soma de quadrados, outros) e interesse do problema (acadêmico, modelagem, aplicação real). Todos os problemas têm apenas restrições de limite, são duas vezes continuamente diferenciáveis e possuem mais de 50 variáveis.

Classificação dos problemas:

1. Seis problemas de interesse acadêmico;
2. Dois problemas de modelagem;
3. Um problema de aplicação real;
4. Seis problemas de interesse acadêmico e tipo de função objetivo sendo soma de quadrados;
5. Um problema de interesse da modelagem e tipo de função objetivo sendo soma de quadrados;
6. Vinte e um problemas de interesse acadêmico e tipo de função objetivo sendo quadrática;
7. Doze problemas de interesse da modelagem e tipo de função objetivo sendo quadrática;
8. Um problema de aplicação real e tipo de função objetivo sendo quadrática.

CUTEst function	Dim	SPG1	SPG2
BDEXP	5000	15	15
EXPLIN	120	1836	1345
EXPLIN2	120	2159	1891
EXPQUAD	120	2164	2160
MCCORMCK	10000	Máx. iteradas	
PROBPENL	500	2	2
QRTQUAD	120	NULL lib. handle	
S368	100	2	2
HADAMALS	1024	21	21
CHEBYQAD	50	Máx. iteradas	
HS110	50	Máx. iteradas	
LINVERSE	1999	400	515
NONSCOMP	10000	57	60
QR3DLS	610	Máx. iteradas	
SCOND1LS	1002	Máx. iteradas	

DECONVB	61	2298	2764
BIGGSB1	1000	4378	7376
BQPGABIM	50	49	47
BQPGASIM	50	49	47
BQPGAUSS	2003	Máx. iteradas	
CHENHARK	1000	NULL lib. handle	
CVXBQP1	10000	Máx. iteradas	
HARKERP2	100	Máx. iteradas	
JNLBRNG1	15625	793	1163
JNLBRNG2	15625	5756	3936
JNLBRNGA	15625	1370	1404
JNLBRNGB	15625	Máx. iteradas	
NCVXBQP1	10000	597	591
NCVXBQP2	10000	3253	1952
NCVXBQP3	10000	2566	5396
NOBNDTOR	14884	715	792
OBSTCLAE	15625	7420	Máx. iteradas
OBSTCLAL	15625	6523	6122
OBSTCLBL	15625	3559	6933
OBSTCLBM	15625	3126	3925
OBSTCLBU	15625	Máx. iteradas	3611
PENTDI	1000	13	13
TORSION1	14884	715	792
TORSION2	14884	1611	6446
TORSION3	14884	691	330
TORSION4	14884	370	310
TORSION5	14884	446	418
TORSION6	14884	68	806
TORSIONA	14884	998	922
TORSIONB	14884	3223	457
TORSIONC	14884	751	268
TORSIOND	14884	1165	269
TORSIONE	14884	663	635
TORSIONF	14884	667	1015
ODNAMUR	11130	Máx. iteradas	
Resumo		37 resuolvidos	37 resuolvidos

Referências

- [1] L. Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of mathematics*, 16(1):1–3, 1966.
- [2] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1):141–148, 1988.
- [3] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000.
- [4] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for newton’s method. *SIAM journal on Numerical Analysis*, 23(4):707–716, 1986.
- [5] A. Izmailov and M. Solodov. *Otimização, Volume 2: métodos computacionais*. IMPA, 2018.
- [6] A. Izmailov and M. Solodov. *Otimização, Volume 1*. Rio de Janeiro: IMPA, 2020.
- [7] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, 1999.